MGRECON 548Q: Empirical Economic Analysis

Fuqua School of Business, Duke University
Fall 2020

Help Section Week#4[1]

Why we want instrument variable?:

OVB Problem

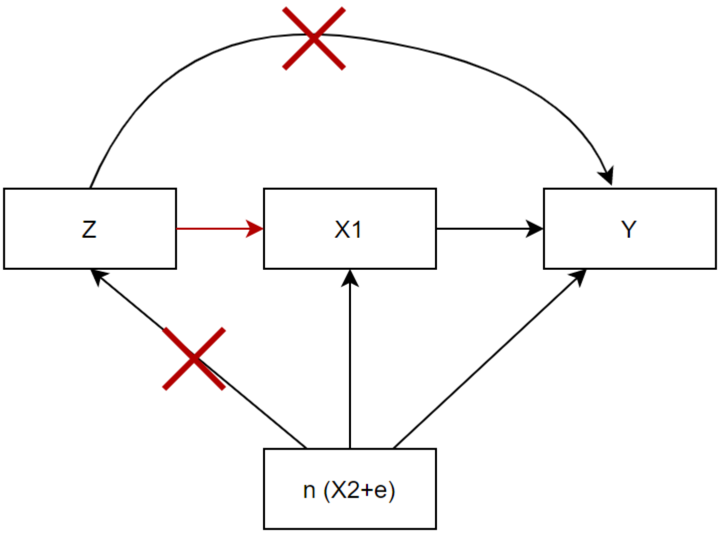
- Consider regression:

$$
\begin{aligned}
&\text{(long)}\ \ \mathrm{Y} = \beta_0^{long} + \beta_1^{long} X_1 + \gamma^{long} X_2 + \epsilon \\
&\text{(short)}\ \mathrm{Y} = \beta_0^{short} + \beta_1^{short} X_1 + \eta \\
\implies &\ \hat{\beta}_1^{\ short} = \hat{\beta}_1^{\ \text{long}} + \hat{\gamma}^{\text{long}} \frac{\operatorname{cov}(\mathrm{X}_1, \mathrm{X}_2)}{\operatorname{var}(\mathrm{X}_1)}
\end{aligned}
$$

- In this case, we are not able to estimate the marginal effect of X_1 on Y:

  If X_1 in crease by one unit, X_2 will change by some unit, thus the change of Y includes both the effect from changes in X_1 and X_2. In order to estimate the marginal effect of X_1 on Y, we want "something" that can change X_1 without affecting other unobserved/ignore variables (errors) or affecting Y directly.

- Consider X_1's instrument variable Z satisfies:



  (i) Relevant: $\mathrm{Cov}(\mathrm{Z}, \mathrm{X}_1) \neq 0$

  (ii) Exogeneity: $\mathrm{Cov}(\mathrm{Z}, \eta) = 0$ and Z should be related to Y only through X

  These assumptions ensure Z to be a valid instrument variable: by (i), we can change the value of X_1 through Z; by (ii), we are able to change X_1 without affecting other variables and the changes of Y is generated purely through the changes on X_1.

[1] Please watch for typos and errors

Variable with Measure Error

- Consider regression:

$$Y = \beta_0 + \beta_1 X^* + \eta$$

where X^* meets all of the standard exogeneity assumptions, however, we do not observe true X^*, instead we observe X:

$$X = X^* + \mu_X$$

where μ_X is the measure error.

We are interested in the marginal effect of X^*, but we can only run regression on X:

$$\begin{aligned} Y &= \beta_0 + \beta_1 X^* + \eta \\ &= \beta_0 + \beta_1 (X - \mu_X) + \eta \\ &= \beta_0 + \beta_1 X + (\eta - \beta_1 \mu_X) \end{aligned}$$

Since X is related to the composite error term $(\eta - \beta_1 \mu_X)$, thus OLS will produce biased estimator of β_1.

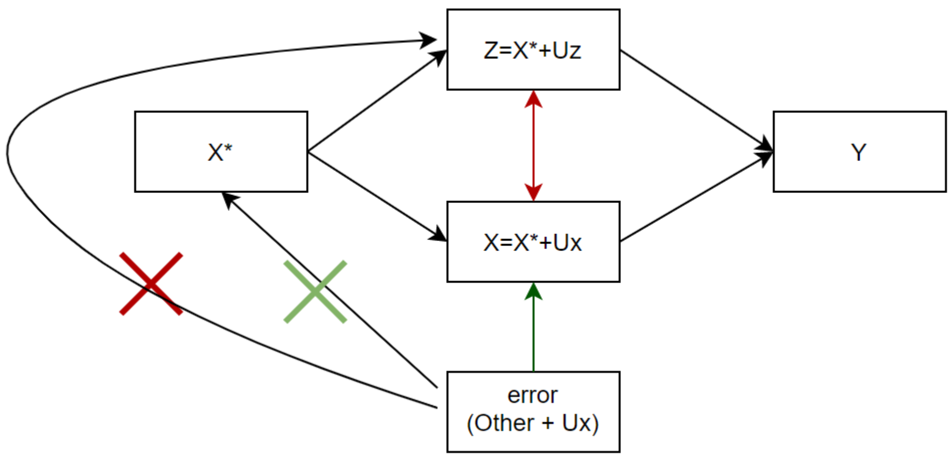
- Suppose Z is another measure of X^* with measure error:

$$Z = X^* + \mu_Z$$

where μ_Z is the measure error and we have $\mathrm{Cov}(\mu_Z, \eta) = 0$ and $\mathrm{Cov}(\mu_Z, \mu_X) = 0$. We can use Z as an instrument if:

  (i) Relevant: $\mathrm{Cov}(Z, X) = \mathrm{Cov}(X^* + \mu_Z, X^* + \mu_X) = \mathrm{Var}(X^*) \neq 0$

  (ii) Exogeneity: $\mathrm{Cov}(Z, \eta) = \mathrm{Cov}(X^* + \mu_Z, \eta) = \mathrm{Cov}(X^*, \eta) + \mathrm{Cov}(\mu_Z, \eta) = 0$



Implement:

- Two Stage Least Squares(2SLS)

$$\begin{aligned}
&\text{(First stage)}\ \ \mathrm{X}_1 = \pi_0 + \pi_1 Z + \mu \\
&\text{(Second stage)}\ \mathrm{Y} = \beta_0 + \beta_1 \hat{X}_1 + \eta
\end{aligned}$$

The estimator of 2SLS:

$$\begin{aligned}
\mathrm{Cov}(\mathrm{Y},\mathrm{Z}) &= \mathrm{Cov}(\beta_0^{\mathrm{long}} + \beta_1^{\mathrm{long}}\mathrm{X}_1 + \gamma^{\mathrm{long}}\mathrm{X}_2 + \epsilon, \mathrm{Z}) \\
&= \mathrm{Cov}(\beta_0^{\mathrm{long}}, \mathrm{Z}) + \mathrm{Cov}(\beta_1^{\mathrm{long}}\mathrm{X}_1, \mathrm{Z}) + \mathrm{Cov}(\gamma^{\mathrm{long}}\mathrm{X}_2, \mathrm{Z}) + \mathrm{Cov}(\epsilon, \mathrm{Z}) \\
&= 0 + \beta_1^{\mathrm{long}}\mathrm{Cov}(\mathrm{X}_1, \mathrm{Z}) + \mathrm{Cov}(\eta, \mathrm{Z}) \\
\Longrightarrow\ \beta_1^{\mathrm{long,IV}} &= \frac{\mathrm{Cov}(\mathrm{Y},\mathrm{Z})/\mathrm{Var}(\mathrm{Z})}{\mathrm{Cov}(\mathrm{X}_1,\mathrm{Z})/\mathrm{Var}(\mathrm{Z})} + \frac{\mathrm{Cov}(\eta,\mathrm{Z})}{\mathrm{Cov}(\mathrm{X}_1,\mathrm{Z})} = \beta_1^{\mathrm{long}} + \frac{\mathrm{Cov}(\eta,\mathrm{Z})}{\mathrm{Cov}(\mathrm{X}_1,\mathrm{Z})} \\
&= \beta_1^{\mathrm{long}}\ \ (\text{if Z is exogenous, i.e. } \mathrm{Cov}(\eta,\mathrm{Z}) = 0)
\end{aligned}$$

- Compare IV and OLS:

  (i) Coefficient β_1^{IV} with β_1^{OLS}[2]:

$$\begin{aligned}
\beta_1^{\mathrm{OLS}} &= \beta_1^{\mathrm{long}} + \frac{\mathrm{Cov}(\eta,\mathrm{X}_1)}{\mathrm{Var}(\mathrm{X}_1)} = \beta_1^{\mathrm{long}} + \mathrm{Corr}(\mathrm{X}_1,\eta) \times \frac{\sigma_\eta}{\sigma_{\mathrm{X}_1}} \\
\beta_1^{\mathrm{IV}} &= \beta_1^{\mathrm{long}} + \frac{\mathrm{Cov}(\eta,\mathrm{Z})}{\mathrm{Cov}(\mathrm{X}_1,\mathrm{Z})} = \beta_1^{\mathrm{long}} + \frac{\mathrm{Corr}(\mathrm{Z},\eta)}{\mathrm{Corr}(\mathrm{Z},\mathrm{X}_1)} \times \frac{\sigma_\eta}{\sigma_{\mathrm{X}_1}}
\end{aligned}$$

  Suppose the Exogeneity assumption is violated (i.e. $\mathrm{Cov}(\mathrm{Z},\eta) \neq 0$), if the instrument is "weak" (i.e. $\mathrm{Corr}(\mathrm{Z},\mathrm{X}_1)$ is small), the bias in the IV estimator will be large.

