Topic 10: Systems of Equations

Dr. Ani Dasgupta
Mass. Maritime Academy and Boston University

1 SUR (or SURE)

1.1 The Model and Its Estimation

SUR (SURE) stands for 'Seemingly Unrelated Regressions' ('Seemingly Unrelated Regression Equations'). As you will soon see this is really unfortunate terminology; typical equations that are estimated by the SURE technique are often *obviously* related!

The set-up is as follows: we have m equations, each featuring a different left hand side endogenous variable with certain predetermined variables on the right hand side(which may or may not be common)

$$\begin{aligned} \mathbf{y}_1 &= \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}_1 \\ &\vdots \\ \mathbf{y}_m &= \mathbf{X}_m\boldsymbol{\beta}_m + \boldsymbol{\varepsilon}_m \end{aligned}$$

where $\mathbf{y}_j = \begin{pmatrix} y_{1j} \\ \vdots \\ y_{nj} \end{pmatrix}$, $\boldsymbol{\varepsilon}_j = \begin{pmatrix} \varepsilon_{1j} \\ \vdots \\ \varepsilon_{nj} \end{pmatrix}$, $\mathbf{X}_j = \begin{pmatrix} x_{1j1} \cdots\cdots x_{1jk_j} \\ \vdots \\ x_{nj1} \cdots\cdots x_{njk_j} \end{pmatrix}$,

where $n = \#$ of observations and k_j is the number of regressors in the jth equation.

We make the standard orthogonality assumptions $E(\boldsymbol{\varepsilon}_j) = \mathbf{0}$ and $E(\mathbf{X}_j\boldsymbol{\varepsilon}_j) = \mathbf{0}$, whence equation by equation OLS is consistent and a perfectly valid procedure.

However, often, there are reasons to believe that the error terms of the equations are correlated when efficiency gains can be had by implementing GLS-type procedures. Moreover, there may be constraints tying the coefficients of equations. If theory says that the second coefficient of the first equation is the same as the third coeficient of the second equation, which equation's estimate will you use to estimate this paramter?

To the end of creating a unified estimation procedure, let us stack the observations on all equations to create one big equation.

$$\begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_m \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{X}_m \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_m \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \vdots \\ \boldsymbol{\varepsilon}_m \end{bmatrix}$$

We can write this entire model as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad \text{with } Var(\boldsymbol{\varepsilon}) = \boldsymbol{\Omega} \text{ (say)}$$

Now let us explore $\boldsymbol{\Omega}$ which clearly can be written as

$$\boldsymbol{\Omega} = E \begin{bmatrix} \boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_1' & \boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_2' & \cdots & \boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_m' \\ \boldsymbol{\varepsilon}_2\boldsymbol{\varepsilon}_1' & \boldsymbol{\varepsilon}_2\boldsymbol{\varepsilon}_2' & \cdots & \boldsymbol{\varepsilon}_2\boldsymbol{\varepsilon}_m' \\ \vdots & \vdots & & \vdots \\ \boldsymbol{\varepsilon}_n\boldsymbol{\varepsilon}_1' & \boldsymbol{\varepsilon}_n\boldsymbol{\varepsilon}_2' & \cdots & \boldsymbol{\varepsilon}_m\boldsymbol{\varepsilon}_m' \end{bmatrix}$$

Also, note that

$$E(\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}_l') = E \begin{bmatrix} \varepsilon_{1j}\varepsilon_{1l} & \varepsilon_{1j}\varepsilon_{2l} & \cdots & \varepsilon_{1j}\varepsilon_{nl} \\ \vdots & \vdots & & \vdots \\ \varepsilon_{nj}\varepsilon_{1l} & \varepsilon_{nj}\varepsilon_{2l} & \cdots & \varepsilon_{nj}\varepsilon_{nl} \end{bmatrix}$$

If we assume that error terms of equations j and l for two different observational units are independent and the observations are a priori identical (this is just the i.i.d. assumption), then

$$E(\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}_l') = \sigma_{jl}\mathbf{I}_{n\times n}$$

Hence

$$\begin{aligned} \Omega &= \begin{bmatrix} \sigma_{11}\mathbf{I} & \sigma_{12}\mathbf{I} & \cdots & \sigma_{1m}\mathbf{I} \\ \sigma_{21}\mathbf{I} & \sigma_{22}\mathbf{I} & \cdots & \sigma_{2m}\mathbf{I} \\ \vdots & \vdots & & \vdots \\ \sigma_{m1}\mathbf{I} & \sigma_{m2}\mathbf{I} & \cdots & \sigma_{mm}\mathbf{I} \end{bmatrix} \\ &= \boldsymbol{\Sigma} \otimes \mathbf{I} \ \text{ where } \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1m} \\ \vdots & & \\ \sigma_{m1} & \cdots & \sigma_{mm} \end{bmatrix} \end{aligned}$$

and '$\otimes$' represents Kronecker Product.[1]

