

Problem równoczesności w MNK

- O problemie równoczesności mówimy, gdy występuje korelacja między wartością oczekiwaną ε_i i równoczesnym x_i
- Model liniowy

$$\begin{aligned} y &= X\beta + \varepsilon, \\ E(u) &= 0 \end{aligned}$$

- Powiedzmy, że występuje w nim zależność między wartością oczekiwaną błędu losowego a zmiennymi objaśniającymi:

$$E(\varepsilon | X) \neq 0$$

- W takiej sytuacji

$$\begin{aligned} E(b) &= E \left[(X'X)^{-1} X'y \mid X \right] = (X'X)^{-1} X' E[X\beta + \varepsilon \mid X] \\ &= \beta + (X'X)^{-1} X' E[\varepsilon \mid X] \neq \beta \end{aligned}$$

- Estymator *MNK* jest więc obciążony!

Wynik Aby estymator nie był obciążony, nie może być związku między wartością oczekiwaną błędu losowego ε_i a wszystkimi $x_1, \dots, x_n$

- Jednak jeśli $E(x'_i \varepsilon_i) = 0$ estymator obciążony może być estymatorem zgodnym

- W dużych próbach

$$\text{plim} \frac{1}{n} \mathbf{X}' \mathbf{X} = \text{plim} \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i' = \text{E}(\mathbf{x}_i \mathbf{x}_i')$$

$$\text{plim} \frac{1}{n} \mathbf{X}' \boldsymbol{\varepsilon} = \text{plim} \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i' \varepsilon_i = \text{E}(\mathbf{x}_i' \varepsilon_i)$$

- Estymator MNK ma granicę według prawdopodobieństwa równą

$$\begin{aligned} \text{plim } \mathbf{b} &= \text{plim} \left[(\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y} \right] = \text{plim} \left[(\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' (\mathbf{X} \boldsymbol{\beta} + \mathbf{u}) \right] \\ &= \boldsymbol{\beta} + \text{plim} \left(\frac{1}{n} \mathbf{X}' \mathbf{X} \right)^{-1} \left(\frac{1}{n} \mathbf{X}' \boldsymbol{\varepsilon} \right) \\ &= \boldsymbol{\beta} + [\text{E}(\mathbf{x}_i \mathbf{x}_i')]^{-1} \text{E}(\mathbf{x}_i' \varepsilon_i) \neq \boldsymbol{\beta} \end{aligned}$$

Wynik Estymator MNK parametru β jest estymatorem zgodnym jeśli $E(x_i' \varepsilon_i) = \text{Cov}(x_i, \varepsilon_i) = 0$.

- Zauważmy, że wymóg ten jest znacznie słabszy niż warunek konieczny do nieobciążoności - wystarczy jedynie brak korelacji między ε_i i równoczesnym x_i .

Przykład Chcemy wyestymować wpływ edukacji na poziom zarobków. Jednak w badaniach nie jesteśmy w stanie uchwycić zdolności danej osoby. Zwykle zdolności wpływają zarówno na poziom edukacji jak i na poziom zarobków. Jeśli zamiast modelu (dane PGSS $lrincome$ - logarytm dochodu z pracy)

$$\log(lrincome) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{sex} + \beta_3 \text{age} + \beta_4 \text{age}^2 + \beta_5 \text{zdolności} + \varepsilon$$

wyestymujemy model

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 sex + \beta_3 age + \beta_4 age^2 + u$$

to przy założeniu, że osiągnięty poziom wykształcenia zależy od zdolności uzyskujemy:

$$\text{Cov}(educ, u) = \text{Cov}(educ, \beta_5 \text{zdolności} + \varepsilon) = \beta_5 \text{Cov}(educ, \text{zdolności}) \neq 0$$

Ponieważ zmienna pominięta jest dodatnio zdolności są dodatnio skorelowane z edukacją i wpływają dodatnio na zarobki więc oczekujemy, że obciążenie oszacowania współczynnika β_4 jest dodatnie.

