SYSTEMS OF EQUATIONS: GENERALIZED LEAST SQUARES

Econometric Analysis of Cross Section and Panel Data, 2e
MIT Press
Jeffrey M. Wooldridge

1. Asymptotic Properties of GLS
2. Feasible GLS
3. FGLS with Incorrect Restrictions on the Variance Matrix
4. Testing Using FGLS

1. ASYMPTOTIC PROPERTIES OF GLS

- By "generalized least squares," we mean exploiting different *unconditional* variances across equation (time, in the panel data case) and nonzero *unconditional* covariances across equations. We do not exploit situations where the variance-covariance matrix is a function of $\mathbf{X}_i$.
- Write the equation in system form (for a random draw i) as

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{u}_i \tag{1.1}$$

where $\mathbf{y}_i$ is $G \times 1$, $\mathbf{X}_i$ is $G \times K$, and $\mathbf{u}_i$ is $G \times 1$. Remember, in panel data case, $G = T$.

• The $G \times G$ unconditional variance-covariance matrix plays a key role.

$$\mathbf{\Omega} \equiv E(\mathbf{u}_i \mathbf{u}_i') = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1G} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2G} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1G} & \sigma_{2G} & \cdots & \sigma_G^2 \end{pmatrix}. \tag{1.2}$$

• The following language is as if $E(\mathbf{u}_i) = \mathbf{0}$, which is maintained in virtually all applications and is no assumption at all when the various equations contain an intercept. If $\mathbf{u}_i$ does not have a zero mean, then $\mathbf{\Omega}$ is not the V-C matrix, but everything goes through.

• We already discussed system OLS. What else might we do? Without additional assumptions, suppose we use "generalized least squares." Assume, for now, that we know $\mathbf{\Omega}$.

• Transform the equation to remove correlations in errors and make variances constant (actually, unity):

$$\mathbf{\Omega}^{-1/2}\mathbf{y}_i = \mathbf{\Omega}^{-1/2}\mathbf{X}_i\boldsymbol{\beta} + \mathbf{\Omega}^{-1/2}\mathbf{u}_i, \tag{1.3}$$

where $\mathbf{\Omega}$ is assumed to be nonsingular and $\mathbf{\Omega}^{-1/2}$ is a symmetric matrix such that $\mathbf{\Omega}^{-1/2}\mathbf{\Omega}^{-1/2} = \mathbf{\Omega}^{-1}$ and $\mathbf{\Omega}^{-1/2}\mathbf{\Omega}\mathbf{\Omega}^{-1/2} = \mathbf{I}_G$. Let $\mathbf{X}_i^* = \mathbf{\Omega}^{-1/2}\mathbf{X}_i$ and similarly for $\mathbf{y}_i^*$, $\mathbf{u}_i^*$. Then $E(\mathbf{u}_i^*\mathbf{u}_i^{*\prime}) = \mathbf{\Omega}^{-1/2}E(\mathbf{u}_i\mathbf{u}_i^{\prime})\mathbf{\Omega}^{-1/2} = \mathbf{I}_G$.

- Apply System OLS to $\mathbf{y}_i^* = \mathbf{X}_i^* \boldsymbol{\beta} + \mathbf{u}_i^*$. The GLS estimator is

$$\begin{aligned}
\boldsymbol{\beta}^* &= \left(N^{-1} \sum_{i=1}^{N} \mathbf{X}_i^{*\prime} \mathbf{X}_i^* \right)^{-1} \left(N^{-1} \sum_{i=1}^{N} \mathbf{X}_i^{*\prime} \mathbf{y}_i^* \right) \\
&= \left(N^{-1} \sum_{i=1}^{N} \mathbf{X}_i^{\prime} \boldsymbol{\Omega}^{-1} \mathbf{X}_i \right)^{-1} \left(N^{-1} \sum_{i=1}^{N} \mathbf{X}_i^{\prime} \boldsymbol{\Omega}^{-1} \mathbf{y}_i \right) \qquad (1.4) \\
&= \boldsymbol{\beta} + \left(N^{-1} \sum_{i=1}^{N} \mathbf{X}_i^{\prime} \boldsymbol{\Omega}^{-1} \mathbf{X}_i \right)^{-1} \left(N^{-1} \sum_{i=1}^{N} \mathbf{X}_i^{\prime} \boldsymbol{\Omega}^{-1} \mathbf{u}_i \right)
\end{aligned}$$

