Heterogeneity in TSCS

I. Notation, an almost most-general (linear) model:

$$y_{it} = \alpha_{it} + \mathbf{x}'_{it}\boldsymbol{\beta}_{it} + \varepsilon_{it};\ \ \boldsymbol{\varepsilon} \sim (\mathbf{0}, \boldsymbol{\Sigma});\ \ i = 1..N,\ t = 1..T,\ n = NT$$

A. Nothing necessarily changes:

1. If willing assume Gauss-Markov:

$$y_{it} = \alpha + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it};\ \ \boldsymbol{\varepsilon} \sim (\mathbf{0}, \sigma^2\mathbf{I})$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon};\ \ \boldsymbol{\varepsilon} \sim (\mathbf{0}, \sigma^2\mathbf{I})$$

$$Cov\left(\mathbf{X}, \boldsymbol{\varepsilon}\right) = \mathbf{0}$$

$$\left\{\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})\right\}$$

a) Last not nec for OLS (if absent: CLT for dists of ests); nec for ML=OLS

b) Nothing new; this C(N)LRM => OLS=BLUE.

(1) BLUE: Best Linear Unbiased Estimator =>

(2) Coefficient-Estimate Properties: unbiased, consistent, efficient

(3) V-Cov(b)-Estimate Properties: unbiased, consistent, efficient

2. Similarly, if want relax to: $\mathbf{y} = \mathbf{XB} + \boldsymbol{\varepsilon};\ \ \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\boldsymbol{\Omega});\ \ \ Cov\left(\mathbf{X}, \boldsymbol{\varepsilon}\right) = \mathbf{0}$

a) (w/ normality again as above) nothing new: G(N)LRM => GLS=BLUE, FGLS=asymptotically BLUE, where asymptotically BLUE means:

b) **FGLS Properties:** Consistent, Asymptotically Efficient

3. TSCS=collection of time-series, so all may know regarding TS models applies (w/ approp care to respect breaks b/w units; e.g., $y_{2,1}$(t-1)$\neq y_{1,T}$).

B. Thus, departure from C(N)LRM lies in plausibility key assumptions:

1. Parameter Stability:

$$y_{it} = \alpha + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it} \Rightarrow \left(\alpha_{it}, \boldsymbol{\beta}_{it}\right) = \left(\alpha, \boldsymbol{\beta}\right)\ \forall i, t$$

2. Spherical Errors (homoskedasticity+uncorrelated):

$$\boldsymbol{\varepsilon} \sim (\mathbf{0}, \sigma^2\boldsymbol{\Omega}) \text{ more plausible than } \boldsymbol{\varepsilon} \sim (\mathbf{0}, \sigma^2\mathbf{I})$$

3. Parameters Vary? Stochastic Component Variance-Covariance Structure Varies? => First Line of Defense, Always & Everywhere: MODEL IT!

C. 1st Defense: ***Model It!***™ (besides, likely this substance, not nuisance)

1. If, for example, expect some pattern nonsphericity, this likely because you expect, e.g., some *systematic*…

a) …variation in effect of x_{it} across i,t

(1) => what looks like heteroskedasticity if model effect as a constant, β

(2) => model the interactive (or group-wise varying) effect:

(a) $\beta_{it} = \gamma_0 + \gamma_1 z_{it} (+\phi_{it})$ (=> linear-interaction model)

(b) $\beta_{it} = \gamma_i (+\phi_{it})$, or $\beta_{it} = \gamma_t (+\phi_{it})$, or $\beta_{it} = \gamma_s (+\phi_{it})$ (dummy-interax)

b) …dependence of y_{it} on $y_{i,t-1}$, &/or y_{it} on y_{jt}

(1) => what looks like serial &/or spatial error-correlation if fail model the temporal &/or spatial dynamics in outcome, y

(2) => model the temporal &/or spatial/spatiotemporal dynamics:

(a) $y_{it} = \alpha_{\{i,t\}} + \beta_{\{i,t\}} x_{i,t} + \rho_{\{i,t\}} y_{i,t-1} + \varepsilon_{it}$

(b) $y_{it} = \alpha_{\{i,t\}} + \beta_{\{i,t\}} x_{i,t} \left(+\rho_{\{i,t\}} y_{i,t-1}\right) + \theta_{\{i,t\}} \sum_{j \neq i} w_{ij} y_{jt} + \varepsilon_{it}$

2. Implications of ***Model It!***™ Strategy; first, call all RHS: **XB**

a)

$$\hat{\boldsymbol{\beta}}_{OLS} = (\mathbf{X'X})^{-1}\mathbf{X'y} = (\mathbf{X'X})^{-1}\mathbf{X'}(\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon})$$

$$= (\mathbf{X'X})^{-1}\mathbf{X'X}\boldsymbol{\beta} + (\mathbf{X'X})^{-1}\mathbf{X'}\boldsymbol{\varepsilon}$$

$$= \boldsymbol{\beta} + (\mathbf{X'X})^{-1}\mathbf{X'}\boldsymbol{\varepsilon} \Rightarrow E\left(\hat{\boldsymbol{\beta}}\right) = \boldsymbol{\beta} \text{ if } E(\mathbf{X'}\boldsymbol{\varepsilon}) = 0 \text{ (as usual)}$$

b)

$$\mathbf{V}\left(\hat{\boldsymbol{\beta}}_{OLS}\right) = \mathbf{V}\left[(\mathbf{X'X})^{-1}\mathbf{X'y}\right] = (\mathbf{X'X})^{-1}\mathbf{X'V}(\boldsymbol{\varepsilon})\mathbf{X}(\mathbf{X'X})^{-1}$$

$$= (\mathbf{X'X})^{-1}\mathbf{X'}\sigma^2\mathbf{IX}(\mathbf{X'X})^{-1}$$

$$= \sigma^2(\mathbf{X'X})^{-1}\mathbf{X'X}(\mathbf{X'X})^{-1}$$

$$= \sigma^2(\mathbf{X'X})^{-1} \text{ (as usual)}$$

c) OLS=BLUE if model right(*spatial); what if imperfect/incomplete?

