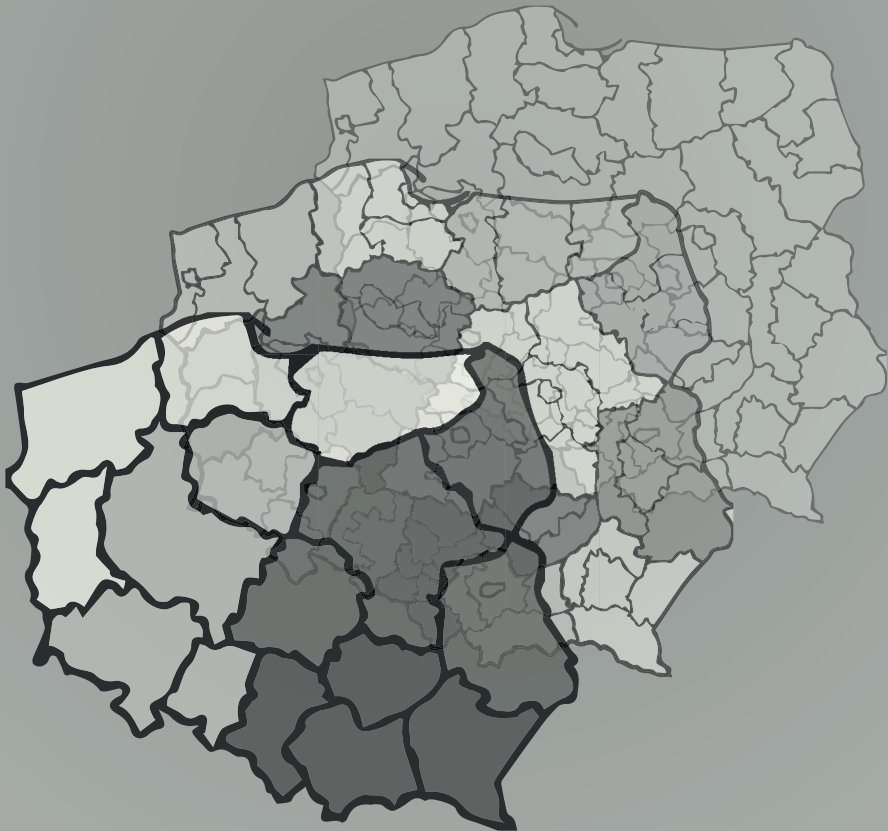


Michał Bernard Pietrzak

PROBLEM ZMIENNEJ JEDNOSTKI ODNIESIENIA
W PRZESTRZENNYCH BADANIACH EKONOMICZNYCH



Michał Bernard Pietrzak

**PROBLEM ZMIENNEJ JEDNOSTKI ODNIESIENIA
W PRZESTRZENNYCH BADANIACH EKONOMICZNYCH**

Polskie Towarzystwo Ekonomiczne
Oddział w Toruniu

Toruń 2018

Recenzenci:

prof. zw. dr hab. Jan Zawadzki
dr hab. Elżbieta Szulc, prof. UMK

Skład, łamanie i projekt okładki:

Ilona Pietryka

© Copyright by Polskie Towarzystwo Ekonomiczne Oddział w Toruniu

ISBN 978-83-62049-32-5

DOI: 10.24136/eep.mon.2018.1

Polskie Towarzystwo Ekonomiczne Oddział w Toruniu
e-mail: pte@umk.pl
www.ptetorun.pl

ul. Kopernika 21
87–100 Toruń

Praca powstała w wyniku realizacji projektu badawczego nr rej. 2015/17/B/HS4/01004 finansowanego ze środków Narodowego Centrum Nauki.

Spis treści

Wstęp	7
Rozdział 1. Agregacja danych w przestrzennych badaniach ekonomicznych: aspekty teoretyczne	13
1.1. Podstawowe zagadnienia dotyczące procesu agregacji danych	13
1.1.1. Agregacja danych a problem zmiennej jednostki odniesienia	13
1.1.2. Agregacja danych w badaniach ekonomicznych	16
1.2. Systematyka agregacji danych	21
1.2.1. Typy agregacji danych	21
1.2.2. System agregacji danych	26
1.2.3. Specyfika agregacji danych przestrzennych	29
Rozdział 2. Poprawność przestrzennych badań ekonomicznych	37
2.1. Etapy przestrzennych badań ekonomicznych	37
2.1.1. Etapy agregacji danych przestrzennych	37
2.1.2. Etap przestrzennej analizy ekonomicznej	42
2.2. Potencjalne błędy badawcze w przestrzennych badaniach ekonomicznych	48
2.2.1. Błąd identyfikacji jednorodnego zbioru układów	48
2.2.2. Błąd wyznaczenia obszaru wniosków	57
2.3. Identyfikacja mikrozależności i makrozależności w badaniach ekonomicznych	60
2.3.1. Mikrozależności a makrozależności ekonomiczne	60
2.3.2. Analiza mikrozależności oraz makrozależności przestrzennych	62
Rozdział 3. Redefinicje problemu sposobu podziału przestrzeni oraz problemu skali	75
3.1. Geneza powstania problemu zmiennej jednostki odniesienia	75
3.2. Problem zmiennej jednostki odniesienia a nowy paradygmat badań naukowych	79
3.2.1. Nowy paradygmat badań naukowych	79

3.2.2. Problem zmienności otrzymywanych rezultatów badań naukowych	83
3.2.3. Problem modyfikowalności jednostek terytorialnych	88
3.2.4. Wyznaczanie granic układu jednostek terytorialnych a problem zmiennej jednostki odniesienia	90
3.2.5. Jednorodność przyczynowa danych przestrzennych a wyniki badań	94
3.2.6. Potencjalne scenariusze badań nad problemem zmiennej jednostki odniesienia	97
3.2.7. Metodyka badań naukowych w ramach nowego paradygmatu nauk empirycznych	99
3.3. Rewizja problemu zmiennej jednostki odniesienia	102
3.3.1. Redefinicja problemu sposobu podziału przestrzeni	102
3.3.2. Redefinicja problemu skali	104
Rozdział 4. Podział przestrzeni dla badań dotyczących rozwoju rolnictwa w Polsce	107
4.1. Rozwiązanie problemu sposobu podziału przestrzeni na poziomie makroregionów	107
4.1.1. Wyznaczenie jednorodnego zbioru układów	107
4.1.2. Wyznaczenie granic układu makroregionów rolniczych	113
4.2. Relacja między mikroparametrami a makroparametrem na przykładzie badania makroregionów rolniczych SGM	121
4.2.1. Identyfikacja mikrozależności i makrozależności przestrzennych dotyczących zużycia nawozów mineralnych w rolnictwie	121
4.2.2. Wyznaczenie makrozależności przestrzennej na podstawie mikrozależności przestrzennych	131
Rozdział 5. Problem skali: identyfikacja na podstawie generowanych danych przestrzennych wyrażonych w wartościach średnich	141
5.1. Procedura symulacyjna w świetle zaproponowanej redefinicji problemu skali	141
5.2. Analiza symulacyjna dla szumów przestrzennych	143
5.3. Identyfikacja problemu skali w przypadku występowania trendu przestrzennego	153
5.4. Analiza wpływu autokorelacji przestrzennej na problem skali	162
5.4.1. Identyfikacja struktury przestrzennej	162
5.4.2. Wybór rozkładu losowego oraz systemu wag w procedurze symulacyjnej	165
5.4.3. Wpływ siły autokorelacji przestrzennej na poziom zmian wynikających z problemu skali	177
Rozdział 6. Symulacyjna analiza problemu skali na podstawie ważonych danych przestrzennych	183
6.1. Identyfikacja problemu skali dla szumów przestrzennych	183
6.2. Badanie problemu skali w warunkach występowania niejednorodności systematycznej	193
6.3. Analiza symulacyjna dla procesów o własności autokorelacji przestrzennej	206

Zakończenie	215
Literatura	221
Spis tabel	233
Spis rysunków	237

Wstęp

Praca dotyczy problemu zmiennej jednostki odniesienia¹. W ramach problemu zmiennej jednostki odniesienia rozpatrywane są osobno dwa zagadnienia, problem sposobu podziału przestrzeni oraz problem skali (poziomu agregacji)². Istota obydwu problemów określona została w literaturze w kontekście możliwości uzyskania odmiennych wyników badań na skutek zmiany układu jednostek terytorialnych lub w wyniku wykonania procesu agregacji danych.

Problem sposobu podziału przestrzeni opisany jest w literaturze jako problem związany z otrzymywaniem różnych wyników na podstawie analizy danych przestrzennych w zależności od wyboru układu jednostek terytorialnych, jednak w ramach tego samego poziomu agregacji (Openshaw, Taylor, 1979, s. 128; Openshaw, 1984b, s. 8). Natomiast problem skali określany jest jako problem związany z uzyskaniem odmiennych wyników badań pod wpływem zmiany poziomu agregacji danych przestrzennych (Openshaw, Taylor, 1979, s. 128; Openshaw, 1984b, s. 8).

Celem pracy jest opracowanie nowego podejścia metodologicznego do identyfikacji problemu zmiennej jednostki odniesienia, w tym problemu podziału przestrzeni oraz problemu skali w przestrzennych badaniach ekonomicznych. Właściwe zrozumienie tych problemów i ich uwzględnienie w prowadzonych przestrzennych badaniach ekonomicznych ma istotny wpływ na otrzymywane wyniki.

¹ Termin „problem zmiennej jednostki odniesienia” przyjęty został z pracy Suhecka (2014, s. 51–52, 56–64). Stanowi on tłumaczenie z języka angielskiego wyrażenia „Modifiable Areal Unit Problem” (MAUP).

² Również terminy „problem sposobu podziału przestrzeni” oraz „problem skali” („problem poziomu agregacji”) przyjęte zostały z pracy Suhecka (2014, s. 56–64) i stanowią one tłumaczenie z wyrażen „Aggregation Problem”, „Zonation Problem” oraz „Scale Problem” (Openshaw, Taylor 1979, s. 128, Openshaw, 1984b, s. 8).

Przeprowadzone w pracy rozważania koncentrują się wokół następujących tez:

1. Problem podziału przestrzeni oraz problem skali wymagają redefinicji, uwzględniających kwestię jednorodności przyczynowej danych przestrzennych.
2. Ustalenie jednorodnego zbioru układów w ramach prowadzonych przestrzennych badań ekonomicznych pozwala na poprawną identyfikację własności jednorodności przyczynowej danych przestrzennych.
3. Błąd identyfikacji jednorodnego zbioru układów oraz błąd wyznaczania obszaru wniosków prowadzą do niepoprawnych wniosków formułowanych na podstawie wyników analiz przestrzennych.
4. Identyfikacja makrozależności przestrzennych często obciążona jest niebezpieczeństwem popełnienia błędu wyznaczenia obszaru wniosków.
5. Zastosowanie *zoning system* lub *grouping system* do losowego wyznaczania granic układów terytorialnych nie gwarantuje, że odniesione do nich dane przestrzenne charakteryzują się własnością jednorodności przyczynowej.
6. Identyfikacja problemu zmiennej jednostki odniesienia na podstawie wyznaczanych w sposób losowy układów jednostek terytorialnych ma charakter pozorny.
7. Rozwiązanie problemu podziału przestrzeni pozwala na otrzymanie nowego układu jednostek terytorialnych, przyporządkowanego do jednorodnego zbioru układów.
8. Uwzględnienie składników struktury wewnętrznej ekonomicznych procesów przestrzennych jest niezbędne do właściwego rozważenia problemu skali.
9. Prawidłowa identyfikacja problemu skali w warunkach autokorelacji przestrzennej wymaga ustalenia właściwej struktury przestrzennej obszaru i wykonania analiz w jej ramach.

Monografia składa się z sześciu rozdziałów.

Zarówno w przypadku problemu sposobu podziału przestrzeni, jak i problemu skali, kluczową kwestią jest proces agregacji danych. W związku z tym rozdział pierwszy pracy rozpoczyna się od teoretycznych rozważań na temat procesu agregacji. Szczegółowo przedstawiony został związek między zagadnieniem agregacji danych oraz problemem zmiennej jednostki odniesienia. Omówiono także rolę agregacji danych w badaniach ekonomicznych. Teoretyczne rozważania dotyczyły definicji agregacji danych, typów agregacji w podziale na agregację podmiotową, agregację dóbr i agregację w czasie (zob. Pawłowski, 1969, s. 234–252), budowy systemu agregacji danych oraz specyfiki agregacji danych przestrzennych.

W rozdziale drugim rozważone zostało zagadnienie poprawności przestrzennych badań ekonomicznych oraz poprawności procesu agregacji. W rozdziale przedstawiony został wyraźny podział przestrzennych badań eko-

nomicznych na trzy etapy: etap postawienia problemu badawczego, etap wykonania przestrzennej analizy ekonomicznej oraz etap rozwiązania problemu badawczego na podstawie wniosków uzyskanych z analizy.

W ramach przeprowadzonych rozważań autor wprowadził dwa ważne pojęcia: pojęcie jednorodnego przyczynowo układu danych przestrzennych oraz pojęcie jednorodnego zbioru układów. Przestrzenne badania ekonomiczne mogą być prowadzone na różnych poziomach agregacji, co wiąże się z koniecznością zbadania jednorodności przyczynowej danych przestrzennych dla każdego z tych poziomów. Rozpatrzenie tej kwestii pozwoliło autorowi na opracowanie pojęcia jednorodnego zbioru układów, który określony został w pracy jako zbiór układów jednostek terytorialnych na różnych poziomach agregacji, gdzie odniesione do tych układów dane przestrzenne posiadają własność jednorodności przyczynowej. Oznacza to, że ustalenie jednorodnego zbioru układów w ramach przestrzennych badań ekonomicznych jest koniecznym warunkiem uzyskania poprawnych wyników badań.

W ostatniej części rozdziału drugiego przedyskutowano podjęty przez Pawłowskiego problem ustalenia relacji między makroparametrem a mikroparametrami (zob. Pawłowski, 1969, 239–244). W wyniku przeprowadzonych rozważań wyprowadzone zostały równania określające relację między mikroparametrami dla trzech makroregionów a makroparametrem dla całego obszaru Polski. Zgodnie z wyznaczoną relacją ocena makroparametru stanowi kombinację liniową ocen mikroparametrów oraz odpowiednio określonych wag. Otrzymane wyniki pokrywają się z wnioskami Pawłowskiego dotyczącymi analizy szeregów czasowych.

W rozdziale trzecim przeprowadzone zostały studia literaturowe nad zagadnieniem zmiennej jednostki odniesienia, a następnie skonfrontowane z rozważaniami z rozdziału pierwszego i drugiego. Na przykładzie konkretnych prac literaturowych wskazano, na czym polegają słabości przedstawianych w nich badań nad zagadnieniem zmiennej jednostki odniesienia. W omawianych pracach dotyczących zarówno problemu sposobu podziału przestrzeni, jak i problemu skali, badania wykonywane były na podstawie wybieranych w dowolny sposób układów jednostek terytorialnych. Ma to duże znaczenie dla przeprowadzanych analiz przestrzennych, ponieważ przyjęcie pojedynczych układów danych przestrzennych, niewchodzących w skład jednorodnego zbioru układów, prowadzić może do niepoprawnych wniosków. Krytyczna ocena prac z literatury przedmiotu skonfrontowana z rozważaniami autora na temat poprawności przestrzennych badań ekonomicznych wskazała na potrzebę redefinicji tych problemów. Autor przedstawia propozycję redefinicji obydwu problemów na końcu rozdziału trzeciego.

W związku z zaproponowaną redefinicją problemu sposobu podziału przestrzeni, w rozdziale czwartym rozważono go na przykładzie przestrzennych badań ekonomicznych dotyczących rozwoju rolnictwa w Polsce na poziomie makroregionów. W wyniku wykonanych badań wyznaczony został nowy układ

czterech makroregionów rolniczych, co stanowiło rozwiązanie podjętego przez autora problemu sposobu podziału przestrzeni.

Następnie w rozdziale czwartym podjęta została ponownie kwestia ustalenia relacji między mikroparametrami a makroparametrem na przykładzie makroregionów rolniczych. W wyniku przeprowadzonych badań ustalono relację między mikroparametrami dla makroregionów rolniczych SGM a makroparametrem dla obszaru Polski, co poszerzyło postawiony przez Pawłowskiego problem o aspekty przestrzenne.

W rozdziałach piątym i szóstym dokonano analizy problemu skali na podstawie danych generowanych, której celem było sprawdzenie wpływu zmiany poziomu agregacji na otrzymywane wyniki w ramach wykonanych analiz symulacyjnych. W rozdziale piątym analizy symulacyjne przeprowadzone zostały na podstawie generowanych danych przestrzennych wyrażonych w wartościach średnich, a w rozdziale szóstym podstawę obliczeń stanowiły generowane dane przestrzenne ważone wielkością obszaru. Zarówno w rozdziale piątym, jak i w rozdziale szóstym w kolejnych scenariuszach procedury symulacyjnej zakładano wybrane własności procesów przestrzennych.

W scenariuszu pierwszym generowano realizacje procesów szumów przestrzennych o stałym poziomie wartości oczekiwanej i wariancji oraz braku autokorelacji przestrzennej. W scenariuszu drugim analizowano sytuację, gdzie w wewnętrznej strukturze procesów przestrzennych zakładano występowanie własności systematycznej niejednorodności. W ostatnim, trzecim scenariuszu założono możliwość występowania autokorelacji przestrzennej w analizowanych procesach.

Na podstawie wykonanych analiz w ramach scenariusza pierwszego i drugiego nie stwierdzono istotnych różnic w ocenach własności procesów przestrzennych oraz łączących je zależnościach przyczynowych, które byłyby wynikiem procesu agregacji. Natomiast w przypadku własności autokorelacji przestrzennej w scenariuszu trzecim ustalono występowanie problemu skali, gdzie analizowane procesy charakteryzowały się dodatnią autokorelacją przestrzenną. Oznacza to, że w sytuacji identyfikacji autozależności przestrzennych zachodzą znaczne zmiany w otrzymywanych rezultatach badań na skutek procesu agregacji danych.

Monografia zakończona została podsumowaniem, w którym przedstawione zostały wnioski z przeprowadzonych rozważań teoretycznych, jak również z wykonanych analiz symulacyjnych oraz badań empirycznych odnoszących się do przestrzennych zjawisk ekonomicznych.

Wszystkie obliczenia numeryczne zostały przeprowadzone w ramach oprogramowania R, przy wykorzystaniu gotowych pakietów lub napisanych przez autora procedur. Część testów statystycznych zostało wykonanych za pomocą narzędzi programu gretl. Natomiast rysunki przygotowane zostały w programie QGIS. Dane statystyczne pozyskane zostały z bazy Bank Danych Lokalnych udostępnianej przez Główny Urząd Statystyczny.

Praca powstała w wyniku realizacji projektu badawczego nr rej. 2015/17/B/HS4/01004 finansowanego ze środków Narodowego Centrum Nauki.

Pragnę wyrazić podziękowania Recenzentom: prof. zw. dr. hab. Janowi Zawadzkiemu oraz dr hab. Elżbiecie Szulc, prof. UMK, za wnikliwe recenzje wydawnicze, w szczególności za uwagi krytyczne, a także cenne wskazówki, których uwzględnienie przyczyniło się do podwyższenia wartości merytorycznej pracy. Pragnę wyrazić również serdeczne podziękowania Koleżankom i Kolegom z Katedry Ekonometrii i Statystyki oraz Katedry Zastosowań Informatyki i Matematyki w Ekonomii, Wydziału Nauk Ekonomicznych i Zarządzania Uniwersytetu Mikołaja Kopernika w Toruniu za wymianę doświadczeń oraz naukowe wsparcie.

Rozdział 1.

Agregacja danych w przestrzennych badaniach ekonomicznych: aspekty teoretyczne

1.1. Podstawowe zagadnienia dotyczące procesu agregacji danych

1.1.1. Agregacja danych a problem zmiennej jednostki odniesienia

W pracy rozpatrzone zostaną zagadnienia związane z problemem zmiennej jednostki odniesienia (*Modifiable Areal Unit Problem*, MAUP), który wskazywany jest w literaturze jako ważny problem badawczy w ekonometrii przestrzennej (Anselin, 1989, s. 26; Tobler, 1989, s. 115; Haining, 2005, s. 150; Suhecka, 2014, s. 56–64). W ramach tego zagadnienia rozpatrywane są dwa jego aspekty, problem sposobu podziału przestrzeni¹ oraz problem skali² (Openshaw, Taylor, 1979, s. 128; Openshaw, 1984b, s. 8; Suhecka, 2014, s. 56–57). W przypadku obydwu problemów podkreślana jest w literaturze przedmiotu możliwość otrzymywania różnych wyników badań, na skutek zmiany układu jednostek terytorialnych lub wykonania agregacji danych przestrzennych (Gehlke, Biehl, 1934, s. 170; Anselin, 1989, s. 26, Arbia, 1989, s. 1–4). Możliwość otrzymania odmiennych wyników badań na skutek zmiany układu jednostek terytorialnych określona została w literaturze polskiej jako problem sposobu podziału przestrzeni (Suhecka, 2014, s. 57). Z kolei agregacja danych przestrzennych i związana z tym zmiana przyjętego poziomu agregacji również

¹ Ang. *Aggregation Problem*, *Zonation Problem* (Openshaw, Taylor, 1979, s. 128; Openshaw, 1984b, s. 8).

² Ang. *Scale Problem* (Openshaw, Taylor, 1979, s. 128, Openshaw, 1984b, s. 8).

mogą prowadzić do uzyskania różnych wyników badań, co określone zostało w literaturze jako problem skali (Suchecka, 2014, s. 56).

Bezpośrednie tłumaczenie wyrażenia *Modifiable Areal Unit Problem* (MAUP) w oparciu o pracę Yule i Kendall (1966, s. 320) jest następujące „problem modyfikowalności jednostek obszarowych”. Yule i Kendall (1966, s. 320–325) podkreślają występowanie dwóch typów jednostek statystycznych. Pierwszy typ stanowią jednostki statystyczne, których nie można zmodyfikować. Jako przykłady takich jednostek można podać konsumenta, przedsiębiorstwo. Drugi typ stanowią jednostki, które można poddać modyfikacji. Należy tutaj wyróżnić jednostki obszarowe, które jako obszary geograficzne można modyfikować za pomocą zmiany ich granic. Najczęściej modyfikacja polega na łączeniu jednostek obszarowych w celu uzyskania nowych jednostek o większej powierzchni lub podziale poszczególnych jednostek obszarowych na nowe jednostki o mniejszej powierzchni (Yule, Kendall, 1966, s. 320). Możliwość modyfikacji jednostek obszarowych wskazuje na konieczność postawienia problemu badawczego dotyczącego otrzymywanych wyników badań w zależności od ustalonych granic tych jednostek. W tym kontekście zaproponowane w pracy Suchecka (2014, s. 51–52) tłumaczenie wyrażenia *Modifiable Areal Unit Problem* jako problemu zmiennej jednostki odniesienia jest trafne, ponieważ wskazuje na główne źródło problematyki, jakim jest możliwość modyfikacji jednostek obszarowych. Trafne są również tłumaczenia *Aggregation Problem* jako problemu sposobu podziału przestrzeni oraz *Scale Problem* jako problemu skali, gdzie obydwa wyrażenia również jednoznacznie wskazują na konsekwencje własności modyfikowalności jednostek obszarowych (Suchecka, 2014, s. 51–52).

Istotnym zagadnieniem w ramach problemu zmiennej jednostki odniesienia jest agregacja danych przestrzennych. Wykonanie procesu agregacji stanowi podstawę zmiany sposobu podziału przestrzeni oraz prowadzi do zmiany poziomu agregacji (Openshaw 1984b, s. 8–9). Dlatego w pracy omówiona zostanie najpierw problematyka związana z agregacją danych, co pozwoli w dalszej części na pogłębione rozważania dotyczące zarówno problemu agregacji danych przestrzennych, jak i problemu zmiennej jednostki odniesienia.

W przypadku problemu agregacji danych należy podkreślić, że agregacja może być rozumiana znacznie szerzej, jako proces, który stanowi pomost pozwalający na przejście między mikroekonomią a makroekonomią (Theil, 1965, s. 1–9). Problem właściwej agregacji polegałby w takim rozumieniu na ustaleniu procedury agregacji danych, pozwalającej na otrzymanie odpowiedniego agregatu na podstawie danych dla pojedynczych obiektów ekonomicznych. Procedura agregacji danych powinna pozwolić na przejście od akceptowalnych mikrozależności w ramach teorii mikroekonomii (analiza danych dla pojedynczych obiektów) do akceptowalnych makrozależności w ramach teorii makro-

ekonomii (analiza danych zagregowanych)³. Jako przykład tak rozumianej agregacji można podać badania Tinbergena (1939, s. 14), który w celu wyznaczenia elastyczności dochodowej agregatu (makrozależności) zaproponował wykorzystanie średniej uzyskanej na podstawie zbioru elastyczności dochodowych dla pojedynczych obiektów ekonomicznych (mikrozależności). O ważności agregacji danych w badaniach ekonomicznych świadczy stanowisko Pawłowskiego, który teorię agregacji określił jako jeden z działów ekonometrii (Pawłowski, 1969, s. 235). Dodatkowo Pawłowski ustalił cztery, kluczowe zagadnienia (problemy badawcze), mogące być przedmiotem rozważań w ramach teorii agregacji (Pawłowski, 1969, s. 235).

Jako pierwsze zagadnienie badawcze, Pawłowski wymienia ogólne rozważania nad istotą i rodzajami agregacji (Pawłowski, 1969, s. 235). W ramach tego zagadnienia, w pracy rozpatrzone zostaną podstawowe pojęcia z teorii agregacji danych oraz potencjalne problemy badawcze związane z procesem agregacji danych. Rozważania te pozwolą na zaproponowanie procedury budowy systemu agregacji danych, która omówiona zostanie w kolejnym podrozdziale. Pokazane zostanie, jak ważne jest stworzenie systemu agregacji danych oraz jak taki system może funkcjonować w dowolnym obiekcie ekonomicznym, w którym gromadzone są dane. W związku z propozycją systemu agregacji danych omówione zostaną również potencjalne rodzaje agregacji. Podkreślona zostanie istotna rola odbiorcy danych, od którego potrzeb informacyjnych powinien zależeć wybór danych na najniższym poziomie agregacji oraz rodzaj ich agregacji.

Drugie zagadnienie badawcze stanowi kwestia ustalenia relacji, jakie zachodzą między zależnościami dla jednostek niższego rzędu, a zależnościami dla jednostek wyższego rzędu. Pawłowski określa to zagadnienie jako problem ustalenia relacji między mikroparametrami a makroparametrem (Pawłowski, 1969, s. 235; zob. także Bołt, Krauze, Kulawczuk, 1985, s. 33–34). Rozwiązanie tego problemu sprowadza się do identyfikacji badanych zależności na różnych poziomach agregacji, a następnie próbie oceny związku między ustalonymi zależnościami. Należy podkreślić, że Pawłowski nie ogranicza tutaj rozważań do podejścia prezentowanego przez Theila, gdzie istotą problemu było określenie relacji między parametrami określającymi zależności mikroekonomiczne a parametrami określającymi zależności makroekonomiczne. Pawłowski proponuje ustalenie charakteru parametrów w oparciu o relację między obiektami niższego i wyższego rzędu. Podejście takie oznacza możliwość zmiany charakteru parametru z makroparametru na mikroparametr w związku ze zmianą poziomem agregacji (Pawłowski, 1969, s. 239–240).

Kolejne, trzecie zagadnienie dotyczy kwestii możliwości uwzględnienia w modelach dla jednostek wyższego rzędu własności, którymi charakteryzują

³ Należy podkreślić, że w literaturze proponowane były odmienne podejścia metodyczne dotyczące przejścia od mikroekonomii do makroekonomii (Klein 1946, s. 93–108; May, 1946, s. 285–298, Bołt, Krauze, Kulawczuk, 1985, s. 8–9)

się jednostki niższego rzędu (Pawłowski, 1969, s. 244). Wymieniony problem badawczy jest ważny z praktycznego punktu widzenia, ponieważ uwzględnienie własności jednostek niższego rzędu może przyczynić się do znacznej poprawy jakości badań w skali makro. Pawłowski wskazuje na potrzebę rozwijania metod uwzględniających rozkład zmiennych dla jednostek niższego rzędu w badaniach typu makroekonomicznego, jako jedno z zadań teorii ekonometrii (Pawłowski, 1969, s. 245).

Jako czwarte, ostatnie zagadnienie Pawłowski wyróżnia problem predykcji na podstawie modeli dla jednostek wyższego rzędu w sytuacji, gdy zmieniają się istotnie własności jednostek niższego rzędu (Pawłowski, 1969, s. 235). Pawłowski wskazuje na jedno z rozwiązań, którym jest zaproponowana przez Theila metoda agregacji doskonałej (Theil, s. 133–145, Bołt, Krauze, Kulawczuk, 1985, s. 27–28). Metoda ta polega na poszukiwaniu takiej procedury agregacji, która nie prowadziłaby do błędów prognoz w przypadku istotnych zmian w strukturze jednostek niższego rzędu.

W pracy rozważone zostaną szczegółowo dwa pierwsze zagadnienia, ponieważ okażą się one istotne z perspektywy badań nad problemem zmiennej jednostki odniesienia. Najpierw przedstawione zostaną ogólne rozważania nad istotą i rodzajami agregacji, a następnie podjęty zostanie problem ustalenia relacji między mikroparametrami a makroparametrem. Istotną wartością dodaną w pracy będzie rozpatrzenie zagadnienia relacji między mikroparametrami a makroparametrem, których oceny wyznaczane są na podstawie danych przestrzennych.

1.1.2. Agregacja danych w badaniach ekonomicznych

Agregacja danych określona została przez Pawłowskiego jako „proces łączenia danych liczbowych dotyczących zbiorów jednostek niższego rzędu, wskutek czego otrzymuje się informacje liczbowe dotyczące jednostek wyższego rzędu” (Pawłowski, 1969, s. 24). Pawłowski jako przykład procesu agregacji danych przedstawia możliwość sumowania wartości produkcji piwa w poszczególnych browarach, w wyniku czego uzyskiwana jest wartość produkcji piwa dla całego przemysłu piwowarskiego. Poszczególne browary wchodzą w skład zbioru jednostek niższego rzędu, a przemysł piwowarski stanowi jednostkę wyższego rzędu.

Ważne podkreślenia jest, że wykonanie procesu agregacji danych nie jest ograniczone do określonego typu podmiotów ekonomicznych, czy rodzaju działalności gospodarczej. Inny przykład procesu agregacji może dotyczyć sieci sklepów z materiałami biurowymi. Również w tym przypadku agregacja polega na sumowaniu wartości sprzedawanych materiałów biurowych w poszczególnych sklepach. Pojedyncze sklepy mogą być traktowane jako jednostki niższego rzędu, a cała sieć jako jednostka wyższego rzędu. Agregacja danych może być również przeprowadzana osobno dla każdego ze sklepów (wybranej

jednostki niższego rzędu) i dotyczyć sumowania wybranego asortymentu dla pojedynczego sklepu. W związku z tym proces agregacji może dotyczyć sumowania wartości sprzedaży artykułów kreślarskich, długopisów, piór, markerów oraz ołówków, dzięki czemu uzyskiwana jest wartość sprzedaży artykułów piśmienniczych. Opiszana agregacja wybranego asortymentu może dotyczyć również całej sieci sklepów (jednostki wyższego rzędu).

Możliwa jest agregacja danych w czasie (dotycząca wybranego asortymentu, jak np. artykułów papierniczych). Wartość sprzedaży artykułów papierniczych w danym miesiącu łatwo ustalić poprzez wykonanie sumy danych dla poszczególnych dni, a dla okresu rocznego poprzez wykonanie sumy danych dla miesięcy. Biorąc pod uwagę zróżnicowanie asortymentów w ramach sieci sklepów z artykułami biurowymi oraz możliwe opcje przyjętych okresów (np. dzień, tydzień, miesiąc, kwartał, rok), widoczna jest mnogość wariantów agregacji danych. Wybór właściwej agregacji danych jest trudny, jednak zaletą agregacji danych jest możliwość otrzymania czytelnych informacji na podstawie zagregowanych danych, które byłyby niemożliwe do uzyskania na podstawie zbioru szczegółowych danych przed agregacją (Pawłowski, 1969, s. 237–238).

Z definicji przedstawionej przez Pawłowskiego wynika, że dane liczbowe stanowią odzwierciedlenie zmierzonych zdarzeń ekonomicznych⁴. Zdarzenie ekonomiczne określone zostało przez Zielińskiego jako „każde zdarzenie związane z obiektem (systemem) ekonomicznym, poprzez które przejawiają się cele gospodarczej działalności obiektu (systemu) oraz charakteryzują sposób działania obiektu” (Wiśniewski, Zieliński, 1998, s. 5). Jako przykłady zdarzeń ekonomicznych Zieliński podaje: płacę pracownika, zakup dobra w ramach gospodarstwa domowego, czy wielkość produkcji w przedsiębiorstwie, a jako przykłady obiektów podane zostały: pracownik, gospodarstwo domowe, przedsiębiorstwo, gospodarka danego kraju, czy gałąź gospodarki narodowej (Wiśniewski, Zieliński, 1998, s. 5). Dodatkowo stwierdzona została dla obiektów następująca własność: „obiekty (systemy) niższego rzędu są elementami obiektów wyższego rzędu” (Wiśniewski, Zieliński, 1998, s. 5), co tożsame jest z określeniem jednostek niższego i wyższego rzędu w podanej przez Pawłowskiego definicji agregacji danych.

W przedstawionym wcześniej przykładzie agregacji danych dotyczącym produkcji piwa, zdarzeniem ekonomicznym jest zaobserwowana wartość produkcji piwa w ustalonym przedziale czasu. Pojedynczy browar powinno się roz-

⁴ Dodatkowo Pawłowski wyróżnia fakty elementarne, jako najprostsze zdarzenia elementarne. Fakty elementarne rozumiane są jako te, których nie można rozłożyć na inne drobniejsze fakty (Pawłowski, 1969, s. 236). Pawłowski podkreśla, że w ekonomii trudno jest znaleźć sytuację, gdy mamy do czynienia z faktami elementarnymi. Mierzone zdarzenia ekonomiczne są najczęściej ustalonym agregatem faktów elementarnych. W tym sensie mierzone zdarzenia ekonomiczne stanowią dane zagregowane, które mogą być następnie ponownie agregowane. Jest to bardzo ważne z punktu widzenia pozyskiwania danych pierwotnych, które jako zdarzenia ekonomiczne gromadzone są już w postaci ustalonego agregatu faktów elementarnych.

patrywać jako obiekt ekonomiczny niższego rzędu, natomiast cały przemysł piwowarski jako obiekt wyższego rzędu. Należy zwrócić uwagę na jeszcze jeden fakt związany z obiektami niższego oraz wyższego rzędu. W ramach badań ekonomicznych obiekty te powinny być definiowane łącznie, ze względu na ich hierarchię (Wiśniewski, Zieliński, 1998, s. 5). Nawet jeśli analiza dotyczy wyłącznie obiektu niższego rzędu (na przykład produkcji piwa wybranego browaru), to należy w ramach podjętego problemu badawczego określić również obiekty wyższego rzędu (określenie całej branży piwowarskiej). Również w sytuacji, kiedy analiza dotyczy wyłącznie obiektu wyższego rzędu (na przykład produkcji piwa w branży piwowarskiej w Polsce), należy podjąć próbę definicji obiektów niższego rzędu. Dodatkowa analiza obiektów niższego rzędu może znacznie wzbogacić wykonywaną analizę dla obiektu wyższego rzędu i odwrotnie.

Rozważane zdarzenia ekonomiczne przedstawiają wartości wybranej cechy dla ustalonego obiektu ekonomicznego. Mierzalne cechy obiektów ekonomicznych określane są mianem zmiennych ekonomicznych (Zieliński, 1991 s. 22–25, Wiśniewski, Zieliński, 1998, s. 45–47). Należy podkreślić, że zmienne ekonomiczne mogą być rozumiane na różnych poziomach uogólnienia, w zależności od przyjęcia określonych warunków dotyczących obiektu, cechy oraz momentu lub przedziału czasowego obserwacji. Za przykłady zmiennej ekonomicznej można uznać, wychodząc od najwyższego poziomu uogólnienia, płacę pracownika w branży piwowarskiej, następnie płacę pracownika w wybranym przedsiębiorstwie, jak i w wybranym browarze. W związku z tym wartości zmiennych ekonomicznych o najmniejszym poziomie uogólnienia będzie można zmierzyć za pomocą obserwacji zdarzeń ekonomicznych będących prostym agregatem faktów elementarnych⁵ (Pawłowski, 1969, s. 236). Natomiast w celu obserwacji wartości zmiennej ekonomicznej rozumianej na wysokim poziomie uogólnienia konieczna będzie agregacja zdarzeń ekonomicznych.

Wszystkie zdarzenia ekonomiczne stanowią skutek określonego splotu przyczyn (Zieliński, 1991, s. 4–7). Niestety, określenie w badaniach ekonomicznych, pełnego splotu przyczyn dla wybranej grupy zdarzeń jest niemożliwe. Najczęściej teoria ekonomii oraz doświadczenie badacza pozwalają na ustalenie wyłącznie splotu przyczyn głównych, czyli tych, które oddziałują w przypadku wszystkich zdarzeń ekonomicznych z rozpatrywanej grupy zdarzeń (Lange, Banasiński, 1968, s. 15; Zieliński, 1991, s. 5). Należy pokreślić, że w pracy Lange, Banasiński (1968, s. 15) wskazano uwagę również na fakt, że obok przyczyn głównych występują także przyczyny uboczne, które oddziałują dodatkowo, zwiększając lub hamując siłę oddziaływania przyczyn głównych. W takiej sytuacji, jeżeli wystąpiły przyczyny główne, to możemy określić jedynie prawdopodobieństwo wystąpienia określonego zdarzenia ekonomicznego. Może wystąpić sytuacja, że pomimo zajścia splotu przyczyn głównych,

⁵ Na przykład płaca brutto jest sumą wynagrodzenia zasadniczego oraz przysługujących dodatków pracowniczych.

zdarzenie ekonomiczne nie wystąpiło (Wiśniewski, Zieliński, 1998, s. 18). W takim przypadku splot przyczyn głównych należy określić jako niewystarczający zespół przyczyn (Wiśniewski, Zieliński, 1998, s. 17–18)⁶. W związku z tym wszystkie zmienne ekonomiczne należy traktować jako zmienne losowe, które przyjmują różne wartości z określonymi prawdopodobieństwami (Hellwig, 1972, s. 66; Wiśniewski, Zieliński, 1998, s. 8).

Pojęcie ekonomicznej zmiennej losowej może zostać uogólnione do pojęcia ekonomicznego procesu stochastycznego, które wprowadzane jest wówczas, kiedy w analizie zmiennej losowej uwzględniony zostanie wymiar czasu. Formalnie proces stochastyczny określany jest jako losowa funkcja względem nielosowego argumentu czasu (Zieliński, 1991, s. 6). Zbiór zdarzeń ekonomicznych stanowiących realizację ekonomicznego procesu stochastycznego określany jest mianem ekonomicznego szeregu czasowego.

W ekonomii rozważane są również losowe funkcje wielu nielosowych argumentów, które określane są mianem pól losowych (zob. Zieliński, 1990; Szulc, 2007). Przykładem ekonomicznego pola losowego mogą być wydatki na żywność gospodarstw domowych, które można rozważać jako funkcję losową wielu nielosowych argumentów, takich jak czas, lokalizacja, grupa zawodowa, czy liczba osób w rodzinie.

Dodatkowo w badaniach ekonomicznych mogą być również rozpatrywane zbiory zmiennych losowych, procesów stochastycznych czy pól losowych. Mamy wtedy do czynienia z wielowymiarowymi zmiennymi losowymi, wielowymiarowymi procesami stochastycznymi oraz wielowymiarowymi polami losowymi.

Zieliński podkreśla istotność rozróżnienia między pojęciami ekonomicznego pola losowego, ekonomicznego procesu stochastycznego oraz zmiennej ekonomicznej. Ideą Zielińskiego była propozycja metodyki badań ekonomicznych, polegająca na rozpatrywaniu zjawisk społeczno-gospodarczych w postaci ekonomicznych pól losowych (Zieliński, 1990, 1991). Koncepcję tą podjęła w swojej pracy Szulc (2007) dając podstawy rodzącej się nowej ekonometrii pól losowych. Analizowane ekonomiczne pola losowe byłyby funkcjami wielu, nielosowych argumentów, w zależności od podjętego problemu badawczego. W kolejnych krokach, na podstawie wiedzy ekonomicznej oraz uzyskanych wyników, badacz dokonywałby próby redukcji nielosowych argumentów. W ten sposób różne aspekty zjawisk społeczno-gospodarczych, związane z nielosowymi argumentami, byłoby rozważane krok po kroku. Analiza rozpoczynałaby się na najwyższym poziomie z perspektywy wielowymiarowego ekonomicznego pola losowego. W wyniku redukcji argumentów analiza przenoszona byłaby na niższe poziomy związane z ekonomicznymi procesami stochastycznymi, czy ekonomicznymi zmiennymi. Formalne założenia dotyczące metodyki

⁶ Natomiast konieczny splot przyczyn dla wystąpienia zdarzenia ekonomicznego określony został jako wystarczający zespół przyczyn (Wiśniewski, Zieliński, 1998, s. 17–18).

badań ekonomicznych opierającej się na idei redukcji wymiaru ekonomicznych pól losowych zostały sformułowane w pracy Szulc (2007, s. 62–66).

Szczególnym celem badań ekonomicznych jest identyfikacja prawidłowości jako przyczynowych zależności zjawisk społeczno-ekonomicznych, przy założeniu odpowiednich, nielosowych argumentów. Należy podkreślić, że to właśnie wspomniany wyżej Profesor Zieliński wypracował warsztat metodyczny pozwalający na opis oraz analizę przyczynowych zależności zjawisk społeczno-ekonomicznych (Zieliński, 1991). Mierzone zdarzenia ekonomiczne stanowią realizacje zmiennych losowych. Zieliński wskazuje na fakt, że analiza konkretnego zdarzenia ekonomicznego jest próbą wyjaśnienia faktu szczegółowego. Natomiast analiza realizacji zmiennej ekonomicznej w postaci zbioru zdarzeń ekonomicznych (danych statystycznych) pozwala na wykrycie ekonomicznych prawidłowości, których identyfikacja jest głównym celem ekonometrii w ramach prowadzonych badań ekonomicznych (Lange, 1978, s. 54–64; Lange, Banasiński, 1968, s. 15–28; Zieliński, 1991, s. 1–7; Wiśniewski, Zieliński, 1998, s. 6). Lange określa ekonometrię w następujący sposób, „... to nauka zajmująca się ustalaniem za pomocą metod statystycznych konkretnych ilościowych prawidłowości zachodzących w życiu gospodarczym” (Lange, 1961, s. 11). Takie rozumienie roli ekonometrii, w którym nacisk położony jest na rozwiązywanie problemów badawczych w ekonomii, prezentowane jest szeroko w polskiej literaturze ekonomicznej (Pawłowski, 1969, s. 17; Zeliaś, 1991, s. 11, Zieliński, 1991, s. iii–vi; Wiśniewski, Zieliński, 1998, s. 33). Zeliaś utrzymuje takie podejście w przypadku definicji ekonometrii przestrzennej określając ją jako „autonomiczny dział metod ilościowych opracowanych na użytek badań zjawisk społeczno-gospodarczych w ujęciu przestrzennym”⁷ (Zeliaś, 1991, s. 7). Oznacza to, że również w przypadku ekonometrii przestrzennej jej rola jest widziana w rozwiązywaniu problemów badawczych w ramach przestrzennych badań ekonomicznych. Podejście to podkreślane jest w kolejnych pracach z zakresu ekonometrii przestrzennej (np. Suchecki, 2010, s. 18–19, zob. także Szulc, 2007⁸).

Przedstawiona przez Zeliasia definicja ekonometrii przestrzennej nie jest tożsama z definicją zaproponowaną w pracy Anselin (1988, s. 10), która powszechnie przyjęta została w literaturze zachodniej. Anselin definiuje ekonometrię przestrzenną jako zbiór metod oraz technik, które poprzez uwzględnienie własności przestrzennych pozwalają na poprawną specyfikację modeli, predykcję badanych procesów przestrzennych oraz testowanie hipotez (zob. Anselin,

⁷ W dalszej części pracy Zeliaś definiuje przedmiot ekonometrii przestrzennej, jakim jest „wyjaśnienie oraz opis zjawisk i procesów gospodarczych posiadających aspekt przestrzenny za pomocą metod ekonometrycznych” (Zeliaś, 1991, s. 12).

⁸ Pomimo rozszerzenia rozważań do pól losowych, gdzie cel cytowanej pracy określony został jako „przedstawienie teorii pól losowych w takim ujęciu i zakresie, w jakim z powodzeniem może być ona stosowana w badaniach ekonomicznych procesów wielowymiarowych”, wpisuje się ona w nurt nowoczesnej ekonometrii przestrzennej (zob. Szulc, 2007, s. 11).

1988, s. 10; Suhecki, 2010, s. 19). Oznacza to, że zarówno w pracy Anselina (1988), jak i w większości innych prac z literatury zachodniej, rola ekonometrii przestrzennej sprowadzona została do użytecznego zbioru technik pozwalających na uniknięcie problemów metodologicznych⁹. Upowszechnienie stanowiska Anselina odnośnie do rozumienia istoty ekonometrii przestrzennej miało ogromny wpływ na jej rozwój, gdzie rozwiązywanie problemów badawczych w ramach przestrzennych badań ekonomicznych zostało zepchnięte na dalszy plan. Autor podkreśla ten fakt, ponieważ miało to wpływ również na podejście do badań nad problemem zmiennej jednostki odniesienia. Świadczy to również o pozytywnym charakterze polskiej szkoły ekonometrycznej, ponieważ polscy badacze uznają za główną rolę ekonometrii przestrzennej wyjaśnianie zależności przestrzennych zjawisk ekonomicznych. Niewątpliwie autem przedstawionej pracy będzie rozważenie problemu zmiennej jednostki odniesienia przy założeniu definicji, zarówno ekonometrii, jak i ekonometrii przestrzennej, przyjętej przez polskich badaczy.

1.2. Systematyka agregacji danych

1.2.1. Typy agregacji danych

W wyniku pomiaru zdarzeń ekonomicznych uzyskiwane są dane pierwotne. Zgromadzone w odpowiedni sposób dane pierwotne stanowią wiarygodne źródło informacji¹⁰ dla odbiorcy danych, które jest istotnym zasobem ekonomicznym (Oleński, 2001, s. 23–26, 208–209). Dane pierwotne mogą być wykorzystane również do wykonania procesu agregacji i uzyskania danych zagregowanych, które stanowią kolejne źródło informacji (Zeliaś, 1991, s. 27).

Gwarantem wiarygodności wybranych zbiorów danych statystycznych jest fakt, że zmierzone zdarzenia są wynikiem oddziaływania tych samych przyczyn głównych. Zieliński określił zdarzenia, które są wynikiem tego samego splotu przyczyn głównych jako słabo jednorodne (Zieliński, 1991, s. 7–11)¹¹.

⁹ Nie umniejsza to ogromnych zasług Anselina w rozwój i propagowanie ekonometrii przestrzennej (Suhecki, 2010, s. 19).

¹⁰ W literaturze zwraca się szczególną uwagę na to, by dane charakteryzowały się odpowiednią jakością (Suhecka, 2014, s. 11; Zeliaś, 1991, s. 42–44).

¹¹ Zieliński opracował aparat narzędziowy związany z modelowaniem zależności przyczynowo-skutkowych (Zieliński, 1991, 1–76). Ze względu jednak na możliwość pozyskiwania różnego typu danych pierwotnych, dotyczących dowolnych cech obiektów ekonomicznych, a następnie generowania nowych danych w wyniku procesu agregacji, dokonywanie oceny jednorodności przyczynowej danych opartej na rachunku prawdopodobieństwa wydaje się niemożliwe dla każdego przypadku. W związku z tym w pracy założone zostanie, że ocena słabej jednorodności danych ustalana będzie arbitralnie przez specjalistę, badacza. Oczywiście arbitralna ocena powinna opierać się na znajomości badanej cechy oraz obiektu ekonomicznego, jak i na zawodowym doświadczeniu badacza.

Na potrzeby pracy nazwa tej własności zostanie zmieniona na jednorodność przyczynową danych¹². Należy stwierdzić, że zdarzenia masowe takiego typu odzwierciedlają we właściwy sposób oddziaływanie przyczyn. Oznacza to, że ustalenie prawdziwości ekonomicznych powinno odbywać się na podstawie jednorodnych przyczynowo danych. Jeżeli badacz będzie prowadził badania na podstawie danych nie posiadających własności jednorodności przyczynowej, to wyniki badania zawsze będą obciążone pewnym błędem poznawczym. Błąd ten wynika z faktu, że niejednorodne przyczynowo dane są odzwierciedleniem splotu przyczyn głównych zakłóconego przez inne sploty przyczyn albo splot przyczyn głównych nie istnieje. W sytuacji, gdy występuje dodatkowy splot przyczyn, który w niewielkim stopniu zakłóca oddziaływanie splotu przyczyn głównych, to analiza może być obciążona niewielkim błędem poznawczym. Jeżeli natomiast zmierzone dane są odzwierciedleniem różnych splotów przyczyn, z których żaden nie dominuje, to uzyskane wyniki mają bardzo niską wartość poznawczą. Podsumowując należy stwierdzić, że pozyskiwane dane pierwotne powinny posiadać własność jednorodności przyczynowej.

W ramach pozyskiwania danych pierwotnych konieczne jest ustalenie odbiorcy tych danych. Bezpośrednie wykorzystanie danych pierwotnych lub uzyskanie na ich podstawie danych zagregowanych pozwala na uzyskanie pożytecznych informacji dla potencjalnego odbiorcy. Fakt ten komplikuje wybór danych pierwotnych oraz proces agregacji danych, ponieważ dla różnych odbiorców mogą być istotne zupełnie inne informacje (mogą mieć oni odmienne potrzeby informacyjne; zob. Oleński, 2001, s. 226–227). W związku z tym ustalenie odbiorcy danych oraz jego potrzeb informacyjnych staje się nieodzowną czynnością poprzedzającą proces agregacji. Zagadnienia dotyczące typów odbiorców oraz ich potrzeb informacyjnych przedstawione zostaną w kolejnym podrozdziale.

Gromadzenie danych powinno się opierać na wcześniejszym wyborze zdarzeń ekonomicznych, które mają być rejestrowane. Oprócz ustalenia faktu, czy pozyskane dane będą posiadać własność jednorodności przyczynowej, należy także ustalić czy zdarzenia ekonomiczne są istotne z punktu widzenia potrzeb informacyjnych odbiorcy oraz potencjalnych analiz ekonomicznych. Wybór odpowiednich, jednorodnych przyczynowo zdarzeń ekonomicznych jest ważny, ponieważ zarówno przed, jak i po wykonaniu procesu agregacji stanowią one przedmiot analizy ekonomicznej. Należy wybrać tylko te zdarzenia ekonomiczne, które niosą ze sobą użyteczne informacje i na podstawie których możliwe jest wyciągnięcie wniosków. Ważną kwestią stanowi również wybór momentu ich rejestracji. Na przykład w przypadku przedsiębiorstwa produkcyjnego zużycie relatywnie tanich oraz powszechnie stosowanych w produkcji elementów może być rejestrowane w postaci pakietów np. 100, 1000 sztuk. Najczęściej

¹² Zaproponowana zmiana jest uzasadniona specyfiką nazewnictwa stosowanego w przypadku opisu własności pól losowych. Zabieg ten ma na celu ominięcie komplikacji związanych z podobnym nazewnictwem.

możliwy jest zakup oraz przechowywanie zapasu takiego typu elementów i dopiero ich zużycie na odpowiednio wyższym poziomie staje się ważne w ramach wykonywanej produkcji. W związku z tym za istotne można uznać np. zdarzenie zużycia 1000 sztuk tego typu elementów. W każdym przedsiębiorstwie występuje też zużycie elementów relatywnie drogiej oraz o nietypowym zastosowaniu. Powoduje to, że zastosowanie tego typu elementów nie występuje zbyt często. Wystąpienie nawet pojedynczego zużycia tego typu elementów może być uznane za istotne zdarzenie ekonomiczne w ramach procesu produkcji przedsiębiorstwa. W związku z tym należy stwierdzić, że ustalenie istotnych zdarzeń elementarnych, które posiadają własność jednorodności przyczynowej oraz właściwy wybór momentu ich pomiaru pozwala na pozyskanie danych pierwotnych, które następnie stanowią podstawę do wykonywanego procesu agregacji danych.

Na podstawie przedstawionych rozważań dotyczących agregacji danych widoczna jest mnogość potencjalnych jej wariantów. Potrzebne jest zatem ich usystematyzowanie. Pawłowski wyróżnił trzy typy agregacji danych: agregację podmiotową, agregację dóbr oraz agregację w czasie (Pawłowski, 1969, s. 237)¹³. Należy podkreślić, że na podstawie zbioru danych pierwotnych możliwe jest najczęściej wykonanie każdego z wymienionych typów agregacji, dodatkowo możliwe jest również wykonanie różnych typów agregacji po sobie¹⁴. Oznacza to, że istnieje wiele dowolnych ścieżek postępowania z danymi, jeśli chodzi o proces agregacji. Występują tutaj jedynie dwa ograniczenia. Pierwsze ograniczenie dotyczy pozyskiwania oraz przekształcania jedynie takich danych pierwotnych, które posiadają własność jednorodności przyczynowej¹⁵. Drugie ograniczenie dotyczy merytorycznego uzasadnienia procesu agregacji. Każdy proces agregacji powinien być uzasadniony możliwością uzyskania na podstawie danych zagregowanych istotnych informacji z punktu widzenia potencjalnego odbiorcy.

Agregacja podmiotowa polega na agregacji danych na podstawie zbioru obiektów ekonomicznych niższego rzędu, w wyniku czego otrzymywane są

¹³ Oleński z kolei wyróżnia następujące typy agregacji, agregacja w przestrzeni obiektów, agregacja w przestrzeni cech, agregacja w przestrzeni obiektów i cech, agregacja przez sumowanie, agregacja przez transformację algorytmiczną oraz agregacja przez translację (Oleński, 2001, s. 209–218).

¹⁴ Przykładowe dane zagregowane są najczęściej wynikiem kolejnych etapów agregacji. Możliwe jest na przykład wykonanie agregacji dóbr dla wybranych obiektów ekonomicznych niższego rzędu. Następnie na podstawie otrzymanych danych może zostać przeprowadzona agregacja podmiotowa dla obiektu wyższego rzędu i na koniec dane mogą zostać poddane jeszcze agregacji czasowej.

¹⁵ Należy podkreślić, że w wyniku procesu agregacji dane zagregowane mogą utracić własność jednorodności przyczynowej. Kolejne etapy procesu agregacji znacznie zwiększają prawdopodobieństwo uzyskania danych nie posiadających tej własności. Możliwe jest jednak wnioskowanie na podstawie takiego typu danych i sytuacja taka ma miejsce najczęściej przy próbie badania makrozależności.

dane dla obiektu ekonomicznego wyższego rzędu (Pawłowski, 1969, s. 236–237). Agregacja tego typu może dotyczyć dowolnych obiektów ekonomicznych, takich jak przedsiębiorstwa, gospodarstwa domowe, czy konsumenci. Jedynym ograniczeniem jest fakt, że dla obiektów ekonomicznych musi być ustalona hierarchia, co zapewnia zdefiniowanie obiektów niższego i wyższego rzędu. Przykładowa agregacja podmiotowa może dotyczyć agregacji produkcji artykułów biurowych, wytworzonych przez grupę zakładów produkcyjnych, funkcjonujących w ramach jednego przedsiębiorstwa. Na podstawie ustalonej hierarchii wyróżnić można obiekty ekonomiczne niższego rzędu (zakłady produkcyjne) oraz obiekt ekonomiczny wyższego rzędu, przedsiębiorstwo. W tym przypadku agregacja podmiotowa dotyczy procesu agregacji danych, w wyniku którego wyznaczana jest całkowita wartość wytworzonego asortymentu biurowego dla przedsiębiorstwa, jako suma wartości produkcji dla pojedynczych zakładów produkcyjnych.

Agregacja dóbr polega na agregacji danych dotyczących zbioru cech pojedynczego obiektu ekonomicznego, w wyniku czego otrzymywane są dane zagregowane dla cechy agregatowej (Pawłowski, 1969, s. 237–238). Agregacja tego typu może dotyczyć dowolnych obiektów ekonomicznych i stosowana jest najczęściej w przypadku występowania zbyt wielu cech obiektu ekonomicznego. W tym przypadku dla wybranej cechy (zbioru cech) obiektu ekonomicznego tworzona jest hierarchia. W związku z tym zastosowanie agregacji dóbr ma na celu redukcję zbyt dużej liczby danych szczegółowych. Wykrycie prawidłowości na podstawie danych odniesionych do wielu, bliskich sobie cech obiektu ekonomicznego jest bardzo utrudnione. W związku z tym przyjęcie odpowiedniego stopnia agregacji dóbr znacznie ułatwia przeprowadzenie analiz ekonomicznych. Ponownie przedsiębiorstwo produkcyjne materiałów biurowych jest tutaj dobrym przykładem dla pokazania zastosowania agregacji dóbr. Już w przypadku pojedynczego zakładu produkcyjnego dysponujemy całym wachlarzem asortymentu biurowego. Agregacja dóbr może zatem dotyczyć procesu agregacji towarów tworzących wybraną grupę asortymentową. Przyjęta grupa asortymentowa stanowi zbiór dóbr zebranych razem w ramach ustalonego kryterium. Materiały piśmiennicze w postaci piór, długopisów, ołówków mogą tworzyć grupę materiałów piśmienniczych. W wyniku agregacji dóbr otrzymane zostaną dane dotyczące wartości produkcji dla ustalonej grupy asortymentowej, wyznaczonej jako suma wartości produkcji poszczególnych dóbr należących do tej grupy.

W przypadku agregacji w czasie Pawłowski zwraca uwagę na fakt, że dane zagregowane dotyczą zazwyczaj dłuższych okresów (np. miesiąc, kwartał, rok), a nie wybranych momentów czasu (Pawłowski, 1969, s. 238). Wyróżnić należy tutaj dwa najważniejsze problemy. Pierwszym problemem jest wybór właściwego okresu agregacji danych na wyższym poziomie agregacji. Im dłuższe okresy, tym mniej informacji dotyczących dynamiki badanych zdarzeń

ekonomicznych zawierają dane¹⁶. Dla bardzo krótkich okresów, dane mogą charakteryzować się zbyt dużą zmiennością, co może uniemożliwić ocenę systematycznych zmian w czasie.

Drugi problem związany jest ze sposobem wyznaczania wartości danych dla ustalonych okresów. Możliwe jest wyznaczenie wartości danych w postaci zasobów lub strumieni (Pawłowski, 1969, s. 255; Zeliaś, 1991, s. 33; Zieliński, 1991, s. 6). Na przykład dla zakładu produkcyjnego wielkość zużytych surowców, półfabrykatów, jak i wielkość wytworzonej produkcji w wybranym okresie powinno się przedstawiać w postaci strumieni. Natomiast wartość uzbrojenia stanowisk pracy, zaangażowanego w procesie produkcji, przedstawiana powinna być jako zasób.

W związku z potencjalnymi typami agregacji danych Pawłowski określa również pojęcie stopnia agregacji danych (Pawłowski, 1969, s. 236; Bołt, Krauze, Kulawczuk, 1985, s. 20). W najprostszym przypadku agregacja może być dwustopniowa, gdzie na podstawie danych niższego rzędu wykonywany jest proces agregacji, w wyniku czego otrzymywane są dane wyższego rzędu. Agregacja dwustopniowa wymaga wyboru dwóch poziomów agregacji, poziomu niższego rzędu i poziomu wyższego rzędu. Dla obydwu poziomów agregacji konieczne jest określenie obiektów ekonomicznych oraz dóbr, których dane dotyczą, a także zakresu czasowego. Agregacja wielostopniowa dotyczy przypadku, gdzie proces agregacji danych wykonywany jest po raz kolejny. Jeśli na podstawie danych wynikających z agregacji dwustopniowej wykonana zostanie kolejna agregacja, to mamy do czynienia z agregacją trójstopniową. W takiej sytuacji badacz powinien być świadomy hierarchii poziomów agregacji. Dla każdego sąsiadujących ze sobą dwóch poziomów agregacji, jeden stanowi poziom niższego rzędu, a drugi poziom wyższego rzędu. Największym stopniem czytelności danych charakteryzuje się agregacja dwustopniowa. W przypadku agregacji wielostopniowych może dojść do zamazania się powiązań między danymi z różnych poziomów agregacji. Dlatego należy podkreślić, że zastosowanie agregacji wielostopniowej oraz związanych z nią kolejnych poziomów agregacji powinno wynikać przede wszystkim z potrzeb informacyjnych odbiorcy. Zastosowanie agregacji trójstopniowej można zaprezentować na przykładzie dowolnej branży przedsiębiorstw. Na pierwszym poziomie agregacji, dane dotyczące wartości produkcji mogą być przedstawione dla wszystkich przedsiębiorstw osobno. Na drugim poziomie agregacji dane mogą być odniesione do grup przedsiębiorstw, gdzie podział na grupy powinien wynikać ze specyfiki branży. Natomiast na trzecim poziomie agregacji, dane zagregowane dotyczą wartości produkcji dla całej branży.

¹⁶ Dodatkowy problem powstaje, gdy ustalono występowanie istotnych wahań periodycznych dla wybranych okresów, a dane zostały zagregowane dla dłuższych okresów. W takim przypadku niemożliwa jest analiza potencjalnych wahań periodycznych.

1.2.2. System agregacji danych

Podstawowym celem pozyskiwania danych pierwotnych oraz procesu ich agregacji jest próba otrzymania danych zgodnych z potrzebami informacyjnymi potencjalnego odbiorcy. Dla potencjalnego odbiorcy dane są przede wszystkim źródłem istotnych informacji. Dlatego wybór zbioru danych pierwotnych, a następnie typu, rodzaju, jak i poziomu agregacji danych uzależniony jest od tego, jakich informacji mają dostarczyć dane, ponieważ potrzeby informacyjne różnych odbiorców, jak i sposób wykorzystania przez nich danych mogą być odmienne (Oleński, 2001, s. 221–240).

W związku z kwestią różnych potrzeb informacyjnych oraz sposobów wykorzystania danych przedstawiona zostanie przez autora propozycja podziału odbiorców danych na trzy grupy, przy wykorzystaniu kryterium charakteru prowadzonej przez nich działalności¹⁷. Zgodnie z tym kryterium pozyskane dane mogą być wykorzystane przez odbiorców danych w ramach prowadzonej działalności gospodarczej, działalności naukowej lub działalności informacyjnej.

Pierwszą grupę odbiorców danych stanowią odbiorcy zajmujący się działalnością gospodarczą. Zbiory danych pierwotnych oraz dane zagregowane powinny zapewnić odbiorcom uzyskanie takich informacji, które pozwolą na podjęcie właściwych decyzji ekonomicznych. Można stwierdzić, że w tym przypadku gromadzenie danych pierwotnych oraz agregacja danych wykonywane są na potrzeby procesu decyzyjnego (np. w przedsiębiorstwie).

Drugą grupę stanowią odbiorcy prowadzący działalność naukową. W przypadku tego typu odbiorców posiadane dane powinny pozwolić na wyciągnięcie wniosków w ramach prowadzonych badań naukowych. Oznacza to, że rejestracja zdarzeń ekonomicznych oraz agregacja danych dotyczy obiektów ekonomicznych oraz ich cech, które są przedmiotem wykonywanych analiz.

Trzecią, ostatnią grupę stanowią odbiorcy danych prowadzący działalność informacyjną¹⁸. Celem działalności tych odbiorców jest uzyskanie takich danych, które mogą zostać powszechnie udostępnione lub sprzedane w celach informacyjnych innym finalnym odbiorcom. Działalność taką prowadzą przede wszystkim podmioty statystyki publicznej oraz niepubliczne agencje badań rynkowych. Potrzeby informacyjne w ramach prowadzonej działalności informacyjnej są zdeterminowane przez potrzeby informacyjne potencjalnych odbiorców finalnych. Oznacza to, że podmioty te mają bardzo trudne zadanie w postaci ustalenia zarówno potencjalnych odbiorców danych, jak i ich potrzeb informacyjnych. Ostatecznie bowiem, to od finalnego odbiorcy danych zależy,

¹⁷ Charakter działalności ustalony został przy założeniu, że dla odbiorców zbiory informacji powinny pełnić dwie główne funkcje, funkcję informacyjną oraz funkcję decyzyjną (Oleński, 2001, s. 221–230).

¹⁸ Odbiorców danych prowadzących działalność informacyjną można określić jako odbiorców bezpośrednich, natomiast odbiorców finalnych, którym dane zostały udostępnione, jako odbiorców pośrednich.

w jakich celach wykorzystane zostaną dane. Dla finalnego odbiorcy dane te mogą stać się źródłem informacji w ramach prowadzonej działalności gospodarczej, czy też działalności naukowej. Przedsiębiorca wykorzysta tego typu dane jako podstawę procesu decyzyjnego, a pracownik naukowy jako podstawę do otrzymania wyników w ramach prowadzonych badań.

Potencjalne warianty agregacji usystematyzowane zostały w poprzednim podrozdziale poprzez wyróżnienie trzech typów agregacji danych: agregacji podmiotowej, agregacji dóbr oraz agregacji w czasie. Jednoczesne rozważenie wymienionych typów agregacji oraz zaproponowanych trzech grup odbiorców danych pozwala na propozycję kolejnych pojęć dotyczących procesu agregacji w postaci układu odbiorców danych, układu podmiotowego agregacji, układu przedmiotowego agregacji oraz układu czasowego agregacji.

Zaproponowane pojęcia omówione zostaną na przykładzie hipotetycznego przedsiębiorstwa produkującego materiały biurowe i posiadającego kilka zakładów produkcyjnych. W celu zaplanowania procesu pozyskania danych pierwotnych oraz ich agregacji w ramach hipotetycznego przedsiębiorstwa, należy wyjść od ustalenia zaproponowanych przez autora pojęć, tj.: układu odbiorców danych, układu podmiotowego agregacji, układu przedmiotowego agregacji oraz układu czasowego agregacji. Pozwoli to następnie na uszczegółowienie potrzeb informacyjnych oraz ustalenie zbioru pojedynczych zdarzeń ekonomicznych, które staną się źródłem danych pierwotnych.

W hipotetycznym przedsiębiorstwie wyróżnić można wielu odbiorców danych, od dyrektorów, kierowników wybranych działów, po poszczególne osoby odpowiedzialne za pojedyncze zadania. Ustalenie układu odbiorców polegać będzie na identyfikacji wszystkich potencjalnych odbiorców oraz ich potrzeb informacyjnych w kontekście procesu decyzyjnego w przedsiębiorstwie. Część potrzeb informacyjnych może się pokrywać dla większej grupy odbiorców, a część może się znacznie różnić. Wyznaczony układ odbiorców danych determinował będzie dalszy proces pozyskiwania danych pierwotnych oraz proces agregacji danych.

Po wyznaczeniu układu odbiorców danych, należy następnie ustalić potencjalne obiekty ekonomiczne oraz ich hierarchię. Zbiór obiektów ekonomicznych wraz z hierarchią utworzy układ podmiotowy agregacji i będzie stanowił podstawę do wykonania agregacji danych typu podmiotowego. Kolejnym krokiem będzie wyznaczenie zbioru cech w ramach ustalonego układu podmiotowego oraz ustalenie hierarchii cech w ramach różnych kryteriów¹⁹. Zbiór cech wraz z ustaloną hierarchią utworzy układ przedmiotowy. W ostatnim kroku należy ustalić, dla jakich potencjalnych okresów dane powinny być pozyskiwane lub agregowane. Ustalony zbiór okresów, od najkrótszych do najdłuższych, utworzy układ czasowy agregacji. Należy podkreślić, że przyjęcie kolejnych układów ma charakter rekurencyjny. Oznacza to, że budowa układu podmiotowego

¹⁹ Pojedyncze cechy mogą stać się elementem różnych schematów hierarchii w zależności od przyjętego kryterium.

powinna zostać przeprowadzona na podstawie analizy układu odbiorców danych. Wyznaczenie układu przedmiotowego powinno opierać się na układzie odbiorców danych oraz układzie podmiotowym agregacji. Natomiast przy budowie układu czasowego agregacji powinno się uwzględniać wszystkie trzy poprzednie układy. Zaproponowane cztery układy agregacji powinny stanowić podstawę organizacji procesu gromadzenia danych pierwotnych. Dodatkowo przyjęcie układu odbiorców danych, układu podmiotowego, przedmiotowego i czasowego pozwala na ustalenie potencjalnych wariantów agregacji danych.

Ogromna ilość gromadzonych danych pierwotnych oraz danych zagregowanych wymaga ich formalnego usystematyzowania, ponieważ do każdego odbiorcy z układu odbiorców danych powinny zostać przyporządkowane odpowiednie dane. Oznacza to potrzebę budowy systemu agregacji danych, który pozwalałby, za pomocą ustalonych procedur, na rejestrację danych, ich gromadzenie oraz agregację²⁰. System agregacji danych powinien funkcjonować w taki sposób, żeby do wybranych odbiorców trafiały takie dane, które niosą ze sobą użyteczne z ich punktu widzenia informacje.

