9 Seemingly Unrelated Regression & Simultaneous Equations

9.1 Seemingly Unrelated Regression

Seemingly Unrelated Regression (dfn): A set of equations that may be related not because they interact, but because their error terms are related.

Note: Seemingly Unrelated Regression (SURE) is a natural analytical step from last week's single equation setup to this week's set of equations that are directly related to simultaneous equations.

Examples

- Demand functions for a commodity — exogenous shocks affect the demand for both X and Y. The equations are estimated as a "set" to increase efficiency.
- The behavior of two political parties is determined by certain exogenous variables and the connection between them lies only in the shocks that both parties experience.

9.1.1 Basic Setup

SURE is based on the idea of a ***set*** of equations of the form:

$$\mathsf{y} = \mathsf{X}\beta + \epsilon$$

where the disturbances are correlated across equations (e.g. countries, parties).

Various methods have been employed to estimate such a set of equations. All attempt to exploit the information in the correlated errors, either contemporaneously or autoregressively, in order to achieve greater efficiency in the estimates.

OLS will yield unbiased & consistent estimates for each separate equation. However, because the approach ignores the correlation of the disturbances the estimates will ***not*** be efficient.

9.1.2 Estimation via Generalized Least Squares

This approach takes into account the variability across equations and will yield BLU (best, linear, unbiased) estimates.

The system has the following general form (see Kmenta, p.636):

$$\begin{aligned} y_{1t} &= \beta_{11}x_{1t,1} + \beta_{12}x_{1t,2} + \cdots + \beta_{1k_1}x_{1t,k_1} + \epsilon_{1t} \\ &\vdots \\ y_{mt} &= \beta_{m1}x_{mt,1} + \beta_{m2}x_{mt,2} + \cdots + \beta_{mk_m}x_{mt,k_m} + \epsilon_{mt} \end{aligned} \tag{1}$$

for $t = 1, 2, 3, \ldots, T$

The system can be expressed more parsimoniously using matrix notation:

$$\begin{aligned} \mathsf{y_1} &= \mathsf{X_1}\beta_\mathsf{1} + \epsilon_\mathsf{1} \\ &\vdots \\ \mathsf{y_m} &= \mathsf{X_m}\beta_\mathsf{m} + \epsilon_\mathsf{m} \end{aligned} \tag{2}$$

Each equation is expected to satisfy the assumptions of the Classical Linear Regression model (CLRM). However, if the regression disturbances in the different equations are mutually correlated then we have that:

$$E[\epsilon_m, \epsilon_p'] = \sigma_{mp}\mathsf{I}_T$$

for $m, p = 1, 2, 3, \ldots, m$

Therefore, σ_{mp} is the covariance of the disturbances of the m^{th} and p^{th} equations, which is assumed to be constant over all observations, and is the only link between the m^{th} and p^{th} equations.

Equation (2) can be re-written as:

$$\begin{vmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{vmatrix} = \begin{vmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & X_m \end{vmatrix} \begin{vmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{vmatrix} + \begin{vmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_m \end{vmatrix}$$

Where each,
y_i is a $T \times 1$ vector of sample values on the dependent variable(s)
X_i is a $T \times k_i$ matrix of sample values on the k_i independent variables
β_i is a $k_i \times 1$ vector of coefficients

Assume that ϵ_i is normally distributed with:

$E[\epsilon_i] = 0$
$E[\epsilon_i \epsilon_i'] = \sigma_{ii}\mathsf{I}_T$

Then, the variance-covariance matrix ($\Omega = E[\epsilon\epsilon']$) can be defined as:

