Quantitative Analysis of Finance I

ECON90033

THE NATURE AND IDENTIFICATION OF SIMULTANEOUS EQUATION MODELS

Dr László Kónya
January 2023

SIMULTANEOUS EQUATION MODELS (*SEM*)

- Single-equation models are based on the implicit assumption that there is a one-way relationship running from the set of independent variables (X's) to a single dependent variable (Y).

  $\longleftrightarrow$ In many situations this assumption is unrealistic because there is a two-way or simultaneous relationship, that is some of the X's are also affected by Y.

  In these cases, instead of modelling each variable separately with single-equation models, they should be modelled simultaneously in a multi-equation system.

  Doing so, it does not make sense any more to categorize the variables as independent and dependent.

  Instead, the variables that are jointly determined within the model are called endogenous variables, the remaining non-stochastic variables are referred to as predetermined (exogenous and lagged endogenous) variables, and there must be one equation for each endogenous variable.

- A simultaneous equation model (*SEM*) with *M* endogenous and *K* exogenous variables takes the following form:

$$
\begin{aligned}
y_{1t} &= \delta_{10} + \delta_{12} y_{2t} + \ldots + \delta_{1M} y_{Mt} + \lambda_{11} x_{1t} + \ldots + \lambda_{1K} x_{Kt} + \varepsilon_{1t} \\
y_{2t} &= \delta_{20} + \delta_{21} y_{1t} + \delta_{23} y_{3t} + \ldots + \delta_{2M} y_{Mt} + \lambda_{21} x_{1t} + \ldots + \lambda_{2K} x_{Kt} + \varepsilon_{2t} \\
&\vdots \\
y_{Mt} &= \delta_{M0} + \delta_{M1} y_{1t} + \ldots + \delta_{M,M-1} y_{M-1,t} + \lambda_{M1} x_{1t} + \ldots + \lambda_{MK} x_{Kt} + \varepsilon_{Mt}
\end{aligned}
$$

There are many simultaneous-equation models in economics, like

i. *Classical demand-supply model*

In equilibrium

$$
\begin{aligned}
Q_t^d &= \alpha_0 + \alpha_1 P_t + u_t \quad , \quad \alpha_1 < 0 \\
Q_t^s &= \beta_0 + \beta_1 P_t + w_t \quad , \quad \beta_1 > 0 \\
Q_t^d &= Q_t^s
\end{aligned}
\longrightarrow
\begin{aligned}
Q_t &= \alpha_0 + \alpha_1 P_t + u_t \\
P_t &= \frac{-\beta_0}{\beta_1} + \frac{1}{\beta_1} Q_t + \frac{-w_t}{\beta_1}
\end{aligned}
$$

where Q_t, P_t are quantity and price, and *s* and *d* denote supply and demand, respectively.

ii. *Phillips-type model of wage and price*

$$\dot{W}_t = \alpha_0 + \alpha_1 U_t + \alpha_2 \dot{P}_t + u_t$$
$$\dot{P}_t = \beta_0 + \beta_1 \dot{W}_t + \beta_2 \dot{R}_t + \beta_3 \dot{M}_t + w_t$$

where the dots above the variables refer to rates of change,

W_t is money wage, U_t is unemployment rate, P_t is price, R_t is cost of capital, and M_t is price of imported raw material.

iii. *Keynesian model of income determination*

$$C_t = \beta_0 + \beta_1 Y_t + u_t \quad , \quad 0 < \beta_1 < 1$$
$$Y_t = C_t + I_t$$

where C_t is consumption, Y_t is income and I_t is investment (savings).

- In simultaneous-equation models the error terms are typically correlated with some of the predetermined variables. Consequently, the application of the classical OLS method to the equations individually results in inconsistent estimators.

This can be illustrated with the Keynesian model of income determination as follows.

To show that Y_t and u_t are correlated, substitute the first equation into the second and rearrange the terms.

$$Y_t = \beta_0 + \beta_1 Y_t + I_t + u_t \longrightarrow Y_t = \frac{\beta_0}{1-\beta_1} + \frac{1}{1-\beta_1} I_t + \frac{1}{1-\beta_1} u_t$$

$$\longrightarrow E(Y_t) = \frac{\beta_0}{1-\beta_1} + \frac{1}{1-\beta_1} I_t \longrightarrow Y_t - E(Y_t) = \frac{1}{1-\beta_1} u_t$$

$$\longrightarrow Cov(Y_t, u_t) = E\left[\left(Y_t - E(Y_t)\right) u_t\right] = E\left(\frac{u_t^2}{1-\beta_1}\right) = \frac{\sigma_u^2}{1-\beta_1} \neq 0$$

Hence, Y_t and u_t are correlated, violating a crucial OLS assumption.