- The $K \times K$ matrix average converges in probability to $E(\mathbf{X}_i'\mathbf{\Omega}^{-1}\mathbf{X}_i)$; assume this is nonsingular. Then, consistency of $\boldsymbol{\beta}^*$ holds if

$$E(\mathbf{X}_i'\mathbf{\Omega}^{-1}\mathbf{u}_i) = \mathbf{0}. \tag{1.5}$$

- In general, (1.5) is not implied by SOLS.1,

$$E(\mathbf{X}_i'\mathbf{u}_i) = \mathbf{0}. \tag{1.6}$$

GLS transforms the orthogonality conditions; it may not be consistent when SOLS is.

- Rather than assume (1.5), use

Assumption SGLS.1 (Exogeneity):

$$E(\mathbf{X}_i \otimes \mathbf{u}_i) = \mathbf{0}. \ \square \tag{1.7}$$

- The Kronecker product is used so that every element of $\mathbf{X}_i$ is uncorrelated with every element of $\mathbf{u}_i$, so any linear combination of $\mathbf{X}_i$ is uncorrelated with $\mathbf{u}_i$. In particular, (1.5) holds.

- In some special cases $E(\mathbf{X}_i'\mathbf{\Omega}^{-1}\mathbf{u}_i) = \mathbf{0}$ can hold when $E(\mathbf{X}_i \otimes \mathbf{u}_i) = \mathbf{0}$, but the latter assumption implies that a variety of GLS estimators, even with a misspecified variance matrix, will be consistent. Plus, in the next section we rely on (1.7) to justify ignoring estimation of $\mathbf{\Omega}$.

Assumption SGLS.2 (Rank Condition): $\mathbf{\Omega}$ is nonsingular and $E(\mathbf{X}_i'\mathbf{\Omega}^{-1}\mathbf{X}_i)$ is nonsingular. $\square$

THEOREM: Under SGLS.1 and SGLS.2, $\boldsymbol{\beta}^*$ is consistent for $\boldsymbol{\beta}$ as $N \rightarrow \infty$. $\square$

• Must take the distinction between SGLS.1 and SOLS.1 seriously. If only $E(\mathbf{X}_i'\mathbf{u}_i) = \mathbf{0}$ holds, GLS is generally inconsistent.

EXAMPLE: Suppose $G = 2$, so that in the SUR case we can write

$$\mathbf{\Omega}^{-1} = \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{12} & \omega_{22} \end{pmatrix}$$

$$\mathbf{\Omega}^{-1}\mathbf{X}_i = \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{12} & \omega_{22} \end{pmatrix} \begin{pmatrix} \mathbf{x}_{i1} & \mathbf{0} \\ \mathbf{0} & \mathbf{x}_{i2} \end{pmatrix} = \begin{pmatrix} \omega_{11}\mathbf{x}_{i1} & \omega_{12}\mathbf{x}_{i2} \\ \omega_{12}\mathbf{x}_{i1} & \omega_{22}\mathbf{x}_{i2} \end{pmatrix}. \qquad (1.8)$$

Then

$$E[(\mathbf{\Omega}^{-1}\mathbf{X}_i)'\mathbf{u}_i] = \begin{pmatrix} \omega_{11}E(\mathbf{x}_{i1}'u_{i1}) + \omega_{12}E(\mathbf{x}_{i1}'u_{i2}) \\ \omega_{12}E(\mathbf{x}_{i2}'u_{i1}) + \omega_{22}E(\mathbf{x}_{i2}'u_{i2}) \end{pmatrix}. \tag{1.9}$$

Unless $\omega_{12} = 0$, which is true if and only if $\sigma_{12} = 0$, we need the covariates in each equation to be uncorrelated with the errors in each equation.

• If $\sigma_{12} = 0$, only need $E(\mathbf{x}_{ig}'u_{ig}) = \mathbf{0}$, $g = 1, 2$. The GLS estimator in this case is OLS equation-by-equation. (More general result later.)

