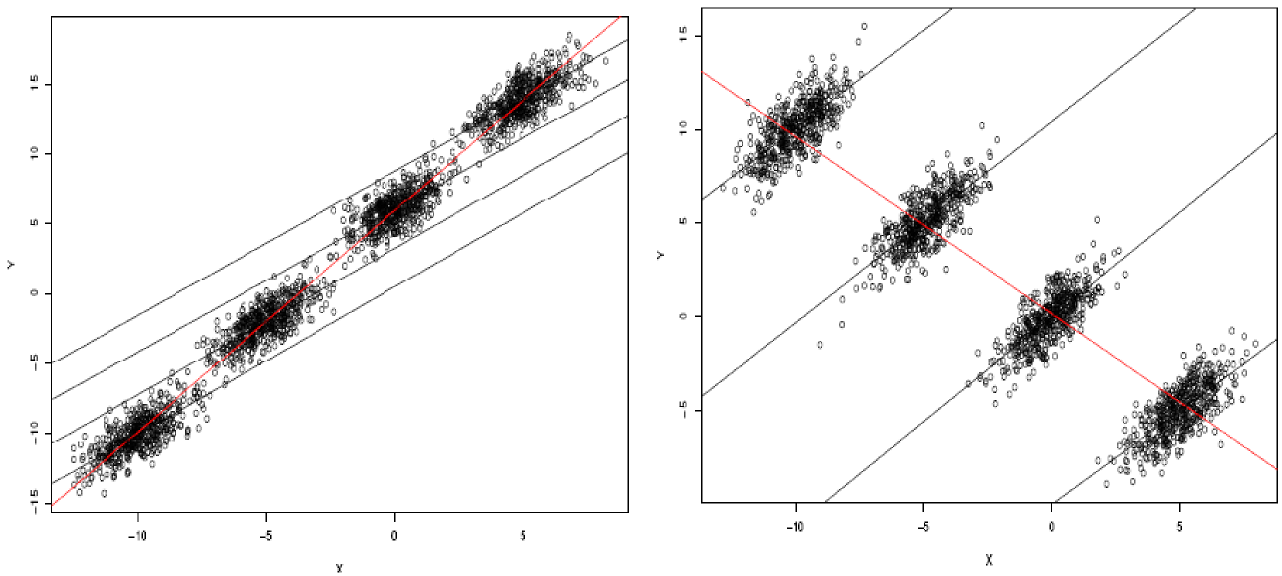
RE Models/Estimators in TSCS

I. Issue: Unit (or Time-Period) Heterogeneity: $y_{it} = \alpha_i + \alpha_t + \mathbf{x\beta} + \varepsilon_{it}$

II. Examples: graphs of heterogeneity in α



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

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

C. Random-effects estimators address essentially to the former.

III. Substantive/Conceptual issues:

A. Fixed Effects (FE):

1. Philosophically apt if see unit-specific effects as fixed (repeated samples), estimable amounts, for each cross-section unit. (Sweden always have intercept 1.2 units if somehow took other samples.)
2. Crucial if unit-specific effects correlate with $\bar{\mathbf{x}}_i$.
3. Costs of FE:
   - a) Cannot include regressors that not vary over time w/in units (e.g., individuals' demographics, countries' institutions, …).
   - b) Highly inefficient:
     - (1) Estimates MANY parameters. Uses $1/T$ degrees freedom. Small T...
     - (2) Often cross-sectional var dom, so worse yet in useful-variation terms.
   - c) Slowly/rarely changing variables highly co-linear w/ *fixed* effects => their effects very imprecisely, unstably, unsurely estimated.
   - d) Costs of FE are one reason that people use random effects.
   - e) Only intercept shift; no other heterogeneity (well) addressed.
4. Note: many these args weigh both ways more in Panel than TSCS

B. Random Effects (RE):

1. However, if model est'd in any TSCS, regardless size *N*, aims generalize to how types of units work, not estimate factuals for specif units, then these unobserved unit-specific effects should be random.

   a) We cannot estimate intercept for each country because we don't have, indeed logically could not have, all of them.

2. May instead wish estimate coefficients on substantive regressors well, accounting possible country-specific effects that would enter as a random shock from a known distribution. Logic ⇒ **random effects**.

   a) Very efficient relative to FE: need only distributional parameters.

3. Costs:

   a) Must assume a lot: no correlation of random effects with regressors.

   b) Must assume a lot: no correlation among random components.

   c) Addresses only parts of concerns re: unmodeled heterogeneity: the lesser parts in some ways (as per Troeger's comments…)

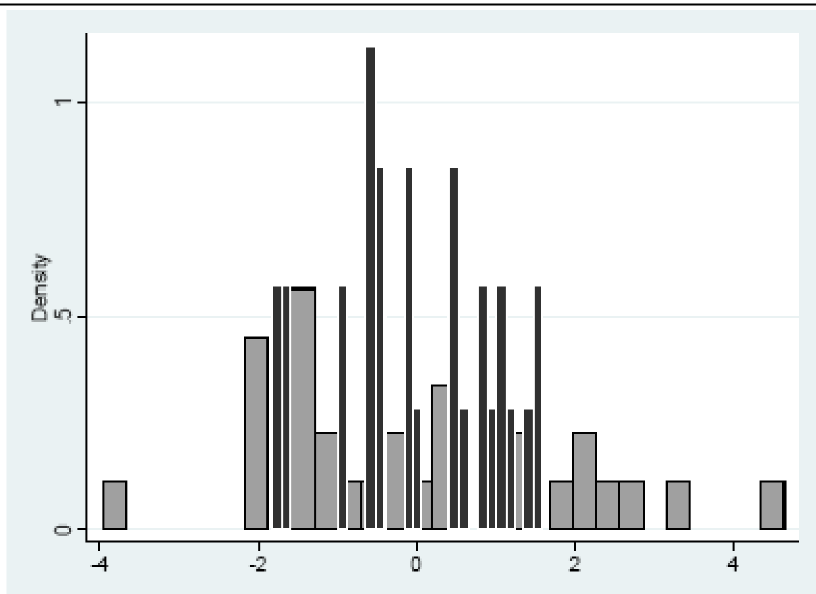
   d) MLE of RE/C, perform best, also adds strong Normality assumpt

4. Part of how FE manifests is tendency to *pick up* too much heterogeneity and call it unit-fixed &, in LSDV case, part systematic.