$$\Omega = \begin{vmatrix} E[\epsilon_1\epsilon_1'] & E[\epsilon_1\epsilon_2'] & \cdots & E[\epsilon_1\epsilon_m'] \\ E[\epsilon_2\epsilon_1'] & E[\epsilon_2\epsilon_2'] & & E[\epsilon_2\epsilon_m'] \\ \vdots & & \ddots & \vdots \\ E[\epsilon_m\epsilon_1'] & E[\epsilon_m\epsilon_2'] & \cdots & E[\epsilon_m\epsilon_m'] \end{vmatrix} \begin{vmatrix} \sigma_{11}\mathsf{I}_T & \sigma_{12}\mathsf{I}_T & \cdots & \sigma_{1m}\mathsf{I}_T \\ \sigma_{21}\mathsf{I}_T & \sigma_{22}\mathsf{I}_T & & \sigma_{2m}\mathsf{I}_T \\ \vdots & & \ddots & \vdots \\ \sigma_{m1}\mathsf{I}_T & \sigma_{m2}\mathsf{I}_T & \cdots & \sigma_{mm}\mathsf{I}_T \end{vmatrix}$$

Where the relevant information is contained in:

$$E[\epsilon_m\epsilon_p'] = \sigma_{mp}\mathsf{I}_T$$

σ_{mp} is the covariance of disturbances between the m^{th} and p^{th} equations, contemporaneously (which is assumed to be constant across all equations).

Estimating via GLS yields:

$$\tilde{\beta} = (\mathsf{X}'\Omega^{-1}\mathsf{X})^{-1}(\mathsf{X}'\Omega^{-1}\mathsf{y})$$

$$VCV = (\tilde{\beta} - \beta)(\tilde{\beta} - \beta)^{-1} = (\mathsf{X}'\Omega^{-1}\mathsf{X})^{-1}$$

The inclusion of Ω^{-1} improves the efficiency of the estimates, especially when the disturbances are highly correlated, but the independent variables are not.

Caveats: The GLS estimator will equal the OLS estimator when:

1. $\sigma_{mp} = 0$
2. The equations have exactly the same values on all the independent variables (i.e. $x_m = x_p \; \forall \; m,p$)

In these instances the regressions are not "seemingly," but actually, unrelated.

Calculation — Proceeds in 2-stages (see Kmenta, pp.643-5).

1. Estimate via OLS, obtain residuals.
2. Estimate $\hat{\Omega}$ (must estimate because the variance-covariance of regression disturbances will generally be unknown).

$$\hat{\Omega} = \begin{vmatrix} s_{11}\mathsf{I}_T & s_{12}\mathsf{I}_T & \cdots & s_{1m}\mathsf{I}_T \\ s_{21}\mathsf{I}_T & s_{22}\mathsf{I}_T & & s_{2m}\mathsf{I}_T \\ \vdots & & \ddots & \vdots \\ s_{m1}\mathsf{I}_T & s_{m2}\mathsf{I}_T & \cdots & s_{mm}\mathsf{I}_T \end{vmatrix}$$

9.1.3 AR Disturbances

Returning to the basic model:

$$\begin{vmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{vmatrix} = \begin{vmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & X_m \end{vmatrix} \begin{vmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{vmatrix} + \begin{vmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_m \end{vmatrix}$$

Where we now have:

$$\begin{vmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \vdots \\ \epsilon_{mt} \end{vmatrix} = \begin{vmatrix} \rho_{11} & \rho_{12} & \cdots & \rho_{1m} \\ \rho_{21} & \rho_{22} & & \rho_{2m} \\ \vdots & & \ddots & \vdots \\ \rho_{m1} & \rho_{m2} & \cdots & \rho_{mm} \end{vmatrix} \begin{vmatrix} \epsilon_{1,t-1} \\ \epsilon_{2,t-1} \\ \vdots \\ \epsilon_{m,t-1} \end{vmatrix} + \begin{vmatrix} \nu_{1t} \\ \nu_{2t} \\ \vdots \\ \nu_{mt} \end{vmatrix} \tag{3}$$

Which may be re-written as:

$$\epsilon_{\mathsf{t}} = \mathsf{R}\epsilon_{\mathsf{t-1}} + \nu_{\mathsf{t}} \tag{4}$$

where, ν_t are IID random vectors with mean zero (i.e. $E[\nu_t] = 0$)

And, the covariance matrix is given by:

$$E[\nu_t \nu_t'] = \Sigma = \begin{vmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1m} \\ \sigma_{21} & \sigma_{22} & & \sigma_{2m} \\ \vdots & & \ddots & \vdots \\ \sigma_{m1} & \sigma_{m2} & \cdots & \sigma_{mm} \end{vmatrix} \tag{5}$$

That is,

$E[\nu_{it}] = 0$
$E[\nu_{it}\nu_{js}] = \sigma_{ij}$ for $t = s$ and 0 otherwise.

