Ec485 Lecture 1, LT2023

PANEL DATA MODELS
(or Longitudinal or Time-Series/Cross-Section)
and an INTRODUCTION TO SIMULATION-BASED INFERENCE

1 Preliminary Issues:

1.1 At least double-indexed data

Panel or Longitudinal or Time-series/Cross-section Data are such where a unit of observation s subsumes at least two indices/dimensions of sampling. E.g.,

$$\begin{array}{llllllll} s = 1, \cdots, S & & & & & & & \\ y_s & x_s' & \epsilon_s & s = it & where & i = 1, \cdots, N & & \text{cross-section side} \\ & & & & & t = 1, \cdots, T_i & & \text{time-series side} \end{array}$$

NB: Throughout our discussion, we will focus on "Large N, small T" asymptotics with $N \to \infty$ while $\max_i T_i \approx$ small and finite.

1.2 Organization of the data — three alternatives with two dimensions:

1.2.1 t "fastest":

$$\{y_s\} = \begin{pmatrix} y_1 \\ \vdots \\ y_s \\ \vdots \\ y_S \end{pmatrix} = \begin{pmatrix} y_{11} \\ \vdots \\ y_{it} \\ \vdots \\ y_{NT_N} \end{pmatrix} = \begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1t} \\ \vdots \\ y_{1T_1} \\ - - - \\ y_{21} \\ y_{22} \\ \vdots \\ y_{2t} \\ \vdots \\ y_{2T_2} \\ - - - \\ \vdots \\ y_{N1} \\ \vdots \\ y_{NT_N} \end{pmatrix} \cdots \begin{pmatrix} IID \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \end{pmatrix} \cdots \begin{pmatrix} TID \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \end{pmatrix}$$

1.2.2 i "fastest":

$$\{y_s\} = \begin{pmatrix} y_1 \\ \vdots \\ y_s \\ \vdots \\ y_S \end{pmatrix} = \begin{pmatrix} y_{11} \\ \vdots \\ y_{it} \\ \vdots \\ y_{NT_N} \end{pmatrix} = \begin{pmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{i1} \\ \vdots \\ y_{N1} \\ --- \\ y_{12} \\ y_{22} \\ \vdots \\ y_{i2} \\ \vdots \\ y_{N2} \\ --- \\ \vdots \\ y_{N1} \\ \vdots \\ y_{NT_N} \end{pmatrix} \ldots \begin{pmatrix} IID \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \end{pmatrix} \ldots \begin{pmatrix} TID \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \end{pmatrix}$$

1.2.3 data organized as they come but double-indexed ID variables:

$$
\begin{aligned}
\{y_s\} &= \begin{pmatrix} y_1 \\ \vdots \\ y_s \\ \vdots \\ y_S \end{pmatrix} ... \begin{pmatrix} iid(1) \\ \vdots \\ iid(s) \\ \vdots \\ iid(S) \end{pmatrix} ... \begin{pmatrix} tid(1) \\ \vdots \\ tid(s) \\ \vdots \\ tid(S) \end{pmatrix} \\
&= \text{Sx1 vector } y \text{ ... Sx1 vector } IID \text{ ... Sx1 vector } TID
\end{aligned}
$$

.

1.3 (3) Balanced ($T_i = T$) vs. Unbalanced Data Sets (T_i varies with i)

Balanced: $S = N \times T$:

$$\begin{pmatrix} y_1 \\ \vdots \\ y_s \\ \vdots \\ y_S \end{pmatrix} = \begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1t} \\ \vdots \\ y_{1T} \\ --- \\ y_{21} \\ y_{22} \\ \vdots \\ y_{2t} \\ \vdots \\ y_{2T} \\ --- \\ \vdots \\ y_{N1} \\ \vdots \\ y_{NT} \end{pmatrix}$$

Unbalanced: $S = \sum_{i=1}^{N} T_i$:

$$\begin{pmatrix} y_1 \\ \vdots \\ y_s \\ \vdots \\ y_S \end{pmatrix} = \begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1t} \\ \vdots \\ y_{1T_1} \\ --- \\ y_{21} \\ y_{22} \\ \vdots \\ y_{2t} \\ \vdots \\ y_{2T_2} \\ --- \\ \vdots \\ y_{N1} \\ \vdots \\ y_{NT_N} \end{pmatrix}$$

1.3.1 (3b) (related issue) Use PADDING with Missing Data Code (MDC) — Then every Unbalanced PDS becomes Balanced

New single constant $T = \max_i T_i$.