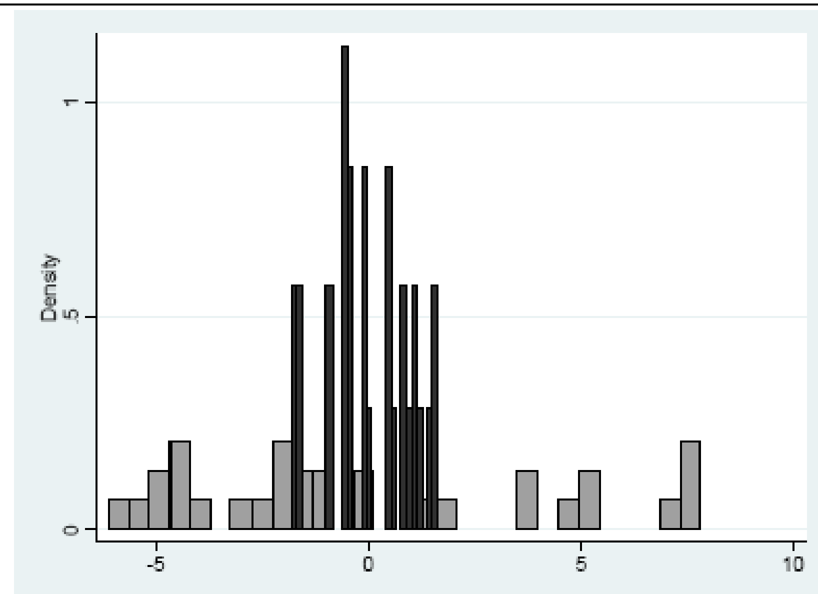
a) I.e., the *sweep* sweeps both fixed & stochastic unit-specific effects.

b) I.e., classic *overfitting* = another way see incidental-param problem

c) Troeger's MC's illustrate problem: note severe overdispersion of estimated relative to actual unit-specific effects:

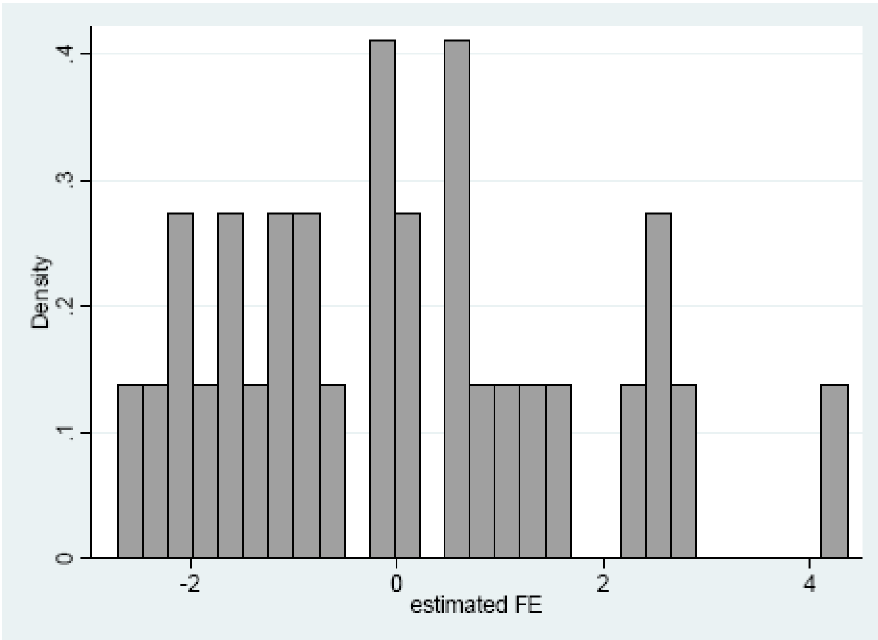


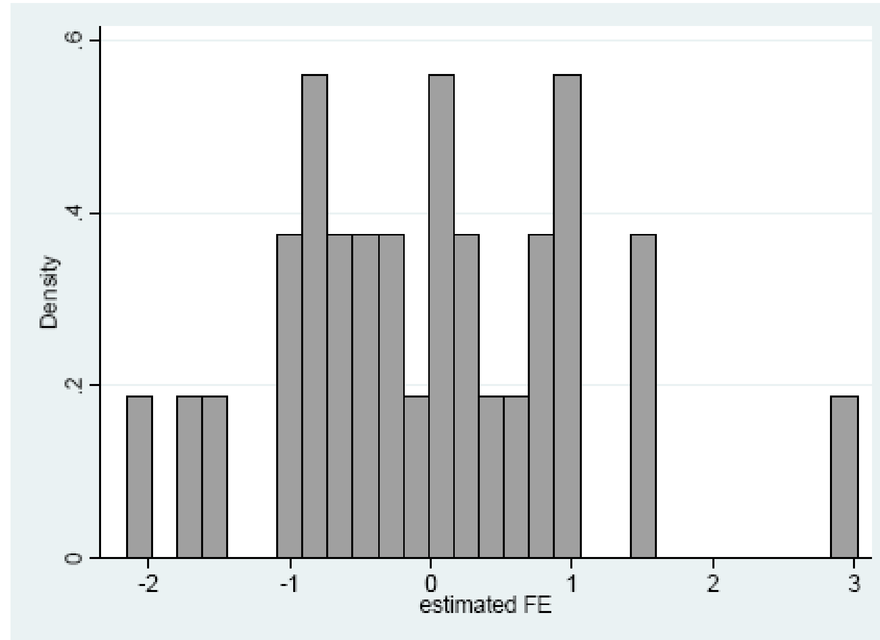
black: true FE, grey: estimated FE
Settings: FE in DGP ~ N(0,1), 1 RHS variable, SD(within)=SD(between)=1



black: true FE, grey: estimated FE
Settings: FE in DGP ~N(0,1), 1 RHS variable SD(within)=1 SD(between)=3

d) Can be even worse. Will even find fixed-effects where they ain't:





Settings: no FE in DGP, 1 RHS variable, SD(within)=SD(between)=1

Settings: no FE in DGP, 3 RHS variables, SD(within)=SD(between)=1

Notice: in both this & previous case, not obviously biased, but highly inefficient. In limited (in *T*) samples, "mere inefficiency" issue can be "not so *mere*".

