http://dx.doi.org/10.18778/7969-073-2

Agata Żółtaszek

MODELE MIKROSYMULACYJNE

40 LAT
WYDAWNICTWA
UNIWERSYTETU
ŁÓDZKIEGO

Agata Żółtaszek

MODELE MIKROSYMULACYJNE

Teoria i zastosowania ekonomiczno-społeczne

WYDZIAŁ EKONOMICZNO-SOCJOLOGICZNY
UNIWERSYTETU ŁÓDZKIEGO

Łódź 2013

Agata Żółtaszek – Uniwersytet Łódzki, Wydział Ekonomiczno-Socjologiczny
Katedra Ekonometrii Przestrzennej, 90-214 Łódź, ul. Rewolucji 1905 r. nr 39

RECENZENT
Jerzy W. Wiśniewski

SKŁAD KOMPUTEROWY
Monika Wolska

PROJEKT OKŁADKI
Adam Mateusz Suchecki

Wydrukowano z gotowych materiałów dostarczonych do Wydawnictwa UŁ
przez Wydział Ekonomiczno-Socjologiczny

© Copyright by Uniwersytet Łódzki, Łódź 2013

Wydane przez Wydawnictwo Uniwersytetu Łódzkiego
Wydanie I. W.06469.14.0.D

ISBN (wersja drukowana) 978-83-7969-073-2
ISBN (ebook) 978-83-7969-376-4

Wydawnictwo Uniwersytetu Łódzkiego
90-131 Łódź, ul. Lindleya 8
www.wydawnictwo.uni.lodz.pl
e-mail: ksiegarnia@uni.lodz.pl
tel. (42) 665 58 63, faks (42) 665 58 62

SPIS TREŚCI

Wstęp 7

Rozdział 1. Koncepcja mikrosymulacji i modeli mikrosymulacyjnych 11
1.1. Wprowadzenie 11
1.2. Geneza mikrosymulacji 12
1.3. Przegląd wybranej literatury krajowej i zagranicznej 14
1.4. Podstawowe aspekty mikrosymulacji 19
1.4.1. Podstawowe pojęcia i definicje 21
1.4.2. Populacja wstępna i startowa baza mikrodanych 24
1.4.3. Model mikrosymulacyjny 26
1.4.4. Model (mikro)ekonometryczny – submodel mikrosymulacyjny 28
1.4.5. Klasyfikacja modeli mikrosymulacyjnych 29
1.4.6. Walidacja i kalibracja modelu mikrosymulacyjnego 36
1.5. Podsumowanie 39

Rozdział 2. Wybrane ekonomiczne i społeczne modele mikrosymulacyjne 41
2.1. Wprowadzenie 41
2.2. Modele świadczeń emerytalnych i rentowych 42
2.2.1. Model Dynamicznej Symulacji Dochodu (DYNASIM) 43
2.2.2. Mikrosymulacyjny Model Cornell (CORSIM) 44
2.2.3. Model Symulacji Dochodów z Świadczeń Socjalnych i Emerytur (PRISM) 45
2.2.4. Kanadyjski Mikrosymulacyjny Model Analiz Świadczeń Socjalnych (DYNA-CAN) 47
2.2.5. System Modelowania Ubezpieczeń Emerytalnych (PIMS) 48
2.2.6. Porównanie modeli świadczeń emerytalnych i rentowych 51
2.3. Modele podatkowo-zasiłkowe 51
2.3.1. EUROMOD – Podatkowo-zasiłkowy model dla Unii Europejskiej 51
2.3.2. SIMPL – Podatkowo-zasiłkowy model dla Polski 52
2.3.3. Model Ministerstwa Finansów RP 53
2.3.4. Porównanie modeli podatkowo-zasiłkowych 54

2.4. Modele społeczne i demograficzne 54
2.4.1. Symulacja Społeczna (SOCSIM) 56
2.4.2. Ścieżki Życia (LifePaths) 56
2.4.3. Demosim – model symulacji demograficznych 58
2.4.4. Porównanie modeli społecznych i demograficznych 61
2.5. Modele systemu ochrony zdrowia 61
2.5.1. Modele Statistics Canada 61
2.5.2. Modele NATSEM (University of Canberra, Australia) 62
2.5.3. Porównanie modeli systemów ochrony zdrowia 67
2.6. Podsumowanie 67

Rozdział 3. Modele mikroekonometryczne w mikrosymulacjach 69
3.1. Wprowadzenie 69
3.2. Specyfika mikrodanych 71
3.3. Estymacja przekrojowych modeli mikroekonometrycznych 75
3.3.1. Estymatory parametrów strukturalnych jednorównaniowych modeli liniowych danych przekrojowych 76
3.3.2. Estymatory parametrów strukturalnych jednorównaniowych modeli nieliniowych funkcji różniczkowalnych danych przekrojowych 81
3.3.3. Estymatory parametrów strukturalnych jednorównaniowych wybranych modeli nieliniowych funkcji nieróżniczkowalnych 84
3.3.4. Estymatory parametrów strukturalnych modeli wielorównaniowych danych przekrojowych 95
3.4. Specyfika mikrodanych panelowych 97
3.5. Estymacja panelowych modeli mikroekonometrycznych 98
3.5.1. Estymatory parametrów strukturalnych jednorównaniowych modeli liniowych danych panelowych 98
3.5.2. Estymatory parametrów strukturalnych jednorównaniowych modeli nieliniowych danych panelowych 102
3.5.3. Estymatory parametrów strukturalnych modeli wielorównaniowych danych panelowych 103
3.6. Podsumowanie 105

Zakończenie 107
Bibliografia 111

WSTĘP

Dynamiczny rozwój gospodarczy, społeczny, kulturowy i technologiczny ostatnich dekad zainicjował ekspansję wymagań dotyczących jakości życia i poziomu zaspokajania potrzeb. Rosnące oczekiwania społeczeństw co do jakości, szybkości, efektywności i ilości świadczonych usług i konsumowanych dóbr wyznaczają standardy współczesnego życia. W czasach postępującej globalizacji, rozwoju przedsiębiorczości i konkurencyjności wzrost tych wymagań oznacza oczekiwanie bezwarunkowej, systematycznej i trwałej poprawy jakości życia oraz zaspokajania ciągle rosnącego popytu.

Jednocześnie jednak ograniczoność zasobów naturalnych, kapitałowych i ludzkich nakłada nieprzekraczalne restrykcje na możliwości zaspokojenia wszystkich potrzeb. Dlatego szczególnie istotne stało się efektywne wykorzystanie posiadanych środków poprzez ich optymalną alokację. Oznacza to zazwyczaj maksymalizację zysków lub minimalizację ryzyka i strat strony podażowej albo popytowej, przy jednoczesnym zapewnieniu akceptowalnej sytuacji dla drugiej z nich. Niestety ze względu na wielowymiarowość i wieloprzedmiotowość takich problemów są one trudne do rozwiązania. Często niełatwo określić w nich satysfakcjonujący poziom jakości życia czy oferowanych dóbr i usług, a co za tym idzie skonstruować reguły „efektywnego" i „optymalnego" postępowania nie tylko w chwili obecnej, ale także w przyszłości. Dlatego zainicjowano badania ekonomiczne i społeczne, które przyczyniły się do ewolucji interdyscyplinarnych i wielokryterialnych narzędzi prowadzenia analiz wspomagających proces podejmowania decyzji na różnych szczeblach zarządzania. Niezbędne stało się też zastosowanie metod umożliwiających ocenę przyszłych skutków prowadzonych i planowanych polityk.

Najpopularniejszą grupę narzędzi stanowią szeroko rozumiane metody prognozowania i symulacji. Pozwalają one na „przewidywanie" konsekwencji działań i decyzji na poziomie makro, często pomijając strukturę zależności na poziomie mikro. Jednym z nielicznych rozwiązań prowadzenia analiz na wysokim poziomie agregacji na podstawie prognoz i projekcji dla obiektów indywidualnych stały się mikrosymulacje. Ich rozwój rozpoczął się w latach 60. XX w. i był silnie uwarunkowany postępem technologicznym w zakresie mocy obliczeniowych oraz oprogramowania. Postęp w technologii informatycznej i wzrost

zainteresowania decydentów szczegółowymi informacjami o przebiegu procesów ekonomicznych i społecznych przyczyniły się do rozwoju badań wykorzystujących dane indywidualne, a zatem także analiz mikrosymulacyjnych. Połączenie modelowania informacji niezagregowanych i eksperymentów symulacyjnych, przy jednoczesnym wzroście możliwości obliczeniowych i dynamicznym rozwoju oprogramowania, umożliwiło prowadzenie analiz wielowymiarowych rozkładów efektów prowadzonych i planowanych polityk. W rezultacie mikrosymulacje stały się skutecznym narzędziem służącym do podejmowania decyzji nie tylko gospodarczych, ale też społecznych.

Obecnie wykorzystuje się wiele modeli mikrosymulacyjnych, zróżnicowanych tematycznie, wiekowo i strukturalnie. Większość z nich stanowią modele: świadczeń emerytalno-rentowych, zasiłkowo-podatkowe i demograficzno-społeczne. Nowe trendy modeli mikrosymulacji dotyczą nie tylko nowoczesnych technologii obliczeniowych i metod ilościowych, ale także rozwoju dziedzin interdyscyplinarnych.

Modele mikrosymulacyjne są zazwyczaj własnością instytucji rządowych, akademickich i ubezpieczeniowych, dlatego szczegóły dotyczące ich konstrukcji i eksploatacji nie są upubliczniane. Wiele modeli mikrosymulacyjnych zawiera submodele ekonometryczne, jednak zastosowane metody estymacji nie zawsze są znane lub odpowiednie dla specyfiki mikrodanych. Brak informacji o strukturze modeli, w szczególności ekonometrycznych, utrudnia ocenę jakości wyników badań z zastosowaniem tych modeli. Dobór odpowiedniej postaci funkcyjnej modelu i metod estymacji silnie determinuje jakość oszacowań parametrów strukturalnych, a zatem także rezultaty prowadzonych mikrosymulacji. Istniejąca literatura przedmiotu obfituje w opracowania dotyczące aplikacji metodyki mikrosymulacyjnej w zakresie szeroko rozumianej logistyki, ekonomii, socjologii i medycyny. Jednak publikacje te rzadko wspominają o metodologii konstrukcji modeli mikrosymulacyjnych. Również nieliczne są dysertacje teoretyczne przybliżające podstawowe aspekty mikrosymulacji, takie jak definicje, klasyfikacje czy etapy konstrukcji modeli mikrosymulacyjnych. Dlatego w niniejszej pracy podjęto próbę wypełnienia luki w opracowaniach teoretycznych dotyczących metod konstrukcji modeli mikrosymulacyjnych i syntezy wiedzy o eksploatowanych modelach.

W Polsce metody mikrosymulacyjne nie są jeszcze powszechnie znane i stosowane. Istnieją tylko dwa główne modele: SIMPL, submodel EUROMOD, oraz model Ministerstwa Finansów RP, oba o tematyce podatkowo-zasiłkowej. Na świecie jednak modele mikrosymulacyjne, szczególnie w badaniach gospodarczych i socjologicznych, zyskują coraz więcej zwolenników. Równocześnie część badaczy podchodzi sceptycznie do metodologii mikrosymulacyjnej. Ta niejednoznaczność postrzegania mikrosymulacji sprawia, że należy rozważyć: czy modele mikrosymulacyjne stosowane w badaniach ekonomiczno-społecznych mogą być skutecznym narzędziem wspomagającym proces podej-

mowania decyzji, na różnych szczeblach zarządzania? Rozważając niniejszy problem badawczy zdecydowano, że celem głównym jest prezentacja teoretycznych i aplikacyjnych aspektów mikrosymulacji w badaniach ekonomicznych i społecznych, zaś celami szczegółowymi są: (1) synteza dotychczasowych osiągnięć w zakresie genezy, konstrukcji i klasyfikacji modeli mikrosymulacyjnych, (2) przegląd istniejących modeli mikrosymulacyjnych, ze szczególnym uwzględnieniem metodologii mikroekonometrycznej i (3) zaprezentowanie całościowego i syntetycznego zestawienia modeli i metod mikroekonometrycznych, które mogą być przydatne w konstrukcji modelu ekonometrycznego stanowiącego część modelu mikrosymulacyjnego. W procesie realizacji przyjętych celów podjęto próbę weryfikacji (1) przydatności modeli mikrosymulacyjnych do oceny rozkładu efektów prowadzonych polityk ekonomicznych i społecznych oraz (2) wpływu wyboru metody estymacji parametrów strukturalnych submodelu ekonometrycznego na wyniki mikrosymulacji.

W opracowaniu zaprezentowano całościowy zbiór podstawowych pojęć, zintegrowany system klasyfikacji modeli mikrosymulacyjnych oraz etapów przebiegu mikrosymulacji (rozdział 1). Dokonano przeglądu funkcjonujących obecnie ekonomiczno-społecznych modeli mikrosymulacyjnych, ze szczególnym uwzględnieniem zastosowanych w nich metod ekonometrycznych (rozdział 2). Przedstawiono także syntetyczne zestawienie modeli i metod mikroekonometrycznych, które mogą być przydatne w konstrukcji submodelu ekonometrycznego stanowiącego część modelu mikrosymulacyjnego (rozdział 3).

Bardzo nieliczne opracowania literaturowe dotyczą aspektów metodologicznych mikrosymulacji, a zwłaszcza ich podstaw i koncepcji. Dlatego zaproponowano spójny system definicji dla zasadniczych pojęć związanych z mikrosymulacjami, m. in. modelu mikrosymulacyjnego, populacji startowej, bazy mikrodanych i parametrów zewnętrznych. Sprecyzowano różnice pomiędzy modelem mikrosymulacyjnym a modelem mikroekonometrycznym oraz współzależność między mikrosymulacjami a mikroekonometrią. Dokonano także wielokryterialnej klasyfikacji modeli mikrosymulacyjnych, uwzględniając ich cel, zakres i strukturę. Opracowanie teoretyczne zawarte w pierwszym rozdziale stanowi jedną z nielicznych syntetycznych i całościowych prac dotyczących metodologii mikrosymulacji.

W rozdziale 2 przeprowadzono przegląd wybranych ekonomiczno-społecznych modeli mikrosymulacyjnych, w szczególności modeli: świadczeń emerytalnych i rentowych, podatkowo-zasiłkowych, demograficzno-społecznych i z zakresu ekonomii zdrowia. Podjęto próbę konsolidacji informacji o poszczególnych rodzajach modeli ze szczególnym uwzględnieniem aspektów ekonometrycznych.

W ostatnim rozdziale zestawiono informacje o rodzajach modeli mikroekonometrycznych oraz metodach estymacji ich parametrów, wynikających ze specyfiki mikrodanych stosowanych w mikrosymulacjach. Model ekonometryczny

jako submodel mikrosymulacyjny musi, z jednej strony, być kompatybilny z modelem mikrosymulacyjnym, co oznacza zazwyczaj ograniczenie zbioru potencjalnych zmiennych egzogenicznych do łatwo symulowanych czynników społeczno-demograficznych. Z drugiej strony konieczne jest uwzględnienie specyfiki danych indywidualnych, m. in. nieciągłości rozkładu, heteroskedastyczności składnika losowego, nieobjaśnionej niejednorodności jednostek badania. Dlatego specyfikacja i estymacja parametrów submodeli mikroekonometrycznch musi być dostosowana do charakteru badanych zmiennych.

Przeprowadzona analiza dowodzi, że modele mikrosymulacyjne stanowią uniwersalne narzędzie prowadzenia analiz dla makrosystemów poprzez symulację zachowań i zdarzeń dla mikroobiektów. Specyfika ich konstrukcji pozwala na pewną elastyczność w zakresie ich struktury i interdyscyplinarnego potencjału aplikacyjnego. Możliwość dostosowania modeli do celu oraz restrykcji finansowych i technologicznych badania sprawia, że mikrosymulacje zyskują nowych zwolenników. Równocześnie jednak proces upowszechniania mikrosymulacji spowalniany jest ograniczonością syntetycznych i kompleksowych opracowań dotyczących podstaw merytorycznych oraz warunków ich implementacji. Dlatego w niniejszej pracy podjęto próbę przybliżenia aspektów teoretycznych i aplikacyjnych modeli mikrosymulacyjnych stosowanych w badaniach ekonomicznych i społecznych. Opracowanie to może okazać się przydatne w promowaniu metodyki mikrosymulacyjnej w Polsce oraz zachęcić badaczy do konstrukcji spersonalizowanych narzędzi analiz symulacyjnych, jakimi są modele mikrosymulacyjne.

Rozdział 1

KONCEPCJA MIKROSYMULACJI I MODELI MIKROSYMULACYJNYCH

1.1. Wprowadzenie

Dynamiczny rozwój badań ekonomiczno-społecznych końca XX w. i początku XXI w. przyczynił się do ewolucji merytorycznej i metodologicznej prowadzonych analiz. Coraz częściej miały one charakter interdyscyplinarny, asymilując aspekty demografii, medycyny, geografii, transportu, ekologii i zarządzania. Także podejście do prowadzonych badań i uzyskiwanych wyników ulegało modyfikacji metodologicznej. W miejsce wcześniejszych dociekań makroekonomicznych zaczęto stosować mikroanalizy oraz kompilacje metod i rezultatów dla różnych poziomów agregacji.

Ewolucja tematyczna i koncepcyjna w badaniach ekonomiczno-społecznych stała się więc katalizatorem dla rozwoju nowych oraz modernizacji istniejących narzędzi interdyscyplinarnych i wielowymiarowych analiz. Ogromna różnorodność i konkurencyjność dostępnych instrumentów badawczych utrudnia dokonywanie jednoznacznych wyborów metod analitycznych. Selekcja narzędzi ma charakter subiektywny i zależy często od preferencji analityków, których decyzje uwarunkowane są atrakcyjnością oraz prostotą implementacji i uzyskiwanych wyników. Dlatego popularność różnorodnych podejść analitycznych odzwierciedla ich elastyczność wdrożeniową, wymagania sprzętowo-programowe oraz zaplecze teoretyczne.

Wśród szybko rozwijających się narzędzi prowadzenia badań ekonomiczno-społecznych na szczególną uwagę zasługują mikrosymulacje. Umożliwiają one prowadzenie badań na poziomie makro, poprzez symulacje i agregacje wyników dla mikroobiektów, a w konsekwencji uzyskanie wielowymiarowych rozkładów wyników reformowanych polityk i podejmowanych decyzji. Dlatego wiele analiz prowadzonych jest z zastosowaniem tej właśnie metodyki, nie tylko w zakresie ekonomii i socjologii, ale też w medycynie, logistyce oraz takich dziedzinach interdyscyplinarnych jak ekonomia zdrowia. Na popularność mikrosymulacji ogromny wpływ ma także postęp technologiczny. Obecnie modele mikrosymu-

lacyjne można programować w wielu językach programowania, a eksperymenty przeprowadzać często na powszechnie dostępnych komputerach osobistych. Redukuje to znacznie wymagania sprzętowo-programowe mikrosymulacji podnosząc atrakcyjność tej metody. Niestety, równocześnie z dynamicznym rozwojem tematyki badań mikrosymulacyjnych i łatwiejszym dostępem do narzędzi technologicznych, próby zapewnienia zaplecza teoretycznego, szczególnie istotnego dla nowych użytkowników podejmowane są sporadycznie. Nie oznacza to, że teoria mikrosymulacji nie istnieje, jest jednak niespójna i nieusystematyzowana. Wiele podstawowych aspektów, np. etapy konstrukcji modeli mikrosymulacyjnych i ich klasyfikacje są bardzo rozproszone w istniejącej literaturze. Inne elementy jak model, populacja startowa czy kalibracja traktowane są prawie jak pojęcia pierwotne, których nie trzeba definiować. Ponieważ mikrosymulacje to narzędzie wieloelementowe, wieloetapowe i interdyscyplinarne, zrozumienie koncepcji i fundamentalnych elementów metodologii badań mikrosymulacyjnych jest kluczowe nie tylko dla nowicjuszy, ale także dla specjalistów w tej dziedzinie.

Z tego powodu celem rozdziału jest próba skonstruowania spójnego, syntetycznego i całościowego systemu definicji dla podstawowych pojęć mikrostymulacji oraz prezentacja zintegrowanego systemu klasyfikacji i etapów mikrosymulacji. Zaprezentowane w niniejszym opracowaniu aspekty teoretyczne powinny okazać się pomocne przy planowaniu i konstrukcji modeli oraz eksperymentów mikrosymulacyjnych.

1.2. Geneza mikrosymulacji

Mikroekonomia, socjologia i ekonometria dały początek koncepcji mikrosymulacji. Jednak ewolucja mikrosymulacji determinowana była nie tyle postępem w aspektach teoretycznych, co technologicznych. Za początek mikrosymulacji uznaje się rok 1957 r., tzn. wydanie przez Guy'a Orcutta pracy: *A new type of socio-economic system*. Autor zwraca w nim uwagę na ograniczone możliwości predykcyjne modeli ekonomiczno-społecznych opartych na danych zagregowanych. Modele te traktowały marginalnie cechy jednostek decyzyjnych (gospodarstw domowych, przedsiębiorstw). Orcutt podkreślał, że jakość agregacji danych i opartego na niej modelowania bez znajomości podstawowych reguł zachowań tych jednostek jest wątpliwa. Sugerował, że opisywanie, również w sposób zmatematyzowany, aktywności sektora bez zrozumienia czynników reakcji i wyborów poszczególnych elementów na poziomie mikro może prowadzić do błędnych wniosków. Orcutt wysunął więc śmiałą tezę, że współczesne mu modele makroekonomiczne nie pozwalały na przewidywanie rozkładu efektów polityk ze względu na wysoki stopień agregacji przy równoczesnym braku

wystarczającej informacji o powiązaniach na poziomie mikro [Orcutt 1957: 116 i n.]. Wskazał on też, że wyniki makroanaliz często zależą od sposobu grupowania, które można przeprowadzić wykorzystując różne kryteria. Tymczasem podejście mikro ma tę zaletę, że jednostki funkcjonują jako osobne obiekty. Orcutt, będący niejako ojcem mikrosymulacji, sformułował ogólne idee konstrukcji i wykorzystania mikrosymulacji do analizy i porównań efektów alternatywnych scenariuszy polityk, wariantów zdarzeń lub przyjętych parametrów. W 1961 r. Orcutt wraz z Greenbergeriem, Korbelem i Rivlinem dopracowali nowe zasady budowy ekonomiczno-społecznych modeli mikroanalitycznych [Orcutt i in. 1978: 83; Orcutt i in. 1976: 4; Anderson 1997a: 3].

Początkowo mikrosymulacje nazywano symulacjami mikroekonomicznymi lub symulacjami mikroanalitycznymi. Były one postrzegane jako symulacje ekonometryczne lub ekonometryczne analizy scenariuszowe, dokonywane na danych indywidualnych lub o niskim poziomie agregacji, dotyczące procesów gospodarczych. Z czasem jednak spektrum aplikacji tych metod rozszerzyło się na dziedziny z obrzeża ekonomii, jej pokrewne lub całkowicie z nią niezwiązane. W rezultacie nazwa „symulacje mikroekonomiczne" stała się nieadekwatna, a nawet myląca, stąd obecnie powszechnie stosuje się pojęcie „mikrosymulacja", które nie ogranicza i nie określa zakresu zastosowań tych metod.

Następne dwie dekady XX w. sprzyjały intensywnemu rozwojowi tej nowej metodyki, m. in. dzięki dotacjom rządowym, np. w USA. Był to okres tworzenia wielkich modeli (DYNASIM, CORSIM) oraz ich dynamizacji. W latach 60. i początku 70. XX w. powstało kilka modeli mikrosymulacyjnych, głównie na potrzeby analizy sytemu świadczeń socjalnych w USA. Były one tworzone lub współtworzone przez komisje i agencje rządowe, nieliczne zaś samodzielnie przez uniwersytety i ośrodki naukowe. Z czasem aplikacje rozszerzono na inne obszary, np. energetykę czy opiekę medyczną.

Początkowo jednostkami decyzyjnymi były gospodarstwa domowe lub ich członkowie, z czasem zauważono jednak nowe możliwości dla mikrosymulacji tworzonych dla przedsiębiorstw. W ten sposób powstały symulacyjne modele produkcji, które pozwalały na wprowadzanie licznych założeń dotyczących zachowań na rynku. Umożliwiały one analizy koncentracji przemysłu, inflacji cenowej i planowania produkcji.

W drugiej połowie lat 70. dokonano pierwszych prób połączenia modeli skonstruowanych dla analiz gospodarstw domowych z modelami produkcji, których celem było m. in. wyznaczenie cen i płac równowagi [Haveman i Hollenback 1980: XXII i XXIII]. Modele tego okresu były rozbudowane i wymagały obszernych baz danych. Zwiększało to czasochłonność obliczeniową i zmniejszało precyzję wyników. Stało się to podstawą ich krytyki. Sformułowano podstawowe zarzuty pod adresem mikrosymulacji: zbyt duże wymagania technologiczne (zarówno sprzętowe jak i programowe) i ograniczone możliwości aplikacyjne (niedostatek baz danych indywidualnych).

Powyższa krytyka spowodowała, że pod koniec lat 70. i w pierwszej połowie lat 80. zrezygnowano ze skomplikowanych modeli dynamicznych, dostosowując poziom mikrosymulacji do ówczesnego rozwoju technologicznego. Skoncentrowano się na bardzo uproszczonych modelach statycznych, których wymagania sprzętowe, programowe oraz informacyjne były dużo mniejsze, dzięki czemu znacznie redukowały czas obliczeń. Równocześnie ograniczało to możliwość oceny wyników mikrosymulacyjnych analiz ekonomiczno-społecznych tylko do efektów krótkofalowych [Baroni i Richiardi 2007: 5–6].

W tym samym okresie, na przełomie lat 70. i 80., nastąpiła rewolucja w technologii informacyjnej. Rozwój mikroprocesorów przyczynił się do ewolucji komputerowej. Rozpoczęła się era komputerów 4 generacji, tzn. komputerów osobistych (PC) oraz towarzyszącego im oprogramowania, np. MS DOS i Windows. Ich rozpowszechnienie objęło również użytkowników prywatnych, co wraz ze wzrostem mocy obliczeniowej sprzyjało popularyzacji technik prowadzenia badań ekonomicznych i społecznych w oparciu o wyniki modeli mikrosymulacyjnych [Kruk 2009].

Postęp technologiczny oraz łatwiejszy dostęp do danych indywidualnych, doprowadził do redukcji kosztów związanych z tworzeniem i eksploatowaniem modeli mikrosymulacyjnych. W drugiej połowie lat 80. XX w. ponownie zainteresowano się modelami dynamicznymi, oferującymi szerszy wachlarz możliwości analitycznych niż podejście statyczne. Większość modeli była jednak wtedy własnością jednostek akademickich. Powoli zaczęły interesować się nimi także instytucje rządowe i międzynarodowe, głównie w zakresie ubezpieczeń i emerytur. Jednak pomimo ciągłego rozwoju technologicznego, wzrostu możliwości obliczeniowych i oprogramowania, pozostają one nadal narzędziem wymagającym dużych nakładów pracy, czasu i środków finansowych, zarówno na etapie budowy jak i eksploatacji. Wysokie koszty, czasochłonność i pracochłonność konstrukcji i eksploatacji oraz konieczność wykorzystywania dużych baz danych indywidualnych, wciąż trudno dostępnych, sprawia, że większość modeli mikrosymulacyjnych pozostaje w rękach instytucji rządowych, organizacji międzynarodowych i ośrodków akademickich [Baroni i Richiardi 2007: 7–8].

1.3. Przegląd wybranej literatury krajowej i zagranicznej

Istnieje bardzo bogata literatura anglojęzyczna dotycząca problematyki mikrosymulacji. Przeważającą większość stawią artykuły, książki i rozdziały o charakterze aplikacyjnym, opisujące głównie wyniki przeprowadzonych badań. Na szczególną uwagę zasługuje publikacja G. H. Orcutta z 1957 r. *A new type of socio-economic system*, która dała początek mikrosymulacjom oraz dwutomowe opracowanie *Microeconomic simulation model for public*

policy analysis, pod redakcją R. H. Havemana i K. Hollenbecka z 1978 r., ukazujące rosnące zainteresowanie nowatorskimi technikami symulacji mikroekonomicznych w analizach polityk społecznych. W latach 90. XX i w XXI w. W czasopismach specjalistycznych z różnych dziedzin nauki ukazywały się liczne publikacje, ale także dotyczące przeważnie zastosowania modeli mikrosymulacyjnych.

Powszechność zastosowania mikrosymulacji w transporcie i logistyce *sensu largo* spowodowała, że duża część artykułów dotyczyła planowania i oceny ruchu samochodowego na wybranych odcinkach tras, często z uwzględnieniem reakcji na zmieniające się warunki drogowe. Wyniki badań zaprezentowano m. in. w opracowaniach: Pendyala i in. [1997: 183–192] dotyczącym analiz dla Waszyngtonu, D.C w USA; Dupuis i Chopard [2003: 9–21] poświęconym modelowaniu ruchu w Genewie, Szwajcaria; Roorda i in. [2008: 360–375], weryfikującym poprawność symulacji modelu TASHA[1] dla okręgu Toronto w Kanadzie.

Podjęto także pierwsze próby rozszerzenia zastosowań logistycznych w zakresie ruchu pieszych [zob. m. in. Blue i Adler 2001: 293–312] oraz symulacji efektywności planów ewakuacji na wypadek występienia naturalnych katastrof [Chen i in. 2005: 321–338]. Mikrosymulacje zastosowano także w zagadnieniach transportowo-ekologicznych dotyczących emisji zanieczyszczeń w ruchu drogowym, uwzględniając różne ograniczenia prędkości [Int Panis i in. 2006: 270–285] oraz ceny pojazdów o alternatywnych źródłach zasilania [Kazimi 1997: 163–185]. Prowadzone mikrosymulacyjne analizy transportowe mają duże znaczenie dla funkcjonowania ośrodków miejskich i metropolii ze względu na możliwość poprawy płynności ruchu drogowego, redukcję korków i wypadków, a zatem usprawnienie działania gospodarki i polepszenie jakości życia obywateli.

Inne opracowania naukowe wskazują na przydatność mikrosymulacji w analizach medycznych. Większość artykułów dotyczy analizy skutków (m. in. oczekiwanego czasu życia, oczekiwanej długości pobytu w szpitalu, występowania powikłań) przeprowadzonych zabiegów, w szczególności: wszczepienia zastawki serca (z uwzględnieniem rodzaju zastawki, wieku i schorzeń powiązanych [np. Puvimanasinghe i in. 2003: 688–695; Puvimanasinghe i in. 2006: 374–379], próby konstrukcji modelu rezultatów interwencji u chorych na osteoporozę z uwzględnieniem biologicznych uwarunkowań choroby [Vanness i in. 2005: 353–358], przewidywania sytuacji pacjentów przyjmowanych na oddział intensywnej terapii (w 11 wybranych krajach) w zależności od diagnozy wstępnej [Hout i Angus 2004: 2237–2244]. Innym aspektem medycznym badanym w wyniku eksperymentów mikrosymulacyjnych okazało się być rozprzestrzenianie się chorób zakaźnych [zob. Fischer i in. 2008: 185–190] oraz nieza-

[1] Organizator Aktywności Podróży dla Agentów Będących Gospodarstwami Domowymi (ang. *Travel Activity Scheduler for House-hold Agents*).

kaźnych (np. szacowanie oczekiwanej liczby osób niepełnosprawnych [Lymer i in. 2008: 99–116]. Próbne analizy epidemiologiczne przeprowadzono m. in. dla Wysp Brytyjskich [Hoshen i in. 2007: 26–34]. Zaprezentowana literatura wskazuje, że mikrosymulacje medyczne stanowią metodę optymalizacji wyboru procedur medycznych ze względu na ryzyko powikłań lub zapobiegania i zwalczania szybko rozprzestrzeniających się chorób zakaźnych. Ze względu na swój nieinwazyjny charakter, szybkość uzyskiwania wyników i elastyczność prowadzonych eksperymentów, mikrosymulacje dostarczają potrzebnych informacji do kreowania polityki zdrowotnej i ratujących życie procedur medycznych.

Najwięcej jednak publikacji poświęconych jest szeroko rozumianej tematyce ekonomiczno-społecznej oraz dziedzinom pokrewnym. Ze względu na charakter metodologii mikrosymulacyjnej często trudno jest jednoznacznie sklasyfikować prowadzone badania, dlatego najogólniej można określić je jako analizy ekonomiczno-społeczne. Socjologiczno-demograficzno-kulturowe implementacje mikrosymulacji najczęściej poświęcone były długookresowym projekcjom populacji ze szczególnym uwzględnieniem: pokrewieństwa pomiędzy osobami [zob. Wachter i in. 1997: 89–104 – zastosowanie modelu SOCSIM], zmieniających się parametrów demograficznych [np. współczynnika dzietności we Włoszech – Tomassini i Wolf 2000: 353–372] lub innych czynników [np. zachorowania i umieralności w Tajlandii spowodowanej epidemią HIV/AIDS – Wachter i in. 2003: 193–206]. Do nowotarskich aplikacji zaliczyć można próby prowadzenia analiz antagonizmów społeczno-kulturowych, porównujących przyczyny i skutki sytuacji konfliktowych wewnątrz i pomiędzy jednorodnymi grupami etnicznymi, na podstawie badania kontaktów wirtualnych, tzn. rozpowszechnionych w ostatnich latach portali społecznościach [Mueller 2011: 21–34]. Podsumowując, analizy powiązań społecznych i zmian demograficznych prowadzone z zastosowaniem mikrosymulacji, nie tylko pozwalają nakreślić obraz przyszłych populacji, ale przede wszystkim z dużym wyprzedzeniem ostrzegają przed problemami w makroskali, którym można próbować zapobiegać. Starzenie się społeczeństw, alienacja, kryzys rodziny, to kwestie, które poza aspektem społecznym mają czysto pragmatyczne konsekwencje jak redukcja wpływów do budżetu państwa z tytułu podatków dochodowych oraz konieczność zapewnienia opieki osobom starszym i chorym, nieposiadającym żyjących krewnych. Dlatego prowadzenie systematycznych społeczno-demograficzno-kulturowych analiz mikrosymulacyjnych jest kluczowe dla kreowania przyszłych polityk społecznych.

Zastosowania o profilu bezpośrednio lub pośrednio ekonomicznym, z elementami społeczno-demograficznymi, stanowią najbardziej heterogeniczną grupę aplikacji mikrosymulacji w literaturze zagranicznej. Prowadzone w ostatnich latach badania poświęcono przede wszystkim problemom rozkładu dochodów

w społeczeństwie, polityce podatkowo-zasiłkowej oraz szybko rozwijającym się nowym dziedzinom, takim jak ekonomia zdrowia.

Rozkład dochodu badany był jako konsekwencja zmian demograficznych, m. in. współczynnika dzietności w Argentynie [Marchionni i Gasparini 2007: 97–114] oraz reformowanych polityk ekonomiczno-społecznych, ukierunkowanych na walkę z nierównościami ekonomicznymi i ubóstwem. Skuteczność mikrosymulacji w tej dziedzinie potwierdzili Bourguignon i Spadaro [2006: 77–106]. Prowadzone badania dotyczyły wielu działów gospodarki i różnych krajów. W Bangladeszu zanalizowano makroekonomiczne skutki liberalizacji handlu, szczególnie na nierówności dochodowe, z uwzględnieniem krótko- i długofalowych skutków, przy rozróżnieniu sytuacji w regionach wiejskich i miejskich [Raihan 2010: 123–126]. Redukcja podatków pośrednich jako metoda wyrównywania sytuacji ekonomicznej w Brazylii stanowiła podstawę badania de Souza Ferreira Filho i in. [2010: 114–117]. Ahmed i O'Donoghue [2010: 127–129] zweryfikowali wpływ kryzysu światowego na sytuację ekonomiczną w Pakistanie, z uwzględnieniem zmian w średnich dochodach ludności, zaś Grimm [2005: 231–247] zbadał wpływ polityki edukacyjnej na redukcję nierówności ekonomicznych i ubóstwa na Wybrzeżu Kości Słoniowej. Mikrosymulacyjne analizy rozkładu dochodów w społeczeństwie wskazują, że problem nierówności i ubóstwa dotyczy bardzo wielu krajów na świecie, a proponowane reformy ekonomiczno-społeczne nastawione na ich niwelowanie nie zawsze są skuteczne w krótkim lub długim okresie. Eksperymenty mikrosymulacyjne stanowią syntetyczne narzędzie prognozowania i symulacji konsekwencji wprowadzanych i proponowanych reform nie tylko na najbliższe lata, ale także całe dekady.

Analizy reform podatkowych dotyczyły m. in. konsekwencji przeniesienia obciążenia podatkowego z pracy na formy podatków pośrednich [Decoster i in., 2011: 41–56]. Ponadto symulowano efekty wprowadzenia Podwójnego Podatku Dochodowego (ang. *Dual Income Tax*, DIT) w Niemczech [Wagenhals 2011: 3–13] oraz we Włoszech [Balzano i in. 2011: 14–26]. Z drugiej strony zbadano skutki podwyższenia świadczeń socjalnych (Minimalnych Dodatków Dochodowych, ang. *Minimum Insertion Income*) na podaż pracy i rozkład dochodów we Włoszech [Berliri i Parisi 2006: 213–227], oraz długookresowe konsekwencje wprowadzonej w latach 90. reformy planów obowiązkowych emerytur dla pracowników we Włoszech [Sartor 2001: 83–111]. Powyższe publikacje stanowią przykład mikrosymulacyjnych analiz podatkowych służących ocenie istniejących i proponowanych polityk ekonomiczno-społecznych.

Mikrosymulacyjne badania z zakresu ekonomii zdrowia dotyczą skutków reform obowiązkowych i dobrowolnych ubezpieczeń zdrowotnych [np. dla USA Zabinski i in. 1999: 195–218]. Van Sonsbeek i Gradus [2006: 427–456] zanalizowali długookresowe skutki wprowadzonej w 2006 r. w Danii zmiany sposobu przyznawania zasiłków dla osób niepełnosprawnych. Równocześnie mikrosymulacje posłużyły do oceny strony podażowej usług medycznych. Moreno i in.

[1999: 373–393] zastosowali model mikrosymulacyjny do zbadania przebiegu procesów w szpitalach w celu wspomagania podejmowania decyzji menedżerskich. Zaprezentowane artykuły potwierdzają przydatność mikrosymulacji w badaniach z zakresu ekonomii zdrowia.

Również książki opublikowane w ostatnich latach miały charakter aplikacyjny. Zaidi i in. [2009] stanowi zbiór rozdziałów poświęconych wynikom niezależnych badań z zakresu modelowania przestrzennego, podaży pracy, demografii i opieki socjalnej oraz modelowania hybrydowego łączącego analizy mikro- i makroekonomiczne. W opracowaniu Klevmarken i Björn [2008: 31–38] poza prezentacją wyników badań społeczno-demograficznych dla Szwecji zaprezentowano kilkustronicowe wprowadzenie do metodologii mikrosymulacyjnej. Podobnie w innych pracach, m. in. Creedy i in. [2002]; Creedy i Kalb [2006] dokładnie opisane są aspekty ekonomiczno-społeczne prowadzonych badań oraz ich wyniki, jednak teoria mikrosymulacji została jedynie zasygnalizowana.

W Polsce metodologia mikrosymulacyjna jest nowa i stosunkowo mało rozpowszechniona, choć zdobywa coraz więcej zwolenników. Na stronie internetowej http://www.mikrosymulacje.pl/ można znaleźć informacje o prowadzonych mikrosymulacjach transportowych ruchu drogowego, które zyskują na popularności przy prowadzeniu analiz ruchu drogowego w Polsce. CenEA – Centrum Analiz Ekonomicznych [źródło: CenEA] prezentuje krótkie opisy modeli EUROMOD i SIMPL, z uwzględnieniem zarysu charakterystyki mikrosymulacji. SIMPL – podatkowo-zasiłkowy model mikrosymulacyjny dla Polski, poza kilkoma wyjątkami w prezentacjach multimedialnych [zob. *Analiza reform...*, 2010] jest znacznie lepiej opisany w języku angielskim. Dotyczy to zarówno aspektów metodologicznych [Bargain i in. 2007a; Bargain i in. 2007b], jak i aplikacyjnych [Morawski i Myck 2009; Haan i Myck 2009]. Inne, nieliczne opracowania, przede wszystkim dostępne w Internecie, wprowadzają pojęcia mikrosymulacji, modeli mikrosymulacyjnych i ich aplikacji. W szczególności można się zapoznać z zamieszonym na stronach Ministerstwa Finansów RP opisem nowego modelu podatkowo-zasiłkowego: Konopczak i Skibicki [2012]. Publikacja zawiera schemat budowy i charakterystykę modelu oraz krótki zarys historii i idei mikrosymulacji. Do najnowszej polskojęzycznej literatury dotyczącej mikrosymulacji zaliczyć można prace autorki niniejszej monografii. Opracowanie Żółtaszek [2011a] zawiera rys historyczny modelowania mikrosymulacyjnego, propozycję definicji podstawowych pojęć jak mikrosymulacja, populacja startowa i model mikrosymulacyjny, najważniejsze klasyfikacje modeli oraz wykaz wybranych zastosowań w badaniach zagranicznych. W artykułach Żółtaszek [2012] oraz Żółtaszek [2011b] zamieszono przebieg konstrukcji i wyniki badania przeprowadzonego z zastosowaniem próbnego modelu mikrosymulacyjnego, stanowiącego istotny krok w konstrukcji MESMEZ (Mikrosymulacyjnego Ekono-

miczno-Społecznego Modelu Ekonomii Zdrowia), którego zmodyfikowaną wersję FARMMES zaprezentowano w dalszej części rozprawy.

Literatura polskojęzyczna, choć nieliczna w stosunku do osiągnięć zagranicznych, wskazuje na rosnące wśród analityków zainteresowanie metodologią mikrosymulacyjną. Równocześnie popularyzacja ta hamowana jest niedoborem publikacji dotyczących teorii mikrosymulacji i technicznych wskazówek ich wykorzystania. W większości zaprezentowanych prac brak jest szczegółowego opisu procesu konstrukcji modelu mikrosymulacyjnego, stosowanych metod ekonometrycznych oraz przebiegu eksperymentu. Oficjalna dokumentacja modeli pozostających w użytku, jeśli istnieje i została upubliczniona, również zawiera zazwyczaj jedynie podstawowe informacje. W szczególności wzmianki dotyczące estymacji parametrów i weryfikacji submodeli ekonometrycznych wydają się niewystarczające do dokonania wiarygodnej oceny uzyskiwanych w eksperymentach mikrosymulacyjnych wyników.