1.3.2 (3c) (related issue) DROP OBSERVATIONS to make Balanced

Example: new single constant $T = \min_i T_i$.

(4) Lagged variables in Panel Data

$$LAG1\begin{pmatrix} y_0 \\ \vdots \\ y_{s-1} \\ \vdots \\ y_{S-1} \end{pmatrix} = \begin{pmatrix} y_{10} \\ y_{11} \\ \vdots \\ y_{1,t-1} \\ \vdots \\ y_{1,T_1-1} \\ --- \\ y_{1T_1} \\ y_{21} \\ \vdots \\ y_{2,t-1} \\ \vdots \\ y_{2,T_2-1} \\ --- \\ \vdots \\ y_{N-1,T_N} \\ \vdots \\ y_{N,T_N-1} \end{pmatrix} = \begin{pmatrix} MDC \\ y_{11} \\ \vdots \\ y_{1,t-1} \\ \vdots \\ y_{1,T_1-1} \\ --- \\ y_{1T_1} \\ y_{22} \\ \vdots \\ y_{2,t-1} \\ \vdots \\ y_{2,T_2-1} \\ --- \\ \vdots \\ y_{N-1,T_N} \\ \vdots \\ y_{N,T_N-1} \end{pmatrix} \quad \text{vs.} \quad XTLAG1\begin{pmatrix} y_0 \\ \vdots \\ y_{s-1} \\ \vdots \\ y_{S-1} \end{pmatrix} = \begin{pmatrix} y_{10} \\ y_{11} \\ \vdots \\ y_{1t} \\ \vdots \\ y_{1,T_1-1} \\ --- \\ y_{20} \\ y_{22} \\ \vdots \\ y_{2,t-1} \\ \vdots \\ y_{2,T_2-1} \\ --- \\ \vdots \\ --- \\ y_{N0} \\ \vdots \\ y_{N,T_N-1} \end{pmatrix} = \begin{pmatrix} MDC \\ y_{11} \\ \vdots \\ y_{1t} \\ \vdots \\ y_{1,T_1-1} \\ --- \\ MDC \\ y_{22} \\ \vdots \\ y_{2,t-1} \\ \vdots \\ y_{2,T_2-1} \\ --- \\ \vdots \\ --- \\ MDC \\ \vdots \\ y_{N,T_N-1} \end{pmatrix}$$

In sum, the LAG1 variable will contain a single Missing Value, whereas the XTLAG1 variable will contain N Mising Values.

1.4 (5) Linear vs. Nonlinear models (additive vs nonadditive, index vs general)

$$\begin{array}{ccc} & s = 1, \cdots, S & i = 1, \cdots, N \text{ and } t = 1, \cdots, T_i \\ Linear & y_s = x_s'\beta + \epsilon_s & y_{it} = x_{it}'\beta + \epsilon_{it} \\ Additively\ Nonlinear\ Index & y_s = f(x_s'\beta) + \epsilon_s & y_{it} = f(x_{it}'\beta) + \epsilon_{it} \\ Additively\ Nonlinear & y_s = g(x_s', \beta) + \epsilon_s & y_{it} = g(x_{it}', \beta) + \epsilon_{it} \\ Non\text{-}additively\ Nonlinear & y_s = h(x_s', \beta, \epsilon_s) & y_{it} = h(x_{it}', \beta, \epsilon_{it}) \end{array}$$

1.5 (6) Combination of (1) and (3): Endogenous Data Availability

NB: even an apparently Linear model is in fact Nonlinear if Endogenous Data Availability — Distinction between Latent and (observed) Limited Dependent Variables.