The AR(1) process is represented in Equations (3) and (4) and extends the single equation case by allowing the current disturbance for a given equation to depend on previous periods' disturbances in all equations. This is known as ***Vector Autoregressive Errors***.

When R is Diagonal

Equation (4) may e re-written as:

$$\epsilon_{it} = \rho_{ii}\epsilon_{i,t-1} + \nu_{it}$$

So, the value of the the disturbance ϵ_{it} does not depend on the lagged disturbance in the other equations.

The goal remains the same — estimation using GLS.

$$\hat{\beta} = (\mathsf{X}'\Omega^{-1}\mathsf{X})^{-1}(\mathsf{X}'\Omega^{-1}\mathsf{y})$$

where,

$$E[\epsilon_i\epsilon_j'] = \Omega_{ij} = \frac{\sigma_{ij}}{1-\rho_i\rho_j}\begin{vmatrix} 1 & \rho_j & \cdots & \rho_j^{T-1} \\ \rho_i & 1 & & \rho_j^{T-2} \\ \vdots & & \ddots & \vdots \\ \rho_i^{T-1} & \rho_i^{T-2} & \cdots & 1 \end{vmatrix}$$

There are certain computational costs associated with this approach. You need to transform each variable using standard procedures. Basically performing first differences on all but the first observation — the transformation of the first observation is slightly more complex requiring a 5-step procedure (see Judge et al., pp.488-9)

Higher order AR processes can also be modelled, however, the process is more complex (see Judge et al., Section 12.3.2)

General Observations

1. Finite Sample Properties
   (a) There is some evidence that asymptotics hold in finite samples.
   (b) The omission of the first observation doesn't appear to matter a great deal.
2. There exists a test of $\mathsf{R} = 0$
   (a) Assuming a diagonal R appears to be the best approach even if you know that it is not.
3. General Recommendations
   (a) If there are a sufficient number of observations it is always best to allow for contemporaneous correlation of the disturbances.
   (b) Some form of autocorrelation process may also be included. However, if the sample has less than 20 observations it will be difficult to get accurate parameter estimates (even if you are sure an AR process exists).

Note: For a good example of SURE in political science, see Franklin's 1991 APSR article (Maximum Likelihood Heteroscedastic SURE).

9.2 Simultaneous Equations

Simultaneous Equations (dfn): The dependent variable is determined by the simultaneous interaction of several relationships.

If all of the relationships involved are needed for determining the value of at least one of the endogenous variables, then we have a ***simultaneous equation(s) problem***.

Predetermined Variables

1. Exogenous Variables: Variables that are completely determined outside of the system of equations

2. Lagged Endogenous Variables: Variables that represent *past* values of the endogenous variables in the system.

Endogenous Variables: Variables which are to be explained by the model and which are stochastic.

The unique feature simultaneous equations is the fact that the dependent variable in one equation may be an explanatory variable in another. The problem then becomes that the dependent variable is now stochastic and may be correlated with the disturbances in that equation.

There are different forms of the basic model (see Kmenta, pp.655-7). Variants include:

- Block Diagonal
- Recursive
- Block Triangular

The form of a particular model depends on the positions of the zero elements in the B matrix. This determines which endogenous variables appear (or do not appear) in the different structural equations.