- Asymptotic normality is also straightforward:

$$
\begin{aligned}
\sqrt{N}\,(\boldsymbol{\beta}^* - \boldsymbol{\beta}) &= \left(N^{-1} \sum_{i=1}^{N} \mathbf{X}_i' \boldsymbol{\Omega}^{-1} \mathbf{X}_i \right)^{-1} \left(N^{-1/2} \sum_{i=1}^{N} \mathbf{X}_i' \boldsymbol{\Omega}^{-1} \mathbf{u}_i \right) \\
&= \mathbf{A}^{-1} \left(N^{-1/2} \sum_{i=1}^{N} \mathbf{X}_i' \boldsymbol{\Omega}^{-1} \mathbf{u}_i \right) \\
&\quad + \left[\left(N^{-1} \sum_{i=1}^{N} \mathbf{X}_i' \boldsymbol{\Omega}^{-1} \mathbf{X}_i \right)^{-1} - \mathbf{A}^{-1} \right] \left(N^{-1/2} \sum_{i=1}^{N} \mathbf{X}_i' \boldsymbol{\Omega}^{-1} \mathbf{u}_i \right) \\
&= \mathbf{A}^{-1} \left(N^{-1/2} \sum_{i=1}^{N} \mathbf{X}_i' \boldsymbol{\Omega}^{-1} \mathbf{u}_i \right) + o_p(1)
\end{aligned}
$$

Now

$$\mathbf{A} = E(\mathbf{X}_i'\mathbf{\Omega}^{-1}\mathbf{X}_i). \tag{1.11}$$

• Using the same argument as for SOLS,

$$\sqrt{N}\,(\boldsymbol{\beta}^* - \boldsymbol{\beta}) \xrightarrow{d} Normal(\mathbf{0}, \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}) \tag{1.12}$$

$$\mathbf{B} = Var(\mathbf{X}_i'\mathbf{\Omega}^{-1}\mathbf{u}_i) = E(\mathbf{X}_i'\mathbf{\Omega}^{-1}\mathbf{u}_i\mathbf{u}_i'\mathbf{\Omega}^{-1}\mathbf{X}_i) \tag{1.13}$$

• Important: At this point, we cannot simplify **B** further. We are not assuming $\mathbf{X}_i$ is nonrandom, and we do not have enough assumptions about the distribution of $\mathbf{u}_i$ given $\mathbf{X}_i$ to reduce **B**.

• One consequence of the complicated expression for **B**: under SGLS.1, SOLS.2, and SGLS.2, GLS need *not* be more efficient than SOLS!

2. FEASIBLE GLS

2.1. The Estimator and Asymptotic Properties

• Now we study the estimator from the previous section but were an estimator of $\mathbf{\Omega}$ is used in place of $\mathbf{\Omega}$. Generally, let $\hat{\mathbf{\Omega}}$ be a $G \times G$ matrix such that

$$\operatorname{plim}_{N\to\infty}\hat{\mathbf{\Omega}} = \mathbf{\Omega}. \tag{2.1}$$

• This only makes sense when $\mathbf{\Omega}$ has fixed dimension. (In the panel data case, T is fixed.)

- In SUR analysis, we almost always use

$$\hat{\boldsymbol{\Omega}} = N^{-1} \sum_{i=1}^{N} \check{\mathbf{u}}_i \check{\mathbf{u}}_i' \tag{2.2}$$

where $\check{\mathbf{u}}_i \equiv \mathbf{y}_i - \mathbf{X}_i \check{\boldsymbol{\beta}}$ are the $G \times 1$ SOLS residuals ($\check{\boldsymbol{\beta}}$ is the SOLS estimator).

- Same $\hat{\boldsymbol{\Omega}}$ can be used for panel data.

- Write

$$
\begin{aligned}
\check{\mathbf{u}}_i &= \mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{X}_i(\check{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \mathbf{u}_i - \mathbf{X}_i(\check{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\
\check{\mathbf{u}}_i\check{\mathbf{u}}_i' &= \mathbf{u}_i\mathbf{u}_i' - \mathbf{u}_i(\check{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{X}_i' - \mathbf{X}_i(\check{\boldsymbol{\beta}} - \boldsymbol{\beta})\mathbf{u}_i' \\
&\quad + \mathbf{X}_i(\check{\boldsymbol{\beta}} - \boldsymbol{\beta})(\check{\boldsymbol{\beta}} - \boldsymbol{\beta})'\mathbf{X}_i'
\end{aligned} \tag{2.3}
$$

can show that

$$
\hat{\boldsymbol{\Omega}} = N^{-1}\sum_{i=1}^{N}\mathbf{u}_i\mathbf{u}_i' + o_p(1). \tag{2.4}
$$

• In fact, can even show under SGLS.1 and SOLS.2,

$$\sqrt{N}\left(\hat{\mathbf{\Omega}} - N^{-1}\sum_{i=1}^{N}\mathbf{u}_i\mathbf{u}_i'\right) = o_p(1), \tag{2.5}$$

so, for performing inference about the elements of $\mathbf{\Omega}$, we can ignore the estimation error in $\check{\boldsymbol{\beta}}$ (in large samples). Very useful for testing zero covariances, constant variances, and no serial correlation.

