ECON 480
LECTURE 7: PANEL DATA

Ivan A. Canay
Northwestern University

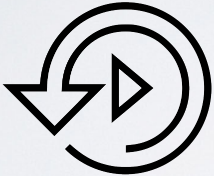
Past & Future

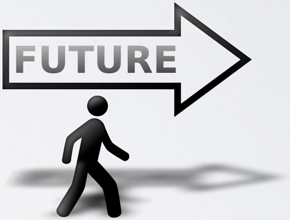
Last Class

- Generalized Method of Moments
- Empirical Likelihood
- Asymptotic Properties

TODAY

- Panel Data: intuition
- Fixed Effects: FD
- Fixed Effects: Demeaning
- Random Effects





- Let (Y, X, η, U) be a random vector where Y, η, and U take values in $\mathbf{R}$ and X takes values in $\mathbf{R}^k$.
- We are not assuming that the first component of X is a constant equal to one.
- Let $\beta = (\beta_1, \ldots, \beta_k)' \in \mathbf{R}^k$ be such that

$$Y = X'\beta + \eta + U \ ,$$

  where we assume both η and U are unobserved.
- We also allow for X and η to be correlated, so that $E[X\eta] \neq 0$.
- Given this, combining $\eta + U$ into a single unobservable would require an IV to get an estimator of β, even if we assume $E[XU] = 0$.
- **Today**: when we observe the same units (individuals, firms, families, etc) multiple times (across time, regions, etc) we may identify and consistently estimate β without an IV, at least under certain restrictions on η and U.

Two Period Model

- **Model**: Suppose that we observe the same unit at two different points in time, and that the unobservable η captures unobserved heterogeneity that is unit specific and constant over time,

$$Y_1 = X_1'\beta + \eta + U_1$$
$$Y_2 = X_2'\beta + \eta + U_2 \ .$$

  We are also assuming that β is a constant parameter that does not change over time.

- **Simple Approach**: take first differences:

Two Period Model

$$E[\Delta X \Delta U] = E[X_2 U_2] + E[X_1 U_1] - E[X_2 U_1] - E[X_1 U_2] \ .$$

- For the expression above to be equal to zero, we need
  - (1) $E[X_2 U_2] = E[X_1 U_1] = 0$ - the standard orthogonality assumption.
  - (2) $E[X_2 U_1] = E[X_1 U_2] = 0$ - covariates are uncorrelated with the unobservables in other time periods.
- (1) + (2) is called **strict exogeneity**.
- Under (1) + (2): Least squares of ΔY on ΔX delivers a consistent estimator of β since

$$\beta = E[\Delta X \Delta X']^{-1} E[\Delta X \Delta Y] \ , \qquad (1)$$

provided that $E[\Delta X \Delta X']$ is invertible.

Remarks

- **Constant Unobserved factors**: observing the same units over multiple time periods (the so-called *panel data*) allow us to control for unobserved factors that are **constant** over time (the η).

  The trick we just used would not work if η was allowed to change over time.

- **Time-varying covariates**: the requirement "$E[\Delta X \Delta X']$ is invertible" implies X changes over time. The trick does not allow us to estimate coefficients of variables that are constant over time.

  Indeed, such variables are removed by the transformation in the same way η is removed.

- **Strict exogeneity**: is arguably stronger than simply assuming $E[X_t U_t] = 0$ for all t. Cases where X_2 is a decision variable of an agent in a context where U_1 is known at $t = 2$ may seriously question the validity of $E[X_2 U_1] = 0$.

  This type of dynamic argument is distinct from the one in omitted variables bias.

QUESTIONS?

