From C&G(N)LRM to Models for TSCS

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

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

A. Nothing necessarily changes (all data are TSCS data):

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:** Consistent, Asymptotically Efficient

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

some $\boldsymbol{\varepsilon} \sim (\mathbf{0}, \sigma^2\boldsymbol{\Omega})$ 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: $\mathbf{x}^1 = 1$; $\boldsymbol{\beta}^1 = \alpha$; $\mathbf{x}$ 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*() 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 distribution with same variance-covariance 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 **way** too many.

2. May be stronger than needed, can allow: $\boldsymbol{\beta}_{it} = g(\mathbf{z}, \boldsymbol{\gamma}, \eta_{it})$, 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 $V(\varepsilon_{it}) = \sigma_i^2$ => 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 $T \Rightarrow 2N+½N(N-1) \Rightarrow$ needs LOTS of T.

e) *Many other plausible parameterizations…*

Ω=

$$
\begin{bmatrix}
\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{bmatrix}
$$

III. From simplest model upward (parsimony principle):

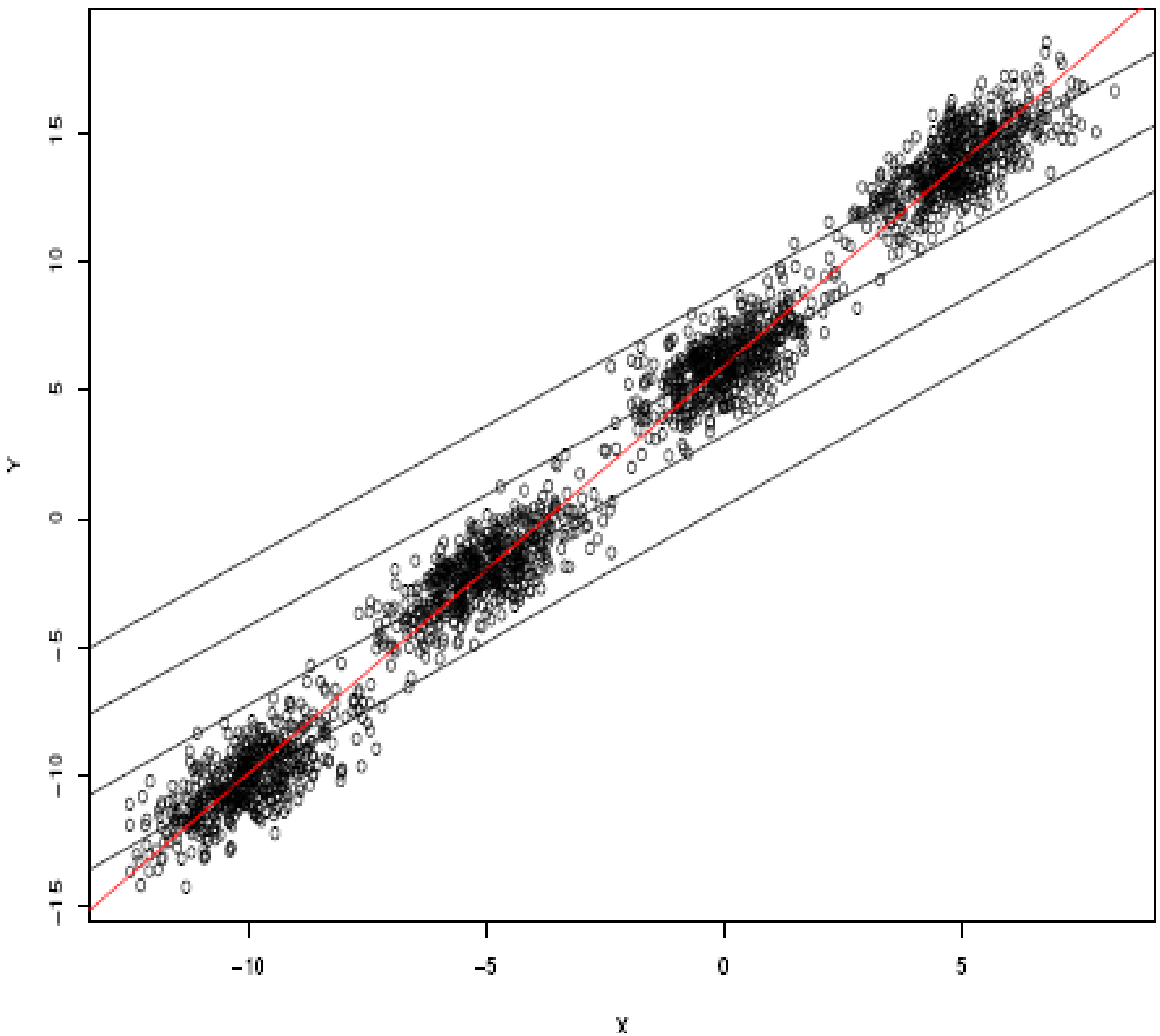
$$y_{it} = \alpha + \beta x_{it} + \varepsilon_{it}, \quad \mathrm{V}\left(\varepsilon_{it}\right) = \sigma^2$$

