

Simultaneous Equation Model

Introduction

Consider the following system (1)(2)

$$q_t = \alpha_1 p_t + \alpha_2 x_t + \alpha_3 y_t + u_t \quad (1)$$

$$p_t = \beta_1 q_t + \beta_2 x_t + \beta_3 y_t + v_t \quad (2)$$

suppose that u are strictly independent of x, y, p , while v is strictly independent of x, y, q .

Apparently, separately estimating these two equations leads to the problem of endogeneity. q_t and p_t are endogeneous, while x and y are exogeneous.

To address this problem, one solution is to solve the expression for q_t and p_t from these two equations, which can be rearranged as

$$p_t = \frac{\beta_1}{1 - \beta_1 \alpha_1} (\alpha_2 x_t + \alpha_3 y_t + u_t) + \frac{1}{1 - \beta_1 \alpha_1} (\beta_2 x_t + \beta_3 y_t + v_t)$$

$$q_t = \frac{1}{1 - \beta_1 \alpha_1} (\alpha_2 x_t + \alpha_3 y_t + u_t) + \frac{\alpha_1}{1 - \beta_1 \alpha_1} (\beta_2 x_t + \beta_3 y_t + v_t)$$

or

$$p_t = \gamma_1 x_t + \gamma_2 y_t + error_q$$

$$q_t = \theta_1 x_t + \theta_2 y_t + error_p$$

in which

$$\gamma_1 = \frac{\beta_1 \alpha_2 + \beta_2}{1 - \beta_1 \alpha_1}, \gamma_2 = \frac{\beta_1 \alpha_3 + \beta_3}{1 - \beta_1 \alpha_1}, \theta_1 = \frac{\alpha_2 + \alpha_1 \beta_2}{1 - \alpha_1 \beta_1}, \theta_2 = \frac{\alpha_3 + \alpha_1 \beta_3}{1 - \alpha_1 \beta_1}$$

In this equation system, the left hand side only contains the dependent variable (and endogeneous variables), p_t and q_t , while the right hand side only contains the exogeneous variables. We call this simplified equation system. Separately estimating each equation in this simplified system brings consistent estimation on $\theta_1, \theta_2, \gamma_1, \gamma_2$

But what we care more about is the original parameters $\beta_1, \beta_2, \beta_3, \alpha_1, \alpha_2, \alpha_3$. This is impossible: We only have four new parameters (i.e., four equations for the original parameters) in the simplified system, but we have six original parameters ! This makes the original parameters unidentifiable. Unless two additional constraints (equations) are imposed on original parameters, we cannot estimate the original parameters.

Identification Strategy

Then how to impose another two restrictions? A simple way is to directly put some parameters to zero. For example, we can impose $\alpha_2 = 0$, which means we 'exclude' the x in the equation (1). But of course only one additional restriction is not enough; we need one more restriction. We next discuss different possible cases.

$$\alpha_2 = 0, \alpha_3 = 0$$

How about further making $\alpha_3 = 0$, which means we ‘exclude’ the y in the equation (1)? This means that

$$(1 - \beta_1\alpha_1)p_t = \beta_2x_t + \beta_3y_t, \quad q_t = \alpha_1p_t$$

Therefore we can indentify $\frac{\beta_2}{1-\beta_1\alpha_1}, \frac{\beta_3}{1-\beta_1\alpha_1}, \frac{\alpha_1\beta_2}{1-\beta_1\alpha_1}, \frac{\alpha_1\beta_3}{1-\beta_1\alpha_1}$. But we still cannot identify $\beta_1, \beta_2, \beta_3$; α_1 is identifiable but overidentified: we can get the value of α_1 from the first and third expression, but we can also get it from the second and forth expression.

A summary : - for equation (1), the exogeneous variables that we exclude are x and y . In other words, we exclude two exogeneous variables in the equation (1), which is more than the number of endogeneous variables in (1), 1. Also, the exogeneous variables we exclude in (1) shows up in equation (2). We find out that the the parameter in equation (1), i.e., α_1 , is over identified. - On the other hand, in equation (2), we exclude 0 exogeneous variable, which is less than the number of endogeneous variables in (2). We find out that the parameters in equation (2) are unidentifiable.

$$\alpha_2 = 0, \beta_2 = 0$$

Another strategy is to make $\alpha_2 = 0$ and $\beta_2 = 0$. This gives

$$p_t = \gamma_2y_t + error_p$$

$$q_t = \theta_2y_t + error_q$$

in which

$$\gamma_2 = \frac{\beta_1\alpha_3 + \beta_3}{1 - \alpha_1\beta_1}, \theta_2 = \frac{\alpha_3 + \alpha_1\beta_3}{1 - \alpha_1\beta_1}$$

A summary : - notice that when $\alpha_2 = 0, \beta_2 = 0$. For equation (1), the exogeneous variables that we exclude are x . In other words, we exclude one exogeneous variable in the equation (1), which equals the number of endogeneous variables in (1), 1. Also, the exogeneous variables we exclude in (1) does not show up in equation (2). We find out that the the parameter in equation (1), i.e., α_1 and α_3 , is unidentified. - On the other hand, in equation (2), we exclude 1 exogeneous variables (x), which equals the number of endogeneous variables in (2), 1. Also, the exogeneous variables we exclude in (2) does not show up in equation (1). We find out that the the parameter in equation (1), i.e., β_1 and β_3 , are unidentified.

