Panel Data Analysis
Part I – Classical Methods:
Background Material

James J. Heckman
University of Chicago

Econ 312, Spring 2021

THE UNIVERSITY OF CHICAGO

Review of the Early Econometric Literature on the Problem

- $Y_{it} = X_{it}\beta + \varepsilon_{it} \qquad i = 1, \ldots, I$, and $t = 1, \ldots, T$
- $Y_{i\cdot} = X_{i\cdot}\beta + \varepsilon_{i\cdot} \qquad i = 1, \ldots, I$
- $Y_{i\cdot} = \sum_{t=1}^{T} \frac{Y_{it}}{T}, \quad X_{i\cdot} = \sum_{t=1}^{T} \frac{X_{it}}{T}$
- $Y_i = (Y_{i1}, \ldots, Y_{iT}), \quad X_i = (X_{i1}, \ldots, X_{iT})$
- $\varepsilon_{it} = f_i + U_{it}$

THE UNIVERSITY OF CHICAGO

- $E(f_i) = E(U_{it}) = 0$
- $E(f_i^2) = \sigma_f^2 \qquad E(U_{it}^2) = \sigma_U^2$
- $E(f_i f_{i'}) = 0 \qquad i \neq i'$
- U_{it} is iid for all i, t.
- Assume that X_{it} is strictly exogenous:
- $E(U_{it} + f_i \mid X_{i1}, \ldots, X_{iT}) = 0$ all t.
- Also distribution of the $X_i = (X_{i1}, \ldots, X_{iT})'$ does not depend on β.

$$Cov(\varepsilon_{i,t}, \varepsilon_{i,t'}) = \sigma_f^2$$

$$\rho = \frac{\sigma_f^2}{\sigma_f^2 + \sigma_u^2} \quad \text{(intraclass correlation coefficient)}$$

- Look at covariance matrix for $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{iT})'$

$$E(\varepsilon_i' \varepsilon_i) = (\sigma_f^2 + \sigma_u^2) \begin{pmatrix} 1 & \rho & \rho \\ \rho & \ddots & \rho \\ \rho & \rho & 1 \end{pmatrix} = A.$$

THE UNIVERSITY OF CHICAGO

- Now stack all of the disturbances (in groups of T)
- $\varepsilon = (\varepsilon_1, \varepsilon_2 \ldots \varepsilon_I)$

$$E(\varepsilon\varepsilon') = \Omega \neq I = \begin{pmatrix} A & 0 & 0 & 0 \\ 0 & A & 0 & 0 \\ 0 & 0 & A & 0 \\ 0 & 0 & 0 & A \end{pmatrix}$$

- $T \times T$ blocks
- stack X_i into a supervector X
- stack Y_i into a supervector Y
- Then, we have that GLS estimator is

$$\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}Y)$$

- OLS applied to the data yields *unbiased* but inefficient estimators of β (because of exogeneity of X_{it} with respect to ε_{it}).
- But computer program produces the wrong standard errors.
- Correct standard errors are:

$$\sigma^2(X'X)^{-1}(X'\Omega X)(X'X)^{-1}$$

- OLS standard errors to assume to be $\sigma^2(X'X)^{-1}$.
- $\therefore$ inferences based on OLS models are incorrect.

- Write

$$A = (1 - \rho)I + \rho\iota\iota'$$

- ι is a $T \times 1$ vector of 1's

$$A = \begin{pmatrix} 1\text{-}\rho & 0 & 0 & \vdots \\ 0 & 1\text{-}\rho & 0 & \vdots \\ 0 & 0 & 1\text{-}\rho & \vdots \\ \dots & \dots & \dots & \dots \end{pmatrix} + \begin{pmatrix} \rho & \rho & \rho & \rho \\ \rho & \rho & \rho & \rho \\ \rho & \rho & \rho & \rho \\ \dots & \dots & \dots & \dots \end{pmatrix}$$

$$A^{-1} = \lambda_1 \iota\iota' + \lambda_2 I$$

- $\rho \neq 1$

THE UNIVERSITY OF CHICAGO

- Where

$$\lambda_1 = \frac{-\rho}{(1-\rho)(1-\rho+T\rho)}$$

$$\lambda_2 = \frac{1}{1-\rho}.$$

- **Prove this result on A^{-1}.**

- **Proof:** $AA^{-1} = I$ (direct multiplication). What is the GLS estimator doing?