Fixed Effects: FD

- Let (Y, X, η, U) be distributed as described above and denote by P the distribution of
$$(Y_{i,1}, \ldots, Y_{i,T}, X_{i,1}, \ldots, X_{i,T}) \ .$$
- We assume that we have a random sample of size n, so that the observed data is given by
$$\{(Y_{i,t}, X_{i,t}) : 1 \leqslant i \leqslant n, 1 \leqslant t \leqslant T\} \ .$$
- **Sampling process**: i.i.d. across i, but agnostic about dependence over time.
- Consider
$$Y_{i,t} = X'_{i,t}\beta + \eta_i + U_{i,t}, \ i = 1, \ldots, n \quad t = 1, \ldots, T \ , \tag{2}$$
and define
$$\Delta X_{i,t} = X_{i,t} - X_{i,t-1} \quad \text{for} \quad t \geqslant 2 \ ,$$
and proceed analogously with the other random variables. Note again that $\Delta\eta_i = 0$.
- Applying this transformation to (2), we get
$$\Delta Y_{i,t} = \Delta X'_{i,t}\beta + \Delta U_{i,t} \ , \qquad i = 1, \ldots, n \quad t = 2, \ldots, T \ .$$

FD: Assumptions

- A regression of $\Delta Y_{i,t}$ on $\Delta X_{i,t}$ leads to a consistent estimator of β under the following assumptions:

  FD1. $E[U_{i,t}|X_{i,1},\ldots,X_{i,T}] = 0$ for all $t = 1,\ldots,T$,

  FD2. $\sum_{t=2}^{T} E[\Delta X_{i,t}\Delta X_{i,t}'] < \infty$ is invertible.

- **FD1**: sufficient for $E[\Delta U_{i,t}\Delta X_{i,t}] = 0$. **FD2**: fails if some component of $X_{i,t}$ does not vary over time.

- **First-difference estimator**:

$$\hat{\beta}_n^{\mathrm{fd}} = \left(\sum_{i=1}^{n}\sum_{t=2}^{T}\Delta X_{i,t}\Delta X_{i,t}'\right)^{-1}\left(\sum_{i=1}^{n}\sum_{t=2}^{T}\Delta X_{i,t}\Delta Y_{i,t}\right) .$$

- **On Efficiency**: under the assumption that $\mathrm{Var}[U_{i,t}|X_{i,1},\ldots,X_{i,T}]$ is constant (homoskedasticity), together with the assumption of no serial correlation in $U_{i,t}$, it is possible to show that $\hat{\beta}_n^{\mathrm{fd}}$ is **not** asymptotically efficient and that a different transformation of the data delivers an estimator with a lower asymptotic variance under those assumptions.

Deviations from Means: Assumptions

- **De-meaning**: an alternative transformation to remove the individual effects η_i
- Let

$$\dot{X}_{i,t} = X_{i,t} - \bar{X}_i \quad \text{where} \quad \bar{X}_i = \frac{1}{T}\sum_{t=1}^{T} X_{i,t} \ ,$$

  and define $\dot{Y}_{i,t}$ and $\dot{U}_{i,t}$ analogously.

- Note that $\dot{\eta}_i = 0$ for all $i = 1, \dots, n$. We then obtain:

$$\dot{Y}_{i,t} = \dot{X}'_{i,t}\beta + \dot{U}_{i,t}, \quad i = 1, \dots, n \quad t = 1, \dots, T \ . \tag{3}$$

- A regression of $\dot{Y}_{i,t}$ on $\dot{X}_{i,t}$ provides a consistent estimator of β under the following two assumptions,

FE1. $E[U_{i,t}|X_{i,1}, \dots, X_{i,T}] = 0$ for all $t = 1, \dots, T$,

FE2. $\sum_{t=1}^{T} E[\dot{X}_{i,t}\dot{X}'_{i,t}] < \infty$ is invertible.

