Econometric Analysis of Cross Section and Panel Data

Lecture 7: Panel Data

Zhian Hu

Central University of Finance and Economics

Fall 2024

This Lecture

- Hansen (2022): Chapter 17
- **Panel data or longitudinal data**: data structures consisting of observations on individuals for multiple time periods.
- It allows us to control for unobserved time-invariant endogeneity.

Introduction

- **Micro panel:** a large number of individuals (often in the 1000's or higher) and a relatively small number of time periods (often 2 to 20 years).
- **Macro panel:** a moderate number of individuals (e.g. 7-20) and a moderate number of time periods (20-60 years).
- **Key assumption:** the individuals are mutually independent while the observations for a given individual are correlated across time periods.

Time Indexing and Unbalanced Panels

- Y_{it} or X_{it} denotes variables for individual i in period t, where individuals $i = 1, \ldots, N$ and time periods as $t = 1, \ldots T$.
- Thus N is the number of individuals in the panel and T is the number of time series periods.
- Balanced panel: an equal number T of observations for each individual and the total number of observations is $n = NT$.
- Unbalanced panel: not balanced panel. Each individual is observed for a subset of T_i periods.

Time Indexing and Unbalanced Panels

Table 17.1: Observations from Investment Data Set

Firm Code Number	Year	I_{it}	$\overline{I}_i$	$\dot{I}_{it}$	Q_{it}	$\overline{Q}_i$	$\dot{Q}_{it}$	$\hat{e}_{it}$
32	1970	0.122	0.155	-0.033	1.17	0.62	0.55	.
32	1971	0.092	0.155	-0.063	0.79	0.62	0.17	-0.005
32	1972	0.094	0.155	-0.061	0.91	0.62	0.29	-0.005
32	1973	0.116	0.155	-0.039	0.29	0.62	-0.33	0.014
32	1974	0.099	0.155	-0.057	0.30	0.62	-0.32	-0.002
32	1975	0.187	0.155	0.032	0.56	0.62	-0.06	0.086
32	1976	0.349	0.155	0.194	0.38	0.62	-0.24	0.248
32	1977	0.182	0.155	0.027	0.57	0.62	-0.05	0.081
209	1987	0.095	0.071	0.024	9.06	21.57	-12.51	.
209	1988	0.044	0.071	-0.027	16.90	21.57	-4.67	-0.244
209	1989	0.069	0.071	-0.002	25.14	21.57	3.57	-0.257
209	1990	0.113	0.071	0.042	25.60	21.57	4.03	-0.226
209	1991	0.034	0.071	-0.037	31.14	21.57	9.57	-0.283

Notation

- Observations are pairs (Y_{it}, X_{it}) where Y_{it} is the dependent variable and X_{it} is a k-vector of regressors.
- It will be useful to cluster the observations at the level of the individual.
- Write $\boldsymbol{Y}_i$ as the $T_i \times 1$ stacked observations on Y_{it} for $t \in S_i$. Similarly, we write $\boldsymbol{X}_i$ as the $T_i \times k$ matrix of stacked X_{it}' for $t \in S_i$.
- In matrix, let $\boldsymbol{Y} = (\boldsymbol{Y}_1', \ldots, \boldsymbol{Y}_N')'$ denote the $n \times 1$ vector of stacked $\boldsymbol{Y}_i$, and set $\boldsymbol{X} = (\boldsymbol{X}_1', \ldots, \boldsymbol{X}_N')'$ similarly.

Pooled Regression

- The simplest model in panel regresion is pooled regresssion

$$
\begin{aligned}
Y_{it} &= X_{it}'\beta + e_{it} \\
\mathbb{E}\left[X_{it}e_{it}\right] &= 0
\end{aligned}
$$

  where β is a $k \times 1$ coefficient vector and e_{it} is an error.
- The standard estimator of β in the pooled regression model is least squares, which can be written as

$$
\begin{aligned}
\widehat{\beta}_{\text{pool}} &= \left(\sum_{i=1}^{N}\sum_{t\in S_i} X_{it}X_{it}'\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t\in S_i} X_{it}Y_{it}\right) \\
&= \left(\sum_{i=1}^{N} \boldsymbol{X}_i'\boldsymbol{X}_i\right)^{-1}\left(\sum_{i=1}^{N} \boldsymbol{X}_i'\boldsymbol{Y}_i\right) \\
&= \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{Y}\right)
\end{aligned}
$$

- $\widehat{\beta}_{\text{pool}}$ is called the pooled regression estimator.

