

PIOTR WÓJCIK*

Metody *quasi*-eksperymentalne i ich zastosowanie w badaniu wpływu historii na współczesność¹

Wprowadzenie

Ekonomiści od zawsze próbują odpowiedzieć na pytanie, dlaczego niektóre kraje czy regiony są bogatsze lub biedniejsze od innych. W drugiej połowie XX w. wiele modeli teoretycznych i analiz empirycznych dotyczących wzrostu gospodarczego poświęcono poszukiwaniu jego determinant. Liczne badania empiryczne pokazują, że czynniki geograficzne, takie jak szerokość geograficzna czy klimat, są istotnie powiązane z poziomem zamożności krajów i regionów, jednak interpretacja tej zależności nie jest jednoznaczna. Wielu badaczy wskazuje, że relacja ta nie ma charakteru bezpośredniego, ale jest skutkiem uwarunkowań historycznych – wpływu początkowych (odległych w czasie) warunków geograficznych na przestrzenne zróżnicowanie cech kulturowych, kapitału ludzkiego, kapitału społecznego, instytucji czy technologii, mających bezpośredni wpływ na poziom dochodu i produktywność w długim okresie (patrz np. Acemoglu 2005; Galor 2005, 2011).

W ciągu ostatnich piętnastu lat pojawiło się wiele opracowań, które koncentrują się na badaniu fundamentalnych czynników wpływających na bieżące zróżnicowanie zamożności krajów i regionów, ale mających swoje źródła w odległej czasem przeszłości, czyli historycznych wydarzeń lub instytucji (interesujący przegląd badań historycznych można znaleźć np. w: Nunn (2014)). Powstaje pytanie, jak daleko w przeszłość powinniśmy sięgać próbując zrozumieć źródła współczesnych zróżnicowań (Comin i in. 2010; Spolaore, Wacziarg 2013). W jaki sposób te odległe w czasie czynniki geograficzne czy historyczne determinują dzisiejsze różnice w wartościach wskaźników społeczno-ekonomicznych?

* Dr Piotr Wójcik – Uniwersytet Warszawski, Wydział Nauk Ekonomicznych; e-mail: pwojcik@wne.uw.edu.pl

¹ Niniejsza praca została wykonana w ramach projektu „Historyczne podziały terytorialne i instytucjonalne a osiągnięcia edukacyjne. Podejście interdyscyplinarne”, finansowanego ze środków Narodowego Centrum Nauki (2014/14/E/HS4/00089).

Jednym z najistotniejszych wydarzeń historycznych, które zdaniem wielu badaczy wpłynęły na długookresowy wzrost gospodarczy była kolonizacja odległych terytoriów przez mieszkańców krajów europejskich (m.in. Feyrer, Sacerdote 2009; Bruhn, Gallego 2012; Easterly, Levine 2016). Wśród innych ważnych wydarzeń z odległej historii rozważano m.in.: handel niewolnikami z Afryki, średniowieczny handel dalekomorski, protestancką reformację, rewolucję francuską, rewolucję meksykańską, zastosowanie pługa w rolnictwie, wynalezienie prasy drukarskiej, rewolucję neolityczną czy różne katastrofy przyrodnicze (patrz np. Nunn 2014).

Badanie wpływu historii na współczesność nie musi dotyczyć wyjaśniania aktualnego zróżnicowania wskaźników społeczno-ekonomicznych, ale także np. preferencji wyborczych (Gerber i in. 2011; Eggers i in. 2014). W pojawiających się w ostatnich latach badaniach dla Polski analizowany jest głównie wpływ uwarunkowań historycznych, powiązanych z trwającymi przez ponad stulecie zaborami ziem polskich przez trzy imperia, na współczesne zróżnicowanie wskaźników społeczno-ekonomicznych (np. Zarycki 2007; Bukowski 2015; Grosfeld, Zhuravskaya 2015; Wysokińska 2015; Becker i in. 2016).

Wiele badań empirycznych tropiących wpływ historii na współczesność obejmuje analizę unikalnych danych dotyczących historycznych charakterystyk wybranych obszarów geograficznych. Często stosowanym podejściem jest wnioskowanie na poziomie mikro, polegające na analizie zróżnicowania badanego zjawiska w pobliżu aktualnych lub historycznych granic administracyjnych. Zwykle w takim przypadku stosowane są specyficzne *quasi*-eksperymentalne metody badawcze. Metodologia jest w tym przypadku podobna do wykorzystywanej w badaniach ewaluacyjnych, w których ocenie podlega efektywność wdrażanych programów, polityk czy interwencji. Badania te, nazywane *quasi*-eksperymentalnymi albo naturalnymi eksperymentami, polegają na porównaniu pod względem wybranej miary efektywności grupy jednostek objętych programem czy interwencją z odpowiednio dobraną grupą kontrolną. Jedynym czynnikiem, który różnicuje obie grupy, jest udział (lub nie) w ocenianym programie. Jednak w *quasi*-eksperymentach, w odróżnieniu od standardowych eksperymentów, podział badanych jednostek na grupę objętą specjalnym traktowaniem i grupę kontrolną jest nielosowy – przeprowadzający badanie nie ma ten podział wpływu. Z tego względu wymagają one zastosowania odpowiedniej metodologii korygującej te różnice, zwykle z wykorzystaniem dodatkowych zmiennych kontrolnych. W przypadku analiz tropiących wpływ historii na obecne zróżnicowanie zjawiska społeczno-gospodarczego czynnikiem, który różnicuje dwie porównywane grupy może być położenie po jednej lub drugiej stronie historycznej granicy.

Metody stosowane w tego typu ewaluacjach obejmują podejście polegające na dopasowywaniu obserwacji z dwóch różnych grup poprzez analizę podobieństwa wybranych charakterystyk (*matching*, *propensity score matching*), wykorzystanie zmiennych instrumentalnych (*instrumental variables*) czy nieciągłej regresji (*regression discontinuity*). „Ewaluacje” historyczne nie wykorzystują metody „przed versus po” (*before vs. after*) i jej rozszerzenia, jakim jest metoda nazwana różnicą w różnicach (*difference-in-difference*), które wymagają zaobserwowania dla anali-

zowanych jednostek badanego zjawiska przed i po objęciu interwencją. Metody te są stosowane w sytuacji, gdy na ewaluację przeznaczone są ograniczone fundusze. W takim przypadku grupę kontrolną tworzy się z osób uczestniczących w programie (Maksim 2007). Ta sama grupa jest więc badana dwukrotnie – przed i po udziale w programie. Oczywiście nie jest to możliwe, gdy stosowaną „interwencją” jest lokalizacja po jednej bądź drugiej stronie historycznej granicy.

Celem opracowania niniejszego artykułu jest omówienie sposobów empirycznego badania wpływu historii na współczesne zróżnicowanie wskaźników społeczno-gospodarczych za pomocą metod *quasi*-eksperymentalnych, ze szczególnym uwzględnieniem nieciągłej analizy regresji. Dalsze części artykułu zawierają omówienie kolejnych metod *quasi*-eksperymentalnych, w tym w szczególności najobszerniej przedstawionej analizy nieciągłej regresji, uwzględniającej także przestrzenną nieciągłą regresję (*spatial regression discontinuity design*). Ostatnia część artykułu zawiera krótki przegląd badań, w których zastosowano omówione metody do empirycznego badania wpływu historii na współczesność.

1. Metody *quasi*-eksperymentalne

1.1. Wprowadzenie

Badania ewaluacyjne, mające na celu ocenę efektywności programów czy polityk, nazywane są *quasi*-eksperymentalnymi albo naturalnymi eksperymentami. Ich istota polega na porównaniu grupy jednostek objętych jakimś programem, polityką czy interwencją z właściwie dobraną grupą kontrolną w celu oceny efektywności analizowanej polityki czy programu (np. Moffit 1991). Jedynym czynnikiem, który różnicuje obie grupy, powinien być udział (lub nie) w ocenianym programie. W przypadku analiz tropiących wpływ historii na obecne zróżnicowanie zjawiska społeczno-gospodarczego czynnikiem, który różnicuje dwie porównywane grupy może być położenie po jednej lub drugiej stronie historycznej granicy. W odróżnieniu od eksperymentów losowych, w eksperymentach naturalnych podział analizowanych jednostek na grupę objętą specjalnym traktowaniem i grupę kontrolną jest nielosowy, grupy więc mogą się między sobą różnić także w odniesieniu do innych charakterystyk i dlatego wymagane jest zastosowanie odpowiedniej metodologii eliminującej te różnice. Efekt ocenianego programu jest obliczany jako różnica przeciętnej wartości analizowanego zjawiska w grupie osób objętych specjalnym traktowaniem² oraz w grupie kontrolnej³ – określa go akronim ATE (*Average Treatment Effect*).

² Wiele pojęć w literaturze eksperymentalnej i *quasi*-eksperymentalnej jest zapożyczone z literatury medycznej.

³ Bardzo często badana w ten sposób jest np. efektywność różnego typu aktywnych polityk rynku pracy (*Active Labour Market Policies, ALMP*) i ich wpływu na obniżenie stopy bezrobocia np. Frölich i Lechner (2010).

Szczególne potraktowanie obserwacji (objęcie programem, polityką, interwencją) zwykle oznacza się za pomocą zmiennej zero-jedynkowej D . Przyjmie ona wartość 1 dla jednostek potraktowanych w specjalny sposób oraz 0 dla obserwacji z grupy kontrolnej. Jeśli zmienną objaśnianą (np. prawdopodobieństwo bycia osobą bezrobotną czy wielkość osiąganego wynagrodzenia) oznaczymy jako Y , efekt specjalnego traktowania danej obserwacji będzie wyrażony następującą formułą:

$$(Y_i|D_i = 1) - (Y_i|D_i = 0). \quad (1)$$

Fundamentalny problem polega jednak na tym, że porównywane wartości nie są obserwowane równocześnie dla tej samej obserwacji (Imbens, Lemieux 2008). Jednostka i zostaje potraktowana w specjalny sposób i wtedy obserwowane jest $Y_i|D_i = 1$ (oznaczane często także jako Y_{1i}) albo też jest traktowana standardowo i wtedy można zaobserwować jedynie $Y_i|D_i = 0$ (inaczej Y_{0i}). Tak więc wyzwaniem w tego typu porównaniach jest skonstruowanie właściwej obserwacji kontrfaktycznej⁴ (*counterfactual*) dla każdej jednostki i , tzn. próba odtworzenia, co by się stało, gdyby obserwacja i nie była potraktowana w specjalny sposób, podczas gdy w rzeczywistości była objęta ocenianą interwencją lub na odwrót. W większości przypadków skonstruowanie kontrfaktycznego reprezentanta dla każdej analizowanej jednostki wymaga silnych założeń.

Jeśli uda się skonstruować odpowiednią grupę kontrolną, przeciętny efekt specjalnego traktowania (ATE) może być oszacowany jako różnica wartości oczekiwanych (średnich) w obu porównywanych grupach:

$$ATE = E(Y_i|D_i = 1) - E(Y_i|D_i = 0) = E(Y_{1i} - Y_{0i}), \quad (2)$$

gdzie i jest indeksem poszczególnych obserwacji. Gdy w analizie uwzględnione są dodatkowe charakterystyki porównywanych obiektów, oznaczone jako wektor zmiennych X , powyższe równanie będzie miało następującą postać:

$$ATE = E(Y_i|D_i = 1, X_i) - E(Y_i|D_i = 0, X_i). \quad (3)$$

Inną często analizowaną miarą efektywności ocenianej interwencji jest przeciętny efekt specjalnego traktowania dla tych, którzy zostali nim objęci (*average treatment effect on the treated* – *ATET* lub *ATT*), wyznaczany zgodnie z formułą:

$$ATET = E(Y_{1i} - Y_{0i}|D_i = 1) = E(Y_{1i}|D_i = 1) - E(Y_{0i}|D_i = 1). \quad (4)$$

Powyższe wyrażenie dobrze ilustruje kontrfaktyczną naturę analizowanego efektu przyczynowego. Odjemna (element po lewej stronie różnicy) to przeciętna wartość zmiennej objaśnianej dla grupy jednostek objętych specjalnym traktowaniem, co jest wielkością obserwowalną. Z kolei odjemnik (element odejmowa-

⁴ Podejście kontrfaktyczne polega na wykorzystaniu specjalnie dobranej grupy kontrolnej, mającej symulować nieobserwowalną w rzeczywistości sytuację alternatywnego potraktowania analizowanej grupy jednostek.