Modelling Framework: Sample Selection or Selectivity or Endogenous Data Availability or Endogenous Attrition

Two-equation Latent variables model:

$$\begin{aligned} y_{it}^* &= x_{it}'\beta + \epsilon_{it} \\ d_{it}^* &= z_{it}'\gamma + u_{it} \end{aligned}$$

Observation LDV Rule:

$$\begin{aligned} D_{it} &= \begin{cases} 1 & iff \quad d_{it}^* = z_{it}'\gamma + u_{it} > 0 \\ 0 & iff \quad d_{it}^* = z_{it}'\gamma + u_{it} \leq 0 \end{cases} \quad \text{and} \\ y_{it} &= \begin{cases} y_{it}^* & iff \quad d_{it}^* = z_{it}'\gamma + u_{it} > 0 \\ MDC & iff \quad d_{it}^* = z_{it}'\gamma + u_{it} \leq 0 \end{cases} \end{aligned}$$

NB: Distinction between Censored Selectivity and Truncated Selectivity:

Selectivity with Censoring

$$\begin{aligned} y_{it} &= \begin{cases} y_{it}^* & iff \quad D_{it} = 1 \\ MDC & iff \quad D_{it} = 0 \end{cases} \quad \text{and} \\ & D_{it}, x_{it}, \text{ and } z_{it} \text{ always observed} \end{aligned}$$

Selectivity with Truncation

$$\begin{aligned} y_{it} &= \left\{ y_{it}^* \quad iff \quad D_{it} = 1 \quad \text{and} \right. \\ & D_{it}, x_{it}, \text{ and } z_{it} \text{ observed *only* when } D_{it} = 1 \end{aligned}$$

NB: Fundamental Point: If $u_{it} \& \epsilon_{it}$ are *not* *independent*, then

$$\begin{aligned} E(y_{it}|X) &\neq x'_{it}\beta \text{ and } E(y_{it}|X,Z) \neq x'_{it}\beta \text{ *BUT*} \\ E(y_{it}|X,Z) &= g(x'_{it}, z'_{it}, \delta) \end{aligned}$$

where the parameter vector δ is related to $\beta, \gamma, \sigma^2_\epsilon, \sigma^2_u$, and $\rho_{\epsilon u}$.

1.6 (7) Types of variables w.r.t. i and t indices:

$$\underset{\text{default}}{x_s^j = x_{it}^j} \quad vs. \quad \underset{\text{time-invariant}}{z_s^j = z_i^j} \quad vs. \quad \underset{\text{individual-invariant (e.g., economy-wide/macro)}}{w_s^j = w_t^j}$$

.

1.7 (8) Error-Components/Factor-Analytic structures:

1.7.1 Error-components with single time-invariant factor:

$$\epsilon_s = \epsilon_{it} = \alpha_i + \nu_{it} = \alpha_s + \nu_s$$

NOTE: α_i is termed the "unobserved persistent heterogeneity".

Basic assumptions:

$$\alpha_i \underset{iid \text{ over } i}{\sim} ?(0, \sigma_\alpha^2)$$

$$\nu_{it} \underset{iid \text{ over } i \text{ and } t}{\sim} ?(0, \sigma_\nu^2)$$

and α_i,$\nu_{\ell t}$ independent/uncorrelated for all i, ℓ, t

NB: Key conclusion: $VCov(\epsilon|regressors)$ is a Block-Diagonal matrix with Diagonal blocks equal to:

$$\begin{pmatrix} \sigma_\alpha^2 + \sigma_\nu^2 & \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 \\ & \sigma_\alpha^2 + \sigma_\nu^2 & \ddots & \vdots \\ & & \ddots & \sigma_\alpha^2 \\ & & & \sigma_\alpha^2 + \sigma_\nu^2 \end{pmatrix}$$

and Off-Diagonal blocks between individuals i and n equal to $0_{T_i \times T_n}$. This is called the "equi-correlated" error components model.