• (2.5) does not go through under SOLS.1, even though (2.4) does.

- The FGLS estimator is

$$\hat{\boldsymbol{\beta}} = \left(N^{-1} \sum_{i=1}^{N} \mathbf{X}_i' \hat{\boldsymbol{\Omega}}^{-1} \mathbf{X}_i \right)^{-1} \left(N^{-1} \sum_{i=1}^{N} \mathbf{X}_i' \hat{\boldsymbol{\Omega}}^{-1} \mathbf{y}_i \right) \tag{2.6}$$
$$= \boldsymbol{\beta} + \left(N^{-1} \sum_{i=1}^{N} \mathbf{X}_i' \hat{\boldsymbol{\Omega}}^{-1} \mathbf{X}_i \right)^{-1} \left(N^{-1} \sum_{i=1}^{N} \mathbf{X}_i' \hat{\boldsymbol{\Omega}}^{-1} \mathbf{u}_i \right)$$

Write

$$N^{-1} \sum_{i=1}^{N} \mathbf{X}_i' \hat{\boldsymbol{\Omega}}^{-1} \mathbf{X}_i - N^{-1} \sum_{i=1}^{N} \mathbf{X}_i' \boldsymbol{\Omega}^{-1} \mathbf{X}_i = N^{-1} \sum_{i=1}^{N} \mathbf{X}_i' (\hat{\boldsymbol{\Omega}}^{-1} - \boldsymbol{\Omega}^{-1}) \mathbf{X}_i.$$

- Fact from matrix algebra: for conformable matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$,

$$\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A})\mathrm{vec}(\mathbf{B})$$

where “vec” is the vectorization of a matrix (stacking the columns).

- Under SGLS.1, SGLS.2, and $\hat{\boldsymbol{\Omega}} = N^{-1}\sum_{i=1}^{N}\mathbf{u}_i\mathbf{u}_i' + o_p(1)$,

$$\operatorname{vec}\left[N^{-1}\sum_{i=1}^{N}\mathbf{X}_i'(\hat{\boldsymbol{\Omega}}^{-1} - \boldsymbol{\Omega}^{-1})\mathbf{X}_i\right] = \left[N^{-1}\sum_{i=1}^{N}(\mathbf{X}_i' \otimes \mathbf{X}_i')\right]\operatorname{vec}(\hat{\boldsymbol{\Omega}}^{-1} - \boldsymbol{\Omega}^{-1})$$
$$= O_p(1) \cdot o_p(1)$$

and

$$N^{-1/2}\sum_{i=1}^{N}(\mathbf{X}_i'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{u}_i - \mathbf{X}_i'\boldsymbol{\Omega}^{-1}\mathbf{u}_i) = N^{-1/2}\sum_{i=1}^{N}\mathbf{X}_i'(\hat{\boldsymbol{\Omega}}^{-1} - \boldsymbol{\Omega}^{-1})\mathbf{u}_i$$
$$= \left[N^{-1/2}\sum_{i=1}^{N}(\mathbf{u}_i \otimes \mathbf{X}_i)'\right]\operatorname{vec}(\hat{\boldsymbol{\Omega}}^{-1} - \boldsymbol{\Omega}^{-1})$$
$$= O_p(1) \cdot o_p(1).$$

- Combining the above two results gives

$$\sqrt{N}\,(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \left(N^{-1}\sum_{i=1}^{N}\mathbf{X}_i'\boldsymbol{\Omega}^{-1}\mathbf{X}_i\right)^{-1}\left(N^{-1/2}\sum_{i=1}^{N}\mathbf{X}_i'\boldsymbol{\Omega}^{-1}\mathbf{u}_i\right) + o_p(1) \tag{2.7}$$

$$= \sqrt{N}\,(\boldsymbol{\beta}^* - \boldsymbol{\beta}) + o_p(1). \tag{2.8}$$

By the asymptotic equivalence lemma, the asymptotic distribution of $\sqrt{N}\,(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is the same as that of $\sqrt{N}\,(\boldsymbol{\beta}^* - \boldsymbol{\beta})$.

