

MIKROEKONOMETRIA

ZAJĘCIA 5

**WIKTOR
BUDZIŃSKI**

ZADANIE 1

Wykorzystując dane *me_wine_update.dta* przygotuj model regresji hedonicznej dla cen wina

- I. Przeanalizuj problem heteroskedastyczności wykorzystując test Breusha-Pagana w różnych wersjach
 - Wynik testu wskazujący na heteroskedastyczność może być także spowodowany niepoprawną formą funkcyjną modelu lub 'nienormalnymi' resztami

HETEROSKEDASTYCZNOŚĆ

- Pomimo heteroskedastyczności estymator MNK wciąż jest zgodny i nieobciążony. Przestaje być on jednak efektywny (wartości błędów standardowych mogą być mylące)
 - Najprostsze rozwiązanie – tzw. odporne macierze kowariancji:
 - Macierz White'a: $(\mathbf{X}'\mathbf{X})^{-1} \times \sum_{i=1}^n e_i^2 X_i X_i' \times (\mathbf{X}'\mathbf{X})^{-1}$
 - Bądź inne alternatywy dostępne w STATA

ZADANIE 2

Wykorzystując dane *me_wine_update.dta* przygotuj model regresji hedonicznej dla cen wina

- I. Dokonaj estymacji modelu regresji liniowej wykorzystując odporne macierze kowariancji
 - A. Czy wnioski z tak uzyskanych wyników się zmieniają?

HETEROSKEDASTYCZNOŚĆ

- Innym rozwiązaniem problemu heteroskedastyczności może być jego bezpośrednie modelowanie
- Możemy estymować równanie
$$Y_i = \mathbf{X}_i\boldsymbol{\beta} + \sigma(\mathbf{X}_i, \lambda)\varepsilon$$
- Gdzie zazwyczaj $\sigma(\mathbf{X}_i, \lambda) = \exp(\mathbf{X}_i\boldsymbol{\lambda})$
- **Problem:** Nie jest to możliwe przy pomocy MNK

ALE PO CO?

- Odporne macierze kowariancji nie zmieniają oszacowań parametrów, a regresja z heteroskedastycznością tak
 - Teoretycznie, dla dużych prób, powinny być bardzo zbliżone, ale w praktyce różnice mogą być znaczne
- Ma to znaczenie jeśli interesuje nas elastyczność (model na logarytmach)
 - Jeśli heteroskedastyczność występuje, to jej uwzględnienie ma znaczenie dla prognozowania
- Np. rozważmy model w postaci: $\log(y_i) = \mathbf{X}_i\boldsymbol{\beta} + \sigma(\mathbf{X}_i)\varepsilon_i$
 - Jeśli chcielibyśmy prognozować średnie y_i przy pewnym poziomie zmiennych objaśnianych to musimy także uwzględnić ich wpływ na średnie wartości składnika losowego: $E(y_i | \mathbf{X}_i^*) = \exp(\mathbf{X}_i^*\boldsymbol{\beta} + 0,5\sigma^2(\mathbf{X}_i^*))$
 - Bo średnia rozkładu lognormalnego to $\exp(\mu + \sigma^2/2)$

ALE PO CO?

- W MNK interpretacja modelu dość prosta

- Np. dla formy funkcyjnej log-log mamy:

$$\frac{d(\log(Y))}{d\log(X)} = \frac{\frac{dY}{Y}}{\frac{dX}{X}} = \frac{dY}{dX} \frac{X}{Y} = \frac{\exp(\log(X)\beta + 0.5\sigma^2)\beta}{X} \frac{X}{\exp(\log(X)\beta + 0.5\sigma^2)} = \beta$$

- Ale jeśli mamy heteroskedastyczność:

$$\begin{aligned} \frac{d(\log(Y))}{d\log(X)} &= \frac{\frac{dY}{Y}}{\frac{dX}{X}} = \frac{dY}{dX} \frac{X}{Y} = \exp(\log(X)\beta + 0.5\sigma(X)^2) \left(\frac{\beta}{X} + \sigma(X)\sigma'(X) \right) \frac{X}{\exp(\log(X)\beta + 0.5\sigma(X)^2)} \\ &= X \left(\frac{\beta}{X} + \sigma(X)\sigma'(X) \right) \end{aligned}$$

METODA NAJWIĘKSZEJ WIARYGODNOŚCI

- Założenia MNK są dość restrykcyjne i w praktyce często niespełnione
 - Dane mikroekonomiczne często cechują się nieliniową formą funkcyjną i heteroskedastycznością
 - Często zmienne w ogóle nie mają rozkładów ciągłych (np. zmienne binarne)
- Potrzebny jest nam estymator, który pozwoliłby na oszacowanie bardziej ogólnych zależności, który moglibyśmy dostosować do konkretnego rozkładu zmiennej objaśnianej
- MNW jest właśnie takim estymatorem