W 2005 r. założono międzynarodowe stowarzyszenie *The International Microsimulation Association* (IMA), które od 2007 r. wydaje co pół roku, a ostatnio co kwartał, *The International Journal of Microsimulation*, poświęcony aplikacjom mikrosymulacji, głównie z zakresu ekonomii i socjologii oraz nowatorskim rozważaniom teoretycznym. Aspekty techniczne dotyczą jednak zazwyczaj zaawansowanych szczegółów przeznaczonych dla badaczy z wieloletnim doświadczeniem w stosowaniu mikrosymulacji [zob. Zinn 2012: 31–51; Rahman i in. 2010: 3–22; Cumpston 2010: 34–45] istotnych dla rozwoju metodologii, jednak mało pomocnych dla potencjalnych nowych konstruktorów i użytkowników. IMA w ostatnich latach rozpoczęło promocję podejścia „pełnej jawności" konstrukcji modeli mikrosymulacyjnych, w przeciwieństwie do dominującego stanowiska „czarnej/magicznej skrzynki" (ang. *black/magic box*). Na razie nie podjęto prób aktualizacji dokumentacji istniejących modeli, ale coraz więcej badaczy zaczyna doceniać znaczenie systematycznej i spójnej teorii oraz jawnych i wyczerpujących informacji technicznych modeli.

1.4. Podstawowe aspekty mikrosymulacji

Mikrosymulacje stanowią narzędzie analiz i weryfikacji konsekwencji polityk oraz decyzji ekonomiczno-społecznych poprzez agregacje symulowanych rezultatów dla podmiotów indywidualnych. Pozwalają na ocenę korzyści i kosztów reform oraz wskazanie „wygranych" i „przegranych" proponowanych reform. W rezultacie możliwe jest wyznaczanie rozkładu efektów (według różnych kryteriów) konsekwencji alternatywnych propozycji polityk, a zatem bardziej precyzyjne staje się wnioskowanie i podejmowanie decyzji.

Często mikrosymulacje postrzega się błędnie, ze względu na przedrostek, jako narzędzia mikroanaliz, w rzeczywistości stanowią one symulacje makrosys-

temów, opierając się jedynie na danych indywidualnych. W wielu przypadkach liczba czynników i zależności związanych z danym zagadnieniem uniemożliwia zrozumienie lub prognozowanie funkcjonowania całego systemu. Tymczasem w skali mikro zależności mogą okazać się znacznie prostsze do objaśnienia [Spielauer 2009: 3–4].

Mikrosymulacje stanowią przykład analiz scenariuszowych[2] będących metodą heurystyczną[3] wspomagania procesu podejmowania decyzji. Zasadniczym celem wprowadzenia tej grupy metod wspomagania procesów decyzyjnych jest redukcja niepewności i wzrost elastyczności wariantów strategii [Gierszewska i Romanowska 2004]. Podstawową różnicą pomiędzy klasycznie rozumianymi metodami scenariuszowymi a mikrosymulacjami jest to, że celem badań opartych na modelach mikrosymulacyjnych jest analiza porównawcza skutków różnych scenariuszy, a często także wspomaganie wyboru, który scenariusz ma zostać zrealizowany. W przypadku przedsiębiorstw, dla których stworzono podejście scenariuszowe, ich funkcjonowanie ma często charakter gry z naturą[4], gdzie rynek nie reaguje na strategię pojedynczej firmy i nie jest zainteresowany wynikiem starcia [Miszczyńska i Miszczyński 2002: 5–6]. Analizy mikrosymulacyjne wprowadzane są jako narzędzie wspomagania podejmowania decyzji dotyczących reform politycznych, a zatem ich wyniki mają wpływ na „rynek" (podatników, pacjentów, ubezpieczeniodawców, ubezpieczeniobiorców, emerytów, itp.), który jest zainteresowany rezultatem gry. Wynika to m. in. z uwzględnienia w mikrosymulacji charakterystyk nie jednego podmiotu (firmy, gospodarstwa domowego, osoby), ale ich zbioru, a często pełnej populacji generalnej.

Koncepcja, cele i budowa eksperymentu mikrosymulacyjnego nie różnią się od tych stosowanych dla symulacji ogółem. Zazwyczaj jednak pod pojęciem symulacje (ekonometryczne) rozumie się eksperyment na danych makroekonomicznych, tymczasem mikrosymulacje:

– bazują na danych indywidualnych lub o niskim poziomie agregacji,

– mogą dotyczyć dowolnej dziedziny nauki, nie tylko ekonomii (zatem nie muszą być rozumiane jako przypadek symulacji ekonometrycznej, ale raczej statystycznej),

– nie muszą zawierać równań regresji (tzn. modelu ekonometrycznego lub statystycznego – w zależności od tematyki).

---

[2] Metoda scenariuszowa to metoda wspomagająca podejmowanie decyzji, oparta na określeniu najbardziej prawdopodobnych przyszłych sytuacji, stworzeniu dla nich oddzielnych „scenariuszy" (zdarzeń, zachowań, decyzji, itp.) i wyborze tego, który wydaje się najbardziej prawdopodobny lub rzeczywiście zachodzi.

[3] Heurystyka to nauka obejmująca metody podejmowania decyzji oparte na twórczym myśleniu. Znana była już w starożytności, jednak jako nauka rozwinęła się w XX w., szczególnie w drugiej połowie [Antoszkiewicz 1990: 73 i n.; Osborn 1953].

[4] Oznacza to grę ze środowiskiem, w którym istnieje gracz. Często decyzje podejmowane są w warunkach ryzyka (znane są rozkłady prawdopodobieństwa poszczególnych stanów natury) lub niepewności (rozkłady prawdopodobieństwa są nieznane) [Trzaskalik 2003: 217 i n.].

Ponadto symulacje (makroekonomiczne) często mają charakter deterministyczny. Stanowią zespół alternatywnych prognoz, tzw. projekcji, dla odmiennych warunków początkowych i zewnętrznych. Z tego względu stosowane są przede wszystkim w analizach krótkookresowych [Welfe 2003: 295 i n.]. Mikrosymulacje najczęściej uwzględniają cechy losowe o zadanych rozkładach, mają więc charakter stochastyczny. Eksperymenty symulacyjne często są wielokrotnie powtarzane, w rezultacie otrzymuje się nie pojedyncze wyniki dla poszczególnych scenariuszy, ale tyle, ile założono powtórzeń. Otrzymane wartości służą do wyznaczenia miar opisujących rozkład wyników.

W celu prezentacji postaw teoretycznych mikrosymulacji konieczne jest zdefiniowanie podstawowych pojęć, zaprezentowanie specyfiki i konstrukcji baz mikrodanych i modeli mikrosymulacyjnych, wraz z ich klasyfikacją, a także metod ich oceny.

1.4.1. Podstawowe pojęcia i definicje

W dostępnej literaturze dotyczącej mikrosymulacji brak jest precyzyjnych definicji podstawowych pojęć. Przyjęło się rozumieć je intuicyjnie i poprzez analogie do innych dziedzin wiedzy, których aspekty teoretyczne są dobrze znane. Jednak stanowi to jedną z wad tej metodologii, ograniczających jej popularyzację wśród badaczy i analityków. Dlatego podjęto próbę wprowadzania dokładnego i ujednoliconego systemu nazewnictwa i zaproponowano definicje, które stosowane będą w niniejszej pracy.

Najogólniejszą kategorią jest mikrosymulacja, którą poprzez odwołanie się do pojęcia symulacji można określić następująco:

mikrosymulacja *to metoda numeryczna służąca do prowadzenia eksperymentów na modelach mikroekonometrycznych lub matematycznych opartych na mikrodanych, zwanych mikrosymulacyjnymi.*

W takim rozumieniu mikrosymulacja oznacza proces konstrukcji modelu mikrosymulacyjnego, jego ewentualną walidację i kalibrację, wykonanie eksperymentu (badania) mikrosymulacyjnego i uzyskanie wyników na poziomie indywidualnym i zagregowanym. Dwoma najważniejszymi elementami potrzebnymi do przeprowadzenia mikrosymulacji są baza mikrodanych populacji startowej i model mikrosymulacyjny.

Proponuje się, aby populację wstępną rozumieć następująco:

zbiór jednostek (elementów, obiektów, aktorów), indywidualnych lub o niskim poziomie agregacji, poddawanych eksperymentowi mikrosymulacyjnemu i stanowiących podstawę bazy mikrodanych nazywa się ***populacją startową (wstępną)****.*

Populacja wstępna mikrosymulacji składa się z dużej liczby obiektów (kilkaset do kilkuset tysięcy). Najczęściej jednostkami są osoby lub gospodarstwa domowe, rzadziej przedsiębiorstwa, regiony, pojazdy mechaniczne, przedmioty nieożywione. Często populacja startowa jest niejednorodna, tzn. eksperyment przeprowadzany jest dla osób i gospodarstw domowych, osób i przedsiębiorstw czy ubezpieczeniobiorców i ubezpieczeniodawców. Każda grupa homogenicznych jednostek opisana jest oddzielną, choć często powiązaną z innymi, bazą mikrodanych. Bazę taką można spróbować zdefiniować następująco:

baza mikrodanych populacji startowej (startowa baza mikrodanych) *to zbiór wartości cech statystycznych dla homogenicznych jednostek populacji wstępnej, w okresie startowym; każdy obiekt ma unikatowy numer umożliwiający dokonywanie jednoznacznej identyfikacji oraz przypisany sobie zbiór wartości cech.*

Baza mikrodanych to stabelaryzowany zbiór cech opisujących obiekty i ich wartości. Cechy (zmienne) charakteryzujące aktorów rozumie się w kategoriach statystycznych. Ze względu na niski poziom agregacji zmienne często wyrażone są na różnych skalach pomiaru, co wpływa zarówno na strukturę modelu mikrosymulacyjnego i jego parametrów zewnętrznych.

Model mikrosymulacyjny na potrzeby niniejszej pracy zdefiniowano następująco:

model mikrosymulacyjny *to heterogeniczny, interdyscyplinarny model definiujący wyizolowane środowisko w jakim dokonuje się eksperyment mikrosymulacyjny oraz zarządzający zdarzeniami w oparciu o zadane parametry w trakcie jego trwania;*

gdzie:

zbiór informacji opisujących, (1) które obiekty, (2) w jakich okolicznościach i czasie, (3) z jakim prawdopodobieństwem oraz (4) w jaki sposób podlegają zmianie wartości lub wariantów zmiennej nazywany jest ***parametrem modelu mikrosymulacyjnego lub parametrem zewnętrznym****;*

zdarzenie *to proces aktualizacji wartości lub wariantu zmiennej opisującej elementy populacji startowej w oparciu o parametry modelu mikrosymulacyjnego.*

Model ma charakter wieloprzedmiotowy, ponieważ uwzględnia elementy: informatyczne (eksperyment musi być oprogramowany w pakiecie komputerowym, np. MS Excel, STATA, Modgen), demograficzne (parametry modelu odzwierciadlają rozkłady zmiennych w populacji generalnej lub ich aproksymacje) i dziedzin wiedzy związanych z celem badania (często ekonomia i socjologia, ale również medycyna, technika). W przypadku kiedy tematyka modelu dotyczy kilku powiązanych aspektów, często są one grupowane w moduły (bloki) konstruowane osobno:

moduł (blok) *to spójna i tematycznie homogeniczna część modelu mikrosymulacyjnego.*

Celem mikrosymulacji jest przeprowadzenie eksperymentu często utożsamianego z badaniem, tzn.:

eksperyment mikrosymulacyjny (badanie mikrosymulacyjne) *to proces wykonywania wszystkich procedur modelu mikrosymulacyjnego na wartościach i wariantach zmiennych opisujących elementy populacji startowej, aż do uzyskania wartości i wariantów zmiennych dla populacji końcowej*;

gdzie:

populacja końcowa (wynikowa) *to zbiór wszystkich jednostek jakie otrzymywane są po zakończeniu eksperymentu mikrosymulacyjnego*.

Teoretycznie populacja startowa i jej baza mikrodanych są niezależne od modelu mikrosymulacyjnego. Oznacza to, że w eksperymentach można zastosować model dla różnych populacji wstępnych i odwrotnie: na danej populacji startowej możliwe jest przeprowadzenie badań w oparciu o różne modele mikrosymulacyjne. W praktyce ograniczony dostęp do odpowiednio licznych danych indywidualnych sprawia, że jedna baza danych (np. ze spisu powszechnego) stanowi podstawę kilku populacji startowych, choć struktury baz mikrodanych mogą być różne, ze względu na konieczność dostosowania ich do wymagań modeli mikrosymulacyjnych i celów badań. Rzadko zdarza się, by jeden model mikrosymulacyjny uruchamiano dla różnych baz danych. Z reguły baza danych jest zintegrowana z modelem na poziomie programowania komputerowego. Wyjątek stanowi aktualizacja populacji wstępnej przez programistów i twórców modelu. Zazwyczaj oznacza to wymianę starszej bazy na analogiczną, nowszą, np. pochodzącą z innego spisu powszechnego lub nowszej fali ankiety.

Oceny jakości modelu mikrosymulacyjnego dokonuje się poprzez jego walidację:

w modelach mikrosymulacyjnych proces oceny jakości dopasowania otrzymywanych wyników do danych rzeczywistych, zgodności z założeniami, teorią i logiką nazywany jest ***walidacją***.

Ze względu na stochastyczny charakter eksperymentu oraz brak przyszłych informacji dla elementów populacji startowej trudno jest dokonać pełnej oceny modelu mikrosymulacyjnego.

W przypadku niektórych modeli przeprowadza się korektę wyników mającą przybliżyć je do empirycznych wartości dla wybranego okresu. Jest to tzw. kalibracja, gdzie:

kalibracja *(ang. calibration) lub* ***wyrównanie*** *(ang. alignment) oznacza wprowadzenie arbitralnej korekty wartości zmiennych i/lub parametrów w okresie startowym i/lub końcowym, w celu zapewnienia lepszego dopasowania modelu mikrosymulacyjnego do danych rzeczywistych, założeń, teorii i logiki.*

Zaprezentowany system definicji pozwala zrozumieć istotę poszczególnych aspektów mikrosymulacji, kluczowych dla zrozumienia koncepcji i pomocnych

w konstrukcji modeli mikrosymulacyjnych. Przebieg mikrosymulacji i jakość jej wyników są silnie determinowane własnościami jej dwóch najważniejszych elementów: bazy mikrodanych populacji startowej i modelu mikrosymulacyjnego. Specyfika bazy startowej i modelu jest kluczowa, ponieważ znajduje odzwierciadlenie w strukturze eksperymentu, doborze parametrów zewnętrznych i zastosowanej technologii informatycznej.

1.4.2. Populacja wstępna i startowa baza mikrodanych

Populacja startowa może pochodzić z badania pełnego lub częściowego, źródeł pierwotnych lub wtórnych. Czasami, jeśli istnieją różne bazy danych tej samej grupy obiektów, możliwe jest ich zintegrowanie (np. poprzez agregację informacji dla osób do poziomu ich gospodarstw domowych) do jednego zbioru mikrodanych. Populacja startowa nie musi być jednorodna, w takim przypadku niezbędne jest jednak skonstruowanie oddzielnych baz mikrodanych dla elementów homogenicznych grup populacji wstępnej.

W bazie mikrodanych każdy obiekt musi posiadać jednoznaczny identyfikator (np. numer porządkowy). Wszystkie elementy populacji wstępnej, gdy jest ona jednorodna lub jej homogenicznego podzbioru, gdy nie jest, opisywane są zbiorem cech statystycznych, wyrażonych na dowolnej skali pomiarowej. Struktura zmiennych jest bardzo istotna, ponieważ należy do niej dostosować konstrukcję modelu mikrosymulacyjnego.

Ze względu na sposób aktualizacji zmiennych w eksperymencie mikrosymulacyjnym można sklasyfikować zmienne jako cechy (1) stałe w całym okresie badania (np. płeć), (2) zmieniające się według znanego lub przybliżonego rozkładu prawdopodobieństwa (np. umieralność) oraz (3) zmieniające się losowo, ale o nieznanym rozkładzie lub nieznanych parametrach statystycznych rozkładu (np. dochód gospodarstwa domowego).

Rozważając jakość wyników eksperymentu mikrosymulacyjnego, należy postrzegać populację startową (zbiór wszystkich elementów poddanych mikrosymulacji) w kontekście populacji generalnej (zbiór wszystkich elementów stanowiących obiekt zainteresowania badania mikrosymulacyjnego). W wielu przypadkach, szczególnie modeli mikrosymulacyjnych stanowiących własność rządową lub urzędów statystycznych, eksperymenty mikrosymulacyjne są badaniami pełnymi. Baza mikrodanych pochodzi zazwyczaj ze spisu powszechnego, czasami sprzed kilku dekad.

W sytuacji kiedy dane pełne nie istnieją lub nie są dostępne, populacja startowa stanowi próbę pochodzącą ze zbiorowości generalnej. Niezbędne jest skorzystanie z badań częściowych, czasami zaprojektowanych dla potrzeb eksperymentu mikrosymulacyjnego. Należy jednak pamiętać, że podnosi to znacząco koszty ich prowadzenia, stąd preferowane są źródła wtórne. W celu zapewnienia

najwyższej jakości wnioskowania o zbiorowości generalnej z częściowego badania mikrosymulacyjnego, najczęściej stosowane są próby reprezentatywne. Zapewnieniu zgodności struktury populacji wstępnej pochodzącej z próby reprezentatywnej ze zbiorowością generalną służą wagi indywidulne nadawane poszczególnym elementom. Często stanowią one integralną część bazy danych (szczególnie panelowych), wyznaczoną w procesie zbieranie informacji [Spadaro 2007: 20]. Wagi nie mają wpływu na eksperyment mikrosymulacyjny. Wykorzystywane są w procesie agregacji wyników i wyznaczania statystyk ich rozkładów (np. średnich). W innym przypadku, gdy próba ma charakter celowy[5], uzyskane rezultaty stanowią opis jedynie elementów populacji wstępnej, czyli próby. Ogranicza to znacząco jakość i efektywność modelu oraz eksperymentu mikrosymulacyjnego.

Specyficznym rodzajem populacji wstępnej jest populacja syntetyczna, zawsze utożsamiana ze zbiorowością generalną. Wszystkie obiekty wraz z opisującymi je zmiennymi są generowane sztucznie. Zaletą takiego podejścia jest uniezależnienie się od pytań zawartych w ankietach. Podstawowa wada to obniżenie realizmu symulowanych elementów i ich cech.

Należy także zauważyć, że jeżeli w modelu mikrosymulacyjnym zawarty jest submodel mikroekonometryczny, to zazwyczaj próba, przekrojowa lub panelowa, dla której oszacowano parametry strukturalne, jest niezależna od mikrosymulacyjnej bazy danych. Nie ma reguł określających sposób konstrukcji i estymacji parametrów takiego modelu ekonometrycznego w odniesieniu do startowej bazy mikrodanych.

Z populacją startową ściśle powiązana jest populacja końcowa. W zależności od zaplanowanych i zaprogramowanych w modelu mikrosymulacyjnym procesów, konkretne obiekty mogą opuścić populację startową (umrzeć, wyemigrować, zbankrutować itp.) i/lub nowe mogą zasilić liczebność populacji wstępnej (urodzenia, rejestracja nowych podmiotów gospodarczych itp.). Nie istnieje jednoznaczna zależność między tymi populacjami. W każdym przypadku (oddzielnym eksperymencie czy nawet powtórzeniu) populacja końcowa zależy od tego:

– jakie zmiany liczebności populacji, *in plus* i *in minus* przewidziano w modelu,

– czy procesy mają charakter stochastyczny – losowość sprawia, że powtórzenia w eksperymencie mogą generować różne populacje wynikowe.

Jeżeli nie zaprojektuje się eksperymentu tak, by zapewnić tożsamość (co do składu jednostek, a nie cech je opisujących) populacji startowej i końcowej, to, pomimo że są one powiązane, niemożliwe jest przewidzenie różnic między nimi.

[5] „Dobór celowy oznacza, że o wyborze jednostek decyduje badacz, opierając się na merytorycznej znajomości problematyki badawczej" – cyt. Zeliaś [2000: 15].

1.4.3. Model mikrosymulacyjny

Konstrukcja modelu jest silnie determinowana celami badania, strukturą startowej bazy mikrodanych oraz ograniczeniami sprzętowymi i programowymi. Jednak każdy model musi realizować 5 podstawowych funkcji: zarządzanie czasem, zarządzenie zdarzeniami, integracja obiektów populacji startowej, integracja funkcjonowania modułów i prezentacja wyników.

Zarządzanie czasem

Mikrosymulacja to symulacja obiektów populacji startowej w zadanym przedziale czasu. Konieczne staje się zapętlenie funkcjonowania procedur modelu, zazwyczaj powtarzających się cyklicznie lub losowo, aby umożliwić „starzenie się" populacji. W modelu musi być określony rozkład czasu jako dyskretny (zmiany następują w stałych interwałach) lub ciągły (zmiany zachodzą w dowolnych momentach). Ponadto konieczne jest zdefiniowanie okresu początkowego, zazwyczaj utożsamianego z rokiem, dla którego zgromadzono startową bazę mikrodanych, choć zdarza się, że populacja startowa jest najpierw „odmładzana" (okres startowy wcześniejszy niż okres, dla którego zgromadzono dane) lub „postarzana" (okres startowy późniejszy niż okres, dla którego zgromadzono dane). Populacja wstępna może być również zbiorem obiektów syntetycznych (wygenerowanym sztucznie) i okres startowy przyjmowany jest arbitralnie. Analogicznie należy określić okres końcowy, czyli zakończenie eksperymentu. Z reguły kryterium stopu jest upływ określonego czasu, rzadziej jest to konkretne zdarzenie.

Zarządzanie zdarzeniami

Ponieważ w procesie mikrosymulacji jednostki populacji wstępnej podlegają zmianom w czasie poprzez aktualizację wartości lub wariantu zmiennych, niezbędne jest kontrolowanie wszystkich aspektów tych zdarzeń. Często „zdarzenie" opiera się na zadeklarowanych w modelu parametrach zewnętrznych. Najważniejszym kryterium dywersyfikującym aktualizacje w modelu mikrosymulacyjnym jest ich losowość. Cechy mogą być **deterministyczne** (zmieniają się cyklicznie lub warunkowo) lub **stochastyczne** (dla określonego rozkładu prawdopodobieństwa poszczególnych wartości lub wariantów). Oznacza to także zawiadowanie występowaniem zdarzeń tzn. częstością występowania oraz ich kolejnością (tzw. kolejka zdarzeń)[6], które wpływają na wyniki eksperymentu. Zdarzenia są determinowane parametrami modelu mikrosymulacyjnego. Parametry są elementami całkowicie egzogenicznymi dla modelu mikrosymulacyj-

[6] Porządkowanie zdarzeń związane jest nie tylko z logiką, ale także wpływa na wyniki eksperymentu.

nego, odzwierciadlają czynniki środowiskowe, np. demograficzne, makroekonomiczne. Często są one warunkowe względem pewnych zmiennych, np. wieku, płci. Model określa, jakie zmienne, dla których obiektów i kiedy, zostaną aktualizowane oraz nową wartość cechy. Zdarzenia, zachodzące według zadanych parametrów mogą być:

– **powszechne** (zachodzić dla wszystkich elementów badania) lub **warunkowe** (zachodzić dla wybranych elementów, np. tylko dla kobiet),

– **jedno-, dwu- lub wielowariantowe** (w zależności od liczby możliwych realizacji cechy),

– **jednorazowe** (mogą nastąpić jeden raz w całym eksperymencie, np. śmierć) lub **powtarzalne** (mogą zajść więcej niż raz, zazwyczaj zdefiniowane są cyklicznie).

Wybór parametrów, ich wartości i charakterystyk wpływa na wyniki eksperymentu mikrosymulacyjnego.

Integracja obiektów populacji startowej

Jeżeli model mikrosymulacyjny uwzględnia więcej niż jeden rodzaj obiektów, np. jednostki odpowiadające za popyt i podaż na rynku, osoby i gospodarstwa domowe, oddziały i firmy, to model mikrosymulacyjny musi być zaprogramowany do kontroli interakcji pomiędzy nimi. Czasami oznacza to konieczność agregacji wartości lub koligacji elementów. Takie sterowanie może mieć charakter deterministyczny (oddziały są jednoznacznie przypisane do firm) lub stochastyczny (klient może pójść do wielu punktów podaży, ale z pewnym zadanym prawdopodobieństwem). Często także obiekty jednorodne są ze sobą powiązane. Może występować element dominujący, np. głowa rodziny, matka, któremu podległe są inne jednostki (małżonek, dzieci). W takim wypadku model, zarządzając czasem i zdarzeniami, musi uwzględniać takie interakcje między aktorami.

Integracja funkcjonowania modułów

Heterogeniczny tematycznie i interdyscyplinarny charakter modelu mikrosymulacyjnego odzwierciedlony jest w jego strukturze, często poprzez wydzielenie modułów (bloków). Należy zapewnić możliwość funkcjonowania całego systemu oraz poszczególnych bloków. Kolejność wykonywania procedur zgrupowanych w poszczególnych modułach i interakcje pomiędzy nimi mają istotne znaczenie dla zapewnienia poprawnego procesu symulacji.

Prezentacja wyników

Obiekty z populacji startowej są „przeniesione” do świata opisanego modelem mikrosymulacyjnym. Pozwala to symulować skutki wprowadzanych do

modelu zmian, reform i zaburzeń oraz porównywać wyniki alternatywnych rozwiązań. Rezultaty pośrednie i końcowe wyznaczane są dla jednostek populacji (startowej, końcowej lub pośredniej). Wyniki na poziomie mikrodanych nie stanowią efektywnego narzędzia analiz porównawczych, gdyż można na ich podstawie wnioskować jedynie o różnicach między poszczególnymi osobami, gospodarstwami domowymi lub firmami, heterogenicznymi ze względu na wiele cech statystycznych. Dlatego wyniki indywidualne agregowane są według wybranego kryterium lub kryteriów i porównywane pomiędzy kategoriami, np. wyznaczany jest średni zysk firm według: województw w celu przeprowadzenia analiz regionalnych, wielkości przedsiębiorstwa (liczby zatrudnionych), kraju pochodzenia kapitału (krajowy i zagraniczny) itp. W przypadku populacji startowej pochodzącej z próby o elementach ważonych, agregacja wyników musi uwzględnić wagi, w celu zapewnienia ich poprawności.

1.4.4. Model (mikro)ekonometryczny – submodel mikrosymulacyjny

Zasadniczą funkcją modelu mikrosymulacyjnego jest zarządzanie zdarzeniami, które do eksperymentu wprowadzają zmiany i aktualizacje zmiennych opisujących obiekty lub warunków zewnętrznych. W zależności od charakteru zmiennych, można wyróżnić następujące cechy badanych jednostek:

1) stałe w całym okresie badania (nie muszą być aktualizowane),
2) zmieniające się według znanego lub przybliżonego rozkładu prawdopodobieństwa,
3) zmieniające się losowo, ale o nieznanym rozkładzie lub nieznanych parametrach statystycznych rozkładu.

Cechy (2) uwzględniają zmienne oprogramowane w modelu mikrosymulacyjnym zdarzeniami deterministycznymi (o prawdopodobieństwie równym 1) lub stochastyczne, dla których możliwe jest uwzględnienie parametrów prawdopodobieństwa (bezwarunkowych – jedna wartość lub warunkowanych względem innych cech – dwu- i wielowymiarowa tablica parametrów) występowania poszczególnych wartości i wariantów. Zazwyczaj parametry te odzwierciadlają częstość występowania poszczególnych wartości w zbiorowości generalnej lub w próbie reprezentatywnej. Zmienne (3) nie mogą być jednak aktualizowane na podstawie parametrów mikrosymulacji, gdyż są one nieznane. Rozwiązaniem problemu aktualizacji zmiennych o nieznanych rozkładach jest najczęściej wprowadzenie do modelu mikrosymulacyjnego metodologii szacowania wartości zmiennych z zastosowaniem równania regresji mającego charakter przyczynowo-skutkowy lub indykatorowy. Oznacza to zastosowanie modelu ekonometrycznego[7], a dokładnie mikroekonometrycznego. W ostatnich latach nastąpił

[7] W przypadku kiedy model regresji nie dotyczy tematyki ekonomicznej, powinien być nazywany modelem statystycznym. Ponieważ jednak (1) modele mikrosymulacyjne bardzo często mają

dynamiczny rozwój metod estymacji parametrów modeli mikroekonometrycznych dla danych przekrojowych i panelowych o różnorodnych rozkładach zmiennych endogenicznych. Mimo to konstruktorzy modeli mikrosymulacyjnych nadal zbyt często marginalnie traktują problem doboru postaci funkcyjnej i adekwatnego, ze względu na własności estymatora, sposobu szacowania parametrów strukturalnych modeli mikroekonometrycznych. Stosują metody makroekonometryczne, nie weryfikując czy odpowiadają one specyfice danych indywidualnych. W ich opracowaniach nie są prezentowane szczegóły dotyczące sposobu konstrukcji, estymacji i weryfikacji równań regresji, co uniemożliwia ocenę jakości wyników modelu mikroekonometrycznego, a zatem także mikrosymulacyjnego. Jest to efekt polityki „czarnej skrzynki" zapoczątkowanej w okresie rozwoju mikrosymulacji przez agencje rządowe (lata 60. –80. XX w.).

Połączenie modelu mikrosymulacyjnego i mikroekonometrycznego zalecane jest w szczególności, gdy:

– **celem eksperymentu mikrosymulacyjnego** jest projekcja zmiennych o nieznanym rozkładzie prawdopodobieństwa lub jego parametrach,

– **celem badania (mikro)ekonometrycznego** jest prognozowanie lub symulowanie wartości zmiennych endogenicznych na podstawie modelu przekrojowego lub panelowego w przypadku, jeżeli wartości zmiennych egzogenicznych są trudne do przewidzenia w przyszłości.

1.4.5. Klasyfikacja modeli mikrosymulacyjnych

Każdy model mikrosymulacyjny determinowany jest **celami**, **zakresem** i **metodami** badania. W rezultacie istnieje wiele kryteriów klasyfikacji modeli mikrosymulacyjnych (zob. rys. 1.1).

Klasyfikacja modeli mikrosymulacyjnych według celu

Modele mikrosymulacyjne projektowane są w celu **objaśniania** rzeczywistości lub jej **przewidywania**.

Chcąc zrozumieć jak funkcjonuje dany system, **objaśniający model mikrosymulacyjny** musi być silnie oparty na teorii danej dziedziny. Często poza prostymi zasadami funkcjonowania obiekty potrafią „uczyć się" na doświadczeniach. Może to wiązać się z koniecznością generowania syntetycznych populacji, posiadających indywidualne historie. Takie podejście jest przydatne w kreo-

---

charakter ekonomiczno-społeczny, (2) zakres metod estymacji i weryfikacji modelu nie jest zależny od tematyki oraz (3) większość metodologii mikroanaliz została stworzona dla potrzeb ekonometrycznych, nie będzie wprowadzone rozróżnienie pomiędzy modelem ekonometrycznym a statystycznym. Metody i modele będą nazywane ekonometrycznymi lub mikroekonometrycznymi.

waniu teorii funkcjonowania systemów według zasady *bottom-up*, czyli zrozumienia interakcji na poziomie mikro w celu ich uogólnienia na poziom makro. Obecnie są stosunkowo rzadko stosowane.

CEL	Cel modelu	• model objaśniający • model przewidywania
ZAKRES	Populacja startowa	• model populacji • model próby
	Selekcja z populacji startowej	• model (pełnej) populacji (startowej) • model kohortny
	Specjalizacja	• model ogólny • model wyspecjalizowany
STRUKTURA	Okres pomiaru efektów	• model statyczny • model dynamiczny
	Rozkład czasu	• model w czasie dyskretnym • model w czasie ciągłym
	Spektrum badania	• model przypadku • model czasu
	Izolacja populacji startowej	• model otwarty • model zamknięty
	Pochodzenie populacji startowej	• model o populacji przekrojowej • model o populacji syntetycznej
	Układ modułowy	• model jednomodułowy • model wielomodułowy
	„Czystości”	• model mikrosymulacyjny *sensu stricte* • model mieszany (hybrydowy)

Rys. 1.1. Klasyfikacja modeli mikrosymulacyjnych

Ź r ó d ł o: opracowanie własne na podstawie dostępnej literatury.

W celu przewidywania skutków wprowadzania nowych lub zreformowanych polityk do systemu nie zawsze konieczna jest dogłębna znajomość teorii. Istotniejsze jest badanie zależności empirycznych, do czego potrzebne są dane rzeczywiste, nie zaś syntetyczne. Od prognozy (gdy poszukuje się skutków naj-

bardziej prawdopodobnego scenariusza) lub projekcji (gdy scenariuszy jest więcej) oczekuje się dokładności, nie zaś objaśniania aspektów teoretycznych. **Mikrosymulacyjne modele przewidywania** służące predykcji konsekwencji są coraz bardziej popularne [Spielauer 2009: 1–4].

Klasyfikacja modeli mikrosymulacyjnych według zakresu

a) Populacja startowa mikrosymulacji

Populacja wstępna może stanowić pełną zbiorowość generalną lub jedynie jej próbę, wyróżniamy zatem model **populacji** i **próby**. Podział ten ma znaczenie dla: agregacji wyników – ponieważ w przypadku próby niezbędne jest uwzględnienie wag poszczególnych elementów – oraz wnioskowania z badania – w modelu populacji agregacja pozwala na wnioskowanie, w modelu próby musi ona być reprezentatywna, by móc uogólnić rezultat na całą zbiorowość generalną.

b) Selekcja z populacji startowej

Mając populację wstępną, niezależnie od tego czy jest tożsama z pełną zbiorowością generalną, czy pochodzi z próby, możliwe jest przeprowadzenie symulacji na wszystkich jednostkach z zastosowaniem **modelu (pełnej) populacji (startowej)** lub wyselekcjonowanie z niej elementów według pewnego kryterium w **modelu kohorty**. Dla **modeli populacji** generowana jest „przyszłość" wszystkich elementów populacji. Populacja końcowa nie musi być zgodna co do liczby czy charakterystyki z populacją startową, ale jest jej bezpośrednim produktem. **Model kohorty** wybiera do symulacji jedynie pewne jednostki odmienne od pozostałych elementów populacji. Najczęściej kryterium dywersyfikacji jest wiek lub grupa przynależności (np. etnicznej). Cała populacja zostaje podzielona na jednorodne pod względem wybranej cechy grupy, tzn. kohorty, które stanowią podstawę dalszych analiz mikrosymulacyjnych.

c) Specjalizacja modelu mikrosymulacyjnego

Chcąc odwzorować fragment rzeczywistości, można wybrać **model ogólny** lub **wyspecjalizowany**. **Model ogólny** zakłada, że poza elementami związanymi z bezpośrednimi celami eksperymentu niezbędne staje się uwzględnienie aspektów pośrednio powiązanych, ale zapewniających realizm modelowanemu systemowi. Często ogólny charakter modelu wymusza jego wieloblokowość. Umożliwia to także rozszerzenie początkowych celów badań, stąd staje się bardziej uniwersalny. **Modele** małe mogą być **wyspecjalizowane**, szczególnie jeśli mają służyć jednorazowym i krótkotrwałym analizom. Mogą one być jednoblokowe i/lub skupiać się na wybranych kohortach. Często modele wyspecjalizowane rozrastają się jednak do modeli ogólnych poprzez ich rozbudowę tematyczną i strukturalną [Spielauer 2009: 4–6].

Klasyfikacja modeli mikrosymulacyjnych według ich struktury

a) Okres pomiaru efektów mikrosymulacji [Baroni i Richiardi 2007: 16–24]

Modele mikrosymulacyjne podzielić można ze względu na **okres pomiaru efektów na modele statyczne i dynamiczne.** Klasyfikacja jest podobna do podziału modeli i symulacji ekonometrycznych. Jednak nieco odmienna charakterystyka i konsekwencje tego podziału w mikroskali bardzo silnie determinują możliwości ich wykorzystania, zarówno w aspekcie aplikacji jak i celów analiz scenariuszowych.

Model statyczny nie uwzględnia istnienia opóźnionych zmiennych endogenicznych lub trendu, a wartości zmiennych zależnych ustalane są za każdym razem na poziomie rzeczywistej realizacji. Ogranicza to znacząco horyzont czasowy wyników mikrosymulacji. Model pozwala na ocenę jedynie natychmiastowych efektów analizowanych polityk. Niemożliwe jest uwzględnienie skutków w dłuższej perspektywie, tzn. późniejszych i/lub opóźnionych reakcji badanych jednostek. Ponadto utrudnia to porównywanie scenariuszy, przy zachodzących zmianach parametrów systemu, gdyż w takiej sytuacji wyniki modyfikacji polityk gospodarczych rozłożone są w czasie dłuższym niż spektrum zalecane dla mikrosymulacji statycznej. Prognozy i projekcje statycznej mikrosymulacji uwzględniają wyłącznie procesy tzw. *statycznego starzenia populacji startowej*. Oznacza to, że zmiany zachodzące w strukturze populacji wstępnej oraz aktualizacje parametrów, niezbędne do stworzenia prognozy, mają charakter deterministyczny. Mimo licznych ograniczeń aplikacyjnych modeli statycznych są one popularne wśród analityków. Ich prostota kalkulacyjna znajduje odzwierciedlenie w stosunkowo niedużych nakładach czasu, pracy i kapitału, niezbędnych do prowadzenia statycznych symulacji mikroekonomicznych.

Modele dynamiczne uwzględniają opóźnienia zmiennej endogenicznej lub trend. Ponadto wprowadzają „endogenizację" zmiennych egzogenicznych (tzn. nie związanych bezpośrednio z celem badania) elementów populacji. Jedynie parametry opisujące system pozostają całkowicie egzogeniczne względem modelu[8]. Mikrosymulacja dynamiczna jest znacznie bardziej skomplikowana niż statyczna. Dla każdej z jednostek populacji startowej, w procesie tzw. *dynamicznego starzenia*, symulowana jest trajektoria dla całego okresu symulacji. Możliwe jest także włączanie do analizy nowych obiektów, z wygenerowaną historią i wartościami cech, oraz wyłączanie (eliminowanie) tych, które umarły, emigrowały, zbankrutowały itp. Wszystkie elementy populacji, rzeczywiste i sztuczne, doświadczają zaprogramowanych procesów, które mogą mieć charakter deterministyczny lub stochastyczny. Podstawowymi zaletami dynamicznych modeli mikrosymulacyjnych są: (1) symulowanie historii elementów populacji, niezbędne do dokonywania analiz, (2) symulowanie reakcji elementów populacji na wprowadzane do systemu zmiany polityk oraz, co jest charaktery-

[8] Wprowadzenie endogenizacji parametrów mających charakter demograficzny lub makroekonomiczny oznaczałoby, że model nie jest mikrosymulacyjny *sensu stricte*, ale hybrydowy.

styczne także dla niektórych symulacji statycznych, (3) symulowanie decyzji elementów populacji w oparciu o dane indywidualne. Podstawowymi zarzutami przeciwko mikrosymulacjom dynamicznym są: (a) niedoskonałość tych modeli wynikająca z niepełnej wiedzy dotyczącej zachowani ludzkich (np. społecznych, ekonomicznych), (b) danochłonność tzn. zapotrzebowanie na olbrzymie bazy danych, (c) czasochłonność, pracochłonność i kapitałochłonność przy tworzeniu i eksploatacji modeli oraz (d) brak ogólnie przyjętych metod oceny modeli.

Te ograniczenia wymuszają tworzenie modeli hybrydowych, poprzez wykorzystywanie np. modułów makroekonomicznych lub tworzenie wyspecjalizowanych mikrosymulacji zamiast podejść wszechstronnych. Modele dynamiczne, ze względu na długookresowy charakter analiz zazwyczaj zawierają elementy modelowania ekonometrycznego.

b) Rozkład czasu w mikrosymulacji

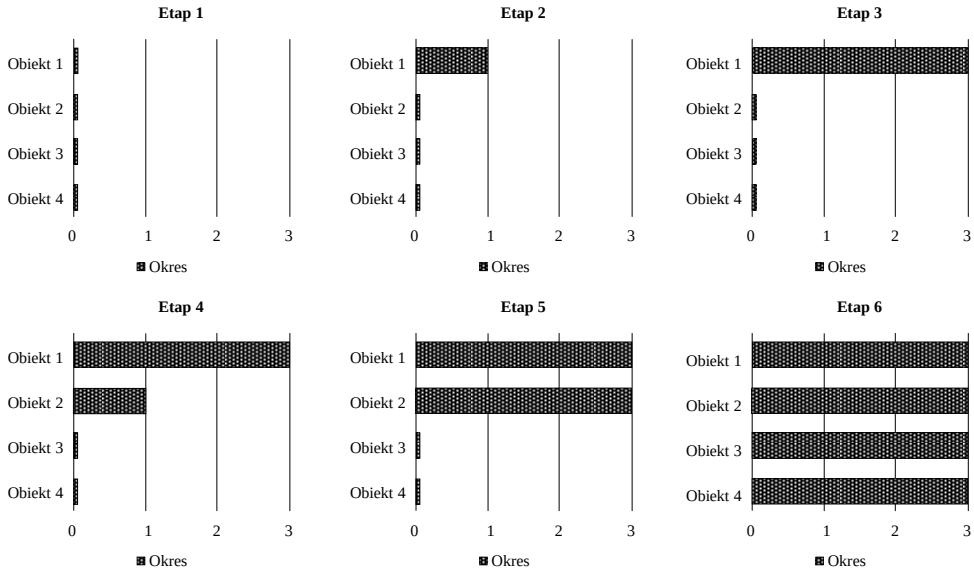
Podział ze względu na charakter rozkładu czasu w symulacji dotyczy zazwyczaj modeli dynamicznych, gdyż pozwalają one na analizy długookresowe. W modelach statycznych problem wyboru rozkładu rzadko występuje ze względu na symulacje krótkookresowe. Znacząca większość istniejących dynamicznych modeli mikrosymulacyjnych służy do symulacji w **czasie dyskretnym**. Oznacza to, że wszystkie zdarzenia (aktualizacja wartości i wariantów zmiennych, podejmowanie decyzji przez obiekty, agregacja wyników pośrednich itp.) następują w stałych przedziałach (odstępach) czasowych, np. co rok. Pomiędzy wyznaczonymi momentami system pozostaje w stagnacji. Po ukończeniu każdego cyklu wartości charakterystyk są weryfikowane: niektóre ulegają zmianie, a inne pozostają stałe. Kluczowym zagadnieniem jest zaprogramowanie kolejności zdarzeń, gdyż w modelu dyskretnym ma ona charakter egzogeniczny. W niektórych przypadkach uszeregowanie zdarzeń i decyzji okazuje się oczywiste i wynika z przyczynowo-skutkowej zależności między zmiennymi. Należy jednak zwrócić uwagę, iż nawet zdarzenia niezależne lub zachodzące symultanicznie muszą być uporządkowane według pewnego klucza. Dyskretne ujęcie czasu pozwala na uwzględnienie jedynie pozornej równoczesności zdarzeń. W przypadku krótkich przedziałów czasowych model dyskretny zbliża się do ciągłego.