IV. General Random-Effects (Error-Components, HLM) Model

$$y_{it} = \mathbf{x\beta} + \alpha_i + \lambda_t + u_{it} = \mathbf{x\beta} + v_{it}$$

$$E[\alpha_i] = E[\lambda_t] = E[u_{it}] = 0$$

$$E[\alpha_i \lambda_t] = E[\alpha_i u_{it}] = E[\lambda_t u_{it}] = 0$$

$$E[\alpha_i \alpha_j] = \begin{cases} \sigma_\alpha^2 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

$$E[\lambda_t \lambda_s] = \begin{cases} \sigma_\lambda^2 & \text{if } t = s \\ 0 & \text{if } t \neq s \end{cases}$$

$$E[u_{it} u_{js}] = \begin{cases} \sigma_u^2 & \text{if } i = j, t = s \\ 0 & \text{otherwise} \end{cases}$$

$$E[\alpha_i \mathrm{x}'_{it}] = E[\lambda_t \mathrm{x}'_{it}] = E[u_{it} \mathrm{x}'_{it}] = 0$$

1. Mean-zero for each component.
2. No correlation across components.

3-5. Constant variance each component.

6. Regressors exog. to each component.

Note: second & last lines are crucial; w/o them (or some sufficient replacement), model (becomes) inestimable.

Note: last line implies "only" a GLRM departure from CLRM.

Note: the sphericity w/i & across error components implies following:

$$\mathrm{var}[y_{it} | \mathrm{x}_{it}] = \sigma_y^2 = \sigma_\alpha^2 + \sigma_\lambda^2 + \sigma_u^2$$

V. The Typical RE Model (not as general as above):

- Let's add a general intercept to our model and set $\lambda_t = 0 \; \forall \; t$:

$$y_{it} = \mu + \beta' \mathbf{x}_{it} + \alpha_i + u_{it} \tag{4.1}$$

- We can rewrite this in vector form:

$$\mathbf{y}_i = \tilde{\mathbf{X}}_i \delta + \mathbf{v}_i \tag{4.2}$$

where

$$\underset{T\times(K+1)}{\tilde{\mathbf{X}}_i} = \begin{bmatrix} \mathrm{e} & \mathbf{X}_i \end{bmatrix}, \quad \underset{(K+1)\times 1}{\delta} = \begin{bmatrix} \mu \\ \beta \end{bmatrix}, \quad \underset{T\times 1}{\mathbf{v}_i} = \begin{bmatrix} v_{i1} \\ v_{i2} \\ \vdots \\ v_{iT} \end{bmatrix}, \quad v_{it} = \alpha_i + u_{it}$$

I.e., the typical RE model has just two error components: a unit-specific component and the unit-time unique component.

VI. RE is an example of the G(N)LRM, where:

- The variance-covariance matrix of the T disturbance terms $\mathbf{v}_i$ is:

$$\mathbf{V} = E[\mathbf{v}_i\mathbf{v}_i'] = \begin{bmatrix} (\sigma_u^2+\sigma_\alpha^2) & \sigma_\alpha^2 & \sigma_\alpha^2 & \dots & \sigma_\alpha^2 \\ \sigma_\alpha^2 & (\sigma_u^2+\sigma_\alpha^2) & \sigma_\alpha^2 & \dots & \sigma_\alpha^2 \\ \vdots & & & \ddots & \vdots \\ \sigma_\alpha^2 & \sigma_\alpha^2 & \sigma_\alpha^2 & \dots & (\sigma_u^2+\sigma_\alpha^2) \end{bmatrix}$$

$$= \sigma_u^2\mathbf{I}_T + \sigma_\alpha^2\mathbf{e}\mathbf{e}'$$

- Note that

$$\mathbf{V}^{-1} = \frac{1}{\sigma_u^2}\left[\mathbf{I}_T - \frac{\sigma_\alpha^2}{\sigma_u^2+T\sigma_\alpha^2}\right]\mathbf{e}\mathbf{e}'.$$

- The full variance-covariance matrix for all the NT observations is:

$$\Omega = \begin{bmatrix} \mathbf{V} & 0 & 0 & \dots & 0 \\ 0 & \mathbf{V} & 0 & \dots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \mathbf{V} \end{bmatrix} = \mathbf{I}_N \otimes \mathbf{V}$$

A. FE estimator: consistent, but inefficient & wrong V(**b**) est.

B. Since RE Model is G(N)LRM, suggests FGLS estimator:

- The normal equations for the GLS estimator are

$$\left[\sum_{i=1}^{N} \tilde{X}_i' V^{-1} \tilde{X}_i\right] \hat{\delta}_{\text{GLS}} = \sum_{i=1}^{N} \tilde{X}_i' V^{-1} y_i$$

- We could write the GLS estimator simply as

$$\hat{\delta}_{\text{GLS}} = \left[\sum_{i=1}^{N} \tilde{X}_i' V^{-1} \tilde{X}_i\right]^{-1} \sum_{i=1}^{N} \tilde{X}_i' V^{-1} y_i$$

, where

$$V^{-1} = \frac{1}{\sigma_u^2}\left[I_T - \frac{1}{T}ee' + \psi \cdot \frac{1}{T}ee'\right] = \frac{1}{\sigma_u^2}\left[Q + \psi \cdot \frac{1}{T}ee'\right]$$ and $$\psi = \frac{\sigma_u^2}{\sigma_u^2 + T\sigma_\alpha^2}$$.

