

## ŁĄCZENIE DANYCH Z DYNAMICZNYM OBCIĘCIEM. WYNIKI WSTĘPNE<sup>1</sup>

**Paweł Strawiński**

Katedra Statystyki i Ekonometrii UW  
e-mail: pstrawinski@wne.uw.edu.pl

**Streszczenie:** Problem ewaluacji działań jest problemem braku danych. Zatem ważnym zagadnieniem jest konstrukcja adekwatnej kontrfaktycznej grupy porównawczej. W artykule zaproponowane modyfikację procedury pozwalającej uznać obserwację za dobrze dobrany stan kontrfaktyczny. Proponowana zmiana polega na wprowadzeniu nowego mechanizmu sprawdzania czy jednostka z grupy kontrolnej jest bliskim odpowiednikiem jednostki z grupy porównawczej. W ostatniej części artykułu zaprezentowano wyniki symulacji badającej właściwości statystyczne proponowanego rozwiązania. Uzyskane wyniki wskazują, że metoda pozwala osiągnąć lepsze rezultaty niż dotychczas wykorzystywane techniki

**Słowa kluczowe:** łączenie danych, matching, propensity score, caliper, ewaluacja.

### WSTĘP

W literaturze dotyczącej oceny różnego typu działań o charakterze zorganizowanych programów w ostatnich latach jako narzędzie badawcze wykorzystywane są metody quasi-eksperymentalne. Polegają one na odtworzeniu na podstawie danych nie eksperymentalnych warunków eksperymentu.

Badacz obserwuje efekt działania programu i dokonuje oceny jego skuteczności na podstawie oceny pewnych wielkości. W danej chwili rozpatrywana jednostka może znajdować się w jednym z dwóch alternatywnych stanów, ale nie może być w obu jednocześnie. Może być uczestnikiem programu, bądź w nim nie uczestniczyć. Gdyby dane pochodziły z eksperymentu wystarczyłoby porównać wynik jednostek z grupy eksperymentalnej i grupy kontrolnej. Zatem problem oceny programów jest problemem braku danych [Heckman i in. 1997].

---

<sup>1</sup> Badanie finansowane ze środków na naukę w latach 2008-2009.

Ocena wpływu działania na obserwowany rezultat zawiera spekulację w jaki sposób jednostka by funkcjonowała, gdyby nie została poddana oddziaływaniu [Caliendo, Kopeinig 2008]. W literaturze statystycznej i ekonometrycznej opisanych jest szereg metod pozwalających na utworzenie grupy porównawczej. Występują one pod nazwą metod łączenia. Najpopularniejszą jest metoda wykorzystująca wartości wektora *propensity score*. Wykorzystywane są różne warianty tej metody różniące się sposobami łączenia obserwacji w pary, warunkami uznania stopnia podobieństwa obserwacji za wystarczający, oraz sposobami obliczenia efektu oddziaływania. Każda z implementacji z jednej strony umożliwia wyznaczenie kontrfaktycznego zachowania jednostki, z drugiej, pomimo tego, że metody są asymptotycznie równoważne [Hirano i in. 2003], prowadzi do uzyskania innego wyniku.

W dalszej części artykułu zostaną w skrócie omówione wady i zalety dotychczas wykorzystywanych podejść do konstrukcji grupy porównawczej i wyznaczania na jej podstawie stanu kontrfaktycznego. Głównym celem artykułu jest udoskonalenie jednej z metod stosowanych do łączenia danych. Proponowana zmiana polega na wykorzystaniu zmodyfikowanego mechanizmu sprawdzania czy jednostka z grupy kontrolnej jest bliskim odpowiednikiem jednostki z grupy poddanej oddziaływaniu. Następnie za pomocą symulacji porównamy dotychczas stosowane w badaniach metody z zaproponowaną modyfikacją. Celem dodatkowym jest zbadanie właściwości statystycznych, szczególnie małopróbkowych, różnych metod łączenia danych oraz identyfikacja źródeł wariancji.