Is the OLS estimator of β_1 unbiased, i.e. $E\left(\hat{\beta}_1\right) \overset{?}{=} \beta_1$

$$\hat{\beta}_1 = \frac{\sum (C_t - \bar{C})(Y_t - \bar{Y})}{\sum (Y_t - \bar{Y})^2} = \frac{\sum C_t (Y_t - \bar{Y})}{\sum (Y_t - \bar{Y})^2}$$

$$= \frac{\sum (\beta_0 + \beta_1 Y_t + u_t)(Y_t - \bar{Y})}{\sum (Y_t - \bar{Y})^2} = \frac{\sum \beta_1 Y_t (Y_t - \bar{Y})}{\sum (Y_t - \bar{Y})^2} + \frac{\sum u_t (Y_t - \bar{Y})}{\sum (Y_t - \bar{Y})^2}$$

$$= \beta_1 + \frac{\sum (Y_t - \bar{Y}) u_t}{\sum (Y_t - \bar{Y})^2}$$

We should now take the expected value of both sides.

However, the expectation operator is a linear operator, so the expected value of the second term is not equal to the ratio of the expected value of the numerator to the expected value of the denominator.

$\longrightarrow$ We cannot simplify the second term on the right side. Consequently, we cannot explicitly evaluate the expected value of the β_1-hat estimator and to establish its finite sample properties.

It is possible, though, to study its behaviour under the assumption that the sample size increases indefinitely, i.e. to see whether it is consistent.

← An estimator θ-hat is a consistent estimator of θ if the probability that the deviation of θ-hat from θ is less than any arbitrarily small positive quantity (δ) approaches unity, that is

$$\lim_{n\to\infty} P\left\{\left|\hat{\theta}-\theta\right| < \delta\right\} = 1 \ , \ \delta > 0$$

Alternatively, this can be written as

$$\underset{n\to\infty}{\text{plim}}\, \hat{\theta} = \theta$$ where plim denotes the probability limit

and has the following properties:

i. For a constant *c*, $\text{plim}\, c = c$

ii. If *h* is a continuous function of a consistent estimator θ-hat,

$$\text{plim}\, h(\hat{\theta}) = h(\theta)$$

iii. If θ_1-hat and θ_2-hat are consistent estimators,

$$\text{plim}\left(\hat{\theta}_1 + \hat{\theta}_2\right) = \text{plim}\,\hat{\theta}_1 + \text{plim}\,\hat{\theta}_2$$

$$\text{plim}\left(\hat{\theta}_1 \hat{\theta}_2\right) = \text{plim}\,\hat{\theta}_1 \times \text{plim}\,\hat{\theta}_2$$

$$\text{plim}\left(\frac{\hat{\theta}_1}{\hat{\theta}_2}\right) = \frac{\text{plim}\,\hat{\theta}_1}{\text{plim}\,\hat{\theta}_2}$$

Note:
The expectation operator does not have these two properties in general.

It can be shown that the sample variance of *Y* and the sample covariance of *Y* and *u*

$$s_Y^2 = \frac{\sum (Y_t - \bar{Y})^2}{n} \quad , \quad s_{Y,u} = \frac{\sum (Y_t - \bar{Y}) u_t}{n}$$

are consistent estimators.

Therefore, the application of the properties of plim to the β_1-hat OLS estimator yields

$$\operatorname*{plim}_{n\to\infty} \hat{\beta}_1 = \beta_1 + \frac{\operatorname{plim}\left(\sum (Y_t - \bar{Y})u_t\right)}{\operatorname{plim}\left(\sum (Y_t - \bar{Y})^2\right)} = \beta_1 + \frac{\operatorname{plim}\left(ns_{Y,u}\right)}{\operatorname{plim}\left(ns_Y^2\right)}$$

$$= \beta_1 + \frac{1}{1-\beta_1}\frac{\sigma_u^2}{\sigma_y^2}$$

→ No matter how large the sample is, the probability limit of β_1-hat is always different from β_1 (in fact, greater than β_1 because $0 < \beta_1 < 1$).

→ β_1-hat is a biased estimator and its bias does not disappear as the sample size increases, so β_1-hat is an inconsistent estimator.