3. If model not right, or not enough, some decent properties may still hold. From theory-evaluation perspective, worry about unmodeled heterogeneity only insofar as failure to model adequately biases or otherwise worsens estimates of what can model/understand or certainty-estimates thereof.

4. For example, suppose use just linear-interaction, when linear-interaction w/ error (=random coefficients)

Truth: $y_{it} = x_{it}\left(\gamma_0 + \gamma_1 z_{it} + \phi_{it}\right) + \varepsilon_{it}$

Model: $y_{it} = x_{it}\left(\gamma_0^* + \gamma_1^* z_{it}\right) + \varepsilon_{it}^*$

$$\Rightarrow \hat{\boldsymbol{\beta}}_{OLS} = \left(\begin{bmatrix}\mathbf{x} & \mathbf{xz}\end{bmatrix}'\begin{bmatrix}\mathbf{x} & \mathbf{xz}\end{bmatrix}\right)^{-1}\begin{bmatrix}\mathbf{x} & \mathbf{xz}\end{bmatrix}'\mathbf{y}$$

$$= \left(\begin{bmatrix}\mathbf{x} & \mathbf{xz}\end{bmatrix}'\begin{bmatrix}\mathbf{x} & \mathbf{xz}\end{bmatrix}\right)^{-1}\begin{bmatrix}\mathbf{x} & \mathbf{xz}\end{bmatrix}'\left[\gamma_0\mathbf{x} + \gamma_1\mathbf{x}\cdot\mathbf{z} + \boldsymbol{\phi}\cdot\mathbf{x} + \boldsymbol{\varepsilon}\right]$$

a)

$$= \begin{bmatrix}\gamma_0 \\ \gamma_1\end{bmatrix} + \left(\begin{bmatrix}\mathbf{x} & \mathbf{xz}\end{bmatrix}'\begin{bmatrix}\mathbf{x} & \mathbf{xz}\end{bmatrix}\right)^{-1}\begin{bmatrix}\mathbf{x} & \mathbf{xz}\end{bmatrix}'\left[\boldsymbol{\phi}\cdot\mathbf{x} + \boldsymbol{\varepsilon}\right]$$

$$\Rightarrow E\left(\hat{\boldsymbol{\beta}}\right) = \boldsymbol{\beta} \text{ if } E\left(\boldsymbol{\phi}\cdot\mathbf{x}\right) = \mathbf{0},\ E\left(\mathbf{X}'\boldsymbol{\varepsilon}\right) = 0$$

$$\mathbf{V}\left(\hat{\boldsymbol{\beta}}_{OLS}\right) = \mathbf{V}\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}\right] = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{V}\left(\mathbf{x}\cdot\boldsymbol{\phi} + \boldsymbol{\varepsilon}\right)\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}$$

b)

$$= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\sigma^2\boldsymbol{\Omega}\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}$$

c) OLS=unbiased, but inefficient coefficients; wrong s.e.'s

d) However, some easy fixes (reviewed later…)

5. Example 2: unit-specific effects or coefficients, but manage to model only part of that parameter heterogeneity (mis-specification: OVB)

$$\text{Truth: } \mathbf{y} = \mathbf{X\beta} + \mathbf{Z\gamma} + \boldsymbol{\varepsilon}$$

$$\text{Model: } \mathbf{y} = \mathbf{X\beta}^* + \boldsymbol{\varepsilon}^*$$

$$\hat{\boldsymbol{\beta}}_{OLS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left(\mathbf{X\beta} + \mathbf{Z\gamma} + \boldsymbol{\varepsilon}\right)$$

a)
$$= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z\gamma} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}$$

$$= \boldsymbol{\beta} + \mathbf{F_{ZX}}\boldsymbol{\gamma} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon} \text{ where } \mathbf{F_{ZX}} \text{ is OLS } \mathbf{Z} \text{ on } \mathbf{X}$$

$$\mathbf{V}\left(\hat{\boldsymbol{\beta}}_{OLS}\right) = \mathbf{V}\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\right)$$

b)
$$\text{but } \widehat{\mathbf{V}\left(\hat{\boldsymbol{\beta}}_{OLS}\right)} = \hat{\mathbf{V}}\left(\hat{\mathbf{F}}_{\mathbf{ZX}}\hat{\boldsymbol{\gamma}} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\right)$$

$$\text{and } \widehat{\mathbf{V}\left(\hat{\boldsymbol{\beta}}_{OLS}\right)}_{OLS} = \widehat{\sigma^2}(\mathbf{X}'\mathbf{X})^{-1}$$

c) I.e., completely-standard omitted-variable bias (OVB).