De-meaning: Remarks

- **FE1**: same condition as FD1, which is sufficient for $E[\dot{U}_{i,t}\dot{X}_{i,t}] = 0$.
- **FE2**: fails if some component of $X_{i,t}$ does not vary over time.
- **Fixed effect estimator**: The de-meaning estimator (commonly known as the fixed effect estimator or dummy variable estimator) takes the form

$$\hat{\beta}_n^{\text{fe}} = \left(\sum_{i=1}^{n}\sum_{t=1}^{T}\dot{X}_{i,t}\dot{X}_{i,t}'\right)^{-1}\left(\sum_{i=1}^{n}\sum_{t=1}^{T}\dot{X}_{i,t}\dot{Y}_{i,t}\right) . \tag{4}$$

- **On Efficiency**: under the assumption that $\text{Var}[U_{i,t}|X_{i,1},\ldots,X_{i,T}]$ is constant (homoskedasticity), together with the assumption of no serial correlation in $U_{i,t}$, it is possible to show that $\hat{\beta}_n^{\text{fe}}$ is asymptotically efficient.

Asymptotic Properties

- **Asymptotic approximation in panel data models**: two elements that were not present with cross-sectional data: (1) the data is i.i.d. across i but may be dependent across time; (2) the data has two indices now: the number of units (n) and the number of time periods (T).

- **Required**: $nT \to \infty$, but we may achieve this by all sort of different assumptions about how n and/or T grow.

- Short Panels: $n \to \infty$ and T fixed vs Long Panels: $n \to \infty$ and $T \to \infty$.

- Under asymptotics where $n \to \infty$ and fixed T, we can show that $\hat{\beta}_n^{\text{fe}}$ and $\hat{\beta}_n^{\text{fd}}$ are asymptotically normal using similar arguments to those we use before, provided we assume

$$(Y_{i,1}, \ldots, Y_{i,T}, X_{i,1}, \ldots, X_{i,T}, U_{i,1}, \ldots, U_{i,T})$$

are i.i.d. across $i = 1, \ldots, n$.

Proof

$$\sqrt{n}(\hat{\beta}_n^{\text{fe}} - \beta) = \left(\frac{1}{n} \sum_{i=1}^{n} \sum_{t=1}^{T} \dot{X}_{i,t} \dot{X}_{i,t}' \right)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{t=1}^{T} \dot{X}_{i,t} \dot{U}_{i,t} \right) .$$

Proof

$$\sqrt{n}(\hat{\beta}_n^{\text{fe}} - \beta) = \left(\frac{1}{n}\sum_{i=1}^{n} \dot{X}_i'\dot{X}_i\right)^{-1} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \dot{X}_i' U_i\right) .$$

Estimating V: CCE

$$\sqrt{n}(\hat{\beta}_n^{\text{fe}} - \beta) \xrightarrow{d} N(0, \mathbb{V}^{\text{fe}}) \qquad \text{where} \qquad \mathbb{V}^{\text{fe}} = \Sigma_{\dot{X}}^{-1} \Omega \Sigma_{\dot{X}}^{-1} .$$

- Historically, researchers often assumed that $U_{i,t}$ was serially uncorrelated with variance independent of $X_{i,t}$ (i.e. homoskedastic). Default standard errors in Stata are still based on these assumptions!

- Most common strategy: use the fully robust consistent estimator of the asymptotic variance,

$$\hat{\mathbb{V}}^{\text{fe}} = \left(\frac{1}{n} \sum_{i=1}^{n} \dot{X}_i' \dot{X}_i \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} \dot{X}_i' \hat{U}_i \hat{U}_i' \dot{X}_i \right) \left(\frac{1}{n} \sum_{i=1}^{n} \dot{X}_i' \dot{X}_i \right)^{-1} ,$$

  where $\hat{U}_i = \dot{Y}_i - \dot{X}_i \hat{\beta}_n^{\text{fe}}$. This is what Stata computes when one uses the `cluster(unit)` option to `xtreg` where unit is the variable that indexes i.

- This estimator generalizes the HC estimator and is known as a cluster covariance estimator (CCE) and is consistent as $n \to \infty$, i.e., $\hat{\mathbb{V}}^{\text{fe}} \xrightarrow{P} \mathbb{V}^{\text{fe}}$.

