Stochastic-Term Heterogeneity in Panel/TSCS

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

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

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

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

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

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

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

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

C. OLS=BLUE if model right\(^{*}\text{temporal, } **\text{spatial}\)…
1. * Even if right RHS time-dynamic model, LDV covaries conditional mean, $\alpha_i$, proportionally to $1/T = \text{Nickell-Hurwicz bias of } O(1/T)$. 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 we use just linear-interaction, when truth is linear-interaction w/ error (=random coefficients):

Truth: \( y_{it} = x_{it} (\gamma_0 + \gamma_1 z_{it} + \phi_i) + \epsilon_{it} \)

Model: \( y_{it} = x_{it} (\gamma_0^* + \gamma_1^* z_{it}) + \epsilon_{it}^* \)

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

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

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

\[ \Rightarrow E(\hat{\beta}) = \beta \text{ if } E(\phi \cdot x) = 0, \ E(X'\epsilon) = 0 \]

\[ \mathbb{V}(\hat{\beta}_{OLS}) = \mathbb{V}\left( \begin{bmatrix} X'X \end{bmatrix}^{-1} X' y \right) = \begin{bmatrix} X'X \end{bmatrix}^{-1} X' \mathbb{V}(x \cdot \phi + \epsilon) X \begin{bmatrix} X'X \end{bmatrix}^{-1} \]

2. \[ = \begin{bmatrix} X'X \end{bmatrix}^{-1} X' \sigma^2 \Omega X \begin{bmatrix} X'X \end{bmatrix}^{-1} \]

3. OLS=unbiased, but inefficient \( b \); wrong \( \text{v-cov}(b) \); 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 → BLUE.
2. If insufficient, OLS inconsistent, but still:
   a) Magnitude of the problem: 
      \[ E(\hat{\rho}_y) = \rho_y + \rho_\varepsilon \left(1 - \rho_y^2\right)/(1 + \rho_\varepsilon \rho_y) \]
   b) And can (partially) address s.e. part of problem (as shall see…)

G. Spatial Dependence:

1. If all in stochastic component, & \( \text{Cov}(X,\varepsilon) = 0 \), 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)

<table>
<thead>
<tr>
<th><strong>Model It!™ Adequacy</strong></th>
<th>Second-Moment &amp; Inadequacy</th>
<th>Implications for OLS Properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model E(y) Sufficient</td>
<td>V(e) Spherical</td>
<td>OLS is BLUE</td>
</tr>
<tr>
<td>V(e) Nonspherical</td>
<td>Nonsphericity Unrelated (xx')</td>
<td>OLS (b) unbiased, inefficient; OLS (V(b)) unbiased, inefficient</td>
</tr>
<tr>
<td>V(e) Nonspherical</td>
<td>Nonsphericity Related (xx')</td>
<td>OLS (b) unbiased, inefficient; OLS (V(b)) biased, inefficient</td>
</tr>
<tr>
<td>Unmodeled (b) het unrelated (X)</td>
<td>V(e) Will be Nonspherical &amp; Related (xx')</td>
<td>OLS (b) unbiased, inefficient; OLS (V(b)) biased, inefficient</td>
</tr>
<tr>
<td>Unmodeled (b) het related to (X)</td>
<td>V(e) Will be Nonspherical &amp; Related (xx')</td>
<td>OLS (b) biased, inefficient; OLS (V(b)) biased, inefficient</td>
</tr>
</tbody>
</table>
II. Some Tests for Various Forms Non-Sphericity

A. White’s General Test:

c. White test if form of Heteroskedasticity is unknown:

\[ H_0: \ V(\varepsilon_i \mid x_i) = \sigma^2 \]
\[ H_a: \ V(\varepsilon_i \mid x_i) = \sigma_i^2 \]

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 \( R^2 \) of auxiliary regression: \( n^* R^2 \xrightarrow{a} \chi^2_{(q)} \)

Use one-sided test and check if \( n^* R^2 \) is larger than 95% quantile of \( \chi^2 \)

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:

\[ S = \frac{\sum_{i=1}^{NT} h(x_i) e_i^2}{\sum_{i=1}^{NT} e_i^2} \]

where \( h(x_i) \) is some weight increasing in \( x_i \)

a) King (1982) (another King, not Gary) suggests \( h(x) = \text{rank}(x) \).

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, \( e_1 \) & \( e_2 \), (some evidence more power discard middle set), & stat:

\[
\frac{e_1'e_1/(n_1-k)}{e_2'e_2/(n_2-k)} = F \sim F_{n_1-k,n_2-k}
\]

3. **Glesjer’s**: Wald or \( NT \times R^2 \) test coefficients:

\[ \ln e_i^2 = \delta_0 + \delta_1 z_{1i} + ... + \delta_k z_{ki} + \nu_i \]

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 \( \sim \) Chi\(^2\) 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:

\[-2 \ln(\lambda) = (NT) \ln(\sigma^2) - \sum(T \ln(\sigma_i^2)) \sim \chi^2(dF = n - 1)\]

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

\[ e_t^2 = \delta_0 + \delta_1 x_{1t} + \ldots + \delta_k x_{kt} + \sum_{s=1}^{S} \lambda_s e_{t-s}^2 \]
D. **Spatial LM tests:** (elab Thursday)

- "Testing" (measuring) Spatial Association:
  
  Moran's I: \( I = \frac{N \mathbf{\varepsilon}^\prime \mathbf{W} \mathbf{\varepsilon}}{S \mathbf{\varepsilon}^\prime \mathbf{\varepsilon}} \), where \( S = \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij} \), if row-nrmlzd: \( I = \frac{\mathbf{\varepsilon}^\prime \mathbf{W} \mathbf{\varepsilon}}{\mathbf{\varepsilon}^\prime \mathbf{\varepsilon}} \)

- See also Anselin’s LISA...