1. Recall that Q here is the FE estimator “sweep” matrix.

2. So, RE doesn’t sweep all the unit-averages out, only 1-ψ of them.

   a) So, as $T\sigma_\alpha^2$ grows relative σ_u^2, →sweeps all: FE. As to 0, toward pool.

   b) Typical GLS intuition for correlated obs: partial diff the corr (see page 98 of review notes for analogous partial-difference for AR(1)).

C. As usual, then, FGLS applies OLS to transformed **y** & **X**. Here:

$$y_{it}^{*} = y_{it} - \theta \overline{y}_i \text{ , where } \theta \equiv \frac{T\sigma_\alpha^2}{T\sigma_\alpha^2 + \sigma_u^2}$$

1.Again: as $T\sigma_\alpha^2$ grows rel. σ_u^2, →sweeps all: FE. As to 0, → pool.

D. Can show that either expression implies RE estimator is wtd average of pooled & within estimates [VT's (u,ε)=GW's (α,u)]:

$$\hat{\boldsymbol{\beta}}_{\text{Model}} = \left[\Sigma_{i=1}^{N}\mathbf{X}_i'\left(\mathbf{I} - \frac{\hat{\gamma}_{i,\text{Model}}}{T_i}\mathbf{ii}'\right)\mathbf{X}_i\right]^{-1}\left[\Sigma_{i=1}^{N}\mathbf{X}_i'\left(\mathbf{I} - \frac{\hat{\gamma}_{i,\text{Model}}}{T_i}\mathbf{ii}'\right)\mathbf{y}_i\right]$$

$\hat{\gamma}_{\text{Model}} = 1$ for fixed effects.

$\hat{\gamma}_{i,\text{Model}} = \sqrt{\dfrac{T_i\hat{\sigma}_u^2}{\hat{\sigma}_\varepsilon^2 + T_i\hat{\sigma}_u^2}}$ for random effects.

As $T_i \to \infty, \hat{\gamma}_{i,RE} \to 1,$ random effects becomes fixed effects

As $\hat{\sigma}_u^2 \to 0, \hat{\gamma}_{i,RE} \to 0,$ random effects becomes OLS (of course)

As $\hat{\sigma}_u^2 \to \infty, \hat{\gamma}_{i,RE} \to 1,$ random effects becomes fixed effects

E. In fact, all these various estimators are related thus:

(The following is from Greene, *Econometric Analysis*, ch. 13.)

$$\mathbf{S}_{xx}^{total} = \mathbf{S}_{xx}^{within} + \mathbf{S}_{xx}^{between} \quad \text{and} \quad \mathbf{S}_{xy}^{total} = \mathbf{S}_{xy}^{within} + \mathbf{S}_{xy}^{between}.$$

Therefore, there are three possible least squares estimators of β corresponding to the decomposition. The least squares estimator is Pooled

$$\mathbf{b}^{total} = \left[\mathbf{S}_{xx}^{total}\right]^{-1}\mathbf{S}_{xy}^{total} = \left[\mathbf{S}_{xx}^{within} + \mathbf{S}_{xx}^{between}\right]^{-1}\left[\mathbf{S}_{xy}^{within} + \mathbf{S}_{xy}^{between}\right]. \quad \textbf{(13-11)}$$

The within-groups estimator is

$$\mathbf{b}^{within} = \left[\mathbf{S}_{xx}^{within}\right]^{-1}\mathbf{S}_{xy}^{within}. \quad \textbf{(13-12)}$$

This is the LSDV estimator computed earlier. [See (13-4).] An alternative estimator would be the between-groups estimator,

$$\mathbf{b}^{between} = \left[\mathbf{S}_{xx}^{between}\right]^{-1}\mathbf{S}_{xy}^{between} \quad \textbf{(13-13)}$$

(sometimes called the **group means estimator**). This least squares estimator of (13-10c)

$$\mathbf{S}_{xy}^{within} = \mathbf{S}_{xx}^{within}\mathbf{b}^{within} \quad \text{and} \quad \mathbf{S}_{xy}^{between} = \mathbf{S}_{xx}^{between}\mathbf{b}^{between}.$$

Inserting these in (13-11), we see that the least squares estimator is a **matrix weighted average** of the within- and between-groups estimators:

$$\mathbf{b}^{total} = \mathbf{F}^{within}\mathbf{b}^{within} + \mathbf{F}^{between}\mathbf{b}^{between}, \tag{13-14}$$

where From below, weight on each in above is its share of total variation.

$$\mathbf{F}^{within} = \left[\mathbf{S}_{xx}^{within} + \mathbf{S}_{xx}^{between}\right]^{-1}\mathbf{S}_{xx}^{within} = \mathbf{I} - \mathbf{F}^{between}$$

The form of this result resembles the Bayesian estimator in the classical model discussed in Section 16.2. The resemblance is more than passing; it can be shown [see, e.g., Judge (1985)] that