D. Implications of *Model It!*™ Strategy (read flow-chart left-to-right)

Model It!™ Adequacy	Second-Moment & Inadequacy Variance-Covariance Structure		Implications for OLS Properties
Model E(**y**) Sufficient	V(**e**) Spherical		OLS is BLUE
	V(**e**) Nonspherical	Nonsphericity Unrelated **xx'**	OLS **b** unbiased, inefficient; OLS V(**b**) unbiased, inefficient
		Nonsphericity Related **xx'**	OLS **b** unbiased, inefficient; OLS V(**b**) biased, inefficient
Model E(**y**) Insufficient	Unmodeled **b** het unrelated **X**	V(**e**) Will be Nonspherical & Related **xx'**	OLS **b** unbiased, inefficient; OLS V(**b**) biased, inefficient
	Unmodeled **b** het related **X**	V(**e**) Will be Nonspherical & Related **xx'**	OLS **b** biased, inefficient; OLS V(**b**) biased, inefficient

1. Advantages:
   a) Gives maximal leverage estimating parameters
   b) Consistent w/ general theories
   c) BLUE, *iff* this right model…
2. Disadvantages:
   a) Might not be right model.
   b) What might go wrong (specification error; omitted-variables):
      (1) Heterogeneity [SEE TROEGER's LIST AT END]
      (2) Nonsphericity: $V(\boldsymbol{\varepsilon}) = \sigma^2\boldsymbol{\Omega} \neq \sigma^2\mathbf{I}$

White's General Test:

c. White test if form of Heteroskedasticity is unknown:

H0: $V\left[\varepsilon_i \mid x_i\right] = \sigma^2$

Ha: $V\left[\varepsilon_i \mid x_i\right] = \sigma_i^2$

Estimate the model under H0
Compute squared residuals: e_i^2

Use squared residuals as dependent variable of auxiliary regression: RHS: all regressors, their quadratic forms and interaction terms:

$$e_i^2 = \delta_0 + \delta_1 x_{i2} + \ldots + \delta_{k-1} x_{ik} + \delta_{k-1} x_{i2}^2 + \delta_{k+1} x_{i2} x_{i3} + \ldots + \delta_q x_{ik}^2 + \xi_i$$

Compute White statistic from R² of auxiliary regression: $n * R^2 \xrightarrow{a} \chi^2_{(q)}$

Use one-sided test and check if n*R² is larger than 95% quantile of Chi²

Many, Many Heteroskedasticity-Test Strategies:

<u>Szroeter's</u> class of tests: $S = \frac{\sum_{i=1}^{NT} h(x_i) e_i^2}{\sum_{i=1}^{NT} e_i^2}$ where $h(x_i)$ is some weight increasing in x_i

King (1982) suggests h(x) = rank(x).

`htest` (downloadable enhanced `hettest`) `szroeter` gives a Q-statistic transformation of S such that, under homoscedasticity, Q is approximately N(0,1) distributed (Judge 1985:452).

<u>Goldfeld-Quandt</u>: sort e_i by some var. Take high & low sets, $\mathbf{e}_1$ & $\mathbf{e}_2$, (some evidence more power discard middle set), & stat: $\frac{\mathbf{e}_1'\mathbf{e}_1/(n_1-k)}{\mathbf{e}_2'\mathbf{e}_2/(n_2-k)} = F \sim F_{n_1-k,n_2-k}$

<u>Glesjer's</u>: Wald or $NT \times R^2$ set of coefficients in: $\ln e_i^2 = \delta_0 + \delta_1 z_{1i} + \ldots + \delta_{kz} z_{ki} + \upsilon_i$

a. Breusch-Pagan LM test for known form of Heteroskedasticity: groupwise

$$LM = \frac{T}{2}\sum_{i=1}^{n}\left(\frac{s_i^2}{s^2} - 1\right)^2$$

s_i^2 =sum of group-specific squared residuals

s^2 = OLS residuals

H0: homoskedasticity ~ Chi² with n-1 degrees of freedom
LM-test assumes normality of residuals, not appropriate if assumption not

b. Likelihood Ratio Statistic

Residuals are computed using MLE (e.g. iterated FGLS, OLS loss of power)

$$-2\ln(\lambda) = (NT)\ln\left(\sigma^2\right) - \Sigma\left(T\ln\left(\sigma_i^2\right)\right) \sim \chi^2\left(dF = n-1\right)$$

LM ARCH-Tests (test-statistic $T \times R^2 \sim \chi^2_{(S)}$. For example:

$$e_t^2 = \delta_0 + \delta_1 x_{1_t} + \ldots + \delta_k x_{k_t} + \sum_{s=1}^{S} \lambda_s e_{t-s}^2$$

Tests Spatial Patterns Correlation (**ELAB TOMORROW…**)

- "Testing" (measuring) Spatial Association:

$$\text{Moran's I: } I = \frac{N}{S}\frac{\boldsymbol{\varepsilon}'\mathbf{W}\boldsymbol{\varepsilon}}{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}, \text{ where } S = \sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}, \text{ if row-nrmlzd: } I = \frac{\boldsymbol{\varepsilon}'\mathbf{W}\boldsymbol{\varepsilon}}{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}$$

