

**Stanisław Halkiewicz**

e-mail: 184031@ue.wroc.pl

ORCID: 0009-0000-7344-7522

Uniwersytet Ekonomiczny we Wrocławiu

## **Modele *Difference-in-Differences* w ewaluacji wpływu zdarzeń ekonomicznych na poprawę edukacji**

DOI: 10.15611/2023.89.3.06

JEL Classification: C31, C21, C26, I250, I240, O110

**Streszczenie:** Artykuł porusza tematykę zastosowania modeli *Difference-in-Differences* w badaniach z zakresu ekonomii rozwoju, skupiając się na dziedzinie szkolnictwa. Opisywane modele nabierają coraz większego znaczenia w XXI wieku. Celem badawczym opracowania jest analiza ich użyteczności do mierzenia efektów zdarzeń ekonomicznych, w tym głównie programów i projektów rządowych i społecznych, a także opis wpływu na poprawę jakości i dostępności edukacji w porównaniu z innymi metodami ekonometrycznymi. W tym celu przeanalizowane zostaną działania w tym obszarze na terenie Indonezji. Zostaną dokonane obliczenia z wykorzystaniem DD w ujęciu sprzed propozycji Esther Duflou i z jej uwzględnieniem w celu oceny roli poprawki zaproponowanej przez jej zespół badawczy w dokładności mierzenia efektów poprawy edukacji.

**Słowa kluczowe:** *Difference-in-Differences*, ekonometria, edukacja, ekonomia rozwoju

### **1. Wstęp**

Modele *Difference-in-Differences* (często określane skrótowo jako modele DD, DID lub DiD, w naukach społecznych również jako modele *Before-and-After Study*) są szeroko wykorzystywane w analizach danych panelowych. Dzięki przejrzystości metody i jej względnej łatwości w zastosowaniu są one bardzo chętnie używane przez naukowców w wielu badaniach. Modele te oparte są na porównywaniu międzygrupowych średnich, co sprawia, że metoda ta nadaje się do oceny nagłych zmian w otoczeniu ekonomicznym – w tym na przykład katastrof naturalnych, wprowadzonych zmian prawnych, programów rządowych czy inicjatyw społecznych (Angrist i Krueger, 1999). W celu estymacji takiego modelu wyodrębnia się zdarzenie ekonomiczne, zwane *treatment*, a następnie narzędziami ekonometrycznymi porównuje efekty „przed” i „po”, dla grupy objętej efektem badanego zdarzenia oraz grupy niedotkniętej jej efektem. Za prekursora tej metody uważa się Johna Snowa, w którego badaniach nad przyczynami epidemii cholery w Londynie w 1854 r. zastosowany został model stawiający podwaliny pod *Difference-in-Difference* (Caniglia i Murray, 2020).

W niniejszym artykule skupiono się na zastosowaniu modeli DD w ewaluacji wpływu zdarzeń ekonomicznych (tutaj: projektów i programów, ale w szerszej perspektywie wyciągnięte wnioski można przenieść na grunt również zdarzeń losowych) na jakość i dostępność edukacji. Dokonano przeglądu literatury. Jako przykład ilustrujący porównanie metody *Difference-in-Differences* z metodą zmiennych in-

strumentalnych (zwany skrótoowo modelami IV od angielskiego terminu *Instrumental Variable*) rozpatrzono wpływ programu budowy szkół w Indonezji na wykształcenie tamtejszej populacji (Duflo, 2001). Niniejsza praca ma odpowiedzieć na pytanie, dlaczego modele *Difference-in-Differences* są używane w tej ewaluacji i czy jest to najlepsza opcja z dostępnych metod ekonometrycznych. Przedstawiono również przemyślenia autora i obliczenia dotyczące estymacji DD w świetle poprawki zaproponowanej przez zespół badawczy National Bureau of Economic Research (NBER) (Bertrand i in., 2002) jako sposób na ocenianie efektywności metody DD w badanym przypadku. Wszystkie istotności rozpatrywano na poziomie  $\alpha = 0,05$ .

## 2. Przegląd literatury

W XXI wieku, wraz ze wzrostem znaczenia ekonometrii i szerszym zainteresowaniem badaczy problematyką ekonomii rozwoju, modele *Difference-in-Differences* są coraz chętniej używane jako narzędzie do oceny efektów wielu programów, projektów czy też ewaluacji pokłosi zdarzeń losowych. Edukacja jako jeden z najczęściej adresowanych politycznie i społecznie sektorów również skorzystała na rozwoju tych metodyk, czego rezultatem było wiele prac naukowych adaptujących metody DD w szeroko rozumianych badaniach nad szkolnictwem. Za dowód tego można uznać przyznanie w 2019 roku nagrody *The Sveriges Riksbank Prize in Economic Sciences*, zwanej potocznie Noblem z ekonomii, grupie badaczy zajmujących się m.in. właściwościami i zastosowaniami modeli DD. Wśród laureatów znaleźli się Esther Duflo, Abhijit Banerjee z Instytutu Technologicznego Massachusetts i Michael Kremer z Uniwersytetu Harvarda, którzy swoją karierę naukową poświęcili badaniom na temat przeciwdziałania biedzie. W szerszym kontekście omówione zostaną w następnych rozdziałach artykułu założenia, metodologia i wnioski dwóch z ich prac (Bertrand i in., 2002; Duflo, 2001), gdyż to one będą stanowiły podstawę dalszych rozważań w niniejszej pracy.

