$y = X\beta + \varepsilon$

i can use $\hat{\beta}_{OLS} = (X'X)^{-1}X'y$

or $\hat{\beta}_{TSLS} = (\hat{X}'\hat{X})^{-1}\hat{X}'y$

base on the 2 potential estimator

$\mathrm{var}\,\hat{\beta}_{OLS} = \hat{\sigma}^2_\varepsilon (X'X)^{-1}$

$\mathrm{var}\,\hat{\beta}_{TSLS} = \hat{\sigma}^2_\varepsilon (\hat{X}'\hat{X})^{-1}$

Ideally i want to compare them

$(\beta_{IV} - \hat{\beta}_{OLS})$ the test is based on this difference

$H_0: \hat{\beta}_{IV} - \hat{\beta}_{OLS} = 0$

$H_1: \hat{\beta}_{IV} - \hat{\beta}_{OLS} \neq 0$

ENDOGENEITY TEST

if the difference is $= 0$ then there is not exogeneity

HAUSMAN

$$\underbrace{(\hat{\beta}_{TSLS} - \hat{\beta}_{OLS})'}_{1\times k}\Big[\underbrace{\mathrm{AVAR}(\hat{\beta}_{TSLS} - \hat{\beta}_{OLS})}_{k\times k}\Big]^{-1}\underbrace{(\hat{\beta}_{TSLS} - \hat{\beta}_{OLS})}_{k\times 1}$$

AVAR: Asymptotic

this converges in distribution to $\chi^2(k_1)$

k_1: only the coefficients associated with endogenous variables change

HAUSMAN RESULT FOR THE ASYMPTOTIC VARIANCE OF THE DIFFERENCE

$$Cov(\hat{\beta}_{OLS}, \hat{\beta}_{IV} - \hat{\beta}_{OLS}) = 0$$

$\hat{\beta}_{TSLS}$ → robust and inefficient if compared to OLS that for Gauss Markov is the most efficient → but it is robust

$\hat{\beta}_{OLS}$ → is efficient → it is the best under the null but it is not robust under endogeneity

$$Cov(\hat{\beta}_{OLS}, \hat{\beta}_{IV} - \hat{\beta}_{OLS}) = 0$$

$$Cov(\hat{\beta}_{OLS}, \hat{\beta}_{IV}) - Var(\hat{\beta}_{OLS}) = 0$$

$$Cov(\hat{\beta}_{OLS}, \hat{\beta}_{IV}) = Var(\hat{\beta}_{OLS})$$

variance of the most efficient estimator

Now $AVAR(\hat{\beta}_{TSLS} - \hat{\beta}_{OLS})$

$$= AVAR(\hat{\beta}_{IV}) + AVAR(\hat{\beta}_{OLS}) - 2Cov(\hat{\beta}_{IV}, \hat{\beta}_{OLS})$$

$$= AVAR\,\hat{\beta}_{IV} + AVAR(\hat{\beta}_{OLS}) - 2AVAR(\hat{\beta}_{OLS})$$

$$= AVAR\,\hat{\beta}_{IV} - AVAR(\hat{\beta}_{OLS})$$

$$H = \underbrace{(\hat{\beta}_{IV} - \hat{\beta}_{OLS})'}_{\text{the problem is that a lot of } \beta \text{ does not change, the exogenous variable bring to the same estimates in iv and ols}} \underbrace{\left[\hat{\sigma}^2_\varepsilon(\hat{X}'\hat{X})^{-1} - \hat{\sigma}^2_\varepsilon(X'X)^{-1}\right]^{-1}}_{\text{this quantity is positive by Gauss Markov because } \hat{\beta}_{OLS} \text{ is efficient}} (\hat{\beta}_{IV} - \hat{\beta}_{OLS})$$

CONTROL REGRESSION

$$\begin{cases} y = x_1\beta_1 + x_2\beta_2 + \varepsilon \\ x_1 = z_1\Gamma_{11} + x_2\Gamma_{21} + u \\ x_2 = x_2 \end{cases}$$

we have seen strategy number 1

strategy n. 2 is estimating the following regression

$$y = x_1\beta_1 + x_2\beta_2 + \alpha\hat{u} + \varepsilon$$

i add extra control that is the residual from the first equation