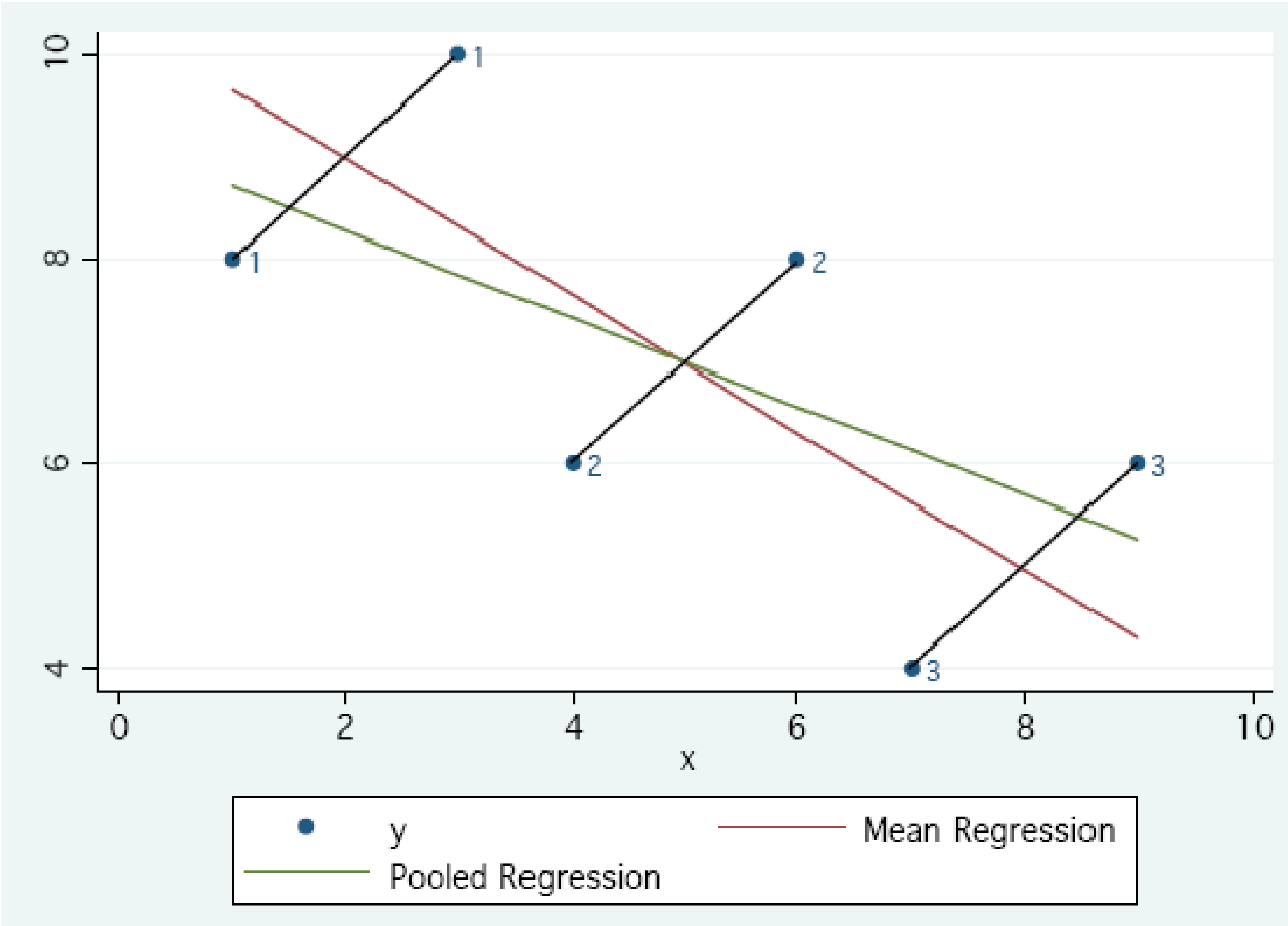
$$\mathbf{F}^{within} = \left\{[\text{Asy. Var}(\mathbf{b}^{within})]^{-1} + [\text{Asy. Var}(\mathbf{b}^{between})]^{-1}\right\}^{-1}[\text{Asy. Var}(\mathbf{b}^{within})]^{-1},$$

which is essentially the same mixing result we have for the Bayesian estimator. In the weighted average, the estimator with the smaller variance receives the greater weight.

--So, e.g., fully-pooled OLS weighs *between* relative to *within* proportionately to shares of total variation at those aggregations.

--That can yield (seeming) too much weight on *between*, as next fig.:

(Figure courtesy of Marco Steenbergen)



RE

It can be shown that the GLS estimator is, like the OLS estimator, a matrix weighted average of the within- and between-units estimators:

$$\hat{\beta} = \hat{\mathbf{F}}^{within}\mathbf{b}^{within} + (\mathbf{I} - \hat{\mathbf{F}}^{within})\mathbf{b}^{between},^{12} \quad \textbf{(13-23)}$$

where now,

$$\hat{\mathbf{F}}^{within} = \left[\mathbf{S}_{xx}^{within} + \lambda\mathbf{S}_{xx}^{between}\right]^{-1}\mathbf{S}_{xx}^{within},$$

$$\lambda = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_u^2} = (1-\theta)^2.$$

I.e., RE weighs BW less & WI (FE) more relative to POOL as *T* times cross-unit variation rises

To the extent that λ differs from one, we see that the inefficiency of least squares will follow from an inefficient weighting of the two estimators. Compared with generalized least squares, ordinary least squares places too much weight on the between-units variation. It includes it all in the variation in $\mathbf{X}$, rather than apportioning some of it to random variation across groups attributable to the variation in u_i across units.

--I like to say *seems* here, because this is assuming RE right model.

--Could say (VT would say): FE, conversely, due to what is essentially overfitting, seems place too much weight on *within* variation (by placing 0 wt on *between*).

To repeat (in Wawro's terms (& font) now):

In words: the GLS estimator is a weighted average of the b/t group estimator and the w/in group estimator, w/ ψ indicating the weight given to b/t group variation. Recall

$$\psi = \frac{\sigma_u^2}{\sigma_u^2 + T\sigma_\alpha^2}$$

Recall: Wawro's *u* is the *it*-unique error-component; α is the *i*-specific effect.

- As $\psi \rightarrow 1$, $\hat{\boldsymbol{\delta}}_{\text{GLS}} \rightarrow T_{\tilde{x}\tilde{x}}^{-1} T_{\tilde{x}y}$ (i.e., the OLS estimator). This means that little variance is explained by the unit effects.
- As $\psi \rightarrow 0$, $\hat{\boldsymbol{\beta}}_{\text{GLS}} \rightarrow$ the w/in estimator. This happens as either
  1. the unit-specific effects dominate the disturbance u_{it}.
  2. $T \rightarrow \infty$ (intuition: the α_i are like fixed parameters since we have so much data on the T dimension).
- GLS then is an intermediate approach b/t OLS and FE (which uses no b/t group variation).