Zestawienie przykładowych publikacji zawierających podejście DD wraz z problemem, zdarzeniem ekonomicznym i wynikiem opisywanymi w pracy zaprezentowano w tabeli 1 (Kelchen i in., 2019). W większości z nich autorzy opisują badania na terenie Stanów Zjednoczonych, a te nie są dokładne w obszarze ekonomii rozwoju. Dotyczą one m.in. zmian w politykach dotyczących dofinansowywania czesnego po wydarzeniach z 9 września 2011 roku (Baird i in., 2018), programu stypendialno-mentoringowego Knox Achieves na terenie Tennessee (Carruthers i Fox, 2016), wpływu zwiększonych zapisów ośmioklasistów w Kalifornii na kursy z algebry na ich ukierunkowanie (Domina i in., 2015) czy wpływu traumy związanej z atakami snajperskimi w Waszyngtonie na oceny studentów (Gershenson i Tekin, 2018). Wśród tych spoza Stanów Zjednoczonych możemy wyróżnić m.in. badania wpływu programu zwalczania malarii na udział kobiet w edukacji w Paragwaju i na Sri Lance (Lucas, 2010) czy też wpływu nieodpłatnej, podstawowej edukacji w Kenii na ukończenie szkoły (Lucas i Mbiti, 2012). Modelami DD ewaluowano również program odrobaczania

nia dzieci w wieku szkolnym w Kenii, za mierzony efekt obierając na przykład absencję szkolną (Hamory i in., 2021; Miguel i Kremer, 2004). W literaturze można spotkać wiele publikacji opisujących wszystkie najnowsze dokonania i badania w obszarze modeli DiD (np. Roth i in., 2023), jak również w obszarze analiz dotyczących wpływu zdarzeń ekonomicznych na edukację (np. Psacharopoulos i Patrinos, 2018).

**Tabela 1.** Przykłady badań z zakresu edukacji używających podejścia DD

Study	Policy context	Treatment	Outcomes
Baird, Kofoed, Miller, and Wenger (2018)	Post 9-11 GI Bill changes	Dollar amount change in maximum benefit	Tuition at for-profit higher education institutions
Carruthers and Fox (2016)	Knox Achieves (college coaching and financial aid program in Tennessee)	Share of high school's senior class participating in program	High school graduation, college enrollment, and college persistence
Domina, McEachin, Penner, and Penner (2015)	California's Algebra for All effort	Changes in eighth grade algebra enrollment rates	10th grade mathematics achievement
Gershenson and Tekin (2018)	Beltway sniper attacks	Distance from traumatic event	School-level proficiency rates
Lucas (2010)	Malaria eradication campaigns in Paraguay and Sri Lanka	Changes in malaria infection rates	Female educational attainment
Lucas and Mbiti (2012a, 2012b)	Kenyan free primary education program	Percentage increase in primary school completers based on preprogram rates	Primary school participation, completion, and achievement
Shores and Steinberg (2017)	Great Recession	School district-level recession intensity and length of time exposed to recession	Student achievement

Źródło: (Kelchen i in., 2019, Table 5).

Temat zastosowania modeli DD w badaniu szkolnictwa wyższego podjęto w pracy *A Primer for Interpreting and Designing Difference-in-Differences Studies in Higher Education Research* (Furquim i in., 2020).

### 3. Problemy szkolnictwa w Indonezji i program INPRES

Indonezja jest krajem bardzo chętnie opisywanym przez ekonomistów zajmujących się ekonomią rozwoju. Powodem może być fakt, że kraj ten, jeszcze do niedawna nazywany „najbardziej liczną demokracją trzeciego świata”, dzięki wielu inicjatywom rządowym, szerokiej demokratyzacji państwa i dynamicznie rozwijającej się ekonomii, teraz określany jest jako oaza rozwoju (Leggei in., 2023). Doprowadziły do tego nie tylko demokratyzacja państwa i jego rozszerzający się eksport, ale też intensywność i wielość prospołecznych reform i działań, w tym również w dziedzinie edukacji (Martinez-Bravo, 2017; Purnastuti i in., 2015).

Przykładem programu celowego skierowanego na poprawę szkolnictwa był rządowy program budowy szkół, zwany INPRES, z lat 1973-1978. Jego głównymi celami i miernikami sukcesu było zwiększenie średniej długości odbywania edukacji przez młodzież mierzonej w latach i podniesienie przeciętnego wynagrodzenia otrzymywanego przez pracowników, którzy przeszli edukację (Martinez-Bravo, 2017). Program ten był wielokrotnie poddawany dyskusji w publikacjach naukowych, najczę-

ściej w szerszym kontekście problemów systemu edukacji w Indonezji czy szerzej – w krajach Azji Południowo-Wschodniej. Warto wspomnieć, że w ramach finansowania INPRES zostały przeprowadzone również działania dodatkowe, bez których korzystanie z nowo wybudowanej infrastruktury byłoby niemożliwe, m.in. zatrudniono nauczycieli i zabezpieczono budżet na ich przyszłe wynagrodzenia.

