

# **MIKROEKONOMETRIA**

## **ZAJĘCIA 12**

**WIKTOR  
BUDZIŃSKI**

# ENDOGENICZNOŚĆ

- W regresji liniowej estymujemy następujące równanie:

$$Y_i = \mathbf{X}_i\boldsymbol{\beta} + \varepsilon_i$$

- KMRL zakłada, że wszystkie zmienne objaśniające są egzogeniczne tj.  $E(\varepsilon|\mathbf{X}) = 0$
- Jeżeli ten warunek jest niespełniony mówimy o endogeniczności zmiennych objaśniających
  - Złamanie tego założenia ma poważne konsekwencje – estymator MNK przestaje być zgodny
- Może to wynikać z różnych przyczyn:
  - Korelacja ze zmienną pominiętą
  - Sprzężenie zwrotne między zmienną objaśnianą i objaśniającą

# ENDOGENICZNOŚĆ

- Rozwiązanie – zastosowanie zmiennych instrumentalnych
  - Znalezienie zmiennych, które są silnie skorelowane ze zmienną, którą podejrzewamy o endogeniczność, ale nie są skorelowane z błędem losowym
  - Nie jest to zadanie łatwe ...
- Podstawowym estymatorem w przypadku zmiennych ciągłych jest tzw. Dwustopniowa Metoda Najmniejszych Kwadratów (2MNK)

# 2MNK

- W pierwszym kroku liczymy regresję, w której wyjaśniamy wszystkie zmienne z podstawowego modelu zmiennymi instrumentalnymi

$$\mathbf{B} = (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'\mathbf{X}$$

- Wśród zmiennych instrumentalnych  $\mathbf{Z}$  mogą być egzogeniczne zmienne z  $\mathbf{X}$
- W drugim kroku wyjaśniamy  $\mathbf{Y}$  przy użyciu wartości dopasowanych z pierwszych regresji  $\mathbf{X} = \mathbf{ZB}$  :

$$\boldsymbol{\beta}_{2MNK} = \left( \mathbf{X}'\mathbf{X} \right)^{-1} \mathbf{X}'\mathbf{Y}$$

- Estymator jest zgodny jeżeli instrumenty są nieskorelowane z błędem losowym
- Aby model skonwergował potrzebujemy co najmniej tyle samo zmiennych instrumentalnych co zmiennych endogenicznych

# ZADANIE 1

Korzystając z danych *me\_twins.dta* przeprowadź regresję wyjaśniającą jak posiadanie więcej niż dwójki dzieci wpływa na dochód kobiet

- A. Przeprowadź zwykłą regresję liniową, w której logarytm dochodu jest objaśniany przez to czy, kobieta ma więcej niż dwójkę dzieci (*morekids*), wiek (*agem*), wiek przy pierwszym porodzie (*agefstm*), oraz zmienne mówiące o kolorze skóry (*blackm*, *hispm*, *othracem*)
- B. Porównaj wyniki z estymatorem 2MNK, gdzie zmienna *morekids* jest endogeniczna, a jako instrument wykorzystana jest zmienna *twins*. Czy wyniki różnią od tych uzyskanych z MNK?
- C. Sprawdź czy wykorzystane instrumenty są dobrymi predyktorami zmiennej *morekids*
- D. Porównaj wyniki z modelem, w którym wykorzystany jest instrument *samesex*

# TESTY

## Test Hausmana

- Jeżeli w modelu nie ma endogeniczności to zastosowanie estymatora 2MNL nie powinno specjalnie zmienić oszacowań parametrów
- $H_0$ : endogeniczność nie występuje
- Jeśli  $H_0$  spełnione – estymator 2MNL jest nieefektywny, ale nie powinien powodować niezgodności
- Statystyka testu

$$H = (\beta_{2MNL} - \beta_{MNL})' (V_{2MNL} - V_{MNL})^{-1} (\beta_{2MNL} - \beta_{MNL})$$

- Statystyka ma rozkład chi-kwadrat z liczbą stopni swobody równą liczbie parametrów

# TESTY

## Test Wu-Hausmana

- $H_0$ : endogeniczność nie występuje
- Zakłada się, że mamy następujący model

$$Y_i = \mathbf{X}_i\boldsymbol{\beta} + \alpha X_i^e + \rho v_i + \varepsilon_i$$