Source	SS	df	MS	Number of obs = 4227		
Model	192.55502	4	48.1387549	F(4, 4222) = 42.20		
Residual	4815.96513	4222	1.14068336	Prob > F = 0.0000		
Total	5008.52015	4226	1.18516804	R-squared = 0.0384		
				Adj R-squared = 0.0375		
				Root MSE = 1.068		

lrincome	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	.0415494	.0052952	7.85	0.000	.0311681	.0519307
sex	-.3073812	.033083	-9.29	0.000	-.3722412	-.2425211
age	.052235	.0096668	5.40	0.000	.0332831	.071187
age2	-.0005858	.0001157	-5.06	0.000	-.0008127	-.0003589
_cons	4.788569	.2033733	23.55	0.000	4.389851	5.187288

Metoda Zmiennych Instrumentalnych

- Załóżmy, że jesteśmy w stanie znaleźć takie zmienne, od których nie zależy równoczesny błąd losowy:

$$E(\varepsilon_i | z_i) = 0$$

- Zmiennne takie będziemy nazywać instrumentami
- Instrumentami mogą być również elementy x_i
- Estymator *MZI* (*IV* - Instrumental Variable) jest dany wzorem:

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

- **Zgodność** estymatora MZI

$$\begin{aligned}
 \text{plim} \left(\hat{\beta} \right) &= \text{plim} \left[(\mathbf{Z}'\mathbf{X})^{-1} \mathbf{Z}'\mathbf{y} \right] = \text{plim} \left[(\mathbf{Z}'\mathbf{X})^{-1} \mathbf{Z}' (\mathbf{X}\beta + \varepsilon) \right] \\
 &= \beta + \text{plim} \left[(\mathbf{Z}'\mathbf{X})^{-1} \mathbf{Z}'\varepsilon \right] \\
 &= \beta + \left(\text{plim} \frac{1}{n} \mathbf{Z}'\mathbf{X} \right)^{-1} \left(\text{plim} \frac{1}{n} \mathbf{Z}'\varepsilon \right) \\
 &= \beta + [\text{E}(\mathbf{z}_i\mathbf{x}_i)]^{-1} \underbrace{\text{E}(\mathbf{z}_i'\varepsilon_i)}_0 = \beta
 \end{aligned}$$

Uwaga Wyprowadzenie to jest prawidłowe jedynie, gdy $\text{E}(\mathbf{z}_i\mathbf{x}_i)$ jest macierzą nieosobliwą: estymator MZI jest zgodny jeśli nie ma korelacji między instrumentami i równoczesnymi błędami losowymi, ale *jest* korelacja między instrumentami i zmiennymi objaśnianymi.

Uogólniony estymator estymator MZI

- Uogólniony estymator MZI stosujemy, gdy liczba zmiennych instrumentalnych przekracza liczbę zmiennych

$$\widehat{W} = Z (Z'Z)^{-1} Z'X = P_Z X$$

- Postać estymatora można parametru $\hat{\beta}$ ma teraz postać:

$$\begin{aligned} b_{MZI} &= \left(\widehat{W}' X \right)^{-1} \widehat{W}' y = \left(X' Z (Z'Z)^{-1} Z' X \right)^{-1} X' Z (Z'Z)^{-1} Z' y \\ &= \left(X' P_Z X \right)^{-1} X' P_Z y = \left(\widehat{W}' \widehat{W} \right)^{-1} \widehat{W}' y \end{aligned}$$

gdzie

$$P_Z = Z (Z' Z)^{-1} Z'$$

- Pokazanie zgodności i postaci macierzy wariancji przebiega tak samo jak dla zwykłego estymatora MZI

Testowanie w *MZI*

- Test egzogeniczności jest wariantem testu Hausmana
- Test Hausmana można przeprowadzić, jeśli mamy dwa estymatory o następujących właściwościach
 1. jeśli H_0 jest prawdziwa to $\hat{\theta}$ jest zgodny a $\tilde{\theta}$ jest asymptotycznie efektywny
 2. jeśli H_0 jest fałszywa jest to $\hat{\theta}$ jest dalej zgodny, ale $\tilde{\theta}$ jest asymptotycznie obciążony
- Test Hausmana polega na sprawdzeniu, czy oszacowania parametrów uzyskanych za pomocą estymatora $\hat{\theta}$ i $\tilde{\theta}$ są do siebie zbliżone.

- W naszym przypadku H_0 brzmi, że X jest egzogeniczne.
- W takim przypadku rolę estymatora $\hat{\theta}$ pełni estymator MZI a estymatora $\tilde{\theta}$ estymator MNK .
- Jeśli H_0 jest prawdziwe i X jest egzogeniczne to łatwo pokazać, że

$$\sqrt{n} (\mathbf{b}_{MZI} - \mathbf{b}_{MNK}) \xrightarrow{D} N(0, \Sigma_{\mathbf{b}_{MZI} - \mathbf{b}_{MNK}})$$

co oznacza, że statystyka

$$(\mathbf{b}_{MZI} - \mathbf{b}_{MNK})' \left(\tilde{\Sigma}_{\mathbf{b}_{MZI} - \mathbf{b}_{MNK}} \right)^{-1} (\mathbf{b}_{MZI} - \mathbf{b}_{MNK}) \xrightarrow{D} \chi_g^2$$

gdzie g jest ilością zmiennych, których egzogeniczność badamy.

