Lecture 2: System OLS and Cross-Equation Restrictions

January 13, 2025

Seemingly Unrelated Regressions

The SUR Case

- A G-equation SUR system is written in the population as

$$\begin{aligned} y_1 =& \mathbf{x}_1 \boldsymbol{\beta}_1 + u_1 \\ y_2 =& \mathbf{x}_2 \boldsymbol{\beta}_2 + u_2 \\ &\vdots \\ y_G =& \mathbf{x}_G \boldsymbol{\beta}_G + u_G \end{aligned} \tag{1.1}$$

  where y_g is a response variable, $g = 1, \ldots, G$. The explanatory variables, $\mathbf{x}_g$, can be different across equations. For now, think if the $\boldsymbol{\beta}_g$ as being unrestricted across equations.

Seemingly Unrelated Regressions

EXAMPLE: Suppose each worker in a population receives three kinds of compensation: wage, pension, and health:

$$\begin{aligned} wage &= \beta_{10} + \beta_{11} educ + \beta_{12} tenure + \beta_{13} age + \beta_{14} union + u_1 \\ pension &= \beta_{20} + \beta_{21} educ + \beta_{22} tenure + \beta_{23} age + \beta_{24} union + u_2 \\ health &= \beta_{30} + \beta_{31} educ + \beta_{32} tenure + \beta_{33} age + \beta_{34} union + u_3 \end{aligned} \tag{1.2}$$

Then $G = 3$, and our random sample consists of workers from the specified population. We have three response variables in our population model.

- In some applications, especially to consumer and firm theory, the coefficients are restricted across equations (later).

Seemingly Unrelated Regressions

- What might we assume about exogeneity of the explanatory variables? In terms of conditional means, two possibilities:

$$E(u_g \mid \mathbf{x}_g) = 0, g = 1, \ldots, G, \tag{1.3}$$

which means

$$E(y_g \mid \mathbf{x}_g) = \mathbf{x}_g\boldsymbol{\beta}_g, g = 1, \ldots, G. \tag{1.4}$$

- We could instead use the weaker condition

$$E(\mathbf{x}_g' u_g) = \mathbf{0}, g = 1, \ldots, G. \tag{1.5}$$

Key point is that neither (1.3) nor (1.5) restricts the relationship between $\mathbf{x}_h$ and u_g for $g \neq h$.

Seemingly Unrelated Regressions

- A stronger assumption, implicitly or explicitly maintained by most SUR analyses, is

$$E(u_g \mid \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_g, \ldots, \mathbf{x}_G) = 0, g = 1, \ldots, G, \tag{1.6}$$

which implies

$$E(y_g \mid \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_g, \ldots, \mathbf{x}_G) = \mathbf{x}_g \boldsymbol{\beta}_g, g = 1, \ldots, G. \tag{1.7}$$

- This means that, if $\mathbf{x}_h$ for $h \neq g$ includes elements not in $\mathbf{x}_g$, then those elements of $\mathbf{x}_h$ are assumed to have no partial effect on the expected value of y_g once $\mathbf{x}_g$ is controlled for.

- Unless $\mathbf{x}_g = \mathbf{x}_h$ for all g and h,

$$E(u_g \mid \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_g, \ldots, \mathbf{x}_G) = E(u_g \mid \mathbf{x}_g)$$

imposes substantive exclusion restrictions. Treating the explanatory variables as fixed in repeated samples is operationally the same as this assumption (at least in terms of obtaining statistical properties).

- If $\mathbf{x}_g = \mathbf{x}_h$ for all g and h, there is no difference between (1.3) and (1.7).

System OLS Estimation

2.1. Consistency

- First estimation method uses OLS on the system of equations. We will be interested in what this entails for the SUR and panel data cases.
- It is notationally useful to use an i subscript when writing the system, to distinguish between unit-specific observations and full data matrices. We write the model for a random draw i as

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{u}_i \tag{2.1}$$

  where $\mathbf{y}_i$ is $G \times 1, \mathbf{X}_i$ is $G \times K, \boldsymbol{\beta}$ is $K \times 1$, and $\mathbf{u}_i$ is $G \times 1$. The observed data is $\{(\mathbf{X}_i, \mathbf{y}_i) : i = 1, \ldots, N\}$, where N is the sample size.
- We want to estimate the population parameter vector $\boldsymbol{\beta}$.

