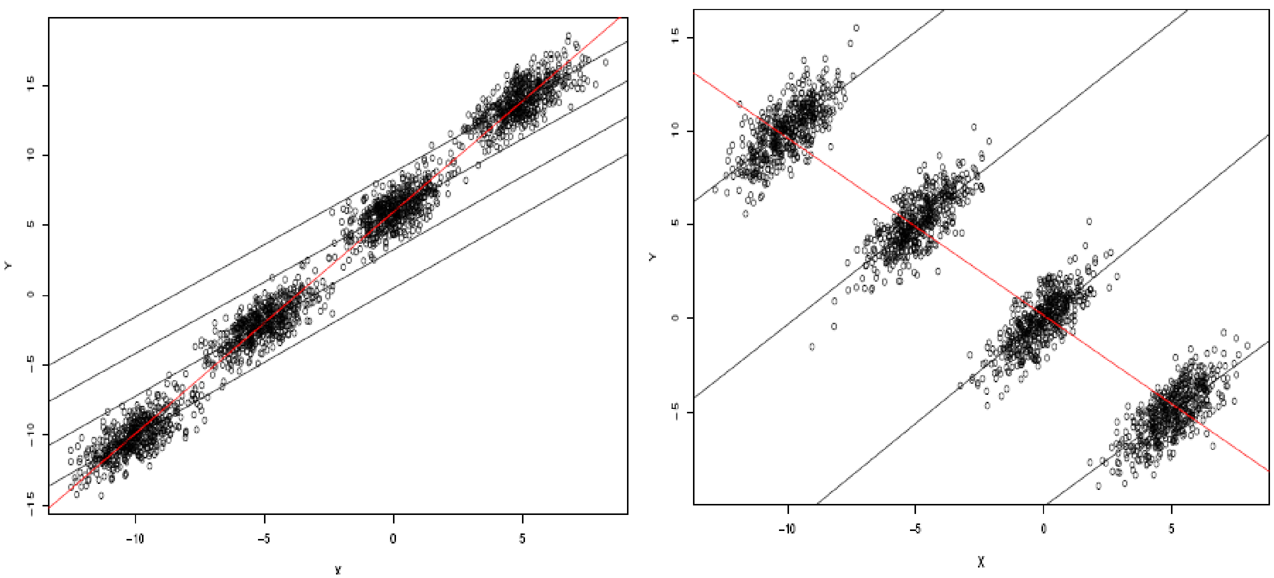
LSDV, a.k.a., FE Models/Estimators in TSCS

I. (Unmodeled) Unit-Specific "Effects":

A. Unobserved (unmodeled) Unit (e.g., country) Effects: $\alpha_i \neq \alpha$

B. Unobserved (unmodeled) Time (sub-unit) Effects: $\alpha_t \neq \alpha$

II. Examples: graphs of heterogeneity in α



A. If Cov($\bar{x}_i$,u_i)=0 no bias, though still ineff & s.e.'s likely wrong.

B. If Cov($\bar{x}_i$,u_i)≠0, then biased, inconsistent, & inefficient.

1.2 Example: Unobserved Country Effects and LSDV

- In the model:

$$Govt\ Spending_{it} = \beta_0 + \beta_1 Openness_{it} + \beta_2 Z_{it} + \varepsilon_{it}$$

  It might be argued that the level of government spending as a percentage of GDP differs for reasons that are specific to each country (e.g., solidaristic values in Sweden). This is also known as cross-sectional heterogeneity.

- If these unit-specific factors are correlated with other variables in the model, we will have an instance of omitted variable bias. Even if not, we will get larger standard errors because we are not incorporating sources of cross-country variation into the model.

- We could try to explicitly incorporate all the systematic factors that might lead to different levels of government spending across countries, but places high demands in terms of data gathering.

- Another way to do this, which may not be as demanding data-wise, is to introduce a set of country dummies into the model.

$$Govt\ Spending_{it} = \alpha_i + \beta_1 Openness_{it} + \beta_2 Z_{it} + \varepsilon_{it}$$

  This is equivalent to introducing a country-specific intercept into the model. Either include a dummy for all the countries but one, and keep the intercept term, or estimate the model with a full set of country dummies and no intercept.

3: *should* try…

4. Equiv, so which…

1.2.1 Time Effects

- There might also be time-specific effects (e.g., government spending went up everywhere in 1973–74 in OECD economies because the first oil shock led to unemployment and increased government unemployment payments). Once again, if the time-specific factors are not accounted for, we could face the problem of bias.