The result is the same for $\alpha_3 = 0$ and $\beta_3 = 0$.

$$\alpha_2 = 0, \beta_3 = 0$$

A final strategy is to make $\alpha_2 = 0$ and $\beta_3 = 0$. This gives

$$\gamma_1 = \frac{\beta_2}{1 - \beta_1\alpha_1}, \gamma_2 = \frac{\beta_1\alpha_3}{1 - \alpha_1\beta_1}, \theta_1 = \frac{\alpha_1\beta_2}{1 - \alpha_1\beta_1}, \theta_2 = \frac{\alpha_3}{1 - \alpha_1\beta_1}$$

We can see that all parameters are idendtifiable.

A summary : - notice that when $\alpha_2 = 0, \beta_3 = 0$. For equation (1), the exogeneous variables that we exclude are x . In other words, we exclude one exogeneous variable in the equation (1), which equals the number of endogeneous variables in (1), 1. Also, the exogeneous variables we exclude in (1) shows up in equation (2). We find out that the the parameter in equation (1), i.e., α_1 and α_3 , are identified. - On the other hand, in equation (2), we exclude 1 exogeneous variables (y), which equals the number of endogeneous variables in (2), 1. Also, the exogeneous variables we exclude in (2) shows up in equation (1). We find out that the the parameter in equation (1), i.e., β_1 and β_2 , are identified.

The result is the same for $\alpha_3 = 0$ and $\beta_2 = 0$.

Conclusion (Important)

in the equation system, for an equation (i), denote its dependent variable as $y(i)$, its set of exogenous variable as $EX(i)$, its set of endogenous variables as $ED(i)$. Denote total exogenous variables in the whole equation system as EX . We need the following conditions: 1. The # of exogenous variables excluded in (i), i.e., # of $EX/EX(i)$, must equal or be more than the # of $ED(i)$. 2. for $\forall x \in EX/EX(i)$, it must appear in - either in equation j s.t., $y(j) \in ED(i)$ - or in equation j such that $y(j)$ appears in the equation (k) s.t. $y(k) \in ED(i)$

If both two conditions are satisfied, then all the parameters in the equation (i), can be identified. If the number of excluded exogenous variables is strictly larger than (equals) the # of endogenous variables, then the parameters in this equation is over identified (just identified).

To see the application of this example, we next see a simple example (omit the error terms):

$$y_1 = \beta_{12}y_2 + \beta_{13}y_3 + \theta z$$

$$y_2 = \beta_2 y_1 + \gamma z$$

$$y_3 = \beta_3 y_1 + \alpha x$$

- For first equation, the excluded exogenous variable is x , while there are two endogenous variables. therefore, the parameters in the first equation is not identified, since the first condition in 1.1.4 is violated. - For second equation, the excluded exogenous variable is x , and there is one endogenous variable. Also, the excluded variable x appears in the third equation, where the dependent variable y_3 appears in the first equation where the dependent variable y_1 is an endogenous variable of the second equation. (This is exactly the second case in the second condition!). therefore, the parameters in the second equation are all identified. - For third equation, the excluded exogenous variable is z , and there is one endogenous variable. Also, the excluded variable z appears in the first equation, where dependent variable y_1 is the endogenous variable of the third equation (This is exactly the first case in the second condition !). Therefore, the parameters in the third equation are all identified.

Estimation Strategy

Having addressed the issues of the identification, we next turn to the estimation strategy. R package 'systemfit' is a library for doing simultaneous equation estimation. For example of using this package, see [Here](#)

Error terms are uncorrelated across equations

ILS

Intuitively, if all the equations are just identified, we can first estimate the simplified equation system and then back out the original parameters. Apparently, this strategy does not work under the case of over identification: for one original parameter, we may have multiple equations that can determine its value, and we actually do not know which one to choose.

2SLS

Check the second condition in the conclusion part, 1.1.4. If this condition is satisfied, then the x , i.e., the excluded exogenous variables, are actually instrument variables of $EG(i)$. Then we can estimate each equation separately using the 2SLS. The merit of using 2SLS is that it can deal with over-identified situation. Since we can use the excluded EX as IV and apply 2SLS to estimate the equation.

Error terms are correlated across equations

Intuitively, when error terms are correlated across equations, estimating them together would be efficient.
3SLS We can first estimate each equation using 2sls. We then get the residuals vectors for each equations, and we can estimate the covariance matrix of the error terms. Using this covariate matrix, we can use GLS to estimate..(?)

MLE

We can use the MLE to estimate. This can be easily done using Rstan.

The following is some code using systemfit package.

```
library (systemfit)
library(readstata13)
df_raw = read.dta13('klein.dta')

# SUR. No endogeneous issues, but allow the correlation of error terms across equations
eqs1 = consump~ wagepriv + wagegovt
eqs2 = wagepriv ~ govt + capital1
system = list(eqs1, eqs2)
fitsur <- systemfit( system, "SUR", data = df_raw, maxit = 100 )

# There is endogeneous variable issues. So use 2sls
# assume the error terms are uncorrelated.
eqs1 = consump~ wagepriv + wagegovt
eqs2 = wagepriv ~ consump + govt + capital1
system = list(eqs1, eqs2)
inst = ~ govt+ capital1 + wagegovt # it seems safe to write ALL the exogeneous variable into the instru

fit2sls <- systemfit( system, "2SLS", inst = inst, data = df_raw )

# 3sls: i.e. using 2sls, and allow the correlation of error terms across equations.
fit3sls <- systemfit( system, "3SLS", inst = inst, data = df_raw , maxit = 500)
```