Próby ewaluacji efektów programu INPRES podjęła chociażby Duflo (2001). Badała ona, jak inwestycja w infrastrukturę szkolną (szerzej – szkolnictwo rozumiane jako całość) wpływa na przeciętną długość otrzymywanej edukacji. W kraju, gdzie dotychczas edukacja była często trudno dostępnym luksusem, budowa szkoły w biednym regionie mogła bowiem zarówno zachęcić nowych uczniów z biedniejszych regionów do nauki, jak i wpłynąć na dłuższy cykl edukacyjny dziecka, którego rodziny mogłoby nie być stać na postanie go do odległej placówki oświatowej na pełen cykl edukacyjny. Drugim badanym efektem programu był logarytm wynagrodzeń beneficjentów. Jako narzędzie do przeprowadzenia obliczeń Duflo wykorzystowała model DD. Charakterystyka badania wymagała wybrania grupy kontrolnej, co było utrudnione, ponieważ INPRES z założenia obejmował teren całego państwa. Stąd też jej funkcję pełnił podzbiór tych uczniów, którzy zamieszkiwali tereny, na których budowa nowych placówek była mniej intensywna niż na tych których mieszkańcy wchodziłi do grupy uznanej za objętą efektem. Jako drugi podział kontrolny został wyróżniony rok urodzenia. Duflo wyróżnia dwie grupy – osoby w wieku od 12 do 17 lat oraz od 2 do 6 lat (wiek w roku 1974). Ludzie z tej pierwszej grupy nie byli beneficjentami programu albo też skorzystali z niego w niewielkim stopniu. Druga grupa odbyła prawie całą swoją edukację lub jej znaczącą część po rozpoczęciu programu (edukacja w Indonezji ustawowo rozpoczyna się w 7. roku życia). Czynnikiem zaburzającym grupę kontrolną może być powtarzanie klas przez niektórych uczniów, co w tamtych latach było znacznym problemem (Behrman i Deolalikar, 1991), jednak Duflo szacuje, że nie powinno to mieć statystycznego wpływu na wyniki badania.

### **3.1. Efekt wpływu programu INPRES na średnią długość edukacji**

Jako pierwszy aspekt przeanalizowano wpływ programu INPRES na długość edukacji odbytej przez beneficjentów. Przeanalizowano, jak ten wpływ jest estymowany przez model DD, a jak przez model ekonometryczny ze zmienną instrumentalną i estymator Walda.

#### **Model *Difference-in-Differences***

W dalszej części tekstu odtworzono badanie wpływu programu INPRES na średnią długość edukacji (zob. tab. 2). Przy obu estymacjach – zarówno dla średniej długości otrzymanej edukacji, jak i dla średniego wynagrodzenia – zastosowano pewne uogólnienia, głównie ujednoczenie badanych grup wiekowych w regionach kontrolnych. Nie wpływa to jednak na ogólny sens i wnioski odtworzenia.

**Tabela 2.** Wpływ INPRES na przeciętną długość edukacji – estymacja DD średnimi międzygrupowymi

Estymacja DD			
	intensywność programu		różnica
	wysoka	niska	
wiek w 1974 r.			
Od 2 do 6 r. ż.	8,937881	10,118392	-1,180510
Od 12 do 17 r. ż.	8,585727	9,856342	-1,270615
Różnica	0,3521544	0,2620496	0,09010479

Źródło: opracowanie własne na podstawie (Duflo, 2001).

Zaprezentowane obliczenia wskazują na to, że dla osoby objętej programem w pełni (tj. poniżej 6. roku życia, urodzonej w regionie o wysokiej intensywności), program INPRES przeciętnie zwiększył długość otrzymanej edukacji o ok. 0,09 roku ( $DD_e = 0,09010479$ ). Wynik ten potwierdzono, budując model DD w formie równania ekonometrycznego:

$$R_e = \alpha_0 + \alpha_1 \cdot \text{Wiek} + \beta \cdot (\text{Wiek} \cdot \text{Intensywność}) + \alpha_3 \cdot \text{Intensywność}, \quad (1)$$

gdzie *Wiek* i *Intensywność* to zmienne typu 0-1, oznaczające odpowiednio przynależność do grupy wieku objętej efektem programu oraz intensywność programu w regionie,  $\alpha$ , to parametry stojące przy zmiennych 0-1, a  $\beta$  to parametr DD odpowiadający interpretacji różnicy *Difference-in-Differences*, wyliczonej wcześniej w tabeli 2. W tabeli 3 zawarto charakterystyki powyższej estymacji z równania (1) (tutaj parametry zmiennych znajdują się w kolumnie „Beta” w wierszu odpowiadającym danej zmiennej).

**Tabela 3.** Parametry równania modeli DD dla wpływu INPRES na przeciętną długość edukacji

Charakterystyki	Beta	95% CI <sup>1</sup>	p-value
Wiek	0,26	(0,15; 0,37)	<0,001
Intensywność	-1,3	(-1,4; -1,2)	<0,001
Wiek *Intensywność	0,09	(-0,08; 0,26)	0,3

<sup>1</sup> CI = Przedział ufności

Źródło: opracowanie własne na podstawie (Duflo, 2001).