  (ii) Variance $\mathrm{Var}(\beta_1^{\mathrm{OLS}})$ and $\mathrm{Var}(\beta_1^{\mathrm{IV}})$[3]:

$$\begin{aligned}
\mathrm{Var}(\beta_1^{\mathrm{OLS}}) &= \frac{\sigma_\eta^2}{\sum_{\mathrm{i}=1}^{\mathrm{N}}(\mathrm{X}_{1\mathrm{i}} - \bar{\mathrm{X}}_1)^2} = \frac{\sigma_\eta^2}{\mathrm{SST}_{\mathrm{X}_1}} \\
\mathrm{Var}(\beta_1^{\mathrm{IV}}) &= \frac{\sigma_\eta^2}{\mathrm{SST}_{\mathrm{X}_1} * \mathrm{R}^2_{\mathrm{X}_1,\mathrm{Z}}}
\end{aligned}$$

  $\mathrm{SST}_{\mathrm{X}_1}$ is the sum of residual $\mathrm{R}^2_{\mathrm{X}_1,\mathrm{Z}}$ comes from an auxiliary regression of X_1 upon Z and captures the relationship instrument relevance (first stage). The variance of the IV estimator is (almost) always larger than that of OLS, suggesting that we should run OLS when possible.

[2] Estimate from : $\mathrm{Y} = \beta_0^{\mathrm{short}} + \beta_1^{\mathrm{short}}\mathrm{X}_1 + \eta$

[3] $\mathrm{Var}(\beta_1^{\mathrm{IV}}) = \sigma_\eta^2(\mathrm{X}_1'\mathrm{P}_\mathrm{Z}\mathrm{X}_1)^{-1}$

Testing:

Test for Exogeneity

- IV might generate estimators with large standard errors, so we should not apply IV unless necessary. Consider regression:

$$\text{(short) } \mathrm{Y} = \beta_0^{\text{short}} + \beta_1^{\text{short}}\mathrm{X}_1 + \lambda^{\text{short}}\text{Exogenous} + \eta$$

  If we have $\mathrm{Cov}(\mathrm{X}_1, \eta) = 0 \implies$ no endogeneity problem (indicating there is no OVB) $\implies$ OLS can give us an unbiased estimator $\implies$ we should estimate the short via OLS.

- (i) A quick diagnostic is to estimate via OLS and via IV. If they are similar,then you may not have an endogeneity problem

  (ii) A formal rest – Hausman test

$$\begin{aligned} \mathrm{X}_1 &= \pi_0 + \pi_1\mathrm{Z} + \pi_2\text{Exogenous} + \mathrm{v}_1 \\ \implies \mathrm{Cov}(\mathrm{X}_1, \eta) &= \mathrm{Cov}(\pi_0 + \pi_1\mathrm{Z} + \pi_2\text{Exogenous} + \mathrm{v}_1, \eta) = \mathrm{Cov}(\mathrm{v}_1, \eta) \\ \eta &= \delta\mathrm{v}_1 + \mathrm{v}_2 \implies \mathrm{Cov}(\mathrm{X}_1, \eta) = 0 \text{ if } \delta = 0 \\ \mathrm{Y} &= \beta_0^{\text{short}} + \beta_1^{\text{short}}\mathrm{X}_1 + \lambda^{\text{short}}\text{Exogenous} + \delta\hat{\mathrm{v}}_1 + \text{error} \end{aligned}$$

  $\mathrm{H}_0 : \delta = 0$(Exogeneity) v.s. $\mathrm{H}_0 : \delta \neq 0$(Endogeneity).

Weak Instruments Test

- Consider the auxiliary regression:

$$\mathrm{X}_1 = \pi_0 + \pi_1\mathrm{Z} + \pi_2\text{Exogenous} + \text{error}$$

  We want $\mathrm{Cov}(\mathrm{X}_1, \mathrm{Z}) \neq 0$. Since $\pi_1 \approx \frac{\mathrm{Cov}(\mathrm{X}_1,\mathrm{Z})}{\mathrm{Var}(\mathrm{Z})}$ so Z is relevant if and only if $\pi_1 \neq 0$. $\mathrm{H}_0 : \pi_1 = 0$(Weak) v.s. $\mathrm{H}_0 : \pi_1 \neq 0$(Relevant).