## METODY ŁĄCZENIA DANYCH

Przeprowadzając analizę ekonomiczną działania pewnej polityki jako jego efekt obserwowana jest dla każdej jednostki  $i$  wartość zmiennej wynikowej. Wynosi ona  $Y_{1i}$  jeżeli jednostka jest poddana działaniu, lub  $Y_{0i}$  jeżeli polityka na nią nie oddziałuje. Jeżeli można by było jednocześnie obserwować obie wartości dla każdej poddanej badaniu jednostki problem związany z oceną programu zniknąłby.

W wielu pracach empirycznych dokonuje się oceny efektów polityki opisując ją jako program. Celem jest pomiar przeciętnego efektu zastosowania programu. Niech zmienna  $P$  opisuje stan uczestnictwa w programie i przyjmuje wartość  $P_i=1$ , gdy jednostka  $i$  jest uczestnikiem programu (należy do grupy eksperymentalnej), a  $P_i=0$  oznacza, że znajduje się poza programem. Wówczas przeciętny efekt działania programu (ang. *average treatment effect*) można zapisać jako:

$$Y_{ATE} = \frac{1}{N} \left[ \sum_{i=1}^N (P_i Y_{1i} + (1 - P_i) Y_{0i}) \right] \quad (1)$$

Efekt programu obliczony w sposób opisany przez wzór (1) jest efektem dla przeciętnej jednostki pochodzącej z badanej próby, bez uwzględnienia czy jest ona poddana oddziaływaniu czy nie. W praktyce jako wielkość efektu przyjmuje się różnicę w średnim wyniku dla grupy eksperymentalnej i wszystkich jednostek nie-

uczestniczących w eksperymencie. Może to spowodować niedokładność, ponieważ jednostki z grupy eksperymentalnej mogą w sposób systematyczny odróżniać się od pozostałych, co przy pomiarze nie jest uwzględnione.

W praktyce badawczej dąży się do obliczenia innej miary, niosącej więcej informacji o skuteczności działania. Jest nią przeciętny efekt oddziaływania programu na jednostki poddane oddziaływaniu (ang. *average treatment on treated*). Jest to efekt ograniczony do grupy, na którą oddziaływano.

$$Y_{ATT} = E(Y_{1i} - Y_{0i} | P_i = 1) = E(Y_{1i} | P_i = 1) - E(Y_{0i} | P_i = 1) \quad (2)$$

Pokazuje on ilościowy efekt oddziaływania programu wobec jednostek w nim uczestniczących. Jest on równy przeciętnej zmianie wartości zmiennej wynikowej wskutek wystąpienia oddziaływania. Informuje ona jaki jest zysk bądź strata z faktu uczestniczenia w programie dla jego beneficjentów. Należy zaznaczyć, że wielkość  $E(Y_{1i}|P_i=1)$  jest oczekiwaną wartością zmiennej wynikowej dla beneficjentów programu i jest wartością obserwowaną, natomiast  $E(Y_{0i}|P_i=1)$  jest wielkością dla nieobserwowanego stanu kontrfaktycznego.

Łączenie obserwacji polega na obliczeniu miary podobieństwa, a następnie określeniu według wybranego algorytmu, która obserwacja z grupy kontrolnej jest najbardziej zbliżona do obserwacji z grupy eksperymentalnej. Celem łączenia obserwacji jest skonstruowanie poprawnej grupy porównawczej zastępującej brakujące dane, bez konieczności przyjmowania dodatkowych założeń *a-priori* [Blundell i Costa Diás 2000]. Jego zadaniem jest odtworzenie warunków eksperymentu poprzez taki wybór jednostek do grupy kontrolnej, by była jak najbardziej zbliżona pod względem obserwowanych charakterystyk do grupy poddanej oddziaływaniu. Metody łączenia koncentrują się na sposobach konstrukcji grupy referencyjnej wykorzystując techniki nieparametryczne w celu transformacji rzeczywistych danych zebranych podczas badania przekrojowego w zbiór danych quasi-eksperymentalnych. Dzięki wykorzystaniu metod nieparametrycznych maksymalne i efektywne wykorzystywane są dostępne informacje zawarte w danych empirycznych. Poprawna konstrukcja sztucznej grupy odniesienia zapewnia zredukowanie obciążenia i zwiększenie precyzji uzyskanych wyników [Rubin 1973]. Podobnie jak szereg innych elementów teorii ekonometrii, jej korzenie i pierwsze zastosowania tkwią w biologii i naukach medycznych.

