From C&G(N)LRM to Models Panel/TSCS
Heterogeneity in Panel/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: $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon};\ \ \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\boldsymbol{\Omega});\ \ \ Cov(\mathbf{X}, \boldsymbol{\varepsilon}) = \mathbf{0}$

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

b) **FGLS Properties:** Given consistent $\hat{\boldsymbol{\Omega}}$, 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:

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

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!

II. From the Most-General (& Inestimable) Form Down:

A. Most-General Form:

$$y_{it} = f_{it}\left(\mathbf{x}_{it}, \boldsymbol{\beta}_{it}, \varepsilon_{it}\right);\ \boldsymbol{\varepsilon} \sim (\mathbf{0}, \boldsymbol{\Sigma}_{it});\ i = 1..N,\ t = 1..T,\ n = NT$$

Notes: $\mathbf{x}^1 = \mathbf{i};\ \boldsymbol{\beta}_0 = \alpha;\ \mathbf{x}$ may contain time-space lags $\mathbf{x}$ or $\mathbf{y}$.

1. Parameters=K+½$(NT)^2$+½NT *per function*, ***per observation***!

2. MASSIVELY under-identified => impose structure to reduce parameterization; from where? Theory & Substance (& Assumption)

B. Virtually always assume:

1. $f_{it}(\bullet) = f(\bullet)\ \forall i,t$: same $f(\cdot)$ relates $\mathbf{X}_{it}, \boldsymbol{\beta}_{it}, \boldsymbol{\varepsilon}_{it}$ to $\mathbf{y}$ in all obs; may be stronger than needed; could parameterize changes $f_{it}(\bullet)$ or allow it to vary across but not within groups of obs $\{it\}$.

2. $\boldsymbol{\Sigma}_{it} = \boldsymbol{\Sigma}\ \forall i,t$: each obs draw from a common multivariate distribution, i.e., a MV dist w/ same var-covar across obs; may be stronger than needed…

3. Parameters: K(NT)+½$(NT)^2$+½NT per NT obs.=>K+½(NT+1) per obs. Still **way**, **way** too many.

C. Next, can assume constant coefficient-vector: $\boldsymbol{\beta}_{it} = \boldsymbol{\beta} \ \forall i, t$

1. Parameters: K/(NT)+½(NT+1) per obs. Still **<u>way</u>** too many.

2. May be stronger than needed, can allow: $\boldsymbol{\beta}_{it} = g\left(\mathbf{z}, \boldsymbol{\gamma}, \eta_{it}\right)$, with # of parameters<NT-K-parameters($\boldsymbol{\Sigma}$).

D. Still must reduce parameterization $\boldsymbol{\Sigma}$.

1. Fully general var-covar structure not estimable; nothing to learn from history &/or comparison if insist all unique.

2. Plausible/practically-realistic variance-covariance structures:

a) *Sphericity*: from $\sigma^2\boldsymbol{\Omega}$ to $\sigma^2\mathbf{I}$ => from ½(NT)^2+½NT to 1 parameter.

b) *Panel Heteroskedasticity*: from $V\left(\varepsilon_{it}\right) = \sigma_i^2$ => N parameters.

c) *Serial Correlation*: (AR1) $\varepsilon_{it} = \rho\varepsilon_{i,t-1} + \upsilon$ (2 params), or $\varepsilon_{it} = \rho_i\varepsilon_{i,t-1}$ $(N+1$ pars)

d) *Parks-Kmenta*: panel het + each TS unique AR1 + unique $\sigma_{ij} = \sigma_{ji} \ \forall ij$, though common for all $T \Rightarrow 2N+½N(N\text{-}1) \Rightarrow$ needs <u>LOTS</u> of T.

