Panel Econometrics

Paulo M. M. Rodrigues

February 2022

Today's outline

Estimation of linear panel data models:

Pooled OLS and fixed effects estimation

Outline

The one-way error components model

A model to start with

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + v_{it}$$

Random sample of individuals (or households, firms, ...) $i = 1, \ldots, N$ for time periods $t = 1, \ldots, T$.

The error v_{it} contains *unobservable* individual specific effects c_i (intellectual ability, gender etc.) and a remainder disturbance:

$$v_{it} = c_i + u_{it}$$

This is called a one-way error components structure: both c_i and u_{it} are unobservable random variables.

Before finding estimators, we have to think about the assumptions.

Strict exogeneity I

Strict exogeneity assumption:

$$\mathrm{E}(y_{it}|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT},c_i) = \mathrm{E}(y_{it}|\mathbf{x}_{it},c_i)$$

Interpretation: once $\mathbf{x}_{it}$ and c_i are controlled for, $\mathbf{x}_{is}$ has no partial effect on y_{it} for all $s \neq t$.

(We may explicitly separate *variables of interest* from *controls* at times.)

Adding linearity:

$$\mathrm{E}(y_{it}|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT},c_i) = \mathbf{x}_{it}\boldsymbol{\beta} + c_i.$$

Error form:

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}.$$

Strict exogeneity II

Implication of the strict exogeneity assumption for the disturbance:

$$\mathrm{E}(u_{it}|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT},c_i)=0 \qquad t=1,\ldots,T.$$

This implies in turn

$$\mathrm{E}(\mathbf{x}_{is}'u_{it})=0 \qquad s,t=1,\ldots,T.$$

But it leaves

$$\mathrm{E}(\mathbf{x}_{it}'c_i)$$

fully unrestricted (which is nice for us).

Large N asymptotics

In the following we use large N asymptotics.

This means, to derive an asymptotic distribution we keep T fixed and let $N \to \infty$.

Whether this leads to an asymptotic distribution that approximates the (unknown) finite-sample distribution well, obviously depends on the sample relation between N and T. Whenever $N \geq T$ it should work nicely.

Panels with $N \approx T$ or $N < T$ typically require different asymptotics but we will not cover them in this course.

We'll typically let u_{it} be iid (or so).

Random effects or fixed effects?

In earlier literature, there was a discussion whether c_i should be treated as a fixed parameter (fixed effect) or as a random variable (random effect).

Since we typically assume we have a random sample of individuals (or households, firms, etc.), c_i should be treated as a random variable.

But as we shall see, there is still an important distinction between the fixed effects **estimator** and the random effects **estimator**:

- Fixed effects estimator: the c_i are allowed to correlate with $\mathbf{x}_{it}$.
- Random effects estimator: the c_i need to be uncorrelated with $\mathbf{x}_{it}$, $t = 1, \ldots, T$.

Questions to be answered by the researcher

(1) Are the c_i uncorrelated with $\mathbf{x}_{it}$, for all $t = 1, \ldots, T$?

(2) Is the strict exogeneity assumption $\mathrm{E}(u_{it}|\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}, c_i) = 0$ reasonable?

This may be difficult at times; see below.

Example: program evaluation

Model:

$$\log(\texttt{wage}_{it}) = \theta_t + \mathbf{z}_{it}\boldsymbol{\gamma} + \delta_1 \texttt{prog}_{it} + c_i + u_{it} \qquad t = 1, 2$$

$t = 1$: no-one has participated $\rightarrow$ $\texttt{prog}_{it} = 0$ for all i
$t = 2$: treatment group has participated, control group not

Role of c_i: participation may depend on personal characteristics (self selection, non-random assignment to one of the groups)

Discussion:

(1) The c_i are probably correlated with $\mathbf{z}_{it}$.

(2) $\mathrm{Cov}(u_{it}, \mathbf{x}_{it}) = 0$ may be uncritical as long as we include all important control variables (and because we include c_i)

(3) What about $\mathrm{Cov}(u_{i\mathbf{1}}, \mathbf{x}_{i\mathbf{2}}) = 0$?

Think of a feedback: Negative income shock u_{i1} may induce people to participate in the program or program administrators to choose these people to participate. Then $\mathrm{Cov}(u_{i1}, \texttt{prog}_{i2}) \neq 0$.