System OLS Estimation

SUR

In the SUR case, $\mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{iG})', \mathbf{u}_i = (u_{i1}, u_{i2}, \ldots, u_{iG})'$,

$$\mathbf{X}_i = \begin{pmatrix} \mathbf{x}_{i1} & \mathbf{0} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{x}_{i2} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} & \vdots \\ \vdots & \vdots & \mathbf{0} & \mathbf{x}_{i,G-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{x}_{iG} \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_{G-1} \\ \boldsymbol{\beta}_G \end{pmatrix}. \tag{2.2}$$

If $\mathbf{x}_{ig}$ is $1 \times K_g$, define $K = K_1 + K_2 + \ldots + K_G$, and then $\mathbf{X}_i$ is $G \times K$ and $\boldsymbol{\beta}$ is $K \times 1$.

System OLS Estimation

Assumptions for System OLS (SOLS)
Assumption SOLS.1:

$$E\left(\mathbf{X}_i'\mathbf{u}_i\right) = \mathbf{0} \tag{2.3}$$

- This is the weakest possible assumption without moving into instrumental variables territory.
- In the SUR case,

$$\mathbf{X}_i'\mathbf{u}_i = \begin{pmatrix} \mathbf{x}_{i1}'u_{i1} \\ \mathbf{x}_{i2}'u_{i2} \\ \vdots \\ \mathbf{x}_{iG}'u_{iG} \end{pmatrix} \tag{2.4}$$

  and so SOLS.1 is equivalent to

$$E\left(\mathbf{x}_{ig}'u_{ig}\right) = \mathbf{0}, g = 1, \ldots, G. \tag{2.5}$$

- SOLS.1 does not restrict relationship between u_{ig} and covariates in other equations.

System OLS Estimation

Assumption SOLS.2:

$$\text{rank}\, E(\mathbf{X}_i'\mathbf{X}_i) = K. \tag{2.6}$$

SUR

- In the SUR case,

$$\mathbf{X}_i'\mathbf{X}_i = \begin{pmatrix} \mathbf{x}_{i1}'\mathbf{x}_{i1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{x}_{i2}'\mathbf{x}_{i2} & \mathbf{0} & \vdots \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{x}_{iG}'\mathbf{x}_{iG} \end{pmatrix} \tag{2.7}$$

- SOLS. 2 holds if and only if

$$\text{rank}\, E(\mathbf{x}_{ig}'\mathbf{x}_{ig}) = K_g, g = 1, \ldots, G, \tag{2.8}$$

  which simply says that the single-equation OLS rank condition (OLS.2) holds for each equation.

System OLS Estimation

- The SOLS estimator looks just like the single-equations OLS estimator, but $\mathbf{X}_i$ is a matrix and $\mathbf{y}_i$ is a vector:

$$\hat{\boldsymbol{\beta}}_{SOLS} = \left(N^{-1} \sum_{i=1}^{N} \mathbf{X}_i' \mathbf{X}_i \right)^{-1} \left(N^{-1} \sum_{i=1}^{N} \mathbf{X}_i' \mathbf{y}_i \right), \tag{2.9}$$

  which can be written as $\hat{\boldsymbol{\beta}}_{SOLS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ where $\mathbf{X}$ is $NG \times K$ and $\mathbf{Y}$ is $NG \times 1$.