e) *Many other plausible parameterizations…*

$\boldsymbol{\Omega}=$

$$
\left[\begin{array}{ccccccccccccccccccccc}
\omega_{11}^{2} & \omega_{11,12} & \omega_{11,13} & \cdots & \omega_{11,1T} & \omega_{11,21} & \omega_{11,22} & \omega_{11,23} & \cdots & \omega_{11,2T} & \omega_{11,31} & \omega_{11,32} & \omega_{11,33} & \cdots & \omega_{11,3T} & & \omega_{11,N1} & \omega_{11,N2} & \omega_{11,N3} & \cdots & \omega_{11,NT} \\
\omega_{12,11} & \omega_{12}^{2} & & & \vdots & \omega_{12,21} & \omega_{12,22} & & & \vdots & \omega_{12,31} & \omega_{12,32} & & & \vdots & & \omega_{12,N1} & \omega_{12,N2} & & & \vdots \\
\omega_{13,11} & & \omega_{13}^{2} & & \vdots & \omega_{13,21} & & \omega_{13,23} & & \vdots & \omega_{13,31} & & \omega_{13,33} & & \vdots & \cdots & \omega_{13,N1} & & \omega_{13,N3} & & \vdots \\
\vdots & & & \ddots & \vdots & \vdots & & & \ddots & \vdots & \vdots & & & \ddots & \vdots & & \vdots & & & \ddots & \vdots \\
\omega_{1T,11} & \cdots & \cdots & \cdots & \omega_{1T}^{2} & \omega_{1T,21} & \cdots & \cdots & \cdots & \omega_{1T,2T} & \omega_{1T,31} & \cdots & \cdots & \cdots & \omega_{1T,3T} & & \omega_{1T,N1} & \cdots & \cdots & \cdots & \omega_{1T,NT} \\
\omega_{21,11} & \omega_{21,12} & \omega_{21,13} & \cdots & \omega_{21,1T} & \omega_{21}^{2} & \omega_{21,22} & \omega_{21,23} & \cdots & \omega_{21,2T} & & & & & & & & & & & \\
\omega_{22,11} & \omega_{22,12} & & & \vdots & \omega_{21,22} & \omega_{22}^{2} & & & \vdots & & & & & & & & & & & \\
\omega_{23,11} & & \omega_{23,13} & & \vdots & \omega_{21,23} & & \omega_{23}^{2} & & \vdots & & & & & & & & & \vdots & & \\
\vdots & & & \ddots & \vdots & \vdots & & & \ddots & \vdots & & & & & & & & & & & \\
\omega_{2T,11} & \cdots & \cdots & \cdots & \omega_{2T,1T} & \omega_{21,2T} & \cdots & \cdots & \cdots & \omega_{2T}^{2} & & & & & & & & & & & \\
\omega_{31,11} & \omega_{31,12} & \omega_{31,13} & \cdots & \omega_{31,1T} & & & & & & \omega_{31}^{2} & \omega_{31,32} & \omega_{31,33} & \cdots & \omega_{31,3T} & & & & & & \\
\omega_{32,11} & \omega_{32,12} & & & \vdots & & & & & & \omega_{32,31} & \omega_{32}^{2} & & & \vdots & & & & & & \\
\omega_{33,11} & & \omega_{33,13} & & \vdots & & & & & & \omega_{33,31} & & \omega_{33}^{2} & & \vdots & & & & \vdots & & \\
\vdots & & & \ddots & \vdots & & & & & & \vdots & & & \ddots & \vdots & & & & & & \\
\omega_{3T,11} & \cdots & \cdots & \cdots & \omega_{3T,1T} & & & & & & \omega_{3T,31} & \cdots & \cdots & \cdots & \omega_{3T}^{2} & & & & & & \\
 & & \vdots & & & & & & & & & & & & & \ddots & & & \vdots & & \\
\omega_{N1,11} & \omega_{N1,12} & \omega_{N1,13} & \cdots & \omega_{N1,1T} & & & & & & & & & & & & \omega_{N1}^{2} & \omega_{N1,N2} & \omega_{N1,N3} & \cdots & \omega_{N1,NT} \\
\omega_{N2,11} & \omega_{N2,12} & & & \vdots & & & & & & & & & & & & \omega_{N2,N1} & \omega_{N2}^{2} & & & \vdots \\
\omega_{N3,11} & & \omega_{N3,13} & & \vdots & & & \cdots & & & & & \cdots & & & \cdots & \omega_{N3,N1} & & \omega_{N3}^{2} & & \vdots \\
\vdots & & & \ddots & \vdots & & & & & & & & & & & & \vdots & & & \ddots & \vdots \\
\omega_{NT,11} & \cdots & \cdots & \cdots & \omega_{NT,1T} & & & & & & & & & & & & \omega_{NT,N1} & \cdots & \cdots & \cdots & \omega_{NT}^{2}
\end{array}\right]
$$

