

3.3 Hausman's Specification Error (特定化誤差) Test

Regression model:

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

Suppose that X is stochastic.

If $E(u|X) = 0$, OLSE $\hat{\beta}$ is unbiased because of $\hat{\beta} = (X'X)^{-1}X'y = \beta + (X'X)^{-1}X'u$ and $E((X'X)^{-1}X'u) = 0$.

However, If $E(u|X) \neq 0$, OLSE $\hat{\beta}$ is biased and inconsistent.

Therefore, we need to check if X is correlated with u or not.

⇒ **Hausman's Specification Error Test**

The null and alternative hypotheses are:

- H_0 : X and u are independent, i.e., $\text{Cov}(X, u) = 0$,
- H_1 : X and u are not independent.

Suppose that we have two estimators $\hat{\beta}_0$ and $\hat{\beta}_1$, which have the following properties:

- $\hat{\beta}_0$ is consistent and efficient under H_0 , but is not consistent under H_1 ,
- $\hat{\beta}_1$ is consistent under both H_0 and H_1 , but is not efficient under H_0 .

Under the conditions above, we have the following test statistic:

$$(\hat{\beta}_1 - \hat{\beta}_0)' \left(\text{V}(\hat{\beta}_1) - \text{V}(\hat{\beta}_0) \right)^{-1} (\hat{\beta}_1 - \hat{\beta}_0) \rightarrow \chi^2(k).$$

Example: $\hat{\beta}_0$ is OLS, while $\hat{\beta}_1$ is IV such as 2SLS.

Hausman, J.A. (1978) "Specification Tests in Econometrics," *Econometrica*, Vol.46, No.6, pp.1251–1271.

3.4 Choice of Fixed Effect Model or Random Effect Model

3.4.1 The Case where X is Correlated with u — Review

The standard regression model is given by:

$$y = X\beta + u, \quad u \sim N(0, \sigma^2 I_n)$$

OLS is:

$$\hat{\beta} = (X'X)^{-1}X'y = \beta + (X'X)^{-1}X'u.$$

If X is not correlated with u , i.e., $E(X'u) = 0$, we have the result: $E(\hat{\beta}) = \beta$.

However, if X is correlated with u , i.e., $E(X'u) \neq 0$, we have the result: $E(\hat{\beta}) \neq \beta$.

$\implies \hat{\beta}$ is biased.

Assume that in the limit we have the followings:

$$\begin{aligned}\frac{1}{n}(X'X)^{-1} &\longrightarrow M_{xx}^{-1}, \\ \frac{1}{n}X'u &\longrightarrow M_{xu} \neq 0 \text{ when } X \text{ is correlated with } u.\end{aligned}$$

Therefore, even in the limit,

$$\text{plim } \hat{\beta} = \beta + M_{xx}^{-1}M_{xu} \neq \beta,$$

which implies that $\hat{\beta}$ is not a consistent estimator of β .

Thus, in the case where X is correlated with u , OLSE $\hat{\beta}$ is neither unbiased nor consistent.

3.4.2 Fixed Effect Model or Random Effect Model

Usually, in the random effect model, we can consider that v_i is correlated with X_{it} .

[Reason:]

v_i includes the unobserved variables in the i th individual, i.e., ability, intelligence, and so on.

X_{it} represents the observed variables in the i th individual, i.e., income, assets, and so on.

The unobserved variables v_i are related to the observed variables X_{it} .

Therefore, we consider that v_i is correlated with X_{it} .

Thus, in the case of the random effect model, usually we cannot use OLS or GLS.

In order to use the random effect model, we need to test whether v_i is uncorrelated with X_{it} .

Apply Hausman's test.

- H_0 : X_{it} and e_{it} are independent (\longrightarrow Use the random effect model),
- H_1 : X_{it} and e_{it} are not independent (\longrightarrow Use the fixed effect model),

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

Note that:

- We can use the random effect model under H_0 , but not under H_1 .
- We can use the fixed effect model under both H_0 and H_1 .
- The random effect model is more efficient than the fixed effect model under H_0 .

Therefore, under H_0 we should use the random effect model, rather than the fixed effect model.