So, recall the GLS estimator of the RE model (in Wawro's notation):

$$\hat{\boldsymbol{\delta}}_{\text{GLS}} = \left[\sum_{i=1}^{N} \tilde{\mathbf{X}}_i' \mathbf{V}^{-1} \tilde{\mathbf{X}}_i\right]^{-1} \sum_{i=1}^{N} \tilde{\mathbf{X}}_i' \mathbf{V}^{-1} \mathbf{y}_i$$

, where

$$\mathbf{V}^{-1} = \frac{1}{\sigma_u^2}\left[\mathbf{I}_T - \frac{1}{T}\mathbf{e}\mathbf{e}' + \psi \cdot \frac{1}{T}\mathbf{e}\mathbf{e}'\right] = \frac{1}{\sigma_u^2}\left[\mathbf{Q} + \psi \cdot \frac{1}{T}\mathbf{e}\mathbf{e}'\right]$$

We didn't talk about the V-Cov of these GLS estimates yet:

The variance of the GLS estimator is

$$\mathrm{var}\left[\hat{\boldsymbol{\beta}}_{\mathrm{GLS}}\right] = \sigma_u^2 \left[\sum_{i=1}^{N} \mathbf{X}_i'\mathbf{Q}\mathbf{X}_i + \psi T \sum_{i=1}^{N} (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})'\right]^{-1}$$

Recall the variance for the w/in group estimator:

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

The difference b/t these var-cov matrices is a p.d. matrix (assuming $\psi > 0$).

Thus, as $T \to \infty$, $\psi \to 0$, and $\mathrm{var}\left[\sqrt{T}\hat{\boldsymbol{\beta}}_{\mathrm{GLS}}\right] \to \mathrm{var}\left[\sqrt{T}\hat{\boldsymbol{\beta}}_{\mathrm{CV}}\right]$ (assuming our cross-product matrices converge to finite p.d. matrices).

--That is, the first is "smaller" than the second (efficiency), under RE being the right model. And (his last statement): RE→FE as $T\sigma_\alpha^2 \to \infty$.

FGLS: How estimate RE model?

--need estimates of *sweep* (partial-differencing) terms, which are :

$$y_{it}^{*} = y_{it} - \theta \overline{y}_i \text{ , where } \theta \equiv \frac{T\sigma_\alpha^2}{T\sigma_\alpha^2 + \sigma_u^2}$$

--to implement 2-stage FGLS, need consistent first-stage.

--since RE model just non-spherical errors, plenty of options for consistent 1st-stages: between (avg), within (FE), pooled (OLS).

-- within estimator (LSDV or unit-mean-differenced) purges α_i by construction, so it's estimated residual-variance is used for $\hat{\sigma}_u^2$.

-- between estimator (group-means regression) contains $\hat{\sigma}_u^2 + \hat{\sigma}_\alpha^2$, so it minus LSDV-based $\hat{\sigma}_u^2$ produces the desired $\hat{\sigma}_\alpha^2$ estimate.

--plug into above, transform, and OLS.

--can get $\hat{\sigma}_\alpha^2$<0, but taken to mean no "effects" worry about [*disc.*]

- RE estimates can also be computed by ML.
- To obtain the MLE, assume u_{it} and α_i are normally dist'd and start w/ the log of the likelihood function:

Totally standard likelihood for GNLRM, which, recall, is all the RE

$$\ln L = -\frac{NT}{2}\ln 2\pi - \frac{N}{2}\ln|\mathbf{V}| - \frac{1}{2}\sum_{i=1}^{N}(\mathbf{y}_i - \mathbf{e}\mu - \mathbf{X}_i\boldsymbol{\beta})'\mathbf{V}^{-1}(\mathbf{y}_i - \mathbf{e}\mu - \mathbf{X}_i\boldsymbol{\beta})$$

ln|**V**| & $\mathbf{V}^{-1}$ terms differ GNLRM & CNLRM likelihoods

- To obtain the MLE $\hat{\boldsymbol{\delta}}' = (\mu, \boldsymbol{\beta}', \sigma_u^2, \sigma_\alpha^2)$, we take partial derivatives wrt each of these parameters, set to zero and solve.

Note: multivariate normality, combined with the uncorrelated error-component assumptions, amount to a strong independence assumpt. This what meant earlier by *strong normality* assumptions of ML-RE.

- This gives four equations that we must solve simultaneously, which can be difficult.

- Instead we can use a sequential iterative procedure, alternating back and forth b/t μ and $\boldsymbol{\beta}$ and the variance components σ_u^2 and σ_α^2.

- For N fixed and $T \rightarrow \infty$, the MLEs of $\mu, \boldsymbol{\beta}'$, and σ_u^2 are consistent and $\rightarrow$ CV i.e., FE est. The MLE of σ_α^2 is inconsistent (insufficient variation b/c of fixed N).

- With simultaneous solution of σ_α^2, it's possible to obtain a negative value. It's also possible to obtain a boundary solution, although the prob. of this $\rightarrow 0$ as either T or $N \rightarrow \infty$.

MLE generally outperforms FGLS in these kinds of settings; this is very noticeable as the problem grows in complexity, such as in the move from RE to RC… Of course, this relative efficiency of MLE comes to some extent from its stronger assumptions & so presumably comes also at price of distributional fragility (lack of robustness to alternative distributions).

Also estimable by Bayesian MCMC (see Gill). Possible allow some nonsphericity in error-component assumpts of model: makes even more complex estimation. Bayesian MCMC.

Note the big differences FE vs. RE—

--FE addresses the potential bias issue from unit-specific effects, but:

--inefficient; relatedly, it overfits;

--in a sense, it's inefficient *because* it overfits, potentially massively so, in fact even infinitely so, relative to some questions (those re: time-invariant regressors, some of which are often of considerable substantive interest: e.g., effects of institutions.)