  - See also Anselin's LISA…

- LM's appropriate for LS resids (F&H OxfHndbk):

$$LM_{\rho} = \frac{\hat{\sigma}_{\varepsilon}^{2}\left(\hat{\boldsymbol{\varepsilon}}'\mathbf{W}\mathbf{y}/\hat{\sigma}_{\varepsilon}^{2}\right)^{2}}{G+T\hat{\sigma}_{\varepsilon}^{2}}, \text{ and } LM_{\lambda} = \frac{\left(\hat{\boldsymbol{\varepsilon}}'\mathbf{W}\hat{\boldsymbol{\varepsilon}}/\hat{\sigma}_{\varepsilon}^{2}\right)^{2}}{T}$$

  - But these test v. *iid*, & over-reject…

- Robust LM tests appropriate for SAR v. SAE:

$$LM_{\rho}^{*} = G^{-1}\hat{\sigma}_{\varepsilon}^{2}\left(\hat{\varepsilon}'Wy/\hat{\sigma}_{\varepsilon}^{2} - \hat{\varepsilon}'W\hat{\varepsilon}/\hat{\sigma}_{\varepsilon}^{2}\right)^{2}, \text{ and } LM_{\lambda}^{*} = \frac{\left(\hat{\varepsilon}'W\hat{\varepsilon}/\hat{\sigma}_{\varepsilon}^{2} - \left[T\hat{\sigma}_{\varepsilon}^{2}\left(G+T\hat{\sigma}_{\varepsilon}^{2}\right)^{-1}\right]\hat{\varepsilon}'Wy/\hat{\sigma}_{\varepsilon}^{2}\right)^{2}}{T\left[1-\frac{T\hat{\sigma}_{\varepsilon}^{2}}{G+T\hat{\sigma}_{\varepsilon}^{2}}\right]}$$

- Robust Joint Test: $LM_{\rho\lambda} = LM_{\lambda} + LM_{\rho}^{*} = LM_{\rho} + LM_{\lambda}^{*}$

Tests Temporal Correlation (**ELAB TOMORROW…**)

Many, many of these also.

My favorite…

LM (Time-)Serial-Correlation Tests (test-statistic $T \times R^2 \sim \chi^2_{(S)}$):

$$e_t = \delta_0 + \delta_1 x_{1_t} + \ldots + \delta_k x_{k_t} + \sum_{s=1}^{S} \lambda_s e_{t-s}$$

…for same reasons Glesjer's/White's favorite among hettests:

Uses familiar, intuitive strategy. Constructive. Powerful.
Plus: valid following LDV model (include among *x*) unlike, e.g., DW; flexible (detects either MA or AR).

E. Redresses, mostly partial &/or imperfect, of non-spherical V(**e**) deficiencies in implementation of the ***Model It!***™ strategy

1. "Robust" or "Sandwich" Variance-Covariance Estimators

a) Key Insight:

(1) **Ω** has ½n(n+1)>n elements/parameters; however, for consistent v-cov est:

(2) need consistent est only of **X'ΩX**, which only ½k(k+1) elements total.

b) Proper $\mathbf{V}\left(\hat{\boldsymbol{\beta}}_{LS}\right)=\sigma^2(\mathbf{X'X})^{-1}\mathbf{X'\Omega X}(\mathbf{X'X})^{-1}$ differs from OLS $\mathbf{V}\left(\hat{\boldsymbol{\beta}}_{OLS}\right)=\sigma^2(\mathbf{X'X})^{-1}$, only insofar as **X'ΩX** differs from **X'X**, which is only insofar as elements of **Ω** covary with elements of **xx'**, i.e. ω_{ij} w/ x's, x^2's, &/or cross-products of x's.

c) Visualizing the matrix multiplication, we see that v-cov estimates using LS formula are off by factor of: $\sum_{i,j,s,t} e_{it}e_{js}\left(\mathbf{x}_{it}\mathbf{x}'_{js}\right)-\sum_{i,t} e_{it}^2\mathbf{I}_k$.

d) ∴, we can *fix* our v-cov estimates, i.e. render them *robust*, i.e., consistent, to presence of certain pattern of nonsphericity by replacing **X'ΩX** in formula w/ some $\sum_{i,j,s,t} e_{it}e_{js}\left(\mathbf{x}_{it}\mathbf{x}'_{js}\right)$ configured to reflect that nonsphericity pattern.

$$\hat{\mathbf{V}}_s\left(\hat{\boldsymbol{\beta}}\right)=(\mathbf{X'X})^{-1}\widehat{\mathbf{X'\Omega X}}(\mathbf{X'X})^{-1}\equiv(\mathbf{X'X})^{-1}\hat{\mathbf{Q}}(\mathbf{X'X})^{-1}$$

2. Cases of (Robust/Sandwich) Consistent V-Cov(**b**) Estimators:

a) Pure Heteroskedasiticity (White's): $\hat{\mathbf{Q}} = \sum_i \sum_t e_{it}^2 \left(\mathbf{x}_{it} \mathbf{x}'_{it} \right)$

b) Pure Het. & (Time) Auto-Correlation (HAC) (Newey-West):

$$\hat{\mathbf{Q}} = \sum_i \sum_t e_{it}^2 \left(\mathbf{x}_{it} \mathbf{x}'_{it} \right) + \sum_i \sum_t \left(\sum_{s=1}^{L} w_t e_{it} e_{i,t-s} \left(\mathbf{x}_{it} \mathbf{x}'_{i,t-s} + \mathbf{x}_{i,t-s} \mathbf{x}'_{i,t} \right) \right)$$