Example: models of lagged adjustment

Distributed lag model to study the relationship between patents and current and past levels of R&D spending:

$$\texttt{patents}_{it} = \theta_t + \mathbf{z}_{it}\boldsymbol{\gamma} + \delta_0 \texttt{RD}_{it} + \delta_1 \texttt{RD}_{it-1} + \delta_2 \texttt{RD}_{it-2} + \cdots + c_i + u_{it}$$

Hence, interest rests on the δ_j's.

Role of c_i: unobserved firm heterogeneity (firm culture, risk attitudes, productivity)

Discussion:

(1) The c_i probably correlated with $\texttt{RD}_{it}$ and its lags unless all important factors of firm heterogeneity are controlled for.

(2) Feedback: Negative patent shock u_{it} may induce firm to spend more on future R&D. Then $\mathrm{Cov}(u_{it}, \texttt{RD}_{it+\tau}) \neq 0$.

Example: lagged dependent variable

Model of dynamic wage adjustment:

$$\log(\texttt{wage}_{it}) = \beta_1 \log(\texttt{wage}_{it-1}) + c_i + u_{it} \qquad t = 1, \ldots, T$$

Interest is on speed of wage adjustment.

Role of c_i: unobserved heterogeneity (e.g., individual productivity)

Reasonable assumption: $\mathrm{E}(u_{it} | \log(\texttt{wage}_{it-1}), \ldots, \log(\texttt{wage}_{i0}), c_i) = 0$.

Discussion:

(1) The c_i by construction correlated with $\log(\texttt{wage}_{it-1})$.

(2) "Feedback" shows up by construction: strict exogeneity fails.

Outline

Some LS stuff for starters

Stacking observations $t = 1, \dots, T$ for individual i yields

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{v}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\iota}_T c_i + \mathbf{u}_i,$$

where $\boldsymbol{\iota}_T$ is a $T \times 1$ vector of ones. Stacking individuals $i = 1, \dots, N$ yields

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{v},$$

where

$$\underset{(NT\times 1)}{\mathbf{y}} = \begin{bmatrix} y_{11} \\ \vdots \\ y_{1T} \\ \vdots \\ y_{N1} \\ \vdots \\ y_{NT} \end{bmatrix} \qquad \underset{(NT\times K)}{\mathbf{X}} = \begin{bmatrix} \mathbf{x}_{11} \\ \vdots \\ \mathbf{x}_{1T} \\ \vdots \\ \mathbf{x}_{N1} \\ \vdots \\ \mathbf{x}_{NT} \end{bmatrix} \qquad \underset{(NT\times 1)}{\mathbf{v}} = \begin{bmatrix} v_{11} \\ \vdots \\ v_{1T} \\ \vdots \\ v_{N1} \\ \vdots \\ v_{NT} \end{bmatrix} = \begin{bmatrix} c_1 \\ \vdots \\ c_1 \\ \vdots \\ c_N \\ \vdots \\ c_N \end{bmatrix} + \begin{bmatrix} u_{11} \\ \vdots \\ u_{1T} \\ \vdots \\ u_{N1} \\ \vdots \\ u_{NT} \end{bmatrix}$$

Pooled OLS estimator

The classical estimator is:

$$\hat{\boldsymbol{\beta}}_{POLS} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y} = \left(\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{X}_i\right)^{-1}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{y}_i$$

Condition for consistency:

$$\mathrm{E}(\mathbf{x}_{it}'v_{it}) = \mathbf{0}$$

which is satisfied if

$$\mathrm{E}(\mathbf{x}_{it}'u_{it}) = \mathbf{0} \qquad \text{and} \qquad \mathrm{E}(\mathbf{x}_{it}'c_i) = \mathbf{0}.$$

Hence, omitted variable problem **only avoided if** c is not confounding.

Variance of the pooled OLS estimator

Even if the error component u_{it} of

$$v_{it} = c_i + u_{it}$$

is white noise with time-invariant variance, the overall disturbances v_{it} and v_{is} are correlated due to c_i:

$$\mathrm{E}(v_{it}v_{js}) = \mathrm{E}[(c_i + u_{it})(c_j + u_{js})] = \begin{cases} \mathrm{Var}(c_i) + \mathrm{Var}(u_{it}) & \text{if } i = j \text{ and } t = s \\ \mathrm{Var}(c_i) & \text{if } i = j \text{ and } t \neq s \\ 0 & \text{if } i \neq j \end{cases}$$

Hence, the standard assumption that elements are independent is violated.

