Lecture 15
Panel Data Models

(For private use, not to be posted/shared online).

Panel Data Sets

• A panel, or longitudinal, data set is one where there are repeated observations on the same units: individuals, households, firms, countries, or any set of entities that remain stable through time.

• Repeated observations create a potentially very large panel data sets. With N units and T time periods ⇒ Number of observations: NT.

– Advantage: Large sample! Great for estimation.

– Disadvantage: Dependence! Observations are, very likely, not independent.

• Modeling the potential dependence creates different models.

Panel Data Sets

• The National Longitudinal Survey (NLS) of Youth is an example. The same respondents were interviewed every year from 1979 to 1994. Since 1994 they have been interviewed every two years.

• Panel data allows us a researcher to study cross section effects –i.e., along *N*, variation across the firms- & time series effects –i.e., along *T*, variation across time.

$$
\begin{array}{cc}
 & \text{Cross section} \\
\begin{array}{c}\text{Time} \\ \text{series}\end{array} & \begin{bmatrix} y_{11} & y_{21} & \ldots & y_{i1} & \cdots & y_{N1} \\ y_{12} & y_{22} & \ldots & y_{i2} & \cdots & y_{N2} \\ \ldots & \ldots & \ldots & \vdots & & \ldots \\ y_{1t} & y_{2t} & \ldots & y_{it} & \cdots & y_{Nt} \\ \ldots & \ldots & & \ldots & \ldots & \ldots \\ y_{1T} & y_{2T} & \ldots & y_{iT} & \cdots & y_{NT} \end{bmatrix} = [\mathbf{y}_1\ \mathbf{y}_2 \ldots \mathbf{y}_i \ldots \mathbf{y}_N]
\end{array}
$$

Panel Data Sets

• Notation:

$$
y_1 = \begin{bmatrix} y_{11} \\ y_{12} \\ \ldots \\ y_{1t} \\ \ldots \\ y_{1T} \end{bmatrix};\ldots; y_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \ldots \\ y_{it} \\ \ldots \\ y_{iT} \end{bmatrix} \quad X_1 = \begin{bmatrix} x_{11} & x_{21} & \ldots & x_{k1} \\ x_{12} & x_{22} & \ldots & x_{k2} \\ \ldots & \ldots & \ldots & \ldots \\ x_{1t} & x_{2t} & \ldots & x_{k2} \\ \ldots & \ldots & \ldots & \ldots \\ x_{1T_1} & x_{2T_1} & \ldots & x_{kT_1} \end{bmatrix}; \ldots; X_i = \begin{bmatrix} w_{11} & w_{21} & \ldots & w_{k1} \\ w_{12} & w_{22} & \ldots & x_{k2} \\ \ldots & \ldots & \ldots & \ldots \\ w_{1t} & w_{2t} & \ldots & w_{k2} \\ \ldots & \ldots & \ldots & \ldots \\ w_{1T_j} & w_{2T_j} & \ldots & w_{kT_j} \end{bmatrix}
$$

• A standard panel data set model stacks the $y_i's$ and the $\boldsymbol{x}_i's$:

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{c} + \boldsymbol{\varepsilon}$$

$\mathbf{X}$ is a $\sum_{i=1}^{N} T_i$xk matrix

$\boldsymbol{\beta}$ is a kx1 matrix

$\mathbf{c}$ is $\sum_{i=1}^{j} T_i$x1 matrix, associated with unobservable variables.

$\boldsymbol{y}$ and $\boldsymbol{\varepsilon}$ are $\sum_{i=1}^{N} T_i$x1 matrices

Balanced and Unbalanced Panels

• Notation:
$y_{i,t}$, $i = 1,\ldots, N$; $t = 1,\ldots, T_i$
• Mathematical and notational convenience:
- Balanced: NT
(that is, every unit is surveyed in every time period.)
- Unbalanced: $\sum_{i=1}^{N} T_i$
Q: Is the fixed T_i assumption ever necessary? SUR models.

• The NLS of Youth is *unbalanced* because some individuals have not been interviewed in some years. Some could not be located, some refused, and a few have died. CRSP is also *unbalanced,* some firms are listed from 1962, others started to be listed later.

Panel Data Models

- With panel data we can study different issues:
  - *Cross sectional variation* (unobservable in time series data) vs. *Time series variation* (unobservable in cross sectional data)
  - Heterogeneity (observable and unobservable individual heterogeneity)
  - Hierarchical structures (say, zip code, city and state effects)
  - Dynamics in economic behavior
  - Individual/Group effects (individual effects)
  - Time effects

Panel Data Models: Example 1 - SUR

• In Zellner's SUR formulation (no linear dependence on $y_{i,t}$) we have:

(**A1**) $y_{i,t} = \boldsymbol{x}_{i,t}'\beta_i + \varepsilon_{i,t}$ – the DGP

(**A2**) $\mathrm{E}[\boldsymbol{\varepsilon}_i|\mathbf{X}] = \mathbf{0},$

(**A3'**) $\mathrm{Var}[\varepsilon_i|\mathbf{X}] = \sigma_i^2\mathbf{I}_T = \sigma_{ii}\,\mathbf{I}_T$ – groupwise heteroscedasticity.

$\mathrm{E}[\varepsilon_{it}\varepsilon_{jt}|\mathbf{X}] = \sigma_{ij}$ – contemporaneous correlation

$\mathrm{E}[\varepsilon_{it}\varepsilon_{js}|\mathbf{X}] = 0$ when $t{\neq}s$

(**A4**) Rank($\mathbf{X}$) = full rank

• In (**A1**) – (**A4**), we have the a GR model with heteroscedasticity. OLS in each equation is OK, but not efficient. GLS is efficient.

• We are not taking advantage of pooling –i.e., using *NT* observations!

• Use LR or *F* tests to check if pooling (aggregation) can be done.

Panel Data Models: Example 2 - Pooling

• Assumptions

(A1) $y_{i,t} = \boldsymbol{x}_{i,t}'\beta + \boldsymbol{z}_i'\gamma + \varepsilon_{i,t}$ – the DGP

$i = 1, 2, ..., N$ - we have N *individual, groups* or *firms.*

$t = 1, 2, ..., T_i$ - usually, $N >> T_i$.

(**A2**) $\mathrm{E}[\boldsymbol{\varepsilon}_i|\mathbf{X}, \boldsymbol{z}] = \mathbf{0},$ – **X** and $\boldsymbol{z}$: exogenous

(A3) $\mathrm{Var}[\boldsymbol{\varepsilon}_i|\mathbf{X}, \boldsymbol{z}] = \sigma^2\mathbf{I}.$ – Heteroscedasticity can be allowed.

(**A4**) Rank($\mathbf{X}$) = full rank

• We think of **X** as a vector of observed characteristics. For example, firm size, Market-to-book, Z-score, R&D expenditures, etc.

• We think of $\boldsymbol{z}$ as a vector of unobserved characteristics *(individual effects*). For example, quality of management, growth opportunities, etc.

Panel Data Models: Basic Model

• The DGP (**A1**) is linear:

$$y_{i,t} = \beta_1 + \sum_{j=1}^{k} \beta_j\, \boldsymbol{x}_{ij,t} + \sum_{p=1}^{s} \gamma_p\, \boldsymbol{z}_{ip} + \delta\, t + \varepsilon_{i,t}$$

– Indices:

- i*:* individuals –i.e., the unit of observation–,
- t*:* time period,
- j*:* observed explanatory variables,
- p *:* unobserved explanatory variables.

– Time trend t allows for a shift of the intercept over time, capturing time effects –technological change, regulations, etc. But, if the implicit assumption of a constant rate of change is strong (=δ), we use a set of dummy variables, one for each time period except reference period.

25

Panel Data Models: Basic Model

– $\boldsymbol{X}$: The variables of interest –β is the vector of parameter of interest.

– $\boldsymbol{Z}$: The variables responsible for unobserved heterogeneity (& dependence on the y_i's). Usually, a *nuisance* component of the model.

• The Z_p variables are unobserved: Impossible to obtain information about the $\sum_{p=1}^{s} \gamma_p\, \boldsymbol{z}_{ip}$ component of the model. We define a term c_i, *the unobserved effect*, representing the joint impact of the Z_p variables on y_i – like an index of unobservables for individual i:

$$c_i = \sum_{p=1}^{s} \gamma_p\, \boldsymbol{z}_{ip}$$

• We can rewrite the regression model as:

$$y_{i,t} = \beta_1 + \sum_{j=1}^{k} \beta_j\, \boldsymbol{x}_{ij,t} + c_i + \delta\, t + \varepsilon_{i,t}$$

31

Panel Data Models: Basic Model

$$y_{i,t} = \beta_1 + \sum_{j=1}^{k} \beta_j \, \boldsymbol{x}_{ij,t} + c_i + \delta\, t + \varepsilon_{i,t}$$

Note: If the X_j's are so comprehensive that they capture all relevant characteristics of individual i, c_i can be dropped and, then, pooled OLS may be used. But, this is situation is very unlikely.

• In general, dropping c_i leads to missing variables problem: bias!

• We usually think of c_i as *contemporaneously exogenous* to the conditional error. That is, $E[\varepsilon_{it} \,|\, c_i] = 0$, $t = 1,..., T$

A stronger assumption: *Strict exogeneity* can also be imposed. Then,

$E[\varepsilon_{it} \,|\, \boldsymbol{x}_{i,1}, \boldsymbol{x}_{i,2},..., \boldsymbol{x}_{i,T}, c_i] = 0$, $\quad t = 1,..., T$

30

Panel Data Models: Basic Model

• Strict exogeneity conditions on the whole history of $\boldsymbol{x}_i$. Under this assumption:

$$E[y_{i,t}\varepsilon_{it} | \boldsymbol{x}_{i,t}, c_i] = \beta_1 + \sum_{j=1}^{k} \beta_j \, \boldsymbol{x}_{ij,t} + c_i + \delta\, t$$

⇒ The β_j's are partial effects holding c_i constant.

• Violations of strict exogeneity are not rare. For example, if $\boldsymbol{x}_{i,t}$ contains lagged dependent variables or if changes in ε_{it} affect $\boldsymbol{x}_{i,t+1}$ (a "feedback" effect).

• But to estimate $\boldsymbol{\beta}$ we still need to say something about the relation between $\boldsymbol{x}_{i,t}$ and c_i. Different assumptions will give rise to different models.

30

Panel Data Models: Types

• The basic DGP: $y_{i,t} = \boldsymbol{x}_{i,t}'\beta_i + \boldsymbol{z}_i'\gamma + \varepsilon_{i,t}$
& (**A2**)-(**A4**) apply.

Depending on how we model the heterogeneity in the panel, we have different models.

• Four Popular Models:

(1) ***Pooled (Constant Effect) Model***

$\boldsymbol{z}_i'\gamma$ is a constant. $\boldsymbol{z}_i = \alpha$ (and uncorrelated with $\boldsymbol{x}_{i,t}$!). Dependence on the $y_{i,t}$ may enter through the variance. That is, repeated observations on individual *i* are linearly independent. In this case,

$$y_{i,t} = \boldsymbol{x}_{i,t}'\beta_i + \alpha + \varepsilon_{i,t}$$

⇒ OLS estimates α and β consistently. We estimate k+1 parameters. 31

Panel Data Models: Types

(2) ***Fixed Effects Model*** (FEM)

The $\boldsymbol{z}_i$'s are correlated with $\mathbf{X}_i$ Fixed Effects:

$$\text{E}[\boldsymbol{z}_i \,|\, \boldsymbol{X}_i] = \text{g}(\boldsymbol{X}_i) = \alpha_i^*;$$

the unobservable effects are correlated with included variables –i.e., pooled OLS will be inconsistent.

Assume $\boldsymbol{z}_i'\gamma = \alpha_i$ (constant; it does not vary with t). Then,

$$y_{i,t} = \boldsymbol{x}_{i,t}'\beta_i + \alpha_i + \varepsilon_{i,t}$$

⇒ the regression line is raised/lowered by a fixed amount for each individual i (the dependence created by the repeated observations!). In econometrics terms, this is the source of the *fixed-effect*s.

⇒ We have a lot of parameters: $k + N$. We have N individual effects! OLS can be used to estimates α and β consistently.

Panel Data Models: Types

(3) ***Random Effects Model*** (REM)

The differences between individuals are random, drawn from a given distribution with constant parameters. We assume the $\boldsymbol{z}_i$'s are uncorrelated with the $\boldsymbol{X}_i$. That is,

$$E[\boldsymbol{z}_i | \boldsymbol{X}_i] = \mu \qquad \text{(if } \boldsymbol{X}_i \text{ contains a constant term, } \mu\text{=0 WLOG).}$$

Add and subtract $E[\boldsymbol{z}_i' \boldsymbol{\gamma}] = \mu^*$ from (*):

$$y_{i,t} = \boldsymbol{x}_{i,t}'\boldsymbol{\beta} + E[\boldsymbol{z}_i' \boldsymbol{\gamma}] + (\boldsymbol{z}_i' \boldsymbol{\gamma}) - E[\boldsymbol{z}_i' \boldsymbol{\gamma}] + \varepsilon_{i,t}$$
$$= \boldsymbol{x}_{i,t}'\boldsymbol{\beta} + \mu^* + u_i + \varepsilon_{i,t}$$

We have a compound ("composed") error –i.e., $u_i + \varepsilon_{i,t} = w_{i,t}$. This $w_{i,t}$ introduces contemporaneous cross-correlations across the *i* group.

⇒ OLS estimates μ and $\boldsymbol{\beta}$ consistently, but GLS will be efficient.

Panel Data Models: Types

(4) ***Random Parameters /Coefficients Model***

We introduce heterogeneity through $\boldsymbol{\beta}_i$. But, this may introduce additional *N* parameters. A solution is to model $\boldsymbol{\beta}_i$. For example,

$$y_{i,t} = \boldsymbol{x}_{i,t}'(\boldsymbol{\beta} + h_i) + \alpha_i + \varepsilon_{i,t}$$

h_i is a random vector that induces parameter variation, where h_i ~ D(0, $\sigma^2_{h_i}$). That is, we introduce heteroscedasticity.

Now, the coefficients are different for each individual. It is possible to complicate the model by making them different through time:

$$\boldsymbol{\beta}_{it} = (\boldsymbol{\beta} + h_i) + \theta_t \qquad \text{where } \theta_t \sim D(0, \sigma^2_t).$$

Estimation: GLS, MLE.

Long history: Rao (1965) and Chow (1975) worked on these models.

Compact Notation

- Compact Notation: $\boldsymbol{y}_i = \boldsymbol{X}_i\beta + \boldsymbol{c}_i + \boldsymbol{\varepsilon}_i$

  $\boldsymbol{X}_i$ is a T_ixk matrix

  β is a kx1 matrix

  $\boldsymbol{c}_i$ is a T_ix1 matrix

  $\boldsymbol{y}_i$ and $\boldsymbol{\varepsilon}_i$ are T_ix1 matrices

- Recall we stack the $\boldsymbol{y}_i$'s and $\boldsymbol{X}_i$'s: $\boldsymbol{y} = \boldsymbol{X}\,\beta + \boldsymbol{c} + \varepsilon$

  X is a $\sum_{i=1}^{N} T_i$xk matrix

  β is a kx1 matrix

  $\boldsymbol{c}, \boldsymbol{y}$ and ε are $\sum_{i=1}^{N} T_i$x1 matrices

Or $\boldsymbol{y}$ = **X*** β* + ε, with **X** * = [**X** ι] - $\sum_{i=1}^{N} T_i$x$(k+1)$ matrix.

β* = [β $\boldsymbol{c}$]' - $(k+1)$x1 matrix

Assumptions for Asymptotics (Greene)

- Convergence of moments involving cross section $\boldsymbol{X}_i$.

Usually, we assume N increasing, T or T_i assumed fixed.

  – "*Fixed-T* asymptotics" (see Greene)

  – Time series characteristics are not relevant (may be nonstationary)

  – If T is also growing, need to treat as multivariate time series.

- Rank of matrices. **X** must have full column rank.

  ⇒ $\boldsymbol{X}_i$ may not, if $T_i < k$.

- Strict exogeneity and dynamics. If $\boldsymbol{x}_{i,t}$ contains $y_{i,t}$, then $\boldsymbol{x}_{i,t}$ cannot be strictly exogenous. $\boldsymbol{x}_{i,t}$ will be correlated with the unobservables in period t – 1. Inconsistent OLS estimates! (To be revisited later.)

Panel Data Models: (A3') - No Homoscedasticity

• We can relax assumption (**A3**). The new DGP model:

$\boldsymbol{y} = \mathbf{X}^* \boldsymbol{\beta}^* + \boldsymbol{\varepsilon}$, with $\mathbf{X}^* = [\mathbf{X}\ \iota]$ – $\sum_{i=1}^{N} T_i \text{x}(k+1)$ matrix.