ny po prawej stronie różnicy) oznacza przeciętną wartość zmiennej objaśnianej dla osób objętych specjalnym traktowaniem w sytuacji, gdyby nie byli nim objęci (wielkość nieobserwowalna). Z tego powodu wymagana jest odpowiednia grupa kontrolna i/lub właściwa metoda estymacji efektu. Można pokazać, że ATE i ATET są ze sobą powiązane:

$$\begin{aligned}ATE &= E(Y_i|D_i = 1) - E(Y_i|D_i = 0) = E(Y_{1i}|D_i = 1) - E(Y_{0i}|D_i = 0) = \\&= E(Y_{1i}|D_i = 1) - E(Y_{0i}|D_i = 1) + E(Y_{0i}|D_i = 1) - E(Y_{0i}|D_i = 0) = \\&= E(Y_{1i} - Y_{0i}|D_i = 1) + [E(Y_{0i}|D_i = 1) - E(Y_{0i}|D_i = 0)] = \\&= ATET + \text{obciążenie selekcji}\end{aligned}\tag{5}$$

Obciążenie selekcji (*selection bias*) wynika z tego, że przeciętna wartość zmiennej objaśnianej dla osób nieobjętych specjalnym traktowaniem, czyli $E(Y_{0i}|D_i = 0)$, niekoniecznie jest właściwym odpowiednikiem przeciętnej wartości zmiennej objaśnianej dla osób objętych interwencją w sytuacji, gdyby nie zostali poddani specjalnemu traktowaniu, czyli $E(Y_{0i}|D_i = 1)$. Problem ten nie występuje w przypadku losowego przydziału do grupy objętej losowaniem. Tymczasem przyporządkowanie jednostki do jednej bądź drugiej grupy (udział w programie, objęcie interwencją) nie jest losowe, zależy także od nieobserwowalnych jej charakterystyk. Skutkiem tego estymator metody najmniejszych kwadratów (MNK) dla ATE jest niezgodny z powodu występowania korelacji między zmienną objaśniającą, definiującą szczególne potraktowanie obserwacji (D), i składnikiem losowym⁵. Tak więc wszystkie specyficzne metody estymacji efektu specjalnego potraktowania polegają *de facto* na eliminacji tej korelacji (ortogonalizacji D i składnika losowego). Do najczęściej stosowanych technik statystycznych, które pozwalają skorygować problem pominiętych nieobserwowalnych charakterystyk, należą metody oparte na dopasowaniu obserwacji (*matching*) oraz metoda zmiennych instrumentalnych (*instrumental variables*) opisane w kolejnych częściach artykułu.

1.2. Metody oparte na dopasowaniu obserwacji

U podstaw metod opartych na dopasowaniu obserwacji stoi założenie, że jeśli podział na grupy testowaną i kontrolną zależy od cech analizowanych obiektów, to wszystkie cechy decydujące o objęciu specjalnym traktowaniem powinny być obserwowane. Jest ono znane w literaturze jako założenie warunkowej niezależności (*conditional independence assumption – CIA*) lub jako dobór na podstawie cech obserwowanych (*selection on observables*).

⁵ Nieobserwowalne charakterystyki nieuwzględnione w modelu będą miały wpływ na składnik losowy, a tym samym będzie on skorelowany ze zmienną D , co spowoduje niegodność estymatora. Brak korelacji między zmiennymi objaśniającymi i składnikiem losowym jest warunkiem zgodności estymatora MNK. Założenie to jest bardzo rzadko spełnione w przypadku danych pochodzących z obserwacji (nielosowe próby).

Metody oparte na dopasowaniu, nazywane także w polskiej literaturze metodami łączenia danych (np. Strawiński 2014), polegają na tym, że każdej jednostce z grupy objętych specjalnym traktowaniem przyporządkowana zostaje bardzo do niej podobna w odniesieniu do analizowanych obserwowalnych charakterystyk (najlepiej do niej dopasowana) obserwacja z grupy nieobjętej interwencją. Rolą łączenia danych jest skonstruowanie grupy kontrolnej składającej się z obserwacji niepoddanych interwencji w taki sposób, aby była ona jak najbardziej podobna do grupy jednostek objętych specjalnym traktowaniem. Ma to na celu jak największe ograniczenie różnic w obserwowalnych charakterystykach obu grup (np. Rubin 1974; Trzeciński 2009). Utworzone w ten sposób pary obserwacji są nazywane „bliźniakami statystycznymi” i na próbie wszystkich dopasowanych par wykonywana jest estymacja przeciętnego efektu specjalnego potraktowania części obserwacji (ATE).

Istotnym ograniczeniem metod opartych na dopasowaniu obserwacji jest zjawisko znane jako „przekleństwo wielowymiarowości” – obserwowane charakterystyki analizowanych jednostek mogą być liczne, a ich zróżnicowanie między obserwacjami duże, przez co znalezienie właściwego „bliźniaka statystycznego” może być praktycznie niemożliwe, zwłaszcza w mniejszych próbach. Możliwym rozwiązaniem tego problemu może być zastosowanie łączenia obserwacji w pary według oceny ich skłonności do (lub prawdopodobieństwa) objęcia specjalnym traktowaniem (*propensity score matching* – PSM) zaproponowane przez Rosenbauma i Rubina (1983). W ten sposób zamiast dopasowywania obserwacji w wielu wymiarach problem zostaje zredukowany do jednego wskaźnika, upraszczając tym samym znacznie procedurę łączenia danych. Oczywiście wskaźnik ten – prawdopodobieństwo objęcia specjalnym traktowaniem warunkowane wektorem zmiennych objaśniających – musi zostać dla każdej obserwacji oszacowany na podstawie danych (np. z wykorzystaniem regresji logistycznej lub probitowej). Oprócz oczywistych zalet, uproszczenie dopasowania do jednego wskaźnika ma również wady. Rzeczywisty proces generujący dane (decydujący o przypisaniu specjalnego traktowania) nie jest zazwyczaj znany. Stosując metodę PSM, rozważa się zwykle wiele specyfikacji przed wybraniem ostatecznej postaci modelu. Ważną słabością metody PSM jest więc próba przybliżenia eksperymentu całkowicie losowego zamiast bardziej efektywnego (często w większym stopniu eliminującego różnice między grupami) blokowego eksperymentu losowego, jak to ma miejsce przy innych metodach łączenia danych. Przez to wykorzystywanie PSM może więc dawać efekt przeciwny do spodziewanego – prowadzić do jeszcze większego niezbilansowania porównywanych grup (patrz King i Nielsen 2016).

Łączenie jednostek w pary może odbywać się bez zwracania – wtedy obserwacja przyporządkowana do swojego odpowiednika z grupy objętej interwencją nie może być już wykorzystana (połączona) w kolejnej parze – lub ze zwracaniem, gdzie takiego ograniczenia nie ma i każda obserwacja nieobjęta specjalnym traktowaniem może być wielokrotnie wykorzystana jako „statystyczny bliźniak”. Oba podejścia mają swoje wady – w dopasowywaniu bez zwracania może się zdarzyć, że w niektórych parach będą obserwacje dość od siebie odległe. Z kolei w przypadku łączenia ze zwracaniem w skrajnym przypadku może się okazać, że w utworzonej grupie kontrolnej znajdzie się jedynie kilka powtarzających się wielokrotnie obserwacji.

Dopasowywanie obserwacji w pary na podstawie oceny skłonności do objęcia interwencją może opierać się na algorytmie najbliższego sąsiada (*nearest neighbour matching*). Metoda najbliższego sąsiada oznacza, że dla każdej obserwacji z grupy objętej specjalnym traktowaniem jest dobierana jednostka z grupy kontrolnej, której ocena skłonności jest najbardziej zbliżona (zgodnie z przyjętą metryką odległości). Aby zapobiec dobieraniu zbyt odległego sąsiada, dopuszczalną odległość między łączonymi obserwacjami można ograniczyć do przyjętego progu tolerancji (*caliper matching*)⁶. W przypadku gdy takich najbliższych obserwacji z grupy kontrolnej jest kilka, ostateczny „statystyczny bliźniak” jest wybierany spośród nich za pomocą losowania prostego. *Caliper matching* daje podobny efekt jak dopuszczenie łączenia ze zwracaniem – pozwala unikać złych połączeń, dzięki czemu wzrasta jakość łączenia danych. Wariantem tego rozwiązania jest łączenie wg promienia (*radius matching*). W tym przypadku, zamiast opierać łączenie na pojedynczym najbliższym sąsiedzie w ramach przyjętego progu tolerancji, wykorzystuje się wszystkich sąsiadów mieszczących się w przyjętym zakresie, przypisując im wagi wg założonego schematu (np. jednostajny lub trójkątny) – patrz np. Caliendo i Kopeining (2008).

Z kolei nieparametryczne łączenie danych wykorzystujące funkcję jądrową (*kernel matching*) tworzy „statystycznego bliźniaka”, biorąc pod uwagę wszystkie obserwacje z próby i ważąc je za pomocą przyjętej funkcji jądra (*kernel*). Wagi zależą od odległości poszczególnych obserwacji od grupy kontrolnej i od jednostki, dla której jest konstruowana obserwacja kontrfaktyczna. Jeśli wykorzystywana funkcja jądrowa jest symetryczna, nieujemna i jednomodalna, obserwacjom bliższym porównywanej jednostce w odniesieniu do wartości *propensity score* przypisywane są wyższe wagi, a najbardziej niepodobnym wagi najniższe. Przy estymacji nieparametrycznej dobór funkcji jądra nie ma kluczowego wpływu na wynik estymacji. Kluczowe jest określenie właściwej szerokości pasma estymacji (*bandwidth, window*) – patrz np. Silverman (1986).

Alternatywnie obliczone dla całej próby oceny skłonności do objęcia specjalnym traktowaniem można wykorzystać do skonstruowania odwrotnie do nich proporcjonalnych wag poszczególnych obserwacji, a następnie oszacowania na danych ważonych modelu pozwalającego ocenić efekt interwencji. Podejście to nazywane jest ważeniem odwrotnością prawdopodobieństwa (*inverse probability weighting*) – patrz Strawiński (2014).

1.3. Metoda zmiennych instrumentalnych

Założenie, że selekcja jednostek do grupy objętej specjalnym traktowaniem zależy wyłącznie od obserwowalnych cech obserwacji, które jest podstawą omówionych wcześniej metod opartych na dopasowaniu obserwacji, często w praktyce jest mało wiarygodne. Wiele ważnych czynników selekcji jest w praktyce niemierzal-

⁶ Smith i Todd (2005) zwracają uwagę, że wadą tej metody jest konieczność określenia sensownej wartości progu tolerancji – trudno określić *a priori*, jaki jego poziom będzie optymalny.

nych albo zwyczajnie nieznanymi. W takim przypadku, jak już wspomniano wcześniej, wystąpi korelacja składnika losowego ze zmienną objaśniającą kontrolującą objęcie specjalnym traktowaniem, przez co estymator MNK równania regresji będzie niezgodny. Jednym z częściej stosowanych podejść empirycznych, które ten problem rozwiązują, jest metoda zmiennych instrumentalnych (*instrumental variables*). Środkiem do uzyskania poprawnych ocen wpływu interwencji na analizowane zjawisko jest znalezienie zastępczej zmiennej (instrumentu), która będzie skorelowana ze zmienną kontrolującą selekcję do próby testowej, ale nieskorelowana ze składnikiem losowym. Przy założeniu prostej liniowej zależności model szacujący efekt objęcia specjalnym traktowaniem można sobie wyobrazić jako:

$$Y_i = \alpha + D_i\gamma + \varepsilon_i, \quad (6)$$

gdzie oznaczenia są identyczne jak stosowane wcześniej (Y oznacza analizowane zjawisko, D jest zmienną zero-jedynkową wskazującą, czy obserwacja została potraktowana w specjalny sposób, czy nie, a ε to składnik losowy). Zmienna D nie jest w tym przypadku egzogeniczna ze względu na korelację ze składnikiem losowym. Jeśli uda się znaleźć zmienną Z , która będzie skorelowana z D , ale nieskorelowana z ε , można jej użyć jako instrumentu dla D , w celu uzyskania zgodnych oszacowań parametru γ . Jeśli Z jest także zmienną zero-jedynkową, to można pokazać (patrz np. Angrist, Pischke 2009), że estymator efektu specjalnego traktowania ma postać:

$$\gamma_{IV} = \frac{E(Y_i | Z_i = 1) - E(Y_i | Z_i = 0)}{E(D_i | Z_i = 1) - E(D_i | Z_i = 0)}. \quad (7)$$

Wyrażenie to jest często nazywane estymatorem Walda, gdyż po raz pierwszy pojawiło się w artykule Walda (1940), dotyczącym błędów pomiaru zmiennych. W przypadku więcej niż jednej zmiennej instrumentalnej albo pojedynczego instrumentu przyjmującego więcej niż dwa poziomy formuła na powyższy estymator jest bardziej złożona. Może on być jednak zwykle przedstawiony jako kombinacja liniowa estymatorów Walda (Angrist 1991).