On Efficiency

- Traditional arguments in favor of the fixed effects (or within-group) estimator $\hat{\beta}_n^{\mathrm{fe}}$ over the first-difference estimator $\hat{\beta}_n^{\mathrm{fd}}$ rely on the fact that under homoskedasticity and no-serial correlation of $U_{i,t}$, $\hat{\beta}_n^{\mathrm{fe}}$ has a lower asymptotic variance than $\hat{\beta}_n^{\mathrm{fd}}$.
- **Intuition**: taking first differences introduces correlation in $\Delta U_{i,t}$ as

$$\begin{aligned} E[\Delta U_{i,t}\Delta U_{i,t-1}] &= E[U_{i,t}U_{i,t-1} - U_{i,t-1}U_{i,t-1} - U_{i,t}U_{i,t-2} + U_{i,t-1}U_{i,t-2}] \\ &= -\,\mathrm{Var}(U_{i,t-1}) \ . \end{aligned}$$

- **Other extreme**: $U_{i,t}$ follows a random walk, i.e., $U_{i,t} = U_{i,t-1} + V_{i,t}$ for some i.i.d. sequence $V_{i,t}$, then $\Delta U_{i,t} = V_{i,t}$.
- These arguments still rely on homoskedasticity, so it is advised to simply use a robust standard error as above and forget about efficiency considerations.
- $T = 2$: these two estimators are numerically the same. In addition, first differences are used in dynamic panels and difference in differences, as we will discuss later.

QUESTIONS?

Random Effects

Panel Data

$$Y_{i,t} = X_{i,t}'\beta + \eta_i + U_{i,t}, \quad i = 1, \ldots, n \quad t = 1, \ldots, T \, ,$$

- Random effect models add the following assumption: **RE1**. $E[\eta_i | X_{i,1}, \ldots, X_{i,T}] = 0$.
- **Meaning**: Unobservable time-invariant factors that were being controlled for in the fixed effects approach are now assumed to be mean independent (ergo, uncorrelated) with the covariates at all time periods.
- The strict exogeneity condition of the fixed effects approach (i.e. FE1) is still maintained, so that the aggregate error term

  $$V_{it} = \eta_i + U_{i,t}$$

  now satisfies $E[V_{it} | X_{i1}, \ldots, X_{iT}] = 0$ for all $t = 1, \ldots, T$.
- **Immediate implication**: We can just estimate β by OLS!

Assumptions in Random Effects

- **Idea behind random effects**: exploit the serial correlation in V_{it} that is generated by the common shock η_i under some fairly strong assumptions with the goal of improving efficiency.

  RE2. (i) $\text{Var}[U_{i,t}|X_{i,1},\ldots,X_{i,T}] = \sigma_U^2$,
  (ii) $\text{Var}[\eta_i|X_{i,1},\ldots,X_{i,T}] = \sigma_\eta^2$,
  (iii) $E[U_{i,t}U_{i,s}|X_{i,1},\ldots,X_{i,T}] = 0$ for all $t \neq s$,
  (iv) $E[U_{i,t}\eta_i|X_{i,1},\ldots,X_{i,T}] = 0$ for all $t = 1,\ldots,T$.

  Under these assumptions,

  $$\text{Var}[V_{i,t}|X_{i,1},\ldots,X_{i,T}] = E[\eta_i^2 + U_{i,t}^2 + \eta_i U_{i,t} \mid X_{i,1},\ldots,X_{i,T}] = \sigma_\eta^2 + \sigma_U^2 ,$$

  and

  $$E[V_{i,t}V_{i,s}|X_{i,1},\ldots,X_{i,T}] = E[\eta_i^2 + U_{i,t}U_{i,s} + \eta_i U_{i,t} + \eta_i U_{i,s} \mid X_{i,1},\ldots,X_{i,T}] = \sigma_\eta^2 .$$

- **Assumptions lead to**:

ASSUMPTIONS LEAD TO

$$\Omega = E[V_i V_i' | X_i] = \sigma_U^2 \mathbb{I}_T + \sigma_\eta^2 \iota_T \iota_T' ,$$

where $\mathbb{I}_T$ is the $T \times T$ identity matrix and ι_T is a T-dimensional vector of ones.