$\boldsymbol{\beta}^* = [\boldsymbol{\beta}\ \text{c}]'$ – $(k+1)$x1 matrix

Now, we assume (**A3'**) $E[\boldsymbol{\varepsilon}\,\boldsymbol{\varepsilon}'|\mathbf{X}] = \boldsymbol{\Sigma} \neq \sigma^2\, \mathbf{I}_{\sum_{i=1}^{N} T_i}$

• Potentially, a lot of different elements in $E[\boldsymbol{\varepsilon}\,\boldsymbol{\varepsilon}'|\mathbf{X}]$ in a panel:

- *Individual heteroscedasticity*. Usual groupwise heteroscedasticity.
- *Autocorrelation (Individual/group/firm) effects*. Errors have arbitrary correlation across time for a particular individual *i*:
- *Temporal correlation (Time) effects*. Errors have arbitrary correlation across individuals at a moment in time (SUR-type correlation).
- *Persistent common shocks*: Errors have some correlation between different firms in different time periods (but, these shocks are assumed to die out over time, and may be ignored after *L* periods).

Panel Data Models: (A3') – Error Structures

• To understand the different elements in $\boldsymbol{\Sigma}$, consider the following DGP for the errors, $\varepsilon_{i,t}$'s:

$\varepsilon_{i,t} = \boldsymbol{\theta}_i'\mathbf{f}_t + \eta_{i,t}$, $\mathbf{f}_t \sim D(0, \sigma_f^2)$

& $\eta_{i,t} = \phi\, \eta_{i,t-1} + \varsigma_{i,t}$, $\varsigma_{i,t} \sim D(0, \sigma_{\varsigma i}^2)$

$\mathbf{f}_t$: vector of random factors common to all individuals/groups/firms.

$\boldsymbol{\theta}_i$: vector of factor loadings, specific to individual i.

$\varsigma_{i,t}$: random shocks to individual i, uncorrelated across both i and t.

$\eta_{i,t}$: random shocks to i. This generates autocorrelation effects in i.

• $\boldsymbol{\theta}_i'\mathbf{f}_t$ generates both contemporaneous (SUR) and time-varying cross-correlations between i and j. (Autocorrelations die out after *L* periods.)

- If $\mathbf{f}_t$ is uncorrelated across t ⇒ only contemporaneous (SUR) effects.
- If $\mathbf{f}_t$ is persistent in t ⇒ both SUR and persistent common effects.

Panel Data Models: (A3') – Error Structures

• Different forms for E[$\boldsymbol{\varepsilon}\,\boldsymbol{\varepsilon}'$|**X**]:

- *Individual heteroscedasticity.* E[ε_i^2|**X**] = σ_i^2

⇒ standard groupwise heteroscedasticity *driven by* $\varsigma_{i,t}$.

- *Autocorrelation (Individual) effects*: E[$\varepsilon_{i,t}\varepsilon_{i,s}$|**X**] ≠ 0 ($t \neq s$)

⇒ auto-/time-correlation for errors, $\varepsilon_{i,t}$ *driven by* $\eta_{i,t}$.

- *Temporal correlation effects:* E[$\varepsilon_{i,t}\varepsilon_{j,t}$|**X**] ≠ 0 ($i \neq j$)

⇒ contemporary cross-correlation for errors *driven by* $\mathbf{f}_t$.

- *Persistent common shocks*: E[$\varepsilon_{i,t}\varepsilon_{j,s}$|**X**] ≠ 0 ($i \neq j$) and $|t - s| < L$

⇒ time-varying cross-correlation for errors *driven by* $\mathbf{f}_t$.

• Remark: Heteroscedasticity points to GLS efficient estimation, but, as before, for consistent inferences we can use OLS with (adjusted for panels) White or NW SE's.

Panel Data Models: (A3') – Clustered SE

• For consistent inferences, we can use OLS with White or NW SE's:

- White SE's adjust only for heteroscedasticity:

$$\mathbf{S_0} = (1/T)\sum_{i=1}^{T} e_i^2 \boldsymbol{x}_i\, \boldsymbol{x}_i'$$

- NW SE's adjust for heteroscedasticity and autocorrelation:

$$\mathbf{S}_T = \mathbf{S}_0 + (1/T)\sum_{l=1}^{L} k(l) \sum_{t=l+1}^{T} (\mathbf{x}_{t-l} e_{t-l} e_t \mathbf{x}_t' + \mathbf{x}_t e_t e_{t-l} \mathbf{x}_{t-l}')$$

• But, cross-sectional (SUR) or "spatial" dependencies are ignored. If present, the White's or NW's HAC need to be adjusted.

• Simple intuition: Repeating a dataset 10 times should not increase the precision of parameter estimates. However, the *i.i.d.* assumption will do this: Now, we divide by NT, not T or N.

⇒ We cannot ignore the dependence in the data.

Obvious solution: Aggregate the repeated data –i.e., aggregate in groups

Panel Data Models: (A3') – Clustered SE

• In general, the observations are not identical, but correlated within a *cluster* –i.e., a group that share certain characteristic. Depending on the data, the clusters may correspond to firms, industries, years, cities, etc.

• Simple idea: Aggregate over the clusters. The key is how (& when) to cluster.

Canonical example: We want to study the effect of class size on 1st graders' grades, the *unobservables* of 1st graders belonging to the same classroom will be correlated (say, teachers' quality, recess routines) while will not be correlated with 1st graders in far away classrooms. Then, we can cluster by school/teacher.

• In finance, it is reasonable to expect that shocks to firms in the same industry are not independent. Then, we can cluster by industry.

Panel Data Models: (A3') – Clustered SE

• We assume the existence of G disjoint clusters. Within the cluster, any pattern of dependence and/or heteroscedasticity is allowed; but, there is independence across the G clusters.

• Under the above assumption, it is easy to compute a counterpart to White or NW SE in panels. These SE are usually referred as **PCSE** – panel clustered SE- or, more general, just **clustered SE** or **Liang-Zeger**, **LZ**, **SE**.

Panel Data Models: (A3') – Clustered SE

• We remove the dependence by assuming correlation within a *cluster,* but independence across clusters. That is, we think of the data as

$$y_{ig} = \boldsymbol{x}_{ig}{}'\,\beta + c_{ig} + \delta\, t + \varepsilon_{ig} \qquad g = 1, ..., G.$$

$$E[\varepsilon_{ig}\varepsilon_{jg'}|\boldsymbol{X}_g] = 0 \qquad g \neq g'$$

$$= \sigma_{ijg} \qquad g = g'$$

Or stacking the data by cluster:

$$\boldsymbol{y}_g = \boldsymbol{x}_g{}'\,\beta + c_g + \boldsymbol{\varepsilon}_g \qquad g = 1, ..., G.$$

• Let $\boldsymbol{w}_g = \boldsymbol{x}_g{}'\,\boldsymbol{\varepsilon}_g$. Then, let E[$\boldsymbol{w}_g\boldsymbol{w}_g{}'$] = $\boldsymbol{W}_g$ be the kxk "meat" for the g cluster. We make inferences with OLS using the sandwich.

$$\text{Var}_T[\mathbf{b}|\mathbf{X}] = (\mathbf{X'X})^{-1}\,(\textstyle\sum_{g=1}^{G}\boldsymbol{W}_g)\,(\mathbf{X'X})^{-1}$$

$\boldsymbol{W}_g$ needs to be estimated. There are different ways to do it.

Panel Data Models: (A3') – Clustered SE

• $\boldsymbol{W}_g$ needs to be estimated. There are different ways to do it. But, a natural estimator is to just replace $\boldsymbol{w}_g = \boldsymbol{x}_g{}'\boldsymbol{\varepsilon}_g$ by $\widehat{\boldsymbol{w}}_g = \boldsymbol{x}_g{}'\boldsymbol{e}_g$

$$\text{Est Var}[\mathbf{b}|\mathbf{X}] = (\mathbf{X'X})^{-1}\,(\textstyle\sum_{g=1}^{G}\widehat{\boldsymbol{w}}_g\widehat{\boldsymbol{w}}_g{}')\,(\mathbf{X'X})^{-1}$$

Corrections using degrees of freedom, using transformed residuals, etc., are common. We think of these clustered SE structures as *clustered White SE*.

• Similarly, if we allow for autocorrelation in the structure of $\boldsymbol{w}_g$, then we can also have clustered NW SE.

• Driscoll and Kraay (1998) provide an easy extension to estimate robust NW SE's in panels with cross-sectional dependencies: Average the $\boldsymbol{x}_t\, e_t$ over the clusters we suspect cause dependence.

Panel Data Models: (A3') – PCSE

• Recall that, within a cluster, we assume that the correlations within a cluster are the same for different observations.

We define $h_{it}(\mathbf{b}) = \boldsymbol{x}_{i,t}\ e_{i,t}$, which we average over the cluster N_g:

$$h_t(\mathbf{b}) = \sum_{i=1}^{N_g} h_{it}(\mathbf{b})$$

• The kxk meat for cluster g in the sandwich matrix is estimated as

$$\widehat{\mathbf{X'\Sigma X}} = (\sum_{t=J+1}^{T} h_t(\mathbf{b})\ h_{t-j}(\mathbf{b})')$$

• Remark: The NW method is applied to the time series of cross-sectional averages of $h_{it}(\mathbf{b})$. We average over the G clusters to get G $h_t(\mathbf{b})$. We use the sandwich matrix to estimate *clustered NW SE*, using w(*l*) as usual Bartlett or QS weights –other weights are OK.

Panel Data Models: (A3') – PCSE

• The NW method is applied to the time series of cross-sectional averages of $h_{it}(\mathbf{b})$. By using cross-sectional averages, estimated SE are consistent independently of the panel's cross-sectional dimension *N*.

• Clearly, these *clustered NW SE* reduce to the usual NW SE if each cluster only has one observation.

• If we do not suspect autocorrelation problems –not rare, given that many panel data sets have heavy temporally spaced observations-, we can rely on White SE ($\mathbf{S}_0$).

• These clustered standard errors are called *Driscoll & Kraay SE* (DK SE's). The clustered White-style SE are, sometimes, called *Rogers SE*. They can all be just referred as LZ SE!

Panel Data Models: (A3') – PCSE

• These PCSE's are robust to very general forms of cross-sectional (and temporal) dependence. But, the usual problems with NW SE apply (downwards biased, poor performance in finite sample, etc.)

• PCSE's using HAR estimators (based on KVB SE) are possible, see Hansen (2007). Bootstrapping SE is also possible -many approaches; see Goncalves and Perron (2017) for a factor model application.

• Consistency of the PCSE is discussed by White (1984), Liang and Zegger (1986) for panels with finite number of observations per cluster, as G (or N) $\rightarrow \infty$. Hansen (2007) shows that PCSE can be used with $N_g \rightarrow \infty$ –i.e., long panels–, in addition to $G \rightarrow \infty$.

Note: Asymptotic inference can be affected by small G, cluster heterogeneity (different sizes, dominant cluster), and experiments where "treatment" occurs only for a small number of clusters.

Panel Data Models: (A3') – PCSE

• Lots of potential issues when G is not large. PCSE can be very poor. Angrist and Pischke (2008) popularized "$G > 42$" for reliable inferences. Some evidence that in many situations is a "generous" rule.

• Technical point: In practice, clusters vary greatly in size, Hansen and Lee (2019) restrict the variation of N_g relative to the total sample size, N. Now, not all $N_g \rightarrow \infty$, as $N \rightarrow \infty$ (in addition to $G \rightarrow \infty$.)

Under this situation, the "normalizing factor," for the application of the CLT for OLS **b**, is not $\sqrt{G}$, but "unknown." This justifies the widely use of the $t_{G\text{-}1}$ distribution for tests, which is more conservative.

Panel Data Models: PCSE – Clustering

• Before calculating the NW SE, we cluster the data to remove the dependence caused by the within group correlation of the data.

• We can cluster the SE by one variable (say, industry) or by several variables (say, year and industry) –"*multi-level clustering*." If these several variables are nested (say, industry and state), cluster at highest level.

• We assume that the correlations within a cluster (a group of firms, a region, different years for the same firm, different years for the same region) are the same for different observations.

• Different clusters can produce very different SE. We want to cluster in groups that produce correlated errors. Usually, we cluster using economic theory (clustering by industry, year, industry and year).

Panel Data Models: PCSE – Clustering Remarks

• Since we allow for correlation between observations, clustered SE will increase CIs. The higher the clustering level, the larger the resulting SE.

• In simulations, MacKinnon and Webb (2020) show that clustering at the wrong level has serious implications: "too fine" clustering leads to serious over-rejections; while "too coarse" clustering leads to some over-rejection and loss of power, especially when G is small.

• Ibragimov and Muller (2016) have a test for the appropriate clustering level. It is based on the observed variation across different clusters.

• When G is small, PCSE tend to be small.

• The asymptotics of multi-level clustering (say, by firm and by year), popular in economics and finance, are not well established.

Panel Data Models: PCSE – Clustering Remarks

• Practical rules

- Usual rules of thumb for picking the clustering level:

(1) Use the coarsest feasible level (Cameron and Miller, 2015), but this can be not reasonable when G is small or clusters very different in size.

(2) Try different ways of defining clusters and see how the estimated SE are affected. Be conservative, use the cluster with the largest SE (Angrist & Pischke, 2008).

- If aggregate variables (say, by industry, or zip code) are used in the model, clustering should be done at that level.

- When the data correlates in more than one way, we have two cases:
- If nested (say, city and state), cluster at highest level of aggregation
- If not nested (e.g., time and industry), use "*multi-level clustering*."

Pooled Model

• General DGP $\quad y_{i,t} = \boldsymbol{x}_{i,t}'\, \boldsymbol{\beta} + c_i + \varepsilon_{i,t}$ $\quad$ & (**A2**)-(**A4**) apply.

• The pooled model assumes that unobservable characteristics are uncorrelated with $\boldsymbol{x}_{i,t}$. We can rewrite panel DGP as:

$$y_{i,t} = \boldsymbol{x}_{i,t}'\, \boldsymbol{\beta} + v_i, \qquad \text{where} \quad v_i = c_i + \varepsilon_{i,t} \quad \text{(compound error)}$$

To get a consistent estimator of $\boldsymbol{\beta}$, we need E[$\boldsymbol{x}_{i,t}'\, v_i$] = 0.

<u>Note</u>: E[$\boldsymbol{x}_{i,t}'\varepsilon_{i,t}$] is derived from **(A2)** E[$\varepsilon_{i,t} | \boldsymbol{x}_{i,t}, c_i$] = 0. Then, to get consistency, we need E[$\boldsymbol{x}_{i,t}' c_i$] = 0 for all t.

• Given the assumptions, we can assume $c_i = \alpha$ –a constant, independent of i. That is, no heterogeneity. Then:

$$y_{i,t} = \boldsymbol{x}_{i,t}'\, \boldsymbol{\beta} + \alpha + \varepsilon_{i,t} \qquad \Rightarrow \text{CLM, with } k + 1 \text{ parameters.}$$

Pooled Model

• We have the CLM, estimating $k + 1$ parameters :

$$y_{i,t} = \boldsymbol{x}_{i,t}{}' \, \beta + \alpha + \varepsilon_{i,t} \quad \Rightarrow \text{Pooled OLS is BLUE \& consistent.}$$

• Stacking the variables in matrices, we have:

$$\boldsymbol{y} = \mathbf{X}\,\boldsymbol{\beta} + \alpha\,\iota + \boldsymbol{\varepsilon}$$

Dimensions:

- $\boldsymbol{y}$, ι and $\boldsymbol{\varepsilon}$ are $\sum_{i=1}^{N} T_i$x1
- $\mathbf{X}$ is $\sum_{i=1}^{N} T_i$xk
- $\boldsymbol{\beta}$ is kx1

• We can re-write the pooled equation model as:

$\boldsymbol{y} = \mathbf{X}^* \,\boldsymbol{\beta}^* + \boldsymbol{\varepsilon}$, $\quad \mathbf{X}^* = [\mathbf{X}\ \iota]$ $\quad - \sum_{i=1}^{N} T_ix(k+1)$ matrix:

$\quad \boldsymbol{\beta}^* = [\boldsymbol{\beta}\ \alpha]'$ $\quad - (k+1)$x1 matrix

Pooled Model

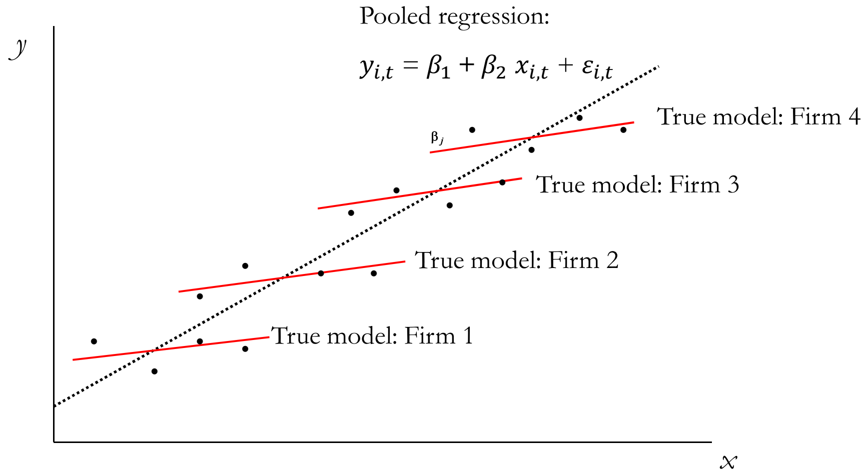
• In this context, OLS produces BLUE and consistent estimator. In this model, we refer to *pooled OLS estimation*

• Of course, if our assumption regarding the unobservable variables is wrong, we are in the presence of an omitted variable, c_i.