This implies that conventional standard errors computed for $\hat{\boldsymbol{\beta}}_{POLS}$ are incorrect, no matter whether they are heteroskedasticity robust or not.

But there is more

Even without c_i, we're still not back to the cross-section case anymore.

To see the problem, let us number the $N \cdot T$ pooled observations as $l = 1, \ldots, NT$ as if it was a single cross section.

One may be tempted to base the asymptotic distribution of the pooled OLS estimator on a CLT for the cross section

$$(NT)^{-\frac{1}{2}} \sum_{l=1}^{NT} \mathbf{x}_l' v_l = (NT)^{-\frac{1}{2}} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{x}_{it}' v_{it} \xrightarrow{\text{d}} \mathcal{N}(\mathbf{0}, \mathbf{B})$$

where

$$\mathbf{B} \equiv \text{Cov}\left(\mathbf{x}_l' v_l\right) = \text{Cov}\left(\mathbf{x}_{it}' v_{it}\right).$$

In fact, this is what a regression software would do unless you tell it that this is *not* a single cross section.

Problem: this CLT requires the (here: invalid) assumption that $\text{E}(\mathbf{x}_l' v_l v_m \mathbf{x}_m) \neq 0$ only if $l = m$ and thus $i = j$ and $t = s$.

For fixed-T panels

The asymptotic distribution of the pooled OLS estimator is based on

$$N^{-\frac{1}{2}} \sum_{i=1}^{N} \mathbf{X}_i' \mathbf{v}_i \xrightarrow{\text{d}} \mathcal{N}(\mathbf{0}, \mathbf{B})$$

where $\mathbf{B} \equiv \mathrm{E}\left(\mathbf{X}_i' \mathbf{v}_i \mathbf{v}_i' \mathbf{X}_i\right) \equiv \mathrm{Cov}\left(\mathbf{X}_i' \mathbf{v}_i\right)$.

The pooled OLS estimator is asymptotically normally distributed,

$$\sqrt{N}\left(\hat{\boldsymbol{\beta}}_{POLS} - \boldsymbol{\beta}\right) \xrightarrow{\text{d}} \mathcal{N}\left(\mathbf{0}, \mathbf{A}^{-1} \mathbf{B} \mathbf{A}^{-1}\right),$$

where $\mathbf{A} \equiv \mathrm{E}\left[\mathbf{X}_i' \mathbf{X}_i\right]$.

We estimate the asymptotic variance by sample counterparts:

$$\hat{\mathbf{A}} = N^{-1} \sum_{i=1}^{N} \mathbf{X}_i' \mathbf{X}_i \quad \text{and} \quad \hat{\mathbf{B}} = N^{-1} \sum_{i=1}^{N} \mathbf{X}_i' \hat{\mathbf{v}}_i \hat{\mathbf{v}}_i' \mathbf{X}_i.$$

Note: the variance estimator leaves the correlation between different time periods of the same individual fully unrestricted. (May think of GLS.)

Software/Stata hints

What all this means when you estimate a pooled regression using a regression package:

- Applying OLS to your pooled data will produce the correct $\hat{\beta}$ but the wrong standard errors. Your regression package does not know that there is correlation between certain elements (same individual, different time periods) *unless you tell it*.
- Example: Stata command `regress y x1 x2 x3, vce(robust)` generates the wrong standard errors.
- Some regression packages have an option to take the correlation within groups of observations into account. This is called clustering and applicable in many situations.
- Example: Data set has identifier `id` for each unit. Then you should use the Stata command `regress y x1 x2 x3, vce(cluster id)`

Outline

Strict exogeneity assumption

Pooled OLS is just the start. Let's take more advantage of the panel structure.

Assumption FE.1:

- $\mathrm{E}(u_{it}|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT},c_i)=0$ for all $t=1,\ldots,T$

Discussion:

- This is the strict exogeneity assumption as discussed above.
- However, the correlation between c_i and any $\mathbf{x}_{it}$, $t=1,\ldots,T$ is left unrestricted. Hence, the "omitted variable problem" used to motivate panel analysis can be handled here.