W związku z tym autor proponuje następującą definicję systemu agregacji danych. System agregacji danych tworzyć będą: układ odbiorców, układ podmiotowy agregacji, układ przedmiotowy agregacji i układ czasowy agregacji oraz zbiór reguł dotyczących gromadzenia danych pierwotnych, procesu agregacji danych oraz udostępniania danych potencjalnym odbiorcom. Należy podkreślić, że dla każdej z trzech grup odbiorców danych, utworzone systemy agregacji danych mogą się istotnie różnić od siebie. W przypadku dwóch pierwszych grup odbiorców danych, system agregacji danych powinien zawierać reguły dotyczące rejestrowania pierwotnych zdarzeń ekonomicznych, ich agregacji lub agregacji posiadanych danych wtórnych, które wykorzystane zostaną wyłącznie w ramach prowadzonych przez nie działalności. System musi zostać tak opracowany, żeby posiadane dane pozwalały na otrzymanie istotnych informacji z punktu widzenia procesu podejmowania decyzji lub prowadzonych badań naukowych. Należy podkreślić, że zbiór reguł składający się na system agregacji danych zależeć będzie w głównym stopniu od indywidualnych celów decydentów prowadzących działalność gospodarczą lub podjętych problemów naukowych.

Znacznie trudniejszym zadaniem jest opracowanie systemu agregacji danych dla odbiorców danych prowadzących działalność informacyjną. W tym przypadku podmiot udostępniający dane musi przewidywać potencjalne potrzeby informacyjne odbiorców finalnych i przewidywania te muszą znaleźć swoje odzwierciedlenie w funkcjonującym systemie agregacji danych.

Za przykład zaproponowanego przez autora systemu agregacji danych, może posłużyć system funkcjonujący w Głównym Urzędzie Statystycznym,

²⁰ Autor zakłada, że ze względu na istniejące technologie informacyjne nie ma technicznych problemów z rejestracją danych pierwotnych, ich agregacją oraz udostępnianiem.

który utworzony został na potrzeby procesu pozyskiwania danych statystycznych. Główny Urząd Statystyczny, jako podmiot statystyki publicznej udostępnia zebrane dane działając zgodnie z przepisami wynikającymi z *Ustawy o Statystyce Publicznej* (1995). Udostępniane dane dotyczą różnych kategorii ekonomicznych oraz prezentowane mogą być na różnych poziomach agregacji i w różnych okresach. Dla każdego roku działalności Główny Urząd Statystyczny przygotowuje *Program Badań Statystycznych Statystyki Publicznej* (PBSSP), gdzie zawarte są założenia odnośnie do procesu pozyskania danych, ich agregacji oraz sposobu udostępniania. Przygotowywany corocznie program badań statystycznych statystyki publicznej można uznać za przykład zaproponowanego w pracy systemu agregacji danych.

Program ten składa się z dwóch części. W części pierwszej znajdują się informacje odnośnie do przyjętej metodologii badania oraz metodyki pozyskiwania pierwotnych danych statystycznych. W programie badań statystycznych statystyki publicznej zamieszczone są dla każdego z prowadzonych badań informacje dotyczące tematu badania, cykliczności badania, określony jest prowadzący badanie, cel badania, zakres podmiotowy, zakres przedmiotowy, źródła danych, rodzaje wynikowych informacji statystycznych oraz formy i terminy udostępniania wynikowych informacji statystycznych (Załącznik do *Rozporządzenia Rady Ministrów w sprawie programu badań statystycznych statystyki publicznej*, 2017, s. 20–451).

Natomiast w części drugiej znajdują się informacje dotyczące pozyskania danych statystycznych. Informacje te określają źródło danych, podmiot przekazujący dane, metodę obserwacji, zakres danych, w tym szczegółowy zakres danych osobowych, częstotliwość i termin przekazania danych, obligatoryjność albo dobrowolność przekazania, przyjęty poziom agregacji, postać i formę danych oraz miejsce przekazania uzyskanych danych (Załącznik do *Rozporządzenia Rady Ministrów w sprawie programu badań statystycznych statystyki publicznej*, 2017, s. 452–813).

1.2.3. Specyfika agregacji danych przestrzennych

W poprzednich podrozdziałach wskazano, że proces agregacji danych może dotyczyć danych różnego typu danych (np. danych przestrzennych, danych przekrojowych itd.). Wybór danego typu danych oraz sposobu ich agregacji powinien wynikać z potrzeb informacyjnych odbiorcy oraz być podejmowany w ramach przyjętego systemu agregacji danych. W dalszej części pracy uwaga skupiona zostanie wyłącznie na procesie agregacji danych przestrzennych, co pozwoli na skoncentrowanie się na problemie zmiennej jednostki odniesienia²¹.

²¹ Dane przestrzenne stanowią realizacje przestrzennych procesów ekonomicznych, które stanowią dwuwymiarowe pole losowe. Dane tego typu są przedmiotem przestrzennych analiz ekonomicznych wykonywanych w ramach ekonometrii przestrzennej. Własności danych przestrzennych zaprezentowane zostaną w dalszej części dysertacji.

Rozważane w ramach ekonometrii przestrzennej dane przestrzenne można podzielić na dwa typy, dane punktowe oraz dane obszarowe (Cressie, 1993, s. 7–13; Suhecki, 2010, s. 38–39; Suhecka, 2014, s. 21). Dane punktowe są to dane przestrzenne przyporządkowane do punktów w przestrzeni, natomiast dane obszarowe przyporządkowane są do obszarów regularnych lub nieregularnych²² (Suhecka, 2014, s. 52–55). Pierwotne dane przestrzenne odniesione są najczęściej do punktów w przestrzeni (dane punktowe). Dane te obrazują zaobserwowane wartości zdarzeń ekonomicznych w konkretnych punktach w przestrzeni. W przypadku ekonomicznych danych przestrzennych, dane tego typu pozyskiwane są w większości przez instytucje statystyki publicznej. Dane te są następnie agregowane do danych przestrzennych odniesionych do obszarów nieregularnych. Zagregowane w ten sposób dane przyporządkowywane są do obszarów w ramach ustalonych układów jednostek terytorialnych. Agregacja danych punktowych do danych przestrzennych odniesionych do obszarów nieregularnych jest konieczna. Dane punktowe ze względu na swoją szczegółowość byłyby niemożliwe do interpretacji. Dodatkowo, obserwowane są tylko zdarzenia ekonomiczne dla wybranych punktów w przestrzeni i zgodnie z zasadami statystyki matematycznej dane te są uogólniane na cały obszar nieregularny. Oznacza to, że w ramach pozyskania ekonomicznych danych przestrzennych nie wykonuje się spisu wszystkich jednostek, a przeprowadzane jest badanie reprezentatywne na losowej próbie. Agregacja danych wynika również z konieczności zachowania tajemnicy statystycznej, gdzie musi zostać utrzymana anonimowość obiektów badania odniesionych do punktów w przestrzeni. Biorąc pod uwagę konieczność uogólniania informacji z próby na cały obszar nieregularny, podstawową kwestią jest problem wiarygodności otrzymanych w ten sposób danych przestrzennych. Wiarygodność danych przestrzennych zapewnić ma ich pozyskanie przez wyspecjalizowane do tego celu jednostki, najczęściej jednostki statystyki publicznej. Dane takie są wiarygodne dzięki zachowaniu odpowiedniej metodologii i metodyki badania. Są również reprezentatywne dla badanych regionów poprzez odpowiedni dobór próby. Dodatkowo, w przypadku jednostek statystyki publicznej, pozyskiwanie danych jest powtarzalne, co również jest ogromnym atutem tych danych.

Jednostki statystyki publicznej pozyskują i udostępniają dane przestrzenne w krajach członkowskich Unii Europejskiej w ramach administracyjnych układów jednostek terytorialnych zgodnych z klasyfikacją NUTS opracowanej przez Eurostat (Nomenclature of Units for Territorial Statistics, zob. Eurostat,

²² Wyróżniane są także dane powierzchniowe, jednak w ekonomii dane tego typu nie mają powszechnego zastosowania (Suhecki, 2010, s. 38). Dane powierzchniowe charakteryzują się ciągłą zmiennością i mogą dotyczyć na przykład powierzchni terenu. Należy zwrócić uwagę na fakt, że dane powierzchniowe uzyskiwane są najczęściej na podstawie danych przestrzennych odniesionych do punktów w przestrzeni, gdzie w kolejnym kroku wykonywana jest ekstrapolacja trendu powierzchniowego lub analiza potencjału, w wyniku czego uzyskiwane są dane powierzchniowe (techniki analizy danych powierzchniowych zob. Suhecka, 2014, s. 53).

2015)²³. Celem wprowadzenia klasyfikacji NUTS było zapewnienie zbierania, opracowywania i udostępniania na obszarze UE porównywalnych danych dla państw członkowskich. Układ jednostek terytorialnych NUTS 0 określa kraje Unii Europejskiej. W przypadku Polski, w ramach układu jednostek terytorialnych NUTS 1 zdefiniowane są regiony, w ramach układu NUTS 2 województwa, w ramach układu NUTS 3 podregiony, w ramach układu NUTS 4 powiaty oraz w ramach układu NUTS 5 gminy. Kolejne poziomy NUTS nie są przypadkowe i analiza większości zjawisk ekonomicznych oraz zależności między nimi w ramach klasyfikacji NUTS prowadzi do uzyskania poprawnych wyników badań. W konfrontacji do obszarów nieregularnych (np. obszary z układu NUTS) możliwe jest przeprowadzenie przestrzennych badań ekonomicznych na podstawie układu obszarów regularnych. W literaturze wskazywane jest, że dane przestrzenne odniesione do obszarów regularnych zawierają najbardziej pożądane informacje w ramach prowadzonych przestrzennych analiz ekonomicznych (zob. Szulc 2007, s. 54). Jednak dane przestrzenne udostępniane przez jednostki statystyki publicznej przyporządkowywane są do obszarów nieregularnych i nie jest sprawą łatwą ich przekształcenie dla obszarów regularnych. Dodatkowo zgromadzone przez jednostki statystyki publicznej punktowe dane przestrzenne nie mogą być publicznie udostępnione. Oznacza to, że pozyskanie pierwotnych, punktowych danych przestrzennych jest najczęściej niemożliwe ze względu na tajemnicę statystyczną, obowiązującą zgodnie z aktualnym ustawodawstwem (*Ustawa o statystyce publicznej*, 1995). W związku z tym należy podkreślić dwie ważne kwestie związane z danymi odniesionymi do obszarów regularnych. Pierwszą stanowi konieczność samodzielnego pozyskania punktowych danych przestrzennych, na podstawie których wykonany byłby proces agregacji danych dla obszarów regularnych. Pozyskanie punktowych danych przestrzennych jest bardzo kosztowne i dodatkowo konieczne jest ustalenie właściwej metodyki badania, pozyskane dane muszą być uogólnione zgodnie z zasadami statystyki matematycznej. Ze względu na wymienione kwestie, pozyskanie danych punktowych w samodzielnym zakresie jest bardzo trudnym zadaniem. W przypadku drugiej kwestii występuje trudność agregacji zakupionych danych punktowych do danych obszarowych odniesionych do obszarów regularnych. Z powodu wymienionych trudności oraz powszechnej dostępności danych odniesionych do obszarów nieregularnych, dane odniesione do obszarów regularnych nie znajdują szerszego zastosowania w przestrzennych badaniach ekonomicznych. Jednak pomimo ogromnych trudności z ich pozyskaniem, mogłyby w znacznym stopniu wzbogacić przestrzenne analizy ekonomiczne. Ostateczna decyzja o wyborze, czy wykonywane przestrzenne badania ekonomiczne oparte powinny być na danych statystycznych odniesionych do obszarów regularnych, czy też obszarów nieregularnych (np. układy NUTS) należy do badacza. Jednak ze względu na koszt oraz czas pozyskania danych, badania przestrzenne dotyczące Polski, czy Unii Europejskiej prawie zawsze oparte są na obszarach nieregularnych zgodnych z klasyfikacją NUTS,

²³ Pojedyncze jednostki terytorialne układu NUTS są obszarami nieregularnymi.

co wynika z dostępności²⁴ oraz wiarygodności udostępnianych przez jednostki statystyki publicznej danych statystycznych.

Na podstawie przeprowadzonych rozważań dotyczących dostępności wybranych typów ekonomicznych danych przestrzennych, kolejnym założeniem w niniejszej pracy będzie rozpatrywanie wyłącznie danych przestrzennych odniesionych do obszarów nieregularnych. W związku z tym w dalszej części pracy uwaga zostanie skoncentrowana tylko na procesie agregacji danych obszarowych, gdzie wyjściowymi pierwotnymi danymi będą dane przestrzenne odniesione do obszarów nieregularnych.

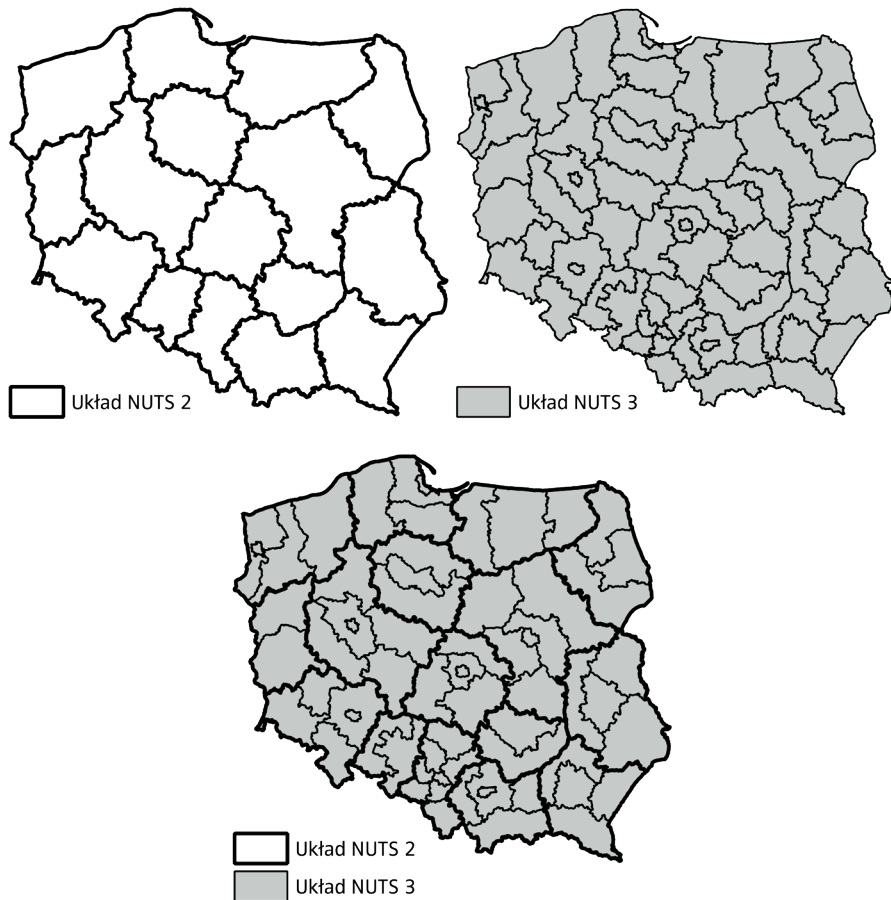
Agregację danych przestrzennych należy przyporządkować do typu agregacji podmiotowej. W tym przypadku zamiast kryterium hierarchicznego obiektów ekonomicznych stosowane jest kryterium geograficzne (Pawłowski, 1969, s. 237). Przy założeniu, że rozważania dotyczyć będą wyłącznie obszarów nieregularnych, system hierarchii oparty jest na granicach regionów. Obiekty niższego rzędu (regiony) zawierają się przestrzennie w granicach obiektów wyższego rzędu (makroregionów). Najczęściej hierarchia przestrzenna regionów przyjmowana jest na podstawie granic układów jednostek terytorialnych. W ramach prowadzonych badań można przyjąć układy jednostek terytorialnych na różnych poziomach agregacji. Przykładem takich układów mogą być układ podregionów (NUTS 3) oraz układ województw (NUTS 2). Dla obszaru Polski układ NUTS 3 jest układem jednostek terytorialnych na niższym poziomie agregacji, a układ jednostek terytorialnych NUTS 2 układem na wyższym poziomie agregacji. W tym przypadku przyjęte układy jednostek terytorialnych tworzą hierarchię przestrzenną. Układy NUTS 3 oraz NUTS 2 przedstawione zostały na rysunku 1.1, gdzie widoczna jest hierarchia przestrzenna, zgodnie z którą regiony niższego rzędu zawierają się w regionach wyższego rzędu.

Opisane wyżej układy danych terytorialnych NUTS 3 i NUTS 2 pozwalają na analizę zjawisk na niższym i wyższym poziomie agregacji. Poziom agregacji oznacza przyjęcie układu jednostek terytorialnych o odpowiednich charakterystykach, do którego zostaną przyporządkowane dane przestrzenne. Wybrany układ posiada ustaloną liczbę jednostek terytorialnych określonego kształtu oraz wielkości. Układy jednostek terytorialnych mogą charakteryzować się różną liczbą jednostek terytorialnych. Biorąc pod uwagę obszar o ustalonych granicach (np. obszar Polski) istnieje możliwość ustalenia wielu potencjalnych układów jednostek terytorialnych o różnej liczbie n podobszarów. Najniższy poziom agregacji dotyczyć będzie układu jednostek terytorialnych o największej liczbie n podobszarów. Natomiast najwyższy poziom agregacji dotyczyć będzie układu składającego się z najmniejszej liczby n podobszarów.

²⁴ Część gromadzonych przez jednostki statystyki publicznej danych statystycznych jest odpłatna. Koszt zakupu takich danych jest nieporównywalnie niższy od próby samodzielnego pozyskania danych lub skorzystania z usług komercyjnych instytucji.

Rysunek 1.1.

Hierarchia przestrzenna oparta na układach NUTS 3 oraz NUTS 2



Źródło: opracowanie własne.

W przypadku danych przestrzennych odniesionych do obszarów nieregularnych pojawia się problem porównywalności wybranych obszarów. O poziomie większości zjawisk ekonomicznych na pojedynczym obszarze decydują w głównym stopniu dwa aspekty, wielkość obszaru (powierzchnia, potencjał ludności, zasobów pracy) oraz poziom rozwoju ekonomicznego przekładający się na poziom tych zjawisk. Obszary nieregularne różnią się między sobą pod względem wielkości, a zatem należy wyeliminować wpływ wielkości obszaru podczas pomiaru zjawisk ekonomicznych. Po eliminacji wpływu wielkości obszaru, poziom badanego zjawiska powinien wynikać wyłącznie z poziomu rozwoju ekonomicznego regionu. Oznacza to, że większość danych przestrzen-

nych wyrażających poziom zjawisk ekonomicznych należy zważyć wielkością obszaru (Zeliaś, 1991, s. 26–28, 32–33, 37–38).

W celu zapewnienia porównywalności obszarów (regionów) stosuje się dwa sposoby wyrażenia wartości danych przestrzennych odniesionych do obszarów nieregularnych. Pierwszym sposobem zapewnienia porównywalności obszarów (regionów) jest policzenie średniej wartości rozpatrywanego zjawiska ekonomicznego dla wybranego obszaru, co pozwoli na uniknięcie wpływu wielkości obszaru na wyniki analiz. Otrzymane w ten sposób dane przestrzenne na niższym poziomie agregacji wyrażone są w postaci wartości średnich. Przykładem takiego działania może być udostępnianie danych przestrzennych dotyczących poziomu płac, których wartości stanowią oszacowanie średniej płacy dla każdego z obszarów.

Drugi sposób polega na tym, że wyjściowe dane przestrzenne są ważone w stosunku do odpowiednio rozumianej wielkości obszaru. W wyniku zważenia wyjściowych danych przestrzennych otrzymywane są ważone dane przestrzenne o innym mianie (Zeliaś, 1991, s. 37–38). Taka sytuacja występuje na przykład podczas pomiaru wielkości Produktu Krajowego Brutto dla wybranego obszaru. Wartość PKB uzależniona jest w znacznym stopniu od wielkości obszaru. Oznacza to, że dane przestrzenne dotyczące wielkości PKB w wyrażeniu bezwzględny nie powinny być wykorzystywane w analizach dotyczących porównań obszarów. Dopiero wyrażenie wielkości PKB w stosunku do liczby mieszkańców (wielkości regionu) pozwala na uzyskanie danych przestrzennych, które można wykorzystać przy porównywaniu regionów ze względu na kryterium rozwoju gospodarczego. Uwzględnienie wagi obszarów oraz porównanie regionów na podstawie odpowiednio przekształconych danych przestrzennych jest ważne, ponieważ zapewnia poprawność wykonywanych przestrzennych badań ekonomicznych.

Przestrzenne badania ekonomiczne powinny być wykonywane na podstawie danych przestrzennych odniesionych do wartości średnich lub ważonych odpowiednio rozumianą wielkością regionu, co zapewni możliwość porównywania obszarów oraz ich oceny. W związku z tym również w wyniku agregacji danych przestrzennych powinny być otrzymywane tego typu dane. Proces agregacji danych przestrzennych odniesionych do obszarów nieregularnych polega na zagregowaniu wyjściowych danych przestrzennych na wybranym poziomie agregacji, w wyniku czego otrzymane zostaną dane przestrzenne na wyższym poziomie agregacji. Oznacza to, że w procesie agregacji danych należy wybrać co najmniej dwa układy jednostek terytorialnych, układ niższego rzędu, dla którego posiadamy dane przestrzenne oraz układ wyższego rzędu, dla którego wykonany zostanie proces agregacji danych. Na potrzeby agregacji danych przestrzennych wybierane są najczęściej dane dostępne na możliwie najniższym poziomie agregacji. Dane takie uzyskiwane są w wyniku agregacji danych punktowych, gdzie proces agregacji wykonywany jest przez odpowiednie jednostki statystyki publicznej lub inne wyspecjalizowane

podmioty. Zgodnie z przyjętymi poziomami agregacji, grupy regionów z układu jednostek terytorialnych na niższym poziomie agregacji tworzą pojedyncze regiony z układu na wyższym poziomie agregacji. Zależność ta przedstawiona została na przykładzie układów jednostek terytorialnych NUTS 3 oraz NUTS 2 (zob. rysunek 1.1).

Wykonanie procesu agregacji oznacza zebranie danych z odpowiedniej grupy regionów dla układu niższego poziomu agregacji oraz ich sumaryczne przedstawienie dla odpowiednich regionów dla układu na wyższym poziomie agregacji. Należy jednak zwrócić uwagę na fakt, że zarówno w przypadku danych przestrzennych wyrażonych w wartościach średnich, jak i danych ważonych wielkością obszarów pojawia się problem ich prawidłowej agregacji. W pracach dotyczących problemu zmiennej jednostki odniesienia proces agregacji danych przestrzennych oparty jest w większości przypadków na wykonaniu operacji sumy (Openshaw, Taylor, 1979, s. 129), co jest zabiegiem niewystarczającym. W związku z tym istnieje konieczność wypracowania prawidłowej procedury agregacji danych przestrzennych. Autor pracy proponuje następujące rozwiązanie. W przypadku, gdy dane przestrzenne na niższym poziomie agregacji wyrażone są w postaci wartości średnich, procedura agregacji danych polega na wyborze odpowiedniej grupy regionów i policzeniu na ich podstawie średniej ważonej. Otrzymana wartość średnia przyporządkowywana jest następnie do odpowiedniego regionu z układu na wyższym poziomie agregacji.

Proces agregacji tego typu danych przestrzennych określony został zgodnie z równaniem:

$$\bar{x}_j = \sum_{i=1}^{n_j} w_i \bar{x}_i, \quad (1.1)$$

gdzie \bar{x}_j stanowi wartość średnią dla j -tego regionu z układu na wyższym poziomie agregacji, \bar{x}_i stanowi wartość średnią dla i -tego regionu z układu na niższym poziomie agregacji, w_i są to wagi przyporządkowane do i -tych regionów, n_j stanowi liczbę jednostek terytorialnych niższego rzędu tworzących j -tą jednostkę wyższego rzędu. Waga dla i -tego regionu wyznaczana jest na podstawie liczby jednostek wykorzystanych do obliczenia średniej \bar{x}_i .

W przypadku ważonych danych przestrzennych proces agregacji się komplikuje, ponieważ w celu porównywalności obszarów należy zważyć dane wielkością regionu. W takiej sytuacji agregacja danych musi być wykonana w sposób pośredni. W związku z tym wyjściowe, ważne dane przestrzenne muszą zostać rozbite na niższym poziomie agregacji na licznik x_i^1 i mianownik x_i^2 zgodnie z równaniem:

$$x_i = \frac{x_i^1}{x_i^2}, \quad (1.2)$$

gdzie x_i oznacza wartość dla i -tego regionu z układu na niższym poziomie agregacji, a x_i^1, x_i^2 są to dane cząstkowe na wybranym poziomie agregacji.

W wyniku tego otrzymane zostaną dwa zbiory danych przestrzennych. Proces agregacji powinien dotyczyć każdego zbioru danych osobno, gdzie agregacja polegać będzie na sumowaniu danych dla odpowiedniej grupy regionów z układu na niższym poziomie agregacji i odniesieniu tej sumy zgodnie z równaniem (1.3) do odpowiedniego regionu z układu na wyższym poziomie agregacji. W ten sposób otrzymywane są dwa zbiory zagregowanych danych przestrzennych na wyższym poziomie agregacji y_j^1, y_j^2 . W ostatnim kroku, na podstawie równania (1.4), wyznaczane są zagregowane wartości ważonych danych przestrzennych na wyższym poziomie agregacji jako iloraz odpowiednich wartości y_j^1 oraz y_j^2 :

$$\frac{y_j^1}{y_j^2} = \frac{\sum_{i=1}^{n_j} x_i^1}{\sum_{i=1}^{n_j} x_i^2}, \quad (1.3)$$

$$y_j = \frac{y_j^1}{y_j^2}, \quad (1.4)$$

gdzie x_i oznacza wartość dla i -tego regionu z układu na niższym poziomie agregacji, y_j oznacza wartość dla j -tego regionu z układu na wyższym poziomie agregacji, $x_i^1, x_i^2, y_j^1, y_j^2$ są to dane cząstkowe na wybranych poziomach agregacji, a n_j stanowi liczbę jednostek terytorialnych niższego rzędu tworzących tą jednostkę wyższego rzędu.

Rozdział 2.

Poprawność przestrzennych badań ekonomicznych

2.1. Etapy przestrzennych badań ekonomicznych

2.1.1. Etapy agregacji danych przestrzennych

W poprzednim podrozdziale przedstawione zostały podstawowe kwestie dotyczące roli agregacji danych w przestrzennych badaniach ekonomicznych. Kolejnym krokiem powinno być rozpatrzenie zagadnienia poprawności wykonywanego procesu agregacji. Problem ten wymaga jednak rozważenia zagadnienia agregacji w szerszym kontekście¹. Wynika to z faktu, że agregacja danych przestrzennych stanowi integralną część wykonywanych przestrzennych badań ekonomicznych. Integralność procesu agregacji z całością przestrzennych badań ekonomicznych oznacza, że decyzje dotyczące przyjęcia układów jednostek terytorialnych, wyboru danych przestrzennych, czy procesu agregacji muszą być podejmowane wyłącznie w ramach wykonywanych badań.

W związku z tym zachodzi konieczność rozważenia zagadnienia poprawności przestrzennych badań ekonomicznych, dla których zaproponowany zostanie podział na trzy, następujące po sobie etapy. Etapy przestrzennych badań ekonomicznych dotyczą kolejno, postawienia problemu badawczego, wykonania przestrzennej analizy ekonomicznej oraz próby rozwiązania problemu

¹ Zeliaś proponuje ustalenie na potrzeby badań ekonomicznych „struktur gospodarczych”, które powinny być rozpatrywane w czterech aspektach, rodzajowym, typologicznym, przestrzennym i czasowym. W prowadzonych badaniach każdy aspekt struktury gospodarczej może być wykorzystany do utworzenia zagregowanych struktur wyższego rzędu (Zeliaś, 1991, s. 33–37).

badawczego na podstawie wniosków uzyskanych z analizy (zob. tabela 2.1). Przedstawiona propozycja podziału przestrzennych badań ekonomicznych stanowi rozwinięcie badań z pracy Pietrzak (2014a). Jak wynikać będzie z przeprowadzonego w rozdziale trzecim przeglądu literatury dotyczącej problemu zmiennej jednostki odniesienia, zaproponowane etapy badań związane z postawieniem problemu badawczego i próbą jego rozwiązania (etap pierwszy i trzeci) są najczęściej pomijane lub wykonywane w niewystarczającej części. W przypadku większości przedstawianych w literaturze wyników dotyczących problemu zmiennej jednostki odniesienia, badacze koncentrują się głównie na wykonaniu etapu drugiego, co ma negatywny wpływ na otrzymywane wyniki, zarówno w kontekście rozważań teoretycznych, jak i empirycznych analiz². Pomijanie wymienionych etapów w publikacjach dotyczących przestrzennych badań ekonomicznych oraz problemu zmiennej jednostki odniesienia stanowi najważniejsze uzasadnienie dla treści niniejszego podrozdziału.

Tabela 2.1.

Etapy przestrzennych badań ekonomicznych

Etapy
Etap 1. Postawienie problemu badawczego
Etap 2. Przestrzenna analiza ekonomiczna
Etap 3. Próba rozwiązania problemu badawczego na podstawie wniosków uzyskanych z analizy

Źródło: opracowanie własne.

Zgodnie z informacjami zamieszczonymi w tabeli 2.1, pierwszy etap przestrzennych badań ekonomicznych związany jest z postawieniem problemu badawczego. Postawienie problemu badawczego powinno być punktem wyjścia dla przestrzennych badań ekonomicznych. Wykonanie tego etapu stanowi najważniejszy warunek zapewniający poprawność wykonywanych przestrzennych badań ekonomicznych. Dopiero w ramach postawionego problemu badawczego powinna być podejmowana decyzja o wyborze analizowanych zjawisk ekonomicznych oraz formułowane powinny być hipotezy odnośnie do własności zjawisk, czy istniejących zależności między nimi. Wszystkie decyzje dotyczące badań powinny być podejmowane w ramach postawionego problemu badawczego, gdzie badacz wykorzystuje odpowiednio dostępną wiedzę ekonomiczną oraz własne doświadczenie naukowe. Należy podkreślić, że różne problemy badawcze mogą wymagać innego aspektu wiedzy i doświadczenia badacza. Warunkiem otrzymania przez badacza właściwych wyników jest przeprowadzenie badań w taki sposób, jak wymaga tego postawiony problem

² Literatura dotycząca problematyki zmiennej jednostki odniesienia omówiona zostanie w rozdziale trzecim, należy jednak już tutaj podkreślić, że większość prac opiera się wyłącznie na analizie własności dowolnie wybranych lub symulowanych danych bez odniesienia się do problematyki ekonomicznej.

badawczy. Osiągane cele badania oraz stawiane hipotezy muszą wynikać z postawionego problemu badawczego. Niedopuszczalna jest sytuacja, kiedy badacz sam wyznacza cel badania w oderwaniu od problemu badawczego. Taka sytuacja niestety zachodzi w pracach dotyczących problemu zmiennej jednostki odniesienia, gdzie poprzez funkcję celu określa się na przykład, jaką wartość ma przyjąć współczynnik korelacji Persony w ramach wykonywanej analizy. Następnie wykonywane są odpowiednie zabiegi badawcze (przyjmowanie różnych układów jednostek terytorialnych oraz agregacja danych przestrzennych) pozwalające na uzyskanie wyników zbliżonych się do zamierzonego celu badania (Openshaw, 1984b, s. 34–35).

Drugi etap przestrzennych badań ekonomicznych stanowi analiza ekonomiczna. Celem wykonywanej analizy jest uzyskanie rezultatów, na podstawie których badacz odniesie się do postawionego problemu badawczego. Natomiast w trzecim etapie badań powinno się dokonać próby rozwiązania problemu badawczego na podstawie otrzymanych wyników z analizy. Może wystąpić również sytuacja, że wykonana analiza nie pozwoli na wyciągnięcie odpowiednich wniosków, potrzebnych do rozwiązania postawionego problemu badawczego. W takiej sytuacji badacz powinien być świadomy takiej sytuacji i po rozważeniu otrzymanych wyników powinna zostać zaproponowana alternatywna analiza lub ostatecznie problem badawczy powinien zostać przeformułowany. Kolejna analiza powinna dostarczyć dodatkowych rezultatów, które w większym stopniu przełożyłyby się na rozwiązanie podjętego problemu badawczego.

W porównaniu, z pierwszym i trzecim etapem przestrzennych badań ekonomicznych, etap przestrzennej analizy ekonomicznej jest najbardziej rozbudowany. Odnosząc się do procesu agregacji danych przestrzennych należy stwierdzić, że proces ten jest składową etapem analizy (etapu drugiego). W wyniku procesu agregacji danych otrzymywane są dodatkowe dane przestrzenne, które mogą w znaczący sposób przyczynić się do wzbogacenia formułowanych wniosków w ramach wykonywanych analiz. W treści aktualnego podrozdziału zostaną omówione tylko te elementy etapu analizy, które są niezbędne do rozpatrzenia problemu poprawności agregacji danych przestrzennych. Dopiero w kolejnym podrozdziale omówiony zostanie całościowo etap przestrzennej analizy ekonomicznej.

Pierwszy podetap przestrzennej analizy ekonomicznej stanowi ustalenie granic obszaru analizy. Analiza przestrzenna wykonana w zakresie tego obszaru powinna pozwolić na próbę rozwiązania postawionego problemu badawczego. Najważniejszą konsekwencją przyjęcia obszaru analizy jest fakt, że uzyskane wnioski na podstawie wykonanej analizy mogą odnosić się wyłącznie do granic tego obszaru lub ewentualnie do obszaru o mniejszej powierzchni. Należy podkreślić, że za pomocą obszaru analizy wyznaczane są przestrzenne granice prowadzonych badań, natomiast czynności badawcze mogą dotyczyć każdego obszaru zawierającego się w obszarze analizy.

Kolejnym podetapem, po ustaleniu granic obszaru analizy, powinien być wybór układów jednostek terytorialnych na wybranych poziomach agregacji oraz wybór danych przestrzennych, odniesionych do tych układów. Wybrane układy jednostek terytorialnych wraz z odniesionymi do nich danymi przestrzennymi określane będą w pracy jako układy danych przestrzennych. Przestrzenna analiza ekonomiczna wykonywana będzie na podstawie układów danych przestrzennych. Wykorzystane układy jednostek terytorialnych mogą wykraczać poza granice ustalonego obszaru analizy. W takiej sytuacji analiza powinna zostać ograniczona do wybranej części układu jednostek terytorialnych, zgodnie z granicami obszaru analizy. Wybór układów jednostek terytorialnych powinien wynikać z podjętego problemu badawczego oraz z dostępności danych przestrzennych. Ostatecznie, w ramach drugiego podetapu analizy, powinien zostać wybrany zbiór układów danych przestrzennych, gdzie każdy układ dotyczyć będzie innego poziomu agregacji.

Wszystkie, przyjęte układy danych przestrzennych (układy jednostek terytorialnych wraz z danymi przestrzennymi) mogą zostać wykorzystane w przestrzennej analizie ekonomicznej. Możliwa jest jednak sytuacja, w której pozyskane dane przestrzenne na niższym poziomie agregacji warto poddać analizie na wyższym poziomie agregacji. W przypadku, gdy dane przestrzenne nie są dostępne na wyższym poziomie agregacji, konieczne jest wykonanie agregacji danych przestrzennych, którą należy potraktować jako kolejny podetap przestrzennej analizy ekonomicznej. Również w przypadku agregacji danych przestrzennych zaproponowany został jej podział na cztery, kolejne etapy (zob. tabela 2.2).

Tabela 2.2.

Etapy agregacji danych przestrzennych

Etapy
Etap 1. Wybór wyjściowego, jednorodnego przyczynowo układu danych przestrzennych na niższym poziomie agregacji
Etap 2. Wybór układów jednostek terytorialnych na poziomach agregacji wyższych rządów
Etap 3. Proces agregacji danych
Etap 4. Ustalenie własności jednorodności przyczynowej dla uzyskanych układów danych przestrzennych

Źródło: opracowanie własne.

Pierwszy etap agregacji danych przestrzennych stanowi wybór wyjściowego układu danych przestrzennych na niższym poziomie agregacji. Wybrany układ danych przestrzennych powinien zawierać układ jednostek terytorialnych oraz odniesione do niego dane przestrzenne, które charakteryzują się własnością jednorodności przyczynowej. Taki układ określany będzie jako jednorodny przyczynowo układ danych przestrzennych. Rozwijając pojęcie jednorodno-

ści przyczynowej Zielińskiego w stosunku do danych przestrzennych, należy stwierdzić, że dane posiadające tę własność są wynikiem tego samego splotu przyczyn głównych dla każdego z obszarów (regionów) tworzących układ jednostek terytorialnych. Dane przestrzenne jednorodnie przyczynowo powinny właściwie odzwierciedlać oddziaływanie przyczyn w ramach przyjętego układu danych przestrzennych. Oznacza to, że wykorzystanie jednorodnego przyczynowo układu danych przestrzennych zapewnia właściwą identyfikację przestrzennych prawidłowości ekonomicznych (Tobler, 1989, s. 115–116; Anselin, 1988, s. 27; Hainig, 2003, s. 150–151). Jeżeli dane przestrzenne nie będą posiadały własności jednorodności przyczynowej, to w każdym obszarze z układu jednostek terytorialnych oddziaływać mogą różne sploty przyczyn głównych lub może dochodzić do mieszania się różnych splotów przyczyn głównych. Otrzymywane na podstawie takiego układu danych przestrzennych wyniki badań będą obciążone błędem poznawczym, którego waga będzie uzależniona od stopnia zakłóceń oddziaływania splotu przyczyn głównych. Należy podkreślić, że w ramach etapu agregacji danych przestrzennych, wybór jednorodnego przyczynowo układu danych przestrzennych daje możliwość uzyskania na wyższym poziomie agregacji układu danych przestrzennych również o takiej własności³.

W drugim etapie agregacji danych przestrzennych należy dokonać wyboru układów⁴ jednostek terytorialnych dla wyższych poziomów agregacji. Wybór konkretnych układów powinien być uzasadniony możliwością wzbogacenia wyników przestrzennej analizy ekonomicznej na podstawie dodatkowych danych zagregowanych.

W kolejnym kroku wykonywany jest proces agregacji danych przestrzennych, gdzie należy wybrać prawidłowy sposób agregacji, w zależności od tego, czy dane przestrzenne wyrażone są w wartościach średnich, czy są to dane mianowane⁵. Wykonany proces agregacji pozwala na otrzymanie danych przestrzennych, które można odnieść dla wybranego wcześniej układu (układów) jednostek terytorialnych na wyższych poziomach agregacji.

W ostatnim etapie należy ustalić, czy otrzymany w wyniku procesu agregacji układ danych przestrzennych na wyższym poziomie agregacji posiada własność jednorodności przyczynowej. Ustalenie faktu, czy mamy do czynienia z jednorodnym przyczynowo albo niejednorodnym przyczynowo układem da-

³ Jeżeli przyjęty układ danych przestrzennych na niższym poziomie agregacji nie posiada własności jednorodności przyczynowej, to również utworzony w ramach procesu agregacji układ danych przestrzennych na wyższym poziomie agregacji nie będzie posiadał tego typu własności.

⁴ Wybór układu (układów) uzależniony jest od tego, czy agregacja jest dwustopniowa, czy też wielostopniowa.

⁵ Odpowiednie procedury agregacji danych przestrzennych zostały opisane na końcu rozdziału pierwszego.

nych przestrzennych⁶, kończy czwarty etap agregacji danych. Należy podkreślić, że istnieje prawdopodobieństwo, że nowy układ danych przestrzennych na wyższym poziomie agregacji może nie posiadać własności jednorodności przyczynowej. W takiej sytuacji, podczas wyciągania wniosków z analizy, należy mieć na uwadze, czy wyniki zostały uzyskane na podstawie niejednorodnego przyczynowo układu danych przestrzennych, ponieważ mogą być one obciążone błędem poznawczym.

2.1.2. Etap przestrzennej analizy ekonomicznej

Po omówieniu zagadnień związanych z etapami agregacji danych przestrzennych dokończone zostaną rozważania dotyczące kwestii związanych z poprawnością wykonywanych przestrzennych badań ekonomicznych. W podrozdziale 2.1.1 zaproponowane zostały trzy etapy badań, przy czym omówiono etap pierwszy i trzeci oraz rozpoczęto omawianie kolejnych podetapów przestrzennej analizy ekonomicznej (etapu drugiego). W tabeli 2.3 przedstawiono ponownie trzy etapy przestrzennych badań ekonomicznych, gdzie dodatkowo przedstawiono wszystkie podetapy przestrzennej analizy ekonomicznej.

Tabela 2.3.

Etapy przestrzennych badań ekonomicznych: rozwinięcie

Etapy
Etap 1. Postawienie problemu badawczego
Etap 2. Analiza ekonomiczna
2.1. Ustalenie obszaru analizy
2.2. Wybór układu (układów) danych przestrzennych
2.3. Agregacja danych przestrzennych
2.4. Ustalenie jednorodnego zbioru układów
2.5. Identyfikacja wewnętrznej struktury danych przestrzennych
2.6. Ustalenie obszaru (podobszarów) wniosków
2.7. Wybór właściwych narzędzi badawczych
2.8. Analiza właściwa
Etap 3. Próba rozwiązania problemu badawczego na podstawie wniosków uzyskanych z analizy

Źródło: opracowanie własne.

Do tej pory omówione zostały trzy pierwsze podetapy przestrzennej analizy ekonomicznej (zob. tabela 2.3). Pierwszy podetap stanowi ustalenie granic

⁶ Układ jednostek terytorialnych wraz z odniesionymi do niego danymi przestrzennymi nie posiadającymi własności jednorodności przyczynowej tworzy niejednorodny przyczynowo układ danych przestrzennych.

obszaru analizy, którego dotyczyć będzie wykonana analiza. W drugim podetapie następuje wybór układów jednostek terytorialnych wraz z odniesionymi do nich danymi przestrzennymi, które utworzą układy danych przestrzennych. Możliwe jest również uzyskanie dodatkowych układów danych przestrzennych w ramach wykonania agregacji danych (trzeciego podetapu). Należy podkreślić, że zarówno decyzja odnośnie do wyboru obszaru analizy, jak i decyzja związana z ustaleniem układów danych przestrzennych powinny wynikać z postawionego w ramach badania problemu badawczego.

Kolejny czwarty podetap przestrzennej analizy ekonomicznej dotyczy ustalenia jednorodnego zbioru układów (zob. tabela 2.3)⁷. Jednorodny zbiór układów zostanie określony jako zbiór układów jednostek terytorialnych na różnych poziomach agregacji, gdzie odniesione do tych układów dane przestrzenne posiadają własność jednorodności przyczynowej. W wyniku wyznaczenia jednorodnego zbioru układów, badacz ustala, dla każdego z przyjętych poziomów agregacji, jeden układ jednostek terytorialnych wraz z odniesionymi do niego jednorodnymi przyczynowo danymi przestrzennymi. Wynika to z faktu, że w ramach rozpatrywanego problemu badawczego, na wybranym poziomie agregacji istnieje tylko jeden, jednorodny przyczynowo układ danych przestrzennych⁸. Każdy inny układ jednostek terytorialnych na tym samym poziomie agregacji nie powinien być wykorzystywany w badaniach ze względu na brak jednorodności przyczynowej odniesionych do niego danych przestrzennych. Ustalenie jednorodnego zbioru układów w ramach etapu przestrzennej analizy ekonomicznej jest kluczowe, ponieważ pozwala na wyciągnięcie poprawnych wniosków, które wykorzystane zostaną do rozwiązania postawionego problemu badawczego. Ostatecznie należy podsumować, że w ramach postawionego problemu badawczego istnieje tylko jeden jednorodny zbiór układów, który pozwala na identyfikację oraz opis zależności dla analizowanych zjawisk społeczno-ekonomicznych. Określony, w ramach przestrzennej analizy ekonomicznej, jednorodny zbiór układów jest koniecznym warunkiem uzyskania właściwych wyników badań. Należy mieć jednak świadomość, że przyjęte na wybranych poziomach agregacji układy danych przestrzennych nigdy nie będą odzwierciedlały w sposób idealny oddziaływania rzeczywistych splotów przyczyn głównych.

W przypadku jednorodnego zbioru układów można stwierdzić, że wybór układów jednostek terytorialnych jest ograniczony od dołu i z góry, pod względem własności jednorodności przyczynowej danych przestrzennych. Ograniczenie to polega na tym, że układy danych przestrzennych najczęściej nie są jednorodne przyczynowo na bardzo niskim lub na bardzo wysokim poziomie agregacji. Dlatego jednorodny zbiór układów powinien zostać ograniczony tylko

⁷ Pojęcie to rozwinęte zostało na podstawie wstępnych rozważań przedstawionych w pracy Pietrzak (2014a).

⁸ Na układ danych przestrzennych składa się układ jednostek terytorialnych wraz z odniesionymi do niego danymi przestrzennymi.

do jednorodnych przyczynowo układów danych przestrzennych na odpowiednich poziomach agregacji. Ograniczenie z dołu wynika z faktu, że w przypadku zjawisk społeczno-ekonomicznych identyfikacja oddziaływania splotu przyczyn głównych jest możliwa dopiero w ramach całego regionu, który stanowi złożony system ekonomiczny. Wybór najniższego poziomu agregacji w postaci analizy przestrzennych danych punktowych (wybranych konsumentów, przedsiębiorstw) nie pozwoli na ocenę zjawisk społeczno-ekonomicznych, których charakter ujawnia się dopiero w ramach funkcjonowania regionu. Z kolei ograniczenie z góry wynika z faktu, że dla zbyt dużych obszarów dochodzi do przestrzennego oddziaływania kilku splotów przyczyn głównych. Każdy ze splotów przyczyn głównych powstaje na skutek funkcjonowania względnie samodzielnych regionów, znajdujących się w granicach większego obszaru. Oznacza to, że dane, które posiadają na wybranym poziomie agregacji własność jednorodności przyczynowej, mogą utracić tę własność na wyższym poziomie agregacji.

Należy podkreślić fakt, że w przypadku danych przestrzennych ocena ich jednorodności przyczynowej jest bardzo trudna. Właściwie każde powiększenie obszaru analizowanych regionów (przyjęcie wyższego poziomu agregacji) może spowodować zmianę w oddziaływaniu przyczyn głównych. Dodatkowo, istniejące współzależności między regionami powodują, że trudnym jest określenie jakiegokolwiek regionu, jako w pełni samodzielnego systemu. Oznacza to, że ilościowa ocena własności jednorodności przyczynowej danych przestrzennych na etapie tworzenia jednorodnego zbioru układów za pomocą aparatu badawczego Zielińskiego jest trudnym zadaniem (Zieliński, 1991, s. 7–17). W każdym przypadku decyzja co do wyboru układu jednostek terytorialnych oraz ustalenia własności jednorodności przyczynowej dla danych przestrzennych musi zostać podjęta przez badacza w sposób arbitralny, jednak wyłącznie w ramach postawionego problemu badawczego. W związku z tym badacz powinien ustalić jednorodny zbiór układów w oparciu o dostępną wiedzę ekonomiczną, wyniki wcześniejszych badań oraz własne doświadczenie naukowe.

Przykładowo dla większości problemów badawczych poruszanych w ramach przestrzennych badań ekonomicznych dotyczących Polski można przyjąć na potrzeby wyznaczenia jednorodnego zbioru układów sześć potencjalnych układów jednostek terytorialnych zgodnych z klasyfikacją NUTS, od układu NUTS 5 do układu NUTS 0. Wybór ten jest uzasadniony co najmniej z dwóch powodów. Pierwszym, najważniejszym powodem jest fakt, że układy jednostek terytorialnych NUTS nie powstały w sposób przypadkowy. Granice układów zgodnych z klasyfikacją NUTS wynikają z nakładających się na siebie silnych współzależności historycznych, społecznych i ekonomicznych. Odniesione do tych układów jednostek terytorialnych dane przestrzenne o charakterze społeczno-ekonomicznym powinny charakteryzować się własnością jednorodności przyczynowej.

Analiza własności większości zjawisk społeczno-ekonomicznych oraz zależności między nimi, wykonana na podstawie danych przestrzennych odnie-

sionych do układów zgodnych z klasyfikacją NUTS, powinna pozwolić na ich prawidłową identyfikację. Jednak w zależności od podjętego problemu badawczego, badacz powinien każdorazowo dokonać wyboru, które z układów utworzą jednorodny zbiór układów. Należy uwzględnić fakt, że również dla układów jednostek terytorialnych zgodnych z klasyfikacją NUTS może występować ograniczenie od dołu i z góry związane z utratą własności jednorodności przyczynowej danych przestrzennych. W przypadku tych układów ograniczenie takie występuje najczęściej z góry. Dla części zjawisk społeczno-ekonomicznych dane przestrzenne odniesione do układów jednostek terytorialnych NUTS 2 (województwa), NUTS 1 (regiony) lub NUTS 0 (obszar kraju) mogą nie posiadać własności jednorodności przyczynowej. Ograniczenie od dołu dla większości badanych zjawisk występuje w znacznie mniejszym stopniu.

Drugi powód wykorzystania układów jednostek terytorialnych zgodnych z klasyfikacją NUTS w przestrzennych badaniach ekonomicznych, wynika z kwestii możliwości pozyskania wiarygodnych danych przestrzennych, które są udostępniane przez statystykę publiczną dla układów z klasyfikacji NUTS. Należy jednak podkreślić, że na najniższym poziomie agregacji NUTS 5 udostępnianych jest najmniej zbiorów danych przestrzennych. Większość publikowanych w ramach statystyki publicznej danych przestrzennych udostępniana jest dopiero na wysokich poziomach agregacji, począwszy od układu NUTS 2. Może to oznaczać konieczność wykorzystania w przestrzennych analizach ekonomicznych układów danych przestrzennych z poza jednorodnego zbioru układów. Należy jednak pamiętać, że wykorzystane do rozwiązania postawionego problemu badawczego rezultaty badań, mogą być obciążone błędem poznawczym. Ustalenie braku własności jednorodności przyczynowej dla danych przestrzennych, które jako jedyne są dostępne, nie powinno jednak wstrzymywać prowadzonych badań. Wynika to z faktu, że korzyści informacyjne z wykorzystania tego typu danych przestrzennych najczęściej przewyższają znacznie potencjalne zagrożenia związane z niepełną identyfikacją oddziaływania przyczyn głównych.

Piąty podetap przestrzennej analizy ekonomicznej dotyczy identyfikacji wewnętrznej struktury danych przestrzennych (zob. tabela 2.3). Wykonanie tego podetapu związane jest z faktem, że w przypadku zjawisk społeczno-ekonomicznych, dane przestrzenne będące realizacją tych zjawisk, powinny być traktowane jako realizacja dwuwymiarowego pola losowego $X(\mathbf{s})$ ⁹ (Arbia, 1989, s. 43–92; Szulc, 2007, s. 17–65). Oznacza to, że badane na podstawie dostępnych danych przestrzennych, zjawiska społeczno-ekonomiczne należy traktować jako przestrzenne procesy stochastyczne. W związku z tym po ustaleniu jednorodnego zbioru układów następnym podetapem przestrzennej analizy ekonomicznej stanowi zbadanie wewnętrznej struktury wybranych procesów

⁹ Wektor argumentów $\mathbf{s}=[s_1, s_2]$ zawiera współrzędne na płaszczyźnie (zob. Szulc, 2007, s. 51–52). W przypadku argumentów dotyczących lokalizacji w przestrzeni geograficznej, pole losowe nazywane będzie przestrzennym procesem stochastycznym.

przestrzennych (Pietrzak, 2010a, s. 84–90; Pietrzak, 2010b, s. 323–330). Analiza wewnętrznej struktury procesu przestrzennego polega na identyfikacji jego własności deterministycznych i stochastycznych. Ponieważ ekonomiczne procesy przestrzenne są nieekperymentalne, stąd o własnościach tych procesów wnioskuje się na podstawie ich realizacji w postaci danych przestrzennych (Wiśniewski, Zieliński, 1998, s. 7).

W przypadku procesów przestrzennych wyróżnić można następujące składniki ich wewnętrznej struktury: składnik związany z niejednorodnością niesystematyczną, składnik związany z niejednorodnością systematyczną oraz składnik struktury jednorodny przestrzennie. Identyfikacja wewnętrznej struktury procesu przestrzennego dokonywana jest poprzez ustalenie własności procesu, w odniesieniu do kolejnych składników tej struktury. Należy podkreślić, że nie wszystkie wymienione składniki muszą wchodzić w skład wewnętrznej struktury procesu przestrzennego.

Opis wewnętrznej struktury procesów przestrzennych rozpoczęty zostanie od składnika związanego z jednorodnością przestrzenną. Jednorodność przestrzenna rozumiana jest w pracy jako stacjonarność w szerszym sensie procesu przestrzennego. W przypadku zachodzenia tej własności, spełnione są następujące własności dla procesu przestrzennego (zob. Arbia, 2006, s. 43–49; Szulc, 2007, s. 28):

$$E(X(\mathbf{s})) = E(X(\mathbf{s} + \boldsymbol{\delta})) = \mu \quad \forall \mathbf{s}, \boldsymbol{\delta}, \quad (2.1)$$

$$E(X(\mathbf{s}))^2 = E(X(\mathbf{s} + \boldsymbol{\delta}))^2 = \delta^2 \quad \forall \mathbf{s}, \boldsymbol{\delta}, \quad (2.2)$$

$$E[X(\mathbf{s}), X(\mathbf{v})] = \gamma(d) \quad \forall \mathbf{s}, \mathbf{v}, \quad (2.3)$$

gdzie $E(X(\mathbf{s}))$, $E(X(\mathbf{s}))^2$ i $\gamma(d)$ oznaczają odpowiednio funkcje wartości oczekiwanej, wariancji oraz kowariancji procesu przestrzennego, $\mathbf{s} = [s_1, s_2]$, $\mathbf{v} = [v_1, v_2]$ i $\boldsymbol{\delta} = [\delta_1, \delta_2]$ są punktami w przestrzeni, d jest odległością między punktami.

W przypadku własności jednorodności przestrzennej czynione są założenia odnośnie do wartości oczekiwanej oraz funkcji kowariancji procesu przestrzennego, wyrażone w równaniach (2.1)–(2.3). W przypadku tej własności zakłada się, że proces charakteryzują się stałą wartością oczekiwaną w przestrzeni. Oznacza to, że wartość oczekiwana, niezależnie od wyboru punktu $\mathbf{s} = [s_1, s_2]$ w przestrzeni, jest taka sama i przyjmuje wartość μ . Natomiast wartość funkcji kowariancji zależy wyłącznie od odległości d między parą punktów w przestrzeni. W ramach składnika jednorodności przestrzennej, proces przestrzenny posiada stałą wartość oczekiwaną, stałą wariancję oraz może charakteryzować się własnością autokorelacji przestrzennej.

Identyfikacja składnika niejednorodności systematycznej polega na odkryciu własności związanych z systematycznymi zmianami w wartości oczekiwanej, wariancji lub kowariancji. Składnik ten modelować można na przykład za pomocą modeli trendów przestrzennych, czy modeli przełącznikowych (Szulc, 2007, s. 205–236; Suchecki, 2010, s. 255–266). Ostatnim składnikiem wewnętrznej struktury danych jest niejednorodność niesystematyczna. Składnik ten wynika z niemożliwości określenia przez badacza systematycznych zmian w wartości oczekiwanej, wariancji lub funkcji kowariancji. Jediną możliwością objaśnienia zmienności procesu przestrzennego charakteryzującego się własnością niejednorodności niesystematycznej jest wykorzystanie innego procesu przestrzennego, charakteryzującego się podobnym przebiegiem przestrzennym tej własności.

Analiza wewnętrznej struktury rozważanego procesu przestrzennego może doprowadzić do identyfikacji wszystkich trzech składników tej struktury albo wybranej kombinacji tych składników. Badanie kolejnych składników wewnętrznej struktury rozpoczynane jest od identyfikacji niejednorodności systematycznej procesu przestrzennego. W przypadku ustalenia charakteru tej własności, należy dokonać jej eliminacji poprzez transformację procesu przestrzennego za pomocą odpowiedniego przekształcenia. W kolejnym kroku należy przejść do identyfikacji własności jednorodności przestrzennej. Również w przypadku stwierdzenia tej własności należy dokonać jej eliminacji poprzez odpowiednią transformację procesu. W ostatnim kroku należy ustalić, czy badany proces jest białym szumem przestrzennym¹⁰.

Identyfikacja wewnętrznej struktury procesów przestrzennych powinna być podstawą do przeprowadzania kolejnego, szóstego podetapu przestrzennej analizy, którego celem jest ustalenie obszaru lub podobszarów wniosków (zob. tabela 2.3). Etap ten polega na wyznaczeniu granic obszaru, w stosunku do którego wyciągane będą wnioski na podstawie wykonanej analizy. Granice obszaru wniosków powinny być ustalone w taki sposób, żeby przy przesunięciach w przestrzeni nie dochodziło do zmian własności procesu przestrzennego.

Wnioski odniesione do obszaru wniosków, formułowane są na podstawie wyników otrzymanych w ramach zastosowania wybranego narzędzia badawczego. Dlatego siódmy podetap przestrzennej analizy ekonomicznej polega na wyborze narzędzia badawczego (zob. tabela 2.3). Podczas podejmowania decyzji o wyborze narzędzia badawczego należy uwzględnić ustalone przez badacza własności procesu przestrzennego. Brak zapewnienia wymaganych własności badanego procesu przestrzennego w ramach zastosowania wybranego narzędzia badawczego powoduje, że jego wartość poznawcza ulega obniżeniu. Fakt ten jest istotny, ponieważ w wyniku zastosowania narzędzi badawczych otrzymane zostaną rezultaty, które są poprawne jedynie w granicach właściwie dobranego obszaru wniosków. Oznacza to, że wybór właściwego narzędzia ba-

¹⁰ Własności białego szumu przestrzennego przedstawione zostały np. w pracy Szulc (2007, s. 30).

dawczego stanowi kolejny, konieczny warunek otrzymania poprawnych rezultatów dla wykonywanej przestrzennej analizy ekonomicznej.

Wykonanie wszystkich, przedstawionych kolejno podetapów pozwala na przeprowadzenie ostatniego etapu, jakim jest analiza właściwa. Realizacja podetapu analizy właściwej powinna pozwolić na wyciągnięcie wniosków na podstawie otrzymanych wyników, które następnie mogą zostać wykorzystane do rozwiązania postawionego problemu badawczego.

Podetap analizy właściwej polega na zastosowaniu wybranych narzędzi badawczych do układów danych przestrzennych tworzących jednorodny zbiór układów. W wyniku tego otrzymane zostaną rezultaty, które pozwolą na wyciągnięcie wniosków w ramach przyjętego obszaru wniosków. Ustalone przez badacza wnioski powinny pozwolić na przejście do etapu trzeciego badań i podjęcie próby rozwiązania postawionego problemu badawczego (zob. tabela 2.3). Należy podkreślić, że pominięcie wcześniejszych podetapów przestrzennej analizy ekonomicznej może doprowadzić do obniżenia wartości podetapu analizy właściwej poprzez wyciągnięcie niewłaściwych wniosków lub nawet niemożliwość wyciągnięcia wniosków. Jeżeli wnioski, otrzymane na podstawie przestrzennej analizy ekonomicznej, nie pozwolą na rozwiązanie postawionego problemu badawczego, to należy przeformułować etap pierwszy badania lub zaproponować alternatywną przestrzenną analizę ekonomiczną.

2.2. Potencjalne błędy badawcze w przestrzennych badaniach ekonomicznych

2.2.1. Błąd identyfikacji jednorodnego zbioru układów

W pracy zaproponowane zostanie sformułowanie dwóch, kluczowych błędów badawczych, na które narażony jest badacz w ramach pominięcia lub niewłaściwego wykonania etapów przestrzennego badania ekonomicznego (zob. tabela 2.3). Błędy badawcze określone zostały w pracy jako błąd identyfikacji jednorodnego zbioru układów oraz błąd wyznaczenia obszaru wniosków. Wstępne rozważania na temat tych błędów przeprowadzone zostały w pracy Pietrzak (2014a), a inspiracją do podjęcia problemu kluczowych błędów badawczych stanowi praca Toblera (1989). Pomimo, że Tobler nie sformułował wprost pojęć obydwu błędów badawczych, to treść jego pracy jednoznacznie wskazuje na dwa podstawowe problemy, które stanowią źródło powstawania błędów w przestrzennych badaniach ekonomicznych. Są to problem jednorodności przyczynowej danych przestrzennych oraz problem wyboru narzędzi badawczych. Kwestie dotyczące wymienionej problematyki poruszane były również w pracach (Anselin, 1988, s. 27; Hainnig, 2003, s. 150–151).

Pierwszy błąd badawczy, błąd identyfikacji jednorodnego zbioru układów polega na przyporządkowaniu do tego zbioru układu jednostek terytorial-

nych, gdzie odniesione do niego dane przestrzenne nie posiadają własności jednorodności przyczynowej. Popełnienie tego błędu następuje w sytuacji, gdy przestrzenna analiza ekonomiczna wykonana zostanie na podstawie danych przestrzennych, które nie odzwierciedlają poprawnie oddziaływania przyczyn. Otrzymane wyniki, dotyczące własności procesów oraz zależności między nimi, będą odzwierciedleniem różnych splotów przyczyn. W wyniku popełnienia błędu identyfikacji jednorodnego zbioru układów, wartość poznawcza otrzymanych wyników będzie zawsze obniżona. Należy jednak podkreślić, że popełnienie błędu identyfikacji jednorodnego zbioru układów nie przekreśla otrzymanych wyników badań, ponieważ wartość poznawcza wyników uzależniona jest od tego, w jakim stopniu przyjęty w badaniu układ danych przestrzennych różni się od jednorodnego przyczynowo układu danych przestrzennych, który powinien zostać wykorzystany w badaniu.

Należy podkreślić, że zgodnie z omówionymi podetapami przestrzennej analizy ekonomicznej, właściwa analiza powinna zostać wykonana tylko na podstawie danych odniesionych do jednorodnego zbioru układów, ustalonego w ramach postawionego problemu badawczego. W związku z tym jednym z kluczowych kroków w ramach prowadzonych przestrzennych badań ekonomicznych powinno być ustalenie jednorodnego zbioru układów. Czynność ta pozwala bowiem na uniknięcie błędu identyfikacji jednorodnego zbioru układów, ponieważ na wybranym poziomie agregacji badacz ma do dyspozycji ustalony układ jednostek terytorialnych oraz odniesione do niego dane przestrzenne, które charakteryzują się własnością jednorodności przyczynowej.

W związku z możliwością popełnienia błędu identyfikacji jednorodnego zbioru układów wyróżnić należy trzy sytuacje niewłaściwego wyboru układu jednostek terytorialnych. Sytuacja pierwsza polega na tym, że układ jednostek terytorialnych, do którego przyporządkowane dane przestrzenne charakteryzowałyby się własnością jednorodności przestrzennej, nie został nigdy wcześniej wyznaczony. Oznacza to konieczność samodzielnego wyznaczenia właściwego układu jednostek terytorialnych przez badacza. Należy podkreślić, że istnieje wiele problemów badawczych, gdzie konieczna jest ponowna delimitacja obszarów i wyznaczenie nowych układów jednostek terytorialnych, które następnie badacz może wykorzystać do ustalenia jednorodnego zbioru układów. Do problematyki tej zaliczyć można między innymi badania dotyczące rozwoju rolnictwa, wymiany handlowej, rozwoju obszarów metropolitalnych, czy rozwoju obszarów o szczególnych walorach turystycznych. W takiej sytuacji badacz musi samodzielnie wyznaczyć układ jednostek terytorialnych oraz zgromadzić i przyporządkować do niego dane przestrzenne. Jest to konieczne, ponieważ wykorzystanie istniejących układów (np. układów z klasyfikacji NUTS) prowadziłoby do błędnych wniosków ze względu na inaczej kształtującą się przyczynowość związaną z rozwojem wybranych zjawisk społeczno-ekonomicznych. W przypadku ustalenia granic nowego układu jednostek terytorialnych szczególnego znaczenia nabiera agregacja danych przestrzennych,

ponieważ tylko na podstawie procesu agregacji możliwe jest otrzymanie zagregowanych danych dla nowego układu.

Sytuacja pierwsza przedstawiona zostanie w świetle problemu badawczego, jakim była konieczność wyznaczenia nowego układu makroregionów rolniczych SGM¹¹ w Polsce. W ramach podjętego problemu badawczego zakłada się, że wyznaczone makroregiony rolnicze powinny wyraźnie różnić się pod względem podstawowych zmiennych dotyczących rolnictwa. Dodatkowo makroregiony SGM powinny być homogeniczne wewnętrznie, pod względem rozwoju i kultury rolnictwa. Przed 2000 rokiem w Polsce obowiązywał podział na cztery makroregiony SGM. Konieczność wyznaczenia nowego układu makroregionów rolniczych SGM związana była z przygotowaniem się Polski do członkostwa w Unii Europejskiej (Skarżyńska, Goraj, Ziętek, 2005, s. 7, 12, 16). W 2000 roku przyjęta została *Ustawa o zbieraniu i wykorzystywaniu danych rachunkowych gospodarstw rolnych* (2000). Przyszła akcesja nakładała na Polskę obowiązek utworzenia unijnego systemu zbierania danych rachunkowych z gospodarstw rolnych (z ang. *Farm Accountancy Data Network, FADN*). W ramach systemu FADN w Polsce mierzona jest wielkość ekonomiczna gospodarstw rolnych (ich dochodowość), która wyrażana jest w postaci Europejskiej Jednostki Wielkości (ESU) (Skarżyńska, Goraj, Ziętek, 2005, s. 7, 69–70). Wielkość ekonomiczna gospodarstw rolnych oraz ich przestrzenne zróżnicowanie stanowią istotny element racjonalnego gospodarowania w rolnictwie. Gospodarstwa rolne o małej powierzchni nie są w stanie wygenerować odpowiedniego poziomu dochodu, aby sprawnie funkcjonować. Zbyt duże rozdrobnienie gospodarstw oznacza brak możliwości postępu i zwiększenie jednostkowych kosztów produkcji (Michna, 2007, s. 5–13). Wartość ESU dla indywidualnych gospodarstw powinna być liczona w oparciu o standardową produkcję, której poziom zależy od przynależności gospodarstwa do jednego z makroregionów rolniczych SGM (Skarżyńska, Goraj, Ziętek, 2005, s. 69–76).

Makroregiony rolnicze SGM powinny wyraźnie różnić się między sobą rozmiarami produkcji rolniczej oraz czynnikami, które mają decydujący wpływ na efekty produkcyjne uzyskiwane przez gospodarstwa. Punktem wyjścia dla wyznaczenia granic makroregionów był już istniejący układ czterech makroregionów SGM (Skarżyńska, Goraj, Ziętek, 2005, s. 10–19). W celu zapewnienia jednorodności makroregionów pod względem rozwoju i kultury rolnictwa dokonano klasyfikacji województw (NUTS 2) biorąc pod uwagę stopień ich podobieństwa pod względem istotnych cech dotyczących rolnictwa. W celu wyodrębnienia makroregionów wykonano analizę skupień w odniesieniu do województw, gdzie przyjęto dziewięć zmiennych diagnostycznych. Przyjęte zostały następujące zmienne: wartość Produktu Krajowego Brutto na miesz-

¹¹ Współczynniki Standardowych Nadwyżek Bezpośrednich (współczynniki SGM) stosowano do wyznaczania wartości ESU (Europejska Jednostka Wielkości) dla indywidualnych gospodarstw w Polsce. Wyznaczony na potrzeby ustalenia wartości współczynników SGM układ jednostek terytorialnych określany był jako układ makroregionów rolniczych SGM (zob. Skarżyńska, Goraj, Ziętek, 2005, s. 7–11).

kańca w 2001 roku, udział rolnictwa w tworzeniu Produktu Krajowego Brutto w 2001 roku, udział trwałych użytków zielonych w użytkach rolnych w 2001 roku, średni plon zbóż z lat 1999–2001, zużycie nawozów NPK na 1 ha użytków rolnych w 2001 roku, średnia roczna produkcja mleka od 1 krowy z lat 1999–2001, średnia powierzchnia użytków rolnych indywidualnego gospodarstwa rolnego w 1996 roku, średnia liczba krów dojnych w indywidualnych gospodarstwach rolnych w latach 1999–2001 (stan na koniec roku), średnia liczba trzody ogółem w indywidualnych gospodarstwach rolnych w latach 1999–2001 (stan na koniec roku) (Skarżyńska, Goraj, Ziętek, 2005, s. 13). Ostatecznie w wyniku przeprowadzonej analizy taksonomicznej zmieniono granice makroregionów SGM. Województwo lubuskie przesunięte zostało do makroregionu Pomorze i Mazury, a województwo świętokrzyskie do makroregionu Małopolska i Pogórze. Granice makroregionów Wielkopolska i Śląsk oraz Mazowsze i Podlasie pozostały bez zmian. W ten sposób układ makroregionów dostosowany został do aktualnej sytuacji polskiego rolnictwa, która od 1989 do 2000 roku uległa zmianie. Jest to ważne spostrzeżenie, ponieważ wynika z niego, że w ramach postawionego problemu badawczego, granice układów jednostek terytorialnych¹² mogą zmieniać się w czasie. Korekta granic układu jednostek terytorialnych będzie konieczna, jeżeli nastąpi zmiana w przestrzennym zróżnicowaniu rozpatrywanych zjawisk oraz zależności między nimi. Jednak w wybranym momencie czasu na ustalonym poziomie agregacji poprawny jest tylko jeden układ jednostek terytorialnych.