THEOREM: Under SOLS. 1 and SOLS.2, OLS on a random sample is consistent:

$$\text{plim}_{N\to\infty}(\hat{\boldsymbol{\beta}}_{SOLS}) = \boldsymbol{\beta} \tag{2.10}$$

System OLS Estimation

- What is the SOLS estimator in the SUR case? Use

$$\sum_{i=1}^{N} \mathbf{X}_i'\mathbf{X}_i = \begin{pmatrix} \sum_{i=1}^{N} \mathbf{x}_{i1}'\mathbf{x}_{i1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \sum_{i=1}^{N} \mathbf{x}_{i2}'\mathbf{x}_{i2} & \mathbf{0} & \vdots \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \sum_{i=1}^{N} \mathbf{x}_{iG}'\mathbf{x}_{iG} \end{pmatrix} \tag{2.11}$$

$$\sum_{i=1}^{N} \mathbf{X}_i'\mathbf{y}_i = \begin{pmatrix} \sum_{i=1}^{N} \mathbf{x}_{i1}'y_{i1} \\ \sum_{i=1}^{N} \mathbf{x}_{i2}'y_{i2} \\ \vdots \\ \sum_{i=1}^{N} \mathbf{x}_{iG}'y_{iG} \end{pmatrix},$$

System OLS Estimation

$$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \sum_{i=1}^{N} \mathbf{x}_{i1}'\mathbf{x}_{i1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \sum_{i=1}^{N} \mathbf{x}_{i2}'\mathbf{x}_{i2} & \mathbf{0} & \vdots \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \sum_{i=1}^{N} \mathbf{x}_{iG}'\mathbf{x}_{iG} \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^{N} \mathbf{x}_{i1}' y_{i1} \\ \sum_{i=1}^{N} \mathbf{x}_{i2}' y_{i2} \\ \vdots \\ \sum_{i=1}^{N} \mathbf{x}_{iG}' y_{iG} \end{pmatrix}$$

$$= \begin{pmatrix} \left(\sum_{i=1}^{N} \mathbf{x}_{i1}'\mathbf{x}_{i1}\right)^{-1} \sum_{i=1}^{N} \mathbf{x}_{i1}' y_{i1} \\ \left(\sum_{i=1}^{N} \mathbf{x}_{i2}'\mathbf{x}_{i2}\right)^{-1} \sum_{i=1}^{N} \mathbf{x}_{i2}' y_{i2} \\ \vdots \\ \left(\sum_{i=1}^{N} \mathbf{x}_{iG}'\mathbf{x}_{iG}\right)^{-1} \sum_{i=1}^{N} \mathbf{x}_{iG}' y_{iG} \end{pmatrix} = \begin{pmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \\ \vdots \\ \hat{\boldsymbol{\beta}}_G \end{pmatrix}$$

- Each $\hat{\boldsymbol{\beta}}_g$ is just the OLS estimator on equation g.
- In this case, system OLS is **ordinary least squares equation-by-equation**.

System OLS Estimation

2.2 Asymptotic Normality and Inference

- The proof of asymptotic normality is similar to the single-equation case, but we have to be more careful with the linear algebra. Write

$$N^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \left(N^{-1} \sum_{i=1}^{N} \mathbf{X}_i' \mathbf{X}_i \right)^{-1} \left(N^{-1/2} \sum_{i=1}^{N} \mathbf{X}_i' \mathbf{u}_i \right). \tag{2.12}$$

- By the CLT for i.i.d. random vectors,

$$N^{-1/2} \sum_{i=1}^{N} \mathbf{X}_i' \mathbf{u}_i \xrightarrow{d} \text{Normal}(\mathbf{0}, \mathbf{B}) \tag{2.13}$$

$$\mathbf{B} = \text{Var}\left(\mathbf{X}_i' \mathbf{u}_i\right) = E\left(\mathbf{X}_i' \mathbf{u}_i \mathbf{u}_i' \mathbf{X}_i\right). \tag{2.14}$$

System OLS Estimation

- Define

$$\mathbf{A} = E\left(\mathbf{X}_i'\mathbf{X}_i\right). \tag{2.15}$$

$$\begin{aligned} N^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) =& \mathbf{A}^{-1}\left(N^{-1/2}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{u}_i\right) \\ &+ \left[\left(N^{-1}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{X}_i\right)^{-1} - \mathbf{A}^{-1}\right]\left(N^{-1/2}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{u}_i\right) \\ =& \mathbf{A}^{-1}\left(N^{-1/2}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{u}_i\right) + o_p(1)\cdot O_p(1). \end{aligned} \tag{2.16}$$