Pooled Regression

- The pooled regression model is ideally suited for the context where the errors e_{it} satisfy strict mean independence:

$$\mathbb{E}\left[e_{it} \mid \boldsymbol{X}_i\right] = 0$$

- Strict mean independence requires that neither lagged nor future values of X_{it} help to forecast e_{it}.

$$\begin{aligned}\widehat{\beta}_{\mathrm{pool}} &= \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{X}_i\right)^{-1} \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \left(\boldsymbol{X}_i \beta + \boldsymbol{e}_i\right)\right) \\ &= \beta + \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{X}_i\right)^{-1} \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{e}_i\right)\end{aligned}$$

Pooled Regression

- Then we can show

$$\mathbb{E}\left[\widehat{\beta}_{\text{pool}} \mid \boldsymbol{X}\right] = \beta + \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{X}_i\right)^{-1} \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \mathbb{E}\left[\boldsymbol{e}_i \mid \boldsymbol{X}_i\right]\right) = \beta$$

- We expect the errors e_{it} to be correlated across time t for a given individual.
- Thus, we use a cluster-robust covariance matrix estimator which allows arbitrary within-cluster dependence, using the Stata command *regress cluster(id)* where *id* indicates the individual.
- When strict mean independence fails the pooled least squares estimator $\widehat{\beta}_{\text{pool}}$ is not necessarily consistent for β.

One-Way Error Component Model

- One approach to panel data regression is to model the correlation structure of the regression error e_{it}.
- The most common choice is an error-components structure. The simplest takes the form

$$e_{it} = u_i + \varepsilon_{it}$$

where u_i is an individual-specific effect and ε_{it} are idiosyncratic (i.i.d.) errors. This is known as a oneway error component model.
- In vector notation we can write $\boldsymbol{e}_i = \mathbf{1}_i u_i + \boldsymbol{\varepsilon}_i$ where $\mathbf{1}_i$ is a $T_i \times 1$ vector of 1's.
- The one-way error component regression model is

$$Y_{it} = X_{it}'\beta + u_i + \varepsilon_{it}$$

or $\boldsymbol{Y}_i = \boldsymbol{X}_i\beta + \mathbf{1}_i u_i + \boldsymbol{\varepsilon}_i$ written at the level of the individual.

Random Effects

- Random effects model assumes $e_{it} = u_i + \varepsilon_{it}$ and

$$
\begin{aligned}
\mathbb{E}\left[\varepsilon_{it} \mid \boldsymbol{X}_i\right] &= 0 \\
\mathbb{E}\left[\varepsilon_{it}^2 \mid \boldsymbol{X}_i\right] &= \sigma_\varepsilon^2 \\
\mathbb{E}\left[\varepsilon_{it}\varepsilon_{js} \mid \boldsymbol{X}_i\right] &= 0, s \neq t \\
\mathbb{E}\left[u_i \mid \boldsymbol{X}_i\right] &= 0 \\
\mathbb{E}\left[u_i^2 \mid \boldsymbol{X}_i\right] &= \sigma_u^2 \\
\mathbb{E}\left[u_i\varepsilon_{it} \mid \boldsymbol{X}_i\right] &= 0
\end{aligned}
$$

Random Effects

- The random effects model implies that the vector of errors $\boldsymbol{e}_i$ for individual i has the covariance structure

$$\begin{aligned}
\mathbb{E}\left[\boldsymbol{e}_i \mid \boldsymbol{X}_i\right] &= 0 \\
\mathbb{E}\left[\boldsymbol{e}_i \boldsymbol{e}_i' \mid \boldsymbol{X}_i\right] &= \mathbf{1}_i \mathbf{1}_i' \sigma_u^2 + \boldsymbol{I}_i \sigma_\varepsilon^2 \\
&= \begin{pmatrix} \sigma_u^2 + \sigma_\varepsilon^2 & \sigma_u^2 & \cdots & \sigma_u^2 \\ \sigma_u^2 & \sigma_u^2 + \sigma_\varepsilon^2 & \cdots & \sigma_u^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_u^2 & \sigma_u^2 & \cdots & \sigma_u^2 + \sigma_\varepsilon^2 \end{pmatrix} \\
&= \Omega_i
\end{aligned}$$

where $\boldsymbol{I}_i$ is an identity matrix of dimension T_i. The matrix Ω_i depends on i since its dimension depends on the number of observed time periods T_i.