- Gdzie  $v$  jest zmienną, której nie obserwujemy i która jest skorelowana z  $X_i^e$
- Chcemy zbadać istotność tej zmiennej tj. czy  $\rho = 0$
- Robimy to dwu stopniowo:
  - Estymujemy  $X_i^e = \mathbf{Z}_i\boldsymbol{\theta} + \mathbf{X}_i\boldsymbol{\beta} + u_i$ , z czego liczymy reszty  $\hat{u}_i$
  - Estymujemy  $Y_i = \mathbf{X}_i\boldsymbol{\beta} + \alpha X_i^e + \rho\hat{u}_i + \varepsilon_i$ , i testujemy  $\rho = 0$
- Statystyka ma rozkład F

# TEST SARGANA

- Jeśli posiadamy więcej niż jedną zmienną instrumentalną dla każdej zmiennej endogenicznej możliwe jest przetestowanie czy instrumenty faktycznie nie wpływają na zmienną objaśnianą
- Najczęściej stosowanym testem jest test Sargana
  - Sprawdza korelacje reszt z regresji z instrumentami
  - Hipotezą zerową jest, że zmienne instrumentalne są warunkowo niezależne od zmiennej objaśnianej

## ZADANIE 1

Korzystając z danych *me\_twins.dta* przeprowadź regresję wyjaśniającą jak posiadanie więcej niż dwójki dzieci wpływa na dochód kobiet

- E. Przetestuj model pod kątem występowania endogeniczności wykorzystując testy Hausmana i Wu-Hausmana
- F. Wykorzystaj test Sargana aby sprawdzić poprawność zmiennych instrumentalnych

# SŁABE INSTRUMENTY

- Pomimo, że estymator 2MNK jest zgodny w nieskończonej próbie, w skończonej próbie jest on obciążony
  - Więcej zmiennych instrumentalnych może powodować większe obciążenie
- Kiedy zmienne instrumentalne są słabo skorelowane ze zmienną endogeniczną mówimy o problemie 'słabych instrumentów' (ang. *weak instruments*)
  - Słabe instrumenty mogą również zwiększać obciążenie

# SŁABE INSTRUMENTY

- Mniej formalne testowanie czy instrumenty są 'słabe' można przeprowadzić na dwa sposoby
  1. Licząc korelacje zmiennej endogenicznej ze zmiennymi instrumentalnymi
  2. Sprawdzając łączną istotność instrumentów w regresji pomocniczej w 2MNK
    - Reguła kciuka mówi, że jeśli statystyka F jest mniejsza niż 10 to instrumenty są słabe

# STOCK & YOGO

Formalny test na słabość instrumentów można przeprowadzić wykorzystując procedurę Stock i Yogo

- Tak jak wcześniej wykorzystują oni test F na łączną istotność instrumentów
- Jeżeli jest więcej niż jedna zmienna endogeniczna wykorzystywane są wartości własne macierzowego odpowiednika statystyki F
- W tej procedurze wartości krytyczne są jednak inne niż w standardowym teście F
  - Liczone na dwa sposoby

# STOCK & YOGO

Jeżeli mamy o co najmniej 2 instrumenty więcej niż zmiennych endogenicznych to wartość krytyczna może zostać policzona na podstawie

- Największego obciążenia (relatywnie do MNK) jakie jesteśmy w stanie zaakceptować
- Liczby zmiennych instrumentalnych
- Liczby zmiennych endogenicznych

Dla dowolnej liczby zmiennych instrumentalnych wartość krytyczna może zostać policzona na podstawie

- Największego obciążenia testu Walda na istotność zmiennych endogenicznych jakie jesteśmy w stanie zaakceptować
- Liczby zmiennych instrumentalnych + liczby zmiennych endogenicznych

## ZADANIE 1

Korzystając z danych *me\_twins.dta* przeprowadź regresję wyjaśniającą jak posiadanie więcej niż dwójki dzieci wpływa na dochód kobiet

- G. Przeprowadź test na słabość instrumentów wykorzystując metody formalne i nieformalne

