Stochastic-Term Heterogeneity in Panel/TSCS

I. Recall Implications of *Model It!*™ Strategy; first, call all RHS: $\mathbf{X\beta}$, and suppose it's rightly specified.

A.

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

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

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

B.

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

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

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

$$=\sigma^2(\mathbf{X}'\mathbf{X})^{-1} \text{ if } V(\boldsymbol\varepsilon)=\sigma^2\mathbf{I}, \text{ (as usual)}$$

C. OLS=BLUE if model right$^{(\text{*temporal, **spatial})}$…

1. * Even if right RHS time-dynamic model, LDV covaries conditional mean, α_i, proportionally to $1/T$ = Nickell-Hurwicz bias of $O(1/T)$. Small-T sample bias; still consistent in T.

2. ** Even if right RHS spatial-dynamic model, spatial lag covaries with residual because classic endogeneity: $y_i \Rightarrow y_j$ but also $y_j \Rightarrow y_i$.

D. If model not rt., or insuff., some decent properties may still hold

1. Argument of Franzese *PA* (2005): From theory-eval. 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.

2. Demonstrated there that hierarchical structure of error components *per se* does not bias OLS estimates of cross-level interactive models, as follows. Does bias OLS v-cov(**b**) estimates, however.

   a) *Note:* demonstration is for linear model (separable error-components); may not apply for nonlinear (nonseparable stochastic components is real issue).

E. For example, suppose use just linear-interaction, when truth is linear-interaction w/ error (=random coefficients):

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

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

$$\Rightarrow \hat{\boldsymbol{\beta}}_{OLS} = \left(\left[\mathbf{x} \quad \mathbf{xz}\right]'\left[\mathbf{x} \quad \mathbf{xz}\right]\right)^{-1}\left[\mathbf{x} \quad \mathbf{xz}\right]'\mathbf{y}$$

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

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

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

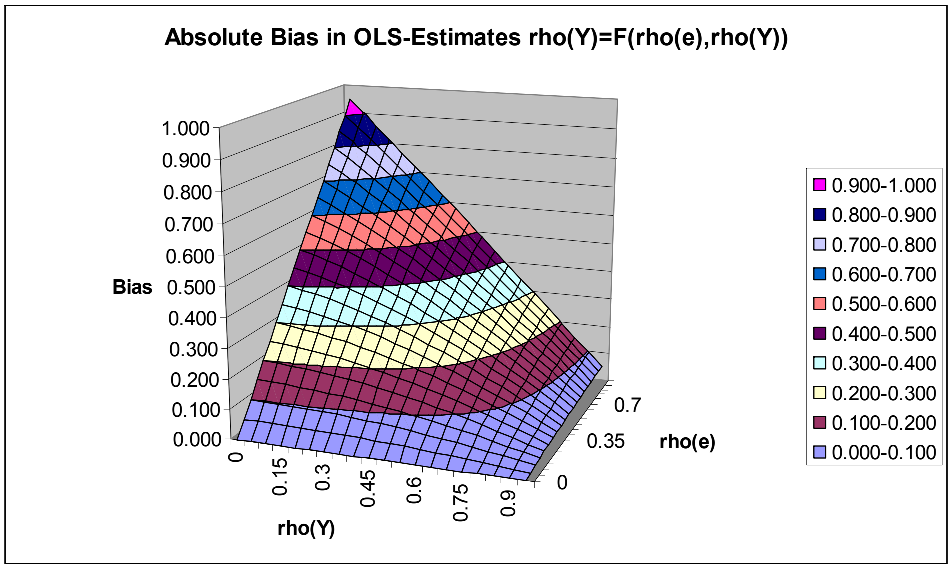
$$\mathbf{V}\left(\hat{\boldsymbol{\beta}}_{OLS}\right) = \mathbf{V}\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}\right] = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'V\left(\mathbf{x}\cdot\boldsymbol{\phi} + \boldsymbol{\varepsilon}\right)\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}$$

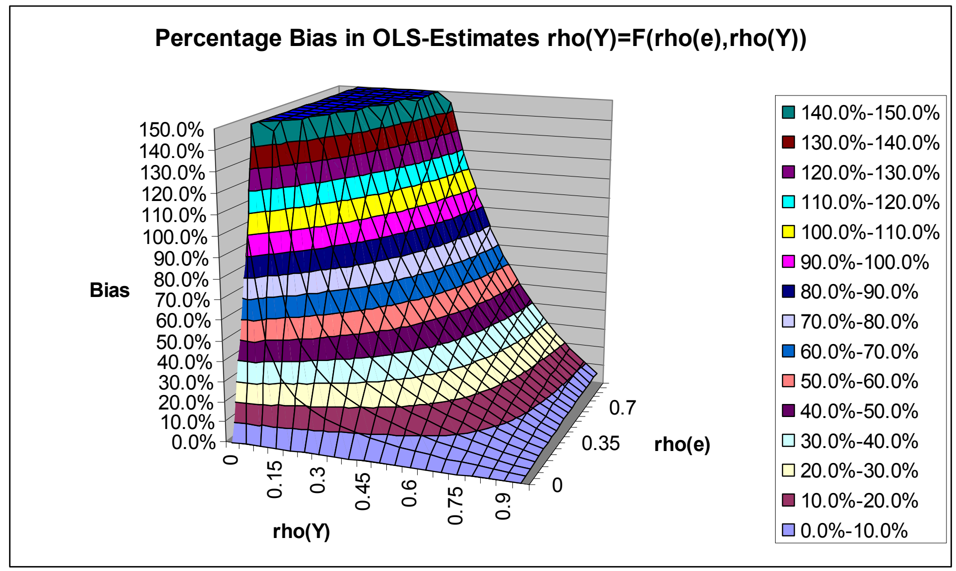
2. $$= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\sigma^2\boldsymbol{\Omega}\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}$$