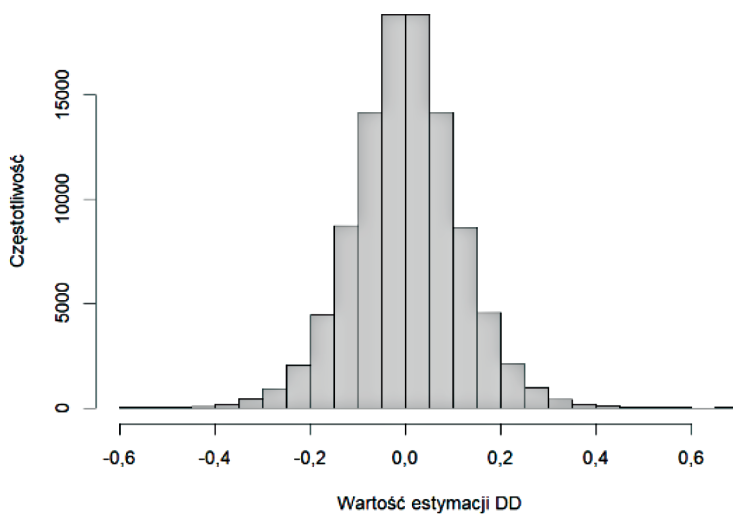
Na podstawie uzyskanych wyników można podejrzewać, że program ten nie miał istotnych efektów dla średniej długości otrzymanej edukacji. Te same przypuszczenia potwierdza *p-value* oraz przedział ufności odpowiadający parametrowi  $\beta$ . Również Duflo otrzymała podobne wnioski ze swojego modelu DD (2001). Może to znaczyć zarówno, że program INPRES miał nieznaczny wpływ na średnią długość otrzymywanego wykształcenia, jak i że model DD nie jest efektywny w ocenie tychże

rezultatów. W celu dodatkowej oceny efektywności tego modelu zbadano istotność błędu zbudowanego modelu, stosując poprawkę zaproponowaną przez zespół badawczy NBER (Bertrand i in., 2002).

### Test na istotność błędu modelu DD

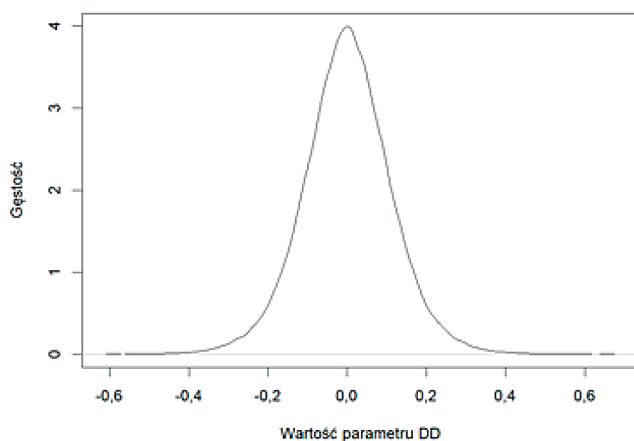
Warto zauważyć, że praca, w której Duflo dokonała uprzednio przedstawionych obliczeń, została wydana w roku 2001 – rok wcześniej niż praca, w której zespół badawczy NBER zaproponował poprawkę do testowania istotności błędu modelu. Z powodu braku wzmianki o użyciu takiego testu można założyć, że wobec modelu efektów programu INPRES Duflo nie zastosowała tej poprawki. Poprawka ta zaleca wykonanie serii estymacji DD, gdzie moment zdarzenia i objęte nim jednostki będą wybierane w sposób losowy. W ten sposób, zgodnie z teorią, odpowiednio  $\alpha/2$  i  $1-\alpha/2$  percentyle wyznaczą dwa przedziały istotności dla wyniku estymacji DD. Gdy między populacją objętą efektem a nieobjętą jest grupa wieku, której w badaniu nie rozpatruje się, samo przeprowadzenie symulacji można wykonać w różny sposób – zależnie od interpretacji. W niniejszej pracy zdecydowano się zachować odstęp między grupami. Rozkład wartości DD z symulacji (wielkość próby testowej wynosi 100 000 estymacji) przedstawiono na histogramie (zob. rys. 1) i wykresie gęstości (zob. rys. 2).

Już wizualna analiza wykazuje spodziewaną zależność – większość przeprowadzonych symulacji wskazała na brak efektu.



**Rysunek 1.** Histogram rozkładu testowego DD wpływu INPRES na przeciętną długość edukacji

Źródło: opracowanie własne.



**Rysunek 2.** Wykres gęstości rozkładu testowego DD wpływu INPRES na przeciętną długość edukacji

Źródło: opracowanie własne.

Skonstruowano test opierający się na odkryciach i obliczeniach zespołu badawczego NBER (Bertrand i in., 2002), ale uogólniony i sformalizowany przez autora w niniejszej pracy, w którym wyznaczono następujące hipotezy:

- $H_0 : DD_e = 0$ ,
- $H_A : DD_e \neq 0$ .

Przedział odrzucenia wyznaczają odpowiednio  $\alpha/2$  i  $1-\alpha/2$  percentyle. Dla  $\alpha = 0,05$  odczytać można z wygenerowanej próbki testowej, korzystając z wbudowanych funkcji  $R$  (R Core Team, 2021), wartości:  $P_{0,025} = -0.2246644$ ;  $P_{0,975} = -0.2241286$ . Obszar odrzucenia to  $R/(P_{0,025}; P_{0,975})$ , a obszar przyjęcia  $H_0$  wynosi  $(P_{0,025}; P_{0,975})$ .

Parametr  $DD_e$  eksperymentu właściwego wynosi, gwoli przypomnienia,  $DD_e = 0,09010479$ . Wynika z tego, że  $DD_e \in (P_{0,025}; P_{0,975})$ , a więc należy przyjąć  $H_0$  o nieistotności statystycznej parametru  $DD_e$  estymacji właściwej.

### 3.2. Wpływ na logarytm średnich wynagrodzeń beneficjentów

Przeprowadzono analogiczne postępowanie w celu oceny wpływu programu INPRES na średnie wynagrodzenia (mierzone logarytmem). W związku z bardzo podobnym przebiegiem badania zdecydowano się nie opisywać poszczególnych kroków procesu aż tak szczegółowo jak w poprzednim punkcie.