# ENDOGENICZNOŚĆ I MNW

- 2MKN jest rozszerzeniem regresji liniowej
- Zakłada, że zarówno zmienna objaśniana jak i zmienna endogeniczna są ciągłe
  - Jeśli tak nie jest estymator nie będzie zgodny
- W przypadku gdy któraś (bądź obie) z tych zmiennych nie są ciągłe, możemy skonstruować model wykorzystując MNW
- Zazwyczaj będzie to model z dwoma równaniami
  - Jednym dla zmiennej objaśnianej
  - Drugim dla zmiennej endogenicznej
  - Korelacja błędów losowych w tych równaniach będzie traktowana jako parametr modelu

# ENDOGENICZNOŚĆ I MNW

- 2MKN jest rozszerzeniem regresji liniowej
- Zakłada, że zarówno zmienna objaśniana jak i zmienna endogeniczna są ciągłe
  - Jeśli tak nie jest estymator nie będzie zgodny
- W przypadku gdy któraś (bądź obie) z tych zmiennych nie są ciągłe, możemy skonstruować model wykorzystując MNW
- Zazwyczaj będzie to model z dwoma równaniami
  - Jednym dla zmiennej objaśnianej
  - Drugim dla zmiennej endogenicznej
  - Korelacja błędów losowych w tych równaniach będzie traktowana jako parametr modelu

# ENDOGENICZNE DZIAŁANIE

Jeśli zmienna objaśniana jest ciągła, ale zmienna endogeniczna jest binarna możemy zastosować tzw. regresję z endogenicznym działaniem

Model ma dwa równania:

- Regresja liniowa

$$Y_i = \mathbf{X}_i \boldsymbol{\beta} + \alpha X_i^e + \varepsilon_i$$

- Regresja probitowa:

$$X_i^{e*} = \mathbf{Z}_i \boldsymbol{\theta} + \mathbf{X}_i \boldsymbol{\beta} + u_i$$

- Przy czym obserwujemy  $X_i^e = 1$ , jeśli  $X_i^{e*} > 0$
- Korelacja między błędami losowymi jest bezpośrednio estymowana
  - Testując jej istotność, testujemy endogeniczność

## ZADANIE 2

Korzystając z danych *me\_medexp3.dta* sprawdź czy logarytm wydatków na leki zależy od tego czy ktoś posiada dodatkowe ubezpieczenie od pracodawcy

- A. Porównaj modele MNK oraz 2MNK
- B. Wykorzystaj model z endogenicznym działaniem

# ENDOGENICZNOŚĆ DLA ZMIENNYCH DYSKRETNYCH

- W przypadku kiedy nasza zmienna objaśniana nie jest ciągła, również możemy skonstruować model z dwoma równaniami przy pomocy MNW
- Często niestety takie modele nie są dostępne w pakietach statystycznych
- To co można wykorzystać jako relatywnie proste rozwiązanie to tzw. metoda funkcji kontrolnej
  - Działa podobnie do testu Wu-Hausmana
  - Zakłada, że zmienna endogeniczna jest ciągła
  - Wykorzystuje zmienne instrumentalne

# METODA FUNKCJI KONTROLNEJ

Najpierw liczymy regresję pomocniczą:

$$X_i^e = \mathbf{Z}_i \boldsymbol{\theta} + \mathbf{X}_i \boldsymbol{\beta} + u_i$$

Przewidujemy z niej reszty:  $\hat{u}_i$

Liczymy główną regresję wykorzystując reszty jako dodatkową zmienną objaśniającą:

$$Y_i^* = \mathbf{X}_i \boldsymbol{\beta} + \alpha X_i^e + \lambda \hat{u}_i + \varepsilon_i$$

Oszacowane współczynniki będą zgodne

- Błędy standardowe będą jednak niepoprawne
- Najlepiej wykorzystać dodatkowo bootstrap

## ZADANIE 3

Wykorzystując *me\_sim.dta* oszacuj elastyczność cenową prawdopodobieństwa zakupu produktu

- A. Wykorzystując zwykłą regresję probitową
- B. Wykorzystując regresję probitową oraz metodę funkcji kontrolnej
- C. Policz błędy standardowe dla oszacowania elastyczności cenowej uzyskanej przy pomocy metody funkcji kontrolnej używając bootstrapu