- So

$$N^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} \text{Normal}\left(\mathbf{0}, \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\right). \tag{2.17}$$

$$\hat{\mathbf{A}} = N^{-1}\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{X}_i \quad (\text{ a } K\times K \text{ matrix }) \tag{2.18}$$

- The SOLS residuals are

$$\hat{\mathbf{u}}_i = \mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}} \text{ (each is } G\times 1) \tag{2.19}$$

System OLS Estimation

- A fully robust estimator of $\mathbf{B}$ - that is, an estimator valid under SOLS. 1 and SOLS.2, without any second moment assumptions on $\mathbf{u}_i$ - is

$$\hat{\mathbf{B}} = (N-K)^{-1}\sum_{i=1}^{N}\mathbf{X}_i'\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i'\mathbf{X}_i. \tag{2.20}$$

- The degrees-of-freedom adjustment is common but unecessary with large N. The resulting sandwich estimator is

$$\widehat{\text{Avar}}(\hat{\boldsymbol{\beta}}) = \frac{N}{(N-K)}\left(\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{X}_i\right)^{-1}\left(\sum_{i=1}^{N}\mathbf{X}_i'\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i'\mathbf{X}_i\right)\left(\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{X}_i\right)^{-1} \tag{2.21}$$

Generalized Least Squares

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 $\mathbf{X}_i$.
- The $G \times G$ unconditional variance-covariance matrix plays a key role.

$$\Omega \equiv E\left(\mathbf{u}_i \mathbf{u}_i'\right) = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1G} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2G} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1G} & \sigma_{2G} & \cdots & \sigma_G^2 \end{pmatrix}. \tag{3.1}$$

Generalized Least Squares

- 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 Ω.

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

$$\Omega^{-1/2}\mathbf{y}_i = \Omega^{-1/2}\mathbf{X}_i\beta + \Omega^{-1/2}\mathbf{u}_i, \tag{3.2}$$

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

Generalized Least Squares

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

$$\begin{aligned}\beta^* &= \left(N^{-1}\sum_{i=1}^{N}\mathbf{X}_i^{*\prime}\mathbf{X}_i^*\right)^{-1}\left(N^{-1}\sum_{i=1}^{N}\mathbf{X}_i^{*\prime}\mathbf{y}_i^*\right) \\ &= \left(N^{-1}\sum_{i=1}^{N}\mathbf{X}_i'\Omega^{-1}\mathbf{X}_i\right)^{-1}\left(N^{-1}\sum_{i=1}^{N}\mathbf{X}_i'\Omega^{-1}\mathbf{y}_i\right) \qquad (3.3) \\ &= \beta + \left(N^{-1}\sum_{i=1}^{N}\mathbf{X}_i'\Omega^{-1}\mathbf{X}_i\right)^{-1}\left(N^{-1}\sum_{i=1}^{N}\mathbf{X}_i'\Omega^{-1}\mathbf{u}_i\right)\end{aligned}$$

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

$$E\left(\mathbf{X}_i'\Omega^{-1}\mathbf{u}_i\right) = 0 \qquad (3.4)$$

- In general, (3.4) is not implied by SOLS.1,

$$E\left(\mathbf{X}_i'\mathbf{u}_i\right) = 0. \qquad (3.5)$$

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

Generalized Least Squares

Assumption SGLS.1 (Exogeneity):

$$E\left(\mathbf{X}_i \otimes \mathbf{u}_i\right) = 0 \tag{3.6}$$

- 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, (3.4) holds.

Assumption SGLS.2 (Rank Condition): Ω is nonsingular and $E\left(\mathbf{X}_i'\Omega^{-1}\mathbf{X}_i\right)$ is nonsingular.
THEOREM: Under SGLS. 1 and SGLS. 2, β^* is consistent for β as $N \to \infty$.

