Economics 361
Seemingly Unrelated Regressions (SUR)

Jun Ishii *
Department of Economics
Amherst College

Fall 2023

1 Overview

Thus far, we have focused on what some refer to as **single regression equation** models. Here, "regression equation" refers to the equation that can be derived from the Linearity Condition

$$E[Y|X] \;=\; X\beta \qquad \Longrightarrow \qquad Y \;=\; X\beta + \epsilon$$

where $\epsilon \equiv Y - X\beta$ by convention.

But some empirical work involves examining more than one Linearity Condition and, thus, more than one regression equation.

$$\begin{aligned}
E[Y_1|X_1] \;&=\; X_1\beta_1 & \Longrightarrow \qquad Y_1 \;&=\; X_1\beta_1 + \epsilon_1 \\
E[Y_2|X_2] \;&=\; X_2\beta_2 & \Longrightarrow \qquad Y_2 \;&=\; X_2\beta_2 + \epsilon_2 \\
&\;\;\vdots & \vdots \qquad\qquad &\;\;\vdots \\
E[Y_M|X_M] \;&=\; X_M\beta_M & \Longrightarrow \qquad Y_M \;&=\; X_M\beta_M + \epsilon_M
\end{aligned}$$

Each Y_j is a $(N_j \times 1)$ vector and each X_j a $(N_j \times k_j)$ full rank matrix. So equation j has N_j observations and k_j explanatory variables.

The above implies that the Linearity and Full Rank conditions are satisfied for each equation. So $\beta_1 \dots \beta_M$ could each be estimated by simply applying OLS to each $\{Y_j, X_j\}$ (sub)sample. This is referred to as **equation-by-equation OLS**. The resulting linear estimator of β_j would be unbiased.

Furthermore, if each equation *individually* satisfies the spherical errors condition, $V[Y_j|X_j] = \sigma_j^2 I_N$, then the estimates from equation-by-equation OLS will be the "best" linear unbiased estimates of β_j, where linear refers to linear in Y_j.

*Office: Converse Hall 315 Phone: (413) 542-2901 E-mail: jishii@amherst.edu

But it may still be possible to do *better* than the estimates from equation-by-equation OLS, *even if each equation individually satisfies the Gauss Markov assumptions.* This is because equation-by-equation OLS only uses the information available *within each equation* to estimate the parameters of that equation. It does not consider information that may be available *across equations.*

In general, there are two forms of "cross equation" information

1. Restrictions in the parameter values across regression equations
2. Correlation between observations across equation (sub)samples

The first refers to the idea that the values of parameters in one equation may necessarily restrict the value of the parameters in other equations. The extreme version of this restriction is when parameters from different equations must equal each other.

The second refers to the idea that while there may be no correlation in observations within each (sub)sample – so each equation may still satisfy Spherical Errors – there may be correlation in observations across (sub)samples. This correlation may be exploited in a manner similar to our earlier Generalized Least Squares (GLS) discussion.

To explore these issues, it is useful first to derive the "stacked" regression model. This is a single equation regression model that, effectively, estimates all of the regression equations in a joint fashion.

2 Stacked Regression

Let each equation be associated with the size N_j (sub)sample $\{Y_j, X_j\}$. X_j consists of k_j variables and is full rank. Moreover, the following Linearity Condition holds for each (sub)sample: $E[Y_j|X_j] = X_j\beta_j$. Equation-by-equation OLS would result in $b_j^{ols} = (X_j'X_j)^{-1}X_j'Y_j$. Given the above Linearity Condition, $E[b_j^{ols}|X_j] = \beta_j$.

These same estimates could have been obtained if we applied OLS, once, to the appropriate "stacked" vectors and matrices. Consider the following example involving 3 equations ($M = 3$):

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} \qquad X = \begin{pmatrix} X_1 & O_{N_1,k_2} & O_{N_1,k_3} \\ O_{N_2,k_1} & X_2 & O_{N_2,k_3} \\ O_{N_3,k_1} & O_{N_3,k_2} & X_3 \end{pmatrix}$$

where $O_{N,k}$ refers to a $(N \times k)$ matrix consisting entirely of zeroes.

Y is a $([N_1 + N_2 + N_3] \times 1)$ vector consisting of the Y_j vectors "stacked" on top of each other. X is a $([N_1 + N_2 + N_3] \times [k_1 + k_2 + k_3])$ matrix consisting of the X_j matrices arranged in a "diagonal" fashion, with zero valued elements in the "off-diagonal" spots.