Wśród metod łączenia danych można wyróżnić dwa podejścia. Historycznie pierwszą metodą było łączenie według wartości współzmiennych (ang. *covariates matching*). Współzmiennymi nazywa się zmienne używane w procedurze łączenia. Ich nazwa wywodzi się stąd, że różne ich wartości determinują zarówno jeden ze stanów oddziaływania programu, jak również wartość zmiennej mierzącej efekt ilościowy programu. Zatem ich wartość zmienia się wspólnie z wartością zmiennej określającej stan oddziaływania.

Łączenie według wartości współzmiennych polega na łączeniu w pary obserwacji o identycznych lub zbliżonych wartościach współzmiennych. Jeżeli war-

tości zmiennych zostaną przedstawione jako punkty przestrzeni wielowymiarowej, to statystyczna metoda łączenia polega na podziale tej przestrzeni na klatki według wartości współzmiennych. Następnie łączone są nie obserwacje a utworzone klatki w pary. Niestety, jeżeli wśród współzmiennych występują zmienne ciągle albo jest duża liczba współzmiennych to metoda okazuje się niepraktyczna w zastosowaniu, gdyż duża liczba klatek pozostaje pusta, lub klatki są zbyt duże. Zmniejsza to efektywność łączenia i zwiększa wariancję estymatora wielkości oddziaływania.

Problem znalezienia elementu porównawczego w bliskim sąsiedztwie jest problemem wielowymiarowym. Metodą pozwalającą na uniknięcie konieczności porównywania wszystkich cech badanych jednostek jest użycie zamiast cech pewnej miary podobieństwa obserwacji. Najpopularniejsza i najszerzej stosowaną jest wektor *propensity score* [Rosenbaum i Rubin 1983]. *Propensity score* jest wartością pochodnej funkcji prawdopodobieństwa doboru obserwacji do programu obliczona względem wektora cech jednostki. Wartość wektora dla jednostki jest równoważna *a-priori* prawdopodobieństwu zdarzenia, że jednostka należy do grupy eksperymentalnej poddanej działaniu [Coniffe i in. 2000].

Metoda łączenia danych oparta jest na dwóch założeniach [Rosenbaum i Rubin 1983]. Pierwsze z nich jest nazywane warunkową niezależnością (ang. *Conditional Independence Assumption*). Zakłada się, że badacz dysponuje zbiorem czynników warunkujących  $W$ , nazywanych współzmiennymi, takich, że w każdej warstwie<sup>2</sup> zdefiniowanej przez podzbiór  $w \in W$  rozkład obserwowanych cech jednostek jest taki sam w grupie na którą oddziaływano i w grupie kontrolnej.

$$(Y_0, Y_1) \perp P \mid W \quad (4)$$

Oznacza to, że realizacje zmiennej wynikowej są warunkowo względem zastosowanych w procedurze łączenia współzmiennych  $W$  niezależne od statusu uczestnictwa w programie. Znajomość stanu uczestnictwa w programie warunkowo względem cech  $w$  nie powinna determinować wartości zmiennej wynikowej. Drugim założeniem jest, że wszystkie jednostki poddane działaniu programu posiadają swój odpowiednik znajdujący się poza programem.

$$0 < \Pr(P_i = 1 \mid W) < 1 \quad (5)$$

Oznacza to, że dla każdej wartości ze zbioru zmiennych warunkujących  $W$ , i dla każdego uczestnika programu, istnieje z niezerowym prawdopodobieństwem co najmniej jedna jednostka znajdująca się poza programem o identycznych cechach. Ponadto, cechy jednostek nie determinują w sposób jednoznaczny czy jednostka należy do grupy eksperymentalnej czy do grupy kontrolnej.