Error-components with two factors (one time-, one individual-invariant):

$$\epsilon_s = \epsilon_{it} = \alpha_i + \zeta_t + \nu_{it} = \alpha_s + \zeta_s + \nu_s$$

.where

$$\alpha_i \underset{iid \text{ over } i}{\sim} ?(0, \sigma_\alpha^2)$$

$$\nu_{it} \underset{iid \text{ over } i \text{ and } t}{\sim} ?(0, \sigma_\nu^2)$$

$$\zeta_t \underset{iid \text{ over } t}{\sim} ?(0, \sigma_\zeta^2)$$

and α_i,$\nu_{\ell t}$,ζ_q mutually independent/uncorrelated for all i, ℓ, t, q

The $VCov(\epsilon|regressors)$ matrix has a similar block structure with $\sigma_\alpha^2 + \sigma_\nu^2 + \sigma_\zeta^2$ on the main diagonal, and either σ_α^2, σ_ζ^2, or $\sigma_\alpha^2 + \sigma_\zeta^2$ in the elements of the off-diagonal blocks depending on the values of i, ℓ, t.

.

2 Random Effect "vs." Fixed Effects

Common misconception: the approaches are frequently thought of as *alternative* DGPs. A much more appropriate framework is to think of them as the *same* DGP, but alternative Estimation Approaches

Common DGP with one-factor error-components model as in (1.8) above:

$$y_{it} = x'_{it}\beta + z'_i\gamma + \epsilon_{it} = x'_{it}\beta + z'_i\gamma + \alpha_i + \nu_{it}$$

RE Approaches: in *RED*: [.]+[.]

$$y_{it} = x'_{it}\beta + z'_i\gamma + \epsilon_{it} = [x'_{it}\beta + z'_i\gamma] + [\alpha_i + \nu_{it}]$$

FE Approaches in *BLACK*: (.) + (.)

$$y_{it} = x'_{it}\beta + z'_i\gamma + \epsilon_{it} = (x'_{it}\beta + z'_i\gamma + \alpha_i) + (\nu_{it})$$

FE-(BLACK): The four classic regression assumptions A1, A2, A3, A4 take the form:

$A1$	no perfect multicollinearity among the regressors X and Z	$rank(X, Z) = k_x + k_z$
$A2$	linear additive model	$y = X\beta + Z\gamma + \epsilon$
$A3$	regressor exogeneity	X and Z exogenous w.r.t. ϵ
$A4$	$VCov(error\|regressors)$	$VCov(\epsilon\|X, Z)$

RE-[RED]: Now the four classic regression assumptions A1, A2, A3, A4 take the form: (D is the full set of N variable intercepts dummies, one for each individual)

$A1$	no perfect multicollinearity among the regressors X and D	$rank(X, D) = k_x + k_z + N$ $NB:$ Z is dropped since perfectly collinear with D
$A2$	linear additive model	$y = X\beta + Z\gamma + \epsilon = X\beta + D\alpha + \nu$
$A3$	regressor exogeneity	X and D exogenous w.r.t. ν (no Z $regressors$)
$A4$	$VCov(error\|regressors)$	$VCov(\nu\|X, D)$

2.1 *FE-TYPE estimators: the α_i's are eliminated through suitable transformation or conditioned upon or estimated through sufficient statistics

Key fact: Parameters estimated (either explicitly or implicitly): β (k_x) and $a_1, \cdots, a_N$ (N), σ^2_ν (1)

2.1.1 FE1: FD

***Apply OLS on FD model:

$$\Delta y_{it} = \Delta x'_{it}\beta + 0 + 0 + \Delta\nu_{it}$$

NB1: No estimates of γ are possible by the approach since Z has dropped out.

NB2: $\Delta\nu_{it}$ is a non-invertible MA(1) process, with known parameter -1. Hence OLS will not be BLUE and we will need to calculate Robust SEs/VCovs

2.1.2 FE2: Quasi-differencing/Within

***Apply OLS on Quasi-Differenced model:

$$Qy = QX\beta + QZ\gamma + Q\alpha + Q\nu = QX\beta + Q\nu$$

where Qy has typical element

$$\{Qy\}_{it} = y_{it} - \bar{y}_{i\cdot} \equiv y_{it} - \sum_{t=1}^{T_i} y_{it}$$

Consequently, the Q transformation eliminates all time-invariant terms — in particular α and Z.