- To account for this, introduce a set of dummies for each time period.

$$Govt\ Spending_{it} = \alpha_i + \delta_t + \beta_1 Openness_{it} + \beta_2 Z_{it} + \varepsilon_{it}$$

- The degrees of freedom for the model are now $NT - k - N - T$. The statistical significance of the country-specific and time-specific effects can be tested by using an F-test to see if the country (time) dummies are jointly significant.

- The general approach of including unit-specific dummies is known as *Least Squares Dummy Variables* model, or *LSDV*.

- Can also include $(T-1)$ year dummies for time effects. These give the difference between the predicted causal effect from $\mathbf{x}_{it}\beta$ and what you would expect for that year. There has to be one year that provides the baseline prediction.

DLMFE: model?

1.3 Consequences of not accounting for heterogeneity

- If the α vary over individuals and we pool the data we can get bias in estimates of the slope and intercepts—see Figs. 1.1 and 1.2.

1.4 Testing for unit or time effects

- For LSDV (including an intercept), we want to test the linear hypothesis that

$$\alpha_1 = \alpha_2 = ... = \alpha_{N-1} = 0$$

- Can use an F-test:

$$F(N-1, NT-N-K) = \frac{(R^2_{UR} - R^2_R)/(N-1)}{(1-R^2_{UR})/(NT-N-K)}$$

In this case, the unrestricted model is the one with the cross-sectional dummies (and hence different intercepts); the restricted model is the one with just a single intercept. A similar test could also be performed on the year dummies.

➢ Note that $N-1$ represents the number of new regressors in the unrestricted model and $NT-N-K$ represents the total number of data points minus the total number of parameters in the unrestricted model.

ΔR^2 test…

1.4.1 How to do this test in Stata?

- After the `regress` command you do:

  1. If there are $(N-1)$ cross-sectional dummies and an intercept
     `test dummy1=dummy2=dummy3=dummy4=...=dummyN-1=0`
  2. If there are N cross-sectional dummies and no intercept
     `test dummy1=dummy2=dummy3=dummy4=...=dummyN`

1.4.2 An alternative test

- Beck and Katz ('01 *IO*) argue that the Schwartz Criterion (SC) is superior to the standard F test for the presence of unit effects, because the SC imposes a higher penalty for including more explanatory variables.

- The SC provides a difficult test for the LSDV model where N is particularly large and separate dummies for each cross-sectional unit are specified.

- Assume a prior probability of the true model being K_1 and a prior conditional distribution of the parameters given that K_1 is the true model. Then choose the a posteriori most probable model.

- We choose the model that minimizes

$$SC = \ln(\mathbf{u}'\mathbf{u}/NT) + \frac{K \ln NT}{NT} \tag{1.1}$$

  where $\mathbf{u}$ is the NT vector of estimated residuals.

- Just choose the model that has the lowest SC.

Note: Stata's `testparm` useful…

Fixed Effects Estimators

3.1 LSDV as Fixed Effects

- Least squares dummy variable estimation is also known as **fixed effects**, because it assumes that the unobserved effect for a given cross-sectional unit or time period can be estimated as a given, *fixed* effect.
- Can think of this as fixed in repeated samples (e.g., France is France) as opposed to randomly drawn.
- Let the original model be

$$y_{it} = \alpha_i^* + \beta' \mathbf{x}_{it} + u_{it} \tag{3.1}$$

Note: In this sense, fixed-effect model philosophically less consistent with hyper-population view than random-effects.

- We can rewrite this in vector form as

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_N \end{bmatrix} = \begin{bmatrix} \mathbf{e} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} \alpha_1^* + \begin{bmatrix} \mathbf{0} \\ \mathbf{e} \\ \vdots \\ \mathbf{0} \end{bmatrix} \alpha_2^* + \ldots + \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{e} \end{bmatrix} \alpha_N^* + \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_N \end{bmatrix} \beta + \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_N \end{bmatrix} \quad (3.2)$$

where

$$\underset{T\times 1}{\mathbf{y}_i} = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{bmatrix}, \quad \underset{T\times K}{\mathbf{X}_i} = \begin{bmatrix} x_{1i1} & x_{2i1} & \ldots & x_{Ki1} \\ x_{1i2} & x_{2i2} & \ldots & x_{Ki2} \\ \vdots & \vdots & & \vdots \\ x_{1iT} & x_{2iT} & \ldots & x_{KiT} \end{bmatrix}, \quad \underset{T\times 1}{\mathbf{e}} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix},$$

$$\underset{T\times 1}{\mathbf{u}_i} = \begin{bmatrix} u_{i1} \\ u_{i2} \\ \vdots \\ u_{iT} \end{bmatrix},$$

$$E[\mathbf{u}_i] = \mathbf{0}, \quad E[\mathbf{u}_i \mathbf{u}_i'] = \sigma_u^2 \mathbf{I}_T, \quad E[\mathbf{u}_i \mathbf{u}_j'] = \mathbf{0} \text{ if } i \neq j.$$

- These assumptions regarding u_{it} mean that the OLS estimator for eq. 3.2 is BLUE.