III. From simplest model upward (parsimony principle):

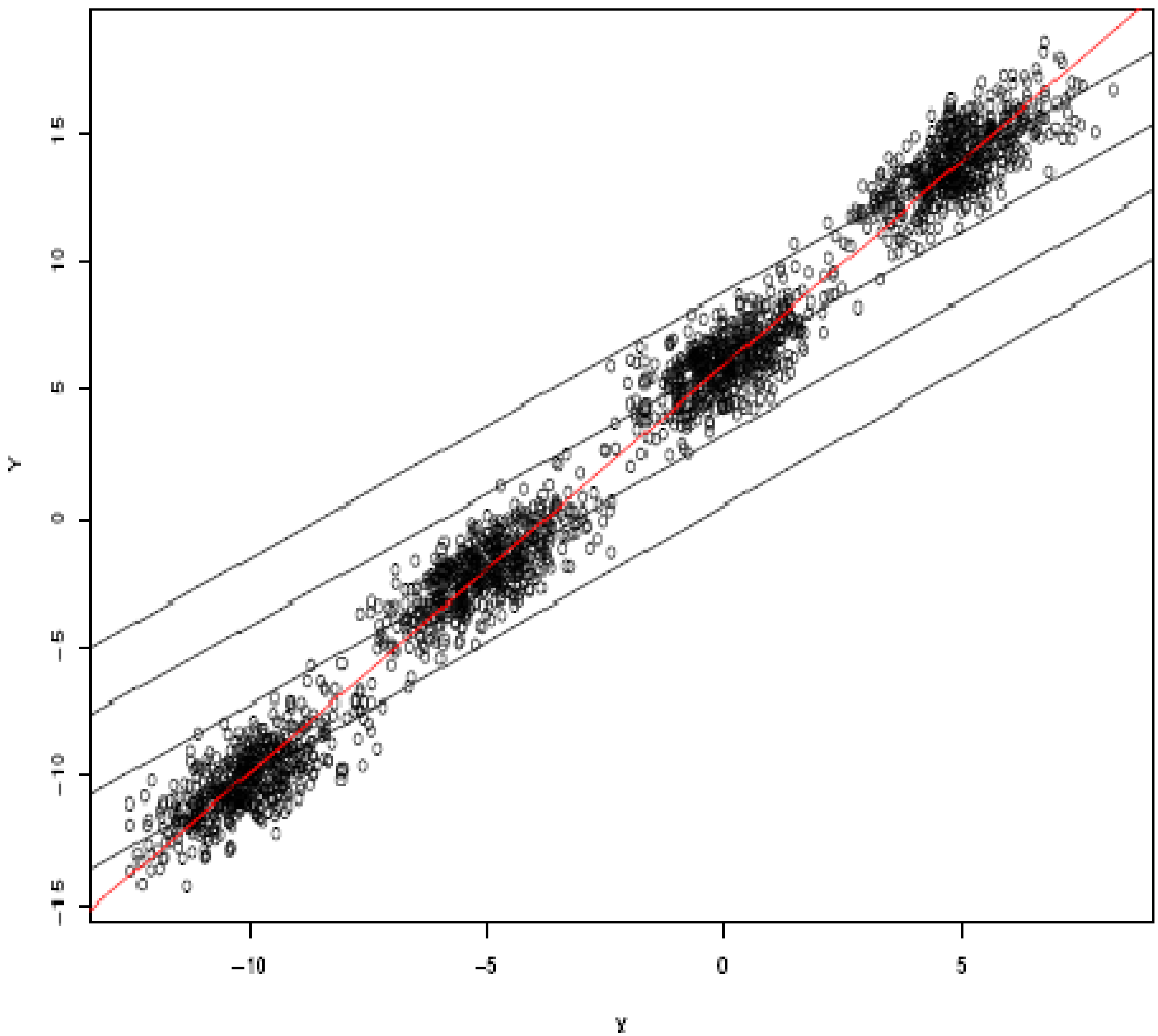
$$y_{it} = \alpha + \beta x_{it} + \varepsilon_{it}, \quad \mathrm{V}(\varepsilon_{it}) = \sigma^2$$

A. Pool all data and estimate by OLS

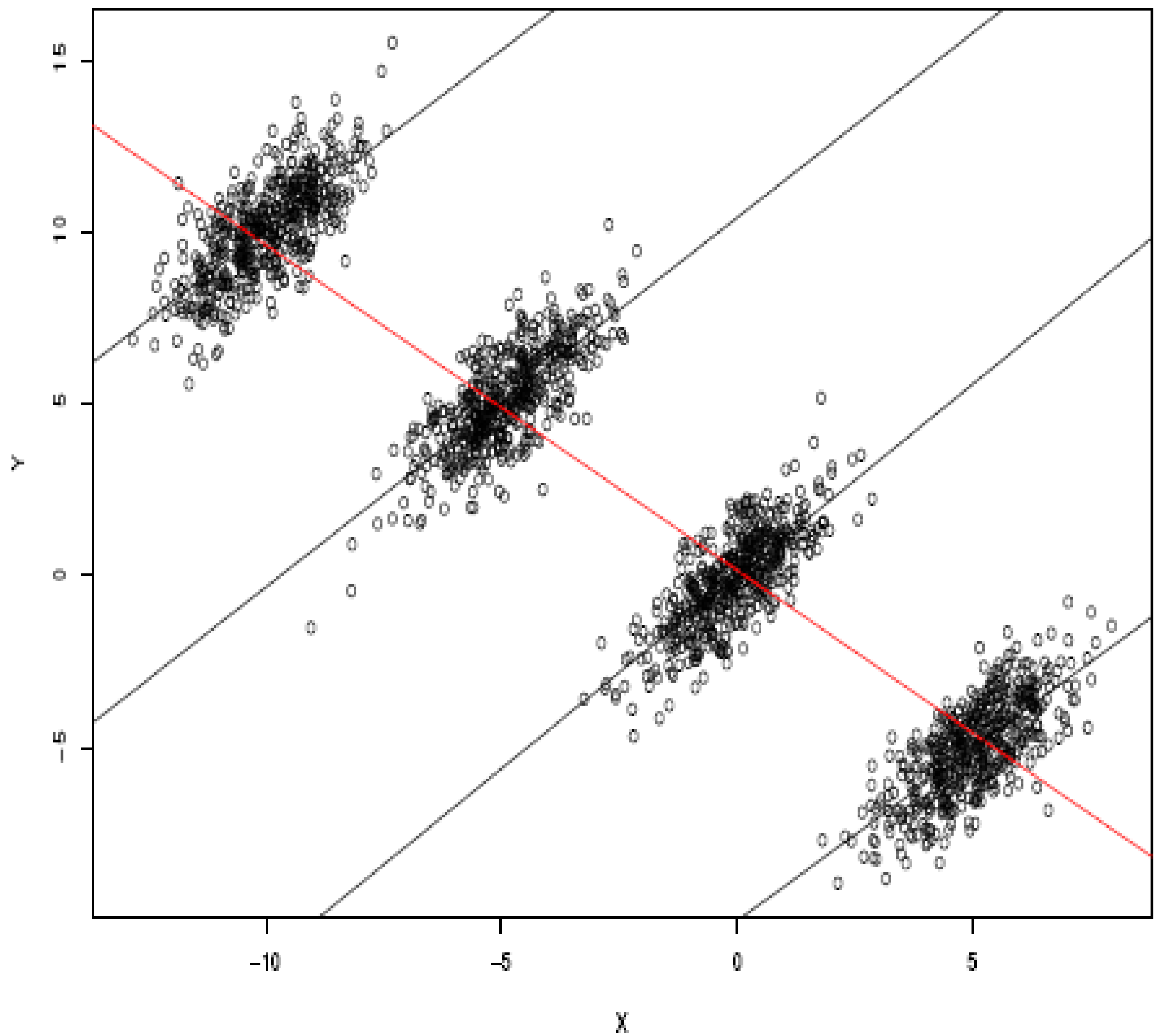
1. Advantages:
   a) Gives maximal leverage estimating parameters
   b) Consistent w/ general theories
   c) BLUE, *iff* this right model…
2. Disadvantages:
   a) Might not be right model.
   b) What might go wrong (specification error; omitted-variables):
      (1) Nonsphericity: $\mathrm{V}(\boldsymbol{\varepsilon}) = \sigma^2\boldsymbol{\Omega} \neq \sigma^2\mathbf{I}$
      (2) Unobserved (unmodeled) Unit (e.g., country) Effects: $\alpha_i \neq \alpha$
      (3) Unobserved (unmodeled) Time (sub-unit) Effects: $\alpha_t \neq \alpha$
      (4) Unobserved (unmodeled) Coefficient Variability: $\beta_{i \text{ \&/or } t} \neq \beta$