where L=max lag-length considered appreciable & $w_t = 1 - \frac{t}{L-1}$

c) Panel Heteroskedasiticity (Rob): $\hat{\mathbf{Q}} = \sum_i \left(\sum_t e_{it}^2 \left(\mathbf{x}_{it} \mathbf{x}'_{it} \right) \right)$

d) Panel Het. & (Time) Auto-Correlation (Rob):

$$\hat{\mathbf{Q}} = \sum_i \left(\sum_t e_{it}^2 \left(\mathbf{x}_{it} \mathbf{x}'_{it} \right) \right) + \sum_i \left(\sum_t \left(\sum_{s=1}^{L} w_t e_{it} e_{i,t-s} \left(\mathbf{x}_{it} \mathbf{x}'_{i,t-s} + \mathbf{x}_{i,t-s} \mathbf{x}'_{i,t} \right) \right) \right)$$

where L=max lag-length considered appreciable & $w_t = 1 - \frac{t}{L-1}$

e) Panel Het. & Contemporaneous (Spatial) Corr. (Beck-Katz PCSE):

$$\hat{\mathbf{Q}} = \mathbf{X}'\left(\frac{\mathbf{E}'\mathbf{E}}{T} \otimes \mathbf{I}_T\right)\mathbf{X}$$

where **E**=the $T \times N$ matrix estimated residuals

f) Many others possible. Several "cluster" types, e.g., designed for various multilevel/hierarchical data structures (i.e., a panel 'random-effect' structure)

$$\hat{V}\left(\hat{\boldsymbol{\beta}}\right) = \frac{1}{N-k}(\mathbf{X}'\mathbf{X})^{-1}\left(\sum_{j=1}^{n_c}\left\{\left[\sum_{i=1}^{n_j} e_i\mathbf{x}_i\right]\left[\sum_{i=1}^{n_j} e_i\mathbf{x}_i\right]'\right\}\right)(\mathbf{X}'\mathbf{X})^{-1}$$

where n_j=# obs. i in macro-level (cluster) j, & n_c=# clusters

g) Note: excepting Newey-West, asymptotics for these tend to be in *N* or in some function of *N* and *T*, not in *T*. Many may not work well in TSCS.

h) Some not assuredly "well-behaved" in estimation, so various *kludges*.

i) Small-sample adjusts been suggested for each; may be key in TSCS.

j) Small-sample adjusts been suggested for each; may be key in TSCS.

(1) E.g., "[For White's,] Davidson and MacKinnon (1993: 554) strongly suggest a finite-sample correction of replacing e_i^2 by $e_i^2/(1-\mathbf{x}_i(\mathbf{X'X})^{-1}\mathbf{x}_i)$, which scales estimated squared residuals by their variance, or of multiplying by $N/(N-k)$, which inflates estimates by a factor reflecting the number of regressors as a percentage of degrees of freedom. Accumulating simulation work favors their suggestion."

(2) E.g., "As with [White's], a finite-sample (degrees-of-freedom) correction, $[n_c/(n_c-1)][(N-1)/(N-k)]$, is suggested. This inflates standard errors as there, but now multiplicatively further, by a declining function of J. Again, simulations strongly support using such adjustments.

(3) Not clear to me (from `help vcetype`) if and which of these adjustments Stata applies as defaults or options (need manuals).

(4) From many MC's I've seen on Cluster, PCSE, etc., more attention to these small-sample adjustments would be a good thing in TSCS esp.

3. FGLS: Feasible Generalized-Least-Squares

a) Consistent V-Cov Ests only address "inconsistency" of s.e.'s, do not address bias or efficiency of coeff estimates (although require consistent coefficient-estimates for formal properties) or "unbiasedness" and "efficiency" of s.e.'s.

b) To improve efficiency coeff (& s.e.) estimates—still not directly or formally redress any bias concerns arising from other problems, OVB e.g., and still reliant on "first-stage" consistency—we can parameterize and estimate $\hat{\mathbf{\Omega}}$, use it to transform the data to such that C(N)LRM applies.

c) Example: *Parks-Kmenta FGLS* for TSCS:

(1) Panel-specific AR(1) in residuals=> *N* parameters

(2) Panel-specific σ_i^2 => *N* parameters

(3) Dyad-specific σ_{ij} => *N(N-1)* parameters (n.b., symmetric)

(4) =>N(N+1) pars=>unless T>>2N, inadvisable (Beck-Katz '95)

(5) NOTE: Could offer more theoretically structured (& thereby parametrically reduced) structure non-sphericity pattern => greater efficiency & better small-sample properties. E.g., just contemp corr. => N(N-1) parameters needs T>>N.

d) FGLS properties: consistent & asymptotically efficient.