$$\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}Y)$$

$$\Omega^{-1} = \begin{pmatrix} A^{-1} & 0 & 0 & 0 & \vdots \\ 0 & A^{-1} & 0 & 0 & \vdots \\ 0 & 0 & A^{-1} & 0 & \vdots \\ 0 & 0 & 0 & A^{-1} & \vdots \end{pmatrix}$$

$$X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_I \end{pmatrix} \qquad Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_I \end{pmatrix}.$$

- Thus, we conclude that

$$\hat{\beta}_{GLS} = \left(\sum_{i=1}^{I} X_i' A^{-1} X_i \right)^{-1} \left(\sum_{i=1}^{I} X_i' A^{-1} Y_i \right)$$

- Use the expression for A^{-1} given above:

$$\sum_{i=1}^{I} X_i' A^{-1} X_i = T\lambda_1 \left(\sum_{i=1}^{I} \frac{X_i' \iota \iota' X_i}{T} \right) + \lambda_2 \sum_{i=1}^{I} X_i' X_i$$

- Look at

$$\frac{\iota' X_i}{T} = X_{i\cdot}$$

a K vector of means.

- Now look at the GLS estimator it is a function of within and between variation.

- Total sum of squares

$$T_{XX} = \sum_{i=1}^{I} X_i' X_i$$

Within Deviations ($T \times K$)

$$\begin{aligned} X_i - \iota X_{i\cdot}' &= X_i - \frac{\iota\iota'}{T} X_i \\ &= [I - \frac{\iota\iota'}{T}] X_i \end{aligned}$$

- Observe:

$$\left[I - \frac{\iota\iota'}{T}\right]\left[I - \frac{\iota\iota'}{T}\right] = I - \frac{\iota\iota'}{T} \quad \text{(idempotent)}$$

X_i is $T \times K$, X_{i}' is $1 \times K$.

- Thus, we have that within variation is given by

$$\sum_{i=1}^{I} X_i'(I - \frac{\iota\iota'}{T})X_i = W_{XX}$$

$$B_{XX} = T\sum_{i=1}^{I} X_{i\cdot}'X_{i\cdot} = \sum_{i=1}^{I} \frac{X_i'\iota\iota'X_i}{T}$$

$$T_{XX} = \underbrace{W_{XX}}_{\text{within}} + \underbrace{B_{XX}}_{\text{between}}$$

- GLS estimator is given by taking

$$\begin{aligned} & T\lambda_1 \left(\sum_{i=1}^{I} \frac{X_i' \iota \iota' X_i}{T} \right) + \lambda_2 \left[\sum_{i=1}^{I} \frac{X_i' \iota \iota' X_i}{T} + W_{XX} \right] \\ = \; & \lambda_2 W_{XX} + (\lambda_2 + T\lambda_1) \left(\sum_{i=1}^{I} \frac{X_i' \iota \iota' X_i}{T} \right) \\ = \; & \lambda_2 W_{XX} + (\lambda_2 + T\lambda_1) B_{XX}. \end{aligned}$$

- There is a similar decomposition for other term and we get that

$$\begin{aligned}\hat{\beta}_{GLS} &= \left[\lambda_2 W_{XX} + \left(\lambda_2 + T\lambda_1\right) B_{XX}\right]^{-1} \\ &\quad \cdot [\lambda_2 W_{XY} + (\lambda_1 + T\lambda_2) B_{XY}]\end{aligned}$$

- Define

$$\theta = 1 + T\frac{\lambda_1}{\lambda_2} = 1 - T\frac{\rho}{(1-\rho+T\rho)}$$

$$\theta = \frac{1-\rho+T\rho-T\rho}{1-\rho+T\rho} = \frac{1-\rho}{1-\rho+T\rho}$$

$$\hat{\beta}_{GLS} = [W_{XX}+\theta B_{XX}]^{-1}[W_{XY}+\theta B_{XY}].$$

- 2 estimators are averaged together:
  Take first estimator the *within* estimator.