• Then, we have potential bias and inconsistency of pooled OLS. The magnitude of these problems depends on how the true model behaves: 'fixed' or 'random.'

Pooled Regression: Heterogeneity Bias

• In the pooled model, there is no model for group/individual *i* heterogeneity. Thus, pooled regression may result in *heterogeneity bias*:



Pooled Regression: Within Transformation

• We can estimate $\boldsymbol{\beta}$ ($\beta_1 = \alpha$) by centering the observations around their group/individual means. That is,

$$y_{i,t} = \beta_1 + \sum_{j=2}^{k} \beta_j\, \boldsymbol{x}_{ij,t} + \delta\, t + \varepsilon_{i,t}$$

Subtracting the mean:

$$y_{i,t} - \bar{y}_i = \sum_{j=2}^{k} \beta_j\, (\boldsymbol{x}_{ij,t} - \overline{\boldsymbol{x}}_{ij}) + \delta + (\varepsilon_{i,t} - \bar{\varepsilon}_i)$$

• This method is called the **within-groups estimation** because the model explains the variations about the mean of the dependent variable in terms of the variations about the means of the explanatory variables for the set of observations relating to a given unit, i.

• That is, this estimator reflects the time-series or within-individual i information reflected in the changes within individuals across time.

$\Rightarrow \boldsymbol{\beta}$ is estimated using the time-series information in the data.

Pooled Regression: Within Transformation

• There is a cost in the simplicity of the within-groups estimation. First, the intercept β_1 and any $\boldsymbol{x}_{ij}$ variable that remains constant for each individual (say, gender or College degree) will drop out of the model.

The elimination of the intercept may not matter, but the loss of the unchanging explanatory variables may be frustrating.

⇒ Obviously, if we are interested on the effect of gender on CEO compensation, within transformation will not work. But it will work well if we are interested on the effect of an independent Board of Directors, by looking at the compensation pre-/post-BOD.

Pooled Regression: Between Transformation

• There is an additional alternative to estimate $\boldsymbol{\beta}$, by expressing the model in terms of group/individual means. That is,

$$y_{i,t} = \beta_1 + \sum_{j=2}^{k} \beta_j \, \boldsymbol{x}_{ij,t} + \delta \, t + \varepsilon_{i,t}$$

Computing the mean:

$$\bar{y}_i = \beta_1 + \sum_{j=2}^{k} \beta_j \, \bar{\boldsymbol{x}}_{ij} + \delta^* + \bar{\varepsilon}_i$$

• It is called the *between estimator* because it relies on variations between individuals (say, i & j). We are estimating $\boldsymbol{\beta}$ using the cross-sectional information in the data (the time-series individual *i* variation is gone!).

⇒ Obviously, if we are interested on the effect of a new independent BOD during the tenure of a CEO on that CEO's compensation, between transformation will not work. But it will work well if we study the effect of gender on CEO's compensation.

Pooled Regression: Between Transformation

• We lose observations (and power!): we have only N data points.

Remark: Under the usual assumptions, pooled OLS using the between transformation is consistent and unbiased.

Useful Analysis of Variance Notation (Greene)

• The variance (total variation) quantifies the idea that each individual i –say, each firm– differs from the overall average. We can decompose the variance into two parts: a within-group/individual part and a between group/individual part:

$$.\sum_{i=1}^{N}\sum_{t=1}^{T_i}(z_{it} - \bar{\bar{z}})^2 = \sum_{i=1}^{N}\sum_{t=1}^{T_i}(z_{it} - \bar{z}_{i*})^2 + \sum_{i=1}^{N} T_i(\bar{z}_{i*} - \bar{\bar{z}})^2$$

Total variation = Within groups variation + Between groups variation

• Interpretation:

- Within group variation: Measures variation of individuals over time.

- Between group variation: Measures variation of the means across individuals.

WHO Data (Greene)

The model used by the researchers at WHO was

$$\ln DALE_{it} = \alpha_i + \beta_1 \ln \textit{Health Expenditure}_{it} + \beta_2 \ln \textit{Education}_{it} + \beta_3 \ln \textit{Education}_{it}^2 + \varepsilon_{it}$$

The analysis of variance for a variable x_{it} is based on the decomposition

$$\sum_{i=1}^{n}\sum_{t=1}^{T_i}(x_{it} - \bar{\bar{x}})^2 = \sum_{i=1}^{n}\sum_{t=1}^{T_i}(x_{it} - \bar{x}_{i.})^2 + \sum_{t=1}^{n} T_i(\bar{x}_{i.} - \bar{\bar{x}})^2 .$$

Analysis of Variance for WHO Data on Health Care Attainment

Variable	Within Groups Variation	Between Groups Variation
DALE	5.645%	94.355%
COMP	0.150%	99.850%
Expenditure	0.635%	99.365%
Education	0.178%	99.822%

Note: The variability is driven by between groups variation

Pooled Model: Living with (A3')

- We start with the pooled model:

$$\boldsymbol{y} = \mathbf{X}^* \beta^* + \varepsilon, \quad \text{with } \mathbf{X}^* = [\mathbf{X}\ \iota] \quad - \textstyle\sum_{i=1}^{N} T_i \text{x}(k+1) \text{ matrix.}$$

$$\beta^* = [\beta\ \alpha]' \quad - (k+1)\text{x1 matrix}$$

Now, we allow $\text{E}[\boldsymbol{\varepsilon}_i\ \boldsymbol{\varepsilon}_j'|\mathbf{X}_i] = \sigma_{ij}\ \Omega_{ij}$

- Potentially a lot of different forms for $\text{E}[\boldsymbol{\varepsilon}_i\ \boldsymbol{\varepsilon}_j'|\boldsymbol{X}_i]$ in a panel:

- *Individual heteroscedasticity*. $\quad \text{E}[\varepsilon_i^2|\boldsymbol{X}_i] = \sigma_i^2$
- *Individual/group effects*: $\quad \text{E}[\varepsilon_{it}\varepsilon_{is}|\boldsymbol{X}_i] \neq 0\ (t \neq s)$
- *Time (SUR or spatial) effects:* $\quad \text{E}[\varepsilon_{it}\varepsilon_{jt}|\boldsymbol{X}_i] \neq 0\ (i \neq j)$
- *Persistent common shocks*: $\quad \text{E}[\varepsilon_{it}\varepsilon_{js}|\boldsymbol{X}_i] \neq 0\ (i \neq j)$ and $|t-s| < L$

- Heteroscedasticity points to GLS efficient estimation, but, for consistent inferences we can use OLS with *clustered* White/NW SE.

Pooled OLS: Cornwell and Rupert Data (Greene)

Cornwell and Rupert Returns to Schooling Data, 595 Individuals, 7 Years
Variables in the file are

EXP = work experience
WKS = weeks worked
OCC = occupation, 1 if blue collar,
IND = 1 if manufacturing industry
SOUTH = 1 if resides in south
SMSA = 1 if resides in a city (SMSA)
MS = 1 if married
FEM = 1 if female
UNION = 1 if wage set by unioin contract
ED = years of education
BLK = 1 if individual is black
LWAGE = log of wage = dependent variable in regressions

These data were analyzed in Cornwell, C. and Rupert, P., "Efficient Estimation with Panel Data: An Empirical Comparison of Instrumental Variable Estimators," Journal of Applied Econometrics, 3, 1988, pp. 149-155. See Baltagi, page 122 for further analysis. The data were downloaded from the website for Baltagi's text.

Pooled OLS: Clustered SE – Results (Greene)

```
Ordinary     least squares regression ............
LHS=LWAGE    Mean                 =        6.67635
Residuals    Sum of squares       =      522.20082
             Standard error of e  =         .35447
Fit          R-squared            =         .41121
Model test   F[  8,   4156] (prob) =   362.8(.0000)
Panel Data Analysis of LWAGE               [ONE way]
               Unconditional ANOVA (No regressors)
Source         Variation  Deg. Free.   Mean Square
Between        646.25374        594.       1.08797
Residual       240.65119       3570.        .06741
Total          886.90494       4164.        .21299
--------+----------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]
--------+----------------------------------------------------
     EXP|     .04085***         .00219       18.693   .0000
   EXPSQ|    -.00069***     .480428D-04     -14.318   .0000
     OCC|    -.13830***         .01480       -9.344   .0000
    SMSA|     .14856***         .01207       12.311   .0000
      MS|     .06798***         .02075        3.277   .0010
     FEM|    -.40020***         .02526      -15.843   .0000
   UNION|     .09410***         .01253        7.509   .0000
      ED|     .05812***         .00260       22.351   .0000
Constant|    5.40160***         .04839      111.628   .0000
```

Pooled OLS: Clustered SE – Results (Greene)

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]
Constant	5.40159723	.04838934	111.628	.0000
EXP	.04084968	.00218534	18.693	.0000
EXPSQ	-.00068788	.480428D-04	-14.318	.0000
OCC	-.13830480	.01480107	-9.344	.0000
SMSA	.14856267	.01206772	12.311	.0000
MS	**.06798358**	**.02074599**	**3.277**	**.0010**
FEM	-.40020215	.02526118	-15.843	.0000
UNION	.09409925	.01253203	7.509	.0000
ED	.05812166	.00260039	22.351	.0000
Clustered SE				
Constant	5.40159723	.10156038	53.186	.0000
EXP	.04084968	.00432272	9.450	.0000
EXPSQ	-.00068788	.983981D-04	-6.991	.0000
OCC	-.13830480	.02772631	-4.988	.0000
SMSA	.14856267	.02423668	6.130	.0000
MS	**.06798358**	**.04382220**	**1.551**	**.1208**
FEM	-.40020215	.04961926	-8.065	.0000
UNION	.09409925	.02422669	3.884	.0001
ED	.05812166	.00555697	10.459	.0000

Note: Clustered SE's tend to be bigger. The more correlation allowed, the higher the SE.

Pooled Model: PCSE – Remarks

• All the remarks that we have done before, apply. Driscoll and Kraay SE –i.e., cluster NW SE- are almost universally applied. Then, the bigger the cross-sectional correlation, the bigger the SE.

• In simulations, it is found (as expected) that the PCSE perform better when there is cross-sectional dependence in the data. But, when there is no dependence in the cross-section, the standard White or NW SE do better. In some cases, these differences can be significant

• Testing for cross-sectional dependence may be a good idea, especially when results are not robust to different SE. LM tests can be easily implemented. Pesaran (2004) proposes an easy test.

Pooled Model: PCSE – Remarks

• The computational issues are straightforward for balanced data. We need only the vector of residuals, the model matrix (**X**), and indicators for group (and, usually, time) to form the clusters.

• But for unbalanced there are two approaches

- Create a balanced subset of the panel to estimate Ω.

    Advantage: Computationally simple.

- Loop over $e_{i,t}$ $e_{j,t}$ pairs to estimate covariances over available overlapping time frames (loop over all pairs can take a long time).

    Advantage: Information is not thrown out.

Application: Bid-Ask Spread (Hoechle)

Table 2: Comparison of standard error estimates for pooled OLS estimation

SE	OLS	White	Rogers	Newey-West	Driscoll-Kraay
aVol	-0.0018***	-0.0018**	-0.0018*	-0.0018*	-0.0018
	(-4.006)	(-2.043)	(-1.831)	(-1.760)	(-1.627)
Size	-0.1519***	-0.1519***	-0.1519***	-0.1519***	-0.1519***
	(-17.412)	(-12.496)	(-6.756)	(-10.717)	(-14.789)
TRMS2	0.0033***	0.0033***	0.0033***	0.0033***	0.0033***
	(5.295)	(5.520)	(5.495)	(5.582)	(3.773)
TRMS	-0.0018	-0.0018	-0.0018	-0.0018	-0.0018
	(-0.370)	(-0.353)	(-0.381)	(-0.340)	(-0.351)
Const.	1.4591***	1.4591***	1.4591***	1.4591***	1.4591***
	(25.266)	(18.067)	(9.172)	(14.883)	(10.775)
# obs.	11775	11775	11775	11775	11775
# clusters			219		219
R^2	0.029	0.029	0.029	0.029	0.029

This table provides the coefficient estimates from the regression model in (10) estimated by pooled OLS. The t-stats (in parentheses) are based on standard error estimates obtained from the covariance matrix estimators in the column headings. The dataset contains monthly data from December 2000 to December 2005 for a panel of 219 stocks that have been randomly selected from the MSCI Europe constituents list as of December 31, 2000. The dependent variable in the regression is the relative bid-ask spread *BA*. *aVol* is the abnormal trading volume, *Size* contains the stock's size decile, *TRMS* denotes the monthly return in % of the MSCI Europe total return index and *TRMS2* is the square of it. *, **, and *** imply statistical significance on the 10, 5, and 1% level, respectively.

Pooled Model with (A3') - GLS

• We start with the pooled model:

$$\boldsymbol{y} = \mathbf{X}^* \boldsymbol{\beta}^* + \boldsymbol{\varepsilon},$$

where $\mathbf{X}^* = [\mathbf{X}\ \iota]$ $\quad - \sum_{i=1}^{N} T_i \text{x}(k+1)$ matrix.

$\boldsymbol{\beta}^* = [\boldsymbol{\beta}\ \alpha]'$ $\quad - (k+1)\text{x}1$ matrix

Now, we allow $E[\boldsymbol{\varepsilon}_i\ \boldsymbol{\varepsilon}_j' | \boldsymbol{X}_i] = \sigma_{ij}\ \Omega_{ij}$

• We can use OLS with PCSE's or we can do GLS.

Note: Why GLS? Efficiency.

• Suppose $\Omega_{ij} = \mathbf{I}_T$. Then, we only have cross-equation correlation, not time correlation. We are back in the (aggregation) SUR framework

Pooled Model with SUR - GLS

• Suppose $\Omega_{ij} = \mathbf{I}_T$. We are in the (aggregation) SUR framework:

$$\hat{\beta}_{GLS} = (X'V^{-1}X)^{-1} X'V^{-1}y = (X'[\Sigma \otimes I]^{-1} X)^{-1} X'[\Sigma \otimes I]^{-1} y$$
$$= (X'[\Sigma^{-1} \otimes I]X)^{-1} X'[\Sigma^{-1} \otimes I]y$$

• For FGLS, use the pooled OLS residuals $\boldsymbol{e}_i$ and $\boldsymbol{e}_j$ to estimate the covariance σ_{ij}. Note that

$$\hat{\Sigma} = \frac{1}{T}\sum_{t=1}^{T} e_t e_t' = \frac{1}{T} E'E$$

where $\boldsymbol{E}$ is a TxN matrix and $\boldsymbol{e}_t = [e_{1,t}\ e_{2,t}\ \dots\ e_{N,t}]'$ is Nx1 vector. We need to invert $\hat{\Sigma}$ (NxN matrix).

Note: In general, the rank($\boldsymbol{E}$) $\leq T$. Then, rank($\hat{\Sigma}$) $\leq T < N \Rightarrow$ singularity, FGLS cannot be computed. This is a problem of the data, not the model.

Pooled Model with Heteroscedasticity - GLS

• Now, suppose we have groupwise heteroscedasticity. That is,

$E[\boldsymbol{\varepsilon}_i\ \boldsymbol{\varepsilon}_j' | \boldsymbol{X}_i] = 0$ for $i \neq j$

$E[\varepsilon_i^2 | \boldsymbol{X}_i] = \text{Var}[\varepsilon_i | \boldsymbol{X}_i] = \sigma_i^2\ \mathbf{I}_T$

• We do FGLS, as usual, using the pooled OLS residuals e_i to estimate the variance σ_i^2 and, thus, to estimate Σ:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & \ldots & 0 \\ 0 & \sigma_2^2 & \ldots & 0 \\ .. & .. & \ldots & .. \\ 0 & 0 & \ldots & \sigma_N^2 \end{bmatrix}$$

• We can test this model with H_0: $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_N^2$. We can use:

$$W = \sum_{j=1}^{N} (s_i^2 - s_{pooled}^2)/\text{Var}(s_i^2) \xrightarrow{d} \chi_N^2$$

where s_i^2 is computed using the pooled OLS e_i residuals.