Modele w czasie ciągłym nadal stanowią mniejszość wśród modeli dynamicznych, ich rozwój determinuje poziom technologii informatycznej. Każda zmiana, aktualizacja i decyzja mają przypisane przedziały lub rozkłady czasu, tzw. *czas oczekiwania*, w których zachodzą. Jednak każdorazowa realizacja dowolnego zdarzenia powoduje w odpowiedzi konieczność przeszacowania wszystkich pozostałych czasów oczekiwania. Podobnie jak w modelu dyskretnym, zdarzenia nie mogą zachodzić równocześnie, ale w tym przypadku odzwierciedla to rzeczywisty charakter wydarzeń. Aktualizacja wartości zachodzi tylko wtedy, jeżeli nastąpi jakieś zdarzenie. Model ciągły nie likwiduje skokowego charakteru pewnych zmian [Baroni i Richiardi 2007: 24–25].

W modelach w czasie ciągłym łatwiej zbadać dokładność modelu ponieważ są one bliższe rzeczywistości, jednak to modele dyskretne mają silniejsze podstawy w teorii matematyki (określony rozkład cech statystycznych). Ponadto w małych modelach wprowadzenie zmian w czasie ciągłym jest kłopotliwe i czasochłonne, w dużych modelach o małej liczbie aktualizowanych cech przewagę mają modele w czasie ciągłym [Charette 2011].

c) Spektrum badania mikrosymulacyjnego

Klasyfikacja według spektrum badania ma charakter praktyczny, jednak jej wpływ na wyniki eksperymentu jest stosunkowo niewielki. **Modele przypadku** symulują kolejne obiekty populacji na cały okres badania. Oznacza to, że pierwszy obiekt przechodzi przez cały cykl (wszystkie zmiany dla wszystkich okresów), po nim następny aż do ostatniego kończącego mikrosymulację (zob. rys.1.2) [Baroni i Richiardi 2007: 26].

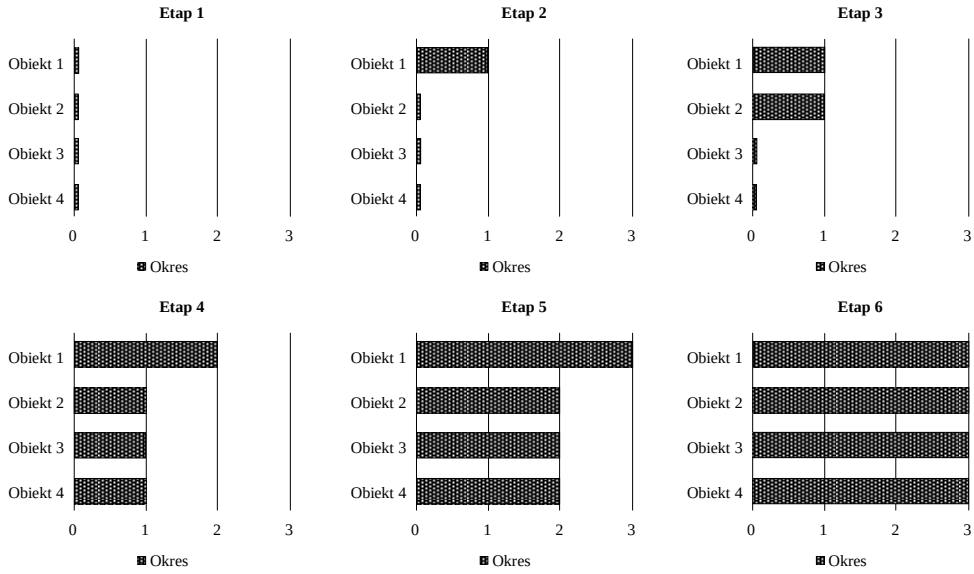


Rys. 1.2. Schemat mikrosymulacyjnego modelu przypadku

Ź r ó d ł o: opracowanie własne.

W **modelu czasu** wszystkie obiekty są symulowane dla kolejnych okresów. Oznacza to, że pierwszy obiekt jest „przeprowadzany” do pierwszego okresu, następnie drugi i kolejne aż ostatni element zostanie wyznaczony. Potem procedurę powtarza się dla drugiego i późniejszych okresów do zakończenia symulacji (zob. rys.1.3) [Baroni i Richiardi 2007: 26].

Jako, że w rezultacie obu podejść otrzymuje się wszystkie obiekty we wszystkich okresach, zatem ma to znaczenie jedynie dla oprogramowania eksperymentu. Czasami wybór uwarunkowany jest liczbą koniecznych do wykonania operacji. Jeżeli jakiś parametr (np. zmieniająca się z okresu na okres inflacja) powiązany jest z konkretnym okresem, to wybór modelu czasu skraca procedurę. Jeśli obiekty charakteryzują się pewnym parametrem (np. nadaną wagą), różnym dla kolejnych elementów, to model przypadku zmniejsza liczba niezbędnych operacji.



Rys. 1.3. Schemat mikrosymulacyjnego modelu czasu

Ź r ó d ł o: jak do rys. 1.2.

d) Izolacji populacji startowej

Ze względu na izolację elementów populacji wstępnej można wyróżnić **modele otwarte** i **zamknięte**. Analitycy nie są jednomyślni, która opcja dokładniej odwzorowuje rzeczywistość. Populacja światowa (ludzi, firm) jest zamknięta, ponieważ nie istnieją interakcje z analogicznymi obiektami spoza niej. Jednak zmniejszając skalę do populacji kraju, regionu, sektora czy kohorty zaczyna dominować system otwarty. Elementy mogą wejść do populacji startowej spoza „środowiska” modelu mikrosymulacyjnego. Dlatego oba podejścia znajdują zastosowanie.

W praktyce model otwarty oznacza, że niezbędne jest zaprogramowanie możliwości interakcji z elementami zewnętrznymi. W modelach zamkniętych

nie istnieje możliwość „wejścia” do eksperymentu. Jest to łatwiejsze, choć uproszczone podejście [Spielauer 2009: 6].

e) Pochodzenie populacji startowej

Populację startową mogą stanowić realne obiekty o charakterze **przekrojowym** lub być sztucznie (**syntetycznie**) wygenerowane. Wykorzystanie danych rzeczywistych wydaje się najlepiej sprzyjać celowi przewidywania skutków polityk i ich reform, a także objaśniania zależności pomiędzy obiektami i cechami. Jednak w niektórych przypadkach niezbędne jest uwzględnienie informacji historycznych, niedostępnych w bazach danych. W takim przypadku populacja wstępna generowana jest od początku (np. od narodzin), co zapewnia pełną informację biograficzną. Brak realizmu stanowi *trade-off* dla rozszerzenia analiz ponad zbiór cech i obiektów z próby czy ankiety.

W zależności od celu badania i dostępności danych oba rodzaje modeli są stosowane.

f) Układ modułowy (blokowy) modelu mikrosymulacyjnego

Cele i zakres tematyczny modelu mikrosymulacyjnego determinują, czy ma on charakter **jedno-** czy **wielomodułowy (-blokowy)**. Wielkie modele mikrosymulacyjne z reguły mają charakter wieloblokowy, zapewniający szersze możliwości tematyczne prowadzonych badań. Modele mniejsze, wyspecjalizowane mogą mieć strukturę jednoblokową.

g) „Czystość” modelu mikrosymulacyjnego

Ze względu na niedoskonałości **modeli mikrosymulacyjnych *sensu stricte*** oraz zapotrzebowania na narzędzia jak najdokładniej odzwierciedlające rzeczywistość, coraz częstsze są próby kompilacji ich z innymi podejściami. Tworzy się w ten sposób **modele mieszane** lub **hybrydowe**. Do najpowszechniejszych metod „wzbogacania” modeli mikrosymulacyjnych należy „endogenizacja” parametrów zewnętrznych, tzn. wprowadzenie submodelu makroekonomicznego [Bourguignon i in. 2010: 1–7; Chitiga 2010: 104–108] oraz implementacja modelu równowagi ogólnej [Colombo 2010: 72–91; Hérault 2010: 35–42; Cororaton i Cockburn 2007: 141–163] (ang. *Computable General Equilibrium*, CGE)[9].

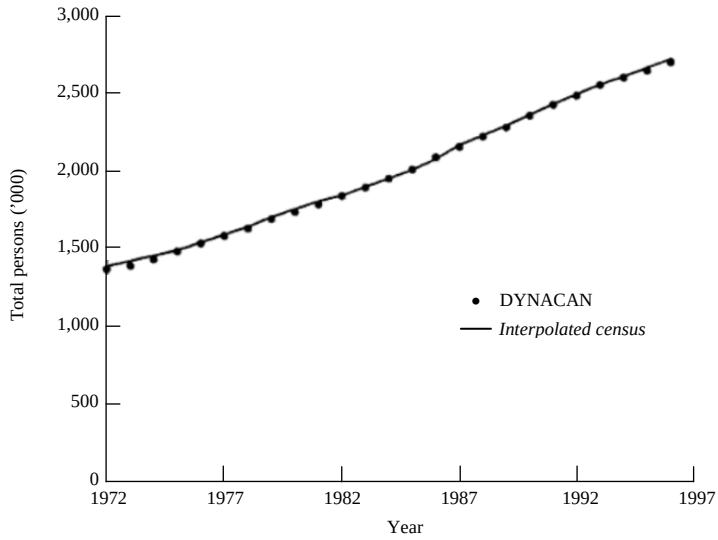
1.4.6. Walidacja i kalibracja modelu mikrosymulacyjnego

W każdym modelu istotny element stanowi ocena jego poprawności. Determinuje ona, w jakim stopniu uzyskane narzędzie mikrosymulacyjne jest wiarygodne, a co za tym idzie przydatne dla prowadzanych analiz oraz podejmowania na ich podstawie decyzji gospodarczych i politycznych. Niezbędne jest nie

[9] „Model CGE jest zbiorem relacji zachowań (równań) które w warunkach doskonałej konkurencji wyznaczają równowagę pomiędzy produkcją i handlem produktami i usługami poprzez zmiany w relacjach cen tych produktów i usług” – cyt. *Raport 2030* [2008: 3].

tylko zweryfikowanie czy model spełnia swoje zadania, ale także stwierdzenie, czy uzyskiwane rezultaty są sensowne z czysto logicznego punktu widzenia.

Walidacja (ang. *validation*) modelu mikrosymulacyjnego opiera się na weryfikacji wyników indywidualnych lub zagregowanych z rzeczywistymi wartościami lub pochodzącymi z niezależnych źródeł, jeśli istnieją. Zarówno wartości zmiennych jak i parametrów muszą być zgodne z logiką, założeniami i teorią (np. dochód osobisty netto nie może być ujemny) [Caldwell i Morrison 2000: 200–203; Hiligsmann i in. 2009: 691–694; *Guidelines for the Use*..., 2007: 40–41]. Procedura walidacyjna nie jest jednoznacznie określona. Cele, zakres i metody modelu oraz subiektywna opinia analityka i/lub zleceniodawcy determinują, które aspekty są najważniejsze dla potwierdzenia jego wiarygodności. Walidacja jest procesem pomiaru jakości, jednak często wykorzystujące miary ilościowe, testy i porównania. Problemem staje się wybór narzędzi lub baz danych służących do weryfikacji wyników modelu. Jeżeli istnieją dane historyczne dla okresu eksperymentu mikrosymulacyjnego, stanowią one jednoznaczny punkt odniesienia otrzymanych rezultatów. Jeśli nie, niezbędne jest wykorzystanie analogicznych informacji z niezależnego źródła (zob. rys. 1.4) [Caldwell i Morrison 2000: 200–203; Hiligsmann i in. 2009: 691–694; *Guidelines for the Use*..., 2007: 40–41; Lymer 2011: 2].



Rys. 1.4. Liczba osób w wieku 65+ w Kanadzie w latach 1972–1997 – porównanie danych interpolowanych ze spisu powszechnego i wyników modelu DYNACAN

Ź r ó d ł o: Caldwell i Morrison [2000: 215].

W przypadku kiedy walidacja pozwala stwierdzić, że wyniki mikrosymulacji nie są wystarczająco zbliżone do rzeczywistych, niezębne jest przeprowadzenie korekty. Kalibracja poprzez „dostrojenie" do rzeczywistych wartości podnosi jakość prognostyczną modeli mikrosymulacyjnych, szczególnie statycznych [*Guidelines for the Use...*, 2007: 24]. Zasadniczym celem wyrównania jest minimalizacja rozbieżności między wartościami empirycznymi a symulacyjnymi. Oznacza to: (1) niwelowanie skutków nieidealnej specyfikacji modelu (braku danych i/lub zmiennych i/lub zależności), (2) wprowadzanie analiz wieloscenariuszowych (np. przez przyjęcie różnych parametrów), (3) powiązanie modelu micro i makro oraz (4) redukcję zmienności rezultatów wynikającą ze stochastycznego charakteru mikrosymulacji. Niekiedy, szczególnie w przypadku danych wejściowych (wartości cech statystycznych obiektów populacji wstępnej) niezbędne jest ich zmodyfikowanie a *priori*, zatem również przed przeprowadzeniem eksperymentu, a zatem także walidacji [Li i O'Donoghue 2011: 4–5; Walker 2007: 114].

Kalibracja może mieć charakter:

– **arbitralny** – wybór przez analityka dowolnej wartości parametru z dopuszczalnego zakresu;

– **automatyczny** – poprzez:

- proste transformacje matematyczne na zmiennych, np.:

$$y_{it}^{K} = \hat{y}_{it} + e_{it} \tag{1.1}$$

$$e_{it} = y_{it} - \hat{y}_{it} \tag{1.2}$$

gdzie:

y_{it}^{K} – wartość zmiennej *Y* w *t*-tym okresie dla *i*-tego obiektu po kalibracji,
$\hat{y}_{it}$ – wartość teoretyczna (wyznaczona przez model mikrosymulacyjny) zmiennej *Y* w *t*-tym okresie dla *i*-tego obiektu,
y_{it} – wartość rzeczywista (empiryczna) lub prognozowana w niezależnym badaniu zmiennej *Y* w *t*-tym okresie dla *i*-tego obiektu,
e_{it} – reszta z modelu dla zmiennej *Y* w *t*-tym okresie dla *i*-tego obiektu;

- wykorzystanie narządzi statystycznych lub ekonometrycznych (np. optymalizacja);
- zastosowanie algorytmów (szczególnie w modelach dynamicznych).

Należy podkreślić, że niezależnie od wybranej postaci kalibracji uzyskane wyniki muszą spełniać teoretyczne i logiczne założenia nałożone na wartości cech statystycznych i parametrów [Menneni i in. 2007; Hollander i Liu 2008; Duncan i Weeks 2000: 298].

W przypadku modeli dynamicznych walidacja i kalibracja są utrudnione, a często wręcz niemożliwe. Istnieje niewiele alternatywnych i niezależnych dla eksperymentu mikrosymulacyjnego długookresowych prognoz danych umożliwiających weryfikacje wyników. Występujące elementy stochastyczne (np. zmienne losowe, parametry) generują różne wyniki, nawet przy niezmienionych warunkach, przy każdej symulacji.

Rozwój mikrosymulacji przyczynia się do szerszego udostępniania danych indywidualnych przez agencje statystyczne, co ma posłużyć m. in. umożliwieniu łatwiejszej i dokładniejszej walidacji i kalibracji w modelach [Gupta i Harding 2007: 3].

1.5. Podsumowanie

Rozwój nowoczesnych narzędzi prowadzenia wielowymiarowych analiz wspomagających proces podejmowania decyzji zależy przede wszystkim od ich możliwości aplikacyjnych. Oznacza to łatwość implementacji i uniwersalność zastosowań. Mikrosymulacje to zbiór narzędzi wspomagających proces podejmowania decyzji na różnych szczeblach zarządzania. Umożliwiają one ocenę przyszłych skutków implementacji poszczególnych wariantów polityk i decyzji poprzez symulowanie sytuacji indywidualnych podmiotów gospodarczych oraz agregację wyników do rezultatów dla makrosystemów. Przełomem w popularyzacji modelowania mikrosymulacyjnego okazał się postęp technologiczny, który przyczynił się do podwyższenia mocy obliczeniowej i redukcji wielkości sprzętu, a zatem także do wzrostu dostępności komputerów osobistych i specjalistycznego oprogramowania. W rezultacie użytkownicy mikrosymulacji mogli rozszerzyć spektrum tematyczne badań na zagadnienia transportowe, medyczne, ekologiczne i interdyscyplinarne. Zamieszczony w rozdziale przegląd literatury krajowej i zagranicznej ostatnich lat stanowi dowód elastyczności aplikacyjnej i prostoty implementacyjnej mikrosymulacji. Jednak nie są to warunki wystarczające do upowszechnienie tej metody badawczej. Konieczne jest zapewnienie zaplecza teoretycznego, by nowi użytkownicy mogli zrozumieć podstawy proponowanego podejścia, tzn. zasadnicze pojęcia, klasyfikacje i sposób konstrukcji modelu mikrosymulacyjnego oraz zaadoptować je do własnych analiz. Niestety dotychczas stosunkowo niewielu autorów, głównie zagranicznych, dostarcza informacji o metodologii mikrosymulacji. Istniejące publikacje są dość rozproszone. Utrudnia to zapoznanie się z całością problematyki, prowadzenie badań i może zniechęcać potencjalnych zwolenników.

W rozdziale podjęto próbę prezentacji spójnej koncepcji mikrosymulacji, która byłaby pomocna zarówno w ewaluacji istniejących modeli i badań, jak również w rozpoczęciu konstrukcji własnych narzędzi. Zaproponowano zinte-

growany system definicji i pojęć kluczowych dla zrozumienia istoty mikrosymulacji i procesu modelowania. Rozróżniono kategorie powiązane ze sobą i zbliżone definicyjnie, jak mikrosymulacja, eksperyment mikrosymulacyjny, model mikrosymulacyjny, populacja startowa i baza mikrodanych. Ponadto zwrócono szczególną uwagę na współzależność i współistnienie modeli ekonometrycznych i mikrosymulacyjnych. Korzystając z istniejącej literatury, dokonano skompilowanej wielokryterialnej klasyfikacji modeli mikrosymulacyjnych, która umożliwia określenie rodzaju modelu ze względu na jego cel, zakres i strukturę.

Rozdział 2

WYBRANE EKONOMICZNE I SPOŁECZNE MODELE MIKROSYMULACYJNE

2.1. Wprowadzenie

Rozwój technologii informatyczno-obliczeniowej przyczynił się do popularyzacji eksperymentów mikrosymulacyjnych. Obszerna literatura przedmiotu zawiera opisy licznych badań ekonomiczno-społecznych oraz dotyczących innych dziedzin nauki, np. medycyny, logistyki. Autorzy publikacji koncentrują się na podstawach teoretycznych i przyczynach prowadzonych analiz oraz na uzyskanych rezultatach eksperymentów mikrosymulacyjnych. Niektóre badania opierają się na wynikach pochodzących z wykorzystania modeli „wielokrotnego użytku", inne bazują na modelach „jednorazowych" zaprojektowanych na potrzeby pojedynczego eksperymentu. W pierwszym przypadku najczęściej narzędziami analiz są modele skonstruowane przez ośrodki naukowe i udostępniane osobom, zespołom oraz instytucjom zewnętrznym. Wykorzystuje się je wielokrotnie, często przez wiele lat, w bardzo różnych celach naukowych. Badacze i autorzy publikacji przytaczają wprawdzie nazwę stosowanego modelu, np. DYNASIM, lecz nie opisują sposobu jego funkcjonowania. Dzieje się tak m. in. dlatego, że konstruktorzy tych modeli nie udostępniają pełnej dokumentacji użytkownikom, pozwalając im korzystać jedynie z „interfejsu" oraz podstawowych informacji. W przypadku modeli projektowanych na „zamówienie" badacz sam jest wykonawcą lub współwykonawcą modelu, co zapewnia personalizacje i dostosowanie narzędzi do celu i możliwości badania. Jednak również w takich opracowaniach brakuje szczegółowych opisów założeń, struktury i funkcjonowania modelu mikrosymulacyjnego. Udzielane informacje są dość ogólne i wybiórcze. Obecnie wykorzystywanych jest wiele modeli mikrosymulacyjnych („wielokrotnego użytku" i „jednorazowych"), zróżnicowanych tematycznie, strukturalnie oraz względem okresu powstania. Niestety większość z nich to „czarne/magiczne skrzynki", dla których niemożliwe jest dokonanie oceny poprawności teoretycznej, merytorycznej, strukturalnej, programistycznej czy estymacyjnej. W szczególności dostępne dokumentacje wnoszą bardzo niewiele

informacji o aspektach ekonometrycznych. Nie zawierają dokładnych informacji o ocenach parametrów strukturalnych oraz metodach estymacji i weryfikacji zastosowanych w submodelach mikroekonometrycznych modeli mikrosymulacyjnych. W przypadku modeli „jednorazowych" publikowanie takich informacji zależy jedynie od dobrej woli autorów i powoli polityka jawności *International Microsimulation Association* (Międzynarodowego Stowarzyszenia Mikrosymulacji) skłania ich do prezentacji wyników ekonometrycznych. Na szczególną jednak uwagę zasługują te modele, które można wielokrotnego zastosować, stanowią podstawę licznych eksperymentów. Oznacza to, że ewentualne błędy konstrukcyjne i estymacyjne tych narzędzi rzutują na wyniki wszystkich przeprowadzonych z ich zastosowaniem badań. Szczególnie, że są one bardziej interdyscyplinarne, wielotematyczne i wymagające większych nakładów sprzętowych niż modele „jednorazowe". Ponadto odzwierciedlają one główne idee modelowania mikrosymulacyjnego zarówno w czasie jak i przekroju aplikacyjnym. Dlatego celem rozdziału jest prezentacja wybranych modeli mikrosymulacyjnych stosowanych w badaniach ekonomicznych i społecznych ze szczególnym uwzględnieniem aspektów ekonometrycznych. Różnorodność prezentowanych modeli pokazuje wszechstronność aplikacyjną i strukturalną mikrosymulacji oraz ich elastyczność implementacyjną. Do przeglądu wybrano modele reprezentujące główne trendy w rozwoju mikrosymulacji. Podstawowym kryterium doboru i klasyfikacji była tematyka modeli.

2.2. Modele świadczeń emerytalnych i rentowych

DYNASIN (*Dynamic Simulation of Income Model* – Model Dynamicznej Symulacji Dochodu), **CORSIM (***Cornell Microsimulation Model* – Mikrosymulacyjny Model Cornell), **DYNACAN** (*Canadian Microsimulation Model for Pension Analysis* – Kanadyjski Mikrosymulacyjny Model Analiz Świadczeń Socjalnych), **PRISM** (*Pension and Retirement Income Simulation Model* – Model Symulacji Dochodów z Świadczeń Socjalnych i Emerytur) i **PIMS** (*Pension Insurance Modeling System* – System Modelowania Ubezpieczeń Emerytalnych) należą do najstarszych modeli mikrosymulacyjnych, odzwierciedlających stan wiedzy i technologii z okresu początków eksperymentów mikrosymulacji. Są to bardzo duże i wszechstronne modele dynamiczne, w większości o strukturze modułowej. Uwzględniają liczne zmienne dotyczące osób i gospodarstw domowych. Dane populacji startowych pochodzą w przeważającej liczbie ze spisów powszechnych, często sprzed kilku dekad. Ze względu na budowę i interdyscyplinarny charakter postrzegane są jako klasyczne modele ekonomiczno-społeczne. Aspekty technologiczne, informatyczne i ekonometryczne tych modeli nie były nigdy w pełni upublicznione. Modele te to tzw. czarne pudełka, czarne

skrzynki (ang. *black box*), które opierają się na zasadzie „włóż dane, wyjmij wyniki i nie zadawaj pytań”. Pełna dokumentacja modeli, a szczególnie ich równań ekonometrycznych i metod estymacji parametrów, jak już wcześniej wspomniano, nigdy nie została upubliczniona.

2.2.1. Model Dynamicznej Symulacji Dochodu (DYNASIM)

Model Dynamicznej Symulacji Dochodu (*Dynamic Simulation of Income Model* – DYNASIM) to pierwszy duży dynamiczny model mikrosymulacyjny. Stanowi on klasyczny, sztandarowy przykład wieloobszarowego, wielomodułowego modelu z początku ery mikrosymulacji, wymagającego znacznych nakładów pracy, sprzętu i czasu. Prace nad nim rozpoczęto w the Urban Institute (USA, Washington) w 1969 r. pod kierunkiem Guy’a Orcutta, a pierwsza wersja została oddana do użytku w 1975 r. W latach 1979–1983 udoskonalono poprawioną wersję modelu – DYNASIM2. Zredukowała ona koszty symulacji i poprawiła jakość analiz odnoszących się do świadczeń emerytalnych i rentowych. DYNASIM stanowi efekt pracy wielu osób i niezależnych zespołów zadaniowych, które przyczyniły się do interdyscyplinarnego charakteru i złożonej budowy blokowej modelu. Wsparcie przy konstrukcji zapewniało wiele instytucji rządowych i pozarządowych, m. in. U.S. Office of Economic Opportunity, the Department of Health, Education and Welfare (HEW), the National Science Foundation, the U.S. Treasury Department, Fundacja Forda i U.S. Department of Labor.

DYNASIM skonstruowano jako narzędzie oceny polityk oraz prowadzenia analiz ekonomiczno-społecznego statusu jednostek badania, tzn. osób indywidualnych, rodzin i gospodarstw domowych. Jego celem jest prowadzenie prognoz i projekcji przyszłego stanu obiektów i warunków ekonomiczno-społecznych, związanych z planowanymi politykami z zakresu transferów publicznych, ubezpieczeń publicznych i emerytur. Dane dotyczące aktorów o charakterze demograficznym, ekonomicznym, społecznym i medycznym pochodzą ze spisu ludności w USA w roku 1973 oraz z bazy ubezpieczeń społecznych w latach 1951–1972.

Model składa się z 3 następujących submodeli:

– historia rodziny i zarobków (ang. *The Family and Earnings History*, FEH) – jest to dynamiczny model symulujący zmienne demograficzne i zachowania siły roboczej; jednostką eksperymentu jest rodzina;

– historia zatrudniania i zasiłków (ang. *The Jobs and Benefits History*, JBH) – plik wyjściowy historii rodziny i zarobków (FEH) stanowi dane wejściowe historii zatrudnienia i zasiłków (JBH), model symuluje zmiany wielkości i struktury zatrudnienia, wartości ubezpieczeń społecznych, emerytur i podatków; jednostką eksperymentu jest rodzina;

– przypisania przekrojowe (ang. *The Cross-Section Imputation*, CSIM) – model statyczny, pliki wyjściowe pierwszych dwóch submodeli FEH i JBH stanowią dane wejściowe modelu CSIM; jednostkami badania są osoby w wieku 60 lat i starsze, ich rodziny i gospodarstwa domowe; uwzględnia informacje zdrowotne i społeczne.

Większość zdarzeń w czasie dyskretnym (np. narodziny, śmierć) symulowana jest metodami Monte Carlo[1]. Oznacza to, że wydarzenia dla każdego obiektu opisywane są równaniami określającymi prawdopodobieństwo ich zajścia. Do estymacji parametrów równań ekonometrycznych postaci logit, probit i tobit zastosowano metodę najmniejszych kwadratów (MNK).

Zaletami modelu DYNASIM są jego wielodyscyplinarny charakter oraz elastyczność w implementacji parametrów. Jest on jednym z najbardziej wszechstronnych modeli ze względu na szeroki zakres uwzględnionych w nim zagadnień ekonomiczno-społecznych. Stanowi on próbę symulacji całego systemu przy minimalnych uproszczeniach. Niestety w rezultacie podnosi to wymagania czasowe i sprzętowe DYNASIM. Model jest także trudny w użytkowaniu, szczególnie dla analityków niezwiązanych z jego konstrukcją, choć jest on udostępniany niezależnym osobom i instytucjom. Szczegółowa dokumentacja dotycząca jego funkcjonowania jest nieaktualna. Ze względu na swoją strukturę oraz wiek danych, model powinien zostać „odnowiony", uwzględniając reestymację parametrów modeli ekonometrycznych. Niestety proces ten byłby bardzo kosztowny, dlatego dotychczas nie podjęto się tego zadania [Anderson 1997c: 1–24].

2.2.2. Mikrosymulacyjny Model Cornell (CORSIM)

Mikrosymulacyjny Model Cornell (*Cornell Microsimulation Model* – CORSIM) to duży dynamiczny model skonstruowany i obsługiwany w Cornell University. Jego profil nastawiony jest bardziej na aspekty socjologiczne niż ekonomiczne, a celem konstrukcji było umożliwienie projekcji pełnej populacji obiektów badania, tzn. osób i rodzin. Model postarza osoby ze spisu powszechnego w USA z roku 1960 do roku 2030, uwzględniając najważniejsze charakterystyki demograficzne i wybrane gospodarcze.

Model został skonstruowany w 1986 r. pod kierownictwem Stevena Caldwella, współtwórcy wcześniej omówionego modelu DYNASIM. Pierwsza wersja CORSIM była podobna do konstrukcji modelu DYNASIM, jednak w latach 1990–1995 model został zaktualizowany i rozszerzony. Obecnie używana jest wersja CORSIM 3.0. CORSIM jest modelem zintegrowanym, składającym się

[1] Zdarzenia (np. umieralność) opisane są tabelami prawdopodobieństw, według zmiennych (np. płeć i wiek). Dla każdego obiektu losowana jest liczba z przedziału ⟨0; 1⟩, jeżeli jest ona nie większa niż prawdopodobieństwo zdarzenia (określone zgodnie z cechami) to zdarzenie zajdzie (np. osoba umiera) – Anderson [1997b: 5].

z licznych powiązanych modułów (ich liczba przekracza 30). Taka wielość bloków jest nietypowa, nawet dla wczesnych dynamicznych modeli mikrosymulacyjnych. Wiąże się to z próbą uwzględnienia w symulacji jak największej liczby czynników ekonomiczno-społecznych. Zdarzenia i aktualizacje wartości charakterystyk dotyczące różnych kategorii są umieszczone w oddzielnych modułach i zachodzą w cyklach rocznych. Nie wszystkie moduły muszą być uruchomiane przy każdej aktywacji modelu, co daje badaczom dużą elastyczność i personalizację w projektowaniu wysłanych eksperymentów. Symulacje w CORSIM postępują według modelu czasu w cyklu rocznym. Model wspomaga analizy dotyczące kształtowania się dochodów osobistych, polityk transferowych, ubezpieczeń społecznych, ryzyka zdrowotnego, ubezpieczeń i opieki dentystycznej. Poza celami badawczymi model miał zapewnić możliwość prowadzenia eksperymentów na próbach z zastosowaniem komputerów osobistych oraz służyć do poprawy efektywności technik mikrosymulacyjnych. Większość zaprogramowanych zdarzeń ma charakter dyskretny i jest symulowana metodami Monte Carlo. Ponieważ eksperyment przeprowadzany jest od roku 1960 oraz dostępne są liczne bazy danych historycznych, walidacja i kalibracja wyników stanowią istotny element mikrosymulacji. Wyniki społeczne i niektóre ekonomiczne uzyskane z eksperymentów w oparciu o CORSIM porównywane są z empirycznymi realizacjami zmiennych. Umożliwia to sprawdzenie trafności prowadzonych symulacji, czyli ich walidację. Ponadto znając prawdziwe wartości cech można skalibrować model tak, by w eksperymentach dla krótkich okresów rezultaty były arbitralnie dopasowane, redukując błędy szacunków.

CORSIM jest dostępny dla badaczy spoza *Cornell University*, ale odpłatnie. Ze względu na akademicki charakter, zarówno procesu konstrukcji jak i przeznaczenia, niektóre aspekty ekonomiczno-społeczne zaimplementowano do modelu w wersji uproszczonej. Dokumentacja CORSIM jest przestarzała i niepełna. Istnieją informacje techniczne dla wersji 3.0, jednak nie ma dostępu do wyników estymacji parametrów modeli ekonometrycznych, co uniemożliwia ocenę jakości oszacowań [Anderson 1997d: 1–35].

2.2.3. Model Symulacji Dochodów z Świadczeń Socjalnych i Emerytur (PRISM)

Model Symulacji Dochodów z Świadczeń Socjalnych i Emerytur (*Pension and Retirement Income Simulation Model* – PRISM) to dynamiczny model mikrosymulacyjny służący analizom rozkładu dochodów emerytalnych, długotrwałej opieki i finasowania osób starszych (w wieku 65 lat i więcej). Został skonstruowany w 1980 r. dla *U.S. Department of Labor*. Stosowany jest przez instytucje publiczne i prywatne w celu weryfikacji rozkładu efektów różnorodnych polityk emerytalnych. Jest bardziej wyspecjalizowanym modelem niż DYNASIM i CORSIM, a jego tematyka została ograniczona w porównaniu

z wszechstronnymi modelami do zagadnień powiązanych z bezpośrednimi celami.

Do podstawowych funkcji należą dynamiczne postarzanie próby 28 000 osób (baza danych powstała z połączenia różnych źródeł z końca lat 70. XX w.) oraz dostarczenie syntetycznej przekrojowej bazy danych przyszłej populacji USA. Model mikrosymulacyjny umożliwia prowadzenie analiz publicznych polityk związanych z emeryturą i długotrwałą opieką medyczną. Jednostki badania (aktorzy) w eksperymencie to osoby indywidualne i małżeństwa. Ich charakterystyki demograficzne i ekonomiczne są symulowane na okres 1979–2020 (lub 2030).

Model składa się z 4 modułów:

– wydarzenie demograficzne osób i rodzin – symuluje wydarzenia demograficzne osób dla wszystkich lat eksperymentu;

– doświadczenie na rynku pracy, wraz z charakterystyką planu emerytalnego i przyjmowaniu świadczeń – symuluje indywidualną historię zatrudnienia osób od 1979 r. lub od momentu rozpoczęcia pracy, rozpoczynając od statusu wyjściowego (dla osób pełnoletnich w 1979 r.) lub syntetycznego (wyznaczanego poprzez podobieństwo do innych osób), uwzględnia czas pracy, zmiany zatrudnienia, plany emerytalne, ubezpieczenia społeczne;

– dochody emerytalne: źródło i poziom – szacuje dochody osób starszych z różnych źródeł (m. in. pensja, emerytura, ubezpieczenia społeczne) oraz podatki stanowe i lokalne; charakterystyki polityk emerytalnych i planów ubezpieczeniowych zostały zaktualizowane w 1984 r.;

– finasowanie długotrwałej opieki – dane wyjściowe 3 pierwszych modułów stanowią pliki wejściowe modułu finansowania długotrwałej opieki, symuluje on zdarzenia związane z wykorzystaniem i finansowaniem długotrwałej opieki nad osobami starszymi (uwzględnia opiekę zinstytucjonalizowaną, np. domu opieki).

Większość zdarzeń symulowana jest metodami Monte Carlo. Dane o populacji startowej zaczerpnięto z banku informacji o transferach publicznych i ubezpieczeniach społecznych dla reprezentatywnej próby 28 000 osób pełnoletnich w latach 1978–1979 oraz z bazy danych prowizji planów emerytalnych z 1984 r. W przeciwieństwie do DYNASIM i CORSIM, gdzie przydział planów emerytalnych jest losowy, PRISM wykorzystuje rzeczywiste informacje. Zarówno dane z populacji jak i od dostarczycieli planów emerytalnych są stosunkowo stare. Od połowy lat 80. XX w. nastąpiły zmiany legislacyjne w USA. W rezultacie zaostrzono politykę ochrony danych osobowych i utajniono m. in. informacje o ubezpieczonych. Te względy prawne uniemożliwiają ponowne przeprowadzenie badań ankietowych, a zatem także aktualizacje baz mikrodanych populacji startowej.

Dokumentacja tego modelu mikrosymulacyjnego jest przestarzała, chociaż była poddawana aktualizacji. Nie istnieje żaden „podręcznik” ani dokument techniczny dotyczący funkcjonowania PRISM. Pod koniec XX w. model został

dostosowany do technologii komputerów osobistych (PC), co zredukowało koszty prowadzonych eksperymentów. Większość komponentów nie jest dostępna poza centralą organizacji *The Lewin Group*, będącą właścicielem modelu. PRISMM nie jest udostępniany innym podmiotom w celach badawczych [Anderson 2001: 1–34].

2.2.4. Kanadyjski Mikrosymulacyjny Model Analiz Świadczeń Socjalnych (DYNACAN)

Kanadyjski Mikrosymulacyjny Model Analiz Świadczeń Socjalnych (*Canadian Microsimulation Model for Pension Analysis* – DYNACAN) z 1998 r. to dynamiczny model mikrosymulacyjny stworzony przez federalny rząd Kanady jako narzędzie do prowadzenia analiz efektów alternatywnych polityk planów emerytalnych. Przy zastosowaniu DYNACAN możliwe jest postarzanie próby (max. 1%) z danych spisu powszechnego z roku 1971 i kolejnych. Model jest modelem czasu w cyklu rocznym powtarzanym do 2100 r. Jednostkami badania są osoby i rodziny. Głównymi funkcjami modelu jest szacowanie składek i transferów emerytalnych, uwzględnia on charakterystyki demograficzne i ekonomiczne. DYNACAN wzorowany na modelu CORSIM 2.0 i oddany do użytku w 1998 r. składa się z 3 głównych modułów:

– DYNACAN–A – łączy dane indywidualne i rodzinne tworząc populację startową,

– DYNACAN–B – symuluje w czasie historie demograficzne, zarobków i zatrudnienia wszystkich jednostek populacji startowej,

– DYNACAN–C – wylicza wkład i korzyści Kanadyjskiego Planu Emerytalnego.

Większość zdarzeń, szczególnie w DYNACAN–B, symulowana jest metodami Monte Carlo. Wartość pensji jako zmiennej ciągłej szacowana jest na podstawie równania ekonometrycznego, a następnie zaburzana czynnikiem losowym. Niestety specyfikacja równania dochodu ani oceny parametrów strukturalnych nie są znane. W publikacjach dotyczących DYNACAN wspomina się jedynie o stosowaniu dwóch metod estymacji, metody najmniejszych kwadratów (MNK) oraz metody największej wiarygodności (MNW). Nie znając jednak szczegółów procesu estymacji, trudno stwierdzić, czy zastosowany submodel ekonometryczny jest poprawny. Szczególnie, że estymator MNK, w klasycznej postaci, nie jest wskazany do szacowania parametrów strukturalnych modeli mikroekonometrycznych.

DYNACAN ma znaczne wymagania sprzętowe oraz jest modelem kosztoi czasochłonnym. Podobnie jak w innych dużych modelach, m. in. DYNASIM, CORSIM, ze względu na koszty unika się powtarzania eksperymentu przy niezmienionych warunkach i parametrach.

Dokumentacja modelu jest stosunkowo dokładna i dostępna. Brak jest informacji o estymacji parametrów modeli ekonometrycznych w DYNACAN. Wiadomo, że niektóre oszacowania zostały zaczerpnięte bezpośrednio z CORSIM. Model jest elastyczny, co umożliwia łatwą rozbudowę. DYNACAN funkcjonuje na komputerach osobistych o wystarczających parametrach, jednak zainstalowany jest wyłącznie w siedzibie Office of the Superintendent of Financial Institutions (OSFI) [Anderson 1998: 1–30].

2.2.5. System Modelowania Ubezpieczeń Emerytalnych (PIMS)

System Modelowania Ubezpieczeń Emerytalnych (*Pension Insurance Modeling System* – PIMS) to dynamiczny model mikrosymulacyjny stworzony w latach 80. i 90. XX w. na potrzeby *Pension Benefit Guaranty Corporation* (PBGC)[2] w celu analizy sytuacji rynkowej ubezpieczeń emerytalnych. Główną jego funkcją jest szacowanie ryzyka związanego z ubezpieczeniami oraz wysokości roszczeń wobec PBGC w założonym okresie. Jednostkami eksperymentu są firmy ubezpieczające pracowników (ubezpieczający), ich plany ubezpieczeniowe oraz *Pension Benefit Guaranty Corporation* (PBGC). PIMS ma budowę modułową, a do najważniejszych bloków należą moduły:

– ekonomiczny – kontroluje stopy zwrotu, zwroty z giełdy jako procesy stochastyczne,

– firma – na podstawie informacji o sytuacji firmy oraz planach ubezpieczeniowych i zatrudnieniu wyznacza prawdopodobieństwo bankructwa firmy,

– plan – na podstawie informacji demograficznych oraz danych o obciążeniach i kapitale określa sposób finansowania planu,

– IRS – definiuje zasady, przy których wyznaczane są minimum i maksimum opłat planu,

– PBGC – rozważa sytuację bankructwa firmy i skutki dla korporacji PBGC, w tym wypłaty wynikające z planów bankrutującej firmy.

Symulacja dokonywana jest na 20 lat. Dane pochodzą z różnych źródeł, m. in. dane o bankrutujących firmach, trzydziestoletnie stopy zwroty bonów państwowych i spółek akcyjnych, roczne raporty firm. PIMS wykorzystuje modele ekonometryczne o parametrach estymowanych metodą najmniejszych kwadratów (MNK) i największej wiarygodności (MNW), w tym modele logit i probit oraz modele wielorównaniowe o równaniach pozornie niezależnych (SUR). Brak jest bardziej szczegółowych informacji o konstrukcji, estymacji i wersyfikacji submodeli ekonometrycznych. Część zdarzeń symulowana jest metodami Monte Carlo. Model jest dostosowany do komputerów osobistych, zatem koszty PIMS są stosunkowo niskie [Anderson 1999: 1–15].