A. Pool all data and estimate by OLS

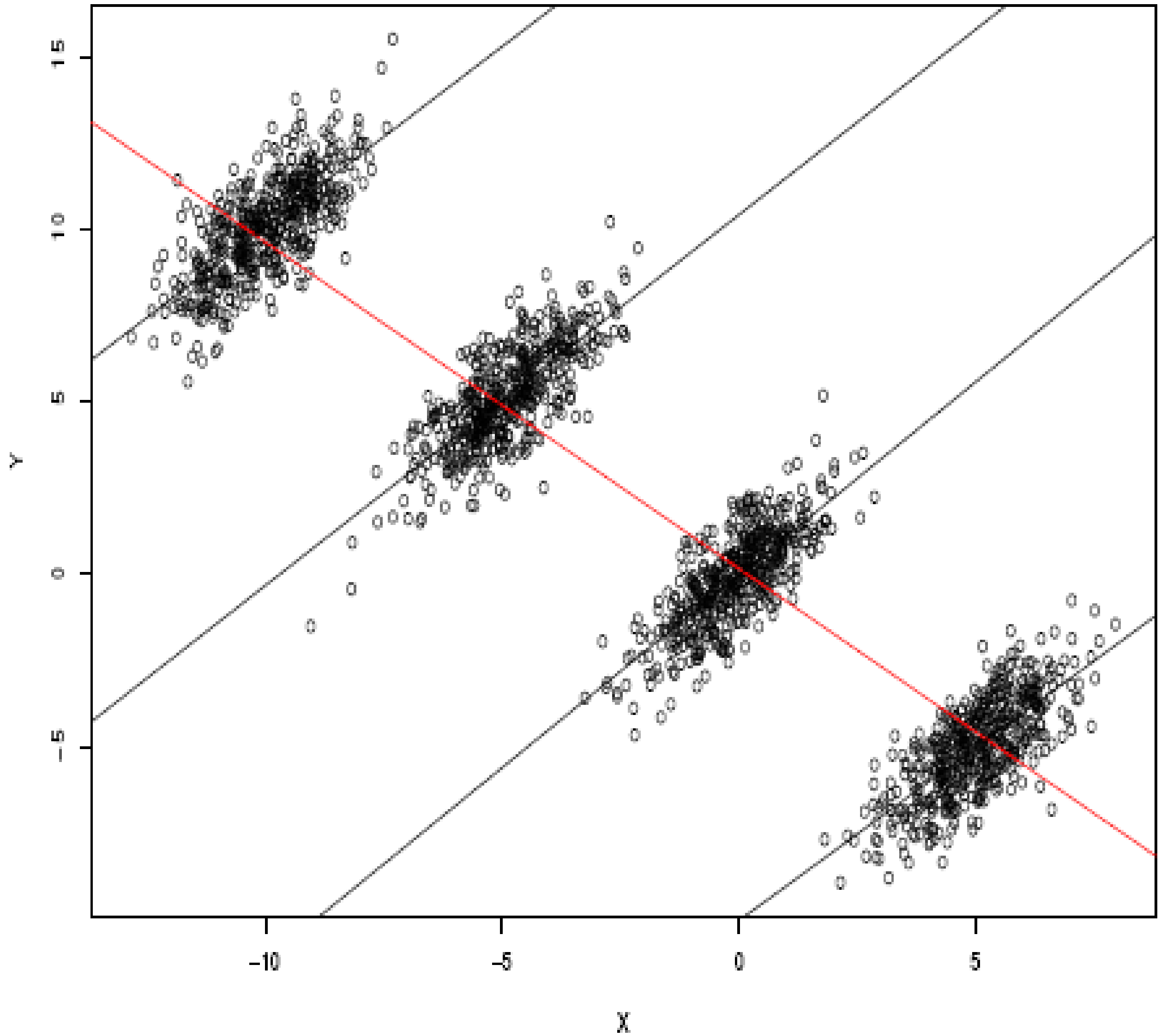
1. Advantages:
    a) Gives maximal leverage estimating parameters
    b) Consistent w/ general theories
    c) BLUE, *iff* this right model…
2. Disadvantages:
    a) Might not be right model.
    b) What might go wrong (specification error; omitted-variables):
        (1) Nonsphericity: $\mathrm{V}(\boldsymbol{\varepsilon}) = \sigma^2\boldsymbol{\Omega} \neq \sigma^2\mathbf{I}$
        (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) *Inflation bias* (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. $\overline{x}_i$ & 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 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: $\mathbf{X\beta}$