Since, we have a (grand) linear model where the disturbances are not 'spherical' clearly the OLS estimator (which coincides with equation by equation OLS) as can be readily verified (do it!), cannot be efficient. Now, noting that $(\mathbf{\Sigma} \otimes \mathbf{I})^{-1}$ is $\mathbf{\Sigma}^{-1} \otimes \mathbf{I}$ (verify this!), one can derive the GLS estimator which is $\mathbf{b}_{GLS} = (\mathbf{X}'(\mathbf{\Sigma}^{-1} \otimes \mathbf{I})\mathbf{X})^{-1}(\mathbf{X}'(\mathbf{\Sigma}^{-1} \otimes \mathbf{I})\mathbf{y})$. This is Zellner's SURE estimator.

In practice, we do not know $\mathbf{\Sigma}$ and so will need to estimate it. This is done via the following 2 step procedure.

Step 1. Use equation by equation OLS. Collect residual vector $\mathbf{e}_j'$s Calculate $\widehat{\sigma}_{jl} = \frac{\mathbf{e}_j'\mathbf{e}_l}{n}$ (usually $\frac{\mathbf{e}_j'\mathbf{e}_l}{\sqrt{(n-k_j)(n-k_l)}}$ works better) and hence form $\widehat{\mathbf{\Sigma}}$.

Step 2. Evaluate $\mathbf{b}_{FGLS} = (\mathbf{X}'(\widehat{\mathbf{\Sigma}}^{-1} \otimes \mathbf{I})\mathbf{X})^{-1}(\mathbf{X}'(\widehat{\mathbf{\Sigma}}^{-1} \otimes \mathbf{I})\mathbf{y})$

Note: If you suspect a lack of conditional homoskedasticity for the error term of an equation, then $E(\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}_l')$ may not have the same terms along the diagonal. In particular, it is possible that $E(\varepsilon_{ij}\varepsilon_{il}) = f(\mathbf{z}_i)$, say $\exp(\mathbf{z}_i'\boldsymbol{\alpha}_{jl})$. In this case, estimate $\boldsymbol{\alpha}_{jl}$ & form $E(\widehat{\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}_l'})$ and use this as the $(j-l)$th block of $\widehat{\mathbf{\Omega}}$.

1.2 Restrictions

If you wish to test $H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$, the way to think about the issue is to transform the original (grand) model with non-spherical errors to a standard one with spherical errors, by multiplying $\mathbf{X}, \mathbf{Y}$ and $\boldsymbol{\varepsilon}$ by $\mathbf{\Omega}^{-1/2}$, a trick that we have used earlier in our discussion of GLS. Thereafter, standard formulas apply for the test statistic and the distribution is standard as well.

$$\frac{(\mathbf{R}\mathbf{b}_{FGLS} - \mathbf{r})' \left[\mathbf{R}(\mathbf{X}'\widehat{\mathbf{\Omega}}^{-1}\mathbf{X})^{-1}\mathbf{R}'\right]^{-1} (\mathbf{R}\mathbf{b}_{FGLS} - \mathbf{r})/q}{(e'\widehat{\mathbf{\Omega}}^{-1}e)/(N-k)} \sim F_{q,N-k}$$

where q, of course is the number of restrictions. $N = nm$ and $k = k_1 + \cdots + k_m$.

Now assuming the model 'passes' the test, one will need to estimate the restricted model which can be obtained via the formula we learned in Topic 3.

$$\mathbf{b}_R = \mathbf{b}_{FGLS} + (\mathbf{X}'\widehat{\mathbf{\Omega}}^{-1}\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\widehat{\mathbf{\Omega}}^{-1}\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{r} - \mathbf{R}\mathbf{b}_{FGLS})$$

[1] In STATA, use A # B to create the Kroneeker product of matrices A and B

1.3 An Example: Estimating Parameters of A Demand System

As an example of how SURE systems occur naturally in Economics, consider the estimation of a demand system, i.e. we wish to estimate the Marshallian demand functions of several goods simultaneously. A popular approach is to specify an indirect utility function[2] called the ADDILOG function.

$$V(\mathbf{p}, I) = \sum_{i=1}^{m} \alpha_i \left(\frac{I}{p_i}\right)^{\beta_i} \qquad \alpha_i \geq 0,\ \beta_i \geq -1$$

where $\mathbf{p}$ is the price vector and I is income. The Addilog function allows for certain essential properties required of indirect utility functions, namely that a) it is decreasing in prices, b) it is quasiconvex in prices[3] and c) it is homogeneous of degree zero in prices and income. While the first and the third properties should be obvious, for the proof of the second, please see Hal Varian's text "Microeconomic Analysis".