Within transformation

Model:

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}$$

Time average (bars denote time averages like $\bar{y}_i = T^{-1}\sum_{t=1}^{T} y_{it}$):

$$\bar{y}_i = \bar{\mathbf{x}}_i\boldsymbol{\beta} + c_i + \bar{u}_i$$

Within transformation (subtract individual time averages) wipes out the c_i:

$$y_{it} - \bar{y}_i = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)\boldsymbol{\beta} + u_{it} - \bar{u}_i$$

Defining $\ddot{y}_{it} = y_{it} - \bar{y}_i$ etc yields the within-transformed equation

$$\ddot{y}_{it} = \ddot{\mathbf{x}}_{it}\boldsymbol{\beta} + \ddot{u}_{it}$$

Discussion

- The within transformation wipes out all time-invariant regressors. To circumvent zero columns in the within-transformed regressor matrix, do not include time-invariant regressors such as an intercept.
- This is the price we have to pay for the weaker assumptions compared to the RE estimator: parameters for time-invariant regressors are not identified.
- The within transformed equation can be estimated by pooled OLS as discussed below.
- Interpretation of parameters is based on the original *structural equation* $y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}$.

The FE or within estimator

The FE or within estimator applies pooled OLS to

$$\ddot{y}_{it} = \ddot{\mathbf{x}}_{it}\boldsymbol{\beta} + \ddot{u}_{it}.$$

Stacking all observations $1, \ldots, T$ of one individual into $\ddot{\mathbf{y}}_i$, $\ddot{\mathbf{X}}_i$, and $\ddot{\mathbf{u}}_i$ yields

$$\ddot{\mathbf{y}}_i = \ddot{\mathbf{X}}_i\boldsymbol{\beta} + \ddot{\mathbf{u}}_i.$$

Hence, the estimator is

$$\hat{\boldsymbol{\beta}}_{FE} = \left(\sum_{i=1}^{N} \ddot{\mathbf{X}}_i'\ddot{\mathbf{X}}_i\right)^{-1} \sum_{i=1}^{N} \ddot{\mathbf{X}}_i'\ddot{\mathbf{y}}_i = \left(\sum_{i=1}^{N}\sum_{t=1}^{T} \ddot{\mathbf{x}}_{it}'\ddot{\mathbf{x}}_{it}\right)^{-1} \sum_{i=1}^{N}\sum_{t=1}^{T} \ddot{\mathbf{x}}_{it}'\ddot{y}_{it}.$$

To guarantee invertibility, we add assumption FE.2: rank $\mathrm{E}(\ddot{\mathbf{X}}_i'\ddot{\mathbf{X}}_i) = K$

Consistency

Is $\hat{\boldsymbol{\beta}}_{FE}$ unbiased and consistent?

A sufficient condition is strict exogeneity of the transformed regressors:

$$\mathrm{E}(\ddot{u}_{it}|\ddot{\mathbf{x}}_{i1},\ldots,\ddot{\mathbf{x}}_{iT}) = 0.$$

Since each $\ddot{\mathbf{x}}_{it}$ is a function of $\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT}$, this is implied by

$$\mathrm{E}(\ddot{u}_{it}|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT}) = \mathrm{E}(u_{it}|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT}) - \mathrm{E}(\bar{u}_{i}|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT}) = 0$$

This is satisfied if the strict exogeneity assumption FE.1 holds because it implies

$$\mathrm{E}(u_{it}|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT}) = 0$$

and

$$\mathrm{E}(\bar{u}_{i}|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT}) = T^{-1}\sum_{s=1}^{T}\mathrm{E}(u_{is}|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT}) = 0$$

Some matrix algebra for the FE model $\star$

Here is the matrix transformation ("within transformation") that turns $\mathbf{X}_i$ into $\ddot{\mathbf{X}}_i$:

$$\ddot{\mathbf{X}}_i = \mathbf{Q}_T \mathbf{X}_i = (\mathbf{I}_T - \mathbf{J}_T)\mathbf{X}_i,$$

where the time-demeaning matrix is defined as

$$\mathbf{Q}_T = \mathbf{I}_T - \mathbf{J}_T,$$

$\mathbf{I}_T$ is a $T \times T$ identity matrix and $\mathbf{J}_T$ is a $T \times T$ projection matrix on a column of ones:

$$\mathbf{J}_T = \boldsymbol{\iota}_T(\boldsymbol{\iota}_T'\boldsymbol{\iota}_T)^{-1}\boldsymbol{\iota}_T' = \boldsymbol{\iota}_T\boldsymbol{\iota}_T'/T = T^{-1}\begin{bmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{bmatrix}.$$

Hence, $\mathbf{J}_T$ is a $T \times T$ of ones divided by T.