a)

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

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

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

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[(*spatial)]; what if imperfect/incomplete?

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

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

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

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

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

a)

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

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

b)

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

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

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

d) However, some easy (and now familiar…) fixes…

4. E.g., unit-specific effects or coefficients, but 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)
$$= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z\gamma} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}$$

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

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

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

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

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

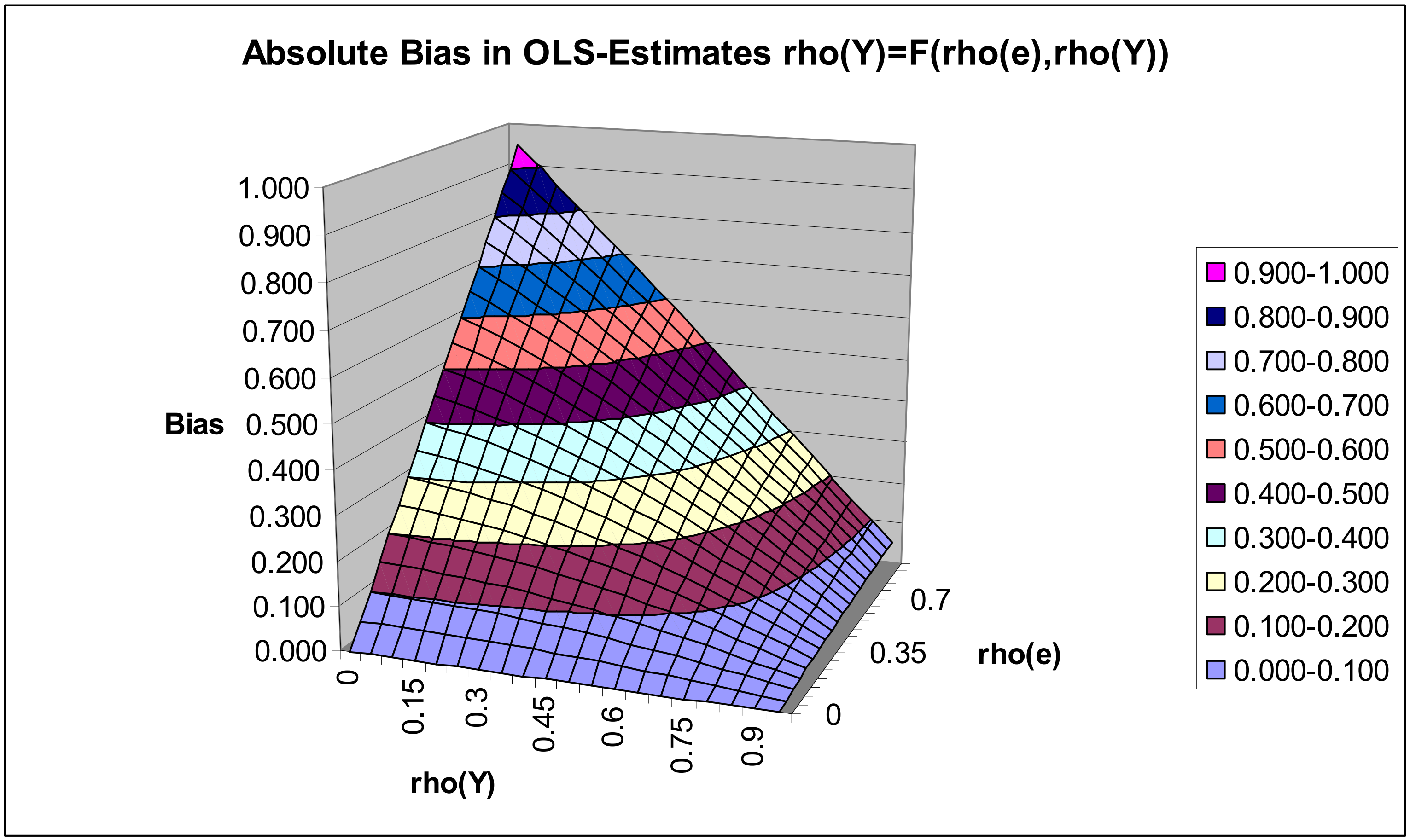
5. (Time-)Serial Dependence:

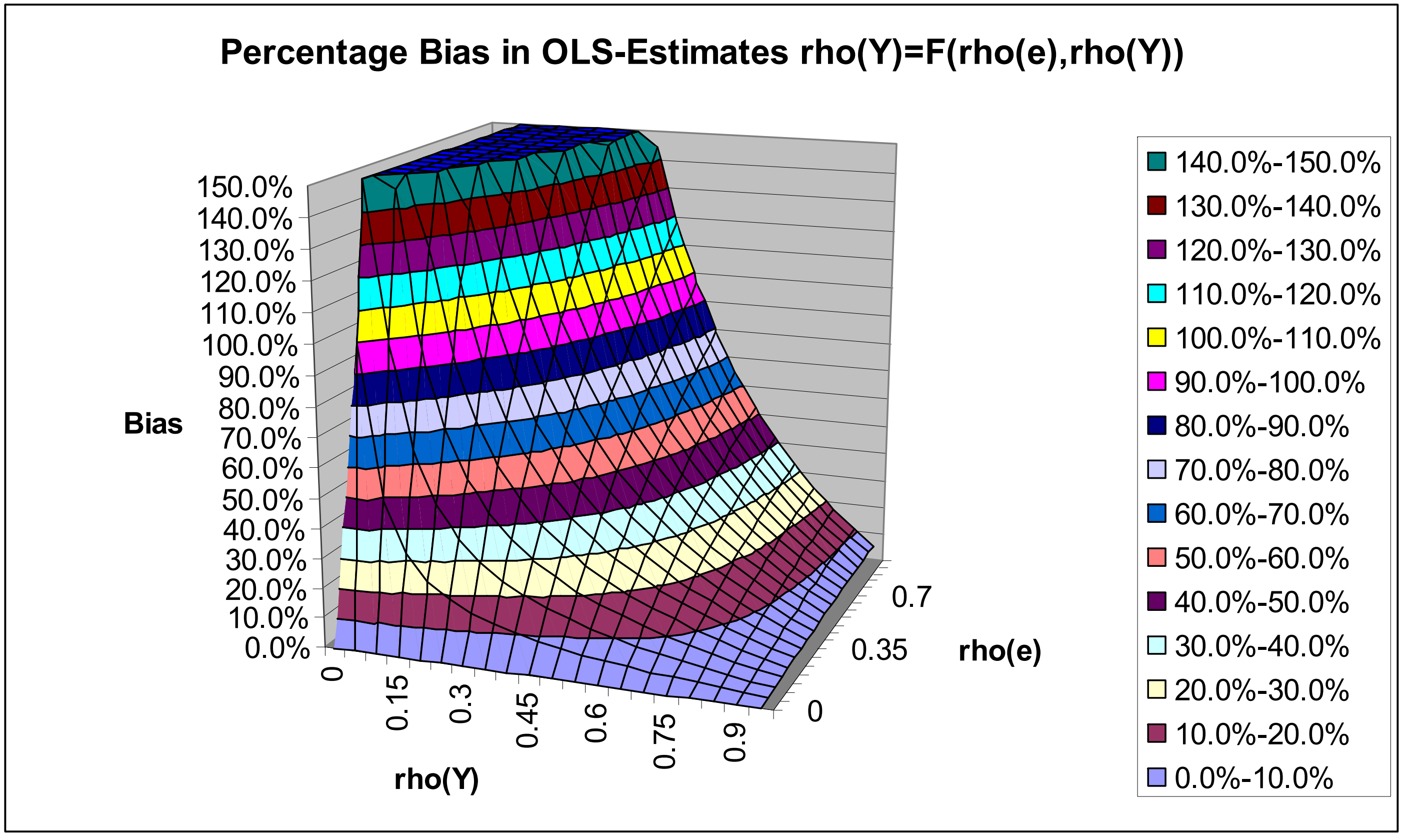
a) If temporal dynamics specified in systematic component sufficiently (no residual/stochastic-component corr. remains, which testable), OLS $\rightarrow^{A}$ BLUE.

b) If insufficient, OLS inconsistent, but still:

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

(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.

6. Spatial Dependence:

a) Situation more complic. OLS inconsistent even if model spatial-dep fully.

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

7. 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 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 => $V(\varepsilon)\sim 1/\#$.

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

(c) Still "Model It!" (in 2nd moment, so, i.e., model reduced parameterization of $\mathbf{\Omega}$).

(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. Redresses, mostly partial &/or imperfect, of deficiencies in implementation of the *Model It!*™ strategy

1. “Robust” or “Sandwich” Variance-Covariance Estimators

a) Key Insight:

(1) $\mathbf{\Omega}$ has $\frac{1}{2}n(n+1)>n$ elements/parameters; however, for consistent v-cov est:

(2) need consistent est only of $\mathbf{X'\Omega X}$, which only $\frac{1}{2}k(k+1)$ elements.

b) Proper $\mathbf{V}\left(\hat{\boldsymbol{\beta}}_{LS}\right)=\sigma^2\mathbf{(X'X)}^{-1}\mathbf{X'\Omega X(X'X)}^{-1}$ differs from OLS $\mathbf{V}\left(\hat{\boldsymbol{\beta}}_{OLS}\right)=\sigma^2\mathbf{(X'X)}^{-1}$, only insofar as $\mathbf{X'\Omega X}$ differs from $\mathbf{X'X}$, which is only insofar as elements of $\mathbf{\Omega}$ covary with elements of $\mathbf{X'X}$, 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 $\mathbf{X'\Omega X}$ 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:

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.

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 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 => inadvisable unless T>>2N (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:

$$\mathbf{Py} = \mathbf{PX\beta} + \mathbf{P\varepsilon} \Rightarrow$$

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

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

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

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

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

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

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

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