Since the inconsistency of β_1-hat is due to simultaneous nature of the system, this bias is often referred to as simultaneous equation bias.

THE IDENTIFICATION OF *SEM*

- As it was already mentioned on slide #3, a simultaneous equation model in M endogenous variables (Y) and K predetermined variables (X) in general looks like

$$
\begin{aligned}
y_{1t} &= \delta_{10} + \delta_{12} y_{2t} + \ldots + \delta_{1M} y_{Mt} + \lambda_{11} x_{1t} + \ldots + \lambda_{1K} x_{Kt} + \varepsilon_{1t} \\
y_{2t} &= \delta_{20} + \delta_{21} y_{1t} + \delta_{23} y_{3t} + \ldots + \delta_{2M} y_{Mt} + \lambda_{21} x_{1t} + \ldots + \lambda_{2K} x_{Kt} + \varepsilon_{2t} \\
&\vdots \\
y_{Mt} &= \delta_{M0} + \delta_{M1} y_{1t} + \ldots + \delta_{M,M-1} y_{M-1,t} + \lambda_{M1} x_{1t} + \ldots + \lambda_{MK} x_{Kt} + \varepsilon_{Mt}
\end{aligned}
$$

These equations are supposed to describe the structure of an economic system, so they are known as structural (or behaviour) equations and their coefficients are referred to as structural coefficients.

Note:

On the right side, each equation might have at most M -1 endogenous variables and all K predetermined variables. If a given variable is not included in an equation, then the corresponding coefficient is set to zero.

Unlike the structural equations, the so-called reduced-form equations express each endogenous variable solely in terms of the predetermined variables and the stochastic error terms.

To illustrate structural and reduced-form equations, consider again the Keynesian model of income determination:

$$C_t = \beta_0 + \beta_1 Y_t + u_t$$
$$Y_t = C_t + I_t$$

Structural equations

$$\longrightarrow$$

$$Y_t = \frac{\beta_0}{1-\beta_1} + \frac{1}{1-\beta_1} I_t + \frac{u_t}{1-\beta_1}$$

Reduced-form equation for Y_t

$$Y_t = \pi_0 + \pi_1 I_t + w_t$$

In this model I_t is an exogenous variable, i.e. it is determined outside this model. Consequently, w_t and I_t are independently distributed and the reduced-form equation for Y_t can be estimated by OLS.

But, once the reduced-form equation is estimated, is it possible to estimate the β_0, β_1 structural coefficients from the estimates of the π_1, π_2 reduced-form coefficients?

This is the so called identification problem.

In this case, the answer is yes because $$\beta_0 = \frac{\pi_0}{\pi_1} \; , \; \beta_1 = 1 - \frac{1}{\pi_1}$$

What about the classical demand-supply model (slide #3)?

$$\begin{aligned} Q_t^d &= \alpha_0 + \alpha_1 P_t + u_t \\ Q_t^s &= \beta_0 + \beta_1 P_t + w_t \\ Q_t^d &= Q_t^s \end{aligned} \longrightarrow \begin{aligned} P_t &= \frac{\alpha_0 - \beta_0}{\beta_1 - \alpha_1} + \frac{u_t - w_t}{\beta_1 - \alpha_1} \\ Q_t &= \frac{\alpha_0 \beta_1 - \alpha_1 \beta_0}{\beta_1 - \alpha_1} + \frac{\beta_1 u_t - \alpha_1 w_t}{\beta_1 - \alpha_1} \end{aligned}$$

Structural equations

Reduced-form equations

$$\begin{aligned} P_t &= \pi_{10} + e_{1t} \\ Q_t &= \pi_{20} + e_{2t} \end{aligned}$$

These reduced-form equations can also be estimated by OLS, producing estimates of the π_{10}, π_{20} reduced-form coefficients.The structural equations, however, have four unknown parameters (α_0, α_1, β_0, β_1).

$\longrightarrow$ It is impossible to recover four structural parameters from two reduced-form parameters.

$\longleftarrow$ Every (P_t, Q_t) pair of observations represents a point of equilibrium where the demand and supply relations are observationally equivalent.

Hence, infinite number of combinations of demand and supply curves could have generated the observations, and without some additional information this structural system cannot be identified from the reduced-form system.

Systems like this are said to be unidentified.