#### Model *Difference-in-Differences*

Wyniki tego procesu zostały przedstawione w tabeli 4 (porównania międzygrupowe). Również tutaj wykorzystano uogólnienie zastosowane w punkcie 3.1.

W wyniku estymacji parametru  $\beta$  modelu z równania (2) oraz patrząc na wyniki estymacji z tabeli średnich międzygrupowych, otrzymuje się informację, że logarytm płac absolwentów wzrósł dzięki programowi INPRES przeciętnie o 0,001 rupii indonezyjskiej. Estymacja parametru  $\beta$  równania (2), która została ujęta w tabeli 5, pokazuje się jako 0,00, jednak w rzeczywistości jest ona taka sama jak w tabeli 4, jedynie została zaokrąglona do części setnych – tak samo, jak miało to miejsce w punkcie 3.1.

$$R_e = \alpha_0 + \alpha_1 \cdot \text{Wiek} + \beta \cdot (\text{Wiek} \cdot \text{Intensywność}) + \alpha_3 \cdot \text{Intensywność}. \quad (2)$$

**Tabela 4.** Wpływ INPRES na logarytm wynagrodzeń – estymacja DD średnimi międzygrupowymi

Estymacja DD			
	intensywność programu		różnica
	wysoka	niska	
wiek w 1974 r.			
6 r. ż. lub mniej	11,83948	11,97481	-0,1353317
Od 12 do 17 r. ż.	12,10723	12,24148	-0,1342488
Różnica	-0,2677537	-0,2666708	-0,001082898

Źródło: opracowanie własne na podstawie (Duflo, 2001).

**Tabela 5.** Parametry estymacji DD wpływu INPRES na logarytm wynagrodzeń

Charakterystyki	Beta	95% CI <sup>1</sup>	<i>p-value</i>
Wiek	-0,27	(-0,29; -0,25)	<0,001
Intensywność	-0,13	(-0,15; -0,11)	<0,001
Wiek * Intensywność	0,00	(-0,03; 0,03)	>0,9

<sup>1</sup>CI = Przedział ufności

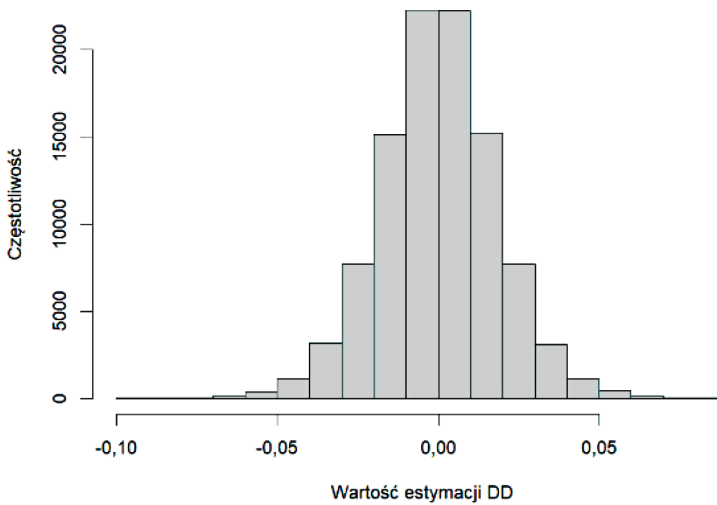
Źródło: opracowanie własne na podstawie (Duflo, 2002).

Tak samo jak w przypadku wpływu na średnią długość edukacji tak i tutaj można mieć podejrzenia o nieistotności wpływu programu. Sugeruje to zarówno bardzo mała wartość estymacji, jak i skrajna nieistotność parametru  $\beta$  (*p-value* > 0,9).

### Test na istotność błędu modelu DD

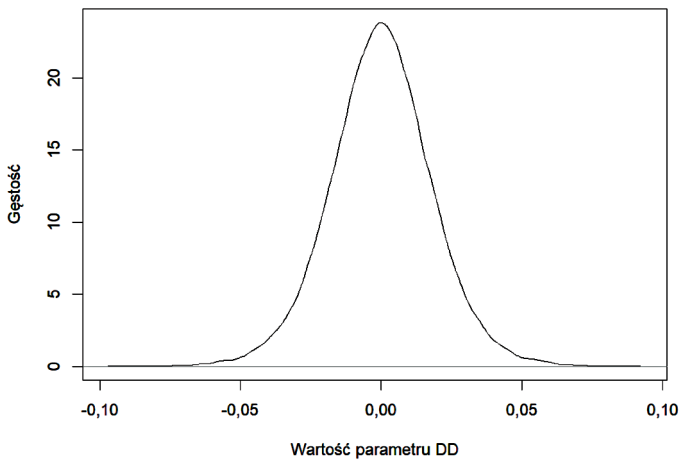
Przeprowadzono test na istotność błędu modelu DD, tym razem uwzględniając jako zmienną objaśnianą logarytm średnich wynagrodzeń. Rozkład testowy, wygenerowany na próbie równej 100 000 symulacji, przedstawiony jest na rysunkach 3 i 4.





**Rysunek 3.** Histogram rozkładu testowego DD wpływu INPRES na logarytm przeciętnych wynagrodzeń

Źródło: opracowanie własne.



**Rysunek 4.** Wykres gęstości rozkładu testowego DD wpływu INPRES na logarytm przeciętnych wynagrodzeń

Źródło: opracowanie własne.