- Test na poprawność instrumentów Sargana (możliwy do przeprowadzenia tylko jeśli liczba instrumentów większa od liczby zmiennych w modelu)

- Przeprowadzamy go badając

$$\frac{\tilde{e}' P_W \tilde{e}}{\tilde{\sigma}^2} \xrightarrow{D} \chi_{p-K}^2$$

gdzie $\tilde{e} = y - Xb_{MZI}$, $\tilde{\sigma}^2 = \frac{\tilde{e}'\tilde{e}}{n}$

- Statystyka ta służy do badania prawidłowości doboru instrumentów.
- Jeśli H_0 jest odrzucone to znaczy, że założenie o braku skorelowania zmiennych instrumentalnych i zaburzeń losowych jest fałszywe.

Przykład kontynuacja. Za instrumenty dla edukacji osiągnięty poziom przez matkę i ojca poziom wykształcenia. Zakładamy, że edukacja rodziców nie mają znaczenia dla poziomu płac (jest nieskorelowana z u) Uzyskane za pomocą MZI oszacowania mają postać:

Instrumental variables (2SLS) regression

Source	SS	df	MS	Number of obs =	3862
Model	36.0635538	4	9.01588846	F(4, 3857) =	37.77
Residual	4503.18032	3857	1.16753444	Prob > F =	0.0000
				R-squared =	0.0079
				Adj R-squared =	0.0069
Total	4539.24387	3861	1.17566534	Root MSE =	1.0805

lrincome	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	.1005262	.0131433	7.65	0.000	.0747576	.1262948
sex	-.3317927	.0358404	-9.26	0.000	-.4020608	-.2615247
age	.0496855	.0103803	4.79	0.000	.0293342	.0700369
age2	-.0005467	.0001253	-4.36	0.000	-.0007924	-.0003011
_cons	4.200114	.2384509	17.61	0.000	3.732612	4.667616

Instrumented: educ

Instruments: sex age age2 _Ipadeg_1 _Ipadeg_2 _Ipadeg_3 _Ipadeg_4 _Ipadeg_5
_Ipadeg_6 _Ipadeg_7 _Ipadeg_8 _Ipadeg_9 _Ipadeg_98 _Imadeg_1
_Imadeg_2 _Imadeg_3 _Imadeg_4 _Imadeg_5 _Imadeg_6 _Imadeg_7
_Imadeg_8 _Imadeg_9 _Imadeg_98

- Zauważmy, że oszacowany współczynnik dla edukacji jest teraz znacznie wyższy!
- Wynik testu sargana

Tests of overidentifying restrictions:

Sargan N*R-sq test	12.918	Chi-sq(19)	P-value = 0.8427
Basman test	12.881	Chi-sq(19)	P-value = 0.8446

- Instrumenty wydają się być poprawne.

- Wynik testu Hausmana:

	---- Coefficients ----			
	(b)	(B)	(b-B)	sqrt (diag (V_b-V_B))
	MZI	MNK	Difference	S.E.
educ	.1005262	.0415494	.0589768	.0118632
sex	-.3317927	-.3073812	-.0244116	.0126693
age	.0496855	.052235	-.0025495	.0034388
age2	-.0005467	-.0005858	.0000391	.0000441

b = consistent under H_0 and H_a ; obtained from ivreg
 B = inconsistent under H_a , efficient under H_0 ; obtained from regress

Test: H_0 : difference in coefficients not systematic

chi2(4) = $(b-B)' [(V_b-V_B)^{-1}] (b-B)$
 = 28.59
 Prob>chi2 = 0.0000

- Wyniki uzyskane dla *MZI* i *MNK* istotnie się różnią - odrzucamy H_0 , że w modelu nie występuje równoczesność

- Uwaga: uzyskany wynik nie jest zgodny z intuicją - obciążenie oszacowania MNK współczynnika przy edukacji powinno być dodatnie. Zgodny estymator MZI dał jednak jeszcze wyższe oszacowanie tego współczynnika.