- **Random effect estimator**: the estimator with the lowest asymptotic variance is

$$\hat{\beta}_n^{\text{re}} = \left(\sum_{i=1}^{n} X_i' \Omega^{-1} X_i \right)^{-1} \left(\sum_{i=1}^{n} X_i' \Omega^{-1} Y_i \right) , \qquad (5)$$

where $X_i = (X_{i,1}, \ldots, X_{i,T})'$ is the $T \times k$ vector of stacked observations for unit i, and similarly for Y_i.

- **Note**: this is just a generalized least squares (GLS) estimator of β. This GLS estimator is, nevertheless, unfeasible, since Ω depends on the unknown parameters σ_U^2 and σ_η^2. However, these two can be easily estimated to form $\hat{\Omega}$ and deliver a feasible GLS estimator of β.

Random Effects: Remarks

- **First**: the efficiency gains hold under the additional structure imposed by **RE1**.
- **Second**: the efficiency gains hold under the homoskedasticity and independence assumptions in **RE2** and do not hold more generally.
- **Third**: unlike the fixed effects estimator, the random effects approach allows to estimate regression coefficients associated with time-invariant covariates.
- **Fourth**: under RE1 and RE2 β is identified in a single cross-section. The parameters that require panel data for identification in this model are the variances of the components of the error σ_η^2 and σ_U^2, which are needed for the GLS approach.
- **Finally**: the terminology "fixed effects" and "random effects" is arguably confusing as η_i is random in both approaches.

Hausman Test

- **Hausman specification test**: compares $\hat{\beta}_n^{\text{fe}}$ with $\hat{\beta}_n^{\text{re}}$ to test the validity of RE1 (assuming RE2 holds).
- Under the null hypothesis that RE1 holds, both estimators are consistent but $\hat{\beta}_n^{\text{re}}$ is efficient.
- Under the alternative hypothesis, $\hat{\beta}_n^{\text{fe}}$ is consistent while $\hat{\beta}_n^{\text{re}}$ is not.
- Suppose we were to define a new estimator $\hat{\beta}_n^*$ as follows

$$\hat{\beta}_n^* = \hat{\beta}_n^{\text{fe}} I\{\text{Hausman test rejects}\} + \hat{\beta}_n^{\text{re}} I\{\text{Hausman test accepts}\} .$$

- The problem with this new estimator is that its finite sample distribution looks **very different** from the usual normal approximations (*uniformity* issue).
- The use of $\hat{\beta}_n^*$ should be **avoided**.

- One benefit of panel data: allows to analyze relationships that are inherently **dynamic**.
- **Model**: let $\{Y_{i,t} : 1 \leqslant i \leqslant n, 1 \leqslant t \leqslant T\}$ be a sequence of random variables and consider

$$Y_{i,t} = \rho Y_{i,t-1} + \eta_i + U_{i,t}, \quad i = 1, \ldots, n \quad t = 2, \ldots, T \ ,$$

  where $Y_{i,t-1}$ has a direct effect on $Y_{i,t}$, a feature sometimes referred to as **state dependence**.
- We assume that $|\rho| < 1$ and that the model is dynamically complete, i.e.,

$$E[U_{i,t}|Y_{i,t-1}, Y_{i,t-2}, \ldots] = 0 \text{ for all } t = 1, \ldots, T.$$

- **First Differences**:

$$\Delta Y_{i,t} = \rho \Delta Y_{i,t-1} + \Delta U_{i,t}, \quad i = 1, \ldots, n \quad t = 2, \ldots, T \ ,$$

- **Problem**:

Dynamic Models

$$\Delta Y_{i,t} = \rho \Delta Y_{i,t-1} + \Delta U_{i,t}, \quad i = 1, \ldots, n \quad t = 2, \ldots, T\ ,$$

- Similar problem with de-meaning transformations. This inherent endogeneity is a generic feature of models that have both **state dependence** and **time-invariant heterogeneity**.
- **IV approach**: use lagged outcomes $(Y_{i,t-2})$ as instruments.

THE END!