--Asymptotics best in T.

--RE addresses the efficiency issue, but:

--must assume away the bias issue to get that traction.

--Asymptotics best in N.

--Can make rather large difference to our estimates and conclusions (even though, as T increases, RE approaches FE). For example:

```
. xtreg sstran L.sstran LL.sstran unem left lftun growthpc depratio
cdem trade lowwage fdi, fe

Fixed-effects (within) regression               Number of obs      =       506
Group variable: cc                              Number of groups   =        17
R-sq:  within  = 0.9756                         Obs per group: min =        19
       between = 0.9952                                        avg =      29.8
       overall = 0.9853                                        max =        31
                                                F(11,478)          =   1737.13
corr(u_i, Xb)  = 0.6209                         Prob > F           =    0.0000
      sstran |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      sstran |
         L1. |   1.115156   .0372883    29.91   0.000     1.041887    1.188426
         L2. |    -.18442   .0372824    -4.95   0.000    -.2576776   -.1111624
        unem |   .0356034   .0172374     2.07   0.039     .0017329    .0694738
        left |  -.0005694   .0014259    -0.40   0.690    -.0033712    .0022325
       lftun |   .0002063   .0002349     0.88   0.380    -.0002552    .0006678
    growthpc |  -.1965328   .0107684   -18.25   0.000    -.2176922   -.1753735
    depratio |   .0411629   .0192749     2.14   0.033      .003289    .0790368
        cdem |  -.0023557   .0020135    -1.17   0.243     -.006312    .0016007
       trade |  -.0070589   .0037508    -1.88   0.060    -.0144291    .0003113
     lowwage |   .0005358   .0062747     0.09   0.932    -.0117936    .0128651
         fdi |   .0136313   .0216196     0.63   0.529    -.0288499    .0561124
       _cons |   .5041075   .8362199     0.60   0.547    -1.139014    2.147229
-------------+----------------------------------------------------------------
     sigma_u |   .5312598
     sigma_e |   .5109317
         rho |  .51949773   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(16, 478) =     4.79             Prob > F = 0.0000
. estimates store FEmodel
```

. xtreg sstran L.sstran LL.sstran unem left lftun growthpc depratio cdem trade lowwage fdi, re

```
Random-effects GLS regression                   Number of obs      =       506
Group variable: cc                              Number of groups   =        17
R-sq:  within  = 0.9738                         Obs per group: min =        19
       between = 0.9990                                        avg =      29.8
       overall = 0.9897                                        max =        31
Random effects u_i ~ Gaussian                   Wald chi2(11)      =  47476.44
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000
      sstran |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      sstran |
         L1. |   1.212013   .0359035    33.76   0.000     1.141644    1.282383
         L2. |  -.2324231   .0361611    -6.43   0.000    -.3032975   -.1615487
        unem |  -.0265803   .0096796    -2.75   0.006    -.0455519   -.0076086
        left |  -.0005572   .0011681    -0.48   0.633    -.0028466    .0017321
       lftun |   .0002636   .0002115     1.25   0.213    -.0001509     .000678
    growthpc |  -.1780914   .0108472   -16.42   0.000    -.1993515   -.1568313
    depratio |   .0208042   .0107735     1.93   0.053    -.0003114    .0419199
        cdem |     .00122   .0011412     1.07   0.285    -.0010167    .0034566
       trade |   .0035654   .0013232     2.69   0.007      .000972    .0061588
     lowwage |   .0080255   .0040181     2.00   0.046     .0001501    .0159009
         fdi |  -.0308592   .0178517    -1.73   0.084    -.0658479    .0041295
       _cons |   .0557606   .4460443     0.13   0.901    -.8184702    .9299914
-------------+----------------------------------------------------------------
     sigma_u |          0
     sigma_e |   .5109317
         rho |          0   (fraction of variance due to u_i)
```

. estimates store Remodel

<u>Hausman test</u>: (one use of which is RE v. FE)

--In a (really) large sample, can distinguish pair of estimates where, under null, both are consistent but one more efficient, and, under alternative, the first remains consistent but the second is biased. One is always-consistent, but inefficient if unnecessary; other is more efficient if assumptions right, but biased if not.

--Test basically follows the Wald logic in its structure:

$$Haus = \left(\hat{\boldsymbol{\theta}}_c - \hat{\boldsymbol{\theta}}_e\right)' \left[{}^A\widehat{\text{var}}\left(\hat{\boldsymbol{\theta}}_c\right) - {}^A\widehat{\text{var}}\left(\hat{\boldsymbol{\theta}}_e\right)\right]^{-1} \left(\hat{\boldsymbol{\theta}}_c - \hat{\boldsymbol{\theta}}_e\right) \sim^A \chi^2_k$$

--null in this case is that unit-specific shocks uncorrelated with regressors.

--under that null, both FE & RE are consistent, but RE is efficient.

--under alternative, FE is consistent still, but RE is biased.

Rejection: (supposed to) imply RE rejected in favor of FE….

---Problem 1: can get negative est'd V-Cov diff in samples. Assumed to mean very strong rejection of null.

---*Problem 2*: weak in limited samples.

---**Problem 3**: lots of reasons FE may differ RE (both biased, e.g.).