Random Effects

- Given the error structure the natural estimator for β is GLS. Suppose σ_u^2 and σ_ε^2 are known. The GLS estimator of β is

$$\widehat{\beta}_{\mathrm{gls}} = \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \Omega_i^{-1} \boldsymbol{X}_i \right)^{-1} \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \Omega_i^{-1} \boldsymbol{Y}_i \right)$$

- A feasible GLS estimator replaces the unknown σ_u^2 and σ_ε^2 with estimators.
- By linearity

$$\mathbb{E}\left[\widehat{\beta}_{\mathrm{gls}} - \beta \mid \boldsymbol{X} \right] = \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \Omega_i^{-1} \boldsymbol{X}_i \right)^{-1} \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \Omega_i^{-1} \mathbb{E}\left[\boldsymbol{e}_i \mid \boldsymbol{X}_i \right] \right) = 0$$

- You should remember that $\boldsymbol{V}_{\mathrm{gls}} \leq \boldsymbol{V}_{\mathrm{pool}}$, so the random effects estimator $\widehat{\beta}_{\mathrm{gls}}$ is more efficient than the pooled estimator $\widehat{\beta}_{\mathrm{pool}}$.

Fixed Effect Model

- Consider the one-way error component regression model

$$Y_{it} = X_{it}'\beta + u_i + \varepsilon_{it}$$

  or

$$\boldsymbol{Y}_i = \boldsymbol{X}_i\beta + \mathbf{1}_i u_i + \boldsymbol{\varepsilon}_i$$

- In many applications it is useful to interpret the individual-specific effect u_i as a time-invariant unobserved missing variable. For example, in a wage regression u_i may be the unobserved ability of individual i.
- **When u_i is interpreted as an omitted variable it is natural to expect it to be correlated with the regressors X_{it}. This is especially the case when X_{it} includes choice variables.**

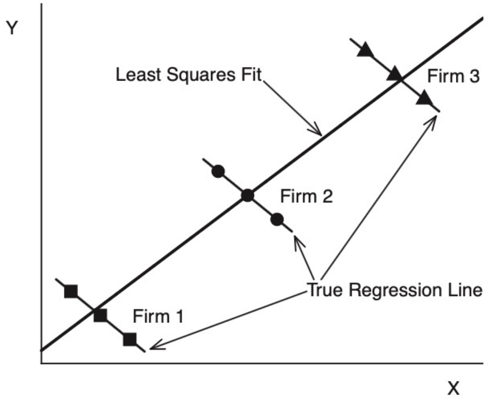
Fixed Effect Model

- In the econometrics literature if the stochastic structure of u_i is treated as unknown and possibly correlated with X_{it} then u_i is called a fixed effect.
- Correlation between u_i and X_{it} will cause both pooled and random effect estimators to be biased. This is due to the classic problems of omitted variables bias and endogeneity.

Fixed Effect Model

- Consider the scatter plot of three observations (Y_{it}, X_{it}) from three firms.
- The true model is $Y_{it} = 9 - X_{it} + u_i$. (The true slope coefficient is -1.)
- The variables u_i and X_{it} are highly correlated so the fitted pooled regression line has a slope close to $+1$.
- Conditional on u, however, the slope is -1. Thus regression techniques which do not control for u_i will produce biased and inconsistent estimators.

Fixed Effect Model



Fixed Effect Model

- To identify β, we need the strict exogeneity assumption.
- The regressor X_{it} is strictly exogenous for the error ε_{it} if

$$\mathbb{E}\left[X_{is}\varepsilon_{it}\right] = 0$$

  for all $s = 1, \ldots, T$.
- This assumption is much weaker than the assumption in random effects model, which requires that the individual effect u_i is also strictly mean independent.

