

Ocena skutków interwencji

Wykład 3

Piotr Dybka

Ekonometria praktyczna

- Zmiana strukturalna może też mieć charakter **interwencji**, czyli sytuacji, w której podjęto działanie mające zmienić badane zjawisko.
- W ekonomii przykładem takiej sytuacji jest zmiana prawa lub wprowadzenie nowego programu gospodarczego; w marketingu – zmiana charakterystyk produktu; w medycynie – podjęcie nowego sposobu leczenia.
- Tego typu zmiany często są analizowane na podstawie eksperymentów (np. pewna próbka konsumentów korzysta z produktu w wersji A, inna grupa w wersji B itp.).
- Jednak w ekonomii przeprowadzanie eksperymentów jest mocno utrudnione (w szczególności dotyczy to makroekonomii), dlatego czasami poszukuje się sytuacji, gdzie można zaobserwować naturalne warunki przypominające eksperyment.
- Poniżej prezentuję tylko kilka wybranych przykładów (ilustracyjnie).

Estymator difference-in-differences

- Załóżmy, że mamy do czynienia z sytuacją stanowiącą eksperyment naturalny, np. losowe osoby objęte nową polityką.
- Przykładem takiego badania jest analiza Carda i Kruegera (1994), którzy badali wpływ płacy minimalnej na zatrudnienie. Autorzy porównywali poziom zatrudnienia w branży *fast food* w New Jersey (gdzie podniesiono płacę minimalną) z sytuacją w sąsiadującej Pensylwanii (gdzie nie wprowadzono takiej regulacji).
- Możemy zacząć od utworzenia zmiennej binarnej S_i , określającej, czy dana osoba była objęta działaniem polityki (*treatment*):

$$S_i = \begin{cases} 1 & \text{gdy osoba znajduje się w grupie objętej działaniem} \\ 0 & \text{gdy osoba znajduje się w grupie kontrolnej} \end{cases}$$

Zobacz: David Card, Alan B. Krueger, *Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania*, „American Economic Review” 84 (4), 1994, s. 772–793.

Estymator difference-in-differences

- Efekt objęcia działaniem można przedstawić jako:

$$y_i = \beta_0 + \beta_1 S_i + \epsilon_i$$

- Gdzie ϵ_i obejmuje pozostałe czynniki wpływające na y_i . Rezultaty regresji dla grupy objętej działaniem i grupy kontrolnej można przedstawić jako:

$$E(y_i) = \begin{cases} \beta_0 + \beta_1 & \text{gdy } S_i = 1 \\ \beta_0 & \text{gdy } S_i = 0 \end{cases}$$

- Naszym celem jest oszacowanie efektu oddziaływania (β_1), który można oszacować KMNK:

$$b_1 = \frac{\sum_{i=1}^N (S_i - \bar{S})(y_i - \bar{y})}{\sum_{i=1}^N (S_i - \bar{S})^2} = \bar{y}_1 - \bar{y}_0$$

Gdzie $\bar{y}_1 = \sum_{i=1}^{N_1} y_i / N_1$, czyli średnia wartość y w grupie objętej oddziaływaniem ($S_i = 1$), a $\bar{y}_0$ to średnia wartość y w grupie kontrolnej. Estymator b_1 nazywamy **estymatorem różnicy** (*difference estimator*).

- Oznaczenie b_1 ma podkreślić, że nie jest to jeszcze nieobciążony estymator efektu oddziaływania β_1 !

Estymator difference-in-differences

- Estymator różnicy możemy też zapisać jako:

$$b_1 = \beta_1 + \frac{\sum_{i=1}^N (S_i - \bar{S})(\epsilon_i - \bar{\epsilon})}{\sum_{i=1}^N (S_i - \bar{S})^2} = \beta_1 + (\bar{\epsilon}_1 - \bar{\epsilon}_0)$$

