Elements of Econometrics. Lecture 13. Simultaneous Equations

FCS, 2022-2023

Example: *IS-LM* model

1) $Y_t = C_t + I_t + G_t + X$

2) $C_t = \alpha + \beta \cdot Y_t (+u_t)$

3) $I_t = \delta + \nu \cdot Y_t + \varepsilon \cdot R_t (+v_t)$

4) $X_t = \rho + \sigma \cdot Y_t + \tau \cdot ER_{t-1} (+w_t)$

5) $M_t = \lambda + \mu \cdot Y_t + \theta \cdot R_t (+z_t)$

Endogenous variables: $Y_t,\ C_t,\ I_t,\ X_t,\ R_t$

Exogenous variables: $G_t,\ ER_{t-1},\ M_t$

SINULTANEUS EQUATIONS:
ASSUMPTION B7 VIOLATION:

1. Reasons: Endogeneity, dependence of explanatory variable(s) and disturbance term: disturbance term → dependent variable → endogenous regressor(s) through other equation(s)

2. Consequences: Simultaneous Equations Bias: biased and inconsistent OLS estimators, standard statistics wrongly calculated, tests invalid.

3. Detection: Durbin-Wu-Hausman test

4. Remedial measures: Instrumental Variables.
Two Stage Least Squares.

SIMULTANEOUS EQUATIONS:

STRUCTURAL FORM

Economic models often include several relationships to hold simultaneously.
If each equation is estimated separately by OLS, the B7 (Gauss - Markov 4) condition is violated which leads to biased and inconsistent estimates (Simultaneous Equations Bias).

Example: Macroeconomic Equilibrium model with Consumption function and Income Identity (time index *t* skipped):

$$C = \alpha + \beta Y + u \quad (1)$$

$$Y = C + I \quad (2)$$

C **is aggregate consumption,** ***Y*** **is aggregate income (endogenous variables),** ***I*** **is aggregate investment (exogenous), and** ***u*** **is the disturbance term.**

The estimation of (1) will be considered. Since u is included in C, and C is included in Y, hence Y and u are related, and B7 assumption is violated.

SIMULTANEOUS EQUATIONS:

REDUCED FORM

System (1)-(2) is called Structural Form of the model. If we express Y and C through exogenous variable(s), parameters and disturbance term, we get the Reduced Form (3)-(4):

$$Y = \frac{\alpha}{1-\beta} + \frac{I}{1-\beta} + \frac{u}{1-\beta} \qquad (3)$$

$$C = \frac{\alpha}{1-\beta} + \frac{\beta}{1-\beta} \cdot I + \frac{u}{1-\beta} \qquad (4)$$

In the Reduced Form the variable I (the only one on the right side) is exogenous and hence independent from u. The B7 assumption is satisfied here.

The equations of the Reduced Form can be used for the Instruments' construction in the TSLS estimation.

SIMULTANEOUS EQUATIONS: LARGE SAMPLE BIAS

The OLS estimator of the slope coefficient in (1) is:

$$C = \alpha + \beta Y + u \quad (1)$$

$$\hat{\beta} = \frac{\widehat{\mathrm{Cov}}\,(Y, C)}{\widehat{\mathrm{Var}}\,(Y)} = \beta + \frac{\widehat{\mathrm{Cov}}(Y, u)}{\widehat{\mathrm{Var}}(Y)}$$

$$\mathrm{plim}\left(\frac{\widehat{\mathrm{Cov}}\,(Y, C)}{\widehat{\mathrm{Var}}\,(Y)}\right) = \beta + \mathrm{plim}\left(\frac{\widehat{\mathrm{Cov}}(Y, u)}{\widehat{\mathrm{Var}}(Y)}\right) = \beta + \frac{\mathrm{plim}\widehat{\mathrm{Cov}}(Y, u)}{\mathrm{plim}\widehat{\mathrm{Var}}(Y)}) = \beta + \frac{\sigma_{Y,u}}{\sigma_Y^2}$$