Note: Hsiao's unfortunate notational choice of **e** for **i**.

Note: could easily combine with panel-heteroskedasticity by $\sigma_{u_i}^2$.

- To obtain the OLS estimators of α_i^* and $\boldsymbol{\beta}$, we minimize:

$$S = \sum_{i=1}^{N} \mathbf{u}_i' \mathbf{u}_i = \sum_{i=1}^{N} (\mathbf{y}_i - \mathbf{e}\alpha_i^* - \mathbf{X}_i \boldsymbol{\beta})'(\mathbf{y}_i - \mathbf{e}\alpha_i^* - \mathbf{X}_i \boldsymbol{\beta}).$$

- Take partial derivatives wrt to α_i^*, set equal to zero and solve to get:

$$\hat{\alpha}_i^* = \bar{y}_i - \boldsymbol{\beta}' \bar{\mathbf{x}}_i \qquad (3.3)$$

where

$$\bar{y}_i = \sum_{t=1}^{T} y_{it}/T, \quad \bar{\mathbf{x}}_i = \sum_{t=1}^{T} \mathbf{x}_{it}/T.$$

- Substitute our estimate for $\hat{\alpha}_i^*$ in S, take partial derivatives wrt $\boldsymbol{\beta}$, set equal to zero and solve:

$$\hat{\boldsymbol{\beta}}_{\mathrm{CV}} = \left[\sum_{i=1}^{N} \sum_{t=1}^{T} (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)' \right]^{-1} \left[\sum_{i=1}^{N} \sum_{t=1}^{T} (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(y_{it} - \bar{y}_i) \right]$$

- Including separate dummies for each cross-sectional unit will produce estimates of the unit-specific effects.

- While this may be desirable, it does come at some cost—possibly inverting a large matrix of 0s and 1s.

Note why this therefore also sometimes called “within estimator”

- Another way to compute this estimator w/o including dummies is to subtract off the time means:

$$\bar{y}_i = \alpha_i^* + \beta' \bar{\mathbf{x}}_i + \bar{u}_i \tag{3.4}$$

- If we estimated β in this equation by OLS (constraining $\alpha_i^* = \alpha^* \ \forall\ i$), it will produce what is known as the "Between Effects" estimator, or β_{BE}, which shows how the mean level of the dependent variable for each cross-sectional unit varies with the mean level of the independent variables.

- Subtracting eq. 3.4 from eq. 3.1 gives

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

or

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

- Running OLS on this equation gives results identical to LSDV.

- Sometimes called the *within-group estimator*, because it uses only the variation in y_{it} and $\mathbf{x}_{it}$ within each cross-sectional unit to estimate the β coefficients.

- Any variation between cross-sectional units is assumed to spring from the unobserved fixed effects.

Note Troeger's point about limited nature this heterogeneity.

- Another way to approach this is to pre-multiply each cross-sectional unit equation ($\mathbf{y}_i = \mathbf{e}\alpha_i^* + \mathbf{X}_i\boldsymbol{\beta} + \mathbf{u}_i$) by a $T \times T$ idempotent "sweep" matrix:

$$\mathbf{Q} = \mathbf{I}_T - \frac{1}{T}\mathbf{e}\mathbf{e}'$$

- This has the effect of sweeping out the α_i^*s and transforming the variables so that the values for each individual are measured in terms of deviations from their means over time:

$$\begin{aligned} \mathbf{Q}\mathbf{y}_i &= \mathbf{Q}\mathbf{e}\alpha_i^* + \mathbf{Q}\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Q}\mathbf{u}_i & (3.5) \\ &= \mathbf{Q}\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Q}\mathbf{u}_i & (3.6) \end{aligned}$$

- Running OLS on this regression gives

$$\hat{\boldsymbol{\beta}}_{\text{CV}} = \left[\sum_{i=1}^{N} \mathbf{X}_i'\mathbf{Q}\mathbf{X}_i\right]^{-1} \left[\sum_{i=1}^{N} \mathbf{X}_i'\mathbf{Q}\mathbf{y}_i\right]$$

