



# Mikroekonometria

## 13



Mikołaj Czajkowski  
Wiktor Budziński

# Ocena efektów działań

---

- ▶ Efekty oszacowane przy pomocy regresji zazwyczaj opisują korelację między zmienną objaśnianą, a zmiennymi objaśniającymi
- ▶ Często to co byśmy chcieli zidentyfikować to przyczynowość
  - ▶ Np. znaczenie wyższego wykształcenia dla zarobków, wpływ karmienia piersią/palenia na IQ dziecka, działania leku, dania łąpówki
  - ▶ Istotne dla decyzji, regulacji, polityki
- ▶ Ocena efektów działań (ang. *treatment effects*) to grupa metod mających na celu określenie przyczynowości i oszacowanie wpływu działania
  - ▶ W większości opierają się o proste modele regresji liniowej lub regresji logistycznej

# Ocena efektów działań

---

- ▶ Formalnie, zastanawiamy się jak pewne działanie (ang. *treatment*),  $D_i$ , wpływa na reakcję (ang. *response*),  $Y_i$ .
  - ▶ Można o tym myśleć w kategoriach zmiennej objaśnianej / zmiennej objaśniającej. Po prostu w tej literaturze używa się trochę innej nomenklatury.

- ▶ Jeśli  $D_i = 1$  to obserwujemy  $Y_i^1$ , w przeciwnym wypadku obserwujemy  $Y_i^0$ . Co oznacza, że:

$$Y_i = D_i Y_i^1 + (1 - D_i) Y_i^0$$

- ▶ Interesuje nas wielkość wpływu działania, tzn.  $Y_i^1 - Y_i^0$
- ▶ Problemem jest, że dla danej osoby obserwujemy tylko jeden stan – nie znamy stanu kontrfaktycznego (ang. *counterfactual*)

# Ocena efektów działań

---

- ▶ Zazwyczaj upraszcza się analizę i estymuje się średni efekt działania, tzn.  $\mathbf{E}(Y^1 - Y^0)$ 
  - ▶ Można też skupić się na kwantylach, jak w regresji kwantylowej
- ▶ W idealnych warunkach przeprowadzilibyśmy randomizowany eksperyment
- ▶ Wtedy  $Y^d \perp D$ , i zachodzi:

$$\mathbf{E}(Y^1 - Y^0) = \mathbf{E}(Y \mid D = 1) - \mathbf{E}(Y \mid D = 0)$$

- ▶ Tzn. możemy policzyć średnią w grupie poddanej działaniu oraz w grupie niepoddanej działaniu, a następnie odjąć je od siebie.
  - ▶ Możliwe w naukach medycznych, np. testowanie leków
  - ▶ Raczej niemożliwe w ekonomii

# Ocena efektów działań

---

- ▶ W ekonomii zazwyczaj występuje pewien stopień auto-selekcji
  - ▶ Selekcja ze względu na zmienne obserwowalne
    - ▶ Tzn. zachodzi  $Y^d \perp D | X$ 
      - Np. chcemy zbadać wpływ podjęcia wyższej edukacji na zarobki. Jeśli z jakichś powodów mniej kobiet decyduje się na edukację i jednocześnie płeć wpływa na zarobki niezależnie od poziomu edukacji, to porównanie średnich w dwóch grupach nie wystarczy.
      - Zazwyczaj obserwujemy płeć, więc możemy to kontrolować
  - ▶ Selekcja ze względu na zmienne nieobserwowalne
    - ▶ Tzn. zachodzi  $Y^d \perp D | \varepsilon$ 
      - Np. ludzie którzy decydują się na edukację mają jakieś cechy, które sprawiają, że będą więcej zarabiali niezależnie od podjęcia edukacji (np. zdolności, pracowitość)
      - Trudniejsze do kontrolowania – analogiczne do problemu endogeniczności

# Parowanie współczynników skłonności

---

- ▶ Jedną z metod, które wykorzystuje się kiedy występuje pierwszy rodzaj auto-selekcji jest parowanie współczynników skłonności (ang. *propensity score matching*).
- ▶ Szacujemy prawdopodobieństwo że osoba zostanie poddana działaniu:  $p(D | X)$
- ▶ Dla każdej osoby, która została poddana działaniu możemy znaleźć „parę”, która nie została poddana działaniu, ale ma podobne prawdopodobieństwo. I analogicznie w drugą stronę.
- ▶ Średni efekt działania można wtedy policzyć jako

$$ATE = \frac{1}{N} \sum_{i=1}^{N_1} (Y_i^1 - \tilde{Y}_i^0) + \frac{1}{N} \sum_{i=1}^{N_0} (\tilde{Y}_i^1 - Y_i^0)$$