c) Examples: graphs of heterogeneity bias

(1) <u>*Inflation bias*</u> (positive β_x, positive correlation of $\overline{x}_i$ & the omitted u_i)



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



(3) If no corr. $\overline{\mathbf{x}}_i$ & u_i, no bias, though still inefficient & s.e.'s likely wrong.

B. 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 (+rndm coeff.))

(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: $\mathbf{X\beta}$

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

b)
$$\mathbf{V}(\hat{\boldsymbol{\beta}}_{OLS}) = \mathbf{V}\left[(\mathbf{X'X})^{-1}\mathbf{X'y}\right] = (\mathbf{X'X})^{-1}\mathbf{X'V}(\boldsymbol{\varepsilon})\mathbf{X}(\mathbf{X'X})^{-1}$$
$$= (\mathbf{X'X})^{-1}\mathbf{X'}\sigma^2\mathbf{IX}(\mathbf{X'X})^{-1}$$
$$= \sigma^2(\mathbf{X'X})^{-1}\mathbf{X'X}(\mathbf{X'X})^{-1}$$
$$= \sigma^2(\mathbf{X'X})^{-1} \text{ (as usual)}$$

c) OLS=BLUE if model right(*temporal, **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)

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

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

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

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

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)

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

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

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

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

but is instead: $\mathbf{V}\left(\hat{\boldsymbol{\beta}}_{OLS}^{mis}\right) = \mathbf{V}\left(\mathbf{F_{ZX}}\boldsymbol{\gamma} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\right)$,

b) and OLS estimates merely: $\widehat{\mathbf{V}\left(\hat{\boldsymbol{\beta}}_{OLS}^{mis}\right)}_{OLS} = \widehat{\sigma^2}(\mathbf{X}'\mathbf{X})^{-1}$.

