TA session 5

Mincheol Kim
mincheol.kim.21@siai.org

1 QnA

Question 1. Can we usually do the Hausman Test regardless of types of endogeneity?

Question 2. What is the Hausman test???

Question 3. Why could the sign of the estimate change due to simultaneity and omitted variable unlike measurement error?

2 Recap

2.1 Endogeneity

There may be data scientific reasons why we might expect that the **errors and regressors and correlated.**

$$\mathbb{E}(u_i|x_j) \neq 0 \;\; \forall i, \; j \quad \Longleftrightarrow \quad Cov(u,x) \neq 0$$

It will make the OLS estmator biased and **inconsistent**. Then The regression is useless.

Solutions

(a) Collect good controls that makes the endogenous variable exogenous. Or add additional lag (explanatory) variables

(b) Find one or more instrumental variables to deal with the endogenous explanatory variable.

2.1.1 Omitted Variables in a Simple Regression Model

Omitting a relevant variable causes this correlation between the errors and regressors.

$$\begin{aligned} y &= \beta_1 + x_2\beta_2 + \cdots + x_k\beta_k + \gamma q + v \\ y &= \beta_1 + x_2\beta_2 + \cdots + x_k\beta_k + u, \quad u \equiv \gamma q + v \\ y &= (\beta_1 + \gamma\delta_1) + x_2(\beta_2 + \gamma\delta_2) + \cdots + x_k(\beta_k + \gamma\delta_k) + \gamma r + v, \quad \because q = \delta_1 + \delta_2 x_2 + \cdots + \delta_k x_k + r \end{aligned}$$

2.1.2 Measurement Error

Please refer to DbDM

2.1.3 Simultaneity

Some explanatory varialbes are jointly determined with the dependent variable. These behavioral relations are also called structural equations.

- Structural Form: listing the equations

$$\begin{aligned} q_i &= \alpha_1 p_i + \beta_1 z_i + u_{1i} \\ p_i &= \alpha_2 q_i + u_{2i} \end{aligned}$$

- Bias in OLS

$$\mathbb{E}(pu_1) = \mathbb{E}((\alpha_2 q + u_2)\cdot u_1) = \mathbb{E}(\alpha_2(\alpha_1 p + \beta_1 z + u_1)u_1) = \alpha_2\sigma^2$$

- Reduced Form: Expressing the endogenous variables in terms of exogenous variables and errors only.

2.2 IV estimation in the Multiple Regression Model

Condition

1. Intrument Validity: $\mathbb{E}(\epsilon_i z_i) = 0$
2. Instrument Relevance: $Cov(x_i, z_i) \neq 0$

When we use the IV, the s.e will be more large but coefficient will be consistent.

Question. What are the exact identification and over identification? the number of IVs.
A. We prefer the IV that has high correlation with explanatory variables. We can check this by standard error of estimate

2.3 Two Stage Least Squres(2SLS)

What is the difference from IV? IV is the special case of the 2SLS.

We can also choose the linear combination of IVs that has to be high correlated with explanatory variable. But notice that weak instrument bias tends to get worse as we add more weak instruments.

If you do the second step manually, you will need to correct the standard errors. But most stat packages provide automatic correction. $\hat{x}_i$ is not x_i.

2.3.1 Choice of Instruments

Cases when we need to choose the IVs

1. Lagged endogenous variables and autocorrelation
2. Simultaneous equation models

2.3.2 Properties of OLS with Serially Correlated Errors

Lagged endogenous variables and AR(p) error make endogeneity.

When there is MA(1) error, lagged variables might make endogeneity.

2.4 Hausman test

To check the exogeneity we use the IV. Compare our OLS and IV parameter estimates.

$$H_0 : \text{plim}(\hat{\beta}_{OLS} - \hat{\beta}_{IV}) = 0 \text{ (Exogenous)}$$
$$H_1 : \text{plim}(\hat{\beta}_{OLS} - \hat{\beta}_{IV}) \neq 0 \text{ (Endogenous)}$$
$$\text{t-value} = (\hat{\beta}_{OLS} - \hat{\beta}_{IV})' Var(\hat{\beta}_{OLS} - \hat{\beta}_{IV})^{-1} (\hat{\beta}_{OLS} - \hat{\beta}_{IV}) \sim \chi^2_k$$

	H_0	H_1
OLS	consistent	inconsistent
2SLS	consistent	consistent