Generalized Least Squares

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

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

$$\begin{aligned} \Omega^{-1} &= \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{12} & \omega_{22} \end{pmatrix} \\ \Omega^{-1}\mathbf{X}_i &= \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{12} & \omega_{22} \end{pmatrix} \begin{pmatrix} \mathbf{x}_{i1} & \mathbf{0} \\ \mathbf{0} & \mathbf{x}_{i2} \end{pmatrix} = \begin{pmatrix} \omega_{11}\mathbf{x}_{i1} & \omega_{12}\mathbf{x}_{i2} \\ \omega_{12}\mathbf{x}_{i1} & \omega_{22}\mathbf{x}_{i2} \end{pmatrix}. \end{aligned} \tag{3.7}$$

Then

$$E\left[\left(\Omega^{-1}\mathbf{X}_i\right)'\mathbf{u}_i\right] = \begin{pmatrix} \omega_{11}E\left(\mathbf{x}_{i1}'u_{i1}\right) + \omega_{12}E\left(\mathbf{x}_{i1}'u_{i2}\right) \\ \omega_{12}E\left(\mathbf{x}_{i2}'u_{i1}\right) + \omega_{22}E\left(\mathbf{x}_{i2}'u_{i2}\right) \end{pmatrix}. \tag{3.8}$$

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.

Unobserved VC Matrix: FGLS

- In SUR analysis, we almost always use

$$\hat{\Omega} = N^{-1} \sum_{i=1}^{N} \check{\mathbf{u}}_i \check{\mathbf{u}}_i' \tag{4.1}$$

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

- Write

$$\begin{aligned} \check{\mathbf{u}}_i =& \mathbf{y}_i - \mathbf{X}_i \beta - \mathbf{X}_i(\check{\beta} - \beta) = \mathbf{u}_i - \mathbf{X}_i(\check{\beta} - \beta) \\ \check{\mathbf{u}}_i \check{\mathbf{u}}_i' =& \mathbf{u}_i \mathbf{u}_i' - \mathbf{u}_i(\check{\beta} - \beta)' \mathbf{X}_i' - \mathbf{X}_i(\check{\beta} - \beta)\mathbf{u}_i' \\ &+ \mathbf{X}_i(\check{\beta} - \beta)(\check{\beta} - \beta)' \mathbf{X}_i' \end{aligned} \tag{4.2}$$

can show that

$$\hat{\Omega} = N^{-1} \sum_{i=1}^{N} \mathbf{u}_i \mathbf{u}_i' + o_p(1) \tag{4.3}$$

Unobserved VC Matrix: FGLS

- The FGLS estimator is

$$\hat{\beta} = \left(N^{-1}\sum_{i=1}^{N}\mathbf{X}_i'\hat{\Omega}^{-1}\mathbf{X}_i\right)^{-1}\left(N^{-1}\sum_{i=1}^{N}\mathbf{X}_i'\hat{\Omega}^{-1}\mathbf{y}_i\right)$$
$$= \beta + \left(N^{-1}\sum_{i=1}^{N}\mathbf{X}_i'\hat{\Omega}^{-1}\mathbf{X}_i\right)^{-1}\left(N^{-1}\sum_{i=1}^{N}\mathbf{X}_i'\hat{\Omega}^{-1}\mathbf{u}_i\right) \tag{4.4}$$

- Some algebra can show the following

$$\sqrt{N}(\hat{\beta}-\beta) = \left(N^{-1}\sum_{i=1}^{N}\mathbf{X}_i'\Omega^{-1}\mathbf{X}_i\right)^{-1}\left(N^{-1/2}\sum_{i=1}^{N}\mathbf{X}_i'\Omega^{-1}\mathbf{u}_i\right) + o_p(1) \tag{4.5}$$
$$= \sqrt{N}\,(\beta^* - \beta) + o_p(1). \tag{4.6}$$

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

Unobserved VC Matrix: FGLS

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

$$\widehat{\text{Avar}}(\hat{\beta}) = \left(\sum_{i=1}^{N} \mathbf{X}_i'\hat{\Omega}^{-1}\mathbf{X}_i\right)^{-1} \left(\sum_{i=1}^{N} \mathbf{X}_i'\hat{\Omega}^{-1}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i'\hat{\Omega}^{-1}\mathbf{X}_i\right)^{-1} \cdot \left(\sum_{i=1}^{N} \mathbf{X}_i'\hat{\Omega}^{-1}\mathbf{X}_i\right)^{-1} \tag{4.7}$$