- Na podstawie powyższego równania możemy zauważyć, że estymator różnicy jest równy efektowi oddziaływania (β_1) powiększonemu o różnicę średnich wartości czynników nieobserwowanych w modelu w grupie poddanej oddziaływaniu i w grupie kontrolnej ($\bar{\epsilon}_1 - \bar{\epsilon}_0$).
- Oznacza to, że jeżeli estymator KMNK ma być nieobciążony, to musi być spełniony warunek:

$$E(\bar{\epsilon}_1 - \bar{\epsilon}_0) = E(\bar{\epsilon}_1) - E(\bar{\epsilon}_0) = 0$$

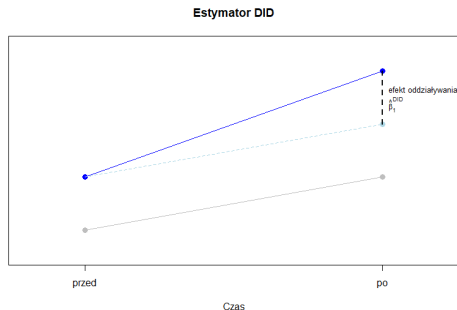
- Innymi słowy: wartość oczekiwana wszystkich pozostałych czynników wpływających na badaną zmienną (poza efektem oddziaływania) musi być **równa** dla grupy poddanej oddziaływaniu i grupy kontrolnej.
- Dlatego przynależność do poszczególnych grup musi mieć charakter **losowy**; w przeciwnym razie estymator będzie obciążony (*selection bias*).

Estymator difference-in-differences

- Estymator ten może być także rozszerzony o wymiar czasowy, wówczas:

$$b_1 = (\bar{y}_{11} - \bar{y}_{10}) - (\bar{y}_{01} - \bar{y}_{00})$$

- Gdzie $\bar{y}_{i,t}$ oznacza średnią wartość dla grupy i ($1 =$ objęta oddziaływaniem, $0 =$ kontrolna) w czasie t ($0 =$ przed, $1 =$ po). Odejmujemy więc zmianę w grupie kontrolnej od zmiany w grupie objętej działaniem.
- Założenie o losowym doborze próby pozostaje kluczowe, ale jego postać jest rozszerzona do **założenia o równoległym trendzie**: wartość y zmieniałaby się dokładnie w taki sam sposób (z takim samym trendem) w obu grupach przy założeniu braku oddziaływania.



DiD – na co uważać we współczesnych zastosowaniach

- Powyższy schemat to klasyczny przypadek 2×2 : dwie grupy, dwa okresy, jeden moment wejścia w życie.
- W praktyce polityki wchodzi w życie **w różnych momentach w różnych jednostkach** (*staggered adoption*). Standardowa regresja z efektami stałymi (TWFE) przestaje wtedy szacować sensowny średni efekt: jako „grupa kontrolna” wykorzystywane są m.in. jednostki *już objęte* działaniem.
- Goodman-Bacon (2021) pokazał, że estymator TWFE jest średnią ważoną wszystkich możliwych porównań 2×2 – a część wag może być **ujemna**.
- Estymatory odporne na ten problem: Callaway i Sant’Anna, Sun i Abraham, de Chaisemartin i D’Haultfœuille – zob. przypis. W R m.in. pakiety *did*, *fixest* (*sunab*), *didimputation*.
- Standardem raportowania jest dziś także wykres *event study* – oszacowania efektu w kolejnych okresach przed i po interwencji. Brak istotnych efektów *przed* interwencją to podstawowa (choć niewystarczająca) przesłanka na rzecz założenia równoległych trendów.

Zobacz: A. Goodman-Bacon, *Difference-in-differences with variation in treatment timing*, „Journal of Econometrics” 225(2), 2021, s. 254–277; B. Callaway, P.H.C. Sant’Anna, *Difference-in-Differences with multiple time periods*, „Journal of Econometrics” 225(2), 2021, s. 200–230; L. Sun, S. Abraham, *Estimating dynamic treatment effects in event studies with heterogeneous treatment effects*, „Journal of Econometrics” 225(2), 2021, s. 175–199; C. de Chaisemartin, X. D’Haultfœuille, *Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects*, „American Economic Review” 110(9), 2020, s. 2964–2996.