More details ⋆

Applying the $\mathbf{J}_T$-projection yields time averages:

$$\mathbf{J}_T\mathbf{X}_i = T^{-1}\begin{bmatrix}1 & \dots & 1\\ \vdots & \ddots & \vdots\\ 1 & \dots & 1\end{bmatrix}\begin{bmatrix}x_{i,11} & \dots & x_{i,K1}\\ \vdots & \ddots & \vdots\\ x_{i,1T} & \dots & x_{i,KT}\end{bmatrix} = \begin{bmatrix}\bar{x}_{i,1} & \dots & \bar{x}_{i,K}\\ \vdots & & \vdots\\ \bar{x}_{i,1} & \dots & \bar{x}_{i,K}\end{bmatrix} = \boldsymbol{\iota}_T\bar{\mathbf{x}}$$

Hence, the within-transformation yields the deviation from time averages:

$$\begin{aligned}(\mathbf{I}_T - \mathbf{J}_T)\mathbf{X}_i &= \begin{bmatrix}x_{i,11} & \dots & x_{i,K1}\\ \vdots & \ddots & \vdots\\ x_{i,1T} & \dots & x_{i,KT}\end{bmatrix} - \begin{bmatrix}\bar{x}_{i,1} & \dots & \bar{x}_{i,K}\\ \vdots & & \vdots\\ \bar{x}_{i,1} & \dots & \bar{x}_{i,K}\end{bmatrix}\\ &= \begin{bmatrix}\ddot{x}_{i,11} & \dots & \ddot{x}_{i,K1}\\ \vdots & \ddots & \vdots\\ \ddot{x}_{i,1T} & \dots & \ddot{x}_{i,KT}\end{bmatrix} = \ddot{\mathbf{X}}_i\end{aligned}$$

... which helps ⋆

Note that $\mathbf{Q}_T$ is a symmetric, idempotent $T \times T$ matrix of rank $T-1$, $\mathbf{Q}_T\mathbf{Q}_T' = \mathbf{Q}_T\mathbf{Q}_T = \mathbf{Q}_T$.

This helps us to re-write the FE estimator as a direct function of the $\mathbf{u}_i$. First note that (like always) we can represent the estimator as

$$\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta} = \left(\sum_{i=1}^{N} \ddot{\mathbf{X}}_i'\ddot{\mathbf{X}}_i\right)^{-1} \sum_{i=1}^{N} \ddot{\mathbf{X}}_i'\ddot{\mathbf{u}}_i$$

Next observe that

$$\ddot{\mathbf{X}}_i'\ddot{\mathbf{u}}_i = \mathbf{X}_i'\mathbf{Q}_T'\mathbf{Q}_T\mathbf{u}_i = \mathbf{X}_i'\mathbf{Q}_T\mathbf{u}_i = \ddot{\mathbf{X}}_i'\mathbf{u}_i.$$

Substitute this into the above expression:

$$\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta} = \left(\sum_{i=1}^{N} \ddot{\mathbf{X}}_i'\ddot{\mathbf{X}}_i\right)^{-1} \sum_{i=1}^{N} \ddot{\mathbf{X}}_i'\mathbf{u}_i.$$

Asymptotic distribution of the FE estimator

Multiplying the result of the previous page by $\sqrt{N}$ yields

$$\sqrt{N}(\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}) = \left(N^{-1} \sum_{i=1}^{N} \ddot{\mathbf{X}}_i' \ddot{\mathbf{X}}_i \right)^{-1} N^{-1/2} \sum_{i=1}^{N} \ddot{\mathbf{X}}_i' \mathbf{u}_i$$

This structure is equivalent to system OLS with regressor matrix $\ddot{\mathbf{X}}_i$ and disturbance vector $\mathbf{u}_i$. Hence,

$$\sqrt{N} \left(\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta} \right) \xrightarrow{\text{d}} \mathcal{N} \left(\mathbf{0}, \mathbf{A}^{-1} \mathbf{B} \mathbf{A}^{-1} \right),$$

where

$$\mathbf{A} \equiv \mathrm{E}(\ddot{\mathbf{X}}_i' \ddot{\mathbf{X}}_i) = \mathrm{E}(\mathbf{X}_i' \mathbf{Q}_T \mathbf{X}_i) \quad \text{and} \quad \mathbf{B} \equiv \mathrm{E} \left(\ddot{\mathbf{X}}_i' \mathbf{u}_i \mathbf{u}_i' \ddot{\mathbf{X}}_i \right).$$