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

When FGLS is more efficient

When is the "Usual" Variance Matrix Estimator for FGLS Valid?
Assumption SGLS. 3 (System Homoskedasticity):

$$E\left(\mathbf{X}_i'\Omega^{-1}\mathbf{u}_i\mathbf{u}_i'\Omega^{-1}\mathbf{X}_i\right) = E\left(\mathbf{X}_i'\Omega^{-1}\mathbf{X}_i\right). \tag{4.8}$$

- 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

$$\widehat{\text{Avar}}(\hat{\beta}) = \left(\sum_{i=1}^{N}\mathbf{X}_i'\hat{\Omega}^{-1}\mathbf{X}_i\right)^{-1} \tag{4.9}$$

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

FGLS vs SOLS

- If the same regressors appear in each equation, that is, $\mathbf{x}_{ig} = \mathbf{x}_i$, $g = 1, \ldots, G$, the OLS equation-by-equation is numerically the same as FGLS for any structure of $\hat{\Omega}$.

  The matrix of regressors can be written as $\mathbf{X}_i = \mathbf{I}_G \otimes \mathbf{x}_i$.
- FGLS is more efficiency, if assumptions (1) - (3) hold
- Otherwise, it would not be consistent
- The standard efficiency vs bias tradeoff

Christensen and Greene (1976)

- Economics of scale in US electric power
    - Increase in competition
    - Decrease in scale economies
- Empirical question depends on cost function estimation
- Using data of U.S electric power industry from 1955 - 1970 (the paper was written in 70s)
- Found no evidence of scale economies in 70s

Use Traditional Steam Plants

PERCENTAGE OF TOTAL PRODUCTION BY TYPE OF PLANT*

Year	Steam	Hydroelectric	Internal Combustion	Nuclear
1950	69.8	29.1	1.1	0
1970	90.7	6.2	1.5	1.6

* The 1950 figures are from Caywood (1956) and refer to total U.S. production. The 1970 figures are from the Federal Power Commission (1971*a*). These figures are for investor-owned utilities only and neglect the large federal projects. Inclusion of these would increase the percentage of hydroelectric power somewhat. For example, TVA produces 6 percent of the nation's power, and 20 percent of its capacity is hydroelectric (*Forbes Magazine* 1975, p. 25).

Cost Function

- Cost function approach is attractive, especially in regulated industries, where output is not determined by firm.
- Duality theory and a flexible modeling of translog cost function (Christensen, Jorgenson, and Lau 1971,73)

Symmetry, HOD1 in Factor Prices

The translog cost function can be written

$$\left.\begin{aligned}\ln C = \quad & \alpha_0 + \alpha_Y \ln Y + \frac{1}{2}\gamma_{YY}(\ln Y)^2 \\ & + \sum_i \alpha_i \ln P_i + \frac{1}{2}\sum_i \sum_j \gamma_{ij} \ln P_i \ln P_j \\ & + \sum_i \gamma_{Y_i} \ln Y \ln P_i,\end{aligned}\right\} \text{Model A}$$

$$\gamma_{ij} = \gamma_{ji} \quad \text{Symmetry}$$

Homogeneous of Degree One in Factor Prices:

$$\sum_i \alpha_i = 1$$

$$\sum_i \gamma_{Y_i} = 0$$

$$\sum_i \gamma_{ij} = \sum_j \gamma_{ji} = \sum_i \sum_j \gamma_{ij} = 0$$

Shephard's Lemma

$$\partial C/\partial P_i = X_i. \quad \partial \ln C/\partial \ln P_i = \frac{P_i X_i}{C} = S_i$$