- Zamiast jednego parametru β_1 szacujemy efekt **osobno dla każdego okresu** względem momentu interwencji:

$$y_{it} = \alpha_i + \lambda_t + \sum_{k \neq -1} \theta_k D_{it}^k + \epsilon_{it}$$

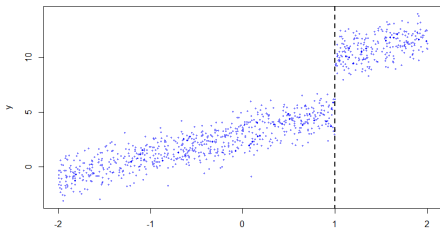
gdzie $D_{it}^k = 1$, gdy jednostka i znajduje się k okresów od momentu wejścia interwencji w życie. Okres $k = -1$ jest pomijany jako kategoria bazowa.

- Oszacowania $\hat{\theta}_k$ dla $k < 0$ służą ocenie założenia o równoległych trendach: powinny być **nieistotne statystycznie**. Istotne efekty „przed” oznaczają, że grupy różniły się już wcześniej.
- Oszacowania dla $k \geq 0$ pokazują **dynamikę** efektu – czy narasta, wygasa, czy pojawia się z opóźnieniem.
- **Uwaga:** brak istotnych efektów przed interwencją jest przesłanką *konieczną*, ale *niewystarczającą* – równoległość trendów w przeszłości nie gwarantuje jej w okresie po interwencji.
- W R: `fixest::feols()` z `i(czas_do_zdarzenia, grupa, ref=-1)` oraz `iplot()`.

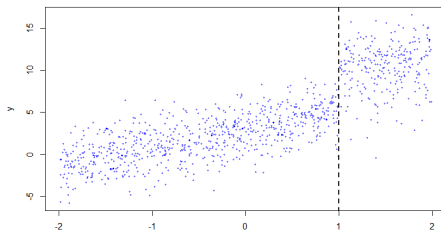
Regression discontinuity design (RDD)

Idea RDD polega na tym, że szukamy takich obszarów, gdzie występuje znacząca zmiana badanej zmiennej w efekcie przekroczenia pewnego progu (np. wartości dochodu, liczby punktów z egzaminu, obszaru geograficznego itp.).

„ostre” (*sharp*) przejście



„nieostre” (*fuzzy*) przejście



Sharp regression discontinuity design (SRDD)

- Załóżmy, że mamy model:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 S_i + \epsilon_i$$

- gdzie x_i to **zmienna progowa** (*forcing / running variable*), r to wartość progu, a S_i to zmienna binarna opisująca objęcie oddziaływaniem:

$$S_i = \begin{cases} 1 & \text{gdy } x_i \geq r \\ 0 & \text{gdy } x_i < r \end{cases}$$

- Czyli wystąpienie efektu oddziaływania ma miejsce, gdy zostanie przekroczony próg r ciągłej zmiennej x_i ; szukanym efektem jest β_2 .
- Idea RDD polega na tym, aby wykorzystać obserwacje znajdujące się **w otoczeniu** r . W wariancie ostrym zakładamy, że wszystkie obserwacje poniżej progu nie są objęte oddziaływaniem, a wszystkie obserwacje powyżej progu są nim objęte.
- Szacujemy tzw. LATE – *local average treatment effect* (lokalny efekt oddziaływania): lokalny, bo dotyczy jednostek znajdujących się blisko progu.

Fuzzy regression discontinuity design (FRDD)

- W przypadku nieostrej (*fuzzy*) RDD zakładamy, że obserwacje są poddane oddziaływaniu z **pewnym prawdopodobieństwem**, np. po przekroczeniu progu prawdopodobieństwo podlegania oddziaływaniu wynosi 80% (np. wskutek wyjątku, który zwalnia 20% osób powyżej progu).
- Trzeba wtedy odróżnić *przekroczenie progu* od *faktycznego oddziaływania*. Wprowadzamy zmienną:

$$Z_i = \begin{cases} 1 & \text{gdy } x_i \geq r \\ 0 & \text{gdy } x_i < r \end{cases}$$

- Z_i mówi, czy obserwacja **przekroczyła próg** (czyli czy *powinna* być objęta działaniem), natomiast S_i nadal opisuje, czy **faktycznie** została nim objęta.
- Z_i jest **zmienną instrumentalną** dla S_i : wpływa na objęcie oddziaływaniem, ale nie wpływa bezpośrednio na y_i .
- Idea FRDD polega na uwzględnieniu efektu tylko dla tych obserwacji, gdzie oddziaływanie faktycznie wystąpiło. Gdybyśmy potraktowali takie dane jak sharp RDD, oszacowanie efektu byłoby **zaniżone** (zmniejszy się jego skala).