Estimating the robust variance matrix

To estimate $\mathbf{A}$ and $\mathbf{B}$, use sample equivalents:

$$\hat{\mathbf{A}} \equiv N^{-1} \sum_{i=1}^{N} \ddot{\mathbf{X}}_i' \ddot{\mathbf{X}}_i \quad \text{and} \quad \hat{\mathbf{B}} \equiv N^{-1} \sum_{i=1}^{N} \ddot{\mathbf{X}}_i' \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i' \ddot{\mathbf{X}}_i,$$

where $\hat{\mathbf{u}}_i \equiv \ddot{\mathbf{y}}_i - \ddot{\mathbf{X}}_i \hat{\boldsymbol{\beta}}_{FE}$ is the residual vector of the FE estimator.

This is sometimes called the Arellano-White variance estimator for the FE model.

Note that this estimator is not only robust to heteroskedasticity. It is also robust to autocorrelation within individuals, i.e., it allows

$$\mathrm{E}(u_{it} u_{is}) \neq 0 \quad \text{for all } s, t = 1, \ldots, T.$$

The classical FE variance estimator

The classical FE estimator assumes homoscedasticity and lack of serial correlation.

Assumption FE.3: $\mathrm{E}(\mathbf{u}_i\mathbf{u}_i'|\mathbf{x}_{i1},\ldots,\mathbf{x}_{iT},c_i) = \mathrm{E}(\mathbf{u}_i\mathbf{u}_i') = \sigma_u^2\mathbf{I}_T$

Discussion:

- The first part means that conditional and unconditional variance matrix are equal. This rules out heteroskedasticity.
- The second part implies constant variances over time and lack of serial correlation:

$$\mathrm{E}(u_{it}^2) = \sigma_u^2 \text{ and } \mathrm{E}(u_{it}u_{is}) = 0.$$

- If this assumption is correct, the variance estimator simplifies greatly. The simplified estimator is more efficient than the robust estimator presented above but inconsistent if the assumptions fails.

... continued

Under FE.3 we obtain by the LIE

$$\begin{aligned}\mathbf{B} &= \mathrm{E}\left(\ddot{\mathbf{X}}_i'\mathbf{u}_i\mathbf{u}_i'\ddot{\mathbf{X}}_i\right) = \mathrm{E}\left(\ddot{\mathbf{X}}_i'\sigma_u^2\mathbf{I}_T\ddot{\mathbf{X}}_i\right) = \sigma_u^2\,\mathrm{E}(\ddot{\mathbf{X}}_i'\ddot{\mathbf{X}}_i) = \sigma_u^2\,\mathrm{E}(\mathbf{X}_i'\mathbf{Q}_T\mathbf{X}_i)\\ &= \sigma_u^2\mathbf{A}.\end{aligned}$$

Thus the asymptotic variance matrix simplifies to

$$\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1} = \mathbf{A}^{-1}\sigma_u^2\mathbf{A}\mathbf{A}^{-1} = \sigma_u^2\mathbf{A}^{-1}.$$

It can be estimated by

$$\hat{\mathbf{A}} \equiv N^{-1}\sum_{i=1}^{N}\ddot{\mathbf{X}}_i'\ddot{\mathbf{X}}_i$$

and the consistent estimator

$$\hat{\sigma}_u^2 = \frac{1}{N(T-1)-K}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{u}_{it}^2.$$

Why division by $N(T-1)-K$ and not $NT-K$?

Why division by $N(T-1)-K$ and not $NT-K$?

Recall that $\hat{u}_{it} \equiv \ddot{y}_{it} - \ddot{\mathbf{x}}_{it}\hat{\boldsymbol{\beta}}_{FE}$ is the residual vector of the FE estimator and thus the sample equivalent of $\ddot{u}_{it}$.