3. OLS=unbiased, but inefficient **b**; wrong v-cov(**b**); some easy redresses

F. (Time-)Serial Dependence:

1. If temporal dynamics specified in systematic component sufficiently (no residual/stoch.-component corr. remains, which testable), OLS $\rightarrow^A$ BLUE.
2. If insufficient, OLS inconsistent, but still:
   a) Magnitude of the problem: $E\left(\hat{\rho}_y\right)=\rho_y+\rho_\varepsilon\left(1-\rho_y^2\right)/\left(1+\rho_\varepsilon\rho_y\right)$
   b) And can (partially) address s.e. part of problem (as shall see…)





G. Spatial Dependence:

1. If all in stochastic component, & Cov(**X**,**ε**)=**0**, then remains unbiased-but-inefficient coefficient-estimation, wrong s.e.
2. Recall that, if in systematic component, simultaneity bias even if rt RHS.

H. Limited & Qualitative Dep-Vars:

1. All these situations are more complicated…

2. Recall: issue is nonseparable stochastic component, not or more than nonlinearity or limited nature *per se*.

   a) For instance, can't do likelihood NT obs as NT conditionally independent binary outcomes, so likelihood not just product of NT unidimensional distributions, but one NT-dimensional multivariate distribution.

I. Implications *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

II. Some Tests for Various Forms Non-Sphericity

A. 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

1. Can also do Wald or loss-of-fit F-test on RHS (asym. normality); Results intuitive, constructive, and directly relate to "more pernicious" form het

B. Many, Many Heteroskedasticity-Test Strategies:

`help regress postestimation (estat [command])`

1. **Szroeter's** class of tests:

$$S = \frac{\sum_{i=1}^{NT} h(x_i) e_i^2}{\sum_{i=1}^{NT} e_i^2} \text{ where } h(x_i) \text{ is some weight increasing in } x_i$$

   a) King (1982) (another King, not Gary) suggests h(x) = rank(x).

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

2. **Goldfeld-Quandt**: sort e_i by some var. Take high & low sets, $\mathbf{e}_1$ & $\mathbf{e}_2$, (some evidence more power discard middle set), & stat:

$$\frac{\mathbf{e}_1'\mathbf{e}_1/(n_1-k)}{\mathbf{e}_2'\mathbf{e}_2/(n_2-k)} = F \sim F_{n_1-k,n_2-k}$$

3. **Glesjer's**: Wald or $NT \times R^2$ test coefficients: $\ln e_i^2 = \delta_0 + \delta_1 z_{1i} + \ldots + \delta_{kz} z_{ki} + \upsilon_i$; can think White's important special case Glesjer's.

4. **Breusch-Pagan** LM test for known form: groupwise (i.e., panel) het.

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

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

5. **Likelihood Ratio Test**: $-2\Delta\ln(L) \sim^{A} \chi^2_{\Delta k}$, for any pair estimable models. For panel(groupwise)-heteroskedasticity, test is:

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

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

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

D. **Spatial LM tests:** (elab Thursday)

- "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: $LM_{\rho\lambda} = LM_\lambda + LM_\rho^* = LM_\rho + LM_\lambda^*$

E. Tests of Temporal Correlation (ELAB THURSDAY…)

1. Many, many of these also.

2. My favorite…

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

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

3. (a.k.a., Breusch-Godfrey)……for same reasons Glesjer's/White's favorite among hettests:

a) Uses familiar, intuitive strategy. Constructive. Powerful.

b) ***Plus:*** valid following LDV model (include among x in auxilary regression) unlike, e.g., DW; flexible (works either MA or AR).

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

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

1. ***Key Insight:*** $\mathbf{\Omega}$ has ½$n(n+1)>n$ parameters; however, for consistent v-cov est: need consistent est only of $\mathbf{X'\Omega X}$, which only ½k(k+1) elements total.

2. Proper $V\left(\hat{\boldsymbol{\beta}}_{LS}\right)=\sigma^2(\mathbf{X'X})^{-1}\mathbf{X'\Omega X}(\mathbf{X'X})^{-1}$ differs from OLS $V\left(\hat{\boldsymbol{\beta}}_{OLS}\right)=\sigma^2(\mathbf{X'X})^{-1}$, 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 $\mathbf{xx'}$, i.e. ω_{ij} w/ x's, x^2's, &/or cross-products of x's.

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

4. ∴, we can *fix* our v-cov estimates, i.e. render them *robust*, i.e., <u>consistent</u>, to presence of certain pattern 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.

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

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

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

2. 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 & $w_t = 1 - \frac{t}{L-1}$

3. Panel Heteroskedasiticity: i.e., constant variance within unit, but differing across units: could write, but would be subsumed by White's

4. Panel Het. & (Time) Auto-Correlation: panel heteroskedasticity + unit-specific temporal correlation: could write, but Newey-West would subsume.

5. 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

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

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

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

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

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

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

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

b) E.g., for clustered: "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."

c) Need note (using `help vcetype`, manuals) which of these adjustments Stata applies as defaults or options.

d) 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.

IV. FGLS: Feasible Generalized-Least-Squares

A. Consistent V-Cov ests only address inconsistency of s.e.'s, not bias or efficiency coefficient estimates (although require consistent coefficient-estimates for formal properties); nor address s.e. unbiasedness, efficiency.

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—can parameterize and estimate $\hat{\Omega}$, 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:** (assuming consistent 1st-stage) consistent & asymptotically efficient. **Estimation**:

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

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

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

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

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

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

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

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

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