9.2.1 Estimation

Pitfalls of OLS estimation

Example: Keysian Model of Consumption

$$C_t = \beta_0 + \beta_1 \mathsf{y_t} + u_t \tag{6}$$

$$\mathsf{y_t} = C_t + I_t (= S_t) \tag{7}$$

Where;

C = Consumption
y = Income
I = Investment = S = Savings
t = Time

u = Stochastic Disturbance
β_1 = Marginal Propensity to Consume (MPC)

Under the assumptions of the Classical Linear Regression Model (CLRM) it can be shown that $\mathsf{y_t}$ and u_t are correlated and thus $\hat{\beta}_1$ is inconsistent.

Assume:

$E[u_t] = 0$
$E[u_t^2] = \sigma^2$
$E[u_t, u_{t+j}] = 0, \forall j \neq 0$
$Cov[I_t, u_t] = 0$

Substituting (6) into (7) we have:

$$\mathsf{y_t} = \beta_0 + \beta_1 \mathsf{y_t} + u_t + I_t \tag{8-a}$$

OR,

$$\mathsf{y_t} = \frac{\beta_0}{1-\beta_1} + \frac{1}{1-\beta_1} I_t + \frac{1}{1-\beta_1} u_t \tag{8-b}$$

So,

$$E[\mathsf{y_t}] = \frac{\beta_0}{1-\beta_1} + \frac{1}{1-\beta_1} I_t \tag{9}$$

Subtracting (9) from (8-b) yields:

$$\mathsf{y_t} - E[\mathsf{y_t}] = \frac{u_t}{1-\beta_1} \tag{10}$$

<u>Note</u>:

$u_t - E[u_t] = u_t$

Therefore,

$$\begin{aligned} Cov[\mathsf{y_t}, u_t] &= E\{(\mathsf{y_t} - E[\mathsf{y_t}])(u_t - E[u_t])\} \\ &= \frac{E[u_t^2]}{1-\beta_1} \\ &= \frac{\sigma^2}{1-\beta_1} \end{aligned} \tag{11}$$

Since σ^2 is positive (by assumption), the covariance between $\mathsf{y_t}$ and u_t will be different from 0.

$Cov > 0$ for $0 < \beta_1 < 1$
$Cov < 0$ for $\beta_1 > 1$

Note: A value of $\beta_1 > 1$ does make much economic sense, therefore, $Cov(\mathsf{y_t}, u_t)$ is expected to be positive.

The correlation of $\mathsf{y_t}$ *and* u_t *violates the assumptions of the CLRM and implies that the OLS estimates will be inconsistent.*

Structural Equations Approach

From a system of structural equations we can derive reduced form equations and coefficients.

Reduced Form Equations (dfn): Expressing the endogenous variables (only) in terms of predetermined variables and stochastic disturbances.

Returning to the previous example:

$$C_t = \beta_0 + \beta_1 \mathsf{y_t} + u_t \tag{12}$$

$$\mathsf{y_t} = C_t + I_t (= S_t) \tag{13}$$

C_t and $\mathsf{y_t}$ are endogenous and I_t is exogenous.

Substituting (12) into (13) yields the reduced form equation:

$$\mathsf{y_t} = \pi_0 + \pi_1 I_t + \omega_t \tag{14}$$

Equation (14) is a reduced form equation because $\mathsf{y_t}$ is written purely in terms of predetermined and exogenous variables.

The reduced form coefficients are given by:

$$\pi_0 = \frac{\beta_0}{1-\beta_1}$$

$$\pi_1 = \frac{1}{1-\beta_1}$$

$$\omega_t = \frac{u_t}{1-\beta_1}$$

OLS can now be used to estimate (14) because I_t (the predetermined or exogenous variable) is assumed to be uncorrelated with the disturbances.

Indirect Least Squares (IDL): Will yield the structural coefficients from the reduced form coefficients

Key Question: Can the structural coefficients *always* be retrieved from the reduced form coefficients?

Answer: No

9.2.2 The Identification Problem

Unidentified or Under-Identified: It is impossible to get structural coefficients back from the reduced form coefficients.

Exactly Identified: Estimation yields unique structural coefficients.

Over-Identified: More than one unique numerical value is produced for some of the structural coefficients. This occurs because a particular reduced form equation may be compatible with more than one set of structural equations.