Pooled Model with Autocorrelation - GLS

• Now, suppose we have individual autocorrelation. That is,

$E[\varepsilon_{it}\ \varepsilon_{js} | \mathbf{X}_i] = 0$ for $i \neq j$

$E[\varepsilon_{it}\ \varepsilon_{it-p} | \mathbf{X}_i] \neq 0$ -for example, $\varepsilon_{i,t} = \rho_i\ \varepsilon_{i,t-1} + u_{i,t}$

$\text{Var}[\varepsilon_{it} | \mathbf{X}_i] = \sigma^2$

• We do FGLS, as usual, using the pooled OLS residuals e_i to estimate the ρ_i and, thus, to estimate Σ_i:

$$\Sigma_i = \frac{\sigma_u^2(i)}{1-\rho_i^2} \begin{bmatrix} 1 & \rho_i & \ldots & \rho_i^{T-1} \\ \rho_i & 1 & \ldots & 0 \\ .. & .. & \ldots & .. \\ \rho_i^{T-1} & \rho_i^{T-2} & \ldots & 1 \end{bmatrix}$$

• We can test this model with H_0: $\rho_1 = \rho_2 = \ldots = \rho_N = 0$. We can use an LM test to test H_0.

Pooled OLS with First Differences

• From the general DGP:

$$y_{i,t} = \boldsymbol{x}_{i,t}{}'\,\beta + c_i + \varepsilon_{i,t} \qquad \text{\& (A2)-(A4) apply.}$$

It may still be possible to use OLS to estimate β, when we have individual heterogeneity. We can use OLS if we eliminate the cause of heterogeneity: c_i.

We can do this by taking first differences of the DGP. That is,

$$\Delta y_{i,t} = y_{i,t} - y_{i,t-1} = (\boldsymbol{x}_{i,t} - \boldsymbol{x}_{i,t-1})'\,\beta + \Delta\, c_i + \Delta\, \varepsilon_{i,t}$$
$$= \Delta\boldsymbol{x}_{i,t}{}'\beta \; + u_{i,t}$$

Note: All time invariant variables, including c_i, disappear from the model (one "*diff*"). If the model has a time trend –economic fluctuations –, it also disappear, it become the constant term (the other "*diff*"). Thus, this method is usually called "*diffs in diffs*" (*DD* or *DiD*).

Pooled OLS with First Differences

• With strict exogeneity of ($\boldsymbol{X}_i$, c_i), the OLS regression of $\Delta y_{i,t}$ on $\Delta\boldsymbol{x}_{i,t}$ is unbiased and consistent, but inefficient.

• Why? The error is not longer $\varepsilon_{i,t}$, but $u_{i,t}$. The Var[**u**] is given by:

$$\text{Var}\begin{pmatrix} \varepsilon_{i,2} - \varepsilon_{i,1} \\ \varepsilon_{i,3} - \varepsilon_{i,2} \\ \vdots \\ \varepsilon_{i,T_i} - \varepsilon_{i,T_i-1} \end{pmatrix} = \begin{bmatrix} 2\sigma_\varepsilon^2 & -\sigma_\varepsilon^2 & 0 & 0 \\ -\sigma_\varepsilon^2 & 2\sigma_\varepsilon^2 & -\sigma_\varepsilon^2 & \vdots \\ 0 & -\sigma_\varepsilon^2 & \ddots & -\sigma_\varepsilon^2 \\ 0 & \cdots & -\sigma_\varepsilon^2 & 2\sigma_\varepsilon^2 \end{bmatrix} \text{ (Toeplitz form)}$$

• That is, first differencing produces heteroscedasticity. Efficient estimation method: GLS.

• It turns out that GLS is complicated. Use OLS in first differences and use Newey-West SE/PCSE with one lag.

OLS with First Diffs: Treatment Application

• Suppose there is random assignment to treatment and control groups, like in a typical medical experiment.

• We compare the change in outcomes across the treatment and control groups to estimate the treatment effect. (We used this method –"*natural experiment*"– in Lecture 8 to deal with endogeneity.)

• With two periods –i.e., before and after– and strict exogeneity:

$$\Delta y_{i,t} = y_{i,2} - y_{i,1} = \delta_0 + \delta_1\ Treatment_i + (\boldsymbol{x}_{i,2} - \boldsymbol{x}_{i,1})'\ \boldsymbol{\beta}\ + u_{i,t}$$

(This is a CLM. OLS is consistent and unbiased).

Then,

$E[\Delta y_{i,t} \mid Treatment_i = 1] = \delta_0 + \delta_1 + E[\Delta \boldsymbol{x}_{i,t}' \mid Treatment_i = 1]\ \boldsymbol{\beta}$

$E[\Delta y_{i,t} \mid Treatment_i = 0] = \delta_0 + E[\Delta \boldsymbol{x}_{i,t}' \mid Treatment_i = 0]\ \boldsymbol{\beta}$

OLS with First Diffs: Treatment Application

• Assuming that controls are orthogonal to Treatment:

$$\delta_1 = E[\Delta y_{i,t} \mid Treatment_i = 1] - E[\Delta y_{i,t} \mid Treatment_i = 0]$$

δ_1 is the difference in average change in the two periods –i.e., before and after– between the treated and control groups. This is the *diffs in diffs* (DD, DiD) estimator.

• Typical problem: Exogeneity (randomness) of treatment. That is,

(**A2**) $E[u_{i,t} \mid Treatment_i] = 0$.

• In medical experiments, *diffs in diffs* estimation is routinely used to evaluate the effectiveness of a new treatment and/or medication. Usual H_0: $\delta_1 = 0$. It can be tested with a *t-test* (using HAR/PCSEs).

OLS with First Diffs: Treatment Application

• Same result can be derived by looking at levels DGP ($y_{i,t}$, $\boldsymbol{x}_{i,t}$) including two dummies: One for *Treatment,* Tr_i, and one for after *Treatment* ($Post_i$):

$y_{i,t} = \boldsymbol{x}_{i,t}'\beta + c_i + \gamma_0 + \gamma_1 Tr_i + \gamma_2 Post_i + \delta_1 Tr_i \text{ x } Post_i + u_{i,t}$

Now, it is easy to separate *cross-sectional differences* from *time-series differences*.

- Cross-sectional difference

$\text{E}[y_{i,t} \; |Tr_i = 1, Post_i = 1] = \boldsymbol{x}_{i,t}'\beta + c_i + \gamma_0 + \gamma_1 + \gamma_2 + \delta_1$

$\text{E}[y_{i,t} \, |Tr_i = 0, Post_i = 1] = \boldsymbol{x}_{i,t}'\beta + c_i + \gamma_0 + \gamma_2$

Then, the cross sectional difference is:

$\text{E}[y_{i,t} \, |Tr_i=1, Post_i= 1] - \text{E}[y_{i,t} \, |Tr_i = 0, Post_i= 1] = \gamma_1 + \delta_1$

OLS with First Diffs: Treatment Application

- Cross-sectional difference (continuation)

$\text{E}[y_{i,t} \; |Tr_i= 1, Post_i= 1] - \text{E}[y_{i,t} \; |Tr_i= 0, Post_i= 1] = \gamma_1 + \delta_1$

Note: Unbiased if $\gamma_1 = 0 \Rightarrow$ No permanent difference between the treatment and control groups.

$\text{E}[y_{i,t} \; |Tr_{i_i}= 1, Post_i= 0] - \text{E}[y_{i,t} \; |Tr_i= 0, Post_i= 0] = \gamma_1$

OLS with First Diffs: Treatment Application

- Time-sectional difference

E[$y_{i,t}$ |Tr_i= 1, $Post_i$= 1] – E[$y_{i,t}$ |Tr_i= 1, $Post_i$= 0] = γ_2 + δ_1

Note: Unbiased if $\gamma_2 = 0 \Rightarrow$ No common trend over the pre- and post-treatment times.

E[$y_{i,t}$ |Tr_i= 0, $Post_i$= 1] – E[$y_{i,t}$ |Tr_i= 0, $Post_i$= 0] = γ_2

Note: From Lecture 8, we need to make sure that Treatment is the only difference between the two groups. Thus, in the absence of treatment, the average change in $y_{i,t}$ would have been the same for both groups.

This is a key assumption behind the DD estimator, tested with *t-tests* or, more usual, by looking at a graph of the behavior of both groups before treatment –see Redding & Sturm (2008) in Lecture 8.

OLS with First Diffs: Natural Experiment

• In Finance & Economics, especially in Corporate Finance, we apply the DD method when we use *natural experiments* (change in a law, policy or a regulation) to study the effect of x_t on y_t. (Recall Lecture 8.)

• We have two periods: Before and after the natural experiment (the *treatment*).

• If we also have a well-defined control group, where the treatment was not administered –i.e., the natural experiment never occurred–, then, we can use DD estimation.

• The number of groups, S, (treated & not treated) under consideration is usually small –typically 2. *N* is usually very large.

Diffs in Diffs: Natural Experiment - 1

Example 1: We are interested in the effect of labor shocks on wages and employments. Natural experiment: The 1980 *Mariel boatlifs*, a temporary lifting of emigration restrictions in Cuba. Most of the *marielitos* (the 1980 Cuban immigrants) settled in Miami.

• Two periods: Before and after the 1980 Mariel boatlifs.

• Control group: Low skilled workers in Houston, LA and Atlanta.

• Calculate unemployment and wages of low skilled workers in both periods. Then, regress $\Delta y_{i,t}$ against a set of control variables (industry, education, age, etc.) and a treatment dummy:

$$\Delta y_{i,t} = y_{i,2} - y_{i,1} = \delta_0 + \delta_1\, Tr_i + (\boldsymbol{x}_{i,2} - \boldsymbol{x}_{i,1})'\,\boldsymbol{\beta} + u_{i,t}$$

• H_0: $\delta_1 = 0$. Card (1990) found no effect of massive immigration.

Diffs in Diffs: Natural Experiment - 2

Example 2: Suppose we are interested in the effect of a substantial increase in bank deposits on lending practices. We can use the shale revolution, which started around 2011, as a natural experiment.

• Two periods: Before and after shale revolution (say, 2011).

• Control group: Banks in counties outside shale formation areas.

• Measure lending practices (amount lent, FICO scores of loans, etc.), $y_{i,t}$, in both periods & regress $\Delta y_{i,t}$ against a set of control variables (size of county, size of bank, experience of bank employees, etc.) and a treatment dummy:

$$\Delta y_{i,t} = y_{i,2} - y_{i,1} = \delta_0 + \delta_1\, Tr_i + (\boldsymbol{x}_{i,2} - \boldsymbol{x}_{i,1})'\,\boldsymbol{\beta} + u_{i,t}$$

• H_0: $\delta_1 = 0$. Glije (2011) rejects H_0, especially for counties dominated by small banks.

Diffs in Diffs: Remarks

• We express the DGP in terms of i (individuals), s (groups), and t (time): $y_{i,s,t} = \delta_s + \delta_t + \delta_1\, Tr_{s,t} + \boldsymbol{x}_{i,s,t}{}'\beta + \varepsilon_{i,s,t}$

• Usually, we have small S and T; but large N. Since, in general, we have within group correlation (treated individuals show similar errors), the asymptotics of the *t-test* are driven by $S{*}T$.

• Donald and Lang (2004): Under the usual (generous) assumptions, it converges to a normal distribution (a $t_{ST\text{-}K}$ may work better).

• Intuition: Suppose that within s, t groups the errors are perfectly correlated. Then, we only have $S{*}T$ independent observations!

• Given the potential (time-varying) correlations in the errors, OLS SE can be terrible. PCSE tend to do better.

Dealing with Attrition

• Attrition problem: If an unbalanced panel is a result of some selection process related to $\varepsilon_{i,t}$, then endogeneity is present and need to be dealt with using some correction methods. Otherwise, we have *attrition bias*.

• **Example**: In the "Quality of Life for cancer patients" study discussed in Greene, appearance for the second interview was low for people with initial low QOL (death or depression) or with initial high QOL (do not need the treatment).

• Solutions to the attrition problem
– Heckman selection model (used in the study)
  • Prob[Present at exit|covariates] = $\Phi(z'\theta)$ (Probit model)
  • Additional variable added to difference model $\lambda_i = \phi(\boldsymbol{z}_i{}'\theta)/\Phi(\boldsymbol{z}_i{}'\theta)$
– The FDA solution: fill with zeros. (!)

Pooled Model: ML Estimation

• In the pooled model, $\boldsymbol{y} = \mathbf{X}\,\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, we assume $\boldsymbol{\varepsilon}_t \sim N(0, \boldsymbol{\Sigma})$, where $\boldsymbol{\varepsilon}_t = [\varepsilon_{1,t}, \varepsilon_{2,t},..., \varepsilon_{N,t}]'$ and $\boldsymbol{\Sigma}$ is an NxN matrix.

• We can write the log likelihood function as:

$$L = \log L(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\mathbf{X}) = -NT/2 \ln(2\pi) - T/2 \ln|\boldsymbol{\Sigma}| - ½ \sum_{t=1}^{T} \boldsymbol{\varepsilon}_t'\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}_t$$

• The ML estimator is equal to the iterated FGLS estimator.

• Testing is straightforward with likelihood ratio test.

Example: H_0: No cross correlation across equations: The off-diagonal elements of $\boldsymbol{\Sigma}$ are zero.

$$LR = T\,(\ln|\hat{\Sigma}_R| - \ln|\hat{\Sigma}_U|) = T\,(\sum_{j=1}^{T} \ln(s_i^2) - \ln|\hat{\Sigma}|) \xrightarrow{d} \chi^2_{N(N-1)/2}$$

Main Models: FEM and REM

• Two main approaches to fitting models using panel data:

(1) Fixed effects regressions.

(2) Random effects regressions.

• The key difference between these two approaches is how the unobservable characteristics –the *individual effects*– are modeled.

• Terminology from experimental design (say, psychology or medicine), where the emphasis was on the kind of sample at hand and inferences:

- FE: The individuals are fixed. The differences between them are not of interest, only $\boldsymbol{\beta}$ is interesting. No intent on generalizing the results.

- RE: The individuals come from a random sample drawn from a larger population, and the variance between them is interesting and can be informative about the larger population.

The Fixed Effects Model (FEM)

• The fixed effects (FE) model

(A1) $y_{i,t} = \boldsymbol{x}_{i,t}{}'\, \beta + c_i + \varepsilon_{i,t}$ –observation for individual i at time t.

(A2) $E[\varepsilon_{i,t} \,|\, \boldsymbol{X}_{i,s}, c_{i,s}] = 0$, for all t, s. –$\boldsymbol{X}_i$ and c_i strict exogenous.

• The unobserved component, c_i, is arbitrarily correlated with $\boldsymbol{x}_{i,t}$:

$$E[c_i \,|\, \boldsymbol{X}_i] = g(\boldsymbol{X}_i) = constant_i \qquad \Rightarrow \text{Cov}[\boldsymbol{x}_{i,t}, c_i] \neq \mathbf{0}.$$

<u>Note 1</u>: Under the FEM, pooled OLS omits c_i ⇒ biased & inconsistent.

• We summarize ("*control for*") these unobservable effects with α_i, a constant. All time invariant characteristic of individual i (location, gender, nationality, etc.) are swept away under this formulation.

<u>Note 2</u>: In a FEM, individuals serve as their own controls.

Estimation with Fixed Effects

• Whatever effects the omitted variables have on the individual *i* at one time, they will also have the same effect at a later time, thus, their effects will be constant, or "*fixed*."

• For this, we need the omitted variables to have time-invariant values with time-invariant effects. Typical example, a CEO's IQ/gender. We expect this variable to have the same effect at $t=1$ or $t=10$.

• As we will see, FEM are estimated using the *within* transformation. Thus, if individuals do not change much (or at all) across time, a FEM may not work very well. We need within-individuals variability in the variables if we are to use individuals as their own controls.

Estimation with Fixed Effects

• Matrix notation

- In matrix notation for individual i :

$\boldsymbol{y}_i = \boldsymbol{x}_i{}'\,\beta + c_i + \boldsymbol{\varepsilon}_i$ –c_i is a T_ix1 vector. (Each individual has T_i observations.)

- In matrix notation for all individuals –i.e., stacking:

$\boldsymbol{y} = \mathbf{X}\,\boldsymbol{\beta} + \mathbf{c} + \boldsymbol{\varepsilon}$ –Now, $\mathbf{c}$, $\boldsymbol{y}$, and $\boldsymbol{\varepsilon}$ are $\Sigma_i T_i$x1 vectors.

• Dummy variable representation:

$y_{i,t} = \boldsymbol{x}_{i,t}{}'\,\beta + \sum_{j=1}^{N} c_j\, d_{ij,t} + \varepsilon_{i,t}$ with $d_{ij,t} = 1$ if $i = j$

FEM: Estimation

• The FE model assumes $c_i = \alpha_i$ (constant; it does not vary with t):

$\boldsymbol{y}_i = \boldsymbol{X}_i\,\beta + \mathbf{d}_i\,\alpha_i + \boldsymbol{\varepsilon}_i,$ for each individual i.