Od razu można zauważyć pewne podobieństwo do wykresów wykonanych w analizie do punktu 3.1. Wykres ten jest podobnie smukły, jednak przy innej skali, co jest wynikiem przyjęcia innego wyznacznika efektu programu.

Obliczono obszary przyjęcia i odrzucenia  $H_0$  testu skonstruowanego w punkcie 3.1, wyznaczając percentyle z rozkładu testowego dla badanej zależności:  $P_{0,025} = -0.03658100$ ,  $P_{0,975} = 0.03655956$ .

Obszar odrzucenia to  $R/(P_{0,025}; P_{0,975})$ , a obszar przyjęcia  $H_0$  wynosi  $(P_{0,025}; P_{0,975})$ .

Wartość parametru  $DD$  wynosi  $DD = 0,001082898$ , a więc  $DD \in (P_{0,025}; P_{0,975})$ . Parametr  $DD$  estymacji oryginalnej pozwala zatem na przyjęcie hipotezy zerowej o nieistotności statystycznej parametru  $DD$ .

### 3.3. Model IV i estymator Walda

Innym ze sposobów estymacji często używanych w dziedzinie ekonomii rozwoju czy też w modelach opisujących efekty zdarzeń ekonomicznych dla badanego stanu są modele ze zmienną instrumentalną. Założeniem, od którego wychodzą, jest niemożność ujęcia wszystkich czynników wpływających na zmienną zależną czy też na badany stan po zajściu zdarzenia ekonomicznego w modelu ekonometrycznym, a także istnienie trudnej do wykrycia korelacji, która może fałszować postrzeganie modelu i prowadzić do błędnych jego interpretacji. W rzeczywistości estymacja modelu IV przebiega dwuetapowo (stąd też inna nazwa – dwuetapowa metoda najmniejszych kwadratów) (Stock, 2015).

#### Edukacja warunkowana intensywnością programu

Stosując pakiet *ivreg* w języku R (Fox i in., 2023), przeprowadzono estymację pierwszego modelu ekonometrycznego, gdzie zmienną objaśnianą był logarytm wynagrodzeń absolwentów, a jako zmienną objaśniającą przyjęto zmienną instrumentalną – liczbę lat edukacji warunkowaną intensywnością programu w regionie. Zaprezentowano to w równaniu (3):

$$\text{Wynagrodzenia} = \alpha_0 + \beta \cdot \text{Edukacja} \mid \text{Intensywność}. \quad (3)$$

Wartość parametru  $\beta$  wyniosła 0,115392. Test *t*-Studenta pokazuje, że ocena tego parametru jest istotna statystycznie, co w poprzednich metodach nie miało miejsca. Mówi to jednak jedynie o części zagadnienia – jednostka mieszkająca bowiem w regionie o wysokiej intensywności programu, ale urodzona zbyt późno, by z niego skorzystać, nie jest objęta programem, a ujęta jest w modelu jako 1, na równi z osobą korzystającą z programu, co stoi w sprzeczności z celem analizy – oceną efektywności programu INPRES. Ukazuje to jednak dodatni wpływ edukacji na wynagrodzenia.

Estymator Walda policzono ze średnich zgodnie ze wzorem (4).

$$\text{Wald}_1 = \frac{(\text{Wynagrodzenia} \mid \text{Intensywność} = 1) - (\text{Wynagrodzenia} \mid \text{Intensywność} = 0)}{(\text{Edukacja} \mid \text{Intensywność} = 1) - (\text{Edukacja} \mid \text{Intensywność} = 0)}. \quad (4)$$

Obliczono wartość parametru Walda dla zbudowanego równania, otrzymując wartość  $\text{Wald}_1 = 0.115392$ . Wartość ta jest zgodna z wartością parametru odpowiadającego zmiennej instrumentalnej.

### Edukacja warunkowana ekspozycją na program

Drugą z estymacji zbudowano w sposób lepiej odzwierciedlający istotę oceny eksperymentu poprzez warunkowanie edukacji iloczynem zmiennych sygnalizujących intensywność programu w regionie urodzenia osoby oraz czy odbywała ona edukację w latach przeprowadzania programu. Powstały model wyraża się równaniem (5):

$$\text{Wynagrodzenia} = \alpha_0 + \beta \cdot \text{Edukacja} \mid (\text{Intensywność} \cdot \text{Wiek}). \quad (5)$$

Obliczono, że  $\beta = 0.69253$ . Jest to wynik zarówno większy niż przy mniej rygorystycznym warunkowaniu, jak i statystycznie istotny, a jednocześnie odzwierciedla on w lepszym stopniu sens oceny efektywności programu, ponieważ warunkowanie rzeczywiście odbywa się na jednostkach w pełni wyeksponowanych na wysoką intensywność programu.

### Potencjalne problemy ze zmienną instrumentalną

Jednym z wymagań modeli ze zmienną instrumentalną jest wysoka zależność między zmienną warunkującą i warunkowaną (tymczasem w badanym przypadku niekoniecznie musi zostać ono spełnione). Intuicja ekonomiczna podpowiada, że program mający m.in. wydłużyć czas edukacji otrzymywanej przez jednostkę powinien wpływać na swój cel. Jednak poprzednie szacunki DD nie potwierdziły tego przypuszczenia, sugerując brak wpływu lub zły dobór metody.

Drugim z założeń jest brak korelacji zmiennej warunkującej ze zmienną objaśnianą (niestety i tutaj w przypadku rozpatrywanego problemu nie jest ono spełnione). Kierując się bowiem logiką, która zakładałaby wpływ programu na cele przez niego założone, czego wymaga pierwsze założenie, należałoby przyjąć też zależność logarytmu wynagrodzeń, czyli zmiennej zależnej, od programu INPRES. W obu przypadkach więc jedno z założeń niezbędnych do zbudowania modelu ze zmienną instrumentalną jest niespełnione.