SIMULTANEOUS EQUATIONS: LARGE SAMPLE BIAS

$$\sigma_{Y,u} = \mathrm{Cov}\{(\frac{\alpha}{1-\beta} + \frac{1}{1-\beta} \cdot I + \frac{u}{1-\beta}), u\} = \mathrm{Cov}(\frac{u}{1-\beta}, u) = \frac{1}{1-\beta}\sigma_u^2$$

$$\sigma_Y^2 = \mathrm{Var}(\frac{\alpha}{1-\beta} + \frac{1}{1-\beta} \cdot I + \frac{u}{1-\beta}) = \mathrm{Var}(\frac{1}{1-\beta} \cdot I + \frac{u}{1-\beta}) = \frac{1}{(1-\beta)^2}(\sigma_I^2 + \sigma_u^2)$$

$$\mathrm{plim}\hat{\beta} = \beta + (1-\beta) \cdot \frac{\sigma_u^2}{\sigma_I^2 + \sigma_u^2}$$ **i.e. in large samples the estimator is biased.**

The value of large sample bias is specific for particular equation in the particular model.

SIMULTANEOUS EQUATIONS: INDIRECT LEAST SQUARES

The method consists in the following steps:

1) Deriving the reduced form equations;
2) Estimating the reduced form equations;
3) Expressing the parameters of the structural form equations in terms of the reduced form coefficients estimated. For this, it is necessary to solve the system of equations. **Those coefficients of the structural form which can be derived, turn to be consistent.**

Problems with ILS:

1) If the number of endogenous variables in the model exceeds the number of equations, it is impossible to express some of them in terms of exogenous variables.
2) It is necessary to solve a system of equations which may be complicated.

The ILS method gives better understanding of the simultaneous equations problem though in practice the TSLS method is usually applied to solve it.

DEMAND AND SUPPLY MODEL

$$y_t = \alpha + \beta \cdot p_t + \gamma \cdot x_t + u_t^d,$$ **Demand function**

$$y_t = \delta + \varepsilon \cdot p_t + u_t^s,$$ **Supply function**

The variables y_t (quantity) and p_t (price) are endogenous; the variable x_t (income) is exogenous.

The reduced form is:

$$p = \frac{\alpha - \delta}{\varepsilon - \beta} + \frac{\gamma}{\varepsilon - \beta} \cdot x + \frac{u_d - u_s}{\varepsilon - \beta}$$ **i.e.** $$p = \alpha' + \beta' \cdot x + \nu_p$$

$$y = \frac{\alpha \cdot \varepsilon - \beta \cdot \delta}{\varepsilon - \beta} + \frac{\gamma \cdot \varepsilon}{\varepsilon - \beta} \cdot x + \frac{\varepsilon \cdot u_d - \beta \cdot u_s}{\varepsilon - \beta}$$ **i.e.** $$y = \delta' + \varepsilon' \cdot x + \nu_y$$

There is one potential instrument for the endogenous variable p_t in the Supply function (x_t), and no potential instruments for the Demand function.

IDENTIFICATION

The concept of identification to be applied to a particular equation, not to the whole Simultaneous Equations Model.

Generally, an equation in Simultaneous Equations Model is exactly identified if the number of potential instruments is equal to the number of endogenous variables to be instrumented.

An equation in Simultaneous Equations Model is underidentified if no set of consistent estimators can be provided (generally, the number of potential instruments is less than the number of endogenous variables to be instrumented).

An equation in Simultaneous Equations Model is overidentified if (generally) the number of potential instruments is greater than the number of endogenous variables to be instrumented. TSLS to be applied.

IDENTIFICATION: ORDER CONDITION.

Let the numbers of equations and endogenous variables in a Simultaneous Equations model be equal to G.

Suppose that j endogenous variables are missing from the equation. Then ($G - j - 1$) are available on the right side, and at least (G–1–j) instruments are needed. So the minimum number of variables missing from the equation is

$(j+(G-1-j))=G - 1.$

Order condition for identification: equation is likely to be identified if (G–1) or more variables are missing from it.

If exactly G–1 variables are missing, the equation is likely to be exactly identified.

If more than G–1 variables are missing, it is likely to be overidentified, and TSLS to be used.

Example: *IS-LM* model

$$1)\ Y_t = C_t + I_t + G_t + X$$

$$2)\ C_t = \alpha + \beta \cdot Y_t (+u_t)$$

$$3)\ I_t = \delta + \nu \cdot Y_t + \varepsilon \cdot R_t (+v_t)$$

$$4)\ X_t = \rho + \sigma \cdot Y_t + \tau \cdot ER_{t-1} (+w_t)$$

$$5)\ M_t = \lambda + \mu \cdot Y_t + \theta \cdot R_t (+z_t)$$

Endogenous variables: $Y_t,\ C_t,\ I_t,\ X_t,\ R_t$

Exogenous variables: $G_t,\ ER_{t-1},\ M_t$

In this model the equations (2)-(4) are overidentified, as the number of exogenous variables missing from each equations exceeds the number of endogenous variables appearing on its right side.