• Stacking

$$\begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_N \end{pmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{d_1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{X}_2 & \mathbf{0} & \mathbf{d_2} & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{X}_N & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{d_N} \end{bmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\alpha} \end{pmatrix} + \boldsymbol{\varepsilon}$$

$$= [\mathbf{X}, \mathbf{D}] \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\alpha} \end{pmatrix} + \boldsymbol{\varepsilon}$$

$$= \mathbf{Z}\boldsymbol{\delta} + \boldsymbol{\varepsilon}$$

• The FEM is the CLM, but with many independent variables: $k + N$.

⇒ OLS is unbiased, consistent, efficient, but impractical if N is large.

FEM: Estimation

• The OLS estimates of β and α are given by:

$$\begin{pmatrix} \mathbf{b} \\ \mathbf{a} \end{pmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{D} \\ \mathbf{D}'\mathbf{X} & \mathbf{D}'\mathbf{D} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{D}'\mathbf{y} \end{bmatrix}$$

Using the Frisch-Waugh theorem

$$\mathbf{b} = [\mathbf{X}'\mathbf{M}_D\mathbf{X}]^{-1}[\mathbf{X}'\mathbf{M}_D\mathbf{y}]$$

• In practice, we do not estimate **a** –the c_i–, they are not very interesting. Moreover, since we are in a *fixed-T* situation, **a** is unbiased, but not consistent. In addition, there is the potential *incidental parameter problem.*

Note (Greene): LS is an estimator, not a model. Given the formulation with a lot of dummy variables, this particular LS estimator is called Least Squares Dummy Variable (LSDV) estimator.

FEM: Estimation

$$\mathbf{M}_D = \begin{bmatrix} \mathbf{M}_D^1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_D^2 & \mathbf{0} \\ & & \\ \mathbf{0} & \mathbf{0} & \mathbf{M}_D^N \end{bmatrix} \quad \text{(The dummy variables are orthogonal)}$$

$$\mathbf{M}_D^i = \mathbf{I}_{T_i} - \mathbf{d}_i(\mathbf{d}_i'\mathbf{d}_i)^{-1}\mathbf{d}' = \mathbf{I}_{T_i} - (1/T_i)\mathbf{d}_i\mathbf{d}'$$

$$\mathbf{X}'\mathbf{M}_D\mathbf{X} = \Sigma_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_D^i\mathbf{X}_i, \quad \left\{\mathbf{X}_i'\mathbf{M}_D^i\mathbf{X}_i\right\}_{k,l} = \Sigma_{t=1}^{T_i}(x_{it,k}-\bar{x}_{i.,k})(x_{it,l}-\bar{x}_{i.,l})$$

$$\mathbf{X}'\mathbf{M}_D\mathbf{y} = \Sigma_{i=1}^{N}\mathbf{X}_i'\mathbf{M}_D^i\mathbf{y}_i, \quad \left\{\mathbf{X}_i'\mathbf{M}_D^i\mathbf{y}_i\right\}_{k} = \Sigma_{t=1}^{T_i}(x_{it,k}-\bar{x}_{i.,k})(y_{it}-\bar{y}_{i.})$$

• That is, we subtract the group mean from each individual observation. Then, the individual effects disappear. Now, OLS can easily be used to estimate the k β parameters, using the demeaned data.

• We know this method: The *within-groups* estimation.

FEM: Within Transformation Removes Effects

• The within-groups method estimates the parameters using demeaned data. That is,

$$y_{i,t} - \bar{y}_i = \sum_{j=2}^{k} \beta_j \, (\boldsymbol{x}_{ij,t} - \bar{\boldsymbol{x}}_{ij}) + \delta^* + (\varepsilon_{i,t} - \bar{\varepsilon}_i)$$

<u>Recall</u>: It is called *within-groups/individuals* method because because it relies on variations *within* individuals rather than *between* individuals.

• For the usual asymptotic results, we need:
– **(A2)** E[$\Delta\varepsilon_{i,t} \,|\, \boldsymbol{X}_i$] = 0.
– **(A3')** E[$\boldsymbol{\varepsilon}_i{}' \, \boldsymbol{\varepsilon}_i \,|\, \boldsymbol{X}_i, c_i$] = $\boldsymbol{\Sigma}$ –different formulations OK.
– **(A4)** E[$\Delta\boldsymbol{X}_i{}'\Delta\boldsymbol{X}_i$] has full rank.

FEM: Within Transformation Removes Effects

• There are costs in the simplicity of the within-groups estimation:

1) All *time-invariant* variables (including constant) for each individual *i* drop out of the model. This eliminates all between-individuals variability (which may be contaminated by omitted variable bias) and leaves only the within-subject variability to analyze.

2) Dependent variables are likely to have smaller variances than in the original specification (measured as deviations from the *i* mean).

3) The manipulation involves the loss of *N* degrees of freedom (we are estimating *N* means!).

FEM: LS Dummy Variable (LSDV) Estimator

- $\mathbf{b}$ is obtained by *within-groups* least squares (group mean deviations).
- Then, we can use the normal equations to estimate $\mathbf{a}$:

$$\mathbf{D'Xb} + \mathbf{D'Da} = \mathbf{D'}\boldsymbol{y}$$

$$\mathbf{a} = (\mathbf{D'D})^{-1}\mathbf{D'}\,(\boldsymbol{y} - \mathbf{Xb})$$

$$\mathrm{a}_i = \frac{1}{T_i}\sum_{t=1}^{T_i}(\boldsymbol{y}_{i,t} - \boldsymbol{x}'_{i,t}\,\mathbf{b}) = \bar{e}_i$$

<u>Note</u>:

– This is simple algebra –the estimator is just OLS

– Again, LS is an estimator, not a model.

– Note what a_i is when T_i=1. Follow this with $y_{i,t} - \mathrm{a}_i - \boldsymbol{x}_{i,t}'\mathbf{b} = 0$ if $T_i = 1$.

FEM: LSDV Estimator

- Avoid dummy variable trap: If a constant is present in the model, the number of dummy variable should be $N - 1$. The omitted individual or group becomes the reference category.

- However, the choice of reference category is often arbitrary and, thus, the interpretation of the α_i will not be particularly interesting.

- Alternatively, we can drop the β_1 intercept and define dummy variables for all of the individuals. This is the more common approach. The α_i now become the intercepts for each of the i's.

- If $\mathrm{E}[\varepsilon_{i,t}\,|\,\boldsymbol{x}_{i,s}, c_i] \neq 0$, then LSDV cannot be used. It is inconsistent. In this case, we need to use IVs. Or a good natural experiment.

FEM: First-Difference (FD) Method

- We can also eliminate the individual FE using the *first-difference method*.

- The unobserved effect is eliminated by subtracting the observation for the previous time period from the observation for the current time period, for all time periods:

$$y_{i,t} - y_{i,t-1} = \sum_{j=2}^{k} \beta_j \, (\boldsymbol{x}_{ij,t} - (\boldsymbol{x}_{ij,t-1}) + \delta(t-(t-1)) + \varepsilon_{i,t} - \varepsilon_{i,t-1}$$

$$\Delta y_{i,t} = \sum_{j=2}^{k} \beta_j \, \Delta\boldsymbol{x}_{ij,t} + \delta + \Delta\varepsilon_{i,t}$$

- The error term is now $(\varepsilon_{i,t} - \varepsilon_{i,t-1})$. As before, differencing induces a moving average autocorrelation if $\varepsilon_{i,t}$ satisfies the CLM assumptions.

Note: If $\varepsilon_{i,t}$ is subject to AR(1) autocorrelation and ρ is close to 1, taking first differences may approximately solve the problem.

FEM: Estimation – FE or FD?

- Summary:

- **Fixed-effects (or Within) Estimator**

– Each variable is demeaned –i.e., subtracted by its average.

– Dummy Variable Regression –i.e., put in a dummy variable for each cross-sectional unit, along with other explanatory variables. This may cause estimation difficulty when N is large.

- **FD Estimator**

– Each variable is differenced once over time, so we are effectively estimating the relationship between changes of variables.

FEM: Estimation – FE or FD?

• When N is large and T is small but greater than 2 (for T=2, FE=FD)
− FE is more efficient when $\varepsilon_{i,t}$ are serially uncorrelated while FD is more efficient when $\varepsilon_{i,t}$ follows a random walk (ρ=1).

• When T is large and N is small
– FD has advantage for processes with large positive autocorrelation. (If ρ is near 1, FD solves the nonstationary problem!)
– FE is more sensitive to nonnormality, heteroskedasticity, and serial correlation in $\varepsilon_{i,t}$.
– On the other hand, FE is less sensitive to violation of the strict exogeneity assumption. Then, FE is preferred when the processes are weakly dependent over time

FEM: Calculation of Var[b|X]

• Since we have assumed strict exogeneity: Cov[$\varepsilon_{i,t}$,($\boldsymbol{x}_{j,t}$, c_j)] = 0, we have OLS in the CLM. That is,

$$\text{Asy.Var}[\mathbf{b}|\mathbf{X}] = (\sigma^2_\varepsilon / \Sigma_{i=1}^N T_i)\text{plim}[(\sigma^2_\varepsilon / \Sigma_{i=1}^N T_i)\Sigma_{i=1}^N \mathbf{X}_i' \mathbf{M}_D^i \mathbf{X}_i]^{-1}$$

which is the usual estimator for OLS

$$\hat{\sigma}^2_\varepsilon = \frac{\Sigma_{i=1}^N \Sigma_{t=1}^{T_i} (y_{it} - a_i - x_{it}'\mathbf{b})^2}{\left(\Sigma_{i=1}^N T_i \ - \ N \ - \ K\right)}$$

(Note the degrees of freedom correction)

PCSE Remark: All previous remarks apply to the FEM.
• We build the SE according to the type of data we have:
- If we do not suspect autocorrelated errors –not a strange situation–, we can rely on clustered White SE's ($\mathbf{S}_0$).
- If we suspect autocorrelated errors, then the Driscoll and Kraay SE should be used.

FEM: Testing for Fixed Effects

• Under H_0 (No FE): $\alpha_i = \alpha$ for all i.

⇒ That is, we test whether to pool or not to pool the data.

• Different tests:

– *F*-test based on the LSDV dummy variable model: constant or zero coefficients for $\mathbf{D}$. Test follows an $F_{N-1,NT-N-k}$ distribution.

– *F*-test based on FEM (the unrestricted model) vs. pooled model (the restricted model). Test follows an $F_{N-1,NT-N-k}$ distribution.

– A LR can also be done –usually, assuming normality. Test follows a χ^2_{N-1} distribution.

FEM: Hypothesis Testing

• Based on estimated residuals of the fixed effects model.

(1) Estimate FEM:

$$y_{i,t} = \boldsymbol{x}_{i,t}'\beta + \alpha_i + \varepsilon_{i,t}, \qquad \Rightarrow \text{Keep residuals } e_{FE,i,t}$$

(2) Tests as usual:

– Heteroscedasticity

  • Breusch and Pagan (1980)

– Autocorrelation: AR(1)

  • Breusch and Godfrey (1981)

$$LM = \frac{NT^2}{T-1}\left(\frac{e_{FE}'e_{FE-1}}{e_{FE}'e_{FE}}\right) \xrightarrow{d} \chi^2_1$$

Application: Cornwell and Rupert Data (Greene)

Cornwell and Rupert Returns to Schooling Data, 595 Individuals, 7 Years
Variables in the file are: (Not used in regressions)

EXP = work experience, EXPSQ = EXP^2
WKS = weeks worked
OCC = occupation, 1 if blue collar,
(IND = 1 if manufacturing industry)
(SOUTH = 1 if resides in south)
SMSA = 1 if resides in a city (SMSA)
MS = 1 if married
FEM = 1 if female
UNION = 1 if wage set by unioin contract
ED = years of education
(BLK = 1 if individual is black)
LWAGE = log of wage = dependent variable in regressions (Y)

These data were analyzed in Cornwell, C. and Rupert, P., "Efficient Estimation with Panel Data: An Empirical Comparison of Instrumental Variable Estimators," Journal of Applied Econometrics, 3, 1988, pp. 149-155.

Application: Cornwell and Rupert (Greene)

(1) Returns to Schooling – Pooled OLS Results

```
+----------------------------------------------------+
| Panel Data Analysis of LWAGE      [ONE way]         |
|           Unconditional ANOVA (No regressors)       |
| Source     Variation   Deg. Free.     Mean Square   |
| Between     646.254        594.       1.08797       |
| Residual    240.651       3570.       .674093E-01   |
| Total       886.905       4164.       .212994       |
+----------------------------------------------------+

+----------------------------------------------------+
| OLS Without Group Dummy Variables                   |
| LHS=LWAGE    Mean                  =    6.676346    |
|              Standard deviation    =    .4615122    |
| Model size   Parameters            =           5    |
|              Degrees of freedom    =        4160    |
| Residuals    Sum of squares        =    651.7870    |
|              Standard error of e   =    .3958277    |
| Fit          R-squared             =    .2650993    |
|              Adjusted R-squared    =    .2643927    |
| Model test   F[  4,  4160] (prob) = 375.16 (.0000) |
+----------------------------------------------------+
+--------+--------------+----------------+--------+--------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+--------+--------------+----------------+--------+--------+----------+
 OCC         -.29227536      .01259221    -23.211   .0000     .51116447
 SMSA         .17712491      .01327104     13.347   .0000     .65378151
 MS           .35695474      .01610229     22.168   .0000     .81440576
 EXP          .00746892      .00057035     13.095   .0000   19.8537815
 Constant    6.27095389      .02041864    307.119   .0000
```

K

RSS & R^2 X only

Application: Cornwell and Rupert (Greene)

(2) Returns to Schooling – LSDV Results

```
+----------------------------------------------------+
| Least Squares with Group Dummy Variables           |
| LHS=LWAGE     Mean                  =   6.676346   |
|               Standard deviation    =   .4615122   |
| Model size    Parameters            =        599   |
|               Degrees of freedom    =       3566   |
| Residuals     Sum of squares        =   83.88505   |
|               Standard error of e   =   .1533740   |
| Fit           R-squared             =   .9054182   |
|               Adjusted R-squared    =   .8895573   |
| Model test    F[598,  3566] (prob) =  57.08 (.0000) |
+----------------------------------------------------+
+----------------------------------------------------+
| Panel:Groups    Empty       0,   Valid data    595 |
|                 Smallest    7,   Largest         7 |
|                 Average group size            7.00 |
+----------------------------------------------------+
```

N+K

RSS & R^2 X and group effects

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
OCC	-.02021384	.01374007	-1.471	.1412	.51116447
SMSA	-.04250645	.01950085	-2.180	.0293	.65378151
MS	-.02946444	.01913652	-1.540	.1236	.81440576
EXP	.09665711	.00119162	81.114	.0000	19.8537815

FEM: Testing for FE (and other formulations)

```
+---------------------------------------------------------------------+
|                Test Statistics for the Classical Model              |
|                                                                     |
|        Model              Log-Likelihood    Sum of Squares  R-squared |
| (1)  Constant term only    -2688.80597   .8869049390D+03  .0000000  |
| (2)  Group effects only       27.58464   .2406511943D+03  .7286618  |
| (3)  X - variables only    -2047.35445   .6517870323D+03  .2650993  |
| (4)  X and group effects    2222.33376   .8388505089D+02  .9054182  |
|                                                                     |
|                          Hypothesis Tests                           |
|               Likelihood Ratio Test              F Tests             |
|           Chi-squared   d.f.   Prob.         F    num.  denom.  Prob value |
| (2) vs (1)  5432.781    594    .00000    16.140   594   3570     .00000 |
| (3) vs (1)  1282.903      4    .00000   375.157     4   4160     .00000 |
| (4) vs (1)  9822.279    598    .00000    57.085   598   3566     .00000 |
| (4) vs (2)  4389.498      4    .00000  1666.054     4   3566     .00000 |
| (4) vs (3)  8539.376    594    .00000    40.643   594   3566     .00000 |
+---------------------------------------------------------------------+
```

Pooled

FEM

- <u>Calculations</u>:

  $\text{F-test}_{594,3566} = [(651.78 - 83.89)/594]/[83.89/3566] = 40.64$ (reject H_0)

The Random Effects Model (REM)

• Recall the general DGP:

$y_{i,t} = \boldsymbol{x}_{i,t}{}'\,\beta + \boldsymbol{z}_i{}'\gamma + \varepsilon_{i,t}$ –observation for individual i at time t.

When the observed characteristics are constant for each individual, a FEM is not an effective tool because such variables cannot be included.

• An alternative approach, known as a random effects (REM) model that, subject to two conditions, provides a solution to this problem.

• Conditions:

(1) It is possible to treat each of the unobserved Z_p variables as being drawn randomly from a given distribution.

(2) The Z_p variables are distributed independently of all of the $\boldsymbol{X}_i$ variables. $\Rightarrow$ E[$\boldsymbol{z}_i{}'\boldsymbol{X}_j$] = 0.