## 4. Wnioski i proponowane modyfikacje

Modele *Difference-in-Differences* są często używane do ewaluacji wpływu zdarzeń ekonomicznych (Callaway, 2022). Celem niniejszej pracy była ocena takiego postępowania w świetle innych metod ekonometrycznych oraz zaprezentowanie poprawki NBER (Bertrand i in., 2002) jako metody na określenie wiarygodności uzyskanej estymacji DD dla badanego przypadku.

Modele Klasycznej Metody Najmniejszych Kwadratów (KMNK) nie są przystosowane do oceny ewaluacji zdarzeń ekonomicznych, dlatego też nie stosuje się ich szeroko w tym zakresie. Można natomiast je wykorzystywać do innych działań wspierających analizy DD, np. do określenia zależności między różnymi miernikami badanego efektu.

Kwestia modeli ze zmienną instrumentalną rozbija się o charakter zmiennych – objaśnianej i warunkowanej. Jeśli chce się mierzyć efekt programu lub zdarzenia ekonomicznego na zmiennej  $Y$ , z definicji potencjalnie zależnej od programu/zdarzenia, każda zmienna rozpatrywana jako potencjalna zmienna warunkowana również będzie zależna od programu/zdarzenia. Jedyne przypadki, w którym model można estymować przy przyjętych założeniach, to taki, gdy  $Y$  nie jest zależne od efektów programu, co niweluje potrzebę estymacji modelu, a czego i tak nie można wiedzieć *a priori*.

Jak pokazano na przykładzie estymacji efektów programu INPRES, modele DD nie zawsze przynoszą wyniki istotne statystycznie. Nie znaczy to, że są one złym sposobem na taką ewaluację. W przeciwieństwie do modeli IV założenia wymagane do ich estymacji nie kłócą się z formą i charakterem danych, którymi badacz zazwyczaj dysponuje, chcąc dokonać takowej. Są one również stworzone i przystosowane do badania danych panelowych. Proponuje się natomiast, by każdorazowo przeprowadzić postępowanie testowe, na wzór wykonanego w punkcie 3, w celu oceny, czy obliczony parametr DD jest istotny i czy może być analizowany w dalszym postępowaniu.

## Literatura

- Angrist, J. D. i Krueger, A. B. (1999). Empirical Strategies in Labor Economics. *Handbook of Labor Economics*, 3(1), 1277-1366. [https://doi.org/10.1016/S1573-4463\(99\)03004-7](https://doi.org/10.1016/S1573-4463(99)03004-7)
- Baird, M., Kofoed, M., Miller, T. i Wenger, J. (2018). For-Profit Higher Education Responsiveness to Price Shocks: An Investigation of Changes in Post 9-11 GI Bill Allowed Maximum Tuitions. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3174763>
- Bal-Domańska, B. (2011). Wybrane problemy autokorelacji w modelowaniu na podstawie danych panelowych. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, (200), 178-187. [http://www.dbc.wroc.pl/Content/118831/Bal\\_Domanska\\_Wybrane\\_problemy\\_autokorelacji.pdf](http://www.dbc.wroc.pl/Content/118831/Bal_Domanska_Wybrane_problemy_autokorelacji.pdf)
- Behrman, J. R. i Deolalikar, A. B. (1991). School Repetition, Drop-Outs, and the Rates of Return to Schooling: The Case of Indonesia. *Oxford Bulletin of Economics and Statistics*, 53(4). <https://doi.org/10.1111/j.1468-0084.1991.mp53004007.x>
- Bertrand, M., Duflo, E. i Mullainathan, S. (2002). *How Much Should We Trust Differences-in-Differences Estimates?*. NBER. <http://www.nber.org/papers/w8841>
- Callaway, B. (2022). Difference-in-Differences for Policy Evaluation. W: K.F. Zimmermann (red.), *Handbook of Labor, Human Resources and Population Economics*. Springer, Cham. [https://doi.org/10.1007/978-3-319-57365-6\\_352-1](https://doi.org/10.1007/978-3-319-57365-6_352-1)
- Caniglia, E. C. i Murray, E. J. (2020). Difference-in-Difference in the Time of Cholera: A Gentle Introduction for Epidemiologists. *Current Epidemiology Reports*, 7(4). <https://doi.org/10.1007/s40471-020-00245-2>
- Carruthers, C. K. i Fox, W. F. (2016). Aid for All: College Coaching, Financial Aid, and Post-Secondary Persistence in Tennessee. *Economics of Education Review*, (51). <https://doi.org/10.1016/j.econedurev.2015.06.001>
- Domina, T., McEachin, A., Penner, A. i Penner, E. (2015). Aiming High and Falling Short: California's Eighth-Grade Algebra-for-All Effort. *Educational Evaluation and Policy Analysis*, 37(3). <https://doi.org/10.3102/0162373714543685>