Metoda zmiennych instrumentalnych nie pozwoli oszacować przeciętnego efektu specjalnego traktowania (ATE) ani ATT, jeśli specjalne potraktowanie części obserwacji zależy od więcej niż jednej charakterystyki (zero-jedynkowej zmiennej instrumentalnej). Można jednak zdefiniować podgrupy obserwacji, dla których o przypisaniu specjalnego traktowania decyduje konkretna zmienna instrumentalna (zawsze należą do grupy kontrolnej, jeśli zmienna ta jest równa 0, natomiast jeśli zmienna jest równa 1, to zawsze jest im przypisane specjalne traktowanie). Tak zdefiniowaną podgrupę Angrist Imbens i Rubin (1996) nazywają obserwacjami zgodnymi (*compliers*), tzn. (przestrzegającymi zasad)⁷.

⁷ Ponownie jest to termin zaczerpnięty z badań medycznych. W badaniach eksperymentalnych dotyczących terapii medycznych nie wszyscy pacjenci stosują się do ustalonych reguł, np. przyjmują zapisane im lekarstwa. Grupa jednostek zgodnych może być więc przez analogię zdefiniowana jako „przyjmujący zapisane lekarstwa”.

Zdefiniujmy D_{0i} i D_{1i} jako zmienne opisujące przypisanie do specjalnego traktowania w sytuacji, gdy Z_i jest równe odpowiednio 0 i 1, oraz dwa dodatkowe założenia: 1) warunkową niezależność: łączny rozkład $\{Y_{1i}, Y_{0i}, D_{1i}, D_{0i}\}$ jest niezależny od Z_i ; 2) monotoniczność: $D_{1i} \geq D_{0i}$ dla każdej obserwacji i ⁸. Jeśli te dwa założenia są spełnione, estymator metody zmiennych instrumentalnych daje zgodne oszacowanie przeciętnego efektu objęcia interwencją dla jednostek, które przestrzegają reguły. Jest on określany w literaturze jako lokalny przeciętny efekt specjalnego traktowania (*local average treatment effect* – LATE) – patrz np. Imbens i Angrist (1994). Formalnie można go zapisać jako $E[Y_{1i} - Y_{0i} | D_{1i} > D_{0i}]$. Generalnie więc obserwacje zgodne są podgrupą wszystkich jednostek poddanych specjalnemu traktowaniu. W szczególnym przypadku, gdy D_{0i} jest równe zero dla każdego i , LATE = ATT.

Modele liniowe z wykorzystaniem zmiennych instrumentalnych (również wielowartościowych czy nawet ciągłych) i innych charakterystyk analizowanych jednostek są zwykle szacowane za pomocą dwustopniowej metody najmniejszych kwadratów (*two-stage least squares* – TSLS, 2SLS) – patrz np. Angrist i Imbens (1995), Angrist i Krueger (2001), choć możliwa jest także estymacja metodami nieparametrycznymi – np. Frölich (2007).

1.4. Analiza nieciągłej regresji

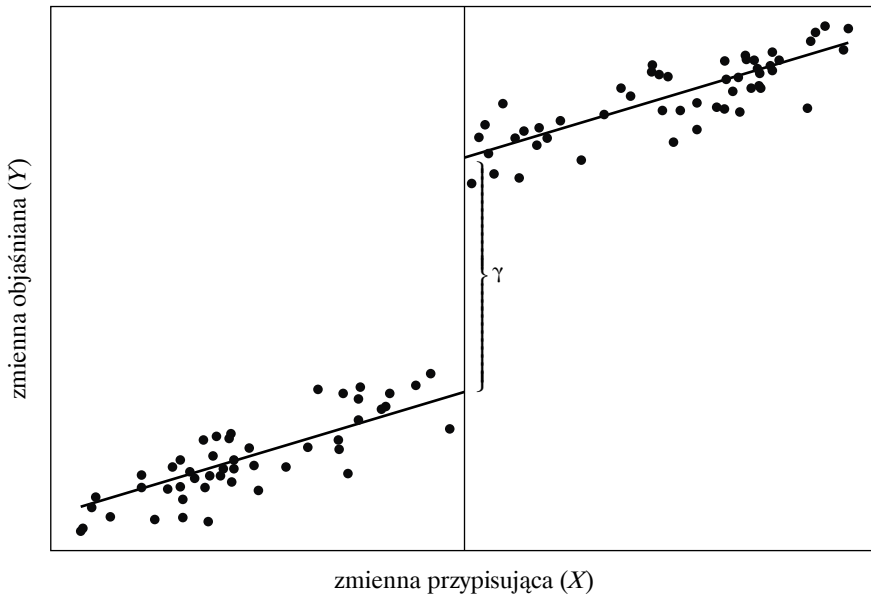
Choć nieciągła analiza regresji została zaproponowana przez Thistlethwaite’a i Campbella (1960) ponad pięćdziesiąt lat temu, jej wykorzystanie w pracach empirycznych rośnie bardzo dynamicznie dopiero w ostatnich piętnastu latach. Historia rozwoju metody nieciągłej regresji i jej wykorzystania w badaniach psychologicznych, poświęconych edukacji, statystycznych i ekonomicznych opisana została przez Cooka (2008). Lee i Lemieux (2010) naliczyli 78 zastosowań tej metody w badaniach ekonomicznych. Liczba zastosowań nieciągłej regresji rośnie bardzo szybko także w naukach politycznych (np. Brookman 2009; Caughey, Sekhon 2011; Gerber i in. 2011; Eggers i in. 2014).

Thistlethwaite i Campbell (1960) analizowali wpływ nagród za wyniki w nauce otrzymywanych przez studentów na ich przyszłe osiągnięcia akademickie. W nieciągłej regresji zakłada się, że istnieje zmienna, która ma nieciągły wpływ na badane zjawisko, przy czym nieciągłość występuje tylko w jednym punkcie (*cut-off*). Tym punktem jest wartość graniczna wspomnianej zmiennej, która decyduje o potraktowaniu wybranych obserwacji w specyficzny sposób – zaliczeniu ich do jednej z dwóch grup. Przykładem może być analiza, w której wynik punktowy na egzaminie wstępnym na studia wpływa na rozwój kariery zawodowej absolwenta (np. na uzyskiwane wynagrodzenie) i wpływ ten jest nieciągły w jednym punkcie – dla wartości granicznej, która decyduje o przyjęciu lub nieprzyjęciu kandydata na prestiżowy kierunek studiów. Oznacza to, że w punkcie odcięcia (przy przejściu

⁸ Monotoniczność może mieć alternatywnie przeciwny znak zależności.

z grupy osób nieprzyjętych na studia do grupy osób zakwalifikowanych), następuje nieciągły skok linii opisującej relację między wynikiem na egzaminie wstępnym a późniejszym wynagrodzeniem (patrz rysunek 1).

Rysunek 1
Przykład liniowej nieciągłej regresji



Źródło: opracowanie własne.

Skok w punkcie progowym wywołuje rozbieżność w traktowaniu poszczególnych obserwacji, który może być traktowany (przy pewnych założeniach) jako niezwiązany z innymi potencjalnymi zmiennymi zakłócającymi (*confounders*). Zakłada się więc, że owe rozbieżne potraktowanie poszczególnych jednostek jest wynikiem wpływu wyłącznie osiągnięcia przez analizowany czynnik X pewnego poziomu granicznego x_0 . Dlatego wnioskowanie w analizie nieciągłej regresji jest zwykle przeprowadzane z wykorzystaniem wyłącznie obserwacji położonych blisko punktu odcięcia czy progu rozgraniczającego grupy potraktowane w różny sposób, w którym następuje nieciągła zmiana relacji. Ze względu na lokalną naturę tego podejścia oszacowanie przeciętnego wpływu różnego potraktowania grup obserwacji na badane zjawisko są przeprowadzane z wykorzystaniem lokalnej wielomianowej regresji nieparametrycznej (*local-polynomial nonparametric regression*) – patrz np. Calonico i in. (2015).

Opisywane tu podejście do nieciągłej regresji nazywane jest „ostrym” (*sharp RD*), ponieważ występuje w nim ostre rozróżnienie na dwie podgrupy w zależności od przekroczenia przez analizowaną zmienną pewnej wartości progowej. Wszystkie jednostki, dla których zmienna przypisująca do jednej z grup (*forcing variable*, *assignment variable* albo *score*) przyjmuje wartość powyżej progu, są poddane spe-

cialnemu traktowaniu, co z kolei nie dotyczy żadnej obserwacji z wartością tej zmiennej poniżej progu.

Oznaczmy przez X zmienną objaśniającą (np. wynik na egzaminie wstępnym na studia), a jako x_0 wartość progową decydującą o zakwalifikowaniu. Możemy także oznaczyć szczególne potraktowanie (np. przyjęcie na studia na prestiżowym kierunku) za pomocą zmiennej zero-jedynkowej D . Przyjmie ona wartość 1, gdy $X \geq x_0$ oraz 0, gdy $X < x_0$. Jeśli zmienną objaśnianą (np. wynagrodzenie osiągnięte przez absolwenta) oznaczmy jako Y , możemy zapisać następujące równanie:

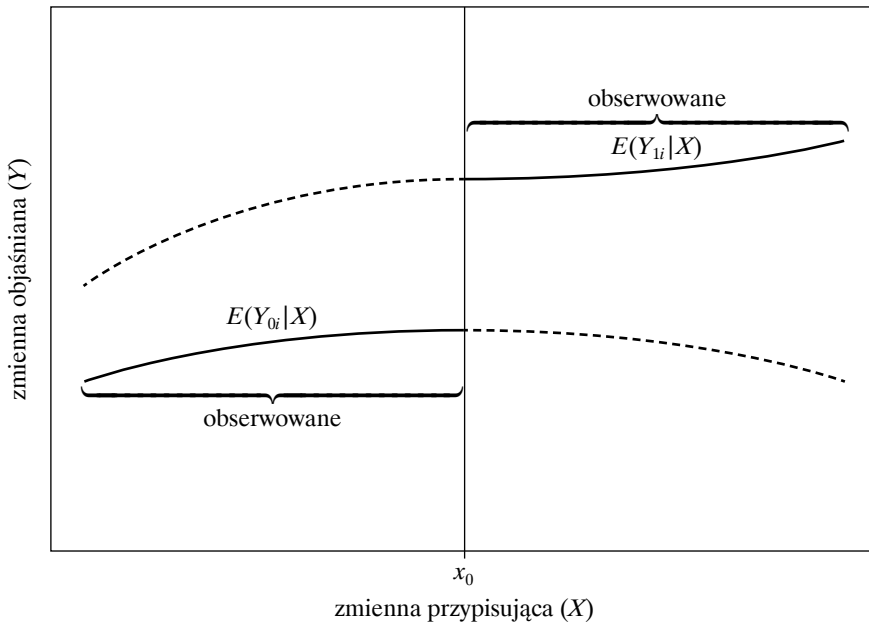
$$Y_i = \alpha + D_i\gamma + X_i\beta + \varepsilon_i, \quad (8)$$

gdzie ε jest składnikiem losowym, a α , β oraz γ parametrami opisującymi relację między zmienną zależną a zmiennymi objaśniającymi. W tym przypadku D jest nie tylko skorelowane z X , ale jest wręcz deterministyczną funkcją X . Równanie zakłada, że poza analizowanym czynnikiem X (np. wynikiem na egzaminie wstępnym) nie ma żadnej przyczyny dla nieciągłości opisywanej relacji. Efekt nieciągłości jest wychwytywany przez parametr γ , który różnicuje stałą w modelu między obiema grupami obserwacji, a zależność między X i Y jest poza tym liniowa (patrz ponownie rys. 1).

Można sobie wyobrazić, że dla każdej jednostki i mamy dwa potencjalnie możliwe wyniki: Y_{1i} , czyli wartość zmiennej zależnej (np. uzyskane wynagrodzenie) w przypadku specjalnego potraktowania (np. zakwalifikowania się na prestiżowe studia) oraz Y_{0i} , czyli wartość zmiennej zależnej w przypadku braku specjalnego potraktowania (np. niezakwalifikowaniu się na wybrane studia). Jak już opisano powyżej, przyczynowy wpływ specjalnego potraktowania może być przedstawiony jako różnica: $Y_{1i} - Y_{0i}$. Fundamentalny problem polega jednak na tym, że porównywane wartości nie są obserwowane równocześnie (Imbens, Lemieux 2008). Jednostka i albo zostaje potraktowana w specjalny sposób i wtedy obserwowane jest Y_{1i} albo też jest traktowana standardowo i wtedy można zaobserwować jedynie Y_{0i} . Dlatego zwykle jest analizowany przeciętny efekt specjalnego potraktowania, czyli różnica średniej (oczekiwanej) wartości Y_{1i} i średniej wartości Y_{0i} w całej analizowanej próbie lub podgrupach.