What if we add an exogenous variable, namely the average income of the consumers (I_t), to the demand equation?

$$Q_t^d = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + u_t$$
$$Q_t^s = \beta_0 + \beta_1 P_t + w_t$$
$$Q_t^d = Q_t^s$$

$\longrightarrow$

$$P_t = \frac{\alpha_0 - \beta_0}{\beta_1 - \alpha_1} + \frac{\alpha_2}{\beta_1 - \alpha_1} I_t + \frac{u_t - w_t}{\beta_1 - \alpha_1}$$
$$Q_t = \frac{\alpha_0 \beta_1 - \alpha_1 \beta_0}{\beta_1 - \alpha_1} + \frac{\alpha_2 \beta_1}{\beta_1 - \alpha_1} I_t + \frac{\beta_1 u_t - \alpha_1 w_t}{\beta_1 - \alpha_1}$$

$$\longrightarrow \quad \begin{aligned} P_t &= \pi_{10} + \pi_{11} I_t + e_{1t} \\ Q_t &= \pi_{20} + \pi_{21} I_t + e_{2t} \end{aligned}$$

Again, the structural system has more parameters than the reduced-form system, so it is unidentified.

Yet, due to the additional variable in the demand equation, it is possible to identify the supply equation at least:

$$\beta_0 = \pi_{20} - \beta_1 \pi_{10} \quad , \quad \beta_1 = \frac{\pi_{21}}{\pi_{11}}$$

As a general rule, a structural equation can be identified if it differs from all possible linear combinations of the other structural equations in the system.

In the current example this linear combination is

$$\begin{aligned} \lambda Q_t^d + (1-\lambda) Q_t^s &= \lambda\left(\alpha_0 + \alpha_1 P_t + \alpha_2 I_t + u_t\right) + (1-\lambda)\left(\beta_0 + \beta_1 P_t + w_t\right) \\ &= \left(\lambda\alpha_0 + (1-\lambda)\beta_0\right) + \left(\lambda\alpha_1 + (1-\lambda)\beta_1\right) P_t + \lambda\alpha_2 I_t + \left(\lambda u_t + (1-\lambda) w_t\right) \end{aligned}$$

In terms of the right-hand side variables, this mongrel equation looks like demand equation, but it has one more variable than the supply equation. Consequently, only the supply equation can be identified.

Suppose this time that the supply equation is also augmented by a predetermined variable, say by $P_{t\text{-}1}$, that is

$$Q_t^d = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + u_t$$
$$Q_t^s = \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + w_t$$
$$Q_t^d = Q_t^s$$

Structural equations

The linear combination of these two structural equations would have three variables on its right side, P_t, I_t and $P_{t\text{-}1}$, so it would be different observationally from both the demand and supply equations. Consequently, these structural equations ca be identified.

Reduced-form equations

$$P_t = \frac{\alpha_0 - \beta_0}{\beta_1 - \alpha_1} + \frac{\alpha_2}{\beta_1 - \alpha_1} I_t + \frac{-\beta_2}{\beta_1 - \alpha_1} P_{t-1} + \frac{u_t - w_t}{\beta_1 - \alpha_1}$$
$$Q_t = \frac{\alpha_0\beta_1 - \alpha_1\beta_0}{\beta_1 - \alpha_1} + \frac{\alpha_2\beta_1}{\beta_1 - \alpha_1} I_t + \frac{-\alpha_1\beta_2}{\beta_1 - \alpha_1} P_{t-1} + \frac{\beta_1 u_t - \alpha_1 w_t}{\beta_1 - \alpha_1}$$

In general, a system like this has a unique solution, implying that both structural equations are exactly identified.

Finally, suppose that the demand equation is further augmented by an extra exogenous variable, say by R_t : average wealth of the consumers.

$$Q_t^d = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + \alpha_3 R_t + u_t$$
$$Q_t^s = \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + w_t$$
$$Q_t^d = Q_t^s$$

Structural equations

The linear combination of the two stochastic structural equations would have four variables on its right side (P_t, I_t, P_{t-1}, R_{t-1}), so again it is observationally different from both the demand and supply equations.

