Identification

Consider how we have analysed the consumption function. Start with a theory and this gives us the variables to use and expected signs etc

Have to decide on functional form consistent with theory. Simplest is linear

$$C_t = \alpha + \beta Y_t + u_t$$
$$\alpha > 0; 0 < \beta < 1$$

Complete the system by an identity

$$Y_t = C_t + I_t$$

Have two structural equations that determine the value of the variable I and the parameters α and β

Endogenous variables are those determined in the system: C_t and Y_t

Exogenous variables are those external to the system I_t

For this system to be consistent i.e. to have a determinate solution giving unique values of the endogenous variables in terms of parameters and exogenous variables, the number of equations must be the same as the number of endogenous variables.

$$C_t = \alpha + \beta(C_t + I_t) + u_{1t}$$
$$C_t = \frac{\alpha}{1-\beta} + \frac{\beta}{1-\beta} I_t + u_{1t}$$

Likewise

$$Y_t = \frac{\alpha}{1-\beta} + \frac{1}{1-\beta} I_t + u_{2t}$$

which we can write as

$$C_t = \Pi_1 + \Pi_2 I_t + u_{1t}$$
$$Y_t = \Pi_3 + \Pi_4 I_t + u_{2t}$$

There are four reduced form coefficients and only two structural parameters so expect some inter-relationships

$$\Pi_1 = \Pi_3 \text{and } \Pi_4 - \Pi_2 = 1$$
$$\text{as } \frac{\alpha}{1-\beta} = \frac{\alpha}{1-\beta} \text{ and } \frac{1}{1-\beta} - \frac{\beta}{1-\beta} = 1$$

Now we can use these reduced form equations to predict the levels of income and consumption associated with a given level of investment. We can do this without any reference to the structural form

To get back to the structural form we would use the equations

$$\Pi_1 = \Pi_3 = \frac{\alpha}{1-\beta}\,;\, \Pi_4 = \frac{1}{1-\beta}\,;\, \Pi_2 = \frac{\beta}{1-\beta}$$

so

$$\Pi_4 = \frac{1}{1-\beta} \Rightarrow \beta = 1 - \frac{1}{\Pi_4}$$
$$\Pi_1 = \frac{\alpha}{1-\beta} \Rightarrow \alpha = \frac{\Pi_1}{\Pi_4}$$

but sometimes we can't do it and this is the IDENTIFICATION PROBLEM

It is important to be able to get back to the structural parameters because this is where the theory determines the nature of the model and how allow for a change in the structure

If the structure changes need more than just the reduced form to make a prediction

This topic would be more easily discussed in matrix algebra, but wont do that here.

Take the example of demand and supply

$$Q_t^D = \alpha_{10} + \alpha_{11}p + \alpha_{12}y + u_1$$
$$Q_t^S = \alpha_{20} + \alpha_{21}p + \alpha_{22}R + u_2$$

We can solve for p and Q in terms of Y and R

$$\alpha_{10} + \alpha_{11}p + \alpha_{12}y + u_1 = \alpha_{20} + \alpha_{21}p + \alpha_{22}R + u_2$$

$$(\alpha_{10} - \alpha_{20}) + (\alpha_{11} - \alpha_{21})p + \alpha_{12}y - \alpha_{22}R + (u_1 - u_2) = 0$$

$$(\alpha_{10} - \alpha_{20}) + \alpha_{12}y - \alpha_{22}R + (u_1 - u_2) = (\alpha_{21} - \alpha_{11})p$$

$$\Rightarrow p = \frac{\alpha_{10} - \alpha_{20}}{\alpha_{21} - \alpha_{11}} + \frac{\alpha_{12}}{\alpha_{21} - \alpha_{11}}y - \frac{\alpha_{22}}{\alpha_{21} - \alpha_{11}}R + v_1$$

Now take one of the equations and substitute for p

$$Q = \alpha_{10} + \alpha_{11}\left[\frac{\alpha_{10} - \alpha_{20}}{\alpha_{21} - \alpha_{11}} + \frac{\alpha_{12}}{\alpha_{21} - \alpha_{11}}y - \frac{\alpha_{22}}{\alpha_{21} - \alpha_{11}}R + error\right] + \alpha_{12}y + u_1$$
$$\Rightarrow Q = \alpha_{10} + \frac{\alpha_{11}(\alpha_{10} - \alpha_{20})}{\alpha_{21} - \alpha_{11}} + \frac{\alpha_{11}\alpha_{12}}{\alpha_{21} - \alpha_{11}}y - \frac{\alpha_{11}\alpha_{22}}{\alpha_{21} - \alpha_{11}}R + \alpha_{12}y + error$$
$$= \frac{\alpha_{10}(\alpha_{21} - \alpha_{11}) + \alpha_{11}(\alpha_{10} - \alpha_{20})}{\alpha_{21} - \alpha_{11}} + \frac{\alpha_{12}(\alpha_{21} - \alpha_{11}) + \alpha_{11}\alpha_{12}}{\alpha_{21} - \alpha_{11}}y - \frac{\alpha_{11}\alpha_{22}}{\alpha_{21} - \alpha_{11}}R + error$$