W podejściu nieciągłej regresji modeluje się dwie relacje między średnią zmienną Y a zmienną X , tzn. $E(Y_{1i}|X)$ oraz $E(Y_{0i}|X)$. Jednak z definicji podejścia nieciągłej regresji wszystkie jednostki osiągające wartość analizowanego czynnika X powyżej punktu odcięcia są traktowane w specjalny sposób, natomiast wszystkie jednostki, dla których zmienna X przyjmuje wartości poniżej progu są traktowane w standardowy sposób. Dlatego też wartości $E(Y_{1i}|X)$ są obserwowane jedynie po prawej stronie punktu odcięcia, a wartości $E(Y_{0i}|X)$ tylko po lewej stronie wartości granicznej – patrz rysunek 2. Co więcej, zależności te nie muszą być liniowe (np. Hahn i in. 2001), a nieobserwowana część każdej z relacji (zaznaczona na wykresie linią przerywaną) nie musi powielać obserwowanego wzorca. Prosty sposób zastosowania nieciągłej regresji w praktyce jest estymacja dwóch osobnych równań regresji po każdej ze stron punktu odcięcia.

Rysunek 2
Przykład nieliniowej nieciągłej regresji



Źródło: opracowanie własne.

Integralną częścią każdej analizy nieciągłej regresji powinna być graficzna weryfikacja zależności między Y i X ⁹. Wartości zmiennej przypisującej X można podzielić na względnie wąskie przedziały i na wykresie przedstawić wartości średnie zmiennej zależnej Y policzone w tych przedziałach. Należy zwrócić uwagę, aby żaden przedział nie obejmował wartości granicznej x_0 , tak aby obserwacje po obu stronach progu były w osobnych przedziałach. Prezentacja graficzna pozwala ocenić formę funkcyjną zależności¹⁰, sprawdzić, czy w punkcie progowym rzeczywiście występuje nieciągłość oraz wstępnie określić wielkość skoku w punkcie nieciągłości. Jeśli wykres pokaże podobne skoki także w innych punktach poza x_0 , trudne do uzasadnienia teoretycznego, może to podważyć wnioskowanie, że dyskretna zmiana w x_0 jest pochodną specjalnego traktowania części obserwacji¹¹. Oprócz wykresu średnich zmiennej Y w przedziałach wg wartości X warto też przedstawić graficznie zależność między zmiennymi Y i X dla indywidualnych obserwacji. Pozwoli to ocenić, czy ewentualna skokowa zmiana wartości zmien-

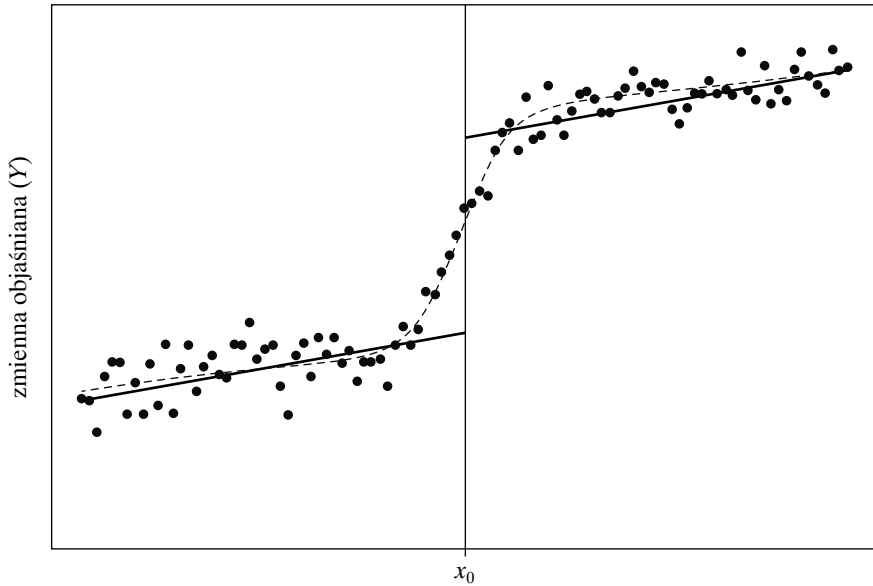
⁹ Możliwość przedstawienia relacji w formie graficznej, w celu wstępnej oceny występowania zależności, jest ważną zaletą analizy nieciągłej regresji w porównaniu z metodami alternatywnymi.

¹⁰ Wykres średnich w przedziałach można potraktować jako wstępny estymator jądrowy zależności między Y i X dla kwadratowej (jednostajnej) funkcji jądra o szerokości pasma estymacji równej szerokości przedziałów, dla których liczone są średnie.

¹¹ Istnieje możliwość bezpośredniego testowania występowania skoków zależności w punktach innych niż próg x_0 – patrz np. Lee i Lemieux (2015).

nej Y w punkcie x_0 , sugerująca (na podstawie wykresu średnich w przedziałach) nieciągłość w punkcie granicznym, nie jest po prostu efektem nieliniowej relacji między Y i X , niemającej nic wspólnego z nieciągłością (patrz rys. 3).

Rysunek 3
Nieliniowa zależność błędnie potraktowana jako nieciągłość regresji



Źródło: opracowanie własne.

W przypadku nieliniowej zależności między Y i X opisanej funkcją $f(\cdot)$ równanie zależności może być opisane następującą formułą:

$$Y_i = D_i\gamma + f(X_i) + \varepsilon_i, \quad (9)$$

a funkcja $f(\cdot)$ może być oszacowana za pomocą nieparametrycznej estymacji jądrowej (*kernel density estimation*) albo regresji parametrycznej z wykorzystaniem wielomianu rzędu p , np.

$$Y_i = \alpha + \beta_1 X_i + \beta_2 X_i^2 + \dots + \beta_p X_i^p + D_i\gamma + \varepsilon_i. \quad (10)$$

Dobór właściwego przybliżenia formy funkcyjnej ma kluczowe znaczenie dla poprawnego oszacowania efektu specjalnego potraktowania (γ). Dla modelu parametrycznego ważny jest właściwy wybór rzędu wykorzystywanego wielomianu, którego można dokonać, np. porównując wartości kryteriów informacyjnych dla poszczególnych wartości p (Black i in. 2007) lub stosując uogólnioną procedurę walidacji krzyżowej (van der Klaauw 2002). Jednak Gelman i Imbens (2014) sugerują, aby nie wykorzystywać wielomianów rzędu wyższego niż drugi, gdyż uzyskane w ich efekcie oszacowania wpływu specjalnego traktowania na badane zjawisko mogą być

mylące. Jak już wspomniano przy omawianiu metod łączenia danych, w przypadku zastosowania nieparametrycznej estymacji jądrowej szczególnie ważny jest dobór optymalnej szerokości pasma estymacji (*bandwidth*)¹², natomiast wybór stosowanej funkcji jądra (*kernel*) nie ma istotnego wpływu na uzyskany wynik (Imbens, Lemieux 2008)¹³. W praktyce warto przeprowadzić analizę wrażliwości uzyskiwanych wyników i ich odporności na dodawanie wyższych rzędów wielomianu (w przypadku parametrycznym) lub zmiany szerokości pasma estymacji (w przypadku nieparametrycznym). Na podstawie obserwowanych wartości zmiennych można obliczyć wartość skoku w punkcie nieciągłości jako różnicę granic wartości oczekiwanej Y przy zbieganiu przez X do progu x_0 odpowiednio z góry i z dołu:

$$\gamma = B - A = \lim_{\varepsilon \downarrow 0} E(Y_i | X_i = x_0 + \varepsilon) - \lim_{\varepsilon \downarrow 0} E(Y_i | X_i = x_0 - \varepsilon), \quad (11)$$

co będzie równe:

$$E(Y_{1i} - Y_{0i} | X = x_0). \quad (12)$$

Jest to lokalny przeciętny efekt specjalnego potraktowania (*local average treatment effect* – LATE) w punkcie odcięcia (dla wartości zmiennej objaśniającej dokładnie równej wartości progowej decydującej o specjalnym lub standardowym potraktowaniu – patrz Imbens i Angrist 1994). Co ważne, wartość ta nie może być uogólniona na całą populację jako przeciętny efekt specjalnego potraktowania dla całej dziedziny zmiennej X , gdyż skok obserwowany w punkcie x_0 może mieć zupełnie inną skalę dla pozostałych wartości X , co jest ważnym ograniczeniem omawianej metody.

Wnioskowanie w podejściu nieciągłej regresji jest możliwe pod warunkiem założenia o ciągłości funkcji $E(Y_{1i}|X)$ oraz $E(Y_{0i}|X)$ odpowiednio po prawej i po lewej stronie punktu odcięcia. Założenie to oznacza, że wszystkie pozostałe czynniki wpływające na badane zjawisko Y są ciągłe względem zmiennej X (Hahn i in 2001). Umożliwia to wykorzystanie przeciętnej wartości zmiennej zależnej Y dla jednostek o wartościach zmiennej objaśniającej X nieco poniżej progu x_0 jako właściwych reprezentantów stanu kontrfaktycznego dla jednostek uzyskujących wartości zmiennej X nieco powyżej progu. Założenie o ciągłości obu funkcji nie może być bezpośrednio testowane¹⁴. Jest ono jednak spełnione, jeśli rozkład analizowanych jednostek w otoczeniu przyjętego punktu odcięcia jest (lokalnie) losowy. Z tego punktu widzenia podejście nieciągłej regresji jest bliższe eksperym-

¹² Patrz m.in. Imbens i Lemieux (2008), Imbens i Kalyanaraman (2012), Otsu i in. (2015). Hahn i in. (2001) pokazują, że optymalna szerokość pasma estymacji jest proporcjonalna do $N^{-0.2}$, gdzie N jest liczebnością próby.

¹³ Najczęściej wykorzystuje się kwadratową (jednostajną) funkcję jądra ze względu na jej prostotę i związaną z tym większą niż w przypadku alternatywnych funkcji jądra efektywność obliczeniową.

¹⁴ Istnieją jednak testy dające podstawy twierdzić, że założenie o ciągłości jest spełnione. Można przetestować hipotezę o ciągłości rozkładu zmiennej X w punkcie nieciągłości x_0 (McCrary 2008) albo sprawdzić, czy inne zmienne (użyte jako zmienne objaśniane w modelu z czynnikiem X) wykazują skok w punkcie nieciągłości x_0 .

mentom losowym niż inne omawiane wcześniej metody oparte na dopasowaniu czy wykorzystujące zmienne instrumentalne. Różnica między nieciągłą regresją a eksperymentami losowymi polega na tym, że o ile w przypadku eksperymentów losowych efekt specjalnego potraktowania grupy jednostek może być obliczony jako zwykła różnica między przeciętną wartością zmiennej objaśnianej w grupie objętej specjalnym traktowaniem oraz przeciętną wartością w grupie kontrolnej, o tyle w przypadku nieciągłej regresji należy najpierw zastosować metody regresji w celu oszacowania lokalnych średnich analizowanej zmiennej w bezpośrednim sąsiedztwie punktu odcięcia. Poza tym stosując nieciągłą regresję, można, analogicznie jak w eksperymentach losowych, przeprowadzić analizę poprawnego dobrania grupy traktowanej specjalnie i grupy „kontrolnej”, sprawdzając ich podobieństwo pod względem zestawu dodatkowych charakterystyk.

W nieciągłej regresji (znów podobnie jak w eksperymencie losowym) nie ma potrzeby włączania do modelu dodatkowych zmiennych kontrolnych, jeśli obie grupy nie różnią się od siebie (lokalnie) pod względem dodatkowych charakterystyk. Jeśli jednak różnice występują (co może się zdarzyć zwłaszcza w mniejszych próbach), możliwe jest włączenie do modelu dodatkowych zmiennych objaśniających (kontrolnych, *covariates*). Mogą one posłużyć do wyeliminowania obciążenia spowodowanego małą próbą i zwiększyć dokładność estymacji. Jeśli rozkład dodatkowych zmiennych warunkowany zmienną przypisującą X jest ciągły w punkcie x_0 , ich dodanie do modelu będzie miało niewielki wpływ na estymator efektu specjalnego traktowania γ , ponieważ dodatkowe zmienne będą niezależne od zmiennej D . Nawet jeśli dodanie zmiennych kontrolnych do modelu nie zmieni znacząco relacji między Y i X w pobliżu punktu progowego x_0 , ich uwzględnienie może poprawić oszacowania (wyeliminować obciążenie) analizowanej relacji dla obserwacji położonych dalej od progu, które często także znajdują się w analizowanej próbie danych. Dodatkowo uwzględnienie zmiennych kontrolnych może zwiększyć dokładność estymacji, jeśli są one skorelowane ze zmienną objaśnianą. W praktyce jednak zwiększenie precyzji oszacowania estymatorów (zmniejszenie ich wariancji) nie będzie duże, o ile dodatkowe zmienne nie wyjaśniają dużej części zmienności badanego zjawiska (Imbens, Lemieux 2008).