- LM’s appropriate for LS resid (F&H OxfHndbk):
  
  \[
  LM_\rho = \frac{\hat{\sigma}_\varepsilon^2 \left( \mathbf{\varepsilon}^\prime \mathbf{W} \mathbf{Y} / \hat{\sigma}_\varepsilon^2 \right)^2}{G + T \hat{\sigma}_\varepsilon^2}, \text{ and } LM_\lambda = \frac{\left( \mathbf{\varepsilon}^\prime \mathbf{W} \hat{\mathbf{\varepsilon}} / \hat{\sigma}_\varepsilon^2 \right)^2}{T}
  \]

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

- Robust LM tests appropriate for SAR v. SAE:
  
  \[
  LM_\rho^* = G^{-1} \hat{\sigma}_\varepsilon^2 \left( \mathbf{\varepsilon}^\prime \mathbf{W} \mathbf{Y} / \hat{\sigma}_\varepsilon^2 - \mathbf{\varepsilon}^\prime \mathbf{W} \hat{\mathbf{\varepsilon}} / \hat{\sigma}_\varepsilon^2 \right)^2, \text{ and } LM_\lambda^* = \frac{\left( \mathbf{\varepsilon}^\prime \mathbf{W} \hat{\mathbf{\varepsilon}} / \hat{\sigma}_\varepsilon^2 - \left[ T \hat{\sigma}_\varepsilon^2 \left( G + T \hat{\sigma}_\varepsilon^2 \right)^{-1} \right] \mathbf{\varepsilon}^\prime \mathbf{W} \mathbf{Y} / \hat{\sigma}_\varepsilon^2 \right)^2}{T \left[ 1 - \frac{T \hat{\sigma}_\varepsilon^2}{G + T \hat{\sigma}_\varepsilon^2} \right]} \]

- 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 $\frac{T \times R^2}{\chi^2_{(S)}}$):

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

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


   b) **Plus:** valid following LDV model (include among $x$ in auxiliary 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 \textit{Model It!}™ strategy:

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

1. \textit{Key Insight}: $\Omega$ has $\frac{1}{2}n(n+1) > n$ parameters; however, for consistent v-cov est: need consistent est only of $X'\Omega X$, which only $\frac{1}{2}k(k+1)$ elements total.

2. Proper $V(\hat{\beta}_{LS}) = \sigma^2 (X'X)^{-1} X'\Omega X (X'X)^{-1}$ differs from OLS $V(\hat{\beta}_{OLS}) = \sigma^2 (X'X)^{-1}$, only insofar as $X'\Omega X$ differs from $X'X$, which is only insofar as elements of $\Omega$ covary with elements of $xx'$, i.e. $\omega_{ij}$ 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: $\sum_{i,j,s,t} e_{it}e_{js} (x_{it}x'_{js}) - \sum_{i,t} e_{it}^2 I_k$.

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

$$\hat{V}_S (\hat{\beta}) = (X'X)^{-1} \hat{X}'\hat{\Omega} \hat{X} (X'X)^{-1} \equiv (X'X)^{-1} \hat{Q}(X'X)^{-1}$$
B. Cases of (Robust/Sandwich) Consistent V-Cov(b) Estimators:

1. **Pure Heteroskedasticity** (White’s):  
   \[
   \hat{Q} = \sum_i \sum_t e_{it}^2 (x_{it} x_{it}')
   \]

2. **Pure Het. & (Time) Auto-Correlation** (**HAC**) (Newey-West):  
   \[
   \hat{Q} = \sum_i \sum_t e_{it}^2 (x_{it} x_{it}') + \sum_i \sum_t \left( \sum_{s=1}^{L} w_t e_{it} e_{i,t-s} (x_{it} x_{i,t-s} + x_{i,t-s} x_{i,t}') \right)
   \]
   where \( L = \text{max lag-length considered appreciable} \) & \( w_t = 1 - \frac{t}{L-1} \)

3. **Panel Heteroskedasticity**: 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):  
   \[
   \hat{Q} = X' \left( \frac{E' E}{T} \otimes I_T \right) X , \text{ where } E = \text{the } T \times N \text{ matrix estimated residuals}
   \]
6. Many others possible. Several “cluster” types, e.g., designed for various multilevel/hierarchical data structures (i.e., panel ‘random-effect’ structure):

$$\hat{V}(\hat{\beta}) = \frac{1}{N - k} (X'X)^{-1} \left\{ \sum_{j=1}^{n_c} \left\{ \sum_{i=1}^{n_j} e_i x_i \right\} \left\{ \sum_{i=1}^{n_j} e_i x_i \right\}' \right\} (X'X)^{-1}$$

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

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 \( e_i^2/(1-x_i(X'X)^{-1}x_i) \), 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, \( \frac{n_c}{(n_c-1)} \frac{(N-1)}{(N-k)} \), 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 \( \sigma_i^2 \) => \( N \) parameters
3. Dyad-specific \( \sigma_{ij} \) => \( N(N-1) \) 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 1\textsuperscript{st}-stage) consistent & asymptotically efficient. Estimation:

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

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

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

\( \Rightarrow \) consistent and asympt'ly efficient if \( C(X, \epsilon) = 0 \)

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

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

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

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

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