Now the first term is

$$\frac{\alpha_{10}(\alpha_{21} - \alpha_{11}) + \alpha_{11}(\alpha_{10} - \alpha_{20})}{\alpha_{21} - \alpha_{11}} = \frac{\alpha_{10}\alpha_{21} - \alpha_{10}\alpha_{11} + \alpha_{11}\alpha_{10} - \alpha_{11}\alpha_{20})}{\alpha_{21} - \alpha_{11}} = \frac{\alpha_{10}\alpha_{21} - \alpha_{11}\alpha_{20})}{\alpha_{21} - \alpha_{11}}$$

and the second

$$\frac{\alpha_{12}(\alpha_{21}-\alpha_{11})+\alpha_{11}\alpha_{12}}{\alpha_{21}-\alpha_{11}} = \frac{\alpha_{12}\alpha_{21}-\alpha_{12}\alpha_{11}+\alpha_{11}\alpha_{12}}{\alpha_{21}-\alpha_{11}} = \frac{\alpha_{12}\alpha_{21}}{\alpha_{21}-\alpha_{11}}$$

Giving

$$Q = \frac{\alpha_{10}\alpha_{21}-\alpha_{11}\alpha_{20}}{\alpha_{21}-\alpha_{11}} + \frac{\alpha_{12}\alpha_{21}}{\alpha_{21}-\alpha_{11}}y - \frac{\alpha_{11}\alpha_{22}}{\alpha_{21}-\alpha_{11}}R + error$$

Which we can write as

$$\begin{aligned} Q &= \Pi_1 + \Pi_2 y + \Pi_3 R + v_1 \\ p &= \Pi_4 + \Pi_5 y + \Pi_6 R + v_2 \end{aligned}$$

Can estimate this with Πs being the reduced form parameters. This is called indirect least squares with

$$\begin{aligned} \alpha_{11} &= \frac{\Pi_6}{\Pi_3} \\ \alpha_{21} &= \frac{\Pi_5}{\Pi_2} \\ \alpha_{22} &= \Pi_6(\alpha_{11}-\alpha_{21}) \\ \alpha_{12} &= -\Pi_5(\alpha_{11}-\alpha_{21}) \\ \alpha_{10} &= \Pi_1 - \alpha_{11}\Pi_4 \\ \alpha_{20} &= \Pi_1 - \alpha_{21}\Pi_4 \end{aligned}$$

Will leave you to work through the algebra as an exercise (see Madalla)

Now if

$$\begin{aligned} Q_t^D &= \alpha_{10} + \alpha_{11}p + \alpha_{12}y + u_1 \\ Q_t^S &= \alpha_{20} + \alpha_{21}p + u_2 \end{aligned}$$

Then

$$\begin{aligned} Q &= \frac{\alpha_{10}\alpha_{21}-\alpha_{20}\alpha_{11}}{\alpha_{21}-\alpha_{11}} + \frac{\alpha_{12}\alpha_{21}}{\alpha_{21}-\alpha_{11}}y + v_1 \\ p &= \frac{\alpha_{10}-\alpha_{20}}{\alpha_{21}-\alpha_{11}} + \frac{\alpha_{12}}{\alpha_{21}-\alpha_{11}}y + v_2 \end{aligned}$$

which we can write as

$$\begin{aligned} Q &= \Pi_1 + \Pi_2 y + v_1 \\ p &= \Pi_3 + \Pi_4 y + v_2 \end{aligned}$$

Now

$$\alpha_{21} = \frac{\Pi_2}{\Pi_4}$$

and

$$\alpha_{20} = \Pi_1 - \alpha_{21}\Pi_3$$

but we cannot get estimates of the other parameter $\alpha_{10};\alpha_{11};\alpha_{12}$

That is the demand function is NOT IDENTIFIED though the supply one is

Now consider these equations and take

$Q_t^D = \alpha_{10} + \alpha_{11}p + u_1$ identified

$Q_t^S = \alpha_{20} + \alpha_{21}p + \alpha_{22}R + u_2$ not identified

and

$Q_t^D = \alpha_{10} + \alpha_{11}p + \alpha_{12}y + \alpha_{13}R + u_1$ under identified

$Q_t^S = \alpha_{20} + \alpha_{21}p + u_2$ over identified

in this case we can get two estimates of $\alpha_{21}$from the reduced form but we get no estimates of the parameters of the demand function.