$$S_i = \alpha_i + \gamma_{Y_i} \ln Y + \sum_j \gamma_{ij} \ln P_j.$$

Define Economies of Scale

$$\mathrm{SCE} = 1 - \partial \ln C/\partial \ln Y.$$

Homothetic

$$\gamma_{Y_i} = 0$$

Homogenous

$$\gamma_{Y_i} = 0, \quad \gamma_{Y_Y} = 0.$$

Unity

$$\gamma_{ij} = 0.$$

Scale Economics

From the paper

Hanoch (1975) has pointed out that it is more appropriate to represent scale economies by the relationship between total cost and output along the expansion path - where input prices are constant and costs are minimized at every level of output. A natural way to express the extent of scale economies is as the proportional increase in cost resulting from a small proportional increase in the level of output, or the elasticity of total cost with respect to output. We will define scale economies (SCE) as unity minus this elasticity:

$$\mathrm{SCE} = 1 - \partial \ln C / \partial \ln Y.$$

TABLE 2

SCALE ECONOMIES FOR MODELS A–F

$$\begin{aligned}
\text{SCE (A)} &= 1 - (\alpha_Y + \gamma_{YY} \ln Y + \textstyle\sum_i \gamma_{Yi} \ln P_i) \\
\text{SCE (B)} &= 1 - (\alpha_Y + \gamma_{YY} \ln Y) \\
\text{SCE (C)} &= 1 - \alpha_Y \\
\text{SCE (D)} &= 1 - (\alpha_Y + \gamma_{YY} \ln Y + \textstyle\sum_i \gamma_{Yi} \ln P_i) \\
\text{SCE (E)} &= 1 - (\alpha_Y + \gamma_{YY} \ln Y) \\
\text{SCE (F)} &= 1 - \alpha_Y
\end{aligned}$$

Detour: how to impose restrictions

- Suppose a two-equation system in the population is

$$y_1 = \gamma_{10} + \gamma_{11}x_{11} + \gamma_{12}x_{12} + \alpha_1 x_{13} + \alpha_2 x_{14} + u_1$$
$$y_2 = \gamma_{20} + \gamma_{21}x_{21} + \alpha_1 x_{22} + \alpha_2 x_{23} + \gamma_{24}x_{24} + u_2$$

- Let the vector of all parameters be the 8×1 vector

$$\boldsymbol{\beta} = (\gamma_{10}, \gamma_{11}, \gamma_{12}, \alpha_1, \alpha_2, \gamma_{20}, \gamma_{21}, \gamma_{24})'.$$

- Then we can define the matrix of regressors as

$$\mathbf{X}_i = \begin{pmatrix} 1 & x_{i11} & x_{i12} & x_{i13} & x_{i14} & 0 & 0 & 0 \\ 0 & 0 & 0 & x_{i22} & x_{i23} & 1 & x_{i21} & x_{i24} \end{pmatrix}$$

Detour: avoid singular VC matrix

- In expenditure and cost share systems, the G responses, if the categories are exhaustive and mutually exclusive, sum to unity.
- For firm i let s_{iK}, s_{iL}, and s_{iM} be the cost shares for capital, labor, and materials, respectively, and assume that $s_{iK} + s_{iL} + s_{iM} = 1$. A popular cost-share system is

$$s_{iK} = \gamma_{10} + \gamma_{11} \log(p_{iK}) + \gamma_{12} \log(p_{iL}) + \gamma_{13} \log(p_{iM}) + u_{iK}$$
$$s_{iL} = \gamma_{20} + \gamma_{21} \log(p_{iK}) + \gamma_{22} \log(p_{iL}) + \gamma_{23} \log(p_{iM}) + u_{iL}$$
$$s_{iM} = \gamma_{30} + \gamma_{31} \log(p_{iK}) + \gamma_{32} \log(p_{iL}) + \gamma_{33} \log(p_{iM}) + u_{iM}$$

Detour: avoid singular VC matrix

- The restriction on the sum implies

$$\gamma_{10} + \gamma_{20} + \gamma_{30} = 1, \gamma_{11} + \gamma_{21} + \gamma_{31} = 0, \gamma_{12} + \gamma_{22} + \gamma_{32} = 0$$
$$\gamma_{13} + \gamma_{23} + \gamma_{33} = 0, u_{iK} + u_{iL} + u_{iM} = 0$$

  and that last restriction implies that $\Omega = E\left(\mathbf{u}_i \mathbf{u}_i'\right)$, a 3×3 matrix, has rank two, not three.