[2] Instytucja rządu USA mająca na celu nadzorowanie dobrowolnych prywatnych planów emerytalnych oraz zapewniania wypłat nawet w przypadku bankructwa firm [źródło: PBGC].

Tabela 2.1

Charakterystyka modeli świadczeń emerytalnych i rentowych

Nazwa modelu	Rodzaj modelu	Funkcje	Jednostki badania	Dane	Moduły	Zakres tematyczny	Dostępność informacji o modelach ekonometrycznych i metodach estymacji	Modele ekonometryczne i metody estymacji
1	2	3	4	5	6	7	8	9
DYNASIM	dynamiczny	• symulacja skutków alternatywnych polityk • prognozy 1973–2030	• osoba • rodzina • gospodarstwo domowe	• spis ludności USA 1973 • dane o dochodach z Ubezpieczeń Społecznych 1951–1972	• model historii rodziny i zarobków • model historii zatrudniania i zasiłków • model przepisania przekrojowego	• ekonomiczny • społeczny • demograficzny • medyczny	ograniczona	1) modele: – logit – probit – tobit 2) estymacja: – MNK
CORSIM	dynamiczny	• badanie wewnątrz- i międzpokoleniowej mobilności ekonomiczno-społecznej • analiza rozkładu dochodów • symulacja skutków różnych polityk podatkowych	• osoba • rodzina • gospodarstwo domowe	• spis ludności USA 1960	więcej niż 30, m. in.: • procesy demograficzne • pokrewieństwo • mobilność geograficzna • wykształcenie • praca i zarobki • transfery i podatki • niepełnosprawność • konsumpcja i oszczędności	• ekonomiczny • społeczny • demograficzny • medyczny	ograniczona	1) modele: – logit – probit – tobit 2) estymacja: –MNK

1	2	3	4	5	6	7	8	9
PRISM	dynamiczny	• stworzenie syntetycznej przekrojowej bazy danych przyszłej populacji USA • analiza efektów polityk publicznych • prognozy do 2020/2030	• osoba • rodzina • gospodarstwo domowe	• historia zarobków z 1978 • różne ankiety i spisy	• wydarzenie demograficzne osób i rodzin • doświadczenie na rynku pracy i plan emerytalny • dochody emerytalne: źródło i poziom • finansowanie długotrwałej opieki	• ekonomiczny • społeczny • demograficzny	brak	
DYNACAN	dynamiczny	• symulacja demografii i zarobków Kanadyjczyków • analiza zmian w Kanadyjskim Planie Emerytalnym	• osoba • rodzina	• spis ludności Kanady 1971	• DYNACAN–A • DYNACAN–B • DYNACAN–C	• ekonomiczny • społeczny • demograficzny • medyczny	ograniczona	1) estymacja: – MNK – MNW
PIMS	dynamiczny	• wyznaczanie poziomu niepewności związanego z ubezpieczeniami emerytalnymi • wyznaczanie poziomów i prawdopodobieństw przyszłych żądań wypłat z tytułu ubezpieczeń emerytalnych • prognozowanie przyszłej sytuacji finansowej netto *Pension Benefit Guaranty Corporation* (PBGC) • analiza efektów obecnych i przyszłych państwowych polityk ubezpieczeń emerytalnych.	• firma–płatnik składek ubezpieczeń emerytalnych • plan emerytalny	• dane o dochodach dane o bankrutujących firmach • 30-letnie stopy zwrotu bonów państwowych i giełdy • roczne raporty firm	• ekonomiczny • firma • plan • IRS (zakres opłat planu ubezpieczeniowego) • PBGC (konsekwencje bankructwa ubezpieczyciela)	• ekonomiczny • społeczny	ograniczona	1) modele: – logit – probit – SUR 2) estymacja: – MNK – MNW

MNK – metoda najmniejszych kwadratów, MNW – największej wiarygodności, SUR – model o równaniach pozornie niezależnych.

Ź r ó d ł o: opracowanie własne.

2.2.6. Porównanie modeli świadczeń emerytalnych i rentowych

W celu porównania struktury oraz tematyki modeli świadczeń emerytalnych i rentowych zestawiono ich charakterystyki w tab. 2.1. Modele świadczeń emerytalnych i rentowych to modele dynamiczne, wielomodułowe i interdyscyplinarne o heterogenicznym zbiorze populacji wstępnej. Startowa baza mikrodanych pochodzi zazwyczaj ze spisów powszechnych z lat 60. i 70. XX w. Wyjątkiem jest PIMS, którego obiektami są firmy i plany emerytalne, jednak dane startowe również pochodzą sprzed kilku dekad. Wszystkie modele zawierają submodele mikroekonometryczne, ale informacje o nich są ograniczone. Nie jest znana specyfikacja modeli, oszacowania parametrów ani jakość dopasowania do danych empirycznych (zob. tab. 2.1). W efekcie nie można ocenić poprawności wyników badań mikrosymulacyjnych prowadzonych w oparciu o te modele.

2.3. Modele podatkowo-zasiłkowe

Obecnie *International Microsimulation Association* promuje odchodzenie od idei „czarnych skrzynek" na rzecz jawnych i transparentnych modeli oraz pełnej dokumentacji. Jednak nie wszystkie nowe modele spełniają te warunki. Modele zasiłkowo-podatkowe dla Unii Europejskiej EUROMOD i jego podmodel dla Polski – SIMPL oraz model Ministerstwa Finansów RP nie posiadają pełnej, jawnej dokumentacji. Wszystkie są oprogramowane w pakietach Microsoft Excel i Visual Studio. Posiadają modele ekonometryczne, jednak nie można zweryfikować zastosowanych w nich metod estymacji i uzyskanych oszacowań.

Mikrosymulacyjny model podatkowo-zasiłkowy Ministerstwa Finansów RP to alternatywny model dla SIMPL, całkowicie niezależny od inicjatywy EUROMOD. Został on ukończony w roku 2011 i na razie nie istnieje jego oficjalna dokumentacja. W roku 2012 pojawiła się pierwsza publikacja dotycząca struktury i funkcjonowania modelu [Konopczak i Skibicki 2012]. Większość informacji o nim pochodzi jednak bezpośrednio od konstruktorów. Modele zasiłkowo-podatkowe są modelami statycznymi, o ograniczonej tematyce, bardziej ekonomicznej niż społecznej.

2.3.1. EUROMOD – Podatkowo-zasiłkowy model dla Unii Europejskiej

EUROMOD to model podatkowo-zasiłkowy zapoczątkowany w 1998 r. dla 15 krajów członkowskich Unii Europejskiej, tzn. Austrii, Belgii, Danii, Finlandii, Francji, Grecji, Holandii, Hiszpanii, Irlandii, Luksemburga, Niemiec, Portu-

galii, Szwecji, Wielkiej Brytanii i Włoch [Atkinson i in. 1999: 3]. Projekt został sfinansowany z IV Programu Ramowego UE – *Targeted Socio-Economic Research* (TSER). Zasadniczą część oprogramowania modelu stanowią pakiety Microsoft. Podstawowymi założeniami przy konstrukcji EUROMOD były:

– transparentność metod,
– elastyczność i możliwość adaptacyjna,
– zgodność i porównywalność pomiędzy krajami.

Do roku 2006 dodano moduły dla Estonii, Polski (SIMPL), Słowacji i Węgier. Trwa także rozbudowa EUROMOD o kolejne kraje UE oraz aktualizacja wprowadzanych polityk podatkowo-zasiłkowych [źródło: ISER].

Model ma charakter statyczny i umożliwia prowadzenie analiz natychmiastowych efektów zmian polityk dotyczących m. in.: zasiłków rodzinnych, świadczeń socjalnych, podatków dochodowych i pośrednich, ubezpieczeń zdrowotnych. Wyniki są wyznaczane tylko na poziomie makroekonomicznym (niektóre wyniki mikroekonomiczne są dostępne, ale jedynie w centrali projektu w Cambridge), co umożliwia badanie skutków polityk w zakresie: dystrybucji dochodów w społeczeństwie, efektów według grup (np. płeć, region, wiek), efektów marginalnych stawek podatkowych, efektów zmian wartości miar ubóstwa i nierówności, różnic w kosztach i korzyściach pomiędzy państwami członkowskimi. EUROMOD ma także zapewnić możliwość porównywania wartości statystyk pomiędzy krajami członkowskimi.

Jednostkami badania są osoby i gospodarstwa domowe. Mikrodane zbierane są niezależnie w poszczególnych krajach. Brak jest ujednoliconych standardów i źródeł. Znormalizowany został jedynie minimalny zbiór zmiennych oraz ich definicje. Model ma budowę modułową według krajów członkowskich, stąd łatwość w adaptacji kolejnych bloków dla nowych państw.

Ze względu na statyczny charakter modelu, EUROMOD jest szczególnie użyteczny do analizy rozkładu efektów krótkookresowych reform polityk podatkowo-zasiłkowych. Nie uwzględniane są zmiany behawioralne ani efekty długookresowe [Sutherland 2001: 1–8].

2.3.2. SIMPL – Podatkowo-zasiłkowy model dla Polski

SIMPL to pierwszy w pełni funkcjonalny model mikrosymulacyjny dla Polski. Stanowi on moduł modelu podatkowo-zasiłkowego Unii Europejskiej EUROMOD. SIMPL skonstruowano w ramach rozbudowy EUROMOD o kolejne kraje członkowie UE. Jest to model statyczny, mający na celu umożliwienie prowadzenia analiz rozkładu natychmiastowych efektów zmian polityk dotyczących kształtowania się zasiłków rodzinnych, świadczeń socjalnych, podatków dochodowych i pośrednich, ubezpieczeń zdrowotnych [Frejlich i Nicińska 2010: 1–7].

Model mikrosymulacyjny jest oprogramowany z wykorzystaniem pakietaów: MS Office (Excel i Access) oraz Visual Studio. Obiektami w eksperymencie są osoby oraz gospodarstwa domowe. Dane pochodzą z Badania Budżetów Gospodarstw Domowych [Bargain i in. 2007a: 1–5].

Pierwszą wersję modelu zbudowano dla danych Badania Budżetów Gospodarstw Domowych w 2003 r. Wyniki symulacji okazały się zbieżne z wartościami empirycznymi. SIMPL jest dostosowywany do baz danych z kolejnych wersji badania. Aktualna wersja pozwala na symulacje wysokości większości podatków bezpośrednich i zasiłków [Bargain i in. 2007b: 1–15]. Specyfika bazy danych gospodarstw domowych odróżnia SIMPL od wielu modułów EUROMOD. Nie wszystkie zmienne dostępne są na poziomie indywidualnych jednostek, a jedynie dla gospodarstw domowych. Problemem do rozwiązania okazała się konieczność dostosowania cech jakościowych do analogicznych kategorii w innych krajach. Informacje o dochodach i wydatkach w Badaniu Budżetów Gospodarstw Domowych dotyczą jednego miesiąca. W państwach członkowskich UE często zbierane są informacje o rocznych dochodach oraz o liczbie przepracowanych miesięcy, co umożliwia wyznaczenie średnich miesięcznych dochodów. W SIMPL dochody są symulowane przy założeniu stałej stawki wynagrodzenia w ciągu całego roku [Frejlich i Nicińska 2010: 7].

2.3.3. Model Ministerstwa Finansów RP

Mikrosymulacyjny model podatkowo-zasiłkowy Ministerstwa Finansów (MF) RP to statyczny model niezależny od SIMPL, choć o analogicznej tematyce. Model MF został oddany do użytku w 2011 r. i nie istnieje jeszcze jego oficjalna pełna dokumentacja techniczna oraz metodologiczna. W 2012 r. opublikowano pierwsze dane dotyczące jego konstrukcji, struktury i funkcjonowania [Konopczak i Skibicki 2012]. Dodatkowe informacje, na potrzeby niniejszej pracy, uzyskano dzięki uprzejmości jego twórców.

Model MF ma na celu umożliwienie prowadzenia krótkookresowych analiz efektów finansowych i redystrybucyjnych zmian w systemie podatkowo-zasiłkowym Polski. Jednostkami badania są gospodarstwa domowe, a bazę mikrodanych wejściowych zaczerpnięto z Badania Budżetów Gospodarstw Domowych.

Model ma strukturę modułową. Jego podstawowe bloki to:

- ubruttowienie danych dochodowych,
- moduł podatkowy,
- moduł ulg podatkowych,
- moduł zasiłków rodzinnych,
- moduł pomocy społecznej,

– moduł VAT,
– moduł podaży pracy.

Do podstawnych charakterystyk gospodarstw domowych należą: dochód brutto, dochód po opodatkowaniu, dochód do dyspozycji przy uwzględnieniu ulg podatkowych, dochód do dyspozycji przy uwzględnieniu transferów socjalnych. W modelu wykorzystano oszacowania mikroekonometrycznych modeli specjalnych zmiennych jakościowych.

2.3.4. Porównanie modeli podatkowo-zasiłkowych

Modele podatkowo-zasiłkowe są modeli statycznymi i pozwalają na prowadzenie krótkookresowych eksperymentów. Ich tematyka jest ściśle ekonomiczna. Obiektami badań są osoby i/lub gospodarstwa domowe, a dane dla Polski pochodzą z Badania Budżetów Gospodarstw Domowych mających charakter ankietowy i reprezentatywny. W modelach podatkowo-zasiłkowych występują submodele mikroekonometryczne, jednak informacje o ich konstrukcji i estymacji parametrów nie są upublicznione. Obniża to możliwości oceny wyników badań z zastosowaniem modeli podatkowo-zasiłkowych, mimo że są to jedne z najnowszych modeli mikrosymulacyjnych (zob. tab. 2.2).

2.4. Modele społeczne i demograficzne

Modele społeczne i demograficzne, często pozbawione submodeli ekonometrycznych[3], ukierunkowane są na szczegółowo określoną problematykę. SOCSIM to model społeczny. Nie zawiera elementów ekonomicznych, a więc i ekonometrycznych. Służy symulacji struktury demograficznej całej populacji danych regionów w okresie do kilkuset lat. Wyniki takich eksperymentów służą nie tylko analizom socjologicznym, ale także antropologicznym, demograficznym i ekonomicznym. LifePaths to jeden z najbardziej wszechstronnych modeli mikrosymulacyjnych, zawierający elementy ekonomiczne, choć nie ekonometryczne. Jako jeden z niewielu symuluje syntetyczną historię obiektów, korzystając z licznych baz danych (spisów powszechnych i ankiet). Demosim ma podobną strukturę do LifePaths, jest jednak nieco prostszy i ukierunkowany na strukturę etniczno-kulturową populacji. Uwzględnia nieliczne czynniki ekonomiczne i żadnych modeli ekonometrycznych, jednak wykorzystuje metody estymacji dla danych społeczno-demograficznych.

[3] Nie musi to być równoznaczne z brakiem statystycznych modeli regresji.

Tabela 2.2

Charakterystyka modeli podatkowo-zasiłkowych

Nazwa modelu	Rodzaj modelu	Funkcje	Jednostki badania	Dane	Moduły	Zakres tematyczny	Dostępność informacji o modelach ekonometrycznych i metodach estymacji	Modele ekonometryczne i metody estymacji
EUROMOD	statyczny	• analiza natychmiastowych efektów na zmiany polityk podatków i zasiłków • porównania pomiędzy państwami członkowskimi	• osoba • gospodarstwo domowe	• dane gospodarstw domowych zbierane w różnych krajach (brak jednego standardu i źródła danych)	podział według krajów członkowskich UE	• ekonomiczny	brak	x
SIML	statyczny	• analiza natychmiastowych efektów na zmiany polityk podatków i zasiłków • porównania pomiędzy państwami członkowskimi	• osoba • gospodarstwo domowe	• Badania Budżetów Gospodarstw Domowych	brak	• ekonomiczny	brak	x
Mikrosymulacyjny model podatkowo-zasiłkowy Ministerstwa Finansów RP	statyczny	• analiza finansowych i redystrybucyjnych skutków zmian w systemie podatkowo--zasiłkowym	• gospodarstwo domowe	• Badania Budżetów Gospodarstw Domowych	• ubruttowienie danych dochodowych • podatkowy • ulg podatkowych • zasiłków rodzinnych • pomocy społecznej • VAT • podaży pracy	• ekonomiczny	ograniczona	1) Modele: – zmiennych jakościowych (m. in. logit)

Źródło: jak to tab. 2.1.

2.4.1. Symulacja Społeczna (SOCSIM)

Symulacja Społeczna (*Social Simulation* – SOCSIM) to dynamiczny model mikrosymulacyjny o zdarzeniach w czasie ciągłym, którego aktorami są osoby. Został on skonstruowany na Uniwersytecie Berkeley w Kalifornii w latach 70. XX w. na zlecenie *National Science Fundation*[4]. Podstawowym celem modelu jest analiza wpływu czynników demograficznych na strukturę pokrewieństwa i formowanie się gospodarstw domowych [źródło: ADSRI]. SOCSIM wykorzystywany jest do badań społecznych, demograficznych i antropologicznych, a pośrednio także ekonomicznych. Umożliwia symulację pokrewieństwa, koligacji i małżeństw pomiędzy osobami z populacji. Model ma charakter zamknięty, co oznacza, że populacja startowa, końcowa lub pośrednia odzwierciedla zbiorowość generalną osób lub gospodarstw domowych zamieszkujących dany teren w określonym punkcie czasu. Zatem nie wprowadza się partnerów (małżonków) spoza populacji startowej, ale dobiera się ich na podstawie podobieństwa cech społecznych i demograficznych. W celu zapewnienia stabilnej struktury społeczności wywodzącej się z populacji wstępnej niezbędne jest przeprowadzenie symulacji długookresowych, nawet na kilkaset lat wstecz. Stąd mikrodane pochodzą często z preindustrialnych spisów ludności, m. in. Anglii i Walii z 1751 r.

Mikrosymulacje w oparciu o SOCSIM pozwalają badać zmienność w populacji i wskazywać długookresowe wzorce demograficzne. Stosowane są w analizach i ocenach alternatywnych sposobów utrzymania stałej wielkości populacji i niewymierania rodów.

SOCSIM jest zaprogramowany w pakiecie R. Kod źródłowy modelu, skrócona dokumentacja techniczna oraz wybrane wyniki badań są ogólnodostępne, m. in. na stronie internetowej *Uniwersytetu Berkeley w Kalifornii* [źródło: UC Berkeley]. Interfejs SOCSIM nie jest przyjazny dla użytkownika. SOCSIM potrzebuje bardzo wielu parametrów demograficznych, żeby adekwatnie symulować zmiany struktury społecznej [Citro i Hanushek 1991: 286–287; Murphy 2009; Mason 2011: 1–30].

2.4.2. Ścieżki Życia (LifePaths)

LifePaths (Ścieżki Życia) to dynamiczny model mikrosymulacyjny skonstruowany przez *Statistics Canada*[5]. Został on oparty na platformie Modgen[6].

[4] Agencja rządowa USA mająca na celu promocje nauki i inżynierii poprzez programy naukowe i projekty edukacyjne [źródło: NSF].

[5] *Statistics Canada to* Główny Urząd Statystyczny Kanady.

[6] Modgen to powszechnie dostępna platforma skonstruowana w *Statistics Canada na* bazie języka programowania C++ i pakietu MS Visual Studia mająca wspomagać konstrukcje modeli mikrosymulacyjnych.

Jest to model przypadku o zdarzeniach w czasie ciągłym, co zwiększa realizm struktury społeczno-demograficznej populacji jednostek (osób i rodzin). LifePaths symuluje syntetyczne historie (tzw. ścieżkę życia, ang. *life path*) osób od narodzin lub imigracji do Kanady do śmierci. Dane wejściowe zawierają jedynie datę i miejsce urodzenia lub migracji oraz płeć. Kolejne osoby dołączają w momencie ich rzeczywistego narodzenia (spis narodzin) lub przyznania im prawa do stałego pobytu (osoby pozostające w kraju jedynie czasowo nie są uwzględnione w modelu). Model ma strukturę otwartą. W każdej rodzinie jest osoba dominująca, dla której partner oraz dzieci są generowani syntetycznie spoza populacji. Aktor dominujący w rodzinie pochodzi z populacji wstępnej (narodziny lub imigracja) i tylko on reprezentuje swoją rodzinę w wynikach końcowych.

Generowanie całej ścieżki życia jednostek badania według zadanych rozkładów prawdopodobieństw pozwala na uzyskanie wielu informacji niedostępnych w spisach i ankietach. Wymaga to jednak wykorzystania różnorodnych baz danych pozwalających określić niezbędne prawdopodobieństwa. Stąd konieczność wprowadzenia licznych źródeł informacji odzwierciedlających poszczególne aspekty społeczne, demograficzne i ekonomiczne populacji Kanady.

LifePaths został skonstruowany do prowadzenia analiz struktury społecznej i demograficznej populacji Kanady w latach 1971–2026. Umożliwia on weryfikację długookresowych skutków wprowadzania reform dotyczących podstawowych dziedzin życia. Struktura modelu odzwierciedla jego wszechstronny charakter. Poszczególne moduły odpowiadają za konkretne aspekty społeczne:

- populacja i migracja,
- umieralność,
- płodność,
- wykształcenie,
- małżeństwa, związki nieformalne i rozwody,
- zmiany w składzie gospodarstwa domowego,
- urlop macierzyński,
- zatrudnienie,
- zarobki.

Moduły zostały skonstruowane oddzielnie i połączone. Struktura LifePaths jest elastyczna, co umożliwia systematyczną ewolucję modelu.

Model LifePaths jest ogólnodostępny na stronie internetowej *Statistics Canada*, wymaga jedynie instalacji pewnych komponentów platformy Modgen. Jest to model wszechstronny, stosunkowo prosty w obsłudze, o szczegółowej dokumentacji technicznej [źródło: StatCan_c; *The LifePaths*...].

2.4.3. Demosim – model symulacji demograficznych

Demosim to dynamiczny model demograficzny skonstruowany w *Statistics Canada*. Jego celem nadrzędnym jest analiza struktury zróżnicowania etniczno-kulturowego w Kanadzie, jednostkami analiz są osoby, a prognozy wyznacza się na okres 2006–2031. Badanie uwzględnia dynamikę oraz dywersyfikację przestrzenną według prowincji i terytoriów, m. in. w zakresie mniejszości etnicznych, religii, statusu pokoleniowego, języka narodowego, a także wykształcenia i zatrudnienia. Dane wejściowe pochodzą ze spisu powszechnego z 2006 r. rozszerzonego o informacje z niektórych rejestrów. Populacja wstępna to próba 20% (około 7 mln osób) zbiorowości generalnej, tzn. populacji Kanady. Baza mikrodanych zawiera zmienne dotyczące: płci, miejsca urodzenia, religii, grup mniejszościowych, statusu imigracyjnego, statusu pokoleniowego, języka narodowego, poziomu wykształcenia i zatrudnienia. Dodatkowe informacje pomocnicze to stan cywilny, rok i wiek w chwili imigracji, stwierdzony status Indian, liczba i wiek dzieci w gospodarstwie domowym. Zdarzenie i zmiany w charakterystykach zachodzą według zadanego prawdopodobieństwa z zastosowaniem metod Monte Carlo. Demosim skonstruowany jest z 11 modułów odzwierciedlających różne aspekty społeczno-demograficzne:

- płodność,
- umieralność,
- imigracja,
- emigracja,
- migracje wewnętrzne,
- najwyższy poziom wykształcenia,
- charakterystyka noworodków,
- stan cywilny,
- udział w rynku pracy,
- zmiana religii w trakcie ścieżki życia,
- dzieci opuszczają dom rodzinny.

Demosim jest zaprogramowany na platformie Modgen, podobnie jak inne modele skonstruowane przez *Statistics Canada*. Jego dokumentacja techniczna i metodologiczna jest szczegółowa. Model nie jest ogólnodostępny, ale wyniki prognoz i analiz są umieszczane na stronie Urzędu Statystycznego Kanady. Demosim nie zawiera modelowania ekonometrycznego, ale uwzględnia oszacowania nieekonomicznych modeli regresji dla danych społeczno-demograficznych [źródło: StatCan_a; StatCan_d; StatCan_e].

Tabela 2.3

Charakterystyka modeli społecznych i demograficznych

Nazwa modelu	Rodzaj modelu	Funkcje	Jednostki badania	Dane	Moduły	Zakres tematyczny
1	2	3	4	5	6	7
SOCSIM	dynamiczny	• symulacja pokrewieństwa i związków pomiędzy obiektami • wyznaczanie zmienności w populacji • analiza alternatywnych sposobów utrzymania stałej wielkości populacji i niewymierania rodów • wskazywanie długookresowych wzorców demograficznych • prognoza do 600 lat	• osoba	• preindustrialne spisy powszechne, m. in. populacji Anglii i Walii z 1751 r.,	brak	• społeczny • demograficzny
LifePaths	dynamiczny	• symulacja historii życia • symulacja efektów polityk społecznych • analiza zjawisk społecznych i demograficznych w czasie • prognozy demograficzne 1971–2026	• osoba • rodzina	• spisy ludności Kanady z 1911, 1921, 1971, 1981, 1991, 1996 • rejestr urodzeń 1921–1998 • rejestry zgonów 1801–1991 • ankiety: School Leavers Survey 1991–1995, Labour Force Survey i National Graduate Surveys, Family History Survey z 1984 i General Social Survey z 1995	• populacja i migracja • umieralność • wykształcenie podstawowe i średnie • wykształcenie wyższe • płodność • małżeństwa, związki nieformalne i rozwody • skład gospodarstwa domowego • zatrudnienie • urlop macierzyński • zarobki	• społeczny • demograficzny • ekonomiczny

1	2	3	4	5	6	7
Demosim	dynamiczny	• analiza zróżnicowania etniczno--kulturowego • prognozy demograficzne 2006–2031	• osoba	• spis ludności Kanady z 2006 • rejestry	• płodność • umieralność • imigracja • emigracja • migracje wewnętrzne • najwyższy poziom wykształcenia • charakterystyka noworodków • stan cywilny • udział w rynku pracy • zmiana religii w trakcie ścieżki życia • dzieci opuszczają dom rodzinny	• społeczny • demograficzny

Ź r ó d ł o: jak do tab. 2.1.

2.4.4. Porównanie modeli społecznych i demograficznych

W tab. 2.3 zestawiono charakterystyki wybranych modeli społecznych i demograficznych. Modele społeczne i demograficzne są dynamiczne, zazwyczaj wielomodułowe. Przeznaczono je do długookresowych analiz. Poszczególne bloki zarządzają konkretnymi cechami demograficznymi o znanych rozkładach i historycznych częstościach występowania. Startowe bazy mikrodanych i parametry pochodzą z licznych źródeł, głównie spisów powszechnych i rejestrów. Obiektami są zazwyczaj osoby. Ze względu na tematykę, brak w nich submodeli ekonometrycznych, jedynie Demosim zawiera proste modele regresji, a LifePaths bardzo podstawowe elementy ekonomiczne. Modele społeczne i demograficzne są jednorodne w odniesieniu do struktury i funkcji.

2.5. Modele systemu ochrony zdrowia

Wspominane już wcześniej modele DYNASIM, CORSIM i LifePaths zawierają elementy związane ze zdrowiem i systemem ochrony zdrowia. W szczególności dotyczy to występowania i stopnia niepełnosprawności. Istnieją także modele ukierunkowane wyłącznie na czynnik zdrowotny. Należą do nich modele czysto medyczne, ale zazwyczaj nie uwzględniają one bezpośrednio czynników ekonomicznych i społecznych, dlatego nie są rozważane w niniejszym przeglądzie modeli [Spielauer 2007: 35]. W ostatnich latach powstało kilka modeli z zakresu ekonomii zdrowia, które łączą elementy medycyny, ekonomii i badań społecznych.

2.5.1. Modele *Statistics Canada*

Statistics Canada jest liderem i pionierem w konstrukcji modeli mikrosymulacyjnych sektora ochrony zdrowia. Modele te nie traktują aspektu zdrowotnego i medycznego marginalnie, ale przeciwnie, jako główny przedmiot zainteresowania. Czynniki ekonomiczne i społeczne są postrzegane jako efekty stanu zdrowia populacji Kanady.

Modele mikrosymulacyjne *Statistics Canada* wśród determinantów stanu zdrowia uwzględniają czynniki ryzyka, indywidualną historię chorobową aktorów (tzn. osób) i strukturę demograficzną Kanady. W eksperymencie symulowane są ścieżki życia poszczególnych osób, a nowe obiekty dodaje się w trakcie trwania procesu, w celu zapewnienia realistycznej struktury populacji. Zdarzenia zaprogramowano w czasie ciągłym, a uzyskane wyniki są agregowane. Tak

skonstruowany model mikrosymulacyjny pozwala na otrzymanie realistycznych projekcji przyszłych wyników oraz testowanie efektów alternatywnych scenariuszy zmian polityk związanych z ochroną zdrowia.

Population Health Model **(POHEM)** to dynamiczny model mikrosymulacyjny sektora zdrowia. Służy on do projekcji ścieżki życia osób (model przypadku) uwzględniając najczęstsze zdarzenia społeczne i demograficzne, ze szczególnym uwzględnieniem chorób (np. choroby serca, osteoporoza, rak, cukrzyca) i czynników ryzyka chorobowego (rozpoczęcie i zaprzestanie palenia, zmiana wagi ciała itp.). Dane populacji startowej to reprezentatywna próba zbiorowości generalnej Kanady, zaś niezbędne parametry pochodzą z licznych rejestrów i baz danych o hospitalizacjach. Model umożliwia porównywanie efektów alternatywnych polityk związanych bezpośrednio lub pośrednio z systemem ochrony zdrowia. POHEM posłużył m. in. do analiz długookresowych kosztów bezpośrednich i pośrednich zachorowania na raka i leczenia.

Statistics Canada na zlecenie *Canadian Partnership Against Cancer (Kanadyjskie Partnerstwo Przeciw Rakowi)* skonstruował model mikrosymulacyjny ***Cancer Risk Management Platform***. Jest to nowoczesna platforma, której model mikrosymulacyjny zajmuje się ewaluacją strategii związanych z prewencją, odsiewem (ang. *screening*) i leczeniem najczęstszych nowotworów. Model pozwala na porównywanie dwudziestoletnich projekcji uwzględniających występowanie chorób, umieralność oraz bezpośrednie i pośrednie koszty.

Statistics Canada we współpracy z *Public Health Agency of Canada* skonstruował także dwa mikrosymulacyjne modele aktywności fizycznej. ***Physical Activity Simulation Static Model*** **(PASSM)** to model statyczny służący analizie codziennej aktywności fizycznej w systemie 24h. Uwzględnia on wpływ ćwiczeń na umieralność i oczekiwaną długość życia. **Dynamiczny model** bada aktywność fizyczną w cyklu życia oraz jej związek ze stanem zdrowia. Oba podejścia pozwalają na porównywanie efektów alternatywnych scenariuszy aktywności fizycznej na sytuację przyszłej populacji Kanady.

Obecnie trwają prace nad dynamicznym modelem chorób neurologicznych. Ma on uwzględniać występowanie i rozprzestrzenianie się schorzeń, ich wpływ na umieralność, oczekiwaną długość życia, niepełnosprawność, bezpośrednie i pośrednie koszty leczenia [źródło: StatCan_b].

2.5.2. Modele NATSEM (*University of Canberra*, Australia)

NATSEM (ang. *The National Centre for Social and Economic Modelling*) to centrum badawcze University of Canberra, Australia, będące jednym z wiodących ośrodków zajmujących się konstrukcją modeli mikrosymulacyjnych oraz powiązanych z nimi analiz ekonomiczno-społecznych. Wśród dorobku NAT-

SEM na szczególną uwagę zasługują modele dotyczące sektora zdrowia [źródło: Canberra_a].

Health&WealthMOD, tzn. **Model Zdrowia i Dobrotytu**, to mikrosymulacyjny model umożliwiający prowadzenie analiz ekonomicznych konsekwencji chorób u starszych pracowników. Model zaprojektowano w języku programowania SAS, tak, by współpracował z komputerami osobistymi i systemem Windows.

Populacja startowa aktualnej wersji Health&WealthMOD pochodzi z ankiety *Survey of Disability, Ageing and Carers* (SDAC) przeprowadzonej przez *Australian Bureau of Statistics* (ABS)[7] w 2003 r. dla osób w wieku 45–64 lat (wagi zapewniające reprezentatywność próby zostały zaktualizowane, by reprezentować strukturę populacji w roku 2009). Baza danych zawiera liczne informacje o stanie zdrowia i chorobach.

Elementy populacji z ankiety są syntetycznie[8] przypisywane do obiektów mikrosymulacyjnego modelu podatkowo-zasiłkowego **STINMOD**. Jest to wiodący statyczny mikrosymulacyjny model dla Australii, utrzymywany i rozbudowywany w NATSEM, służący ocenie efektów zmian wartości dochodów i ich rozkładów wynikających z reform podatkowych i monetarnych, ale także systemu opieki zdrowotnej. Jego obiektami są osoby i rodziny. Symulacje STINMOD mogą być prowadzone na okres 4 lat. Model dobrobytu oparty został na metodologii i technologii, skonstruowanego w NATSEM, dynamicznego modelu mikrosymulacyjnego **DYNAMOD**. Jest to złożone i wszechstronne narzędzie pozwalające na uwzględnienie sytuacji ekonomicznej osób w kolejnych miesiącach, tzn. wartość majątku netto, depozyty gotówkowe, plany emerytalne, własność miejsca zamieszkania i wynajem posesji.

Health&WealthMOD służy szacowaniu liczby osób w populacji Australii przechodzących na rentę i wcześniejszą emeryturę ze względu na przewlekłe choroby. Ponadto analizowane są związane z tym koszty pośrednie i bezpośrednie na poziomie indywidualnym (utrata dochodu, oszczędności) oraz centralnym (utrata wpływu z podatków, wypłata świadczeń i zasiłków). Możliwe jest także ocenianie konsekwencji finansowych wprowadzenia reform dotyczących leczenia i prewencji poszczególnych schorzeń [Schofieldl i in. 2009: 58–60; Schofieldl i in. 2011a: 32–33; Schofieldl i in. 2011b: 37–38].

CAREMOD to statyczny przestrzenny model mikrosymulacyjny, zaprojektowany w celu zbadania zapotrzebowania na opiekę wśród osób starszych w Nowej Południowej Walii (największy stan w Australii). Jest to pierwszy

[7] Urząd Statystyczny Australii.

[8] Dokładne przyporządkowanie danych z dwóch różnych ankiet jest zakazane przez prawo Australii. W celu przeprowadzenia najefektywniejszego przyporządkowania syntetycznego wykorzystano liniową regresję dla dochodu i innych zmiennych określających podobieństwo obiektów w obu populacjach startowych. Estymowano także równanie przychodów emerytalnych w zależności od wybranego planu emerytalnego.

model NATSEM, który uwzględnia potencjalny regionalny popyt na usługi opieki zdrowotnej dla osób powyżej 65 lat. Populacja startowa to populacja generalna Nowej Południowej Walii, klasyfikowana według regionów. Osoby opisane są charakterystykami ekonomiczno-społeczno-demograficzno-medycznymi, m. in. dochodem, strukturą rodziny i gospodarstwa domowego, płcią, wiekiem, wykształceniem, statusem niepełnosprawności, stanem zdrowia i dostępnością opieki w rodzinie. Dane pochodzą z różnych źródeł (spisów i ankiet). Ze względu na brak danych dotyczących zdrowia w danych powszechnych, wyznaczono wskaźniki dla małych obszarów geograficznych na podstawie losowych prób. Reprezentatywność zapewniono konstruując odpowiednie wagi [źródło: Canberra_b; Lymer i in. 2009: 27–31].

Model cukrzycy lub ekonomiczny model prewencji cukrzycy Typu 2 w Australii (ang. *Diabetes Model, Economic Modelling of the Prevention of Type 2 Diabetes in Australia*) to mikrosymulacyjny model skonstruowany we współpracy z *International Diabetes Institute*[9] i *University of Sydney*. Jest to model zapobiegania cukrzycy Typu 2 oraz pomiaru ekonomicznych efektów zachorowań. Symulacja dostarcza projekcji na okres 45 lat. Mierzone są potencjalne efekty, głównie koszty, programów mających zapobiegać rozprzestrzenianiu się choroby oraz stany zagrożenia zachorowaniem [źródło: Canberra_c; Thurecht i Brown 2011: 71–75].

MediSim (*Static Microsimulation Model of the Australian Pharmaceutical Benefits Scheme*[10]) to statyczny model mikrosymulacyjny, oparty na STINMOD, skonstruowany do prognozowania zmian rozkładu i wartości przychodów związanych z Planem Korzyści Farmaceutycznych w Australii (ang. *Australian Pharmaceutical Benefits Scheme*). Plan wprowadzono w celu zapewnienia dostępu do niezbędnych i kosztowoefektywnych leków na receptę, tzn. na czas, po przystępnej cenie. W praktyce oznacza on współpłacenie za prawie wszystkie produkty farmaceutyczne dostępne na rynku, przy jednoczesnym ograniczeniu wydatków ogółem za leki poszczególnych osób i rodzin (system „asekuracyjny"). MediSim modeluje Plan Korzyści Farmaceutycznych poprzez symulacje aktualnych i przyszłych wartości zapotrzebowania na leki i ich koszty, przy zmieniających się założeniach polityki zdrowotnej oraz samego Planu (tzn. zmiana: listy leków, współpłacenia, systemu „asekuracji", zachowań lekarzy przepisujących leki itp.). Możliwe jest także wykorzystanie MediSim do analizy rozkładu efektów tych modyfikacji na gospodarstwa domowe [źródło: Canberra_d; Abello i in. 2008].

[9] Międzynarodowy Instytut Cukrzycy.
[10] Statyczny Mikrosymulacyjny Model Australijskich Planów Korzyści Farmaceutycznych.

Tabela 2.4

Charakterystyka modeli systemów ochrony zdrowia

Nazwa modelu	Rodzaj modelu	Funkcje	Jednostki badania	Dane	Moduły	Zakres tematyczny	Dostępność informacji o modelach ekonometrycznych i metodach estymacji	Modele ekonometryczne i metody estymacji
1	2	3	4	5	6	7	8	9
POHEM	dynamiczny	• symulacja historii życia uwzględniając choroby i czynniki ryzyka chorobowego • symulacja efektów polityk związanych z ochroną zdrowia • analiza różnorodnych kosztów zachorowań na raka	• osoba	• spisy ludności • dane o hospitalizacji • reprezentatywne badania czasowe i przekrojowe • rejestry	brak	• ekonomiczny • medyczny	brak	
Cancer Risk Management Platform	dynamiczny	• ewaluacja strategii związanych z prewencją odsiewem (*screening*) i leczeniem najczęstszych nowotworów • porównywanie dwudziestoletnich projekcji uwzględniających występowanie chorób umieralność oraz bezpośrednie i pośrednie koszty	• osoba	brak informacji	brak informacji	• ekonomiczny • medyczny • społeczny	brak	
Modele aktywności fizycznej	statyczny	• analiza codziennej aktywności fizycznej w systemie 24h • uwzględnia wpływ ćwiczeń na umieralność i oczekiwaną długość życia	• osoba	brak informacji	brak informacji	• ekonomiczny • medyczny	brak	

1	2	3	4	5	6	7	8	9
Modele aktywności fizycznej	dynamiczny	• badanie aktywności fizycznej w cyklu życia oraz jej związku ze stanem zdrowia					brak	dyna-miczny
Health&Wealth MOD	dynamiczny	• szacowanie liczby osób w populacji Australii przechodzących na rentę i wcześniejszą emeryturę ze względu na przewlekłe choroby • analiza kosztów bezpośrednich i pośrednich (w skali mikro i makro) przechodzenia na rentę lub wcześniejszą emeryturę • analiza kosztów i zysków finansowych wynikających z reform dotyczących leczenia i prewencji chorób przewlekłych	• osoba, • rodzina,	• ankieta *Survey of Disability, Ageing and Carers* (SDAC)	• STIN-MOD • dobrobyt • choroby przewle-kłe	• ekonomicz-ny • medyczny • społeczny	ograniczona	1) Modele: – regresja liniowa
CAREMOD	statyczny	• zbadanie zapotrzebowania na opiekę wśród osób starszych w Nowej Południowej Walii • analiza popytu na usługi opieki zdrowotnej dla osób powyżej 65 lat w ujęciu regionalnym	• osoba	• spisy powszechne Nowej Południowej Walii • ankiety	brak informacji	• ekonomicz-ny • medyczny • społeczny	brak	
Model cukrzycy	dynamiczny	• zapobieganie cukrzycy Typu 2 • pomiar ekonomicznych efektów zachorowań na cukrzycę Typu 2 • symulacja na okres 45 lat	• osoba	brak informacji	brak informacji	• ekonomicz-ny • medyczny • społeczny	brak	
Medi Sim	statyczny	• prognozowanie zmian rozkładu i wartości przychodów związanych z Planem Korzyści Farmaceutycznych w Australii	brak informacji	brak informacji	brak informacji	• ekonomicz-ny • medyczny • społeczny	brak	

Ź r ó d ł o: jak do tab. 2.1.

2.5.3. Porównanie modeli systemów ochrony zdrowia

Mikrosymulacyjne modele sektora zdrowia są modelami wyspecjalizowanymi, dotyczącymi aspektów konkretnego schorzenia lub metody prewencyjnej. Jednostkami badania są zazwyczaj osoby. Funkcje modeli są zróżnicowane ze względu na tematykę i cele poszczególnych badań. Wszystkie modele łączą aspekt ekonomiczny (zazwyczaj analiza kosztów) i medyczny, a niektóre także społeczny. Ponieważ rozkład kosztów często jest nieznany, można podejrzewać, że w modelach występują submodele mikroekonometryczne, jednak dla większości z nich brak jest jakiejkolwiek informacji na temat aspektów ilościowych. Również opisy struktury blokowej i źródeł danych są niepełne (zob. tab. 2.4).

