We begin our discussion of systems of equations with a system that is not simultaneous in nature. Consider two dependent variables that are considered as a group because they bear a close conceptual relationship to one another. Aside from this conceptual relationship, the two linear regression models have, outwardly, no connection with one another. Each equation models a different dependent variable and the regressors in each equation need not be the same. The two equations are

\[ Y_{t,1} = X_{t,1}' \beta_1 + U_{t,1} \]

\[ Y_{t,2} = X_{t,2}' \beta_2 + U_{t,2}. \]

Because the two equations appear unrelated, we would think of estimating them separately. (And the system is referred to as a system of seemingly unrelated regression (SUR) equations.) Of course, if the two equations actually are unrelated, then we should estimate them separately. But it may be the case that there is a relation between the equations, brought forward by correlation between the two error terms. If the two error terms are correlated, then we can gain a more efficient estimator by estimating the two equations jointly, as was shown by Zellner in 1962. The intuition mirrors the intuition for a single equation with serial correlation. If \( U_{t,1} \) is correlated with \( U_{t,2} \), then knowledge of \( U_{t,2} \) can help us reduce our predicted value of \( U_{t,1} \).

The formal assumption is that \( U_t = (U_{t,1}, U_{t,2})' \) is an i.i.d. random variable with mean zero and covariance matrix

\[ E(U_t U'_t) = \Omega = \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{12} & \omega_{22} \end{bmatrix}. \]

To proceed in detail, consider the special case in which the elements of \( \Omega \) are known. The system (1) is combined as

\[ Y = X \beta + U, \]

in which

\[ Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}, \quad X = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}, \quad \text{and} \quad U = \begin{bmatrix} U_1 \\ U_2 \end{bmatrix}. \]
The covariance matrix for the system is $\Gamma = E(UU')$ where

$$\Gamma = \begin{bmatrix} \omega_{11} I & \omega_{12} I \\ \omega_{12} I & \omega_{22} I \end{bmatrix} = \Omega \otimes I.$$ 

The SUR estimator is

$$B_S = (X'\Gamma^{-1}X)^{-1}X'\Gamma^{-1}Y.$$ 

Our task is made simpler if we recognize in advance that

$$B_S = (X'd\Gamma^{-1}X)^{-1}X'd\Gamma^{-1}Y,$$

where the scalar $d = \omega_{11}\omega_{22} - \omega_{12}^2$, for

$$d\Gamma^{-1} = (d\Omega^{-1} \otimes I) = \begin{bmatrix} \omega_{22} I & -\omega_{12} I \\ -\omega_{12} I & \omega_{11} I \end{bmatrix}.$$ 

To simplify calculation, let each model be expressed in deviation-from-mean form with a single regressor, so

$$X'\Gamma^{-1}X = \begin{bmatrix} -\omega_{12} \sum_{t=1}^n X_{t,1}^2 & -\omega_{12} \sum_{t=1}^n X_{t,1}X_{t,2} \\ -\omega_{12} \sum_{t=1}^n X_{t,1}X_{t,2} & \omega_{11} \sum_{t=1}^n X_{t,2}^2 \end{bmatrix},$$

so

$$(X'\Gamma^{-1}X)^{-1} = \frac{1}{C} \begin{bmatrix} \omega_{11} \sum_{t=1}^n X_{t,1}^2 & \omega_{12} \sum_{t=1}^n X_{t,1}X_{t,2} \\ \omega_{12} \sum_{t=1}^n X_{t,1}X_{t,2} & \omega_{22} \sum_{t=1}^n X_{t,2}^2 \end{bmatrix},$$

with $C = \omega_{11}\omega_{22} \sum_{t=1}^n X_{t,1}^2 \cdot \sum_{t=1}^n X_{t,2} - \omega_{12}^2 (\sum_{t=1}^n X_{t,1}X_{t,2})^2$. Also

$$X'd\Gamma^{-1}Y = \begin{bmatrix} \omega_{22} \sum_{t=1}^n X_{t,1}Y_{t,1} - \omega_{12} \sum_{t=1}^n X_{t,1}X_{t,2}Y_{t,2} \\ \omega_{11} \sum_{t=1}^n X_{t,2}Y_{t,1} - \omega_{12} \sum_{t=1}^n X_{t,1}X_{t,2}Y_{t,2} \end{bmatrix}.$$ 

