Lecture 15

Model 3:

$$\begin{cases} Q = \alpha_0 + \alpha_1 P + \alpha_2 Y + u & \text{(D)} \\ Q = \beta_0 + \beta_1 P + \beta_2 r + v & \text{(S)} \end{cases}$$

- Endogenous: Q and P.
- Exogenous: Y, r and constants.

Reduced form (for P): $P = \lambda_0 + \lambda_1 Y + \lambda_2 r + \varepsilon_1$.

- $\lambda_0 = \dfrac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}$.
- $\lambda_1 = -\dfrac{\alpha_2}{\alpha_1 - \beta_1}$.
- $\lambda_2 = \dfrac{\beta_2}{\alpha_1 - \beta_1}$.
- $\varepsilon_1 = \dfrac{1}{\alpha_1 - \beta_1}(u - v)$.

Reduced form (for Q): $Q = \gamma_0 + \gamma_1 Y + \gamma_2 r + \varepsilon_2$.

- $\gamma_0 = \alpha_0 + \alpha_1 \lambda_0$.
- $\gamma_1 = \alpha_1 \lambda_1 + \alpha_2$.
- $\gamma_2 = \alpha_1 \lambda_2$.

Now, we estimate the reduced-form equations:

- That produces: $\hat{\lambda}_0, \hat{\lambda}_1, \hat{\lambda}_2, \hat{\gamma}_0, \hat{\gamma}_1, \hat{\gamma}_2$.
- Can we recover $\alpha_0, \alpha_1, \alpha_2, \beta_0, \beta_1, \beta_2$
    - $\gamma_2 = \alpha_1 \lambda_2 \Rightarrow \hat{\alpha}_1 = \dfrac{\hat{\gamma}_2}{\hat{\lambda}_2}$.
    - $\gamma_0 = \alpha_0 + \alpha_1 \lambda_0 \Rightarrow \hat{\alpha}_0 = \hat{\gamma}_0 - \hat{\alpha}_1 \hat{\lambda}_0$.
    - $\gamma_1 = \alpha_1 \lambda_1 + \alpha_2 \Rightarrow \hat{\alpha}_2 = \hat{\gamma}_1 - \hat{\alpha}_1 \hat{\lambda}_1$.
    - $\lambda_1 = -\dfrac{\alpha_2}{\alpha_1 - \beta_1} \Rightarrow \hat{\beta}_1 = -\dfrac{\hat{\alpha}_2 - \hat{\alpha}_1 \hat{\lambda}_1}{\hat{\lambda}_1}$.
    - $\lambda_0 = \dfrac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} \Rightarrow \hat{\beta}_0 = \hat{\alpha}_0 + \hat{\lambda}_0(\hat{\alpha}_1 - \hat{\beta}_1)$.
    - $\lambda_2 = \dfrac{\beta_2}{\alpha_1 - \beta_1} \Rightarrow \hat{\beta}_2 = \hat{\lambda}_2(\hat{\alpha}_1 - \hat{\beta}_1)$.
- The model is (exactly) <u>identified</u>.

Model 4:

$$\begin{cases} Q = \alpha_0 + \alpha_1 P + \alpha_2 Y + u & \text{(D)} \\ Q = \beta_0 + \beta_1 P + \beta_2 r + \beta_3 f + v & \text{(S)} \end{cases}$$

- f stands for fertilizer.

Reduced form (for P): $P = \lambda_0 + \lambda_1 Y + \lambda_2 r + \lambda_3 f + \varepsilon_1$.

- $\lambda_0 = \dfrac{\beta_0 - \alpha_0}{\beta_1 - \alpha_1}$.

- $\lambda_1 = -\dfrac{\alpha_2}{\beta_1 - \alpha_1}$.
- $\lambda_2 = \dfrac{\beta_2}{\beta_1 - \alpha_1}$.
- $\lambda_3 = \dfrac{\beta_3}{\beta_1 - \alpha_1}$.

Reduced form (for Q): $Q = \gamma_0 + \gamma_1 Y + \gamma_2 r + \gamma_3 f + \varepsilon_2$.

- $\gamma_0 = \alpha_0 + \alpha_1 \lambda_1$.
- $\gamma_1 = \alpha_1 \lambda_1 + \alpha_2$.
- $\gamma_2 = \alpha_1 \lambda_2$.
- $\gamma_3 = \lambda_3$.

The model has 7 unknowns: $\alpha_0, \alpha_1, \alpha_2, \beta_0, \beta_1, \beta_2, \beta_3$.

- However, there are 8 equations: $\lambda_0, \lambda_1, \lambda_2, \lambda_3, \gamma_0, \gamma_1, \gamma_2, \gamma_3$.
- Consequently, we get two measures for the slope: $\hat{\alpha}_1 = \dfrac{\hat{\gamma}_2}{\hat{\lambda}_2}$ and $\hat{\alpha}_1 = \dfrac{\hat{\gamma}_3}{\hat{\lambda}_3}$.
    - Implicitly, there is a restriction to be imposed $\dfrac{\hat{\gamma}_2}{\hat{\lambda}_2} = \dfrac{\hat{\gamma}_3}{\hat{\lambda}_3}$, but the estimation does not guarantee it.
    - To solve it, we average the solutions (because we assume the mismatch is due to noise in the data).
- The system is over-identified.

Conditions for identification:

Let G be the number of endogenous variables.