Z założeń (4) i (5) wynika, że gdy znana jest wartości wektora zmiennych warunkujących  $w$ , to oczekiwana wartość zmiennej wynikowej jest warunkowo niezależna od statusu uczestnictwa w programie. Zatem obliczając różnicę w śred-

<sup>2</sup> Warstwa jest odpowiednikiem klatki.

nich wartościach pomiędzy grupami można uzyskać oszacowanie przeciętnego efektu oddziaływania na jednostki poddane oddziaływaniu.

Jeżeli przydzielenie do grupy eksperymentalnej lub grupy kontrolnej jest warunkowo niezależne od obserwowanych cech jednostki, czyli spełnione jest założenie (4), to obciążenie powodowane przez porównanie rezultatu w dwóch grupach wynikające z różnych wartości współzmiennych może zostać wyeliminowane poprzez ich dopasowanie [Hirano i in. 2003]. Rosenbaum i Rubin [1983] pokazali, że zamiast cech  $W$  można zastosować wektor *propensity score*  $PS(w)$ .

$$PS(w) = \Pr(P = 1 | w) \quad (6)$$

Użycie wektorów redukuje problem wielowymiarowy do jednowymiarowego. Zatem analogicznie do (4) zachodzi

$$(Y_0, Y_1) \perp P | PS(w) \quad (7)$$

W takim przypadku uzyskana wartość oszacowania będzie nadal nieobciążonym estymatorem efektu, jednak w ten sposób uzyskany estymator nie musi być efektywny [Heckman i in. 1997]. Nieefektywność metody *propensity score* pojawia się w przypadku ewaluacji danych pochodzących z eksperymentu [Hahn 1998]. Jest ona związana z niepotrzebnym zastosowaniem przekształcenia danych.

Łączenie danych z wykorzystaniem wybranej techniki w większym stopniu wpływa na charakterystyki grupy kontrolnej niż eksperymentalnej. Grupa eksperymentalna ulega modyfikacji tylko w przypadku, gdy dla poszczególnych jednostek należących do tej grupy nie udaje się znaleźć bliskiego odpowiednika w grupie kontrolnej. Z kolei charakterystyki grupy kontrolnej przed i po wykonaniu łączenia różnią się w sposób znaczący. Celem łączenia jest wybranie ze zbioru potencjalnych jednostek kontrolnych tych najbardziej zbliżonych do jednostek z grupy eksperymentalnej. Zatem grupa kontrolna po łączeniu powinna być bardziej zbliżona pod względem charakterystyk użytych do łączenia do grupy eksperymentalnej niż przed zastosowaniem procedury łączenia

Jest to ogólna metoda posiadająca wiele implementacji, różniących się sposobem obliczenia miary podobieństwa. Po pierwsze, w metodzie *propensity score* można wybrać różne postaci funkcyjne dla prawdopodobieństwa poddania oddziaływaniu. Po drugie, można zastosować jedną z dostępnych alternatywnych miar podobieństwa obserwacji. Typowy estymator oparty o połączone obserwacje przybiera formę [Smith, Todd 2005]:

$$\frac{1}{N} \sum [Y_{1i} - E(Y_{0i} | P_i = 1)] \quad (8)$$

gdzie:  $E(Y_0 | P_i = 1) = \sum W(i, j) Y_{0j}$  jest oszacowaniem stanu kontrfaktycznego, a  $W(i, j)$  jest macierzą odległości między jednostkami  $i$  oraz  $j$ .