The Random Effects Model (REM)

• Conditions:

(1) Randomly drawn unobserved Z_p variables.

$\Rightarrow$ the c_i may be treated as RV (thus, the name of this approach) drawn from a given distribution. Let's call it u_i. Then,

$$y_{i,t} = \beta_1 + \sum_{j=2}^{k} \beta_j\, \boldsymbol{x}_{ij,t} + u_i + \delta t + \varepsilon_{i,t}$$

$$= \beta_1 + \sum_{j=2}^{k} \beta_j\, \boldsymbol{x}_{ij,t} + \delta t + w_{i,t} \qquad w_{i,t} = u_i + \varepsilon_{i,t}$$

• We deal with the unobserved effect by subsuming it into a compound disturbance term, $w_{i,t}$. We assume that $u_i \sim$ D(0, σ_u^2). Then,

$$E[w_{i,t}] = \text{E}[u_i] + E[\varepsilon_{i,t}] = 0$$

The zero mean assumption –$E[u_i] = 0$– is not crucial, any nonzero component is being absorbed by the intercept, β_1.

The Random Effects Model (REM)

(2) Z_p is independently of all of the X_j variables.

⇒ Otherwise, u_i (& $w_{i,t}$) will not be uncorrelated with $\boldsymbol{X}_j$. The RE estimation will be biased and inconsistent.

Note: We would have to use the FEM, even if the first condition seems to be satisfied.

• If (1) and (2) are satisfied, we can use the REM, and OLS will work, but there is a complication: $w_{i,t}$ is heteroscedastic.

REM: Error Components Model

• REM Assumptions:

$$y_{i,t} = \boldsymbol{x}_{i,t}'\,\beta + c_i + \varepsilon_{i,t} = \boldsymbol{x}_{i,t}'\,\beta + u_i + \varepsilon_{i,t} = \boldsymbol{x}_{i,t}'\,\beta + w_{i,t}$$

$E[\varepsilon_{i,t}\,|\,\boldsymbol{X}_i] = 0$

$E[\varepsilon_{i,t}^2\,|\,\boldsymbol{X}_i] = \sigma_\varepsilon^2$

$E[u_i\,|\,\boldsymbol{X}_i] = 0$

$E[u_i^2\,|\,\boldsymbol{X}_i] = \sigma_u^2$

$E[u_i\,\varepsilon_{j,t}\,|\,\boldsymbol{X}_i] = E[u_i\varepsilon_{jt}\,|\,\boldsymbol{X}_i] = 0$ – u & ε are independent.

$E[u_i\,u_j\,|\,\boldsymbol{X}_i] = 0$ $(i \neq j)$ – no cross-correlation of RE.

$E[\varepsilon_{i,t}\,\varepsilon_{j,t}\,|\,\boldsymbol{X}_i] = 0$ $(i \neq j)$ – no cross-correlation for errors, $\varepsilon_{i,t}$.

$E[\varepsilon_{i,t}\,\varepsilon_{i,s}\,|\,\boldsymbol{X}_i] = 0$ $(t \neq s)$ – there is no autocorrelation for $\varepsilon_{i,t}$.

$$\sigma_{w_{it}}^2 = \sigma_{u_i+\varepsilon_{it}}^2 = \sigma_{u_i}^2 + \sigma_{\varepsilon_{it}}^2 + 2\sigma_{u_i,\varepsilon_{it}} = \sigma_u^2 + \sigma_\varepsilon^2$$

$$\sigma_{w_{it_1}w_{it_2}} = \sigma_{(u_i+\varepsilon_{it_1})(u_i+\varepsilon_{it_2})} = \sigma_u^2$$

REM: Notation (Greene)

$$\begin{bmatrix} \mathbf{y_1} \\ \mathbf{y_2} \\ \vdots \\ \mathbf{y_N} \end{bmatrix} = \begin{bmatrix} \mathbf{X_1} \\ \mathbf{X_2} \\ \vdots \\ \mathbf{X_N} \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_N \end{bmatrix} + \begin{bmatrix} u_1\mathbf{i_1} \\ u_2\mathbf{i_2} \\ \vdots \\ u_N\mathbf{i_N} \end{bmatrix} \quad \begin{matrix} T_1 \text{ observations} \\ T_2 \text{ observations} \\ \vdots \\ T_N \text{ observations} \end{matrix}$$

$$= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} + \mathbf{u} \quad \Sigma_{\mathbf{i=1}}^{\mathbf{N}} T_i \text{ observations}$$

$$= \mathbf{X}\boldsymbol{\beta} + \mathbf{w}$$

In all that follows, except where explicitly noted, X, $\mathbf{X}_i$ and $\mathbf{x}'_{it}$ contain a constant term as the first element. To avoid notational clutter, in those cases, $\mathbf{x}'_{it}$ etc. will simply denote the counterpart without the constant term. Use of the symbol K for the number of variables will thus be context specific but will usually include the constant term.

REM: Notation (Greene)

$$\text{Var}[\boldsymbol{\varepsilon}_i + u_i\mathbf{i}] = \begin{bmatrix} \sigma_\varepsilon^2 + \sigma_u^2 & \sigma_u^2 & \cdots & \sigma_u^2 \\ \sigma_u^2 & \sigma_\varepsilon^2 + \sigma_u^2 & \cdots & \sigma_u^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_u^2 & \sigma_u^2 & \cdots & \sigma_\varepsilon^2 + \sigma_u^2 \end{bmatrix}$$

$$= \sigma_\varepsilon^2 \mathbf{I}_{\mathbf{T_i}} + \sigma_u^2 \mathbf{ii}' \quad T_i \times T_i$$

$$= \sigma_\varepsilon^2 \mathbf{I}_{\mathbf{T_i}} + \sigma_u^2 \mathbf{ii}'$$

$$= \boldsymbol{\Omega}_i$$

$$\text{Var}[\mathbf{w} \mid \mathbf{X}] = \begin{bmatrix} \boldsymbol{\Omega}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Omega}_N \end{bmatrix} \text{(Note these differ only in the dimension } T_i\text{)}$$

• Note: If $E[\varepsilon_{i,t}\, \varepsilon_{j,t} \mid \boldsymbol{X}_i] = 0$ $(i \neq j)$ or $E[\varepsilon_{i,t}\, \varepsilon_{i,s} \mid \boldsymbol{X}_i] = 0$ $(t \neq s)$, we no longer have this nice diagonal-type structure for Var[**w** | **X**].

REM: Assumptions - Convergence of Moments

$$\frac{\mathbf{X'X}}{\Sigma_{i=1}^{N} T} = \Sigma_{i=1}^{N} f_i \frac{\mathbf{X_i'X_i}}{T_i} = \text{a weighted sum of individual moment matrices}$$

$$\frac{\mathbf{X'\Omega X}}{\Sigma_{i=1}^{N} T} = \Sigma_{i=1}^{N} f_i \frac{\mathbf{X_i'\Omega_i X_i}}{T_i} = \text{a weighted sum of individual moment matrices}$$

$$= \sigma_\varepsilon^2 \Sigma_{i=1}^{N} f_i \frac{\mathbf{X_i'X_i}}{T_i} + \sigma_u^2 \Sigma_{i=1}^{N} f_i \overline{\mathbf{x}}_i \overline{\mathbf{x}}_i'$$

Note asymptotics are with respect to N. Each matrix $\frac{\mathbf{X_i'X_i}}{T_i}$ is the moments for the T_i observations. Should be 'well behaved' in micro level data. The average of N such matrices should be likewise.
T or T_i is assumed to be fixed (and small).

REM: Pooled OLS Estimation (Greene)

- Standard results for the pooled OLS estimator **b** in the GR model
  - Consistent and asymptotic normal
  - Unbiased
  - Inefficient

- We can use pooled OLS, but for inferences we need the true variance –i.e., the sandwich estimator:

$$\text{Var}[\mathbf{b}\,|\,\mathbf{X}] = \frac{1}{\Sigma_{i=1}^{N} T_i} \left[\frac{\mathbf{X'X}}{\Sigma_{i=1}^{N} T_i}\right]^{-1} \frac{\mathbf{X'\Omega X}}{\Sigma_{i=1}^{N} T_i} \left[\frac{\mathbf{X'X}}{\Sigma_{i=1}^{N} T_i}\right]^{-1}$$

$$\rightarrow \mathbf{0} \times \rightarrow \mathbf{Q}^{-1} \times \rightarrow \mathbf{Q}^* \times \rightarrow \mathbf{Q}^{-1}$$

$$\rightarrow \mathbf{0} \text{ as } N \rightarrow \infty \text{ with our convergence assumptions}$$

REM: Sandwich Estimator for OLS (Greene)

$$\text{Var}[\mathbf{b} \mid \mathbf{X}] = \frac{\mathbf{1}}{\Sigma_{i=1}^{N} T_i}\left[\frac{\mathbf{X}'\mathbf{X}}{\Sigma_{i=1}^{N} T_i}\right]^{-1}\left(\frac{\mathbf{X}'\mathbf{\Omega}\mathbf{X}}{\Sigma_{i=1}^{N} T_i}\right)\left[\frac{\mathbf{X}'\mathbf{X}}{\Sigma_{i=1}^{N} T_i}\right]^{-1}$$

$$\frac{\mathbf{X}'\mathbf{\Omega}\mathbf{X}}{\Sigma_{i=1}^{N} T} = \Sigma_{i=1}^{N} f_i \frac{\mathbf{X}_i'\mathbf{\Omega}_i\mathbf{X}_i}{T_i}, \text{ where } = \mathbf{\Omega}_i = E[\mathbf{w}_i\mathbf{w}_i' \mid \mathbf{X}_i]$$

In the spirit of the White estimator, use

$$\frac{\mathbf{X}'\mathbf{\Omega}\mathbf{X}}{\Sigma_{i=1}^{N} T} = \Sigma_{i=1}^{N} f_i \frac{\mathbf{X}_i'\hat{\mathbf{w}}_i\hat{\mathbf{w}}_i'\mathbf{X}_i}{T_i}, \quad \hat{\mathbf{w}}_i = \mathbf{y}_i - \mathbf{X}_i\mathbf{b}$$

Hypothesis tests are then based on Wald statistics.

THIS IS THE 'CLUSTER' ESTIMATOR

• Recall: *Clustered standard errors* or PCSE

There is a grouping, or "cluster," within which the error term is possibly correlated, but outside of which (across groups) it is not.

REM: Sandwich Estimator – Mechanics (Greene)

$$\text{Est.Var}[\mathbf{b} \mid \mathbf{X}] = [\mathbf{X}'\mathbf{X}]^{-1}\left(\Sigma_{i=1}^{N}\mathbf{X}_i'\hat{\mathbf{w}}_i\hat{\mathbf{w}}_i'\mathbf{X}_i\right)[\mathbf{X}'\mathbf{X}]^{-1}$$

$\hat{\mathbf{w}}_i'$= set of T_i OLS residuals for individual i.

$\mathbf{X}_i$ = T_ixK data on exogenous variable for individual i.

$\mathbf{X}_i'\hat{\mathbf{w}}_i$ = K x 1 vector of products

$(\mathbf{X}_i'\hat{\mathbf{w}}_i)(\hat{\mathbf{w}}_i'\mathbf{X}_i)$ = KxK matrix (rank 1, outer product)

$\left(\Sigma_{i=1}^{N}(\mathbf{X}_i'\hat{\mathbf{w}}_i)(\hat{\mathbf{w}}_i'\mathbf{X}_i)\right)$ = sum of N rank 1 matrices. Rank $\leq$ K.

We could compute this as $\left(\Sigma_{i=1}^{N}\mathbf{X}_i'\left(\hat{\mathbf{w}}_i\hat{\mathbf{w}}_i'\right)\mathbf{X}_i\right) = \left(\Sigma_{i=1}^{N}\mathbf{X}_i'\left(\hat{\mathbf{\Omega}}_i\right)\mathbf{X}_i\right)$.

Why not do it that way?

REM: GLS

- Standard results for GLS in a GR model
  - Consistent
  - Unbiased
  - Efficient (if functional form for **Ω** correct)

$$\hat{\boldsymbol{\beta}} = [\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}]^{-1}[\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y}]$$
$$= [\Sigma_{i=1}^{N}\mathbf{X}_i'\boldsymbol{\Omega}_i^{-1}\mathbf{X}_i]^{-1}[\Sigma_{i=1}^{N}\mathbf{X}_i'\boldsymbol{\Omega}_i^{-1}\mathbf{y}_i]$$
$$\boldsymbol{\Omega}_i^{-1} = \frac{1}{\sigma_\varepsilon^2}\left[\mathrm{I}_{T_i} - \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T_i\sigma_u^2}\mathbf{ii}'\right]$$

(note, depends on i only through T_i)

- As usual, the matrix $\boldsymbol{\Omega}^{-1/2} = \mathbf{P}$ will be used to transform the data.

REM: GLS

- The matrix $\boldsymbol{\Omega}^{-1/2} = \mathbf{P}$ is used to transform the data. That is,

$$y_{it} - \theta\bar{y}_i = (x_{it} - \theta\bar{x}_i)\beta + \nu_{it}$$
$$\text{where} \quad \theta_i = 1 - \sqrt{\sigma_\varepsilon^2/(\sigma_\varepsilon^2 + T_i\sigma_u^2)}$$
$$Asy.Var[\hat{\beta}_{GLS}] = (X'\Omega^{-1}X)^{-1} = \sigma_\varepsilon^2(X^{*\prime}X^*)^{-1}$$

- We call the transformed data: *quasi time-demeaned data*. As expected, GLS is just pooled OLS with the transformed data.

Note: The RE can be seen as mixture of two estimators:
- when $\theta = 0$ ($\sigma_u = 0$) ⇒ Pooled OLS estimator
- when $\theta = 1$ ($\sigma_\varepsilon = 0$ or $\sigma_u \to \infty$) ⇒ LSDV estimator (u_i's become the FE)

Then, the bigger (smaller) the variance of the unobserved effect –i.e., individual heterogeneity is bigger–, the closer it is to FE (pooled OLS). Also, when *T* is large, it becomes more like FE.

REM: FGLS - Estimators for the Variances

• To transform the data, we need to estimate σ_ε^2 and σ_u^2, consistently.

• Usual steps (assume a balanced panel):

(1) Start with a consistent estimator of β. For example, pooled OLS, **b**.

(2) Compute $\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{i,t} - \boldsymbol{x}_{i,t}'\,\mathbf{b})^2$ –estimates $\sum_{i=1}^{N}\sum_{t=1}^{T}(\sigma_\varepsilon^2 + \sigma_u^2)$

(3) Divide by a function of NT. For example: $NT - K - 1$

$\Rightarrow$ We estimate σ^2, $s_{pooled}^2 = \boldsymbol{e}_{pooled}{}'\boldsymbol{e}_{pooled}/(NT - K - 1)$

We will use s_{pooled}^2 to estimate the sum: $\sigma_\varepsilon^2 + \sigma_u^2$

(4) Use LSDV estimation to get a_i and $\boldsymbol{b}_{LSDV}$. Keep residuals, $e_{FE,i,t}$.

(5) Compute $\Sigma_i\, \Sigma_t\, (y_{i,t} - a_i - \boldsymbol{x}_{i,t}'\boldsymbol{b}_{LSDV})^2$ – estimates $\sum_{i=1}^{N}\sum_{t=1}^{T}(\sigma_\varepsilon^2)$

(6) To estimate σ_ε^2, divide by $NT - K - N$:

$$s_\varepsilon^2 = \sum_{i=1}^{N}\sum_{t=1}^{T}\left(e_{FE,i,t}\right)^2/(NT - K - N)$$

(7) Estimate σ_u^2 as $s_u^2 = s_{pooled}^2 - s_\varepsilon^2$

REM: FGLS - Estimators for the Variances

Feasible GLS requires (only) consistent estimators of σ_ε^2 and σ_u^2.

Candidates:

From the robust LSDV estimator: $\hat{\sigma}_\varepsilon^2 = \dfrac{\Sigma_{i=1}^{N}\Sigma_{t=1}^{T_i}(y_{it} - a_i - \mathbf{x}_{it}'\mathbf{b}_{LSDV})^2}{\Sigma_{i=1}^{N}T_i - K - N}$

From the pooled OLS estimator: $\widehat{\sigma_\varepsilon^2 + \sigma_u^2} = \dfrac{\Sigma_{i=1}^{N}\Sigma_{t=1}^{T_i}(y_{it} - a_{OLS} - \mathbf{x}_{it}'\mathbf{b}_{OLS})^2}{\Sigma_{i=1}^{N}T_i - K - 1}$

From the group means regression: $\widehat{\sigma_\varepsilon^2 / \overline{T} + \sigma_u^2} = \dfrac{\Sigma_{i=1}^{N}(\overline{y}_{it} - \tilde{a} - \overline{\mathbf{x}}_i'\tilde{\mathbf{b}}_{MEANS})^2}{N - K - 1}$

(Wooldridge) Based on $E[w_{it}w_{is} \mid \mathbf{X}_i] = \sigma_u^2$ if $t \neq s$, $\hat{\sigma}_u^2 = \dfrac{\Sigma_{i=1}^{N}\Sigma_{t=1}^{T_i-1}\Sigma_{s=t+1}^{T_i}\hat{w}_{it}\hat{w}_{is}}{\Sigma_{i=1}^{N}T_i - K - N}$

There are many others.