Within Transformation

- The first way to consistently estimate β is to eliminate u_i using within transformation.
- Define the mean of a variable for a given individual as

$$\bar{Y}_i = \frac{1}{T_i} \sum_{t \in S_i} Y_{it}$$

- Subtracting the individual-specific mean from the variable we obtain the deviations

$$\dot{Y}_{it} = Y_{it} - \bar{Y}_i$$

which is known as within transformation. We also refer to $\dot{Y}_{it}$ as the demeaned values or deviations from individual means.

Within Transformation

- We can also write

$$\begin{aligned}\dot{\boldsymbol{Y}}_i &= \boldsymbol{Y}_i - \mathbf{1}_i \bar{Y}_i \\ &= \boldsymbol{Y}_i - \mathbf{1}_i \left(\mathbf{1}_i'\mathbf{1}_i\right)^{-1} \mathbf{1}_i' \boldsymbol{Y}_i \\ &= \boldsymbol{M}_i \boldsymbol{Y}_i\end{aligned}$$

  where $\boldsymbol{M}_i = \boldsymbol{I}_i - \mathbf{1}_i \left(\mathbf{1}_i'\mathbf{1}_i\right)^{-1} \mathbf{1}_i'$ is the individual-specific demeaning operator.
- Similarly, define

$$\begin{aligned}\bar{X}_i &= \frac{1}{T_i} \sum_{t \in S_i} X_{it} \\ \dot{X}_{it} &= X_{it} - \bar{X}_i \\ \dot{\boldsymbol{X}}_i &= \boldsymbol{M}_i \boldsymbol{X}_i.\end{aligned}$$

Within Transformation

- Taking individual-specific averages we obtain

$$\bar{Y}_i = \bar{X}_i'\beta + u_i + \bar{\varepsilon}_i$$

where $\bar{\varepsilon}_i = \frac{1}{T_i}\sum_{t\in S_i}\varepsilon_{it}$.

- Subtracting from $Y_{it} = X_{it}'\beta + u_i + \varepsilon_{it}$ we obtain

$$\dot{Y}_{it} = \dot{X}_{it}'\beta + \dot{\varepsilon}_{it}$$

where $\dot{\varepsilon}_{it} = \varepsilon_{it} - \bar{\varepsilon}_{it}$. The individual effect u_i has been eliminated!

- We can alternatively write this in vector notation.

$$\dot{\boldsymbol{Y}}_i = \dot{\boldsymbol{X}}_i\beta + \dot{\boldsymbol{\varepsilon}}_i$$

Within Transformation

- Another consequence, however, is that all time-invariant regressors are also eliminated.
- In this framework, it will be impossible to estimate (or identify) a coefficient on any regressor which is time invariant.
- The within transformation can greatly reduce the variance of the regressors.

Fixed Effects Estimator

$$\widehat{\beta}_{\mathrm{fe}} = \left(\sum_{i=1}^{N} \sum_{t \in S_i} \dot{X}_{it} \dot{X}_{it}' \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t \in S_i} \dot{X}_{it} \dot{Y}_{it} \right)$$

$$= \left(\sum_{i=1}^{N} \dot{\boldsymbol{X}}_i' \dot{\boldsymbol{X}}_i \right)^{-1} \left(\sum_{i=1}^{N} \dot{\boldsymbol{X}}_i' \dot{\boldsymbol{Y}}_i \right)$$

$$= \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{M}_i \boldsymbol{X}_i \right)^{-1} \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{M}_i \boldsymbol{Y}_i \right).$$

- This is known as the fixed-effects or within estimator of β.

Fixed Effects Estimator

- Let us describe some of the statistical properties of the estimator under strict mean independence $\mathbb{E}\left[\varepsilon_{it} \mid \boldsymbol{X}_i\right] = 0$.
- We can write

$$\widehat{\beta}_{\mathrm{fe}} - \beta = \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{M}_i \boldsymbol{X}_i\right)^{-1} \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{M}_i \boldsymbol{\varepsilon}_i\right)$$

- Then

$$\mathbb{E}\left[\widehat{\beta}_{\mathrm{fe}} - \beta \mid \boldsymbol{X}\right] = \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{M}_i \boldsymbol{X}_i\right)^{-1} \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{M}_i \mathbb{E}\left[\boldsymbol{\varepsilon}_i \mid \boldsymbol{X}_i\right]\right) = 0$$

Thus $\widehat{\beta}_{\mathrm{fe}}$ is unbiased for β.