2.6. Podsumowanie

Zaprezentowany przegląd wskazuje na różnorodność zastosowań modeli mikrosymulacyjnych w badaniach ekonomicznych i społecznych. Jednocześnie pokazuje on także marginalne i nieadekwatne podejście do problemu konstrukcji, estymacji i weryfikacji submodeli mikroekonometrycznych.

Najstarsze modele mikrosymulacyjne, przede wszystkim z zakresu świadczeń emerytalnych i rentowych, traktowane są nadal przez wielu analityków jako narzędzia uniwersalne i ponadczasowe, pomimo licznych braków informacji. Modele te nie mają upublicznionej pełnej dokumentacji, brak jest szczegółowego opisu specyfikacji, estymacji i weryfikacji submodeli ekonometrycznych. Dostępne dane wskazują, że zarówno startowa baza mikrodanych jak i próba danych zastosowana do szacowania parametrów strukturalnych pochodzą sprzed kilkudziesięciu lat, a niektóre zastosowane narzędzia ekonometryczne i statystyczne są nieadekwatne do specyfiki danych indywidualnych.

W nowszych modelach zasiłkowo-podatkowych, wbrew promowanej polityce jawnych dokumentacji modeli mikrosymulacyjnych, większość aspektów konstrukcyjnych, w tym także ekonometrycznych, jest utajniona. Podobnie w przypadku modeli sektora zdrowia bardzo wiele informacji o strukturze modelu i źródłach danych jest niedostępnych.

Chociaż wiele modeli mikrosymulacyjnych zawiera elementy ekonometryczne, zastosowane metody estymacji nie zawsze są znane lub adekwatne do specyfiki mikrodanych, a konstrukcja modelu nie w każdym przypadku pozwala na wprowadzanie ewentualnych korekt. Jednak postęp ostatnich dekad w dziedzinie metodologii mikroekonometrycznej wskazuje, że dobór odpowiedniej postaci funkcyjnej modelu i metod estymacji silnie determinuje jakość oszacowań parametrów strukturalnych. Własności estymatorów parametrów oraz sto-

pień dopasowania modelu do danych empirycznych znajdują odzwierciedlenie w wynikach eksperymentów mikrosymulacyjnych, szczególnie w ich aspektach ekonomicznych, gdyż najczęściej są one modelowane. W licznych modelach nie tylko starszych, ale i najnowszych, brak jest jakiejkolwiek informacji o specyfikacji, estymacji i weryfikacji równań mikroekonometrycznych. Uniemożliwia to ocenę jakości wyników badań z zastosowaniem tych modeli mikrosymulacyjnych. Dlatego dobór postaci funkcyjnej i metody estymacji ma kluczowe znaczenie dla poprawności funkcjonowania mikrosymulacji.

Rozdział 3

MODELE MIKROEKONOMETRYCZNE W MIKROSYMULACJACH

3.1. Wprowadzenie

Mikrosymulacje umożliwiają prowadzenie badań ekonomiczno-społecznych poprzez prognozowanie i symulację wartości zmiennych opisujących jednostki badania. Model mikrosymulacyjny zapewnia projekcję wartości zmiennych deterministycznych oraz stochastycznych o znanych rozkładach, zaś model mikroekonometryczny pozwala na ich podstawie oszacować wartości cech o nieznanych rozkładach, ale znanych determinantach. Taki charakter ma wiele zmiennych ekonomicznych, np. dochody i wydatki. Uwzględnienie ich w mikrosymulacji wymaga wprowadzenia modelu ekonometrycznego, którego specyfikacja i jakość oszacowań parametrów determinują przydatność całego modelu mikrosymulacyjnego. Jednak istotnym ograniczeniem w konstrukcji submodeli mikroekonometrycznych jest zbiór dostępnych zmiennych egzogenicznych. Niemożliwe jest uwzględnienie zmiennych niezależnych, które nie są symulowane przez model mikrosymulacyjny ani nie pochodzą ze źródeł zewnętrznych. Może to powodować obniżenie jakości wyników prowadzonych badań. Tym ważniejsze staje się zagadnienie doboru odpowiedniej postaci funkcyjnej równań modelu mikroekonometrycznego i najlepszych możliwych (ze względu na własności estymatorów paramentów) metod estymacji, kompensujących potencjalne błędy. W procesie modelowania ekonometrycznego, ze względu na specyfikę mikrodanych, nieuzasadnione jest wykorzystanie modeli innych niż mikroekonometryczne. Często rozkłady zmiennych objaśnianych dla obiektów indywidualnych są dyskretne, ucięte i nieliniowe, przeważają małe wartości lub występują liczne zera. Nawet w modelach liniowych zmiennych ciągłych często pojawia się problem heteroskedastyczności i autokorelacji składnika losowego. Niezbędne jest więc wykorzystanie modelowania mikroekonometrycznego. Niestety w większości starszych modeli mikrosymulacyjnych zastosowane submodele ekonometryczne nie są zgodne z obecnie pojmowanym podejściem mikroekonometrycznym, co obniża jakość wyników mikrosymulacyjnych badań ekonomicznych,

zarówno dawnych jak i obecnych. W nowszych badaniach obserwuje się brak informacji o analizie rozkładu zmiennych oraz sposobie doboru adekwatnych metod estymacji parametrów modelu. Utrudnia to ocenę wiarygodności rezultatów przeprowadzonych eksperymentów.

Wśród analityków stosujących modele mikroekonometryczne nie ma zgodności co do zakresu metodologicznego mikroekonometrii. Niektórzy badacze postrzegają ten dział ekonometrii jako zarezerwowany wyłącznie dla zmiennych o rozkładach binarnych, wielonomialnych i uciętych. Takie podejście promują m. in. autorzy książki pod redakcją M. Gruszczyńskiego *Mikroekonometria. Modele i metody analiz danych indywidualnych*, prezentując modele i metody estymacji dla zmiennych jakościowych i dyskretnych. Podkreśla się w tym podręczniku duży wpływ mikroekonometrii na rozwój teorii ekonomii poprzez wprowadzenie nowych metod (1) analiz dużych zbiorów mikrodanych oraz (2) uwzględniających w badaniach ekonomicznych czynniki indywidualnych zachowań (wraz z ich heterogenicznością) podmiotów gospodarczych. Autorzy tej publikacji uzasadniają dobór metod ekonometrycznych zaprezentowanych w monografii specyfiką danych indywidualnych, m. in. zmiennymi jakościowymi, rozkładami ograniczonymi, heterogenicznością wartości cech i zachowań jednostek badania [Gruszczyński 2010a: 13–37; Wiśniowski i Owczarczuk 2010: 37–50].

Jednak nie wszystkie zmienne danych niezagregowanych mają rozkłady nieciągłe. J. W. Wiśniewski w *Mikroekonometrii* prezentuje liczne zastosowania liniowych i potęgowo-wykładniczych modeli ekonometrycznych dla małych przedsiębiorstw. Wskazuje na ich przydatność w analizie płynności finansowej, procesów produkcyjnych, wykorzystania zasobów pracy oraz popytu. Autor pominął jednak metodykę związaną z endogenicznymi zmiennymi jakościowymi i ograniczonymi [Wiśniewski 2009: 13–180].

Prof. J. Hozer, autor opracowania dotyczącego mikroekonometrii z 1993 r. [Hozer 1993], w przemówieniu inauguracyjnym konferencji *Mikroekonometria w teorii i praktyce 2011* postulował, aby nie ograniczać analiz mikroekonometrycznych ani do zmiennych o nietypowych rozkładach, ani do modeli liniowych zaczerpniętych z badań makroekonometrycznych. Sugerował, że mikroekonometria to dział ekonometrii zorientowany na analizy mikrodanych, niezależnie od ich źródła czy specyfiki rozkładu. W takim rozumieniu najszerszym opracowaniem metodologii mikroekonometrycznej jest pozycja Cameron i Trivedi [2009]. Zawiera ona bardzo szczegółowy i rozbudowany opis modeli, metod estymacji i weryfikacji dla danych przekrojowych oraz znacznie krótszy i mniej obszerny dla danych panelowych. W pozycji tej jedynie wspomniano o modelach wielorównaniowych, a w monografii pod red. M. Gruszczyńskiego całkowicie je pominięto.

Żadna z powyższych pozycji nie oferuje całościowego i syntetycznego zestawienia modeli i metod mikroekonometrycznych, które mogą być przydatne w konstrukcji modelu ekonometrycznego stanowiącego część modelu mikrosy-

mulacyjnego, przeznaczonego m. in. do prowadzenia badań ekonomiczno--społecznych. Celem tego rozdziału jest skonstruowanie właśnie takiego wykazu, który zawiera informacje o poszczególnych rodzajach modeli oraz najlepszych ze względu na własności estymatorów metodach estymacji wynikających ze specyfiki mikrodanych. Zaprezentowane w rozdziale podejście do modeli mikroekonometrycznych i metod estymacji stanowi kompleksowe i zbiorcze zestawienie, które powinno być pomocne w konstruowaniu modeli mikroekonometrycznych zawartych w modelach mikrosymulacyjnych, wykorzystywanych w badaniach ekonomiczno-społecznych. Klasyfikację przeprowadzono najpierw według rodzaju mikrodanych (przekrojowe i panelowe), następnie według struktury modelu (jednorównaniowe i wielorównaniowe, a wśród nich liniowe i nieliniowe). Podstawę literaturową stanowiły publikacje M. Gruszczyńskiego, J. W. Wiśniewskiego oraz A. C. Camerona i P. K. Trivedi'ego, rozszerzone o pozycje dotyczące modeli panelowych, m. in. B. Dańskiej, O. Torres-Reyna'go i F. Heiss oraz modeli wielorównaniowych J. B. Gajdy i B. H Baltagi'ego.

3.2. Specyfika mikrodanych

Specyfika mikrodanych determinuje specyfikację modelu mikroekonometrycznego i jakość oszacowań jego parametrów. Konstrukcja modelu ekonometrycznego i estymacja jego parametrów wpływają silnie na poprawność funkcjonowania zawierającego go modelu mikrosymulacyjnego. Ten zaś ma kluczowe znaczenie dla wyników eksperymentu, które stanowią podstawę do prowadzenia badań i analiz ekonomicznych i społecznych. Dlatego konieczne staje się zrozumienie specyfiki mikrodanych stanowiącej fundament doboru postaci funkcyjnej modelu mikroekonometrycznego i adekwatnych ze względu na własności estymatorów metod szacowania parametrów strukturalnych [Wiśniewski 2009: 8].

Dane indywidualne charakteryzują się innymi własnościami niż ich zagregowane odpowiedniki. Mikrodane należą do informacji empirycznych i mają zazwyczaj charakter przekrojowy (np. osoby, firmy, regiony). Można również spotkać bazy przekrojowo-czasowe gromadzące informacje o przekroju w kolejnych okresach, jednak niekoniecznie dotyczące tych samych obiektów. Niektóre bazy zawierają dane panelowe (ang. *micro panels*), czyli informacje o tych samych jednostkach w czasie. Zazwyczaj mikrodane grupowane są w krótkim panelu (ang. *short panel*), tzn. wiele obiektów obserwowanych w kilku okresach. Specyfikę mikrodanych można rozpatrywać w trzech wzajemnie współzależnych kategoriach: **rozkład**, **heterogeniczność** oraz **modelowanie**.

• Specyfika rozkładu mikrodanych •

Mikrodane, podobnie jak makro, mogą mieć charakter **ciągły** lub **dyskretny**. W przypadku zmiennych ciągłych o nieskończonym zbiorze wartości, a często także ograniczonych np. do liczb nieujemnych, uzasadnione jest wykorzystanie metod analiz zbliżonych do tych stosowanych w makroekonometrii.

Kiedy zmienna jest ograniczona do przedziału (np. opisuje prawdopodobieństwo [0;1]), liczby wariantów (binarna – płeć, wielomianowa uporządkowana – wykształcenie, lub nieuporządkowana – stan cywilny) czy liczb naturalnych (zmienna licznikowa – liczba dzieci), analiza jest utrudniona i konieczne staje się stosowanie metodologii mikroekonometrycznej. Nawet rozkłady ciągłe (również ograniczone do liczb nieujemnych) mogą stać się wyzwaniem dla analiz i modelowania w przypadku występowania częstych wartości zerowych (ang. *Frequent Zeros*) [Winkelmann i Boes 2006: 16]. Może to wynikać z krótkiego okresu objętego badaniem (tydzień, miesiąc), w którym zjawisko nie występowało [Lazaridis 2004: 416]. Należy podkreślić, że cechy statystyczne, których rozkłady uwzględniają powyższe utrudnienia, występują znacznie częściej dla danych niezagregowanych niż w skali makro [Gruszczyński 2010a: 17–19].

Rozkłady cech przed i po agregacji mogą być inne, co utrudnia modelowanie i analizy mikroekonometryczne. W wyniku agregacji wartości zmiennych są sumowane lub uśredniane. W rezultacie analogiczne zmienne:

– zmieniają definicje – ze zmiennej binarnej „płeć” na poziomie indywidualnym, na poziomie makro tworzy się „liczbę mężczyzn” i „liczbę kobiet”, czyli zmienne z przedziału liczb naturalnych,

– zmieniają skalę pomiarową lub dopuszczalny przedział wartości – „liczba dzieci na utrzymaniu w gospodarstwie domowym/rodzinie” na poziomie indywidualnym ma rozkład wielomianowy uporządkowany, po agregacji poprzez uśrednienie przyjmuje wartości rzeczywiste nieujemne (np. 2,5 dziecka w rodzinie).

Zmienne w bazach mikrodanych wykazują większe zróżnicowanie niż makrodane, co przyczynia się do ich niejednorodności. Duże znaczenie dla wysokiej dyspersji ma indywidualny czynnik losowy obiektu. Indywidualne preferencje lub charakterystyki, niekiedy skrajnie zróżnicowane, wpływają na występowanie wartości i wariantów, które można uznać za nietypowe. Stąd wariancja dla zmiennych mikroekonomicznych wykazuje wyższą wartość niż dla analogicznych zmiennych zagregowanych. Zazwyczaj mikrodane niosą większą ilość informacji o podmiotach i procesach gospodarczych niż informacje zagregowane. Nie oznacza to jednak redukcji czynnika losowego. Ostatnie badania pokazują, że nawet podejmowanie decyzji nie jest całkowicie zdeterminowane ustalonymi preferencjami, ale wykazuje element losowości [Loomes i in. 2002: 103].

Agregacja przyczynia się do wygładzania rozkładów zmiennych, stąd redukcja zmienności dla danych makroekonomicznych. Może to prowadzić do

powstawania heteroskedastyczności[1] składnika losowego modeli ekonometrycznych dla danych zagregowanych [Lee 2010: 19]. Modelowanie danych indywidualnych eliminuje to źródło niejednorodności wariancji błędów losowych. Nie oznacza jednak, że mikrodane są wolne od problemu heteroskedastyczności. W modelach mikroekonometrycznych jej przyczyną jest często heterogeniczność, zwłaszcza nieobjaśniona, cech lub zachowań elementów baz mikrodanych. Stąd konieczność uwzględnienia w modelach i analizach czynników indywidualnych obiektów badania [Cameron i Trivedi 2009: 5, 8].

Mikrodane zawierają także więcej informacji o własnościach i zależnościach ekonomicznych niż informacje zagregowane, co sprzyja rozwojowi mikrosymulacji i ich wykorzystaniu w badaniach ekonomiczno-społecznych. Zazwyczaj dane indywidualne są rzeczywiście nośnikiem niemodyfikowanej wiedzy o podmiotach, jednak pewne procesy, np. anonimizacja[2], wiążą się ze świadomym wprowadzeniem błędów do baz mikrodanych. Czym większa ingerencja w dane, tym mniej użyteczne się one stają. Dane indywidualne nie zawsze mogą być jednak udostępnione w postaci niezmienionej, np. ze względu na ich poufność. Procedury anonimizacji muszą zapobiegać możliwości identyfikacji obiektów badania. W niektórych krajach niezmieniane mikrodane udostępniane są analitykom pod groźbą bardzo wysokich kar w przypadku zastosowania ich w sposób nielegalny [Pohlmeier i Ronning 2008: 351–352]. Do najczęściej stosowanych sposobów anonimizacji danych zalicza się cztery metody.

1. Mikro-agregacja – tworzy się małe grupy jednostek i ich wartości zastępuje się średnią grupową.

2. Dodanie szumu (zaburzeń) – można stosować dla wszystkich lub wybranych zmiennych. Zmiana może mieć postać ustalonej wartości lub procentu.

3. „Czyszczenie" (usuwanie, kasowanie) danych – usuwanie wartości dla wszystkich lub wybranych zmiennych u obiektów, które można łatwo zidentyfikować [Pohlmeier i Ronning 2008: 353–355].

4. Multiple Imputation (wykorzystywana w wypełnianiu braków danych) – oznacza uzupełnianie każdego braku danych (lub w przypadku anonimizacji istniejących wartości) dwoma lub więcej wartościami losowymi o zadanym rozkładzie [Rubin 2004: 1–2].

Anonimizacja obniża jakość danych, może przyczyniać się do sztucznego zwiększania lub zmniejszania zmienności, jednak jest ona wymagana w celu zapewnienia respondentom gwarantowanej przez prawo, także polskie, poufności.

Specyfika rozkładu zmiennych dla danych indywidualnych, a w szczególności jego ograniczoność, rozproszenie i anonimizacja, wpływa na jakość oszacowań parametrów modeli mikroekonometrycznych, wyniki eksperymentów mikrosymulacyjnych, a zatem także na rezultaty opartych na nich badań ekonomiczno-społecznych.

---

[1] „(…) niejednorodność wariancji składników losowych w obrębie próby" – cyt. Welfe [2003: 116].

[2] Anonimizacja najogólniej to proces zapewniania anonimowości obiektów badania.

• Heterogeniczność jednostek •

W analizach statystycznych i ekonometrycznych mikrodanych wyzwaniem jest uwzględnienie **heterogeniczności elementów** badania. Może ona prowadzić do występowania heteroskedastyczności składnika losowego modeli mikroekonometrycznych, a w przypadku danych panelowych także jego autokorelacji. **Niejednorodność** obiektów dzieli się na **obserwowalną** i **nieobserwowalną**. Dane indywidualne zazwyczaj obciążone są heterogenicznością, której wpływ można minimalizować wprowadzając do badania elementy specyficzne dla obiektów poprzez ich historię i obserwowalne charakterystyki [Kluve i in. 1999: 87]. Zmienne zawarte w zbiorze mikrodanych uwzględniają *explicite* cechy indywidualne jak: wiek, płeć, wykształcenie, status ekonomiczno-społeczny, region itp. Wiele z tych cech uwzględnianych jest w modelach mikrosymulacyjnych oraz aktualizowanych w trakcie trwania symulacji (ich zmiany zachodzą deterministycznie lub znany jest ich rozkład). Zmienne te stanowią niejednorodność obserwowalną. Jednocześnie to zmienność ich wartości i wariantów stanowi jedną z przyczyn heterogeniczności.

Nieobserwowalna niejednorodność wiąże się z niedostrzeganymi, niemierzalnymi lub po prostu nieuwzględnionymi w badaniu cechami obiektów, np. motywacją, zdolnościami, inteligencją. Zignorowanie tej heterogeniczności i włączenie jej w całości do składnika losowego, choć praktykowane, wiąże się ze wzrostem nieobjaśnionej części zmienności cech endogenicznych modelu. Ponadto prowadzi do błędnej oceny wpływu poszczególnych czynników egzogenicznych uwzględnionych *explicite w* równaniu. Brak informacji o niedostrzegalnych lub niemierzalnych cechach jednostek uniemożliwia zbadanie kierunku i siły ich korelacji ze zmiennymi uwzględnionymi w analizie, a zatem zjawisko to staje się trudne do wychwycenia i przeciwdziałania [Cameron i Trivedi 2009: 8–9].

Do najczęściej stosowanych metod eliminacji heterogeniczności, szczególnie nieobserwowalnej, w procesie modelowania ekonometrycznego, należą:

– zignorowanie jej – wprowadzenie silnych założeń o braku korelacji między zmiennymi objaśniającymi a składnikiem losowym,

– wyznaczenie stałych efektów dla poszczególnych jednostek w próbie – przypisanie każdej jednostce sztucznej zmiennej lub wyrazu wolnego, odzwierciadlającego jej indywidualne własności (podeście to możliwe jest, gdy dane są panelowe, zbiór obiektów jest nieliczny, a liczba okresów jest większa niż liczba jednostek badania),

– wyznaczenie losowych efektów dla poszczególnych jednostek w próbie – nieobserwowalna heterogeniczność ma przynajmniej po części charakter przypadkowy (losowy); dekompozycja składnika losowego umożliwia wyizolowanie dodatkowej zmiennej stochastycznej opisującej czynniki indywidualne (podejście to można zastosować dla danych przekrojowych i panelowych).

Nieobserwowalna heterogeniczność, powszechna w mikrodanych, jest jedną z przyczyn występowania heteroskedastyczności składnika losowego w modelach mikroekonometrycznych, a zatem wpływa na wyniki badań ekonomicznych z zastosowaniem tych modeli i związanych z nimi mikrosymulacji.

• Specyfika w modelowaniu mikrodanych •

Modelowanie ekonometryczne z zastosowaniem mikrodanych jest silnie powiązane ze specyfiką rozkładu i heterogenicznością obiektów badania. Na poziomie mikrodanych możliwe jest wykrycie zależności niedostępnych po agregacji. Czynniki indywidualne, np. demograficzne, mogą mieć duży wpływ na sytuację ekonomiczną podmiotów. Jednak w wyniku agregacji i wiążącego się z nią „wyrównywania", estymatory parametrów modeli dla danych na poziomie makro tracą własność zgodności [Kalwij i in. 1998: 551]. Wspomniane we wcześniejszych podrozdziałach zmienność i heterogeniczność, przyczyniają się do obniżenia stopnia, w jakim model objaśnia wartości zmiennej endogenicznej. Stąd w modelach mikroekonometrycznych należy się więc spodziewać niższych niż w makroekonometrycznych wartości miar oceny statystycznej modelu, np. współczynnika determinacji, statystyki t-Studenta. Heterogeniczność obiektów i niejednorodność wariancji składnika losowego są konsekwencją stosowania mikrodanych i często nie ma możliwości ich zredukowania. Nieuzasadnione jest jednak zignorowane ich w procesie modelowania ekonometrycznego, ponieważ wpływa to niekorzystnie na jakość oszacowań parametrów modeli mikroekonometrycznych, wyniki mikrosymulacji oraz praktyczne wykorzystanie badań. Niezbędne jest uwzględnienie tych czynników w specyfikacji modelu, zarówno dobierając postać funkcji jak i zbiór zmiennych egzogenicznych. W rezultacie nie wszystkie metody estymacji parametrów modeli makroekonometrycznych mogą znaleźć zastosowanie w mikroekonometrii.

3.3. Estymacja przekrojowych modeli mikroekonometrycznych

Metody estymacji ekonometrycznej, a zatem także mikroekonometrycznej, zależą przede wszystkim od struktury modelu. W doborze procedury estymacji parametrów strukturalnych istotny jest typ modelu (jednorównaniowy czy wielorównaniowy) wraz z postacią zależności między równaniami. Kluczowe znaczenie w procesie podejmowania decyzji ma także postać funkcyjna równań. W przypadku mikrodanych należy również uwzględnić ich specyfikę tzn. heterogeniczność jednostek, rozkłady endogenicznych zmiennych jakościowych i ograniczonych oraz niesferyczność składnika losowego modeli. Jedynie uwzględnienie wszystkich powyższych informacji umożliwia optymalny dobór

metody estymacji. Własności estymatorów i stopień dopasowania modelu do danych empirycznych wpływa bezpośrednio na jakość powiązanych modeli mikrosymulacyjnych i wyniki eksperymentów mikrosymulacyjnych wykorzystywanych w badaniach ekonomicznych i społecznych.

Zaprezentowana w rozdziale metodologia nie jest zarezerwowana jedynie dla danych przekrojowych. Większość specyfikacji modeli i metod estymacji stosuje się również dla danych czasowych. Należy jednak zauważyć, że w przypadku mikroekonometrii rzadko ma się do czynienia z informacjami dotyczącymi jednego obiektu obserwowanego w czasie. Zazwyczaj badane są bardzo liczne jednostki, setki lub tysiące w jednym okresie lub w nielicznych punktach czasu (metodologię danych panelowych zaprezentowano w dalszej części pracy). Szczególnie w przypadku modeli mikroekonometrycznych będących submodelami modeli mikrosymulacyjnych, które rozważane są w niniejszym rozdziale, nie prowadzi się analiz danych czasowych. Dlatego metodologia dla danych jednowymiarowych będzie utożsamiana z danymi przekrojowymi.

Modele jednorównaniowe stanowią najliczniejszą grupę submodeli mikroekonometrycznych modeli mikrosymulacyjnych. Ze względów praktycznych analitycy preferują najprostsze specyfikacje modeli. Umożliwiają one łatwą estymację parametrów strukturalnych, weryfikację uzyskanych rezultatów i implementację do modelu mikrosymulacyjnego. Wybór modeli jednorównaniowych nie determinuje jednak doboru metody estymacji.

3.3.1. Estymatory paramentów strukturalnych jednorównaniowych modeli liniowych danych przekrojowych

Parametry jednorównaniowego modelu liniowego:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{3.1}$$

gdzie: $\mathbf{y}$ to wektor kolumnowy wartości zmiennej endogenicznej, $\mathbf{X}$ to macierz wartości zmiennych egzogenicznych, w której numer kolumny oznacza numer zmiennej, a wiersza – numer obserwacji, $\boldsymbol{\beta}$ to wektor kolumnowy rzeczywistych i nieznanych parametrów strukturalnych, a $\boldsymbol{\varepsilon}$ to wektor kolumnowy składników losowych [Wiśniewski 2009: 187–190], którego macierzą wariancji –kowariancji jest $\boldsymbol{\Omega} = \mathbf{D}^2(\varepsilon_i)$

mogą być szacowane popularną w modelowaniu ekonometrycznym metodą najmniejszych kwadratów (MNK, ang. *ordinary least squares*, OLS) tylko wtedy, kiedy macierzy $\boldsymbol{\Omega}$ jest sferyczna, tzn.:

$$\mathbf{\Omega} = \mathbf{D}^2(\varepsilon_i) = \sigma^2\mathbf{I} \tag{3.2}$$

gdzie: σ^2 to wartość wariancji składnika losowego stała dla wszystkich *i* obserwacji, zaś **I** to macierz jednostkowa o wymiarze równym liczbie obserwacji *N*, a *i* to numer obserwacji, *i* = 1,…, *N*[3].

Jednak w mikroekonometrii metoda ta nie jest zalecana, ponieważ struktura mikrodanych i powiązań pomiędzy zmiennymi często nie zapewnia spełniania założeń MNK, m. in. dotyczących sferyczności składnika losowego. Wiąże się to z wysokim prawdopodobieństwem występowania heteroskedastyczności składnika losowego, które sprawia, że estymator MNK traci efektywność, choć pozostaje nieobciążony i zgodny [Gruszczyński i Podgórska 1996: 76]. Niezbędne staje się wykorzystanie metod odpornych na niejednorodność wariancji. Metodologia mikroekonometryczna wskazuje kilka najczęściej stosowanych podejść, które umożliwiają uzyskanie nie tylko nieobciążonych i zgodnych, ale także efektywnych oszacowań parametrów modelu ekonometrycznego. Dotyczy to modeli, których zmienne endogeniczne są ciągłe i zależność pomiędzy zmienną objaśnianą a objaśniającymi jest liniowa względem parametrów lub sprowadzalna do liniowej. W modelach mikrosymulacyjnych taki charakter mają często cechy typowo ekonomiczne dotyczące m. in. dochodów, kosztów, wielkości produkcji lub służące np. badaniu sytuacji gospodarczej osób, gospodarstw domowych, firm lub regionów.

• Uogólniona Metoda Najmniejszych Kwadratów (UMNK) •

Jednym z podejść uzyskania odpornego na heteroskedastyczność estymatora jest wykorzystanie **uogólnionej metody najmniejszych kwadratów** (UMNK, ang. *generalised least squares*, GLS). Zapewnia ona większą efektywność estymatorów niż MNK. Najczęściej stosowane są dwa szczególne przypadki UMNK: **ważona metoda najmniejszych kwadratów** (WMNK, ang. *weighted least squares*, WLS) **lub osiągalna uogólniona metoda najmniejszych kwadratów** (OUMNK, ang. *feasible generalised least squares*, FGLS). Do wykorzystania UMNK niezbędna jest znajomość macierzy wariancji–kowariancji składnika losowego modelu.

W przypadku kiedy warunek (3.2) nie jest spełniony[4], pogwałcone jest jedno z założeń MNK [Welfe 2003: 65]. Powoduje to utratę efektywności i estymator

[3] Sferyczność oznacza, że składnik losowy jest homoskedastyczny (jego wariancja jest stała w próbie $\sigma^2 =$ const.) oraz nie występuje jego autokorelacja (poszczególne składniki losowe nie są skorelowane, tzn. kowariancje równe są 0, $cov(\varepsilon_i\varepsilon_j) = 0$, dla $i \neq j$, a $\mathbf{\Omega}$ jest macierzą diagonalną) – zob. Welfe [2003: 65].

[4] Może to wynikać z heteroskedastyczności ($\sigma^2 \neq$ const.) i/lub autokorelacji ($cov(\varepsilon_i\varepsilon_j) = 0$, dla $i \neq j$, (tzn. $\mathbf{\Omega}$ nie jest diagonalna) składnika losowego. W przypadku modeli mikroekonometrycznych dla danych przkrojowych niestałość wariancji w próbie jest bardziej prawdopodobna niż

metody najmniejszych kwadratów nie jest już najlepszym z liniowych, nieobciążonych estymatorów [Osińska i Kośko 2007: 161–163]. Jeżeli macierz $\mathbf{\Omega} = \mathbf{D}^2(\varepsilon_i)$ jest **nieosobliwa**[5] i **znana**, możliwe staje się jej wprowadzenie do modelu (3.1)[6].

W takim modelu składnik losowy ma rozkład o średniej równej 0 oraz sferycznej macierzy warinacji [Cameron i Trivedi 2009: 81–82]. Estymator UMNK parametrów równania (3.1) postaci:

$$\widehat{\boldsymbol{\beta}}_{UMNK} = (\mathbf{X}^T\mathbf{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{\Omega}^{-1}\mathbf{y} \tag{3.3}$$

jest BLUE (ang. *best linear unbiased estimator* – najlepszym liniowym nieobciążonym estymatorem) [Welfe 2003: 94–95]. W przypadku, kiedy macierz $\mathbf{\Omega}$ spełnia warunek (3.2) UMNK upraszcza się do MNK.

Ponieważ rzeczywista macierz wariancji–kowariancji składników losowych jest rzadko znana, możliwe jest wykorzystanie oszacowanej macierzy. W takim przypadku stosuje się **osiągalną uogólnioną metodę najmniejszych kwadratów:**

$$\widehat{\boldsymbol{\beta}}_{OUMNK} = (\mathbf{X}^T\widehat{\mathbf{\Omega}}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\widehat{\mathbf{\Omega}}^{-1}\mathbf{y} \tag{3.4}$$

gdzie: $\widehat{\mathbf{\Omega}}$ **to** estymator macierzy wariancji–kowariancji $\mathbf{\Omega}$. Zakłada się, że $\widehat{\mathbf{\Omega}} = \mathbf{\Omega}$.

Należy jednak zauważyć, że choć estymator OUMNK jest zgodny, to jego efektywność jest silnie skorelowana z efektywnością estymatora $\widehat{\mathbf{\Omega}}$ [Gajda 2004: 162–164]. Dlatego często rezygnuje się z osiągalnej UMNK na rzecz **ważonej metody najmniejszych kwadratów**[7], kiedy składnik losowy jest heteroskedastyczny, ale nie występuje jego autokorelacja. WMNK oznacza użycie estymatora $\widehat{\mathbf{\Sigma}}$ **„roboczej" macierzy wariancji–kowariancji** $\mathbf{\Sigma}$, o której nie zakłada się, że jest ona równa rzeczywistej macierzy $\mathbf{\Omega}$**,** ani że jest jej bezbłędnie dobranym estymatorem. WMNK nie opiera się więc na restrykcyjnych założeniach. We wzorze (3.3) macierz $\mathbf{\Omega}$ zostaje zastąpiona $\widehat{\mathbf{\Sigma}}$:

$$\widehat{\boldsymbol{\beta}}_{WMNK} = (\mathbf{X}^T\widehat{\mathbf{\Sigma}}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\widehat{\mathbf{\Sigma}}^{-1}\mathbf{y} \tag{3.5}$$

---

skorelowanie obserwacji, występujące częściej w modelach opartych na danych czasowych lub panelowych.

[5] Istnieje macierz odwrotna do $\mathbf{\Omega}$, tzn. $\mathbf{\Omega}^{-1}$.

[6] $\mathbf{\Omega}^{-1/2}\mathbf{y} = \mathbf{\Omega}^{-1/2}\mathbf{X}\boldsymbol{\beta} + \mathbf{\Omega}^{-1/2}\boldsymbol{\varepsilon}$, gdzie $\mathbf{\Omega} = \mathbf{\Omega}^{1/2}\mathbf{\Omega}^{1/2}$ stąd:

$$\mathbf{D}^2\left(\mathbf{\Omega}^{-1/2}\boldsymbol{\varepsilon}\right) = \mathbf{E}\left(\left(\mathbf{\Omega}^{-1/2}\boldsymbol{\varepsilon}\right)\left(\mathbf{\Omega}^{-1/2}\boldsymbol{\varepsilon}\right)^T\right) = \mathbf{\Omega}^{-1/2}\mathbf{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T)(\mathbf{\Omega}^{-1/2})^T = \mathbf{\Omega}^{-1/2}\sigma^2\mathbf{\Omega}(\mathbf{\Omega}^{-1/2})^T = \sigma^2\mathbf{I}$$

[7] Ważona Metoda Najmniejszych Kwadratów może u niektórych autorów oznaczać szczególny przypadek UMNK lub OUMNK, kiedy macierz wariancji–kowariancji jest diagonalna (a więc nie występuje autokorelacja składnika losowego). W tym opracowaniu będzie ona jednak oznaczała użycie estymatora $\widehat{\mathbf{\Sigma}}$ **„roboczej" macierzy wariancji–kowariancji** $\mathbf{\Sigma}$.

Estymator WMNK ma mniejszą efektywność niż UMNK, jednak jest też mniej restrykcyjny i łatwiejszy w aplikacji [Cameron i Trivedi 2009: 82–84]. Większość pakietów ekonometrycznych, oferujących możliwość szacowania parametrów modeli UMNK w rzeczywistości wyznacza estymatory OUMNK lub WMNK z wykorzystaniem macierz wariancji–kowariancji reszt modelu. UMNK i jej transformacje to najczęściej wykorzystywane metody estymacji parametrów modeli mikroekonometrycznych ze względu na prostotę, brak restrykcyjnych założeń i możliwość uwzględnienia specyfiki mikrodanych.

• Metoda zmiennych instrumentalnych (MZI) •

Estymując parametry modelu mikroekonometrycznego, istnieje zagrożenie wystąpienia endogenizacji zmiennych objaśniających. Wiąże się to z niespełnieniem założenia MNK o nielosowości regresorów oraz ich nieskorelowaniu ze składnikiem losowym [Osińska i Kośko 2007: 47]. Ryzyko to wynika m. in. z nieobjaśnionej heterogeniczności elementów baz mikrodanych. Zmienne nieobserwowalne lub niemierzalne, a zatem niewyspecyfikowane *explicite* w modelu znajdują odzwierciedlenie w składniku losowym. Jednak często są one skorelowane ze zmiennymi egzogenicznymi modelu, czasami wpływając na kształtowanie się ich wartości – powoduje to ednogenizację zmiennych objaśniających. W takiej sytuacji estymator parametrów MNK traci zgodność. W rezultacie oszacowania parametrów nie odzwierciedlają siły i kierunku relacji przyczynowej pomiędzy zmiennymi modelu. Niemożliwe staje się wyizolowanie wpływu poszczególnych regresorów na zmienną endogeniczną. Metoda zmiennych instrumentalnych (MZI, ang. *instrumental variables*, IV) pozwala na uzyskanie zgodnych estymatorów parametrów strukturalnych równania (3.1) [Cameron i Trivedi 2009: 95–96].

Jeżeli estymator MNK parametrów równania jest niezgodny w wyniku endogenizacji (wszystkich lub niektórych) zmiennych egzogenicznych, niezbędne jest znalezienie dla każdej z nich zmiennej instrumentalnej (instrumentu) *z* spełniającej dwa warunki:

- trafności instrumentu:

$$cov(z_i x_i) \neq 0^{8} \tag{3.6}$$

- egzogeniczności instrumentu:

$$cov(z_i \varepsilon_i) = 0^{9}. \tag{3.7}$$

[8] Warunek trafności instrumentu oznacza, że zmienna instrumentalna musi być powiązana ze zmienną objaśniającą. Jeżeli zależność jest niska to z stanowi słaby instrument, a gdy (3.6) nie jest spełnione – nietrafny instrument.

[9] Cecha całkowicie egzogeniczna oznacza, że z jest powiązana ze zróżnicowaniem **x**, ale nie bezpośrednio z **y**. Skoro zmienna z nie jest skorelowana ze zmienną endogeniczną, to jej pominię-

Konieczne jest określenie co najmniej jednego instrumentu dla każdej zmiennej objaśniającej[10]. MZI stosuje się kiedy liczba zmiennych instrumentalnych jest równa liczbie zmiennych objaśniających[11]. Macierz zmiennych instrumentalnych (uporządkowana zgodnie z kolejnością odpowiadających zmiennych objaśniających macierzy **X**) oznacza się **Z** [Wooldridge 2002: 83–84]. Estymator MZI ma postać [Gajda 2004: 165–166]:

$$\widehat{\boldsymbol{\beta}}_{MZI} = (\mathbf{Z}^T\mathbf{X})^{-1}\mathbf{Z}^T\mathbf{y} \tag{3.8}$$

W przypadku kiedy liczba instrumentów jest większa niż liczba regresorów i jej redukcja oznaczałaby rezygnację z istotnych dla badanej zależności informacji, możliwe jest wykorzystanie transformacji MZI, tzn. **podwójnej metody najmniejszych kwadratów** (2MNK, ang. *two-stage least squares*, 2SLS). Macierz **Z** nie ma tego samego wymiaru co **X**. Liczba kolumn (zmiennych) macierzy **Z** jest większa niż macierzy **X**, ale liczba wierszy (obserwacji) jest w obu przypadkach równa. Estymator 2MNK sprowadza się do dwukrotnej estymacji metodą najmniejszych kwadratów, w pierwszym kroku MNK dla modelu postaci:

$$\mathbf{Z}^T\mathbf{y} = \mathbf{Z}^T\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}^T\boldsymbol{\varepsilon} \tag{3.9}$$

a następnie UMNK[12] dla:

$$\boldsymbol{\Omega} = \mathbf{Z}^T\mathbf{Z} \tag{3.10}$$

stąd:

$$\widehat{\boldsymbol{\beta}}_{2MNK} = [\mathbf{X}^T\mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1}\boldsymbol{Z}^T\mathbf{X}]^{-1}\mathbf{X}^T\mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1}\boldsymbol{Z}^T\mathbf{y}] \tag{3.11}$$

[zob. Gajda 2004: 168–169].

Estymator (3.11) w przypadku heteroskedastyczności składnika losowego może tracić efektywność.

Należy podkreślić, że własności estymatorów metody zmiennych instrumentalnych są asymptotyczne, zatem zaleca się stosowanie go w przypadku

---

cie w wyjściowym modelu (3.1) nie przyczynia się do zwiększenia składnika losowego, a więc jest z nim nieskorelowana. Kiedy nie zachodzi (3.7) to *z* jest instrumentem błędnym.

[10] Kiedy liczba instrumentów mniejsza jest od liczby regresorów, niemożliwe jest skorzystanie z MZI lub jej transformacji.

[11] Jeżeli zmienna egzogeniczna sama spełnia warunek (3.7), to sama może być swoim instrumentem. Macierz $\mathbf{X} = [\mathbf{X}_1\ \mathbf{X}_2]$, gdzie $\mathbf{X}_1$ to zmienne objaśniające zendogenizowane, a $\mathbf{X}_2$ to zmienne objaśniane całkowiecie egzogeniczne. Analogicznie $\mathbf{Z} = [\mathbf{Z}_1\ \mathbf{Z}_2]$, $\mathbf{Z}_1$ to instrumenty dla $\mathbf{X}_1$, a $\mathbf{Z}_2$ to instrumenty dla $\mathbf{X}_2$. Ponieważ $\mathbf{X}_2$ są egzogeniczne, zatem mogą być wysłanymi zmiennymi instrumentalnymi, tzn. $\mathbf{Z}_2 = \mathbf{X}_2$, stąd $\mathbf{Z} = [\mathbf{Z}_1\ \mathbf{X}_2]$.

[12] Jest to UMNK, nie zaś OUMNK lub WMNK, ponieważ macierz **Z**, a zatem i $\boldsymbol{\Omega}$, są znane.

dużych prób. Ponadto warunku (3.7) nie można zweryfikować statystycznie, a (3.6) często ma arbitralny charakter. Niespełnienie warunków może prowadzić do niezgodności estymatora.

3.3.2. Estymatory paramentów strukturalnych jednorównaniowych modeli nieliniowych funkcji różniczkowalnych danych przekrojowych

Zależności mikroekonomiczne często wykazują się nieliniową postacią funkcyjną. Może to oznaczać inną niż liniowa funkcję ciągła, ale także rozkłady skokowe, ucięte, binarne, wielonominalne. Nieliniowość może też być własnością wtórną, wynikającą np. z anionimizacji danych. W mikrosymulacjach modele mikroekonometryczne stosowane są do szacowania wartości zmiennych o nieznanych rozkładach lub ich parametrach. Większość z tych cech ma charakter nieliniowy i dobór odpowiedniej postaci modelu i metod estymacji wpływa na wyniki prowadzonych eksperymentów mikrosymulacyjnych.

Zazwyczaj próby mikroekonomiczne są duże, co pozwala na stosowanie Praw Wielkich Liczb[13]. Większość estymatorów nieliniowych jest zgodna i asymptotycznie normalna. Wśród metod szacowania parametrów strukturalnych modeli dużą popularnością cieszą się estymatory ekstremum. Jest to bardzo ogólna klasa estymatorów opierająca się na optymalizacji zadanej funkcji celu. Do tej grupy metod estymacji należy zarówno **metoda największej wiarygodności** jak i **nieliniowa metoda najmniejszych kwadratów**.