- ▶ Gdzie  $\tilde{Y}_i^1, \tilde{Y}_i^0$  to stany kontrfaktyczne oszacowane na podstawie „pary”
  - ▶ Może to być np. średnia z kilku osób, które mają bliskie p-stwo

# Parowanie współczynników skłonności

---

- ▶ Aby ta metoda poprawnie działała musimy założyć, że  $Y^d \perp D | X$ , oraz  $0 < p(D | X) < 1$ 
  - ▶ Drugi warunek mówi, że żadna ze zmiennych nie może w pełni determinować, czy osoba została poddana działaniu
- ▶ W praktyce  $p(D | X)$  jest nieznane i szacuje się je używając modeli dla zmiennej binarnej, np. logit, probit
- ▶ Parowanie może być przeprowadzone używając innych współczynników

# Zadanie 1. Nowy program nauczania matematyki

---

1. Wczytaj zbiór danych `me.math.dta`
2. Sprawdź jak nowy program nauczania matematyki w szkole wpływa na wynik uczniów na teście z algebry (American College Testing Program)
  1. Wykorzystaj Parowanie współczynników skłonności
  2. Porównaj wyniki ze zwykłą regresją liniową





# Endogeniczne działanie

---

- ▶ Co zrobić kiedy działanie jest endogeniczne?
- ▶ Używając regresji można postępować w analogiczny sposób jak w przypadku zwykłego problemu endogeniczności
- ▶ Ponieważ działanie jest zazwyczaj zmienną binarną można wykorzystać bardziej skomplikowany model niż 2MNK, który bierze to pod uwagę:
  - ▶ Model logitowy/probitowy dla równania objaśniającego czy ktoś został poddany działaniu
  - ▶ Regresja liniowa dla zmiennej, która nas interesuje ( $Y$ )
  - ▶ Modelujemy korelację między błędami losowymi tych równań (podobnie jak w modelu Heckmana)

# Endogeniczne działanie

- ▶ Model można uogólnić pozwalając na różne błędy losowe w równaniu regresji w zależności od tego czy osoba została poddana działaniu czy nie

$$D_i^* = \mathbf{y}'_i \mathbf{z} + u_i, \quad D_i = 1 \text{ jeśli } D_i^* > 0, \quad 0 \text{ w p.p.}$$

$$Y_i = \boldsymbol{\beta}' \mathbf{X}_i + D_i \varepsilon_{i1} + (1 - D_i) \varepsilon_{i0}, \quad j = 0, 1$$

$$\begin{bmatrix} u_i \\ \varepsilon_{i0} \\ \varepsilon_{i1} \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_0 \theta_0 & \rho_1 \theta_1 \\ \rho_0 \theta_0 & \theta_0^2 & 0 \\ \rho_1 \theta_1 & 0 & \theta_1^2 \end{bmatrix} \right)$$

## Zadanie 2. Posiadanie ubezpieczenia, a wydatki na leki

---

1. Wczytaj zbiór danych me. medexp3.dta
2. Sprawdź czy logarytm wydatków na leki zależy od tego czy ktoś posiada dodatkowe ubezpieczenie od pracodawcy
  1. Porównaj modele MNK oraz 2MNK
3. Wykorzystaj model z endogenicznym działaniem
4. Rozszerz powyższy model o osobne błędy losowe dla osób poddanych działaniu oraz osób niepoddanych działaniu

