University of Illinois | Department of Economics
Fall 2016 | Roger Koenker

Economics 536
Lecture 14

Estimation of Systems of Simultaneous Equation Model

In this brief lecture we try to introduce estimation methods for simultaneous equation models which apply to the entire system rather than treating the models one equation at a time as we have done thus far with two stage least squares.

Consider the model,

$$\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = \begin{bmatrix} Z_1 & 0 & \cdots & \cdots & 0 \\ 0 & & \ddots & & \\ \vdots & & & \ddots & \\ 0 & & & & Z_m \end{bmatrix} \begin{pmatrix} \delta_1 \\ \vdots \\ \delta_m \end{pmatrix} + \begin{pmatrix} u_1 \\ \vdots \\ n_m \end{pmatrix}$$

which we will write simply write

$$y = Z\delta + u$$

We will assume as in the SUR model that

$$Euu^\top = \Omega \otimes I_n$$

so there is contemporaneous correlation across equation errors, but each equation has a classical, spherical error structure. The model differs from SUR in that each Z_i has the structure

$$Z_i = [Y_i \vdots X_i]$$

with the possible inclusion of endogenous variables Y_i in each equation. This obviously necessitates some form of instrumental variables estimation method in addition to the problem of dealing with the correlation introduced with Ω.

To motivate the simultaneous treatment of both problems, let's consider how to deal with them separately. The SUR solution for the Ω problem introduced a weighting matrix $\Omega^{-1} \otimes I$ so if Z were orthogonal to u we could use

$$\hat{\delta}_{SUR} = (Z^\top(\Omega^{-1} \otimes I)Z)^{-1} Z^\top(\Omega^{-1} \otimes I)y.$$

On the other hand, suppose $\Omega = \sigma^2 I$ then the 2SLS estimator for the whole system could be written as,

$$\hat{\delta}_{2SLS} = (Z^\top(I \otimes P_X)Z)^{-1} Z^\top(I \otimes P_X)y$$

Note that this is equivalent to doing m separate 2SLS estimations

$$\hat{\delta}^{(i)}_{2SLS} = (Z_i' P_X Z_i)^{-1} Z_i' P_X y_i \quad i = 1, \ldots, m.$$

Can we do both of these steps together? Yes, consider the following estimator,

$$\hat{\delta}_{3SLS} = \text{ argmin}_\delta\{(y - Z\delta)(\Omega^{-1} \otimes P_X)(y - Z\delta)\}.$$

One way to see how this works is to consider the construction of IV"s for the whole system of equations as

$$\begin{aligned} \tilde{Z} &= (\Omega^{-1/2} \otimes I_n)(I_m \otimes P_X)Z \\ &= (\Omega^{-1/2} \otimes I_n)\hat{Z} \end{aligned}$$

where $\hat{Z} = (I_m \otimes P_X)Z$. In effect, this transformation first creates predicted Z's using the instrument set X and then reweights the equations to get the Ω effect. This can be viewed as motivating the name "three stage least squares," since first we run 2SLS and then we run SUR.

Having seen how this works in estimating systems of equations, it is perhaps useful to go back and review how it is connected to the single equation theory. Recall that in the classical single equation setting

$$y = X\beta + u \quad \text{ with } Euu^\top = \Omega$$

the GLS estimator

$$\begin{aligned} \hat{\beta} &= \text{ argmin } \{(y - X\beta)^\top \Omega^{-1}(y - X\beta)\} \\ &= (X^\top \Omega^{-1} X)^{-1} X^\top \Omega^{-1} y \end{aligned}$$

is optimal among linear unbiased estimator for general error distributions, and optimal among unbiased estimators for Gaussian errors. here, the Ω^{-1} reweights the usual orthogonal projection of ordinary least squares to accommodate the nonspherical error structure.

In the case of two stage least squares we have the model

$$y = Z\delta + u \quad \text{with } Euu^\top = \sigma^2 I$$

but $Z \not\perp u$. This is resolved by the estimator,

$$\hat{\delta} = \text{ argmin } (y - Z\delta)^\top P_X (y - Z\delta) = (Z^\top P_X Z)^{-1} Z^\top P_X y$$

where $P_X = X(X^\top X)^{-1}X^\top$ is the projection onto the column space of the full set of available instrumental variables, X. Thus, here P_X plays somewhat the same role as Ω^{-1} in the GLS problem.

This leads naturally to the question what should we do in single equation situations in which we have need of both 2SLS and GLS? Consider the model

$$y = Z\delta + u \quad \text{with } Euu^\top = \Omega$$

and $Z \not\perp u$, but $X \perp u$ as in the 2SLS case. Clearly, the 2SLS estimator is inefficient in this case and it is easy (please verify!) to show that

$$V(\hat{\delta}_{2SLS}) = (Z^\top P_X Z)^{-1} Z^\top P_X \Omega P_X Z (Z^\top P_X Z)^{-1}$$

This is a particular form of "sandwich formula" which we gradually learn to associate with asymptotic covariance matrices which are inefficient. The efficient estimator for this situation is,

$$\check{\delta} = (Z^\top P_X^* Z)^{-1} Z^\top P_X^* y$$

where $P_X^* = X(X^\top \Omega X)^{-1} X^\top$. As a final exercise prove $\mathrm{Var}(\check{\delta}) = (Z^\top P_X^* Z)^{-1}$. Note that P_X^* is *not* a projection matrix so we should regard $\check{\delta}$ as a proper IV estimator, but not a proper 2SLS estimator. Hendry calls it a GIVE estimator, for generalized IV estimator.

Another way to motivate the foregoing results is via "generalized method of moments." Much of econometric theory can be viewed as an application of this old idea, which originated in the minimum χ^2 ideas of Karl Pearson. The general idea is to minimize some form of discrepancy between observed, or empirical moments and population moments. The latter are simply functions that express population counterparts of the observed moments as functions of the unknown parameters of the problem.

For example, in the classical linear regression model we have the moment condition

$$EX^\top u = 0,$$

which in effect says that the observed moments $X^\top y$ can be expected to be close to their population counterparts $X^\top X\beta$. If we equate these quantities, we obtain the OLS estimator $\hat{\beta}$. The classical IV estimator can be similarly motivated. However, a question arises when the number of moment conditions exceeds the number of parameters, since the discrepancy can't be eliminated, only minimized in such cases. How should we go about defining the discrepancy in such cases? Standard theory suggests that the observed moments will be asymptotically normal random variables, so it is natural to consider minimum χ^2 criteria, i.e. minimizing a quadratic form that represents a weighted sum of squares that would have χ^2 behavior under the hypothesis represented by the moment conditions themselves. For the case of 2SLS we may consider a vector of q moment conditions

$$EZ^\top u = EZ^\top (y - X\beta) = 0$$

where $q \geq p$, the dimension of β. To standardize we consider solving

$$(*) \qquad \min_{\beta} u^\top Z V Z^\top u$$

where V is chosen to induce χ^2 behavior of the objective function. This entails choosing V to be the inverse of the covariance matrix of $Z^\top u$. Thus, if $Euu^\top = \sigma^2 I$,

$$V = \sigma^{-2}(Z^\top Z)^{-1}$$

and $(*)$ becomes

$$\min_{\beta} u^\top P_Z u$$

which yields the usual 2SLS estimator. If $Euu^\top = \Omega$, then $V = (Z^\top \Omega Z)^{-1}$ and $(*)$ takes the form

$$\min_{\beta} u^\top P_{Z^*} u$$

and yields the GIVE estimator discussed earlier.