Fixed Effects Estimator

- Let $\Sigma_i = \mathbb{E}\left[\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i' \mid \boldsymbol{X}_i\right]$ denote the $T_i \times T_i$ conditional covariance matrix of the idiosyncratic errors. The variance of $\widehat{\beta}_{\mathrm{fe}}$ is

$$\boldsymbol{V}_{\mathrm{fe}} = \operatorname{var}\left[\widehat{\beta}_{\mathrm{fe}} \mid \boldsymbol{X}\right] = \left(\sum_{i=1}^{N} \dot{\boldsymbol{X}}_i' \dot{\boldsymbol{X}}_i\right)^{-1} \left(\sum_{i=1}^{N} \dot{\boldsymbol{X}}_i' \Sigma_i \dot{\boldsymbol{X}}_i\right) \left(\sum_{i=1}^{N} \dot{\boldsymbol{X}}_i' \dot{\boldsymbol{X}}_i\right)^{-1}$$

- It simplifies when the idiosyncratic errors are homoskedastic and serially uncorrelated:

$$\mathbb{E}\left[\varepsilon_{it}^2 \mid \boldsymbol{X}_i\right] = \sigma_\varepsilon^2$$
$$\mathbb{E}\left[\varepsilon_{ij}\varepsilon_{it} \mid \boldsymbol{X}_i\right] = 0$$

for all $j \neq t$. In this case, $\Sigma_i = I_i \sigma_\varepsilon^2$ and $\boldsymbol{V}_{\mathrm{fe}}$ simplifies to

$$\boldsymbol{V}_{\mathrm{fe}}^0 = \sigma_\varepsilon^2 \left(\sum_{i=1}^{N} \dot{\boldsymbol{X}}_i' \dot{\boldsymbol{X}}_i\right)^{-1}$$

Differenced Estimator

- Another important transformation which eliminates the individual-specific effect is first-differencing.
- The first-differencing transformation is $\Delta Y_{it} = Y_{it} - Y_{it-1}$. This can be applied to all but the first observation (which is essentially lost).
- At the level of the individual this can be written as $\Delta \boldsymbol{Y}_i = \boldsymbol{D}_i \boldsymbol{Y}_i$ where $\boldsymbol{D}_i$ is the $(T_i - 1) \times T_i$ matrix differencing operator

$$\boldsymbol{D}_i = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & & 0 & 0 \\ \vdots & & & \ddots & & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix}$$

Differenced Estimator

- Applying the transformation Δ to $Y_{it} = X_{it}'\beta + u_i + \varepsilon_{it}$ we obtain $\Delta Y_{it} = \Delta X_{it}'\beta + \Delta\varepsilon_{it}$ or
$$\Delta \boldsymbol{Y}_i = \Delta \boldsymbol{X}_i \beta + \Delta \boldsymbol{\varepsilon}_i$$
- We can see that the individual effect u_i has been eliminated.
- Least squares applied to the differenced equation is

$$\begin{aligned}
\widehat{\beta}_\Delta &= \left(\sum_{i=1}^{N} \sum_{t\geq 2} \Delta X_{it} \Delta X_{it}' \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t\geq 2} \Delta X_{it} \Delta Y_{it} \right) \\
&= \left(\sum_{i=1}^{N} \Delta \boldsymbol{X}_i' \Delta \boldsymbol{X}_i \right)^{-1} \left(\sum_{i=1}^{N} \Delta \boldsymbol{X}_i' \Delta \boldsymbol{Y}_i \right) \\
&= \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{D}_i' \boldsymbol{D}_i \boldsymbol{X}_i \right)^{-1} \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{D}_i' \boldsymbol{D}_i \boldsymbol{Y}_i \right)
\end{aligned}$$

Differenced Estimator

- $\widehat{\beta}_{\Delta}$ is called the differenced estimator. For $T = 2, \widehat{\beta}_{\Delta} = \widehat{\beta}_{\text{fe}}$ equals the fixed effects estimator.
- When the errors ε_{it} are serially uncorrelated and homoskedastic then the error $\Delta\boldsymbol{\varepsilon}_i = \boldsymbol{D}_i\boldsymbol{\varepsilon}_i$ has covariance matrix $\boldsymbol{H}\sigma_{\varepsilon}^2$ where

$$\boldsymbol{H} = \boldsymbol{D}_i\boldsymbol{D}_i' = \begin{pmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & \ddots & 0 \\ 0 & \ddots & \ddots & -1 \\ 0 & 0 & -1 & 2 \end{pmatrix}$$