FGLS: given consistent est $\hat{\mathbf{\Omega}}$, let $\mathbf{P} \equiv \hat{\mathbf{\Omega}}^{-\frac{1}{2}}$, then:

$$\mathbf{Py} = \mathbf{PX\beta} + \mathbf{P\varepsilon} \Rightarrow \hat{\boldsymbol{\beta}}_{FGLS} = \left[(\mathbf{PX})' (\mathbf{PX}) \right]^{-1} (\mathbf{PX})' \mathbf{Py}$$

$$\hat{\boldsymbol{\beta}}_{FGLS} = \left[\mathbf{X'P'PX} \right]^{-1} \mathbf{XP'Py} = \left[\mathbf{X'}\hat{\mathbf{\Omega}}^{-1}\mathbf{X} \right]^{-1} \mathbf{X'}\hat{\mathbf{\Omega}}^{-1}\mathbf{y}$$

$\Rightarrow$ consistent and asympt'ly efficient if $\mathrm{C}(\mathbf{X}, \boldsymbol{\varepsilon}) = 0$

$$\widehat{V\left(\hat{\boldsymbol{\beta}}_{FGLS}\right)}_{FGLS} = \left[\mathbf{X'}\hat{\mathbf{\Omega}}^{-1}\mathbf{X} \right]^{-1} \mathbf{X'}\hat{\mathbf{\Omega}}^{-1}\mathbf{V}(\mathbf{y})\hat{\mathbf{\Omega}}^{-1}\mathbf{X} \left[\mathbf{X'}\hat{\mathbf{\Omega}}^{-1}\mathbf{X} \right]^{-1}$$

$$\widehat{V\left(\hat{\boldsymbol{\beta}}_{FGLS}\right)}_{FGLS} = \sigma^2 \left[\mathbf{X'}\hat{\mathbf{\Omega}}^{-1}\mathbf{X} \right]^{-1} \mathbf{X'}\hat{\mathbf{\Omega}}^{-1}\hat{\mathbf{\Omega}}\hat{\mathbf{\Omega}}^{-1}\mathbf{X} \left[\mathbf{X'}\hat{\mathbf{\Omega}}^{-1}\mathbf{X} \right]^{-1}$$

$$\widehat{V\left(\hat{\boldsymbol{\beta}}_{FGLS}\right)}_{FGLS} = \sigma^2 \left[\mathbf{X'}\hat{\mathbf{\Omega}}^{-1}\mathbf{X} \right]^{-1} \mathbf{X'}\hat{\mathbf{\Omega}}^{-1}\mathbf{X} \left[\mathbf{X'}\hat{\mathbf{\Omega}}^{-1}\mathbf{X} \right]^{-1}$$

$$\widehat{V\left(\hat{\boldsymbol{\beta}}_{FGLS}\right)}_{FGLS} = \sigma^2 \left[\mathbf{X'}\hat{\mathbf{\Omega}}^{-1}\mathbf{X} \right]^{-1}$$

$\Rightarrow$ "consistent and asympt'ly efficient" (as above)

Troeger's List of Locuses of Model Heterogeneity

1. Different intercepts:
   first difference models, fixed effects model

2. Different coefficients:
   random coefficient model, SUR model, IA effects

3. Time dependent slopes:
   IA effects

4. Different lag structures:
   no textbook solution available

5. Different dynamics:
   no textbook solution available

II. Textbooks emphasize only 1-3 (A-C) below, and 2 (B) only indirectly:

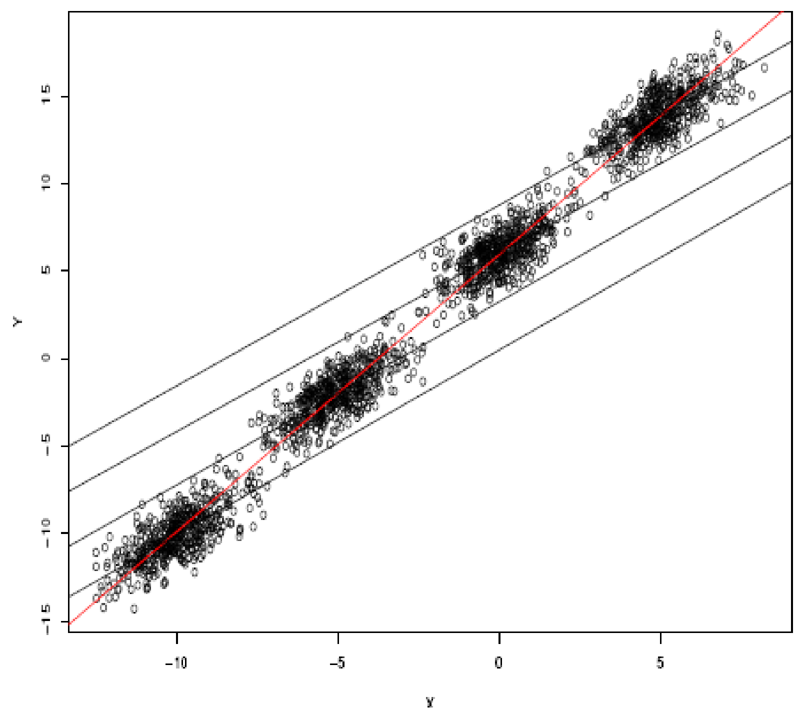
A. Unobserved (unmodeled) Unit (e.g., country) Effects: $\alpha_i \neq \alpha$