METODA NAJWIĘKSZEJ WIARYGODNOŚCI

- Idea:

- Zaobserwowaliśmy dane: $(Y_i, \mathbf{X}_i)$, dla $i = 1, \dots, N$
- Zakładamy pewien rozkład dla zmiennej objaśnianej $f(Y_i | \mathbf{X}_i, \theta)$
 - Dla zmiennej dyskretnej będzie to po prostu funkcja prawdopodobieństwa, dla zmiennych ciągłych gęstość
- Rozkład ten zależy od pewnych zmiennych egzogenicznych oraz nieznanymi nam parametrów
- Zakładając, że obserwacje są niezależne między respondentami, to rozkład dla całej próby będzie dany przez następującą funkcję wiarygodności (ang. *likelihood function*)

$$L = \prod_{i=1}^N f(Y_i | \mathbf{X}_i, \theta)$$

METODA NAJWIĘKSZEJ WIARYGODNOŚCI

- Chcemy znaleźć takie θ , żeby zmaksymalizować tę funkcję
 - Czyli, chcemy żeby parametry modelu były takie, żeby było jak najbardziej prawdopodobne że zaobserwujemy takie dane jakie mamy przy założonym rozkładzie
- W praktyce łatwiej maksymalizuje się logarytm funkcji wiarygodności

$$LL = \log \left(\prod_{i=1}^N f(Y_i | \mathbf{X}_i, \theta) \right) = \sum_{i=1}^N \log (f(Y_i | \mathbf{X}_i, \theta))$$

- Taki estymator jest zgodny i asymptotycznie efektywny
 - Zakładając, że wybrany przez nas rozkład jest poprawny

PRZYKŁAD

Regresja liniowa przy pomocy MNW

- Regresję liniową można oszacować przy użyciu MNW, zakładając:

$$f(Y_i | \mathbf{X}_i, \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-0.5 \frac{(Y_i - \mathbf{X}_i\boldsymbol{\beta})^2}{\sigma^2}\right)$$

- Przy czym $\theta = [\boldsymbol{\beta}, \sigma^2]$
- W przypadku regresji liniowej MNW będzie dawało mniej efektywne oszacowania parametrów niż MNK

PRZYKŁAD

- W MNW heteroskedastyczność można łatwo uwzględnić w modelu:

$$f(Y_i | \mathbf{X}_i, \theta) = \frac{1}{\sqrt{2\pi\sigma(\mathbf{X}_i, \lambda)^2}} \exp\left(-0.5 \frac{(Y_i - \mathbf{X}_i\boldsymbol{\beta})^2}{\sigma(\mathbf{X}_i, \lambda)^2}\right)$$

- Zakładając $\sigma(\mathbf{X}_i, \lambda) = \exp(\mathbf{X}_i\lambda)$
- Teraz $\theta = [\boldsymbol{\beta}, \lambda]$

ZADANIE 3

Wykorzystując dane `me_wine.dta` przygotuj model regresji hedonicznej dla cen wina

1. Dokonaj estymacji modelu regresji liniowej przy użyciu MNW
2. Rozszerz model o bezpośrednie modelowanie heteroskedastyczności

EFEKTY KRAŃCOWE

- W modelach nieliniowych w których niemożliwa jest prosta interpretacja parametrów wykorzystuje się tzw. efekty krańcowe
 - Jest to numeryczne przybliżenie szukanej wielkości
- Jeśli X zmieni się o jednostkę to jak zmieni się Y : $\frac{\partial E(Y|X)}{\partial X}$
- Jeśli X zmieni się o 1% to o ile % zmieni się Y : $\frac{\partial E(\log(Y)|X)}{\partial \log(X)}$
 - Można też liczyć semi-elastyczności itd.
- Zależy od przyjętego rozkładu błędu losowego
- Zależy od aktualnego poziomu X (funkcja jest nieliniowa)
 - Można wyznaczyć średnie efekty krańcowe
 - Dla średnich X w próbie lub średnia dla poziomów X obserwacji w próbie (powszechniejsza druga metoda)

ZADANIE 4

1. Policz efekt krańcowy mówiący jak wzrost zmiennej Age o jednostkę wpłynie na zmianę ceny
2. Policz efekt krańcowy mówiący jak wzrost zmiennej Cases o jeden procent wpłynie na procentową zmianę ceny
3. *Wykorzystaj metody symulacyjne by sprawdzić czy kontrolowanie heteroskedastyczności faktycznie może wpłynąć na nasze wnioskowanie