Po obliczeniu miary podobieństwa jednostek są one łączone za pomocą wybranego przez badacza algorytmu. Łączenie może odbywać się według zasady 1 do 1 (w pary), 1 do wielu (do każdej jednostki poddanej oddziaływaniu dobiera się więcej niż jedną jednostkę z grupy kontrolnej) lub wykorzystując jądro estymator gęstości (dla każdej obserwacji obliczana jest wartość dystrybuanty i dobierany jest odpowiednik z drugiej grupy o identycznej lub najbardziej zbliżonej wartości). W ostatnim kroku obliczane są różnice w zmiennej wynikowej dla połączonych obserwacji. Te różnice są oszacowaniami szukanych efektów.

Problem dobierania obserwacji porównawczej pochodzącej z najbliższego sąsiedztwa (ang. *nearest neighbour matching*) jest wielowymiarowy. Tradycyjnym i najbardziej intuicyjnym sposobem jego rozwiązania jest wskazanie, dla każdej jednostki z grupy poddanej oddziaływaniu, jej najbliższego sąsiada z grupy kontrolnej. Definiujemy zbiór  $A_i$  w taki sposób, że dokładnie jeden sąsiad jednostki  $i$  do niego należy

$$A_j = \left\{ j \mid j \in \{1..n\}: \min \|w_i - w_j\| \right\} \quad (8)$$

gdzie  $\|\cdot\|$  jest metryką. Wówczas macierz wag  $W(i,j)$  jest kwadratowa i składa się wyłącznie z zer i jedynek. Jedyńska jest przypisywana najbliższemu sąsiadowi, a zera pozostałym jednostkom grupy kontrolnej.

Estymator najbliższego sąsiada działa dobrze, jeżeli  $W_i$  oraz  $W_j$  są określone na tym samym zbiorze, lub część wspólna tych zbiorów jest stosunkowo duża. Badacz musi podjąć decyzję zrobić z niedokładnie połączonymi obserwacjami (Lee, str. 89). Jeżeli duża liczba słabo połączonych par zostanie odrzucona, wówczas zmniejszy się liczebność grupy kontrolnej, i dla pewnych obserwacji z grupy eksperymentalnej nie będzie dobrej pary, zatem zostaną one pominięte w analizie. Zredukuje to problem zbilansowania, ale równocześnie zmniejszy się efektywność, z uwagi na pominięcie istotnej części informacji. Z drugiej strony uwzględnienie wszystkich par, bez względu na jakość połączenia pogorszy zbilansowanie.

Aby zapobiec porównywaniu elementów bardzo różniących się między sobą wprowadza się poprawkę tzw. *caliper matching* [Cochran i Rubin, 1975]. Działanie poprawki można porównać z działaniem przesłony w aparacie fotograficznym. Jeżeli uwaga jest skoncentrowana na wybranym punkcie, to inne punkty leżące od niego daleko zostają przesłonięte. Procedura usuwa obserwacje, dla których nie ma bliskiego odpowiednika.

$$A_j = \left\{ j \mid j \in \{1..n\}: \min \|w_i - w_j\| < \delta \right\} \quad (9)$$

Do zbioru  $A_i$  należą takie punkty  $j$ , dla których odległość od najbliższego odpowiednika jest nie większa niż  $\delta$ . Zwiększa to precyzję estymatora. Gdy rozkład prawdopodobieństwa uczestnictwa w programie jest równomierny w grupie eksperymentalnej i kontrolnej obciążenie usuwa skrajnie źle dopasowane jednostki zmniejszając obciążenie estymatora. Gdy rozkłady w obu grupach znacznie różnią

się, stała wartość obciążenia inaczej traktuje połączenia jednostek o niskim prawdopodobieństwie znalezienia się w grupie eksperymentalnej a inaczej o wysokim. Te pierwsze częściej podlegają obciążeniu. W efekcie po eliminacji źle połączonych par obciążenie estymatora może wzrosnąć. Jest to szczególnie prawdopodobne, gdy zachodzi zależność między prawdopodobieństwem uczestnictwa w eksperymencie a jego wynikiem.