Podział terytorium Polski na makroregiony rolnicze SGM przedstawiony został na rysunku 2.1. W skład nowego układu jednostek terytorialnych wchodzi Region Pomorze i Mazury (województwa: lubuskie, pomorskie, warmińsko-mazurskie i zachodniopomorskie), Region Wielkopolska i Śląsk (województwa: dolnośląskie, kujawsko-pomorskie, opolskie i wielkopolskie), Region Mazowsze i Podlasie (województwa: lubelskie, łódzkie, mazowieckie i podlaskie), Region Małopolska i Pogórze (województwa: małopolskie, podkarpackie, śląskie i świętokrzyskie). Wszystkie makroregiony rolnicze SGM są względnie jednorodne pod względem poziomu rozwoju i kultury rolnictwa (Skarżyńska, Goraj, Ziętek, 2005, s. 12–19).

Utworzenie nowego układu makroregionów SGM w związku z przystąpieniem Polski do Unii Europejskiej jest przykładem utworzenia nowego układu jednostek terytorialnych. Układ ten może zostać wykorzystany do ustalenia jednorodnego zbioru układów w ramach badań dotyczących rolnictwa w Polsce. Zarówno wykonywane przestrzenne analizy ekonomiczne, jak i prowadzona statystyka regionów na podstawie makroregionów SGM powinny prowadzić do poprawnych wyników.

¹² Układów jednostek terytorialnych, które przyporządkowane zostały do jednorodnego zbioru układów.

Rysunek 2.1.

Podział terytorium Polski na makroregiony rolnicze SGM



Źródło: opracowanie własne.

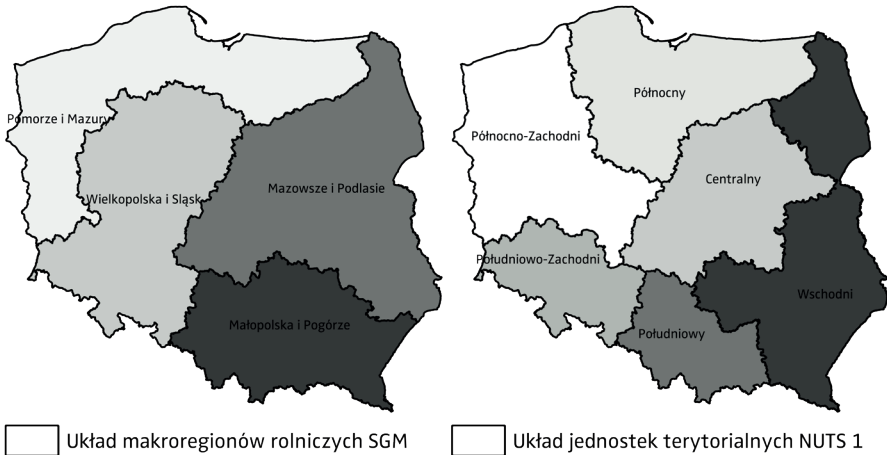
Druga sytuacja dotycząca możliwości popełnienia błędu identyfikacji jednorodnego zbioru układów polega na przyporządkowaniu do jednorodnego zbioru układów istniejącego układu jednostek terytorialnych, gdzie odniesione do niego dane przestrzenne nie mają własności jednorodności przyczynowej. W tym przypadku błąd polega na wyborze niewłaściwego układu jednostek terytorialnych ze zbioru dostępnych układów. Rozwiązaniem w takiej sytuacji jest odrzucenie przyjętego układu jednostek terytorialnych i dokonanie przebudowy jednorodnego zbioru układów poprzez wybór właściwego układu.

Sytuacja druga przedstawiona zostanie również na przykładzie analizy zjawisk zachodzących w polskim rolnictwie. W przypadku badań dotyczących problemów rolnictwa w Polsce, może wystąpić potrzeba poszerzenia jednorodnego zbioru układów o układ jednostek terytorialnych, gdzie jednostki układu miałyby charakter makroregionów. Na takim poziomie agregacji wyróżnić można co najmniej dwa istniejące układy makroregionów, układ NUTS 1 oraz układ makroregionów SGM. Jak zostało wcześniej stwierdzone większość danych przestrzennych dotyczących procesów w rolnictwie i przyporządkowanych do układu makroregionów rolniczych SGM powinno charakteryzować się własnością jednorodności przyczynowej. Możliwe jest jednak przyjęcie w badaniach innego układu jednostek terytorialnych na tym poziomie agregacji, na przykład układu regionów (NUTS 1). Wybór tego układu jednostek terytorialnych

doprowadzi do popełnienia błędu identyfikacji jednorodnego zbioru układów. Wynika to z faktu, że przyporządkowane do układu NUTS 1 dane przestrzenne dotyczące rolnictwa nie utworzą jednorodnego układu danych przestrzennych.

Rysunek 2.2.

Porównanie układu SGM z układem NUTS 1



Źródło: opracowanie własne.

Na rysunku 2.2 zaprezentowane zostały dwa układy jednostek terytorialnych, układ NUTS 1 oraz układ makroregionów rolniczych SGM. Na przedstawionym rysunku widoczne są różnice w kształtowaniu się granic regionów dla obydwu układów. Dodatkowo rozważenie wszystkich regionów pod względem poziomu i kultury rolnictwa pozwala na stwierdzenie, że dla układu jednostek terytorialnych NUTS 1 w regionie wschodnim województwo lubelskie odróżnia się od pozostałych województw, w regionie północno-zachodnim od pozostałych odróżnia się województwo wielkopolskie, a w regionie północnym województwo kujawsko-pomorskie. W związku z powyższym należy stwierdzić, że wykorzystanie układu NUTS 1 do analizy zagadnień rolniczych prowadzi będzie do niepoprawnych wyników. Oznacza to, że układ NUTS 1 należy odrzucić i przyjąć inny układ jednostek terytorialnych. Jak zostało wcześniej stwierdzone, w przypadku analizy zjawisk zachodzących w polskim rolnictwie poprawnym układem jednostek terytorialnych, który można przyporządkować do jednorodnego zbioru układów, jest układ makroregionów rolniczych SGM.

W przypadku trzeciej sytuacji do jednorodnego zbioru układów przyporządkowano układ danych przestrzennych na zbyt niskim albo zbyt wysokim poziomie agregacji, co powoduje brak jednorodności przyczynowej danych przestrzennych. Biorąc pod uwagę specyfikę zjawisk społeczno-ekonomicznych, należy stwierdzić, że w przypadku większości z nich nie ma możliwości

ustalenia jednorodnego przyczynowo układu danych przestrzennych na zbyt niskim albo zbyt wysokim poziomie agregacji. Rozwiązaniem tej sytuacji i jednocześnie sposobem na uniknięcie błędu identyfikacji jednorodnego zbioru układów jest podwyższenie lub obniżenie wybranego poziomu agregacji w ramach wykonywanej analizy do takiego, w którym możliwe jest ustalenie jednorodnego przyczynowo układu danych przestrzennych.

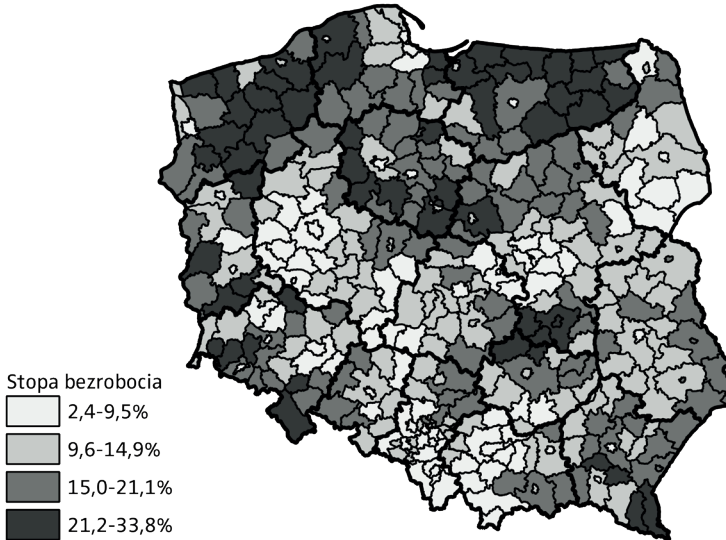
Przykład z wyborem zbyt niskiego poziomu agregacji przedstawiony zostanie na podstawie wyników zaprezentowanych w pracy Robinson, 1950, gdzie przedmiotem badania była analiza poziomu analfabetyzmu w Wielkiej Brytanii w zależności od pochodzenia etnicznego. Poruszony w pracy problem analfabetyzmu jest bardzo istotny w ramach prowadzonej przez państwo polityki edukacyjnej oraz społecznej. Problem ten związany jest głównie z hermetycznym funkcjonowaniem wybranych społeczności tworzonych przez grupy obcego pochodzenia etnicznego na terenie wybranych obszarów zamieszkania (np. dzielnica miasta). Uniemożliwia to proces asymilacji kulturowej z całym społeczeństwem w ramach większego obszaru (np. teren miasta). Oznacza to, że analiza takiego zjawiska jak analfabetyzm powinna być wykonana w odniesieniu do jednostek terytorialnych, w granicach których dochodzi do tworzenia się tego zjawiska. W przypadku zjawiska analfabetyzmu, należałoby przyjąć układy jednostek terytorialnych na niskim poziomie agregacji i dla tak przyjętego układu badać zjawisko analfabetyzmu, które w zależności od udziału hermetycznych społeczności w poszczególnych jednostkach administracyjnych będzie miało różne nasilenie. Zaobserwowana zostanie dodatnia zależność pomiędzy poziomem analfabetyzmu wśród osób obcego pochodzenia etnicznego a udziałem społeczności hermetycznych zamieszkujących danych region. Tymczasem w pracy Robinson (1950) wskazuje się na konieczność analizy zjawiska analfabetyzmu wyłącznie na podstawie analizy indywidualnych osób. Przy tak niskim poziomie agregacji nie uwzględnia się w żadnym stopniu przyczyn wynikających z przynależności tych osób do obszarów zdominowanych przez grupy obcego pochodzenia etnicznego, które oddziałują na pojedyncze osoby i powodują, że część z nich nie osiągnie minimalnego poziomu szkolnej edukacji.

Z kolei przykład popełnienia błędu identyfikacji jednorodnego zbioru układów, wynikający z przyjęcia układu jednostek terytorialnych na zbyt wysokim poziomie agregacji, przedstawiony zostanie w odzwierciedleniu do przestrzennego kształtowania się bezrobocia w Polsce. Do opisu przestrzennego zróżnicowania zjawiska bezrobocia wykorzystane zostaną dane przestrzenne dotyczące stopy bezrobocia rejestrowanego w Polsce. Dane przestrzenne dotyczące stopy bezrobocia rejestrowanego udostępniane są przez Główny Urząd Statystyczny dla układu jednostek terytorialnych NUTS 4 oraz na wyższych poziomach agregacji. Autor uważa, że w przypadku dwóch układów jednostek terytorialnych: układu NUTS 4 oraz układu NUTS 3, można założyć własność jednorodności przyczynowej danych przestrzennych odzwierciedlających poziom bezrobocia

w Polsce¹³. Natomiast dane przestrzenne odniesione do układów jednostek terytorialnych na wyższym poziomie agregacji (układ NUTS 2, układ NUTS 1), nie posiadają własności jednorodności przyczynowej.

Rysunek 2.3.

Przestrzenne zróżnicowanie stopy bezrobocia rejestrowanego



Źródło: opracowanie własne.

Na rysunku 2.3 przedstawiono kształtowanie się zjawiska bezrobocia w Polsce w 2007 roku dla układu jednostek terytorialnych NUTS 4 oraz dodatkowo naniesione zostały granice województw (układ NUTS 2). Wartości stopy bezrobocia przyporządkowane zostały do czterech klas przy zastosowaniu metody podziału naturalnego¹⁴. Na rysunku 2.3 widoczna jest własność niejednorodności przyczynowej danych przestrzennych przedstawiających poziom stopy bezrobocia dla układu NUTS 2. Na przykład dla województwa mazowieckiego powiaty wokół Warszawy charakteryzują się bardzo niskim poziomem stopy bezrobocia i przyporządkowane zostały do klasy 1. Natomiast powiaty z północnej i południowej części województwa mazowieckiego charakteryzują się znacznie wyższym poziomem stopy bezrobocia i przyporządkowane zostały

¹³ W tym przypadku przyjęcie własności jednorodności przestrzennej dla danych przyporządkowanych do układu jednostek terytorialnych NUTS 3 (układ podregionów) można uznać za dyskusyjne.

¹⁴ Idea metody podziału naturalnego polega na przyporządkowaniu obiektów do klas (grup) w taki sposób, żeby wartość wariancji wewnątrzgrupowej była minimalna a wartość wariancji międzygrupowej była maksymalna (zob. Jenks, 1967).

do klasy 3 i klasy 4. Podobną sytuację można zaobserwować w przypadku innych województw.

Liczne badania wskazują na fakt, że zjawisko bezrobocia charakteryzuje się silnymi, dodatnimi zależnościami przestrzennymi (np. Szulc, 2007, s. 205–239). Oznacza to, że najczęściej dochodzi do tworzenia się osobnych skupień regionów o słabej sytuacji społeczno-gospodarczej oraz skupień o wysokim poziomie rozwoju społeczno-gospodarczego. Regiony o słabej sytuacji społeczno-gospodarczej charakteryzują się wysokim poziomem stopy bezrobocia, w przeciwieństwie do regionów o dobrej sytuacji. W związku z tym w przedstawionym przykładzie przyjęte zostało założenie, w którym zjawisko bezrobocia w Polsce charakteryzuje się własnością dodatnich autozależności przestrzennych. Założenie to jest dużym uproszczeniem przestrzennego kształtowania się zjawiska bezrobocia, jednak uzasadnia wykonanie testowania występowania autokorelacji przestrzennej za pomocą testu statystycznego Morana (Moran, 1950, s. 17–23). W związku z tym wykonany został test Morana na podstawie danych dotyczących kształtowania się stopy bezrobocia rejestrowanego dla układu NUTS 4 przy założeniu, że dane posiadają własność jednorodności przyczynowej. Poprawność przyjętego założenia powinna potwierdzić identyfikacja dodatniej autokorelacji przestrzennej. Uzyskane wyniki testu Morana przedstawione zostały w tabeli 2.4.

Tabela 2.4.

Wyniki testu Morana dla lat 2004–2007

Statystyka	2004	2005	2006	2007
Układ NUTS 4 (powiaty)				
Statystyka Morana I	0,678	0,561	0,521	0,515
Wartość p	~0,000	~0,000	~0,000	~0,000
Układ NUTS 2 (województwa)				
Statystyka Morana I	0,004	-0,051	-0,139	-0,215
Wartość p	0,317	0,545	0,313	0,159

Źródło: opracowanie własne.

Obliczone wartości statystyk testu wskazują na występowanie silnych, dodatnich autozależności przestrzennych dla zjawiska bezrobocia w Polsce na poziomie powiatów. Następnie ponownie wykonany został test Morana, jednak na podstawie danych przestrzennych przyporządkowanych do układu jednostek terytorialnych NUTS 2. W tym przypadku, zgodnie z wizualną oceną przestrzennego zróżnicowania bezrobocia, założony zostanie brak własności jednorodności przyczynowej dla tak określonego układu danych przestrzennych (zob. rysunek 2.3). Słuszność przyjętego założenia powinna zostać

potwierdzona przez otrzymane wyniki dotyczące testowania własności autokorelacji przestrzennej na poziomie województw.

Obliczone wartości statystyki testu przedstawione zostały również w tabeli 2.4, gdzie dla wszystkich lat należy podjąć decyzję o braku podstaw do odrzucenia hipotezy zerowej, mówiącej o braku autokorelacji przestrzennej¹⁵. Otrzymany wynik jest niewłaściwy, ponieważ w układzie województw, analizowane dane przestrzenne odzwierciedlają różne sploty przyczyn. Oznacza to, że przyjęty poziom agregacji województw jest zbyt wysoki i wyznaczony dla niego układ danych przestrzennych nie może być przyporządkowany do jednorodnego zbioru układów, ponieważ nie pozwala na przeprowadzenie poprawnych analiz.

2.2.2. Błąd wyznaczenia obszaru wniosków

W pracach Tobler (1989, s. 115–116), Hainnig (2003, s. 150–151) zwrócono uwagę na kwestię niewłaściwego wykorzystania narzędzi badawczych w badaniach przestrzennych. Autor podjął się rozważenia tego problemu w niniejszej pracy, jednak problem zastosowania niewłaściwych narzędzi badawczych potraktowany został w inny sposób. Główna uwaga została skupiona na własnościach analizowanych procesów przestrzennych. Oznacza to, że ciężar badania położony jest na konieczność zbadania własności procesów przestrzennych w granicach ustalonego obszaru wniosków, a dobór narzędzia badawczego wynika z ustalonych własności. Dla każdego narzędzia badawczego zakłada się bowiem występowanie określonych własności procesów przestrzennych. W przypadku nie spełniania wymaganych własności procesów w granicach ustalonego obszaru wniosków, zastosowane narzędzie badawcze traci swoją moc poznawczą. Oznacza to, że sformułowane na podstawie etapu analizy właściwej wnioski będą obarczone błędem poznawczym. W takiej sytuacji na uniknięcie błędu wyznaczenia obszaru wniosków powinno pozwolić zmniejszenie granic obszaru wniosków w celu uzyskania odpowiednich własności procesów przestrzennych lub przyjęciu innych, odpowiednich narzędzi badawczych.

W związku z tym autor określił drugi błąd badawczy jako błąd wyznaczenia obszaru wniosków. Błąd ten polega na niewłaściwym powiązaniu przyjętego narzędzia badawczego oraz obszaru wniosków¹⁶. W wyniku tego analizowane procesy nie posiadają wymaganych własności dla stosowanego narzędzia badawczego, co ma wpływ na mniejszą moc poznawczą wybranego narzędzia. Uzyskane w ten sposób wnioski wyciągane są w stosunku do ustalonego obszaru wniosków, co również ma negatywny wpływ na jakość wykonywanej analizy. Określony w ten sposób błąd wyznaczenia obszaru wniosków doty-

¹⁵ Również w tym wypadku powinny zostać zidentyfikowane dodatnie autozależności przestrzenne dla zjawiska bezrobocia.

¹⁶ Najczęściej zachodzi sytuacja przyjęcia obszaru wniosków o zbyt dużej powierzchni (na przykład obszar Polski).

czy w głównej mierze ustalenia właściwych granic obszaru (podobszarów) wniosków, ponieważ zmiana wybranego narzędzia badawczego jest często niemożliwa.

Wykorzystywane w badaniach ekonomicznych dane przestrzenne powinny być odniesione do układów jednostek terytorialnych tworzących jednorodny zbiór układów. Jednak wykorzystanie w analizie układu jednostek terytorialnych należącego do jednorodnego zbioru układów jest warunkiem koniecznym, ale nie wystarczającym do sformułowania poprawnych wniosków z analizy. Kolejnym warunkiem jest odpowiedni wybór obszaru wniosków oraz narzędzia badawczego, co wiąże się właśnie z niebezpieczeństwem popełnienia błędu wyznaczenia obszaru wniosków. W badaniach ekonomicznych przyjmowany jest najczęściej maksymalny, możliwy rozmiar obszaru wniosków, co znacznie zwiększa prawdopodobieństwo wystąpienia tego błędu badawczego. W przypadku układów jednostek terytorialnych NUTS dla Polski, będzie to obszar całego kraju. W celu uniknięcia błędu wyznaczenia obszaru wniosków analizowane procesy przestrzenne powinny charakteryzować się w granicach obszaru wniosków takimi własnościami, które są wymagane do zastosowania wybranego narzędzia badawczego.

Do stosowanych najczęściej narzędzi badawczych należy wykorzystanie statystyk opisowych do określenia parametrów rozkładów analizowanych procesów, czy wykorzystanie modelu regresji do identyfikacji zależności ekonomicznych zachodzących między procesami. W obydwu przypadkach wymagana jest własność jednorodności przestrzennej dla analizowanych procesów¹⁷. Odnosząc to do badań wykonywanych dla obszaru całej Polski należy podkreślić, że analizowane procesy przestrzenne (w tym proces objaśniany) posiadają w większości złożoną strukturę wewnętrzną (Pietrzak 2010a), gdzie jednorodność przestrzenna jest jedną z własności tej struktury¹⁸. W takiej sytuacji, w wyniku zastosowania wymienionych narzędzi badawczych popełniany jest błąd wyznaczenia obszaru wniosków. W przypadku identyfikacji niejednorodności systematycznej zamiast modelu regresji liniowej powinna zostać zaproponowana rozszerzona specyfikacja modelu pozwalająca na opis owej własności wewnętrznej struktury, co uchroni badacza przed popełnieniem tego błędu badawczego. W tym celu wykorzystywane są najczęściej modele trendu przestrzennego, czy modele przełącznikowe (Szulc, 2007, s. 205–236; Suchecki, 2010, s. 255–266).

W celu zobrazowania sytuacji, w której może zostać popełniony błąd wyznaczenia obszaru wniosków przedstawiony zostanie przykład wyznaczenia war-

¹⁷ W przypadku wyznaczania wartości statystyk opisowych wymagany jest stały poziom wartości oczekiwanej i wariancji analizowanego procesu oraz brak autokorelacji przestrzennej. W przypadku modelu regresji, identyfikacja autokorelacji przestrzennej dla procesu objaśnianego wymaga rozszerzenia modelu o autoregresję przestrzenną.

¹⁸ Własność jednorodności przestrzennej może być w takiej sytuacji zidentyfikowana dopiero po wcześniejszej eliminacji własności niejednorodności systematycznej.

tości średniej dowolnego procesu przestrzennego dla hipotetycznego obszaru wniosków. Obszar wniosków ustalony zostanie jako obszar całej Polski, zgodny z układem NUTS 0. W takim wypadku badany proces przestrzenny powinien charakteryzować się w granicach obszaru wniosków własnością jednorodności przestrzennej. W ramach własności jednorodności przestrzennej należy założyć, że proces przestrzenny posiada stały poziom wartości oczekiwanej i wariancji na tym obszarze oraz charakteryzuje się brakiem autokorelacji przestrzennej. Możemy jednak założyć, że proces przestrzenny w ramach wyznaczonego obszaru wniosków charakteryzuje się własnością niejednorodności systematycznej. W takim przypadku obserwowane są systematyczne zmiany wartości średniej w ramach obszaru wniosków. W związku z powyższym założeniem obszar Polski można hipotetycznie podzielić na dwa podobszary o podobnych rozmiarach, na przykład podobszar południowy oraz podobszar północny. Dla każdego podobszaru badany proces będzie charakteryzował się jednorodnością przestrzenną w postaci stałego poziomu wartości średniej. Natomiast zgodnie z przyjętym założeniem, w ramach granic całego kraju proces ten charakteryzował się będzie niejednorodnością systematyczną. Oznacza to, że będzie występować istotna różnica między wartościami średnimi wyznaczonymi na podstawie danych dla obydwu podobszarów. Sytuacja taka może wynikać na przykład z zaszczości historycznych, czy różnic w poziomie rozwoju społeczno-ekonomicznego regionów, w wyniku których może występować znaczny wzrost lub spadek średniej wartości badanego procesu w ramach jednego z podobszarów.

W celu wyznaczenia wartości średniej badanego procesu należy przyjąć układ jednostek terytorialnych przyporządkowany do jednorodnego zbioru układów. Przykładowy układ danych przestrzennych może być utworzony na podstawie układu jednostek terytorialnych NUTS 4 oraz odniesionych do niego jednorodnych przyczynowo danych przestrzennych, będących realizacją badanego procesu. Ustalenie dwóch podobszarów wniosków (podobszaru południowego oraz podobszaru północnego) i wyznaczenie dla nich wartości średniej pozwoli na otrzymanie poprawnych wyników. Natomiast wyznaczona wartość średnia dla obszaru całego kraju będzie obciążona pewnym błędem, ponieważ wymagane własności dla zastosowanego narzędzia badawczego nie są spełnione. Wybrane narzędzie zostanie zastosowane do zbyt dużego obszaru i moc poznawcza narzędzia zostanie ograniczona. W tym przypadku wyznaczona wartość średnia dla obszaru całego kraju będzie ważoną średnią dla obydwu podobszarów. W konsekwencji, przyjmując za obszar wniosków obszar całego kraju i wyznaczając dla tego obszaru wartość średnią procesu, badacz popełnia błąd wyznaczenia obszaru wniosków.

2.3. Identyfikacja mikrozależności i makrozależności w badaniach ekonomicznych

2.3.1. Mikrozależności a makrozależności ekonomiczne

W kontekście wykonywanych badań ekonomicznych Pawłowski zwrócił uwagę na fakt, że zwyczajowo przyjęło się używanie terminu makroekonomiczny w analizie jednostek wyższego rzędu, a terminu mikroekonomiczny w przypadku analizy jednostek niższego rzędu (Pawłowski, 1969 s. 239). Dodatkowo Pawłowski dokonuje podziału prawidłowości (relacji) ekonomicznych oraz parametrów strukturalnych tych relacji z perspektywy makro i mikroekonomicznej. Prawidłowości dotyczące jednostek wyższego rzędu określone zostały jako makrorelacje, a odpowiadające im parametry jako makroparametry. Natomiast mikrorelacje oraz mikroparametry określone zostały w odniesieniu do obiektów niższego rzędu. Ze względu na stopień złożoności procesu agregacji (np. agregacja dwu, trzystopniowa) Pawłowski wskazuje na względność terminów „mikroekonomiczny” oraz „makroekonomiczny” (Pawłowski, 1969, s. 239–240). Względność ta polega na tym, że w zależności od wybranego poziomu agregacji relacje makroekonomiczne mogą stać się relacjami mikroekonomicznymi przy przejściu na wyższy poziom agregacji¹⁹. W takiej sytuacji dla uniknięcia istniejącej względności tych pojęć, część badaczy stosuje termin makroekonomiczny wyłącznie do badań dotyczących gospodarki całego kraju, a termin mikroekonomiczny wyłącznie dla badań dotyczących podstawowych obiektów ekonomicznych (Pawłowski, 1969, s. 240, Bołt, Krauze, Kulawczuk, 1985, s. 33–34).

Wskazane wcześniej, jedno z głównych zagadnień teorii agregacji, polegające na ustaleniu relacji między makroparametrami a mikroparametrami Pawłowski opisuje na podstawie modelu popytu na mięso Y_t względem średniej ceny realnej mięsa X_t (Pawłowski, 1969, s. 241–242). Do oszacowania parametrów modelu wykorzystano dane statystyczne w postaci szeregów czasowych dotyczących zbiorowości populacji całego kraju oraz w rozbiciu na trzy podzbiorowości, ludności utrzymującej się wyłącznie z pracy w rolnictwie, ludności nierolniczej oraz ludności czerpiącej swe dochody częściowo z pracy w rolnictwie oraz częściowo z pracy poza rolnictwem. W planowaniu gospodarczym niezwykle cenna jest analiza makrozależności Y_t względem X_t dla całej gospodarki, w wyniku czego otrzymujemy ocenę makroparametru (Pawłowski, 1969, s. 225–229). Należy jednak zwrócić uwagę na fakt, że dane dotyczące popytu na mięso dla zbiorowości ludności całego kraju nie posiadają własności jednorodności przyczynowej. Wynika to z faktu, że zależność popytu na mięso od ceny kształtuje się w odmienny sposób dla osób różniących się co do źródeł

¹⁹ Podejście to wyróżnia się na tle podejścia przedstawianego w literaturze zachodniej, gdzie granica między mikroekonomią a makroekonomią była traktowana jako niezmienna (Theil, 1965, s. 1–9).

dochodu. W związku z tym wyodrębnienie trzech wymienionych podpopulacji ludności ze względu na kryterium źródła dochodu pozwala na otrzymanie zbiorowości, gdzie dane dotyczące popytu na mięso powinny charakteryzować się własnością jednorodności przyczynowej. W takim przypadku analizowane zależności dla jednej z trzech podpopulacji ludności określić można jako mikro-zależności, a estymowane parametry jako mikroparametry.

Należy podkreślić, że również w przypadku badania wybranej podpopulacji ludności ze względu na kryterium źródła dochodu (np. ludności utrzymującej się wyłącznie z pracy w rolnictwie) może wystąpić problem braku jednorodności przyczynowej dla danych dotyczących popytu na mięso. Istnieje wówczas możliwość wydzielenia kolejnych podpopulacji w ramach zbiorowości ludności utrzymującej się wyłącznie z pracy w rolnictwie, w przypadku których dane będą posiadały własność jednorodności przyczynowej. W takiej sytuacji zależność dla wyjściowej podpopulacji ludności utrzymującej się wyłącznie z pracy w rolnictwie, która była wcześniej rozpatrywana jako mikroekonomiczna należy określić jako makroekonomiczną. Natomiast zależności dla nowo wydzielonych podpopulacji zostaną określone jako mikroekonomiczne. Świadczy to o ogromnej roli badacza, który w zależności od ogólności postawionego problemu badawczego ustala jednostki niższego i wyższego rzędu, identyfikuje własność jednorodności przyczynowej na podstawie posiadanych danych, a następnie ustala czy analizowane prawidłowości traktować jako makrozależności, czy mikrozależności.

W omawianym przykładzie zależności popytu na mięso Y_t względem średniej ceny realnej mięsa X_t dla zbiorowości ludności całego kraju oraz dla trzech podzbiorowości Pawłowski wyprowadza równanie opisujące relację między makroparametrem a mikroparametrami (Pawłowski, 1969, s. 239–244). Ponieważ dla każdej z podpopulacji zależność ulega zmianom, to w wyniku estymacji parametrów modeli regresji otrzymany jest zbiór trzech różnych ocen mikroparametrów. Pawłowski dowodzi, że ocena makroparametru dla całej zbiorowości równa jest ważonej sumie mikroparametrów. Ustalony system wag zależny jest zarówno od udziałów wartości procesu Y_t z poszczególnych podpopulacji w stosunku do całej zbiorowości, jak i od udziałów wartości procesu X_t . Oznacza to, że ocena makroparametru wynika z ocen mikroparametrów oraz struktury całej zbiorowości ze względu na procesy ekonomiczne Y_t i X_t . Analizując szeregi czasowe Pawłowski zwraca uwagę na możliwość zmian mikrozależności dla poszczególnych podpopulacji w czasie, co wpływać będzie na zmiany mikroparametrów. Dodatkowo na skutek zmian strukturalnych w całej zbiorowości, może dojść do zmian wartości wag. Wymienione czynniki wpływają na zmiany makroparametru w czasie. Zmiany te mogą charakteryzować się ustaloną systematycznością, co wskazywałoby na występowanie określonych tendencji w rozwoju badanego zjawiska.

Pawłowski podkreśla również, że terminy makroekonomiczny i mikroekonomiczny stosowane są wyłącznie w ramach agregacji typu przedmiotowego

(Pawłowski, 1969, s. 239). Uwaga ta jest istotna w kontekście agregacji danych przestrzennych, ponieważ jest to agregacja typu przedmiotowego, gdzie za kryterium hierarchiczne obiektów ekonomicznych służy kryterium geograficzne. Dodatkowo obszary składające się na układ danych terytorialnych są najczęściej samodzielnymi regionami, które należy rozważać jako obiekty ekonomiczne²⁰. Regiony te posiadają złożone relacje wewnętrzne, jak i złożone relacje z innymi regionami. W związku z tym w pracy przedyskutowane zostaną terminy makroekonomiczny i mikroekonomiczny również w przypadku przestrzennych badań ekonomicznych, co stanowić będzie rozszerzenie dotychczasowych rozważań o aspekty przestrzenne.

Dyskusja rozpoczęta zostanie od rozszerzenia zagadnienia makroekonomicznych oraz mikroekonomicznych zależności o problematykę przestrzenną. Wskazany przez Pawłowskiego problem ustalenia związków, jakie zachodzą między relacjami mikroekonomicznymi a relacjami makroekonomicznymi (Pawłowski, 1969, s. 235) poruszony zostanie w kontekście analizy zależności przestrzennych. Uwzględnienie aspektu przestrzennego w problemie ustalenia relacji między mikrozależnościami a makrozależnością wskazało na możliwość powiązania tego zagadnienia z problemem skali²¹. Oznacza to, że otrzymane przez autora rezultaty badań nad zagadnieniem mikro i makrozależności przestrzennych wpisywać się będą nie tylko w tematykę teorii agregacji, ale również w tematykę problemu zmiennej jednostki odniesienia.

2.3.2. Analiza mikrozależności oraz makrozależności przestrzennych

Rozważenie problematyki przestrzennej w tematyce makroekonomicznych oraz mikroekonomicznych relacji rozpoczęte zostanie od określenia pojęć mikrozależności oraz makrozależności przestrzennych. W celu rozróżnienia między mikrozależnościami a makrozależnościami przestrzennymi wykorzystane zostanie kryterium oparte na możliwości zmiany charakteru analizo-

²⁰ W różnych aspektach regiony jako obiekty ekonomiczne mogą posiadać nawet silniejszą pozycję niż państwo, w którego granicach się znajdują. Regiony takie, pomimo, że podlegają polityce państwa, prowadzą równocześnie samodzielną politykę samorządową i za jej pomocą mogą wzmacniać politykę państwa lub mogą jej przeciwdziałać. Na przykład w aspekcie rozwijania usług technologii informacyjnych może występować region, gdzie rozwój tych usług jest na najwyższym poziomie w kraju. Biorąc pod uwagę, że polityka państwa może polegać na wzmacnianiu rozwoju usług technologii informacyjnych na terenie całego kraju, to region taki może stać się skutecznym konkurentem dla państwa. Region taki za pomocą własnej polityki samorządowej może ograniczać oddziaływanie państwa i przyczynić się do jeszcze bardziej intensywnego rozwoju usług technologii informacyjnych w jego obszarze, powodując jednocześnie wolniejszy rozwój tych usług w innych regionach.

²¹ Problem skali przedstawiony zostanie szczegółowo w rozdziale trzecim. Jego istotą jest możliwość uzyskania odmiennych wyników badań, w zależności od tego, czy badania te dotyczą jednostek (regionów) niższego rzędu, czy jednostek wyższego rzędu.

wanej zależności dla wybranej grupy jednostek terytorialnych należących do wyznaczonego obszaru wniosków. Zmiana charakteru zależności między procesami rozumiana jest jako istotna²² zmiana jej siły lub zmiana postaci funkcyjnej badanej zależności przy przejściu pomiędzy wybranymi jednostkami terytorialnymi.

Makrozależności przestrzenne określone zostaną jako zależności, podczas wyznaczania których dochodzi do zmiany ich charakteru w ramach przyjętego obszaru wniosków. Oznacza to, że w przypadku identyfikacji makrozależności przestrzennych rośnie prawdopodobieństwo popełnienia błędu wyznaczenia obszaru wniosków, ponieważ badany proces może nie posiadać odpowiednich własności dla zastosowanego narzędzia badawczego. Natomiast w przypadku wyznaczania mikrozaszależności przestrzennych, możliwe jest ustalenie granic obszaru wniosków w taki sposób, by zależności przestrzenne między analizowanymi procesami miały stały charakter dla wszystkich jednostek terytorialnych tworzących obszar wniosków. W takiej sytuacji zidentyfikowaną mikrozaszależność można odnieść do całego obszaru wniosków.

Przedstawiony przez Pawłowskiego sposób ustalenia relacji między makroparametrami a mikroparametrami na przykładzie analizy szeregów czasowych przeniesiony zostanie obecnie do analizy danych przestrzennych. Ponownie rozważona zostanie zależność popytu na mięso *per capita* $Y(\mathbf{s})$, względem średniej ceny realnej mięsa $X(\mathbf{s})$ ²³. Założona relacja analizowana będzie dla podzbiorowości ludności utrzymującej się wyłącznie z pracy w rolnictwie. Przyjętym obszarem analizy będzie całe terytorium Polski. Punktem wyjścia powinno być ustalenie, czy analizowana zależność ma podobny charakter na przestrzeni całego kraju. Wydaje się, że można przyjąć takie założenie tylko w sytuacji, gdy całe terytorium Polski będzie homogeniczne pod względem rozwoju i kultury rolnictwa. W związku z tym zaprezentowana zostanie analiza, której celem będzie wykazanie, czy rozwój i kultura rolnictwa są na podobnym poziomie na terytorium całej Polski. W przypadku negatywnej odpowiedzi wyznaczone zostaną podobszary, w ramach których założyć będzie można podobieństwo kultury rolnictwa. Ustalenie tego typu podobszarów jest ważne, ponieważ oznaczać może występowanie przestrzennych zmian w charakterze zależności popytu na mięso *per capita* względem średniej ceny realnej mięsa w przypadku podzbiorowości ludności utrzymującej się wyłącznie z pracy w rolnictwie. Ustalone granice podobszarów można przyjąć jako granice podobszarów wniosków, gdzie identyfikowane będą mikrozaszależności przestrzenne w ramach wykonywanej analizy ekonomicznej. Na przykład, w przypadku ustalenia dwóch podobszarów, występować będą dwa mikroparametry dla każdego z nich i je-

²² Zmiana rozumiana jest za istotną, jeśli według badacza niesie za sobą ekonomiczne konsekwencje w ramach podjętego problemu badawczego.

²³ Realizacjami ekonomicznych procesów przestrzennych $Y(\mathbf{s})$, $X(\mathbf{s})$ są dane przestrzenne $y(\mathbf{s}_i)$, $x(\mathbf{s}_i)$, gdzie wektor argumentów $\mathbf{s}_i = [s_{1i}, s_{2i}]$ zawiera współrzędne na płaszczyźnie.

den makroparametr odniesiony do terytorium całego kraju. Oznaczać to będzie również potrzebę ustalenia relacji między makroparametrem a dwoma mikroparametrami.

W celu zbadania, czy terytorium Polski jest homogeniczne pod względem rozwoju i kultury rolnictwa, przeprowadzona została przestrzenna analiza zróżnicowania struktury agrarnej²⁴. Struktura agrarna jest najprostszą i zarazem poprawną miarą poziomu rozwoju rolnictwa, ponieważ stanowi ona podstawowy element racjonalnego gospodarowania w rolnictwie (Michna, 2007, s. 5–13). Ponadto, ocena stanu struktury agrarnej jest bardzo ważna z punktu widzenia polityki rolnej, ponieważ determinuje sytuację ekonomiczną rolnictwa. Od wielkości gospodarstw rolnych zależy dochód ich właścicieli oraz wydajność produkcyjna gospodarstwa. W celu określenia przestrzennej zmienności struktury agrarnej w Polsce wykorzystana została analiza koncentracji powierzchni użytków rolnych. Wartość koncentracji zmierzono za pomocą indeksu Giniego²⁵ (Suchecki, 2010, s. 132–142; Ceriani, Verme, 2012, s. 421–426).

Pierwszy krok analizy stanowić będzie wybór wyjściowego układu danych przestrzennych na niższym poziomie agregacji. W celu wyznaczenia wartości indeksu Giniego wykorzystano układ jednostek terytorialnych NUTS 4 (powiaty) i dane przestrzenne dotyczące liczby gospodarstw rolnych oraz powierzchni użytków rolnych²⁶. Układ NUTS 4 jest układem na najniższym poziomie agregacji, dla którego dostępne są wymienione dane przestrzenne. Obliczenie dla każdego z powiatów wartości indeksu Giniego pozwoliło na otrzymanie danych przestrzennych dotyczących koncentracji przestrzennej użytków rolnych. Posiadane dane przestrzenne wraz z układem jednostek terytorialnych NUTS 4 utworzyły układ danych przestrzennych. Założone zostało również, że wyznaczone wartości indeksu Giniego na poziomie powiatów są danymi przestrzennymi, które powinny charakteryzować się własnością jednorodności przyczynowej, a przyjęty układ danych przestrzennych jest jednorodny przy-

²⁴ W przedstawianej wcześniej części pracy dotyczącej wyznaczania układu makroregionów rolniczych SGM wyróżnione zostały zmienne diagnostyczne, których analiza może pozwolić na identyfikację przestrzennej heterogeniczności polskiego rolnictwa (Skarzyńska, Goraj, Ziętek, 2005, s. 13). Analiza struktury agrarnej jest kolejnym, alternatywnym sposobem na zbadanie przestrzennego zróżnicowania rozwoju i kultury rolnictwa w Polsce.

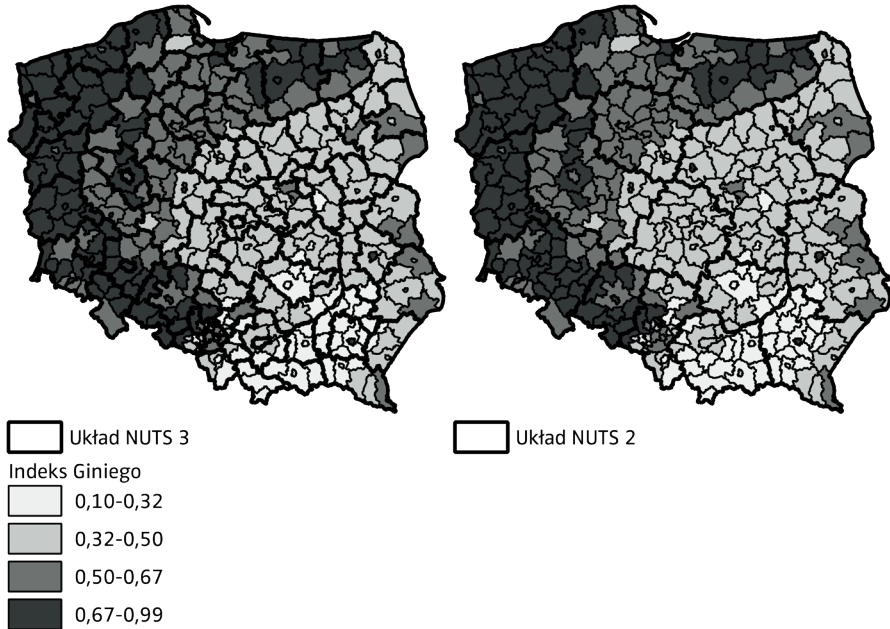
²⁵ Wykorzystane dane pochodzą z Powszechnego Spisu Rolnego przeprowadzonego w 2002 roku. Należy podkreślić, że dane pozwalające na wyznaczenie struktury agrarnej na poziomie powiatów zostały udostępnione przez Główny Urząd Statystyczny wyłącznie dla 2002 roku w związku z publikacją wyników Powszechnego Spisu Rolnego. Dla innych lat, udostępnione przez GUS dane przestrzenne pozwalają na wyznaczenie struktury agrarnej na poziomie województw.

²⁶ Do wyznaczenia wartości indeksu Giniego powierzchnia użytków rolnych policzona została jako całkowita powierzchnia gospodarstw rolnych pomniejszone o grunty w użytkowaniu leśnym oraz nieużytki, gdzie wybrano następujące grupy obszarowe (1–5 ha), (5–10 ha), (10–20 ha), (20–50 ha) oraz (50 ha i więcej).

czynowo. Oznacza to, że dane przestrzenne w postaci wartości indeksu Giniego powinny właściwie odzwierciedlać oddziaływanie przyczyn kształtujących stan struktury agrarnej w ramach przyjętego układu danych przestrzennych.

Rysunek 2.4.

Zróżnicowanie przestrzenne struktury agrarnej w Polsce



Źródło: opracowanie własne.

Otrzymane wartości indeksu Giniego przedstawione zostały na rysunku 2.4, gdzie dokonano podziału powiatów pod względem koncentracji powierzchni użytków rolnych na cztery klasy. Powiaty przyporządkowane zostały do klas na podstawie metody podziału naturalnego. Dodatkowo na rysunku 2.4 przedstawiono układ województw NUTS 2, co pozwoliło na przeprowadzanie dyskusji nad budową jednorodnego zbioru układów.

W związku z otrzymanymi wynikami dokonano oceny przestrzennego zróżnicowania struktury agrarnej, co pozwoliło na ustalenie jednorodnego zbioru układów oraz podobszarów wniosków. W wyniku przeprowadzonej oceny stwierdzono, że w granicach poszczególnych województw dominują powiaty przyporządkowane do jednej z klas. W województwach zachodnio-pomorskim, lubuskim, dolnośląskim i opolskim dominują powiaty z czwartej klasy o najwyższym poziomie koncentracji użytków rolnych. W województwach pomorskim, wielkopolskim, kujawsko-pomorskim oraz warmińsko-mazurskim dominują z kolei powiaty z trzeciej klasy. W województwach mazowieckim, łódzkim,

śląskim, świętokrzyskim, lubelskim i podlaskim większość powiatów należy do drugiej klasy. Natomiast w przypadku województwa małopolskiego oraz podkarpackiego występuje przewaga liczebna powiatów przyporządkowanych do pierwszej klasy o najniższym poziomie koncentracji użytków rolnych. Zaobserwowana dominacja powiatów z jednej, wybranej klasy, świadczy o tym, że na terytorium każdego z województw zachodzi podobieństwo w zakresie struktury agrarnej użytków rolnych. Regiony te (województwa) powinny być homogeniczne wewnątrznie pod względem rozwoju i kultury rolnictwa.

Dodatkowo dokonano oceny zróżnicowania przestrzennego struktury agrarnej dla układu jednostek terytorialnych NUTS 3. W tym przypadku również stwierdzono dla większości podregionów tworzących układ NUTS 3 dominację powiatów z jednej, wybranej klasy. Oznacza to, że dla danych przestrzennych dotyczących koncentracji użytków rolnych można wykorzystać układy jednostek terytorialnych NUTS 3 oraz NUTS 2 w celu utworzenia jednorodnych przyczynowo układów danych przestrzennych. W związku z tym można stwierdzić, że w przypadku przestrzennych analiz ekonomicznych dotyczących struktury agrarnej w Polsce jednorodny zbiór układów tworzą następujące układy jednostek terytorialnych: układ NUTS 4, układ NUTS 3 oraz układ NUTS 2. Tak określony jednorodny zbiór układów powinien być właściwy nie tylko dla badań dotyczących zagadnienia struktury agrarnej, ale również dla badań innych procesów związanych z rozwojem rolnictwa w Polsce.

Wykorzystanie istniejących układów jednostek terytorialnych NUTS 4, NUTS 3 oraz NUTS 2 do wyznaczenia jednorodnego zbioru układów jest dodatkowo uzasadnione udostępnianiem przez służby statystyki publicznej danych przestrzennych dotyczących rolnictwa. Należy jednak zwrócić szczególną uwagę na fakt, że w ramach granic poszczególnych województw można zaobserwować zmiany w charakterze struktury agrarnej. Zmiany te są najbardziej widoczne w przypadku województwa wielkopolskiego oraz województwa kujawsko-pomorskiego. Na wytłumaczenie występujących odchyłeń w charakterze struktury agrarnej w ramach województw może pozwolić porównanie jej przestrzennego zróżnicowania z granicami zaborów w Polsce (zob. rysunek 2.4 oraz rysunek 2.5).

W latach 1772–1795 Rosja, Prusy oraz Austria dokonały zaborów w Polsce. W wyniku zaborów Polska podzielona została następująco: Rosja zagarnęła centrum i wschód Polski, Prusy przejęły Wielkopolskę, Pomorze oraz część Śląska, a Austria południe Polski. Kolejne 123 lata zaborów doprowadziło do wytworzenia na terenach Polski niejednorodnej struktury agrarnej. Po 1918 roku były tereny zaboru pruskiego charakteryzowały się wysokim stopniem koncentracji użytków rolnych, były tereny zaboru rosyjskiego średnim stopniem koncentracji, a były tereny zaboru austriackiego charakteryzowały się niskim stopniem koncentracji użytków rolnych. Po odzyskaniu przez Polskę niepodległości, na terenach byłych zaborów występowały różne czynniki ekonomiczne, polityczne oraz społeczne, które spowodowały utrwalenie się różnic

regionalnych w zakresie poziomu rozwoju rolnictwa (Pietrzak, Walczak, 2013, s. 334–335; Pietrzak, Walczak, 2014, s. 1019–1021, 1036–1038).

Rysunek 2.5.

Mapa granic zaborów w Polsce



Źródło: opracowanie własne.

Analiza rysunków 2.4 i 2.5 pozwala stwierdzić, że zarówno zmienność przestrzenna struktury agrarnej, jak i zaobserwowane zmiany w charakterze tej struktury dla województwa wielkopolskiego oraz województwa kujawsko-pomorskiego pokrywają się z granicami zaborów. Oznacza to, że zmiany w charakterze struktury agrarnej na terytorium Polski wynikają z zaszczości historycznych. Ważne podkreślenia jest również zjawisko trwałości w czasie struktury agrarnej Polski. Pomimo prowadzonej od 1918 roku przez Państwo polityki rolnej, struktura agrarna w 2002 roku posiada wysoki stopień podobieństwa do struktury zastanej po zaborach (Walczak, Pietrzak, 2016, s. 463–470).

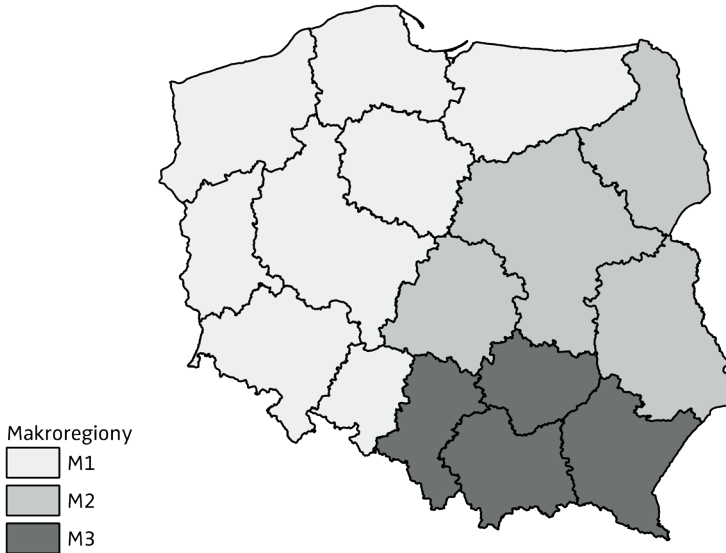
Analiza struktury agrarnej w Polsce jest dobrym przykładem, jak trudne dla badacza jest ustalenie własności jednorodności przyczynowej dla danych przestrzennych i związane z tym wyznaczenie jednorodnego zbioru układów. Dodatkową trudność podczas wyznaczania jednorodnego zbioru układów oraz oceny jego aktualności w kolejnych badaniach powodują również tendencje w przestrzennym kształtowaniu się zjawisk społeczno-ekonomicznych. W przypadku analizy struktury agrarnej użytków rolnych stwierdzono

wysoki stopień trwałości przestrzennego kształtowania się tego zjawiska (Pietrzak, Walczak, 2012, s. 214; Walczak, Pietrzak, 2016, s. 470–471). Oznacza to, że przyjęty poprawnie jednorodny zbiór układów powinien obowiązywać w kolejnych okresach analizy. Natomiast występują również zjawiska, których poziom rozwoju przestrzennego ulega znacznym zmianom nawet w krótkich odstępach czasu. Na przykład, struktura przestrzenna skłonności mieszkańców do zakupów za pomocą Internetu, czy korzystania z bankowości elektronicznej może ulegać takim zmianom i w tym przypadku ustalony jednorodny zbiór układów może nie obowiązywać w kolejnych okresach.

Wizualna ocena zaprezentowanej na rysunku 2.4 koncentracji użytków rolnych pozwoliła na stwierdzenie przestrzennego zróżnicowania struktury agrarnej w Polsce. Świadczy to o istotnych różnicach pod względem powierzchni gospodarstw rolnych i kultury rolnictwa, w zależności od wyboru obszarów o wysokiej lub niskiej koncentracji struktury agrarnej. Oznacza to, że rozpatrywana przez Pawłowskiego zależność popytu na mięso *per capita* względem średniej ceny realnej mięsa nie będzie miała stałego charakteru na terytorium całej Polski dla zbiorowości ludności utrzymującej się wyłącznie z pracy w rolnictwie. W związku z tym biorąc pod uwagę przestrzenne zróżnicowanie struktury agrarnej oraz granice zaborów, terytorium Polski podzielono na trzy makroregiony (zob. rysunek 2.6). Pierwszy makroregion M1 powstał z połączenia województw z zachodniej i północnej części Polski, gdzie występuje wysoka koncentracja użytków rolnych. Drugi makroregion M2 powstał z połączenia województw łódzkiego, mazowieckiego, podlaskiego oraz lubelskiego, gdzie koncentracja użytków rolnych jest na średnim poziomie. Ostatni makroregion M3 wyznaczono poprzez połączenie województw śląskiego, świętokrzyskiego, małopolskiego oraz podkarpackiego. Na obszarze trzeciego makroregionu występuje najniższy poziom koncentracji użytków rolnych.

Wyznaczone granice makroregionów M_i posłużyły do ustalenia trzech podobszarów wniosków. Ze względu na charakter poszczególnych makroregionów założono, że dla wybranych jednostek terytorialnych należących do jednego makroregionu (podobszaru wniosków), zależność popytu na mięso *per capita* względem średniej ceny realnej mięsa dla podzbiorowości ludności utrzymującej się wyłącznie z pracy w rolnictwie będzie na zbliżonym poziomie. Natomiast po przekroczeniu granic wybranego makroregionu M_i dojdzie najprawdopodobniej do zmiany charakteru analizowanej prawidłowości ekonomicznej. Zgodnie z zaproponowanym w pracy określeniem mikro oraz makrozależności przestrzennych, identyfikowana relacja dla każdego z makroregionów M_i będzie mikrozależnością a szacowany parametr mikroparametrem. W związku z tym, dla każdego z trzech makroregionów należy ustalić procedurę wyznaczenia ocen mikroparametrów.

Rysunek 2.6.
Przykładowy podział Polski na trzy makroregiony



Źródło: opracowanie własne.

Procedura wyznaczenia oceny mikroparametru na podstawie modelu regresji przedstawiona zostanie na przykładzie makroregionu M1. Wyznaczenie ocen mikroparametrów dla makroregionów M2 oraz M3 może zostać przeprowadzone w analogiczny sposób. Pierwszy krok procedury stanowi wybór danych przestrzennych, na podstawie których przeprowadzona zostanie estymacja parametrów modelu regresji. W przestrzennych badaniach ekonomicznych, wykorzystane mogą zostać wszystkie układy danych przestrzennych tworzące jednorodny zbiór układów. W tym przypadku, zgodnie z przyjętym jednorodnym zbiorem układów, można wykorzystać dane przestrzenne odniesione do układów jednostek terytorialnych: układu NUTS 4, układu NUTS 3 oraz układu NUTS 2. Następnie z układów danych przestrzennych powinny zostać wybrane wyłącznie te dane przestrzenne, które można jednocześnie przyporządkować do obszaru wniosków, jakim jest makroregion M1. Hipoteza modelowa odniesiona do układu jednostek terytorialnych NUTS 4 określona została wzorem (2.4), hipoteza modelowa odniesiona do układu NUTS 3 wzorem (2.5), a hipoteza modelowa odniesiona do układu NUTS 2 wzorem (2.6)²⁷:

²⁷ Nawet w przypadku zgodności wszystkich trzech parametrów występować mogą różnice w ocenach parametrów wynikające z błędów estymacji oraz różnej liczby stopni swobody.

$$\mathbf{Y}_{M1}^{N4} = \alpha_{0,1}^{N4} + \alpha_{1,1}^{N4} \mathbf{X}_{M1}^{N4} + \boldsymbol{\varepsilon}_{M1}^{N4}, \quad (2.4)$$

$$\mathbf{Y}_{M1}^{N3} = \alpha_{0,1}^{N3} + \alpha_{1,1}^{N3} \mathbf{X}_{M1}^{N3} + \boldsymbol{\varepsilon}_{M1}^{N3}, \quad (2.5)$$

$$\mathbf{Y}_{M1}^{N2} = \alpha_{0,1}^{N2} + \alpha_{1,1}^{N2} \mathbf{X}_{M1}^{N2} + \boldsymbol{\varepsilon}_{M1}^{N2}, \quad (2.6)$$

gdzie \mathbf{Y}_{M1}^{N4} , \mathbf{Y}_{M1}^{N3} , \mathbf{Y}_{M1}^{N2} , \mathbf{X}_{M1}^{N4} , \mathbf{X}_{M1}^{N3} , $\mathbf{X}_{M1}^{N2} \in M1$ są wektorami wartości procesów przestrzennych, $\alpha_{0,1}^{N4}$, $\alpha_{0,1}^{N3}$, $\alpha_{0,1}^{N2}$, $\alpha_{1,1}^{N4}$, $\alpha_{1,1}^{N3}$, $\alpha_{1,1}^{N2}$ stanowią parametry modeli, $\boldsymbol{\varepsilon}_{M1}^{N4}$, $\boldsymbol{\varepsilon}_{M1}^{N3}$, $\boldsymbol{\varepsilon}_{M1}^{N2}$ są wektorami wartości białych szumów przestrzennych, N4, N3 N2 oznaczają odpowiednie układy jednostek terytorialnych NUTS. Równania (2.4)–(2.6) można uogólnić dla procesów przestrzennych $Y(s)$ odniesionych do dowolnego makroregionu M_i .

Należy dodatkowo zwrócić uwagę, że ustalenie trzech układów jednostek terytorialnych w ramach jednorodnego zbioru układów pozwala na estymację mikroparametru za pomocą równań (2.4)–(2.6), na podstawie trzech zbiorów danych przestrzennych odniesionych do układu NUTS 4, układu NUTS 3 oraz NUTS 2. W takiej sytuacji możliwe jest zadanie pytania, czy oceny mikroparametrów $\alpha_{1,1}^{N4}$, $\alpha_{1,1}^{N3}$ i $\alpha_{1,1}^{N2}$ różnią się w zależności od wyboru poziomu agregacji danych w ramach makroregionu $M1$. Pytanie wydaje się być istotne w kontekście zagadnienia problemu skali, gdzie zakładana jest zmienność otrzymywanych wyników pod wpływem zmiany poziomu agregacji danych przestrzennych (Openshaw, 1984b, s. 8; Suchecka, 2014, s. 56–57). Przeprowadzona analiza estymacji mikroparametrów $\alpha_{1,1}^{N4}$, $\alpha_{1,1}^{N3}$, $\alpha_{1,1}^{N2}$ na różnych poziomach agregacji²⁸ nie tylko wpisuje się w problematykę problemu skali, ale może być również skutecznym sposobem na jego rozważenie.

Otrzymane wyniki posłużyć mogą do sformułowania wniosków odnośnie do mikrozależności dla makroregionu $M1$. W związku z tym otrzymane na podstawie równań (2.4)–(2.6) oceny parametrów regresji można wykorzystać do określenia mikrozależności dla całego makroregionu $M1$. W zależności od wyboru układu jednostek terytorialnych NUTS, badana mikrozależność dla makroregionu $M1$ określona jest za pomocą równań (2.7)–(2.9) w ten sposób, że ocena parametru regresji z równań (2.4)–(2.6) uogólniana jest na cały obszar wniosków i staje się oceną mikroparametru dla makroregionu $M1$ zgodnie z równaniami:

²⁸ Zmiana poziomu agregacji związana jest ze zmianą układu danych przestrzennych, tworzącego jednorodny zbiór układów.

$$Y^{M1} = \hat{\alpha}_{0,1}^{N4} + \hat{\alpha}_{1,1}^{N4} X^{M1}, \quad (2.7)$$

$$Y^{M1} = \hat{\alpha}_{0,1}^{N3} + \hat{\alpha}_{1,1}^{N3} X^{M1}, \quad (2.8)$$

$$Y^{M1} = \hat{\alpha}_{0,1}^{N2} + \hat{\alpha}_{1,1}^{N2} X^{M1}, \quad (2.9)$$

gdzie Y^{M1} , Y^{M2} , Y^{M3} , X^{M1} , X^{M2} , X^{M3} są wartościami procesów przestrzennych dla całego makroregionu M1, $\hat{\alpha}_{1,1}^{N4}$, $\hat{\alpha}_{1,1}^{N3}$, $\hat{\alpha}_{1,1}^{N2}$ są ocenami mikroparametru.

Zgodnie z przyjętym rozróżnieniem między mikro i makrozależnościami przestrzennymi, wyznaczanie zależności dla podobszarów wniosków (makroregionów M_i) zawiera się w ramach zagadnienia wyznaczania mikrozależności przestrzennych. Natomiast w przypadku przyjęcia za obszar wniosków całego obszaru Polski, wyznaczaną zależność należy rozumieć jako makrozależność. Uzasadnione jest to faktem, że dla całego obszaru Polski dochodzi do zmiany charakteru zależności popytu na mięso *per capita* względem średniej ceny realnej mięsa dla całego obszaru Polski. Wyznaczone oceny mikroparametrów mogą zostać wykorzystane do policzenia oceny makroparametru zgodnie z postawionym przez Pawłowskiego problemem wyznaczenia relacji między makroparametrem a mikroparametrami (Pawłowski, 1969, s. 235). W tym przypadku relacja dotyczyć będzie związku między makrozależnością popytu na mięso *per capita* względem średniej ceny realnej mięsa a mikrozależnościami określającymi relację popytu do ceny dla makroregionów M_i . Równania przedstawiające modele mikrozależności²⁹ dla trzech makroregionów M_i , jak i oceny mikroparametrów przedstawione zostały za pomocą równań:

$$Y_{M1}^{N4} = \alpha_{0,1}^{N4} + \alpha_{1,1}^{N4} X_{M1}^{N4} + \varepsilon_{M1}^{N4} \rightarrow Y^{M1} = \hat{\alpha}_{0,1}^{N4} + \hat{\alpha}_{1,1}^{N4} X^{M1}, \quad (2.10)$$

$$Y_{M2}^{N4} = \alpha_{0,1}^{N4} + \alpha_{1,1}^{N4} X_{M2}^{N4} + \varepsilon_{M2}^{N4} \rightarrow Y^{M2} = \hat{\alpha}_{0,2}^{N4} + \hat{\alpha}_{1,2}^{N4} X^{M2}, \quad (2.11)$$

$$Y_{M1}^{N4} = \alpha_{0,1}^{N4} + \alpha_{1,1}^{N4} X_{M1}^{N4} + \varepsilon_{M1}^{N4} \rightarrow Y^{M1} = \hat{\alpha}_{0,1}^{N4} + \hat{\alpha}_{1,1}^{N4} X^{M1}, \quad (2.12)$$

gdzie oznaczenia są takie same, jak w przypadku równań (2.4)–(2.9).

Następnie za pomocą równania (2.13) przedstawione zostały dla wybranych makroregionów M_i procesy: proces Y^{M_i} jako wielkość popytu na mięso *per capita* oraz proces X^{M_i} odzwierciedlający średnią cenę realną mięsa. W związku z tym realizację procesu Y^{M_i} stanowią dane przestrzenne mianowane, gdzie

²⁹ Przyjęty został układ danych przestrzennych oparty na układzie jednostek terytorialnych NUTS 4. Możliwy był też alternatywny wybór oparty na układzie jednostek terytorialnych NUTS 3 lub NUTS 2.

wielkość popytu $Y^{1,Mi}$ odniesiona jest do liczby ludności utrzymującej się wyłączenie z pracy w rolnictwie $Y^{2,Mi}$. Natomiast realizację procesu X^{Mi} stanowią dane przestrzenne, które są wynikiem wyznaczenia średniej wartości ceny realnej mięsa dla odpowiednich makroregionów Mi , tj.:

$$Y^{Mi} = \frac{Y^{1,Mi}}{Y^{2,Mi}}, X^{Mi} = \bar{X}_{Mi}^{N4}. \quad (2.13)$$

Popyt na mięso *per capita* dla obszaru Polski Y^{N0} ³⁰ jest ilorazem sumy popytu $Y^{1,Mi}$ oraz sumy liczebności ludności $Y^{2,Mi}$ dla trzech makroregionów Mi , co określone zostało w równaniu (2.14). Popyt na mięso *per capita* Y^{N0} można przedstawić następnie za pomocą równania (2.18), jako ważoną sumę popytów dla makroregionów, gdzie wagi wyznaczane są jako stosunek wielkości popytu wybranego makroregionu Mi oraz liczby ludności całego kraju. Otrzymuje się zatem następujące przekształcenia:

$$Y^{N0} = \frac{Y^{1,N0}}{Y^{2,N0}} = \frac{\sum_{i=1}^3 Y^{1,Mi}}{\sum_{i=1}^3 Y^{2,Mi}} = \frac{Y_1^{1,M1}}{\sum_{i=1}^3 Y^{2,Mi}} + \frac{Y^{1,M2}}{\sum_{i=1}^3 Y^{2,Mi}} + \frac{Y^{1,M3}}{\sum_{i=1}^3 Y^{2,Mi}}, \quad (2.14)$$

$$Y^{N0} = \frac{Y_1^{1,M1}}{\sum_{i=1}^3 Y^{2,Mi}} \frac{Y^{2,M1}}{Y^{2,M1}} + \frac{Y_1^{1,M2}}{\sum_{i=1}^3 Y^{2,Mi}} \frac{Y^{2,M2}}{Y^{2,M2}} + \frac{Y_1^{1,M3}}{\sum_{i=1}^3 Y^{2,Mi}} \frac{Y^{2,M3}}{Y^{2,M3}}, \quad (2.15)$$

$$W_i = \frac{Y^{2,Mi}}{\sum_{i=1}^n Y^{2,Mi}}, \quad (2.16)$$

$$Y^{N0} = \frac{Y^{1,M1}}{Y^{2,M1}} W_1 + \frac{Y^{1,M2}}{Y^{2,M2}} W_2 + \frac{Y^{1,M3}}{Y^{2,M3}} W_3, \quad (2.17)$$

$$Y^{N0} = Y^{M1} W_1 + Y^{M2} W_2 + Y^{M3} W_3. \quad (2.18)$$

³⁰ Indeks N0 oznacza układ jednostek terytorialnych NUTS 0, który tożsamy jest z całym terytorium Polski.

Następnie w równaniu (2.18) można dokonać podstawienia mikrozależności określonych w równaniach (2.10)–(2.12) w miejsce popytu Y^{Mi} dla kolejnych makroregionów M_i , w wyniku czego otrzymano równanie (2.19), tj.:

$$Y^{NO} = (\alpha_{0,1}^{N4} + \alpha_{1,1}^{N4} X^{M1}) w_1 + (\alpha_{0,2}^{N4} + \alpha_{1,2}^{N4} X^{M2}) w_2 + (\alpha_{0,3}^{N4} + \alpha_{1,3}^{N4} X^{M3}) w_3. \quad (2.19)$$

Dodatkowo na podstawie równania:

$$X^{Mi} = \beta_i X^{NO}, \quad (2.20)$$

określono średnią cenę realną dla Polski X^{NO} jako relację średnich cen X^{Mi} dla poszczególnych makroregionów M_i . Uwzględnienie w równaniu (2.19) relacji dla średniej ceny realnej dla Polski X^{NO} pozwoliło na otrzymanie równania:

$$Y^{NO} = (\alpha_{0,1}^{N4} + \alpha_{1,1}^{N4} \beta_1 X^{NO}) w_1 + (\alpha_{0,2}^{N4} + \alpha_{1,2}^{N4} \beta_2 X^{NO}) w_2 + (\alpha_{0,3}^{N4} + \alpha_{1,3}^{N4} \beta_3 X^{NO}) w_3. \quad (2.21)$$

Po odpowiednich przekształceniach równania (2.21) otrzymywano równania (2.22)–(2.25). Wyznaczenie oceny makroparametru α_1^{N4} możliwe jest na podstawie równania (2.24). Ostatecznie otrzymywane równanie (2.25) opisuje makrozależność popytu na mięso *per capita* względem średniej ceny realnej mięsa dla obszaru Polski:

$$Y^{NO} = (\alpha_{0,1}^{N4} w_1 + \alpha_{0,2}^{N4} w_2 + \alpha_{0,3}^{N4} w_3) + (\alpha_{1,1}^{N4} w_1 \beta_1 + \alpha_{1,2}^{N4} w_2 \beta_2 + \alpha_{1,3}^{N4} w_3 \beta_3) X^{NO}, \quad (2.22)$$

$$\alpha_0^{N4} = \alpha_{0,1}^{N4} w_1 + \alpha_{0,2}^{N4} w_2 + \alpha_{0,3}^{N4} w_3, \quad (2.23)$$

$$\alpha_1^{N4} = \alpha_{1,1}^{N4} w_1 \beta_1 + \alpha_{1,2}^{N4} w_2 \beta_2 + \alpha_{1,3}^{N4} w_3 \beta_3, \quad (2.24)$$

$$Y^{NO} = \alpha_0^{N4} = \alpha_1^{N4} X^{NO}. \quad (2.25)$$

Równanie (2.24) przedstawia relację między mikroparametrami dla trzech makroregionów M_i a makroparametrem dla całego obszaru Polski. Wartość relacji zależy od ocen mikroparametrów oraz wag określonych za pomocą równań (2.16) i (2.20). Otrzymane wyniki pokrywają się z wnioskami Pawłowskiego odnoszącymi się do analizy szeregów czasowych (Pawłowski, 1969, s. 239–

244). Należy zwrócić uwagę, że ocena dla makroparametru jest zbliżona do ocen mikroparametrów w przypadku podobieństwa ich wartości oraz podobnych wartości wag. Natomiast, gdy oceny mikroparametrów znacznie różnią się od siebie, może to istotnie wpłynąć na otrzymane wartości oceny makroparametru. Efekt ten może zostać dodatkowo wzmocniony różnicą wartości wag. Należy podkreślić, że w przestrzennych badaniach ekonomicznych najczęściej wyznaczane są makrozależności przestrzenne dla obszaru całego kraju. Pomimo niższej wartości poznawczej w porównaniu z mikrozależnościami przestrzennymi, otrzymane wyniki dotyczące makrozależności przestrzennych mogą być również podstawą wniosków naukowych i stanowić cenny materiał do rozwiązania podjętego problemu badawczego.