We can work this out using a simple rule: order condition: necessary but not sufficient

If:
G=number of endogenous variable
K=number of variables missing from the equation (both endogenous and exogenous)

IDENTIFIED $K = G - 1$

OVERIDENTIFIED $K > G - 1$

UNDERIDENTIFIED $K < G - 1$

So going back to the demand function

$Q_t^D = \alpha_{10} + \alpha_{11}p + u_1$ identified $G = 2; K = 1$

$Q_t^S = \alpha_{20} + \alpha_{21}p + \alpha_{22}R + u_2$ not identified $G = 2; K = 0$

and

$Q_t^D = \alpha_{10} + \alpha_{11}p + \alpha_{12}y + \alpha_{13}R + u_1$ under identified $G = 2; K = 0$

$Q_t^S = \alpha_{20} + \alpha_{21}p + u_2$ over identified $G = 2; K = 2$

An alternative way of looking at the identification problem is to see whether an equation in a system can be obtained as a linear combination of all other equations. Consider

$$Q_t^D = \alpha_{10} + \alpha_{11}p + \alpha_{12}y + u_1$$
$$Q_t^S = \alpha_{20} + \alpha_{21}p + u_2$$

A weighted average:

$$\begin{aligned} Q_t^D &= w(\alpha_{10} + \alpha_{11}p + \alpha_{12}y) + (1 - w)(\alpha_{20} + \alpha_{21}p) \\ &= \alpha_0^* + \alpha_1^* p + \alpha_2^* y \end{aligned}$$

which looks like a demand equation.

So if we estimate an equation of this form we don’t know if we get the demand equation or a weighted sum of both equations

So demand is not identified

but this is not true of supply as would need $\alpha_2^* = 0$ and as $\alpha_{12} \neq 0$ then would need $w = 0$ otherwise you will get y as well as p

For more than 2 equations need more systematic way of dealing with this. We look at what is called the rank condition which is both a necessary and a sufficient condition for identification.

Consider 3 equations, 3 endogenous variables y_i and 3 exogenous variables x_i

$$
\begin{aligned}
y_1 &= \alpha_{11} + \alpha_{12}y_2 + \alpha_{13}y_3 + \alpha_{14}x_1 + \alpha_{15}x_2 + \alpha_{16}x_3 + u_1 \\
y_2 &= \alpha_{21} + \alpha_{22}y_1 + \alpha_{23}y_3 + \alpha_{24}x_1 + \alpha_{25}x_2 + \alpha_{26}x_3 + u_2 \\
y_3 &= \alpha_{31} + \alpha_{32}y_1 + \alpha_{33}y_2 + \alpha_{34}x_1 + \alpha_{35}x_2 + \alpha_{36}x_3 + u_3
\end{aligned}
$$

Now if we write in the following form

$$
\begin{aligned}
y_1 - [\alpha_{11} + \alpha_{12}y_2 + \alpha_{13}y_3 + \alpha_{14}x_1 + \alpha_{15}x_2 + \alpha_{16}x_3 + u_1] &= 0 \\
y_2 - [\alpha_{21} + \alpha_{22}y_1 + \alpha_{23}y_3 + \alpha_{24}x_1 + \alpha_{25}x_2 + \alpha_{26}x_3 + u_2] &= 0 \\
y_3 - [\alpha_{31} + \alpha_{32}y_1 + \alpha_{33}y_2 + \alpha_{34}x_1 + \alpha_{35}x_2 + \alpha_{36}x_3 + u_3] &= 0
\end{aligned}
$$

and then leave out some of the variables in each equation as follows:

equation	y_1	y_2	y_3	x_1	x_2	x_3
1	✓	0	✓	✓	0	✓
2	✓	0	0	✓	0	✓
3	0	✓	✓	✓	✓	0

1. Choose a row
2. Take the columns that have excluded variables and look at the elements in the other rows
3. If the number of both rows and columns, not including the chosen one, that are not all zero is $G-1$ (the number of endogenous variables -1) or more then the equation is identified
4. Otherwise it is not

By the order condition:

Eq		K		$G-1$	*Verdict*
1	No. missing	2	=	2	identified
2	No. missing	3	>	2	overidentified
3	No. missing	2	=	2	identified

By the rank condition:

Eq			No. non zero.	*Verdict*
1	0	0	1 row	Not identified
	✓	✓	2 columns	

Eq				No. non zero.	*Verdict*
2	0	✓	0	2 rows	Identified
	✓	✓	✓	3 columns	

Eq			No. non zero.	*Verdict*
3	✓	✓	2 rows	Identified
	✓	✓	2 columns	

Which gives a different answer to the order condition

The rank condition states whether an equation is identified or not, the order condition lets us know if it is over or underidentified

So the order is necessary, but not sufficient.

With simultaneous equation systems, given the problems with OLS we can use two methods for estimation:

- Single equation or ’limited information’ methods (LIML -TSLS, IV) where estimate each equation separately using only information on restrictions on coefficients of that particular equation
- System or ”Full Information Maximum Likelihood” methods (FIML, 3SLS, SUR) which use all info on all restrictions including cross equation restriction and information on the variances and covariances of all equations.