Order condition: *G* – 1= 4, and we check if the number of variables missing equals to or exceeds 4.

Equation (5) to be transformed to make an endogenous variable the dependent one.

TWO STAGE LEAST SQUARES (TSLS)

$$y_t = \alpha + \beta \cdot p_t + \gamma \cdot x_t + \rho \cdot t + u_t^d,$$ **Demand function**

$$y_t = \delta + \varepsilon \cdot p_t + u_t^s$$ **Supply function**

There are two potential instrumental variables for p_t in the Supply function with the variable t (time) added: variables x_t and t, as well as their linear combinations:

$$z_t = h_0 + h_1 \cdot x_t + h_2 \cdot t$$

The idea of TSLS: to select the linear combination which is the most strongly correlated with p_t from all possible linear combinations of potential instruments (exoneous variables), which is the theoretical value of p_t from the regression:

$$\hat{p}_t = a' + b' \cdot x_t + c' \cdot t \qquad r_{\hat{p},p}^2 = R^2 = 1 - \frac{RSS}{TSS_p} \to \max$$

TWO STAGE LEAST SQUARES (TSLS)

For each identified equation:

Stage 1.

Set the list of potential instruments.

Estimate the regression of each endogenous variable from the right side on the instruments.

Calculate the instruments (theoretical, or predicted, values of explanatory variables).

Stage 2.

Regress the dependent variable on the instruments.

TSLS in EViews, for the Consumption Function (2) in ISLM:

$$C_t = \alpha + \beta \cdot Y_t (+u_t)$$

```
tsls cons c y @ c m g  er(-1)
```

TWO STAGE LEAST SQUARES FOR THE MODEL (1)-(2)

$$C = \alpha + \beta Y + u \quad (1)$$

$$Y = C + I \quad (2)$$

The Equation 1 provides biased and inconsistent OLS estimators of β and α.

Then the Instrument for Y is needed. It is formed at the Stage 1:

$$\hat{Y} = g_0 + g_1 \cdot I \quad (Stage\ 1)$$

$$\hat{\beta}_{TSLS} = \frac{\widehat{\mathrm{Cov}}\ (\hat{Y}, C)}{\widehat{\mathrm{Cov}}(\hat{Y}, Y)} = \frac{\widehat{\mathrm{Cov}}(g_1 I, C)}{\widehat{\mathrm{Cov}}(g_1 I, Y)} = \frac{\widehat{\mathrm{Cov}}(I, C)}{\widehat{\mathrm{Cov}}(I, Y)} = \hat{\beta}_{IV} \quad (Stage\ 2)$$

$$\widehat{\mathrm{Cov}}(\hat{Y}, Y) = \widehat{\mathrm{Cov}}(\hat{Y}, \hat{Y} + \hat{u}) = \widehat{\mathrm{Var}}(\hat{Y}) + \widehat{\mathrm{Cov}}(\hat{Y}, \hat{u})$$
$$= \widehat{\mathrm{Var}}\ (\hat{Y}) + 0 = \widehat{\mathrm{Var}}(\hat{Y})$$
$$\Downarrow$$
$$\hat{\beta}_{TSLS} = \frac{\widehat{\mathrm{Cov}}(\hat{Y}, C)}{\widehat{\mathrm{Var}}(\hat{Y})}$$

So, IV and TSLS give the same estimates in this model (for an equation which is exactly identified). You may regress *C* on $\hat{Y}$

TSLS APPLICATION EXAMPLE

Estimate α_2 *in (2) in the model (for EAEF40):*

$$S=\beta_1+\beta_2 ASVABC+\beta_3 SM+\beta_4 SF+u \qquad (1)$$

$$ASVABC=\alpha_1+\alpha_2 S+v \qquad (2)$$

Instruments for S: SM, SF

Command in EViews:

tsls asvabc c s @ c sm sf

Test: DWH or its Davidson&McKinnon version

OLS and TSLS ESTIMATIONS:

Dependent Variable: ASVABC   Method: Least Squares   Included observations: 570

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	23.03563	1.847766	12.46675	0.0000
S	2.003304	0.133468	15.00960	0.0000