Examples:

- Under-Identified

  Demand Equation: $Q_t^d = \alpha_0 + \alpha_1 P_t + u_{1t}, \quad \alpha_1 < 0$

  Supply Equation: $Q_t^s = \beta_0 + \beta_1 P_t + u_{2t}, \quad \beta_1 > 0$

  Equilibrium Condition: $Q_t^d = Q_t^s$

  Re-arranging yields:

  $$\alpha_0 + \alpha_1 P_t + u_{1t} = \beta_0 + \beta_1 P_t + u_{2t} \tag{15}$$

  The goal is to solve:

  $$P_t = \pi_0 + \nu_t \tag{16}$$

  where,

  $$\pi_0 = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}$$
  $$\nu_t = \frac{u_{2t} - u_{1t}}{\alpha_1 - \beta_1}$$

  Now, substitute (16) into either the Demand Equation or the Supply Equation and solve for the equilibrium quantity:

  $$Q_t = \pi_1 + \omega_t \tag{17}$$

  where,

  $$\pi_1 = \frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1}$$
  $$\omega_t = \frac{\alpha_1 u_{2t} - \beta_1 u_{1t}}{\alpha_1 - \beta_1}$$

  Equations (16) and (17) are the reduced form equations, with two reduced form coefficients. There are, however, four structural coefficients, α_0, α_1, β_0, β_1. Therefore, the structural coefficients cannot be uniquely determined.

- Over-Identified

$$\text{Demand Function:} \qquad Q_t^d = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + \alpha_3 R_t + u_{1t}$$

$$\text{Supply Equation:} \qquad Q_t^s = \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + u_{2t}$$

where all the variables are as previously defined, and:

$I =$ Income

$R =$ Wealth

Equating the Supply and Demand Functions yields the equilibrium price and quantities:

$$P_t = \pi_0 + \pi_1 I_t + \pi_2 R_t + \pi_3 P_{t-1} + \nu_t \tag{18}$$

$$Q_t = \pi_4 + \pi_5 I_t + \pi_6 R_t + \pi_7 P_{t-1} + \omega_t \tag{19}$$

where,

$$\pi_0 = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} \qquad \pi_1 = \frac{-\alpha_2}{\alpha_1 - \beta_1}$$

$$\pi_2 = \frac{-\alpha_3}{\alpha_1 - \beta_1} \qquad \pi_3 = \frac{\beta_2}{\alpha_1 - \beta_1}$$

$$\pi_4 = \frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1} \qquad \pi_5 = \frac{-\alpha_2\beta_1}{\alpha_1 - \beta_1}$$

$$\pi_6 = \frac{-\alpha_3\beta_1}{\alpha_1 - \beta_1} \qquad \pi_7 = \frac{\alpha_1\beta_2}{\alpha_1 - \beta_1}$$

$$\nu_t = \frac{\alpha_1 u_{2t} - \beta_1 u_{1t}}{\alpha_1 - \beta_1} \qquad \omega_t = \frac{u_{2t} - u_{1t}}{\alpha_1 - \beta_1}$$

Unique estimation of the structural coefficients is not possible.

Notice:

$\beta_1 = \frac{\pi_6}{\pi_2}$ or,

$\beta_1 = \frac{\pi_5}{\pi_1}$

There exist two estimates of the price coefficient in the supply equation. Since β_1 cannot be uniquely determined, neither can any of the other coefficients. The reason is simply that β_1 appears in the denominator of *all* the other reduced form coefficients.

There is a straightforward (albeit time-consuming) process for determining whether or not a structural equation is identified.

Order Condition

"A necessary condition for a structural equation to be identified is that the number of predetermined variables excluded from a given equation is at least as large as the number of endogenous variables included, *minus one*."

This is a necessary, but not sufficient condition. It is not sufficient because the predetermined variables excluded from an equation, but not the model, may not be independent. There is not a one-to-one correspondence between reduced from and structural coefficients.