NB1: No estimates of γ are possible by the approach since Z has dropped out.

NB2: The transformation Q is idempotent (and symmetric, hence a projection matrix). Therefore, the

$$VCov(\nu|X) = Q\sigma^2_\nu I_{NT} Q' = \sigma^2_\nu Q$$

which is *singular* (it has deficient rank). Therefore its generalized inverse will be *itself* and so the GLS estimator to take into account the non-spherical distribution of ν will be *identical* to plain OLS! To see this formally:

$$\begin{aligned}
\text{plain OLS} \quad &: \quad \hat{\beta}_{FE2} = \hat{\beta}_W = \left((QX)'(QX)\right)^{-1}(QX)'(Qy) \\
\text{GLS} \quad &: \quad \left((QX)'\left(VCov(\nu|X)\right)^{geninv}(QX)\right)^{-1}(QX)'\left(VCov(\nu|X)\right)^{geninv}(Qy) \\
&= \quad \left((QX)'Q(QX)\right)^{-1}(QX)'Q(Qy) = \hat{\beta}_{FE2} = \hat{\beta}_W
\end{aligned}$$

NB3: The FE2 model is *numerically* *identical* to the Variable Intercepts OLS model:

$$y = X\beta + D\alpha + \nu$$

because by the Frisch-Waugh-Lovell theorem, linear regression partitioning gives that:

$$\begin{aligned}
\hat{\beta}_{VIols} &= \left((M_D X)'(M_D X)\right)^{-1}(M_D X)'(M_D y) : M_D \equiv I_{NT} - D(D'D)^{-1}D' = Q \\
&= \left((QX)'(QX)\right)^{-1}(QX)'(Qy) = \hat{\beta}_{FE2} = \hat{\beta}_W \\
\{\hat{\alpha}_{VIols}\}_i &= \bar{y}_{i\cdot} - \bar{x}'_{i\cdot}\hat{\beta}_{FE2}
\end{aligned}$$

2.2 *RE-TYPE estimators:

Key fact: Parameters estimated: β (k_x), γ (k_z), σ^2_α (1), and σ^2_ν (1)

Consider model

$$y = [X\beta + Z\gamma] + [\alpha + \nu] = [X\beta + Z\gamma] + [\epsilon] \equiv W\theta + \epsilon$$

RE1: pooled OLS

$$\hat{\theta}_{RE1} = \begin{pmatrix} \hat{\beta}_{RE1} \\ \hat{\gamma}_{RE1} \end{pmatrix} = (W'W)^{-1}W'y$$

NB: This will *not* be BLUE and its *Robust* SEs/VCov must be calculated to allow for the Clustering exhibited by the *block-diagonal* $VCov(\epsilon|X, Z) \equiv \sigma^2_\epsilon \Omega$.

RE2: "the RE"-GLS estimator

$$\hat{\theta}_{RE2} = \hat{\theta}_{REgls} = \begin{pmatrix} \hat{\beta}_{REgls} \\ \hat{\gamma}_{REgls} \end{pmatrix} = (W'\Omega^{-1}W)^{-1}W'\Omega^{-1}y$$

NB1: This estimator will be BLUE and will have the correct SEs/VCov.

NB2: In 1972, Fuller and Battese showed that calculating Ω^{-1}, which is computationally burdensome, can be avoided. Instead, the rotation $\Omega^{-1/2\prime}$ yields the equivalent very straightforward expressions:

$$\begin{aligned} \Omega^{-1/2\prime}y &= \{y_{it} - \lambda_i \bar{y}_{i\cdot}\} \\ \Omega^{-1/2\prime}X &= \{x_{it} - \lambda_i \bar{x}_{i\cdot}\} \\ \Omega^{-1/2\prime}Z &= \{(1 - \lambda_i) z_i\} \\ \text{where } \lambda_i &= 1 - \sqrt{\frac{\sigma^2_\nu}{\sigma^2_\nu + T_i \sigma^2_\alpha}} \end{aligned}$$

Hence the RE2-GLS estimator can be obtained by applying plain OLS on the $\Omega^{-1/2\prime}$-transformed variables.