

**Metody radzenia sobie z
heteroskedastycznością i
autokorelacją
MNK dla losowych zmiennych
niezależnych**

Stanisław Cichocki

Natalia Nehrebecka

Wykład 14


Plan wykładu

- ▶ 1. Metody radzenia sobie z heteroskedastycznością i autokorelacją
 - Uogólniona Metoda Najmniejszych Kwadratów
 - Stosowalna Uogólniona Metoda Najmniejszych Kwadratów (SUMNK)
 - Odporne macierze wariancji i kowariancji
- ▶ 2. MNK dla losowych zmiennych niezależnych

Plan wykładu

- ▶ 1. Metody radzenia sobie z heteroskedastycznością i autokorelacją
 - Uogólniona Metoda Najmniejszych Kwadratów
 - Stosowalna Uogólniona Metoda Najmniejszych Kwadratów (SUMNK)
 - Odporne macierze wariancji i kowariancji
- ▶ 2. MNK dla losowych zmiennych niezależnych

UMNK

- ▶ Jednym z możliwych sposobów rozwiązania problemu niesferyczności błędu losowego 
oszacowanie macierzy wariancji błędu losowego
- ▶ Zastosowana metoda estymacji jest uogólnieniem MNK dla przypadku, gdy błędy losowe są niesferyczne
- ▶ Taki uogólniony estymator można uzyskać bezpośrednio z zadania na minimalizację sumy kwadratów reszt
- ▶ Wyprowadzimy go sprowadzając model uogólniony do KMRL

UMNK

- ▶ Założenia analogiczne do KMRL, ale różniące się brakiem założenia o homoskedastyczności i autokorelacji:

1. Model jest liniowy: $y = X\beta + \varepsilon$

2. X jest nielosowe

3. $E(\varepsilon) = 0$

4. $Var(\varepsilon) = \Omega = \sigma^2 V$, gdzie V jest znana

UMNK

- ▶ Dla dodatnio określonej i symetrycznej macierzy V można zawsze znaleźć taką nieosobliwą macierz z elementami rzeczywistymi, L że

$$LVL' = I$$

$$V = L^{-1}L^{-1'}$$

$$L'L = V^{-1}$$

UMNK

- ▶ Mnożąc, $y = X\beta + \varepsilon$ przez L :

$$Ly = LX\beta + L\varepsilon$$

- ▶ Definiujemy nowe zmienne, będące przekształceniem zmiennych pierwotnych:

$$y^* = X^*\beta + \varepsilon^*$$

UMNK

- ▶ Wariancja ε^* :

$$\text{Var}(\varepsilon^*) = \text{Var}(L\varepsilon) = L\text{Var}(\varepsilon)L' = \sigma^2 LVL' = \sigma^2 I$$

- ▶ Błędy losowe są homosekdstyczne i nieskorelowane.
- ▶ Model przekształcony będzie spełniał wszystkie **założenia KMRL**
 —————▶ prawdziwe są wszystkie wnioski na temat **własności MNK**
- ▶ Estymator, który otrzymamy nazywamy **estymatorem UMNK** i ma on wszystkie własności, które posiada estymator MNK

UMNK

- ▶ Estymator Ugólnionej Metody Najmniejszych Kwadratów (Generalized Least Squares – GLS):

$$\begin{aligned} b_{UMNK} &= (X^* ' X^*)^{-1} X^* ' y^* = [(LX)'(LX)]^{-1} (LX)' Ly \\ &= (X ' L' LX)^{-1} X ' L' Ly = (X ' V^{-1} X)^{-1} X ' V^{-1} y \\ &= (X ' \Omega^{-1} X)^{-1} X ' \Omega^{-1} y \end{aligned}$$

UMNK

- ▶ Macierz wariancji kowariancji estymatora b_{UMNK}

$$\begin{aligned} \text{Var}(b_{UMNK}) &= \sigma^2 (X^* ' X^*)^{-1} = \sigma^2 [(LX)'(LX)]^{-1} \\ &= \sigma^2 (X ' L' LX)^{-1} = \sigma^2 (X ' V^{-1} X)^{-1} = (X ' \Omega^{-1} X)^{-1} \end{aligned}$$

UMNK

- ▶ Reszty w przypadku estymatora UMNK:

$$y - \hat{y} = y - Xb_{UMNK} = e_{UMNK}$$

UMNK

- ▶ Estymator s^2 :

$$\begin{aligned} s_{UMNK}^2 &= \frac{e^*{}' e^*}{N - K} = \frac{(y^* - X^* b_{UMNK})' (y^* - X^* b_{UMNK})}{N - K} \\ &= \frac{(y - X b_{UMNK})' V^{-1} (y - X b_{UMNK})}{N - K} = \frac{e'_{UMNK} V^{-1} e_{UMNK}}{N - K} \end{aligned}$$