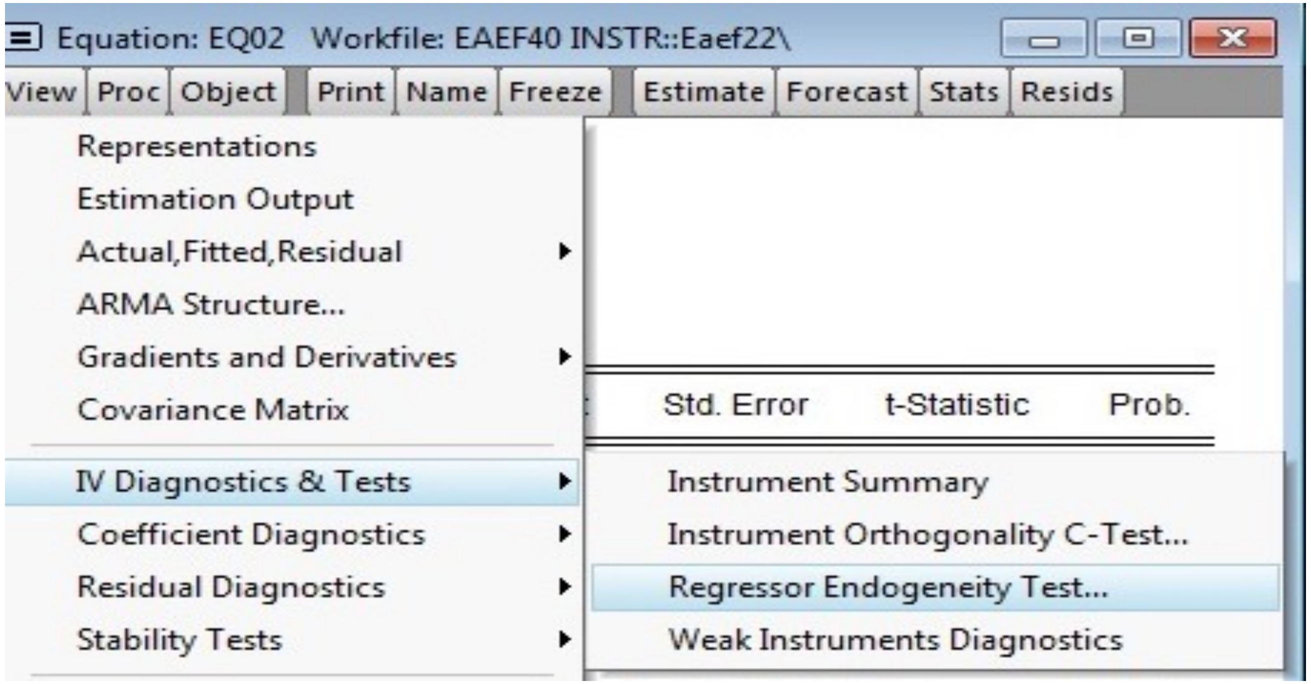
Dependent Variable: ASVABC   Method: Two-Stage Least Squares   Included observations: 570
Instrument list: C SM SF

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.217455	4.801296	0.461845	0.6444
S	3.529327	0.351016	10.05460	0.0000

In Eviews there is the DWH test for regressors' endogeneity:

TSLS estimation → View → IV diagnortics and tests → Regressor endogeneity test, providing with the p-values for the test statistics. Under H_0 of exogeneity, the test statistic has a $\chi 2$ (chi-squared) distribution with degrees of freedom equal to the number of regressors tested for endogeneity.

Durbin-Wu-Hausman test (Eviews)



Endogeneity Test

Endogenous variables to treat as exogenous: S
Specification: ASVABC C S
Instrument specification: C SM SF
Null hypothesis: S are exogenous

	Value	df	Probability
DWH Statistics	28.27992	1	0.0000

DWH test (Davidson-McKinnon)

Dependent Variable: S   Method: Least Squares Included observations: 570

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.230042	0.432887	21.32206	0.0000
SM	0.190291	0.043133	4.411751	0.000
SF	0.182065	0.031782	5.728623	0.0000

Dependent Variable: ASVABC   Method: Least Squares Included observations: 570

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.217455	4.223490	0.525029	0.5998
S	3.529327	0.308774	11.43015	0.0000
RES1	-1.856144	0.340538	-5.450625	0.0000

Since RES1 is significant, the H_o hypothesis of exogeneity of S is rejected