Rozdział 3.

Redefinicje problemu sposobu podziału przestrzeni oraz problemu skali

3.1. Geneza powstania problemu zmiennej jednostki odniesienia

W literaturze dotyczącej problemu zmiennej jednostki odniesienia (*Modifiable Areal Unit Problem*, MAUP), autorzy większości prac w kwestii definicji oraz metodyki badań odwołują się do dwóch pozycji: Openshaw, Taylor (1979)¹ oraz Openshaw (1984b) (zob. Anselin 1988, s. 27; Arbia, 1989, s. 1–2, 7–21; Tobler 1989, s. 115–117; Fotheringharn, Wong, 1991, s. 1026–1028; Reynolds, 1998, s. 1; Paelinck, 2000, s. 157–158; Dark, Bram, 2007, s. 471; Flowerdew, 2011, s. 1–3). W obu wymienionych publikacjach Openshaw i Taylor stwierdzają, że problem zmiennej jednostki odniesienia został zidentyfikowany w pracy Yule, Kendall (1966, s. 320–334). Swój wniosek wysuwają na podstawie przedstawionych przez Yule'a i Kendalla badań statystycznych oraz wprowadzonego przez nich pojęcia modyfikowalności obiektów ekonomicznych² (Openshaw, Taylor, 1979, s. 128; Openshaw, 1984b, s. 3, 8).

¹ Praca ta została opublikowana w monografii Wrigley (1979) zawierającej badania wielu autorów z zakresu statystyki oraz ekonometrii przestrzennej. Dzięki jej umieszczeniu w wymienionej monografii praca Openshawa i Taylora traktowana była jako kluczowy głos na temat problemu zmiennej jednostki odniesienia.

² Yule i Kendall definiują pojęcie modyfikowalności obiektów ekonomicznych jako możliwość ich dezagregacji na obiekty ekonomiczne niższego rzędu (Yule, Kendall, 1966, s. 322–323). W przypadku jednostek terytorialnych oznacza to możliwość modyfikacji przebiegu ich granic.

W Openshaw, Taylor (1979, s. 128) stwierdzono, że omówione w pracach Gehlke, Biehl (1934) oraz Yule, Kendall (1966) zagadnienie zmiany otrzymanych wyników podczas przejścia na inny poziom agregacji określone zostanie jako problem skali. Natomiast problem sposobu podziału przestrzeni, określony został jako problem związany z możliwością utworzenia wielu, potencjalnych układów jednostek terytorialnych na tym samym poziomie agregacji (Openshaw, Taylor, 1979, s. 128; Openshaw, 1984b, s. 8).

W pracy Yule, Kendall (1966, s. 320–323) przedstawione zostały wyniki przestrzennych badań ekonomicznych dla modyfikowalnych jednostek obszarowych (układ jednostek terytorialnych 48 powiatów rolniczych w Anglii). Badania dotyczyły analizy korelacji między plonami pszenicy oraz plonami ziemniaków. Postawione zostało pytanie badawcze, czy dla wybranych jednostek obszarowych wysokim plonom pszenicy towarzyszą również wysokie plony ziemniaków. Wyznaczona na podstawie 48 jednostkach terytorialnych wartość współczynnika korelacji liniowej Pearsona wyniosła 0,218. W kolejnym kroku, na podstawie układu 48 jednostek terytorialnych, utworzono kolejne układy na wyższych poziomach agregacji, uzyskując układy 24 jednostek, 12 jednostek, 6 jednostek oraz 3 jednostek terytorialnych. Na podstawie ustalonych układów jednostek terytorialnych wykonany został proces agregacji danych przestrzennych, w wyniku czego otrzymano cztery nowe układy danych przestrzennych³. Następnie ponownie wyznaczono wartości współczynnika korelacji liniowej Pearsona i otrzymano wartość 0,296 dla układu 24 jednostek terytorialnych, wartość 0,575 dla układu 12 jednostek, wartość 0,764 dla układu 6 jednostek oraz wartość 0,992 dla układu 3 jednostek. Otrzymane wartości współczynnika korelacji liniowej Pearsona systematycznie rosły na skutek zmiany poziomu agregacji, co stanowiło dla Openshawa i Taylora argument wspierający stwierdzenie, że Yule i Kendall zidentyfikowali problem zmiennej jednostki odniesienia. Dodatkowo Openshaw i Taylor uznają wyniki dotyczące uzyskiwania różnych wartości współczynnika korelacji liniowej Pearsona, jako wpisane w charakter badań przestrzennych, które występować będą w przypadku każdej analizy.

Otrzymane przez Yule'a i Kendalla wyniki można jednak wytłumaczyć przy użyciu dwóch argumentów. Pierwszy argument odnosi się do kwestii własności jednorodności przyczynowej wykorzystanych układów danych przestrzennych. Dane przestrzenne przyporządkowane zostały najpierw do układu 48 jednostek terytorialnych na niższym poziomie agregacji, a następnie do kolejnych układów na wyższych poziomach agregacji, układów składających się z 24, 12, 6 oraz 3 jednostek terytorialnych. Oznacza to, że jeżeli układ danych przestrzennych (48 jednostek terytorialnych) posiada własność jednorodności przyczynowej, to dla każdego kolejnego układu danych przestrzennych

³ Cztery nowe układy danych przestrzennych otrzymano na podstawie danych dotyczących plonów pszenicy, kolejne zaś cztery nowe układy danych przestrzennych na podstawie danych dotyczących plonów ziemniaków.

na wyższym poziomie agregacji rośnie prawdopodobieństwo, że taka własność nie występuje. Odniesione do układu na wyższym poziomie agregacji, dane przestrzenne są odzwierciedleniem wielu splotów przyczyn i otrzymane wyniki mogą tylko w części odzwierciedlać zależność między wielkością plonów pszenicy a wielkością plonów ziemniaków. Odzwierciedlenie w danych przestrzennych dodatkowych splotów przyczyn może wpływać na otrzymanie różnych wyników, w zależności od przyjętego poziomu agregacji. W takiej sytuacji należałoby pogłębić badanie w odniesieniu do kultury i rozwoju rolnictwa w Anglii w badanym okresie, co pozwoliłoby na ustalenie jednorodnego zbioru układów dla analizowanych płodów rolnych. Rozważenie występowania własności jednorodności przyczynowej dla danych przestrzennych nie tylko pozwoliłoby na ustalenie zbioru układów jednostek terytorialnych, dla których otrzymane wyniki są wiarygodne, ale również na postawienie hipotezy badawczej, co do wartości parametru analizowanej zależności korelacyjnej.

Drugi argument zawarty jest w dalszej części pracy Yule, Kendall (1966, s. 323–325) i dotyczy zjawiska łagodzenia na skutek procesu agregacji danych. Zjawisko to występuje w analizie korelacyjnej i przyczynia się do systematycznego wzrostu wartości współczynnika korelacji liniowej Pearsona przy przechodzeniu na wyższy poziom agregacji. Yule i Kendall przedstawiają zjawisko łagodzenia na przykładzie dwóch procesów ekonomicznych X oraz Y , których realizacje opisane są za pomocą wzorów:

$$x = \xi + e, \quad (3.1)$$

$$y = \eta + f, \quad (3.2)$$

gdzie ξ , η stanowią składniki systematyczne (czynniki przyczynowe, splot przyczyn głównych) procesów X , Y , natomiast e , f są składnikami przypadkowymi (ubocznymi).

Oznacza to, że na realizacji procesów ekonomicznych składa się składnik systematyczny, za którego zmienność odpowiada splot przyczyn głównych oraz składnik przypadkowy. W sytuacji, gdy sploty przyczyn głównych składników systematycznych ξ , η w znacznym stopniu się pokrywają, to możliwa jest identyfikacja zależności korelacyjnej. Pomiar zależności korelacyjnej między składnikami systematycznymi jest zakłócony przez oddziaływanie składników przypadkowych. W pracy Yule, Kendall (1966, s. 324) wykazane zostało, że wyznaczona wartość współczynnika korelacji liniowej Pearsona dla danych na niskim poziomie agregacji zostaje zmniejszona, złagodzona przez oddziaływanie składników przypadkowych. Otrzymane wartości współczynnika korelacji liniowej Pearsona są mniejsze niż miałyby to miejsce w przypadku oddziaływania wyłącznie głównego splotu przyczyn ξ , η . Podczas przechodzenia na wyższy poziom agregacji czynniki przypadkowe będą miały tendencję do wzajemnego

redukowania, co przy kumulatywnym charakterze czynników przyczynowych ξ , η prowadzić będzie do systematycznego wzrostu otrzymywanych wartości współczynnika korelacji liniowej Pearsona. Podsumowując wymienione argumenty należy stwierdzić, że zmierzony poziom zależności korelacyjnej zależy przede wszystkim od tego, czy dane przyporządkowane do wybranych układów jednostek terytorialnych posiadają własność jednorodności przyczynowej. Tylko w przypadku pozytywnej odpowiedzi badacz jest w stanie ocenić w sposób wiarygodny poziom zależności korelacyjnej. Poziom tej zależności jest zakłócany przez udział składnika przypadkowego, który wraz ze wzrostem poziomu agregacji ulega redukcji. Wpływ składnika przypadkowego ulega złagodzeniu, co przekłada się na systematyczne zmiany w otrzymywanych wartościach współczynnika korelacji liniowej Pearsona.

W pracy Openshaw (1984b, s. 11) pojawia się zarzut wobec Yule'a i Kendalla, że nie doceniają oni w wystarczającym stopniu znaczenia swojego odkrycia w postaci problemu zmiennej jednostki odniesienia. Tymczasem zarzut ten należy odeprzeć, ponieważ zgodnie z otrzymanymi wynikami, Yule i Kendall stwierdzają możliwość zmian w otrzymywanych wynikach w zależności od przyjętego poziomu agregacji danych przestrzennych. Autorzy podkreślają istotność tego faktu, szczególnie w sytuacji, kiedy badanie dotyczy oddziaływania czynników przyczynowych (Yule, Kendall, 1966, s. 323–325). Yule i Kendall sugerują jednak przeprowadzenie uzupełniających badań pozwalających na ustalenie właściwych układów danych przestrzennych oraz granic przedziału wartości współczynnika korelacji liniowej Pearsona, które można uznać za wiarygodne. Autorzy podkreślają również, że nie można twierdzić, że niemożliwa jest interpretacja otrzymanych wyników ze względu na występujące różnice, a przeprowadzona analiza korelacyjna jest bezużyteczna w ocenie zależności przyczynowych. Yule i Kendall nie widzą sprzeczności w otrzymaniu różnych wyników w zależności od przyjętego poziomu agregacji. Źródła zmienności w rezultatach doszukują się w przestrzennym kształtowaniu się istniejących zależności przyczynowych na badanym obszarze oraz własności statystycznych analizowanych procesów.

Dodatkowo Openshaw i Taylor zapożyczają z pracy Yule, Kendall (1966, s. 322) pojęcie modyfikowalności obiektów ekonomicznych, które stanowiło argument na rzecz twierdzenia o arbitralności ustalania granic jednostek terytorialnych (zob. Openshaw, Taylor, 1979, s. 128; Openshaw, 1984b, s. 3). W wyniku przeprowadzonych rozważań w rozdziale pierwszym niniejszej pracy można stwierdzić, że pojęcie modyfikowalności obiektów ekonomicznych oparte jest na kryterium agregacji podmiotowej. Własność niemodyfikowalności posiadać będą elementarne obiekty ekonomiczne, takie jak np. konsument. Natomiast jako modyfikowalne określane są te obiekty ekonomiczne, które powstały z elementarnych obiektów niższego rzędu w wyniku agregacji

podmiotowej⁴ (np. gospodarstwo domowe, grupa pracowników jednego zakładu produkcyjnego, zjednoczenia przedsiębiorstw czy obszar geograficzny). W przypadku jednostek terytorialnych własność modyfikowalności polegałaby na możliwości zmiany przebiegu ich granic lub ich agregacji. W wyniku agregacji podmiotowej jednostek terytorialnych opartej na kryterium geograficznym, uzyskiwana byłaby nowa jednostka terytorialna o powiększonych granicach.

3.2. Problem zmiennej jednostki odniesienia a nowy paradygmat badań naukowych

3.2.1. Nowy paradygmat badań naukowych

Praca Openshaw (1984b) stanowi syntezę wcześniejszych rozważań naukowych Openshawa i Taylora (zob. Openshaw, 1977a, 1977b, 1977c, 1978a, 1978b; Taylor, 1977; Openshaw, Taylor, 1979). Traktowana jest ona w literaturze przedmiotu jako praca podsumowująca dotychczasowe badania nad problemem zmiennej jednostki odniesienia. W przypadku prac Openshawa i Taylora należy zwrócić uwagę na problematyczny fakt, zgodnie z którym problem zmiennej jednostki odniesienia traktują oni instrumentalnie, jako narzędzie pozyskania argumentów uderzających w paradygmat nauk empirycznych. Według autora niniejszej pracy, intencją Openshawa i Taylora nie było prowadzenie badań w celu rozwiązania problemu zmiennej jednostki odniesienia, a wykorzystanie w znacznej części problemu zmiennej jednostki odniesienia do propozycji nowego paradygmatu nauk empirycznych (zob. Openshaw, 1984b, s. 34–38). Wskazanie przez autora faktycznego stanu rzeczy jest istotne, ponieważ podejście metodyczne Openshawa i Taylora jest powszechnie wykorzystywane w badaniach naukowych dotyczących problemu zmiennej jednostki odniesienia (zob. Arbia, 1989; Dudley, 1991; Fotheringharn, Wong, 1991; Amrhein, 1995; Reynolds, 1998; Dark, Bram, 2007; Flowerdew, 2011).

Według Openshawa, rozważania nad problemem zmiennej jednostki odniesienia dotyczące możliwości otrzymywania różnych rezultatów badań w zależności od przyjętego poziomu agregacji lub wyboru układu jednostek terytorialnych ukazują słabość paradygmatu nauk empirycznych ze względu na brak możliwości identyfikacji zależności przyczynowych (zob. Openshaw, 1984b, s. 1–7). W związku z tym Openshaw proponuje nowy paradygmat badań naukowych (zob. Openshaw, 1984b, s. 37–38). Istotą nowego paradygmatu byłaby zmiana formuły stawianego problemu badawczego. Ustalenie problemu badawczego polegałoby na sformułowaniu przez badacza hipotezy

⁴ Obiekty ekonomiczne o własności modyfikowalności byłyby tożsame z określonymi w pracach Pawłowski (1969, s. 24), Wiśniewski, Zieliński (1998, s. 5) obiektami wyższego rzędu. W przypadku dwustopniowej agregacji podmiotowej elementarne obiekty ekonomiczne (obiekty niższego rzędu) posiadałyby własność niemodyfikowalności a obiekty wyższego rzędu własność modyfikowalności.

mówiącej, jakie wyniki badań są przez niego pożądane⁵. Hipoteza weryfikowana byłaby na podstawie testu statystycznego, gdzie pożądany wynik byłby zakładany w hipotezie zerowej, wobec określonej hipotezy alternatywnej. Następnie na podstawie zbioru generowanych losowo układów jednostek terytorialnych uzyskiwane byłyby informacje na temat granic przedziału możliwych do uzyskania wartości wyznaczanej statystyki lub oceny parametru⁶. Uzyskane w ten sposób informacje służyłyby weryfikacji postawionej hipotezy badawczej. W sytuacji, gdy ustalone w hipotezie zerowej, pożądane wartości statystyki lub oceny parametru nie należałyby do wyznaczonego przedziału, założona hipoteza badawcza byłaby odrzucana. W odwrotnej sytuacji, kiedy ustalone w hipotezie pożądane wyniki należałyby do wyznaczonego przedziału, brak byłoby podstaw do odrzucenia hipotezy (Openshaw, 1984b, s. 34–35).

Proces wyznaczania wartości statystyk oraz ocen parametrów miał opierać się według Openshawa na zastosowaniu programu komputerowego, który pozwalałby na generowanie losowych układów jednostek terytorialnych na wyższym poziomie agregacji (*zoning system* lub *grouping system*, zob. Openshaw, Taylor, 1979, s. 128, zob. Openshaw, 1984b, s. 8–16). W wyniku agregacji wyjściowych danych przestrzennych otrzymany byłby zbiór danych przestrzennych odniesionych do wielu losowych układów jednostek terytorialnych. Następnie na podstawie danych przestrzennych odniesionych do każdego z losowych układów obliczano by wartości statystyk lub ocen parametrów, co pozwoliłoby na uzyskanie ich rozkładu.

W wyniku przeprowadzonych przez Openshawa badań okazało się jednak, że proces wyznaczania rozkładu analizowanych statystyk lub parametrów na podstawie zbioru generowanych losowo układów jednostek terytorialnych prowadzi do problematycznej sytuacji, gdzie uzyskiwana wielkość rozstępu rozkładu podważa wiarygodność weryfikacji preferowanej hipotezy (Openshaw, 1984b, s. 35–37). W przedstawionym badaniu Openshaw zakładał w hipotezie zerowej konkretną wartość współczynnika korelacji liniowej ($r=0,8$). W wyniku zaproponowanej metodyki badań otrzymał zbiór ocen współczynnika korelacji z przedziału od $-0,928$ do $0,993$. Sytuacja ta prowadziła do niejednoznaczności, ponieważ według Openshawa nie było podstaw do odrzucenia hipotezy zerowej, ze względu na fakt, że konkretna wartość współczynnika korelacji liniowej należała do uzyskanego przedziału wartości. Jednak ze

⁵ Tłumaczenie na podstawie „*what outcome is desired*” (zob. Openshaw, 1984 b, s. 34).

⁶ Według autora, dane przestrzenne odniesione do różnych losowych układów jednostek terytorialnych nie miałyby własności jednorodności przyczynowej, stąd możliwy jest do uzyskania przedział wyznaczanych ocen o szerokim rozstępie. Nie stanowi to jednak kłopotu dla Openshawa, ponieważ odrzuca on paradygmat, gdzie na podstawie danych przestrzennych identyfikowane są zależności przyczynowe. Dodatkowo Openshaw wyklucza możliwość wykorzystania analizy przyczynowej (w tym analizy korelacji i regresji) do rozwiązania stawianych w ramach badań przestrzennych problemów badawczych.

względu na duży rozstęp przedziału wartości współczynnika korelacji weryfikacja postawionej hipotezy nie wydawała się wiarygodna.

W związku z tym Openshaw zaproponował zastosowanie procedury optymalizacyjnej. W przypadku zbyt szerokiego przedziału uzyskanych wartości Openshaw dopuścił możliwość wykorzystania funkcji celu z ograniczeniami, gdzie wprowadzenie dodatkowych warunków brzegowych pozwoliłoby na zawężenie granic przedziału. Openshaw wprowadza narzędzie badawcze w postaci algorytmu optymalizacyjnego⁷, który pozwalałby na zastosowanie systemu generowania losowych układów jednostek terytorialnych do optymalizacji funkcji celu przy założonych ograniczeniach.

Należy podkreślić, że w ramach tak rozumianego paradygmatu badań przestrzennych, identyfikacja zależności przyczynowych przestaje mieć znaczenie, ponieważ badaczka interesuje wyłącznie fakt, czy w wyniku analizy danych przestrzennych odniesionych do wielu alternatywnych układów jednostek terytorialnych możliwy jest do uzyskania, sformułowany w hipotezie badawczej „pożądany wynik”. Zaproponowany nowy paradygmat stawia pożądane przez badaczka cele na pierwszym miejscu, przy jednoczesnej negacji możliwości odkrywania zależności przyczynowych. Zmienia to zupełnie rolę badacza oraz sens prowadzonych badań naukowych. Autor niniejszej pracy uważa, że prowadzenie badań naukowych w ramach nowego paradygmatu sprowadza się wyłącznie do sytuacji generowania kolejnych rozwiązań metodycznych pozwalających na sprawdzenie przez badacza⁸, czy pożądany wynik jest możliwy do uzyskania.

Oceniany krytycznie przez autora nowy paradygmat badań naukowych został celowo przedstawiony na początku podrozdziału. Związane jest to z faktem, że w dalszej części pracy dyskutowane będą kolejne pojęcia związane z problemem zmiennej jednostki odniesienia i przeprowadzone przez autora rozważania wskażą, w jaki sposób Openshaw i Taylor podporządkowali tor swoich badań (prezentacja prac z literatury przedmiotu, kolejne etapy wykonywanych analiz oraz wprowadzane pojęcia) w celu uzyskania argumentów pozwalających na uzasadnienie propozycji nowego paradygmatu nauki. Problem zmiennej jednostki odniesienia był kluczowy dla Openshawa i Taylora, ponieważ w ramach tego problemu mogli podjąć próbę wskazania słabości paradygmatu nauk empirycznych (Openshaw, Taylor, 1979, s. 142–143; Openshaw, 1984b, s. 34–38).

Jak zostało wskazane w podrozdziale 3.1, Openshaw i Taylor zaczynają rozważania nad problemem zmiennej jednostki odniesienia od wykorzy-

⁷ Oryginalne określenie procedury pracy brzmi „*automatic zoning algorithm*” lub „*automatic zoning problem*” (Openshaw, 1977a, s. 462–464; 1984b, s. 21–25).

⁸ W dalszej części tekstu pokazane zostanie, że w przedstawionych przez Openshawa i Taylora badaniach nie udało się zweryfikować, żadnej z postawionych hipotez. Natomiast liczba proponowanych metod oraz rozwiązań roznosiła się nie prowadząc do żadnego postępu w rozwiązaniu problemu badawczego.

stania wyników pracy Yule, Kendall, (1966, s. 320–334), które traktują jako fundamentalny argument przemawiający za występowaniem tego problemu w badaniach przestrzennych. Następnie jednak przechodzą do wyników innych prac, zapożyczając wyłącznie pojęcie modyfikowalności jednostek terytorialnych oraz krytykując Yule'a i Kendalla za podporządkowanie ich badań naukowych paradygmatowi nauk empirycznych, które stoją w sprzeczności z rozważaniami Openshawa i Taylora.

Poniżej omówione zostaną podstawowe założenia paradygmatu nauk empirycznych, do których krytycznie odnoszą się badania Openshawa i Taylora. W ramach paradygmatu nauk empirycznych ustalane są prawa naukowe w szerszym (prawdowości) i węższym znaczeniu (idealizacyjne prawa ściśle ogólne)⁹. Charakter procesów ekonomicznych rozumiany jest jako masowy, co powoduje wyodrębnienie składnika systematycznego i przypadkowego. W takim rozumieniu ekonomii jako nauki empirycznej, w wyniku pomiaru procesów ekonomicznych zachodzących dla ustalonych obiektów ekonomicznych, otrzymywany jest zbiór zdarzeń. Analiza serii zdarzeń ekonomicznych stanowiących realizację badanych procesów pozwala na identyfikację prawidłowości (zależności), które mają charakter przyczynowy. Zależności przyczynowe odkrywane są na podstawie wykonywanej analizy zdarzeń, które są wynikiem tego samego splotu przyczyn głównych (zdarzenia słabo jednorodne; zob. Wiśniewski, Zieliński, 1998, s. 20). Należy podkreślić, że przyczynowe zależności są obiektywne, ponieważ opierają się na mierzalnych zdarzeniach związanych z obiektami ekonomicznymi (Zieliński, 1991, s. 1–24).

Dodatkowo Zieliński określa charakter zdarzeń ekonomicznych w ramach nauk empirycznych, zarówno jako deterministyczny, jak i stochastyczny. Zdarzenia mają charakter deterministyczny, ponieważ jeżeli wystąpił wystarczający zespół przyczyn to bezwzględnie musi wystąpić zdarzenie skutek (Zieliński, 1991, s. 7–17). Natomiast identyfikowane w ramach badań ekonomicznych zależności przyczynowe mają charakter stochastyczny (Zieliński, 1991, s. v). Wynika to z faktu, że splot przyczyn głównych jest zakłócany przez przyczyny uboczne i najczęściej stanowi niewystarczający zespół przyczyn. Oznacza to, że badacz może określić jedynie prawdopodobieństwo wystąpienia wystarczającego zespołu przyczyn. Należy stwierdzić, że w ramach prowadzonych badań rzeczywistość ekonomiczna jest odkrywana za pomocą aparatu stochastycznego, ponieważ determinizm odpowiadający za zmienność procesów jest zbyt skomplikowany do poznania. W przypadku przedstawionych założeń paradygmatu nauk empirycznych należy stwierdzić, że Openshaw i Taylor odrzucają możliwość przyczynowego charakteru prawidłowości ekonomicznych oraz możliwości ich obiektywnego poznania.

⁹ Zob. Lange (1978, s. 54–61); Zieliński (1991, s. 1–24).

3.2.2. Problem zmienności otrzymywanych rezultatów badań naukowych

W celu wprowadzenia nowego paradygmatu nauk empirycznych Openshaw i Taylor zapożyczają z pracy Yule, Kendall (1966) pojęcie modyfikowalności jednostek terytorialnych, a następnie powołują się na wyniki badań przedstawionych w pracy Gehlke, Biehl (1934). Gehlke i Biehl wykonali analizę korelacyjną na podstawie dwóch zbiorów danych przestrzennych odniesionych do 252 jednostek terytorialnych Cleveland. Dla danych przestrzennych z pierwszego zbioru wyznaczono zależność korelacyjną między liczbą nieletnich przestępców w regionie¹⁰ a medianą czynszu mieszkań. Natomiast dla danych z drugiego zbioru ustalono zależność między udziałem nieletnich przestępców wśród nastolatków oraz medianą czynszu mieszkań (Gehlke, Biehl, 1934, s. 169).

W przypadku obydwu zbiorów danych, odniesione do 252 jednostek terytorialnych dane przestrzenne zostały poddane agregacji i przyporządkowane kolejno do układów składających się z 200, 175, 150, 125, 100 50 oraz 25 jednostek terytorialnych. Nowe układy jednostek terytorialnych uzyskane zostały w wyniku łączenia jednostek terytorialnych na podstawie kryterium wspólnej granicy, przy czym założono dodatkowo, że wynikowe jednostki terytorialne powinny być zbliżonych rozmiarów. Dla obydwu przypadków otrzymano ujemne wartości współczynnika korelacji. W przypadku pierwszego zbioru danych wartości bezwzględne współczynnika systematycznie rosły wskazując na rosnącą siłę zależności korelacyjnej, gdzie dla układu 252 jednostek terytorialnych otrzymano wartość $-0,502$, a dla układu 25 jednostek terytorialnych wartość $-0,763$.

W przypadku analizy zależności korelacyjnej między udziałem nieletnich przestępców a wielkością czynszu mieszkań, wartość współczynnika korelacji liniowej dla układu 252 jednostek terytorialnych wyniosła $-0,516$, następnie uległa zmianie do wartości $-0,480$, $-0,475$ i $-0,621$ dla układów 175, 150 oraz 25 jednostek terytorialnych. Otrzymana wartość współczynnika korelacji liniowej Pearsona dla układu 252 jednostek terytorialnych na poziomie $-0,516$, wydaje się właściwie odzwierciedlać badaną zależność korelacyjną. Zmiana wartości współczynnika korelacji liniowej Pearsona dla układów składających się ze 175 i 150 jednostek wynikać może z faktu, że układy te w mniejszym stopniu odzwierciedlają oddziaływanie istotnych determinant zjawiska przestępczości. Natomiast systematyczny wzrost wartości bezwzględnej współczynnika korelacji dla układów z wyższych poziomów agregacji wynika najprawdopodobniej z efektu łągodzenia wpływu składnika przypadkowego.

¹⁰ Dane dotyczące liczby nieletnich przestępców nie powinny zostać wykorzystane w analizie przestrzennej ze względu na brak porównywalności danych w poszczególnych regionach. Dopiero wyznaczenie udziału nieletnich przestępców wśród nastolatków gwarantuje porównywalność danych.

Gehlke i Biehl wskazują na różnice w obliczonych wartościach współczynnika korelacji liniowej Pearsona, jednak nie podejmują próby ekonomicznej interpretacji lub uzasadnienia otrzymanych wyników. Natomiast na podstawie wykonanej analizy właściwej¹¹, stawiają pytanie badawcze podważające paradygmat nauk empirycznych, czy w ogóle jest możliwa identyfikacja zależności przyczynowych za pomocą analizy korelacyjnej (Gehlke, Biehl, 1934, s. 170). Zdaniem autora, otrzymane wyniki nie stanowią uzasadnienia dla postawionego w cytowanej pracy pytania badawczego odnośnie do możliwości identyfikacji zależności przyczynowych. Natomiast otrzymane w pracy Gehlke, Biehl (1934) wyniki potwierdzały występowanie wprowadzonego przez Openshawa i Taylora problemu zmiennej jednostki odniesienia oraz stanowiły ważny argument za brakiem obiektywności wykonywanych badań naukowych (Openshaw, Taylor, 1979, s. 142–143)

W świetle przedstawionych rozważań można stwierdzić, że otrzymane przez Gehlke'a i Biehl rezultaty są odzwierciedleniem oddziaływania różnych splotów przyczyn, w zależności od przyjętego poziomu agregacji. Badanie dotyczy oceny oddziaływania wybranych determinant na wielkość przestępczości wśród nastolatków, w tym wysokości czynszów w badanym regionie. Zjawisko przestępczości wśród nastolatków związane jest z niskim poziomem zamożności rodzin, co powinno mieć swoje odzwierciedlenie w otrzymaniu ujemnych wartości współczynnika korelacji liniowej Pearsona. Otwartym pozostaje pytanie, jak zjawisko przestępczości kształtuje się przestrzennie. Odpowiedź na to pytanie pozwoliłaby na ustalenie jednorodnego zbioru układów, gdzie dane odniesione do poszczególnych układów jednostek terytorialnych odzwierciedlałyby w sposób właściwy oddziaływanie przyczyn odpowiedzialnych za powstawanie zjawiska przestępczości wśród nastolatków. Należy podkreślić, że ustalenie jednorodnego zbioru układów wymagałoby pogłębionego badania sytuacji społeczno-ekonomicznej poszczególnych jednostkach terytorialnych miasta Cleveland. Występowanie tego zjawiska jest charakterystyczne dla małych jednostek (dzielnica, część dzielnicy), gdzie lokalne, zamknięte społeczności stają się źródłem przestępczości. Wstępnie wydaje się, że układ 252 jednostek terytorialnych miasta Cleveland jest właściwym układem jednostek terytorialnych do analizy tego zjawiska. Niestety kolejne układy jednostek terytorialnych tworzone były przez Gehlke'a i Biehl w subiektywny sposób oparty na zasadzie podobnej wielkości regionów. Ponieważ jednostki łączone były na podstawie kryterium wspólnej granicy, to możliwe jest, że układy dla 200, 175, czy 150 jednostek terytorialnych odzwierciedlają częściowo oddziaływa-

¹¹ Można stwierdzić, że z przedstawionych w rozdziale drugim trzech etapów przestrzennych badań ekonomicznych, wykonana została tylko analiza właściwa. Podetap analizy właściwej polegał jednak wyłącznie na policzeniu wartości współczynnika korelacji liniowej.

nie istotnych determinant zjawiska przestępczości¹². Natomiast kolejne układy, w tym układ 25 jednostek terytorialnych, są zbyt duże powierzchniowo i w ich obrębie nakłada się oddziaływanie wielu splotów przyczyn odpowiedzialnych za zjawiska społeczno-ekonomiczne, w tym za powstawanie przestępczości. Ostatecznie można podsumować, że otrzymany na podstawie liczby nastoletnich przestępców, systematyczny spadek wartości współczynnika korelacji liniowej Pearsona wynikać może z przedstawionej w pracy Yule, Kendall (1966) tendencji do wzrostu wartości współczynnika przy rosnącym poziomie agregacji, co związane jest z kumulacją wartości danych oraz łagodzeniem wpływu składnika przypadkowego.

Dodatkowo Openshaw i Taylor wskazują na pionierskie podejście w pracy Gehlke, Biehl (1934) do sposobu losowego wyznaczania granic układu jednostek terytorialnych¹³. Gehlke i Biehl, na podstawie układu 252 jednostek terytorialnych na niższym poziomie, wyznaczyli losowo kolejne dwa układy na wyższych poziomach agregacji. Pierwszy układ jednostek terytorialnych składał się ze 150 jednostek terytorialnych, a otrzymana na podstawie liczby nastoletnich przestępców wartość współczynnika korelacji liniowej Pearsona wyniosła $-0,434$, natomiast drugi układ utworzony został z 25 jednostek, gdzie wartość współczynnika wyniosła $-0,544$ (Gehlke, Biehl, 1934, s. 169). Losowy sposób łączenia jednostek spowodował, że dane przyporządkowane do takich układów reprezentowały oddziaływanie różnych splotów przyczyn. W tej sytuacji nie powinno dochodzić do systematycznych zmian wartości współczynnika korelacji liniowej Pearsona, a otrzymywane różnice były nieistotne statystycznie, co wynikałoby z losowego charakteru doboru danych. Niewłaściwa jest natomiast interpretacja ekonomiczna wyników, które otrzymane zostały na podstawie losowo dobranego układu jednostek terytorialnych ze względu na pomieszenie oddziaływania różnych splotów przyczyn.

Ostatni przykład empiryczny z pracy Gehlke, Biehl (1934, s. 170), gdzie Openshaw i Taylor wskazują na uzyskane zmiany w wartościach współczynnika korelacji liniowej, dotyczy danych przestrzennych odniesionych do układu 1000 wiejskich jednostek terytorialnych¹⁴. Na podstawie danych przestrzennych wyznaczono zależność korelacyjną pomiędzy wartością produkcji farm a liczbą farmerów¹⁵. Otrzymano wartość współczynnika korelacji liniowej Pearsona równą $0,649$. Następnie wyznaczone zostały granice nowych układów,

¹² W latach 30. XX w. miasto Cleveland liczyło około 900 tysięcy mieszkańców. Dla układu 252 jednostek terytorialnych średnia liczba mieszkańców na region wynosi 3550, dla układu 150 jednostek 6000 mieszkańców, a dla układu 25 jednostek 36 tysięcy mieszkańców.

¹³ Niestety procedura losowego grupowania jednostek nie została szczegółowo omówiona.

¹⁴ Autorzy nie podali jakiego kraju dotyczy wybrany układ wiejskich jednostek terytorialnych.

¹⁵ Autor zakłada, że farmer utożsamiany był z gospodarstwem rolnym.

składających się z 125, 63, 40 oraz 31 jednostek terytorialnych. Wyznaczone wartości współczynnika dla kolejnych poziomów agregacji były równe 0,826, 0,859, 0,725 oraz 0,756. Na podstawie wykonanej analizy korelacyjnej Gehlke i Biehl ponownie stwierdzili występowanie różnic w otrzymywanych wynikach w zależności od przyjętego poziomu agregacji.

Otrzymana dla układu 1000 wiejskich jednostek terytorialnych (regionów) wartość współczynnika korelacji liniowej Pearsona na poziomie 0,649 wydaje się wstępnie wiarygodnym wynikiem ze względu na fakt, że większej liczbie farmerów powinna towarzyszyć większa produkcja rolnicza w regionie. Jednak Gehlke i Biehl nie uzasadnili wyboru układu jednostek terytorialnych i danych przestrzennych oraz nie dokonali oceny poziomu rozwoju i kultury rolnictwa na badanym obszarze. W rozdziale drugim zostało podkreślone, że w dużym stopniu o poziomie rozwoju i kulturze rolnictwa świadczy przestrzenne zróżnicowanie struktury agrarnej. W zależności od charakteru struktury agrarnej, potencjalnie wystąpić mogą dwie przeciwstawne sytuacje. W sytuacji pierwszej występować może niski stopień koncentracji użytków rolnych na podobnym poziomie dla każdego z 1000 regionów wiejskich wybranego układu. Oznaczać to będzie dla wszystkich badanych regionów znaczną przewagę rozdrobnionych gospodarstw rolnych o małej lub średniej powierzchni. Przełoży się to na niski poziom produkcji rolnej, ponieważ gospodarstwa rolne o małej powierzchni generować będą niski dochód, co będzie stanowić barierę dla zwiększenia efektywności produkcji. W takiej sytuacji większa liczba farmerów w regionie przełoży się proporcjonalnie na większą wartość produkcji rolnej, a badana zależność korelacyjna powinna być dodatnia i silna.

W sytuacji drugiej, część regionów może charakteryzować się wysokim poziomem koncentracji użytków rolnych, a pozostałe regiony niskim poziomem. Udział regionów o wysokim poziomie koncentracji gospodarstw rolnych jest najczęściej mniejszy, jednak charakteryzują się one większą średnią powierzchnią użytków rolnych oraz większą produkcją rolną¹⁶. Mniejsza liczba farmerów w przypadku regionów o wysokim poziomie koncentracji użytków rolnych może się przełożyć na znacznie większą wartość produkcji rolnej w regionie, niż miałyby to miejsce w przypadku większej liczby farmerów gospodarujących na rozdrobnionych gospodarstwach rolnych. Wyznaczona w takiej sytuacji wartość współczynnika korelacji liniowej Pearsona może wskazywać nawet na ujemną zależność korelacyjną. Przedstawione rozważania wskazują jednoznacznie, że bez wnikliwej analizy kultury rolnictwa oraz charakteru struktury agrarnej na badanym obszarze, trudno zinterpretować i ocenić wiarygodność otrzymanych przez Gehlke'a i Biehl wyników oraz ich zmian na różnych poziomach agregacji.

¹⁶ Wyższy poziom produkcji rolnej nie wynika wyłącznie z wyższej średniej powierzchni użytków rolnych, ale również z powodu wyższej efektywności produkcyjnej gospodarstw.

Należy podkreślić, że w przedstawionych przez Gehlke'a oraz Biehl badaniach, wykonali oni wyłącznie analizę właściwą polegającą na wyznaczeniu wartości współczynnika korelacji liniowej Pearsona, a następnie wyciągnięte zostały wnioski mówiące jedynie o otrzymanych różnicach w rezultatach. W tabeli 3.1 zaprezentowano oceniony przez autora stopień realizacji w pracy Gehlke, Biehl (1934) rekomendowanych w rozdziale drugim trzech etapów przestrzennych badań ekonomicznych.

Tabela 3.1.

Stopień wykonania kolejnych etapów przestrzennych badań ekonomicznych w pracy Gehlke i Biehl (1934)

Etapy przestrzennego badania ekonomicznego
Etap1. Postawienie problemu badawczego: brak
Etap 2. Analiza ekonomiczna
2.1. Ustalenie obszaru analizy: brak
2.2. Wybór układu (układów) danych przestrzennych: brak
2.3. Agregacja danych przestrzennych
2.3.1. Wybór wyjściowego, jednorodnego przyczynowo układu danych przestrzennych na niższym poziomie agregacji: brak, wybór układu nieuzasadniony, wybór wyniku prawdopodobnie z posiadanych danych
2.3.2. Wybór układów jednostek terytorialnych dla poziomów agregacji wyższego rzędu: brak, układy na wyższych poziomach agregacji ustalone arbitralnie przy założeniu uzyskiwania jednostek o podobnych rozmiarach
2.3.3. Proces agregacji danych: wykonano, ale nie opisano procedury agregacji danych
2.3.4. Ustalenie własności jednorodności przyczynowej dla układów danych przestrzennych wyższych rzędów: brak
2.4. Ustalenie jednorodnego zbioru układów: brak
2.5. Identyfikacja wewnętrznej struktury danych przestrzennych: brak
2.6. Wybór właściwych narzędzi badawczych: brak
2.7. Ustalenie obszaru (podobszarów) wniosków: brak
2.8. Właściwa analiza: analiza korelacyjna
Etap 3. Próba rozwiązania problemu badawczego na podstawie wniosków uzyskanych z analizy: brak

Źródło: opracowanie własne.

Analiza kolejnych punktów zawartych w tabeli 3.1 pozwala na negatywną ocenę jakości przeprowadzonych przez Gehlke'a i Biehl badań naukowych, w których pomijane są najważniejsze etapy. Pomimo tego autorzy przekonani są o zasadności poddania w wątpliwość wykorzystania analizy korelacyjnej, jako narzędzia identyfikacji zależności przyczynowych. Nie poprzestając na tym stawiają pytanie uderzające w paradygmat nauk empirycznych: czy jednostki

terytorialne (regiony) stanowią obiekty ekonomiczne o obiektywnych cechach (zmiennych ekonomicznych) (Gehlke, Biehl, 1934, s. 170).

Po szczegółowym omówieniu wyników zaprezentowanych w pracy Gehlke, Biehl (1934) pojawia się pytanie, dlaczego Openshaw i Taylor odwołują się do tej pracy i określają ją jako jedną z kluczowych w badaniach nad problemem zmiennej jednostki odniesienia. Z punktu widzenia rekomendowanej w niniejszej pracy metodyki badań zjawisk ekonomicznych cytowana praca zawiera istotne uchybienia. Dodatkowo pomimo słabej strony metodycznej badania, Gehlke i Biehl wysuwają nieuprawnione stwierdzenia o niemożliwości identyfikacji zależności przyczynowych na podstawie danych przestrzennych. Możliwym powodem wydaje się być użyteczność dla Openshawa i Taylora końcowych wniosków z pracy Gehlke, Biehl (1934) o braku możliwości obiektywnego poznania analizowanych cech regionów oraz o braku wartości poznawczej dla analizy przyczynowej, w ich argumentacji przeciwko paradygmatowi nauk empirycznych.

3.2.3. Problem modyfikowalności jednostek terytorialnych

Na początku rozdziału trzeciego wyróżnione zostały dwie prace: Openshaw, Taylor (1979) oraz Openshaw (1984b), na podstawie których przyjmowana jest definicja oraz metodyka badań nad problemem zmiennej jednostki odniesienia w większości późniejszych prac naukowych. Ponieważ wymienione prace miały kluczowy wpływ na dalszy rozwój problemu zmiennej jednostki odniesienia, ich treść zostanie omówiona oraz poddana krytycznej dyskusji. Przeprowadzone rozważania pozwolą na wskazanie błędów w określeniu problemu zmiennej jednostki odniesienia oraz na sformułowanie propozycji redefinicji tego problemu.

Pierwszym z podjętych przez Openshawa i Taylora zagadnieniem jest problem modyfikowalności jednostek terytorialnych. Autorzy stawiają pytanie badawcze, w jaki sposób należy wyznaczyć granice układu jednostek terytorialnych oraz jaki wybrać poziom agregacji w badaniach (Openshaw, Taylor, 1979, s. 127). Próba odpowiedzi na to pytanie prowadzi autorów do stwierdzenia o arbitralnej naturze jednostek przestrzennych. W utworzonym arbitralnie przez badacza układzie, jednostki terytorialne mogą posiadać dowolne granice. Oznacza to możliwość utworzenia wielu potencjalnych układów na dowolnym poziomie agregacji. Stwierdzenie to stanowi punkt wyjścia dla Openshawa i Taylora podczas określenia problemu zmiennej jednostki odniesienia, gdzie podstawą problemu jest modyfikowalność jednostek terytorialnych oraz arbitralny charakter układów jednostek terytorialnych (Openshaw, Taylor, 1979, s. 127).

Należy nie zgodzić się z prezentowanym przez Openshawa i Taylora stanowiskiem i stwierdzić, że wybór układu jednostek terytorialnych na wybranym poziomie agregacji nie jest dowolny i powinien wynikać z podjętego problemu

badawczego. Decyzję o wyborze układu jednostek terytorialnych oraz przyporządkowanych do niego danych podejmuje arbitralnie badacz, jednak wybór ten musi wynikać z podjętego problemu badawczego, wyników wcześniejszych badań oraz doświadczenia naukowego badacza¹⁷.

Następnie Openshaw i Taylor przeprowadzają dyskusję nad brakiem argumentów, które by konstytuowały jednostkę terytorialną jako obiekt badania (Openshaw, Taylor 1979, s. 143). Według nich decyzja o wyborze układu jednostek terytorialnych ustalana jest bez możliwości uzasadnienia tego wyboru. W związku z tym Openshaw i Taylor proponują rozwiązanie polegające na uznaniu prawomocności dowolnych wyników, uzyskanych na postawie arbitralnie przyjętych układów jednostek terytorialnych. Dodatkowo pojęcie arbitralności jednostek terytorialnych wykorzystane zostało do autorskiej propozycji systemu generowania układów jednostek terytorialnych (Openshaw, Taylor, 1979, s. 131–134). Zbudowany odpowiednio system pozwalałoby na losowe generowanie granic dla układów jednostek terytorialnych, eliminując przy tym udział badacza¹⁸.

Dodatkowo w pracy Openshaw (1984b, s. 3) formułowane jest twierdzenie, że ze względu na własność modyfikowalności jednostek terytorialnych, źródłem ich arbitralności jest badacz. Ponieważ ustalone przez badacza modyfikowalne jednostki terytorialne mają arbitralny charakter, to wartość prowadzonych na ich podstawie badań przestrzennych jest wątpliwa (Openshaw, 1984b, s. 4–7). Proponowane jest rozwiązanie polegające na ustaleniu zbioru zasad (systemu zasad), przy wykorzystaniu których granice układów jednostek terytorialnych mogłyby być ustalane w obiektywny sposób. Należy pokreślić negatywne skutki tej propozycji, gdzie ciężar zapewnienia obiektywności przestrzennych badań ekonomicznych został błędnie przesunięty. Według Openshawa obiektywności badań nie zapewni postawienie problemu badawczego, odpowiednia metodyka badań, wybór odpowiednich danych przestrzennych i układu jednostek terytorialnych. Natomiast obiektywność badań może zapewnić uniwersalna reguła wyznaczania granic układu jednostek terytorialnych (np. jednostki terytorialne powinny być podobnych rozmiarów lub posiadać taką samą liczbę ludności). Według Openshawa, tylko takie podejście może zagwarantować obiektywizm badań, ponieważ pozwala na wyeliminowanie arbitralnego wpływu badacza na wybór układu jednostek terytorialnych, które mogłyby być wybierane w oparciu o przyjęte powszechnie zasady. Propozycja uniwersalnej reguły wyznaczania granic układów powinna prowadzić do wypracowania naukowego konsensusu, rozumianego jako zbiór reguł pozwalających na zobiektywizowanie wyboru układów jednostek terytorialnych. Problem polega jednak na tym, że przyporządkowane do uprawomocnionych

¹⁷ Oznacza to, że dwóch niezależnych badaczy powinno przyjąć arbitralnie podobne układy jednostek terytorialnych w ramach tego samego problemu badawczego.

¹⁸ System losowego generowania układów jednostek terytorialnych opisany zostanie w dalszej części pracy.

w ten sposób układów dane przestrzenne nie będą miały najprawdopodobniej własności jednorodności przyczynowej. W związku z tym przyjęty powszechnie zbiór reguł określający zasady wyznaczania granic układu jednostek terytorialnych nie może zapewnić poprawnego opisu zależności przyczynowych.

3.2.4. Wyznaczanie granic układu jednostek terytorialnych a problem zmiennej jednostki odniesienia

W przestrzennych badaniach ekonomicznych kluczową kwestią jest wybór odpowiedniego układu jednostek terytorialnych, do którego odniesione zostaną dane przestrzenne. Openshaw i Taylor wprowadzają propozycje metodyczne w postaci budowy dwóch systemów, w ramach których generowany byłby w sposób losowy zbiór potencjalnych układów jednostek terytorialnych (Openshaw, Taylor 1979, s. 127, 131–132; Openshaw, 1984b, s. 8–12). Pierwszy z systemów, *zoning system* pozwala na losowe tworzenie zbioru układów jednostek terytorialnych, jednak tylko takich, których jednostki terytorialne posiadają ciągłą granicę. Dodatkowo w celu zastosowania systemu, należy założyć wyjściowy układ jednostek terytorialnych na niższym poziomie agregacji, który stanowić będzie podstawę do utworzenia jednostek układu na wyższym poziomie agregacji¹⁹.

Na rysunku 3.1 przedstawiono układ jednostek terytorialnych o proporcjonalnej powierzchni wszystkich jednostek (układ NUTS 2)²⁰. Natomiast na rysunku 3.2 przedstawiony został utworzony losowo w ramach *zoning system* układ 16 jednostek terytorialnych, gdzie jedna z jednostek posiada znacznie większą powierzchnię od pozostałych jednostek. Do budowy wymienionych układów wykorzystano układ jednostek terytorialnych NUTS 4, gdzie dla obydwu układów wszystkie jednostki terytorialne utworzone zostały na podstawie graniczących ze sobą powiatów.

Analiza układów przedstawionych na rysunkach 3.1 oraz 3.2 pozwala na stwierdzenie, że w wyniku zastosowania *zoning system* możliwe jest uzyskanie układów o zupełnie odmiennym przebiegu granic²¹. Według autora nie można traktować obydwu układów równorzędnie, ponieważ zgodnie z przeprowadzoną w rozdziale drugim dyskusją, na wybranym poziomie agregacji tylko jeden układ jednostek terytorialnych pozwoli na poprawne uchwycenie oddziaływania przyczyn w ramach podjętego problemu badawczego. Układ

¹⁹ Możliwe jest również wykorzystanie systemu bez założenia wyjściowego układu jednostek terytorialnych na najniższym poziomie agregacji. W takiej sytuacji centroidy jednostek terytorialnych ustalane są na podstawie losowo wybranych punktów.

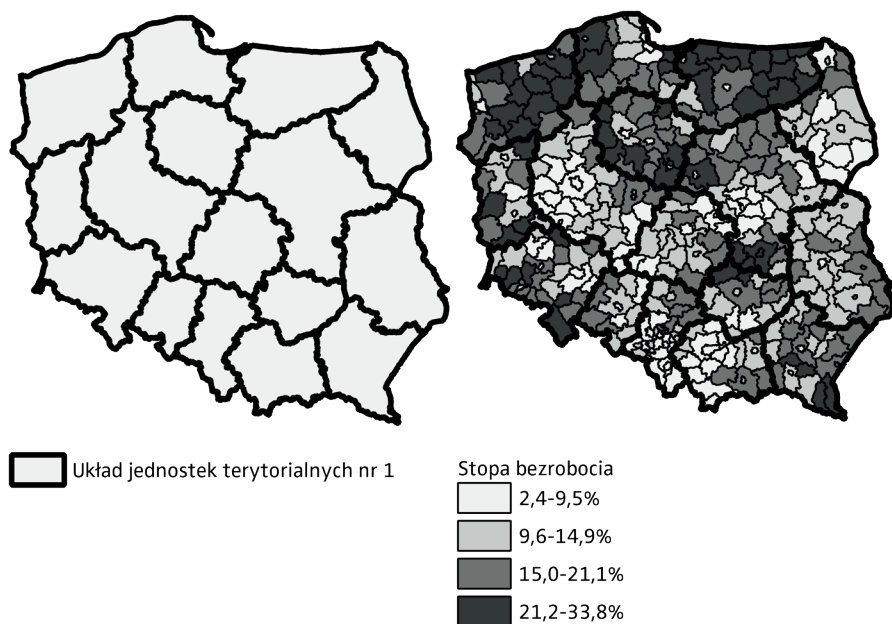
²⁰ Układ jednostek terytorialnych NUTS 2 stanowi przykład potencjalnego układu jednostek terytorialnych, który może zostać wygenerowany losowo w ramach zastosowania *zoning system*.

²¹ Przebieg granic determinowany jest jednak przebiegiem granic układu na niższym poziomie agregacji.

16 jednostek terytorialnych przyjęty został ze względu na czytelność przedstawianych wyników²². Idea stosowania *zoning system* do losowego generowania układów jednostek terytorialnych wydaje się być naukowo atrakcyjna. Jednak porównanie dowolnych dwóch, generowanych losowo, układów jednostek terytorialnych od razu pokazuje negatywny aspekt tego pomysłu. Jeżeli weźmiemy administracyjny układ jednostek dla dowolnego obszaru, to udział przestrzenny jednostek, pomimo występujących różnic, jest proporcjonalny. Tymczasem utworzenie losowego układu w ramach *zoning system* może doprowadzić do sytuacji, gdzie udział powierzchni pojedynczej jednostki obszarowej znacznie przewyższa udziały pozostałych jednostek (np. dla układu 16 jednostek terytorialnych, ponad 90% powierzchni całego obszaru zajmuje jedna jednostka terytorialna, zob. rysunek 3.2).

Rysunek 3.1.

Układ jednostek terytorialnych nr 1

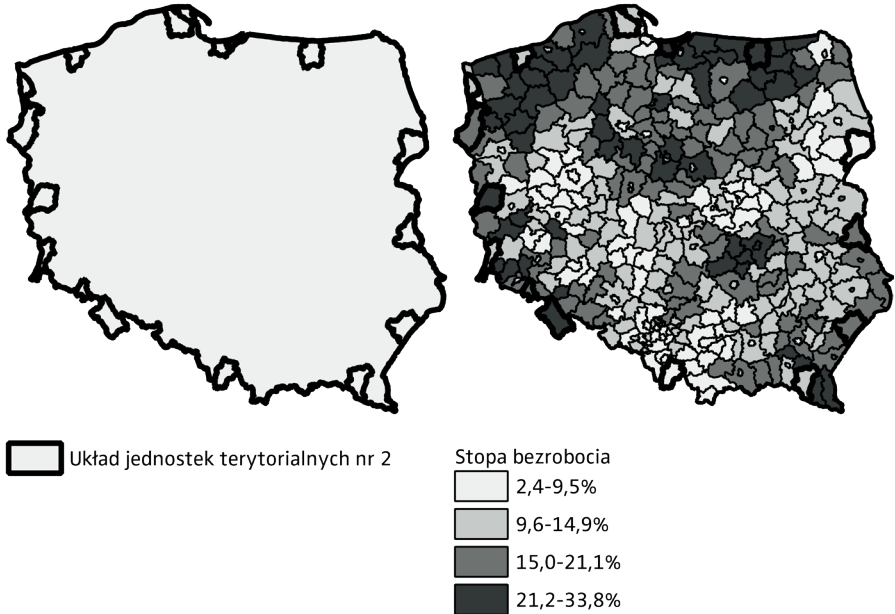


Źródło: opracowanie własne.

²² W przypadku analizy zjawiska bezrobocia układ składający się z 16 jednostek terytorialnych wyznaczany jest na zbyt wysokim poziomie agregacji. Widoczne jest, że odniesione do układu jednostek terytorialnych NUTS 2 dane przestrzenne dotyczące zjawiska bezrobocia nie posiadają własności jednorodności przyczynowej.

Rysunek 3.2.

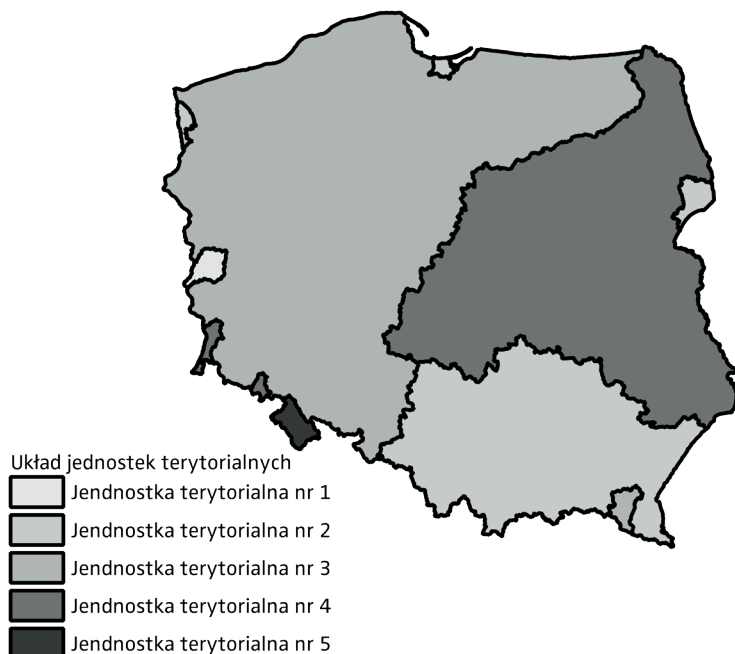
Układ jednostek terytorialnych nr 2



Źródło: opracowanie własne.

Drugi z zaproponowanych przez Openshawa i Taylora systemów określony został jako *grouping system*. Zastosowanie tego systemu również pozwala na losowe generowanie układów jednostek terytorialnych. Tworzone w ramach *grouping system* jednostki terytorialne nie muszą jednak spełniać warunku ciągłości granic (Openshaw, 1984b, s. 8). Oznacza to, że utworzona jednostka terytorialna na wyższym poziomie agregacji może składać się z kilku, niegraniczących ze sobą jednostek terytorialnych (wysp) z niższego poziomu agregacji. Na rysunku 3.3 przedstawiono przykładowy układ 5 jednostek terytorialnych, który mógłby być losowo wygenerowany w ramach *grouping system*. Podobnie jak w przypadku *zoning system* (zob. rysunek 3.1, 3.2), układ ten został zbudowany na podstawie jednostek terytorialnych z układu NUTS 4. Na rysunku 3.3 widoczne są dwie jednostki terytorialne o małej powierzchni (jednostki: nr 1 i 5) oraz trzy jednostki terytorialne o znacznie większej powierzchni, w skład których wchodzi również obszary o charakterze wysp (jednostki: nr 2, 3 i 4).

Rysunek 3.3.

Układ jednostek terytorialnych wygenerowany w ramach *grouping system*

Źródło: opracowanie własne.

Skonfrontowanie idei losowych systemów generowania układów jednostek terytorialnych z rozważaniami przeprowadzonymi w rozdziale drugim, pozwala stwierdzić, że wykorzystanie *zoning system* oraz *grouping system* w ramach prowadzonych przestrzennych badań ekonomicznych jest niewłaściwe. Wykorzystanie *zoning system* oraz *grouping system* stwarza potencjalne niebezpieczeństwo otrzymania niepoprawnych wniosków podczas rozwiązywania postawionego problemu badawczego. Zgodnie z przedstawionymi etapami przestrzennego badania ekonomicznego, badacz powinien utworzyć jednorodny zbiór układów, który z perspektywy podjętego problemu badawczego składać się będzie z układów danych przestrzennych o własności jednorodności przyczynowej. Natomiast w wyniku wykorzystania *zoning system*, wygenerowany może zostać w losowy sposób dowolny układ jednostek terytorialnych, różniący się istotnie w przebiegu granic od układu przyporządkowanego do jednorodnego zbioru układów. Oznacza to, że zagregowane dane przestrzenne dla wygenerowanego losowo układu jednostek terytorialnych nie będą posiadały własności jednorodności przyczynowej, ze względu na pomieszanie oddziaływania różnych splotów przyczyn. W związku z tym, zagregowane w ten sposób dane przestrzenne nie będą poprawnie odzwierciedlać zależności przyczynowych dla badanych procesów. Oznacza to, że jeżeli badacz nie rozważy

problemu w ramach właściwego układu jednostek terytorialnych, analiza będzie niepoprawna.

3.2.5. Jednorodność przyczynowa danych przestrzennych a wyniki badań

Podsumowując dotychczasową dyskusję na temat propozycji metodycznych w pracach Openshawa i Taylora (1979) oraz Openshawa (1984b) należy dokonać krytycznej oceny idei zastosowania *zoning system*, jak również *grouping system* do generowania losowych układów jednostek terytorialnych. Konsekwencją zastosowania tego typu systemów jest uzyskanie zbioru układów jednostek terytorialnych o losowym przebiegu granic. Następnie do zbioru wygenerowanych losowo układów przyporządkowane zostają agregowane empiryczne dane przestrzenne lub symulowane realizacje procesów o określonych własnościach. Dla każdego losowego układu można wyznaczyć wartość statystyki lub oceny parametru na podstawie danych przestrzennych odniesionych do tego układu. Według Openshawa i Taylora otrzymywany jest w ten sposób rozkład badanej statystyki lub parametru. Wykonana następnie analiza rozkładu pozwala na ocenę własności procesów lub zależności między procesami, których realizacje stanowią przyjęte dane przestrzenne.

W przypadku tak przeprowadzonych badań pojawia się jednak problem polegający na tym, że losowe generowanie zbioru układów jednostek terytorialnych może spowodować otrzymanie szerokiego zakresu wartości statystyk czy ocen parametrów na podstawie wykonywanej analizy. W celu wyjaśnienia istoty rozważanej kwestii, poniżej przedstawiony zostanie przykład empirycznej analizy wykonanej przez Openshawa i Taylora, na podstawie której dokonali oni identyfikacji problemu zmiennej jednostki odniesienia. Przeprowadzona analiza dotyczyła badania wpływu wieku wyborców na posiadane przez nich preferencje wyborcze (Openshaw, Taylor, 1979, s. 129). Dane dotyczyły procentowych wyników głosowania na Partię Republikańską w USA oraz udziału procentowego wyborców powyżej 60 lat. Dane te odniesione zostały do 99 jednostek administracyjnych stanu Iowa w USA. Niestety wybór danych przestrzennych uzasadniony został jedynie łatwością ich pozyskania oraz przekształcania (Openshaw, Taylor, 1979, s. 129). Autorzy łączą udział osób starszych z odzwierciedleniem sytuacji demograficzno-ekonomicznej regionów, która z kolei przekłada się na wynik wyborów politycznych w wybranych regionach. Na podstawie danych przestrzennych wyznaczona została wartość współczynnika korelacji liniowej Pearsona na poziomie 0,346, co pozwoliło na identyfikację dodatkowej zależności korelacyjnej między udziałem procentowym poparcia dla partii republikańskiej a udziałem wyborców powyżej 60 lat. Następnie dane przestrzenne odniesione do 99 jednostek zostały poddane

agregacji²³ dla pięciu osobnych układów jednostek terytorialnych na wyższym poziomie agregacji²⁴. Na podstawie zagregowanych danych przestrzennych policzono ponownie wartości współczynnika korelacji liniowej dla każdego z pięciu układów jednostek terytorialnych i otrzymano wartości z przedziału od 0,265 dla oficjalnego układu wyborczego stanu Iowa do 0,862 dla układu jednostek miejskich i wiejskich. Openshaw i Taylor argumentują, że jeżeli badacz jest zainteresowany uzyskaniem silnej zależności korelacyjnej, to powinien wśród dostępnych układów wyborczych preferować układ Partii Demokratycznej, ponieważ dla tego układu otrzymana została najwyższa wartość współczynnika korelacji ($r=0,627$). W przypadku układu Partii Republikańskiej otrzymana wartość współczynnika korelacji wyniosła 0,482, a dla oficjalnego układu wyborczego wartość współczynnika ukształtowała się na poziomie 0,265 (Openshaw, Taylor, 1979, s. 129–130). Autorzy podsumowują badanie komentarzem o identyfikacji problemu zmiennej jednostki odniesienia oraz stwierdzeniem, że analiza korelacyjna jest dyskusyjnym zadaniem ze względu na otrzymany zbiór różnych wartości współczynnika korelacji liniowej w zależności od przyjętego układu.

Następnie Openshaw i Taylor bazując na tych samych danych przestrzennych dla 99 jednostek administracyjnych stanu Iowa, zadają pytanie badawcze o możliwe do uzyskania najmniejsze i największe wartości współczynnika korelacji liniowej Pearsona (Openshaw, Taylor, 1979, s. 130–131). Stanowi to konsekwencję faktu otrzymania przez autorów różnych wartości współczynnika korelacji dla pięciu rozważanych wcześniej układów jednostek terytorialnych. Zgodnie z postawionym pytaniem, Openshaw i Taylor określają procedurę wyznaczania granic przedziału wartości statystyk lub ocen parametrów, jako identyfikację granic dla problemu sposobu podziału przestrzeni oraz problemu skali. Fakt ten pokazuje niebezpieczeństwo polegające na tym, iż w wyniku świadomego przekroczenia zasad obowiązującego paradygmatu naukowego, otrzymane przez autorów wyniki prowadzą do kolejnych propozycji metodycznych (w tym wypadku do identyfikacji granic problemu zmiennej jednostki odniesienia).

W celu empirycznej identyfikacji granic przedziału autorzy proponują wykorzystanie algorytmu optymalizacyjnego, który polega na wykorzystaniu

²³ Autorzy podają, że agregacja polegała na policzeniu średnich na podstawie danych odniesionych do jednostek terytorialnych na niższym poziomie agregacji (Openshaw, Taylor, 1979, s. 129).

²⁴ Były to 3 układy jednostek terytorialnych związanych z procesem wyborczym w Stanach Zjednoczonych: administracyjny układ okręgów wyborczych, układ okręgów wyborczych zaproponowany przez partię Demokratyczną oraz przez Partię Republikańską oraz 2 układy jednostek terytorialnych związanych z rozwojem regionalnym, układ regionów miejskich i wiejskich oraz układ jednostek terytorialnych oparty na zależnościach funkcyjnych regionów. Wszystkie układy składały się z 6 jednostek terytorialnych. Problemem podczas wyznaczania wartości współczynnika korelacji jest również niska liczba stopni swobody.

systemu losowego generowania układów jednostek terytorialnych do optymalizacji zadanej przez badacza funkcji celu²⁵ (Openshaw, Taylor, 1979, s. 130–131). Bazując na danych dla układu 99 jednostek administracyjnych stanu Iowa, autorzy generowali zbiór losowych układów na kolejnych dwunastu poziomach agregacji (układy 72, 66, 60, 54, 48, 42, 36, 30, 24, 18, 12 oraz 6 jednostek), a następnie agregowali dane i wyznaczyli wartości współczynnika korelacji liniowej Pearsona. W pierwszym kroku do generowania układów jednostek terytorialnych zastosowany został *zoning system*, gdzie na podstawie otrzymanych wyników ustalona została minimalna oraz maksymalna wartość współczynnika korelacji liniowej. W drugim kroku do ustalenia minimalnych oraz maksymalnych wartości współczynnika zastosowany został *grouping system*. W wyniku zastosowania *zoning system* otrzymano następujące granice przedziałów: $(-0,059; 0,703)$ dla układu 72 jednostek oraz $(-0,999; 0,999)$ dla układu 6 jednostek. Natomiast w ramach zastosowania *grouping system* uzyskano granice przedziałów o szerszym rozstępie: $(-0,579; 0,927)$ dla układu 72 jednostek oraz $(-0,999; 0,999)$ dla układu 6 jednostek²⁶. Należy podkreślić, że w wyniku przeprowadzonych badań, autorzy uzyskali właściwie cały zakres potencjalnych wartości współczynnika korelacji liniowej Pearsona, od wartości $-0,999$ do wartości $0,999$. Granice przedziałów wartości współczynnika korelacji liniowej były znacznie większe przy wykorzystaniu *grouping system*, szczególnie dla wyższych poziomów agregacji. Na koniec Openshaw i Taylor stwierdzają, że identyfikacja na kolejnych poziomach agregacji różnych granic przedziałów wymaga podjęcia alternatywnych interpretacji zależności korelacyjnej dla każdego poziomu agregacji z osobna.

W kolejnym kroku analizy Openshaw i Taylor podejmują kwestię interpretacji otrzymanych wyników. Interpretują oni minimalne oraz maksymalne wartości współczynnika korelacji liniowej jako granice empirycznych rozkładów tego współczynnika dla alternatywnych poziomów agregacji. W związku z tym autorzy sugerują podjęcie próby ustalenia parametrów empirycznego rozkładu współczynnika korelacji liniowej (zob. Openshaw, Taylor, 1979, s. 131–134). W tym celu Openshaw i Taylor zastosowali *zoning system* oraz *grouping system*, gdzie dla każdego z przyjętych poziomów agregacji (odpowiednio układów 66, 60, 55, 48, 42, 36, 30, 24, 18, 12 oraz 6 jednostek) generowano losowo po 10000 układów. Następnie dla każdego układu jednostek terytorialnych dokonano agregacji danych przestrzennych i wyznaczano wartości

²⁵ Propozycja budowy oraz zasady działania algorytmu optymalizacyjnego przedstawione zostały w pracach Openshaw (1977 a, b, c). W wyniku zastosowania algorytmu optymalizacyjnego zakładana jest funkcja celu (na przykład minimalna wartość współczynnika korelacji), a następnie na podstawie *zoning* lub *grouping system* generowany jest układ jednostek terytorialnych pozwalający na optymalizację funkcji celu.