• Metoda największej wiarygodności (MNW) •

Estymator **metody największej wiarygodności** (MNW, ang. *maximum likelihood*, ML)[14] to najefektywniejszy z estymatorów zgodnych o rozkładzie asymptotycznie normalnym. Zakłada on wybór takich oszacowań $\widehat{\boldsymbol{\theta}}$ rzeczywistych parametrów równania $\boldsymbol{\theta}$[15], które maksymalizują prawdopodobieństwo wystąpienia empirycznych wartości z próby.

Łączna funkcja gęstości (dla zmiennych ciągłych) lub prawdopodobieństwa (dla zmiennych dyskretnych) $f(\mathbf{y}, \mathbf{X}|\boldsymbol{\theta})$[16] nazywana jest funkcją wiarygodności:

---

[13] Szerzej na ten temat np. Domański [2001: 94]; Krysicki i in. [2002: 236–237].

[14] Ponieważ metoda dotyczy raczej warunkowej funkcji wiarygodności (gęstości) pełna nazwa powinna brzmieć warunkowa metoda największej wiarygodności (ang. *conditional ML*), ale przyjęło się upraszczać nazwę; Cameron i Trivedi [2009: 139–141].

[15] Ponieważ w zależności od rozkładu przyjmuje się różne oznaczenia parametrów: β, p, λ, tu przyjęto ogólne oznaczenie $\boldsymbol{\theta}$.

[16] Funkcja ta traktowana jest jako funkcja parametrów $\boldsymbol{\theta}$, przy ustalonych wartościach danych ($\mathbf{y}$, $\mathbf{X}$).

$$L_N(\boldsymbol{\theta}) = f(\mathbf{y}, \mathbf{X}|\boldsymbol{\theta}) = f(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta}) f(\mathbf{X}|\boldsymbol{\theta}) \tag{3.12}$$

gdzie: N to liczba obserwacji, a $f(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta})$ i $f(\mathbf{X}|\boldsymbol{\theta})$ to warunkowe funkcje wiarygodności **y** i **X**, odpowiednio.

Zazwyczaj jednak estymator wyznaczany jest na podstawie funkcji:

$$L_N(\boldsymbol{\theta}) = f(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta}) \tag{3.13}$$

Takie podejście jest uzasadnione tylko wtedy, gdy dobór zmiennych objaśniających do $f(\mathbf{y}|\mathbf{X})$ i $f(\mathbf{X})$ wyklucza współliniowość poprzez zastosowanie rozłącznych zbiorów zmiennych egzogenicznych[17].

Dane przekrojowe dla obserwacji $(y_i, \mathbf{x}_i)$[18] o warunkowej funkcji gęstości lub prawdopodobieństwa $f\left(y_i \middle| \mathbf{x}_{i,}, \boldsymbol{\theta}\right)$ są niezależne względem i, i=1,…, N, zatem w tym przypadku łączna (warunkowa) funkcja gęstości ma postać:

$$L_N(\boldsymbol{\theta}) = f(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta}) = \prod_{i=1}^{N} f(y_i|\mathbf{x}_i, \boldsymbol{\theta}) \tag{3.14}$$

Maksymalizacja funkcji wiarygodności (3.14) lub jej zlogarytmowanej wersji[19] prowadzi do wyznaczenia estymatora największej wiarygodności:

$$\frac{1}{N}\frac{\partial ln L_N(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{1}{N}\sum_{i=1}^{N} \frac{\partial ln\, f(y_i|\mathbf{x}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0}. \tag{3.15}$$

Estymator traci zgodność, jeżeli funkcja gęstości lub prawdopodobieństwa jest błędnie wyspecyfikowana [Cameron i Trivedi 2009: 139–140].

MNW stosowana jest w modelowaniu zmiennych o rozkładach ciągłych nieliniowych oraz skokowych, uciętych, binarnych i wielonominalnych.

• Nieliniowa metoda najmniejszych kwadratów (NMNK) •

Jeżeli dany jest nieliniowy model mikroekonometryczny z addytywnymi zakłóceniami:

$$\mathbf{y} = f(\mathbf{X}, \boldsymbol{\theta}) + \boldsymbol{\varepsilon} \tag{3.16}$$

---

[17] W przypadku wystąpienia problemu endogenizacji niezbędne staje się wyznaczenie pełnej łącznej funkcji gęstości.

[18] $\mathbf{x}_i$ to wektor wartości zmiennych egzogenicznych dla i-tej obserwacji.

[19] $Q_N(\boldsymbol{\theta}) = \frac{ln L_N(\boldsymbol{\theta})}{N} = \frac{1}{N}\sum_{i=1}^{N} ln\, f(y_i|\mathbf{x}_i, \boldsymbol{\theta})$

gdy: $f(\mathbf{X},\boldsymbol{\theta})$ jest funkcją **nieliniową**[20] i **różniczkowalną** względem parametru $\boldsymbol{\theta}$, to możliwe jest wykorzystanie uogólnionej wersji MNK – **nieliniowej metody najmniejszych kwadratów** (NMNK, ang. *nonlinear least squares*, NLS). Podobnie jak klasyczna wersja ma na celu minimalizację sumy kwadratów reszt $\mathbf{e}$[21], co czyni z niej metodę estymacji ekstremum. Poszukiwane są takie oszacowania $\widehat{\boldsymbol{\theta}}$ rzeczywistych parametrów $\boldsymbol{\theta}$, które spełniają warunek [Wiśniowski i Owczarczuk 2010: 43; Gajda 2004: 115–116]:

$$\begin{aligned}
\frac{\partial \mathbf{e}^T\mathbf{e}}{\partial\widehat{\boldsymbol{\theta}}} &= \frac{\partial\left(\mathbf{y}-f(\mathbf{X},\widehat{\boldsymbol{\theta}})\right)^{\boldsymbol{T}}\left(\mathbf{y}-f(\mathbf{X},\widehat{\boldsymbol{\theta}})\right)}{\partial\widehat{\boldsymbol{\theta}}} = \\
&= \frac{\partial\left[(\mathbf{y}^T\mathbf{y})+\left(f(\mathbf{X},\widehat{\boldsymbol{\theta}})\right)^T\left(f(\mathbf{X},\widehat{\boldsymbol{\theta}})\right)-2\mathbf{y}^T f(\mathbf{X},\widehat{\boldsymbol{\theta}})\right]}{\partial\widehat{\boldsymbol{\theta}}} = \\
&= -2\left[\frac{\partial f(\mathbf{X},\widehat{\boldsymbol{\theta}})}{\partial\widehat{\boldsymbol{\theta}}}\right]^T\mathbf{y}+ +2\left[\frac{\partial f(\mathbf{X},\widehat{\boldsymbol{\theta}})}{\partial\widehat{\boldsymbol{\theta}}}\right]^T f(\mathbf{X},\widehat{\boldsymbol{\theta}}) = \\
&= -2\left[\frac{\partial f(\mathbf{X},\widehat{\boldsymbol{\theta}})}{\partial\widehat{\boldsymbol{\theta}}}\right]^T\left(\mathbf{y}-f(\mathbf{X},\widehat{\boldsymbol{\theta}})\right) = \mathbf{0}
\end{aligned} \tag{3.17}$$

Układ równań (3.17) nie ma skończonego rozwiązania. Konieczne jest użycie metody iteracyjnej, np. algorytmu Gaussa–Newtona i przyjęcie wartości startowych. W każdym kolejnym kroku wyznacza się nowe parametry:

$$\widehat{\boldsymbol{\theta}}_{k+1} = \widehat{\boldsymbol{\theta}}_k + \left(\mathbf{J}_k^T\mathbf{J}_{\boldsymbol{k}}\right)^{-1}\mathbf{J}_k^T\mathbf{e}_k \tag{3.18}$$

gdzie: k oznacza numer kroku (iteracji), J_k to Jakobian[22] (macierz pierwszych pochodnych $f(\mathrm{X},\hat{\theta})$ po $\hat{\theta}$ w k-tej iteracji.

---

[20] Jeżeli $f(\mathbf{X},\boldsymbol{\theta})$ jest funkcją liniową to $f(\mathbf{X},\boldsymbol{\theta}) = \mathbf{X}\boldsymbol{\theta}$, zatem podstawiając do wzoru (3.20) $f(\mathbf{X},\widehat{\boldsymbol{\theta}}) = \mathbf{X}\widehat{\boldsymbol{\theta}}$ otrzymuje się:

$$-2\left[\frac{\partial\mathbf{X}\widehat{\boldsymbol{\theta}}}{\partial\widehat{\boldsymbol{\theta}}}\right]^T\left(\mathbf{y}-\mathbf{X}\widehat{\boldsymbol{\theta}}\right) = -2\mathbf{X}^T\mathbf{y}+2\mathbf{X}^T\mathbf{X}\widehat{\boldsymbol{\theta}} = \mathbf{0},$$

a po przekształceniu:

$$\mathbf{X}^T\mathbf{X}\widehat{\boldsymbol{\theta}} = \mathbf{X}^T\mathbf{y}, \text{ a następnie } \widehat{\boldsymbol{\theta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}.$$

Układ równań (3.20) ma więc rozwiązanie, którym jest estymator MNK, dla $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\beta}}$.

[21] $\mathbf{e}^T\mathbf{e} = \left(\mathbf{y}-f(\mathbf{X},\widehat{\boldsymbol{\theta}})\right)^{\boldsymbol{T}}\left(\mathbf{y}-f(\mathbf{X},\widehat{\boldsymbol{\theta}})\right)$

[22] Szerzej na ten temat Krysicki i Włodarski [2002: 78–84].

Iteracje powtarza się aż do osiągnięcia zbieżności. Metoda jest prosta obliczeniowo i może dać estymator zgodny i asymptotycznie normalny[23].

Ze względu na występującą w modelach mikroekonometrycznych danych przekrojowych heteroskedastyczność składnika losowego, efektywniejsza niż NMNK (choć mniej efektywna niż MNW) jest **osiągalna uogólniona nieliniowa metoda najmniejszych kwadratów** (OUNMNK, ang. *feasible generalized nonlinear least squares*, FGNLS). W OUNMNK minimalizuje się nie sumę kwadratów reszt, lecz[24]:

$$\mathbf{e}^T \widehat{\boldsymbol{\Omega}}^{-1}\, \mathbf{e} = \left(\mathbf{y} - f\left(\mathbf{X}, \widehat{\boldsymbol{\theta}}\right)\right)^{\boldsymbol{T}} \widehat{\boldsymbol{\Omega}}^{-1} \left(\mathbf{y} - f\left(\mathbf{X}, \widehat{\boldsymbol{\theta}}\right)\right) \tag{3.19}$$

Estymator OUNMNK jest estymatorem najefektywniejszym, pod warunkiem, że $\widehat{\boldsymbol{\Omega}}$ jest dobrze dobranym estymatorem $\boldsymbol{\Omega}$. W innym przypadku kompromisem staje się **ważona nieliniowa metoda najmniejszych kwadratów** (WNMNK, ang. *weighted nonlinear least squares*, WNLS). We wzorze (3.19) macierz $\widehat{\boldsymbol{\Omega}}$ zastępuje się estymatorem $\widehat{\boldsymbol{\Sigma}}$ macierzy roboczej wariancji błędów, $\boldsymbol{\Sigma} \neq \boldsymbol{\Omega}$**:**

$$\mathbf{e}^T \widehat{\boldsymbol{\Sigma}}^{-1}\, \mathbf{e} = \left(\mathbf{y} - f\left(\mathbf{X}, \widehat{\boldsymbol{\theta}}\right)\right)^{\boldsymbol{T}} \widehat{\boldsymbol{\Sigma}}^{-1} \left(\mathbf{y} - f\left(\mathbf{X}, \widehat{\boldsymbol{\theta}}\right)\right) \tag{3.20}$$

Efektywność zgodnego i asymptotycznie normalnego estymatora WNMNK znajduje się pomiędzy efektywnością NMNK i OUNMNK.

NMNK, OUNMNK i WNMNK stosuje się, gdy rozkład zmiennej endogenicznej jest ciągły i różniczkowalny.

3.3.3. Estymatory paramentów strukturalnych jednorównaniowych wybranych modeli nieliniowych funkcji nieróżniczkowalnych

Bazy mikrodanych często zawierają zmienne o rozkładach nieciągłych, np. skokowych, uciętych, wyrażone na skali nominalnej i porządkowej. Rozkłady tych zmiennych są z reguły znane, ale nie ma informacji o parametrach tych rozkładów. Dlatego w eksperymencie mikrosymulacyjnym cechy te nie mogą być aktualizowane z zadanym prawdopodobieństwem. Konieczne jest szacowanie ich wartości na podstawie modelu regresji, który nie może być traktowany

[23] Nie ma jednak gwarancji, że zbieżność zostanie osiągnięta, istnieje również możliwość występowania niejednoznacznych rozwiązań. Ponadto metoda jest wrażliwa na wybór wartości startowych – zob. Wiśniowski i Owczarczuk [2010: 43–44]; Gajda [2004: 116].

[24] Zakłada się, że $\widehat{\boldsymbol{\Omega}} = \boldsymbol{\Omega}$ i jeżeli warunek jest spełniony to estymator jest zgodny i asymptotycznie normalny. Utrudnieniem jest konieczność wyznaczenia macierzy $\widehat{\boldsymbol{\Omega}}^{-1}$, szczególnie dla dużych *N*. W praktyce $\widehat{\boldsymbol{\Omega}}$ ma często określoną strukturę, np. diagonalną, co prowadzi do możliwości wyznaczenia analitycznego rozwiązania $\widehat{\boldsymbol{\Omega}}^{-1}$ – zob. Cameron i Trivedi [2009: 155–156].

jako nieliniowy postaci (3.16), ponieważ funkcja opisująca zależność jest nieróżniczkowalna. W takim przypadku konieczne jest dobranie postaci funkcyjnej odzwierciedlającej empiryczny rozkład zmiennej endogenicznej i zastosowanie adekwatnej metody estymacji.

• Modele zmiennej dwumianowej •

W przypadku, gdy zmienna objaśniana przyjmuje jedynie dwie wzajemnie wykluczające i dopełniające się wartości lub warianty, niezbędne jest wykorzystanie modeli zmiennej dwumianowej (dychotomicznej, binarnej) [Wiśniewski 2009: 10]. W istniejących modelach mikrosymulacyjnych zmienne binarne stanowią najczęściej modelowane ekonometrycznie cechy. W przeciwieństwie do innych modeli mikroekonometrycznych, występują nie tylko w mikrosymulacjach zawierających elementy ekonomiczne, ale także w czysto społecznych modelach mikrosymulacyjnych. Szacowanie cechy binarnej oznacza estymowanie prawdopodobieństw występowania danego zjawiska, przynależności do grupy, podjęcia danej decyzji. Często analiza zmiennych dwumianowych nie jest celem badania ekonomiczno-społecznego opartego na mikrosymulacji. Ocenie podlegają raczej efekty wystąpienia danego wariantu cechy, np. skutki zdrowotne palenia, a nie sam wariant cechy 1 – pali, 0 – nie pali.

W modelu zmiennej binarnej rozważana jest zmienna *Y* z zerojedynkowego rozkładu Bernoulliego:

$$f(y_i) = p_i^{y_i}(1-p_i)^{1-y_i}, \text{dla } y_i = 0{,}1\ , \tag{3.21}$$

gdzie: $P(y_i = 1) = p_i$ a $P(y_i = 0) = 1 - p_i$.

Jeżeli $y_i = 1\ to\ f(y_i) = p_i$, w przeciwnym przypadku $f(y_i) = 1 - p_i$. Wartość oczekiwana wynosi $E(y_i) = p_i$, a wariancja $D^2(y_i) = p_i(1-p_i)$.

W modelach ekonometrycznych dychotomicznych modeluje się prawdopodobieństwo p_i jako funkcję zmiennych w macierzy **X** przy parametrach **β**:

$$p_i = f(\mathbf{x}_i, \boldsymbol{\beta}) \tag{3.22}$$

Najczęściej stosowane modele binarne to: **liniowy model prawdopodobieństwa (LPM), model logitowy i probitowy**. Różnią się one postacią funkcji (3.22). Wybór pomiędzy modelem LPM, logit i probit często jest arbitralny lub zależy od stopnia dopasowania modelu do danych empirycznych. Model logit wydaje się być najczęściej wykorzystywany, ponieważ zapewnia spełnienie warunku prawdopodobieństwa: $p_i \in \langle 0{,}1 \rangle$ oraz wygodną interpretację [Cameron i Trivedi 2009: 471–473].

Liniowy model prawdopodobieństwa (LPM)

Liniowy model prawdopodobieństwa (ang. *linear probability model*, LPM) zakłada, że funkcja (3.22) ma postać liniową, analogicznie do modelu (3.1). Zatem:

$$p_i = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i \tag{3.23}$$

Model LPM jest atrakcyjny ze względu na prostotę estymacji i interpretacji. Jednak nie ogranicza on wartości prawdopodobieństwa do przedziału $\langle 0,1 \rangle$. Istnieje możliwość wystąpienia heterosedastyczności składnika losowego, a przy jednoczesnym braku jego autokorelacji[25] parametry powinny być estymowane OUMNK [Winkelmann i Boes 2006: 98–99; Gruszczyński 2010: 58–59].

Model logitowy

W modelu logitowym funkcja (3.22) jest dystrybuantą rozkładu logistycznego:

$$p_i = f(\mathbf{x}_i, \boldsymbol{\beta}) = \Lambda(\mathbf{x}_i, \boldsymbol{\beta}) = \frac{exp\left(\mathbf{x}_i^T \boldsymbol{\beta}\right)}{1 + exp\left(\mathbf{x}_i^T \boldsymbol{\beta}\right)} = \frac{1}{1 + exp\left(-\mathbf{x}_i^T \boldsymbol{\beta}\right)} \tag{3.24}$$

która posiada funkcję odwrotną[26], pozwalającą modelować prawdopodobieństwo w postaci modelu liniowego:

$$logit(p_i) = \mathbf{x}_i^T \boldsymbol{\beta} \tag{3.25}$$

gdzie: $logit(p_i) = ln \frac{p_i}{1-p_i}$ nazywany jest **logitem.** Równanie (3.25) zapewnia, że $p_i \in \langle 0,1 \rangle$. Do estymacji parametrów modelu logitowego wykorzystuje się metodę największej wiarygodności (MNW). Łączna funkcja prawdopodobieństwa ze wzoru (3.14) ma postać:

$$L_N = \prod_{i=1}^{N} p_i^{y_i} (1 - p_i)^{1-y_i} \tag{3.26}$$

[25] Składnik losowy modeli mikroekonometrycznych dla danych przekrojowych rzadko charakteryzuje się występowaniem autokorelacji. Heteroskedastyczność stanowi dużo częstszy powód niesferyczności błędu losowego.

[26] $ln \frac{p_i}{1-p_i} = \mathbf{x}_i^T \boldsymbol{\beta}$

a funkcja wiarygodności[27] [Cameron i Trivedi 2009: 469–470]:

$$lnL_N = \sum_{i=1}^{N} \left[y_i \mathbf{x}_i^T \boldsymbol{\beta} - ln\left(1 + exp\left(\mathbf{x}_i^T \boldsymbol{\beta}\right)\right)\right] \tag{3.27}$$

Wzrost wartości zmiennej X_{mi} (m – numer zmiennej m = 1,…, K) o jednostkę, *ceteris paribus*, powoduje zmianę ilorazu szans, że $y_i = 1$ o wartość $exp(\beta_m)$.

Model probitowy

W modelu probitowym funkcja (3.22) jest dystrybuantą rozkładu normalnego standaryzowanego N(0,1) w punktach $\mathbf{x}_i^T \boldsymbol{\beta}$:

$$p_i = f(\mathbf{x}_i, \boldsymbol{\beta}) = \Phi(\mathbf{x}_i, \boldsymbol{\beta}) = \int_{-\infty}^{\mathbf{x}_i^T \boldsymbol{\beta}} \frac{1}{2\pi} exp\left(-\frac{t^2}{2}\right) dt \tag{3.28}$$

która posiada funkcję odwrotną, pozwalającą modelować prawdopodobieństwo

$$f^{-1}(p_i)\, probit(p_i) = \mathbf{x}_i^T \boldsymbol{\beta} \tag{3.29}$$

gdzie: wartości $probit(p_i) = f^{-1}(p_i)$ nazywane są **probitami**. W modelu probitowym niemożliwe jest wyznaczenie tak wygodnej analitycznej zależności pomiędzy p_i a $\mathbf{x}_i^T \boldsymbol{\beta}$ jak w modelu logitowym. Najczęściej stosowanymi estymatorami są estymatory otrzymane metodą największej wiarygodności (MNW) [Gruszczyński 2010: 69].

[27] Po zlogarytmowaniu równania (3.26)

$$Q_N = lnL_N = \sum_{i=1}^{N} [y_i lnp_i + (1 - y_i) ln(1 - p_i)].$$

Po podstawieniu (3.24) oraz $1 - p_i = 1 - \frac{exp(\mathbf{x}_i^T \boldsymbol{\beta})}{1+exp(\mathbf{x}_i^T \boldsymbol{\beta})} = \frac{1+exp(\mathbf{x}_i^T \boldsymbol{\beta})-exp(\mathbf{x}_i^T \boldsymbol{\beta})}{1+exp(\mathbf{x}_i^T \boldsymbol{\beta})} = \frac{1}{1+exp(\mathbf{x}_i^T \boldsymbol{\beta})}$ otrzymuje się:

$$lnL_N = \sum_{i=1}^{N} [y_i lnp_i + (1 - y_i) ln(1 - p_i)] = \sum_{i=1}^{N} \left[y_i ln \frac{exp(\mathbf{x}_i^T \boldsymbol{\beta})}{1+exp(\mathbf{x}_i^T \boldsymbol{\beta})} + (1 - y_i) ln \frac{1}{1+exp(\mathbf{x}_i^T \boldsymbol{\beta})} \right] =$$
$$= \sum_{i=1}^{N} \left[y_i \mathbf{x}_i^T \boldsymbol{\beta} - y_i ln\left(1 + exp\left(\mathbf{x}_i^T \boldsymbol{\beta}\right)\right) + (1 - y_i) \cdot 0 - (1 - y_i) ln\left(1 + exp\left(\mathbf{x}_i^T \boldsymbol{\beta}\right)\right)\right] =$$
$$= \sum_{i=1}^{N} \left[y_i \mathbf{x}_i^T \boldsymbol{\beta} - y_i ln\left(1 + exp\left(\mathbf{x}_i^T \boldsymbol{\beta}\right)\right) - ln\left(1 + exp\left(\mathbf{x}_i^T \boldsymbol{\beta}\right)\right) - y_i ln\left(1 + exp\left(\mathbf{x}_i^T \boldsymbol{\beta}\right)\right)\right] =$$
$$= \sum_{i=1}^{N} \left[y_i \mathbf{x}_i^T \boldsymbol{\beta} - ln\left(1 + exp\left(\mathbf{x}_i^T \boldsymbol{\beta}\right)\right)\right].$$

• Modele zmiennej wielomianowej (wielonomialnej) •

Rozszerzeniem przypadku zmiennej dychotomicznej są zmienne wielomianowe. Również one stanowią integralną część modelowania mikroekonometrycznego i eksperymentów mikrosymulacyjnych. Zmienna wielomianowa, odzwierciedlająca wybory lub przynależność do grupy obiektów, przyjmuje skończoną liczbę możliwych wariantów. Określa się zbiór opcji lub alternatyw, które są wzajemnie wykluczające i dopełniają się do zbioru wszystkich możliwych realizacji. Mogą mieć one charakter **uporządkowany**, np. wykształcenie (posiada ono naturalny porządek umożliwiający ranking: podstawowe, średnie, wyższe) lub **nieuporządkowany** (niepozwalający na stworzenie jednoznacznego i logicznego rankingu, np. region zamieszkania).

W badaniach ekonomiczno-społecznych, wykorzystujących modele mikrosymulacyjne, występują zmienne jakościowe mające więcej niż 2 kategorie. Analogicznie do zmiennych binarnych analityk może nie dysponować informacjami o prawdopodobieństwach poszczególnych kategorii, co uniemożliwia aktualizację cech przez model mikrosymulacyjny na podstawie parametrów zewnętrznych modelu mikrosymulacyjnego. Niezbędne staje się więc skonstruowanie modelu ekonometrycznego, objaśniającego zmienną opisująca wybór opcji. W takim przypadku należy rozważyć czy warianty są wyrażone na skali nominalnej czy porządkowej, tzn. czy rozkład zmiennej endogenicznej jest wielomianowy uporządkowany czy nieuporządkowany. Zazwyczaj wybór ten jest intuicyjny i wynika bezpośrednio z definicji zmiennej w bazie mikrodanych.

Modele zmiennej wielomianowej uporządkowanej

Metodologia tych modeli bazuje na założeniu, że zmienna objaśniana y_i^* jest ciągła, lecz nieobserwowalna[28] i często rozumiana jako skłonność. Oryginalna zmienna y_i^* jest funkcją:

$$y_i^* = \mathbf{x}_i^T \boldsymbol{\beta} + \varepsilon_i \tag{3.30}$$

Przybliżeniem y_i^* jest zmienna wielomianowa y_i przyjmująca jeden z J możliwych wariantów, oznaczanych kolejnymi liczbami naturalnymi, $j = 1, 2, \ldots, J$. Przekształcenie pomiędzy tymi dwoma zmiennymi wymaga podzielenia całego zakresu zmiennej nieobserwowalnej na J przedziałów, poprzez wyznaczenie J+1 punktów odcięcia (ang. *cutting points*) – elementów wektora $\boldsymbol{\kappa}$, stanowiących ich granice. Zazwyczaj przyjmuje się, że $\kappa_0 = -\infty$ a $\kappa_J = +\infty$. Zatem:

[28] Podobne rozumowanie można wykorzystać w modelach zmiennych binarnych.

$$
\begin{aligned}
y_i = 1 &\Leftrightarrow -\infty < y_i^* \leq \kappa_1 \\
y_i = 2 &\Leftrightarrow \kappa_1 < y_i^* \leq \kappa_2 \\
&\vdots \\
y_i = J &\Leftrightarrow \kappa_{J-1} < y_i^* \leq +\infty
\end{aligned}
\tag{3.31}
$$

Specyfikacja ta zależy jednak od nieobserwowalnej zmiennej, dlatego podstawiając (3.30) do (3.31) otrzymuje się po przekształceniach[29]:

$$
\begin{aligned}
y_i = 1 &\Leftrightarrow -\infty < \varepsilon_i \leq \kappa_1 - \mathbf{x}_i^T\boldsymbol{\beta} \\
y_i = 2 &\Leftrightarrow \kappa_1 - \mathbf{x}_i^T\boldsymbol{\beta} < \varepsilon_i \leq \kappa_2 - \mathbf{x}_i^T\boldsymbol{\beta} \\
&\vdots \\
y_i = J &\Leftrightarrow \kappa_{J-1} - \mathbf{x}_i^T\boldsymbol{\beta} < \varepsilon_i \leq +\infty
\end{aligned}
\tag{3.32}
$$

Postać (3.32) jest niezależna od y_i^* i wymaga estymacji parametrów $\boldsymbol{\beta}$. Macierz $\mathbf{X}$ nie zawiera wyrazu wolnego, ponieważ niemożliwe byłoby odseparowanie jego wpływu od wielkości κ_j.

Parametry modelu zmiennej wielomianowej uporządkowanej szacuje się metodą największej wiarygodności (MNW). Funkcja wiarygodności ma postać:

$$
lnL_N = \sum_{i=1}^{N}\sum_{j=1}^{J} d_{ij}\left[F\left(\kappa_j - \mathbf{x}_i^T\boldsymbol{\beta}\right) - F\left(\kappa_{j-1} - \mathbf{x}_i^T\boldsymbol{\beta}\right)\right] \tag{3.33}
$$

gdzie: $d_{ij} = \begin{cases} 1 & gdy\ y_i = j \\ 0 & gdy\ y_i \neq j \end{cases}$ [30].

[29] Po podstawieniu otrzymuje się $y_i = j \Leftrightarrow \kappa_{j-1} < \mathbf{x}_i^T\boldsymbol{\beta} + \varepsilon_i \leq \kappa_j$, a po przekształceniu $y_i = j \Leftrightarrow \kappa_{j-1} - \mathbf{x}_i^T\boldsymbol{\beta} < \varepsilon_i \leq \kappa_j - \mathbf{x}_i^T\boldsymbol{\beta}$.

[30] Niech p_{ij} oznacza prawdopodobieństwo, że *i*-ty obiekt przyjmie wartość:

$$j\ (p_{ij} = P(y_i = j)), \text{a } d_{ij}$$

to wskaźnik oznaczający czy w próbie y_i przyjęła wariant *j*. Wiarygodnością danej obserwacji jest prawdopodobieństwo, że zmienna przyjęła wariant, który został zaobserwowany, tzn.

$$L_i(y_i|\mathbf{x}_i) = \prod_{j=1}^{J} {p_{ij}}^{d_{ij}}.$$

Wartości d_{ij} pochodzą z próby, ale p_{ij} nie są znane. Uwzględniając w prawdopodobieństwie wzór (3.34) można zapisać:

$$
\begin{aligned}
p_{ij} = P(y_i = j) &= P\left(\kappa_{j-1} - \mathbf{x}_i^T\boldsymbol{\beta} < \varepsilon_i \leq \kappa_j - \mathbf{x}_i^T\boldsymbol{\beta}\right) = \\
&= P\left(\varepsilon_i \leq \kappa_j - \mathbf{x}_i^T\boldsymbol{\beta}\right) - P\left(\varepsilon_i < \kappa_{j-1} - \mathbf{x}_i^T\boldsymbol{\beta}\right) = F\left(\kappa_j - \mathbf{x}_i^T\boldsymbol{\beta}\right) - F\left(\kappa_{j-1} - \mathbf{x}_i^T\boldsymbol{\beta}\right),
\end{aligned}
$$

W modelach mikroekonometrycznych najczęściej wykorzystuje się:

▪ **uporządkowany model logitowy** (ang. *ordered multinomial logit*, OMNL)[31] o funkcji prawdopodobieństwa postaci:

$$p_{ij} = \frac{exp\left(\kappa_j - \mathbf{x}_i^T\boldsymbol{\beta}\right)}{1 + exp\left(\kappa_j - \mathbf{x}_i^T\boldsymbol{\beta}\right)} - \frac{exp\left(\kappa_{j-1} - \mathbf{x}_i^T\boldsymbol{\beta}\right)}{1 + exp\left(\kappa_{j-1} - \mathbf{x}_i^T\boldsymbol{\beta}\right)} \tag{3.34}$$

▪ **uporządkowany model probitowy** (ang. *ordered multinomial probit*, OMNP) o funkcji prawdopodobieństwa postaci:

$$p_{ij} = \Phi\left(\kappa_j - \mathbf{x}_i^T\boldsymbol{\beta}\right) - \Phi\left(\kappa_{j-1} - \mathbf{x}_i^T\boldsymbol{\beta}\right) \tag{3.35}$$

gdzie: Φ to dystrybuanta standaryzowanego rozkładu normalnego.

Podobnie jak w binarnym modelu logitowym, wzrost wartości zmiennej X_{mi} o jednostkę, *ceteris paribus*, powoduje zmianę ilorazu szans o wartość $exp(\beta_m)$ [Książek 2010: 103–141].

Modele zmiennej wielomianowej nieuporządkowanej

Model zmiennej wielomianowej nieuporządkowanej zakłada, że użyteczność poszczególnych obiektów z każdej kategorii U_{ij} składa się z części systematycznej (V_{ij}) oraz losowej (ε_{ij}):

$$U_{ij} = V_{ij} + \varepsilon_{ij} \tag{3.36}$$

gdzie: i to numer obiektu, j to numer wariantu cechy.

Część systematyczna to funkcja zmiennych objaśniających (w tym wyrazu wolnego), o parametrach różniących się pomiędzy kategoriami zmiennej endogenicznej:

$$V_{ij} = \mathbf{x}_i^T\boldsymbol{\beta}_j. \tag{3.37}$$

---

gdzie: F to dystrybuanta składnika losowego odpowiadająca funkcji gęstości f. Stąd wiarygodność pojedynczej obserwacji jest funkcją zmiennych egzogenicznych i parametrów:

$$L_i(y_i|\mathbf{x}_i) = \prod_{j=1}^{J} p_{ij}{}^{d_{ij}} = \prod_{j=1}^{J}\left[F\left(\kappa_j - \mathbf{x}_i^T\boldsymbol{\beta}\right) - F\left(\kappa_{j-1} - \mathbf{x}_i^T\boldsymbol{\beta}\right)\right]^{d_{ij}}$$

a zakładając niezależność informacji funkcja wiarygodności próby to:

$$L_N = \prod_{i=1}^{N}\prod_{j=1}^{J}\left[F\left(\kappa_j - \mathbf{x}_i^T\boldsymbol{\beta}\right) - F\left(\kappa_{j-1} - \mathbf{x}_i^T\boldsymbol{\beta}\right)\right]^{d_{ij}},$$

po zlogarytmowaniu otrzymuje się zapis (3.32).

[31] Dystrybuanta rozkładu logistycznego ma postać $F(\varepsilon) = \frac{exp(\varepsilon)}{1+exp(\varepsilon)}$.

W mikroekonometrii najczęściej wykorzystuje się **wielomianowy model logitowy** (ang. *multinomial logit*, MNL). Różni się on od uporządkowanego wielomianowego modelu logitowego jedynie określeniem prawdopodobieństw p_{ij}[32], dla którego zlogarytmowana funkcja wiarygodności ma postać:

$$lnL_N = \sum_{i=1}^{N} \sum_{j=1}^{J} d_{ij} p_{ij} \tag{3.38}$$

Dla zapewnienia identyfikowalności, przyjmuje się, że wektor $\boldsymbol{\beta}_j$ dla dowolnej kategorii (najczęściej $\boldsymbol{\beta}_1$) zwanej bazową, równy jest wektorowi o elementach zerowych. Parametry modelu estymuje się MNW, a interpretuje w odniesieniu do kategorii bazowej [Wiśniowski 2010: 153–167].

Modele zmiennych ograniczonych

Rozważane są zmienne o rozkładzie dyskretno-ciągłym (najczęściej jednopunktowym-normalnym) – dotyczy to **modeli tobit** i **rozwiązań brzegowych** oraz pochodzące z próby nielosowej – **model dwuczęściowy (selekcji próby)** i **regresji uciętej**.

Zmienna jest ograniczona, jeśli przyjmuje wartości z pewnego przedziału $[C_0, +\infty)$, $[-\infty, C_1]$ lub $[C_0, C_1]$[33]. Ograniczenia zmiennych w bazach mikrodanych mogą wynikać z definicji lub ograniczeń technicznych. Sposób gromadzenia danych na podstawie badań ankietowych, sprzyja ograniczaniu zbioru wartości. Kwestionariusz badań ankietowych może być kierowany do wyselekcjonowanej grupy jednostek, np. do osób chorych lub gospodarstw rolnych o powierzchni co najmniej 1ha. W takiej sytuacji populacja startowa mikrosymulacji nie zawiera informacji o wartościach cech obiektów nieobjętych badaniem. Zatem model mikroekonometryczny, stanowiący część modelu mikrosymulacyjnego, musi zostać dostosowany do rozkładu cechy w populacji startowej nie zaś w populacji generalnej. Estymacja MNK parametrów modeli ograniczonej zmiennej endogenicznej nie jest właściwa, ponieważ im więcej informacji znajduje się poza przedziałem lub na jego krańcach, tym większe jest obciążenie estymatora. Należy podkreślić, że kiedy zmienna nie przyjmuje wartości spoza przedziału i rzadko równa jest wartości granicznej, estymator parametrów strukturalnych MNK jest dopuszczalny[34].

---

[32] $p_{ij} = \frac{exp(\mathbf{x}_i^T \boldsymbol{\beta}_j)}{\sum_{r=1}^{J} exp(\mathbf{x}_i^T \boldsymbol{\beta}_r)}$

[33] Rzadziej rozważa się przedziały otwarte.

[34] Stąd w analizach makroekonomicznych stosowanie metody najmniejszych kwadratów jest uzasadnione. Ekonomiczne zmienne dla danych zagregowanych (np. wydatki, koszty, dochody osób, gospodarstw domowych czy firm zsumowane lub uśrednione według wybranych kategorii – sekcji PKD, województwa, okresów) bardzo rzadko przyjmują wartości graniczne (zazwyczaj $C = 0$).

Zmienne endogeniczne o rozkładzie dyskretno-ciągłym szacowane są na podstawie modelu tobit lub modelu rozwiązań brzegowych. W **modelu tobitowym** zakłada się, że zmienna objaśniana ma rozkład ciągły w pewnym przedziale, a dla znacznego odsetka populacji cecha przyjmuje wartość brzegową. Empiryczne wartości niegraniczne zazwyczaj pochodzą z przedziału $(C_0, +\infty), (-\infty, C_1)$ lub (C_0, C_1), gdzie C_0 i C_1 to odpowiednio wartość brzegowa dolna i górna. Przykładem mogą być udziały kosztów bezpośrednich w kosztach ogółem, ograniczone obustronnie $C_0 = 0$ i $C_1 = 1$ lub udział wydatków na ochronę zdrowia w dochodzie, ograniczone lewostronnie $C_0 = 0$. W **modelu rozwiązań brzegowych** przyjmuje się, że wartość brzegowa ma charakter dyskretny i jest wynikiem podjęcia innej decyzji niż te generujące wartość z przedziału o rozkładzie ciągłym. Oznacza to, że choć teoretycznie wartości cechy pochodzą z przedziałów $[C_0, +\infty)$, $(-\infty, C_1]$ lub $[C_0, C_1]$, empiryczne wartości niegraniczne przyjmują wartości z zakresu przedziału $(D_0, +\infty)$ $(-\infty, D_1)$ lub (D_0, D_1), gdzie $C_0 < D_0$, a często $C_0 \ll D_0$, $D_1 < C_1$ lub $D_1 \ll C_1$. Taki charakter mają często wydatki gospodarstw domowych na leczenie czy wydatki przedsiębiorstw ponoszone na restrukturyzację. Mogą one nie wystąpić w okresie badania, przyjmując wartość brzegową 0 zł albo (w wyniku zajścia pewnego zdarzenia czy podjęcia decyzji) przyjmować wartości dużo wyższe od granicy, np. $D_0 = 1$ tys. zł lub $D_0 = 1$ mln zł.

O wyborze modelu tobit lub rozwiązań brzegowych decyduje rozkład cechy w startowej bazie mikrodanych. Estymatory MNK dla obu modeli są niezgodne i obciążone. Najczęściej parametry tych modeli szacuje się metodą największej wiarygodności (MNW) [Maddala 1983: 149–180; Owczarczuk 2010: 197]. W badaniach ekonomiczno-społecznych dla danych indywidualnych powszechniej stosowany jest model tobitowy, ponieważ zmienna endogeniczna często objaśniana jest modelem liniowym o składniku losowym $\varepsilon \sim N(0, \sigma^2)$, a $C_0 = 0$. Stąd funkcja wiarygodności ma postać:

$$lnL_N(\boldsymbol{\beta}, \sigma^2) = \sum_{i=1}^{N} \left\{ d_i \left[-\frac{1}{2} ln2\pi - \frac{1}{2} ln\sigma^2 - \frac{1}{2\sigma^2} \left(y_i - \mathbf{x}_i^T \boldsymbol{\beta} \right)^2 \right] + (1 - d_i) ln \left[1 - \Phi \left(\frac{\mathbf{x}_i^T \boldsymbol{\beta}}{\sigma} \right) \right] \right\} \tag{3.39}$$

gdzie: d_i to indykator przyjmujący wartość 1, gdy y_i różne jest od wartości granicznych i 0 w przeciwnym przypadku.

Funkcja wiarygodności uwzględnia obserwacje przyjmujące wartość brzegową (drugi składnik sumy) oraz te o wartościach różnych od granicy (pierwszy składnik sumy)[35].

[35] W modelu tobit i regresji brzegowej przyjmuje się, że istnieje zmienna ukryta y^*, która przyjmuje wartości ze zbioru liczb rzeczywistych oraz jej ocenzurowane przekształcenie y, takie, że:

W **modelu regresji uciętej** (ang. *truncation model*) jeszcze przed losowaniem próby dochodzi do ucięcia części obserwacji, które w rezultacie nie są znane badaczowi. Powoduje to obciążenie estymatorów MNK. Dlatego parametry modelu szacowane są metodą największej wiarygodności[36]. Stosowanie modelu regresji uciętej w mikrosymulacjach ekonomicznych i społecznych stanowi zazwyczaj konsekwencję sposobu prowadzenia badań ankietowych i konstrukcji populacji startowej.

Rozważając jednorównaniowe modele zmiennych ograniczonych, nie sposób pominąć **model dwuczęściowy** (selekcji próby, ang. *two-part model*, TPM). Wprawdzie jest to model parametryczny dwurównaniowy, jednak dotyczy jednej zmiennej endogenicznej. Dlatego często traktowany jest na równi z modelami jednorównaniowymi. Pierwsze równanie to równanie selekcji (ma postać modelu logit lub probit), a drugie to równanie wynikowe (dla danych dodatnich całkowitoliczbowych, określonych rozkładem prawdopodobieństwa Poissona lub ujemnym binarnym). Parametry modelu można szacować metodami dla modeli wielorównaniowych, np. metodą największej wiarygodności lub sekwencyjnie: równanie selekcji a następnie równanie wynikowe[37].

Model dwuczęściowy umożliwia analizę skutków „podwójnych" decyzji związanych ze zmiennymi warunkowymi i binarnymi, np. stanu zdrowia pod warunkiem, że osoba pali. Inkorporacja modeli selekcji próby do modeli mikrosymulacyjnych, analizujących cechy warunkowe, mogłaby przyczynić się do poprawy jakości uzyskanych wyników [Suchecka i in. 2007: 125–128].