- The variance-covariance matrix is

$$\text{var}[\boldsymbol{\beta}_{\text{CV}}] = \sigma_u^2 \left[\sum_{i=1}^{N} \mathbf{X}_i'\mathbf{Q}\mathbf{X}_i\right]^{-1}$$

- We can compute an estimate of σ_u^2 as

$$\hat{\sigma}_u^2 = \hat{\mathbf{u}}'\hat{\mathbf{u}}/(NT - N - k)$$

where

$$\hat{\mathbf{u}}_i = \mathbf{Q}\mathbf{y}_i - \mathbf{Q}\mathbf{X}_i\hat{\boldsymbol{\beta}}_{\text{CV}}$$

- Properties of β_{CV}: unbiased and consistent whether N or T or both tend to infinity.

- Note that the OLS estimate of α_i^* is unbiased, but is consistent only as $T \to \infty$.

  ➢ With LSDV consistency is an issue: **incidental parameters problem**.

- A key advantages of FE estimators: can have correlation between $\mathbf{x}_{it}$ and α_i^*.

- A key drawback: if time-invariant regressors are included in the model, the standard FE estimator will not produce estimates for the effects of these variables (perfect collinearity in LSDV).

  ➢ There is an IV approach to produce estimates, but requires some exogeneity assumptions that may not be met in practice.

- The effects of slow-moving variables can also be estimated very imprecisely due to collinearity.

(Yet) Another Way to See How/Why Equivalence Differencing & Dummying

1. Break X' into Two (sets of) variables, X_1 & X_2:

$$Y = X_1 \boldsymbol{\beta}_1 + X_2 \boldsymbol{\beta}_2 + e$$

- X_1 is variables (columns) 1-j of X
- X_2 are the remaining (j+1)-k columns (variables)

the $\boldsymbol{\beta}_1$ & $\boldsymbol{\beta}_2$ are the vectors of their coefficients.

2. Recall the "normal equations"

$$(X'X)\,\mathbf{b} = X'\mathbf{Y}$$

, now these can be (are) "broken up" into:

$$\begin{matrix}(a)\\(b)\end{matrix}\quad \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} = \begin{bmatrix} X_1'\mathbf{Y} \\ X_2'\mathbf{Y} \end{bmatrix}$$

, which, solving by pre-multiplying both sides by $[X'X]^{-1}$ gives ..

$$\begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix}^{-1} \begin{bmatrix} X_1'\mathbf{Y} \\ X_2'\mathbf{Y} \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix}$$

$$\begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix}^{-1} \begin{bmatrix} X_1'\underset{\sim}{Y} \\ X_2'\underset{\sim}{Y} \end{bmatrix} = \begin{bmatrix} \underset{\sim}{b}_1 \\ \underset{\sim}{b}_2 \end{bmatrix}$$ gives ...

3. By rules of inverting partitioned matrices (chapt. 2)

$$\underset{\sim}{b}_1 = (X_1'X_1)^{-1}X_1'\underset{\sim}{Y} - (X_1'X_1)^{-1}X_1'X_2 b_2$$

$$= A_1\underset{\sim}{Y} - A_1X_2\underset{\sim}{b}_2$$

$$= \boxed{\underset{\sim}{b}_1} - \text{some adjustment which involves coeffs from } X_2 \text{ on } X_1 \; (A_1X_2)$$

$\hookrightarrow$ from regression of $\underset{\sim}{Y}$ on X_1 only

times coeffs. on X_2 in whole regression

subst. this into equation (b) & you get:

$$[X_2'X_1(A_1\underset{\sim}{Y} - A_1X_2\underset{\sim}{b}_2) + X_2'X_2\underset{\sim}{b}_2] = X_2'\underset{\sim}{Y}$$

$$X_2'X_1A_1\underset{\sim}{Y} - X_2'X_1A_1X_2\underset{\sim}{b}_2 + X_2'X_2\underset{\sim}{b}_2 = X_2'\underset{\sim}{Y}$$

(but $X_1A_1 \equiv N_1$) $$X_2'N_1\underset{\sim}{Y} - X_2'N_1X_2\underset{\sim}{b}_2 + X_2'X_2\underset{\sim}{b}_2 = X_2'\underset{\sim}{Y}$$