A necessary *and* sufficient condition is the

Rank Order Condition

For a model with M equations in M endogenous variables, a structural equation is identified *if and only if* at least one non-zero determinant of order $(M-1)(M-1)$ can be constructed from the coefficients of the variables (endogenous and predetermined) excluded from that particular equation, but included in the other equations of the model.

<u>Rank of a Matrix</u> (dfn): The rank of a matrix is given by the largest number of linearly independent rows or columns

9.2.3 Simultaneous Equation Methods

- If the equation is un-identified then you cannot get consistent estimates
- There are two approaches if the equation is exactly identified or over-identified.
  - Single Equation Methods
  - Systems Methods

Single Equation Methods

(a) Ordinary Least Squares — Recursive Models
These models deal with uni-directional cause & effect among the endogenous variables.

• The first endogenous variable is determined by only the exogenous variables.
• The second endogenous variable is determined by the exogenous variables and the 1^{st} endogenous variable
• Etc. ...

(b) Indirect Least Squares — Use when equations are exactly identified.
The approach is to get reduced form coefficients and work backwards to the structural coefficients.

(c) Two-Stage Least Squares — Designed for equations that are over-identified. The approach is to replace the stochastic endogenous variables by a linear combination of non-stochastic predetermined variables and then use these as the explanatory variables.

The estimation proceeds in two steps:

• Regress the endogenous variables on the predetermined variables in the whole system.
• Re-estimate the initial equation using the fitted values from the first step.

The results from this approach will be consistent *but not efficient*, because it does not take into account the correlation between the disturbances.

Systems Methods

(a) Three-Stage Least Squares —- Parallels SURE via OLS
If you don't take into account the correlation between the disturbances of the different structural equations you are not using all the available information and, therefore, you estimates will not be asymptotically efficient.

This deficiency is overcome, as in SURE, be estimating all the equations simultaneously.

In the case of autocorrelated errors you can apply a modified version of Two-Stage Least Squares (applying GLS to 2SLS). The general estimation approach is as follows:

• Estimate the reduced form equations.
• Use the fitted values of the endogenous variables to get 2SLS estimates.
• Use residuals to estimate the cross-equations variance & covariance (just as in SURE).
•Then get GLS estimates.

The parameter estimates will be more efficient because they take into account the cross-equation variance and covariance.

Example:

$$y_{1t} = \alpha_2 y_{2t} + \nu_{1t} \quad (20)$$

$$y_{2t} = \beta_1 y_{1t} + \beta_3 z_t + \nu_{2t} \quad (21)$$

$$\nu_{1t} = \rho \nu_{1t-1} + \epsilon_{1t} \quad (22)$$

where, ϵ_{1t} and ν_{2t} are not autocorrelated.

The first step is to take the generalized difference of (20). Then solve the reduced form equations to get:

$$y_{1t} = \frac{1}{1-\alpha_2\beta_1}\left(\rho y_{1t-1} - \alpha_2\rho y_{2t-1} + \alpha_2\beta_3 z_t + \nu_{1t}\right) \tag{23}$$

$$y_{2t} = \frac{1}{1-\alpha_2\beta_1}\left(\beta_1\rho y_{1t-1} - \beta_1\alpha_2\rho y_{2t-1} + \alpha_2\beta_3 z_t + \nu_{2t}\right) \tag{24}$$

To estimate (20), get fitted values for y_{2t} by applying OLS to equation (24). Then rewrite the first equation in generalized difference form and substitute in the fitted values for y_{2t}:

$$y_{1t} - \rho y_{1t-1} = \alpha_2(\hat{y}_{2t} - \rho\hat{y}_{2t-1}) + \hat{\epsilon}_{1t} \tag{25}$$

Finally, estimate this modified 2SLS with an autoregressive estimation procedure, such as Cochrane-Orcutt.

(b) Full Information Maximum Likelihood Estimation of Simultaneous Equations

Trade-Offs between Single Equation and Systems Approaches

- Systems Approach provides more efficient estimates, but specification error can be more of a problem (because it tracks through all of the equations).
- Identification is often a problem, although it is usually solved via over-identification.

Next Week... Simulation & Calibration