Modele zmiennych licznikowych

Zmienna endogeniczna może być zmienną licznikową o wartościach dyskretnych (naturalnych). Zbiór wartości zmiennej skokowej nie ma górnej granicy, jednak w rozkładzie przeważają małe wartości [Roning 2006: 156]. W mikrosymulacyjnych badaniach ekonomiczno-społecznych taki charakter mogą

$$y = \begin{cases} y^* & dla\ y^* > C \\ C & dla\ y^* = C \end{cases}.$$

Dla y warunkowa cenzurowana funkcja gęstości ma postać:

$$f(y|\mathbf{x}) = f^*(y|\mathbf{x})^d F^*(C|\mathbf{x})^{1-d}$$

(d to indykator przyjmujący wartość 1 dla y różnego od wartości brzegowych i 0 w przeciwnym przypadku). Stąd funkcję wiarygodności modeli tobit i rozwiązań brzegowych można zapisać najogólniej jako:

$$lnL_N(\boldsymbol{\theta}) = \sum_{i=1}^{N} [d_i lnf^*(y|\mathbf{x},\boldsymbol{\theta}) + (1 - d_i) lnF^*(L|\mathbf{x},\boldsymbol{\theta})]$$

Zob. Cameron i Trivedi [2009: 532–537].

[36] Funkcja wiarygodności jest identyczna jak ogólna funkcja modelu tobit i regresji brzegowej – zob. Cameron i Trivedi [2009: 533–534].

[37] Jeżeli składniki losowe obu równań są skorelowane w specyfikacji funkcji wiarygodności lub estymując parametry równań wynikowych, należy uwzględnić współczynnik korelacji.

mieć zmienne dotyczące gospodarstw domowych (np. liczba osób) lub przedsiębiorstw (liczba oddziałów). W ten sposób określona zmienna nie może być postrzegana ani modelowana jako zmienna ciągła. Najczęściej model zmiennej licznikowej ma postać modelu Poissona, ale tylko w przypadku, gdy spełnione są założenia o rozkładzie Poissona.

W przypadku zmiennych ekonomiczno-społecznych wartość oczekiwana nie jest znana ani stała dla wszystkich obiektów badania. Dlatego przyjmuje się, że wartości zmiennej zależnej są uwarunkowane wartościami $\mathbf{x}_i$. Stąd **model regresji Poissona (MP)** objaśnia wartość oczekiwaną rozkładu jako funkcję wykładniczą zmiennych egzogenicznych, o wartościach nieujemnych:

$$\lambda_i = e^{\mathbf{x}_i^T \boldsymbol{\beta}}. \tag{3.40}$$

Parametry modelu szacowane są metodą największej wiarygodności[38] i interpretowane jako względna (procentowa) zmiana wartości cechy endogenicznej na wyizolowaną bezwzględną jednostkową zmianę zmiennej egzogenicznej [Bazyl 2010: 217].

W przypadku kiedy występuje nieobserwowalna heterogeniczność obiektów, powodująca „nadmierne rozproszenie" (ang. *overdispersion*), tzn. wartość wariancji jest wyższa niż wartość oczekiwana rozkładu, zalecane jest stosowanie uogólnionego modelu Poissona – **modelu regresji ujemnej dwumianowej (MRUD)**:

$$\tilde{\lambda}_i = e^{\mathbf{x}_i^T \boldsymbol{\beta} + \epsilon_i} = e^{\mathbf{x}_i^T \boldsymbol{\beta}} e^{\epsilon_i} = e^{\mathbf{x}_i^T \boldsymbol{\beta}} u_i \tag{3.41}$$

gdzie: ϵ_i to nieobserwowalna część zmienności *Y*.

Estymatory modelu regresji Poissona tracą efektywność, ale nie własność zgodności, ponieważ specyfikacja wartości oczekiwanej jest poprawna. W przypadku kiedy „nadmierne rozproszenie" wynika z błędnej specyfikacji modelu i wartości oczekiwanej, np. nie uwzględniono, że rozkład jest ucięty, estymator MP nie jest zgodny. Parametry MRUD szacowane są MNW i interpretowane analogicznie do modelu Poissona. W przypadku „niedostatecznego rozproszenia" (ang. *underdispersion*), kiedy wartość oczekiwana jest większa od wariancji, nie można stosować modelu regresji ujemnej dwumianowej.

Drugim problemem w przypadku zmiennych licznikowych jest „nadmiar zer" (ang. *excess zeros*), związany z występowaniem wartości zerowych częściej niż przewiduje to model Poissona. W celu uwzględnienia specyfiki takiego rozkładu najczęściej stosuje się model dwuczęściowy [Cameron i Trivedi 2009: 670–681].

[38] Jedynie w przypadku kiedy wartości licznikowe przyjmują duże wartości (powyżej 15–20), ze względu na podobieństwo rozkładu Poissona do normalnego, można stosować regresję liniową i estymację MNK.

3.3.4. Estymatory paramentów strukturalnych modeli wielorównaniowych danych przekrojowych

Szacując parametry strukturalne wielorównaniowych modeli mikroekonometrycznych należy uwzględnić specyfikę rozkładu poszczególnych zmiennych endogenicznych (każde z równań może mieć inną postać funkcyjną) oraz charakter interakcji pomiędzy nimi. Parametry modeli prostych oraz rekurencyjnych[39] mogą być estymowane równanie po równaniu podejściami jednorównaniowymi [Wiśniewski 2009: 215]. Modele rekurencyjne wymagają ustalenia kolejności równań umożliwiającej oszacowanie parametrów strukturalnych, dla modeli prostych uszeregowanie nie ma wpływu na rezultaty estymacji. Ze względu na prostotę i możliwość dobrania adekwatnych metod do założeń i postaci funkcyjnej każdego równania najczęściej stosuje się taką procedurę estymacji sekwencyjnej. W przypadku kiedy składniki losowe poszczególnych równań modeli prostych są skorelowane, a zbiór zmiennych egzogenicznych w każdym równaniu nie jest jednakowy, zalecane jest stosowanie metod estymacji systemowej regresji pozornie niezależnej (czasami nazywanej pozornie zależną, modelu o równaniach pozornie niezależnych ang. *seemingly unrelated regression*, SUR). Pozwala to szacować parametry, uwzględniając informacje o korelacji błędów losowych (w modelach nieprostych struktura interakcji pomiędzy zmiennymi endogenicznymi inkorporuje tę dodatkową wiedzę). Wielorównaniowy model *M* równań, niezależnie od charakteru interakcji między nimi, można zapisać w postaci:

$$\mathbf{y}_j = \mathbf{X}_j \boldsymbol{\beta}_j + \boldsymbol{\varepsilon}_j \tag{3.42}$$

lub:

$$\tilde{\mathbf{y}} = \tilde{\mathbf{X}}\tilde{\boldsymbol{\beta}} + \tilde{\boldsymbol{\varepsilon}} \tag{3.43}$$

gdzie: $\tilde{\mathbf{y}} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_M \end{bmatrix}$, $\tilde{\mathbf{X}} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_M \end{bmatrix}$, $\tilde{\boldsymbol{\beta}} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_M \end{bmatrix}$, $\tilde{\boldsymbol{\varepsilon}} = \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_M \end{bmatrix}$,

j – numer równania, a macierz wariancji–kowariancji $\tilde{\boldsymbol{\Omega}} = \boldsymbol{\Omega} \otimes \mathbf{I}_N$.

Paramenty modelu (3.43) można szacować **metodą największej wiarygodności (SUR–MNW)** lub, kiedy znane jest oszacowanie macierzy $\tilde{\boldsymbol{\Omega}}$ na podstawie reszt z estymacji MNK, **osiągalną uogólnioną metodą najmniejszych kwadratów (SUR–OUMNK)**.

[39] Modele proste – nie występują sprzężenia pomiędzy zmiennymi objaśnianymi; modele rekurencyjne – sprzężenia mają charakter jednokierunkowy – zob. Gajda [2004: 183–184].

Modele współzależne (o równaniach współzależnych) postaci:

$$\mathbf{y}_j = \mathbf{Y}_j\boldsymbol{\alpha}_j + \mathbf{X}_j\boldsymbol{\beta}_j + \boldsymbol{\varepsilon}_j = \mathbf{W}_j\boldsymbol{\delta}_j + \boldsymbol{\varepsilon}_j \tag{3.44}$$

gdzie: $\mathbf{Y}_j$ to macierz wartości zmiennych endogenicznych, w której numery kolumny oznaczają numer zmiennej, a wiersza – numer obserwacji, a $\boldsymbol{\alpha}_j$ to parametry strukturalne przy zmiennych endogenicznych w równaniu j

muszą być traktowane jako układ. Istnieją dwie rodziny podejść estymacji parametrów równań: pojedynczo (sekwencyjnie, nie wykorzystując pełnej informacji) i łącznie (systemowo, wykorzystując pełną informację). Struktura modelu (3.44) jest analogiczna do (3.1), zatem większość zaprezentowanych metod estymacji parametrów modeli jednorównaniowych przebiega analogicznie dla modeli współzależnych. Parametry modelu (3.44) można estymować sekwencyjnie (estymacja pojedynczego równania, niepełnej informacji) metodami szacowania modeli jednorównaniowych, m. in. metodą zmiennych instrumentalnych (3.8), podwójną metodą najmniejszych kwadratów (3.11), metodą największej wiarygodności (3.14, 3.15) podstawiając do wzorów zamiast macierzy $\mathbf{X}$ macierz $\mathbf{W}_j = [\mathbf{Y}_j, \mathbf{X}_j]$. Należy jednak wskazać, że podejście 2MNK dla danych z szeregów czasowych o krótszym okresie pomiaru niż rok może znacząco obniżać stopień dopasowania wartości teoretycznych zmiennych endogenicznych do ich realizacji empirycznych [Wiśniewski 2011: 523–530]. **Metody systemowe** (**pełnej informacji**) wymagają postrzegania modelu M równań postaci (3.44) jako zintegrowanej całości:

$$\tilde{\mathbf{y}} = \widetilde{\mathbf{W}}\widetilde{\boldsymbol{\delta}} + \tilde{\boldsymbol{\varepsilon}} \tag{3.45}$$

zdefiniowanego analogicznie do (3.43) [Gajda 1988: 96–130]. Parametry takiego modelu szacować można potrójną metodą najmniejszych kwadratów lub metodą największej wiarygodności z pełną informacją. **Potrójna metoda najmniejszych kwadratów** (3MNK, ang. *three stage least square*, 3SLS) przebiega w trzech etapach: (1) estymacja MNK parametrów kolejnych równań i wyznaczenie teoretycznych wartości zmiennych endogenicznych; (2) estymacja kolejnych równań MNK na podstawie oszacowania wartości teoretycznych zmiennych endogenicznych w celu wyznaczenia macierzy wariancji–kowariancji reszt; (3) estymacja wszystkich równań z wykorzystaniem oszacowania macierzy wariancji–kowariancji z poprzedniego etapu. Stosując **metodę największej wiarygodności z pełną informacją** (MWPI, ang. *full information maximum likelihood*, FIML) należy we wzorach (3.14) i (3.15) zamienić skalar y_i na wektor $\mathbf{y}_i$ [Cameron i Trivedi 2009: 779–807].

Metody 2MNK i 3MNK łatwo przetransformować do postaci ważonej lub osiągalnej uogólnionej metody najmniejszych kwadratów, analogicznie do wzorów (3.5) i (3.4), co pozwala wyeliminować problemy wynikające ze specyfiki rozkładu mikrodanych.

3.4. Specyfika mikrodanych panelowych

W przypadku mikrodanych panelowych[40] (ang. *micro panel*) zazwyczaj dostępne są zmienne dla bardzo wielu elementów (setek lub tysięcy) w niewielu punktach czasu (kilku lub kilkunastu). Tak zdefiniowany zbiór danych nazywany jest krótkim panelem (ang. *short panel*) [Dańska-Borsiak 2011: 13–15]. Makrodane panelowe z reguły mają odwrotną strukturę: niewiele obiektów (kilka do kilkudziesięciu) w wielu okresach (kilkudziesięciu), tzw. długi panel (ang. *long panel*). Z tego względu również metody analiz mikrodanych panelowych różnią się od narzędzi dla makrodanych panelowych. Przykładem różnic w modelowaniu może być występowanie niestacjonarności szeregów czasowych, charakterystyczne jedynie dla długich paneli. Zjawisko to nie zachodzi w krótkich panelach mikrodanych [Baltagi 2005: 1; Cameron i Trivedi 2009: 697].

Mikrodane panelowe pozwalają na uwzględnienie w analizach (a) **indywidualnej heterogeniczności**. Powtarzająca się w pewnych odstępach czasu informacja dla każdego obiektu stanowi podstawę do wyizolowania czynnika specyficznego. Dane panelowe zawierają (b) **więcej informacji i są bardziej rozproszone** niż dla szeregów czasowych lub przekrojowych. W efekcie rzadziej obserwowana jest współliniowość pomiędzy zmiennymi egzogenicznymi, szczególnie kłopotliwa w makrodanych czasowych. Oszacowania parametrów strukturalnych są bardziej wiarygodne, przy założeniu, że są one stałe dla wszystkich obserwacji. Dane panelowe pozwalają (c) **analizować dynamikę reakcji i zmian** cech elementów badania, której nie uwzględniają dane czasowe (brak rozróżnionych obiektów), przekrojowe (statyczne) ani przekrojowo-czasowe (zmienny zbiór obserwacji). Posiadając informacje dla tych samych obiektów w kolejnych okresach, można zaobserwować zachodzące w nich zmiany oraz ich reakcje na sytuacje ekonomiczne, społeczne i polityczne. W przypadku zjawisk dynamicznych i niejednorodnych przekrojowo dane panelowe (d) **umożliwiają konstrukcję dobrze dopasowanych modeli uwzględniających czynniki czasowe i przekrojowe**. Mikrodane panelowe charakteryzują się (e) **większą precyzją pomiaru** niż informacje zagregowane.

Głównym ograniczeniem wykorzystania mikrodanych panelowych jest ich gromadzenie. Zazwyczaj informacje zbierane są w powtarzalnym badaniu ankie-

[40] Dane przekrojowo-czasowe, w których informacje zgromadzone są dla tych samych obiektów w kolejnych okresach, nazywane są danymi panelowymi.

towym. Prowadzi to do obniżenia jakości mikrodanych, ponieważ: (1) **nie ma informacji o wszystkich elementach** populacji generalnej, (2) **odpowiedzi na pytania są niewiarygodne** (np. nieprecyzyjnie sformułowane pytanie, kłamstwo), (3) **brak odpowiedzi** – współczynnik braku odpowiedzi wzrasta z każdą falą (kolejną edycją badania ankietowego). Czynniki te mogą przyczynić się do wzrostu błędu pomiaru odpowiedzi w badaniu ankietowym. Ponadto w krótkich panelach zachowanie asymptotycznych własności rozkładu jest silnie powiązane z liczebnością obiektów ze względu na małą liczbę okresów.

W przypadkach gdy obiektami panelowej bazy mikrodanych są regiony, niezbędne jest uwzględnienie interakcji przestrzennych poprzez wykorzystanie narzędzi ekonometrii przestrzennej [Dańska-Borsiak 2011: 19–21; Baltagi 2005: 6–11].

Zastosowanie danych panelowych w modelach mikroekonometrycznych stanowiących część modeli mikrosymulacyjnych staje się coraz powszechniejsze. Wynika to z łatwiejszego dostępu do mikrodanych panelowych oraz możliwości uwzględnienia w analizach mikrosymulacyjnych wpływu czynników indywidualnych jednostek i ewentualnej dynamiki zjawisk ekonomicznych.

3.5. Estymacja panelowych modeli mikroekonometrycznych

Podobnie jak w przypadku modeli przekrojowych, modele panelowe w mikrosymulacjach mają zazwyczaj postać jednorównaniową. Ze względu na dwuwymiarowy charakter danych przekrojowo-czasowych preferowane są specyfikacje liniowe. Nieczęste stosowanie modeli nieliniowych, szczególnie o rozkładach nieciągłych i nieróżniczkowalnych, wynika ze skomplikowanej procedury estymacyjnej oraz utrudnionego dostępu do odpowiednich funkcji w pakietach ekonometrycznych.

3.5.1. Estymatory paramentów strukturalnych jednorównaniowych modeli liniowych danych panelowych

Najogólniejszy panelowy model[41] mikroekonometryczny ma postać:

$$y_{it} = \alpha_{it} + \mathbf{x}_{it}\boldsymbol{\beta}_{it} + \varepsilon_{it} \tag{3.46}$$

gdzie: i to indeks obiektu (i = 1, 2…, N), t to indeks czasu (t = 1, 2 …, T), y_{it} to wartość zmiennej endogenicznej i-tego obiektu w okresie t, $\mathbf{x}_{it}$ to wektor

[41] Przez model panelowy rozumie się model ekonometryczny estymowany na podstawie danych panelowych – zob. Dańska [2000: 33].

wierszowy wartości zmiennych egzogenicznych *i*–tego obiektu w okresie *t*, zaś α_{it} to wyraz wolny, a $\boldsymbol{\beta}_{it}$ to wektor kolumnowy parametrów przy zmiennych zależnych.

Czynnik specyficzny każdego obiektu został zaadaptowany do wartości wszystkich paramentów. Ponadto każdy parametr może być zróżnicowany w czasie. W tak sformułowanym modelu niemożliwe jest oszacowanie parametrów, ponieważ ich liczba ($N \cdot T \cdot (K + 1)$, (K to liczba zmiennych w wektorze $\mathbf{x}_{it}$) jest większa niż liczba obserwacji.

Najbardziej restrykcyjnym pod względem liczby parametrów modelem liniowym jest **model przekrojowo-czasowy** (ang. *pooled model*) zakładający, że wszystkie parametry są stałe, zarówno w czasie jak i w przekroju (dla wszystkich obiektów) postaci:

$$y_{it} = \alpha + \mathbf{x}_{it}\boldsymbol{\beta} + \varepsilon_{it} \tag{3.47}$$

Jeżeli specyfikacja (3.47) jest poprawna i nie występuje korelacja pomiędzy zmiennymi objaśniającymi a składnikiem losowym, parametry modelu mogą być estymowane **przekrojowo-czasową metodą najmniejszych kwadratów** (ang. *pooled OLS*), dając zgodne oszacowania.

Zróżnicowanie wyrazu wolnego w czasie i przekroju można wprowadzić do **modelu przekrojowo-czasowego poprzez sztuczne zmienne czasu i obiektów**:

$$y_{it} = \alpha_i + \gamma_t + \mathbf{x}_{it}\boldsymbol{\beta} + \varepsilon_{it} \tag{3.48}$$

gdzie: α_i to wyraz wolny *i*-tego obiektu a γ_t to wyraz wolny czasu ($t = 2,\ldots, T$).

Standardowy wyraz wolny (wektor jedynkowy) nie może być częścią wektora $\mathbf{x}_{it}$. Estymacja $N + (T - 1) + K$ parametrów modelu (3.48) może dać zgodne oszacowania jedynie dla $N \to \infty$ oraz $T \to \infty$. Jednak w krótkim panelu, charakterystycznym dla mikrodanych, gdzie tylko $N \to \infty$, niemożliwe staje się uzyskanie zgodnych estymatorów wszystkich parametrów. Dlatego w modelowaniu najczęściej stosuje się uproszczoną wersję modelu, tzw. **model wyspecyfikowanych efektów indywidualnych**, który różnicuje wyrazy wolne poszczególnych obiektów:

$$y_{it} = \alpha_i + \mathbf{x}_{it}\boldsymbol{\beta} + \varepsilon_{it} \tag{3.49}$$

gdzie sztuczne zmienne czasu, jeśli uwzględnione, znajdują się w wektorze zmiennych objaśniających. Parametr α_i to **zmienna losowa** odzwierciedlająca **nieobserwowalną heterogeniczność** obiektów. Warunkowa wartość oczekiwana składnika losowego modelu względem przeszłych, teraźniejszych i przy-

szłych wartości zmiennych egzogenicznych oraz wyrazu wolnego wynosi 0 (silna egzogeniczność).

Istnieją dwa najczęściej stosowane warianty modelu (3.49) o stałych i losowanych efektach indywidualnych. **Model o stałych efektach** (*ang. fixed effects model*, FE model) zakłada, że α_i to nieobserwowalna zmienna losowa potencjalnie skorelowana z wartościami $\mathbf{x}_{it}$. Największe zainteresowanie analityków budzą jednak oszacowania parametru $\boldsymbol{\beta}$. Niektóre metody estymują parametry $\alpha_1, \dots, \alpha_N$[42], inne traktują je marginalnie uznając za parametry niedogodne (ang. *nuisance parameters*). W przypadku występowania stałych efektów wiele estymatorów, np. przekrojowo-czasowa MNK, traci własność zgodność. Model o stałych efektach stosuje się do zbioru konkretnych i niezbyt licznych obiektów. Do modelu FE nie można wprowadzać zmiennych stałych w czasie, ze względu na możliwość wystąpienia dokładnej współliniowości zmiennych [Baltagi 2005: 14–15]. W **modelu o losowych efektach** (ang. *random effect model*, RE model) przyjmuje się, że nieobserwowalny czynnik indywidualny α_i jest zmienną losową o rozkładzie niezależnym od zmiennych egzogenicznych [Cameron i Trivedi 2009: 698 i n.]. Często model RE zapisuje się podstawiając za α_i składnik losowy u_{it} oraz dodając wyraz wolny, stały w czasie i przekroju, opisujący wartość oczekiwaną rozkładu nieobserwowalnych czynników indywidualnych. Model (3.49) przyjmuje postać:

$$y_{it} = \alpha + \mathbf{x}_{it}\boldsymbol{\beta} + u_{it} + \varepsilon_{it} \tag{3.50}$$

gdzie: α to wyraz wolny, stały w czasie i przekroju, e_{it} to składnik losowy odpowiadający za zróżnicowanie pomiędzy obiektami (ang. *within-entity error*), powiązany z czynnikiem efektów stałych (ang. *fix effects, within effects*), u_{it} to składnik losowy zróżnicowania obiektów (ang. *between-entity error*), związany z czynnikiem specyficznym elementów badania [Torres-Reyna 2012].

Model RE zalecany jest szczególnie w przypadku, kiedy obiekty uwzględnione w analizie są losowo dobrane z populacji. W takim przypadku nie jest ważny czynnik specyficzny poszczególnych elementów próby, a jedynie świadomość, że istnieje i jest losowy [Baltagi 2005: 17]. Wybór między modelem o stałych i losowych efektach jest trudny. W dotychczasowej teorii ekonometrycznej brak jest jednoznacznych warunków ich stosowalności.

W mikroekonometrii najczęściej stosuje się modele przekrojowo-czasowe bez efektów indywidualnych (3.47) oraz o wyspecyfikowanych efektach indywidualnych (3.49) stałych i losowych (również (3.50)). Można je zapisać w ogólnej postaci jako:

[42] W niektórych przypadkach można przypisać poszczególne parametry grupom obiektów, nie zaś każdemu z osobna – zob. Klette i in. [2000: 473].

$$\tilde{y}_{it} = \widetilde{\mathbf{w}}_{it}\boldsymbol{\theta} + \tilde{\varepsilon}_{it} \tag{3.51}$$

gdzie: $\tilde{y}_{it}$, $\widetilde{\mathbf{w}}_{it}$ i $\tilde{\varepsilon}_{it}$ to znane transformacje, odpowiednio, wartości y_{it}, $\mathbf{w}_{it} = [1\ \mathbf{x}_{it}]$ i ε_{it}.

W przypadku modelu przekrojowo-czasowego nie stosuje się żadnej transformacji, tzn. wzór (3.47) i (3.51) są tożsame i $\boldsymbol{\theta} = [\alpha\ \boldsymbol{\beta}^T]^T$. W modelu FE $\tilde{y}_{it} = y_{it} - \bar{y}_{it}$, $\widetilde{\mathbf{w}}_{it} = \mathbf{x}_{it} - \bar{\mathbf{x}}_{it}$, a $\boldsymbol{\theta}$ to wektor parametrów przy zmiennych niestałych w czasie. W modelu RE $\tilde{y}_{it} = y_{it} - \hat{\lambda}\bar{y}_{it}$, $\widetilde{\mathbf{w}}_{it} = \mathbf{w}_{it} - \hat{\lambda}\bar{\mathbf{w}}_{it}$, $\boldsymbol{\theta} = [\alpha\ \boldsymbol{\beta}^T]^T$, a $\hat{\lambda}$ to zgodny estymator:

$$\lambda = 1 - \frac{\sigma_\varepsilon}{\sqrt{\sigma_\varepsilon^2 + T\sigma_u^2}} \tag{3.52}$$

gdzie: σ_ε i $\sigma_u$to odchylenia standardowe składników losowych w modelu (3.50).

Jeżeli $\hat{\lambda} = 0$, to model efektów losowych jest równoznaczny z modelem przekrojowo-czasowym, a jeżeli $\hat{\lambda} = 1$, jest równoważny z modelem o efektach stałych.

W zależności od charakteru czynnika indywidualnego: stałego w modelu przekrojowo-czasowym, losowego i skorelowanego z regresorami w modelu FE czy losowego i niezależnego od zmiennych egzogenicznych w modelu RE, stosuje się różne metody estymacji.

• Przekrojowo-czasowe estymatory metody najmniejszych kwadratów •

Parametry wszystkich modeli opisanych wzorem (3.51) można szacować stosując przekrojowo-czasową metodę najmniejszych kwadratów lub jej modyfikacje. **Przekrojowo-czasowy estymator metody najmniejszych kwadratów** (ang. *pooled ordinary least squares*, POLS) zakłada uszeregowanie danych klasyfikowanych według *i* oraz *t* w jeden ciąg i szacownie wartości parametrów strukturalnych MNK. Jeżeli $cov(\varepsilon_{it}, \mathbf{x}_{it}) = \mathbf{0}$, to $N \to \infty$ lub $T \to \infty$ wystarcza, by estymator był zgodny. Należy zauważyć, że w modelu przekrojowo-czasowym istnieje wysokie prawdopodobieństwo wystąpienia autokorelacji składnika losowego dla tych samych obiektów w czasie, tzn. $cov(\varepsilon_{it}, \varepsilon_{is}) \neq 0$, $s, t = 1,\ldots, T$. Wynika to z nieuwzględniania czynnika specyficznego obiektów (parametr α jest stały w czasie i przekroju). Ponadto w modelach dla mikrodanych istnieje ryzyko pojawienia się heteroskedastyczności składnika losowego. W takich przypadkach uzasadnione jest wykorzystanie przekształconych estymatorów MNK: **przekrojowo-czasowej uogólnionej metody najmniejszych kwadratów** (PUMNK, ang. *pooled generalised least squares*, PGLS), **przekrojowo-**

-czasowej ważonej metody najmniejszych kwadratów (PWMNK, ang. *pooled weighted least squares*, PWLS) **lub przekrojowo-czasowej osiągalnej uogólnionej metody najmniejszych kwadratów** (POUMNK, ang. *pooled feasible generalised least squares*, FGLS), w zależności od przyjętej postaci macierzy wariancji –kowariancji[43]. Aplikacja tych metod przebiega analogicznie do szacowania parametrów modeli danych przekrojowych. Oznacza to wykorzystanie wzorów, odpowiednio, (3.3), (3.4) i (3.5) dla danych panelowych uszeregowanych w jeden ciąg. Wszystkie metody opierają się na założeniu silnej egzogeniczności.

W przypadku niespełnienia założenia o silnej egzogeniczności niezbędne staje się wykorzystanie estymatorów **przekrojowo-czasowej metody zmiennych instrumentalnych** (PMZI, ang. *pooled instrumental variables*, PIV) oraz **przekrojowo-czasowej podwójnej metody najmniejszych kwadratów** (P2MNK, ang. *pooled two-stage least squares*, 2SLS). Parametry modelu szacowane są analogicznie jak we wzorach, odpowiednio, (3.8) i (3.11) po uszeregowaniu obserwacji w jeden ciąg [Cameron i Trivedi 2009: 702 i n.].

3.5.2. Estymatory paramentów strukturalnych jednorównaniowych modeli nieliniowych danych panelowych

Nieliniowe modele panelowe można sklasyfikować analogicznie do przekrojowych modeli nieliniowych zaprezentowanych w poprzednich podrozdziałach. Również proces ich estymacji jest podobny. W przypadku parametrów panelowych modeli nieliniowych, gdy empiryczna funkcja zmiennej endogenicznej jest ciągła i różniczkowalna, najczęściej stosowanymi metodami estymacja są przekrojowo-czasowa metoda zmiennych instrumentalnych i przekrojowo-czasowa podwójna metoda najmniejszych kwadratów. Jeżeli postać funkcyjna modelu jest nieciągła i nieróżniczkowalna (zmienna zależna ma np. charakter binarny lub licznikowy), stosuje się metodę największej wiarygodności. We wzorach (3.14) i (3.15) należy zamienić skalar y_i na wektor $\mathbf{y}_i$ [Cameron i Trivedi 2009: 779–807]. W przypadku modelu nieliniowego ograniczonej zmiennej zależnej zauważono, że dodanie czynnika autoregresyjnego wpływa korzystnie na eliminację autokorelacji składnika losowego wynikającej z czynników indywidualnych [Heiss 2006: 7–10].

Panelowe modele nieliniowe są stosowane w modelach mikrosymulacyjnych bardzo rzadko. Równoczesne uwzględnienie nieliniowości postaci funkcyjnej i dwuwymiarowości rozkładu zmiennej endogenicznej sprawia, że prezentowane powyżej metody estymacji stają się stosunkowo skomplikowane. Ponadto odpowiadające im procedury estymacyjne nie są oferowane przez wszystkie dostępne na rynku komputerowe pakiety ekonometryczne. Dlatego większość badaczy decyduje się na uproszczenie specyfikacji modelu do nieli-

[43] Macierz wariancji–kowariancji ma wymiar *NTxNT*, nie zaś *NxN* jak w przypadku danych przekrojowych.

niowego modelu przekrojowego lub panelowego modelu liniowego. Wraz z rozwojem oprogramowania i wzrostem dostępności adekwatnych opcji estymacyjnych można oczekiwać, że nieliniowe modele panelowe będą stawały się coraz bardziej popularne.

3.5.3. Estymatory paramentów strukturalnych modeli wielorównaniowych danych panelowych

Estymacja parametrów wielorównaniowych modeli panelowych przebiega analogicznie do estymacji współczynników wielorównaniowych modeli przekrojowych opisanych wcześniej, ale z uwzględnieniem specyfiki rozkładu danych panelowych. Modele proste i rekurencyjne można traktować jako układ modeli jednorównaniowych, co umożliwia szacowanie parametrów każdego równania osobno. Parametry modeli prostych, których równania nie mają identycznych zbiorów zmiennych egzogenicznych, można szacować jak w modelu o równaniach pozornie niezależnych SUR, jednak niezbędne jest uwzględnienie zmodyfikowanej macierzy wariancji kowariancji.

Wielorównaniowy model panelowy:

$$\mathbf{y}_j = \mathbf{X}_j\boldsymbol{\beta}_j + \boldsymbol{\varepsilon}_j \tag{3.53}$$

dla N jednostek i T okresów, gdzie: $\mathbf{y}_j$ **to** wektor $NT\ x\ 1$,[44] $\mathbf{X}_j$ **to** macierz $NT\ x\ (K_j+1)$, K_j to liczba zmiennych egzogenicznych w równaniu j, $\boldsymbol{\beta}_j$ to wektor parametrów $(K_j+1)\ x\ 1$ (uwzględniający wyraz wolny), $\boldsymbol{\varepsilon}_j$ **to** wektor $NT\ x\ 1$

można postrzegać jako jedno- lub dwukierunkową regresję z czynnikiem losowym. W modelu **jednokierunkowej regresji z czynnikiem losowym** (ang. *One-way Error Component Model*) składnik losowy modelu (3.53) ma postać:

$$\boldsymbol{\varepsilon}_j = \mathbf{F}_\mu\boldsymbol{\mu}_j + \mathbf{v}_j \tag{3.54}$$

gdzie: $\boldsymbol{\mu}_j$to wektor $N\ x\ 1$ nieobserwowalnych czynników indywidualnych każdego z N obiektów, $\mathbf{F}_\mu = \mathbf{I}_N \otimes \mathbf{j}_T$ ($\mathbf{I}_N$ – macierz jednostkowa wymiaru N, $\mathbf{j}_T$ – wektor jedynkowy wymiaru T), a $\mathbf{v}_j$ to wektor $NT\ x\ 1$ losowej części zakłócenia

jego macierz wariancji–kowariancji pomiędzy równaniami j i l (j, l = 1,…, M):

$$\boldsymbol{\Omega}_{jl} = E\left(\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}_l^T\right) = \sigma_{\mu_{jl}}^2(\mathbf{I}_N \otimes \mathbf{J}_T) + \sigma_{v_{jl}}^2(\mathbf{I}_N \otimes \mathbf{I}_T) \tag{3.55}$$

---

[44] Wektor $\mathbf{y}_j$ uporządkowany jest następująco: obserwacje i = 1, …, N w okresie t = 1, obserwacje i = 1, …, N w okresie t = 2, …, obserwacje i = 1, …, N w okresie t = T. Analogicznie prezentuje się struktura $\mathbf{X}_j$, $\boldsymbol{\beta}_j$ i $\boldsymbol{\varepsilon}_j$.

a dla całego modelu:

$$\mathbf{\Omega} = E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T) = \mathbf{\Sigma}_\mu \otimes (\mathbf{I}_N \otimes \mathbf{J}_T) + \mathbf{\Sigma}_v \otimes (\mathbf{I}_N \otimes \mathbf{I}_T) \tag{3.56}$$

gdzie: $\mathbf{J}_T$ – macierz jedynkowa wymiaru *T*, $\mathbf{I}_T$ – macierz jednostkowa wymiaru *T*, $\boldsymbol{\varepsilon}$ – wektor składników losowych wymiaru *MNK x* 1, wariancja $\boldsymbol{\mu}_j$ to $\mathbf{\Sigma}_\mu \otimes \mathbf{I}_N$, **a** $\mathbf{\Sigma}_\mu = \left[\sigma^2_{\mu_{jl}}\right]$, wariancja $\mathbf{v}_j$ to $\mathbf{\Sigma}_v \otimes \mathbf{I}_{NT}$, **a** $\mathbf{\Sigma}_v = \left[\sigma^2_{v_{jl}}\right]$ (obie wymiaru *M x M*).

W **modelu dwukierunkowej regresji z czynnikiem losowym (**ang. *Two--way Error Component Model*) składnik losowy modelu (3.53) ma postać:

$$\boldsymbol{\varepsilon}_j = \mathbf{F}_\mu \boldsymbol{\mu}_j + \mathbf{F}_\lambda \boldsymbol{\lambda}_j + \mathbf{v}_j \tag{3.57}$$

zdefiniowanego analogicznie do (3.54), gdzie: $\boldsymbol{\lambda}_j$ **to** wektor *T x* 1 czynników indywidualnych każdego z *T* okresów, $\mathbf{F}_\lambda = \mathbf{j}_N \otimes \mathbf{I}_T$ ($\mathbf{I}_T$ – macierz jednostkowa wymiaru *T*, $\mathbf{j}_N$ – wektor jedynkowy wymiaru *N*)

macierz wariancji–kowariancji pomiędzy równaniami *j* i *l* (*j*, *l* = 1,…, *M*):

$$\mathbf{\Omega}_{jl} = E\left(\boldsymbol{\varepsilon}_j \boldsymbol{\varepsilon}_l^T\right) = \sigma^2_{\mu_{jl}}(\mathbf{I}_N \otimes \mathbf{J}_T) + \sigma^2_{\lambda_{jl}}(\mathbf{J}_N \otimes \mathbf{I}_T) + \sigma^2_{v_{jl}}(\mathbf{I}_N \otimes \mathbf{I}_T) \tag{3.58}$$

a dla całego modelu:

$$\mathbf{\Omega} = E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T) = \mathbf{\Sigma}_\mu \otimes (\mathbf{I}_N \otimes \mathbf{J}_T) + \mathbf{\Sigma}_\lambda \otimes (\mathbf{J}_N \otimes \mathbf{I}_T) + \mathbf{\Sigma}_v \otimes (\mathbf{I}_N \otimes \mathbf{I}_T) \tag{3.59}$$

gdzie: $\mathbf{J}_N$ – macierz jedynkowa wymiaru *N*, wariancja $\boldsymbol{\lambda}_j$ to $\mathbf{\Sigma}_\lambda \otimes \mathbf{I}_T$, a $\mathbf{\Sigma}_\lambda = \left[\sigma^2_{\lambda_{jl}}\right]$, wymiaru *M x M*.

Parametry panelowych modeli wielorównaniowych o równaniach zależnych mogą być estymowane analogicznie do przekrojowych z uwzględnieniem specyfiki mikrodanych panelowych [Baltagi 2005: 13, 35, 115–117]. W mikrosymulacjach rzadko stosuje się tak zaawansowane modelowanie mikroekonometryczne jak panelowe modele wielorównaniowe estymowane systemowo metodami pełnej informacji, ze względu na ich skomplikowaną procedurę estymacji.

3.6. Podsumowanie

Dobór właściwych narzędzi mikroekonometrycznych, dostosowanych do specyfiki rozkładu zmiennych endogenicznych wpływa nie tylko na poprawę jakości modelu mikroekonometrycznego, ale także mikrosymulacyjnego. Poprawna specyfikacja modelu, szczególnie wybór postaci funkcyjnej oraz metoda estymacji parametrów strukturalnych mają kluczowe znaczenie dla poprawności wyników w skali mikro i makro badań ekonomiczno-społecznych eksperymentów mikrosymulacyjnych. Ekonometryczne pakiety komputerowe zawierają coraz więcej modułów umożliwiających dostosowanie modelu do specyfiki rozkładu zmiennych dla danych niezagregowanych, przekrojowych i panelowych jak również dobranie najlepszej z punktu widzenia własności estymatora metody szacowania wartości parametrów strukturalnych. Sprzyja to popularyzacji modeli mikroekonometrycznych w modelach mikrosymulacyjnych, a w efekcie prowadzenie zaawansowanych badań ekonomicznych i społecznych z zastosowaniem mikrosymulacji.

Zaprezentowane w rozdziale podejście do modeli mikroekonometrycznych i metod estymacji stanowi kompleksowe i syntetyczne zestawienie, które powinno być pomocne w konstruowaniu modeli mikroekonometrycznych zawartych w modelach mikrosymulacyjnych wykorzystywanych w badaniach ekonomiczno-społecznych.

ZAKOŃCZENIE

Rozwój kulturowy, społeczny i gospodarczy społeczeństw na całym świecie w połączeniu z postępującą globalizacją sprawił, że sieć powiązań różnych dziedzin życia i nauki nie pozwala już na prowadzenie monotematycznych i jednokryterialnych badań ekonomicznych i społecznych. Konieczne stało się konstruowanie nowych interdyscyplinarnych i wielokryterialnych narzędzi prowadzenia analiz wspomagających proces podejmowania decyzji na różnych szczeblach zarządzania. W szczególności decydenci zainteresowani są aplikacją metod umożliwiającey ocenę przyszłych skutków prowadzonych i planowanych polityk. W niniejszej monografii podjęto temat modeli mikrosymulacyjnych jako interdyscyplinarnych i wielowymiarowych narzędzi służących prognozowaniu i symulowaniu sytuacji makrosystemów poprzez agregacją wartości cech dla obiektów indywidualnych. Opracowanie swoim zakresem merytorycznym objęło zagadnienia dotyczące teorii i aplikacji ekonomiczno-społecznych modeli mikrosymulacyjnych. Rozważając zasadniczy problem badawczy, tzn. czy modele mikrosymulacyjne stosowane w badaniach ekonomiczno-społecznych są skutecznym narzędziem wspomagającym proces podejmowania decyzji na różnych szczeblach zarządzania, skoncentrowano się na aspektach merytorycznych i aplikacyjnych. Umożliwiło to realizację celu głównego oraz wszystkich celów szczegółowych. W pierwszej części dokonano syntetycznego i całościowego zestawienia dotychczasowej wiedzy teoretycznej w zakresie mikrosymulacji i modeli mikrosymulacyjnych. Wskazano etapy konstrukcji tych modeli, ich rodzaje oraz wymagania. Następnie dokonano przeglądu istniejących modeli ekonomiczno-społecznych zawracając uwagę na ich zalety, wady, możliwości praktyczne i ograniczenia. Wszystkie zaprezentowane modele nadal funkcjonują wspomagając prowadzenie badań i analiz ekonomicznych i społecznych w sektorach publicznym, prywatnym i akademickim. W przeglądzie zwrócono uwagę na nieco przestarzałe podejście do stosowanych w modelach mikrosymulacyjnych submodeli ekonometrycznych. W większości przypadków brak jest pełnej informacji o konstrukcji, estymacji i weryfikacji tych podmodeli. Jako, że jakość modeli ekonometrycznych wpływa na wiarygodność wyników eksperymentów mikrosymulacyjnych, zdecydowano się na prezentację metodologii mikroeko-

nometrycznej, która mogłaby okazać się przydatna w konstrukcji submodeli ekonometrycznych i estymacji ich parametrów. W syntezie metodologicznej zwrócono szczególną uwagę na własności mikrodanych oraz na niepożądane konsekwencje wykorzystania nieodpowiednich, ze względu na własności estymatorów, metod estymacji parametrów modeli mikroekonometrycznych związanych z danymi indywidualnymi.

Realizując założone cele podjęto próbę walidacji lub falsyfikacji założonych hipotez badawczych dotyczących (1) przydatności modeli mikrosymulacyjnych do oceny rozkładu efektów prowadzonych polityk ekonomicznych i społecznych oraz (2) wpływu wyboru metody estymacji parametrów strukturalnych submodelu ekonometrycznego na wyniki mikrosymulacji. Pierwszą hipotezę przyjęto, ponieważ przegląd istniejących modeli mikrosymulacyjnych stosowanych w badaniach ekonomicznych i społecznych, wskazał na możliwości prowadzenia wielowymiarowych analiz rozkładów efektów prowadzonych i proponowanych polityk społecznych i gospodarczych. W przypadku drugiej hipotezy przeprowadzone syntezy wskazały, że (1) dobór metody estymacji parametrów modelu mikroekonometrycznego wpływa na jakość oszacowań parametrów strukturalnych, co określa przydatność tego modelu oraz (2) jakość submodelu ekonometrycznego wpływa na jakość zawierającego go modelu mikrosymulacyjnego, czyli również wyników eksperymentów mikrosymulacyjnych. Potwierdziło to, że dobór metody estymacji parametrów submodelu mikroekonometrycznego determinuje jakość rezultatów eksperymentów mikrosymulacyjnych.