- This is simply given by taking deviations from mean:

$$\begin{aligned} Y_{it} - Y_{i\cdot} &= (X_{it} - X_{\iota\cdot})\beta + U_{it} - U_{i\cdot} \\ Y_{i\cdot} &= X_{i\cdot}\beta + f_i + \bar{U}_{i\cdot} \end{aligned}$$

- $\therefore$ subtracting $Y_{i\cdot}$ produces an estimator *free* of f_i.
- Doing that eliminates the fixed effects from the model. Thus, we have that

$$\begin{aligned} \hat{\beta}_W &= (W_{XX})^{-1} W_{XY} \\ \hat{\beta}_B &= (B_{XX})^{-1} B_{XY} \end{aligned}$$

- Simply average over the groups and we are done.

$$
\begin{aligned}
(W_{XX})\hat{\beta}_W &= W_{XY} \\
(B_{XX})\hat{\beta}_B &= B_{XY} \\
[W_{XX} + \theta B_{XX}]\hat{\beta}_{GLS} &= W_{XX}\hat{\beta}_W + \theta(B_{XX})\hat{\beta}_B
\end{aligned}
$$

- $\therefore$ we have that

$$
\hat{\beta}_{GLS} = [W_{XX} + \theta B_{XX}]^{-1}[W_{XX}\hat{\beta}_W + \theta B_{XX}\hat{\beta}_B].
$$

- For a scalar regressor $\hat{\beta}_{GLS}$ lies between $\hat{\beta}_W$ and $\hat{\beta}_B$.
- (But not, necessarily so, for the general regressor case).

- Now suppose $\rho = 0$

$$\Longrightarrow \lambda_1 = 0 \Longrightarrow \theta = 1.$$

- Then $\hat{\beta}_{GLS}$ is simply OLS. Suppose that $\rho = 1$.
- A is singular, A^{-1} doesn't exist.
- If we have that regressors are fixed over the spell for the case that the rank of $W_{XX} = 0$ and GLS is between estimator.

- Suppose that $T \to \infty, \rho \neq 0$

$$\lim_{T\to\infty} \frac{(T\lambda_1)}{\lambda_2} = \lim_{T\to\infty} \frac{-\rho T}{(1-\rho+T\rho)}$$
$$= \lim_{T\to\infty} \frac{-\rho}{\frac{1-\rho}{T}+\rho} \to -1$$

- $\therefore \theta = 0$
- $\therefore [\hat{\beta}_{GLS} = \hat{\beta}_W]$.

THE UNIVERSITY OF CHICAGO

- In this case, the *within* estimator is the efficient estimator.
- Now A^{-1} matrix itself can be written in an interesting fashion and provides an example of another interpretation of the estimator.
- $A^{-1} = \lambda_2(I - k\iota\iota')$ where k is given by $-\dfrac{\lambda_1}{\lambda_2}$

$$\begin{aligned} & [I - c\iota\iota'][I - c\iota\iota'] \\ = & \; I - c\iota\iota' - c\iota\iota' + Tc^2\iota\iota' \\ = & \; I - (2c - Tc^2)\iota\iota' \end{aligned}$$

where

$$2c - Tc^2 = -\frac{\lambda_1}{\lambda_2}, \quad F = I - c\iota\iota'.$$

THE UNIVERSITY OF CHICAGO

- Solve to get

$$c = \frac{1}{T}\left[1 - \sqrt{\frac{1-\rho}{1-\rho+\rho T}}\right]$$

- $\therefore$GLS estimator is of the form

$$\hat{\beta}_{GLS} = \left[\sum_{i=1}^{I}(X_i' A^{-1} X_i)\right]^{-1}\left[\sum_{i=1}^{I} X_i' A^{-1} Y_i\right].$$

- But this is $\Leftrightarrow$ to transforming the data in the following way

$$\begin{aligned} Y_i &= X_i\beta + \varepsilon_i \\ FY_i &= FX_i\beta + F\varepsilon_i \\ FY_i &= Y_i - (cT\iota)Y_{i\cdot} \\ Y_{i\cdot} &= \frac{\iota' Y_i}{T} \end{aligned}$$

- The mean value of Y_i for person i over sample period T.

$$FX_i = X_i - (cT)\iota X_{i\cdot}'.$$

- Now, suppose that we have $\rho = 0, c = 0$, GLS is OLS applied to data.