--Still…wide range uses (e.g., comparing IV's or IV's & OLS) & has that wonderful property: *existence*.

```
. hausman FEmodel REmodel

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |    FEmodel      REmodel        Difference          S.E.
-------------+----------------------------------------------------------------
    L.sstran |    1.115156     1.212013       -.0968571        .0100676
   L2.sstran |     -.18442    -.2324231        .0480031        .0090748
        unem |    .0356034    -.0265803        .0621837         .014263
        left |   -.0005694    -.0005572       -.0000121        .0008179
       lftun |    .0002063     .0002636       -.0000573        .0001022
    growthpc |   -.1965328    -.1780914       -.0184414               .
    depratio |    .0411629     .0208042        .0203586        .0159829
        cdem |   -.0023557       .00122       -.0035756        .0016588
       trade |   -.0070589     .0035654       -.0106244        .0035097
     lowwage |    .0005358     .0080255       -.0074897        .0048194
         fdi |    .0136313    -.0308592        .0444905        .0121952
------------------------------------------------------------------------------
                           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(11) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =       27.13
                Prob>chi2 =      0.0044
                (V_b-V_B is not positive definite)
```

Hausman Test:

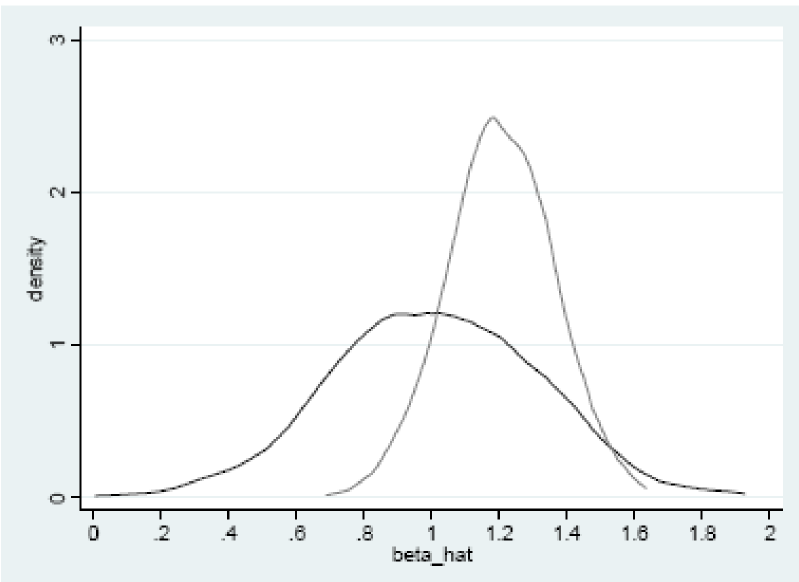
Good asymptotic properties and performs well in large samples, with low frequency of type-1 (false reject) errors (Baltagi 2001).

Many variants of Hausman Test have been developed that allow for serial correlation, non-stationarity, heteroskedasticity, other violations of typical assumptions: power analyses & MC studies show these augmentations of Hausman do work better in detecting differences in two estimators under the conditions for which they were developed.

But…

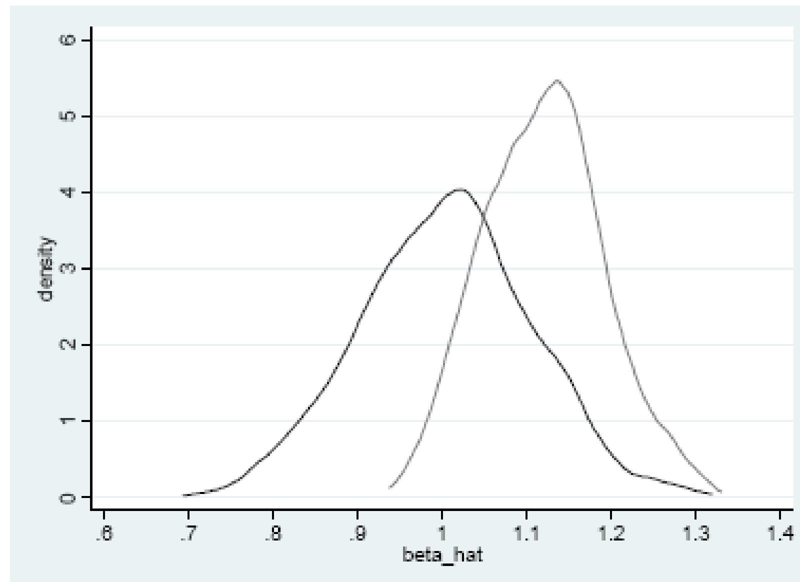
But, in small sample: inefficiency of Hausman test derives in good part from inefficiency of the always-consistent estimator in its tests. (More from Troeger: on how FE can be badly inefficient …)

Trade-off between efficiency and bias



black: FE; grey: RE
Settings : N=T=10, b/w-ratio=3, DGP u~N(0,1), 1000 repetitions, corr (u,x)=0.4, true beta=1

Trade-off between efficiency and bias



black: FE; grey: RE
Settings : N=T=70, b/w-ratio=3, DGP u~N(0,1), 1000 repetitions, corr (u,x)=0.4, true beta=1

Notice the apparent appreciable MSE advantage of RE over FE.

Suggests Hausman of FE vs. RE in trouble b/c FE very badly inefficient…

PART 2: Consequences of large sampling variation

Sampling variation of the FE estimates includes values that are 50 to 80 % smaller or larger than the true value

Only 8 percent of the 1000 coefficients (in the extreme tails) turn out to be statistically insignificant.

In an analysis with real world data these deviations from the true value can be quite substantial and create misleading inferences.

The estimated standard errors of the fixed effects estimates are about one third larger than the standard errors of the random effects estimates.

The fixed effects standard errors remain constant for all point estimates regardless of where in the distribution the point estimate lies.

Thus, point estimates that are far away from the true relationship might still appea statistically significant.

PART 2: Consequences for the Hausman-test

- Test results are influenced by the trade-off between bias and efficiency
- Hausman test is only powerful in the limit, since FE is consistent (in the limit) – difference of RE and FE estimates only result from biased RE estimates i.e., result only from biased RE in the limit $T \to \infty$
- In finite samples: differences can result from two sources: biased RE estimates and unreliable FE estimates (because of inefficient estimation)
- Hausman-test mirrors this trade-off: difference of RE and FE estimates / difference in asymptotic variance of RE and FE estimates

→ Test results should be especially unreliable if regressors are both correlated with the unit specific effects and rarely changing

…which is, of course, exactly where we'd most want them reliable…