Differenced Estimator

- We can reduce estimation variance by using GLS. When the errors ε_{it} are i.i.d. (serially uncorrelated and homoskedastic), this is

$$
\begin{aligned}
\widetilde{\beta}_{\Delta} &= \left(\sum_{i=1}^{N} \Delta \boldsymbol{X}_i' \boldsymbol{H}^{-1} \Delta \boldsymbol{X}_i \right)^{-1} \left(\sum_{i=1}^{N} \Delta \boldsymbol{X}_i' \boldsymbol{H}^{-1} \Delta \boldsymbol{Y}_i \right) \\
&= \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{D}_i' \left(\boldsymbol{D}_i \boldsymbol{D}_i' \right)^{-1} \boldsymbol{D}_i \boldsymbol{X}_i \right)^{-1} \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{D}_i' \left(\boldsymbol{D}_i \boldsymbol{D}_i' \right)^{-1} \boldsymbol{D}_i \boldsymbol{Y}_i \right) \\
&= \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{M}_i \boldsymbol{X}_i \right)^{-1} \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{M}_i \boldsymbol{Y}_i \right)
\end{aligned}
$$

  where $\boldsymbol{M}_i = \boldsymbol{D}_i' \left(\boldsymbol{D}_i \boldsymbol{D}_i' \right)^{-1} \boldsymbol{D}_i$.
- We can find that $\widetilde{\beta}_{\Delta} = \widehat{\beta}_{\mathrm{fe}}$, the fixed effects estimator!

Differenced Estimator

- What we have shown is that under i.i.d. errors, GLS applied to the first-differenced equation precisely equals the fixed effects estimator.
- Since the Gauss-Markov theorem shows that GLS has lower variance than least squares, this means that the fixed effects estimator is more efficient than first differencing under the assumption that ε_{it} is i.i.d.

Dummy Variables Regression

- An alternative way to estimate the fixed effects model is by least squares of Y_{it} on X_{it} and a full set of dummy variables, one for each individual in the sample.
- To see this start with the error-component model without a regressor:

$$Y_{it} = u_i + \varepsilon_{it}$$

- Consider least squares estimation of the vector of fixed effects $u = (u_1, \dots, u_N)'$.
- Now let d_i be a vector of N dummy variables where the i^{th} element indicates the i^{th} individual.
- Thus the i^{th} element of d_i is 1 and the remaining elements are zero.

Dummy Variables Regression

- Notice that $u_i = d_i' u$ and then $Y_{it} = d_i' u + \varepsilon_{it}$.
- This is a regression with the regressors d_i and coefficients u. We can also write this in vector notation at the level of the individual as $\boldsymbol{Y}_i = \mathbf{1}_i d_i' u + \boldsymbol{\varepsilon}_i$ or using full matrix notation as $\boldsymbol{Y} = \boldsymbol{D}\boldsymbol{u} + \boldsymbol{\varepsilon}$ where $\boldsymbol{D} = \operatorname{diag}\{\mathbf{1}_{T_1}, \ldots, \mathbf{1}_{T_N}\}$.
- The least squares estimate of u is

$$\begin{aligned}
\widehat{\boldsymbol{u}} &= \left(\boldsymbol{D}'\boldsymbol{D}\right)^{-1}\left(\boldsymbol{D}'\boldsymbol{Y}\right) \\
&= \operatorname{diag}\left(\mathbf{1}_i'\mathbf{1}_i\right)^{-1}\left\{\mathbf{1}_i'\boldsymbol{Y}_i\right\}_{i=1,\ldots,n} \\
&= \left\{\left(\mathbf{1}_i'\mathbf{1}_i\right)^{-1}\mathbf{1}_i'\boldsymbol{Y}_i\right\}_{i=1,\ldots,n} \\
&= \left\{\bar{Y}_i\right\}_{i=1,\ldots,n}
\end{aligned}$$

- The least squares residuals are $\widehat{\boldsymbol{\varepsilon}} = \left(\boldsymbol{I}_n - \boldsymbol{D}\left(\boldsymbol{D}'\boldsymbol{D}\right)^{-1}\boldsymbol{D}'\right)\boldsymbol{Y} = \dot{\boldsymbol{Y}}$