- Duflo, E. (2001). Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment. *American Economic Review*, 91(4). <https://doi.org/10.1257/aer.91.4.795>
- Fox, J., Kleiber, C., Zeileis, A., i Kuschnig, N. (2023). *\_ivreg: Instrumental-Variables Regression by "2SLS", "2SM", or "2SMM", with Diagnostics*. R Package Version 0.6-2. CRAN. <https://CRAN.R-project.org/package=ivreg>
- Fredriksson, A. i Oliveira, G.M.D. (2019). Impact Evaluation Using Difference-in-Differences. *RAUSP Management Journal*, 54(4). <https://doi.org/10.1108/RAUSP-05-2019-0112>
- Furquim, F., Corral, D. i Hillman, N. (2020). A Primer for Interpreting and Designing Difference-in-Differences Studies in Higher Education. Research. W: L. Perna (red.), *Higher Education: Handbook of Theory and Research*, (35). Springer, Cham. [https://doi.org/10.1007/978-3-030-31365-4\\_5](https://doi.org/10.1007/978-3-030-31365-4_5)
- Gershenson, S. i Tekin, E. (2018). The Effect of Community Traumatic Events on Student Achievement: Evidence from the Beltway Sniper Attacks. *Education Finance and Policy*, 13(4). [https://doi.org/10.1162/edfp\\_a\\_00234](https://doi.org/10.1162/edfp_a_00234)
- Hamory, J., Miguel, E., Walker, M., Kremer, M. i Baird, S. (2021). Twenty-Year Economic Impacts of Deworming. *Proceedings of the National Academy of Sciences of the United States of America*, 118(14). <https://doi.org/10.1073/pnas.2023185118>
- Legge, J. D., McDivitt, J. F., Leinbach, T. R., Mohamad, G. S., Wolters, O. W. i Asvi Warman, A. (2023). Indonesia. W: *Encyclopedia Britannica*. Pobrano 15 czerwca 2023, z <https://www.britannica.com/place/Indonesia>
- Kelchen, R., Rosinger, K. O. i Ortagus, J. C. (2019). How to Create and Use State-level Policy Data Sets in Education Research. *AERA Open*, 5(3). <https://doi.org/10.1177/2332858419873619>
- Lan Ha. (2016). Rising Education Standards in Emerging Market Economies Will Support Income and Economic Growth. *Passport*. <https://www.euromonitor.com/article/rising-education-standards-in-emerging-market-economies-will-support-income-and-economic-growth>
- Lucas, A. M. (2010). Malaria Eradication and Educational Attainment: Evidence from Paraguay and Sri Lanka. *American Economic Journal: Applied Economics*, 2(2). <https://doi.org/10.1257/app.2.2.46>
- Lucas, A. M. i Mbiti, I. M. (2012). Access, Sorting, and Achievement: The Short-run Effects of Free Primary Education in Kenya. *American Economic Journal: Applied Economics*, 4(4). <https://doi.org/10.1257/app.4.4.226>
- Martinez-Bravo, M. (2017). The Local Political Economy Effects of School Construction in Indonesia. *American Economic Journal Applied Economics*, 9(2). <https://doi.org/10.1257/app.20150447>
- Miguel, E. i Kremer, M. (2004). Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities. *Econometrica*, 72(1). <https://doi.org/10.1111/j.1468-0262.2004.00481.x>
- MRTHE (Indonesia). (2022). Government Budget Allocation for Education in Indonesia in 2021, by Program (in Trillion Indonesian Rupiah) [Graph]. W: *Statista*. Pobrano 15 czerwca 2023 z <https://www.statista.com/statistics/1298487/indonesia-government-education-budget-allocation-by-program/>
- Polcyn, J. (2018). Model finansowania systemu edukacji. *Problems of Economics and Law*, (1). <https://bazekon.uek.krakow.pl/rekord/171614877>
- Psacharopoulos, G. i Patrinos, H. A. (2018). Returns to Investment in Education: A Decennial Review of the Global Literature. *Education Economics*, 26(5). <https://doi.org/10.1080/09645292.2018.1484426>
- Purnastuti, L., Salim, R. i Joarder, M. A. M. (2015). The Returns to Education in Indonesia: Post Reform Estimates. *The Journal of Developing Areas*, 49(3). <https://doi.org/10.1353/jda.2015.0174>
- R Core Team. (2022). *R: A language and Environment For Statistical Computing*. R Foundation for Statistical Computing. Vienna. <https://www.R-project.org/>
- Roth, J., Sant'Anna, P. H. C., Bilinski, A. i Poe, J. (2023). What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics' Literature. *Journal of Econometrics*, 235(2). <https://doi.org/10.1016/J.JECONOM.2023.03.008>

- Stock, J. H. (2015). Instrumental Variables in Statistics and Econometrics. W: J. D. Wright (red.), *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)*. <https://doi.org/10.1016/B978-0-08-097086-8.42037-4>
- Wing, C., Simon, K. i Bello-Gomez, R. A. (2018). Designing Difference-in-Difference Studies: Best Practices for Public Health Policy Research. *Annual Review of Public Health*, (39), 453-469. <https://doi.org/10.1016/B978-0-08-097086-8.42037-4>

## **Difference-in-Differences Models in Evaluating Returns of Economic Events to Education**

**Abstract:** This article addresses the application of Difference-in-Differences models in research within the field of development economics, with a particular focus on the education sector, which has been gaining increasing significance in the 21st century. The research objective is to analyze the utility of these models in measuring the effects of economic events, primarily government and social programs and projects, on the improvement of educational quality and accessibility. To achieve this goal, actions in this area will be examined in Indonesia. Calculations will be conducted using the pre-Duflo proposal approach of Difference-in-Differences as well as after its introduction, aiming to assess the role of the correction suggested by the research team in enhancing the accuracy of measuring the effects of educational improvement.

**Keywords:** Difference-in-Differences, econometrics, returns to education, developmental economy