Note: A slight chance in notation, $\boldsymbol{x}_{i,t}$ does not contain the constant term.

REM: Practical Problems with FGLS

All of the preceding regularly produce negative estimates of σ_u^2.
Estimation is made very complicated in unbalanced panels.
A bulletproof solution (originally used in TSP, now LIMDEP and others).

From the robust LSDV estimator: $\hat{\sigma}_\varepsilon^2 = \dfrac{\Sigma_{i=1}^N \Sigma_{t=1}^{T_i} (y_{it} - a_i - \mathbf{x}_{it}'\mathbf{b}_{LSDV})^2}{\Sigma_{i=1}^N T_i}$

From the pooled OLS estimator: $\sigma_\varepsilon^2 + \sigma_u^2 = \dfrac{\Sigma_{i=1}^N \Sigma_{t=1}^{T_i} (y_{it} - a_{OLS} - \mathbf{x}_{it}'\mathbf{b}_{OLS})^2}{\Sigma_{i=1}^N T_i} \geq \hat{\sigma}_\varepsilon^2$

$$\hat{\sigma}_u^2 = \frac{\Sigma_{i=1}^N \Sigma_{t=1}^{T_i} (y_{it} - a_{OLS} - \mathbf{x}_{it}'\mathbf{b}_{OLS})^2 - \Sigma_{i=1}^N \Sigma_{t=1}^{T_i} (y_{it} - a_i - \mathbf{x}_{it}'\mathbf{b}_{LSDV})^2}{\Sigma_{i=1}^N T_i} \geq 0$$

• Bullet proof solution: Do not correct by degrees of freedom. Then, given that the unrestricted RSS (LSDV) will be lower than the restricted (pooled OLS) RSS, σ_u^2 will be positive!

Application: Fixed Effects Estimates (Greene)

```
--------------------------------------------------------------------
Least Squares with Group Dummy Variables..........
LHS=LWAGE     Mean                  =        6.67635
Residuals     Sum of squares        =       82.34912
              Standard error of e   =         .15205
These  2 variables have no within group variation.
FEM      ED
F.E. estimates are based on a generalized inverse.
--------+-----------------------------------------------------------
Variable| Coefficient    Standard Error  b/St.Er. P[|Z|>z]   Mean of X
--------+-----------------------------------------------------------
     EXP|     .11346***       .00247       45.982   .0000      19.8538
   EXPSQ|    -.00042***    .544864D-04     -7.789   .0000      514.405
     OCC|    -.02106          .01373       -1.534   .1251       .51116
    SMSA|    -.04209**        .01934       -2.177   .0295       .65378
      MS|    -.02915          .01897       -1.536   .1245       .81441
     FEM|      .000        ......(Fixed Parameter).......
   UNION|     .03413**        .01491        2.290   .0220       .36399
      ED|      .000        ......(Fixed Parameter).......
--------+-----------------------------------------------------------
```

REM: Computing Variance Estimators (Greene)

Using full list of variables (FEM and ED are time invariant)

OLS sum of squares = 522.2008.

$\hat{\sigma}_\varepsilon^2 + \sigma_u^2$ = 522.2008 / (4165 - 9) = 0.12565.

Using full list of variables and a generalized inverse (same as dropping FEM and ED), LSDV sum of squares = 82.34912.

$\hat{\sigma}_\varepsilon^2$ = 82.34912 / (4165 - 8-595) = 0.023119.

$\hat{\sigma}_u^2$ = 0.12565 - 0.023119 = 0.10253

Both estimators are positive. We stop here. If $\hat{\sigma}_u^2$ were negative, we would use estimators without DF corrections.

REM: Application (Greene)

```
----------------------------------------------------------------------
Random Effects Model: v(i,t)     = e(i,t) + u(i)
Estimates:  Var[e]               =        .023119
            Var[u]               =        .102531
            Corr[v(i,t),v(i,s)] =         .816006
Lagrange Multiplier Test vs. Model (3) =3713.07
( 1 degrees of freedom, prob. value =  .000000)
(High values of LM favor FEM/REM over CR model)
Fixed vs. Random Effects (Hausman)     =    .00  (Cannot be computed)
( 8 degrees of freedom, prob. value = 1.000000)
(High (low) values of H favor F.E.(R.E.) model)
            Sum of Squares           1411.241136
            R-squared                 -.591198
```

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
EXP	.08819204	.00224823	39.227	.0000	19.8537815
EXPSQ	-.00076604	.496074D-04	-15.442	.0000	514.405042
OCC	-.04243576	.01298466	-3.268	.0011	.51116447
SMSA	-.03404260	.01620508	-2.101	.0357	.65378151
MS	-.06708159	.01794516	-3.738	.0002	.81440576
FEM	-.34346104	.04536453	-7.571	.0000	.11260504
UNION	.05752770	.01350031	4.261	.0000	.36398559
ED	.11028379	.00510008	21.624	.0000	12.8453782
Constant	4.01913257	.07724830	52.029	.0000	

Testing for Random Effects: LM Test

• We want to test for RE. That is,

$$H_0: \sigma_u^2 = 0.$$

• We can use the Breusch-Pagan (1980) Test for RE effects. Similar to the LM-BP test for autocorrelation, it is based on the pooled OLS residuals, $\boldsymbol{e}_i$. It is easy to compute – distributed as χ_1^2:

Breusch and Pagan Lagrange Multiplier statistic
Assuming normality (and for convenience now, a balanced panel)

$$LM = \frac{NT}{2(T-1)}\left[\frac{\Sigma_{i=1}^{N}(T\bar{e}_i^2)}{\Sigma_{i=1}^{N}\Sigma_{i=1}^{N}e_{it}^2} - 1\right]^2 = \frac{NT}{2(T-1)}\left[\frac{\Sigma_{i=1}^{N}[(T\bar{e}_i^2) - \mathbf{e}_i'\mathbf{e}_i]}{\Sigma_{i=1}^{N}\mathbf{e}_i'\mathbf{e}_i}\right]^2$$

Converges to chi-squared[1] under the null hypothesis of no common effects. (For unbalanced panels, the scale in front becomes $(\Sigma_{i=1}^{N}T_i)^2/[2\Sigma_{i=1}^{N}T_i(T_i-1)]$.)

REM: LM Test Application – Cornwell-Rupert

```
+----------------------------------------------------+
| Ordinary    least squares regression               |
| LHS=LWAGE    Mean                 =   6.676346      |
|              Standard deviation   =   .4615122      |
| Model size   Parameters           =          7      |
|              Degrees of freedom   =       4158      |
| Residuals    Sum of squares       =   556.3030      |
|              Standard error of e  =   .3657745      |
| Fit          R-squared            =   .3727592      |
|              Adjusted R-squared   =   .3718541      |
+----------------------------------------------------+
+---------+--------------+----------------+--------+---------+----------+
|Variable | Coefficient  | Standard Error |b/St.Er.|P[|Z|>z] | Mean of X|
+---------+--------------+----------------+--------+---------+----------+
 Constant     5.66098218      .04685914    120.808   .0000
 FEM          -.39478212      .02603413    -15.164   .0000      .11260504
 ED            .05688005      .00267743     21.244   .0000    12.8453782
 OCC          -.11220205      .01464317     -7.662   .0000      .51116447
 SMSA          .15504405      .01233744     12.567   .0000      .65378151
 MS            .09569050      .02133490      4.485   .0000      .81440576
 EXP           .01043785      .00054206     19.256   .0000    19.8537815
+--------------------------------------------------+
| Random Effects Model: v(i,t) = e(i,t) + u(i)     |
| Estimates:  Var[e]                =   .235368D-01 |
|             Var[u]                =   .110254D+00 |
|             Corr[v(i,t),v(i,s)] =   .824078      |
| Lagrange Multiplier Test vs. Model (3) = 3797.07 |
| ( 1 df, prob value =  .000000)                   |
| (High values of LM favor FEM/REM over CR model.) |
+--------------------------------------------------+
 Constant     4.24669585      .07763394     54.702   .0000
 FEM          -.34715010      .04681514     -7.415   .0000      .11260504
 ED            .11120152      .00525209     21.173   .0000    12.8453782
 OCC          -.03908144      .01298962     -3.009   .0026      .51116447
 SMSA         -.03881553      .01645862     -2.358   .0184      .65378151
 MS           -.06557030      .01815465     -3.612   .0003      .81440576
 EXP           .05737298      .00088467     64.852   .0000    19.8537815
```

Note: Check the different standard errors from both models.

FE vs. RE: Understanding Differences

• Suppose, we want to study the effect of an MBA on stock trading, controlling for other factors such as income and experience. We have a panel, with individuals measured annually over 10 years. We expect some year-to-year correlation with a given individual i (with unobservable individual-level effects accounting for part of the correlation between yearly trading within the same i).

• To understand the difference between FE & RE, we ask:

- Does a regression coefficient for an MBA represent a comparison of two i's, one with an MBA and one without one? ⇒ *Between Effect* (RE).

- Or does it compare two yearly trading records from the same i who happened to receive an MBA in the interim? ⇒ *Within Effect* (FEM).

FE vs. RE: Understanding Differences

• *Between Effect* (REM): GLS ⇒ consistent and efficient (under H_0)

- Between-individual effects and within-individual effects are identical.

- Very efficient, no data is thrown away.

• *Within Effect* (FEM): OLS ⇒ consistent estimates.

- No confounding due to unmeasured i-level characteristics.

- Cost: All the between-individual comparisons in the data are thrown away.

• Q: Are within and between MBA effects the same?

FE vs. RE

• Q: RE estimation or FE estimation?

• Case for RE:

– Under no omitted variables –or if the omitted variables are uncorrelated with $\boldsymbol{x}_{i,t}$ in the model– then a REM is probably best: It produces unbiased and efficient estimates, & uses all the data available.

– RE can deal with observed characteristics that remain constant for each individual. In FE, they have to be dropped from model.

– In contrast with FE, RE estimates a small number of parameters

– We do not lose N degrees of freedom.

– Philosophically speaking, a REM is more attractive: *Why should we assume one set of unobservables fixed and the other random?*

FE vs. RE

• Case against RE:

- If either of the conditions for using RE is violated, we should use FE.

• *Condition (1):* Randomly drawn unobserved Z_p variables.

This is a reasonable assumption in many cases: Many of the panels are designed to be a random sample (for example, NLSY).

But, it would not be a reasonable assumption if the units of observation in the panel data set were data from the S&P 500 firms.

• *Condition (2):* Z_p is independently of all of the $\boldsymbol{x}_j$ variables.

A violation of condition (2) causes *inconsistency* in the RE estimation

FE vs. RE

• FE estimation is always *consistent*. On the other hand, a violation of condition (2) causes *inconsistency* in the RE estimation.

That is, if there are omitted variables, which are correlated with the $\boldsymbol{x}_{i,t}$ in the model, then the FEM provides a way for controlling for omitted variable bias. In a FEM, individuals serve as their own controls.

• Q: How can we tell if condition (2) is violated?

A: A DHW test can help.

DHW (Hausman) Specification Test: FE vs. RE

Estimator	Random Effects $\text{E}[c_i \mid \boldsymbol{x}_{i,t}] = 0$	Fixed Effects $\text{E}[c_i \mid \boldsymbol{x}_{i,t}] \neq 0$
FGLS (Random Effects)	**Consistent** and Efficient	**Inconsistent**
LSDV (Fixed Effects)	**Consistent** Inefficient	**Consistent** Possibly Efficient

• Under an H_0 (RE is true), we have one estimator that is efficient (RE) and one inefficient (LSDV). We can use a Durbin-Hausman-Wu test.

As in its other applications, the DHW test determines whether the estimates of the coefficients, taken as a group, are significantly different in the two regressions.

DHW (Hausman) Specification Test: FE vs. RE

Basis for the test, $\hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{RE}$

Wald Criterion: $\hat{\mathbf{q}} = \hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{RE}$; $W = \hat{\mathbf{q}}'[\text{Var}(\hat{\mathbf{q}})]^{-1}\hat{\mathbf{q}}$

A lemma (Hausman (1978)): Under the null hypothesis (RE)

$$\sqrt{nT}[\hat{\boldsymbol{\beta}}_{RE} - \boldsymbol{\beta}] \xrightarrow{d} N[\mathbf{0}, \mathbf{V}_{RE}] \text{ (efficient)}$$

$$\sqrt{nT}[\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}] \xrightarrow{d} N[\mathbf{0}, \mathbf{V}_{FE}] \text{ (inefficient)}$$

Note: $\hat{\mathbf{q}} = (\hat{\boldsymbol{\beta}}_{FE} - \boldsymbol{\beta}) - (\hat{\boldsymbol{\beta}}_{RE} - \boldsymbol{\beta})$. The lemma states that in the joint limiting distribution of $\sqrt{nT}[\hat{\boldsymbol{\beta}}_{RE} - \boldsymbol{\beta}]$ and $\sqrt{nT}\ \hat{\mathbf{q}}$, the limiting covariance, $\mathbf{C}_{Q,RE}$ is $\mathbf{0}$. But, $\mathbf{C}_{Q,RE} = \mathbf{C}_{FE,RE} - \mathbf{V}_{RE}$. Then, $\text{Var}[\mathbf{q}] = \mathbf{V}_{FE} + \mathbf{V}_{RE} - \mathbf{C}_{FE,RE} - \mathbf{C}'_{FE,RE}$. Using the lemma, $\mathbf{C}_{FE,RE} = \mathbf{V}_{RE}$. It follows that $\text{Var}[\mathbf{q}] = \mathbf{V}_{FE} - \mathbf{V}_{RE}$. Based on the preceding

$H = (\hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{RE})'\ [\text{Est.Var}(\hat{\boldsymbol{\beta}}_{FE}) - \text{Est.Var}(\hat{\boldsymbol{\beta}}_{RE})]^{-1}(\hat{\boldsymbol{\beta}}_{FE} - \hat{\boldsymbol{\beta}}_{RE})$

Note: β does not contain the constant term.

DHW (Hausman) Specification Test: FE vs. RE

• Then following the structure of the DHW test we saw in Lecture 8:

$$H = (\mathbf{b}_{FEM} - \mathbf{b}_{REM})'\ \mathbf{V}^{-1}(\mathbf{b}_{FEM} - \mathbf{b}_{REM})$$

where $\mathbf{V} = \mathbf{V}_{FEM} - \mathbf{V}_{REM}$.

Note: Columns of zeroes will show in $\mathbf{V}_{FEM}$ if there are time invariant variables in $\boldsymbol{x}_{i,t}$. (Also, β does not contain the constant term.)

Computing the DHW Statistic

$$\text{Est.Var}[\hat{\boldsymbol{\beta}}_{\mathbf{FE}}] = \hat{\sigma}_\varepsilon^2 \left[\Sigma_{i=1}^N \mathbf{X}_i' \left(\mathrm{I} - \frac{1}{T_i} \mathbf{ii}' \right) \mathbf{X}_i \right]^{-1}$$

$$\text{Est.Var}[\hat{\boldsymbol{\beta}}_{\mathbf{RE}}] = \hat{\sigma}_\varepsilon^2 \left[\Sigma_{i=1}^N \mathbf{X}_i' \left(\mathrm{I} - \frac{\hat{\gamma}_i}{T_i} \mathbf{ii}' \right) \mathbf{X}_i \right]^{-1}, \; 0 \;\leq\; \hat{\gamma}_i \;=\; \frac{T_i \hat{\sigma}_u^2}{\hat{\sigma}_\varepsilon^2 + T_i \hat{\sigma}_u^2} \;\leq\; 1$$

As long as $\hat{\sigma}_\varepsilon^2$ and $\hat{\sigma}_u^2$ are consistent, as N $\to \infty$, Est.Var[$\hat{\boldsymbol{\beta}}_{\mathbf{FE}}$] – Est.Var[$\hat{\boldsymbol{\beta}}_{\mathbf{RE}}$] will be nonnegative definite. In a finite sample, to ensure this, both must be computed using the same estimate of $\hat{\sigma}_\varepsilon^2$. The one based on LSDV will generally be the better choice.

Note that columns of zeros will appear in Est.Var[$\hat{\boldsymbol{\beta}}_{\mathbf{FE}}$] if there are time invariant variables in **X**.

Note: Pooled OLS is consistent, but inefficient under H_0. Then, the RE estimation is GLS.

DHW Specification Test: Application (Hoechle)

• Bid-Ask Spread Panel estimation.