Nieciągłość rozmyta

Ostre rozróżnienie decydujące o specjalnym traktowaniu obserwacji po przekroczeniu przez czynnik X wartości progowej nie jest jedynym podejściem stosowanym w nieciągłej regresji. Alternatywą jest założenie, że przydzielenie specjalnego traktowania jednostkom tylko częściowo zależy od przekroczenia przez zmienną X określonego progu. Oznacza to, że o specjalnym potraktowaniu niektórych jednostek decydują także dodatkowe czynniki. Podejście to nazywane jest rozmytą nieciągłą regresją (*fuzzy regression discontinuity*) – patrz np. Trochim (1984, 2001). W odróżnieniu od omówionej wcześniej „ostrej” nieciągłej regresji, gdzie prawdopodobieństwo specjalnego potraktowania obserwacji rośnie punktowo z 0 do 1

na progu x_0 , w podejściu rozmytym występuje w tym punkcie skok prawdopodobieństwa przypisania specjalnego traktowania, ale nie rośnie ono do 1. W tym przypadku prawdopodobieństwo to w zależności od wartości zmiennej przypisywanej może być opisane następującą formułą:

$$Prob(D_i = 1 | X_i) = \begin{cases} g_1(X_i) & \text{dla } X_i \geq x_0 \\ g_0(X_i) & \text{dla } X_i < x_0 \end{cases}, \quad (13)$$

gdzie $g_1(X_i) \neq g_0(X_i)$ mogą być dowolnymi funkcjami, które przyjmują różne wartości w punkcie x_0 . Rozmyta nieciągła regresja wymaga, aby:

$$\lim_{\varepsilon \downarrow 0} Prob(D = 1 | X = x_0 + \varepsilon) \neq \lim_{\varepsilon \downarrow 0} Prob(D = 1 | X = x_0 + \varepsilon). \quad (14)$$

Ponieważ w punkcie granicznym x_0 prawdopodobieństwo objęcia specjalnym traktowaniem rośnie o mniej niż 1, zmiana relacji między Y i X w tym punkcie nie może być w tym przypadku interpretowana jako przeciętny wpływ specjalnego potraktowania na badane zjawisko Y . Efekt ten może być obliczony przez podzielenie wielkości skoku relacji Y i X w punkcie x_0 przez zmianę warunkowego prawdopodobieństwa objęcia specjalnym traktowaniem w tym samym punkcie, zgodnie z następującą formułą (Hahn i in. 2001):

$$\gamma_F = \frac{\lim_{\varepsilon \downarrow 0} E(Y | X = x_0 + \varepsilon) - \lim_{\varepsilon \downarrow 0} E(Y | X = x_0 + \varepsilon)}{\lim_{\varepsilon \downarrow 0} E(D | X = x_0 + \varepsilon) - \lim_{\varepsilon \downarrow 0} E(D | X = x_0 + \varepsilon)}, \quad (15)$$

gdzie indeks F odnosi się do *fuzzy* (rozmytej) nieciągłej regresji.

Parametr γ_F może być interpretowany jako lokalny przeciętny efekt specjalnego potraktowania (*local average treatment effect – LATE*, patrz Imbens i Angrist 1994) dla tzw. obserwacji zgodnych (*compliers*), czyli jednostek, dla których przekroczenie progu x_0 oznacza specjalne traktowanie, którego by nie otrzymały bez przekroczenia tego progu przez zmienną X ¹⁵ (patrz np. Imbens, Lemieux 2008; Otsu i in. 2015):

$$\gamma_F = E(Y_{1i} - Y_{0i} | i \text{ jest obserwacją zgodną}, X_i = x_0). \quad (16)$$

Istnieje bliska analogia między powyższą formułą a wspomnianym wcześniej estymatorem Walda dla metody zmiennych instrumentalnych (porównaj formułę na γ_{IV}) i w związku z tym również dla rozmytej nieciągłej regresji może być on szacowany dwustopniową metodą najmniejszych kwadratów (patrz np. Hahn i in. 2001). Przegląd różnych estymatorów nieciągłej regresji oraz testów specyfikacji można znaleźć w Choi i Lee (2016).

¹⁵ Mianownik powyższej formuły, czyli zmiana warunkowego prawdopodobieństwa objęcia specjalnym traktowaniem w punkcie x_0 , będzie równy udziałowi obserwacji zgodnych (ang. *compliers*) w analizowanej próbie.

Granice geograficzne jako czynnik nieciągłości

Bardzo ważnym zastosowaniem nieciągłości regresji, zwłaszcza w kontekście empirycznego badania wpływu historii na bieżące zróżnicowanie poziomu rozwoju społeczno-gospodarczego między krajami czy regionami, jest wykorzystanie granic geograficznych (w tym granic historycznych) lub administracyjnych jako czynnika determinującego nieciągłość regresji. W tej sytuacji specjalne traktowanie niektórych jednostek analizy będzie sprowadzało się do ich lokalizacji po jednej stronie granicy geograficznej, podczas gdy traktowanie standardowe będzie oznaczało położenie po drugiej stronie granicy. Analiza nieciągłej regresji będzie więc tu polegała na ocenie wpływu istniejącej lub historycznej granicy geograficznej jako jedynego czynnika różnicującego w sposób nieciągły dwie grupy obserwacji pod względem badanego zjawiska. Zakłada się więc, że położenie jednostki po jednej lub drugiej stronie granicy ma charakter losowy. Podejście to jest określane w literaturze geograficzną nieciągłą regresją (*geographic regression discontinuity*, GRD – np. Keele i Titiunik 2015, 2016) lub przestrzenną nieciągłą regresją (*spatial regression discontinuity* – np. Egger i Lassmann 2014; Hidano i in. 2015).

Analiza geograficznej/przestrzennej nieciągłości może być przeprowadzona na dwa sposoby – jednowymiarowy i dwuwymiarowy (*two-dimensional regression discontinuity* – 2DRD, np. Imbens, Zajonc 2011; Hidano i in. 2015). W przypadku jednowymiarowym jako zmienną przypisującą specjalne traktowanie wykorzystuje się odległość euklidesową między lokalizacją obserwacji w przestrzeni a analizowaną granicą. Jako dodatkowe zmienne kontrolne można wykorzystać szczegółowe dane o położeniu obserwacji, czyli długość i szerokość geograficzną, a także np. wysokość nad poziomem morza czy zmienne klimatyczne. W podejściu dwuwymiarowym¹⁶ zmienna przypisująca ma postać dwuwymiarową – nieciągłość analizowanego zjawiska zależy od długości i szerokości geograficznej analizowanych obserwacji (np. Dell 2010; Grosfeld, Zhuravskaya 2015; Bukowski 2015). Kwestią do rozstrzygnięcia jest w tym przypadku postać funkcyjna zależności (rzęd wielomianu). Ze względu na brak wyraźnych wskazań metodologicznych badacze stosują zwykle porównanie kilku wariantów specyfikacji – np. Dell (2010) porównuje trzy specyfikacje: wielomian trzeciego rzędu z długości i szerokości geograficznej, odległość euklidesową od jednego z miast oraz wielomian trzeciego rzędu z odległości od granicy między porównywanymi obszarami.

W przypadku nieparametrycznej analizy geograficznej/przestrzennej nieciągłości dobór szerokości pasma estymacji będzie polegał na wyznaczeniu optymalnej odległości od analizowanej granicy (patrz np. Egger, Lassmann 2014). Dla specyfikacji jednowymiarowej powinna ona być względnie niewielka, aby zależ-

¹⁶ Jest to szczególnie przypadek względnie nowego podejścia nazwanego wielowymiarową analizą nieciągłej regresji (*multivariate regression discontinuity* – MRD). Pojęcie to nie oznacza wykorzystania w analizie dodatkowych zmiennych kontrolnych, lecz odnosi się do modelu, w którym specjalne potraktowanie zależy od przekroczenia ustalonych progów przez więcej niż jedną zmienną przypisującą (np. przyznanie stypendium nagrody za wyniki w nauce zależne od zdania dwóch lub więcej egzaminów, każdego z przekroczeniem odpowiednio wysokiego wyniku punktowego, niekoniecznie identycznego dla wszystkich egzaminów) – więcej np. w Imbens i Zajonc (2011), Papay i in. (2011), Wong i In. (2013).

ność między analizowanym zjawiskiem a odległością od granicy była wystarczająco dobrze przybliżona funkcją liniową.

2. Wybrane zastosowania metod *quasi*-eksperymentalnych do badania wpływu historii na współczesność

Druga część artykułu jest poświęcona omówieniu wybranych opracowań empirycznych wykorzystujących scharakteryzowane powyżej metody do tropienia wpływu historii na współczesność czy też inaczej mówiąc – do analizy wyników historycznych eksperymentów, ze szczególnym uwzględnieniem różnych wariantów przestrzennej analizy nieciągłej regresji. Przegląd ma na celu pokazanie zróżnicowania wykorzystywanej w pracach empirycznych metodologii w przypadku podziału analizowanych jednostek na dwie rozłączne grupy – objętych i nieobjętych historycznym eksperymentem. Znacznie obszerniejszy przegląd wyników badań empirycznych wpływu historii na współczesność, obejmujący wykorzystanie także innych metod niż opisane w części metodologicznej tego artykułu, często zastosowanych do innego niż zero-jedynkowy kontrolowania historycznego eksperymentu, można znaleźć np. w pracy Nunn (2014).

2.1. Łączenie danych i metoda zmiennych instrumentalnych

Hornung (2012) badał na poziomie mikroregionalnym wpływ dostępu do sieci kolejowej na rozwój gospodarczy 978 miast pruskich w XIX wieku. Wykorzystując dane z systemu GIS, autor skonstruował dane geolokalizacyjne pokazujące przestrzenny rozwój sieci kolejowej na terenie Prus oraz lokalizację wszystkich pruskich miast w okresie budowania głównych szlaków kolejowych, czyli od 1838 r. do lat 60. XIX w. Punktem wyjścia jest zwykła regresja liniowa szacowana metodą najmniejszych kwadratów. Zmienną zależną w modelu jest średnioroczne tempo wzrostu populacji miasta (w różnych okresach począwszy od 1831–37 do 1867–71), służące za przybliżenie niedostępnych dla tego okresu danych o tempie rozwoju gospodarczego. Najważniejszą zmienną objaśniającą jest mierzony zero-jedynkowo dostęp do sieci kolejowej w 1848 r. Jako dodatkowe zmienne kontrolne wykorzystano m.in. wielkość populacji cywilnej i wojskowej miasta, udział robotników w populacji, współczynnik skolaryzacji, ale także opóźnioną zmienną zależną czy późniejszy niż w 1848 r. dostęp do sieci kolejowej. Wyniki estymacji pokazują, że dostęp do sieci kolejowej istotnie przyspiesza rozwój miasta, co więcej wpływ ten rośnie z upływem lat, co wskazuje na pozytywny efekt długookresowy. Co więcej, specyfikacja kontryfaktyczna dla okresu 1831–37 pokazuje, że miasta, które miały dostęp do sieci kolejowej w 1848 r., nie rozwijały się uprzednio szybciej od pozostałych miast. W opisanej analizie autor wykorzystuje założenie, że w badanym okresie sieci kolejowe były budowane wyłącznie