W związku z tym uważamy, że należy uzależnić wartość odciążenia (*caliper*) od wartości wektora *propensity score* dla grupy eksperymentalnej. Wówczas zmodyfikowany mechanizm wyboru potencjalnych jednostek do grupy kontrolnej wygląda następująco

$$A_j = \{j \mid j \in \{1..n\}: \min |w_i - w_j| < \delta w_i\} \quad (10)$$

W ten sposób wiążemy wartość odciążenia z oszacowaną wartością wektora *propensity score*. Postulujemy, aby wartość odciążenia  $\delta$  była ustaloną frakcją wartości wektora. Wówczas dla obserwacji o niskim prawdopodobieństwie znalezienia się w grupie eksperymentalnej (wartość wektora bliska 0) będziemy wymuszali bardzo dokładne określenie odpowiednika w stanie kontrfaktycznym. Z kolei, gdy prawdopodobieństwo przydzielenia do programu będzie dążyć do 1 dynamiczny mechanizm obciążenia będzie w niewielkim stopniu różnił się od standardowego. W rezultacie, w obliczeniach zostanie pominiętych więcej połączonych jednostek dla których wartość oszacowanego prawdopodobieństwa znalezienia się w grupie eksperymentalnej jest niewielka.

## EKSPERYMENT SYMULACYJNY

W celu zbadania właściwości statystycznych proponowanej poprawki przeprowadziliśmy symulację według następującego schematu

1. Ustalenie wielkości próby.
2. Generowanie charakterystyk jednostek, oraz dwóch składników losowych.
3. Na podstawie wylosowanych wartości wykorzystując znane równanie o ustalonych współczynnikach generowane są prawdopodobieństwa przynależności do grupy eksperymentalnej.
4. Obliczana jest zmienna wynikowa jako funkcja deterministyczna charakterystyk, indykatora uczestnictwa w grupie eksperymentalnej oraz niewielkiego składnika losowego.
5. Obliczany jest przeciętny efekt oddziaływania za pomocą trzech algorytmów: łączenie 1 do 1, łączenie 1 do 1 z obciążeniem, oraz łączenie 1 do 1 z dynamicznym obciążeniem.
6. Kroki 1-6 powtarzane są 100000 razy

Budując proces generujący dane założono, że prawdziwa oczekiwana wartość oddziaływania wynosi 1. W efekcie eksperymentu uzyskano wyniki zaprezentowane w Tabeli 1.

Wartości parametrów dobrano tak, aby 25% obserwacji znalazło się w grupie eksperymentalnej. Uzyskana przeciętna liczba par wynosząca 241 wynika z losowej natury eksperymentu i obrazuje niedokładność symulacji. Jak widać, dynamiczny sposób uznawania pary za dobrze połączone pozwala w niewielkim stopniu, ale jednak zmniejszyć obciążenie estymatora efektu oddziaływania. W pierwszej chwili wydawało się, że ceną za zmniejszenie obciążenia oszacowania jest mniejsza precyzja uzyskanego wyniku. Ponieważ literatura wskazuje, że estymator ATT ma rozkład normalny, ale tylko asymptotycznie, postanowiliśmy dokładnie przeanalizować uzyskane wyniki. W rezultacie zauważyliśmy, że przyczyną mniejszej dokładności oszacowań może być mniejsza liczba połączonych par. Klasyczny sposób odcięcia powoduje nieuwzględnianie ok. 10% par przy obliczaniu wartości ATT, a metoda dynamiczna odrzuca ok. 46% par, jako niedokładne odpowiedniki.