- **Wybór pasma** (*bandwidth*) to kompromis obciążenie–wariancja: szersze pasmo to więcej obserwacji, ale też większe obciążenie. Standardy: Imbens i Kalyanaraman (2012) oraz Calonico, Cattaneo i Titiunik (2014).
- **Wnioskowanie**: przy paśmie optymalnym dla estymacji punktowej standardowe przedziały ufności są obciążone. Dlatego raportuje się dziś przedziały z korektą obciążenia (*robust bias-corrected*).
- **Testy diagnostyczne**: test McCrary’ego (czy jednostki nie manipulują zmienną progową, gromadząc się tuż za progiem), testy placebo dla fałszywych progów, sprawdzenie ciągłości zmiennych kontrolnych w punkcie r .
- **W R**: pakiet `rddtools` – `rdd_data()`, `rdd_reg_np()` (nieparametrycznie), `rdd_reg_lm()` (parametrycznie), `rdd_bw_ik()`, `dens_test()`. Alternatywa: `rdrobust`.
- **Uwaga**: pakiet `rdd` (funkcja `RDestimate`) został zdjęty z CRAN 10.07.2025 i nie należy go już używać.

Metoda kontroli syntetycznej – idea

- DiD wymaga grupy kontrolnej, która ma *równoległy trend*. A co, jeśli interwencja dotyczy **jednej jednostki** (kraju, regionu, miasta) i żadna pojedyncza jednostka nie jest do niej dostatecznie podobna?
- **Pomysł** (Abadie i Gardeazabal 2003; Abadie, Diamond, Hainmueller 2010): zamiast wybierać jedną grupę kontrolną, **zbudujmy ją** jako średnią ważoną wielu jednostek nieobjętych interwencją (tzw. *donor pool*).
- Wagi dobieramy tak, aby jednostka syntetyczna jak najlepiej odtwarzała jednostkę badaną **przed** interwencją – zarówno pod względem zmiennej objaśnianej, jak i predyktorów.
- Efekt interwencji to **luka** między jednostką rzeczywistą a syntetyczną po momencie interwencji.
- Metoda ta jest naturalnym uogólnieniem DiD: DiD przypisuje wszystkim jednostkom kontrolnym równe wagi, SCM dobiera je optymalnie.

Metoda kontroli syntetycznej – zapis formalny

Mamy $J + 1$ jednostek: jednostkę 1 (objętą interwencją w momencie T_0) oraz J jednostek donor pool. Szukamy wektora wag $\mathbf{W} = (w_2, \dots, w_{J+1})'$:

$$\min_{\mathbf{W}} \|\mathbf{X}_1 - \mathbf{X}_0 \mathbf{W}\|_{\mathbf{V}} = \sqrt{(\mathbf{X}_1 - \mathbf{X}_0 \mathbf{W})' \mathbf{V} (\mathbf{X}_1 - \mathbf{X}_0 \mathbf{W})}$$

przy ograniczeniach $w_j \geq 0$ oraz $\sum_j w_j = 1$.

- $\mathbf{X}_1$ – wektor predyktorów jednostki badanej (średnie z okresu przed interwencją),
 $\mathbf{X}_0$ – macierz predyktorów jednostek donor pool.
- $\mathbf{V}$ – macierz diagonalna wag *predyktorów*; dobierana tak, aby minimalizować błąd dopasowania zmiennej objaśnianej w okresie przed interwencją (RMSPE).
- Ograniczenia $w_j \geq 0$ i $\sum w_j = 1$ są kluczowe: zapobiegają ekstrapolacji poza zakres danych i zwykle dają **rozwiązanie rzadkie** (kilka jednostek z dodatnią wagą).
- Efekt interwencji: $\hat{\tau}_t = Y_{1t} - \sum_{j=2}^{J+1} w_j^* Y_{jt}$ dla $t > T_0$.