²⁶ Zarówno w przypadku *zoning system*, jak i dla *grouping system*, układy jednostek terytorialnych pozwalające na optymalizację funkcji celu posiadały nierównomierną powierzchnię obszarów. Charakter takich układów przedstawiony został na rysunkach 3.2 oraz 3.3.

współczynnika korelacji liniowej Pearsona. W ten sposób dla każdego poziomu agregacji otrzymano empiryczny rozkład 10000 tysięcy wartości współczynnika korelacji. Otrzymane wyniki wskazały na większe rozproszenie rozkładów dla wyższych poziomów agregacji (Openshaw, Taylor, 1979, s. 131–132). Następnie autorzy dokonują porównania rozkładów parametru korelacji liniowej otrzymanych w wyniku zastosowania obydwu systemów i stwierdzają wyższy poziom rozproszenia rozkładu uzyskanego na podstawie *grouping system*. W przypadku losowo generowanych układów o 72 jednostkach terytorialnych, wartość odchylenia standardowego wynosiła 0,051 (*zoning system*) i 0,066 (*grouping system*), a dla układów o 6 jednostkach terytorialnych otrzymano wartości odchylenia standardowego na poziomie 0,218 (*zoning system*) oraz 0,408 (*grouping system*). Jako jeden z powodów otrzymania takich wyników podawany jest potencjalny wpływ autokorelacji przestrzennej, która została zidentyfikowana na podstawie analizowanych danych przestrzennych²⁷.

3.2.6. Potencjalne scenariusze badań nad problemem zmiennej jednostki odniesienia

W poprzednim podrozdziale wskazane zostało, że w wyniku zastosowania systemu losowego generowania układów jednostek terytorialnych Openshaw i Taylor otrzymywali przedział wyznaczanych ocen badanego parametru o szerokim rozstępie. Skutkowało to każdorazowo stwierdzeniem o identyfikacji problemu zmiennej jednostki odniesienia. W aktualnym podrozdziale pokazane zostanie, że próba znalezienia przyczyn zbyt szerokiego rozstępu dla otrzymanego rozkładu prowadzi Openshawa i Taylora do propozycji kolejnych rozwiązań metodycznych. Zastosowanie nowych rozwiązań nie przynosi niestety oczekiwanego rezultatu, ponieważ wielkość rozstępu wynika z błędnego podejścia losowego generowania wielu układów jednostek terytorialnych, gdzie odniesione do wyznaczonych układów dane przestrzenne nie posiadają własności jednorodności przyczynowej. W proponowanych rozwiązaniach testowany jest również wpływ własności procesów na otrzymywany rozkład ocen, co przenosi najczęściej ciężar badań na wykorzystanie symulowanych danych przestrzennych.

Z przykładowych rozwiązań metodycznych, jako pierwsze zaprezentowane zostanie badanie wpływu autokorelacji przestrzennej na rozkład otrzymywanych wyników. W celu sprawdzenia potencjalnego wpływu autokorelacji przestrzennej na własności rozkładu wartości współczynnika korelacji liniowej Pearsona, Openshaw i Taylor generowali dla układu 99 jednostek terytorialnych losowa dane przestrzenne z rozkładu normalnego w dwóch wariantach. W wariantcie pierwszym generowano realizacje dwóch procesów przestrzennych,

²⁷ Statystyka Morana obliczona na podstawie danych odzwierciedlających udział osób powyżej 50 lat wynosiła 0,37, a na podstawie danych odzwierciedlających wyniki wyborcze 0,43. Otrzymane wartości statystyk wskazują na występowanie dodatniej autokorelacji przestrzennej (Openshaw, Taylor, 1979, s. 132).

charakteryzujących się brakiem skośności, poziomem koncentracji zgodnym z rozkładem normalnych, brakiem autokorelacji przestrzennej oraz wartością współczynnika korelacji liniowej równą 0,346²⁸. W wariancie drugim generowano również dwie realizacje procesów przestrzennych, które charakteryzowały się występowaniem silnej, dodatniej autokorelacji przestrzennej.

Następnie przyjęto tylko dwa poziomy agregacji, układ 72 jednostek terytorialnych oraz układ 6 jednostek terytorialnych. Dla każdego z poziomów agregacji wygenerowano losowo po 10000 układów, osobno przy zastosowaniu *zoning system* oraz *grouping system*. Dla każdego z układów jednostek terytorialnych dokonano agregacji danych przestrzennych i obliczono wartości współczynnika korelacji liniowej. Otrzymane wyniki pozwoliły na wyznaczenie empirycznych rozkładów współczynnika dla danych symulowanych. Analiza wyznaczonych rozkładów pozwoliła na stwierdzenie wyższego rozproszenia otrzymywanych ocen w przypadku realizacji procesów charakteryzujących się autokorelacją przestrzenną. Oznacza to istotny wpływ dodatniej autokorelacji przestrzennej na zwiększenie rozproszenia rozkładu współczynnika korelacji liniowej Pearsona (zob. Openshaw, Taylor 1979, s. 138). Należy jednak podkreślić, że w obydwu sytuacjach wielkość rozstępu rozkładu współczynnika korelacji nie pozwalała na ustalenie wiarygodnych wniosków.

Jako drugie przykładowe rozwiązanie metodyczne, zaprezentowane zostanie badanie dotyczące identyfikacji rozkładu współczynnika korelacji liniowej przy nałożonych ograniczeniach (Openshaw, Taylor, 1979, s. 135–137)²⁹. W tym celu autorzy wykorzystali algorytm optymalizacyjny, w ramach którego system losowego generowania układów jednostek terytorialnych zastosowany został do optymalizacji zadanej funkcji celu. Zgodnie z funkcją celu minimalizowany był kwadrat różnicy między założoną wartością różnicy w stracie wariancji ΔS^D a wartością różnicy w stracie wariancji ΔS^D wyznaczoną na podstawie zagregowanych danych przestrzennych³⁰. Za pomocą tego badania Openshaw i Taylor postanowili sprawdzić możliwy wpływ poziomu różnicy w stracie wariancji ΔS^D na rozkład współczynnika korelacji liniowej.

Autorzy przyjęli trzy poziomy agregacji wyższego rzędu, dla których generowane losowo układy składały się z 24, 12 oraz 6 jednostek terytorialnych.

²⁸ Taka sama wartość współczynnika korelacji liniowej Pearsona otrzymana została dla empirycznych danych przestrzennych odniesionych do układu 99 jednostek terytorialnych stanu Iowa.

²⁹ Ponownie wykorzystane zostały empiryczne dane przestrzenne odniesione do układu 99 jednostek terytorialnych Iowa.

³⁰ Openshaw i Taylor definiują różnicę w stracie wariancji ΔS^D jako różnicę pomiędzy stopniem straty wariancji dla procesu objaśnianego Y oraz stopniem straty wariancji procesu objaśniającego X, co określić można jako $\Delta S^D = \Delta S_x^B - \Delta S_y^B$ (Openshaw, Taylor, 1979, s. 135). Z kolei stopień straty wariancji określony został jako $\Delta S_x^B = \frac{S - S^B}{S^B}$, gdzie S jest wartością wariancji całkowitej, a S^B jest wartością wariancji międzygrupowej.

Następnie zakładano kolejne wartości różnicy w stracie wariancji ΔS^{D^*} od -1 do 1 i generowane były układy jednostek terytorialnych, które pozwalały na optymalizację funkcji celu. W ostatnim kroku na podstawie zbioru wartości współczynnika korelacji liniowej, wyznaczano dla kolejnych wartości różnicy w stracie wariancji ΔS^{D^*} postać rozkładów. Ostatecznie dla każdej, założonej wartości różnicy w stracie wariancji ΔS^{D^*} otrzymany został szeroki przedział wartości współczynnika korelacji liniowej. Otrzymane wyniki nie pozwoliły na ustalenie systematycznych zmian w rozkładzie współczynnika pod wpływem zmiany wartości ΔS^D (zob. Openshaw, Taylor, 1979, s. 136–137). Podsumowując omówione badanie należy zwrócić uwagę, że przykład ten stanowi kolejne rozwinięcie proponowanej metodyki, gdzie do optymalizacji funkcji celu wprowadzana jest kolejna statystyka w postaci różnicy w stracie wariancji ΔS^D .

Należy stwierdzić, że w świetle przedstawionych przez Openshawa i Taylora wyników badań uzasadniona jest potrzeba przeprowadzonej przez autora krytycznej dyskusji w tym zakresie. Wynika to z faktu, że Openshaw i Taylor przedstawiają wiele propozycji metodycznych, które mogą być bezkrytycznie stosowane w kolejnych pracach naukowych. Ma to istotne znaczenie, ponieważ prezentowane w pracach Openshaw, Taylor (1979), Openshaw (1984b) podejście metodyczne do problemu zmiennej jednostki odniesienia jest powszechnie wykorzystywane w publikowanych pracach naukowych z tego zakresu (zob. Arbia, 1989; Dudley, 1991; Fotheringharn, Wong, 1991; Amrhein, 1995; Reynolds, 1998; Dark, Bram, 2007; Flowerdew, 2011). Autorzy kolejnych badań, powielając zaproponowaną przez Openshawa i Taylora metodykę dotyczącą problemu zmiennej jednostki odniesienia mogą nie mieć świadomości, że problem ten został w części sztucznie wykreowany na potrzeby forsowanej zmiany paradygmatu badań naukowych. Dodatkowo autorzy proponują nowe rozwiązania metodyczne, w których niestety wykorzystują system losowego generowania układów jednostek terytorialnych oraz algorytmu optymalizacyjny do testowania wpływu założonych ograniczeń.

3.2.7. Metodyka badań naukowych w ramach nowego paradygmatu nauk empirycznych

W niniejszym podrozdziale przedstawiona zostanie metodyka badań naukowych zaproponowana przez Openshawa w ramach nowego paradygmatu nauk empirycznych (Openshaw, 1984b, s. 34). Openshaw był przekonany, że zaproponowana przez niego nowa metodologia badań przestrzennych zastąpi obowiązujący paradygmat nauk empirycznych. W pracy Openshaw (1984b, s. 34–35) zaprezentowano pięć etapów badań oraz przedstawiono przykład badania wykonanego zgodnie z nową metodologią. Kolejne etapy badań przestrzennych przedstawiono w tabeli 3.2.

Tabela 3.2.

Etapy badań przestrzennych zgodnie z nowym paradygmatem nauk empirycznych

Etapy
Etap 1. Przyjęcie hipotezy badawczej mówiącej o wynikach pożądanych przez badacza
Etap 2. Analiza właściwa przy zastosowaniu <i>zoning system</i> lub <i>grouping system</i> , uzupełniona przez wykorzystanie uniwersalnego algorytmu.
Etap 3. Weryfikacja postawionej hipotezy badawczej lub jej odrzucenie
Etap 4. Przyjęcie dodatkowych ograniczeń poprawiających jakość wykonanych badań
Etap 5. Weryfikacja hipotezy przy dodatkowych ograniczeniach

Źródło: opracowanie własne.

Zgodnie z propozycją nowego paradygmatu nauk empirycznych, w etapie pierwszym określona powinna zostać w jasny sposób hipoteza badawcza. Zmiana podejścia polega na tym, że rozważane mogą być dwie kwestie, jaki wynik jest oczekiwany na podstawie dostępnej wiedzy lub jaki wynik jest pożądany przez badacza. Openshaw podaje następujący przykład hipotezy badawczej mówiącej, że ustalona zależność korelacyjna (korelacja między wybraną determinantą a wynikiem wyborczym Partii Republikańskiej) dla stanu Nevada będzie też obowiązywała w stanie Iowa (Openshaw, 1984b, s. 34).

W etapie drugim badań przestrzennych wykonywana jest analiza właściwa na podstawie empirycznych danych przestrzennych lub generowanych realizacji procesów o wybranych własnościach. Celem analizy właściwej jest uzyskanie, założonych w hipotezie badawczej, pożądanych wyników. Najważniejszym elementem etapu drugiego jest zastosowanie *zoning system* lub *grouping system* do losowego generowania zbioru układów jednostek terytorialnych, w wyniku czego otrzymany zostanie rozkład rozpatrywanych w hipotezie badawczej statystyki lub parametru. Dodatkowo dopuszczalne jest zastosowanie algorytmu optymalizacyjnego, jeżeli wystąpi potrzeba minimalizacji funkcji celu lub wprowadzenia warunków ograniczających. Openshaw proponuje wykorzystanie algorytmu optymalizacyjnego w celu poprawy wiarygodności otrzymywanych wyników. Przykładowo, możliwe jest przyjęcie funkcji celu minimalizującej sumę kwadratów reszt modelu regresji, przy założeniu warunku ograniczającego, że suma kwadratów reszt musi być mniejsza lub równa od założonej wartości granicznej. W ten sposób w wyniku wykonania analizy właściwej otrzymywany byłby rozkład współczynnika korelacji liniowej, gdzie dla wszystkich losowo generowanych układów jednostek terytorialnych otrzymano wartość sumy kwadratów reszt spełniającą warunek ograniczający.

Wykonanie analizy właściwej pozwala na przejście do etapu trzeciego, gdzie następuje potwierdzenie postawionej hipotezy badawczej lub jej odrzucenie. Jeśli założone w hipotezie zerowej pożądane wyniki charakteryzują się akceptowalnym poziomem wiarygodności hipoteza jest potwierdzona. Akceptowalny

poziom wiarygodności jest rozumiany przez Openshawa jako zbiór warunków, dla których pożądane wyniki w hipotezie zerowej można uznać za wiarygodne. Na przykład, uzyskanie pożądanej w hipotezie zerowej wartości współczynnika korelacji można uznać za wiarygodne, jeżeli wyznaczony rozkład współczynnika posiada maksimum blisko zakładanej wartości oraz charakteryzuje się akceptowalnym poziomem rozproszenia. W przeciwnym wypadku należy odrzucić hipotezę zerową lub przejść do etapu pierwszego i sformułować hipotezę badawczą w inny sposób. Openshaw zakłada, że przy akceptowalnym poziomie wiarygodności, który pozwala na weryfikację hipotezy, wciąż może być naruszonych szereg własności statystycznych. W związku z tym proponuje przyjęcie dodatkowych ograniczeń i przejście do etapu czwartego.

W etapie czwartym następuje sprawdzenie, czy przyjęcie dodatkowych ograniczeń przyczyni się do poprawy wiarygodności otrzymanych wyników w postaci rozkładu badanej statystyki lub parametru. W etapie tym wykonywana jest ponownie analiza właściwa przy wykorzystaniu algorytmu optymalizacyjnego. Ponownie wyznaczany jest rozkład empiryczny, jednak przy kolejnym, dodatkowym ograniczeniu. Jako przykład dodatkowych ograniczeń Openshaw proponuje przyjęcie warunku braku autokorelacji przestrzennej.

W ostatnim, piątym etapie następuje weryfikacja hipotezy przy dodatkowych ograniczeniach. Jeżeli nowe wyniki charakteryzują się również akceptowalnym poziomem wiarygodności weryfikacja hipotezy kończy się jej potwierdzeniem. W przeciwnym wypadku należy zastanowić się, co jest przyczyną obniżenia poziomu wiarygodności otrzymanych rezultatów. Według Openshawa niepowodzenie związane z wprowadzeniem dodatkowych ograniczeń również stanowi cenne źródło informacji dla wykonywanych badań (Openshaw, 1984b, s. 35).

Na poparcie zasadności zastosowania pięciu etapów badania przestrzennego zgodnie z nowym paradygmatem nauk empirycznych Openshaw prezentuje przykład badania empirycznego (Openshaw, 1984b, s. 35–37). W hipotezie badawczej założona została pożądana wartość współczynnika korelacji liniowej Pearsona między wiekiem wyborców a wynikami wyborczymi Partii Republikańskiej na poziomie 0,8 dla stanu Iowa. W badaniu wykorzystane zostały empiryczne dane przestrzenne dotyczące wyników wyborów w tym stanie. Na podstawie *zoning system* generowano losowo kolejne układy jednostek terytorialnych, w ramach których wyznaczano wartość współczynnika korelacji oraz szacowano parametry modelu regresji. Następnie sformułowane zostały ograniczenia dla wybranych własności modelu regresji w postaci braku autokorelacji przestrzennej reszt modelu, średniej dla reszt równej zero oraz braku korelacji między resztami a zmienną objaśniającą.

Przy tak przyjętych założeniach wykonane zostały obliczenia, polegające na wykorzystaniu algorytmu optymalizacyjnego pozwalającego na wybór tylko tych układów, dla których spełnione są założone ograniczenia. W wyniku zastosowania algorytmu optymalizacyjnego uzyskany został rozkład współ-

czynnika korelacji z przedziału $-0,928$ do $0,993$ (przy spełnionych założeniach co do własności modelu regresji)³¹. Openshaw podkreśla niemożliwość wyciągnięcia wiarygodnych wniosków ze względu na zbyt dużą wartość rozstępu dla współczynnika korelacji, jednak pomimo tego stwierdza, że nie ma wątpliwości co do statystycznej istotności współczynnika korelacji.

W takiej sytuacji Openshaw proponuje wprowadzenie dwóch niezależnych ograniczeń. Pierwsze ograniczenie polegało na tym, że jednostki terytorialne nie mogły się różnić pod względem liczby ludności o więcej niż 15% od średniej. Dla takiego ograniczenia uzyskano wartości współczynnika korelacji z przedziału od $0,28$ do $0,94$. W drugim ograniczeniu założono, że jednostki terytorialne nie mogły się różnić pod względem powierzchni o więcej niż 15% od średniej. W tej sytuacji otrzymano wartości współczynnika korelacji z przedziału od $-0,06$ do $0,81$. Openshaw stwierdza jednak, że nie ma żadnych argumentów dla uzasadnienia zaproponowanych ograniczeń.

W tej sytuacji proponuje, że najlepszym rozwiązaniem będzie znalezienie innego zbioru empirycznych danych przestrzennych, dla którego otrzymano by wartość współczynnika korelacji liniowej Pearsona na poziomie $0,8$. Możliwe jest wówczas zastosowanie uniwersalnego algorytmu jednocześnie do obydwu zbiorów danych i zapożyczanie zidentyfikowanych własności na podstawie innego zbioru danych do ustalenia nowych warunków ograniczających dla analizy danych dotyczących stanu lowa. Na tej propozycji prezentacja empirycznego badania urywa się i Openshaw przechodzi do podsumowania całej pracy, gdzie podkreśla potrzebę odrzucenia obowiązującego paradygmatu nauk empirycznych oraz zalety nowego paradygmatu badań naukowych (Openshaw, 1984b, s. 37–38).

3.3. Rewizja problemu zmiennej jednostki odniesienia

3.3.1. Redefinicja problemu sposobu podziału przestrzeni

Zgodnie z zaprezentowanym przeglądem literatury dotyczącej problemu zmiennej jednostki odniesienia wskazany został fakt, że w większości prac, gdzie rozważany jest ten problem, układy jednostek terytorialnych wyznaczane są arbitralnie przez badacza lub w sposób losowy przez program komputerowy³², bez odniesienia do problemu badawczego. Ma to negatywny wpływ

³¹ Openshaw zakłada, że dostępne są dane przestrzenne na poziomie komisji wyborczych. W związku z tym możliwa jest agregacja danych dla losowych układów o dowolnej liczbie jednostek na obszarze stanu lowa. W ten sposób dla każdego losowego układu jednostek terytorialnych uzyskiwane są inne empiryczne dane przestrzenne. Otwartym pytaniem pozostaje jakość uzyskanych w ten sposób empirycznych danych przestrzennych.

³² Losowe generowanie układów jednostek terytorialnych w oparciu o wykorzystanie *zoning system* lub *grouping system*.

na poprawność wykonywanych przestrzennych badań ekonomicznych, ponieważ ustalenie w taki sposób pojedynczych układów jednostek terytorialnych nie gwarantuje własności jednorodności przyczynowej danych przestrzennych. Wykonywane analizy na podstawie losowych układów jednostek terytorialnych prowadzić będą do uzyskania niepoprawnych wniosków w ramach wykonywanych badań. Oznacza to, że problem zmiennej jednostki odniesienia rozważany był w literaturze przedmiotu w sposób niewłaściwy, ponieważ w przypadku analizy danych empirycznych, dopuszcza się układy jednostek terytorialnych niewchodzące w skład jednorodnego zbioru układów. Wskazuje to na potrzebę redefinicji określonego w literaturze problemu zmiennej jednostki odniesienia, w tym osobno problemu sposobu podziału przestrzeni oraz problemu skali.

Przeprowadzone dotychczas rozważania dostarczyły autorowi argumentów na ponowne rozważenie podstawowych definicji dla problemu zmiennej jednostki odniesienia. Propozycja redefinicji problemu sposobu podziału przestrzeni, jak i problemu skali oparta została na dwóch kwestiach: konieczności uwzględnienia podjętego przez badacza problemu badawczego oraz konieczności ustalenia jednorodnego zbioru układów.

Dzięki uwzględnieniu wskazanych kwestii problem sposobu podziału przestrzeni zostanie określony przez autora jako problem utworzenia pojedynczego układu jednostek terytorialnych na ustalonym przez badacza poziomie agregacji w taki sposób, by w ramach podjętego problemu badawczego należał on do jednorodnego zbioru układów.

Należy podkreślić, że w przestrzennych badaniach ekonomicznych zawsze występuje problem sposobu podziału przestrzeni, ponieważ zgodnie z postawionym problemem badawczym należy ustalić układ jednostek terytorialnych, do którego odniesione zostaną analizowane dane przestrzenne. Wybór odpowiedniego układu jednostek terytorialnych jest istotny, ponieważ warunkiem koniecznym wyciągnięcia poprawnych wniosków jest analiza danych przestrzennych, które charakteryzują się własnością jednorodności przyczynowej.

Tak jak zostało wskazane w rozdziale drugim, istnieją dwa scenariusze dla wystąpienia problemu sposobu podziału przestrzeni. W pierwszym scenariuszu badacz ma możliwość wyboru istniejącego układu jednostek terytorialnych, który wraz z danymi przestrzennymi utworzy jednorodny układ danych przestrzennych³³ w ramach postawionego problemu badawczego. Rozwiązanie problemu sposobu podziału przestrzeni polega na wyborze odpowiedniego układu.

W scenariuszu drugim dostępny apriori pojedynczy układ jednostek terytorialnych nie jest układem właściwym w świetle podjętego problemu badawczego albo układ jednostek terytorialnych nie został wcześniej ustalony i jest niedostępny. W takiej sytuacji rozwiązanie problemu sposobu podziału prze-

³³ Należący do jednorodnego zbioru układów pojedynczy układ jednostek terytorialnych wraz z odniesionymi do niego danymi przestrzennymi tworzy jednorodny układ danych przestrzennych.

strzeni polegać będzie na wyznaczeniu granic właściwego układu jednostek terytorialnych, który wraz z danymi przestrzennymi może zostać przyporządkowany do jednorodnego zbioru układów.

Ponieważ ustalenie właściwego układu jednostek terytorialnych jest ściśle związane z podjętym problemem badawczym, to w każdym przypadku wyjątkowo wiedza i doświadczenie naukowe badacza oraz wyniki wcześniejszych badań pozwolą na prawidłowe wyznaczenie granic obszarów i rozwiązanie problemu sposobu podziału przestrzeni. Oznacza to, że do rozwiązania problemu sposobu podziału przestrzeni nie można wykorzystać analizy symulacyjnej, w ramach której generowane byłyby realizacje procesów przestrzennych. Wykonana analiza symulacyjna nie wniesie w tym przypadku żadnych dodatkowych informacji, które mogłyby pomóc w rozwiązaniu problemu sposobu podziału przestrzeni.

3.3.2. Redefinicja problemu skali

Również w przypadku problemu skali zaproponowana zostanie redefinicja tego problemu, oparta na podstawie przeprowadzonych rozważań w rozdziale drugim oraz krytycznej ocenie literatury przedmiotu w niniejszym rozdziale. W omówionych pracach dotyczących problemu skali, własności procesów przestrzennych oraz zależności między nimi analizowane były na zbiorze arbitralnie przyjętych lub wygenerowanych losowo układów jednostek terytorialnych w ramach jednego poziomu agregacji. Oznacza to, że w prezentowanych w literaturze badaniach wyznaczano zbiór ocen parametrów na podstawie każdego z dowolnie³⁴ przyjętych układów. Ponieważ dane przestrzenne przyporządkowane do takich układów jednostek terytorialnych nie posiadały własności jednorodności przyczynowej, to w konsekwencji wyznaczane oceny parametrów musiały należeć do przedziału o dużym rozstępie. Otrzymywane w ten sposób wyniki nie mogły prowadzić do jednoznacznych wniosków. Dodatkową, negatywną konsekwencją w przypadku większości przeprowadzonych wcześniej badań, była identyfikacja problemu zmiennej jednostki odniesienia na podstawie znacznych różnic w wyznaczanych ocenach parametrów. Według autora, identyfikacja problemu zmiennej jednostki odniesienia w takiej sytuacji miała charakter pozorny.

Wskazane błędy metodyczne pokazują, że tak jak w przypadku problemu sposobu podziału przestrzeni, w ramach rozważania problemu skali konieczne jest również uwzględnienie podjętego przez badacza problemu badawczego oraz ustalenie jednorodnego zbioru układów. W związku z tym problem skali określony zostanie przez autora jako problem istotnych różnic w otrzymywanych ocenach własności procesów przestrzennych oraz łączących je zależności przyczynowych, gdy analiza przeprowadzona została na podstawie układów

³⁴ Dowolność tworzenia układów jednostek terytorialnych ma poważne konsekwencje w postaci niejednorodności przyczynowej danych przestrzennych odniesionych do tych układów.

jednostek terytorialnych przyporządkowanych w ramach podjętego problemu badawczego do jednorodnego zbioru układów. W zaproponowanej redefinicji problemu skali, własności procesów przestrzennych oraz zależności przyczynowe między nimi powinny być analizowane na wybranym przez badacza poziomie agregacji wyłącznie na podstawie jednego układu jednostek terytorialnych, gdzie przyporządkowane do tego układu dane przestrzenne posiadają własność jednorodności przyczynowej. Zmiana poziomu agregacji wymaga od badacza zmiany układu jednostek terytorialnych, jednak tylko w ramach układów należących do jednorodnego zbioru układów.

Zgodnie z zaproponowaną redefinicją można stwierdzić, że problem skali ma charakter empiryczny. Empiryczny charakter problemu skali rozumiany jest w tym sensie, że analiza różnic w otrzymywanych ocenach własności procesów przestrzennych oraz łączących je zależności przyczynowych musi korespondować z realnym problemem badawczym, postawionym w ramach prowadzonych przestrzennych badań ekonomicznych. W przeciwieństwie do problemu sposobu podziału przestrzeni, do badań nad problemem skali może być wykorzystana analiza symulacyjna. Zaproponowana redefinicja problemu skali wymusza jednak zmianę podejścia do stosowanej procedury symulacyjnej, w ramach której generowane byłyby realizacje procesów przestrzennych. Celem wykonania analizy symulacyjnej powinna być identyfikacja różnic w wyznaczanych ocenach parametrów, jednak tylko przy wyborze układów jednostek terytorialnych tworzących jednorodny zbiór układów. Oznacza to konieczność generowania danych przestrzennych na wybranym poziomie agregacji wyłącznie w ramach pojedynczego układu jednostek terytorialnych, który przyporządkowany został do jednorodnego zbioru układów. Zgodnie z zaproponowaną redefinicją problemu skali niedopuszczalne jest generowanie danych przestrzennych na jednym poziomie agregacji dla wielu układów jednostek terytorialnych (np. układów generowanych losowo w ramach *zoning system* lub *grouping system*). Takie podejście nie tylko nie prowadzi do rozwiązania problemu skali, ale wręcz odwrotnie uniemożliwia rozwiązanie tego problemu poprzez uzyskanie szerokiego zakresu niepoprawnych wartości, co wykazane zostało podczas krytycznego przeglądu literatury.

Rozdział 4.

Podział przestrzeni dla badań dotyczących rozwoju rolnictwa w Polsce

4.1. Rozwiązanie problemu sposobu podziału przestrzeni na poziomie makroregionów

4.1.1. Wyznaczenie jednorodnego zbioru układów

W treści podrozdziału poruszony zostanie problem sposobu podziału przestrzeni w przypadku przestrzennych badań ekonomicznych dotyczących rozwoju rolnictwa w Polsce na poziomie makroregionów. Problem ten polega na konieczności wyznaczenia właściwego układu jednostek terytorialnych, którego wykorzystanie zapewni poprawną analizę zjawisk dotyczących rolnictwa. Oznacza to, że dane przestrzenne przyporządkowane do takiego układu jednostek terytorialnych powinny charakteryzować się własnością jednorodności przyczynowej. Tylko w takiej sytuacji analizowane dane przestrzenne odzwierciedlą będą w poprawny sposób oddziaływanie przyczyn wpływających na zmiany zachodzące w rolnictwie. Podjęty problem sposobu podziału przestrzeni dla badań dotyczących rolnictwa w Polsce na poziomie makroregionów będzie można uznać za rozwiązany, jeżeli wyznaczony zostanie układ jednostek terytorialnych, który będzie można przyporządkować do jednorodnego zbioru układów.

Dodatkowo granice pojedynczych jednostek terytorialnych tworzących układ makroregionów rolniczych będą mogły być wykorzystane do utworzenia podobszarów wniosków w ramach prowadzonych badań. Wyznaczony w ten sposób podobszar wniosków przyczyni się do zmniejszenia prawdopodobień-

stwa popełnienia przez badacza błędu wyznaczenia obszaru wniosków¹. Przy założeniu stałego charakteru prawidłowości między procesami w ramach ustalonego podobszaru wniosków, analiza dotyczyć będzie mikrozależności przestrzennych. W przypadku przyjęcia podobszaru wniosków wykraczającego znacznie poza granice makroregionu, należy mieć świadomość, że identyfikowana prawidłowość ekonomiczna ma charakter makrozależności przestrzennej.

W związku z tym podjęta zostanie próba rozwiązania problemu sposobu podziału przestrzeni dla badań dotyczących rolnictwa w Polsce poprzez wyznaczenie granic makroregionów rolniczych dla 2002 roku². Specyfiką utworzonych makroregionów rolniczych³ powinno być to, że wyraźnie różnią się między sobą wartościami zmiennych dotyczących rolnictwa oraz są przestrzennie jednorodne ze względu na poziom rozwoju rolnictwa (zob. Skarżyńska, Goraj, Ziętek, 2005, s. 12–19). Do identyfikacji układu makroregionów rolniczych autor zastosował w pierwszym kroku analizę zróżnicowania przestrzennego struktury agrarnej w Polsce⁴. Charakterystyczną cechą polskiego rolnictwa w aspekcie struktury agrarnej jest rozdrobnienie gospodarstw rolnych oraz jej znaczne zróżnicowanie przestrzenne (Michna, 2007, s. 5–21). Można stwierdzić, że zróżnicowanie przestrzenne struktury agrarnej pozostało niezmiennie pomimo wielu działań polityki rolnej po roku 1990. Trwałość struktury agrarnej wynika m.in. z ekonomicznych, społecznych oraz historycznych czynników (zob. Pietrzak, Walczak, 2012, s. 214; Pietrzak, Walczak, 2013, s. 334–335; Pietrzak, Walczak, 2014, s. 1036–1038; Walczak, Pietrzak, 2016, s. 468–470). W przeszłości czynniki te warunkowały kształt struktury agrarnej w Polsce, utrwalając jej aktualną postać. Oznacza to, że analiza struktury agrarnej może wstępnie zastąpić wielowymiarową analizę zjawisk dotyczących rozwoju rol-

¹ Do uniknięcia błędu wyznaczenia obszaru wniosków konieczne jest jednak, by analizowane procesy przestrzenne posiadały odpowiednie własności dla stosowanych narzędzi badawczych.

² Wybór roku 2002 wynikał zarówno z dostępności danych dla tego roku, jak i merytorycznego uzasadnienia wyznaczenia układu makroregionów rolniczych w tym czasie. W przypadku danych przestrzennych pozwalających na ocenę struktury agrarnej na poziomie powiatów, dane dostępne były wyłącznie dla roku 2002. Merytoryczne uzasadnienie stanowiła konieczność rewizji układu makroregionów SGM przed 2004 rokiem, która wynikała z dostosowania rachunków dotyczących rolnictwa w Polsce do unijnego systemu zbierania danych rachunkowych z gospodarstw rolnych (zob. Skarżyńska, Goraj, Ziętek, 2005, s. 7, 16).

³ Urzędowy układ makroregionów rolniczych SGM został wyznaczony na podstawie narzędzi wielowymiarowej analizy porównawczej i zapisany w aneksie do Traktatu o przystąpieniu Rzeczypospolitej Polskiej do Unii Europejskiej, podpisanego w 2003 roku (zob. Skarżyńska, Goraj, Ziętek, 2005, s. 16).

⁴ Analizę przestrzennej zmienności struktury agrarnej w Polsce rozpoczęto w rozdziale drugim, gdzie do oceny tej struktury wykorzystana została koncentracja powierzchni użytków rolnych. Przeprowadzona analiza stanowi rozwinięcie badania z pracy Pietrzak (2014b), gdzie również na podstawie oceny zróżnicowania przestrzennego struktury agrarnej podjęto próbę wyznaczenia granic układu makroregionów rolniczych.

nictwa, ponieważ przestrzenne zróżnicowanie struktury agrarnej w Polsce powiązane jest w znacznym stopniu z przestrzennym zróżnicowaniem kultury rolnictwa, jak i poziomem jego rozwoju. Dodatkowo związana ze strukturą agrarną wielkość gospodarstw rolnych stanowi istotny element racjonalnego gospodarowania w rolnictwie. W Polsce wielkość gospodarstwa rolnego jest jedną z najważniejszych zmiennych determinujących poziom jego rozwoju, ponieważ gospodarstwa o małej powierzchni nie są w stanie wygenerować odpowiedniego poziomu dochodu, aby sprawnie funkcjonować (zob. Michna, 2007, s. 5–11). Zbyt duże rozdrobnienie gospodarstw oznacza brak możliwości postępu i zwiększenie jednostkowych kosztów produkcji. W związku z tym uzasadnione jest zastosowanie analizy struktury agrarnej w celu wyznaczenia makroregionów rolniczych, które będą jednorodne pod względem rozwoju i kultury rolnictwa.

Analiza rozpoczęta zostanie od ustalenia jednorodnego zbioru układów na podstawie istniejących układów jednostek terytorialnych, który może zostać wykorzystany w przestrzennych badaniach ekonomicznych dotyczących rolnictwa w Polsce. Do tego celu wykorzystane zostaną wstępnie układy jednostek terytorialnych NUTS. Na poziomie agregacji NUTS 5 większość danych przestrzennych dotyczących rozwoju rolnictwa w Polsce jest niedostępna. Oznacza to, że nie ma możliwości wykorzystania w badaniach układów danych przestrzennych na tym poziomie agregacji i układ jednostek terytorialnych NUTS 5 nie zostanie przyporządkowany do jednorodnego zbioru układów.

W przypadku poziomu agregacji NUTS4 statystyka publiczna udostępnia dane przestrzenne, które mogą zostać wykorzystane w badaniach nad zagadnieniami rolniczymi. Autor przyjmuje założenie, że odniesione do układu jednostek terytorialnych NUTS 4 dane przestrzenne dotyczące rolnictwa w Polsce charakteryzują się własnością jednorodności przyczynowej. W związku z tym układ jednostek terytorialnych NUTS 4 przyporządkowany zostanie do jednorodnego zbioru układów.

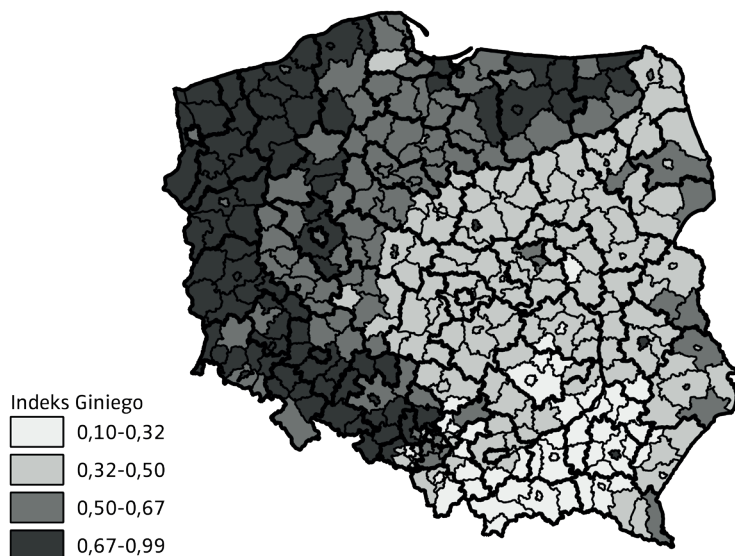
Po przyporządkowaniu układu NUTS 4 do jednorodnego zbioru układów, należy rozważyć kolejne układy na wyższych poziomach agregacji, które również mogą zostać przyporządkowane do tego zbioru. Kolejno rozważone zostaną układy takie jak: układ jednostek terytorialnych NUTS 3, układ jednostek terytorialnych NUTS 2 oraz układ jednostek terytorialnych NUTS 1. W związku z tym na rysunkach 4.1, 4.2 oraz 4.3 zaprezentowano przestrzenne zróżnicowanie koncentracji powierzchni użytków rolnych na poziomie powiatów⁵ oraz zaznaczono granice trzech układów jednostek terytorialnych na kolejnych poziomach agregacji. Wizualna ocena zmian w wartościach indeksu Giniego na rysunkach 4.1–4.3 pozwoliła na stwierdzenie znacznego zróżnicowania przestrzennego struktury agrarnej w Polsce. W związku z tym analiza prze-

⁵ Pomiaru koncentracji powierzchni użytków rolnych dokonano za pomocą indeksu Giniego. Do podziału wartości indeksu Giniego na cztery klasy zastosowano metodę podziału naturalnego.

strzennego zróżnicowania struktury agrarnej powinna pozwolić na ustalenie, które z rozważanych układów jednostek terytorialnych NUTS można przyporządkować do jednorodnego zbioru układów dla przestrzennych badań ekonomicznych dotyczących rozwoju rolnictwa w Polsce.

Rysunek 4.1.

Przestrzenne zróżnicowanie struktury agrarnej w układzie jednostek terytorialnych NUTS 3



Źródło: opracowanie własne.

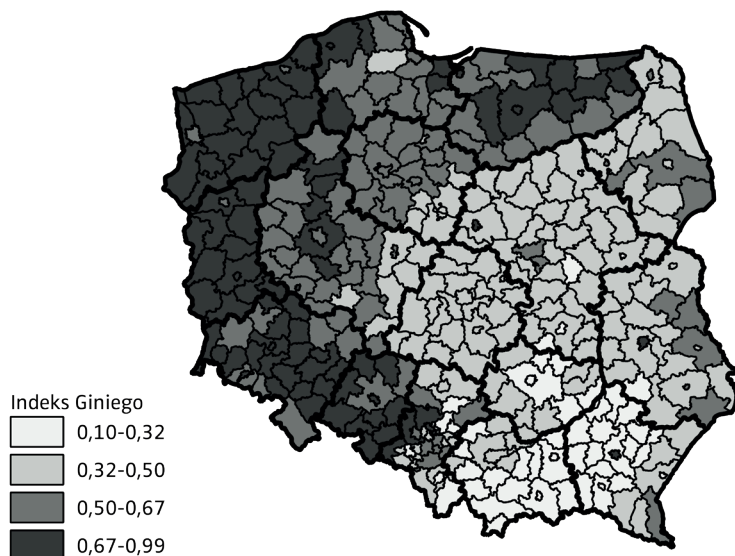
W wyniku analizy przestrzennego zróżnicowania struktury agrarnej dla układu jednostek terytorialnych NUTS 3 stwierdzono, że w granicach każdego z poszczególnych podregionów dominują powiaty przyporządkowane do jednej z czterech klas (zob. rysunek 4.1). Oznacza to, że każdy z podregionów charakteryzuje się w swoich granicach względnie stałą strukturą agrarną. Zaobserwowana dominacja powiatów z jednej klasy w każdym z podregionów pozwala na stwierdzenie, że dane przestrzenne odniesione do układu jednostek terytorialnych NUTS 3 powinny charakteryzować się własnością jednorodności przyczynowej. W związku z tym układ jednostek terytorialnych NUTS 3 również zostanie przyporządkowany do jednorodnego zbioru układów.

Podobna sytuacja ma miejsce w przypadku układu jednostek terytorialnych NUTS 2, gdzie także w większości województw dominują powiaty, które można przyporządkować do jednej z klas (zob. rysunek 4.2). Należy podkreślić, że dla niektórych województw występują odstępstwa w charakterze struktury agrarnej, jednak całościowo można uznać, że województwa są homogeniczne ze

względu na przestrzenne zróżnicowanie struktury agrarnej⁶. W związku z tym do jednorodnego zbioru układów, wraz z układem NUTS 4 oraz układem NUTS 3 przyporządkowany zostanie także układ jednostek terytorialnych NUTS 2.

Rysunek 4.2.

Przestrzenne zróżnicowanie struktury agrarnej w układzie jednostek terytorialnych NUTS 2



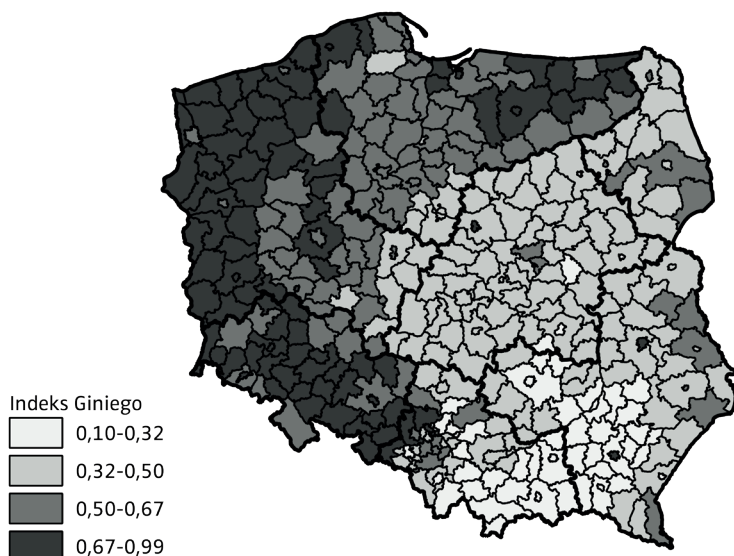
Źródło: opracowanie własne.

Natomiast w przypadku układu jednostek terytorialnych NUTS 1 widoczne są istotne zmiany charakteru struktury agrarnej w granicach makroregionów (zob. rysunek 4.3). W regionie wschodnim, powiaty należące do województwa podkarpackiego i świętokrzyskiego odróżniają się pod względem struktury agrarnej od powiatów z pozostałych województw ze względu na niższy poziom koncentracji użytków rolnych. Z kolei w regionie północno-zachodnim widoczny jest niższy poziom koncentracji użytków rolnych w powiatach województwa wielkopolskiego w porównaniu z powiatami z pozostałych województw. Natomiast w regionie północnym powiaty z województwa warmińsko-mazurskiego charakteryzują się wyższym poziomem koncentracji użytków rolnych. Oznacza to, że układ jednostek terytorialnych NUTS 1 nie może być przyporządkowany do jednorodnego zbioru układów, ponieważ dane przestrzenne odniesione do tego układu nie będą posiadały własności jednorodności przyczynowej. Wykonywane na jego podstawie przestrzenne badania ekonomiczne prowadziłyby do niepoprawnych wniosków, a w konsekwencji do błędnego rozwiązania postawionego problemu badawczego albo niemożliwości jego rozwiązania.

⁶ Odstępstwa te zostały omówione szczegółowo w rozdziale drugim.

Rysunek 4.3.

Przestrzenne zróżnicowanie struktury agrarnej w układzie jednostek terytorialnych NUTS 1



Źródło: opracowanie własne.

Przeprowadzone rozważania pozwoliły na wyciągnięcie wniosku, że wykorzystywany w przestrzennych badaniach ekonomicznych dotyczących rolnictwa w Polsce jednorodny zbiór układów należy utworzyć z układów jednostek terytorialnych NUTS 4, NUTS 3 oraz NUTS 2. Tak określony jednorodny zbiór układów⁷ powinien być wykorzystywany w badaniach dotyczących procesów związanych z rozwojem rolnictwa w Polsce, w tym w badaniach związanych z zagadnieniem zmian w strukturze agrarnej. Dane przestrzenne dotyczące różnych zjawisk w rolnictwie, które przyporządkowane zostaną do wymienionych układów jednostek terytorialnych, powinny pozwolić na utworzenie jednorodnych przyczynowo-układów danych przestrzennych, a ich analiza na wyciągnięcie poprawnych wniosków w ramach podjętego problemu badawczego.

Dodatkowo przeprowadzona analiza wskazała na konieczność utworzenia w miejsce układu jednostek terytorialnych NUTS 1 nowego układu makroregionów rolniczych o zmienionych granicach jednostek terytorialnych. Oznacza to wystąpienie opisanego w rozdziale drugim scenariusza drugiego dla problemu sposobu podziału przestrzeni, gdzie istniejący układ jednostek terytorialnych nie jest właściwy i potrzebne jest wyznaczenie granic nowego układu, który mógłby zostać przyporządkowany do jednorodnego zbioru układów. Jednostki

⁷ Rozważanie dotyczące przyporządkowania wybranych układów jednostek terytorialnych do jednorodnego zbioru układów zgodne są z wynikami uzyskanymi w rozdziale drugim.

terytorialne tworzące nowy układ makroregionów rolniczych powinny różnić się między sobą pod względem istotnych cech dotyczących rolnictwa i jednocześnie powinny być jednorodne wewnątrznie pod względem poziomu rozwoju rolnictwa. Należy podkreślić, że rozwiązanie problemu sposobu podziału przestrzeni na poziomie makroregionów rolniczych pozwoli na rozszerzenie ustalonego przez autora jednorodnego zbioru układów o nowy układ jednostek terytorialnych.

4.1.2. Wyznaczenie granic układu makroregionów rolniczych

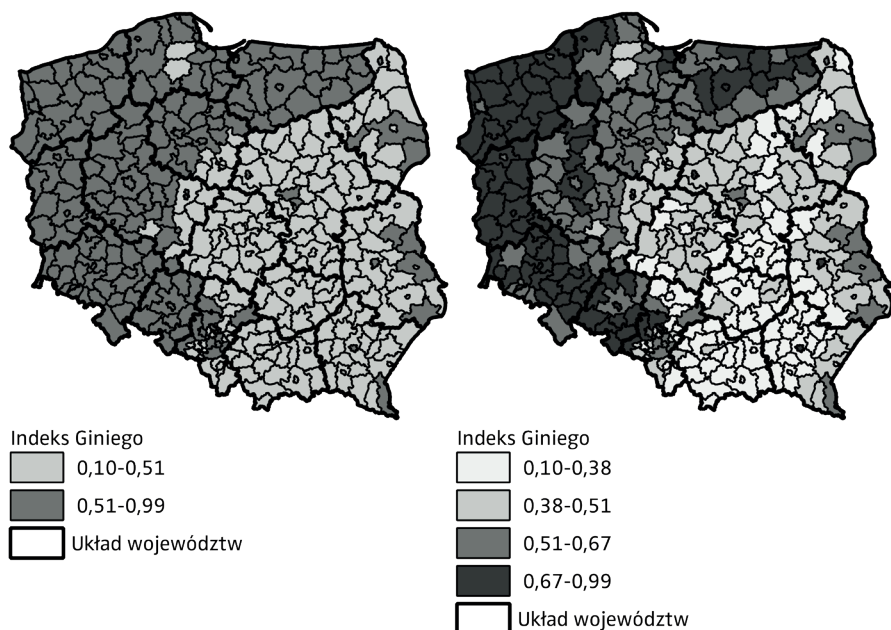
W poprzednim podrozdziale ustalono jednorodny zbiór układów dla przestrzennych badań ekonomicznych dotyczących rolnictwa w Polsce, do którego nie przyporządkowano układu na poziomie makroregionów (układ NUTS 1). W kolejnym kroku badań autor podjął próbę wyznaczenia granic nowego układu jednostek terytorialnych na poziomie makroregionów, w skład którego wchodziłyby cztery makroregiony rolnicze. Do ustalenia granic poszczególnych jednostek terytorialnych układu zastosowano analizę przestrzennego zróżnicowania struktury agrarnej, gdzie wykorzystane dane przestrzenne odniesione zostały do układu jednostek terytorialnych NUTS 4⁸. Dodatkowo przyjęte zostało założenie, że granice makroregionów rolniczych powinny być zgodne z granicami układu jednostek terytorialnych NUTS 2. Założenie to podyktowane jest faktem, że w ramach jednorodnego zbioru układów, jest to układ jednostek terytorialnych o najwyższym poziomie agregacji. Podsumowując, granice makroregionów wyznaczone zostaną na podstawie przestrzennego zróżnicowania struktury agrarnej na poziomie powiatów, ale będą musiały być zgodne z przebiegiem granic województw.

Na rysunku 4.4 przedstawiono zróżnicowanie przestrzenne struktury agrarnej poprzez przyporządkowanie wartości indeksu Giniego do dwóch, a następnie do czterech klas. Podział na klasy oparty został na wartościach kwartyli, gdzie do wyodrębnienia klas wykorzystano wartości mediany, kwartyla pierwszego i trzeciego. Natomiast na rysunku 4.5 przedstawiono kolejny podział koncentracji użytków rolnych na klasy, jednak powiaty zostały przyporządkowane do klas przy wykorzystaniu metody podziału naturalnego (Jenks, 1967).

⁸ Opis danych przestrzennych, na podstawie których wyznaczono wartości indeksu Giniego, zawarty jest w rozdziale drugim.

Rysunek 4.4.

Koncentracja użytków rolnych w Polsce: podział na klasy oparty na wartościach kwartyli

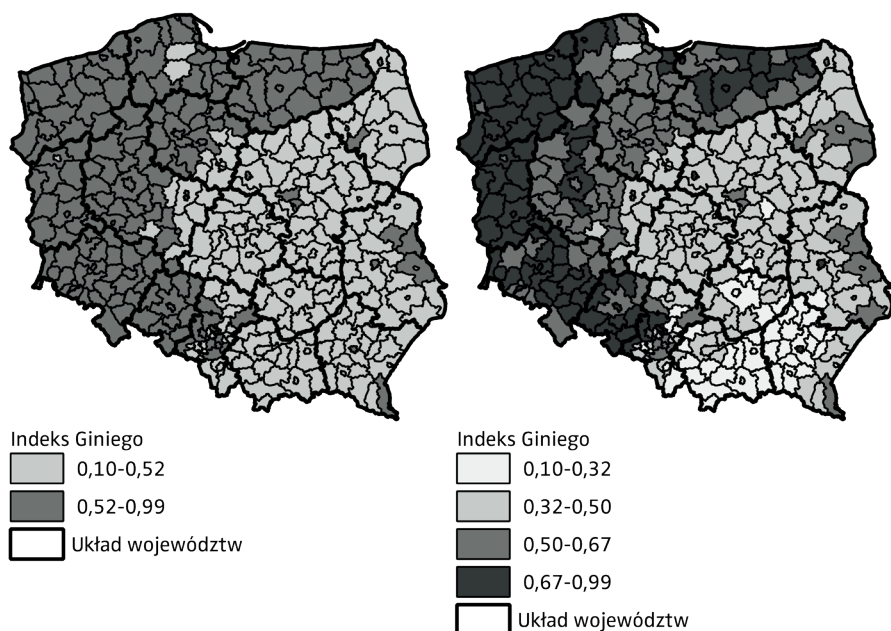


Źródło: opracowanie własne.

Wstępna analiza rysunków 4.4 oraz 4.5 pozwala na wyodrębnienie dwóch obszarów istotnie odmiennych pod względem charakteru struktury agrarnej. Pierwszy obszar nazwany przez autora makroregionem wschodnim może zostać utworzony z części wschodniej terytorium Polski, gdzie zaobserwować można niską koncentrację powierzchni użytków rolnych. Makroregion wschodni mogą utworzyć województwa: podlaskie, lubelskie, podkarpackie, małopolskie, świętokrzyskie, mazowieckie, łódzkie, jednak bez województwa warmińsko-mazurskiego. W porównaniu do pozostałych województw z części wschodniej terytorium Polski, w województwie warmińsko-mazurskim występuje wysoka koncentracja powierzchni użytków rolnych. Drugi obszar nazwany jako makroregion zachodni można wyodrębnić z części zachodniej terytorium Polski i charakteryzuje się on znacznie wyższym poziomem koncentracji użytków rolnych w porównaniu z makroregionem wschodnim. W skład makroregionu zachodniego mogą wejść województwa: zachodnio-pomorskie, pomorskie, lubuskie, kujawsko-pomorskie, wielkopolskie, dolnośląskie, opolskie oraz dodatkowo województwo warmińsko-mazurskie. Ustalone obszary: makroregion wschodni oraz makroregion zachodni utworzą wstępny układ dwóch makroregionów rolniczych, który zaprezentowany został na rysunku 4.6.

Rysunek 4.5.

Koncentracja użytków rolnych w Polsce: podział na klasy oparty na metodzie podziału naturalnego



Źródło: opracowanie własne.

Ocena przestrzennego zróżnicowania struktury agrarnej w ramach układu dwóch makroregionów rolniczych (makroregionu wschodniego oraz makroregionu zachodniego) wskazuje na konieczność dalszego podziału makroregionów ze względu na istotne zmiany w tej strukturze (zob. rysunek 4.4, 4.5, 4.6). W związku z tym, w obrębie każdego z dwóch makroregionów wyodrębnione zostały kolejne dwa podobszary, co pozwoliło na propozycję układu czterech makroregionów rolniczych dla Polski. Z makroregionu wschodniego wyodrębnione zostały makroregion M1 i makroregion M2, a z makroregionu zachodniego makroregion M3 oraz makroregion M4. Makroregion M1 utworzony został z województw: zachodnio-pomorskiego, lubuskiego, dolnośląskiego i opolskiego, gdzie dominują powiaty z czwartej klasy o najwyższym poziomie koncentracji użytków rolnych. Przebieg granicy makroregionu M2 wyznaczono na podstawie granic województw: pomorskiego, wielkopolskiego, kujawsko-pomorskiego oraz warmińsko-mazurskiego, gdzie dominują powiaty z trzeciej klasy. Z kolei z województw: mazowieckiego, łódzkiego, lubelskiego i podlaskiego, gdzie występuje przewaga liczebna powiatów z drugiej klasy, utworzony został makroregion M3. Ostatni makroregion M4 utworzono z województw małopolskiego, śląskiego, świętokrzyskiego oraz podkarpackiego,

ponieważ powiaty należące do tych województw przyporządkowane zostały w większości do pierwszej klasy o najniższym poziomie koncentracji użytków rolnych. W ten sposób ustalony został wstępny układ czterech makroregionów rolniczych M1, M2, M3, M4, których granice przedstawione zostały na rysunku 4.7.

Rysunek 4.6.

Podział Polski na dwa makroregiony rolnicze: makroregion wschodni oraz makroregion zachodni

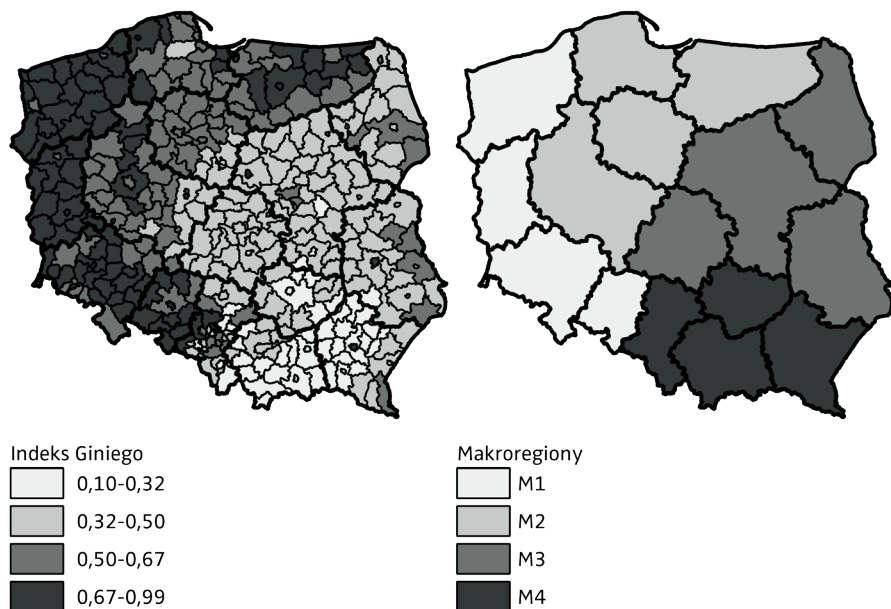


Źródło: opracowanie własne.

Po ustaleniu granic czterech makroregionów rolniczych, zdaniem autora powinna zostać przeprowadzona weryfikacja obszarów pod względem podobieństwa w poziomie rozwoju i kulturze rolnictwa. Celem takiej weryfikacji ma być zatwierdzenie ostatecznych granic układu jednostek terytorialnych i w razie konieczności ewentualna ich korekta. Zmiana przebiegu granic może być konieczna w przypadku braku wewnętrznej jednorodności obszarów ze względu na wybrane kryterium. Jest to ważny etap procesu rozwiązywania problemu sposobu podziału przestrzeni, ponieważ analizowane zjawiska w ramach przestrzennych badań ekonomicznych mają w większości charakter wielowymiarowy i nie jest łatwym zadaniem ustalenie obszarów jednorodnych przestrzennie ze względu na postawiony problem badawczy.

Rysunek 4.7.

Wstępny układ makroregionów rolniczych na tle przestrzennego zróżnicowania struktury agrarnej



Źródło: opracowanie własne.

W związku z tym, w ostatnim kroku badania autor dokonał weryfikacji granic wstępnego układu makroregionów rolniczych, który wyznaczony został wyłącznie w oparciu o analizę struktury agrarnej (zob. rysunek 4.7). W przypadku podjętego problemu wyznaczenia granic układu makroregionów rolniczych, weryfikacja została przeprowadzona na podstawie dodatkowej analizy województw pod względem poziomu rozwoju i kultury rolnictwa. W pracy Skarżyńska, Goraj, Ziętek (2005, s. 7–21) przeprowadzona została ocena sytuacji w rolnictwie w Polsce w okresie 2000–2002 przy zastosowaniu metod wielowymiarowej analizy porównawczej⁹. Do analizy wybrane zostały następujące zmienne diagnostyczne: wartość Produktu Krajowego Brutto na mieszkańca, udział rolnictwa w tworzeniu Produktu Krajowego Brutto, udział trwałych użytków zielonych w użytkach rolnych, średni plon zbóż, zużycie nawozów na 1 ha użytków rolnych, średnia roczna produkcja mleka od 1 krowy, śred-

⁹ Zastosowane metody wielowymiarowej analizy porównawczej powinny dodatkowo uwzględniać aspekty przestrzenne ze względu na występujące, silne zależności przestrzenne w rolnictwie. W pracach Antczak (2013), Pietrzak (2014d, 2016) zaproponowane zostały procedury budowy taksonomicznego miernika rozwoju, gdzie brana jest pod uwagę możliwość występowania w analizowanych procesach własności auto-korelacji przestrzennej.

nia powierzchnia użytków rolnych dla indywidualnego gospodarstwa rolnego, średnia liczba krów dojnych w indywidualnych gospodarstwach rolnych oraz średnia liczba trzody ogółem w indywidualnych gospodarstwach rolnych.

Przeprowadzona w pracy Skarżyńska, Goraj, Ziętek (2005, s. 14–15) analiza przestrzenna wskazała na dysproporcje w wartościach kluczowych procesów dotyczących działalności rolniczej, co w znacznej części koresponduje z przestrzennym zróżnicowaniem struktury agrarnej w Polsce (zob. rysunek 4.7). W makroregionie M1 województwa: dolnośląskie i opolskie różnią się pod względem rozwoju i kultury rolnictwa od województw: lubuskiego i zachodniopomorskiego. Również w przypadku makroregionu M2 występuje odmienność w charakterze rolnictwa pomiędzy województwami: wielkopolskim i kujawsko-pomorskim a województwami: pomorskim i warmińsko-mazurskim. Z obydwu makroregionów M1 i M2, województwa: lubuskie, zachodniopomorskie, pomorskie i warmińsko-mazurskie charakteryzuje najniższa intensywność produkcji rolniczej w kraju. W województwach tych występuje także najniższy udział użytków rolnych znajdujących się w posiadaniu rolników indywidualnych, co ma przełożenie na wysoką koncentrację powierzchni użytków rolnych. Należy jednak podkreślić wysoki udział zbóż w strukturze zasiewów oraz niską obsadę zwierząt w porównaniu z sytuacją innych województw w Polsce. Pozostałe cztery województwa wchodzące w skład makroregionów M1 i M2 województwa: dolnośląskie, kujawsko-pomorskie, opolskie i wielkopolskie są zbliżone do siebie pod względem charakteru rolnictwa. Produkcja rolna prowadzona jest w tych województwach w sposób bardzo intensywny, na co wskazuje najwyższe w Polsce zużycie nawozów mineralnych oraz wielkość zakupów pasz treściwych. Dodatkowo w regionie tym stwierdzony został najwyższy poziom kwalifikacji zawodowej rolników. W związku z tym autor dokonał przekształcenia granic makroregionów M1 i M2 zgodnie z charakterem poszczególnych województw. Zmienione granice makroregionu M1 obejmować będą województwa: lubuskie, pomorskie, zachodniopomorskie oraz warmińsko-mazurskie, a województwa: dolnośląskie, kujawsko-pomorskie, opolskie i wielkopolskie utworzą makroregion M2.

Ustalone wcześniej granice makroregionu M3 należy uznać za poprawne, ponieważ województwa podlaskie, lubelskie, mazowieckie i łódzkie są do siebie podobne pod względem poziomu rozwoju i kultury rolnictwa. Województwa te charakteryzują się średnią wielkością powierzchni gospodarstw rolnych. Makroregion M3 charakteryzuje się słabą jakością gleb oraz niskim poziomem nawożenia. W konsekwencji w kulturze rolnictwa należy podkreślić duży udział trwałych użytków zielonych i związaną z tym wysoką obsadę bydła. Ponadto w wymienionych województwach występuje duży udział zbóż, gdzie ze względu na jakość użytków rolnych uzyskiwane plony są niskie. Wszystko to składa się na niską intensywność produkcji rolniczej w regionie. Również w przypadku makroregionu M4 okazało się, że wykonana analiza przestrzennego zróżnicowania struktury agrarnej pozwoliła na poprawne wyznaczenie

granic tego regionu. Tworzące makroregion M4 województwa: podkarpackie, świętokrzyskie, małopolskie i śląskie charakteryzują się największym w Polsce rozdrobnieniem agrarnym. Gospodarstwa rolne charakteryzują się najmniejszą średnią powierzchnią użytków rolnych, gdzie obsada zwierząt jest najwyższa w Polsce. Województwa te charakteryzują się niską intensywnością produkcji, na co ma wpływ również niski poziom zużycia nawozów mineralnych oraz zakupów pasz treściwych.

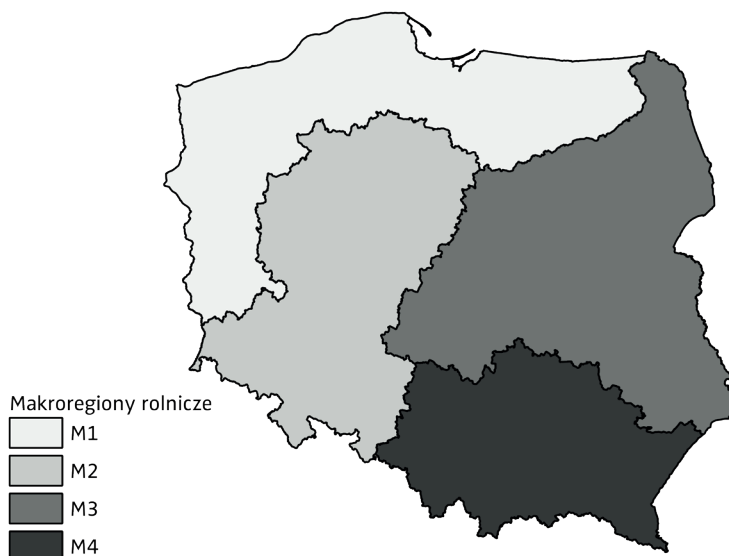
Przeprowadzona przez autora weryfikacja granic wstępnego układu makroregionów rolniczych wskazała na konieczność przebudowy przebiegu jego granic. Wynikało to z faktu, że na podstawie dodatkowej oceny województw pod względem podobieństwa w poziomie rozwoju i kulturze rolnictwa zidentyfikowano występowanie różnic w poszczególnych częściach makroregionów rolniczych. Należy jednak podkreślić, że ustalony na podstawie analizy charakteru struktury agrarnej, wstępny układ makroregionów rolniczych, nie wymagał istotnej korekty granic wszystkich makroregionów. Potwierdza to zasadność zastosowania analizy przestrzennego zróżnicowania struktury agrarnej do wyznaczenia wstępnych granic makroregionów rolniczych. Ostatecznie w wyniku wykonanej weryfikacji granic, przekształcone zostały obszary makroregionów rolniczych M1 i M2. Natomiast granice makroregionów rolniczych M3 i M4 nie wymagały korekty i pozostały niezmienione. Skorygowane na podstawie weryfikacji województw pod względem podobieństwa w poziomie rozwoju i kulturze rolnictwa granice układu makroregionów rolniczych przedstawione zostały na rysunku 4.8.

Wyznaczony na podstawie analizy koncentracji powierzchni użytków rolnych, a następnie zmieniony w wyniku weryfikacji pod względem podobieństwa w poziomie rozwoju i kulturze układ jednostek terytorialnych pokrywa się z oficjalnym układem makroregionów SGM, który został utworzony w związku z przystąpieniem Polski do Unii Europejskiej (zob. rysunek 4.8). Zgodność ta wynika w dużej mierze z powiązania pomiędzy strukturą agrarną, a poziomem rozwoju rolnictwa w regionach. Oznacza to, że wadliwa struktura agrarna staje się silną destymulantą właściwych przemian i rozwoju rolnictwa (zob. Michna, 2007, s. 5–13).

Utworzenie układu makroregionów rolniczych dla celów przestrzennych badań ekonomicznych dotyczących rolnictwa jest przykładem utworzenia nowego układu jednostek terytorialnych. Oznacza to, że przedstawiona procedura wyznaczenia granic układu jednostek terytorialnych doprowadziła do rozwiązania problemu sposobu podziału przestrzeni. Zarówno wykonywane przestrzenne analizy ekonomiczne, jak i prowadzona statystyka regionów na podstawie ustalonego układu makroregionów rolniczych powinny prowadzić do poprawnych wniosków z badań. Oznacza to, że rozwiązanie problemu sposobu podziału przestrzeni na poziomie makroregionów rolniczych pozwala na rozszerzenie jednorodnego zbioru układów o nowy układ jednostek terytorialnych, układ makroregionów rolniczych SGM.

Rysunek 4.8.

Układ makroregionów rolniczych M1, M2, M3, M4



Źródło: opracowanie własne.

Na podstawie przeprowadzonych rozważań można stwierdzić, że w przestrzennych badaniach ekonomicznych dotyczących rolnictwa w Polsce należy przyjmować jednorodny zbiór układów utworzony z następujących układów jednostek terytorialnych: układu NUTS 4, układu NUTS 3, układu NUTS 2 oraz układu makroregionów rolniczych SGM. Dodatkowo należy podkreślić, że prawidłowości między procesami dotyczącymi rozwoju rolnictwa powinny mieć stały charakter w obrębie granic poszczególnych makroregionów rolniczych. W przypadku analizy dla obszaru znacznie wykraczającego poza granice pojedynczego makroregionu należy spodziewać się zmiany charakteru badanej zależności ekonomicznej. Fakt ten powinien mieć przełożenie na decyzję badacza odnośnie do wyznaczenia granic podobszarów wniosków. Podobszary wniosków powinny być zgodne z granicami jednostek terytorialnych należących do układów przyporządkowanych do jednorodnego zbioru układów. W przypadku badań dotyczących rolnictwa w Polsce granice podobszarów wniosków powinny pokrywać się z granicami poszczególnych makroregionów rolniczych lub mniejszych jednostek terytorialnych należących do tych makroregionów (grupy wybranych województw, grupy podregionów, grupy powiatów, czy też pojedynczych województw, podregionów, powiatów). Badacz powinien mieć świadomość, że analiza prawidłowości ekonomicznych dla tak określonego podobszaru wniosków może pozwolić na identyfikację mikrozależności przestrzennych. Natomiast analiza prawidłowości ekonomicznych dla podobszaru wniosków wykraczającego poza granice pojedynczego makroregionu dopro-

wadzi do identyfikacji makrozależności przestrzennej, ze względu na wysokie prawdopodobieństwo zmiany charakteru badanej zależności. Należy podsumować, że procedura wyznaczenia nowego układu jednostek terytorialnych w ramach podjętego problemu badawczego, nie tylko pozwala na rozwiązanie problemu sposobu podziału przestrzeni, ale również dostarcza cennych informacji pozwalających na klasyfikację identyfikowanych zależności jako mikrozależności albo makrozależności przestrzennych. W związku z tym w następnym podrozdziale poruszona zostanie kwestia wykorzystania ustalonych mikrozależności przestrzennych do wyznaczenia przestrzennej makrozależności.