$$P_t = \frac{\alpha_0 - \beta_0}{\beta_1 - \alpha_1} + \frac{\alpha_2}{\beta_1 - \alpha_1} I_t + \frac{-\beta_2}{\beta_1 - \alpha_1} P_{t-1} + \frac{\alpha_3}{\beta_1 - \alpha_1} R_t + \frac{u_t - w_t}{\beta_1 - \alpha_1}$$
$$Q_t = \frac{\alpha_0 \beta_1 - \alpha_1 \beta_0}{\beta_1 - \alpha_1} + \frac{\alpha_2 \beta_1}{\beta_1 - \alpha_1} I_t + \frac{-\alpha_1 \beta_2}{\beta_1 - \alpha_1} P_{t-1} + \frac{\alpha_3 \beta_1}{\beta_1 - \alpha_1} R_t + \frac{\beta_1 u_t - \alpha_1 w_t}{\beta_1 - \alpha_1}$$

Reduced-form equations

However, now the structural system has less parameters than the reduced-form system (7 vs. 8). Consequently, the structural parameters can be estimated from the structural parameters, but not without ambiguity.

For example, denoting the reduced form parameters again as π_{ij} (i = 1, 2 and j = 0, 1, 2, 3), β_1 can be recovered from the following two relationships between the structural and reduced form coefficients:

$$\boxed{\beta_1 = \frac{\pi_{21}}{\pi_{11}}} \quad \text{and} \quad \boxed{\beta_1 = \frac{\pi_{23}}{\pi_{13}}}$$

There is no guarantee that these two estimators of β_1 would produce the same estimates.

⟶ Since it is impossible to identify the structural system uniquely, it is said to be overidentified.

- Given that the reduced-form equations fully determine the (current) endogenous variables, why is the identification problem an issue at all?

  For the purposes of short-term forecasting and policy analysis the reduced-form equations are sufficient, granted that they are correctly specified and estimated. These equations, however, are purely technical and they cannot be used to test the validity of economic theories. Moreover, since all reduced-form equations are affected by the smallest structural change, reduced-form parameters are far less stable than structural parameters.

RULES FOR IDENTIFICATION OF SIMULTANEOUS EQUATION MODELS

- A structural equation is said to be identified if it is possible to estimate its parameters from the estimated reduced-form coefficients. Otherwise it is underidentified.

  An identified equation can be either exactly identified (if the reduced-form coefficients uniquely determine the structural parameters), or overidentified (if the structural parameters cannot be derived without ambiguity from the reduced-form coefficients).

  A structural model as a whole is identified if all of its equations are identified.

- As we have already seen (slide #14), in principle it is possible to determine whether or not a structural equation is observationally different from the rest of the structural system and thus is identified, by comparing it to a mongrel equation. In large systems, however, this procedure can be very cumbersome.

  Fortunately, there are two relatively simple conditions of identification.

- A necessary but not sufficient condition of identification is the so-called order condition:

  The i^{th} equation of a structural system can be identified only if the number of predetermined variables excluded from this equation is not smaller than the number endogenous variables included in it less 1.

In order to write this rule symbolically, let's use the following notations.

M : total number of endogenous variables in the (structural) system,
m_i : the number of endogenous variables in the i^{th} equation,
$m^*_i = M - m_i$: the number of endogenous variables excluded from the i^{th} equation,

K : the total number of predetermined variables in the system,
k_i : the number of predetermined variables in the i^{th} equation,
$k^*_i = K - k_i$: the number of predetermined variables excluded from the i^{th} equation.

If $\boxed{k_i^* > m_i - 1} \longrightarrow \boxed{m_i^* + k_i^* > M - 1} \longrightarrow$ The equation might be overidentified.

$\boxed{k_i^* = m_i - 1} \longrightarrow \boxed{m_i^* + k_i^* = M - 1} \longrightarrow$ The equation might be exactly identified.

$\boxed{k_i^* < m_i - 1} \longrightarrow \boxed{m_i^* + k_i^* < M - 1} \longrightarrow$ The equation is unidentified.

Ex 1: Apply the order condition to the various demand-supply models considered earlier.

a) Eliminating the equilibrium condition, the first structural system is

$$Q_t = \alpha_0 + \alpha_1 P_t + u_t$$
$$Q_t = \beta_0 + \beta_1 P_t + w_t$$

$\longrightarrow$

$$M = 2 \;,\; K = 0$$
$$m_1 = 2 \;,\; m_1^* = 0, \;,\; k_1 = 0 \;,\; k_1^* = 0$$
$$m_2 = 2 \;,\; m_2^* = 0, \;,\; k_2 = 0 \;,\; k_2^* = 0$$

$\longrightarrow$

$$k_1^* < m_1 - 1$$
$$k_2^* < m_2 - 1$$

Hence, both equations (demand and supply) are unidentified.