Tabela 1. Wyniki eksperymentu

Metoda	ATT	se(ATT)	liczba par
Propensity score	0,997	0,176	241
Propensity score z odcięciem	0,997	0,126	220
Propensity score z dynamicznym odcięciem	0,998	0,218	134

Źródło: obliczenia własne

Należy zaznaczyć, że dynamiczny sposób odcięcia odrzuca większą liczbę par, jednak nie należy rozpatrywać tego faktu w kategorii słabości, a zalety metody. Dzięki temu oszacowanie przeciętnego efektu jest uzyskiwane na podstawie lepiej zbilansowanej próby.

Aby rozstrzygnąć problem w sposób jednoznaczny postanowiliśmy przeprowadzić dodatkową symulację. Zbadaliśmy, w jaki sposób wariancja estymatora ATT jest uzależniona od liczby uzyskanych połączonych par. W tym celu dla prób o różnej liczebności przeprowadziliśmy proste łączenie 1 do 1 i obliczyliśmy dla każdej próby estymator ATT i jego odchylenie standardowe. Eksperyment powtórzyliśmy 100000 razy. Wyniki eksperymentu pomocniczego przedstawiono w tabeli 2<sup>3</sup>.

Ewidentnie widać, że wraz ze wzrostem liczby jednostek w grupie kontrolnej, która odpowiada liczbie połączonych par, błąd standardowy oszacowania maleje w tempie wykładniczym. Przedstawia do w innym świetle wyniki poprzedniego eksperymentu. Przy spadku liczby połączonych par z 241 do 134, 25% odchylenia standardowego można przypisać zmniejszonej liczbie par. W rezultacie należy

<sup>3</sup> Z racji na szczupłość miejsca prezentujemy tutaj fragment wyników. Na prośbę osób zainteresowanych wysłamy pełne wyniki.



uznać, że odchylenie standardowe estymatora dynamicznym obciążeniem jest identyczne do odchylenia standardowego estymatora ATT. Tradycyjny mechanizm obciążenia przy tej liczebności próby redukuje odchylenie standardowe oszacowania o 30%. Zatem należy podjąć trudny wybór: czy lepsze jest bardziej precyzyjne oszacowanie o większej wariancji czy mniej precyzyjne o mniejszej wariancji. My opowiadamy się za tym pierwszym.

Tabela 2. Wariancja ATT a wielkość próby

Próba	ATT	se(ATT)	Eksperymentalna	Kontrolna
100	1.004	0.564	24	76
200	0.998	0.397	48	152
300	0.997	0.322	72	228
400	0.997	0.279	96	304
500	0.998	0.249	121	379

Źródło: obliczenia własne

Asymptotycznie wszystkie techniki łączenia danych oparte o wektory *propensity score* powinny dawać takie same rezultaty, gdyż wraz ze wzrostem liczebności próby metoda zbliża się do porównywania obserwacji dokładnie połączonych. Jednakże w małych próbach wybór algorytmu może ważyć na wyniku (Heckman i in. 1997). W takim przypadku należy dokonać wyboru między obciążeniem uzyskanego oszacowania efektu a jego wariancją. Nie ma jednego najlepszego sposobu estymacji sprawdzającego się w każdej sytuacji. Jakość poszczególnych technik i wybór estymatora powinny być uzależnione od statystycznych właściwości zbioru danych.

## PODSUMOWANIE

W artykule przedstawiono estymatory przeciętnego efektu oddziaływania, oraz pokazano wykorzystanie techniki łączenia danych wykorzystującej wektor *propensity score*. Celem metody jest odtworzenie na podstawie danych przekrojowych warunków eksperymentalnych poprzez konstrukcję hipotetycznej grupy odniesienia (stanu kontrfaktycznego) pozwalającego na porównanie wyników działania programu z sytuacją, w której nie był on przeprowadzony. Odbywa się to poprzez taki wybór jednostek do grupy kontrolnej by była jak najbardziej zbliżona pod względem obserwowanych charakterystyk do grupy poddanej oddziaływaniu.