UMNK

- ▶ Przy założeniu normalności rozkładu błędów losowych sposób testowania hipotez w UMNK jest identyczny jak w MNK
- ▶ Prawdziwe będzie też uogólnienie twierdzenia Gaussa Markowa:

Twierdzenie Aitkena: Estymator Uogólnionej Metody Najmniejszych Kwadratów jest najlepszym, liniowym i nieobciążonym estymatorem parametru β o ile macierz Ω lub macierz V jest znana.

Problem: prawie nigdy nie znamy V

UMNK

- ▶ Jednym z przypadków kiedy forma macierzy wariancji jest znana
 ➔ Ważona Metoda Najmniejszych Kwadratów
- ▶ W modelu występuje jedynie **heteroskedastyczność**:

$$\Omega = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_n^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} v_1 & 0 \\ 0 & v_n \end{bmatrix} = \sigma^2 V$$

UMNK

- ▶ Zakładając, że v_i są znane:

$$L = \begin{bmatrix} \frac{1}{\sqrt{v_1}} & 0 \\ 0 & \frac{1}{\sqrt{v_n}} \end{bmatrix}$$

UMNK

- ▶ Można pokazać, iż stosując macierz \mathbf{L} :

$$y_i^* = \frac{y_i}{\sqrt{v_i}}$$

$$x_{k,i}^* = \frac{x_{k,i}}{\sqrt{v_i}}$$

- ▶ v_i - można traktować jako wagi przypisywane obserwacjom

UMNK

- ▶ Im wyższa wariancja błędu losowego, tym niższą wagę przypisujemy obserwacjom

SUMNK

- ▶ Gdy elementy macierzy V są nieznane \longrightarrow używamy estymatora UMNK zastępując V oszacowaniem \hat{V}
 \longrightarrow metoda ta to Stosowalna Uogólniona Metoda Najmniejszych Kwadratów (Feasible Generalized Least Squares)

SUMNK

- ▶ **Istota:** budujemy dodatkowy model dla wariancji i kowariancji tak aby znaleźć funkcję $\Omega(\Theta)$

- ▶ Zakładamy, że:

$$Var(\varepsilon) = \Omega(\theta)$$

- ▶ Wymiar Θ niezależny od liczby obserwacji
- ▶ Dla znanego Θ moglibyśmy zastosować UMNK, jednak zwykle Θ jest nieznane.
- ▶ Wiemy, że estymator MNK jest zgodny nawet wtedy, gdy występuje heteroskedastyczność lub autokorelacja

SUMNK

▶ Etapy estymacji za pomocą SUMNK:

1. Szacujemy model za pomocą MNK uzyskując b_{MNK}
2. Na podstawie wektora reszt szacujemy model pomocniczy, z którego otrzymujemy oszacowania Θ (oszacowania te oznaczamy jako $\hat{\Theta}$)
3. Powtórnie szacujemy parametry modelu stosując estymator UMNK, nieznaną macierz Ω zastępujemy jej oszacowaniem

$$\hat{\Omega} = \Omega(\hat{\Theta})$$

SUMNK

4. Uzyskany estymator SUMNK:

$$b_{SUMNK} = [X' \hat{\Omega}^{-1} X]^{-1} X' \hat{\Omega}^{-1} y$$

Odporne estymatory macierzy wariancji i kowariancji

- ▶ Bazują na formie macierzy wariancji estymatora b_{MNK} dla niesferycznych błędów losowych

$$\text{Var}(b_{MNK}) = \frac{1}{N} \left(\frac{1}{N} X'X \right)^{-1} \left(\frac{1}{N} X' \Omega X \right) \left(\frac{1}{N} X'X \right)^{-1}$$

- ▶ Konkretna postać estymatorów odpornych macierzy wariancji zależy od tego, czy w modelu występuje jedynie heteroskedastyczność, czy też autokorelacja

Odporne estymatory macierzy wariancji i kowariancji

- ▶ Najpopularniejszym odpornym na heteroskedastyczność estymatorem macierzy wariancji i kowariancji **b** jest estymator White'a
- ▶ Estymator White'a umożliwia przeprowadzenie wnioskowania statystycznego bez konieczności szacowania modelu dla wariancji i kowariancji
- ▶ Znacznie łatwiejszy w stosowaniu niż SUMNK