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

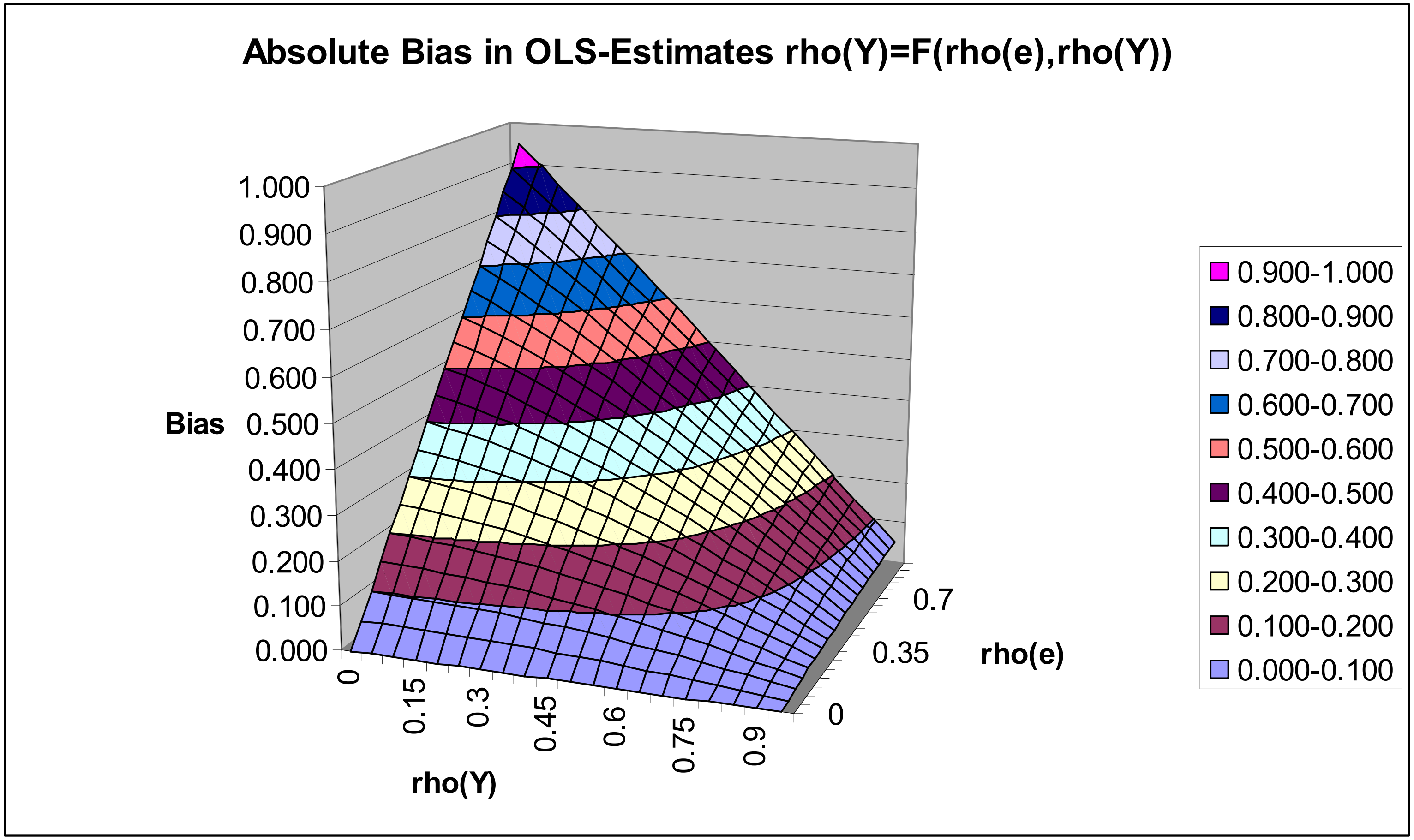
6. (Time-)Serial Dependence:

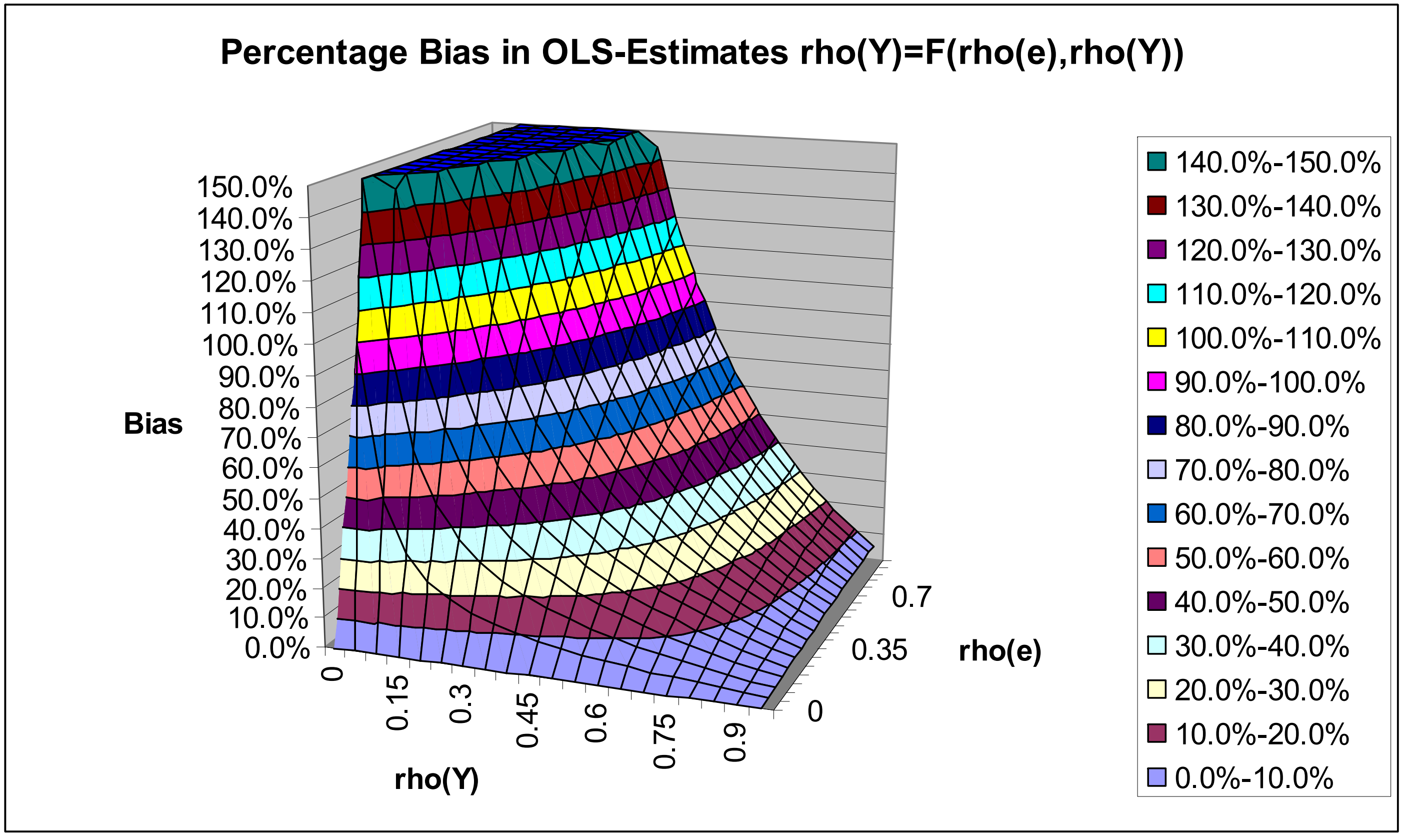
a) If temporal dynamics specified in systematic component sufficiently (no residual/stochastic-component corr. remains, which testable), OLS $\rightarrow^{A}$ BLUE.