Zaprezentowano nową technikę oceny stopnia podobieństwa obserwacji, wykorzystującą zmodyfikowany mechanizm obciążenia. Modyfikacja polega na odnoszeniu wielkości obciążenia do wartości wektora *propensity score* dla jednostki, dla której szukany jest obiekt w stanie kontrfaktyczny. Następnie wykorzystując symulacje komputerowe zbadano właściwości statystyczne metody. Wykazano, że w przypadku znacznych różnic w rozkładach wektora *propensity score* między

grupą poddaną oddziaływaniu a grupą kontrolną dynamiczny mechanizm obciążenia pozwala uzyskać bardziej dokładne oszacowanie. Jednak ceną za mniejsze obciążenie estymatora jest większa wariancja, wynikająca z mniejszej liczby skutecznie połączonych par.

W dalszych pracach zamierzamy po pierwsze uwzględnić w eksperymencie próby o różnej liczebności. Tak jak wariancja estymatora ATT okazała się być nieliniową funkcją liczby połączonych par, także obciążenie estymatora może zależeć w nieliniowy sposób od liczebności próby. Po drugie, zamierzamy tak przeprojektować schemat generowania danych, aby zapewnić stabilną liczbę połączonych par, a przynajmniej chcielibyśmy osiągnąć większy stopień kontroli nad tym parametrem eksperymentu. Po rozwiązaniu tych problemów zamierzamy wykorzystać zmodyfikowaną procedurę w badaniach aplikacyjnych.

## BIBLIOGRAFIA

- Blundell R., Costa-Diás M. (2000) "Evaluation Methods for Non-Experimental Data", *Fiscal Studies*, vol. 21/4, str. 427-468.
- Caliendo M., Kopeinig S. (2008) „Some Practical Guidance for the Implementation of Propensity Score Matching”, *Journal of Economic Surveys*
- Cochrane, Rubin (1973) "Controlling Bias in Observational Studies. A Review", *Sankhya*, vol. 35, str 417-466.
- Hahn J. (1998) "On the Role of the Propensity Score in the Efficient Semiparametric Estimation of Average Treatment Effects", *Econometrica*, vol. 66/2, str. 315-331.
- Heckman J., Ichimura H., Todd P. (1997) "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme", *The Review of Economic Studies*, vol. 64/4, str. 605-654.
- Hirano K., Imbens G., Ridder G. (2003) "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score", *Econometrica* vol. 71/4, str. 1161-1189.
- Lee M-J. (2005) "Micro-Econometrics for Policy, Program, and Treatment Effects", Oxford University Press.
- Rosenbaum P., Rubin D. (1983) „The Central Role of the Propensity Score in Observational Studies for Causal Effects“, *Biometrika*, vol. 70/1, str 41-55.
- Rubin D. (1973) "Matching to Remove Bias in Observational Studies", *Biometrics*, vol. 29, str. 159-183.
- Smith J., Todd P. (2005) "Does Matching Overcome LaLonde's Critique of nonexperimental estimators?", *Journal of Econometrics*, vol. 125, str. 305-353.
- Strawiński P. (2007) "Przyczynowość, selekcja i endogeniczne oddziaływanie", *Przegląd Statystyczny* nr 4/2007, str. 49-61.

### **Matching with dynamic caliper. Preliminary results**

**Abstract:** The evaluation problem is a problem of missing data. Henceforth, the important issue is construction of adequate counterfactuals. In this article modification of procedure that accepts observation as a good counterfactual is proposed. The proposed novelty is a change in the mechanism of observation comparison between treated and control group. In the last section the simulation results are presented. The simulation concerns statistical properties of proposed mechanism. The obtained results indicate that our method is capable of providing better results than techniques that are contemporary used.

**Keywords:** matching, propensity score, caliper, evaluation