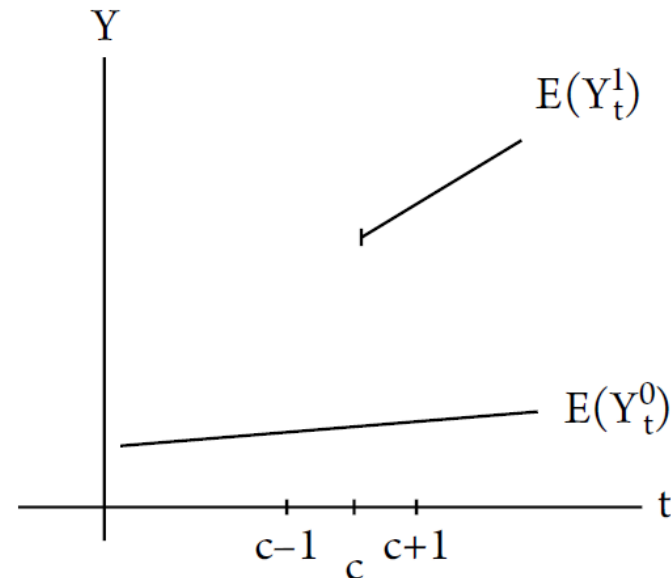
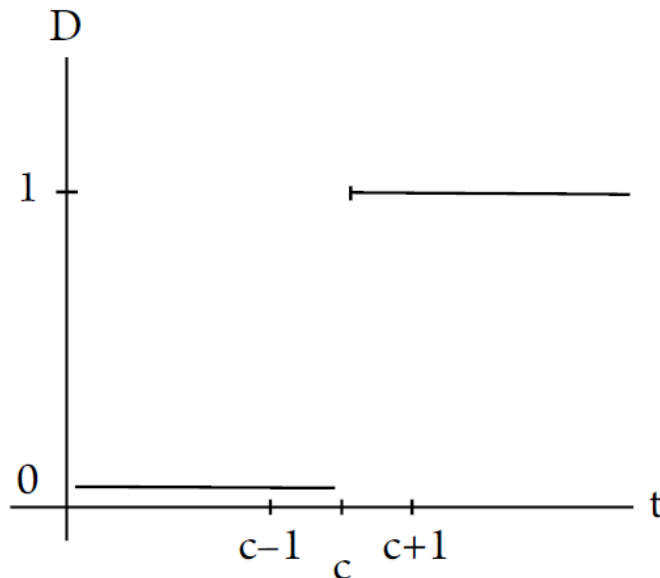
# Regresja z nieciągłością

---

- ▶ Wykorzystanie regresji oraz metod parowania pozwala uodpornić wyniki na selekcję ze względu na zmienne obserwowalne
  - ▶ Nie pomaga jednak z selekcją spowodowaną nieobserwowalnymi czynnikami
- ▶ Jedną z metod, które jest bardziej odporna i zbliżona do randomizowanego eksperymentu jest regresja z nieciągłością (ang. *regression discontinuity*)
- ▶ Metoda ta wymaga, żeby istniała ciągła zmienna  $S_i$ , taka że  $D_i = 1$  tylko jeśli  $S_i > c$ , gdzie  $c$  to jakiś ustalony poziom
  - ▶ Przykładowo jeśli działanie które badamy to pójście na studia, to  $S_i$  to może być wynik matury, a  $c$  to próg od którego przyjmują na studia

# Regresja z nieciągłością

- ▶ Idea tej metody jest taka, że w pobliżu nieciągłości (punktu  $c$ ) obserwacje po obu jego stronach powinny być do siebie zbliżone pod względem innych charakterystyk (obserwowalnych i nie)
- ▶ Np. osoby, które uzyskały 31% na maturze nie powinny się specjalnie różnić od osób, które uzyskały 29% (gdzie 30% to próg zaliczenia)



# Regresja z nieciągłością

---

- ▶ Równanie, które tu estymujemy to

$$E(Y_i | S_i) = \beta D_i + m(S_i)$$

- ▶ gdzie  $m(S_i)$  to nieznana, ciągła funkcja
- ▶ Równanie to jest estymowane dla obserwacji w pobliżu punktu nieciągłości,  $c$
- ▶ Wady regresji z nieciągłością:
  - ▶ Nie do końca wiadomo jak blisko punktu nieciągłości obserwacje powinny się znajdować
    - ▶ Im bliżej tym bardziej niedokładne oszacowanie, im dalej tym większa szansa, że inne czynniki zaczną mieć znaczenie (bias-variance trade-off)
  - ▶ Wyniki trudno generalizować na całą populację