b)

$$Q_t = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + u_t$$
$$Q_t = \beta_0 + \beta_1 P_t + w_t$$

$\longrightarrow$

$$M = 2 \;,\; K = 1$$
$$m_1 = 2 \;,\; m_1^* = 0, \;,\; k_1 = 1 \;,\; k_1^* = 0$$
$$m_2 = 2 \;,\; m_2^* = 0, \;,\; k_2 = 0 \;,\; k_2^* = 1$$

$\longrightarrow$

$$k_1^* < m_1 - 1$$
$$k_2^* = m_2 - 1$$

Hence, the first equation is unidentified, but the second might be exactly identified.

c)

$$Q_t = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + u_t$$
$$Q_t = \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + w_t$$

$\longrightarrow$

$$M = 2 \;,\; K = 2$$
$$m_1 = 2 \;,\; m_1^* = 0, \;,\; k_1 = 1 \;,\; k_1^* = 1$$
$$m_2 = 2 \;,\; m_2^* = 0, \;,\; k_2 = 1 \;,\; k_2^* = 1$$

$\longrightarrow$

$$k_1^* = m_1 - 1$$
$$k_2^* = m_2 - 1$$

Hence, both equations might be exactly identified.

d)

$$Q_t = \alpha_0 + \alpha_1 P_t + \alpha_2 I_t + \alpha_3 R_t + u_t$$
$$Q_t = \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + w_t$$

$\longrightarrow$

$$M = 2 \;,\; K = 3$$
$$m_1 = 2 \;,\; m_1^* = 0, \;,\; k_1 = 2 \;,\; k_1^* = 1$$
$$m_2 = 2 \;,\; m_2^* = 0, \;,\; k_2 = 1 \;,\; k_2^* = 2$$

$\longrightarrow$

$$k_1^* = m_1 - 1$$
$$k_2^* > m_2 - 1$$

Hence, the first equation might be exactly identified, but the second is overidentified.

Two predetermined variables are excluded from the supply equation and β_1 can be estimated in two different ways.

Note:

a) According to the order condition, a structural equation in a simultaneous system is identified if it excludes enough number of variables that are present in some other equations of the system.

The exclusion of a variable from an equation means that its coefficient is set equal to zero (hopefully, on the basis of some solid economic theory).

b) The order condition is a necessary but not sufficient condition for identification.

$\longleftarrow$ Even if an equation satisfies $k^* \geq m - 1$, it might still not be identified because the predetermined variables excluded from this particular equation may not all be independent.

$\longrightarrow$ In order to make the order condition sufficient, it has to be supplemented by the following rule: no linear combination of the other equations in the system can produce the given equation.

(Ex 1)

There are only two equations in each system, and it is clear that this auxiliary condition is violated only in part *a*, when the order condition is also violated.

Ex 2: Consider the following extended Keynesian model of income determination:

$$C_t = \beta_0 + \beta_1 Y_t + \beta_2 T_t + \varepsilon_{1t}$$
$$I_t = \alpha_0 + \alpha_1 Y_{t-1} + \varepsilon_{2t}$$
$$T_t = \gamma_0 + \gamma_1 Y_t + \varepsilon_{3t}$$
$$Y_t = C_t + I_t + G_t$$

where C is consumption expenditures, I is investment, T is taxes, Y is income, and G is government expenditure.

There are four endogenous variables (C_t, I_t, T_t, Y_t), and two predetermined variables (G_t, Y_{t-1}). Hence, $M = 4$ and $K = 2$.

a) By applying the order condition, check the identifiability of each of the equations in the system and of the system as a whole.

The fourth equation is an identity, so it does not have to be identified.

$$m_1 = 3 \ , \ m_1^* = 1 \ , \ k_1 = 0 \ , \ k_1^* = 2$$
$$m_2 = 1 \ , \ m_2^* = 3 \ , \ k_2 = 1 \ , \ k_2^* = 1$$
$$m_3 = 2 \ , \ m_3^* = 2 \ , \ k_3 = 0 \ , \ k_3^* = 2$$

$$\longrightarrow$$

$$k_1^* = m_1 - 1$$
$$k_2^* > m_2 - 1$$
$$k_3^* > m_3 - 1$$

Thus, the 1st equation might be exactly identified, while the 2nd and 3rd equations might be overidentified, so the whole structural system is identified.

b) What would happen if r_t, the exogenous interest rate appeared on the right-hand side of the investment function?