• When

$$\sqrt{N}\,(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) = o_p(1) \tag{2.9}$$

we say that $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}^*$ are “asymptotically equivalent” or, more precisely, “$\sqrt{N}$ –equivalent,” which is much stronger than saying that they are both consistent. (Under SGLS.1, SOLS.2, and SGLS.2, FGLS and SOLS are both consistent but they are not asymptotically equivalent.)

- It is not always true that first-stage estimation of population parameters can be ignored in a second stage (for example, see the control function notes). But, in this case, for $\sqrt{N}$ –asymptotics, we can treat FGLS as if it is GLS.
- If N is "small," the statistical properties of $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}^*$ could be very different (and we would not know, since $\boldsymbol{\beta}^*$ is infeasible).
- FGLS is not unbiased under $E(\mathbf{u}_i|\mathbf{X}_i) = \mathbf{0}$, GLS is (if the moments exist).

• A fully robust sandwich variance matrix estimator can be used under SGLS.1 and SGLS.2: let $\hat{\mathbf{u}}_i \equiv \mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}$ be the FGLS residuals. Then

$$\widehat{\text{Avar}}(\hat{\boldsymbol{\beta}}) = \left(\sum_{i=1}^{N} \mathbf{X}_i'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{X}_i\right)^{-1} \left(\sum_{i=1}^{N} \mathbf{X}_i'\hat{\boldsymbol{\Omega}}^{-1}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{X}_i\right)^{-1} \cdot \left(\sum_{i=1}^{N} \mathbf{X}_i'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{X}_i\right)^{-1}, \tag{2.10}$$

and sometimes with a degrees-of-freedom adjustment, $N - K$.

• This estimator is robust to "system heteroskedasticity." Loosely, the variance-covariance matrix of $\mathbf{u}_i$ conditional on $\mathbf{X}_i$ does not depend on $\mathbf{X}_i$.

2.2. When is the "Usual" Variance Matrix Estimator for FGLS Valid?

Assumption SGLS.3 (System Homoskedasticity):

$$E(\mathbf{X}_i'\boldsymbol{\Omega}^{-1}\mathbf{u}_i\mathbf{u}_i'\boldsymbol{\Omega}^{-1}\mathbf{X}_i) = E(\mathbf{X}_i'\boldsymbol{\Omega}^{-1}\mathbf{X}_i). \quad \square \tag{2.11}$$

- Effectively, all squares and cross products u_{ig}^2, $u_{ig}u_{ih}$, are uncorrelated with the squares and cross products of elements in $\mathbf{X}_i$.
- This assumption simple says that $\mathbf{B} = \mathbf{A}$, which means we can use

$$\widehat{\text{Avar}}(\hat{\boldsymbol{\beta}}) = \left(\sum_{i=1}^{N}\mathbf{X}_i'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{X}_i\right)^{-1}, \tag{2.12}$$

which is the nonrobust ("usual") FGLS variance matrix estimator.

- Sufficient for SGLS.3:

$$E(\mathbf{u}_i\mathbf{u}_i'|\mathbf{X}_i) = E(\mathbf{u}_i\mathbf{u}_i'). \tag{2.13}$$

- Use the law of iterated expectations:

$$\begin{aligned} E(\mathbf{X}_i'\boldsymbol{\Omega}^{-1}\mathbf{u}_i\mathbf{u}_i'\boldsymbol{\Omega}^{-1}\mathbf{X}_i) &= E[E(\mathbf{X}_i'\boldsymbol{\Omega}^{-1}\mathbf{u}_i\mathbf{u}_i'\boldsymbol{\Omega}^{-1}\mathbf{X}_i|\mathbf{X}_i)] \\ &= E[\mathbf{X}_i'\boldsymbol{\Omega}^{-1}E(\mathbf{u}_i\mathbf{u}_i'|\mathbf{X}_i)\boldsymbol{\Omega}^{-1}\mathbf{X}_i] \\ &= E(\mathbf{X}_i'\boldsymbol{\Omega}^{-1}\boldsymbol{\Omega}\boldsymbol{\Omega}^{-1}\mathbf{X}_i) = E(\mathbf{X}_i'\boldsymbol{\Omega}^{-1}\mathbf{X}_i). \end{aligned}$$