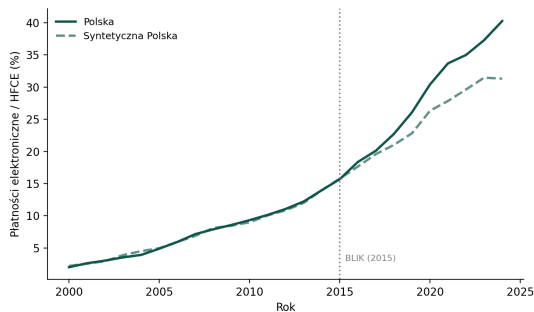
- **Długi okres przed interwencją.** Dobre dopasowanie na krótkiej próbie nic nie znaczy – można je uzyskać przypadkiem. Abadie (2021) zaleca wiele okresów przed interwencją.
- **Dobre dopasowanie przed interwencją.** Jeżeli RMSPE przed interwencją jest duże, luka po interwencji nie ma interpretacji przyczynowej.
- **Brak interferencji.** Jednostki donor pool nie mogą być objęte tą samą (ani podobną) interwencją, ani odczuwać jej efektów ubocznych.
- **Brak szoków idiosynkratycznych** w okresie po interwencji, które dotyczyłyby tylko jednostki badanej.
- **Wnioskowanie** nie opiera się na klasycznych błędach standardowych (mamy jedną jednostkę objętą interwencją!), lecz na **testach permutacyjnych**: przypisujemy interwencję kolejno każdej jednostce donor pool i sprawdzamy, jak nietypowa jest luka dla jednostki rzeczywiście objętej interwencją.

Przykład: wpływ systemu BLIK na cyfryzację płatności

- **Pytanie:** czy krajowy standard płatności mobilnych tylko *przenosi* transakcje z kart na telefon (substytucja), czy *tworzy nowy rynek*, zamieniając płatności gotówkowe na cyfrowe?
- **Interwencja:** uruchomienie BLIK-a w lutym 2015 r. przez konsorcjum sześciu banków (Polski Standard Płatności).
- **Zmienna objaśniana:** wartość elektronicznych płatności detalicznych jako % spożycia gospodarstw domowych (HFCE).
- **Donor pool:** 21 gospodarek europejskich bez własnego standardu płatności mobilnych w badanym okresie. Wykluczono m.in. Danię (MobilePay, 2013) i Szwecję (Swish, 2012).
- **Predyktory (5):** PKB per capita, indeks rozwoju finansowego (IMF), dostęp do internetu, odsetek osób z wyższym wykształceniem, indeks rządów prawa.
- **Okres:** 2000–2024, moment interwencji: 2015.

Źródło: Dybka, Bartha, Łopusiński, Kotkowski, *Substitution or market creation? Causal evidence from Poland's domestic mobile payment scheme* (2026).

BLIK: Polska a Syntetyczna Polska



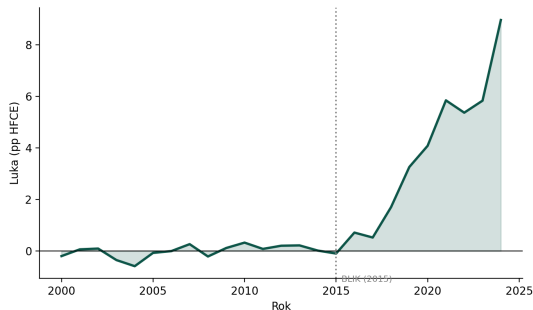
Syntetyczna Polska to:

$$\begin{aligned} &0,388 \cdot \text{Węgry} \\ &+ 0,359 \cdot \text{Łotwa} \\ &+ 0,124 \cdot \text{Bułgaria} \\ &+ 0,071 \cdot \text{Albania} \\ &+ 0,058 \cdot \text{Grecja} \end{aligned}$$

Pozostałe 16 krajów otrzymuje wagę **zero** – to typowa cecha SCM.

RMSPE przed interwencją: 0,24 pp.
Obie serie pokrywają się do 2015 r.,
a następnie systematycznie się
rozchodzą.