The SUR estimators $B_S$ equal

$$B_S = \frac{1}{C} \left[ \begin{array}{c} \omega_{11} \sum_{t=1}^n X_{t,2}^2 (\omega_{22} \sum_{t=1}^n X_{t,1}Y_{t,1} - \omega_{12} \sum_{t=1}^n X_{t,1}X_{t,2}) + \omega_{12} \sum_{t=1}^n X_{t,1}X_{t,2} (\omega_{11} \sum_{t=1}^n X_{t,2}Y_{t,1} - \omega_{12} \sum_{t=1}^n X_{t,1}Y_{t,2}) + \omega_{22} \sum_{t=1}^n X_{t,1}^2 (\omega_{11} \sum_{t=1}^n X_{t,2}Y_{t,1} - \omega_{12} \sum_{t=1}^n X_{t,2}Y_{t,2}) \end{array} \right].$$

There are two cases in which simplification occurs. First, if $\omega_{12} = 0$, then

$$B_S = \frac{1}{\omega_{11}\omega_{22} \sum_{t=1}^n X_{t,1}^2 \cdot \sum_{t=1}^n X_{t,2}^2} \left[ \begin{array}{c} \omega_{11} \sum_{t=1}^n X_{t,2}^2 \omega_{22} \sum_{t=1}^n X_{t,1}Y_{t,1} \\ \omega_{22} \sum_{t=1}^n X_{t,1}^2 \omega_{11} \sum_{t=1}^n X_{t,2}Y_{t,2} \end{array} \right] = \left[ \begin{array}{c} \sum_{t=1}^n X_{t,1}Y_{t,1} \\ \sum_{t=1}^n X_{t,2}Y_{t,2} \end{array} \right].$$
If the errors are uncorrelated across equations, then there is no relation between the equations and the SUR estimators are identical to the OLS estimators for each equation separately.

Second, if $X_1 = X_2$, then $B_s$ equals

$$
B_s = \frac{1}{\omega_{11}\omega_{22} - \omega_{12}^2} \left[ \begin{array}{c} \omega_{11} \sum X_t^2 (\omega_{22} \sum X_t Y_{t,1} - \omega_{12} \sum X_t Y_{t,2}) + \omega_{12} \sum X_t^2 (\omega_{11} \sum X_t Y_{t,2} - \omega_{12} \sum X_t Y_{t,1}) \\ \omega_{12} \sum X_t^2 (\omega_{22} \sum X_t Y_{t,1} - \omega_{12} \sum X_t Y_{t,2}) + \omega_{22} \sum X_t^2 (\omega_{11} \sum X_t Y_{t,2} - \omega_{12} \sum X_t Y_{t,1}) \end{array} \right].
$$

If the regressor is the same in each equation, then there is no additional information from the cross-equation correlation.

To understand why identical regressors should produce such a result, we study the special case in which there is only one observation for each regression. The formula for $B_s$ simplifies to

$$
B_s = \frac{1}{X_1^2 X_2^2 (\omega_{11}\omega_{22} - \omega_{12}^2)} \left[ \begin{array}{c} X^2 (\omega_{11}\omega_{22} - \omega_{12}^2) X Y_1 \\ X^2 (\omega_{11}\omega_{22} - \omega_{12}^2) X Y_2 \end{array} \right].
$$

The SUR estimator is identical to the equation-by-equation OLS estimator, even if $\omega_{12} \neq 0$ and $X_1 \neq X_2$. We see that the SUR estimator gives a different weight to each observation within an equation; for the first equation

$$
B_{s,1} = \frac{wX^2 \cdot X Y_1}{wX^2 \cdot X Y_1},
$$

with $w = \omega_{11}\omega_{22} - \omega_{12}^2$. Yet as there is only one observation for each equation, the SUR estimator cannot give different weights to the observations in an equation and is identical to the OLS estimator, which gives equal weight to each observation within an equation. For $n > 1$, the weights are not simply $wX^2$, the variation in the weights depends on the variation between $X_1$ and $X_2$. With identical regressors, the weight variation disappears and the SUR estimator is identical to OLS.

In practice, $\Omega$ is unknown. The (feasible) implementation of the estimator is achieved in two stages. In the first stage, estimate each equation by OLS, yielding $B_1$, $B_2$, $U_1^P$ and $U_2^P$. We then estimate the elements of the joint error covariance matrix by

$$
S_{i,j} = \frac{1}{n - K} (U_{i,i}^P)' U_{i,j}^P,
$$
The estimated covariance matrix is then used to construct the second-stage GLS estimator.

In the above example, there are only two equations and $n$ observations for each equation. For the first-stage error covariance matrix to be well estimated, $n$ must be substantially larger than the number of equations.