- Most straightforward, and traditional (essentially the same as the fixed regressor assumption) are

$$E(\mathbf{u}_i|\mathbf{X}_i) = E(\mathbf{u}_i) = \mathbf{0} \tag{2.14}$$

$$Var(\mathbf{u}_i|\mathbf{X}_i) = Var(\mathbf{u}_i) = \boldsymbol{\Omega} \tag{2.15}$$

- Given the zero conditional mean assumption, the key conditions are

$$Var(u_{ig}|\mathbf{X}_i) = Var(u_{ig}),\ g = 1,\ldots,G \tag{2.16}$$

$$Cov(u_{ig},u_{ih}|\mathbf{X}_i) = Cov(u_{ig},u_{ih}), \text{ all } g \neq h. \tag{2.17}$$

- By the random sampling assumption, the unconditional variance-covariance matrices $E(\mathbf{u}_i\mathbf{u}_i')$ must be identical across i, and equal to $\mathbf{\Omega}$. The question is whether conditional variances and covariances conditional on $\mathbf{X}_i$ are constant.
- Particularly in panel data applications without strict exogeneity it will not make sense to condition on all of $\mathbf{X}_i$.

2.3 When is FGLS more efficient than SOLS?

• Suppose we start with SGLS.1, $E(\mathbf{X}_i \otimes \mathbf{u}_i) = \mathbf{0}$ (which, of course, implies SOLS.1), add the two rank conditions, SOLS.2 and SGLS.2, and this form of the system homoskedasticity assumption:

$$E(\mathbf{u}_i\mathbf{u}_i'|\mathbf{X}_i) = E(\mathbf{u}_i\mathbf{u}_i').$$

Then

$$\text{Avar}(\hat{\boldsymbol{\beta}}_{FGLS}) = [E(\mathbf{X}_i'\boldsymbol{\Omega}^{-1}\mathbf{X}_i)]^{-1}/N \tag{2.18}$$

$$\begin{aligned}\text{Avar}(\hat{\boldsymbol{\beta}}_{SOLS}) &= [E(\mathbf{X}_i'\mathbf{X}_i)]^{-1}E(\mathbf{X}_i'\mathbf{u}_i\mathbf{u}_i'\mathbf{X}_i)[E(\mathbf{X}_i'\mathbf{X}_i)]^{-1} \\ &= [E(\mathbf{X}_i'\mathbf{X}_i)]^{-1}E(\mathbf{X}_i'\boldsymbol{\Omega}\mathbf{X}_i)[E(\mathbf{X}_i'\mathbf{X}_i)]^{-1}/N \end{aligned} \tag{2.19}$$

- Claim:

$$[E(\mathbf{X}_i'\mathbf{X}_i)]^{-1}E(\mathbf{X}_i'\boldsymbol{\Omega}\mathbf{X}_i)[E(\mathbf{X}_i'\mathbf{X}_i)]^{-1} - [E(\mathbf{X}_i'\boldsymbol{\Omega}^{-1}\mathbf{X}_i)]^{-1} \tag{2.20}$$

is positive semi-definite. Write the difference as $\mathbf{C} - \mathbf{D}$.

It suffices to show $\mathbf{D}^{-1} - \mathbf{C}^{-1}$ is p.s.d., that is

$$E(\mathbf{X}_i'\boldsymbol{\Omega}^{-1}\mathbf{X}_i) - E(\mathbf{X}_i'\mathbf{X}_i)[E(\mathbf{X}_i'\boldsymbol{\Omega}\mathbf{X}_i)]^{-1}E(\mathbf{X}_i'\mathbf{X}_i) \tag{2.21}$$

is p.s.d.