BLIK: oszacowana luka



- Luka rośnie monotonicznie:
 $\approx 5,8$ pp HFCE w 2023 r. i
 $\approx 8,9$ pp w 2024 r.
- Przed 2015 r. luka oscyluje wokół zera – to warunek wiarygodności wyniku.
- Przyspieszenie od 2018 r. zbiega się z rozwojem efektów sieciowych, a od 2021 r. z wprowadzeniem płatności zbliżeniowych.
- **Uwaga:** sama luka to efekt *łączny*. Trzeba jeszcze oddzielić od niej równoległą interwencję – Program Wsparcia Obrotu Bezgotówkowego (dopłaty do terminali od 2018 r.), który odpowiada za $\approx 2,7$ pp.

Kontrola syntetyczna – testy odporności

Ponieważ nie mamy klasycznych błędów standardowych, wiarygodność wyniku opiera się na testach:

- **Placebo „w przestrzeni”** (*in-space*): przypisujemy interwencję kolejno każdemu krajowi donor pool. Jeżeli luka dla Polski jest nietypowo duża na tle rozkładu luk placebo, wynik jest wiarygodny. Daje to p -value permutacyjne, np.
 $p = 1/22 \approx 0,045$.
- Porównujemy *iloraz* RMSPE po i przed interwencją – kraje o słabym dopasowaniu przed interwencją mają duże luki z definicji i nie powinny zaniżać istotności.
- **Placebo „w czasie”** (*in-time*): przesuwamy moment interwencji wstecz (np. na 2012 r.). Luka nie powinna pojawić się przed rzeczywistą datą.
- **Leave-one-out**: usuwamy kolejno każdą jednostkę z dodatnią wagą i sprawdzamy, czy wynik nie zależy od jednego kraju.
- **Zmiana zestawu predyktorów oraz zmiana okresu dopasowania.**

- **Macierz V bywa słabo zidentyfikowana.** Rozwiązanie minimalizujące RMSPE często leży „w rogu” przestrzeni parametrów (jeden predyktor dominuje). Różne pakiety i różne optymalizatory mogą dać różne wagi przy podobnym dopasowaniu – warto raportować wrażliwość wyniku na wybór V .
- **Obciążenie regularyzacyjne:** gdy jednostka badana leży blisko brzegu zbioru danych, dopasowanie jest niedoskonałe i estymator jest obciążony. Stosuje się wtedy warianty z korektą (Abadie 2021; *penalized/bias-corrected* SCM).
- **Jedna jednostka objęta interwencją** oznacza brak klasycznego wnioskowania; p -value permutacyjne ma ograniczoną rozdzielczość ($1/(J + 1)$).
- **Interpolacja vs. ekstrapolacja:** jeżeli jednostka badana ma skrajne wartości predyktorów, żadna kombinacja wypukła jednostek donor pool jej nie odtworzy.
- **Porównywalność pomiaru:** zmienna objaśniana musi być mierzona tak samo we wszystkich krajach – inaczej luka miesza efekt przyczynowy z różnicą definicji.

Kontrola syntetyczna w R

- `Synth` – implementacja referencyjna autorów metody (Abadie, Diamond, Hainmueller); funkcje `dataprep()` i `synth()`.
- `tidysynth` – nowszy interfejs w stylu *tidyverse*; cały proces w jednym potoku, a placebo i p -value permutacyjne generowane automatycznie (`generate_placebos = TRUE`). Tego pakietu używamy na ćwiczeniach.
- `augsynth` – wariant z regularyzacją grzbietową (*ridge-augmented* SCM, Ben-Michael, Feller, Rothstein), przydatny gdy dopasowanie przed interwencją jest niedoskonałe.
- `scpi` – przedziały ufności dla SCM oparte na innej filozofii wnioskowania (Cattaneo, Feng, Titiunik).
- `gsynth` – uogólnienie na wiele jednostek objętych interwencją i różne momenty wejścia w życie (łączy SCM z modelem czynnikowym).
- W Stacie odpowiednikiem jest `synth` oraz `allsynth` (Wiltshire), który dodaje korektę obciążenia i testy permutacyjne.