b) If insufficient, OLS inconsistent, but still:

(1) Magnitude of the problem: $E\left(\hat{\rho}_y\right) = \rho_y + \rho_\varepsilon \left(1 - \rho_y^2\right) \Big/ \left(1 + \rho_\varepsilon \rho_y\right)$

(2) And can (partially) address s.e. part of problem (as shall see…)





c) Possible to model temporal dependence in both y & ε by NLS:

$$\mathbf{y}_t = \mathbf{X}_t\boldsymbol{\beta} + \rho_y \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t; \quad \boldsymbol{\varepsilon}_t = \rho_\varepsilon \boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\upsilon}_t; \quad \boldsymbol{\upsilon}_t \sim \left(\mathbf{0}, \sigma_\upsilon^2 \mathbf{I}\right)$$

$$\Rightarrow \mathbf{y}_t = \mathbf{X}_t\boldsymbol{\beta} + \rho_y \mathbf{y}_{t-1} + \rho_\varepsilon \boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\upsilon}_t$$

$$= \mathbf{X}\boldsymbol{\beta} + \rho_y \mathbf{y}_{t-1} + \rho_\varepsilon \left(\mathbf{y}_{t-1} - \mathbf{X}_{t-1}\boldsymbol{\beta} - \rho_y \mathbf{y}_{t-2}\right) + \boldsymbol{\upsilon}$$

$$= \left(\mathbf{X}_t - \rho_\varepsilon \mathbf{X}_{t-1}\right)\boldsymbol{\beta} + \rho_y \mathbf{y}_{t-1} + \rho_\varepsilon \left(\mathbf{y}_{t-1} - \rho_y \mathbf{y}_{t-2}\right) + \boldsymbol{\upsilon}$$

d) Note, however, indeterminacy in total=systematic+stochastic, so any two of possible lag y, lag x, lag e => third, etc.

7. Spatial Dependence:

a) Situation more complic. OLS inconsistent even if model spatial-dep fully because spatial-lag is endog. (covaries with residual, even asmpt'ly).

b) However, still generally better to model than to omit it, & we'll talk about redressing the simultaneity in this case later.

8. Limited & Qualitative Dependent-Variables:

a) Analogous but more comp b/c systematic & stochastic nonseparable

b) Addresses/redresses also more complicated.

9. Summary:

a) If can model thry/subst reason for deviation from C(N)LRM, in TSCS data or elsewhere, do so, &, if/insofar as successful, strategy optimal in all regards.

b) Insofar as possible, "Model It!" in model of first-moment, $E(\mathbf{y})$, i.e., systematic component; for two reasons:

(1) Usually, the theory/substance in question regards *systematic* component

(2) Observationally, only info we have on stochastic component (i.e., second moment) conditional on info in & model of first moment (systematic component)

(3) May not be possible or theoretically/substantively correct; could have theory/substance info about second moment (variance). E.g.:

(a) DepVar=average varying # lower-level outcomes => V(ε)~1/#.

(b) Thry/Subst e.g.: education (or information) not affect response; rather reliability or accuracy or theoretical-explicability of response=> V(ε)=f(edu)

(c) Still "Model It!" (in 2nd moment, so, i.e., model reduced parameterization of **Ω**).

(4) As seen, insofar as fail model fully deviations CLRM, probs arise essentially as OVB in worst cases, but as "just" inefficiency & wrong s.e.'s else, so…

C. Implications of *Model It!*™ Strategy (read as 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 to **X**	V(**e**) Will be Nonspherical & Related **xx'**	OLS **b** biased, inefficient; OLS V(**b**) biased, inefficient

1. Advantages:
   a) max leverage est 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 VT's list below]
      (2) Nonsphericity: $V(\boldsymbol{\varepsilon}) = \sigma^2\boldsymbol{\Omega} \neq \sigma^2\mathbf{I}$