między ważnymi miastami, a miasta leżące po drodze uzyskiwały dostęp do kolei niejako przy okazji (w sposób losowy). Wynikało to z wysokich kosztów budowy kolei, przez co tory były układane w miarę możliwości w linii prostej. Miasta położone poza prostym korytarzem mogły uzyskać dostęp do kolei z innych powodów, potencjalnie powiązanych (endogenicznych) z rozwojem miasta. Okazuje się, że niektóre połączenia między głównymi miastami nie były zbudowane w linii prostej, co może rodzić problem endogeniczności. Aby ten problem rozwiązać, autor stosuje metodę zmiennych instrumentalnych. Jako zmienną instrumentalną dla uzyskania dostępu do kolei wykorzystano położenie miasta wzdłuż jednej z prostych linii łączących najważniejsze miasta pruskie (przyjęto korytarz $\pm 1,5$ km w każdą stronę). Wyniki estymacji metodą zmiennych instrumentalnych w pełni potwierdzają istotny pozytywny wpływ dostępu do sieci kolejowej na rozwój miast – jest on około dwukrotnie silniejszy niż w przypadku zwykłej regresji liniowej. Wśród dodatkowych testów odporności wyników autor stosuje estymację metodą zmiennych instrumentalnych na różnych próbkach powstałych z łączenia danych. W pierwszym podejściu model jest szacowany na próbie zawierającej miasta mające dostęp do kolei w 1848 r. oraz dla każdego z nich po dwóch najbliższych sąsiadów bez połączenia kolejowego (geograficzne łączenie danych). Oprócz tego została wykorzystana także metoda *propensity score matching*, w której miasta były łączone w pary, które różniły się dostępem do sieci kolejowej w 1848 r. Wartości *propensity score* były wyznaczane na podstawie zmiennych charakteryzujących wielkość miasta i poziom jego rozwoju w okresie przed budową pierwszych linii kolejowych. Należały do nich: wielkość populacji, przyrost liczby mieszkańców, znormalizowana liczba kupców, liczba krosien, liczba protestantów, liczba prywatnych nieruchomości, liczba budynków komercyjnych, wartość ubezpieczeniowa nieruchomości (w ubezpieczeniach przeciwpożarowych). Zastosowane podejście PSM wykorzystywało łączenie danych wg promienia (*radius matching*) oraz łączenie nieparametryczne z użyciem funkcji jądra (*kernel matching*). Wyniki estymacji metodą zmiennych instrumentalnych na łączonych próbkach danych potwierdzają pozytywny wpływ kolei na rozwój ekonomiczny miast pruskich, choć w przypadku niektórych podokresów wzrost błędów standardowych oszacowanych parametrów powoduje utratę istotności statystycznej wyniku.

Iyer (2010) analizowała kolonialne doświadczenia Indii, porównując współczesny poziom rozwoju na obszarach, które przed 1947 r. były pod bezpośrednią i pośrednią władzą Brytyjczyków. Przez władzę pośrednią autorka rozumie obszary (księstwa) pozostające pod brytyjskim protektoratem, ale rządzone przez lokalnych władców, natomiast władza bezpośrednia odnosi się do prowincji administrowanych bezpośrednio przez władze brytyjskie. Księstwa cieszyły się szeroką autonomią w zakresie administracji wewnętrznej. Po zakończeniu brytyjskiego panowania w 1947 r. wszystkie obszary zostały zintegrowane w niepodległe Indie i objęte jednolitą strukturą administracyjną, prawną i polityczną. Analiza przeprowadzona została na poziomie 415 dystryktów z wykorzystaniem modelu liniowego. Jako zmienne zależne wykorzystano współczesne miary dobrostanu, takie jak wielkość inwestycji rolniczych oraz produktywności w rolnictwie (dane o dochodach i konsumpcji nie są

dostępne na poziomie dystryktów), dostępność dóbr publicznych, takich jak szkoły, przychodnie, drogi, miary dotyczące zdrowia i edukacji, a także udział zatrudnionych w sektorze produkcyjnym i różne szacunkowe miary ubóstwa na poziomie dystryktów. Główną zmienną objaśniającą była przynależność do jednej z dwóch grup dystryktów – pod bezpośrednią lub pośrednią władzą brytyjską. Dodatkowo wykorzystano zmienne kontrolne, głównie geograficzne (m.in. szerokość geograficzna, wysokość nad poziomem morza, przeciętne opady, proporcja obszaru piaszczystego czy skalnego, zmienne zero-jedynkowe dla przeważającego rodzaju gleby czy dystryktu położonego na wybrzeżu) oraz demograficzne (np. wielkość populacji, gęstość zaludnienia, udział zatrudnionych w rolnictwie, udział obszarów wiejskich, udział piśmiennej populacji). Wyzwaniem analizy empirycznej była nielosowość (endogeniczność) podziału na obszary objęte bezpośrednim i pośrednim zarządem brytyjskim. Problem ten został rozwiązany przez wykorzystanie wiedzy o tzw. doktrynie Lapse’a, brytyjskiej polityce obowiązującej w latach 1848–1856, która nie uznawała adoptowanych potomków indyjskich książąt. W przypadku bezpotomnej (bez naturalnego potomka) śmierci lokalnego władcy rządzone przez niego księstwo przechodziło pod bezpośrednią władzę Brytyjczyków. Wyniki estymacji modelu dla wszystkich dystryktów wskazują na wyższe inwestycje i produktywność w rolnictwie w regionach zarządzanych bezpośrednio przez Brytyjczyków. Kontrastuje to wyraźnie z analizą wykorzystującą metodę zmiennych instrumentalnych, przeprowadzoną dla grupy 181 dystryktów, które nie były zarządzane bezpośrednio przez Brytyjczyków przed 1848 r. (tylko dla nich możliwe jest wyznaczenie wartości zmiennej instrumentalnej). Oszacowania z wykorzystaniem zmiennej instrumentalnej pokazują, że obszary pod bezpośrednią władzą Imperium Brytyjskiego nie wyróżniają się istotnie pod względem poziomu inwestycji i produktywności w rolnictwie, co więcej, oferują istotnie gorszą dostępność dóbr publicznych. Wynik ten sugeruje, że Brytyjczycy selektywnie anektowali obszary o największej produktywności i największym potencjale dla rolnictwa, ale nie inwestowali tak dużo jak lokalni władcy w kapitał fizyczny i ludzki. Historyczne niedoinwestowanie spowodowało, że dystrykty zarządzane bezpośrednio przez Imperium Brytyjskie charakteryzują się w okresie postkolonialnym wyższymi wskaźnikami ubóstwa i śmiertelności niemowląt. Autorka przeprowadza także wiele testów odporności wyników na zmiany założeń. Testy w pełni potwierdzają wnioski o negatywnym wpływie bezpośredniej władzy Brytyjczyków w czasach kolonialnych na współczesny poziom rozwoju.

Dittmar (2011) analizował długoterminowe efekty wynalezienia prasy drukarskiej, po raz pierwszy zastosowanej w Mainz w Niemczech w latach 1446–1450. Autor konstruuje i wykorzystuje unikalne dane historyczne dotyczące europejskich miast dla okresów stuletnich między rokiem 1300 a 1800, łącząc m.in. dane o wielkości populacji miast z informacjami o zastosowaniu w nich prasy drukarskiej. Zmienną zależną w modelu jest zlogarytmowane tempo wzrostu populacji w kolejnym okresie (1400–1500, 1500–1600, 1500–1700, 1500–1800), a najważniejszym analizowanym czynnikiem zmienna zero-jedynkowa dotycząca zastosowania w mieście druku w okresie 1450–1500. Wśród dodatkowych zmiennych kontrolnych pojawiają się m.in. wielkość populacji na początku analizowanego okresu, liczba

książek wydanych w okresie 1450–1500 na 10000 mieszkańców, a także zmienne zero-jedynkowe dla posiadania uniwersytetu, lokacji na prawie rzymskim, historycznej stolicy, portu bałtyckiego, atlantyckiego lub śródziemnomorskiego, czy położenia przy żeglownej rzece. Wyniki oszacowania regresji liniowej pokazują, że miasta, które przyjęły prasę drukarską między rokiem 1450 a 1500, doświadczyły znacznie szybszego wzrostu liczby ludności w XVI–XIX w. W celu potwierdzenia powyższych wniosków autor stosuje (wspomnianą na początku niniejszego artykułu) metodę różnic w różnicach, porównując tempo wzrostu populacji w analizowanych miastach przed i po przyjęciu przez nie prasy drukarskiej (analiza na zbilansowanych danych panelowych dla 202 miast). Wyniki pokazują, że miasta, które przyjęły prasę drukarską w drugiej połowie XV w., nie rozwijały się uprzednio szybciej od pozostałych miast, natomiast po 1500 r. ich rozwój był istotnie szybszy. Zastosowanie prasy drukarskiej przez poszczególne miasta nie miało charakteru losowego – w pierwszych latach po jej wynalezieniu, do 1500 r., nowa technologia rozchodziła się wokół Mainz w koncentrycznych kołach, gdy drukarze z Mainz zakładali swoje warsztaty w okolicznych miastach. Aby kontrolować problem nielosowości, autor wykorzystuje także metodę zmiennych instrumentalnych, używając odległości od Mainz jako instrumentu dla wczesnego przyjęcia prasy drukarskiej. Wyniki analizy z wykorzystaniem zmiennej instrumentalnej potwierdzają wnioski z poprzednich modeli. Badając prawidłowość uzyskanych wyników, autor wykorzystuje dodatkowo jako potencjalne instrumenty (placebo) odległości od innych ważnych miast z tego okresu: Amsterdamu, Londynu, Paryża, Wenecji i Wittenbergi. Nie mają one jednak istotnego znaczenia dla relacji między zastosowaniem w mieście prasy drukarskiej a późniejszym tempem rozwoju miasta.

2.2. Analizy nieciągłej regresji

Opracowania wykorzystujące metodologię nieciągłej regresji do badania wpływu historii na współczesność są bardzo liczne. Poniżej ograniczono się do omówienia kilku niedawno opublikowanych analiz dotyczących Polski. Wszystkie omówione poniżej opracowania dotyczą analizy wpływu dziewiętnastowiecznych zaborów ziem polskich na różnice we współczesnych wartościach wskaźników społeczno-ekonomicznych. Podział ziem polskich między państwa zaborcze można potraktować jako historyczny eksperyment różnicujący na ponad sto lat warunki instytucjonalne i kulturowe na tych ziemiach. Granice zaborów miały charakter egzogeniczny, zostały wyznaczone arbitralnie, zależały od sprawności militarnej trzech państw zaborczych i nie powielały żadnych wcześniejszych granic. Można więc domniemywać, że czynniki, na które imperia zaborcze nie miały wpływu zachowują na historycznych granicach ciągłość (Grosfeld, Zhuravskaya 2015). Każde z omówionych opracowań wykorzystuje nieco inną metodologię nieciągłej regresji.

Wysokińska (2015) wykorzystuje nieparametryczną analizę nieciągłej regresji do zbadania różnic w wartościach wybranych wskaźników rozwojowych, kulturowych i instytucjonalnych na granicy byłych zaborów rosyjskiego i pruskiego oraz

rosyjskiego i austriackiego. Analiza została przeprowadzona na poziomie gmin i w zależności od badanych wskaźników dotyczy różnych lat z okresu 1997–2012. Jako zmienne kulturowe zostały wykorzystane wielkości frekwencji wyborczej w wyborach parlamentarnych w 1997 r. i w referendum akcesyjnym 2003 r., wielkość poparcia dla wejścia Polski do Unii Europejskiej w referendum w 2003 r. czy wielkość poparcia dla wybranych kandydatów w wyborach prezydenckich (Kwaśniewski w 2000 r., Kaczyński w 2005 r., Tusk w 2005 r., Komorowski w 2010 r.). Jako porównywane wskaźniki rozwojowe wzięto pod uwagę dochody własne gminy na 1 mieszkańca oraz dochody gminy osobno z PIT, z CIT oraz podatku rolnego w 2012 r. Z kolei analizowane zmienne instytucjonalne obejmowały przeciętną wielkość gospodarstwa, liczbę gospodarstw uwzględnionych w spisie rolnym (zajmujących się produkcją rolną) w 2010 r., długość procesu sądowego oraz udział procesów trwających powyżej pięciu lat, wynik sprawdzianu szóstoklasisty (różne części) oraz egzaminu gimnazjalnego. Postać funkcyjna szacowanego modelu była bardzo prosta – obejmuje stałą oraz zmienną zero-jedynkową kontrolującą położenie po jednej ze stron historycznej granicy między zaborami. W tym przypadku testowana więc była *de facto* różnica w średniej wartości wybranych wskaźników po obu stronach historycznej granicy (jako szerokość pasma estymacji przyjęto odległość 50 km). Autorka weryfikowała także odporność wyników na zmiany założeń, dokonując zawężenia granicy odległości do 40, 30, 20 i 10 km, a także stosując test Zivota-Andrewsa na najbardziej prawdopodobną zmianę strukturalną (*structural break*) zastosowaną do wskazania optymalnej szerokości pasma estymacji (odległości od granicy), dla której zachodzi skokowa zmiana wartości zmiennej objaśnianej. Dodatkową formą weryfikacji wyników było zbadanie efektu granicy placebo, polegające na sztucznym przesunięciu historycznej granicy o 25 km w jedną albo drugą stronę, w głąb jednego z porównywanych zaborów i porównanie wyników oszacowań do rezultatów modelu z rzeczywistą historyczną granicą. Autorka stwierdziła istotne statystycznie i odporne na zmianę założeń modelu różnice między byłymi zaborami pruskim a rosyjskim w odniesieniu do zmiennych dotyczących poziomu rozwoju, a także przeciętnej wielkości gospodarstwa rolnego, liczby gospodarstw zaangażowanych w produkcję rolniczą czy wielkości frekwencji wyborczej. Przy porównaniu byłego zaboru rosyjskiego z austriackim autorka także stwierdza występowanie niewielkich różnic w wielkości gospodarstw w pasie 50 km od historycznej granicy, jednak nie są one odporne na zmianę założeń analizy. Istotne statystycznie, jednak ponownie nieodporne na zmianę założeń, okazują się także różnice w odniesieniu do wielkości frekwencji wyborczej i preferencji politycznych. Autorka nie stwierdza natomiast różnic między historycznymi zaborami w odniesieniu do analizowanych wskaźników efektywności systemu prawnego czy osiągnięć edukacyjnych.