Odporne estymatory macierzy wariancji i kowariancji

```
reg wydg dochg los
```

Source	SS	df	MS			
Model	2.3877e+10	2	1.1938e+10	Number of obs =	31705	
Residual	3.4095e+10	31702	1075469.35	F(2, 31702) =	11100.72	
Total	5.7971e+10	31704	1828523.21	Prob > F =	0.0000	
				R-squared =	0.4119	
				Adj R-squared =	0.4118	
				Root MSE =	1037	

wydg	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
dochg	.5683134	.0041262	137.73	0.000	.5602259	.576401
los	65.12119	3.858651	16.88	0.000	57.55809	72.6843
_cons	550.8444	13.90437	39.62	0.000	523.5913	578.0976

Odporne estymatory macierzy wariancji i kowariancji

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity

Ho: Constant variance

Variables: fitted values of wydg

chi2(1) = 121623.16

Prob > chi2 = 0.0000

White's test for Ho: homoskedasticity

against Ha: unrestricted heteroskedasticity

chi2(5) = 5700.95

Prob > chi2 = 0.0000

Odporne estymatory macierzy wariancji i kowariancji

```
reg wydg dochg los, robust
```

Linear regression

```
Number of obs = 31705  
F( 2, 31702) = 1785.95  
Prob > F      = 0.0000  
R-squared     = 0.4119  
Root MSE     = 1037
```

		Robust				[95% Conf. Interval]	
	wydg	Coef.	Std. Err.	t	P> t		
	dochg	.5683134	.0319817	17.77	0.000	.505628	.6309989
	los	65.12119	8.640123	7.54	0.000	48.18621	82.05617
	_cons	550.8444	40.10087	13.74	0.000	472.2452	629.4437

Odporne estymatory macierzy wariancji i kowariancji

- ▶ Estymator Newey'a-Westa macierzy wariancji i kowariancji
 ➔ gdy w modelu występuje heteroskedastyczność i autokorelacja

Plan wykładu

- ▶ 1. Metody radzenia sobie z heteroskedastycznością i autokorelacją
 - Uogólniona Metoda Najmniejszych Kwadratów
 - Stosowalna Uogólniona Metoda Najmniejszych Kwadratów (SUMNK)
 - Odporne macierze wariancji i kowariancji
- ▶ 2. MNK dla losowych zmiennych niezależnych

MNK dla losowych zmiennych niezależnych

- ▶ Do tej pory przyjmowaliśmy założenie o nielosowości zmiennych objaśniających. Jest to sensowne w zastosowaniach eksperymentalnych. Jednak w badaniach empirycznych jest to trudne do obrony.
- ▶ Możemy wprowadzić analogiczny do KMRL układ założeń w przypadku losowej macierzy zmiennych objaśniających:

1. Model ma postać liniową:

$$y = X\beta + \varepsilon$$

2. Warunkowa wartość oczekiwania błędu losowego jest równa zero:

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

MNK dla losowych zmiennych niezależnych

- ▶ 3. Warunkowa macierz wariancji i kowariancji ε nie zależy od X i jest równa:

$$\text{Var}(\varepsilon|X) = \sigma^2 I$$

MNK dla losowych zmiennych niezależnych

- ▶ Warunkowa wartość oczekiwana \mathbf{b} jest równa

$$\begin{aligned} E(b|X) &= \beta + E[(X'X)^{-1}X'\varepsilon|X] \\ &= \beta + (X'X)^{-1}X'E(\varepsilon|X) = \beta \end{aligned}$$

Bezwarunkowa wartość oczekiwana \mathbf{b} jest równa

$$E[E(b|X)] = E(\beta) = \beta$$

MNK dla losowych zmiennych niezależnych

- ▶ Wariancja estymatora:

$$\begin{aligned} \text{Var}(b|X) &= \text{Var}(\beta + (X'X)^{-1}X'\varepsilon|X) \\ &= (X'X)^{-1}X'\text{Var}(\varepsilon|X)X(X'X)^{-1} = \sigma^2(X'X)^{-1} \end{aligned}$$

Pytania teoretyczne

1. Pokazać, w jaki sposób można, w przypadku znanej macierzy Ω , sprowadzić model z niesferycznymi błędami losowymi do modelu spełniającego założenia KMRL.
2. Wyjaśnić różnicę między UMNK i SUMNK.
3. Jakie są zalety stosowania estymatora MNK w połączeniu z estymatorem odpornym macierzy wariancji i kowariancji w porównaniu do stosowania estymatora UMNK?

Dziękuję za uwagę