(gathering up $\underset{\sim}{b}_2$ terms on LHS & $\underset{\sim}{Y}$ terms on RHS) $$X_2'(I - N_1)X_2\underset{\sim}{b}_2 = X_2'(I - N_1)\underset{\sim}{Y}$$

$$(X_2'M_1X_2)\underset{\sim}{b}_2 = X_2'M_1\underset{\sim}{Y}$$

$$\underset{\sim}{b}_2 = (X_2'M_1X_2)^{-1}X_2'M_1\underset{\sim}{Y}$$

V. D. "Partial" Regression Coefficients

$$\underset{\sim}{b}_2 = (X_2' M_1 X_2)^{-1} X_2' M_1 \underset{\sim}{Y}$$

now, M_1 is symmetric & idempotent so $M_1' M_1 = M_1 M_1 = M_1$;
replace M_1 in above by $M_1' M_1$:

$$\underset{\sim}{b}_2 = (X_2' M_1' M_1 X_2)^{-1} X_2' M_1' M_1 \underset{\sim}{Y}$$

(C) $$\underset{\sim}{b}_2 = \left[(M_1 X_2)'(M_1 X_2)\right]^{-1} (M_1 X_2)' M_1 \underset{\sim}{Y}$$

M_1 is "residual-maker" for regression of thing it multiplies on X_1,

and $\underset{\sim}{b}_2$ is coefficient from Regression of $(M_1 \underset{\sim}{Y})$ on $(M_1 X_2)$

(equation (C) is in $(Z'Z)^{-1}Z'y$ form, right? so it's an A matrix times $\underset{\sim}{Y}$, thus regression coefficients)

but $M_1 \underset{\sim}{Y}$ are the residuals from $\underset{\sim}{Y}$ regressed on X_1 & $M_1 X_2$ are the residuals from (each of) the X_2's regressed on (all of) the X_1's.

- Thus, $\underset{\sim}{b}_2$, the coefficient(s) on X_2 are also the coefficients on X_2^* when regressed on $\underset{\sim}{Y}^*$ where the $*$'s mean that all (linear) relationships with X_1 has been netted out. This is what it means to say "the effect(s) of X_2, controlling for X_1".

Partial Coefficients -- one important example

- suppose we consider X_1 to be just the constant (the vector of ones) and X_2 to be all the other variables. Then, just as before:

$$\underset{\sim}{b}_2 = [(M_1 X_2)'(M_1 X_2)]^{-1}(M_1 X_2)' M_1 \underset{\sim}{Y}$$

- M_1 here is the "residual maker" regressing what it multiplies on the constant only. What do you get regressing any (matrix) vector, Z, on a constant? A coefficient of $\bar{Z}$. Why?

We did this way back in Week 4: (indirectly)

$$\underset{a}{Min}\ (z - b \cdot c)'(z - b \cdot c) = \underset{a}{Min} \sum_{i=1}^{n} (z - b)^2$$

(c ↑ vector of ones)

$$\Rightarrow b = \bar{Z}$$

So, then M_1 here nets out the mean of $\underset{\sim}{Y}$ and of all X_2 columns. Thus $\underset{\sim}{b}_1$ (← the coeff on the constant) in OLS mult. regression including a constant, nets out the means of all X's & Y's.

Troeger summarizes well the "To FE or Not To FE" dilemma:

What it does:

- Eliminates the omitted variables bias of time-invariant unit specific variables

What it does not:

- Does not control away other problems of unit heterogeneity: unobserved time varying variables, slope heterogeneity, unit specific dynamics and lag structures

Problems:

- does not allow estimating the effect of time-invariant variables
- estimator inefficient: likely to obtain estimates that largely deviate from truth, EVEN IF estimates are unbiased
- Does only use within information

Yet,

not controlling for unit effects leads to biased estimates if unit effects exist and are correlated with any of the regressors

Note the Estimator Options So Far:

1. POOLED: $y_{it} = a + \mathbf{x}'\mathbf{b}^{p} + e_{it}$

2. BETWEEN: $\bar{y}_i = \bar{a}_i + \bar{\mathbf{x}}_i'\mathbf{b}^{b} + \bar{e}_i$

3. WITHIN (FIXED-EFFECTS (LSDV):

$$\left(y_{it} - \bar{y}_i\right) = \left(\mathbf{x}_{it} - \bar{\mathbf{x}}_i\right)' \mathbf{b}^{w} + \left(e_{it} - \bar{e}_i\right)$$

4. UNIT-BY-UNIT: $y_{it} = a_i + \mathbf{x}'\mathbf{b}_i^{u} + e_{it}$

(Discuss difference between Unit-by-Unit pooled or estimated separately…) A fifth option in this family is to come:

5. RANDOM-EFFECTS

Generally, all can be shown to be weighted averages of each other, each making different compromises. More on that later.