4.2. Relacja między mikroparametrami a makroparametrem na przykładzie badania makroregionów rolniczych SGM

4.2.1. Identyfikacja mikrozależności i makrozależności przestrzennych dotyczących zużycia nawozów mineralnych w rolnictwie

W podrozdziale 4.2 przeprowadzone zostaną rozważania na temat sformułowanego przez Pawłowskiego problemu wyznaczenia relacji między makroparametrem a mikroparametrami (Pawłowski, 1969, s. 235). W tym celu przeprowadzono przestrzenne badanie ekonomiczne dotyczące identyfikacji prawidłowości między wybranymi procesami w rolnictwie. Sformułowany został problem badawczy w postaci ustalenia, czy wielkość gospodarstwa rolnego determinuje poziom wykorzystania nawozów mineralnych w polskim rolnictwie. Hipoteza o istotnym wpływie wielkości gospodarstwa domowego na sposób gospodarowania uzasadniona jest faktem, że wielkość gospodarstwa rolnego stanowi istotny element generowania dochodu przez gospodarstwo. Biorąc pod uwagę znaczenie wielkości gospodarstwa w uzyskaniu dochodu z rolnictwa, dla wszystkich makroregionów rolniczych należy założyć pozytywny wpływ wielkości powierzchni gospodarstw rolnych na zużycie nawozów azotowych. Zmiana charakteru badanej zależności między makroregionami rolniczymi polegać będzie na istotnym wzroście lub spadku jej siły.

W makroregionach o niskiej koncentracji użytków rolnych dominują gospodarstwa o małej powierzchni użytków rolnych. Dla gospodarstw rolnych tego typu spodziewana jest identyfikacja wysokiego wpływu wielkości powierzchni użytków rolnych na poziom stosowanych nawozów mineralnych. Wynika to z faktu, że gospodarstwa o małej powierzchni, w wyniku powiększenia powierzchni użytków rolnych są w stanie wygenerować odpowiedni poziomu dochodu pozwalającego na zakup pasz, surowców, nawozów odpowiedniej jakości, czy też specjalistycznych maszyn rolniczych (przy wyjściowej, małej powierzchni użytków rolnych osiągnąć dochód był zbyt mały na wykonanie odpowiednich zakupów, zob. Michna, 2007, s. 11–21). Nowy poziom dochodu przekłada się wówczas na znaczny wzrost zakupów środków produkcji rolniczej, w tym nawozów mineralnych.

Natomiast rolnicy z makroregionu o wysokiej koncentracji użytków rolnych posiadają gospodarstwa o znacznie wyższej średniej powierzchni użytków rolnych. W takiej sytuacji wpływ zwiększenia powierzchni użytków rolnych na zużycie nawozów jest znacznie niższy w porównaniu z gospodarstwami z makroregionu o niższej koncentracji użytków rolnych. W makroregionie charakteryzującym się wysoką koncentracją użytków rolnych ze względu na wyższą, wyjściową wartość średniej powierzchni użytków rolnych, wzrost średniej powierzchni gospodarstwa nie przełoży się w tak wysokim stopniu na wzrost średniego zużycia nawozów mineralnych¹⁰.

W związku z ustalonym w poprzednim podrozdziale jednorodnym zbiorem układów dla zagadnień związanych z rozwojem rolnictwa, próba identyfikacji zależności ekonomicznych podjęta została dla obszaru całej Polski, a następnie dla kolejnych makroregionów rolniczych SGM. Przeprowadzone rozważania pozwalają na stwierdzenie, że dla obszaru Polski dochodzi do zmian charakteru prawidłowości ekonomicznych opisujących związki między procesami dotyczącymi rozwoju rolnictwa. Oznacza to, że ustalone dla obszaru Polski prawidłowości między procesami należy traktować jako makrozależności przestrzenne. Warto podkreślić, że identyfikacja makrozależności przestrzennych stanowi pierwszy krok większości przestrzennych badań ekonomicznych. Dopiero dalsza, pogłębiona analiza podjętego problemu badawczego prowadzi do ustalenia jednorodnych przyczynowo-układów danych przestrzennych oraz identyfikacji wewnętrznej struktury badanych procesów przestrzennych, co pozwala na zawężenie obszaru wniosków i dobór właściwych narzędzi badawczych.

W przeprowadzonym badaniu granice podobszarów wniosków zawężone zostaną do makroregionów rolniczych SGM. Wyznaczone ponownie zależności ekonomiczne dla podobszarów wniosków określić można jako mikrozależności przestrzenne. Należy podkreślić, że mikrozależności przestrzenne mają wyższą wartość poznawczą w porównaniu z makrozależnościami, ponieważ ich interpretacja gwarantuje wyciągnięcie poprawnych wniosków, co z kolei przekłada się na jakość rozwiązania podjętego problemu badawczego.

Przestrzenne badanie ekonomiczne wykonane zostało w dwóch etapach. W etapie pierwszym dokonano estymacji parametrów pięciu przestrzennych modeli ekonometrycznych opisujących wpływ wielkości gospodarstwa rolnego na poziom zużycia nawozów azotowych. Pierwsze cztery specyfikacje modelowe odniesione zostały do kolejnych makroregionów rolniczych SGM, a piąty model ekonometryczny dotyczył obszaru całej Polski. Otrzymane wyniki pozwoliły na wstępną identyfikację mikrozależności przestrzennych dla kolejnych makroregionów rolniczych oraz makrozależności przestrzennej dla obszaru Polski.

¹⁰ Należy podkreślić, że gospodarstwa rolne z makroregionów o wysokiej koncentracji użytków rolnych charakteryzuje, zarówno najwyższa średnia wartość powierzchni użytków, jak i najwyższe średnie zużycie nawozów mineralnych.

W drugim etapie badania, ustalona została relacja między mikroparametrami opisującymi zależność ekonomiczną między powierzchnią użytków rolnych gospodarstwa a wielkością zużywanych nawozów mineralnych w ramach gospodarowania a makroparametrem. W tym celu wykorzystane zostały wyznaczone w etapie pierwszym oceny mikroparametrów dla poszczególnych makroregionów rolniczych SGM. Realizacja drugiego etapu badania pozwoliła na określenie prawidłowości między mikrozależnościami ekonomicznymi a makrozależnością ekonomiczną dla procesów przestrzennych, co stanowi rozwinięcie problemu Pawłowskiego o aspekty przestrzenne.

Ponieważ badanie dotyczy problematyki rozwoju rolnictwa w Polsce, to zgodnie z wynikami otrzymanymi w poprzednim podrozdziale, do jednorodnego zbioru układów przyporządkowane zostały następujące układy jednostek terytorialnych: układ NUTS 4, układ NUTS 3, układ NUTS 2 oraz układ makroregionów SGM. Każdy pojedynczy obszar należący do wymienionych układów jednostek terytorialnych (w tym pojedynczy makroregion rolniczy) charakteryzuje się podobnym poziomem rozwoju rolnictwa oraz kultury rolniczej. Informacja ta ma istotne znaczenie dla podjętych przez autora badań. Pierwszą kwestią stanowi ustalenie podobszarów wniosków o maksymalnej powierzchni. Granice takiego podobszaru nie powinny wykraczać poza obszar pojedynczego makroregionu rolniczego SGM. Odniesienie przyjętej specyfikacji modelu regresji¹¹ do określonego w ten sposób podobszaru wniosków powinno pozwolić na identyfikację mikrozależności ekonomicznej.

Drugą kwestią stanowi możliwość założenia, że charakter badanej zależności między procesami nie zmienia się przy wyborze kolejnych jednostek terytorialnych, mieszczących się w granicach makroregionów rolniczych. Ma to konsekwencje podczas interpretacji zidentyfikowanych mikrozależności przestrzennych, które można odnieść zarówno do poszczególnych jednostek terytorialnych wchodzących w skład makroregionu (pojedyncze województwo, podregion czy powiat), do grupy jednostek terytorialnych, jak i do całego podobszaru wniosków (pojedynczy makroregion rolniczy).

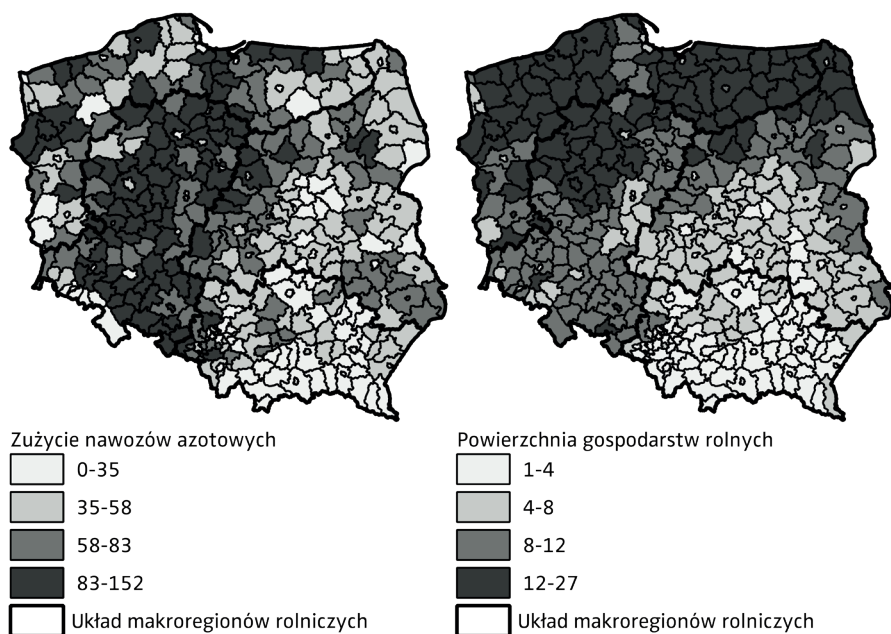
Na rysunkach 4.9 i 4.10 przedstawiono przestrzenne zróżnicowanie średniego zużycia nawozów azotowych na 1 ha oraz średniej powierzchni gospodarstw rolnych w ha na poziomie powiatów. Dane przestrzenne przyporządkowane zostały do czterech klas na podstawie wartości kwantyli. Wybór kwantyli, jako kryterium podziału, zapewnia możliwość porównania przestrzennego zróżnicowania przedstawionych danych. Dodatkowo na rysunku 4.9 przedstawiono granice makroregionów rolniczych SGM, a na rysunku 4.10 granice województw. Pozwoliło to na ocenę kształtowania się średniego zużycia nawozów azotowych oraz średniej powierzchni gospodarstw rolnych w granicach obydwu układów należących do jednorodnego zbioru układów. Wizualna ocena przestrzennego zróżnicowania obydwu procesów wskazuje

¹¹ Modelu zależności między wielkością gospodarstwa rolnego a poziomem stosowanych nawozów azotowych.

na fakt, że dla większości województw i makroregionów rolniczych (makroregionu Wielkopolska i Śląsk, makroregionu Mazowsze i Podlasie oraz makroregionu Małopolska i Pogórze) przyporządkowaniu średniej powierzchni gospodarstw rolnych do jednej z klas w wybranym powiecie odpowiada przyporządkowanie średniego zużycia nawozów mineralnych do tej samej klasy albo klasy sąsiadującej. Występują również różnice, na przykład dla województw lubuskiego, zachodnio-pomorskiego, pomorskiego i warmińsko-mazurskiego (makroregion rolniczy Pomorze i Mazury), co spowodowane jest wynikającą z zaszczości historycznych, dominacją wielohektarowych gospodarstw rolnych w tych województwach. Pomimo zidentyfikowanych różnic można stwierdzić, że przestrzenne zróżnicowanie procesów wskazuje na poprawność postawionej hipotezy o istotnym wpływie wielkości gospodarstwa rolnego na wielkość zużycia nawozów mineralnych.

Rysunek 4.9.

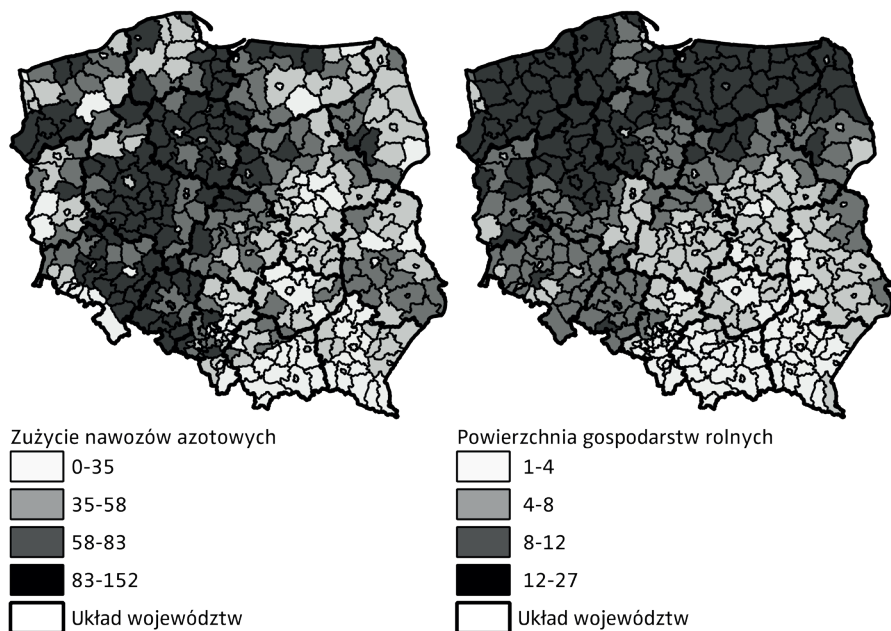
Średnie zużycie nawozów azotowych oraz średnia powierzchnia gospodarstw rolnych w układzie makroregionów rolniczych SGM



Źródło: opracowanie własne.

Rysunek 4.10.

Średnie zużycie nawozów azotowych oraz średnia powierzchnia gospodarstw rolnych w układzie województw



Źródło: opracowanie własne.

Dodatkowo, wizualna ocena przestrzennego zróżnicowania analizowanych danych przestrzennych wskazuje na skupianie się wysokich wartości danych oraz osobno niskich wartości w klastry, zarówno w przypadku średniego zużycia nawozów azotowych (proces Y), jak i średniej powierzchni gospodarstw rolnych (proces X) (zob. rysunek 4.9 i 4.10). W związku z tym zbadana została własność autokorelacji przestrzennej dla obydwu procesów dla obszaru całej Polski oraz obszarów czterech makroregionów rolniczych M_i za pomocą testu Morana (zob. Moran, 1948, s. 243–251; Zeliaś, 1991, s. 102–107; Suchecki, 2010, s. 112–114). Otrzymane wartości statystyki Morana I oraz wartości p testu przedstawione zostały w tabeli 4.1. We wszystkich zbadanych przypadkach stwierdzono dla procesu Y oraz procesu X występowanie silnej, dodatniej autokorelacji przestrzennej. Identyfikacja dodatniej autokorelacji przestrzennej oznacza, że na wartości procesu w wybranym regionie mogą mieć wpływ wartości procesu z regionów sąsiadujących. W takiej sytuacji wartość procesu w wybranym regionie może być objaśniana za pomocą średniej ważonej (wartość opóźnienia przestrzennego), wyznaczonej na podstawie wartości procesu z regionów sąsiadujących (zob. Zeliaś, 1991, s. 97–108, Suchecki, 2010, 102–117).

Tabela 4.1.

Wyniki testu Morana dla zużycia nawozów mineralnych oraz średniej powierzchni gospodarstw rolnych

Proces przestrzenny	Statystyka Morana <i>I</i>	Wartość <i>p</i>
Obszar Polski		
Proces Y	0,53	~0,00
Proces X	0,77	~0,00
Makroregiony rolnicze SGM		
Makroregion M1 (Pomorze i Mazury)		
Proces Y	0,24	~0,00
Proces X	0,39	~0,00
Makroregion M2 (Wielkopolska i Śląsk)		
Proces Y	0,20	~0,00
Proces X	0,45	~0,00
Makroregion M3 (Mazowsze i Podlasie)		
Proces Y	0,42	~0,00
Proces X	0,57	~0,00
Makroregion M4 (Małopolska i Pogórze)		
Proces Y	0,43	~0,00
Proces X	0,42	~0,00

Źródło: opracowanie własne.

W związku z ustaleniem własności autokorelacji przestrzennej dla średniego zużycia nawozów mineralnych (proces Y) oraz średniej powierzchni gospodarstw rolnych (proces X) do pomiaru zależności między procesami wykorzystano specyfikację modelu autoregresji przestrzennej SAR (zob. Suchecki, 2010, s. 248). W modelu SAR wprowadzany jest składnik autoregresji przestrzennej, poprzez uwzględnienie opóźnienia przestrzennego procesu objaśnianego Y, które opisuje uśredniony wpływ sąsiadów na wartości procesu objaśnianego (zob. Zeliaś, 1991, s. 107–116; Arbia 2006, s. 110; Suchecki, 2010, s. 241–254) Ze względu na ustalenie autokorelacji przestrzennej również dla procesu objaśniającego X w specyfikacji modelu uwzględniono również opóźnienie przestrzenne procesu objaśnianego. W związku z tym założona została specyfikacja modelu SDM (zob. Suchecki, 2010, s. 253–255), której postać zapisano za pomocą poniższego równania:

$$\mathbf{Y} = \alpha_0 + q_1 \mathbf{WY} + \alpha_1 \mathbf{X} + q_2 \mathbf{WX} + \boldsymbol{\varepsilon}, \quad (4.1)$$

gdzie \mathbf{Y} , \mathbf{X} są wektorami wartości procesów przestrzennych, α_0 , α_1 są to parametry strukturalne, q_1 , q_2 stanowią parametry autoregresji przestrzennej, \mathbf{W} jest standaryzowaną macierzą sąsiedztwa pierwszego rzędu¹², $\boldsymbol{\varepsilon}$ jest wektorem wartości białego szumu przestrzennego.

Zgodnie z przyjętą specyfikacją modelową zapisaną w równaniu (4.1), oszacowano parametry pięciu przestrzennych modeli ekonometrycznych, pierwszego dla obszaru Polski i czterech kolejnych dla obszarów makroregionów rolniczych. Otrzymane wyniki estymacji parametrów dla modeli końcowych przedstawiono w tabelach 4.2 oraz 4.3. Dla wszystkich modeli stwierdzono istotny wpływ procesu objaśniającego X oraz opóźnienia przestrzennego procesu objaśnianego Y . Z kolei wpływ opóźnienia przestrzennego procesu objaśniającego X okazał się statystycznie nieistotny dla wszystkich przypadków, dlatego w przedstawionych modelach końcowych nie uwzględniono tego opóźnienia. Dodatkowo sprawdzono własność autokorelacji przestrzennej dla procesu resztowego za pomocą testu Morana i w przypadku każdego z modeli nie stwierdzono występowania tej własności.

Otrzymane wyniki estymacji parametrów modeli pozwoliły na identyfikację mikrozależności przestrzennych w przypadku modeli dla wybranych makroregionów rolniczych SGM (zob. tabela 4.2) oraz makrozależności przestrzennej w przypadku modelu odniesionego do obszaru całej Polski (zob. tabela 4.3). Ustalone mikrozależności dla obszarów pojedynczych makroregionów rolniczych SGM powinny poprawnie odzwierciedlać wpływ wielkości gospodarstwa rolnego na poziom zużycia nawozów mineralnych. W przypadku identyfikacji mikrozależności przestrzennych, najwyższe wartości oceny parametru regresji uzyskano dla modelu odniesionego do makroregionu rolniczego M4 (Małopolska i Pogórze, $\hat{\alpha}_1 = 5,24$), który charakteryzuje się najniższym poziomem koncentracji użytków rolnych w Polsce oraz do makroregionu rolniczego M3 (Mazowsze i Podlasie, $\hat{\alpha}_1 = 4,22$), który również posiada niski poziom koncentracji użytków rolnych. Znacznie niższe oceny parametru regresji otrzymano dla modeli empirycznych odniesionych do makroregionu M1 (Pomorze i Mazury, $\hat{\alpha}_1 = 2,52$) oraz makroregionu M2 (Wielkopolska i Śląsk, $\hat{\alpha}_1 = 2,36$), które charakteryzują się wyższym poziomem koncentracji struktury agrarnej w porównaniu z wcześniej rozważanymi makroregionami. Otrzymane wartości ocen parametru autoregresji q_1 dla kolejnych makroregionów tworzą podobny schemat, jak w przypadku ocen parametru regresji. Najwyższe oceny

¹² Za pomocą macierzy \mathbf{W} o wymiarach $n \times n$ wprowadzany jest system wag, gdzie elementy macierzy stanowią odzwierciedlenie istniejącej struktury przestrzennej dla n jednostek terytorialnych (zob. Zeliaś, 2001, s. 98–102; Suchecki, 2010, s. 105–107, 238–241; Suchecka, 2014, s. 145–177). Sąsiedztwo ustalane jest na zasadzie kryterium wspólnej granicy, gdzie element macierzy w_{ij} jest równy jeden, jeśli jednostka i oraz jednostka j posiadają wspólną granicę. Dodatkowo macierz \mathbf{W} jest standaryzowana wierszami do jedności.

parametru autoregresji q_1 uzyskane zostały w empirycznych modelach dla makroregionów: M4 ($\hat{q}_1 = 0,68$) i M3 ($\hat{q}_1 = 0,63$) oraz znacznie niższe wartości ocen w przypadku empirycznych modeli odniesionych do makroregionów: M2 ($\hat{q}_1 = 0,43$) i M1 ($\hat{q}_1 = 0,32$). Oznacza to, że procesy rolnicze na obszarze regionów charakteryzujących się niskim poziomem koncentracji użytków rolnych charakteryzują się silniejszymi dodatnimi zależnościami przestrzennymi w porównaniu z regionami o wysokim poziomie koncentracji użytków rolnych.

Tabela 4.2.

Wyniki estymacji parametrów modeli końcowych dla makroregionów rolniczych

Parametr	Ocena parametru	Statystyka t-Studenta	Wartość p	Parametr	Ocena parametru	Statystyka t-Studenta	Wartość p
Model dla makroregionu M1				Model dla makroregionu M2			
α_0	2,64	0,27	0,79	α_0	-1,15	-0,30	0,77
α_1	2,52	4,47	0,00	α_1	2,36	4,51	0,00
q_1	0,32	2,64	0,01	q_1	0,43	9,79	0,00
Statystyka Morana I	0,03	Wartość p	0,30	Statystyka Morana I	0,03	Wartość p	0,29
Funkcja wiarygodności	-346,31	Kryterium AIC	700,63	Funkcja wiarygodności	-438,96	Kryterium AIC	885,92
Model dla makroregionu M3				Model dla makroregionu M4			
α_0	10,43	1,10	0,27	α_0	-5,34	-1,40	0,16
α_1	4,22	4,47	0,00	α_1	5,24	5,46	0,00
q_1	0,63	3,98	0,00	q_1	0,68	7,81	0,00
Statystyka Morana I	0,03	Wartość p	0,30	Statystyka Morana I	0,05	Wartość p	0,17
Funkcja wiarygodności	-454,52	Kryterium AIC	917,04	Funkcja wiarygodności	-412,20	Kryterium AIC	832,41

Źródło: opracowanie własne.

W przypadku identyfikacji makrozależności przestrzennej (zob. tabela 4.3) należy stwierdzić, że wyznaczone wartości ocen parametru regresji oraz parametru autoregresji ($\hat{\alpha}_1 = 3,63$, $\hat{q}_1 = 0,53$), zawierają się pomiędzy najwyższymi i najniższymi ocenami parametrów dla empirycznych modeli odniesionych do wybranych makroregionów M_i ¹³.

¹³ Należy podkreślić, że występowanie autokorelacji przestrzennej dla procesu objaśnianego ma wpływ na otrzymywane oceny parametru regresji α_1 , przyczyniając się do podwyższenia lub obniżenia wartości tej oceny.

Tabela 4.3.

Wyniki estymacji parametrów modelu końcowego dla obszaru Polski

Parametr	Ocena parametru	Statystyka t-Studenta	Wartość p
α_0	2,33	2,44	0,01
α_1	3,26	6,70	0,00
q_1	0,53	13,95	0,00
Statystyka Morana I	-0,01	Wartość p	0,58
Funkcja wiarygodności	-1 689,84	Kryterium AIC	3 387,70

Źródło: opracowanie własne.

Ustalenie statystycznej istotności parametrów regresji dla czterech makroregionów rolniczych SGM pozwoliło na identyfikację mikrozależności przestrzennych. Jednak przy próbie interpretacji mikrozależności przestrzennych należy zwrócić uwagę na fakt, że istotny wpływ opóźnienia przestrzennego procesu objaśnianego Y dla wszystkich końcowych postaci modeli (zob. tabela 4.2 oraz tabela 4.3) wymusza inną interpretację parametrów regresji (ocenę analizowanej zależności), niż ma to miejsce w przypadku modelu regresji liniowej¹⁴. Problem ten wynika z faktu, że zmiana procesu objaśniającego X w wybranym regionie j będzie miała wpływ na proces objaśniany Y w sąsiednich regionach. W modelu regresji liniowej, odniesionym do układu n jednostek terytorialnych, ocena parametru regresji oznacza taką samą średnią zmianę procesu objaśnianego Y pod wpływem zmiany procesu objaśniającego X , bez względu na fakt, w których regionach zachodzą zmiany. Natomiast w przestrzennym modelu regresji uwzględniającym opóźnienie przestrzenne procesu objaśnianego, należy rozważyć n^2 interpretacji szczegółowych parametru regresji, w zależności od wybranych par jednostek terytorialnych (regionów)¹⁵. Odpowiednie przekształcenie przestrzennego modelu regresji pozwala na wyprowadzenie macierzy $\mathbf{S}(\mathbf{W})$ o liczbie n^2 elementów, gdzie poszczególne elementy $S(W)_{ij}$ macierzy obrazują siłę oddziaływania ze strony procesu objaśnianego X . Wybrany element $S(W)_{ij}$ interpretowany jest zgodnie z równaniem (4.5) jako średni wpływ zmiany procesu objaśnianego X , jaka wystąpiła w regionie i , na wartość procesu objaśnianego Y w regionie j . Za pomocą równań (4.2)–(4.5) przedstawiono specyfikację modelu autoregresji przestrzennej SAR oraz macierz $\mathbf{S}(\mathbf{W})$, zawierającą interpretacje szczegółowe parametru regresji (zob. Pietrzak, 2013, s. 131–133):

$$\mathbf{Y} = \alpha_0 + q\mathbf{W}\mathbf{Y} + \alpha_1\mathbf{X} + \boldsymbol{\varepsilon}, \quad (4.2)$$

¹⁴ Problem interpretacji parametrów modeli przestrzennych poruszony został w pracy LeSage, Pace (2009, s. 33–43), jak również w: Pietrzak (2013, s. 132–138).

¹⁵ Każda pojedyncza interpretacja parametru regresji dotyczy wpływu zmiany procesu objaśnianego X w regionie i na proces objaśniany Y w regionie j .

$$\mathbf{Y} = \frac{\alpha_0}{1-q} + \mathbf{S}(\mathbf{W})\mathbf{X} + \boldsymbol{\varepsilon}, \quad (4.3)$$

$$\mathbf{S}(\mathbf{W}) = (\mathbf{I} - q\mathbf{W})^{-1} \alpha_1, \quad (4.4)$$

$$S(\mathbf{W})_{ij} = \frac{\partial E(y_j)}{\partial x_i}, \quad (4.5)$$

gdzie \mathbf{Y} jest wektorem wartości procesu objaśnianego, \mathbf{X} jest wektorem wartości procesu objaśniającego, α_0, α_1 są to parametry strukturalne, q jest parametrem autoregresji przestrzennej, \mathbf{W} jest standaryzowaną macierzą sąsiedztwa pierwszego rzędu, $\boldsymbol{\varepsilon}$ jest wektorem wartości białego szumu przestrzennego, \mathbf{I} jest macierzą jednostkową, a $\mathbf{S}(\mathbf{W})$ jest macierzą interpretacji szczegółowych parametru regresji.

Identyfikacja makrozależności przestrzennej dla wybranego makroregionu M_i oznacza ustalenie zależności o stałym charakterze na obszarze tego makroregionu. W związku z tym dla wybranych jednostek terytorialnych nie będą występować istotne zmiany siły analizowanej zależności. Siłę zależności kształtują również zależności przestrzenne, co należy uwzględnić podczas interpretacji oddziaływania procesu objaśniającego. Oznacza to, że w przypadku modelu ekonometrycznego rozszerzonego o autoregresję przestrzenną, wpływ na mikrozależność w wybranym regionie mają wpływ również regiony sąsiednie. W związku ze złożonością zagadnienia interpretacji parametru strukturalnego w modelu autoregresji przestrzennej SAR autor proponuje rozważyć kolejno trzy warianty interpretacji mikrozależności dla n jednostek terytorialnych tworzących makroregion M_i . Interpretacja mikrozależności rozważana będzie dla pary zbiorów (O_i, O_j) , gdzie zarówno zbiór O_i , jak i zbiór O_j tworzyć będą wybrane jednostki terytorialne lub grupy jednostek, należące do makroregionu M_i .

W wariancie pierwszym, interpretacja mikrozależności rozważana będzie dla pary zbiorów (O_i, O_j) , do których należą pojedyncze jednostki terytorialne (regiony). Interpretacja mikrozależności będzie dotyczyć wyłącznie wpływu wynikającego ze zmiany procesu objaśnianego w pojedynczym regionie i na proces objaśniany w pojedynczym regionie j . W tym przypadku występuje n^2 interpretacji mikrozależności, które odczytać można na podstawie elementów macierzy $\mathbf{S}(\mathbf{W})$. W związku z tym macierz $\mathbf{S}(\mathbf{W})$ należy traktować jako macierz mikroparametrów dla każdej pary jednostek terytorialnych wchodzących w skład makroregionu M_i .

Z kolei w wariancie drugim, interpretacja mikrozależności rozważana będzie również dla pary zbiorów (O_i, O_j) , jednak zbiory te mogą tworzyć grupy

jednostek terytorialnych (regionów). Interpretacja mikrozależności polegałaby na ocenie wpływu zmiany procesu objaśniającego w grupie jednostek terytorialnych tworzących zbiór O_i na zmiany procesu objaśnianego w grupie jednostek terytorialnych tworzących zbiór O_j . Każda próba interpretacji mikrozależności przestrzennej w wariancie drugim wymaga sumowania odpowiednich elementów macierzy $\mathbf{S}(\mathbf{W})$ zgodnie z przyjętymi zbiorami jednostek terytorialnych O_i oraz O_j . Liczba potencjalnych interpretacji mikrozależności w wariancie drugim znacznie przekracza wartość n^2 .

W wariancie trzecim zastosowanie otrzymanych empirycznych modeli ekonometrycznych może być rozszerzone do analizy zmian w całym makroregionie rolniczym M_i (przyjętym podobszarze wniosków). Jednak wyznaczenie oceny mikroparametru i interpretacja mikrozależności przestrzennej dla całego makroregionu M_i wymaga uśrednienia wpływu wynikającego z n^2 ocen mikroparametrów, co przedstawione zostanie w kolejnym podrozdziale.

4.2.2. Wyznaczenie makrozależności przestrzennej na podstawie mikrozależności przestrzennych

W niniejszym podrozdziale rozważony zostanie problem ustalenia relacji, jaka zachodzi między mikrozależnościami przestrzennymi dotyczącymi wpływu wielkości gospodarstwa rolnego na poziom stosowanych nawozów sztucznych na poziomie makroregionów rolniczych SGM a makrozależnością przestrzenną na poziomie całego obszaru Polski. Jest to wskazane przez Pawłowskiego zagadnienie relacji między mikroparametrami a makroparametrem, przy czym w przedstawionej obecnie analizie relacja między parametrami ustalona została dla procesów przestrzennych. W poprzednim podrozdziale dla obszaru wniosków równemu obszarowi całej Polski zidentyfikowana została makrozależność przestrzenna, a dla podobszarów wyznaczonych zgodnie z granicami makroregionów rolniczych mikrozależności przestrzenne. Przedstawione wnioski co do identyfikacji zależności przestrzennych zgodne są z przyjętym w rozdziale drugim rozróżnieniem między mikrozależnościami i makrozależnościami przestrzennymi, gdzie zidentyfikowana mikrozależność nie zmienia swojego charakteru w ramach podobszarów wniosków.

W celu wyznaczenia równania dla relacji między mikroparametrami a makroparametrem przedstawione zostały najpierw specyfikacje modeli autoregresji przestrzennej SAR, gdzie analizowane dane przestrzenne na poziomie powiatów (układ NUTS 4) ograniczone są do obszarów kolejnych makroregionów rolniczych M_1, M_2, M_3, M_4 :

$$\mathbf{Y}_{M_1}^{N_4} = \alpha_{0,1}^{N_4} + \mathbf{q}_{1,1} \mathbf{W}_{M_1} \mathbf{Y}_{M_1}^{N_4} + \alpha_{1,1}^{N_4} \mathbf{X}_{M_1}^{N_4} + \boldsymbol{\varepsilon}_{M_1}, \quad (4.6)$$

$$\mathbf{Y}_{M_2}^{N_4} = \alpha_{0,2}^{N_4} + \mathbf{q}_{1,2} \mathbf{W}_{M_2} \mathbf{Y}_{M_2}^{N_4} + \alpha_{1,2}^{N_4} \mathbf{X}_{M_2}^{N_4} + \boldsymbol{\varepsilon}_{M_2}, \quad (4.7)$$

$$\mathbf{Y}_{M_3}^{N_4} = \alpha_{0,3}^{N_4} + q_{1,3} \mathbf{W}_{M_3} \mathbf{Y}_{M_3}^{N_4} + \alpha_{1,3}^{N_4} \mathbf{X}_{M_3}^{N_4} + \boldsymbol{\varepsilon}_{M_3}, \quad (4.8)$$

$$\mathbf{Y}_{M_4}^{N_4} = \alpha_{0,4}^{N_4} + q_{1,4} \mathbf{W}_{M_4} \mathbf{Y}_{M_4}^{N_4} + \alpha_{1,4}^{N_4} \mathbf{X}_{M_4}^{N_4} + \boldsymbol{\varepsilon}_{M_4}, \quad (4.9)$$

gdzie $\mathbf{Y}_{M_i}^{N_4}$ jest wektorem wartości procesu objaśnianego, $\mathbf{X}_{M_i}^{N_4}$ jest wektorem wartości procesu objaśniającego, $\alpha_{0,i}^{N_4}$, $\alpha_{1,i}^{N_4}$ oznaczają parametry strukturalne, $q_{1,i}$ to parametr autoregresji przestrzennej, \mathbf{W}_{M_i} oznacza standaryzowaną macierz sąsiedztwa pierwszego rzędu, $\boldsymbol{\varepsilon}_{M_i}$ jest białym szumem przestrzennym, natomiast M_i to i -ty makroregion rolniczy.

Zgodnie z przedstawionymi w poprzednim podrozdziale trzema wariantami interpretacji mikrozależności dla n jednostek terytorialnych tworzących makroregion M_i równania (4.6)–(4.9) pozwalają na sposób interpretacji określony w wariancie pierwszym oraz w wariancie drugim. Zgodnie z wariantem pierwszym, oszacowanie parametrów modeli z równań (4.6)–(4.9) pozwala na ocenę mikrozależności dla par pojedynczych powiatów na podstawie macierzy $\mathbf{S}(\mathbf{W})$, gdzie interpretowany jest wpływ procesu objaśnianego X w powiecie i na proces objaśniany y w powiecie j . Natomiast w ramach wariantu drugiego, zarówno do zbioru O_r , jak i do zbioru O_j mogą zostać przyporządkowane grupy powiatów. Oznacza to, że do interpretacji mikrozależności dla wybranych grup powiatów należy wybrać odpowiednie elementy macierzy $\mathbf{S}(\mathbf{W})$, przy czym w zależności od wybranego modelu powiaty te muszą zawierać się w odpowiednim makroregionie rolniczym M_i .

Zgodnie z wariantem trzecim, w przypadku interpretacji mikrozależności dla całego makroregionu rolniczego M_i składającego się z n powiatów należy uśrednić wpływ zmian procesu objaśnianego X na zmianę procesu objaśnianego Y . W pracy Pietrzak (2013) zaproponowano autorski zestaw czterech miar oddziaływania przestrzennego pozwalających na interpretację uśrednionego wpływu procesu objaśnianego X w modelach ekonometrycznych z autoregresją przestrzenną. Na zestaw ten składa się miara średniego oddziaływania bezpośredniego A_D , miara średniego oddziaływania pośredniego A_I , miara średniego oddziaływania rezydualnego A_R oraz miara średniego oddziaływania całkowitego A_T . Przy znajomości końcowej postaci modelu ekonometrycznego (równania (4.6)–(4.9)) interpretacje mikrozależności z wariantu pierwszego i drugiego mogą być przybliżane za pomocą wymienionych miar oddziaływania przestrzennego (zob. Pietrzak, 2013, s. 138–154).

Miara średniego oddziaływania całkowitego A_T stanowi sumę miary średniego oddziaływania bezpośredniego A_D , miary średniego oddziaływania pośredniego A_I oraz miary średniego oddziaływania rezydualnego A_R . Miara średniego oddziaływania bezpośredniego A_D wyraża średnią wielkość zmiany procesu objaśnianego Y w dowolnej jednostce terytorialnej (regionie) pod wpływem zmiany procesu objaśnianego X w tej samej jednostce. Miara ta obliczana jest na podstawie następującego równania:

$$A_D = n^{-1} \text{tr}(\mathbf{S}(\mathbf{W})), \quad (4.10)$$

gdzie $\mathbf{S}(\mathbf{W})$ określona została za pomocą równania (4.4), symbol tr oznacza ślad macierzy, n jest liczbą jednostek terytorialnych w przyjętym układzie (zob. Pietrzak, 2013, s. 134).

Kolejną miarą jest miara średniego oddziaływania pośredniego A_I , która wyraża średnią zmianę procesu objaśnianego Y w dowolnie wybranej jednostce terytorialnej, wywołaną zmianą procesu objaśniającego X w jednostce sąsiedniej w sensie sąsiedztwa pierwszego rzędu (zgodnie z przyjętą macierzą wag \mathbf{W}), co określa równanie (zob. Pietrzak, 2013, s. 136–137):

$$A_I = n^{-1} \text{tr}(\mathbf{W} \cdot \mathbf{S}(\mathbf{W})^T), \quad (4.11)$$

gdzie \mathbf{W} jest macierzą sąsiedztwa pierwszego rzędu standaryzowaną wierszami, a pozostałe oznaczenia są takie same jak w przypadku równania (4.10).

Ostatnią miarą, miarą średniego oddziaływania rezyduального A_R , wyraża natomiast średnią zmianę procesu objaśnianego Y w dowolnie wybranej jednostce terytorialnej, wywołaną zmianą procesu objaśniającego X , pod warunkiem, że zmiana zmiennej procesu X nastąpiła w jednostce sąsiedniej (w sensie sąsiedztwa wyższych rzędów niż jeden). Miara ta wyznaczana jest za pomocą równań:

$$A_R = n^{-1} \text{tr}(\mathbf{G} \cdot \mathbf{S}(\mathbf{W})^T), \quad (4.12)$$

$$\mathbf{F} = \mathbf{1} - \mathbf{I} - \mathbf{W}_B, \quad (4.13)$$

gdzie $\mathbf{1}$ jest macierzą jedynekową, \mathbf{I} jest macierzą jednostkową, \mathbf{W}_B jest macierzą binarną sąsiedztwa pierwszego rzędu, natomiast \mathbf{G} jest standaryzowaną wierszami macierzą \mathbf{F} (zob. Pietrzak, 2013, s. 137–138).

W rozdziale drugim opisana została procedura wyznaczenia oceny mikroparametru dla całego podobszaru wniosków, gdzie za pomocą równań (2.7)–(2.9) pokazane zostało, że ustalona na podstawie modelu regresji mikrozależność przestrzenna powinna zachodzić zarówno w przypadku pojedynczych jednostek terytorialnych tworzących podobszar wniosków, jak i w przypadku całego makroregionu M_i (podobszaru wniosków). Ponieważ układ NUTS 4 oraz układ makroregionów rolniczych przyporządkowane zostały do jednorodnego zbioru układów, to dla wszystkich powiatów w granicach wybranego makroregionu M_i analizowana mikrozależność powinna mieć podobny charakter. Oznacza to, że ustalony charakter mikrozależności powinien pozostać niezmienny w ramach oceny oddziaływania procesów dla całego obszaru pojedynczego makroregionu M_i . Za pomocą równań (4.6)–(4.9) zapisano specyfikacje modeli

ekonometrycznych SAR odniesionych do układu NUTS 4. Podstawienie miary średniego wpływu całkowitego A_T^{Mi} do wymienionych specyfikacji może posłużyć do sformułowania postaci modeli wyrażających mikrozależność przestrzenną dla całego makroregionu Mi^{16} . W tym celu, na podstawie wyznaczonej z równań (4.6)–(4.9) macierzy $S(W)$, należy obliczyć wartości miar średniego wpływu całkowitego A_T^{Mi} dla kolejnych makroregionów rolniczych Mi . Wartość miary A_T^{Mi} wyraża uśrednioną siłę mikrozależności i może być ona utożsamiona z oceną mikroparametru określającego mikrozależność na poziomie całego makroregionu Mi . W związku z tym oceny mikroparametrów ($\hat{\alpha}_{1,1}$) mogą zostać wyrażone za pomocą miary średniego wpływu całkowitego A_T . Specyfikacje modeli mikrozależności dla kolejnych makroregionów rolniczych Mi przedstawiono za pomocą równań¹⁷:

$$\hat{\alpha}_{0,1} = \frac{\alpha_{0,1}^{N4}}{1 - q_{1,1}}, \hat{\alpha}_{1,1} = A_T^{M1}, \quad (4.14)$$

$$Y^{M1} = \hat{\alpha}_{0,1} + \hat{\alpha}_{1,1} X^{M1}, \quad (4.15)$$

$$\hat{\alpha}_{0,2} = \frac{\alpha_{0,2}^{N4}}{1 - q_{1,2}}, \hat{\alpha}_{1,2} = A_T^{M2}, \quad (4.16)$$

$$Y^{M2} = \hat{\alpha}_{0,2} + \hat{\alpha}_{1,2} X^{M2}, \quad (4.17)$$

$$\hat{\alpha}_{0,3} = \frac{\alpha_{0,3}^{N4}}{1 - q_{1,3}}, \hat{\alpha}_{1,3} = A_T^{M3}, \quad (4.18)$$

$$Y^{M3} = \hat{\alpha}_{0,3} + \hat{\alpha}_{1,3} X^{M3}, \quad (4.19)$$

¹⁶ Dla modeli określonych w równaniach (4.6)–(4.9) analizowane dane przestrzenne odniesione są do układu jednostek terytorialnych NUTS 4 oraz ograniczone przestrzennie do obszarów wybranych makroregionów rolniczych Mi .

¹⁷ Modele ekonometryczne estymowane na poziomie powiatów można wykorzystać do analizy zmian zużycia nawozów mineralnych pod wpływem zmiany wielkości gospodarstwa rolnego w poszczególnych powiatach, natomiast model odnoszący się do makroregionu rolniczego służyć będzie do badania zmian zużycia nawozów w całym makroregionie.

$$\hat{\alpha}_{0,4} = \frac{\alpha_{0,4}^{N4}}{1 - q_{1,4}}, \hat{\alpha}_{1,4} = A_T^{M4}, \quad (4.20)$$

$$Y^{M4} = \hat{\alpha}_{0,4} + \hat{\alpha}_{1,4} X^{M4}, \quad (4.21)$$

gdzie Y^{Mi} i X^{Mi} stanowią wartości procesu objaśnianego oraz objaśniającego dla całego obszaru makroregionu Mi , $\hat{\alpha}_{0,i}$ jest oceną wyrazu wolnego, a $\hat{\alpha}_{1,i}$ jest oceną mikroparametru $\alpha_{1,i}$.

W ostatnim kroku, określone za pomocą równań (4.14)–(4.21) modele mikro zależności dla kolejnych makroregionów rolniczych Mi posłużą do wyznaczenia makro zależności dla obszaru Polski (układ NUTS 0). Ustalona w ten sposób makro zależność przestrzenna pozwoli na syntetyczną ocenę wpływu wielkości gospodarstw rolnych w Polsce na średnie zużycie nawozów azotowych. Makro zależność stanowić będzie średnią ważoną wyznaczoną na podstawie ustalonych mikro zależności, gdzie wartości wag wyznaczone zostaną na podstawie przestrzennego zróżnicowania procesów X oraz Y .

W celu ustalenia zestawu wag założony zostanie najpierw charakter procesów Y^{Mi} oraz X^{Mi} . Proces Y^{Mi} odzwierciedlać będzie średnie zużycie nawozów mineralnych w wybranym makroregionie Mi , a proces X^{Mi} średnią wielkość gospodarstwa rolnego, co określone zostało za pomocą następującego równania:

$$Y^{Mi} = \bar{Y}_{Mi}^{N4}, X^{Mi} = \bar{X}_{Mi}^{N4}. \quad (4.22)$$

Wartości procesów Y^{Mi} i X^{Mi} wyznaczone zostały poprzez agregację danych przestrzennych odniesionych do układu jednostek terytorialnych NUTS 4 na podstawie równania (1.1)¹⁸. W przypadku procesu Y^{Mi} wagi dla i -tego powiatu wyznaczone zostały na podstawie jego udziału w całkowitej powierzchni gospodarstw rolnych makroregionu Mi , a w przypadku procesu X^{Mi} na podstawie udziału i -tego powiatu w całkowitej liczebności gospodarstw rolnych w makroregionie.

Następnie zgodnie z równaniem (4.23) ustalona została relacja między wyznaczonymi wartościami procesu X dla kolejnych makroregionów rolniczych Mi , a wartością procesu X dla obszaru całej Polski:

$$X^{Mi} = \beta_i X^{N0}, \quad (4.23)$$

gdzie $X^{N0} = \bar{X}_{N0}^{N4}$ stanowi średnią powierzchnię gospodarstw rolnych dla Polski.

W kolejnym kroku wyznaczono odwrotną relację dla procesu Y polegającą na wyrażeniu wartości procesu Y^{N0} dla całego obszaru Polski jako średniej wa-

¹⁸ Równanie (1.1) określa proces agregacji danych przestrzennych wyrażonych w wartościach średnich.

żonej z wartości procesów Y^{Mi} odniesionych do czterech makroregionów rolniczych Mi , tj.:

$$Y^{NO} = \sum w_i \bar{Y}_i = w_1 Y^{M1} + w_2 Y^{M2} + w_3 Y^{M3} + w_4 Y^{M4}, \quad (4.24)$$

gdzie wagi w_i wyznaczone zostały na podstawie udziału i -tego makroregionu rolniczego SGM w całkowitej powierzchni gospodarstw rolnych w Polsce.

Następnie do równania (4.24) podstawiono równania: (4.15), (4.17), (4.19), (4.21) modeli mikrozależności dla makroregionów Mi , co pozwoliło na wyprowadzenie równania dla procesu Y^{NO} , odniesionego do obszaru całej Polski:

$$Y^{NO} = (\hat{\alpha}_{0,1} + \hat{\alpha}_{1,1} X^{M1}) w_1 + (\hat{\alpha}_{0,2} + \hat{\alpha}_{1,2} X^{M2}) w_2 + (\hat{\alpha}_{0,3} + \hat{\alpha}_{1,3} X^{M3}) w_3 + (\hat{\alpha}_{0,4} + \hat{\alpha}_{1,4} X^{M4}) w_4. \quad (4.25)$$

Dotatkowo w równaniu (4.25) w miejsce procesów X^{Mi} wprowadzono proces X^{NO} zgodnie z relacją określoną w równaniu (4.23), w wyniku czego otrzymano następujące równanie:

$$Y^{NO} = (\hat{\alpha}_{0,1} + \hat{\alpha}_{1,1} \beta_1 X^{NO}) w_1 + (\hat{\alpha}_{0,2} + \hat{\alpha}_{1,2} \beta_2 X^{NO}) w_2 + (\hat{\alpha}_{0,3} + \hat{\alpha}_{1,3} \beta_3 X^{NO}) w_3 + (\hat{\alpha}_{0,4} + \hat{\alpha}_{1,4} \beta_4 X^{NO}) w_4. \quad (4.26)$$

W ostatnim kroku, odpowiednie przekształcenie równania (4.26) pozwoliło na wyznaczenie równania (4.32) określającego makrozależność przestrzenną dla obszaru całej Polski:

$$Y^{NO} = (\hat{\alpha}_{0,1} w_1 + \hat{\alpha}_{0,2} w_2 + \hat{\alpha}_{0,3} w_3 + \hat{\alpha}_{0,4} w_4) + (\hat{\alpha}_{1,1} w_1 \beta_1 + \hat{\alpha}_{1,2} w_2 \beta_2 + \hat{\alpha}_{1,3} w_3 \beta_3 + \hat{\alpha}_{1,4} w_4 \beta_4) X^{NO}, \quad (4.27)$$

$$\hat{\alpha}_0 = \hat{\alpha}_{0,1} w_1 + \hat{\alpha}_{0,2} w_2 + \hat{\alpha}_{0,3} w_3 + \hat{\alpha}_{0,4} w_4, \quad (4.28)$$

$$\hat{\alpha}_1 = \hat{\alpha}_{1,1} w_1 \beta_1 + \hat{\alpha}_{1,2} w_2 \beta_2 + \hat{\alpha}_{1,3} w_3 \beta_3 + \hat{\alpha}_{1,4} w_4 \beta_4, \quad (4.29)$$

$$\hat{\alpha}_1 = \hat{\alpha}_{1,1} \gamma_1 + \hat{\alpha}_{1,2} \gamma_2 + \hat{\alpha}_{1,3} \gamma_3 + \hat{\alpha}_{1,4} \gamma_4, \quad (4.30)$$

$$\gamma_1 = w_1 \beta_1, \gamma_2 = w_2 \beta_2, \gamma_3 = w_3 \beta_3, \gamma_4 = w_4 \beta_4, \quad (4.31)$$

$$Y^{NO} = \hat{\alpha}_0 + \hat{\alpha}_1 X^{NO}. \quad (4.32)$$

Równanie (4.30) wyraża relację między ocenami mikroparametrów ($\hat{\alpha}_{1,i}$) dla makroregionów rolniczych M_i a oceną makroparametru ($\hat{\alpha}_1$) dla całego obszaru Polski. Jak zostało wykazane w rozdziale drugim, wartość oceny makroparametru ($\hat{\alpha}_1$) zależy od wartości ocen mikroparametrów ($\hat{\alpha}_{1,i}$) oraz od zestawu wag γ_i . Z kolei poszczególne wagi γ_i wyznaczone są na podstawie zestawu wag w_i oraz β_i zgodnie z równaniem (4.31).

Przedstawione kolejno równania (4.14)–(4.32) pozwoliły na wyznaczenie relacji między mikroparametrami a makroparametrem dla rozpatrywanej prawidłowości wpływu wielkości gospodarstwa rolnego na poziom zużycia nawozów azotowych. Relacja ustalona została w dwóch krokach. W pierwszym kroku wyznaczono wartości miary oddziaływania całkowitego A_r^{Mi} , gdzie do obliczeń wykorzystane zostały wyniki estymacji parametrów modeli ekonometrycznych odniesionych do kolejnych makroregionów rolniczych SGM (zob. tabela 4.2). Pozwoliło to na ustalenie wartości ocen mikroparametrów $\alpha_{1,i}$ dla kolejnych makroregionów rolniczych SGM na podstawie równań (4.14)–(4.21). Wyznaczone oceny mikroparametrów ($\hat{\alpha}_{1,i}$) przedstawiono w tabeli 4.4.

Tabela 4.4.

Ocena mikrozależności dla makroregionów rolniczych

Mikroparametr	Ocena parametru	Wartość p^*	Mikroparametr	Ocena parametru	Wartość p
Makroregion M1			Makroregion M2		
$\alpha_{0,1}$	3,88	–	$\alpha_{0,2}$	–2,02	–
$\alpha_{1,1}$	2,58	~0,00	$\alpha_{1,2}$	2,48	~0,00
Makroregion M3			Makroregion M4		
$\alpha_{0,3}$	28,19	–	$\alpha_{0,4}$	–16,69	–
$\alpha_{1,3}$	4,68	~0,00	$\alpha_{1,4}$	5,92	~0,00

Uwagi:

* Możliwa jest ocena istotności statystycznej miary średniego wpływu całkowitego A_r , co pozwala na stwierdzenie istotności statystycznej mikroparametrów $\alpha_{1,i}$ (zob. Le-Sage, Pace, 2009, s. 34–43; Pietrzak, 2013, s. 149).

Źródło: opracowanie własne.

Otrzymane wyniki pozwoliły na interpretację mikrozależności przestrzennych dla kolejnych makroregionów rolniczych SGM. W przypadku analizy mikrozależności, najwyższe wartości oceny mikroparametrów uzyskano dla modelu odniesionego do makroregionu rolniczego M4 (Małopolska i Pogórze, $\hat{\alpha}_{1,4} = 5,92$) oraz do makroregionu rolniczego M3 (Mazowsze i Podlasie,

$\hat{\alpha}_{1,3} = 4,68$), które charakteryzują się niskim poziomem koncentracji użytków rolnych. Z kolei dla makroregionów M1 i M2 o wyższym poziomie koncentracji struktury agrarnej otrzymane zostały niższe oceny mikroparametrów: makroregion M1 (Pomorze i Mazury, $\hat{\alpha}_{1,1} = 2,58$) oraz makroregion M2 (Wielkopolska i Śląsk, $\hat{\alpha}_{1,2} = 2,48$). Oznacza to, że w makroregionach o niższym poziomie koncentracji użytków rolnych wpływ zmiany powierzchni użytków rolnych gospodarstwa rolnego na poziom zużycia nawozów azotowych jest prawie dwukrotnie wyższy.

W ostatnim kroku, na podstawie wyznaczonych ocen mikroparametrów $\alpha_{1,i}$ dla czterech makroregionów rolniczych SGM (zob. tabela 4.4), wyznaczona została ocena makroparametru α_1 na poziomie 3,54 dla obszaru całej Polski zgodnie z równaniem (4.30). Otrzymane wyniki przedstawiono w tabeli 4.5, gdzie dodatkowo zamieszczone zostały także wartości wag w_r , β_i oraz γ_i . Wyznaczoną wartość oceny makroparametru należy interpretować jako uśredniony dla całego obszaru Polski wpływ zmiany wielkości gospodarstwa rolnego na średni poziom zużycia nawozów mineralnych.

Wyznaczenie makroparametru na podstawie kombinacji liniowej mikroparametrów pozwoliło na osiągnięcie postawionego w rozdziale celu badania w postaci ustalenia relacji między mikroparametrami dla makroregionów rolniczych SGM a makroparametrem dla obszaru Polski. Zaprezentowane rozwiązanie poszerza postawiony przez Pawłowskiego problem o aspekty przestrzenne.

Należy zwrócić uwagę, że ocenę makroparametru α_1 można wyznaczyć również drugim sposobem, na podstawie wyników estymacji parametrów modelu ekonometrycznego odniesionego do całego obszaru Polski. Parametry takiego modelu zostały oszacowane w poprzednim podrozdziale 4.2.1 oraz przedstawione w tabeli 4.3. W tym przypadku, zgodnie z równaniem (4.14), ocena makroparametru równa jest wartości miary oddziaływania całkowitego A_r , obliczonej na podstawie końcowej postaci modelu ekonometrycznego (zob. tabela 4.3). Wyznaczone drugim sposobem oceny makroparametru α_1 oraz wyrazu wolnego α_0 przedstawione zostały również w tabeli 4.5.

Tabela 4.5.
Ocena makrozależności dla obszaru Polski

Przyjęte wartości wag w_r , β_i oraz γ_i				
Wagi w_i	w_1	w_2	w_3	w_4
Wartości wag	0,21	0,28	0,36	0,15
Wagi β_i	β_1	β_2	β_3	β_4
Wartości wag	1,73	1,22	0,83	0,40
Wagi γ_i	γ_1	γ_2	γ_3	γ_4
Wartości wag	0,37	0,34	0,30	0,06
Sposoby wyznaczenia oceny makroparametru				
Kombinacja liniowa mikroparametrów		Estymacja parametrów modelu ekonometrycznego		
Parametr	Ocena parametru	Parametr	Ocena parametru*	Wartość p
α_0	4,66	α_0	4,96	–
α_1	3,54	α_1	3,59	~0,00

Uwagi:

* Oceny parametrów α_0 , α_1 wyznaczone zostały sposobem drugim na podstawie równania (4.14), gdzie do obliczenia wartości miary oddziaływania całkowitego A_T wykorzystane zostały oceny parametrów $\hat{\alpha}_0 = 2,33$; $\hat{\alpha}_1 = 3,26$; $\hat{q} = 0,53$ oraz macierz sąsiedztwa \mathbf{W} dla obszaru całej Polski (zob. tabela 4.3).

Źródło: opracowanie własne.

Porównanie oceny makroparametru obliczonej na podstawie równania (4.30) z oceną makroparametru wyznaczoną sposobem drugim na podstawie wyników estymacji parametrów modelu ekonometrycznego nie wskazuje na istotne różnice w otrzymanych wartościach. Należy jednak podkreślić, że wyznaczenie oceny makroparametru pierwszym sposobem wzbogaca wykonywane przestrzenne badania ekonomiczne. Ustalenie relacji między mikroparametrami a makroparametrem wymagało bowiem wyznaczenia jednorodnego zbioru układów, przyjęcia obszaru, podobszarów wniosków oraz identyfikacji i oceny mikrozmależności przestrzennych dla ustalonych podobszarów.

Rozdział 5.

Problem skali: identyfikacja na podstawie generowanych danych przestrzennych wyrażonych w wartościach średnich

5.1. Procedura symulacyjna w świetle zaproponowanej redefinicji problemu skali

Na tle krytycznej oceny prac z literatury przedmiotu oraz rozważań autora na temat poprawności przestrzennych badań ekonomicznych, w rozdziale trzecim zaproponowana została redefinicja problemu skali. Problem skali określony został jako problem istotnych różnic w ocenach własności procesów przestrzennych oraz łączących je zależności przyczynowych, otrzymywanych w wyniku analizy przeprowadzonej na podstawie układów jednostek terytorialnych, które przyporządkowano w ramach podjętego problemu badawczego do jednorodnego zbioru układów. Zgodnie z zaproponowaną nową definicją wybrane układy jednostek terytorialnych dla kolejnych poziomów agregacji nie mogą być ustalone w sposób dowolny, ponieważ muszą być przyporządkowane do jednorodnego zbioru układów. Autor wykazał w rozdziale trzecim, że dopiero rozważenie problemu skali w taki sposób pozwoli na uzyskanie poprawnych wniosków w ramach podjętego problemu badawczego.

W omówionych w rozdziale trzecim pracach dotyczących problemu skali, badania własności procesów przestrzennych prowadzone były najczęściej w jednym lub dwóch etapach. W etapie pierwszym analiza oparta była na empirycznych danych przestrzennych, a w etapie drugim badanie było poszerzane o analizę generowanych danych przestrzennych. W związku z tym w etapie

pierwszym dokonywany był wybór empirycznych danych przestrzennych na niższym poziomie agregacji, a następnie dane te agregowane były na wyższych poziomach agregacji dla wielu, generowanych losowo przez *zoning system* lub *grouping system* układów jednostek terytorialnych (Openshaw, Taylor 1979, s. 128–132; Openshaw, 1984b, s. 8–31; Reynolds, 1998, s. 10–18). Tak określona procedura symulacyjna pozwalała na wyznaczenie zbioru ocen, który utożsamiany był z rozkładem empirycznym badanych parametrów. W wyniku dalszej analizy rozkładu empirycznego wyznaczane były jego statystyki opisowe, których interpretacja determinowała stwierdzenie o identyfikacji problemu skali¹. Jednak poza jednoznacznym stwierdzeniem występowania problemu skali, dalsza interpretacja otrzymanych wyników była problematyczna. Trudność ta wynikała z faktu, że otrzymywany zbiór wartości tworzył przedział o dużym rozstępie (np. dla współczynnika korelacji liniowej, wyznaczone wartości dla wyższych poziomów agregacji zawierały się w przedziale od –1 do 1). Pozorna identyfikacja problemu skali oraz wysokie wartości współczynnika zmienności dla otrzymywanych rozkładów empirycznych wynikały z faktu, że przyporządkowane do generowanych losowo układów jednostek terytorialnych dane przestrzenne nie odzwierciedlały właściwie oddziaływania przyczyn.

W pracach rozważających problem skali wykonywany był także etap drugi oparty na odpowiednio zaprojektowanej procedurze symulacyjnej, w ramach której były generowane dane przestrzenne. Dla celów symulacji, własności procesów zakładane były na podstawie wybranych w etapie pierwszym, empirycznych danych przestrzennych. Stosowana metodyka w etapie drugim była taka sama, jak w etapie pierwszym. W związku z tym również w etapie drugim używany był szeroki rozstęp rozkładu wartości ocen w wyniku agregacji danych, co skutkowało identyfikacją problemu skali i jednocześnie nie pozwalało na wyciągnięcie jednoznacznych wniosków odnośnie do rozwiązania tego problemu.

Zaproponowana przez autora redefinicja problemu skali wymusza odmienne podejście do analizy tego problemu w porównaniu z podejściem reprezentowanym w literaturze przedmiotu. Również w tym przypadku można przeprowadzić badania nad problemem skali w dwóch etapach. W etapie pierwszym następowałoby określenie problemu badawczego, wybór odpowiednich, empirycznych danych przestrzennych oraz ustalenie jednorodnego zbioru układów. Podstawowa różnica polega na tym, że agregacja danych wykonywana byłaby wyłącznie na podstawie układów jednostek terytorialnych tworzących jednorodny zbiór układów. W ten sposób autor nie dopuszcza do sytuacji, w której empiryczne dane przestrzenne byłyby agregowane na wyższych poziomach agregacji na podstawie generowanych losowo układów jednostek terytorialnych. Tak określona procedura pozwala na wyznaczenie na wyższych poziomach agregacji tylko pojedynczych ocen parametrów. Niemożliwe jest natomiast

¹ Autor wskazał w rozdziale trzecim na fakt pozornej identyfikacji problemu skali w przedstawianych w literaturze przedmiotu badaniach.

otrzymanie zbioru ocen, ponieważ empiryczne dane przestrzenne przyporządkowane są na każdym poziomie agregacji tylko do jednego układu jednostek terytorialnych. Sytuacja taka pozwala na stwierdzenie różnicy w otrzymanych ocenach na różnych poziomach agregacji, jednak nie pozwala na stwierdzenie, czy różnice te są statystycznie istotne. Dodatkowo nie jest możliwe ustalenie na podstawie empirycznych danych przestrzennych, w ramach których własności procesu przestrzennego identyfikowany jest problem skali.

Rozwiązaniem w takiej sytuacji jest przejście do etapu drugiego i przeprowadzenie badania opartego na analizie symulacyjnej. Wymaga to jednak odpowiedniego zaprojektowania procedury symulacyjnej, która pozwoliłaby na poprawne ustalenie zmian w wyznaczanych ocenach parametrów pod wpływem procesu agregacji danych. Procedura generowania danych przestrzennych powinna dawać możliwość wyboru określonych własności analizowanych procesów, co z kolei pozwoliłoby na sprawdzenie ich wpływu na występowanie problemu skali. Należy przy tym zauważyć że, tak samo jak w etapie pierwszym generowane dane przestrzenne powinny być agregowane wyłącznie w ramach pojedynczych układów jednostek terytorialnych należących do jednorodnego zbioru układów. Zabieg taki ma pozwolić na poprawną identyfikację deterministycznych i stochastycznych własności procesów przestrzennych. Tak określona procedura symulacyjna daje możliwość otrzymania zbioru wartości ocen parametrów na każdym z poziomów agregacji. Wykonana następnie analiza otrzymanych rozkładów pozwoli na ustalenie zmian w zakresie badanych własności lub zależności. Według autora tylko takie podejście daje możliwość poprawnego rozwiązania problemu skali.

W rozdziale pierwszym stwierdzono, że w celu zapewnienia porównywalności obszarów powinny być przyjmowane dwa typy danych przestrzennych odniesionych do obszarów nieregularnych. W sposobie pierwszym pod uwagę brane są dane przestrzenne wyrażone w postaci wartości średnich, a w drugim przyjmowane są dane przestrzenne, ważone w stosunku do odpowiednio rozumianej wielkości obszaru. W związku z tym w rozdziale piątym analizy symulacyjne przeprowadzone zostaną na podstawie generowanych danych przestrzennych, które wyrażane są w wartościach średnich. Natomiast w rozdziale szóstym wykonana zostanie analiza oparta na generowanych danych przestrzennych ważonych wielkością obszaru.

5.2. Analiza symulacyjna dla szumów przestrzennych

Zgodnie z zaprezentowaną w poprzednim podrozdziale metodyką badań nad problemem skali przeprowadzone zostaną kolejne analizy symulacyjne, których celem będzie sprawdzenie występowania problemu skali w sytuacji, gdy analizowane procesy przestrzenne posiadać będą założone własności. Jako pierwsza, zbadana zostanie możliwość występowania problemu skali w przypadku agregacji danych przestrzennych, będących realizacją procesów stacjonarnych

w szerszym sensie. W związku z tym przyjęto następujące założenia. Analizowany proces charakteryzował się stałym poziomem wartości oczekiwanej i wariancji oraz brakiem występowania autokowariancji przestrzennej. Proces przestrzenny o takich własnościach określany jest mianem szumu przestrzennego, a jego własności opisane są za pomocą równań (Szulc, 2007, s. 27–30):

$$E(X(\mathbf{s})) = \mu, \quad (5.1)$$

$$K(\mathbf{s}, \mathbf{v}) = \begin{cases} \delta^2, & \text{gdy } \mathbf{s} = \mathbf{v} \\ 0, & \text{gdy } \mathbf{s} \neq \mathbf{v} \end{cases} \quad (5.2)$$

gdzie $E(X(\mathbf{s}))$ i $K(\mathbf{s}, \mathbf{v})$ oznaczają funkcję wartości oczekiwanej oraz funkcję kowariancji procesu przestrzennego $X(\mathbf{s})$, $\mathbf{s} = [s_1, s_2]$ oraz $\mathbf{v} = [v_1, v_2]$ są punktami zlokalizowanymi w przestrzeni. Proces szumu przestrzennego jest stacjonarny w szerszym sensie. Szczególny przypadek stanowi proces białego szumu przestrzennego, gdzie wartość oczekiwana jest równa zero a pozostałe własności jak w równaniu (5.2) (zob. np. Arbia, 2006, s. 49; Szulc, 2007, s. 30).

W ramach procedury symulacyjnej generowane były dane przestrzenne stanowiące realizacje dwóch procesów szumu przestrzennego $X_1(\mathbf{s})$, $X_2(\mathbf{s})^2$, gdzie założono dodatkowo, że procesy przestrzenne mogą być wzajemnie skorelowane. Specyfikację modelową dla realizacji tych procesów wyraża równanie:

$$\mathbf{X} = \alpha_0 + \boldsymbol{\varepsilon}, \quad (5.3)$$

gdzie \mathbf{X} jest wektorem wartości generowanego szumu przestrzennego, α_0 jest parametrem odpowiadającym za różny od zera poziom wartości oczekiwanej procesu, a $\boldsymbol{\varepsilon}$ jest wektorem wartości białego szumu przestrzennego.

Wykonana analiza symulacyjna pozwoliła na stwierdzenie, czy w wyniku procesu agregacji danych przestrzennych wystąpiły różnice w otrzymywanych rezultatach na kolejnych poziomach agregacji. W związku z tym sprawdzono możliwość zmian dla wyznaczanych ocen wartości oczekiwanej i wariancji oraz możliwość wystąpienia autokorelacji przestrzennej na skutek zmiany poziomu agregacji. Dodatkowo przeprowadzono analizę korelacyjną oraz analizę regresji, co pozwoliło na ocenę wpływu procesu agregacji również w tych aspektach.

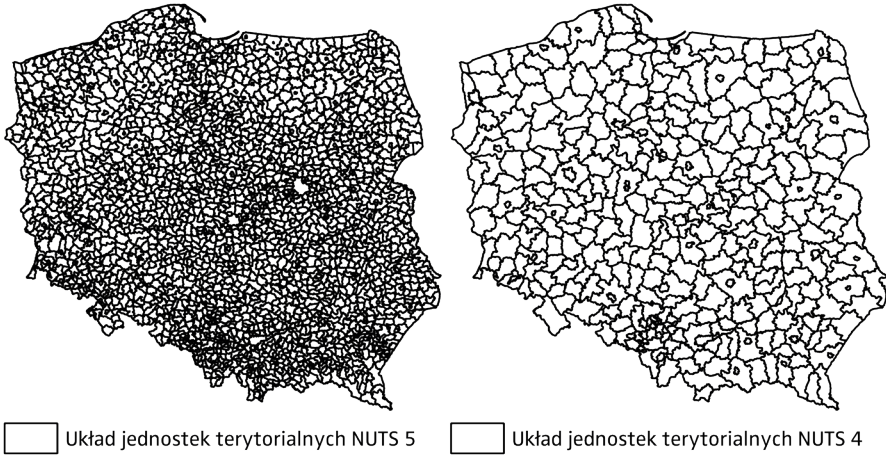
Na potrzeby wykonania procedury symulacyjnej ustalony został hipotetyczny jednorodny zbiór układów, w ramach którego były symulowane oraz agregowane dane przestrzenne. Jednorodny zbiór układów utworzony został z dwóch układów jednostek terytorialnych, układu NUTS 5 oraz układu NUTS 4. Wybór tych układów jednostek terytorialnych umotywowany jest tym, że dla

² W dalszej części pracy dla uproszczenia procesy te będą oznaczane odpowiednio X_1, X_2 .

większości przestrzennych badań ekonomicznych układy te powinny utworzyć jednorodny zbiór układów. Na rysunkach 5.1 i 5.2 przedstawiono wykorzystane w symulacji układy jednostek terytorialnych, które utworzyły hipotetyczny jednorodny zbiór układów³.