THE UNIVERSITY OF CHICAGO

Standard Errors for Fixed Effect; Estimator IS Produced by OLS Formula; GLS is OLS

- **Proof:**

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_N \end{pmatrix} = \begin{pmatrix} \iota \\ 0 \\ \dots \\ 0 \end{pmatrix} f_i + \begin{pmatrix} 0 \\ \iota \\ \dots \\ 0 \end{pmatrix} f_2 + \begin{pmatrix} 0 \\ 0 \\ \dots \\ \iota \end{pmatrix} f_N$$

$$+ \begin{pmatrix} X_1 \\ \dots \\ X_N \end{pmatrix} \beta + \begin{pmatrix} U_1 \\ \dots \\ U_N \end{pmatrix}$$

$$Y_i = \begin{pmatrix} Y_{i1} \\ \dots \\ Y_{iT} \end{pmatrix} \qquad X_i = \begin{pmatrix} X_{i1} \\ \dots \\ X_{iT} \end{pmatrix}$$

$$E(U_i U_i') = \sigma^2 I_T$$

$$E(U_i U_j') = 0, \qquad i \neq j.$$

- Define $F = I - \frac{\iota\iota'}{T}$.
- Partition

$$Y_i = X_i\beta + if_i + U_i$$

- Using

$$(F) + \frac{\iota\iota'}{T} = I$$

$$\left(\frac{\iota\iota'}{T}\right) F = 0$$

$$FY_i = FX_i\beta + FU_i$$

$$\frac{\iota\iota'}{T} Y_i = \frac{\iota\iota'}{T} X_i\beta + \frac{\iota\iota'}{T} U_i$$

$$\hat{\beta}_W = \left[\sum_{i=1}^{I} X_i' F X_i\right]^{-1} \left[\sum_{i=1}^{I} X_i' F Y_i\right].$$

- Let

$$B = \left(\sum_{i=1}^{I} X_i' F X_i \right).$$

- Note: The B here is not the between matrix.

$$\begin{aligned} Var \hat{\beta}_W &= (B)' E \left[\left(\sum_{i=1}^{I} X_i' F U_i \right) \sum_{i=1}^{N} U_i' F X_i \right] B \\ &= B' \left[\sigma_U^2 \sum_{i=1}^{I} (X_i' F) F X_i' \right] B \\ &= \sigma_U^2 (\sum X_i' F X_i)^{-1} \end{aligned}$$

where

$$\left(\frac{\iota \iota'}{T} \right) \left(\frac{\iota \iota'}{T} \right) = \frac{\iota \iota'}{T}.$$

- This is OLS variance covariance.

- Between Estimator:

$$\begin{aligned}
\hat{\beta}_B &= \left[\sum_{i=1}^{N}\left(X_i \frac{\iota\iota'}{T} X_i'\right)\right]^{-1}\left[\sum_{i=1}^{N} X_i \frac{\iota\iota'}{T} Y_i\right] \\
&= \beta + \sum_{i=1}^{N} X_i \frac{\iota\iota'}{T} U_i + \sum_{i=1}^{N} X_i \frac{ii'}{T} f_i
\end{aligned}$$

- Now $\hat{\beta}_B$ uncorrelated with $\hat{\beta}_W$ because

$$f_i \perp\!\!\!\perp U_j \text{ all } i, j;$$

$$
\begin{aligned}
Var\left(\hat{\beta}_B\right) &= \left(\sum X_i \frac{\iota\iota'}{T} X_i'\right)^{-1} \left[\frac{\sigma_U^2}{T} + \sigma_f^2\right] \\
&\quad \cdot \left(\sum \frac{X_i \iota\iota' X_i'}{T}\right) \left(\sum X_i \frac{\iota\iota'}{T} X_i'\right)^{-1} \\
&= \left(\sigma_f^2 + \frac{\sigma_U^2}{T}\right) \left(\sum X_i \frac{\iota\iota'}{T} X_i'\right)^{-1}.
\end{aligned}
$$