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

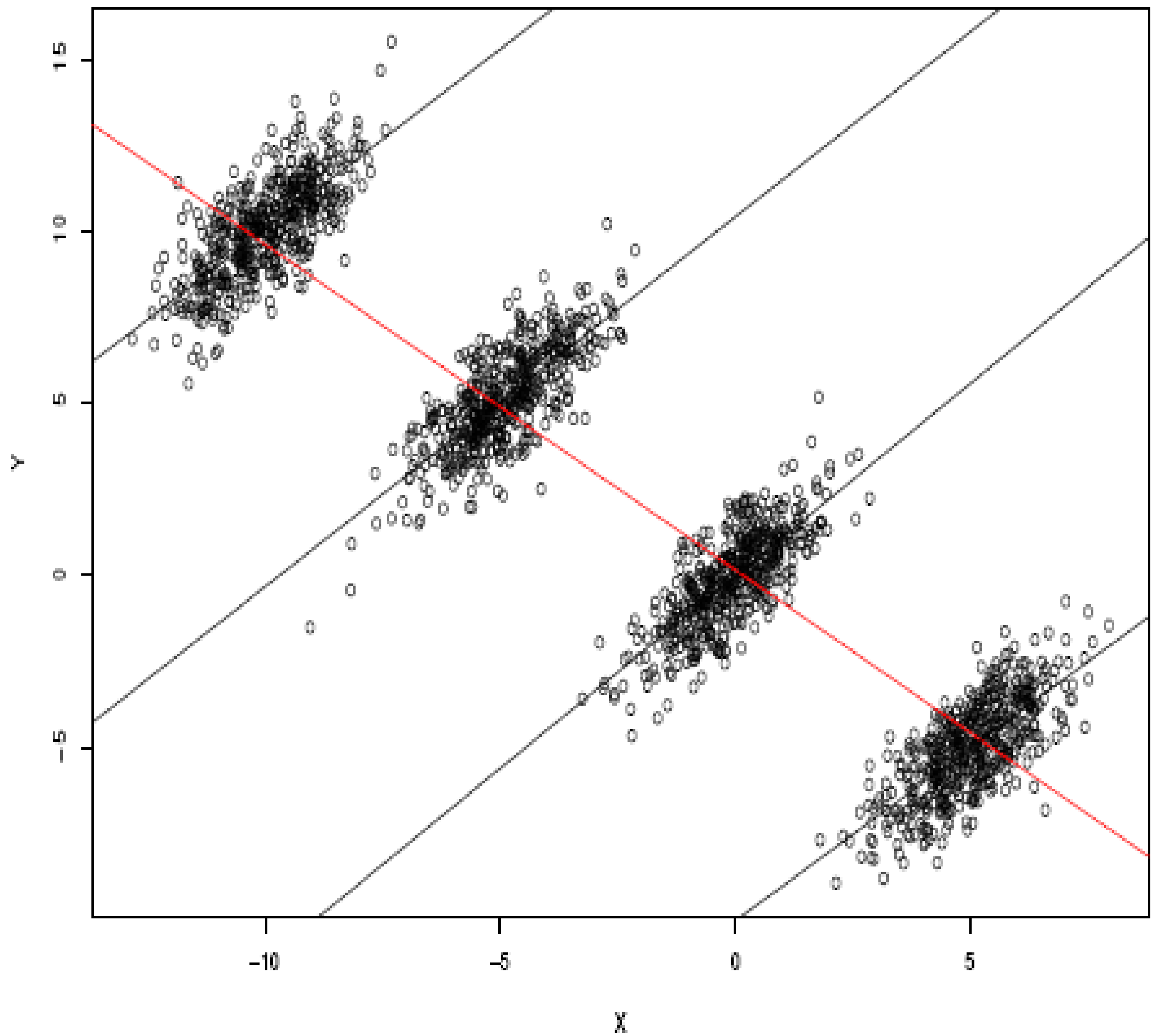
C. Unobserved (unmodeled) Coefficient Variability: $\beta_{i \text{ \&/or } t} \neq \beta$