## Zadanie 3. Wybory do senatu

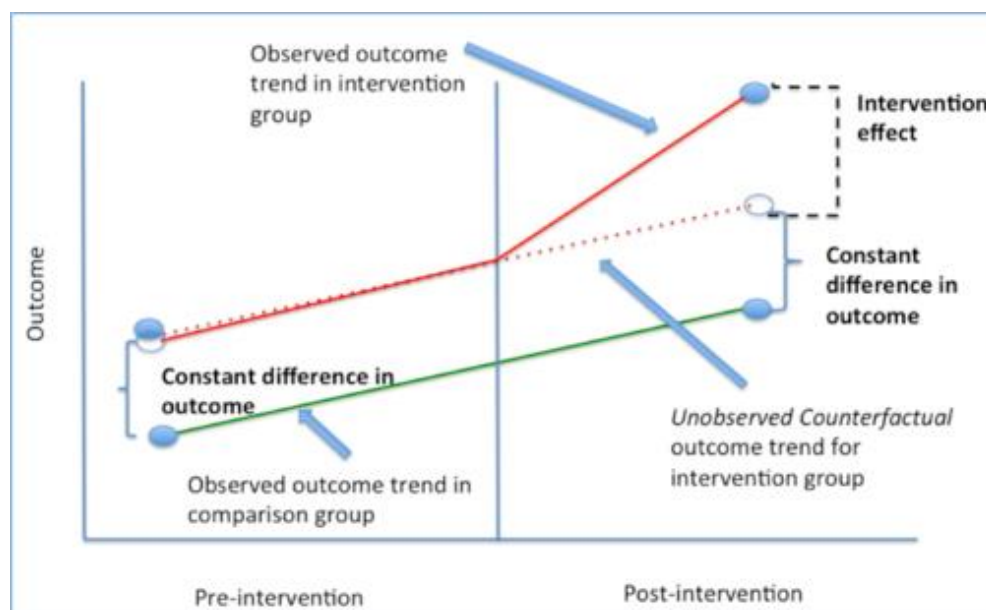
---

1. Wczytaj zbiór danych `me.senate.dta`
2. Sprawdź jak zwycięstwo partii demokratycznej w poprzednich wyborach do senatu wpływa na ich wynik w kolejnych wyborach



# Difference-in-difference

- ▶ Metodę Difference-in-difference można zastosować kiedy mamy dane dla okresu przed działaniem jak i po działaniu zarówno dla osób, które zostały poddane działaniu jak i tych, które nie zostały
  - ▶ Zakładamy, że istnieje taki sam trend dla obu grup





# Difference-in-difference

---

- ▶ Estymujemy model postaci ( $t = \{0, 1\}$ ):

$$Y_{it} = \alpha_1 D_i + \alpha_2 t + \alpha_3 D_i t + \beta X_{it} + u_{it}$$

- ▶ Parametr przy interakcji działania z czasem daje nam pożądany efekt działania:

$$\alpha_3 \approx Y_{post}^{treated} - Y_{pre}^{treated} - \left( Y_{post}^{control} - Y_{pre}^{control} \right)$$

- ▶ Panelowa postać danych pozwala nam się uodpornić na brakujące zmienne, które się nie zmieniają w czasie
- ▶ Wady:
  - ▶ Założenie o takim samym trendzie jest dość silne
    - ▶ Można próbować kontrolować różnice mając dane z większej liczby lat

## Zadanie 4. Płaca minimalna

---

1. Wczytaj zbiór danych `me.minwage.dta`
2. Sprawdź czy wzrost płacy minimalnej w New Jersey w 1992 roku spowodował spadek zatrudnienia w sektorze fast-foodów
  1. Policz model regresji liniowej
  2. Wykorzystaj modele efektów losowych i efektów stałych.



# Praca domowa ME.13 (grupy 2-3-osobowe)

---

1. W wielu państwach muzułmańskich istnieje obawa, że rządy konserwatywnych polityków mogą negatywnie wpływać na równouprawnienie kobiet. Wykorzystaj zbiór danych `me.femedu.dta` żeby sprawdzić czy wygrana polityka z partii Islamskiej w wyborach na prezydenta miasta w Turcji w 1994 (zmienna *Treat*) wpływa na % młodych kobiet, które ukończyły liceum (6 lat później, po zakończeniu kadencji – zmienna *EduWomen*).
  1. Wykorzystaj regresję z nieciągłością, gdzie jako zmienną  $S_i$  użyj zmiennej *Margin*.
    - a. Czy wyniki są wrażliwe na przedział z którego weźmiemy obserwację do regresji z nieciągłością?
  2. Porównaj wyniki z regresją liniową, w której kontrolujesz inne czynniki.
  3. Porównaj wyniki z regresją, w której traktujemy działanie jako endogeniczne (*etregress*)
    - a. Jako dodatkową zmienną w równaniu dla endogenicznego działania wykorzystaj *partygr2* (przyjmuje 1 jeśli było więcej niż dwóch kandydatów na burmistrza)
    - b. Rozszerz ten model o różne błędy losowe w równaniu regresji w zależności od tego czy osoba została poddana działaniu czy nie

```
set seed 10+"Nr indeksu"  
sample 90, by(Treat)
```