- **Order condition**: the number of variables excluded from an equation is greater than or equal to $G - 1$.
    - This is a necessary condition.
        - If it is false, the model is not identified.
        - If it is true, the model *might be* identified.
    - For the aforementioned model 4:
        - The first equation has 2 variables excluded from it: r and f. The order condition is true.
        - The second equation has 1 variable excluded from it: Y. The order condition is also true.
    - If the number of excluded variables is equal to $G - 1$, the equation is exactly identified. If the number of exclusions is bigger, the equation is over-identified.
        - The first equation of model 4 is then over-identified (it has two equations for α_1).

Consider the model:

$$\begin{cases} y_1 = \alpha_0 + \alpha_1 y_2 + \alpha_2 y_3 + \alpha_3 x_1 + \alpha_4 x_2 + u_1 \\ y_2 = \beta_0 + \beta_1 y_3 + \beta_2 x_1 + u_2 \\ y_3 = \gamma_0 + \gamma_1 y_2 + u_3 \end{cases}$$

y_1, y_2, y_3: endogenous ($G = 3$).

x_1, x_2 and constants ($\alpha_0, \beta_0, \gamma_0$): exogenous.

- There is no excluded variable in the y_1 equation ($0 \ngeq G - 1$): it is not identified.
- y_1 and x_2 are missing in the y_2 equation ($2 = G - 1$): it is exactly identified.
- The y_3 equation is missing y_1, y_3, x_1, x_2 ($4 > G - 1$): it is over-identified.

Indirect Least Squares (ILS)

Works **only** if the model is exactly identified.

Consider the model:

$$\begin{cases} C_t = \alpha + \beta Y_t + u_t \\ Y_t = C_t + I_t \end{cases}$$

Y, C: endogenous ($G = 2$).

I, constant (α): exogenous.

- The C_t equation is missing $I_t \to 1 = G - 1 \to$ It is exactly identified.

Reduced form (for C_t): $C_t = \dfrac{\alpha}{1-\beta} + \dfrac{\beta}{1-\beta} I_t + \dfrac{u_t}{1-\beta}$.

Reduced form (for Y_t): $Y_t = \dfrac{\alpha}{1-\beta} + \dfrac{1}{1-\beta} I_t + \dfrac{u_t}{1-\beta}$.

OBS.: The coefficient of I_t in the reduced-form equation of Y_t $\left(\dfrac{1}{1-\beta}\right)$ is a multiplier; it is one divided by one minus the marginal propensity to consume (MPC), the coefficient of Y_t in the structural-form equation of C_t.

- Estimate $C_t = \lambda_0 + \lambda_1 I_t + \varepsilon_t$. That will produce $\hat{\lambda}_0, \hat{\lambda}_1$.
  - $\lambda_1 = \dfrac{\beta}{1-\beta} \Rightarrow \hat{\beta} = \dfrac{\hat{\lambda}_1}{1+\hat{\lambda}_1}$.
  - $\lambda_0 = \dfrac{\alpha}{1-\beta} \Rightarrow \hat{\alpha} = \hat{\lambda}_0(1-\hat{\beta})$.

Instrumental variables

Gives the same answer of ILS when the model is exactly identified and still works when it is not.

- Consider the exactly identified model:

$$\begin{cases} y_1 = \alpha_1 y_2 + \alpha_2 x_1 + u \\ y_2 = \beta_1 y_1 + \beta_2 x_2 + v \end{cases}$$

- Suppose we try OLS in the first equation: what is the problem?
  - y_2 is correlated with u (simultaneity bias): $y_2 = f(y_1), y_1 = g(u) \to y_2 = f(g(u))$.
- We need a variable correlated with y_2 but not with the error term u (exogenous).
  - We can use x_2 as an instrumental variable for y_2 in y_1.

Now, consider the following model:

$$\begin{cases} y_1 = \alpha_1 y_2 + \alpha_2 x_1 + u \\ y_2 = \beta_1 y_1 + \beta_2 x_2 + \beta_3 x_3 + v \end{cases}$$

- The y_1 is over-identified: it is missing x_1 and x_2.
- In the second equation, y_1 is correlated with the error term v: $y_2 = f(y_1), y_1 = g(u) \to y_2 = f \circ g(u)$.
- There are two potential instruments for y_2: x_2 and x_3.
  - We will use a linear combination of the two: $ax_2 + bx_3$ as IV for y_2.
  - This is where two-stage least squares (2SLS).

    - The first stage gives us our instrument.
    - In the second stage, we will use the instrument to do the estimation.

Let us use a different example to illustrate:

$$\begin{cases} Q = \alpha_0 + \alpha_1 P + \alpha_2 Y + u \\ Q = \beta_0 + \beta_1 P + \beta_2 r + \beta_3 f + v \end{cases}$$

Stage 1: Estimate the reduced-form equations.

- Regress each endogenous variable on all the exogenous variables in the system.

$$\begin{cases} \hat{P} = \hat{\lambda}_0 + \hat{\lambda}_1 Y + \hat{\lambda}_2 r + \hat{\lambda}_3 f \\ \hat{Q} = \hat{\gamma}_0 + \hat{\gamma}_1 Y + \hat{\gamma}_2 r + \hat{\gamma}_3 f \end{cases}$$

- These are the IVs.
  - $\hat{P}$ is a linear combination of only exogenous variables, there is no risk of being correlated with an error term. The same applies to $\hat{Q}$.
    - Note: $\hat{P} = P - \text{error term}, \text{error term} = f(u, v)$.

Stage 2: Use $\hat{P}$ and $\hat{Q}$ as IVs.