Dummy Variables Regression

- Now consider the error-component model with regressors, which can be written as

$$Y_{it} = X_{it}'\beta + d_i' u + \varepsilon_{it}$$

- since $u_i = d_i' u$ as discussed above. In matrix notation

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{D}u + \boldsymbol{\varepsilon}$$

- We consider estimation of (β, u) by least squares and write the estimates as $\boldsymbol{Y} = \boldsymbol{X}\widehat{\boldsymbol{\beta}} + \boldsymbol{D}\widehat{\boldsymbol{u}} + \widehat{\boldsymbol{\varepsilon}}$. We call this the dummy variable estimator of the fixed effects model.

Dummy Variables Regression

- By the Frisch-Waugh-Lovell Theorem the dummy variable estimator $\widehat{\beta}$ and residuals $\widehat{\boldsymbol{\varepsilon}}$ may be obtained by the least squares regression of the residuals from the regression of $\boldsymbol{Y}$ on $\boldsymbol{D}$ on the residuals from the regression of $\boldsymbol{X}$ on $\boldsymbol{D}$.
- We learned above that the residuals from the regression on $\boldsymbol{D}$ are the within transformations.
- Thus the dummy variable estimator $\widehat{\beta}$ and residuals $\widehat{\varepsilon}$ may be obtained from least squares regression of the within transformed $\dot{\boldsymbol{Y}}$ on the within transformed $\dot{\boldsymbol{X}}$.
- This is exactly the fixed effects estimator $\widehat{\beta}_{\mathrm{fe}}$. Thus the dummy variable and fixed effects estimators of β are identical.

Fixed Effects Covariance Matrix Estimation

- First consider estimation of the classical covariance matrix $\boldsymbol{V}_{\mathrm{fe}}^{0}$ is

$$\widehat{\boldsymbol{V}}_{\mathrm{fe}}^{0} = \widehat{\sigma}_{\varepsilon}^{2}\left(\dot{\boldsymbol{X}}'\dot{\boldsymbol{X}}\right)^{-1}$$

with

$$\widehat{\sigma}_{\varepsilon}^{2} = \frac{1}{n-N-k}\sum_{i=1}^{n}\sum_{t\in S_i}\widehat{\varepsilon}_{it}^{2} = \frac{1}{n-N-k}\sum_{i=1}^{n}\widehat{\varepsilon}_i'\widehat{\varepsilon}_i$$

- The $N+k$ degree of freedom adjustment is motivated by the dummy variable representation.

Fixed Effects Covariance Matrix Estimation

- A covariance matrix estimator which allows ε_{it} to be heteroskedastic and serially correlated across t is the cluster-robust covariance matrix estimator, clustered by individual

$$\widehat{\boldsymbol{V}}_{\mathrm{fe}}^{\text{cluster}} = \left(\dot{\boldsymbol{X}}'\dot{\boldsymbol{X}}\right)^{-1}\left(\sum_{i=1}^{N}\dot{\boldsymbol{X}}_i'\widehat{\boldsymbol{\varepsilon}}_i\widehat{\boldsymbol{\varepsilon}}_i'\dot{\boldsymbol{X}}_i\right)\left(\dot{\boldsymbol{X}}'\dot{\boldsymbol{X}}\right)^{-1}$$

- where $\widehat{\boldsymbol{\varepsilon}}_i$ as the fixed effects residuals.
- $\widehat{\boldsymbol{V}}_{\mathrm{fe}}^{\text{cluster}}$ can be multiplied by a degree-of-freedom adjustment.

$$\widehat{\boldsymbol{V}}_{\mathrm{fe}}^{\text{cluster}} = \left(\frac{n-1}{n-N-k}\right)\left(\frac{N}{N-1}\right)\left(\dot{\boldsymbol{X}}'\dot{\boldsymbol{X}}\right)^{-1}\left(\sum_{i=1}^{N}\dot{\boldsymbol{X}}_i'\widehat{\boldsymbol{\varepsilon}}_i\widehat{\boldsymbol{\varepsilon}}_i'\dot{\boldsymbol{X}}_i\right)\left(\dot{\boldsymbol{X}}'\dot{\boldsymbol{X}}\right)^{-1}$$