

Heteroskedastyczność i autokorelacja (cz. I)

Wykład 14

Stanisław Cichocki
Natalia Nehrebecka

Plan zajęć

1. Konsekwencje heteroskedastyczności i autokorelacji
2. Uogólniona MNK
3. Stosowalna Uogólniona MNK
4. Odporne macierze wariancji i kowariancji b

Plan zajęć

1. Konsekwencje heteroskedastyczności i autokorelacji
2. Uogólniona MNK
3. Stosowalna Uogólniona MNK
4. Odporne macierze wariancji i kowariancji b

Konsekwencje heteroskedastyczności i autokorelacji

- ▶ Jeżeli założenie o homoskedastyczności lub założenie o braku autokorelacji nie jest spełnione to mówimy o **niesferyczności błędów losowych**
- ▶ W przypadku występowania heteroskedastyczności lub autokorelacji:

$$\text{Var}(\varepsilon) = \Omega = \sigma^2 V$$

Konsekwencje heteroskedastyczności i autokorelacji

- Estymator b jest nadal **nieobciążony** :

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

- Można pokazać, że estymator b , przy odpowiednich założeniach, jest także **zgodny**
- Nie będzie on jednak efektywny \longrightarrow można znaleźć estymator o mniejszej wariancji

Konsekwencje heteroskedastyczności i autokorelacji

- Macierz wariancji i kowariancji b :

$$\begin{aligned} \text{Var}(b) &= E\left((X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}\right) = \\ &(X'X)^{-1}X'\Omega X(X'X)^{-1} = \\ &\sigma^2(X'X)^{-1}X'VX(X'X)^{-1} \end{aligned}$$

- Wzór ten różni się znacznie od prawidłowego wzoru na wariancję MNK:

$$\text{Var}(b) = \sigma^2(X'X)^{-1}$$

Konsekwencje heteroskedastyczności i autokorelacji

- W rezultacie estymator macierzy wariancji i kowariancji b , którym posługiwaliśmy się do tej pory, nie będzie dobrym oszacowaniem macierzy wariancji i kowariancji b
- Konsekwencje braku zgodności estymatora macierzy wariancji i kowariancji b mogą być poważne:
 - ▶ estymatora macierzy wariancji i kowariancji b używamy przy konstruowaniu praktycznie wszystkich statystyk testowych
 - ▶ brak jego zgodności implikuje, że używanie standardowych statystyk testowych może doprowadzić do błędnych wyników wnioskowania statystycznego

Plan zajęć

1. Konsekwencje heteroskedastyczności i autokorelacji
2. Uogólniona MNK
3. Stosowalna Uogólniona MNK
4. Odporne macierze wariancji i kowariancji b

Uogólniona MNK

- ▶ 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

Uogólniona MNK

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

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

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

$$LVL' = I$$

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

Uogólniona MNK

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

- ▶ Zdefiniujmy:

$$y^* = Ly$$

$$X^* = LX$$

$$\varepsilon^* = L\varepsilon$$

- ▶ Dla modelu przekształconego

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

Uogólniona MNK

- Możemy znaleźć pewną macierz L dzięki, której przekształcimy model tak, że błędy losowe w przekształconym modelu będą homoskedastyczne i nieskorelowane
- Model ten będzie **spełniał** wszystkie **założenia KMRL**
- Dla tego modelu prawdziwe są wszystkie wnioski na temat **własności MNK**
- Estymator, który otrzymamy nazywamy **estymatorem UMNK** i ma on wszystkie własności, które posiada estymator MNK

Uogólniona MNK

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

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

Uogólniona MNK

- ▶ Macierz wariancji i kowariancji b_{UMNK} :

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

Uogólniona MNK

- ▶ Estymator UMNK wariancji błędu losowego

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

Uogólniona MNK

- Przy założeniu normalności rozkładu \mathcal{E} sposób testowania hipotez statystycznych w UMNK będzie 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

Ważona UMNK

- Wyjątek, gdy znamy V \longrightarrow **Ważona UMNK** (stosujemy jeżeli w modelu występuje tylko **heteroskedastyczność**)
- Ważona UMNK \longrightarrow przypisujemy wagi obserwacjom przy czym im wyższa wariancja błędu losowego tym niższą wagę przypisujemy

Ważona UMNK

- ▶ W modelu występuje:

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

- ▶ Załóżmy, że elementy v_i są znane.
- ▶ Macierz L ma postać:

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

Ważona UMNK

- ▶ UMNK sprowadza się do MNK z użyciem zmiennych:

$$y^* = Ly$$

$$X^* = LX$$

- ▶ W naszym przypadku zmienne te będą miały postać:

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

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

▶ Wielkości v_i można interpretować jako wagi przypisywane obserwacjom

Przykład

- ▶ Funkcja konsumpcji

$$C_i = a + bY_i + \varepsilon_i$$

- ▶ gdzie:

- C_i - konsumpcja w kraju i ,
- Y_i - poziom GDP w tym kraju.

- ▶ Bardziej realistyczny model na zmiennych wyrażonych *per capita*:

$$C^*_i = a \frac{1}{z_i} + bY^*_i + \varepsilon^*_i$$

- ▶ gdzie:

- $C^*_i = \frac{C_i}{z_i}$
- $Y^*_i = \frac{Y_i}{z_i}$
- z_i – ludność danego kraju.

Pytania teoretyczne

1. Jak niesferyczność błędów losowych wpływa na własności MNK?
2. Pokazać, w jaki sposób można, w przypadku znanej macierzy V , sprowadzić model z niesferycznymi błędami losowymi do modelu spełniającego założenia KMRL.

Dziękuję za uwagę