3.5 Applications

Example of Panel Data in Section 3: Production Function of Prefectures from
2001 to 2010.

pref: 都道府県 (通し番号 1~47)

year: 年度 (2001~2010 年)

y : 県内総生産 (支出側、実質: 固定基準年方式), 出所: 県民経済計算 (平成 13 年度 - 平成 24 年度) (93SNA, 平成 17 年基準計数)

k : 都道府県別民間資本ストック (平成 12 暦年価格, 年度末, 国民経済計算ベース 平成 23 年 3 月時点) 一期前 (2000~2009 年)

l : 県内就業者数, 出所: 県民経済計算 (平成 13 年度 - 平成 24 年度) (93SNA, 平成 17 年基準計数)

```
. tsset pref year
```

```

panel variable:  pref (strongly balanced)
time variable:  year, 2001 to 2010
delta: 1 unit

```

```

. gen ly=log(y)
. gen lk=log(k)
. gen ll=log(l)
. reg ly lk ll

```

Source	SS	df	MS	Number of obs =	470
Model	316.479302	2	158.239651	F(2, 467)	=19374.95
Residual	3.81409572	467	.008167229	Prob > F	= 0.0000
				R-squared	= 0.9881
				Adj R-squared	= 0.9880
Total	320.293398	469	.682928354	Root MSE	= .09037

	ly	lk	ll	_cons	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
	ly	lk	ll	_cons	.0941587	.0081273	11.59	0.000	.0781881 .1101294
					.9976399	.0102641	97.20	0.000	.9774703 1.017809
					.5970719	.0773137	7.72	0.000	.4451461 .7489978

```

. xtreg ly lk ll,fe

```

```

Fixed-effects (within) regression
Group variable: pref

R-sq:  within = 0.1721
        between = 0.9456
        overall = 0.9439

corr(u_i, Xb) = 0.8803

Number of obs   = 470
Number of groups = 47

Obs per group:  min = 10
                  avg  = 10.0
                  max  = 10

F(2,421) = 43.77
Prob > F  = 0.0000

```

```

-----+-----
      ly |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----+-----
      lk |   .2329208   .0252321     9.23   0.000   .1833242   .2825175
      ll |   .3268537   .0810662     4.03   0.000   .1675088   .4861987
     _cons |  7.691145   1.376677     5.59   0.000   4.985128  10.39716
-----+-----
sigma_u |   .41045507
sigma_e |   .03561437
rho     |   .99252757   (fraction of variance due to u_i)
-----+-----

```

```

F test that all u_i=0:      F(46, 421) = 56.22      Prob > F = 0.0000

```

```
. est store fixed
```

```
. xtreg ly lk ll,re
```

```

Random-effects GLS regression
Group variable: pref

Number of obs   = 470
Number of groups = 47

```

R-sq: within = 0.1058
 between = 0.9805
 overall = 0.9787

Obs per group: min = 10
 avg = 10.0
 max = 10

corr(u_i, X) = 0 (assumed)

Wald chi2(2) = 3875.75
 Prob > chi2 = 0.0000

ly	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
lk	.2457767	.0153094	16.05	0.000	.2157708	.2757827
ll	.8105099	.0220256	36.80	0.000	.7673406	.8536793
_cons	.8332015	.2411141	3.46	0.001	.3606265	1.305776
sigma_u	.081609					
sigma_e	.03561437					
rho	.8400205	(fraction of variance due to u_i)				

. hausman fixed

	---- Coefficients ----			
	(b) fixed	(B) .	(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
lk	.2329208	.2457767	-.0128559	.020057
ll	.3268537	.8105099	-.4836562	.0780167

b = consistent under Ho and Ha; obtained from xtreg
B = inconsistent under Ha, efficient under Ho; obtained from xtreg

Test: Ho: difference in coefficients not systematic

chi2(2) = (b-B)'[(V_b-V_B)^(-1)](b-B)
= 144.66
Prob>chi2 = 0.0000

4 Generalized Method of Moments (GMM, 一般化積率法)

4.1 Method of Moments (MM, 積率法)

As $n \rightarrow \infty$, we have the result: $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \rightarrow E(X) = \mu$.

\Rightarrow **Law of Large Number** (大数の法則)

X_1, X_2, \dots, X_n are n realizations of X .

[Review] **Chebyshev's inequality** (チェビシェフの不等式) is given by:

$$P(|X - \mu| > \epsilon) \leq \frac{\sigma^2}{\epsilon^2} \quad \text{or} \quad P(|X - \mu| \leq \epsilon) \geq 1 - \frac{\sigma^2}{\epsilon^2},$$

where $\mu = E(X)$, $\sigma^2 = V(X)$ and any $\epsilon > 0$.

Note that $P(|X - \mu| > \epsilon) + P(|X - \mu| \leq \epsilon) = 1$.

Replace X , $E(X)$ and $V(X)$ by \bar{X} , $E(\bar{X}) = \mu$ and $V(\bar{X}) = \frac{\sigma^2}{n}$.

As $n \rightarrow \infty$,

$$P(|\bar{X} - \mu| \leq \epsilon) \geq 1 - \frac{\sigma^2}{n\epsilon^2} \rightarrow 1.$$

That is, $\bar{X} \rightarrow \mu$ as $n \rightarrow \infty$.

[End of Review]

\bar{X} is an approximation of $E(X) = \mu$.

Therefore, $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ is taken as an estimator of μ .

$\Rightarrow \bar{X}$ is MM estimator of $E(X) = \mu$.