Rysunek 5.1.

Układy jednostek terytorialnych wykorzystane w analizie symulacyjnej



Źródło: opracowanie własne.

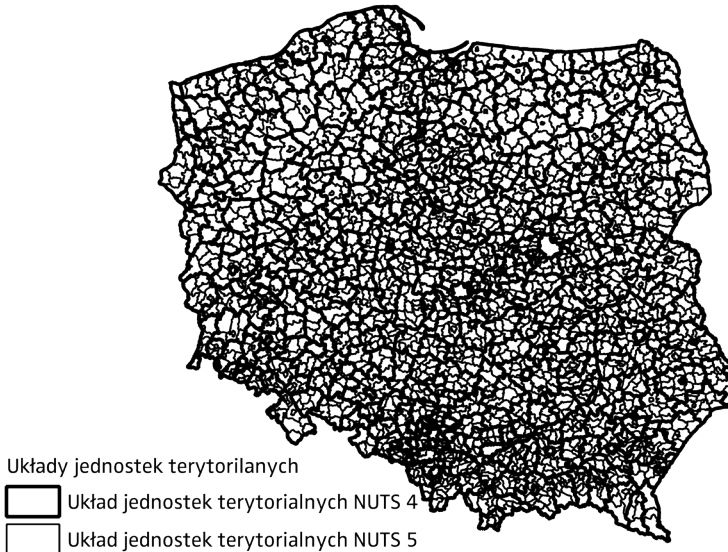
Jak już wspomniano na początku rozdziału problem skali rozpatrzony został najpierw dla procesów ekonomicznych, w przypadku których empiryczne dane przestrzenne wyrażone są w wartościach średnich. Zarówno w przypadku danych przestrzennych wyrażonych w wartościach średnich, jak i danych ważonych wielkością obszarów pojawia się kwestia ich prawidłowej agregacji, która została omówiona w rozdziale pierwszym. Wskazane zostało, że procedura agregacji danych przestrzennych wyrażonych w wartościach średnich polega na odpowiednim wyznaczeniu średniej ważonej dla każdego z regionów należącego do układu jednostek terytorialnych na wyższym poziomie agregacji. W związku z tym procedura symulacyjna przeprowadzona została w następujących krokach. W kroku pierwszym, zgodnie z przyjętym jednorodnym zbiorem układów, generowano dane przestrzenne dla układu jednostek terytorialnych NUTS 5, które stanowiły hipotetyczne realizacje dwóch szumów przestrzennych X_1 , X_2 o ustalonych własnościach. Dla procesu X_1 założono wartość oczekiwaną równą 30 oraz odchylenie standardowe równe 1,5, a dla procesu X_2 wartość oczekiwaną równą 10 oraz odchylenie standardowe równe 0,5. Procesy X_1 oraz X_2 generowane były w pięciu wariantach, w których założono, że współczynnik korelacji liniowej Pearsona dla generowanych procesów przy-

³ Układ jednostek terytorialnych NUTS 5 składa się z 2479 gmin, natomiast w skład układu jednostek terytorialnych NUTS 4 wchodzi 379 powiatów.

muje kolejno wartości: 0,9; 0,45; 0; -0,45 oraz -0,9. Parametry opisujące własności szumów przestrzennych X_1 oraz X_2 przedstawione zostały w tabeli 5.1.

Rysunek 5.2.

Hipotetyczny jednorodny zbiór układów



Źródło: opracowanie własne.

W kroku drugim procedury symulacyjnej, zgodnie z wzorem (1.1) wykonany został proces agregacji generowanych danych przestrzennych dla układu jednostek terytorialnych NUTS 4. Wagi w_i dla i -tego regionu (gminy) z układu NUTS 5 wyznaczone zostały jako udziały procentowe ludności⁴ w odpowiednim j -tym regionie (powiecie) z układu NUTS 4. W ten sposób otrzymane zostały zbiory dwóch układów danych przestrzennych, pierwszy oparty na układzie jednostek terytorialnych NUTS 5, drugi zaś na układzie NUTS 4.

⁴ Do wyznaczenia udziałów procentowych wykorzystane zostały dane statystyczne dotyczące stanu ludności w 2017 roku. Dane pobrano ze strony internetowej Głównego Urzędu Statystycznego. Wybór wag jako udziałów procentowych ludności wynika z faktu, że wyznaczane wartości średnie badanych zjawisk najczęściej ważone są za pomocą stanu ludności w regionie.

Tabela 5.1.

Parametry generowanych szumów przestrzennych

Proces X_1			Proces X_2		
Wartość oczekiwana $E(X_1)$	Odch. std. $S(X_1)$	Wsp. zm. (%) $V(X_1)$	Wartość oczekiwana $E(X_2)$	Odch. std. $S(X_2)$	Wsp. zm. (%) $V(X_2)$
30,0	1,5	5,0	10,0	0,5	5,0
Brak autokorelacji przestrzennej			Brak autokorelacji przestrzennej		
Korelacja wzajemna (-0,9; -0,45; 0; 0,45; 0,9)					

Źródło: opracowanie własne.

Zgodnie z omówioną procedurą symulacyjną wygenerowano po 10000 realizacji⁵ szumów przestrzennych X_1 oraz X_2 przy założonym wariancie odnośnie do zależności korelacyjnej, a następnie dokonano agregacji generowanych danych przestrzennych. Dla każdej z realizacji procesów X_1 i X_2 obliczono średnią, odchylenie standardowe, statystykę Morana I oraz wartość p testu. Dodatkowo wykonano analizę korelacyjną i regresyjną, w ramach której wyznaczono wartość współczynnika korelacji liniowej Pearsona, wartość kowariancji oraz dokonano estymacji parametrów modelu regresji⁶ dla procesu X_1 względem procesu X_2 . Wykorzystano model postaci:

$$\mathbf{X}_1 = \alpha_0 + \alpha_1 \mathbf{X}_2 + \boldsymbol{\varepsilon}, \quad (5.4)$$

gdzie \mathbf{X}_1 , \mathbf{X}_2 stanowią wektory wartości szumów przestrzennych, α_0 , α_1 oznaczają parametry strukturalne modelu regresji liniowej, a $\boldsymbol{\varepsilon}$ jest wektorem wartości białego szumu przestrzennego.

W wyniku przeprowadzonych obliczeń, otrzymano dla każdego parametru zbiór 10000 ocen, na podstawie których obliczono podstawowe statystyki opisowe: średnią, odchylenie standardowe oraz współczynnik zmienności. Wyznaczenie statystyk opisowych pozwoliło na ocenę zmian w otrzymywanych wynikach dotyczących własności procesów X_1 oraz X_2 na skutek procesu agregacji danych. Rezultaty obliczeń dla kolejnych założonych poziomów zależności korelacyjnej między szumami przestrzennymi zamieszczono w tabelach 5.2–5.6.

⁵ Generowanie większej liczby realizacji nie zmieniło dokładności otrzymywanych wyników.

⁶ Wyznaczone zostały oceny parametrów strukturalnych, współczynnika determinacji oraz współczynnika zmienności losowej. Na podstawie wyznaczonych reszt wykonano także test Morana w celu oceny autokorelacji przestrzennej procesu resztowego.

Tabela 5.2.

Analiza problemu skali dla szumów przestrzennych: wariant pierwszy

Układ NUTS 5								
Analiza podstawowych własności								
Statystyki opisowe	$E(X_1)$	$E(X_2)$	$S(X_1)$	$S(X_2)$	$I(X_1)$	p	$I(X_2)$	p
Średnia	30,00	10,00	1,51	0,50	0,00	0,50	0,00	0,47
Odch. std.	0,03	0,01	0,02	0,01	0,01	0,27	0,01	0,29
Wsp. zm. (%)	0,11	0,11	1,63	1,59	–	–	–	–
Analiza korelacji i regresji								
Statystyki opisowe	Kow.	Kor.	α_1	p	R^2	$V(r)$	$I(r)$	p
Średnia	0,68	0,90	2,71	~0,00	0,81	0,03	0,00	0,59
Odch. std.	0,02	0,00	0,03	~0,00	0,01	0,00	0,01	0,28
Wsp. zm. (%)	3,49	0,49	1,05	–	0,97	0,68	–	–
Układ NUTS 4								
Analiza podstawowych własności								
Statystyki opisowe	$E(X_1)$	$E(X_2)$	$S(X_1)$	$S(X_2)$	$I(X_1)$	p	$I(X_2)$	p
Średnia	30,00	10,00	0,90	0,30	-0,01	0,54	0,00	0,50
Odch. std.	0,05	0,02	0,05	0,01	0,03	0,27	0,03	0,29
Wsp. zm. (%)	0,18	0,17	5,00	4,74	–	–	–	–
Analiza korelacji i regresji								
Statystyki opisowe	Kow.	Kor.	α_1	p	R^2	$V(r)$	$I(r)$	p
Średnia	0,25	0,90	2,71	~0,00	0,81	0,02	0,00	0,52
Odch. std.	0,03	0,01	0,09	~0,00	0,02	0,00	0,03	0,29
Wsp. zm. (%)	10,21	1,34	3,36	–	2,68	2,35	–	–

Uwagi:

Symbol $I(\cdot)$ oznacza statystykę Morana.

Źródło: opracowanie własne.

Wstępna ocena wyników zamieszczonych w tabelach 5.2–5.6 pozwala na stwierdzenie, że współczynnik zmienności dla parametrów istotnie różniących się od zera nie przekraczał poziomu 15%, co świadczy o poprawności wykonanych symulacji. Dodatkowo dla większości parametrów współczynnik zmienności jest mniejszy od 5%. Rozpoczynając podsumowanie przeprowadzonej procedury symulacyjnej należy stwierdzić, że dla wszystkich pięciu wariantów zależności korelacyjnej między procesami X_1 i X_2 otrzymano podobne wyniki pod względem zidentyfikowanych zmian w otrzymywanych ocenach parametrów na skutek procesu agregacji danych.

Tabela 5.3.

Analiza problemu skali dla szumów przestrzennych: wariant drugi

Układ NUTS 5								
Analiza podstawowych własności								
Statystyki opisowe	$E(X_1)$	$E(X_2)$	$S(X_1)$	$S(X_2)$	$I(X_1)$	p	$I(X_2)$	p
Średnia	30,00	10,00	1,50	0,50	0,00	0,49	0,00	0,42
Odch. std.	0,03	0,01	0,02	0,01	0,01	0,25	0,01	0,24
Wsp. zm. (%)	0,10	0,11	1,29	1,31	–	–	–	–
Analiza korelacji i regresji								
Statystyki opisowe	Kow.	Kor.	α_1	p	R^2	$V(r)$	$I(r)$	p
Średnia	0,34	0,45	1,35	~0,00	0,20	0,04	0,00	0,53
Odch. std.	0,02	0,02	0,05	~0,00	0,01	0,00	0,01	0,26
Wsp. zm. (%)	4,57	3,43	3,76	–	6,82	0,67	–	–
Układ NUTS 4								
Analiza podstawowych własności								
Statystyki opisowe	$E(X_1)$	$E(X_2)$	$S(X_1)$	$S(X_2)$	$I(X_1)$	p	$I(X_2)$	p
Średnia	30,00	10,00	0,9	0,3	-0,01	0,52	0,00	0,49
Odch. std.	0,04	0,02	0,04	0,01	0,03	0,27	0,03	0,30
Wsp. zm. (%)	0,15	0,17	4,63	4,83	–	–	–	–
Analiza korelacji i regresji								
Statystyki opisowe	Kow.	Kor.	α_1	p	R^2	$V(r)$	$I(r)$	p
Średnia	0,12	0,45	1,33	~0,00	0,20	0,03	-0,01	0,53
Odch. std.	0,02	0,05	0,18	~0,00	0,05	0,00	0,03	0,27
Wsp. zm. (%)	17,81	12,31	13,35	–	24,73	2,14	–	–

Źródło: opracowanie własne.

W przypadku wartości oczekiwanej procesów X_1 i X_2 , obliczone średnie dla zbioru ocen mają zbliżone wartości na obydwu poziomach agregacji, Oznacza to, że nie występowały systematyczne zmiany w otrzymywanych ocenach dla wartości oczekiwanej $E(X_1)$, $E(X_2)$ na skutek procesu agregacji danych przestrzennych.

Tabela 5.4.

Analiza problemu skali dla szumów przestrzennych: wariant trzeci

Układ NUTS 5								
Analiza podstawowych własności								
Statystyki opisowe	$E(X_1)$	$E(X_2)$	$S(X_1)$	$S(X_2)$	$I(X_1)$	p	$I(X_2)$	p
Średnia	30,01	10,00	1,50	0,50	0,00	0,47	0,00	0,55
Odch. std.	0,03	0,01	0,02	0,01	0,01	0,31	0,01	0,29
Wsp. zm. (%)	0,11	0,11	1,19	1,78	–	–	–	–
Analiza korelacji i regresji								
Statystyki opisowe	Kow.	Kor.	α_1	p	R^2	$V(r)$	$I(r)$	p
Średnia	0,00	0,00	0,00	~0,00	–	0,04	0,00	0,48
Odch. std.	0,02	0,02	0,06	~0,00	–	0,00	0,01	0,31
Wsp. zm. (%)	0,00	0,00	–	–	–	0,60	–	–
Układ NUTS 4								
Analiza podstawowych własności								
Statystyki opisowe	$E(X_1)$	$E(X_2)$	$S(X_1)$	$S(X_2)$	$I(X_1)$	p	$I(X_2)$	p
Średnia	30,00	10,00	0,9	0,30	0,00	0,47	-0,01	0,59
Odch. std.	0,05	0,02	0,04	0,02	0,03	0,30	0,03	0,28
Wsp. zm. (%)	0,17	0,17	4,57	6,62	–	–	–	–
Analiza korelacji i regresji								
Statystyki opisowe	Kow.	Kor.	α_1	p	R^2	$V(r)$	$I(r)$	p
Średnia	0,00	0,01	0,0	~0,00	–	0,03	0,00	0,48
Odch. std.	0,02	0,06	0,19	~0,00	–	0,00	0,03	0,30
Wsp. zm. (%)	–	–	–	–	–	2,25	–	–

Źródło: opracowanie własne.

Obliczone średnie dla zbioru ocen odchylenia standardowego $S(X_1)$, $S(X_2)$ wskazują, że jego poziom ulega obniżeniu przy zmianie z niższego poziomu agregacji na wyższy. Wartości średnich dla odchyłeń standardowych $S(X_1)$, $S(X_2)$ wynoszą kolejno 1,5 oraz 0,5 na poziomie agregacji NUTS 5, a następnie przyjmują wartości 0,9 i 0,3 na poziomie agregacji NUTS 4 dla wszystkich pięciu wariantów zależności korelacyjnej. Wskazuje to na 40% spadek poziomu odchylenia standardowego procesów X_1 i X_2 na skutek procesu agregacji.

Tabela 5.5.

Analiza problemu skali dla szumów przestrzennych: wariant czwarty

Układ NUTS 5								
Analiza podstawowych własności								
Statystyki opisowe	$E(X_1)$	$E(X_2)$	$S(X_1)$	$S(X_2)$	$I(X_1)$	p	$I(X_2)$	p
Średnia	29,99	10,00	1,50	0,50	0,00	0,51	0,00	0,43
Odch. std.	0,03	0,01	0,02	0,01	0,01	0,29	0,01	0,27
Wsp. zm. (%)	0,10	0,11	1,47	1,29	–	–	–	–
Analiza korelacji i regresji								
Statystyki opisowe	Kow.	Kor.	α_1	p	R^2	$V(r)$	$I(r)$	p
Średnia	-0,34	-0,45	-1,35	-0,00	0,20	0,04	0,00	0,51
Odch. std.	0,02	0,02	0,06	-0,00	0,02	0,00	0,01	0,29
Wsp. zm. (%)	-5,65	-4,25	-4,77	–	8,54	0,71	–	–
Układ NUTS 4								
Analiza podstawowych własności								
Statystyki opisowe	$E(X_1)$	$E(X_2)$	$S(X_1)$	$S(X_2)$	$I(X_1)$	p	$I(X_2)$	p
Średnia	30,00	10,00	0,90	0,30	0,00	0,51	0,00	0,48
Odch. std.	0,04	0,01	0,05	0,02	0,03	0,29	0,03	0,31
Wsp. zm. (%)	0,15	0,15	5,81	4,92	–	–	–	–
Analiza korelacji i regresji								
Statystyki opisowe	Kow.	Kor.	α_1	p	R^2	$V(r)$	$I(r)$	p
Średnia	-0,12	-0,45	-1,35	-0,00	0,21	0,03	0,00	0,49
Odch. std.	0,02	0,05	0,16	-0,00	0,04	0,00	0,03	0,28
Wsp. zm. (%)	-14,87	-10,96	-11,66	–	21,61	2,60	–	–

Źródło: opracowanie własne.

Wyznaczone wartości średnie dla statystyki Morana I oraz wartości p testu wskazują, że analizowane procesy X_1 i X_2 charakteryzowały się brakiem własności autokorelacji przestrzennej zarówno na poziomie agregacji NUTS 5, jak i na poziomie agregacji NUTS 4. Wynik ten jest bardzo ważny, ponieważ pozwala na stwierdzenie, że agregacja danych przestrzennych nie przyczynia się do powstawania autozależności przestrzennych na wyższych poziomach agregacji.

Tabela 5.6.

Analiza problemu skali dla szumów przestrzennych: wariant piąty

Układ NUTS 5								
Analiza podstawowych własności								
Statystyki opisowe	$E(X_1)$	$E(X_2)$	$S(X_1)$	$S(X_2)$	$I(X_1)$	ρ	$I(X_2)$	ρ
Średnia	30,00	10,00	1,50	0,50	0,00	0,52	0,00	0,47
Odch. std.	0,04	0,01	0,02	0,01	0,01	0,30	0,01	0,31
Wsp. zm. (%)	0,12	0,11	1,08	1,28	–	–	–	–
Analiza korelacji i regresji								
Statystyki opisowe	Kow.	Kor.	α_1	ρ	R^2	$V(r)$	$I(r)$	ρ
Średnia	–0,68	–0,90	–2,70	–0,00	0,81	0,03	0,00	0,50
Odch. std.	0,02	0,00	0,02	–0,00	0,01	0,00	0,01	0,29
Wsp. zm. (%)	2,51	0,43	0,85	–	0,86	0,79	–	–
Układ NUTS 4								
Analiza podstawowych własności								
Statystyki opisowe	$E(X_1)$	$E(X_2)$	$S(X_1)$	$S(X_2)$	$I(X_1)$	ρ	$I(X_2)$	ρ
Średnia	30,00	10,00	0,91	0,30	0,00	0,53	–0,01	0,54
Odch. std.	0,06	0,02	0,04	0,02	0,03	0,28	0,03	0,29
Wsp. zm. (%)	0,19	0,18	4,55	5,08	–	–	–	–
Analiza korelacji i regresji								
Statystyki opisowe	Kow.	Kor.	α_1	ρ	R^2	$V(r)$	$I(r)$	ρ
Średnia	–0,25	–0,90	–2,71	–0,00	0,81	0,02	–0,01	0,57
Odch. std.	0,03	0,01	0,08	–0,00	0,02	0,00	0,03	0,26
Wsp. zm. (%)	–10,26	–1,43	–2,88	–	2,86	2,23	–	–

Źródło: opracowanie własne.

Kolejne warianty zależności korelacyjnych między procesami X_1 i X_2 pozwoliły na sprawdzenie zmian w otrzymywanych ocenach współczynnika korelacji liniowej Pearsona przy różnym kierunku oraz sile zależności korelacyjnych. Podobnie, jak w przypadku wartości oczekiwanej procesów X_1 i X_2 , nie stwierdzono systematycznych zmian dla otrzymywanych ocen współczynnika korelacji liniowej Pearsona na skutek zmiany poziomu agregacji danych. Obliczane wartości kowariancji ulegały zmianie, jednak było to wynikiem obniżenia się poziomu odchylenia standardowego na wyższym poziomie agregacji.

W ramach wykonanej procedury symulacyjnej oszacowano również parametry modelu regresji liniowej dla procesu X_1 względem procesu X_2 . Obliczone wartości średnie na podstawie ocen parametru α_1 , jak i wartość średnia dla

współczynnika determinacji R^2 nie uległy zmianie na skutek procesu agregacji danych w żadnym z pięciu wariantów analizy. Wyznaczone na podstawie reszt oszacowanych modeli regresji wartości średnie statystyki Morana I oraz wartości p wskazują, że proces resztowy charakteryzował się brakiem autokorelacji przestrzennej.

Należy zwrócić szczególną uwagę na fakt, że obniżeniu wartości średniej ocen odchylenia standardowego $S(X_1)$ oraz $S(X_2)$ na wyższym poziomie agregacji towarzyszył wzrost wartości ich odchyłeń standardowych. Procentowy wzrost wyznaczonych wartości odchylenia standardowego na podstawie zbioru ocen dla odchylenia standardowego procesów X_1 i X_2 przedstawiony został w tabeli 5.7. Najniższy procentowy wzrost wartości odchylenia standardowego zaobserwowany został dla zależności korelacyjnej między procesami X_1 i X_2 na poziomie 0,9. Dla kolejnych poziomów zależności korelacyjnej (0,4; 0; -0,45 oraz -0,9) zaobserwowano systematyczne zwiększanie się procentowego wzrostu wartości odchylenia standardowego. Wpłynęło to na wzrost wartości odchylenia standardowego dla pozostałych parametrów: wartości oczekiwanej procesów X_1 i X_2 , współczynnika korelacji liniowej, współczynnika regresji oraz współczynnika determinacji. Fakt ten oznacza, że obliczane oceny parametrów na wyższym poziomie agregacji obciążone są większą zmiennością.

Tabela 5.7.

Procentowy wzrost wartości odchyłeń standardowych procesów X_1 oraz X_2 (%)

Wariant pierwszy		Wariant drugi		Wariant trzeci		Wariant czwarty		Wariant piąty	
$S(X_1)$	$S(X_2)$	$S(X_1)$	$S(X_2)$	$S(X_1)$	$S(X_2)$	$S(X_1)$	$S(X_2)$	$S(X_1)$	$S(X_2)$
83,88	78,43	116,11	125,07	128,48	125,53	136,23	132,93	155,00	140,27

Źródło: opracowanie własne.

W wyniku przeprowadzonej analizy stwierdzono, że w przypadku szumów przestrzennych agregacja danych nie powoduje istotnych zmian w ocenach wartości oczekiwanej poszczególnych procesów, ani też w ocenie korelacyjnej bądź regresyjnej. Należy jednak pamiętać, że na wyższych poziomach agregacji mogą wystąpić pewne różnice ze względu na wyższą zmienność otrzymywanych ocen parametrów.

5.3. Identyfikacja problemu skali w przypadku występowania trendu przestrzennego

W przypadku rozpatrywanych w poprzednim podrozdziale procesów przestrzennych zakładano stały poziom wartości oczekiwanej, wariancji oraz brak autokorelacji przestrzennej. Należy podkreślić, że własności przestrzennych procesów ekonomicznych najczęściej nie spełniają tych założeń i charaktery-

zują się występowaniem systematycznej niejednorodności lub autokorelacji przestrzennej (Zeliaś, 1991, s. 15–23, 97–116; Suchecki, 2010, s. 107–128, 255–266). W pierwszej sytuacji systematyczna niejednorodność może przejawiać się w postaci systematycznych zmian w poziomie wartości oczekiwanej lub systematycznych zmian w poziomie wariancji oraz funkcji autokowariancji przestrzennej. Rozważany wówczas proces jest niestacjonarny przestrzennie. Natomiast w sytuacji drugiej proces przestrzenny jest stacjonarny w szerszym sensie.

W niniejszym podrozdziale autor rozważy problem skali dla procesów przestrzennych, których jedynym deterministycznym składnikiem wewnętrznej struktury jest niejednorodność systematyczna. Autor skupi się na analizie problemu skali wyłącznie dla przypadku możliwości opisu własności niejednorodności systematycznej za pomocą modelu trendu przestrzennego. Problematyka związana z trendami przestrzennymi przedstawiona została kompleksowo w pracach Czyż (1978, s. 10–43, 51–81), Szulc (2007, s. 102–103), Pietrzak (2010b, s. 324–328), Suchecki (2010, s. 214–220).

Trend przestrzenny określany jest jako ogólna tendencja przestrzenna w zmianach poziomu wartości oczekiwanej procesu. W związku z tym w modelu trendu przestrzennego wykorzystywane są funkcje wielomianowe. Ze względu na postać analityczną wyróżnić można przestrzenny trend stopnia pierwszego (trend płaszczyznowy), przestrzenny trend stopnia drugiego (trend kwadratowy) oraz przestrzenne trendy wyższych stopni. Wykorzystanie funkcji wielomianowych posiada dwie podstawowe zalety. Do pierwszej z nich należy zaliczyć możliwość dość dużego stopnia dopasowania funkcji tego typu do danych empirycznych, do drugiej zaś łatwą interpretację estymowanych parametrów w przypadku modelu trendu płaszczyznowego lub modelu trendu kwadratowego.

W przypadku interpretacji parametrów modelu trendu pierwszego stopnia ustalany jest systematyczny wzrost albo spadek poziomu procesu przestrzennego w określonych kierunkach geograficznych lub stały poziom procesu przy przesunięciu przestrzennym w danym kierunku. O kierunku systematycznych zmian średniego poziomu procesu przestrzennego oraz ich sile świadczą otrzymane oceny parametrów strukturalnych modelu trendu przestrzennego. Systematyczny wzrost lub spadek poziomu analizowanego zjawiska w określonym kierunku wynika z długookresowych tendencji w rozwoju społeczno-ekonomicznym analizowanych regionów. W pracy Pietrzak (2010 b, s. 325) model trendu stopnia pierwszego określony został jako model trendu kierunkowego, ponieważ za pomocą tego modelu opisywany jest główny kierunek przestrzennego rozwoju badanego zjawiska.

Systematyczne zmiany średniego poziomu procesu przestrzennego mogą koncentrować się również wokół dominującej jednostki terytorialnej (np. silny ośrodek miejski), do opisu których stosowany jest model przestrzennego trendu kwadratowego. W takiej sytuacji analizowane zjawisko będzie charak-

teryzować się najwyższym poziomem dla regionu dominującego i wraz z rosnącą odległością od centrum poziom zjawiska będzie systematycznie maleć. Możliwa jest również sytuacja odwrotna, gdzie ze względu na sytuację społeczno-ekonomiczną regionów, w dominującej jednostce terytorialnej poziom zjawiska będzie najniższy i wraz ze wzrostem odległości od centrum poziom zjawiska będzie wzrastał. W pracy Pietrzak (2010b, s. 327) model trendu kwadratowego, gdzie opisywane są zmiany przestrzennego rozwoju badanego zjawiska wokół dominującego centrum, określono jako model trendu centralnego.

Wobec perspektywy pojawiających się w strukturze analizowanych procesów trendów przestrzennych, problem skali rozważony został ponownie na podstawie analizy symulacyjnej dla procesów X_1 oraz X_2 . W ramach ustalonej procedury dane generowane były w oparciu o model przestrzennego trendu stopnia pierwszego, określony za pomocą równania:

$$\mathbf{X} = \beta_0 + \beta_1 s_1 + \beta_2 s_2 + \boldsymbol{\varepsilon}, \quad (5.5)$$

gdzie \mathbf{X} jest wektorem wartości procesu przestrzennego o własności niejednorodności systematycznej, $\beta_0, \beta_1, \beta_2$ to parametry trendu przestrzennego, s_1, s_2 to współrzędne lokalizacji na płaszczyźnie, a $\boldsymbol{\varepsilon}$ jest wektorem wartości białego szumu przestrzennego.

Założenie modelu przestrzennego trendu stopnia pierwszego wynika z faktu, że za pomocą tego modelu można w łatwy sposób generować realizacje procesu przestrzennego o określonym kierunku przestrzennych zmian w wartości oczekiwanej oraz o ustalonej sile natężenia tych zmian. Dodatkowo założona zostanie zależność korelacyjna między procesami, co pozwoli na przyjęcie specyfikacji modelu regresji rozszerzonego o trend przestrzenny. Model ten przyjmuje następującą postać:

$$\mathbf{X}_1 = \beta_0 + \beta_1 s_1 + \beta_2 s_2 + \alpha_1 \mathbf{X}_2 + \boldsymbol{\varepsilon}, \quad (5.6)$$

gdzie $\mathbf{X}_1, \mathbf{X}_2$ oznaczają wektory wartości procesów przestrzennych, $\beta_0, \beta_1, \beta_2$ to parametry trendu przestrzennego, α_1 stanowi parametr regresji liniowej, s_1, s_2 to współrzędne lokalizacji na płaszczyźnie, a $\boldsymbol{\varepsilon}$ odnosi się do wartości białego szumu przestrzennego.

W pierwszym kroku procedury symulacyjnej generowane były realizacje procesów przestrzennych X_1, X_2 na poziomie agregacji NUTS 5. Dla obydwu procesów założono własność systematycznej niejednorodności, opisaną przy wykorzystaniu modelu przestrzennego trendu pierwszego stopnia zgodnego z równaniem (5.5). W tabeli 5.8 podane zostały przyjęte wartości parametrów strukturalnych dla trendu przestrzennego. Dla modeli obydwu procesów ustalono wielkość zmienności reszt w takim zakresie, który pozwalał na otrzymanie stopnia dopasowania modelu do danych na poziomie około 80%. Dla każdego z procesów założono również brak własności autorelacji przestrzennej.

Tabela 5.8.

Własności procesów przestrzennych

Parametry przestrzennego trendu liniowego							
Proces X_1				Proces X_2			
$\beta_0(X_1)$	$\beta_1(X_1)$	$\beta_2(X_1)$	R^2	$\beta_0(X_2)$	$\beta_1(X_2)$	$\beta_2(X_2)$	R^2
30,0	1,5	3,0	0,8	10,0	0,5	1,0	0,8
Brak autokorelacji przestrzennej składnika losowego							
Korelacja wzajemna (-0,9; -0,45; 0; 0,45; 0,9)							

Źródło: opracowanie własne.

Realizacje procesów przestrzennych X_1 oraz X_2 generowane były w pięciu wariantach. W wariantach pierwszym założono silną, dodatnią zależność korelacyjną między procesami na poziomie 0,9, a w kolejnych wariantach współczynnik korelacji liniowej Pearsona przyjmował wartości: 0,45; 0; -0,45 oraz -0,9. Dla każdego z pięciu wariantów wygenerowano 10000 realizacji procesów przestrzennych. W ten sposób otrzymano po zbiór układów danych przestrzennych, na które składały się generowane dane przestrzenne odniesione do układu jednostek terytorialnych NUTS 5.

W drugim kroku dokonano agregacji generowanych danych przestrzennych do układu jednostek terytorialnych NUTS 4. Agregacja danych przestrzennych wykonana została zgodnie z równaniem (1.1). Ponownie wartości wag w_i ustalono na podstawie udziałów procentowych ludności. W ten sposób dla każdego z wariantów zależności korelacyjnej otrzymano kolejne 10000 układów danych przestrzennych, na które składały się zagregowane dane przestrzenne odniesione do układu jednostek terytorialnych NUTS 4.

W trzecim kroku procedury symulacyjnej przeprowadzono ocenę własności procesów przestrzennych X_1 , X_2 na podstawie zbiorów generowanych danych przestrzennych (zob. tabele 5.9–5.13). Najpierw dokonano estymacji parametrów modeli przestrzennego trendu stopnia pierwszego na obydwu poziomach agregacji, osobno dla procesu X_1 oraz procesu X_2 . Następnie dla każdego z poziomów agregacji oszacowano parametry modelu regresji poszerzonego o przestrzenne składnik trendu zgodnie z równaniem (5.6). Dla wszystkich modeli sprawdzono również występowanie autokorelacji przestrzennej dla procesu resztowego za pomocą testu Morana. W ten sposób otrzymano dla każdego parametru zbiór ocen, na podstawie których policzono średnią, odchylenie standardowe oraz współczynnik zmienności. Obliczone wartości statystyk dla kolejnych wariantów zależności korelacyjnej między procesami przestrzennymi X_1 oraz X_2 zamieszczono w tabelach 5.9–5.13. Porównanie wyników w zakresie ocen parametrów strukturalnych na obydwu poziomach agregacji, zarówno dla modeli trendu przestrzennego, jak i modeli regresji

z trendem przestrzennym, pozwoliło na stwierdzenie, czy ewentualne zmiany są konsekwencją rozważanego problemu skali.

Tabela 5.9.

Analiza problemu skali dla procesów X_1 i X_2 ; wariant pierwszy

Układ NUTS 5										
Analiza trendów przestrzennych										
Statystyki opisowe	$\beta_1(X_1)$	p	$\beta_2(X_1)$	p	R^2	$\beta_1(X_2)$	p	$\beta_2(X_2)$	p	R^2
Średnia	1,50	-0,00	3,00	-0,00	0,82	0,50	-0,00	1,00	-0,00	0,82
Odch. std.	0,03	-0,00	0,03	-0,00	0,00	0,01	-0,00	0,01	-0,00	0,01
Wsp. zm. (%)	2,16	-	0,95	-	0,61	2,23	-	0,97	-	0,65
Analiza regresji										
Statystyki opisowe	α_1	p	β_1	p	β_2	p	R^2	V	$I(r)$	p
Średnia	2,71	-0,00	0,15	-0,00	0,30	-0,00	0,97	0,01	-0,00	0,52
Odch. std.	0,02	-0,00	0,02	-0,00	0,03	-0,00	0,00	0,00	0,01	0,27
Wsp. zm. (%)	0,79	-	10,24	-	8,77	-	0,13	0,75	-	-
Układ NUTS 4										
Analiza trendów przestrzennych										
Statystyki opisowe	$\beta_1(X_1)$	p	$\beta_2(X_1)$	p	R^2	$\beta_1(X_2)$	p	$\beta_2(X_2)$	p	R^2
Średnia	1,50	-0,00	3,00	-0,00	0,93	0,50	-0,00	1,00	-0,00	0,93
Odch. std.	0,05	-0,00	0,05	-0,00	0,01	0,02	-0,00	0,02	-0,00	0,01
Wsp. zm. (%)	3,14	-	1,59	-	0,57	3,04	-	1,62	-	0,62
Analiza regresji										
Statystyki opisowe	α_1	p	β_1	p	β_2	p	R^2	V	$I(r)$	p
Średnia	2,71	-0,00	0,15	-0,00	0,29	-0,00	0,99	0,01	-0,01	0,53
Odch. std.	0,08	-0,00	0,04	-0,00	0,08	-0,00	0,00	0,00	0,03	0,26
Wsp. zm. (%)	2,89	-	25,52	-	27,33	-	0,16	2,76	-	-

Źródło: opracowanie własne.

W tabelach 5.9.–5.13. przedstawiono wyznaczone na podstawie zbiorów ocen otrzymanych w wyniku wykonanej procedury symulacyjnej wartości średnich tych ocen, ich odchyłeń standardowych oraz współczynników zmienności. Dla wszystkich pięciu wariantów zależności korelacyjnej między procesami X_1 i X_2 otrzymano podobne wyniki w sensie identyfikacji zmian w uzyskiwanych rezultatach lub braku systematycznych zmian.

Tabela 5.10.

Analiza problemu skali dla procesów X_1 i X_2 ; wariant drugi

Układ NUTS 5										
Analiza trendów przestrzennych										
Statystyki opisowe	$\beta_1(X_1)$	p	$\beta_2(X_1)$	p	R^2	$\beta_1(X_2)$	p	$\beta_2(X_2)$	p	R^2
Średnia	1,50	-0,00	3,01	-0,00	0,82	0,50	-0,00	1,00	-0,00	0,82
Odch. std.	0,03	-0,00	0,03	-0,00	0,00	0,01	-0,00	0,01	-0,00	0,01
Wsp. zm. (%)	1,94	-	0,96	-	0,56	2,17	-	1,13	-	0,68
Analiza regresji										
Statystyki opisowe	α_1	p	β_1	p	β_2	p	R^2	V	$I(r)$	p
Średnia	1,36	-0,00	0,82	-0,00	1,65	-0,00	0,85	0,02	-0,00	0,40
Odch. std.	0,06	-0,00	0,04	-0,00	0,06	-0,00	0,00	0,00	0,01	0,27
Wsp. zm. (%)	4,13	-	4,30	-	3,76	-	0,39	0,58	-	-
Układ NUTS 4										
Analiza trendów przestrzennych										
Statystyki opisowe	$\beta_1(X_1)$	p	$\beta_2(X_1)$	p	R^2	$\beta_1(X_2)$	p	$\beta_2(X_2)$	p	R^2
Średnia	1,50	-0,00	3,00	-0,00	0,93	0,50	-0,00	1,00	-0,00	0,93
Odch. std.	0,04	-0,00	0,04	-0,00	0,01	0,02	-0,00	0,02	-0,00	0,01
Wsp. zm. (%)	2,45	-	1,40	-	0,73	3,46	-	1,58	-	0,73
Analiza regresji										
Statystyki opisowe	α_1	p	β_1	p	β_2	p	R^2	V	$I(r)$	p
Średnia	1,39	-0,00	0,81	-0,00	1,62	-0,00	0,94	0,02	0,00	0,50
Odch. std.	0,17	-0,00	0,09	-0,00	0,18	-0,00	0,01	0,00	0,04	0,32
Wsp. zm. (%)	11,91	-	11,70	-	10,98	-	0,59	2,43	-	-

Źródło: opracowanie własne.

W przypadku estymacji modeli trendu przestrzennego osobno dla procesu oraz procesu X_1 , X_2 obliczone średnie ocen parametrów $\beta_1(X_1)$, $\beta_2(X_1)$ oraz $\beta_1(X_2)$, $\beta_2(X_2)$ mają zbliżone wartości na obydwu poziomach agregacji. Oznacza to, że nie występowały systematyczne zmiany otrzymywanych ocen parametrów modeli przestrzennego trendu stopnia pierwszego na skutek procesu agregacji danych przestrzennych. Należy podkreślić, że wartości współczynnika determinacji R^2 dla modeli uległy podwyższeniu z wartości 0,82 na poziomie agregacji NUTS 5 do 0,93 na poziomie agregacji NUTS 4. Jest to wynikiem zaobserwowanego w poprzednim podrozdziale systematycznego spadku poziomu odchylenia standardowego procesów na skutek agregacji danych. Wykonany test Morana w odniesieniu do reszt obydwu modeli trendu przestrzennego

wskazał na brak własności autokorelacji przestrzennej dla procesów resztowych na obydwu poziomach agregacji.

Tabela 5.11.

Analiza problemu skali dla procesów X_1 i X_2 ; wariant trzeci

Układ NUTS 5										
Analiza trendów przestrzennych										
Statystyki opisowe	$\beta_1(X_1)$	p	$\beta_2(X_1)$	p	R^2	$\beta_1(X_2)$	p	$\beta_2(X_2)$	p	R^2
Średnia	1,50	~0,00	3,00	~0,00	0,82	0,50	~0,00	1,00	~0,00	0,82
Odch. std.	0,03	~0,00	0,03	~0,00	0,01	0,01	~0,00	0,01	~0,00	0,01
Wsp. zm. (%)	1,90	-	0,93	-	0,62	1,95	-	0,96	-	0,66
Analiza regresji										
Statystyki opisowe	α_1	p	β_1	p	β_2	p	R^2	V	$I(r)$	p
Średnia	-0,01	~0,00	1,50	~0,00	3,01	~0,00	0,82	0,02	~0,00	0,46
Odch. std.	0,07	~0,00	0,04	~0,00	0,07	~0,00	0,01	0,00	0,01	0,27
Wsp. zm. (%)	-882,94	-	2,87	-	2,37	-	0,62	0,68	-	-
Układ NUTS 4										
Analiza trendów przestrzennych										
Statystyki opisowe	$\beta_1(X_1)$	p	$\beta_2(X_1)$	p	R^2	$\beta_1(X_2)$	p	$\beta_2(X_2)$	p	R^2
Średnia	1,50	~0,00	2,99	~0,00	0,93	0,50	~0,00	1,00	~0,00	0,93
Odch. std.	0,03	~0,00	0,05	~0,00	0,01	0,02	~0,00	0,02	~0,00	0,01
Wsp. zm. (%)	2,32	-	1,72	-	0,75	3,02	-	1,59	-	0,73
Analiza regresji										
Statystyki opisowe	α_1	p	β_1	p	β_2	p	R^2	V	$I(r)$	p
Średnia	-0,03	~0,00	1,51	~0,00	3,02	~0,00	0,93	0,02	-0,01	0,55
Odch. std.	0,22	~0,00	0,12	~0,00	0,23	~0,00	0,01	0,00	0,03	0,25
Wsp. zm. (%)	-690,32	-	7,89	-	7,61	-	0,76	2,24	-	-

Źródło: opracowanie własne.

W przypadku modelu trendowo-regresyjnego dla procesu X_1 względem procesu X_2 , zarówno na poziomie agregacji NUTS 5, jak i na poziomie agregacji NUTS 4 nie stwierdzono różnic w wyznaczanych ocenach współczynnika regresji α_1 oraz parametrów trendu, tj.: β_1 , β_2 . W przypadku współczynnika determinacji R^2 , obliczone średnie wartości ocen uległy podwyższeniu w wyniku agregacji danych, co jest konsekwencją wyższego dopasowania funkcji trendu do danych empirycznych.

Tabela 5.12.

Analiza problemu skali dla procesów X_1 i X_2 ; wariant czwarty

Układ NUTS 5										
Analiza trendów przestrzennych										
Statystyki opisowe	$\beta_1(X_1)$	p	$\beta_2(X_1)$	p	R^2	$\beta_1(X_2)$	p	$\beta_2(X_2)$	p	R^2
Średnia	1,50	-0,00	3,00	-0,00	0,82	0,50	-0,00	1,00	-0,00	0,82
Odch. std.	0,03	-0,00	0,03	-0,00	0,00	0,01	-0,00	0,01	-0,00	0,01
Wsp. zm. (%)	1,74	-	0,98	-	0,56	1,90	-	1,04	-	0,75
Analiza regresji										
Statystyki opisowe	α_1	p	β_1	p	β_2	p	R^2	V	$I(r)$	p
Średnia	-1,34	-0,00	2,17	-0,00	4,34	-0,00	0,85	0,02	-0,00	0,46
Odch. std.	0,06	-0,00	0,04	-0,00	0,07	-0,00	0,00	0,00	0,01	0,29
Wsp. zm. (%)	-4,51	-	1,82	-	1,62	-	0,48	0,76	-	-
Układ NUTS 4										
Analiza trendów przestrzennych										
Statystyki opisowe	$\beta_1(X_1)$	p	$\beta_2(X_1)$	p	R^2	$\beta_1(X_2)$	p	$\beta_2(X_2)$	p	R^2
Średnia	1,50	-0,00	2,99	-0,00	0,93	0,50	-0,00	1,00	-0,00	0,93
Odch. std.	0,04	-0,00	0,05	-0,00	0,01	0,01	-0,00	0,02	-0,00	0,01
Wsp. zm. (%)	2,89	-	1,60	-	0,68	2,75	-	1,56	-	0,66
Analiza regresji										
Statystyki opisowe	α_1	p	β_1	p	β_2	p	R^2	V	$I(r)$	p
Średnia	-1,33	-0,00	2,17	-0,00	4,32	-0,00	0,94	0,02	0,00	0,49
Odch. std.	0,19	-0,00	0,10	-0,00	0,20	-0,00	0,01	0,00	0,03	0,28
Wsp. zm. (%)	-14,50	-	4,76	-	4,57	-	0,57	2,52	-	-

Źródło: opracowanie własne.

Wyznaczone na podstawie reszt wartości średnie ocen statystyki Morana I oraz wartości p nie wskazały na występowanie własności autokorelacji przestrzennej procesu resztowego na obydwu poziomach agregacji. Oznacza to, że agregacja danych przestrzennych nie przyczynia się do powstawania własności autokorelacji przestrzennej w przypadku procesu objaśnianego zawierającego w swej strukturze składnik niejednorodności systematycznej.

Tabela 5.13.

Analiza problemu skali dla procesów X_1 i X_2 ; wariant piąty

Układ NUTS 5										
Analiza trendów przestrzennych										
Statystyki opisowe	$\beta_1(X_1)$	p	$\beta_2(X_1)$	p	R^2	$\beta_1(X_2)$	p	$\beta_2(X_2)$	p	R^2
Średnia	1,50	~0,00	3,00	~0,00	0,82	0,50	~0,00	1,00	~0,00	0,82
Odch. std.	0,03	~0,00	0,03	~0,00	0,00	0,01	~0,00	0,01	~0,00	0,00
Wsp. zm. (%)	1,86	–	0,96	–	0,53	1,82	–	1,04	–	0,57
Analiza regresji										
Statystyki opisowe	α_1	p	β_1	p	β_2	p	R^2	V	$I(r)$	p
Średnia	-2,70	~0,00	2,85	~0,00	5,71	~0,00	0,97	0,01	~0,00	0,49
Odch. std.	0,03	~0,00	0,02	~0,00	0,03	~0,00	0,00	0,00	0,01	0,32
Wsp. zm. (%)	-1,05	–	0,59	–	0,51	–	0,12	0,83	–	–
Układ NUTS 4										
Analiza trendów przestrzennych										
Statystyki opisowe	$\beta_1(X_1)$	p	$\beta_2(X_1)$	p	R^2	$\beta_1(X_2)$	p	$\beta_2(X_2)$	p	R^2
Średnia	1,50	~0,00	2,99	~0,00	0,93	0,50	~0,00	1,00	~0,00	0,93
Odch. std.	0,05	~0,00	0,05	~0,00	0,01	0,01	~0,00	0,02	~0,00	0,01
Wsp. zm. (%)	3,04	–	1,59	–	0,83	2,85	–	1,58	–	0,75
Analiza regresji										
Statystyki opisowe	α_1	p	β_1	p	β_2	p	R^2	V	$I(r)$	p
Średnia	-2,70	~0,00	2,85	~0,00	5,69	~0,00	0,99	0,01	0,00	0,52
Odch. std.	0,09	~0,00	0,05	~0,00	0,09	~0,00	0,00	0,00	0,03	0,25
Wsp. zm. (%)	-3,38	–	1,85	–	1,55	–	0,13	1,99	–	–

Źródło: opracowanie własne.

Na podstawie przeprowadzonej analizy symulacyjnej nie stwierdzono występowania istotnych zmian w otrzymywanych wynikach badań, przy wykorzystaniu modelu regresji poszerzonego o przestrzenny trend stopnia pierwszego. Oznacza to brak identyfikacji problemu skali w tym przypadku. Analiza regresji w sytuacji uwzględnienia własności procesów w postaci trendu przestrzennego powinna gwarantować uzyskanie podobnych rezultatów, bez względu na wybrany poziom agregacji danych przestrzennych. Należy jednak pamiętać, że na różnych poziomach agregacji mogą wystąpić różnice oszacowań ze względu na inną liczbę stopni swobody, w zależności od wybranego układu jednostek terytorialnych.

5.4 Analiza wpływu autokorelacji przestrzennej na problem skali

5.4.1. Identyfikacja struktury przestrzennej

W przestrzennych badaniach ekonomicznych badacz poprzez analizę własności procesów oraz współzależności między nimi identyfikuje strukturę przestrzenną analizowanego obszaru. Struktura przestrzenna obszaru określona została przez autora, jako ogół istniejących zależności przestrzennych na ustalonym obszarze, w szczególności własności analizowanych procesów oraz zależności między nimi. W ramach stawianego problemu badawczego uwaga koncentrowana jest na badaniu zależności ekonomicznych pomiędzy wybranymi procesami dla ustalonego obszaru analizy, co oznacza odkrywanie jedynie części struktury przestrzennej obszaru.

Własności procesów oraz zależności ekonomiczne identyfikowane są na podstawie analizy danych przestrzennych, ponieważ dane te odzwierciedlają oddziaływanie przyczyn wynikających ze struktury przestrzennej obszaru. W takiej sytuacji widoczna jest konieczność przyjęcia jednorodnego zbioru układów, ponieważ tylko dla wybranych układów jednostek terytorialnych, odniesione do nich dane przestrzenne posiadają własność jednorodności przyczynowej. Należy podkreślić, że analizowane w badaniach zależności ekonomiczne stanowią odzwierciedlenie rzeczywistej struktury przestrzennej. Na prawidłową identyfikację własności procesów oraz zależności wynikających z istniejącej struktury przestrzennej obszaru powinno pozwolić zastosowanie modeli ekonometrycznych. W związku z otrzymanymi wynikami estymacji parametrów przyjętych modeli, możliwa jest symulacja na ich podstawie realizacji procesów o ustalonych własnościach. Należy jednak podkreślić fakt, że generowanie realizacji procesów o odpowiednich własnościach nie gwarantuje, że odtworzą one we właściwy sposób strukturę przestrzenną analizowanego obszaru.

Istotą przedstawionych w poprzednich podrozdziałach procedur symulacyjnych była konieczność wielokrotnego generowania realizacji procesów o założonych własnościach. Ważne jest przy tym, by generowane kolejno realizacje procesów odzwierciedlały we właściwy sposób strukturę przestrzenną obszaru. Problem ten podjęty został dopiero teraz, ponieważ generowane realizacje procesów przestrzennych w podrozdziałach 5.2 oraz 5.3 pozwalały na właściwe odzwierciedlenie struktury przestrzennej. W podrozdziale 5.2 generowano realizacje procesów szumów przestrzennych, które charakteryzowały się stałym poziomem wartości oczekiwanej i wariancji oraz brakiem autokorelacji przestrzennej. Zmienność procesu o takich własnościach wynika ze struktury przestrzennej obszaru, której charakter w sensie poziomu wartości oczekiwanej oraz wariancji jest taki sam w granicach rozważanego obszaru, bez względu na wybór jednostki terytorialnej. W związku z tym można stwierdzić, że dla wygenerowanych zgodnie z równaniem (5.3) kolejnych realizacji procesu,

każda z tych realizacji jest odzwierciedleniem struktury przestrzennej o takim samym charakterze. Podobny wniosek można wyciągnąć dla generowanych realizacji dwóch skorelowanych procesów przestrzennych o wymienionych własnościach. Również w tym przypadku można założyć, że symulowane realizacje stanowią odzwierciedlenie struktury przestrzennej, gdzie występuje stały charakter zależności pomiędzy procesami na ustalonym obszarze.

Z kolei w podrozdziale 5.3 generowano realizacje procesów przestrzennych, które charakteryzowały się systematycznymi zmianami przestrzennymi w poziomie wartości oczekiwanej. W ramach procedury symulacyjnej generowane były realizacje procesów przestrzennych w oparciu o model przestrzennego trendu stopnia pierwszego określonego równaniem (5.5). Autor uważa, że również w tym wypadku generowane wielokrotnie realizacje procesów są odzwierciedleniem struktury przestrzennej o takim samym charakterze, która wyraża systematyczny wzrost lub spadek poziomu analizowanego zjawiska w określonym kierunku, wynikający z długookresowych tendencji w rozwoju społeczno-ekonomicznym całego obszaru. W wyniku wykorzystania modelu trendu przestrzennego, zmiany wartości oczekiwanej kształtują się systematycznie w ten sam sposób dla każdej generowanej realizacji procesu.

Należy jednak zwrócić uwagę, że w pracach z zakresu przestrzennych badań ekonomicznych wskazywane jest, że dla części analizowanych zjawisk identyfikowane są autozależności przestrzenne (Zeliaś, 1991, s. 97–116; Suchecki, 2010, s. 107–128). W takiej sytuacji badane procesy charakteryzują się własnością autokorelacji przestrzennej. Dlatego w niniejszym podrozdziale uwaga skupiona została na analizie problemu skali w przypadku występowania tej własności. Na opis procesu X charakteryzującego się własnością autokorelacji przestrzennej pozwala przyjęcie specyfikacji modelu autoregresji przestrzennej SAR (Suchecki, 2010, s. 248–250), tj. :

$$\mathbf{X} = \alpha_0 + q\mathbf{W}\mathbf{X} + \boldsymbol{\varepsilon}, \quad (5.7)$$

gdzie \mathbf{X} jest wektorem wartości procesu przestrzennego, α_0 jest parametrem reprezentującym poziom wartości oczekiwanej procesu, q stanowi parametr autoregresji przestrzennej, \mathbf{W} oznacza standaryzowaną macierz sąsiedztwa pierwszego rzędu, a $\boldsymbol{\varepsilon}$ wektor wartości białego szumu przestrzennego.

W przypadku rozpatrywanych procesów przestrzennych charakteryzujących się własnością autokorelacji przestrzennej, generowanie kolejnych realizacji na podstawie równania (5.7) nie gwarantuje odzwierciedlenia struktury przestrzennej o tym samym charakterze, ponieważ własność autokorelacji przestrzennej jest naturalnym następstwem wcześniejszego występowania niejednorodności systematycznej procesu na analizowanym obszarze. Wspomniana niejednorodność systematyczna rozumiana jest tutaj w tym sensie, że możliwe jest wyróżnienie jednostek terytorialnych, które były dominujące w początkowym okresie rozwoju społeczno-gospodarczego badanego obszaru.

Następnie na przestrzeni lat sąsiadujące z ośrodkami dominującymi jednostki terytorialne upodobały się do nich pod względem rozwoju, a oddalone coraz bardziej się odróżniały. Ostatecznie początkowe występowanie, określonej w ten sposób niejednorodności systematycznej, przekłada się w wyniku długo-okresowego rozwoju obszaru na fakt identyfikacji autozależności przestrzennych. W związku z tym, w przypadku modelowania własności autokorelacji przestrzennej procesów pojawia się problem z generowaniem realizacji procesu na podstawie równania (5.7), ponieważ zidentyfikowana własność autokorelacji wynika z konkretnej postaci struktury przestrzennej. Zastosowanie modelu autoregresji przestrzennej SAR pozwala na identyfikację autokorelacji przestrzennej dla wybranych procesów. Natomiast przy znajomości oszacowanych ocen parametrów, wykorzystanie modelu SAR nie pozwala na generowanie danych przestrzennych, których przestrzenna zmienność zgadzałaby się z rzeczywistą strukturą przestrzenną obszaru. Wielokrotne symulacje na podstawie modelu z autoregresją przestrzenną doprowadzą do otrzymania zbioru realizacji procesu, gdzie każda z realizacji mogłaby wynikać z odmiennej struktury przestrzennej obszaru. Tymczasem analiza wpływu własności autokorelacji przestrzennej na problem skali powinna dotyczyć badania konkretnej struktury przestrzennej. Fakt ten wskazuje na trudność w przygotowaniu procedury symulacyjnej, gdzie konieczne jest, żeby generowane realizacje były odzwierciedleniem tej samej struktury przestrzennej.

W związku z powyższym w celu generowania realizacji procesów posiadających ustalone własności i jednocześnie będących odzwierciedleniem tej samej struktury przestrzennej autor proponuje posłużyć się empirycznymi rozkładami procesów w ramach przyjętego obszaru analizy. W proponowanej procedurze symulacyjnej wzięto pod uwagę dwa procesy przestrzenne na poziomie agregacji NUTS 5: średnie zużycie nawozów azotowych oraz średnią powierzchnię gospodarstw rolnych⁷. Na podstawie empirycznych wartości procesów wyznaczono kolejne centyle: centyl dziesiąty, centyl trzydziesty, centyl pięćdziesiąty, centyl siedemdziesiąty i centyl dziewięćdziesiąty, a następnie na ich podstawie ustalono sześć przedziałów wartości. Osobno dla każdego z dwóch rozważanych procesów, przyporządkowano poszczególne jednostki terytorialne do jednej z sześciu klas zgodnie z przynależnością do odpowiedniego przedziału wartości centyli⁸. W ten sposób, zarówno dla procesu zużycia nawozów azotowych, jak i procesu powierzchni gospodarstw rolnych, uzyskano dyskretne rozkłady jednostek terytorialnych dla układu NUTS 5. Wykorzystanie ustalonych rozkładów jednostek terytorialnych pozwoliło na wygenerowanie realizacji procesów przy wykorzystaniu przyjętego rozkładu losowego (na przykład rozkładu normalnego, czy rozkładu gamma). Generowanie danych przestrzennych

⁷ Dane przestrzenne pochodzą z Banku Danych Lokalnych Głównego Urzędu Statystycznego. Dane pozyskane zostały w związku z przeprowadzonym w 2010 roku Powszechnym Spisie Rolnym.

⁸ Na przykład jednostka terytorialna przyporządkowana została do pierwszej klasy, jeżeli wartość procesu była mniejsza od centyla dziesiątego.

dla wybranej jednostki terytorialnej polegało na sprawdzeniu, do której klasy została ona przyporządkowana, a następnie symulacji wartości zmiennej losowej z odpowiedniego przedziału dystrybuanty wybranego rozkładu losowego⁹. W ten sposób możliwe jest generowanie realizacji dla ustalonego w badaniu procesu przestrzennego przy jednoczesnym uwzględnieniu struktury przestrzennej, z której wynika przestrzenne zróżnicowanie tego procesu.

5.4.2. Wybór rozkładu losowego oraz systemu wag w procedurze symulacyjnej

W ramach zaproponowanej procedury symulacyjnej opartej na empirycznym rozkładzie procesów przestrzennych podjęta została próba oceny wpływu dwóch kwestii na jakość odwzorowywanej struktury przestrzennej. Pierwsza kwestia to wybór rozkładu losowego wykorzystanego do generowania realizacji procesów, druga zaś to wybór systemu wag. Rozpatrzenie wpływu obydwu determinant w podjętym problemie badawczym pozwoli na ustalenie konkretnego rozkładu losowego oraz systemu wag, które gwarantują generowanie poprawnych realizacji. Generowane w takiej sytuacji dane przestrzenne powinny stanowić odzwierciedlenie rzeczywistej struktury zależności przestrzennych.

Problem wyboru rozkładu losowego rozpatrzony został na podstawie dwóch rozkładów: rozkładu normalnego oraz rozkładu gamma. W tabelach 5.14 i 5.15 przedstawiono statystyki opisowe rozkładu zużycia nawozów azotowych oraz powierzchni gospodarstw rolnych. Otrzymane wartości statystyk opisowych wskazują na bardzo wysoki, przekraczający 50% udział odchylenia standardowego w średniej. Dodatkowo występuje dodatnia skośność obydwu rozkładów. Uzyskane statystyki opisowe wskazują, że analizowane rozkłady empiryczne różnią się znacznie od rozkładu normalnego. Na podstawie empirycznych danych przestrzennych wykonano dwa testy na zgodność z rozkładem normalnym oraz rozkładem gamma. W przypadku obydwu procesów stwierdzono brak zgodności z rozkładem normalnym. Dla procesu powierzchni gospodarstw rolnych stwierdzono zgodność z rozkładem gamma. Natomiast rozkład procesu zużycia nawozów azotowych nie był zgodny z rozkładem gamma. Należy podkreślić, że w rozdziale czwartym stwierdzono występowanie dodatniej autokorelacji przestrzennej w przypadku obydwu procesów.

⁹ Jeżeli przykładowa jednostka terytorialna należy do pierwszej klasy, z wybranego rozkładu losowego generowana jest wartość zmiennej losowej, dla której wartość dystrybuanty jest mniejsza od 10%.

Tabela 5.14.

Analiza rozkładu zużycia nawozów azotowych

Statystyki opisowe			
Średnia	57,09	Wsp. zm. (%)	58,28
Odch. std.	33,27	Współczynnik skośności	0,37
Wyniki testów statystycznych			
Rodzaj testu	Statystyka testu		Wartość <i>p</i>
Test na zgodność rozkładu normalnego	79,37		-0,00
Test na zgodność rozkładu gamma	-13,34		-0,00

Źródło: opracowanie własne.

Tabela 5.15.

Analiza rozkładu powierzchni gospodarstw rolnych

Statystyki opisowe			
Średnia	8,69	Wsp. zm. (%)	66,23
Odch. std.	5,76	Współczynnik skośności	1,23
Wyniki testów statystycznych			
Rodzaj testu	Statystyka testu		Wartość <i>p</i>
Test na zgodność rozkładu normalnego	795,31		-0,00
Test na zgodność rozkładu gamma	0,29		0,77

Źródło: opracowanie własne.

Z kolei w celu rozpatrzenia problemu wyboru odpowiednich wartości wag, które wykorzystywane są zgodnie z równaniem (1.1) do agregacji danych przestrzennych, zbudowano cztery systemy wag oparte na liczbie ludności w gminach, kryterium równych wag, powierzchni gospodarstw rolnych oraz liczbie gospodarstw rolnych. Agregacji poddane zostały dane przestrzenne odzwierciedlające zużycie nawozów azotowych oraz powierzchnię gospodarstw rolnych na poziomie agregacji NUTS 5. W wyniku zastosowania każdego z czterech systemów wag otrzymano zbiór kolejnych danych przestrzennych, odniesionych do układu jednostek terytorialnych NUTS 4. W ostatnim kroku wyznaczono poziom skorelowania zagregowanych danych z danymi publikowanymi przez statystykę publiczną na poziomie agregacji NUTS 4. Obliczone wartości współczynnika korelacji liniowej dla każdego z systemów wag przedstawione zostały w tabeli 5.16.

Właściwy system wag dla agregacji procesu zużycia nawozów azotowych powinien być oparty na danych dotyczących powierzchni gospodarstw rolnych, a dla procesu powierzchni gospodarstw rolnych do budowy systemu wag po-

winy zostać wykorzystane dane odzwierciedlające liczbę gospodarstw rolnych w poszczególnych gminach. Widoczna jest zgodność wartości danych zagregowanych oraz publikowanych przez statystykę publiczną przy wykorzystaniu poprawnych systemów wag, gdzie wartości w_i wag dla i -tej gminy wyznaczone zostały jako udziały procentowe powierzchni gospodarstw rolnych lub liczby gospodarstw rolnych w odpowiednim j -tym powiecie.

W przypadku wykorzystania niewłaściwego systemu wag wartości danych przestrzennych nie pokrywają się, jednak poziom skorelowania danych jest wysoki. Najwyższe wartości współczynnika korelacji liniowej ($r=0,964$, $r=0,973$) uzyskane zostały przy wykorzystaniu systemów wag opartych na informacji o powierzchni oraz liczbie gospodarstw rolnych. Wartości współczynnika korelacji liniowej dla systemu opartego na kryterium równych wag wyniosły odpowiednio 0,914 oraz 0,918. Natomiast najniższe wartości współczynnika korelacji otrzymano dla systemu wag, gdzie wykorzystane zostały dane dotyczące liczby ludności w gminach. Wynika to z faktu, że liczba ludności obejmuje zarówno osoby żyjące na wsi, jak i w mieście. Oznacza to, że wykorzystanie systemu wag opartego na liczbie ludności w poszczególnych gminach najslabiej odzwierciedla charakter analizowanych procesów dotyczących rozwoju rolnictwa.

Tabela 5.16.

Wartości współczynnika korelacji liniowej dla wybranych wariantów systemu wag

Podstawa dla systemu wag	Skorelowanie danych publikowanych z danymi zagregowanymi	
	Zużycie nawozów azotowych	Powierzchnia gospodarstw rolnych
Liczba ludności	0,835	0,847
Powierzchnia gospodarstw rolnych	1,000	0,964
Liczba gospodarstw rolnych	0,973	1,000
Równe wagi	0,914	0,918

Źródło: opracowanie własne.

W związku z tym w celu sprawdzenia wpływu rodzaju rozkładu losowego oraz wyboru systemu wag na jakość odwzorowywanej struktury przestrzennej założone zostały dwa scenariusze procedury symulacyjnej. W ramach tych scenariuszy generowano najpierw dane przestrzenne na poziomie agregacji NUTS 5, a następnie agregowano je do poziomu agregacji NUTS 4. W scenariuszu pierwszym wyróżniono trzy warianty, gdzie oceniony został wybór rozkładu normalnego w połączeniu z różnymi systemami wag. W wariantcie pierwszym procedura symulacyjna oparta została na wykorzystaniu rozkładu normalnego oraz systemu wag związanych z danymi dotyczącymi ludności gmin. W wariantcie drugim przyjęto kryterium równych wag. Natomiast w wariantcie trzecim przyjęto poprawne systemy wag, gdzie w przypadku procesu zużycia nawozów

azotowych wykorzystano informację o powierzchni gospodarstw rolnych, a dla procesu powierzchni gospodarstw rolnych informację o liczbie gospodarstw rolnych.

Z kolei w scenariuszu drugim oceniono wybór rozkładu gamma, gdzie również rozpatrzone zostały trzy warianty systemów wag. W ramach kolejnych wariantów przyjęto takie same systemy wag, jak w przypadku scenariusza pierwszego.

Dla każdego z wariantów obydwu scenariuszy przeprowadzona procedura symulacyjna przebiegała w podobny sposób, jak w podrozdziałach 5.2 i 5.3, za wyjątkiem sposobu generowania danych przestrzennych, który oparty był na empirycznych rozkładach rozważanych procesów. W pierwszym kroku generowano po 10000 realizacji dwóch procesów przestrzennych X_1 , X_2 dla układu jednostek terytorialnych NUTS 5. W ramach przedstawionego rozwiązania przyporządkowano jednostki terytorialne do odpowiednich klas na podstawie empirycznych rozkładów, dla procesu X_1 na podstawie rozkładu zużycia nawozów azotowych, a dla procesu X_2 na podstawie rozkładu powierzchni gospodarstw rolnych. Następnie generowano wartości danych przestrzennych na podstawie odpowiedniego przedziału dystrybucyjnego wybranego rozkładu losowego. Należy podkreślić, że przyjęty sposób generowania danych przestrzennych pozwala na uwzględnianie rzeczywistej struktury przestrzennej. Dzięki temu odwzorowane zostały zależności między procesami oraz własność autokorelacji przestrzennej.

W drugim kroku procedury symulacyjnej wykonany został proces agregacji generowanych danych przestrzennych do układu jednostek terytorialnych NUTS 4, zgodnie z równaniem (1.1). W przeprowadzonej procedurze symulacyjnej, wybrany system wag determinował wartości wag w_i dla i -tego regionu z układu NUTS 5. W ten sposób, na podstawie wykonanego procesu agregacji, dla obydwu procesów otrzymywane zostały dwa zbiory danych przestrzennych, odniesionych do układu jednostek terytorialnych NUTS 4. Otrzymane zbiory danych przestrzennych pozwoliły na oszacowanie parametrów modeli autoregresji przestrzennej SAR osobno dla obydwu procesów zgodnie z równaniem (5.7). Stwierdzenie istotności statystycznej dla parametru autoregresji q oznaczało identyfikację własności autokorelacji przestrzennej. Dodatkowo oszacowano parametry modelu regresji dla zużycia nawozów azotowych X_1 względem powierzchni gospodarstw rolnych X_2 poszerzonego o autoregresję przestrzenną oraz wyznaczono miary średniego oddziaływania $A_{D'}$, $A_{I'}$, $A_{R'}$, A_T na podstawie następującego równania:

$$\mathbf{X}_1 = \alpha_0 + \alpha_1 \mathbf{X}_2 + q \mathbf{W} \mathbf{X}_1 + \boldsymbol{\varepsilon}, \quad (5.8)$$

gdzie \mathbf{X}_1 , \mathbf{X}_2 oznaczają wektory wartości procesów przestrzennych, α_0 , α_1 stanowią parametry strukturalne modelu, q zaś parametr autoregresji przestrzen-

nej, natomiast \mathbf{W} jest standaryzowaną macierzą sąsiedztwa pierwszego rzędu, a $\boldsymbol{\varepsilon}$ wektorem wartości białego szumu przestrzennego.

W ostatnim kroku na podstawie zbiorów otrzymanych ocen, obliczono dla kolejnych parametrów średnią, odchylenie standardowe oraz współczynnik zmienności. Wyniki procedury symulacyjnej dla wszystkich wariantów w ramach scenariusza pierwszego przedstawione zostały w tabelach 5.17–5.19. Natomiast w tabelach 5.20–5.22 zaprezentowano wyniki dla kolejnych wariantów otrzymane w ramach scenariusza drugiego.

Tabela 5.17.