IV. 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

A. Textbooks emphasize only 1-3 below, & 2 only indirectly:

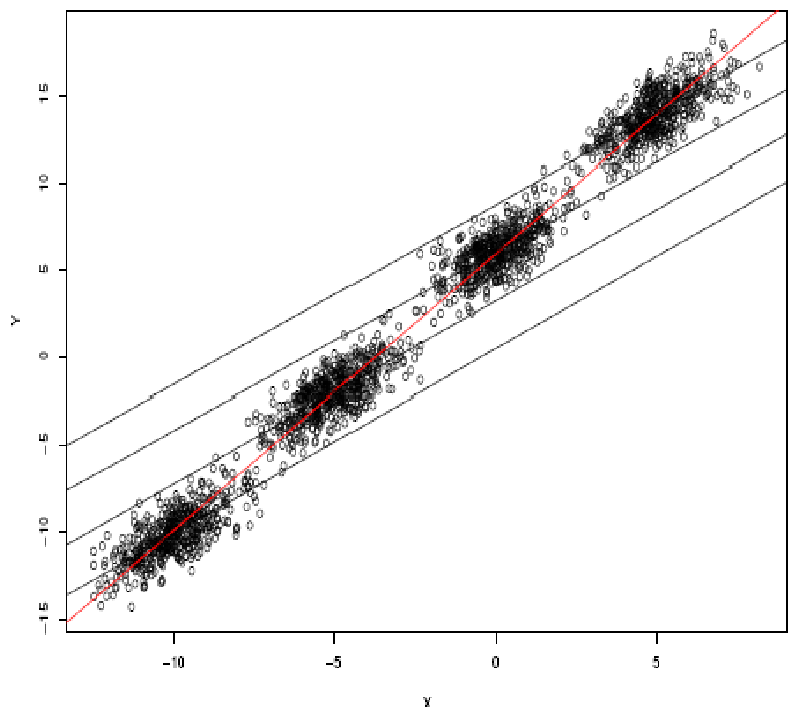
1. Unobserved (unmodeled) Unit (e.g., country) Effects: $\alpha_i \neq \alpha$

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

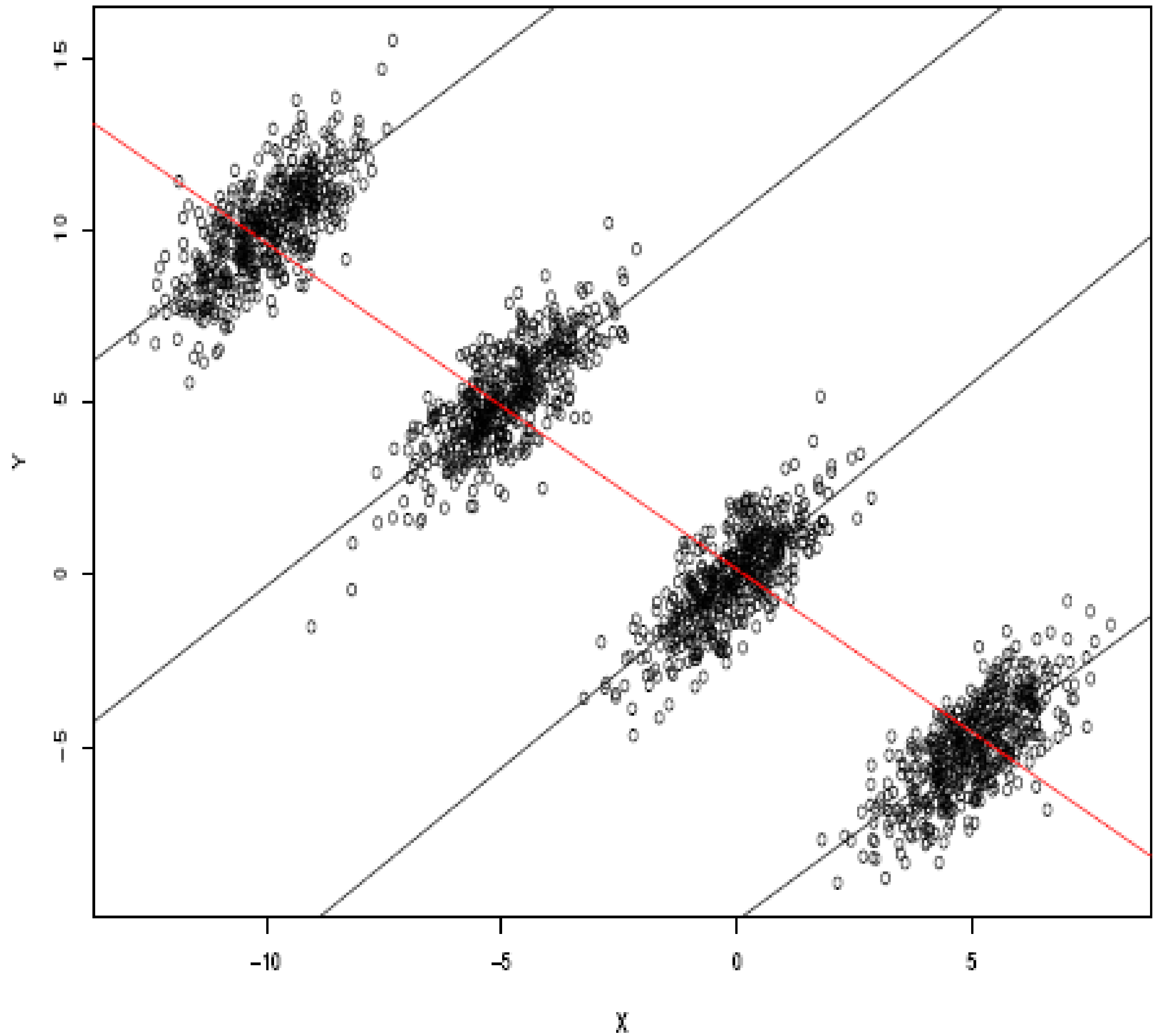
3. Unobserved (unmodeled) Coefficient Variability: $\beta_{i \text{ \&/or } t} \neq \beta$