III. Examples: graphs of heterogeneity in α

A. *Inflation bias* (positive β_x, positive corr. $\bar{x}_i$ & omitted u_i)

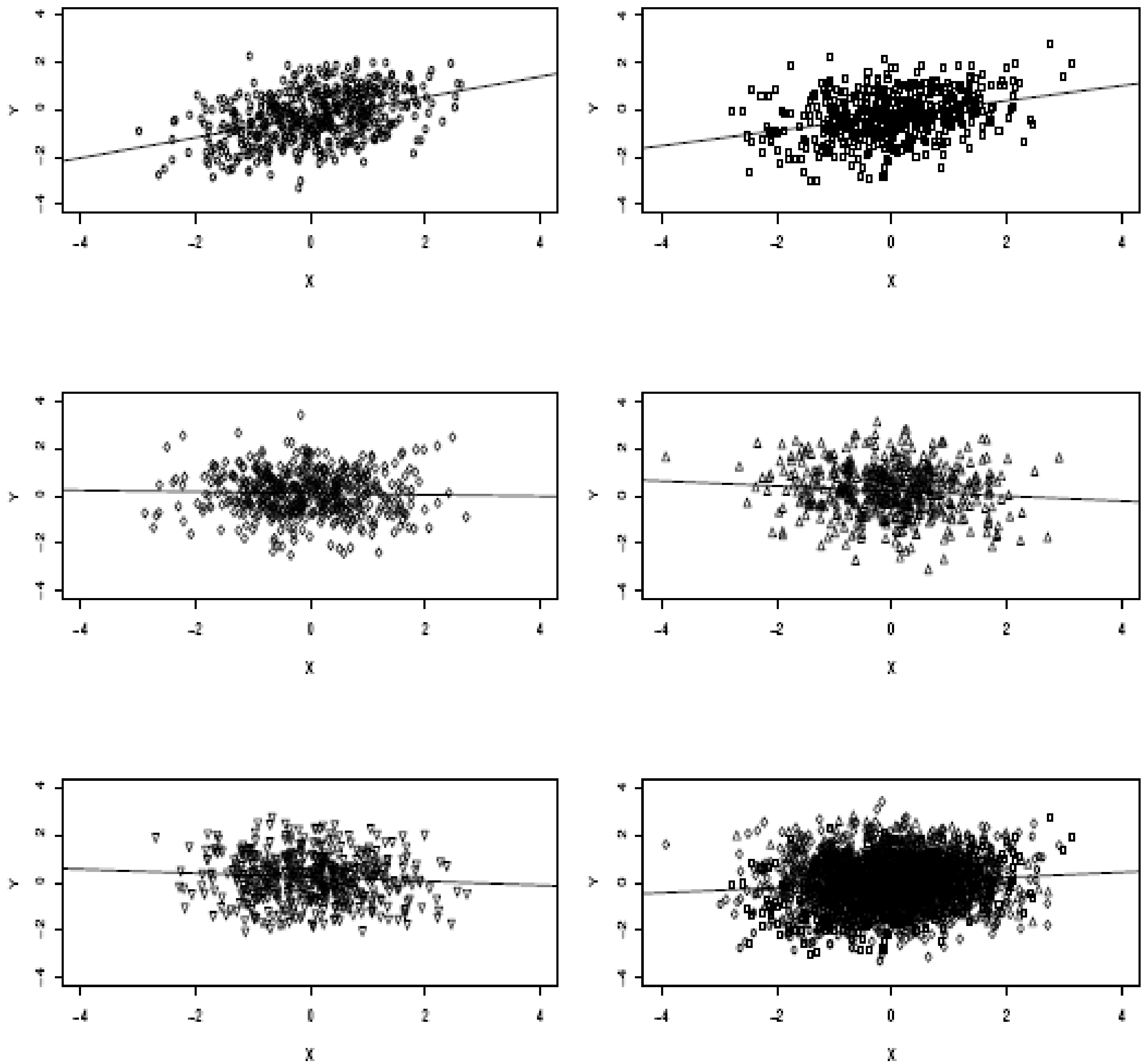


B. *Sign reversal*: positive β_x, strong negative corr. $\overline{x}_i$ & omitted u_i. (*Attenuation bias* if more moderately negative corrproportionately.)



C. If no corr. $\overline{x}_i$ & u_i, no bias, though still ineff & s.e.'s likely wrong.

D. Examples: graphs of heterogeneity in β_i



E. Examples: Heterogeneous Dynamics

1. Unit-Specific AR(IMA) models… May be very important account potentially heterogeneous dynamics:

2. Esp. for slow-moving &/or rarely-changing independent-variables, good estimates depend critically on specifying correct lag structure.

3. Well-known in TS analysis, but, since determining & estimating unit-specific dynamics onerous, most researchers either do not lag IV's or choose arbitrary, uniform lags (mostly one-period). Can have big conseq's:

4. P/T/M (from "Much Left to Do") illustrate using significance of LEFT:

   a) Shifts in value LEFT are frequent but persistent.

5. If institutions matter, may delay policy reactions with less govt autonomy.

6. No generally accepted indicator of opt. lag-length; cands: *t*-, R^2, AIC, BIC

7. P/T/M use un-weighted composite index of these.

8. Result: 11 countries where a change in government has an immediate effect on government spending; 4 countries, Australia, Austria, Germany and Ireland – one-year lag; 2 countries, Italy and Denmark, a two-year lag; 2 countries, Finland and Netherlands, a three-year lag

	model 3.1 uniform lags	model 3.2 optimized lag
Unemployment	1.2146 (0.883) ***	1.2168 (0.0860) ***
GDP per Capita Growth	-0.2604 (0.0403) ***	-0.2608 (0.0401) ***
Dependency Ratio	-0.7384 (0.1267) ***	-0.6655 (0.1229) ***
LEFT (no lags)	0.0002 (0.0041)	
LEFT (optimized lags)		0.0083 (0.0038) **
Christian Democrat Portf.	-0.0418 (0.0093) ***	-0.0338 (0.0088) ***
Trade Openness	0.0507 (0.0290) *	0.0515 (0.0285) *
Low Wage Imports	-0.1488 (0.0436) ***	-0.1567 (0.0420) ***
Foreign Direct Investment	-0.0910 (0.0951)	-0.1217 (0.8604)
N	529	524
R^2	.940	.944
Wald χ^2	13035.32	12310.68
prob> χ^2	0.0000	0.0000
PCSE	yes	Yes
time dummies	no	No
country dummies	yes	Yes

Panel analyses react sensitively to miss-specified lag-structures.

It is theoretically not convincing to assume uniform lags for different units.

Thus:

The precise lag-length should be determined, preferably by theoretical derivation of a hypothesis or empirically by criteria like BIC or AIC

Rejecting a hypothesis because the estimated coefficient turned out to be insignificant while in fact the researcher has wrongly assumed uniform lags would mean blaming the theory for the failures of the methodology.

TROEGER's SUMMARY CONCLUSION

Textbook heterogeneity (FE) is likely to be the least important heterogeneity!

FE "solution" very costly, especially if between-variation is low or theory predicts level effects.

Unit-specific slopes very easy to do. Use fixed coefficients models. (Random coefficient models Beck/Katz 2007 are very a-theoretical.)

Unit-specific dynamics pretty complicated. Perhaps a no go.