Grosfeld i Zhuravskaya (2015) stosują parametryczną analizę nieciągłej regresji do analizy różnic między dawnymi zaborami w poziomie rozwoju, infrastruktury, edukacji, religijności, stosunku do demokracji, zaufania do rządu, policji i sądów, poszanowaniu rządów prawa, a także tolerowaniu korupcji. Wykorzystano kilka źródeł danych, w tym dane ze Spisu Powszechnego z 2002 r. (struktura wykształ-

cenia ludności), z Banku Danych Lokalnych GUS (stopa bezrobocia, płace, produkcja przemysłowa na 1 mieszkańca, udział zatrudnienia w różnych sektorach gospodarki), Państwowej Komisji Wyborczej (frekwencja w wyborach parlamentarnych, procent głosów oddanych na zwycięską partię). Dodatkowo z Diagnozy Społecznej¹⁷ zaczerpnięto dane dotyczące częstości udziału w niedzielnych nabożeństwach, preferowania ustroju demokratycznego względem innych form rządów, zaufania do rządu, policji, sądów oraz wykorzystywania znajomości lub łapówek w kontaktach z administracją publiczną. Z kolei z Instytutu Statystyki Kościoła Katolickiego wzięto udział wiernych chodzących do kościoła. Z innych źródeł pochodziły dane dotyczące gęstości sieci kolejowej, wysokości nad poziomem morza, średniej temperatury, przeciętnych opadów oraz nasłonecznienia w poszczególnych gminach. Autorki analizują historyczną granicę między zaborem rosyjskim i pruskim oraz austriackim i rosyjskim, uznając granicę między zaborem pruskim i austriackim za zbyt krótką, aby możliwe było istotne statystycznie wnioskowanie. Analiza przeprowadzona została na poziomie gmin oraz powiatów leżących nie więcej niż 60 km od analizowanej historycznej granicy (odległość od geometrycznego środka jednostki terytorialnej do najbliższego punktu na granicy). W artykule zostały wykorzystane dwie specyfikacje parametryczne. Podejście jednowymiarowe wykorzystuje jako zmienną przypisującą odległość euklidesową od historycznej granicy między zaborami (zależność liniowa). Do modelu dodano także dodatkowe zmienne kontrolne, takie jak długość i szerokość geograficzna, wysokość nad poziomem morza oraz zmienną zero-jedynkową dla dużego miasta. Specyfikacja dwuwymiarowa zamiast odległości od granicy wykorzystuje wielomian trzeciego rzędu¹⁸ ze współrzędnych geograficznych (długości i szerokości), uwzględniając także dwie pozostałe zmienne kontrolne, czyli wysokość nad poziomem morza oraz zmienną zero-jedynkową dla dużego miasta. W celu zweryfikowania odporności wyników na zmianę założeń przeprowadzono test placebo polegający na przesunięciu historycznej granicy w dwie przeciwne strony o 15, 30, 45, 60 i 75 km. Granicę rosyjsko-pruską przesuwano na wschód bądź zachód, natomiast granicę rosyjsko-austriacką na północ lub południe. Testy placebo zostały przeprowadzone jedynie dla specyfikacji jednowymiarowej. Sprawdzano także wrażliwość modelu jednowymiarowego na zmianę postaci funkcyjnej zależności od zmiennej przypisującej (oprócz liniowej sprawdzono także formę kwadratową i sześcienną) oraz na zmianę szerokości pasma estymacji (100 km zamiast 60 km, a także dodatkowo z wykluczeniem z estymacji gmin/powiatów położonych przy samej historycznej granicy – w odległości do 5 km oraz do 10 km od niej). Wyniki analizy wskazują na trwałe wpływy zaborów na współczesne zróżnicowanie poziomu religijności, wiary w wartości demokratyczne i sieci kolejowej, ale nie na zróżnicowanie dochodów, produkcji przemysłowej, udziału ludności z wykształceniem średnim, korupcji i zaufania do instytucji rządowych. Autorki nie stwierdzają występowania istotnej

¹⁷ Wątpliwości może budzić reprezentatywność danych z Diagnozy Społecznej na poziomie powiatów, a tym bardziej gmin, do czego autorki opracowania się nie odnoszą.

¹⁸ Autorki wzorują się na Dell (2010).

nieciągłości preferencji politycznych na historycznej granicy zaborów rosyjskiego i pruskiego, natomiast odnotowują istotną nieciągłość na dawnej granicy rosyjsko-austriackiej, przypisując to różnicom w poziomie religijności.

Bukowski (2015) wykorzystuje dwuwymiarową przestrzenną nieciągłą regresję do zweryfikowania wpływu historycznych różnic w systemach edukacyjnych ukształtowanych na ziemiach polskich w okresie zaborów na aktualne zróżnicowanie wyników egzaminów na poziomie gmin. Autor bierze pod uwagę zarówno wynik sprawdzianu szóstoklasisty, jak i części matematyczno-przyrodniczej egzaminu gimnazjalnego w okresie 2005–2011. Analizie poddane są gminy wiejskie położone w odległości do 50 km od granicy odpowiednio między byłymi zaborami rosyjskim i pruskim albo rosyjskim i austriackim (z wyłączeniem Śląska i Prus Wschodnich). Specyfikacja modelu nieciągłej regresji obejmuje wielomian rzędu drugiego obliczony z długości i szerokości geograficznej geometrycznego środka gminy, zmienną zero-jedynkową kontrolującą położenie po jednej lub drugiej stronie granicy oraz dodatkowe geograficzne i klimatyczne zmienne kontrolne: wysokość nad poziomem morza, średnie opady oraz średnią temperaturę. Dane za cały okres 2005–2011 są analizowane łącznie (tzw. *pooled regression*). Autor przeprowadza dodatkowo analizę wrażliwości wyników na zmianę przyjętych założeń i rozważa również odległość (pasma estymacji) 75 km i 100 km od historycznej granicy, stosuje wielomian stopnia pierwszego, trzeciego i czwartego, a także szacuje regresję dla każdego roku osobno. Autor stwierdza występowanie dużo lepszych wyników egzaminów na ziemiach byłego zaboru austriackiego w porównaniu z byłym zaborom rosyjskim – różnica jest istotna statystycznie. Z drugiej strony nie odnotowuje istotnej statystycznie różnicy w obecnych osiągnięciach edukacyjnych między byłym zaborom pruskim i rosyjskim. Różne wyniki porównania z zaborom rosyjskim są zastanawiające, gdyż systemy szkolnictwa w zaborach austriackim i pruskim były do siebie bardzo zbliżone. Zdaniem autora różnice mogą wynikać z różnego stosunku żyjących wówczas Polaków do pruskiego i austriackiego systemu edukacyjnego. Prusy prowadziły intensywną politykę germanizacyjną, czego szkoła była ważnym ogniwem, podczas gdy imperium habsburskie zapewniało zamieszkałym na jego obszarze narodom znaczną autonomię. Dzięki temu pozytywny stosunek do edukacji mógł łatwiej wykształcić się w byłym zaborze austriackim i przekazywany z pokolenia na pokolenie może zdaniem autora wciąż wpływać na zachowania współczesnych uczniów i ich rodziców.

Podsumowanie

Powyższy artykuł służył omówieniu metodologii empirycznego badania wpływu historii na współczesność za pomocą metod *quasi*-eksperymentalnych oraz przykładów ich wykorzystania, ze szczególnym uwzględnieniem nieciągłej analizy regresji. Należy podkreślić, że zastosowanie nieciągłej regresji może być nieprawidłowe w sytuacji, gdy analizowane jednostki mogą precyzyjnie sterować wartością zmiennej objaśniającej X , powodując, że rozkład wartości tej zmiennej wokół punktu odcięcia

x_0 będzie nielosowy (będzie zależał od innych charakterystyk tych jednostek). Jeśli jednak, nawet mając pewną kontrolę nad wartością zmiennej X , analizowane jednostki nie mogą precyzyjnie nią manipulować, rozkład zmiennej w otoczeniu prognozy x_0 będzie miał wciąż charakter losowy (Lee, Lemieux 2013). Ta właściwość odróżnia podejście nieciągłej regresji od metody zmiennych instrumentalnych, w której bezwzględnie zakłada się, że instrument jest zmienną egzogeniczną.

Metoda nieciągłej regresji wymaga słabszych założeń niż podejście wykorzystujące zmienne instrumentalne, co jednak ma swoją cenę. Oszacowania w niej uzyskiwane mają bardziej lokalny charakter niż estymacje wykorzystujące zmienne instrumentalne. Jeśli relacja między zmienną zależną a niezależną nie ma charakteru liniowego lub wpływ specjalnego potraktowania na zmienną objaśnianą różni się w pewnym oddaleniu od prognozy x_0 od zależności szacowanej w pobliżu prognozy, wyniki nieciągłej regresji nie będą poprawnym przybliżeniem badanej zależności. Zaletą nieciągłego modelu regresji jest skoncentrowanie się na wąskim paśmie w sąsiedztwie obszaru nieciągłości. Jeżeli przyjęta szerokość tego pasma jest zbyt duża, założenie, że wpływ pozostałych charakterystyk jest stały może nie być spełnione. Rosnąca, zwłaszcza w ostatnich latach, liczba zastosowań analizy nieciągłej regresji, w tym jej wariantu przestrzennego, pokazuje, że zalety tego podejścia przeważają nad wadami.

Tekst wpłynął: 14 listopada 2016 r.
(wersja poprawiona: 6 lutego 2017 r.)

Bibliografia

- Acemoglu D., Johnson S., Robinson J.A., *The Rise of Europe: Atlantic Trade, Institutional Change and Economic Growth*, „The American Economic Review” 2005, nr 95(3).
- Angrist J.D., *Grouped-data Estimation and Testing in Simple Labor-supply Models*, „Journal of Econometrics” 1991, nr 47.
- Angrist J.D., Imbens G., *Two-stage Least Squares Estimates of Average Causal Effects in Models with Variable Treatment Intensity*, „Journal of the American Statistical Association” 1995, nr 90.
- Angrist J.D., Imbens G., Rubin D.B., *Identification of Causal Effects Using Instrumental Variables*, „Journal of the American Statistical Association” 1996, nr 91.
- Angrist J.D., Krueger A., *Instrumental Variables and the Search for Identification: from Supply and Demand to Natural Experiments*, „Journal of Economic Perspectives” 2001, nr 15(4).
- Angrist J.D., Pischke J.S., *Mostly Harmless Econometrics: An Empiricist’s Companion*, Princeton University Press 2009.
- Becker S.O., Boeckh K., Hainz Ch., Woessmann L., *The Empire is Dead, Long Live the Empire! Long-run Persistence of Trust and Corruption in the Bureaucracy*, „The Economic Journal” 2016, nr 126(590).
- Black D.A., Galdo J., Smith J.A., *Evaluating the Worker Profiling and Reemployment Services System Using a Regression Discontinuity Approach*, „American Economic Review” 2007, nr 97(2).