→ In the new augmented system $M = 4$ and $K = 3$.

Although only the second equation has changed, since the equations are interrelated through the simultaneously determined endogenous variables, we have to reconsider the identifiability of each stochastic equation.

$$\begin{array}{l} m_1 = 3 \ , \ m_1^* = 1 \ , \ k_1 = 0 \ , \ k_1^* = 3 \\ m_2 = 1 \ , \ m_2^* = 3 \ , \ k_2 = 2 \ , \ k_2^* = 1 \\ m_3 = 2 \ , \ m_3^* = 2 \ , \ k_3 = 0 \ , \ k_3^* = 3 \end{array} \longrightarrow \begin{array}{l} k_1^* > m_1 - 1 \\ k_2^* > m_2 - 1 \\ k_3^* > m_3 - 1 \end{array}$$

Thus, in the new system all three equations might be overidentified and the structural system itself is identified.

- A necessary condition of identification is the so-called rank condition:

  The i^{th} equation of a structural system is identified if and only if the rank (that is, the largest number of linearly independent rows or columns) of matrix Δ, constructed from the coefficients of all the variables excluded from the given equation but included in the system, is equal to the number of equations less one.

  Symbolically, $\boxed{\rho(\boldsymbol{\Delta}) = M - 1}$

(Ex 2)

c) Do the equations in the original system meet the rank condition?

To make the task easier, move all terms but the error terms to the left side and re-write this system in a tabular form.

	1	C_t	I_t	T_t	Y_t	Y_{t-1}	G_t
C_t	$-\beta_0$	1	0	$-\beta_1$	$-\beta_2$	0	0
I_t	$-\alpha_0$	0	1	0	0	$-\alpha_1$	0
T_t	$-\gamma_0$	0	0	1	$-\gamma_1$	0	0
Y_t	0	-1	-1	0	1	0	-1

Consider the first equation.
It excludes I_t, Y_{t-1} and G_t.

$$\longrightarrow \quad \mathbf{\Delta}_C = \begin{bmatrix} 1 & -\alpha_1 & 0 \\ 0 & 0 & 0 \\ -1 & 0 & -1 \end{bmatrix}$$

The rank of this square matrix is $M - 1 = 3$ if and only if its determinant is nonzero.

However, because of the zero raw, this is clearly not the case.

In general, the rank of an $m \times n$ matrix cannot be greater than m or n (whichever is smaller), and it is equal to the order of the largest square sub-matrix with non-zero determinant.

One of the 2×2 sub-determinants of Δ_C is $\begin{vmatrix} 1 & 0 \\ -1 & -1 \end{vmatrix} = -1 \longrightarrow \rho(\mathbf{\Delta}_C) = 2$

Consider now the second equation. The corresponding Δ_I matrix is

$$\mathbf{\Delta}_I = \begin{bmatrix} 1 & -\beta_1 & -\beta_2 & 0 \\ 0 & 1 & \boxed{-\gamma_1} & 0 \\ -1 & 0 & 1 & -1 \end{bmatrix}$$

The rank of this non-square matrix is at most three. In fact, dropping for example the third column, Laplace expansion by the fourth column yields that the rank of Δ_I is three.

Namely,

$$\begin{vmatrix} 1 & -\beta_1 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & -1 \end{vmatrix} = -\begin{vmatrix} 1 & -\beta_1 \\ 0 & 1 \end{vmatrix} = -1 \longrightarrow \boxed{\rho(\mathbf{\Delta}_I) = 3}$$

Similarly, the Δ_T matrix is

$$\mathbf{\Delta}_T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & \boxed{-\alpha_1} & 0 \\ -1 & -1 & 0 & -1 \end{bmatrix} \longrightarrow \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & -1 \end{vmatrix} = -\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = -1$$

$$\longrightarrow \boxed{\rho(\mathbf{\Delta}_T) = 3}$$

$\longrightarrow$ The first equation is underidentified, but the other two are identified.

Note: The rank condition can be used to determine whether a structural equation is unidentified or identified; and if it is identified, the order condition can tell whether it is exactly identified or overidentified.

WHAT SHOULD YOU KNOW?

- The differences between endogenous, exogenous, and predetermined variables.
- The difference between structural and reduced-form equations.
- The identification problem of simultaneous equation models.
- The order-condition and the rank condition of identification.

For further information see, for example,

Gujarati (2003), § 18.1-18.5, 19.1-19.3