Applying OLS on the size $[N_1 + N_2 + N_3]$ sample $\{Y, X\}$ results in

$$b^{stack} = (X'X)^{-1}X'Y = \begin{pmatrix} (X_1'X_1)^{-1}X_1'Y_1 \\ (X_2'X_2)^{-1}X_2'Y_2 \\ (X_3'X_3)^{-1}X_3'Y_3 \end{pmatrix} = \begin{pmatrix} b_1^{ols} \\ b_2^{ols} \\ b_3^{ols} \end{pmatrix}$$

The equality with the equation-by-equation OLS estimates takes advantage of the "Submatrix of Inverse Theorem" (Goldberger Chapter 17.5, pp.191-192).[1] The result – b^{stack} coinciding with the equation-by-equation OLS estimates – holds for the general M equation case.

$\text{Var}[b^{stack}|X]$ depends on the properties of $\text{Var}[Y_j|X_j]$. If each regression equation satisfies the Spherical Errors condition and the observations are uncorrelated across equations, then

$$\text{Var}[Y|X] = \Sigma = \begin{pmatrix} \sigma_1^2 I_{N_1} & O_{N_1,N_2} & O_{N_1,N_3} \\ O_{N_2,N_1} & \sigma_2^2 I_{N_2} & O_{N_2,N_3} \\ O_{N_3,N_1} & O_{N_3,N_2} & \sigma_3^2 I_{N_3} \end{pmatrix}$$

Note that this implies that the stacked regression may violate homoskedasticity and that the stacked regression may be better estimated using weighted least squares (WLS). But, interestingly, in this particular case, WLS results in exactly the same estimates as OLS. Dividing each of the (sub)samples by σ_j does not alter the results.[2]

The above still implies that $\text{Var}(b^{stack}|X)$ is not generally $\sigma^2(X'X)^{-1}$. Using the "Submatrix of Inverse Theorem" again, we can show that

$$\text{Var}[b^{stack}|X] = (X'X)^{-1}X'\Sigma X(X'X)^{-1} = \begin{pmatrix} \sigma_1^2(X_1'X_1)^{-1} & O_{k_1,k_2} & O_{k_1,k_3} \\ O_{k_2,k_1} & \sigma_2^2(X_2'X_2)^{-1} & O_{k_2,k_3} \\ O_{k_3,k_1} & O_{k_3,k_2} & \sigma_3^2(X_3'X_3)^{-1} \end{pmatrix}$$

Intuition for this coincidence between the above stacked regression and equation-by-equation OLS? The zero matrices (O) that buffer the X_j matrices indicate that one (sub)sample provides no information about the other (sub)samples. The (sub)samples share neither observations nor variables.

3 Seemingly Unrelated Regression

The above system of regression equations – where each equation, individually, satisfies the Gauss-Markov assumptions, is referred to as a **Seemingly Unrelated Regression (SUR)** model. The "seemingly unrelated" refers to the idea that, at first glance, the equations seem to be unrelated to each other. But a closer examination may reveal relationships across the equations that may be exploited. The two main forms of such "hidden" relationship are cross equation parameter restrictions and cross equation correlated observations.

[1] Goldberger Chapter 30.2 explicitly derives the result for the two equation case.

[2] This result is basically the same as the following: consider a single equation regression model where we divide every variable for every observation by the same factor, γ. So $\tilde{Y} = \gamma Y$ and $\tilde{X} = \gamma X$. $\tilde{b} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{Y} = (\gamma^2 X'X)^{-1}\gamma^2 X'Y = (X'X)^{-1}X'Y = b^{ols}$

3.1 Cross Equation Parameter Restrictions

Now, suppose that there are parameter restrictions across equations. Consider the case where $\beta_1 = \beta_2 \neq \beta_3$. For simplicity, assume $k_1 = k_2$. In other words, equations 1 and 2 share the same set of parameters but not equation 3.

But this implies that we are estimating a redundant set of parameters. Instead of estimating $k_1 + k_2 + k_3$ parameters, we should only be estimating $k_1 + k_3$ parameters. The k_2 parameters in equation 2 are the same as the k_1 parameters in equation 1.

This suggests that we can use the following stacked regression model

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} \qquad \tilde{X} = \begin{pmatrix} X_1 & O_{N_1,k_3} \\ X_2 & O_{N_2,k_3} \\ O_{N_3,k_1} & X_3 \end{pmatrix} \qquad \tilde{\beta} = \begin{pmatrix} \beta_1 \\ \beta_3 \end{pmatrix}$$

Note that $E[Y|\tilde{X}] = \tilde{X}\tilde{\beta}$

An alternative way to think about this stacked regression is as follows: instead of three equations, consider this two equations with the first two equations merged together. Instead of one regression involving Y_1 and X_1 and another involving Y_1 and X_2, merge the two into a single regression involving $\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$ and $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ – which is essentially the regression model involving the $N_1 + N_2$ combined (sub)sample.