- Broockman D.E., *Do Congressional Candidates Have Reverse Coattails? Evidence from a Regression Discontinuity Design*, „Political Analysis” 2009, nr 17(4).
- Bruhn M., Gallego F.A., *Good, Bad, and Ugly Colonial Activities: Do they Matter for Economic Development?*, „Review of Economics and Statistics” 2012, nr 94 (2).
- Bukowski P., *Identity as a Determinant of Institutional Persistence: Lessons from the Partitions of Poland*, „BEHL Working Paper Series” 2015, WP2015–10.
- Caliendo M., Kopeinig S., *Some Practical Guidance for the Implementation of Propensity Score Matching*, „Journal of Economic Surveys” 2008, nr 22(1).
- Calonico S., Cattaneo M.D., Titiunik R., *rdrobust: An R Package for Robust Nonparametric Inference in Regression-Discontinuity Designs*, „The R Journal” 2015, nr 7/1.
- Caughey D., Jasjeet S.S., *Elections and the Regression Discontinuity Design: Lessons from Close U.S. House Races, 1942–2008*, „Political Analysis” 2011, nr 19(4).
- Choi J., Lee M., *Regression Discontinuity: Review with Extensions*, „Statistical Papers” 2016, doi: 10.1007/s00362–016–0745-z.
- Comin D., Easterly W., Gong E., *Was the Wealth of Nations Determined in 1000 BC?*, „American Economic Journal: Macroeconomics” 2010, nr 2 (3).
- Cook T.D., *Waiting for Life to Arrive: A History of the Regression-discontinuity Design in Psychology, Statistics and Economics*, „Journal of Econometrics” 2008, nr 142(2).
- Dell M., *The Persistent Effects of Peru’s Mining Mita*, „Econometrica” 2010, nr 78 (6).
- Dittmar J.E., *Information Technology and Economic Change: the Impact of the Printing Press*, „The Quarterly Journal of Economics” 2011, nr 126.
- Easterly W., Levine R., *The European Origins of Economic Development*, „DRI Working Paper”, Development Research Institute NYU 2016, nr 104,.
- Egger P.H., Lassmann A., *The Causal Impact of Common Native Language on International Trade: Evidence from a Spatial Regression Discontinuity Design*, „The Economic Journal” 2015, nr 125(584).
- Eggers A.C. i in., *On the Validity of the Regression Discontinuity Design for Estimating Electoral Effects: New Evidence from over 40,000 Close Races*, „American Journal of Political Science” 2014, doi: 10.1111/ajps.12127.
- Feyrer J.D., Sacerdote B., *Colonialism and Modern Income: Islands as Natural Experiments*, „Review of Economics and Statistics” 2009, nr 91 (2).
- Frölich M., *Nonparametric IV Estimation of Local Average Treatment Effects with Covariates*, „Journal of Econometrics” 2007, nr 139(1).
- Frölich M., Lechner M., *Exploiting Regional Treatment Intensity for the Evaluation of Labour Market Policies*, „Journal of the American Statistical Association” 2010, tom 105, nr 491.
- Galor O., *From Stagnation to Growth: Unified Growth Theory*, w: *Handbook of Economic Growth*, P. Aghion, S. Durlauf, Vol 1A, Elsevier, Amsterdam 2005.
- Galor O., *Unified Growth Theory*, Princeton University Press, Princeton 2011.
- Gelman A., Imbens G., *Why High-order Polynomials Should Not Be Used in Regression Discontinuity Designs*, „NBER Working Paper” 2014, nr 20405.
- Gerber A.S., Kessler D.P., Meredith M., *The Persuasive Effects of Direct Mail: A Regression Discontinuity-based Approach*, „Journal of Politics” 2011, nr 73(1).
- Grosfeld I., Zhuravskaya E., *Cultural vs. Economic Legacies of Empires: Evidence from the Partition of Poland*, „Journal of Comparative Economics” 2015, nr 43.
- Hahn J., Todd P., van der Klaauw W., *Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design*, „Econometrica” 2001, nr 69.

- Hidano N., Hoshino T., Sugiura A., *The Effect of Seismic Hazard Risk Information on Property Prices: Evidence from a Spatial Regression Discontinuity Design*, „Regional Science and Urban Economics” 2015, nr 53.
- Hornung E., *Railroads and Micro-regional Growth in Prussia*, „Ifo Working Paper”, Leibniz Institute for Economic Research, University of Munich 2012, nr 127.
- Imbens G.W., Angrist J.D., *Identification and Estimation of Local Average Treatment Effects*, „Econometrica” 1994, nr 62(2).
- Imbens G.W., Kalyanaraman K., *Optimal Bandwidth Choice for the Regression Discontinuity Estimator*, „Review of Economic Studies” 2012, nr 79(3).
- Imbens G.W., Lemieux T., *Regression Discontinuity Designs: a Guide to Practice*, „Journal of Econometrics” 2008, nr 142.
- Imbens G., Zajonc T., *Regression Discontinuity Design with Multiple Forcing Variables*, Technical Report, Harvard University, Department of Economics, Cambridge, Mass. 2011.
- Iyer L., *Direct versus Indirect Colonial Rule in India: Long-term Consequences*, „Review of Economics and Statistics” 2010, nr 92(4).
- Keele L., Titiunik R., *Geographic Boundaries as Regression Discontinuities*, „Political Analysis” 2015, nr 23(1).
- Keele L., Titiunik R., *Natural Experiments Based on Geography*, „Political Science Research and Methods” 2016, nr 4(01).
- King G., Nielsen R., *Why Propensity Scores Should Not Be Used for Matching*, working paper, Harvard 2016, <http://j.mp/1FQhySn>.
- van der Klaauw W., *Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression-Discontinuity Approach*, „International Economic Review” 2002, nr 43(4).
- Lee D.S., Lemieux T., *Regression Discontinuity Designs in Economics*, „Journal of Economic Literature” 2010, nr 48.
- Lee D.S., Lemieux T., *Regression Discontinuity Designs in Social Sciences, w: The SAGE Handbook of Regression Analysis and Casual Inference*, rozdz. 14, SAGE Publications Ltd., London 2015.
- Maksim M., *Wybrane problemy oceny aktywnych programów rynku pracy*, „Acta Universitatis Lodziensis Folia Oeconomica” 2007, nr 213.
- McCrary J., *Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test*, „Journal of Econometrics” 2008, nr 142(2).
- Moffit R., *Program Evaluation with Nonexperimental Data*, „Evaluation Review” 1991, nr 15.
- Nunn N., *Historical Development*, w: *Handbook of Economic Growth*, tom 2, red. P. Agihion, S. Durlauf, North-Holland 2014.
- Otsu T., Xu K., Matsushita Y., *Empirical Likelihood for Regression Discontinuity Design*, „Journal of Econometrics” 2015, nr 186(1).
- Papay J.P., Willett J.B., Murnane R.J., *Extending the Regression-discontinuity Approach to Multiple Assignment Variables*, „Journal of Econometrics” 2011, nr 161(2).
- Rosenbaum P., Rubin D., *The Central Role of the Propensity Score in Observational Studies for Causal Effects*, „Biometrika” 1983, nr 70(1).
- Rubin D., *Estimating Causal Effects of Treatments in Randomized and Non-randomized Studies*, „Journal of Educational Psychology” 1974, nr 66.
- Silverman B., *Density Estimation for Statistics and Data Analysis*, Chapman & Hall, London 1986.
- Smith J., Todd P., *Does Matching Overcome LaLonde’s Critique of Nonexperimental Estimators?*, „Journal of Econometrics” 2005, nr 125(1–2).

- Spolaore E., Wacziarg R., *How Deep are the Roots of Economic Development?*, „Journal of Economic Literature” 2013, nr 51(2).
- Strawiński P., *Propensity Score Matching. Własności małopróbkowe*, Wydawnictwo Uniwersytetu Warszawskiego, Warszawa 2014.
- Thistlethwaite D.L., Campbell D.T., *Regression-discontinuity Analysis: an Alternative to the Ex Post Facto Experiment*, „Journal of Educational Psychology” 1960, nr 51(6).
- Trochim W., *Research Design for Program Evaluation: The Regression-discontinuity Approach*, Sage, Beverly Hills, CA 1984.
- Trochim W., *Regression-discontinuity Design*, w: *International Encyclopedia of the Social and Behavioral Sciences*, red. N.J. Smelser, P.B. Baltes, North-Holland, Amsterdam 2001, tom 19.
- Trzcíński R., *Wykorzystanie techniki propensity score matching w badaniach ewaluacyjnych*, Polska Agencja Rozwoju Przedsiębiorczości 2009.
- Wald A., *The Fitting of Straight Lines if Both Variables are Subject to Error*, „Annals of Mathematical Statistics” 1940, nr 11.
- Wong V.C., Steiner P., Cook T.D., *Analyzing Regression-discontinuity Designs with Multiple Assignment Variables: a Comparative Study of Four Estimation Methods*, „Journal of Educational and Behavioral Statistics” 2013, nr 38(2).
- Wysokińska A., *Invisible Wall. Role of Institutions and Culture in Long Term Development. Natural Experiment*, „CEP Labour Market Workshop Series”, LSE, London 2015.
- Zarycki T., *History and Regional Development: a Controversy over the „Right” Interpretation of the Role of History in the Development of the Polish Regions*, „Geoforum” 2007, nr 38(3).

METODY QUASI-EKSPERYMENTALNE I ICH ZASTOSOWANIE W BADANIU WPLYWU HISTORII NA WSPÓŁCZESNĄ GOSPODARKE

Streszczenie

Celem artykułu jest omówienie sposobów empirycznego weryfikowania wpływu odległej historii na współczesne zróżnicowanie wskaźników społeczno-gospodarczych w różnych regionach i krajach. Metodologia jest w tym przypadku podobna do wykorzystywanej w badaniach ewaluacyjnych, nazywanych też *quasi*-eksperymentalnymi albo naturalnymi eksperymentami. Polegają one na porównaniu pod względem wybranej miary efektywności grupy jednostek objętych analizowanym programem czy interwencją z odpowiednio dobraną grupą kontrolną. Jedynym czynnikiem, który różnicuje obie grupy, jest udział (lub brak udziału) w ocenianym programie. Czynnikiem, który różnicuje dwie porównywane grupy, może być też położenie po jednej lub drugiej stronie historycznej granicy. W artykule omówiono różne metody *quasi*-eksperymentalne, w tym łączenie badanych jednostek w pary, wykorzystanie zmiennych instrumentalnych oraz nieciągłą regresję. Autor ocenia zalety i wady tych metod. Druga część artykułu zawiera przegląd badań empirycznych wykorzystujących omówione metody do badania wpływu historii na współczesne gospodarki.

Słowa kluczowe: eksperymenty historyczne, metody *quasi*-eksperymentalne, nieciągła regresja, analizy przestrzenne

JEL: C21, B23

QUASI-EXPERIMENTAL METHODS AND THEIR USE IN ANALYSING THE INFLUENCE OF HISTORY ON CONTEMPORARY ECONOMIES

Summary

The aim of this article is to discuss the ways of empirical verification of the impact of a distant history on the contemporary differences of socioeconomic indicators in various regions and states. The methods used in such research are similar to those employed in quasi-experimental research, or natural experiments. They consist in comparing the group covered by the analysed program or intervention with properly selected control group. The only factor differentiating both groups is their participation or non-participation in the evaluated program, or their localization on opposite sides of historical borders. The article presents quasi-experimental methods, including the use of instrumental variables, matching similar units, and discontinuous regression. The author is assessing the virtues and drawbacks of these methods. The last part of the article includes a review of empirical studies using the discussed methods in analysing the impact of history on contemporary economies.

Key words: historical experiments, quasi-experimental methods, discontinuous regression, spacial analysis

JEL: C21, B23

КВАЗИЭКСПЕРИМЕНТАЛЬНЫЕ МЕТОДЫ И ИХ ПРИМЕНЕНИЕ В ИССЛЕДОВАНИИ ВЛИЯНИЯ ИСТОРИИ НА СОВРЕМЕННУЮ ЭКОНОМИКУ

Резюме

Целью статьи является рассмотрение способов эмпирической проверки влияния истории на современную дифференциацию социально-экономических показателей в разных регионах и странах. В этом случае методология похожа на ту, которая используется в эволюционных исследованиях, называемых также квазиэкспериментальными или естественными экспериментами. Они состоят в сравнении с разных точек зрения эффективности группы единиц, охваченных данной программой или интервенцией, с контрольной группой, подобранной соответствующим образом. Единственным фактором, который различает обе группы, является участие (или отсутствие участия) в анализируемой программе. Фактором, который различает две сопоставляемые группы, может быть их положение по одну или по другую сторону исторической границы. В статье рассмотрены различные квазиэкспериментальные методы, в том числе объединение исследуемых единиц в пары, использование инструментальных переменных, а также непостоянная регрессия. Автор оценивает достоинства и недостатки этих методов. Вторая часть статьи содержит обзор эмпирических исследований, использующих рассмотренные методы для исследования влияния истории на современную экономику.

Ключевые слова: исторические эксперименты, квазиэкспериментальные методы, непостоянная регрессия, пространственный анализ

JEL: C21, B23