In population, we obtain

$$\mathrm{E}(\ddot{u}_{it}^2) = \mathrm{E}[(u_{it} - \bar{u}_i)^2] = \mathrm{E}(u_{it}^2) + \mathrm{E}(\bar{u}_i^2) - 2\,\mathrm{E}(u_{it}\bar{u}_i).$$

Under FE.3,

$$\mathrm{E}(u_{it}^2) = \sigma_u^2$$

$$\mathrm{E}(\bar{u}_i^2) = T^{-2}\,\mathrm{E}\left[\left(\sum_{t=1}^{T} u_{it}\right)^2\right] = T^{-2}\sum_{t=1}^{T}\mathrm{E}(u_{it}^2) = T^{-2}\sum_{t=1}^{T}\sigma_u^2 = T^{-1}\sigma_u^2$$

$$\mathrm{E}(u_{it}\bar{u}_i) = T^{-1}\,\mathrm{E}\left(u_{it}\sum_{s=1}^{T} u_{is}\right) = T^{-1}\,\mathrm{E}\left(u_{it}^2\right) = T^{-1}\sigma_u^2$$

Hence,

$$\mathrm{E}(\ddot{u}_{it}^2) = \sigma_u^2 + T^{-1}\sigma_u^2 - 2T^{-1}\sigma_u^2 = (1 - 1/T)\sigma_u^2.$$

... because

Due to $\mathrm{E}(\ddot{u}_{it}^2) = \sigma_u^2(1 - 1/T)$ we have

$$\mathrm{E}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{u}_{it}^2\right) = \sum_{i=1}^{N}\sum_{t=1}^{T}\mathrm{E}(\ddot{u}_{it}^2) = NT(1-1/T)\sigma_u^2 = N(T-1)\sigma_u^2$$

and thus

$$\mathrm{E}\left(\frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=1}^{T}\ddot{u}_{it}^2\right) = \sigma_u^2.$$

Accounting for the loss in degrees of freedom we therefore use the sample equivalent

$$\hat{\sigma}_u^2 = \frac{1}{N(T-1)-K}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{u}_{it}^2.$$

Note that a regression package that applies pooled OLS on within-transformed variables will automatically divide by $NT - K$. Therefore, you should use specialized panel commands.

Efficiency of the FE estimator ⋆

Use the result

$$\ddot{\mathbf{X}}_i'\ddot{\mathbf{y}}_i = \mathbf{X}_i'\mathbf{Q}_T'\mathbf{Q}_T\mathbf{y}_i = \mathbf{X}_i'\mathbf{Q}_T\mathbf{y}_i = \ddot{\mathbf{X}}_i'\mathbf{y}_i$$

to rewrite the FE estimator:

$$\hat{\boldsymbol{\beta}}_{FE} = \left(\sum_{i=1}^{N} \ddot{\mathbf{X}}_i'\ddot{\mathbf{X}}_i\right)^{-1} \sum_{i=1}^{N} \ddot{\mathbf{X}}_i'\ddot{\mathbf{y}}_i = \left(\sum_{i=1}^{N} \ddot{\mathbf{X}}_i'\ddot{\mathbf{X}}_i\right)^{-1} \sum_{i=1}^{N} \ddot{\mathbf{X}}_i'\mathbf{y}_i,$$

Hence, the FE estimator is equivalent to the SURE/SOLS estimator applied to the model

$$\mathbf{y}_i = \ddot{\mathbf{X}}_i\boldsymbol{\beta} + \mathbf{u}_i.$$

Assumptions FE.1 to FE.3 say that the regressors of this model are strictly exogenous and the (conditional and unconditional) variance matrix of the disturbances is $\sigma_u^2\mathbf{I}_T$.

Under these assumptions Gauss-Markov applies.

FE estimator for policy analysis

Consider the model

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + v_{it} = \mathbf{z}_{it}\boldsymbol{\gamma} + \delta w_{it} + v_{it},$$

where w_{it} is a policy variable, $\mathbf{z}_{it}$ contain control variables, and v_{it} may contain an individual effect.

The FE estimator of δ is consistent if

$$\mathrm{E}[\mathbf{x}'_{it}(v_{it} - \bar{v}_i)] = 0.$$

Discussion:

- Consistency requires that w_{it} be uncorrelated with the deviation of v_{it} from its average; correlation of w_{it} with $\bar{v}_i$ is allowed.
- Assume w_{it} measures program participation. I.e., program participation can be systematically related to the persistent component in the error v_{it}. This can be helpful in situations we have to suspect certain kinds of self selection etc.
- Obviously, variation in w_{it} over time is required for at least some i.

Implementation in Stata

Example: Data set has identifier for each individual denoted `id` and for each time period denoted `year`.

You first have to tell Stata that you have panel data:

```
xtset id year
```

The FE estimator with classical (nonrobust) variance matrix is computed using

```
xtreg y x1 x2 x3, fe
```

The FE estimator with robust Arellano-White variance matrix is computed using

```
xtreg y x1 x2 x3, fe vce(robust)
```

… and a small remark

Note that Stata's FE regression results include an intercept even though the within transformation wipes out any time-invariant regressor.