Recall that, with the Gauss-Markov assumptions, OLS is a moment-based estimator of $E[Y|X]$. $E[Y_1|X_1] = X_1\beta_1$ is effectively the same as $E[Y_2|X_2] = X_2\beta_2$ – what differs are only the "names" of the dependent and conditioning variables. So, from a moment-based perspective, the two are equivalent and do not need to be separated.

The above stacked regression model, incorporating the merger between equations 1 and 2, now estimates $k_1 + k_3$ parameters using the $N_1 + N_2 + N_3$ observations. More specifically, $N_1 + N_2$ observations are used to estimate β_1 and N_3 observations to estimate β_3. The estimate of β_1 will improve as we are using more information ($N_1 + N_2 > N_1$).

More generally, linear restrictions in the parameter values across equations imply some redundance in equation-by-equation OLS. Goldberger Chapter 30.6 discusses this more general case. Note that even if each regression equation satisfies Spherical Errors, the stacked regression equation may not (e.g. difference in conditional variance across equations). This is why Goldberger discusses estimating the stacked model using GLS.

Let X be the original stacked "X" (for equation-by-equation OLS) and $\tilde{X}$ the stacked "X" that incorporates the linear cross equation parameter restrictions. $\tilde{X}$ can be decomposed as XT. So the GLS estimator on the stacked Y and stacked $\tilde{X}$ with conditional variance/covariance matrix Σ

can be expressed as

$$
\begin{aligned}
\tilde{b}^{gls} &= (\tilde{X}\Sigma^{-1}\tilde{X})^{-1}\tilde{X}\Sigma^{-1}Y &= (T'X'\Sigma^{-1}XT)^{-1}T'X'\Sigma^{-1}Y \\
\mathrm{Var}(\tilde{b}^{gls}|\tilde{X}) &= (\tilde{X}'\Sigma^{-1}\tilde{X})^{-1} &= (T'X'\Sigma^{-1}XT)^{-1}
\end{aligned}
$$

In the above example where $\beta_1 = \beta_2$

$$
T = \begin{pmatrix} I_{k_1} & O_{k_1,k_3} \\ I_{k_2} & O_{k_2,k_3} \\ O_{k_3,k_1} & I_{k_3} \end{pmatrix}
$$

Note that the Gauss-Markov theorem is *not* invalidated with this SUR model result. Equation-by-equation OLS (via stacked regression) is still "BLUE" ... but BLUE for β (which includes the redundant parameters). The above stacked regression OLS is "BLUE" for $\tilde{\beta}$ (which excludes the redundant parameters).

3.2 Cross Equation Correlation

Even if there are no cross equation parameter restrictions, there may be information across equations if some elements in Y_j are correlated with some elements in Y_k (where j and k index two different regression equations / (sub)samples. This suggests that the dependent variables that are being explored in regression equations j and k are outcomes from related data generating processes.

Using the $M = 3$ equation case, such cross-equation correlation implies that

$$
\mathrm{Var}[Y|X] = \Sigma = \begin{pmatrix} \sigma_1^2 I_{N_1} & \Sigma_{1,2} & \Sigma_{1,3} \\ \Sigma_{2,1} & \sigma_2^2 I_{N_2} & \Sigma_{2,3} \\ \Sigma_{3,1} & \Sigma_{3,2} & \sigma_3^2 I_{N_3} \end{pmatrix}
$$

where $\Sigma_{j,k}$ is a $(N_j \times N_k)$ matrix that may contain some non-zero elements.

Recall that equation-by-equation OLS is equivalent to applying OLS (and also WLS) to $\{Y, X\}$. The above suggests that even if the X is the same (no parameter restrictions), we can obtain a more precise estimate of the stacked β by applying GLS to the same $\{Y, X\}$.[3]

$$
\begin{aligned}
b^{stack,gls} &= (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y \\
\mathrm{Var}(b^{stack,gls}|X) &= (X'\Sigma^{-1}X)^{-1}
\end{aligned}
$$

Even without cross equation parameter restrictions, estimating the stacked regression with GLS may provide more precise estimates than equation-by-equation OLS. The intuition of this result is the same as the general intuition for why GLS improves upon OLS, extended to the cross equation case.

[3]Note: GLS here differs from WLS as this GLS incorporates the non-zero covariances