The demand for various items are given by via the so-called Roy's Identity: $x_j(p, I) = -\frac{\frac{\partial v}{\partial p_j}}{\frac{\partial v}{\partial I}}$ (again, see Varian for a proof). From this the expenditures on these items can be easily derived

$$Z_j(\mathbf{p}, I) = -\frac{\alpha_j \beta_j (I)^{\beta_j} {p_j}^{-\beta_j}}{\sum_{i=1}^{m} \alpha_i \beta_l (I)^{\beta_l - 1} {p_j}^{-\beta_j}}$$

Generally, the expenditures, rather than the demands are treated as endogenous variables since, given the often aggregated nature of consumption categories, it is the expenditure data that are available, not the actual demands.

Hence, taking log and adding an error term

$$\ln Z_j = \text{constant} + A_j + \beta_j \ln\left(I/p_j\right) + u_j \quad \text{where } A_j = \ln(-\alpha_j \beta_j)$$

We can't estimate this directly because the constant is really not a constant; it depends on the observational unit. However, if we consider the difference between pairs of expenditure equations, this washes out.

$$\ln Z_j - \ln Z_l = A_{jl} + \beta_j \ln\left(\frac{I}{p_j}\right) - \beta_l \ln\left(\frac{I}{p_l}\right) + u_j - u_l$$

Note also that if we have m goods, this gives us a system of $(m-1)$ independent equations.Hence,

[2]Indirect utility at a certain price vector and income is the optimized utility the consumer receives having chosen her optimal bundle subject to the budget constraint.

[3]A function f mapping from R^n to R is quasiconvex if for any α, the set $\{x : f(x) \leq \alpha\}$ is convex (such sets are called lower contour sets). For quasiconcave functions, it is the upper contour sets that must be convex.

bringing back i as the first subscript to denote the ith observation, we can write

$$\ln Z_{i1} - \ln Z_{i2} = A_{12} + \beta_1 \ln\left(\frac{I_i}{p_{i1}}\right) - \beta_2 \ln\left(\frac{I_i}{p_{i2}}\right) + \varepsilon_{i2}$$
$$\vdots$$
$$\ln Z_{i1} - \ln Z_{im} = A_{1m} + \beta_1 \ln\left(\frac{I_i}{p_{i1}}\right) - \beta_P \ln\left(\frac{I_i}{p_{im}}\right) + \varepsilon_{im}$$

where ε_{ij} is simply $u_{i1} - u_{ij}$.

However, we now have cross-equation correlations among the disturbance terms! Also note that the coefficient on $\ln(\frac{I_i}{p_{i1}})$ is restricted to be the same for each equation!

2 Simultaneous Equation Systems

2.1 The Model and Identifiability Issues

In a SURE system, the rhs variable in each equation is exogenous (or as some say, 'predetermined'). But this is often not the case in real life situations; our equation system simultaneously determines m endogenous variables and in some of the equations, there may be some endogenous variable on the right hand side. A typical system will have m equations with the $j's$ equation:

$$\beta_{j1}y_{i1} + \cdots \beta_{jm}y_{im} + \gamma_{j1}x_{i1} + \cdots \gamma_{jl}x_{il} = \varepsilon_{ij} \quad i = 1, \cdots n;\ j = 1, \cdots m$$
$$\textit{or}\ \mathbf{B}\mathbf{y_i} + \mathbf{\Gamma}\mathbf{x_i} = \boldsymbol{\varepsilon}_i$$

($\mathbf{B}$ is $m \times m$; $\mathbf{\Gamma}$ is $m \times l$, $y_i, \boldsymbol{\varepsilon}_i$ are $m \times 1$, $\mathbf{x}_i$ is $l \times 1$).

A typical example is the supply-demand system we're seen before.

Not all simultaneous systems are estimable. In order to be estimated, some identifiability conditions are to be satisfied.