Stata estimates the intercept as

$$\hat{\alpha}_{FE} = \bar{y} - \bar{\mathbf{x}}\hat{\boldsymbol{\beta}}_{FE}$$

where $\bar{y}$ and $\bar{\mathbf{x}}$ are the sample means over all N and T.

The individual effects are estimated as

$$\hat{c}_i = \bar{y}_i - \hat{\alpha}_{FE} - \bar{\mathbf{x}}_i\hat{\boldsymbol{\beta}}_{FE}.$$

Note, however, that the c_i's cannot be estimated consistently. (Can you imagine why?)

Example: Effects of job training grants on scrap rates

Example 10.5 taken from Wooldridge's textbook

Question: How do job training grants affect scrap rates?

Sample: 54 firms reported scrap rates for 1987, 1988, and 1989. Some received a grant in one of the years 1988 or 1989 to initiate a training program.

Analysis: Regression of log scrap rates on yearly dummies, grant dummy ("grant") and lagged grant dummy ("grant_1"). (We leave out the union membership dummy as it is time-invariant and include it later on when we use the RE estimator.)

Stata output – descriptive statistics

Load data: `use "jtrain1.dta", clear`

Set panel: `xtset fcode year`

```
. xtsum lscrap grant grant_1
```

Variable		Mean	Std. Dev.	Min	Max	Observations
lscrap	overall	.3936814	1.486471	-4.60517	3.401197	N = 162
	between		1.426381	-3.00934	3.205269	n = 54
	within		.447558	-2.031318	2.121575	T = 3
grant	overall	.1401274	.3474882	0	1	N = 471
	between		.1650666	0	.3333333	n = 157
	within		.305969	-.1932059	.8067941	T = 3
grant_1	overall	.0764331	.2659724	0	1	N = 471
	between		.1405758	0	.3333333	n = 157
	within		.2259731	-.2569002	.7430998	T = 3

Definitions:

- overall $= x_{it}$
- between $= \bar{x}_i$
- within $= x_{it} - \bar{x}_i + \bar{\bar{x}}$

Stata output – FE estimation

```
. xtreg lscrap d88 d89 grant grant_1, fe vce(robust)

Fixed-effects (within) regression               Number of obs      =       162
Group variable: fcode                           Number of groups   =        54

R-sq:  within  = 0.2010                         Obs per group: min =         3
       between = 0.0079                                        avg =       3.0
       overall = 0.0068                                        max =         3

                                                F(4,53)            =      7.07
corr(u_i, Xb)  = -0.0714                        Prob > F           =    0.0001

                                 (Std. Err. adjusted for 54 clusters in fcode)
------------------------------------------------------------------------------
             |               Robust
      lscrap |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         d88 |  -.0802157   .0978408    -0.82   0.416    -.2764594    .1160281
         d89 |  -.2472028   .1967819    -1.26   0.215    -.6418973    .1474917
       grant |  -.2523149   .1434399    -1.76   0.084    -.5400188     .035389
     grant_1 |  -.4215895   .2824604    -1.49   0.141    -.9881333    .1449543
       _cons |   .5974341   .0638746     9.35   0.000     .4693177    .7255504
-------------+----------------------------------------------------------------
     sigma_u |   1.438982
     sigma_e |  .49774421
         rho |  .89313867   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Translation

Notes:

- R-sq within: squared correlation between $(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)\hat{\boldsymbol{\beta}}_{FE}$ and $y_{it} - \bar{y}_i$.
- R-sq between: squared correlation between $\bar{\mathbf{x}}_i\hat{\boldsymbol{\beta}}_{FE}$ and $\bar{y}_i$.
- R-sq overall: squared correlation between $\mathbf{x}_{it}\hat{\boldsymbol{\beta}}_{FE}$ and y_{it}.
- sigma_u: square root of $\mathrm{Var}(c_i) = \sigma_c^2$
- sigma_e: square root of $\mathrm{Var}(u_{it}) = \sigma_u^2$
- rho: variance share $\mathrm{Var}(c_i)/\mathrm{Var}(u_{it}) = \sigma_c^2/\sigma_u^2$
- corr(u_i, Xb): correlation between $\hat{c}_i$ and $\bar{\mathbf{x}}_{it}\hat{\boldsymbol{\beta}}_{FE}$.

Outline

Coming up

The random effects estimator