B. Examples: graphs of heterogeneity in α

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

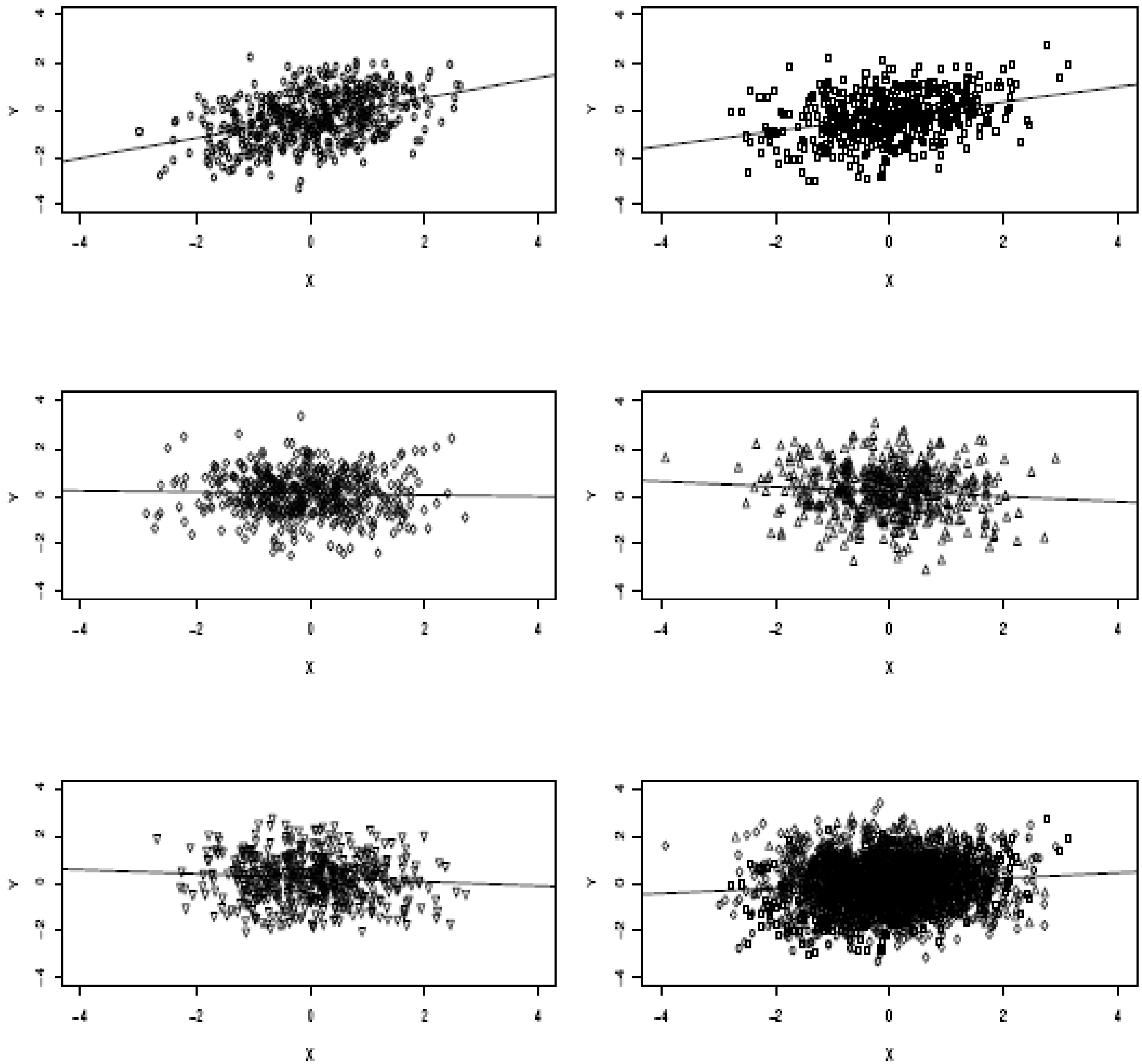


2. *Sign reversal*: positive β_x, strong negative corr. $\bar{x}_i$ & omitted u_i. (*Attenuation bias* if more moderately negative corrproportionately.)



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

4. Examples: graphs of heterogeneity in β_i



5. Examples: Heterogeneous Dynamics

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

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

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

b) P.T.&Manow (*EJPR* 2005) illustrate using significance of LEFT (***next***)

(1) Shifts in value LEFT are frequent but persistent.

(2) If institutions matter, may delay policy reactions with less govt autonomy.

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

(4) P/T/M use un-weighted composite index of these.

(5) 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.

C. VT'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.