- Can drop any of the equations. Make it the last one, and impose the restrictions (i.e. symmetry) on the parameters. Can write

$$s_{iK} = \gamma_{10} + \gamma_{11} \log\left(p_{iK}/p_{iM}\right) + \gamma_{12} \log\left(p_{iL}/p_{iM}\right) + u_{iK}$$
$$s_{iL} = \gamma_{20} + \gamma_{12} \log\left(p_{iK}/p_{iM}\right) + \gamma_{22} \log\left(p_{iL}/p_{iM}\right) + u_{iL}$$

Detour: avoid singular VC matrix

- This two-equation system has a cross equation restriction, too. But the singularity in the variance matrix is gone, so can apply FGLS with

$$\boldsymbol{\beta} = (\gamma_{10}, \gamma_{11}, \gamma_{12}, \gamma_{20}, \gamma_{22})$$

$$\mathbf{X}_i = \begin{pmatrix} 1 & \log(p_{iK}/p_{iM}) & \log(p_{iL}/p_{iM}) & 0 & 0 \\ 0 & 0 & \log(p_{iK}/p_{iM}) & 1 & \log(p_{iL}/p_{iM}) \end{pmatrix}$$

- Can add firm characteristics to the share equations without essential change.

Inputs and Outputs

- Inputs
  - Fuel (state wide average of plant location, plant level type of fuel variation)
  - Capital (?)
  - Labor (labor cost/full time employees reported at firm level)
- Output
  - Electricity generated

The Most General Model

TABLE 4

COST FUNCTION PARAMETER ESTIMATES FOR MODEL A, 1955II AND 1970 DATA
(t-RATIOS IN PARENTHESES)

Parameter	1955II	1970	Parameter	1955II	1970
α_0	8.412 (31.52)	7.14 (32.45)	γ_{YF}	0.024 (5.14)	0.021 (6.64)
α_Y	0.386 (6.22)	0.587 (20.87)	γ_{KK}	0.175 (5.51)	0.118 (6.17)
α_K	0.094 (0.94)	0.208 (2.95)	γ_{LL}	0.038 (2.03)	0.081 (5.00)
α_L	0.348 (4.21)	−0.151 (−1.85)	γ_{FF}	0.176 (6.83)	0.178 (10.79)
α_F	0.558 (8.57)	0.943 (14.64)	γ_{KL}	−0.018 (−1.01)	−0.011 (−0.749)
γ_{YY}	0.059 (5.76)	0.049 (12.94)	γ_{KF}	−0.156 (−6.05)	−0.107 (−7.48)
γ_{YK}	−0.008 (−1.79)	−0.003 (−1.23)	γ_{LF}	−0.020 (−2.08)	−0.070 (−6.30)
γ_{YL}	−0.016 (−10.10)	−0.018 (−8.25)			

Scale Economics: Non-homogenous matters

TABLE 7

ESTIMATED SCALE ECONOMIES UNDER VARIOUS SPECIFICATIONS OF TECHNOLOGY
(t-RATIOS IN PARENTHESES)

	Size Group				
	1	2	3	4	5
Output (million kwh) ..	43	338	1,109	2,226	5,819
			1955I		
Model:					
Homogeneous:					
F	.203 (13.61)	.203 (13.61)	.203 (13.61)	.203 (13.61)	.203 (13.61)
C	.190 (13.11)	.190 (13.11)	.190 (13.11)	.190 (13.11)	.190 (13.11)
Homothetic:					
E	.388 (17.00)	.216 (16.90)	.117 (6.28)	.059 (2.43)	−.020 (−0.62)
B	.369 (16.62)	.208 (16.66)	.113 (6.20)	.059 (2.44)	−.017 (−0.53)
Nonhomothetic:					
D	.418 (18.00)	.258 (18.53)	.153 (7.94)	.096 (3.83)	.026 (0.77)
A	.408 (17.88)	.258 (18.44)	.157 (8.25)	.104 (4.16)	.040 (1.20)
			1955II		
Model A.............	.351 (13.66)	.243 (15.67)	.167 (8.68)	.27 (5.15)	.076 (2.26)