- Use the following trick. Let $\mathbf{Z}_i \equiv \mathbf{\Omega}^{-1/2}\mathbf{X}_i$ and $\mathbf{W}_i \equiv \mathbf{\Omega}^{1/2}\mathbf{X}_i$. Then $E(\mathbf{Z}_i'\mathbf{Z}_i) = E(\mathbf{X}_i'\mathbf{\Omega}^{-1}\mathbf{X}_i)$, $E(\mathbf{W}_i'\mathbf{W}_i) = E(\mathbf{X}_i'\mathbf{\Omega}\mathbf{X}_i)$, and $E(\mathbf{Z}_i'\mathbf{W}_i) = E(\mathbf{X}_i'\mathbf{X}_i)$. Therefore, the difference in (2.21) is

$$E(\mathbf{Z}_i'\mathbf{Z}_i) - E(\mathbf{Z}_i'\mathbf{W}_i)[E(\mathbf{W}_i'\mathbf{W}_i)]^{-1}E(\mathbf{Z}_i'\mathbf{W}_i), \tag{2.22}$$

which looks like a matrix sum of squared residuals in the population. In fact, if we define the matrix residuals $\mathbf{R}_i = \mathbf{Z}_i - \mathbf{W}_i\mathbf{\Pi}$ with $\mathbf{\Pi} \equiv [E(\mathbf{W}_i'\mathbf{W}_i)]^{-1}E(\mathbf{Z}_i'\mathbf{W}_i)$, then it is easily seen that (2.22) is $E(\mathbf{R}_i'\mathbf{R}_i)$, which is necessarily p.s.d.

3. FGLS WITH INCORRECT RESTRICTIONS ON THE VARIANCE MATRIX

- Suppose that, rather than estimate $\mathbf{\Omega}$ in an unrestricted fashion, so that $\text{plim}_{N\to\infty}\hat{\mathbf{\Omega}} = \mathbf{\Omega}$, we impose restrictions on the estimated matrix. This is very common for panel data, as we will see later. Let $\hat{\mathbf{\Lambda}}$ denote an estimator that may be inconsistent for $\mathbf{\Omega}$. Nevertheless, $\hat{\mathbf{\Lambda}}$ usually has a well-defined, nonsingular probability limit: $\mathbf{\Lambda} \equiv \text{plim}_{N\to\infty}\,\hat{\mathbf{\Lambda}}$.
- The FGLS estimator of $\boldsymbol{\beta}$ using $\hat{\mathbf{\Lambda}}$ as the variance matrix estimator is consistent if

$$E(\mathbf{X}_i'\mathbf{\Lambda}^{-1}\mathbf{u}_i) = \mathbf{0} \tag{3.1}$$

(along with the obvious modification of the rank condition SGLS.2).

- Condition (3.1) always holds if Assumption SGLS.1 holds. Therefore, exogeneity of each element of $\mathbf{X}_i$ in each equation (time period) ensures that using an inconsistent estimator of $\mathbf{\Omega}$ does not result in inconsistency of FGLS.
- The $\sqrt{N}$ –asymptotic equivalence between the estimators that use $\hat{\mathbf{\Lambda}}$ and $\mathbf{\Lambda}$ continues to hold under Assumption SGLS.1, and so we can conduct asymptotic inference ignoring the first stage estimation of $\mathbf{\Lambda}$.
- The analog of SGLS.3, namely, $E(\mathbf{X}_i'\mathbf{\Lambda}^{-1}\mathbf{u}_i\mathbf{u}_i'\mathbf{\Lambda}^{-1}\mathbf{X}_i) = E(\mathbf{X}_i'\mathbf{\Lambda}^{-1}\mathbf{X}_i)$, generally fails, even under the system homoskedasticity assumption $E(\mathbf{u}_i\mathbf{u}_i'|\mathbf{X}_i) = E(\mathbf{u}_i\mathbf{u}_i')$.

- Therefore, the sandwich estimator can be needed under system homoskedasticity if incorrect restrictions are imposed on the unconditional variance-covariance matrix.
- We will use this observation later for unobserved effects panel data models, as well as more traditional time series models for the errors.
- Question: When might we want to use a specific form of $\hat{\mathbf{\Lambda}}$ even if we know it is inconsistent for $\mathbf{\Omega}$? (Think about panel data without strictly exogenous regressors.)

4. TESTING USING FGLS

- Let the restrictions be given by

$$H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}, \tag{4.1}$$

where $\mathbf{R}$ is $Q \times K$, $\mathbf{r}$ is $Q \times 1$, $Q \leq K$. A generally available statistic is the **Wald statistic**:

$$\mathcal{W} = (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})'(\mathbf{R}\hat{\mathbf{V}}\mathbf{R}')^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})' \overset{a}{\sim} \chi^2_Q \tag{4.2}$$

under H_0, where $\hat{\mathbf{V}}$ is the fully robust form of Avar($\hat{\boldsymbol{\beta}}$) or the nonrobust form under SGLS.3.