Let $\mathbf{A} = [\mathbf{B}\ \mathbf{\Gamma}]$ and let $\boldsymbol{\alpha}'_j$ denote the $j'th$ row of $\mathbf{A}$. Typically, there will be restrictions on $\boldsymbol{\alpha}'_j$ (mostly exclusion restrictions). Suppose they can be represented as $\boldsymbol{\alpha}'_j\mathbf{\Phi}_j = 0$ where $\boldsymbol{\alpha}'_j$ is $1 \times (m + l)$ and $\mathbf{\Phi}_j$ is $(m + l) \times q$.

For example, in a system of equations with three endogenous and three exogenous variables, if the first equation does not feature the second endogenous variable and the third exogenous variable, then

$$\mathbf{\Phi}_1 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}$$

Fact: A necessary and sufficient condition for the identifiability of the j-th equation is $rank(\mathbf{A\Phi}_j) = m - 1$.

The rank condition may be a bit cumbrous to check directly. A simpler necessary condition (in the case of exclusionary restrictions) is the so-called order condition:

$$\underset{\#\ restrictions}{\underset{\downarrow}{q}} \quad \geq \quad \underset{\#\ of\ endo\ vars}{\underset{\downarrow}{m}} \quad - 1$$

As an example consider the familiar supply-demand system

$$\begin{aligned} q_i &= \alpha_1 + \alpha_2 p_i + \alpha_3 I_i + u_i && \text{(Demand)} \\ q_i &= \beta_1 + \beta_2 p_i + \beta_3 w_i + v_i && \text{(Supply)} \end{aligned}$$

There are two endogenous and three exogenous variables (including the constant). Thus,

$$\mathbf{A} = \begin{bmatrix} \beta_{11} & \beta_{12} & \gamma_{11} & \gamma_{12} & \gamma_{13} \\ \beta_{21} & \beta_{22} & \gamma_{21} & \gamma_{22} & \gamma_{23} \end{bmatrix}$$

As stated, both equations are identified here using the rank condition. For example, for the first equation, only the third exogenous variable is excluded, so

$$\mathbf{\Phi}_1 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

$$\mathbf{A\Phi_1} = \begin{bmatrix} 0 \\ \gamma_{23} \end{bmatrix}$$

which has rank 1 which is also $m-1$ since $m=2$.

However, if we included w_i in the demand equation, the order condition easily shows that it will fail to be identified (0 is not greater than or equal to 1).

2.2 Estimation

Let the jth (identified) equation's data be written as

$$\mathbf{y}_j = \underset{\text{endo vars on RHS}}{\underset{\downarrow}{\mathbf{Y}_j}} \boldsymbol{\beta}_j + \underset{\text{exo vars on RHS}}{\underset{\downarrow}{\mathbf{x}_j}} \gamma_j + \boldsymbol{\varepsilon}_j$$

In the 2SLS approach, we obtain estimates of the coefficients by running OLS on the following model:

$$P_{\mathbf{X}} y_j = P_{\mathbf{X}} \mathbf{Y}_j \boldsymbol{\beta}_j + \underbrace{P_{\mathbf{X}} \mathbf{X}_j}_{\mathbf{X}_j} \gamma_j + P_{\mathbf{X}} \varepsilon_j$$

i.e. we run regressions of each endo on all exo variables (first stage); obtain predicted values and use the predicted values of the endo variables in the equations (as opposed to actual values) to run another regression to recover the coefficients in each equation (second stage).

The 3SLS method implements GLS or the philosophy of SURE on top of this IV approach.

Let $\mathbf{w}_j = P_{\mathbf{X}} \mathbf{y}_j$, $\mathbf{W}_j = P_{\mathbf{X}} [\mathbf{Y}_j \;\; \mathbf{x}_j]$ $\boldsymbol{\delta}_j = \begin{bmatrix} \boldsymbol{\beta}_j \\ \gamma_j \end{bmatrix}$ and $\mathbf{v}_j = P_{\mathbf{X}} \boldsymbol{\varepsilon}_j$

Now consider the (stacked) system

$$\begin{bmatrix} \mathbf{w}_1 \\ \vdots \\ \mathbf{w}_m \end{bmatrix} = \begin{bmatrix} \mathbf{W}_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{W}_m \end{bmatrix} \begin{bmatrix} \boldsymbol{\delta}_1 \\ \vdots \\ \boldsymbol{\delta}_m \end{bmatrix} + \begin{bmatrix} \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_m \end{bmatrix}$$

The FGLS estimates of $\delta's$ from this are the 3SLS estimates of the original equation.

In STATA, the 3SLS procedure can be implemented either via the 'reg3'. The typical command is something like:

reg3 (supply: q p w) (demand: p q i)

Note that you need a new endo variable on the left hand side for each equation. This command also works for SURE. You will append the option "sure" (after a comma at the end).