W procesie realizacji przyjętych celów i weryfikacji hipotez badawczych napotkano ograniczenia i utrudnienia. Próba dokonania syntezy informacji o mikrosymulacjach i funkcjonujących modelach mikrosymulacyjnych okazała się trudniejsza niż zakładano ze względu na nieliczne i często pobieżne pozycje literaturowe. Niektóre szczegóły dotyczące m. in. metod konstrukcji, estymacji i weryfikacji submodeli ekonometrycznych modeli mikrosymulacyjnych objęte są nadal tajemnicą i nigdy nie zostały upublicznione. Dotarcie do tych informacji okazało się niemożliwe. Przestarzała polityka modelu mikrosymulacyjnego jako „czarnego/magicznego pudełka” stała się w ostatnich latach tematem wielu dyskusji wśród konstruktorów i użytkowników mikrosymulacji. Mimo licznych apeli władz Międzynarodowego Towarzystwa Mikrosymulacji (ang. *International Microsimulation Association*) nie udostępniono pełnych dokumentacji funkcjonujących modeli mikrosymulacyjnych. Utrudnia to prowadzenie analiz i wnioskowanie o jakości tych modeli oraz wyników prowadzonych w ich oparciu eksperymentów.

Chociaż w ostatnich latach można zaobserwować szybki proces popularyzacji mikrosymulacji w Polsce i na świecie, to jego kontynuacja w kolejnych okresach będzie uwarunkowana rozwojem w czterech zasadniczych zakresach: **teoria, tematyka, bazy danych** i **oprogramowanie**. Obecnie aspekty metodologiczne mikrosymulacji prezentowane są bardzo pobieżnie, zazwyczaj we wstę-

pach publikacji aplikacyjnych. Często traktowane są one jako elementy oczywiste dla „wtajemniczonych" konstruktorów i użytkowników modeli mikrosymulacyjnych. Stanowi to jednak barierę wejścia dla nowych analityków. Dlatego w niniejszej monografii podjęto próbę syntezy dotychczasowej wiedzy teoretycznej o budowie, etapach, klasyfikacjach i definicjach modeli, parametrów, baz danych i eksperymentów mikrosymulacyjnych. Jednak dalszy rozwój metodologii mikrosymulacyjnej i jej upowszechnianie stanowi warunek konieczny dla popularyzacji mikrosymulacji. Wiele godnych uwagi aspektów modelowania mikrosymulacyjnego jest nadal słabo poznanych i opisanych. Powstanie „instrukcji obsługi" pomogłoby w propagowaniu narzędzi mikrosymulacyjnych w badaniach ekonomicznych i społecznych.

Wielokierunkowość i interdyscyplinarność aplikacji mikrosymulacji stanowi ważny element ich popularności. Są one uniwersalnym narzędziem prowadzenia symulacji w różnych dziedzinach nauki. Można oczekiwać, że spektrum tematyczne modeli mikrosymulacyjnych będzie się rozszerzać, co powinno przyczynić się do rozwoju i upowszechniania metodologii mikrosymulacyjnej. Pierwotnie modele stosowane były w badaniach ekonomiczno-społecznych dotyczących świadczeń społecznych i podatków. Coraz częściej podejmowane są próby konstrukcji zarówno modeli bardzo wyspecjalizowanych jak i ogólnych. W przypadku tych pierwszych pojawiają się próby implementacji metodyki związanej z sektorem zdrowia, ekonometrii przestrzennej z zastawaniem macierzy wag dla kooperujących regionów, zarządzania kryzysowego na wypadek katastrof naturalnych czy elementów kulturowych i lingwistycznych. Można oczekiwać dalszej ekspansji tematycznej i specjalizacji modeli mikrosymulacyjnych. Z drugiej strony trwający rozwój technologiczny zwiększający moce obliczeniową komputerów osobistych zachęca konstruktorów do pracy nad modelami bardziej wszechstronnymi. Podejmowane są działania na rzecz budowy wielomodułowych modeli mikrosymulacyjnych, które w kompleksowy sposób ujmują problemy ekonomiczno-społeczne. Umożliwią one prowadzenie analiz nie tylko interdyscyplinarnych, ale także multidycyplinarnych pomagając zrozumieć i symulować skomplikowane relacje pomiędzy obiektami badania, instytucjami i normami prawnymi. W rezultacie rozwój tematyczny modeli wyspecjalizowanych i ogólnych powinien zapewnić badaczom możliwość optymalizacji właściwości modeli mikrosymulacyjnych ze względu na cel analiz, możliwości finansowe i technologiczne. Taka personalizacja modeli mikrosymulacyjnych stanowić będzie o uniwersalności aplikacyjnej mikrosymulacji.

Trzecim elementem, którego rozwój zapewni postęp w popularyzacji mikrosymulacji są bazy mikrodanych. Czynnik ten jest silnie zróżnicowany względem krajów i norm prawnych. W niektórych krajach, np. w Szwecji, tradycja zbierania i udostępniania informacji ze spisów powszechnych sprzyja stosowaniu mikrosymulacji. W innych, m. in. w Polsce, badania pełne prowadzone są rzadko, a dane pierwotne nie są dostępne. Ze względu na wysoką danochłonność

zastosowanie mikrosymulacji wymaga informacji o pełnych populacjach lub, gdy jest to niemożliwe, o próbach reprezentacyjnych. Zastosowanie baz danych wtórnych dla prób reprezentacyjnych znacząco obniżyłoby koszty zastosowania modeli mikrosymulacyjnych do prowadzenia badań ekonomiczno-społecznych. Zakres tematyczny tych danych powinien być jednak dostosowany do profilu tematycznego badania. W konstrukcji modeli mikrosymulacyjnych niezbędne są również makrodane konieczne do wprowadzenia parametrów zewnętrznych mikrosymulacji. Na szczęście metodologia analiz mikrodanych zyskuje na popularności, co powinno zaowocować rozszerzeniem badań indywidualnych i wzrostem dostępności ich wyników niezagregowanych. Powstawanie nowych baz danych dla osób, gospodarstw domowych, przedsiębiorstw i mikroregionów powinno sprzyjać rozwojowi mikrosymulacji. Jednocześnie nie tylko ilość, ale także jakość banków danych ma istotne znaczenie. Niektóre duże ośrodki badawcze specjalizujące się w konstrukcji modeli mikrosymulacyjnych współpracując z władzami, instytucjami rządowymi i prywatnymi prowadzą badania ankietowe na potrzeby konkretnych mikrosymulacji. Gwarantuje to dostępność wszystkich istotnych zmiennych. Kooperacja akademicko-naukowo-rządowo-komercyjna pozwala rozłożyć koszty oraz uzyskać odpowiedni profil i strukturę baz mikrodanych. Dlatego można oczekiwać, że ten system współpracy otworzy nowe możliwości w konstruowaniu banków informacji, a zatem także w aplikacji modeli mikrosymulacyjnych.

Rozwój teorii i zakresu tematycznego aplikacji mikrosymulacji oraz dostępności baz mikrodanych nie wystarczą jednak, by spopularyzować metodologię modeli mikrosymulacyjnych. Ważnym elementem jest ułatwienie procesu konstrukcji modeli i przebiegu eksperymentów mikrosymulacyjnych. Obecnie budowa modelu oznacza konieczność oprogramowania go samodzielnie w dowolnym języku programowania. Podjęto już pierwszą próbę udostępnienia platformy do konstrukcji modeli mikrosymulacyjnych – MODGEN. Jednak jest to jedynie nakładka w języku C++ ułatwiająca definiowanie obiektów i zdarzeń. Instrukcja obsługi jest mało intuicyjna, stosowanie MODGEN wymaga sporej wprawy w programowaniu komputerowym. Upowszechnienie metodologii mikrosymulacyjnej wymaga dostarczenia użytkownikom prostego narzędzia wspomagającego kolejne etapy konstrukcji modelu i przeprowadzania eksperymentu. Na razie trudno stwierdzić, kto i kiedy udostępni taki program lub *toolbox*, jednak jego powstanie jest jedynie kwestią czasu.

Modele mikrosymulacyjne są uniwersalnym i interdyscyplinarnym narzędziem wspomagającym prowadzenie badań ekonomicznych i społecznych. W ciągu ostatnich lat zyskały wielu zwolenników, a ich dalszy rozwój merytoryczny, aplikacyjny i technologiczny powinien przyczynić się do uproszczenia procesu konstrukcji i oprogramowania modeli mikrosymulacyjnych i ich submodeli mikroekonometrycznych, sprawiając, że staną się bardziej przyjazne dla użytkowników.

BIBLIOGRAFIA

Abello A., Lymer S., Brown L., Harding A., Phillips B. [2008], *Enhancing the Australian National Health Survey Data for Use in a Microsimulation Model of Pharmaceutical Drug Usage and Cost*, [w:] „Journal of Artificial Societies and Social Simulation", vol. 11(3)2 [http://jasss.soc.surrey.ac.uk/11/3/2.html].

Ahmed V., O'Donoghue C. [2010], *Global Economic Crisis and Poverty in Pakistan*, „International Journal of Microsimulation", vol. 3(1).

Analiza reform z uwzględnieniem podaży pracy – reformy podatkowe i zasiłek dla pracujących, Wydział Nauk Ekonomicznych Uniwersytetu Warszawskiego [www.simpl.pl (20.05.2010)].

Anderson J. M. [1997a], *Models for Retirement Policy Analysis, Chapter 1 Modeling for Retirement Policy Analysis* [http://www.soa.org/research/pension/research-models-for-retirement-policy-analysis.aspx (18.03.2009)].

Anderson J. M. [1997b], *Chapter 2. Microanalitic simulation models*, [w:] *Models for Retirement Policy Analysis, Report to the Society of Actuaries*, Capital Research Associates [http://www.soa.org/research/pension/research-models-for-retirement-policy-analysis.aspx (18.03.2009)], [http://www.soa.org/files/pdf/Chapter_2.pdf (18.03.2009)].

Anderson J. M. [1997c], *Chapter 3. DYNASIM*, [w:] *Models for Retirement Policy Analysis, Report to the Society of Actuaries*, Capital Research Associates [http://www.soa.org/research/pension/research-models-for-retirement-policy-analysis.aspx (18.03.2009)], [http://www.soa.org/files/pdf/Chapter_3.pdf (18.03.2009)].

Anderson J. M. [1997d], *Chapter 5. CORSIM*, [w:] *Models for Retirement Policy Analysis, Report to the Society of Actuaries*, Capital Research Associates [http://www.soa.org/research/pension/research-models-for-retirement-policy-analysis.aspx (18.03.2009)], [http://www.soa.org/files/pdf/Chapter_5.pdf (18.03.2009)].

Anderson J. M. [1998], *Chapter 6. DYNACAN*, [w:] *Models for Retirement Policy Analysis, Report to the Society of Actuaries*, Capital Research Associates [http://www.soa.org/research/pension/research-models-for-retirement-policy-analysis.aspx (18.03.2009)], [http://www.soa.org/files/pdf/Chapter_6.pdf (18.03.2009)].

Anderson J. M. [1999], *Chapter 11. PIMS*, [w:] *Models for Retirement Policy Analysis, Report to the Society of Actuaries*, Capital Research Associates [http://www.soa.org/research/pension/research-models-for-retirement-policy-analysis.aspx (18.03.2009)], [http://www.soa.org/files/pdf/Chapter_11.pdf (18.03.2009)].

Anderson J. M. [2001], *Chapter 4. PENSIM*, [w:] *Models for Retirement Policy Analysis, Report to the Society of Actuaries*, Capital Research Associates [http://www.soa.org/research/pension/research-models-for-retirement-policy-analysis.aspx (18.03.2009)], [http://www.soa.org/files/pdf/Chapter_4.pdf (18.03.2009)].

Antoszkiewicz J. [1990], *Metody heurystyczne. Twórcze rozwiązywanie problemów*, PWE, Warszawa.

Atkinson A. B., Bourguignon F., O'Donoghue C., Sutherland H., Utili F. [1999], *Microsimulation and the formulation of policy: A case study of targeting in the European Union*, [w:] Working Papers Series, Working Paper No. EM2/99 [http://www.iser.essex.ac.uk/publications/work-ing-papers/euromod/em0-99.pdf (12.11.2011)].

Baltagi B. H. [2005], *Econometric Analysis of Panel Data*, John Wiley and Sons Ltd.

Balzano S., Oropallo F., Parisi V. [2011], *On the Italian ACE and its impacton enterprise performance: A PLS-path modeling analysis*, „International Journal of Microsimulation", vol. 4(1).

Bargain O., Morawski L., Myck M., Socha M. [2007a], *As SIMPL as that: Introducing a Tax-Benefit Microsimulation Model for Poland*, „Discussion Paper", no. 1 [http://ftp.iza.org/dp2988.pdf (12.11.2011)].

Bargain O., Morawski L., Myck M., Socha M. [2007b], *SIMPL: A microsimulation model for Poland*, Working Paper International Microsimulation Association [http://www.euro.centre.org/ima2007/programme/papers/Morawski_paper.pdf (13.11.2011)].

Baroni E., Richiardi M. [October 2007], *Orcutt's Vision, 50 years on*, Working Paper no. 65.

Bazyl M. [2010], *7. Modele zmiennych licznikowych*, [w:] M. Gruszczyński (red.), *Mikroekonometria. Modele i metody analiz danych indywidualnych*, Oficyna a Wolters Kluwer business, Warszawa.

Berliri C., Parisi V. [2006], *The Distributive and Labor Supply Impact of the Minimum Insertion Income: The Case of Italy*, „International Advances in Economic Research", vol. 12.

Blue V. J., Adler J. L. [2001], *Cellular automata microsimulation for modeling bi-directional pedestrian walkways*, „Transportation Research Part B", vol. 35, Elsevier.

Bourguignon F., Bussolo M., Cockburn J. [2010], *GUEST EDITORIAL Macro-Micro Analytics: Background, Motivation, Advantages and Remaining Challenges*, „International Journal Of Microsimulation", vol. 3(1).

Bourguignon F., Spadaro A. [2006], *Microsimulation as a tool for evaluating redistribution policies*, „Journal of Economic Inequality", vol. 4, Springer.

Caldwell S., Morrison R. J. [2000], *9. Validation of longitudinal dynamic microsimulation models: Experience with CORSIM and DYNACAN*, [w:] L. Mitton, H. Sutherland, M. J. Weeks (red.), *Microsimulation modelling for policy analysis: Challenges and innovations*, Cambridge University Press, Cambridge.

Cameron A. C., Trivedi P. K. [2009], *Microeconometrisc. Methods and Applications*, Cambridge University Press, Cambridge, Nowy Jork, Melbourne, Madryt, Cape Town, Singapur, São Paulo, Delhi.

Chen X., Meaker J. W., Zhan F. B. [2005], *Agent-Based Modeling and Analysis of Hurricane Evacuation Procedures for the Florida Keys*, „Natural Hazards", vol. 38, Springer.

Chitiga M., Cockburn J., Decaluwé B., Fofana I., Mabugu R. [2010], *Case Study: A Gender-focused Macro-Micro Analysis of the Poverty Impacts of Trade Liberalization in South Africa*, „International Journal Of Microsimulation", vol. 3(1).

Citro C. F., Hanushek E. A. (red.) [1991], *Improving Information for Social Policy Decisions, volume I: Review and Recommendations*, National Academy Press, Waszyngton.

Colombo G. [2010], *Linking CGE and Microsimulation Models: A Comparison of Different Approaches*, „International Journal Of Microsimulation", vol. 3(1).

Cororaton C. B., Cockburn J. [2007], *Trade reform and poverty – Lessons from thePhilippines: A CGE-microsimulation analysis*, „Journal of Policy Modeling", vol. 29.

Creedy J., Duncan A. S., Harris M., Scutella R. [2002], *Microsimulation Modelling of Taxation and the Labour Market: The Melbourne Institute Tax and Transfer Simulation*, Edward Elgar Publishing, Cheltenham.

Creedy J., Kalb G. [2006], *Labour supply and microsimulation: The evaluation of tax policy reforms*, Edward Elgar, Cheltenham.

Cumpston J. R. [2010], *Alignment and matching in multi-purpose household microsimulations*, „International Journal of Microsimulation", vol. 3(2).

Dańska B. [2000], *Przestrzenno-czasowe modelowanie zmian w działalności produkcyjnej w Polsce. Zastosowanie modeli panelowych*, [w:] B. Suchecki (red.), *Dane panelowe i modelowanie wielowymiarowe w badaniach ekonometrycznych*, t. 1, Absolwent, Łódź.

Dańska-Borsiak B. [2011], *Dynamiczne modele panelowe w badaniach ekonomicznych*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź.

de Souza Ferreira Filho J. B., dos Santos C. V., do Prado Lima S. M. [2010], *Tax Reform, Income Distribution and Poverty in Brazil: An Applied General Equilibrium Analysis*, „International Journal of Microsimulation", vol. 3(1).

Decoster A., Loughrey J., O'Donoghue C., Verwerft D. [2011], *Microsimulation of indirect taxes*, „International Journal of Microsimulation", vol. 4(2).

Detyna B., Detyna J. [2011], *Jakość usług medycznych. Ocena statystyczna. Podstawy metodyczne*, Difin, Warszawa.

Domański Cz. [2001], *4. Elementy rachunku prawdopodobieństwa*, [w:] Cz. Domański (red.), *Metody statystyczne. Teoria i zadania*, wyd. VI zmienione, Wydawnictwo Uniwersytetu Łódzkiego, Łódź.

Drukker D. M. [2003], *Testing for serial correlation in linear panel-data models*, „The Stata Journal", vol. 3, no. 2.

Duncan A., Weeks M. [2000], *14. Transition estimates in discrete choice models*, [w:] L. Mitton, H. Sutherland, M. J. Weeks (red.), *Microsimulation modelling for policy analysis: Challenges and innovations*, Cambridge University Press, Cambridge.

Dupuis A., Chopard B. [2003], *Cellular Automata Simulations of Traffic: A Model for the City of Geneva*, „Networks and Spatial Economics", vol. 3, Kluwer Academic Publishers.

Fischer E. A. J., De Vlas S. J., Richardus J. H., Habbema J. D. F. [2008], *MUSIDH, multiple use of simulated demographic histories, a novel method to reduce computation time in microsimulation models of infectious diseases*, „Computer methods and programs in biomedicine", vol. 91.

Frejlich J., Nicińska A., *EUROMOD – podatkowo-zasiłkowy europejski model mikrosymulacyjny* [http://www.simpl.pl/dowload/euromod.pdf (16.01.2010)].

Gajda J. B. [1988], *Wielorównaniowe modele ekonometryczne. Estymacja – symulacja – sterowanie*, Wydawnictwo Naukowe PWN, Warszawa.

Gajda J. B. [2004], *Ekonometria. Wykład i łatwe obliczenia w programie komputerowym!*, Wydawnictwo C. H. Beck, Warszawa.

Gierszewska G., Romanowska M. [2004], *Analiza strategiczna przedsiębiorstwa*, PWE, Warszawa.

Grimm M. [2005], *Educational policies and poverty reduction in Côte d'Ivoire*, „Journal of Policy Modeling", vol. 27, Elsevier.

Gruszczyński M. [2010a], *1. Wprowadzenie do mikroekonometrii*, [w:] M. Gruszczyński (red.), *Mikroekonometria. Modele i metody analiz danych indywidualnych*, Oficyna a Wolters Kluwer business, Warszawa.

Gruszczyński M. [2010b], *3. Modele zmiennych jakościowych dwumianowych*, [w:] M. Gruszczyński (red.), *Mikroekonometria. Modele i metody analiz danych indywidualnych*, Oficyna a Wolters Kluwer business, Warszawa.

Gruszczyński M., Podgórska M [1996], *Ekonometria*, Oficyna Wydawnicza SGH, Warszawa.

Guidelines for the Use of Microsimulation Sofware [lipiec 2007] [http://www.highways.gov.uk/knowledge_compendium/assets/documents/Portfolio/Guidelines%20-%20901.pdf (03.07.2011)].

Gupta A., Harding A. [2007], *Chapter 1. Introduction and overview*, [w:] A. Gupta, A. Harding (red.), *Modeling our future: Population ageing health and aged care*, Elsevier.

Haan P., Myck M. [2009], *Multi-family households in a labour supply model: A calibration method with application to Poland*, CenEAWorking Paper Series WP04/09 [http://www.cenea.org.pl/images/stories/pdf/working_papers/cenea_wp_0409.pdf (17.10.2012)].

Haveman R. H., Hollenbeck K. [1980], *Preface*, [w:] R. H. Haveman, K. Hollenbeck (red.), *Microeconomic simulation model for public policy analysis, volume 1: Distributional Impacts*, Academic Press, Nowy Jork, Londyn, Toronto, Sydney, San Francisco.

Heiss F. [2006], *Nonlinear State-Space Models for Microeconometric Panel Data*, Munich Discussion Paper 2006-24, Department of Economics, University of Munich [http://epub.ub.uni-muenchen.de/1157/1/Heiss_WP2006-22.pdf (15.07.2011)].

Hérault N. [2010], *Sequential Linking of Computable General Equilibrium and Microsimulation Models: A Comparison of Behavioural and Reweighting Techniques*, „International Journal Of Microsimulation", vol. 3(1).

Hiligsmann M., Ethgen O., Bruyère O., Richy F., Gathon H. J., Reginster J. Y. [2009], *Development and Validation of a Markov Microsimulation Model for the Economic Evaluation of Treatments in Osteoporosisvhe*, „Value In Health", vol. 12, no. 5.

Hollander Y., Liu R. [2008], *Calibration of a traffic microsimulation model as a tool for estimating the level of travel time variability*, „Advanced OR and AI Methods in Transportation" [http://www.iasi.cnr.it/ewgt/16conference/ID5.pdf (05.07.2011)].

Hoshen M. B., Burton A. H., Bowcock T. J. V. [2007], *Simulating disease transmission dynamics at a multi-scale level*, „International Journal of Microsimulation", vol. 1(1).

Hout B. V., Angus D. C. [2004], *Dynamic microsimulation to model multiple outcomes in cohorts of critically ill patients*, „Intensive Care Med", vol. 30, Springer-Verlag.

Hozer J. [1993], *Mikroekonometria. Analizy. Diagnozy. Prognozy*, PWE, Warszawa.

Int Panis L., Broekx S., Liu R. [2006], *Modelling instantaneous traffic emission and the influence of traffic speed limits*, „Science of the Total Environment", vol. 371, Elsevier.

Kalwij A., Alessie R., Fontein P. [1998], *Household commodity demand and demographics in the Netherlands: A microeconometric analysis*, „Journal of Population Economics", vol. 11, Springer-Verlag.

Kazimi C. [1997], *Evaluating the Environmental Impact of Alternative-Fuel Vehicles*, „Journal of Environmental Economics and Management", vol. 33.

Klette T. J., Jarle Møen J., Griliches Z. [2000], *Do subsidies to commercial R&D reduce market failures? Microeconometric evaluation studies*, „Research Policy", vol. 29.

Klevmarken A., Björn L. [2008], *Simulating an Ageing Population. A microsimulation approach applied to Sweden*, Emerald, Bingley.

Kluve J., Lehmann H., Schmidt C. M. [1999], *Active Labor Market Policies in Poland: Human Capital Enhancement, Stigmatization, or Benefit Churning?*, „Journal of Comparative Economics", vol. 27.

Konopczak K., Skibicki J. [2012], *Mikrosymulacyjny model podatkowo-zasiłkowy Ministerstwa Finansów – dokumentacja*, MF Working Paper Series no. 12–2012 [http://www.mf.gov.pl/_files_/euro/publikacje_naukowe/working_papers/mf_wp_12.pdf (17.10.2012)].

Kruk T. J. [2009], *Systemy operacyjne. Wprowadzenie*, Instytut Automatyki i Informatyki Stosowanej, Politechnika Warszawska [http://www.ia.pw.edu.pl/~tkruk/edu/soi.b/wyklad/w01.pdf (27.12.2009)].

Krysicki W., Bartos J., Dyczka W., Królikowska K., Wasilewski M. [2002], *Rachunek prawdopodobieństwa i statystyka matematyczna w zadaniach. Część I rachunek prawdopodobieństwa*, Wydawnictwo Naukowe PWN, Warszawa.

Krysicki W., Włodarski L. [2002], *Analiza matematyczna w zadaniach, część II*, Wydawnictwo Naukowe PWN, Warszawa.

Książek M. [2010], *4. Modele zmiennych wielomianowych uporządkowanych*, [w:] M. Gruszczyński (red.), *Mikroekonometria. Modele i metody analiz danych indywidualnych*, Oficyna a Wolters Kluwer business, Warszawa.

Lazaridis P. [2004], *Olive Oil Consumption in Greece: A Microeconometric Analysis*, „Journal of Family and Economic Issues", vol. 25(3), Springer Science + Business Media Inc.

Lee M. [2010], *Micro-Econometrisc. Methods of moments and Limited Dependent Variables*, Wyd. II, Springer, Nowy Jork, Dordrecht, Heidelberg, Londyn.

Li J., O'Donoghue C. [2011], *Evaluating Alignment Methods in Dynamic Microsimulation Models* [http://www.scb.se/Grupp/Produkter_Tjanster/Kurser/_Dokument/IMA/li_o_donoghue 2010.pdf (05.07.2011)].

Loomes G., Moffatt P. G., Sugden R. [2002], *A Microeconometric Test of Alternative Stochastic Theories of Risky Choice*, „The Journal of Risk and Uncertainty", vol. 24:2, Kluwer Academic Publishers.

Lymer S. [2011], *Moving beyond comparative validation: Predictive abilities of APPSIM's health module*, [http://www.scb.se/Grupp/Produkter_Tjanster/Kurser/_Dokument/IMA/Lymer_IMA_paper_validation.pdf (03.07.2011)].

Lymer S., Brown L., Harding A., Yap M. [2009], *Predicting the Need for Aged Care Services at the Small Area Level: The CAREMOD Spatial Microsimulation Model*, „International Journal of Microsimulation", vol. 2(2).

Lymer S., Brown L., Yap M., Harding A. [2008], *2001 Regional Disability Estimates for New South Wales, Australia, Using Spatial Microsimulation*, „Applied Spatial Analysis", vol. 1.

Maddala G. S. [1983], *Limited-dependent and qualitative variables in econometrics*, Econometric Society Monographs no. 3, Cambridge University Press, Cambridge, New York, Melbourne, Cape Town, Singapore, Sao Paulo, Delhi.

Marchionni M., Gasparini L. [2007], *Tracing out the effects of demographic changes on the income distribution. The case of Greater Buenos Aires, 1980–1998*, „Journal of Economic Inequality", vol. 5.

Mason C. [2011], *Socsim Oversimplified, skrócona dokumentacja modelu Uniwersytetu Berkeley, Kalifornia* [http://lab.demog.berkeley.edu/socsim/socsimOversimplified.pdf (22.11.2011)].

Menneni S., Sun C., Vortisch P. [2007], *Microsimulation Calibration Using Speed Flow Relationships* [http://www.ptvamerica.com/fileadmin/files_ptvamerica.com/Microsimulation_Calibration_Using_Speed_Flow_Relationships.pdf (05.07.2011)].

Miszczyńska D., Miszczyński M. [2002], *Wybrane metody badań operacyjnych*, Wyższa Szkoła Ekonomiczno-Humanistyczna w Skierniewicach, Skierniewice.

Morawski L., Myck M. [2009], *'Klin'-ing up: Effects of Polish tax reforms on those in and on those out*, CenEAWorking Paper Series WP05/09 [http://www.sciencedirect.com/science/article/B6VFD-4WDGCNJ-2/2/02e62bb9f28ea8e47a14d528baed9e14 (15.11.2011)].

Moreno L., Aguilar R. M., Martin C. A., Piñeiro J. D., Estevez J. I., Sigut J. F., Sanchez J. L., Jimenez V. I. [1999], *Patient-centered simulation tool for aiding in hospital management*, „Simulation Practice and Theory", vol. 7, Elsevier.

Mueller G. P. [2011], *Microsimulation of Virtual Encounters: A New Methodology for the Analysis of Socio-Cultural Cleavages*, „International Journal of Microsimulation", vol. 4(1).

Murphy M. [2009], *A micro simulation model for projecting family and kinship networks in Britain*, prezentacja dla ESRC Micro-simulation Seminar Series, Leeds, 02.07.2009 [http://www.google.pl/url?sa=t&rct=j&q=socsim%20microsimulation%20model&source=web&cd=3&ved=0CC0QFjAC&url=http%3A%2F%2Fpcwww.liv.ac.uk%2F~william%2FESRC-BSPS%2520MSM%2520Seminar%2520Series_files%2FMurphy.ppt&ei=asy_TsLcK43BtAbBpMmZDQ&usg=AFQjCNHtb8lTZ256_P6vCmcPAGyXreTdJg&cad=rja (22.11.2011)].

Orcutt G. H. [1957], *A New Type of Socio-Economic System*, „The Review of Economics and Statistics", vol. 39, no. 2.

Orcutt G. H., Caldwell S., Wertheimer R. F. [1976], *Policy exploration through microanalytic simulation*, The Urban Institute, Washington D.C.

Orcutt G. H., Glazar A., Harris R., Wertheimer III R. [1978], *Microanalytic Modeling and the Analysis of Public Transfer Policies*, [w:] R. H. Haveman, K. Hollenbeck (red.), *Microeconomic simulation model for public policy analysis, volume 1: Distributional Impacts*, Academic Press, Nowy Jork, Londyn, Toronto, Sydney, San Francisco.

Osborn A. F. [1953], *Applied Imagination: Principles and Procedures of Creative Problem Solving*, Charles Scribner's Sons, Nowy Jork.

Osińska M., Kośko M. [2007], *Rozdział 3. Estymacja jednorównaniowego liniowego modelu ekonometrycznego*, [w:] M. Osińska (red.), *Ekonometria współczesna*, Dom Organizatora, Toruń.

Osińska M., Kośko M. [2007], *Rozdział 7. Uogólniona metoda najmniejszych kwadratów i modele zmiennych jakościowych*, [w:] M. Osińska (red.), *Ekonometria współczesna*, Dom Organizatora, Toruń.

Owczarczuk M. [2010], *6. Modele zmiennych ograniczonych*, [w:] M. Gruszczyński (red.), *Mikroekonometria. Modele i metody analiz danych indywidualnych*, Oficyna a Wolters Kluwer business, Warszawa.

Pendyala R. M., Kitamura R., Chen C., Pas E. I. [1997], *An activity-based microsimulation analisys of transportation control measures*, „Transportation Policy", vol. 4, no. 3, Elsevier Science LTD.

Pohlmeier W., Ronning G. [2008], *Microeconometrics and disclosure control*, „AStA Advances in Statistical Analysis", vol. 92, Springer-Verlag.

Puvimanasinghe J. P. A., Takkenberg J. J. M., Eijkemans M. J. C., Steyerberg E. W., van Herwerden L. A., Grunkemeier G. L., Habbema J. D. F., Bogers A. J. J. C. [2003], *Choice of a mechanical valve or a bioprosthesis for AVR: Does CABG matter?*, „European Journal of Cardio-thoracic Surgery", vol. 23, Elsevier.

Puvimanasinghe J. P. A., Takkenberg J. J. M., Eijkemans M. J. C., van Herwerden L. A., Jamieson W. R. E., Grunkemeier G. L., Habbema J. D. F., Bogers A. J. J. C. [2006], *Comparison of Carpentier-Edwards pericardial and supraannular bioprostheses in aortic valve replacement*, „European Journal of Cardio-thoracic Surgery", vol. 29, Elsevier.

Rahman A., Harding A., Tanton R., Liu S. [2010], *Methodological Issues in Spatial Microsimulation Modelling for Small Area Estimation*, „International Journal of Microsimulation", vol. 3(2).

Raihan S. [2010], *Welfare and Poverty Impacts of Trade Liberalization: A Dynamic CGE Microsimulation Analysis*, „International Journal of Microsimulation", vol. 3(1).

Raport 2030, Część 2. Ocena skutków wdrożeń pakietu energetyczno-klimatycznego dla Polski w okresie do roku 2030, Załącznik 6. Opis modelu równowagi ogólnej CGE-PL [wersja z dnia 30.09.2008] [ftp://cire.pl/pliki/2/PKEE_Raport%202030_Czesc_2_Zal_6_2008-09-30.pdf (02.07.2011)].

Ronning G. [2006], *Microeconometric models and anonymized micro data*, „Allgemeines Statistisches Archiv", vol. 90, Physic-Verlag.

Roorda M. J., Miller E. J., Habib K. M. N [2008], *Validation of TASHA: A 24-h activity scheduling microsimulation model*, „Transportation Research Part A", vol. 42, Elsevier.

Rubin D. B. [2004], *Multiple imputation for nonresponse in surveys*, John Wiley and Sons Ltd., Nowy Jork.

Sartor N. [2001], The Long-run Effects of the Italian Pension Reforms, „International Tax and Public Finance", vol. 8, Kluwer Academic Publishers.

Schofield1 D., Passey M., Callander E., Percival R., Kelly S., Shrestha R., Fletcher S. [2011a], *Modelling the cost of ill health in Health&WealthMOD [Version II]: Lost labour force participation, income and taxation, and the impact of disease prevention*, „International Journal of Microsimulation", vol. 4(3).

Schofield1 D., Passey M., Earnest A., Percival R., Kelly S., Shrestha R., Fletcher S. [2009], *Health&WealthMOD: A Microsimulation Model of the Economic Impacts of Diseases on Older Workers*, „International Journal of Microsimulation", vol. 2(2).

Schofield1 D., Passey M., Percival R., Kelly S., Shrestha R., Fletcher S. [2011b], *Projecting the impacts of illness on labour force participation: An application of Health&WealthMOD*, „International Journal of Microsimulation", vol. 4(3).

Spadaro A. [2007], *1. Microsimulation as a tool for the evaluation of public policies*, [w:] A. Spadaro (red.), *Microsimulation as a tool for the evaluation of public policies: Methods and applications*, Fundacion BBVA.

Spielauer M. [2007], *Dynamic microsimulation of health care demand, health care finance and the economic impact of health behaviours: Survey and review*, „International Journal of Microsimulation", vol. 1(1).

Spielauer M. [2009], *Microsimulation approaches*, [w:] *Social science microsimulation using Modgen*, Statistics Canada [http://www.statcan.gc.ca/microsimulation/pdf/chap2-eng.pdf (06.07.2011)].

Spielauer M. [2009], *What is dynamic social science microsimulation?*, [w:] *Social science microsimulation using Modgen*, Statistics Canada [http://www.statcan.gc.ca/microsimulation/pdf/chap1-eng.pdf (06.07.2011)].

Suchecka J., Wierzbicka A., Żółtaszek A. [2007], *Model dwuczęściowy i jego zastosowania w zarządzaniu opieką zdrowotną*, [w:] M. Zawada (red.), *Zastosowanie metod ilościowych w badaniach ekonomiczno-społecznych*, Wydawnictwo Politechniki Częstochowskiej, Częstochowa.

Sutherland H. (red.) [2001], *Final report EUROMOD: An integrated European benefit-tax model*, [w:] Working Papers Series, EUROMOD Working Paper no. EM9/01 [http://www.iser.essex.ac.uk/publications/working-papers/euromod/em9-01.pdf (12.11.2011)].

The LifePaths Microsimulation Model. An Overview [http://www.statcan.gc.ca/microsimulation/pdf/lifepaths-overview-vuedensemble-eng.pdf (29.11.2011)].

Thurecht L., Brown B. [2011], *Economic Modelling of the Prevention of Type 2 Diabetes in Australia – The Diabetes Model*, „International Journal of Microsimulation", vol. 4(3).

Tomassini C., Wolf D. A. [2000], *Shrinking Kin Networks in Italy Due to Sustained Low Fertility*, „European Journal of Population", vol. 1, Kluwer Academic Publishers.

Torres-Reyna O., *Panel Data Analysis. Fixed & Random Effects* [using STATA 10.x], Princeton University [http://dss.princeton.edu/training/Panel101.pdf (11.04.2012)].

Trzaskalik T. [2003], *Wprowadzenie do badań operacyjnych*, PWE, Warszawa.

van Sonsbeek J. M., Gradus R. H. J. M. [2006], *A microsimulation analysis of the 2006 regime change in the Dutch disability scheme*, „Economic Modelling", vol. 23, Elsevier.

Vanness D. J., Tosteson A. N. A., Gabriel S. E., Melton III L. J. [2005], *The need for microsimulation to evaluate osteoporosis interventions*, „Osteoporos Int", vol. 16.

Wachter K. W., Knodel J. E., van Landingham M. [2003], *Parental bereavement: Heterogeneous impacts of AIDS in Thailand*, „Journal of Econometrics", vol. 112, Elsevier.

Wachter K. W., Blackwell D., Hammel E. A. [1997], *Testing the validity of Kinship microsimulation*, „Mathematical and Computer Modelling", vol. 26.

Wagenhals G. [2011], *Dual Income Tax Reform in Germany. A Microsimulation Approach*, „International Journal of Microsimulation", vol. 4(1).

Walker A. [2007], *Chapter 5. Health Status, Health Inequalities and the Ability of Older Australiansto Stay in the Labor Force*, [w:] A. Gupta, A. Harding (red.), *Modelling our future: Population ageing health and aged care*, Elsevier.

Welfe A. [2003], *Ekonometria*, wyd. III zmienione, PWE, Warszawa.

Winkelmann R., Boes S. [2006], *Analysis of Microdata*, Springer-Verlag, Berlin, Heidelberg.

Wiśniewski J. W. [2011], *Dylematy podwójnej metody najmniejszych kwadratów w mikromodelu ekonometrycznym*, [w:] P. Miłobędzki, M. Szreder (red.), *Modelowanie i prognozowanie gospodarki narodowej*, „Prace i Materiały Wydziału Zarządzania Uniwersytetu Gdańskiego", Wydział Zarządzania Uniwersytetu Gdańskiego, Fundacja Uniwersytetu Gdańskiego, Sopot.

Wiśniewski J. W. [2009], *Mikroekonometria*, Wydawnictwo Mikołaja Kopernika, Toruń.

Wiśniowski A. [2010], *5. Modele zmiennych wielomianowych nieuporządkowanych*, [w:] M. Gruszczyński (red.), *Mikroekonometria. Modele i metody analiz danych indywidualnych*, Oficyna a Wolters Kluwer business, Warszawa.

Wiśniowski A., Owczarczuk M. [2010], *2. Metody i modele*, [w:] M. Gruszczyński (red.), *Mikroekonometria. Modele i metody analiz danych indywidualnych*, Oficyna a Wolters Kluwer business, Warszawa.

Wooldridge J. M. [2002], *Econometric analysis of cross section and panel data*, Massachusetts Institute of Technology, USA.

Zabinski D., Selden T. M., Moeller J. F., Banthin J. S. [1999], *Medical savings accounts: Microsimulation results from a model with adverse selection*, „Journal of Health Economics", vol. 18, issue 2.

Zaidi A., Harding A., Williamson P. (red.) [2009], *New Frontiers in Microsimulation Modelling*, Ashgate, Wiedeń.
Zeliaś A. [2000], *Metody statystyczne*, PWE, Warszawa.
Zinn S. [2012], *A Mate-Matching Algorithm for Continuous-Time Microsimulation Models*, „International Journal of Microsimulation", vol. 5(1).
Żółtaszek A. [2011b], *Przestrzenno-czasowa analiza prywatnych wydatków gospodarstw domowych na ochronę zdrowia w Polsce w latach 2009–2018 – eksperyment mikrosymulacyjny*, „Ekonometria. Zastosowania metod ilościowych", nr 34.
Żółtaszek A. [2011a], *Mikrosymulacje – podstawy teoretyczne i wybrane zastosowania*, [w:] *Mikroekonoetria w teorii i praktyce*, Wydawnictwo Naukowe Uniwersytetu Szczecińskiego, Szczecin.
Żółtaszek A. [2012], *Private financing of health care in Poland – microsimulation model*, „Przegląd Statystyczny" [Statistical review], t. 3(59), Warszawa.

Źródła internetowe

strona internetowa Australian Demographic & Social Research Institute [ADSRI]: http://adsri.anu.edu.au/seminars/15nov11 (16.11.2011).
strona internetowa University of Canberra [Canberra_a]: http://www.canberra.edu.au/centres/natsem/home (09.12.2011).
strona internetowa University of Canberra [Canberra_b]: http://www.canberra.edu.au/centres/natsem/research-models/projects_and_models/aged-care-modelling (09.12.2011).
strona internetowa University of Canberra [Canberra_c]: http://www.canberra.edu.au/centres/natsem/research-models/projects_and_models/diabetes_prevention (18.12.2011).
strona internetowa University of Canberra [Canberra_d]: http://www.canberra.edu.au/centres/natsem/research-models/projects_and_models/new-model-australian-pbs (18.12.2011).
strona internetowa Centrum Analiz Ekonomicznych [CenEA]: http://www.cenea.org.pl/Mikrosymulacja/Badania/Mikrosymulacja/mikrosymulacja/menu-id-10.html (16.10.2012).
strona internetowa Institute for Social & Economic Research [ISER]: http://www.iser.essex.ac.uk/euromod/developing-euromod/euromodupdate (12.11.2011).
strona internetowa UC Berkeley [UC Berkeley]: http://www.lab.demog.berkeley.edu/socsim/index.shtml (25.11.2011).
strona internetowa [Mikrosymulacje]: http://www.mikrosymulacje.pl/ (16.10.2012).
strona internetowa National Science Foundation [NSF]: http://www.nsf.gov/ (16.11.2011).
strona internetowa Pension Benefit Guaranty Corporation [PBGC]: http://www.pbgc.gov/ (11.11.2011).
strona internetowa Statistics Canada [StatCan_a]: http://www.statcan.gc.ca/microsimulation/demosim/demosim-eng.htm (29.11.2011).
strona internetowa Statistics Canada [StatCan_b]: http://www.statcan.gc.ca/microsimulation/health-sante/health-sante-eng.htm (10.11.2011).
strona internetowa Statistics Canada [StatCan_c]: http://www.statcan.gc.ca/microsimulation/lifepaths/lifepaths-eng.htm (29.11.2011).
strona internetowa Statistics Canada [StatCan_d]: http://www.statcan.gc.ca/pub/91-551-x/2010001/meth-eng.htm (29.11.2011).
strona internetowa Statistics Canada [StatCan_e]: http://www.statcan.gc.ca/pub/91-551-x/2010001/tbl/tbl001-eng.htm (29.11.2011).