Econometrics II TA Session #6

Royou Kiku

December 1st, 2021

Contents

1.Definition of Hausman's Specification Error (特定化誤差) Test

- OLS VS IV

$$H_0: E(u|X) = 0$$
$$H_1: E(u|X) \neq 0$$

- Fixed Effect Model VS Random Effect Model

$$H_0: E(e|X) = 0$$
$$H_1: E(e|X) \neq 0$$

We use Hausman's Specification Error Test to confirm whether X is related to u and X is related to e.

2.Choice of OLS and IV

2.1 Review of OLS

Regression model:

$$y = X\beta + u$$

$y: n\times 1, X: n\times k, \beta: k\times 1, u: n\times 1.$

- Assumption

(1)$E(u_i|\mathrm{X}) = 0, \mathrm{i} = 1, \ldots, \mathrm{n}.$
(2)$Var(u_i|\mathrm{X}) = \sigma^2$
(3)$\mathrm{E}\left(u_i u_j \middle| X\right) = 0, \mathrm{i} \neq j.$

Under the assumption, $\hat{\beta}_{OLS}$ is consistent and efficient.

2.1 Review of IV(instrumental variables)

$$Y_i = \alpha + \beta X_i + u_i, i = 1,2, \dots, N$$

$Y: wage, X: education,$ $\mathrm{cov}(X_i, u_i) \neq 0$.

Assumption of instrumental variables (Z_i: father education):

$\mathrm{cov}(X_i, Z_i) \neq 0$,$\mathrm{cov}(u_i, Z_i) = 0$.

$$cov(Y_i, Z_i) = cov(\alpha + \beta X_i + u_i, Z_i) = \beta cov(X_i, Z_i)$$

$$\hat{\beta}_{IV} = \frac{cov(Y_i, Z_i)}{cov(X_i, Z_i)} = \frac{\sum_{i=1}^{N}(Y_i - \bar{Y})(Z_i - \bar{Z})}{\sum_{i=1}^{N}(X_i - \bar{X})(Z_i - \bar{Z})}$$

2.2 Two stage least squared method: 2SLS

$$Y = X\beta + u$$

- The first stage:
  endogenous variables are regressed by instrumental variables.

$$x = \pi_0 + \pi_1 z_{i1} + \pi_2 z_{i2} + w_i$$

$$\hat{X} = Z(Z'Z)^{-1}\mathrm{Z}'\mathrm{X}$$

Z: $N \times l, l \geq k$. $P_z = Z(Z'Z)^{-1}\mathrm{Z}'$.

- The second stage:

$$\hat{\beta}_{2SLS} = \left(\hat{X}'\hat{X}\right)^{-1}\hat{X}'Y\,,$$

$$\begin{aligned}\hat{\beta}_{2SLS} &= (X'Z(Z'Z)^{-1}Z'Z(Z'Z)^{-1}Z'\mathrm{X})\mathrm{X}'Z(Z'Z)^{-1}\mathrm{Z}'\mathrm{Y} \\ &= (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}\mathrm{Z}'\mathrm{Y} \\ &= (Z'X)^{-1}(Z'Z)(X'Z)^{-1}X'Z(Z'Z)^{-1}Z'Y \\ &= (Z'X)^{-1}(Z'Z)\ (Z'Z)^{-1}Z'Y \\ &= (Z'X)^{-1}Z'Y\end{aligned}$$

Note:

$$\hat{u}_i = y_i - X\hat{\beta}_{2SLS}$$

Apply Hausman's test

$$H_0: X \text{ and } u \text{ are independent}$$

$$H_1: X \text{ and } u \text{ are not independent}$$

- When we accept H_0, we use OLS.
- When we reject H_0, we use IV.
- $\hat{\beta}_{OLS}$ is consistent and efficient under H_0, but is not efficient under H_1.
- $\hat{\beta}_{IV}$ is consistent under both H_0 and H_1, but is not efficient under H_0.

Test statistic:

$$(\hat{\beta}_{IV} - \hat{\beta}_{OLS})' \left(V(\hat{\beta}_{IV}) - V(\hat{\beta}_{OLS})\right)^{-1} (\hat{\beta}_{IV} - \hat{\beta}_{OLS}) \rightarrow \chi^2(k)$$

3.Choice of Fixed Effect Model or Random Effect Model

3.1 Review of Fixed Effect Model(固定効果モデル)

$$y_{it} = \beta_0 + \beta_1 X_{it} + e_{it}, \qquad (1)$$

$$e_{it} = v_i + u_{it}, \qquad (2)$$

$$y_{it} = \beta_0 + \beta_1 X_{it} + v_i + u_{it}. \qquad (3)$$

$i: individual\ (i = 1,2,\ldots,n),$ $\qquad t: time(t = 1,2,\ldots,T)$

The assumptions are as follows:

$$E(v_i|X_{it}) \neq 0, \qquad V(v_i|X_{it}) = \sigma_v^2,$$

$$E(u_{it}|X_{it}) = 0, \qquad V(u_{it}|X_{it}) = \sigma_u^2,$$

$$E(v_i u_{it}|X_{it}) = 0.$$

3.2 Review of Random Effect Model(ランダム効果モデル)

$$y_{it} = \beta_0 + \beta_1 X_{it} + e_{it}, \qquad (1)$$

$$e_{it} = v_i + u_{it}, \qquad (2)$$

$i: individual\ (i = 1,2,\ldots,n),$ $t: time(t = 1,2,\ldots,T)$

The assumptions are as follows:

$$E(v_i|X_{it}) = 0, \qquad V(v_i|X_{it}) = \sigma_v^2,$$

$$E(u_{it}|X_{it}) = 0, \qquad V(u_{it}|X_{it}) = \sigma_u^2,$$

$$E(v_i u_{it}|X_{it}) = 0.$$

Apply Hausman's test

$$H_0\text{: } X_{it}\ and\ e_{it}\ are\ independent$$

$$H_1\text{: } X_{it}\ and\ e_{it}\ are\ not\ independent$$

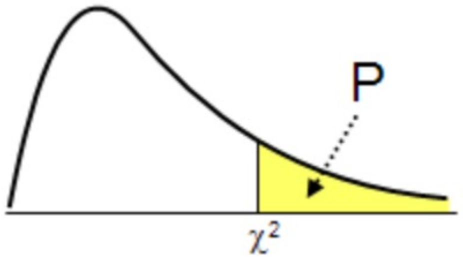
Where $e_{it} = v_i + u_{it}$.

- When we accept H_0, we use the random effect model.
- When we reject H_0, we use the fixed effect model.

Accept or reject the null hypothesis as determined by the p-value

- $P \leq 0.1$, reject the null hypothesis(H_0) at a 10% significance level.
- $P \leq 0.05$, reject the null hypothesis(H_0) at a 5% significance level.
- $P \leq 0.01$, reject the null hypothesis(H_0) at a 1% significance level.

Values of the Chi-squared distribution



	P										
DF	**0.995**	**0.975**	**0.20**	**0.10**	**0.05**	**0.025**	**0.02**	**0.01**	**0.005**	**0.002**	**0.001**
1	0.0000393	0.000982	1.642	2.706	3.841	5.024	5.412	6.635	7.879	9.550	10.828
2	0.0100	0.0506	3.219	4.605	5.991	7.378	7.824	9.210	10.597	12.429	13.816
3	0.0717	0.216	4.642	6.251	7.815	9.348	9.837	11.345	12.838	14.796	16.266
4	0.207	0.484	5.989	7.779	9.488	11.143	11.668	13.277	14.860	16.924	18.467
5	0.412	0.831	7.289	9.236	11.070	12.833	13.388	15.086	16.750	18.907	20.515
6	0.676	1.237	8.558	10.645	12.592	14.449	15.033	16.812	18.548	20.791	22.458
7	0.989	1.690	9.803	12.017	14.067	16.013	16.622	18.475	20.278	22.601	24.322
8	1.344	2.180	11.030	13.362	15.507	17.535	18.168	20.090	21.955	24.352	26.124
9	1.735	2.700	12.242	14.684	16.919	19.023	19.679	21.666	23.589	26.056	27.877
10	2.156	3.247	13.442	15.987	18.307	20.483	21.161	23.209	25.188	27.722	29.588

4.Empirical Example

- We use the following model to analyze what factors determine investment.

$$inv_{it} = \mu_i + \mu_\alpha + \beta_1 value_{it} + \beta_2 capital_{it} + \varepsilon_{it}$$

$i = 1, \dots, 10, t = 1, \dots, 20.$

inv: total investment, value: corporate value, capital: fixed asset.

μ_i: individual effect, μ_α: intercept.

note (assumption):

Fixed effect model: $E(\mu_i) \neq 0,$

Random effect model: $E(\mu_i) = 0.$

Estimate Fixed Effect Model with LSDV(“ Within”)

```
        Estimate Std. Error t-value  Pr(>|t|)
value   0.110124   0.011857  9.2879 < 2.2e-16 ***
capital 0.310065   0.017355 17.8666 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    2244400
Residual Sum of Squares: 523480
R-Squared      :  0.76676
      Adj. R-Squared :  0.72075
F-statistic: 309.014 on 2 and 188 DF, p-value: < 2.22e-16
```

Estimate Random effect Model with GLS

```
              Estimate Std. Error t-value Pr(>|t|)
(Intercept) -57.834415  28.898935 -2.0013  0.04674 *
value         0.109781   0.010493 10.4627  < 2e-16 ***
capital       0.308113   0.017180 17.9339  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    2381400
Residual Sum of Squares: 548900
R-Squared      :  0.7695
     Adj. R-Squared :  0.75796
F-statistic: 328.837 on 2 and 197 DF, p-value: < 2.22e-16
```

Hausman test

- The estimated result of LSDV is " result2", and the estimated result of GLS is " result3".

```
> phtest(result2,result3)

        Hausman Test

data:  inv ~ value + capital
chisq = 2.3304, df = 2, p-value = 0.3119
alternative hypothesis: one model is inconsistent
```

- Conclusion:

P=0.3119>0.05, accept H_0 at a 5% significance level, we choose a fixed effect model to analyze.

Hausman Test

Import Dataset (Grunfeld)

```
> install.packages ("plm")
> library(plm)
> data("Grunfeld",package="plm")
> result2=plm(inv~value+capital,data=Grunfeld,model="within")
> summary(result2)
> result3=plm(inv~value+capital,data=Grunfeld,model="random")
> summary(result3)
> phtest(result2,result3)
```

Reference: http://user.keio.ac.jp/~nagakura/R/R_panel.pdf