• Under SGLS.1 through SGLS.3, can use a statistic based on sums of squared residuals. Given $\hat{\mathbf{\Omega}}$ – usually from the unrestricted SOLS estmation – let $\tilde{\boldsymbol{\beta}}$ denote the restricted FGLS estimator:

$$\tilde{\boldsymbol{\beta}} = \mathrm{argmin}_{\mathbf{b}\in\mathbb{R}^K} \sum_{i=1}^{N} (\mathbf{y}_i - \mathbf{X}_i\mathbf{b})'\hat{\mathbf{\Omega}}^{-1}(\mathbf{y}_i - \mathbf{X}_i\mathbf{b}) \tag{4.3}$$

$$\text{subject to } \mathbf{Rb} = \mathbf{r}$$

and let $\hat{\boldsymbol{\beta}}$ be the unrestricted estimator. Then

$$\sum_{i=1}^{N} \tilde{\mathbf{u}}_i'\hat{\mathbf{\Omega}}^{-1}\tilde{\mathbf{u}}_i \geq \sum_{i=1}^{N} \hat{\mathbf{u}}_i'\hat{\mathbf{\Omega}}^{-1}\hat{\mathbf{u}}_i \tag{4.4}$$

where $\tilde{\mathbf{u}}_i = \mathbf{y}_i - \mathbf{X}_i\tilde{\boldsymbol{\beta}}$ and $\hat{\mathbf{u}}_i = \mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}}$.

- Under SGLS.1 to SGLS.3, can show under H_0 that

$$\left(\sum_{i=1}^{N} \tilde{\mathbf{u}}_i' \hat{\mathbf{\Omega}}^{-1} \tilde{\mathbf{u}}_i - \sum_{i=1}^{N} \hat{\mathbf{u}}_i' \hat{\mathbf{\Omega}}^{-1} \hat{\mathbf{u}}_i \right) \overset{a}{\sim} \chi_Q^2. \tag{4.5}$$

- A small sample adjustment (with justification only via simulations), is

$$\mathcal{F} = \frac{\left(\sum_{i=1}^{N} \tilde{\mathbf{u}}_i' \hat{\mathbf{\Omega}}^{-1} \tilde{\mathbf{u}}_i - \sum_{i=1}^{N} \hat{\mathbf{u}}_i' \hat{\mathbf{\Omega}}^{-1} \hat{\mathbf{u}}_i \right)}{\left(\sum_{i=1}^{N} \hat{\mathbf{u}}_i' \hat{\mathbf{\Omega}}^{-1} \hat{\mathbf{u}}_i \right)} \cdot \frac{(NG - K)}{Q}, \tag{4.6}$$

treated as having an approximate $\mathfrak{F}_{Q,NG-K}$ distribution. Why? $\mathfrak{F}_{Q,NG-K} \overset{a}{\sim} \chi_Q^2/Q$ as $NG - K \to \infty$, so the division by Q makes it roughly valid to use the F distribution.

The other terms are based on

$$\begin{aligned} E(\mathbf{u}_i'\boldsymbol{\Omega}^{-1}\mathbf{u}_i) &= E[tr(\mathbf{u}_i'\boldsymbol{\Omega}^{-1}\mathbf{u}_i)] \\ &= E[tr(\boldsymbol{\Omega}^{-1}\mathbf{u}_i\mathbf{u}_i')] = tr[E(\boldsymbol{\Omega}^{-1}\mathbf{u}_i\mathbf{u}_i')] \\ &= tr[\boldsymbol{\Omega}^{-1}E(\mathbf{u}_i\mathbf{u}_i')] = tr(\mathbf{I}_G) = G. \end{aligned} \tag{4.7}$$

- It follows that

$$(NG)^{-1}\sum_{i=1}^{N}\mathbf{u}_i'\boldsymbol{\Omega}^{-1}\mathbf{u}_i \xrightarrow{p} 1, \tag{4.8}$$

and then insert $\hat{\boldsymbol{\beta}}$ and subtract off K from NG as a degrees-of-freedom adjustment.