	—— Coefficients —— (b) FE	(B) REgls	(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
aVol	-.0017974	-.0017916	-5.81e-06	.0000128
Size	-.1875486	-.1603143	-.0272343	.0337314
TRMS2	.0031042	.0031757	-.0000715	.0000239
TRMS	-.0014581	-.001634	.000176	.0001959

```
                    b = consistent under Ho and Ha; obtained from xtreg
     B = inconsistent under Ha, efficient under Ho; obtained from xtreg

Test:  Ho:  difference in coefficients not systematic

            chi2(4) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                    =       11.64
          Prob>chi2 =      0.0203
```

• Rejection at the 5% level, like in this case, indicates that $\boldsymbol{\beta}_{FE} \neq \boldsymbol{\beta}_{RE}$.
- Usually, this result is taken as an indication of a FEM.

DHW Specification Test: Application (Greene)

```
+--------------------------------------------------+
| Random Effects Model: v(i,t) = e(i,t) + u(i)     |
| Estimates:  Var[e]              =   .235236D-01  |
|             Var[u]              =   .133156D+00  |
|             Corr[v(i,t),v(i,s)] =   .849862      |
| Lagrange Multiplier Test vs. Model (3) = 4061.11 |
| ( 1 df, prob value =  .000000)                   |
| (High values of LM favor FEM/REM over CR model.) |
| Fixed vs. Random Effects (Hausman)     = 2632.34 |
| ( 4 df, prob value =  .000000)                   |
| (High (low) values of H favor FEM (REM).)        |
+--------------------------------------------------+
```

• The DHW statistic is used to tests the difference in coefficients between an RE and FE models
- A rejection, like in this case, indicates that $\beta_{FE} \neq \beta_{RE}$
- But, rejecting H_0 does not imply necessarily H_1 is "accepted."
- Either the model is misspecified or u_i and $\mathbf{x}_{it}$ are correlated
- Q: Is the model misspecified (any variable missing)?

Wu (Variable Addition) Test

• Under the FE assumptions, the common unobserved effect is correlated with the group means.

Add the group means to the RE model. If statistically significant, this suggests that the RE model is inappropriate.

• In a panel context, tests based on a regression can be more computationally more stable, since no problems with non-positive definiteness are encountered.

• Since the errors and the unobserved effect may not be *i.i.d. white noise,* Wooldridge (2009) suggests using PCSE.

Mundlak (Augmented) Regression (Greene)

Variable	Coefficient	Standard Error	b/St.Er.	P[|Z|>z]	Mean of X
EXPBAR	**-.08769*** **	**.00162096**	**-54.099**	**.0000**	**19.853782**
OCCBAR	**-.14806*** **	**.03623348**	**-4.086**	**.0000**	**.5111645**
SMSABAR	**.21707*** **	**.03209640**	**6.763**	**.0000**	**.6537815**
MSBAR	**.14855*** **	**.05087686**	**2.920**	**.0035**	**.8144058**
UNYNBAR	**.07831** **	**.03257465**	**2.404**	**.0162**	**.3639856**
WKSBAR	**.00857** **	**.00362039**	**2.367**	**.0179**	**46.811525**
INDBAR	**.03998**	**.02966215**	**1.348**	**.1777**	**.3954382**
SOUTHBAR	**-.05487**	**.04293224**	**-1.278**	**.2012**	**.2902761**
EXP	.11448***	.00225862	50.684	.0000	19.853782
EXPSQ	-.00045***	.483957D-04	-9.304	.0000	514.40504
OCC	-.02122	.01380348	-1.537	.1243	.5111645
SMSA	-.04237**	.01945829	-2.178	.0294	.6537815
MS	-.02969	.01901293	-1.561	.1184	.8144058
FEM	-.31359***	.05419945	-5.786	.0000	.1126050
UNION	.03268**	.01494574	2.187	.0288	.3639856
ED	.05150***	.00550816	9.349	.0000	12.845378
BLK	-.15768***	.04463738	-3.533	.0004	.0722689
WKS	.00081	.00060031	1.354	.1759	46.811525
IND	.01909	.01546993	1.234	.2171	.3954382
SOUTH	-.00176	.03435229	-.051	.9592	.2902761
Constant	5.15038***	.20122987	25.595	.0000	

Wu Test: Application 1 (Greene)

```
--> matr;bm=b(1:8);vm=varb(1:8,1:8)$
--> matr;list;wutest=bm'<vm>bm$

Matrix WUTEST   has  1 rows and  1 columns.
           1
       +--------------
      1|  3006.13788
--> calc;list;ctb(.95,8)$
+------------------------------------+
| Listed Calculator Results          |
+------------------------------------+
 Result  =      15.507313
```

Wu Test: Application 2 (Hoechle)

• Bid-Ask Spread Wu test estimation with PCSE's. Stata code:

```
. qui xtreg BA aVol Size TRMS2 TRMS, re
. scalar lambda_hat = 1 - sqrt(e(sigma_e)^2/(e(g_avg)*e(sigma_u)^2+e(sigma_e)^2))
. gen in_sample = e(sample)
. sort ID TDate
. qui foreach var of varlist BA aVol Size TRMS2 TRMS   {
.     by ID: egen `var'_bar = mean(`var') if in_sample
.     gen `var'_re = `var' - lambda_hat*`var'_bar if in_sample  // GLS-transform
.     gen `var'_fe = `var' - `var'_bar if in_sample             // within-transform
.     }
. * Wooldridge's auxiliary regression for the panel-robust Hausman test:
. reg BA_re aVol_re Size_re TRMS2_re TRMS_re aVol_fe Size_fe TRMS2_fe
> TRMS_fe if in_sample, cluster(ID)
  (output omitted)

. * Test of the null-hypothesis ``gamma==0'':
. test aVol_fe Size_fe TRMS2_fe TRMS_fe
 ( 1)  aVol_fe = 0
 ( 2)  Size_fe = 0
 ( 3)  TRMS2_fe = 0
 ( 4)  TRMS_fe = 0
       F(  4,   218) =    2.40
            Prob > F =    0.0510
```

Wu Test: Application 2 (Hoechle)

• Bid-Ask Spread Wu test estimation with Driscoll and Kraay SE's. Stata code for auxiliary regression:

```
. xtscc BA_re aVol_re Size_re TRMS2_re TRMS_re aVol_fe Size_fe TRMS2_fe TRMS_fe
> if in_sample, lag(8)
  (output omitted)

. test aVol_fe Size_fe TRMS2_fe TRMS_fe
 ( 1)  aVol_fe = 0
 ( 2)  Size_fe = 0
 ( 3)  TRMS2_fe = 0
 ( 4)  TRMS_fe = 0
       F(  4,   218) =    1.65
            Prob > F =    0.1632
```

• Now, you cannot reject the REM at the 5% level. Here you can say, "after accounting for cross-sectional and temporal dependence, the Hausman test indicates that the coefficient estimates from pooled OLS estimation are consistent."

• Different PCSE's can give different results.

DHW Specification Test: Remarks

• Issues with Hausman tests –as discussed in Wooldridge (2009):

(1) Fail to reject means either:

- FE and RE are similar -i.e., this is great!
- FE estimates are very imprecise
  - Large differences from RE are nevertheless insignificant
  - That can happen if the data are awful/noisy. Be careful.

(2) Watch for difference between "*statistical significance*" and "*practical significance*."

- With a huge sample, the Hausman test may "reject" even though RE is nearly the same as FE
- If differences are tiny, you can feel comfortable using the REM.

(3) PCSE's matter $\Rightarrow$ Q: Which ones to use?

Allison's Hybrid Approach

- Allison (2009) suggests a 'hybrid' approach that provides the benefits of FE and RE
  – Also discussed in Gelman & Hill (2007) textbook
  – Builds on idea of decomposing X into mean, deviation

Steps:

– 1. Compute case-specific mean variables
– 2. Transform X variables into deviations (*within transformation)*
– 3. Do not transform the dependent variable Y
– 4. Include both X deviation & X mean variables
– 5. Estimate with a RE model

Allison's Hybrid Approach

- Benefits of hybrid approach:
  - 1. Effects of "X-deviation" variables (*within effects*) are equivalent to results from a FEM.
    - All time-constant factors are controlled
  - 2. Effects of time-constant X variables (*between effects*)
  - 3. You can build a general multilevel model
    - Random slope coefficients; more than 2 level models…
  - 4. You can directly test FE vs RE
      - No Hausman test needed
    - REM: X-mean and X-deviation coefficients should be equal
    - Conduct a Wald test for equality of coefficients
      - Also differing X-mean & X-deviation coefficents are informative.

Measurement Error

- It can have a severe effect on panel data models.
- It is no longer obvious that a panel data estimator is preferred to a cross-section estimator.
- Measurement error often leads to "attenuation" of signal to noise ratio in panels – biases coefficients towards zero.

Heteroskedasticity - Review

• Given that there is a cross-section component to panel data, there will always be a potential for heteroskedasticity.

• Although there are various tests for heteroskedastcity, as with autocorrelation there is a tendency to automatically use NW's PCSE, which removes the problem. This is fine for the FEM. But not for the REM: REM tell us the structure of heteroscedasitcity => GLS!

• Baltagi (1995) allows u_i to be heteroscedastic. There is an efficiency problem, however, since only one observation --the estimated error u_i, repeated T_i per individual-- will be used to estimate $s^2_u(i)$.

Autocorrelation - Review

• Although different to autocorrelation using the usual univariate models, a version of the Breusch-Pagan LM test can be used.

• To deal with autocorrelated errors, we can use the usual methods, say pseudo-differencing. In general, we will estimate ρ using the LSDV residuals.

• If we allow ρ_i to vary with i, we lose power (in general, T is small).

• We can also reformulate the model, by building a 'Dynamic Model,' which basically involves adding a lagged dependent variable.

• As usual, OLS plus NW's PCSE can help you to avoid a complicated FGLS estimation. (The usual problems with HAC SE apply.)

PCSE - Review

- Key Assumption
  - Correlations within a cluster (a group of firms, a region, different years for the same firm, different years for the same region) are the same for different observations.
- Procedure
  - (1) Identify clusters using economic theory (industry, year, etc.)
  - (2) Calculate clustered standard errors
  - (3) Try different ways of defining clusters and see how the estimated SE are affected. Be conservative, report largest SE.
- Performance
  - Not a lot of studies –some simulations done for simple DGPs.
  - PCSE's coverage rates are not very good (typically below their nominal size).
  - PCSE using HAR estimators is a good idea.

Dynamic Panel Models

- What if we have a dynamic process?
  - Examples
    - Cigarette consumption – lots of inertia.
    - Behavioral finance/momentum models --lagged returns matter.
  - We might consider a model like:

$$y_{it} = y_{it-1}\gamma + x_{it}\beta + c_i + \varepsilon_{it}$$

• Now, y_{it-1} is included as an explanatory variable. Now, => c_i and y_{it-1} are correlated!

- Issue: FE, RE estimators are biased.

• Time-demeaned (or quasi-demeaned) y_{it-1} correlated with error.

• FE is biased for small T. Gets better as T gets bigger (30+).

• RE also biased.

Dynamic Panel Models

- One solution: Use FD and instrumental variables
  - Strategy: If there's a problem between the error, ε_{it}, and lag y_i, let's find a way to calculate a new version of lag y_i that doesn't pose a problem
    - Idea: Further lags of y_i are not an issue in a FD model.
    - Use them as "instrumental variables," as a proxy for lag y_i.
  - Arellano-Bond (1991): GMM estimator
    - A FD estimator.
    - Lag of levels as an instrument for differenced y_i.
  - Arellano-Bover (1995)/Blundell-Bond: "System GMM"
    - Expand on this by using lags of differences and levels as IVs.
    - Generalized Method of Moments (GMM) estimation.

Dynamic Panel Models: GMM

(Arellano/Bond/Bover, Journal of Econometrics, 1995)

$y_{i,t} = \boldsymbol{\beta}'\mathbf{x}_{i,t} + \gamma y_{i,t-1} + \varepsilon_{i,t} + u_i$

Dynamic random effects model for panel data.

Can't use least squares to estimate consistently. Can't use FGLS without estimates of parameters.

Many moment conditions: What is orthogonal to the period 1 disturbance?

$E[(\varepsilon_{i,1} + u_i)\mathbf{x}_{i,1}] = 0$ = K orthogonality conditions, K+1 parameters

$E[(\varepsilon_{i,1} + u_i)\mathbf{x}_{i,2}] = 0$ = K more orthogonality conditions, same K+1 parameters

...

$E[(\varepsilon_{i,1} + u_i)\mathbf{x}_{i,1}] = 0$ = K orthogonality conditions, same K+1 parameters

The same variables are orthogonal to the period 2 disturbance.

There are hundreds, sometimes thousands of moment conditions, even for fairly small models.

Dynamic Panel Models: GMM

- Key usual assumptions / issues
  - – Serial correlation of differenced errors limited to 1 lag
  - – No overidentifying restrictions (No Hansen - Sargan test)
  - – Q: How many instruments?

- Criticisms:
  - Angrist and Pichke (2009): Assumptions are not always plausible.
  - Allison (2009)
  - Bollen and Brand (2010): Hard to compare models.

Dynamic Panel Models: Remarks I

- General remarks:
  - Ignoring dynamics –i.e., lags– not a good idea: omitted variables problem.
  - It is important to think carefully about dynamic processes:
    - How long does it take things to unfold?
    - What lags does it make sense to include?
    - With huge datasets, we can just throw lots in
      - – With smaller datasets, it is important to think things through.

Dynamic Panel Models: IV Framework

- Traditional IV panel estimator:

$$y_{it} = x_{it}\beta + Y_{it}\phi + c_i + \varepsilon_{it}$$

  - **X** = exogenous covariates
  - **Y** = other endogenous covariates (may be related to ε_{it})
  - c_i = unobserved unit-specific characteristic
  - ε_{it} = idiosyncratic error
  - – Treat c_i as random, fixed, or use differencing to wipe it out
  - – Use contemporaneous or lagged **X** and (appropriate) lags of **Y** as instruments in two-stage estimation of y_{it}.

Note: This approach works well if lagged **Y** is plausibly exogenous.

Time Series Cross Section (TSCS) Data

- Time Series Cross Section (TSCS) Data
  - Panel Data with large *T*, small *N*
  - Example I: economic variables for industrialized countries
    Often 10-30 countries
    Often around 30 to 40 years of data
  - Example II: financial variables
    Often more than 1,000 firms
    Often 40-50 years of data for well-established markets (10-30 for emerging markets).

– Beck's (2001) advice:
- No specific minimum for *T*; but be suspicious of *T*<10
- Large *N* is not required (though, it does not hurt)

Time Series Cross Section (TSCS) Data

- Typical complications of TSCS Data
  - Heteroskedasticity and autocorrelation
  - Autocorrelation cannot be ignored.
  - As *N* grows, the probability of cross-correlations (contemporaneous and time-varying) also grows.
    - Correlation at same time point across cases (world factor affecting all markets)
    - Correlation at different time points across cases (contagion effects over time)

TSCS Data: OLS PCSE

- Beck and Katz (2001)
  – "Old" view: Use FGLS to deal with heteroskedasticity & correlated errors.
    - Problem: This underestimates standard errors.

  – New view: Use OLS regression
    - With FE to deal with unit heterogeneity
      – To address panel heteroskedasticity ⇒ With PCSE's.
    - With FE to deal with unit heterogeneity
      – To address serial correlation ⇒ With lagged dependent variable in the model

TSCS Data: Dynamics

- Beck and Katz (2009) examine dynamic models
  - OLS PCSE with lagged **Y** and FE
    - Still appropriate
    - Better than some IV estimators
      - But, did not compare to System GMM.

- Plumper, Troeger, and Manow (2005)
    - FE is not theoretically justified and absorbs theoretically important variance.
    - Lagged **Y** absorbs theoretically important temporal variation
    - Ideally, economic theory must guide model choices.

TSCS Data: Nonstationary Data

- <u>Issue</u>: Analysis of longitudinal (time-series) data is going through big changes
    - Realization that strongly trending data cause problems
      - Random walk / *unit root* (ρ=1) / I(1) / non-stationary or near integrated data.
      - Mixing stationary, I(0) data, with I(1) data.
    - The "spurious regression" problem.

  - Strategies:
    - Tests for *unit roots* in time series & panel data
    - Differencing as a solution
      - A reason to try FD models.

Panel Data: Final Remarks

1. Panel data strategies are taught as "*fixes*"
   – How do I "fix" unobserved effects?
   – How do I "fix" dynamics/serial correlation?
   - But, the fixes really change what you are modeling
   - A FE (within) model is a very different look at your data, compared to pooled OLS.
   - Goal: learn the "fixes." But, think about interpretation.

2. Lots of disagreements in literature
   - What is the best "fix"?
   - Reformulation of model; final model should have "no problems" –LSE approach.

Panel Data: Final Remarks

3. Very important: Try a wide range of models
   - If your findings are robust, you are doing fine.
   - If not, differences may help you figure out a better model.
   - In both cases, you will not get "surprised" when your results go away after following the suggestion of a referee!