Wyniki procedury symulacyjnej: scenariusz pierwszy, wariant pierwszy

Układ NUTS 5								
Statystyki opisowe	Autoregresja przestrzenna procesów				Miary oddziaływania przestrzennego			
	$q(X_1)$	p	$q(X_2)$	p	A_D	A_I	A_R	A_T
Średnia	0,57	~0,00	0,61	~0,00	1,71	0,21	0,00	1,92
Odch. std.	0,00	~0,00	0,01	~0,00	0,07	0,01	0,00	0,08
Wsp. zm. (%)	0,15	–	0,88	–	4,21	4,58	5,02	4,23
Analiza regresji								
Statystyki opisowe	α_1	p	q	p	R^2	$V(r)$	$I(r)$	p
Średnia	1,62	~0,00	0,49	~0,00	0,30	0,04	–0,03	0,99
Odch. std.	0,07	~0,00	0,01	~0,00	0,00	0,00	0,00	0,01
Wsp. zm. (%)	4,06	–	1,14	–	1,40	0,30	–	–
Układ NUTS 4								
Statystyki opisowe	Autoregresja przestrzenna procesów				Miary oddziaływania przestrzennego			
	$q(X_1)$	p	$q(X_2)$	p	A_D	A_I	A_R	A_T
Średnia	0,53	~0,00	0,68	~0,00	1,98	0,22	–0,01	2,19
Odch. std.	0,02	~0,00	0,02	~0,00	0,24	0,03	0,00	0,27
Wsp. zm. (%)	4,36	–	3,50	–	12,05	12,91	–51,02	12,16
Analiza korelacji i regresji								
Statystyki opisowe	α_1	p	q	p	R^2	$V(r)$	$I(r)$	p
Średnia	1,89	~0,00	0,44	~0,00	0,28	0,03	0,02	0,30
Odch. std.	0,23	~0,00	0,03	~0,00	0,02	0,00	0,01	0,14
Wsp. zm. (%)	11,94	–	7,52	–	7,67	3,91	–	–

Źródło: opracowanie własne.

Analizując wyniki otrzymane dla scenariusza pierwszego należy podkreślić, że w pierwszych dwóch wariantach zastosowane zostały niepoprawne systemy

wag. Dopiero w wariancie trzecim przyjęto odpowiednie systemy wag dla obydwu procesów. We wszystkich wariantach proces generowania danych oparty został na rozkładzie normalnym, gdzie konkretna postać rozkładu normalnego szacowana była w programie gretl na podstawie rozkładów empirycznych zużycia nawozów azotowych oraz powierzchni gospodarstw rolnych.

Tabela 5.18.

Wyniki procedury symulacyjnej: scenariusz pierwszy, wariant drugi

Układ NUTS 5								
Statystyki opisowe	Autoregresja przestrzenna procesów				Miary oddziaływania przestrzennego			
	$q(X_1)$	p	$q(X_2)$	p	A_D	A_I	A_R	A_T
Średnia	0,57	~0,00	0,61	~0,00	1,65	0,21	~0,00	1,86
Odch. std.	0,01	~0,00	0,01	~0,00	0,06	0,01	~0,00	0,07
Wsp. zm. (%)	1,49	–	1,06	–	3,59	3,94	3,97	3,63
Analiza korelacji i regresji								
Statystyki opisowe	α_1	p	q	p	R^2	$V(r)$	$I(r)$	p
Średnia	1,58	~0,00	0,49	~0,00	0,31	0,04	~0,03	0,99
Odch. std.	0,05	~0,00	0,01	~0,00	0,01	0,00	~0,00	~0,00
Wsp. zm. (%)	3,26	–	2,41	–	2,06	0,45	–	–
Układ NUTS 4								
Statystyki opisowe	Autoregresja przestrzenna procesów				Miary oddziaływania przestrzennego			
	$q(X_1)$	p	$q(X_2)$	p	A_D	A_I	A_R	A_T
Średnia	0,62	~0,00	0,73	~0,00	1,94	0,25	~0,00	2,19
Odch. std.	0,04	~0,00	0,02	~0,00	0,10	0,02	~0,00	0,12
Wsp. zm. (%)	5,71	–	2,28	–	5,25	6,81	5,67	5,50
Analiza korelacji i regresji								
Statystyki opisowe	α_1	p	q	p	R^2	$V(r)$	$I(r)$	p
Średnia	1,82	~0,00	0,51	~0,00	0,36	0,03	0,03	0,19
Odch. std.	0,09	~0,00	0,04	~0,00	0,03	0,00	0,01	0,07
Wsp. zm. (%)	4,97	–	7,83	–	8,63	4,48	–	–

Źródło: opracowanie własne.

W wyniku estymacji parametrów modelu SAR na poziomie agregacji NUTS 4 dla procesu X_1 otrzymano ocenę parametru autoregresji q równą 0,57, a w przypadku procesu X_2 wartość oceny wyniosła 0,61. Otrzymane wyniki potwierdzają występowanie silnej, dodatniej autokorelacji przestrzennej dla obydwu procesów. We wszystkich wariantach doszło do zmiany w otrzymywanych ocenach parametru autoregresji na skutek procesu agregacji. Dla

każdego z wariantów odnotowano jednak inny charakter zmian. W wariancie pierwszym średnia ocen parametru autoregresji uległa obniżeniu w przypadku procesu X_1 oraz podwyższeniu w przypadku procesu X_2 . W wariancie drugim średnie z ocen parametru autoregresji uległy podwyższeniu dla obydwu procesów, przy czym wzrost średniej z ocen dla procesu X_2 był wyższy w porównaniu do wzrostu odnotowanego w wariancie pierwszym.

Tabela 5.19.

Wyniki procedury symulacyjnej: scenariusz pierwszy, wariant trzeci

Układ NUTS 5								
Statystyki opisowe	Autoregresja przestrzenna procesów				Miary oddziaływania przestrzennego			
	$q(X_1)$	p	$q(X_2)$	p	A_D	A_I	A_R	A_T
Średnia	0,57	~0,00	0,61	~0,00	1,70	0,21	~0,00	1,91
Odch. std.	0,01	~0,00	0,00	~0,00	0,02	0,00	~0,00	0,02
Wsp. zm. (%)	1,03	–	0,72	–	1,23	1,58	2,71	1,23
Analiza korelacji i regresji								
Statystyki opisowe	α_1	p	q	p	R^2	$V(r)$	$I(r)$	p
Średnia	1,61	~0,00	0,49	~0,00	0,30	0,04	~0,03	0,99
Odch. std.	0,02	~0,00	0,01	~0,00	0,01	0,00	0,00	0,00
Wsp. zm. (%)	1,10	–	1,15	–	2,27	0,50	–	–
Układ NUTS 4								
Statystyki opisowe	Autoregresja przestrzenna procesów				Miary oddziaływania przestrzennego			
	$q(X_1)$	p	$q(X_2)$	p	A_D	A_I	A_R	A_T
Średnia	0,02	0,88	0,44	~0,00	5,87	0,65	~0,00	6,53
Odch. std.	0,01	0,09	0,00	~0,00	0,14	0,01	~0,00	0,15
Wsp. zm. (%)	71,01	–	0,71	–	2,33	2,22	9,98	2,30
Analiza korelacji i regresji								
Statystyki opisowe	α_1	p	q	p	R^2	$V(r)$	$I(r)$	p
Średnia	5,64	~0,00	~0,11	~0,00	0,38	0,09	0,07	0,01
Odch. std.	0,13	~0,00	0,01	~0,00	0,01	0,00	0,01	0,01
Wsp. zm. (%)	2,37	–	13,54	–	2,72	1,41	–	–

Źródło: opracowanie własne.

Natomiast w wariancie trzecim opartym na poprawnym systemie wag, po agregacji wygenerowanych danych parametr autoregresji w przypadku procesu X_1 okazał się statystycznie nieistotny, a w przypadku procesu X_2 średnia ocen parametru autoregresji uległa znacznemu obniżeniu. W przypadku analizy regresji dla wariantu pierwszego i drugiego, pomimo różnic w otrzymanych

średnich ocen dla parametru regresji α_1 oraz parametru autoregresji q , wartości miar średniego oddziaływania przestrzennego $A_{D'}$, $A_{R'}$, $A_{R''}$, $A_{T'}$ kształtowały się na zbliżonym poziomie. Natomiast w przypadku oszacowań modelu regresji dla wariantu trzeciego średnia z ocen parametru regresji α_1 była znacznie przeszacowana, a wartość średnia ocen parametru autoregresji q była bliska zeru.

Procedura symulacyjna w scenariuszu pierwszym została świadomie przeprowadzona przy niewłaściwych założeniach w celu zobrazowania trudności w generowaniu realizacji procesów odzwierciedlających rzeczywistą strukturę przestrzenną. Dla wszystkich wariantów scenariusza pierwszego w sposób błędny przyjęty został rozkład normalny. W przypadku wariantu pierwszego oraz drugiego, gdzie dokonano również błędnego założenia odnośnie do systemu wag, otrzymane wyniki wydają się być poprawne. Natomiast w wariantach trzecim, gdzie przyjęto poprawny system wag, otrzymane wyniki na poziomie agregacji NUTS 4 wskazują na istotną zmianę w strukturze przestrzennej w wyniku procesu agregacji, co nie powinno mieć miejsca. Oznacza to, że ustalone wnioski na podstawie wyników wariantu trzeciego wskazywałyby na identyfikację problemu skali, podczas gdy otrzymane zmiany w ocenach parametrów wynikały z niewłaściwego odtworzenia struktury przestrzennej przez generowane realizacje analizowanych procesów.

Następnie autor założył scenariusz drugi, gdzie do generowania realizacji procesów wykorzystany został rozkład gamma, który posiada wyższy stopień zgodności z rozkładami empirycznymi procesów zużycia nawozów azotowych oraz powierzchni gospodarstw rolnych w porównaniu z rozkładem normalnym. Do symulacji realizacji procesów X_1 , X_2 , wykorzystano konkretne postaci rozkładu gamma, które oszacowane zostały ponownie w programie gretl. Zgodnie z założeniami scenariusza drugiego proces agregacji danych przestrzennych oparty został na przyjętych wcześniej systemach wag. Ponownie dokonano estymacji parametrów modelu SAR na poziomie agregacji NUTS 4 zgodnie z równaniem (5.7). W rezultacie otrzymano dla procesu X_1 średnią ocen parametru autoregresji $q(X_1)$ na poziomie 0,54, a w przypadku procesu X_2 wartość średniej ocen wyniosła 0,53. Otrzymane wartości średnie ocen parametru autoregresji były niższe w porównaniu z wynikami ze scenariusza pierwszego, gdzie dla procesu X_1 wyznaczona została średnia ocen parametru autoregresji $q(X_1)$ na poziomie 0,57, a w przypadku procesu X_2 otrzymano wartość średniej ocen na poziomie 0,61. Można jednak stwierdzić, że wyniki są porównywalne w obydwu scenariuszach i jednoznacznie wskazują na występowanie autokorelacji przestrzennej. Podobnie, jak w scenariuszu pierwszym, dla wszystkich wariantów ustalono zmiany w otrzymywanych ocenach parametru autoregresji na wyższym poziomie agregacji.

Tabela 5.20.

Wyniki procedury symulacyjnej: scenariusz drugi, wariant pierwszy

Układ NUTS 5								
Statystyki opisowe	Autoregresja przestrzenna procesów				Miary oddziaływania przestrzennego			
	$q(X_1)$	p	$q(X_2)$	p	A_D	A_I	A_R	A_T
Średnia	0,54	~0,00	0,53	~0,00	1,56	0,20	~0,00	1,76
Odch. std.	0,01	~0,00	0,01	~0,00	0,11	0,01	~0,00	0,12
Wsp. zm. (%)	1,09	–	1,26	–	6,83	5,19	–	6,60
Analiza korelacji i regresji								
Statystyki opisowe	α_1	p	q	p	R^2	$V(r)$	$I(r)$	p
Średnia	1,48	~0,00	0,46	~0,00	0,27	0,50	~0,03	0,99
Odch. std.	0,09	~0,00	0,01	~0,00	0,01	0,01	0,00	0,01
Wsp. zm. (%)	6,36	–	1,81	–	2,76	1,34	–	–
Układ NUTS 4								
Statystyki opisowe	Autoregresja przestrzenna procesów				Miary oddziaływania przestrzennego			
	$q(X_1)$	p	$q(X_2)$	p	A_D	A_I	A_R	A_T
Średnia	0,53	~0,00	0,63	~0,00	1,64	0,18	0,00	1,83
Odch. std.	0,03	~0,00	0,01	~0,00	0,19	0,02	0,01	0,21
Wsp. zm. (%)	5,59	–	1,90	–	11,75	11,02	–	11,36
Analiza korelacji i regresji								
Statystyki opisowe	α_1	p	q	p	R^2	$V(r)$	$I(r)$	p
Średnia	1,58	~0,00	0,46	~0,00	0,28	0,35	0,01	0,33
Odch. std.	0,19	~0,00	0,03	~0,00	0,02	0,01	0,01	0,08
Wsp. zm. (%)	11,78	–	6,95	–	8,57	3,37	–	–

Źródło: opracowanie własne.

Analizując przedstawione w tabelach 5.20–5.22 wyniki dla scenariusza drugiego należy podkreślić, że stwierdzono różnice w otrzymywanych ocenach parametrów pomiędzy kolejnymi wariantami procedury symulacyjnej. W przypadku wariantu pierwszego, gdzie wykorzystano system wag oparty na liczbie ludności, otrzymane rezultaty różniły się znacznie od uzyskanych w wariacie drugim i trzecim. Biorąc pod uwagę wyniki estymacji parametrów modelu SAR, wykonanej osobno dla każdego z procesów na podstawie równania (5.7) należy stwierdzić, że obliczona w wariacie pierwszym średnia z ocen parametru autoregresji $q(X_1)$ dla procesu X_1 nie uległa zmianie, a dla procesu X_2 wyznaczona wartość średniej dla parametru autoregresji $q(X_2)$ uległa podwyższeniu. Interpretując następnie wyniki estymacji parametrów modelu regresji określonego za pomocą równania (5.8) można wskazać, że obliczone dla wa-

riantu pierwszego średnie z ocen parametru regresji α_1 oraz średnie dla miar oddziaływania przestrzennego A_D, A_I, A_R, A_T uległy podwyższeniu, a wartość średnia ocen parametru autoregresji q pozostała na niezmiennym poziomie (zob. tabela 5.20).

Tabela 5.21.

Wyniki procedury symulacyjnej: scenariusz drugi, wariant drugi

Układ NUTS 5								
Statystyki opisowe	Autoregresja przestrzenna procesów				Miary oddziaływania przestrzennego			
	$q(X_1)$	p	$q(X_2)$	p	A_D	A_I	A_R	A_T
Średnia	0,54	~0,00	0,53	~0,00	1,56	0,19	~0,00	1,75
Odch. std.	0,01	~0,00	0,01	~0,00	0,05	0,01	~0,00	0,05
Wsp. zm. (%)	1,58	–	1,67	–	3,09	2,90	–	3,06
Analiza korelacji i regresji								
Statystyki opisowe	α_1	p	q	p	R^2	$V(r)$	$I(r)$	p
Średnia	1,43	~0,00	0,48	~0,00	0,27	0,50	–0,03	0,99
Odch. std.	0,04	~0,00	0,01	~0,00	0,01	0,00	0,00	0,00
Wsp. zm. (%)	2,81	–	1,64	–	3,07	0,97	–	–
Układ NUTS 4								
Statystyki opisowe	Autoregresja przestrzenna procesów				Miary oddziaływania przestrzennego			
	$q(X_1)$	p	$q(X_2)$	p	A_D	A_I	A_R	A_T
Średnia	0,58	~0,00	0,70	~0,00	1,79	0,22	~0,00	2,01
Odch. std.	0,02	~0,00	0,02	~0,00	0,09	0,01	~0,00	0,10
Wsp. zm. (%)	3,81	–	2,81	–	5,23	5,60	8,72	4,97
Analiza korelacji i regresji								
Statystyki opisowe	α_1	p	q	p	R^2	$V(r)$	$I(r)$	p
Średnia	1,70	~0,00	0,49	~0,00	0,33	0,34	0,02	0,27
Odch. std.	0,09	~0,00	0,03	~0,00	0,02	0,01	0,00	0,03
Wsp. zm. (%)	5,28	–	6,14	–	6,47	2,54	–	–

Źródło: opracowanie własne.

Natomiast dla wariantów drugiego i trzeciego stwierdzono podobny charakter zmian w otrzymywanych ocenach parametrów (zob. tabele 5.21, 5.22). Należy podkreślić, że w przeciwieństwie do scenariusza pierwszego nie doszło do istotnej zmiany struktury przestrzennej dla wariantu trzeciego (por. tabele 5.19 oraz 5.22). W tym przypadku, wykorzystanie w wariantcie trzecim rozkładu gamma z właściwymi systemami wag umożliwiło generowanie realizacji

procesów, które odzwierciedlały rzeczywistą strukturę przestrzenną analizowanego obszaru. W związku z tym ustalone wnioski na podstawie wyników wariantu trzeciego powinny pozwolić na poprawne rozważenie występowania problemu skali.

Tabela 5.22.

Wyniki procedury symulacyjnej: scenariusz drugi, wariant trzeci

Układ NUTS 5								
Statystyki opisowe	Autoregresja przestrzenna procesów				Miary oddziaływania przestrzennego			
	$q(X_1)$	p	$q(X_2)$	p	A_D	A_I	A_R	A_T
Średnia	0,54	~0,00	0,53	~0,00	1,55	0,19	~0,00	1,74
Odch. std.	0,01	~0,00	0,01	~0,00	0,11	0,01	~0,00	0,12
Wsp. zm. (%)	2,66	–	1,21	–	6,89	7,76	–	6,98
Analiza korelacji i regresji								
Statystyki opisowe	α_1	p	q	p	R^2	$V(r)$	$I(r)$	p
Średnia	1,48	~0,00	0,46	~0,00	0,27	0,50	–0,03	0,99
Odch. std.	0,09	~0,00	0,02	~0,00	0,01	0,01	0,00	0,00
Wsp. zm. (%)	6,51	–	3,91	–	3,87	1,44	–	–
Układ NUTS 4								
Statystyki opisowe	Autoregresja przestrzenna procesów				Miary oddziaływania przestrzennego			
	$q(X_1)$	p	$q(X_2)$	p	A_D	A_I	A_R	A_T
Średnia	0,63	~0,00	0,69	~0,00	1,84	0,25	0,01	2,10
Odch. std.	0,02	~0,00	0,00	~0,00	0,16	0,02	0,01	0,17
Wsp. zm. (%)	3,16	–	0,58	–	8,59	8,56	–	8,12
Analiza korelacji i regresji								
Statystyki opisowe	α_1	p	q	p	R^2	$V(r)$	$I(r)$	p
Średnia	1,74	~0,00	0,54	~0,00	0,39	0,32	0,05	0,08
Odch. std.	0,15	~0,00	0,02	~0,00	0,02	0,01	0,01	0,04
Wsp. zm. (%)	8,50	–	3,80	–	5,07	1,94	–	–

Źródło: opracowanie własne.

W wariantcie trzecim, w przypadku osobnej estymacji parametrów modelu SAR dla każdego z procesów, średnie z ocen parametrów autoregresji $q(X_i)$ uległy znacznemu podwyższeniu. W wyniku procesu agregacji średnia z ocen parametru autoregresji $q(X_1)$ dla procesu X_1 wzrosła z poziomu 0,54 do poziomu 0,63, a średnia z ocen parametru autoregresji $q(X_2)$ dla procesu X_2 wzrosła z poziomu 0,53 do 0,69. Również w przypadku estymacji parametrów modelu

regresji, w wyniku agregacji danych wzrosły wartości średnich ocen dla parametru regresji α_1 , parametru autoregresji q oraz wartości średnich dla miar oddziaływania przestrzennego A_D , A_I , $A_{R'}$, A_T . Średnia ocen dla parametru regresji α_1 wzrosła z poziomu 1,48 do poziomu 1,74, a dla parametru autoregresji q z poziomu 0,46 do poziomu 0,54. Ustalono w wyniku procesu agregacji wzrosty średnich ocen parametru regresji oraz parametru autoregresji przełożyły się na wzrost średnich dla miar oddziaływania przestrzennego. Obliczona średnia dla miary oddziaływania całkowitego A_T uległa podwyższeniu z wartości równej 1,74 do wartości 2,1 (zob. tabela 5.22).

Otrzymane wyniki analizy symulacyjnej wskazują jednoznacznie na występowanie zmian w wyznaczanych ocenach parametrów modeli w wyniku procesu agregacji danych. Problem skali zidentyfikowany został w sytuacji generowania realizacji procesów charakteryzujących się własnością autokorelacji przestrzennej. Należy podkreślić, że w wykonanych w poprzednich podrozdziałach analizach symulacyjnych dla procesów szumów przestrzennych oraz procesów charakteryzujących się przestrzennym trendem pierwszego stopnia nie stwierdzono istotnych różnic w ocenach własności procesów przestrzennych oraz łączących je zależnościach przyczynowych, w wyniku procesu agregacji.

Tabela 5.23.

Zmiany procentowe dla średnich ocen parametrów w wariancie trzecim scenariusza drugiego

Poziom agregacji	Średnie ocen parametrów						
	$q(X_1)$	$q(X_2)$	α_1	q	A_D	A_I	A_T
Układ NUTS 5	0,54	0,53	1,48	0,46	1,55	0,19	1,74
Układ NUTS 4	0,63	0,69	1,74	0,54	1,84	0,25	2,1
Zmiana procentowa (%)	16,67	30,19	17,57	17,39	18,71	31,58	20,69

Źródło: opracowanie własne.

W tabeli 5.23 przedstawione zostały zmiany procentowe średnich ocen parametrów, wyznaczone na podstawie wyników otrzymanych na poziomie agregacji NUTS 5 oraz NUTS 4. Zmiany procentowe przekraczające 30% uzyskane zostały w przypadku parametru autoregresji $q(X_2)$ oraz miary oddziaływania pośredniego A_I . Dla pozostałych parametrów: parametrów autoregresji $q(X_1)$ i q , parametru regresji α_1 oraz dla miar oddziaływania bezpośredniego A_D i całkowitego A_T , zmiany średnich ocen dochodziły do 20%. Oznacza to, że w przypadku identyfikacji własności autokorelacji przestrzennej zachodzą znaczne zmiany w otrzymywanych rezultatach badań na skutek procesu agregacji danych. Przedstawione wnioski mają znaczenie dla oceny wiarygodności otrzymywanych wyników przestrzennych badań ekonomicznych, ponieważ podstawą analiz mogą być dane przestrzenne poddane procesowi agregacji, które publikowane są przez statystykę publiczną dla wyższych poziomów

agregacji, niż w rzeczywistości zostały zebrane. W takiej sytuacji należy mieć świadomość potencjalnych kierunków zmian oraz ich wielkości w otrzymywanych rezultatach badań.

5.4.3. Wpływ siły autokorelacji przestrzennej na poziom zmian wynikających z problemu skali

W podrozdziale 5.4.2 stwierdzone zostało występowanie problemu skali na podstawie przeprowadzonej procedury symulacyjnej, gdzie analizowane procesy charakteryzowały się własnością autokorelacji przestrzennej. W niniejszym podrozdziale sprawdzono, czy zmiany w sile autokorelacji przestrzennej mogą przełożyć się na wielkość zmian wynikających z problemu skali. W tym celu zaprojektowana została nowa procedura symulacyjna dla procesów X_1^* , X_2^* , w której wykorzystano zbiór 10000 wygenerowanych wcześniej realizacji procesów X_1 , X_2 dla wariantu trzeciego scenariusza drugiego (podrozdział 5.4.2)¹⁰. W ramach procedury, przyjęte dane przestrzenne zostały poddane filtracji w celu wymuszenia konkretnych zmian własności autokorelacji przestrzennej procesów X_1 , X_2 . Kolejne realizacje procesów X_1^* , X_2^* , charakteryzujące się zmienioną siłą autokorelacji przestrzennej, otrzymane zostały na podstawie równania:

$$\mathbf{X}_i^* = (\mathbf{I} - q_{1,i}\mathbf{W})^{-1} \mathbf{X}_i, \quad (5.9)$$

gdzie \mathbf{X}_i^* jest wektorem wartości wynikowego procesu przestrzennego, $q_{1,i}$ stanowi parametr autoregresji przestrzennej, \mathbf{I} oznacza macierz jednostkową, \mathbf{W} jest standaryzowaną macierzą sąsiedztwa pierwszego rzędu, natomiast \mathbf{X}_i to wektor wartości wejściowego procesu przestrzennego.

Dla osiągnięcia celu sprawdzenia wpływu poziomu autokorelacji przestrzennej na wielkość zmian w otrzymywanych ocenach parametrów, przyjęto cztery warianty zmian autokorelacji przestrzennej analizowanych procesów. W wariantie pierwszym poziom autokorelacji przestrzennej obydwu procesów został podwyższony. Realizacje procesów X_1^* , X_2^* otrzymane zostały na podstawie równania (5.9), gdzie przyjęto dodatnie wartości parametrów autoregresji $q_{1,1}=0,2$ oraz $q_{1,2}=0,2$. Przeprowadzona w ten sposób filtracja realizacji procesów X_1 , X_2 spowodowała wzrost siły autokorelacji przestrzennej dla wynikowych procesów X_1^* , X_2^* . W wariantie drugim poziom autokorelacji przestrzennej obydwu procesów został obniżony. Do filtracji realizacji procesów X_1 , X_2 wykorzystano wartości parametrów autoregresji $q_{1,1}=-0,2$ oraz $q_{1,2}=-0,2$, co wpłynęło na spadek siły autokorelacji przestrzennej wynikowych procesów

¹⁰ Dane przestrzenne wygenerowane zostały podrozdział 5.4.2 przy wykorzystaniu rozkładu gamma oraz właściwych systemów wag opartych na danych dotyczących powierzchni gospodarstw rolnych oraz liczby gospodarstw rolnych w poszczególnych gminach.

X_1^* , X_2^* . Z kolei w wariancie trzecim przyjęto wartości parametrów autoregresji $q_{1,1} = -0,2$ i $q_{1,2} = 0,2$, a w wariancie czwartym $q_{1,1} = 0,2$ oraz $q_{1,2} = -0,2$, co przyczyniło się do spadku lub wzrostu siły autokorelacji przestrzennej procesów X_1^* , X_2^* w zależności od znaku przyjętej wartości parametru autoregresji.

Dla każdego z czterech wariantów przeprowadzono kolejne kroki procedury¹¹, polegające na agregacji danych przestrzennych oraz estymacji parametrów modeli określonych za pomocą równań (5.7) i (5.8). Otrzymane wyniki dla kolejnych wariantów przedstawione zostały w tabelach 5.24–5.27.

Tabela 5.24.

Wariant pierwszy: podwyższony poziom autokorelacji przestrzennej procesów X_1^* , X_2^*

Układ NUTS 5								
Statystyki opisowe	Autoregresja przestrzenna procesów				Miary oddziaływania przestrzennego			
	$q(X_1^*)$	p	$q(X_2^*)$	p	A_D	A_I	A_R	A_T
Średnia	0,67	~0,00	0,68	~0,00	1,39	0,23	~0,00	1,62
Odch. std.	0,01	~0,00	0,01	~0,00	0,06	0,01	~0,00	0,08
Wsp. zm. (%)	1,44	–	1,40	–	4,68	5,31	4,92	4,72
Analiza regresji								
Statystyki opisowe	α_1	p	q	p	R^2	$V(r)$	$I(r)$	p
Średnia	1,27	~0,00	0,60	~0,00	0,41	0,49	~0,03	0,99
Odch. std.	0,06	~0,00	0,01	~0,00	0,01	0,01	0,00	0,01
Wsp. zm. (%)	4,62	–	2,13	–	2,22	1,46	–	–
Układ NUTS 4								
Statystyki opisowe	Autoregresja przestrzenna procesów				Miary oddziaływania przestrzennego			
	$q(X_1^*)$	p	$q(X_2^*)$	p	A_D	A_I	A_R	A_T
Średnia	0,66	~0,00	0,76	~0,00	1,68	0,26	~0,00	1,94
Odch. std.	0,02	~0,00	0,01	~0,00	0,06	0,01	~0,00	0,07
Wsp. zm. (%)	2,29	–	1,07	–	3,76	4,12	7,28	3,73
Analiza regresji								
Statystyki opisowe	α_1	p	q	p	R^2	$V(r)$	$I(r)$	p
Średnia	1,55	~0,00	0,57	~0,00	0,40	0,37	0,03	0,17
Odch. std.	0,05	~0,00	0,01	~0,00	0,02	0,01	0,00	0,03
Wsp. zm. (%)	3,29	–	2,51	–	4,59	1,79	–	–

Źródło: opracowanie własne.

¹¹ Kolejne kroki procedury symulacyjnej były zgodne z procedurą przeprowadzoną w podrozdziale 5.4.2.

Tabela 5.25.

Wariant drugi: obniżony poziom autokorelacji przestrzennej procesów X_1^*, X_2^*

Układ NUTS 5								
Statystyki opisowe	Autoregresja przestrzenna procesów				Miary oddziaływania przestrzennego			
	$q(X_1^*)$	p	$q(X_2^*)$	p	A_D	A_I	A_R	A_T
Średnia	0,39	~0,00	0,38	~0,00	1,59	0,13	~0,00	1,72
Odch. std.	0,01	~0,00	0,01	~0,00	0,09	0,01	~0,00	0,10
Wsp. zm. (%)	3,17	–	3,08	–	5,60	5,63	7,01	5,52
Analiza regresji								
Statystyki opisowe	α_1	p	q	p	R^2	$V(r)$	$I(r)$	p
Średnia	1,56	~0,00	0,32	~0,00	0,17	0,51	~0,03	0,99
Odch. std.	0,08	~0,00	0,01	~0,00	0,01	0,01	0,00	0,00
Wsp. zm. (%)	5,16	–	4,35	–	5,68	1,69	–	–
Układ NUTS 4								
Statystyki opisowe	Autoregresja przestrzenna procesów				Miary oddziaływania przestrzennego			
	$q(X_1^*)$	p	$q(X_2^*)$	p	A_D	A_I	A_R	A_T
Średnia	0,51	~0,00	0,65	~0,00	1,92	0,20	~0,00	2,13
Odch. std.	0,02	~0,00	0,02	~0,00	0,09	0,01	~0,00	0,09
Wsp. zm. (%)	4,36	–	3,12	–	4,48	4,92	7,44	4,38
Analiza regresji								
Statystyki opisowe	α_1	p	q	p	R^2	$V(r)$	$I(r)$	p
Średnia	1,85	~0,00	0,42	~0,00	0,28	0,32	0,02	0,30
Odch. std.	0,08	~0,00	0,03	~0,00	0,01	0,00	0,01	0,05
Wsp. zm. (%)	4,39	–	7,94	–	5,30	0,57	–	–

Źródło: opracowanie własne.

Tabela 5.26.

Wariant trzeci: obniżony poziom autokorelacji przestrzennej procesu X_1^* oraz podwyższony poziom autokorelacji przestrzennej procesu X_2^*

Układ NUTS 5								
Statystyki opisowe	Autoregresja przestrzenna procesów				Miary oddziaływania przestrzennego			
	$q(X_1^*)$	p	$q(X_2^*)$	p	A_D	A_I	A_R	A_T
Średnia	0,39	~0,00	0,68	~0,00	1,45	0,12	~0,00	1,57
Odch. std.	0,01	~0,00	0,01	~0,00	0,06	0,00	~0,00	0,06
Wsp. zm. (%)	3,17	–	1,40	–	3,87	4,28	5,16	3,86
Analiza regresji								
Statystyki opisowe	α_1	p	q	p	R^2	$V(r)$	$I(r)$	p
Średnia	1,42	~0,00	0,28	~0,00	0,16	0,51	~0,02	0,91
Odch. std.	0,05	~0,00	0,02	~0,00	0,01	0,01	0,00	0,02
Wsp. zm. (%)	3,73	–	6,04	–	4,35	2,05	–	–
Układ NUTS 4								
Statystyki opisowe	Autoregresja przestrzenna procesów				Miary oddziaływania przestrzennego			
	$q(X_1^*)$	p	$q(X_2^*)$	p	A_D	A_I	A_R	A_T
Średnia	0,49	~0,00	0,76	~0,00	1,38	0,14	~0,00	1,51
Odch. std.	0,03	~0,00	0,01	~0,00	0,08	0,01	~0,00	0,09
Wsp. zm. (%)	6,05	–	1,40	–	6,02	6,06	6,25	5,86
Analiza regresji								
Statystyki opisowe	α_1	p	q	p	R^2	$V(r)$	$I(r)$	p
Średnia	1,34	~0,00	0,39	~0,00	0,25	0,33	0,02	0,27
Odch. std.	0,08	~0,00	0,04	~0,00	0,02	0,01	0,00	0,04
Wsp. zm. (%)	5,86	–	9,60	–	8,04	2,68	–	–

Źródło: opracowanie własne.

W wariantcie pierwszym w wyniku zwiększenia siły autokorelacji przestrzennej otrzymano średnie ocen parametru autoregresji procesów X_1^* , X_2^* wynoszące odpowiednio 0,67 i 0,68 dla poziomu agregacji NUTS 5, w porównaniu do wyjściowych wartości średnich równych 0,54 oraz 0,53 (zob. tabela 5.22, scenariusz drugi, wariant trzeci). W wyniku agregacji danych do poziomu NUTS 4 średnia ocen parametru autoregresji $q(X_1^*)$ dla procesu X_1^* nie uległa zmianie, a średnia ocen parametru autoregresji $q(X_2^*)$ dla procesu X_2^* wzrosła do poziomu 0,76 (zob. tabela 5.24). Z kolei w wariantcie drugim zmniejszenie siły autokorelacji przestrzennej spowodowało otrzymanie dla poziomu agregacji NUTS 5 niższych wartości średnich ocen parametru autoregresji procesów

X_1^* , X_2^* , odpowiednio: 0,39 i 0,38 (zob. tabela 5.25). W tym wypadku proces agregacji wpłynął na zwiększenie średnich ocen obydwu parametrów autoregresji, odpowiednio do poziomu 0,51 oraz 0,68. W zależności od przyjętej wartości parametru autoregresji $q_{1,i}$ w równaniu (5.9), analogiczne zmiany dla średnich ocen parametrów autoregresji $q(X_1^*)$ oraz $q(X_2^*)$ zaszyły dla wariantu trzeciego i wariantu czwartego (zob. tabele 5.26, 5.27).

Tabela 5.27.

Wariant czwarty: podwyższony poziom autokorelacji przestrzennej procesu X_1^* oraz obniżony poziom autokorelacji przestrzennej procesu X_2^*

Układ NUTS 5								
Statystyki opisowe	Autoregresja przestrzenna procesów				Miary oddziaływania przestrzennego			
	$q(X_1^*)$	p	$q(X_2^*)$	p	A_D	A_I	A_R	A_T
Średnia	0,67	~0,00	0,38	~0,00	1,58	0,29	~0,00	1,87
Odch. std.	0,01	~0,00	0,01	~0,00	0,07	0,01	~0,00	0,09
Wsp. zm. (%)	1,44	-	3,08	-	4,67	4,32	2,60	4,59
Analiza regresji								
Statystyki opisowe	α_1	p	q	p	R^2	$V(r)$	$I(r)$	p
Średnia	1,43	~0,00	0,63	~0,00	0,41	0,49	~0,04	0,99
Odch. std.	0,06	~0,00	~0,00	~0,00	0,01	0,01	~0,00,	~0,00
Wsp. zm. (%)	4,03	-	0,53	-	2,21	1,75	-	-
Układ NUTS 4								
Statystyki opisowe	Autoregresja przestrzenna procesów				Miary oddziaływania przestrzennego			
	$q(X_1^*)$	p	$q(X_2^*)$	p	A_D	A_I	A_R	A_T
Średnia	0,65	~0,00	0,62	~0,00	2,00	0,33	~0,00	2,33
Odch. std.	0,01	~0,00	0,02	~0,00	0,20	0,04	~0,00	0,24
Wsp. zm. (%)	1,75	-	3,22	-	9,90	10,88	9,26	10,16
Analiza regresji								
Statystyki opisowe	α_1	p	q	p	R^2	$V(r)$	$I(r)$	p
Średnia	1,83	~0,00	0,59	~0,00	0,39	0,37	0,01	0,35
Odch. std.	0,18	~0,00	0,01	~0,00	0,02	0,01	0,01	0,06
Wsp. zm. (%)	9,73	-	1,52	-	4,69	2,42	-	-

Źródło: opracowanie własne.

Na podstawie otrzymanych wyników można wyciągnąć wniosek, że w zależności od wyjściowej wartości parametru autoregresji $q(X_i^*)$ na poziomie NUTS 5, zmiany ocen tego parametru na skutek procesu agregacji przebiegają

w różny sposób, zarówno dla procesu X_1^* , jak i dla procesu X_2^* . W przypadku procesu X_1^* przy wartości parametru równej 0,39 następuje znaczny wzrost wyznaczanych ocen parametru autoregresji $q(X_1^*)$ na skutek procesu agregacji danych, natomiast przy poziomie 0,67 zmiana wartości ocen nie występuje. Z kolei dla procesu X_2^* występuje znacznie większy wzrost wartości otrzymywanych ocen parametru autoregresji $q(X_2^*)$, gdy wyjściowa wartość parametru jest niższa. W przypadku estymacji parametrów modelu regresji określonego za pomocą równania (5.8), zmiany w wyznaczanych ocenach parametrów przebiegają w analogiczny sposób. W wyniku procesu agregacji danych ustalono podobne poziomy zmian dla średnich ocen parametru regresji α_1 i parametru autoregresji q oraz dla wartości średnich policzonych dla miar oddziaływania przestrzennego A_D, A_I, A_R, A_T w zależności od wyjściowych wartości parametrów autoregresji dla procesów X_1^*, X_2^* (zob. tabele 5.24–5.27).

W wyniku przeprowadzonych badań ustalono występowanie problemu skali w przypadku agregacji danych przestrzennych dla procesów charakteryzujących się własnością autokorelacji przestrzennej. Należy jednak podkreślić, że nie ustalono prawidłowości zmian w wyznaczanych ocenach parametrów w zależności od siły autokorelacji przestrzennej na niższym poziomie agregacji. Odnotowano różne wielkości zmian w ocenach parametrów i dodatkowo dla obydwu procesów X_1^* oraz X_2^* zmiany te przebiegały odmiennie. Otrzymane rezultaty mogą być wynikiem faktu, zgodnie z którym za charakter zmian w otrzymywanych ocenach parametrów w wyniku procesu agregacji danych odpowiada rzeczywista struktura przestrzenna. Oznacza, to że na skutek procesu agregacji wpływ problemu skali na otrzymywane wyniki może być różny w zależności od poziomu autokorelacji przestrzennej analizowanych procesów oraz charakteru struktury przestrzennej rozpatrywanego obszaru. Dlatego istotnym jest uwzględnienie w ramach procedury symulacyjnej, zarówno identyfikacji zależności przestrzennych dla analizowanych procesów, jak i prawidłowego odzwierciedlenia przez generowane dane rzeczywistej struktury przestrzennej dla przyjętego obszaru wniosków.

Rozdział 6.

Symulacyjna analiza problemu skali na podstawie ważonych danych przestrzennych

6.1. Identyfikacja problemu skali dla szumów przestrzennych

W rozdziale piątym rozpatrzony został problem skali w odniesieniu do procesów przestrzennych, których generowane realizacje były wyrażone w wartościach średnich. W niniejszym rozdziale zbadano występowanie problemu skali dla procesów przestrzennych, gdzie generowane dane przestrzenne ważne były wielkością bądź siłą ekonomiczną regionu w celu zapewnienia porównywalności obszarów. Jak zostało opisane w rozdziale pierwszym, w przypadku danych przestrzennych tego rodzaju pojawia się problem ich poprawnej agregacji, która musi być wykonana w sposób pośredni za pomocą równań (1.2)–(1.4). Na niższym poziomie agregacji dane powinny zostać rozbite na licznik oraz mianownik¹. Następnie wykonywana jest osobna agregacja danych przestrzennych z licznika oraz osobna agregacja danych z mianownika. Ostatecznie w wyniku zastosowania operacji dzielenia otrzymuje się zagregowane dane przestrzenne, zapewniające porównywalność obszarów na wyższym poziomie agregacji. W związku z tym procedura symulacyjna pozwalająca na rozważenie problemu skali w obecnym aspekcie musiała zostać rozbudowana².

Analiza symulacyjna rozpoczęta została od wyboru hipotetycznego jednorodnego zbioru układów, do którego ponownie przyporządkowane zostały

¹ W większości sytuacji dane dotyczące licznika, jak i mianownika są udostępniane przez statystykę publiczną.

² Proces agregacji był w tym przypadku bardziej skomplikowany w porównaniu z agregacją danych przestrzennych wyrażonych w wartościach średnich.

dwa układy jednostek terytorialnych, układ NUTS 5 oraz układ NUTS 4. Podobnie, jak w rozdziale piątym, występowanie problemu skali sprawdzone zostało najpierw dla przypadku agregacji danych przestrzennych będących realizacją dwóch szumów przestrzennych X_1 oraz X_2 , stacjonarnych w szerszym sensie. Proces X_1 oraz proces X_2 określone zostały za pomocą następujących równań:

$$X_1 = \frac{X_{1,1}}{X_{1,2}}, X_2 = \frac{X_{2,1}}{X_{2,2}}, \quad (6.1)$$

przy czym

$$X_{1,2} = X_{2,2}, \quad (6.2)$$

$$X_i = \alpha_{0i} + \varepsilon_i, \quad (6.3)$$

$$X_{i,j} = \alpha_{0i,j} + \varepsilon_{i,j}, \quad (6.4)$$

gdzie: $X_1, X_2, X_{1,1}, X_{2,1}, X_{1,2}, X_{2,2}$ to procesy szumu przestrzennego, $X_i, X_{i,j}$ oznaczają wektory wartości odpowiednich szumów przestrzennych, $\alpha_{0i}, \alpha_{0i,j}$ stanowią parametry odpowiadające za poziom wartości oczekiwanej procesu, a $\varepsilon_i, \varepsilon_{i,j}$ są wektorami wartości białych szumów przestrzennych.

Należy zwrócić uwagę na fakt, że zgodnie z równaniem (6.2) mamy do czynienia z pięcioma szumami przestrzennymi $X_1, X_2, X_{1,1}, X_{2,1}, X_{1,2}$, co wynika z przyjęcia wspólnego mianownika dla obydwu procesów X_1 oraz X_2 . Procesy $X_1, X_2, X_{1,1}, X_{2,1}, X_{1,2}$ charakteryzują się stałym poziomem wartości oczekiwanej i wariancji oraz brakiem występowania autokowariancji przestrzennej. Dodatkowo przyjęte jest założenie, że wszystkie procesy przestrzenne mogą być wzajemnie skorelowane.

Pojawia się tutaj jednak problem wynikający z przyjętej w równaniu (6.1) zależności funkcyjnej, wiążącej odpowiednie procesy. Zgodnie z założoną zależnością funkcyjną (np. $X_1 = X_{1,1}/X_{1,2}$) możliwa jest symulacja wzajemnie skorelowanych danych przestrzennych, odniesionych jedynie do dwóch procesów, a wartości danych odniesionych do trzeciego procesu muszą być obliczone na podstawie równania (6.1). Oznacza to, że nie jest możliwe jednoczesne wygenerowanie danych przestrzennych odnoszących się do wszystkich procesów i część danych musi być wyznaczona zgodnie z przyjętą zależnością funkcyjną.

W związku z tym założono scenariusz polegający na generowaniu danych przestrzennych jedynie dla szumów przestrzennych $X_1, X_2, X_{1,2}$, które odniesione zostały do układu jednostek terytorialnych NUTS 5³. Dla procesu X_1 za-

³ Możliwe są inne warianty wyboru procesów przestrzennych w procedurze symulacyjnej. Zagadnienie to będzie jednak przedmiotem kolejnych badań autora.

łożona została wartość oczekiwana równa 10 i odchylenie standardowe równe 0,5, dla procesu X_2 wartość oczekiwana równa 20 i odchylenie standardowe równe 1, a dla procesu $X_{1,2}$ wartość oczekiwana równa 30 i odchylenie standardowe równe 1,5. Założone własności procesów szumów przestrzennych przedstawione zostały w tabeli 6.1. Natomiast wartości danych przestrzennych dla procesów $X_{1,1}$, $X_{2,1}$ policzone zostały zgodnie z równaniem (6.1).

Tabela 6.1.

Własności procesów szumów przestrzennych

Proces X_1			Proces X_2			Proces $X_{1,2}$		
$E(X_1)$	$S(X_1)$	$V(X_1)$	$E(X_2)$	$S(X_2)$	$V(X_2)$	$E(X_{1,2})$	$S(X_{1,2})$	$V(X_{1,2})$
10,0	0,5	5,0	20,0	1,0	5,0	30,0	1,5	5,0
Brak autokorelacji przestrzennej			Brak autokorelacji przestrzennej			Brak autokorelacji przestrzennej		

Źródło: opracowanie własne.

Zgodnie z przyjętym założeniem o wzajemnym skorelowaniu procesów szumów przestrzennych X_1 , X_2 , $X_{1,2}$ ustalono trzy warianty zależności korelacyjnych dla tych procesów. W pierwszym wariantcie założono strukturę korelacyjną przedstawioną w tabeli 6.2. W wariantcie tym przyjęto silną, dodatnią zależność korelacyjną między procesami X_1 , X_2 oraz dodatnie zależności korelacyjne między procesami X_1 , $X_{1,2}$ i X_2 , $X_{1,2}$, jednak o mniejszej sile. W wariantcie drugim założono brak korelacji między procesami X_1 , X_2 , $X_{1,2}$. Natomiast w wariantcie trzecim założono między procesami X_1 , X_2 silną, ujemną zależność korelacyjną, między procesami X_1 , $X_{1,2}$ zależność o dodatnim charakterze, a między procesami X_2 , $X_{1,2}$ zależność ujemną.

Biorąc pod uwagę wcześniejsze założenia odnośnie do własności procesów oraz wariantów zależności korelacyjnych między nimi wykonano procedurę symulacyjną w dwóch krokach. W pierwszym kroku dla każdego wariantu generowano po 10000 realizacji⁴ procesów X_1 , X_2 , $X_{1,2}$ na poziomie agregacji NUTS 5. Następnie w związku z niemożliwością jednoczesnego generowania realizacji procesów $X_{1,1}$, $X_{2,1}$, wartości danych przestrzennych policzone zostały na podstawie równania (6.1). W ten sposób uzyskano po 10000 układów danych przestrzennych dla procesów X_1 , X_2 , $X_{1,1}$, $X_{2,1}$, $X_{1,2}$, gdzie dane przestrzenne odniesione zostały do układu jednostek terytorialnych NUTS 5.

W kroku drugim, wykonano proces agregacji na podstawie wygenerowanych danych przestrzennych dla procesów $X_{1,1}$, $X_{2,1}$, $X_{1,2}$ zgodnie z równaniami (1.2)–(1.4). Agregacja danych przestrzennych do poziomu agregacji NUTS 4 polegała na sumowaniu danych dla odpowiedniej grupy regionów z układu na niższym poziomie agregacji i odniesieniu tej sumy do odpowiedniego regionu

⁴ Generowanie większej liczby realizacji nie zmieniało dokładności otrzymywanych wyników.

z układu na wyższym poziomie agregacji (NUTS 4). Na koniec, za pomocą operacji ilorazu wyznaczone zostały zagregowane dane przestrzenne dla procesów X_1 , X_2 .

Tabela 6.2.

Założona struktura korelacyjna procesów X_1 , X_2 , $X_{1,2}$

Wariant pierwszy			
	X_1	X_2	$X_{1,2}$
X_1	1,0	0,9	0,5
X_2		1,0	0,4
$X_{1,2}$			1,0
Wariant drugi			
	X_1	X_2	$X_{1,2}$
X_1	1,0	0,0	0,0
X_2		1,0	0,0
$X_{1,2}$			1,0
Wariant trzeci			
	X_1	X_2	$X_{1,2}$
X_1	1,0	-0,9	0,5
X_2		1,0	-0,4
$X_{1,2}$			1,0

Źródło: opracowanie własne.

W wyniku wykonania kolejnych kroków procedury symulacyjnej, dla każdego z wariantów otrzymano na obydwu poziomach agregacji po 10000 realizacji procesów X_1 , X_2 , $X_{1,1}$, $X_{2,1}$, $X_{1,2}$. Następnie dla wszystkich procesów przestrzennych, obliczono na podstawie każdej z realizacji średnią, odchylenie standardowe oraz wykonano test Morana⁵. Dodatkowo dla pojedynczych realizacji procesów X_1 , X_2 , wykonana została analiza regresyjna dla procesu X_1 względem procesu X_2 , gdzie przyjęto specyfikację modelową zgodną z równaniem (5.2). W przypadku analizy regresji wykonano również test Morana dla reszt, co pozwoliło na ocenę występowania autokorelacji przestrzennej procesu resztowego. Następnie na podstawie wyznaczonych ocen obliczone zostały podstawowe statystyki opisowe: średnia, odchylenie standardowe oraz współczynnik zmienności, co pozwoliło na ustalenie zmian w otrzymywanych wynikach dotyczących własności procesów X_1 oraz X_2 na skutek procesu agregacji

⁵ Ze względu na czytelność tabel wyniki testu Morana nie zostały przedstawione. Wszystkie procesy przestrzenne, zarówno na poziomie agregacji NUTS 5, jak i na poziomie NUTS 4 charakteryzowały się brakiem autokorelacji przestrzennej.

danych. Uzyskane rezultaty dla kolejnych wariantów zależności korelacyjnych między szumami przestrzennymi zamieszczono w tabelach 6.3–6.8.

Tabela 6.3.

Analiza problemu skali dla szumów przestrzennych: wariant pierwszy

Układ NUTS 5								
Analiza podstawowych własności								
Statystyki opisowe	$E(X_1)$	$E(X_2)$	$E(X_{1,2})$	$E(X_{1,1})$	$E(X_{2,1})$	$S(X_1)$	$S(X_2)$	$S(X_{1,2})$
Średnia	10,00	20,00	30,00	300	600	0,50	1,00	1,50
Odch. std.	0,01	0,02	0,03	0,47	0,94	0,01	0,01	0,02
Wsp. zm. (%)	0,10	0,10	0,10	0,16	0,16	1,55	1,50	1,27
Analiz regresji								
Statystyki opisowe	$S(X_{1,1})$	$S(X_{2,1})$	α_1	p	R^2	$V(r)$	$I(r)$	p
Średnia	25,96	50,21	0,45	~0,00	0,81	0,05	~0,00	0,50
Odch. std.	0,37	0,70	0,00	~0,00	0,01	0,00	0,02	0,33
Wsp. zm. (%)	1,41	1,39	0,97	–	0,81	0,85	–	–
Układ NUTS 4								
Analiza podstawowych własności								
Statystyki opisowe	$E(X_1)$	$E(X_2)$	$E(X_{1,2})$	$E(X_{1,1})$	$E(X_{2,1})$	$S(X_1)$	$S(X_2)$	$S(X_{1,2})$
Średnia	10,01	20,02	196,13	1 964,04	3 927,22	0,27	0,54	112,30
Odch. std.	0,01	0,03	0,20	3,10	6,17	0,01	0,02	0,25
Wsp. zm. (%)	0,14	0,14	0,10	0,16	0,16	4,75	4,33	0,22
Analiza regresji								
Statystyki opisowe	$S(X_{1,1})$	$S(X_{2,1})$	α_1	p	R^2	$V(r)$	$I(r)$	p
Średnia	1 125,86	2 251,19	0,45	~0,00	0,81	0,03	~0,00	0,48
Odch. std.	4,14	7,59	0,01	~0,00	0,02	0,00	0,03	0,29
Wsp. zm. (%)	0,37	0,34	2,99	–	2,68	2,89	–	–

Źródło: opracowanie własne.

Przedstawione w tabelach 6.3, 6.5, 6.7 wartości średniej, odchylenia standardowego oraz współczynnika zmienności, obliczonych na podstawie zbioru ocen szacowanych parametrów, pozwoliły na rozpatrzenie wpływu procesu agregacji danych na otrzymane rezultaty. Wnioski rozdzielone zostały na dwie części: dotyczące analizy własności procesów X_1 i X_2 oraz analizy procesów $X_{1,1}$, $X_{2,1}$, $X_{1,2}$ jako składowych procesów X_1 i X_2 .

Analiza otrzymanych wyników w zakresie ocen wartości oczekiwanej kolejnych procesów prowadzi do następujących wniosków. W przypadku procesów X_1 oraz X_2 obliczona średnia ocen wartości oczekiwanej $E(X_1)$ i $E(X_2)$ ukształto-

wała się na tym samym poziomie dla obydwu poziomów agregacji. Oznacza to, że nie występowały systematyczne zmiany otrzymywanych ocen dla wartości oczekiwanej $E(X_1)$ i $E(X_2)$ na skutek procesu agregacji danych przestrzennych.

Tabela 6.4.

Analiza korelacyjna dla procesów $X_1, X_2, X_{1,2}$: wariant pierwszy

NUTS 5					
	X_1	X_2	$X_{1,2}$	$X_{1,1}$	$X_{2,1}$
X_1	1,00	0,90 (0,4)	0,50 (2,26)	0,86 (0,41)	0,83 (0,51)
X_2		1,00	0,40 (3,38)	0,75 (0,91)	0,84 (0,51)
$X_{1,2}$			1,00	0,86 (0,48)	0,84 (0,64)
$X_{1,1}$				1,00	0,97 (0,13)
$X_{2,1}$					1,00
NUTS 4					
	X_1	X_2	$X_{1,2}$	$X_{1,1}$	$X_{2,1}$
X_1	1,00	0,90 (1,34)	0,02 (339,69)	0,04 (161,14)	0,04 (152,37)
X_2		1,00	0,01 (474,79)	0,04 (188,38)	0,04 (176,17)
$X_{1,2}$			1,00	1,00 (0,00)	1,00 (0,00)
$X_{1,1}$				1,00	1,00 (0,00)
$X_{2,1}$					1,00

Uwagi:

Pogrubione zostały, założone i kontrolowane podczas symulacji wartości zależności korelacyjnej między procesami. W nawiasach podano wartości procentowe współczynnika zmienności.

Źródło: opracowanie własne.

W wyniku procesu agregacji obniżeniu uległy jednak wartości średnich z ocen odchylenia standardowego $S(X_1)$ i $S(X_2)$. Podobnie, jak w przykładzie analizy danych przestrzennych wyrażonych w wartościach średnich z rozdziału piątego, spadek poziomu odchylenia standardowego procesów X_1 i X_2 był wysoki i wyniósł 46%. Dla wszystkich wariantów zależności korelacyjnych między procesami wartości średnich dla odchyłeń standardowych $S(X_1)$, $S(X_2)$ wyno-

siły kolejno 0,5 oraz 1 na poziomie agregacji NUTS 5, a na poziomie agregacji NUTS 4 odpowiednio 0,27 oraz 0,54 (zob. tabele 6.3, 6.5, 6.7).

Tabela 6.5.

Analiza problemu skali dla szumów przestrzennych: wariant drugi

Układ NUTS 5								
Analiza podstawowych własności								
Statystyki opisowe	$E(X_1)$	$E(X_2)$	$E(X_{1,2})$	$E(X_{1,1})$	$E(X_{2,1})$	$S(X_1)$	$S(X_2)$	$S(X_{1,2})$
Średnia	10,00	20,00	30,01	300,08	600,02	0,50	1,00	1,50
Odch. std.	0,01	0,02	0,03	0,38	0,84	0,01	0,01	0,02
Wsp. zm. (%)	0,09	0,09	0,11	0,13	0,14	1,46	1,29	1,47
Analiza regresji								
Statystyki opisowe	$S(X_{1,1})$	$S(X_{2,1})$	α_1	p	R^2	$V(r)$	$I(r)$	p
Średnia	21,23	42,53	0,00	-0,00	0,00	0,07	-0,00	0,46
Odch. std.	0,34	0,61	0,01	-0,00	0,00	0,00	0,01	0,30
Wsp. zm. (%)	1,62	1,44	-	-	1126,79	0,70	-	-
Układ NUTS 4								
Analiza podstawowych własności								
Statystyki opisowe	$E(X_1)$	$E(X_2)$	$E(X_{1,2})$	$E(X_{1,1})$	$E(X_{2,1})$	$S(X_1)$	$S(X_2)$	$S(X_{1,2})$
Średnia	10,00	20,00	196,18	1 961,99	3 923,06	0,27	0,54	112,30
Odch. std.	0,01	0,03	0,21	2,49	5,51	0,02	0,03	0,24
Wsp. zm. (%)	0,11	0,14	0,11	0,13	0,14	6,07	4,89	0,21
Analiza regresji								
Statystyki opisowe	$S(X_{1,1})$	$S(X_{2,1})$	α_1	p	R^2	$V(r)$	$I(r)$	p
Średnia	1 123,79	2 246,58	0,00	-0,00	0,00	0,05	-0,00	0,49
Odch. std.	2,79	6,24	0,03	-0,00	0,01	0,00	0,03	0,28
Wsp. zm. (%)	0,25	0,28	-	-	293,15	3,01	-	-

Źródło: opracowanie własne.

Natomiast w wyniku agregacji danych przestrzennych dla procesów $X_{1,1}$, $X_{2,1}$, $X_{1,2}$ obliczone wartości średnie ocen dotyczących wartości oczekiwanej $E(X_{1,1})$, $E(X_{2,1})$, $E(X_{1,2})$ uległy znacznemu podwyższeniu. Uzyskany rezultat jest konsekwencją zastosowanego sposobu agregacji danych przestrzennych dla procesów $X_{1,1}$, $X_{2,1}$, $X_{1,2}$, gdzie wartości danych na wyższym poziomie agregacji NUTS 4 były wynikiem sumy wybranych wartości danych z niższego poziomu agregacji. Również średnie ocen odchylenia standardowego $S(X_{1,1})$, $S(X_{2,1})$, $S(X_{1,2})$ uległy znacznemu podwyższeniu na skutek zmiany poziomu agregacji.

Podobnie, jak w przypadku wyników dla wartości oczekiwanej, fakt ten jest konsekwencją sposobu agregacji danych (zob. tabele 6.3, 6.5, 6.7).

Tabela 6.6.

Analiza korelacyjna dla procesów $X_1, X_2, X_{1,2}$: wariant drugi

NUTS 5					
	X_1	X_2	$X_{1,2}$	$X_{1,1}$	$X_{2,1}$
X_1	1,00	0,00 –	0,00 –	0,71 (1,54)	0,00 –
X_2		1,00	0,00 –	0,00 –	0,71 (1,51)
$X_{1,2}$			1,00	0,71 (1,27)	0,71 (1,37)
$X_{1,1}$				1,00	0,50 (2,86)
$X_{2,1}$					1,00
NUTS 4					
	X_1	X_2	$X_{1,2}$	$X_{1,1}$	$X_{2,1}$
X_1	1,00	0,00 –	0,00 –	0,00 –	0,00 –
X_2		1,00	0,00 –	0,00 –	0,00 –
$X_{1,2}$			1,00	1,00 (0,00)	1,00 (0,00)
$X_{1,1}$				1,00	1,00 (0,00)
$X_{2,1}$					1,00

Źródło: opracowanie własne.

Otrzymane wyniki dla procesów $X_{1,1}, X_{2,1}, X_{1,2}$ świadczą o konieczności unikania analiz na podstawie danych przestrzennych, dla których proces agregacji polega wyłącznie na sumowaniu odpowiednich wartości. Wykorzystanie tego typu danych przestrzennych nie pozwala na porównywanie jednostek terytorialnych i dodatkowo otrzymywane wyniki dotyczące podstawowych własności analizowanych procesów ulegają znacznym zmianom przy przejściu na wyższy poziom agregacji.

Dodatkowo dla wszystkich procesów $X_1, X_2, X_{1,1}, X_{2,1}, X_{1,2}$ wykonano test Morana. Dla żadnego z procesów nie stwierdzono występowania autokorelacji przestrzennej na wyższym poziomie agregacji NUTS 4. Oznacza to, że agrega-

cja danych przestrzennych nie wpłynęła na pojawienie się autozależności przestrzennych w badanych procesach⁶.

Tabela 6.7.

Analiza problemu skali dla szumów przestrzennych: wariant trzeci

Układ NUTS 5								
Analiza podstawowych własności								
Statystyki opisowe	$E(X_1)$	$E(X_2)$	$E(X_{1,2})$	$E(X_{1,1})$	$E(X_{2,1})$	$S(X_1)$	$S(X_2)$	$S(X_{1,2})$
Średnia	10,00	20,00	30,00	300,44	599,37	0,50	1,00	1,50
Odch. std.	0,01	0,02	0,03	0,53	0,78	0,01	0,01	0,02
Wsp. zm. (%)	0,10	0,11	0,11	0,18	0,13	1,35	1,27	1,50
Analiza regresji								
Statystyki opisowe	$S(X_{1,1})$	$S(X_{2,1})$	α_1	p	R^2	$V(r)$	$I(r)$	p
Średnia	25,98	32,81	-0,45	-0,00	0,81	0,05	-0,00	0,45
Odch. std.	0,34	0,53	0,00	-0,00	0,01	0,00	0,01	0,29
Wsp. zm. (%)	1,31	1,61	-0,95	-	0,78	0,68	-	-
Układ NUTS 4								
Analiza podstawowych własności								
Statystyki opisowe	$E(X_1)$	$E(X_2)$	$E(X_{1,2})$	$E(X_{1,1})$	$E(X_{2,1})$	$S(X_1)$	$S(X_2)$	$S(X_{1,2})$
Średnia	10,01	19,99	196,16	1 964,32	3 918,81	0,27	0,54	112,35
Odch. std.	0,02	0,03	0,21	3,46	5,10	0,01	0,03	0,22
Wsp. zm. (%)	0,16	0,16	0,11	0,18	0,13	5,44	5,32	0,19
Analiza regresji								
Statystyki opisowe	$S(X_{1,1})$	$S(X_{2,1})$	α_1	p	R^2	$V(r)$	$I(r)$	p
Średnia	1 126,72	2 243,69	-0,45	-0,00	0,81	0,03	-0,00	0,47
Odch. std.	3,61	5,01	0,02	-0,00	0,03	0,00	0,03	0,27
Wsp. zm. (%)	0,32	0,22	-3,89	-	3,36	2,83	-	-

Źródło: opracowanie własne.

W tabelach 6.4, 6.6, 6.8 przedstawiono średnie ocen współczynnika korelacji liniowej Pearsona dla procesów X_1 , X_2 , $X_{1,1}$, $X_{2,1}$ oraz podano w nawiasach wartości procentowe współczynnika zmienności. W wyniku analizy uzyskanych rezultatów stwierdzono, że wyznaczone wartości średnich ocen współczynnika korelacji liniowej Pearsona dla pary procesów X_1 , X_2 miały zbliżone wartości na obydwu poziomach agregacji, przy czym nastąpił wzrost wartości współczyn-

⁶ Ze względu na konieczność zapewnienia czytelności prezentowanych tabel wyniki przeprowadzonego testu Morana dla procesów X_1 , X_2 , $X_{1,1}$, $X_{2,1}$, $X_{1,2}$ nie zostały w nich zamieszczone.

nika zmienności na wyższym poziomie agregacji NUTS 4. Oznacza to, że rozkład współczynnika korelacji liniowej Pearsona dla procesów X_1, X_2 na wyższym poziomie agregacji może charakteryzować się większą zmiennością. Z kolei zależności korelacyjne między procesami $X_{1,1}, X_{2,1}, X_{1,2}$ uległy znacznemu podwyższeniu w wyniku procesu agregacji danych przestrzennych (współczynnik korelacji liniowej wynosił 1). Natomiast w pozostałych przypadkach par procesów, zależności korelacyjne okazały się statystycznie nieistotne na wyższym poziomie agregacji (zob. tabele 6.4, 6.6, 6.8). Do otrzymanych wyników należy jednak podejść ostrożnie, ponieważ mogą one po części wynikać z niedoskonałości przeprowadzonej symulacji. Niedoskonałość procedury symulacyjnej wynika z niemożliwości symulacji wszystkich procesów przestrzennych, z których dwa są obliczane na podstawie zależności funkcyjnej. W kolejnych badaniach konieczne jest również rozważenie kwestii ulepszenia procedury symulacyjnej w przypadku generowania realizacji procesów $X_1, X_2, X_{1,1}, X_{2,1}, X_{1,2}$ przy ustalonej strukturze korelacyjnej.

Tabela 6.8.

Analiza korelacyjna dla procesów $X_1, X_2, X_{1,2}$: wariant trzeci

NUTS 5					
	X_1	X_2	$X_{1,2}$	$X_{1,1}$	$X_{2,1}$
X_1	1,00	-0,90 (0,39)	0,50 (3,13)	0,86 (0,57)	-0,37 (5,04)
X_2		1,00	-0,40 (4,29)	-0,75 (1,11)	0,54 (2,80)
$X_{1,2}$			1,00	0,86 (0,65)	0,55 (2,49)
$X_{1,1}$				1,00	0,10 (19,99)
$X_{2,1}$					1,00
NUTS 4					
	X_1	X_2	$X_{1,2}$	$X_{1,1}$	$X_{2,1}$
X_1	1,00	-0,90 (1,68)	0,03 (223,05)	0,06 (125,22)	0,01 (741,23)
X_2		1,00	-0,03 (215,31)	-0,05 (127,39)	-0,01 (926,09)
$X_{1,2}$			1,00	1,00 (0,00)	1,00 (0,00)
$X_{1,1}$				1,00	1,00 (0,02)
$X_{2,1}$					1,00

Źródło: opracowanie własne.

W tabelach 6.3, 6.5, 6.7 przedstawione zostały również wyniki estymacji parametrów modelu regresji liniowej dla procesu X_1 względem procesu X_2 . W przypadku współczynnika regresji α_1 oraz współczynnika determinacji R^2 obliczone średnie dla zbioru ich ocen miały podobne wartości na obydwu poziomach agregacji. Wyznaczone wartości średnich dla statystyki Morana $I(r)$ oraz wartości p pozwalają na stwierdzenie, że proces resztowy na obydwu poziomach agregacji charakteryzuje się brakiem własności autokorelacji przestrzennej. Oznacza to, że w przypadku procesów przestrzennych o stałym poziomie wartości oczekiwanej, wariancji i braku autokorelacji przestrzennej, analiza regresji stanowi poprawne narzędzie badania zależności ekonomicznych.

Na podstawie uzyskanych wyników ustalono występowanie istotnych różnic w ocenach własności procesów przestrzennych $X_{1,1'}$, $X_{2,1'}$, $X_{1,2}$ na skutek procesu agregacji danych przestrzennych. Natomiast w przypadku procesów X_1 , X_2 nie stwierdzono występowania problemu skali. Należy pokreślić fakt występujących różnic w kształtowaniu się zmian w otrzymywanych ocenach parametrów pod wpływem procesu agregacji, pomiędzy procesami X_1 , X_2 oraz procesami $X_{1,1'}$, $X_{2,1'}$, $X_{1,2}$. Oznacza to konieczność uwzględniania zarówno w przestrzennych analizach ekonomicznych, jak i w badaniach nad problemem skali odpowiedniego charakteru zmiennych. Kolejnym krokiem w badaniach będzie rozpatrzenie problemu skali, podobnie jak w rozdziale piątym, przy założeniu dodatkowych własności procesów przestrzennych w postaci trendu przestrzennego oraz autokorelacji przestrzennej.

6.2. Badanie problemu skali w warunkach występowania niejednorodności systematycznej

Podobnie jak w poprzednim rozdziale, po wykonaniu analizy problemu skali dla procesów szumu przestrzennego, rozpatrzone zostaną procesy charakteryzujące się niejednorodnością systematyczną. Autor ponownie skupi się na zastosowaniu modelu trendu przestrzennego pierwszego stopnia do opisu własności niejednorodności systematycznej. W tym przypadku rozważone jednak zostaną procesy przestrzenne, gdzie dla odzwierciedlenia ich zmienności wykorzystywane są dane przestrzenne ważne odpowiednio rozumianą wielkością regionu. Jak zostało opisane w podrozdziale 6.1. w przypadku danych przestrzennych tego typu proces agregacji musi być wykonany w sposób pośredni.

Analizowane procesy X_1 oraz X_2 zostają określone ponownie za pomocą równania (6.1). W tym przypadku nie zostanie założona jednak równość procesów przestrzennych $X_{1,2}$ oraz $X_{2,2'}$, tak jak w równaniu (6.2). W związku z tym, zgodnie z równaniem (6.1) analiza symulacyjna dotyczyć będzie sześciu procesów przestrzennych⁷. Ze względu na ograniczenie wynikające z przyjętej zależności funkcyjnej, możliwa jest symulacja realizacji jedynie czterech procesów przestrzennych. Natomiast wartości realizacji odniesionych do pozostałych

⁷ Zakłada się, że wszystkie procesy przestrzenne mogą być wzajemnie skorelowane.

dwóch procesów muszą być obliczone na podstawie owej zależności funkcyjnej wiążącej odpowiednie procesy. Ustalono, że będą generowane realizacje następujących czterech procesów $X_1, X_2, X_{1,2}, X_{2,2}$, gdzie specyfikacje modelowe dla tych procesów wyrażone zostały za pomocą równań:

$$\mathbf{X}_i = \beta_{0i} + \beta_{1i}s_1 + \beta_{2i}s_2 + \boldsymbol{\varepsilon}_i, \quad (6.5)$$

$$\mathbf{X}_{i,j} = \beta_{0i,j} + \boldsymbol{\varepsilon}_{i,j}, \quad (6.6)$$

gdzie $\mathbf{X}_i, \mathbf{X}_{i,j}$ są wektorami wartości odpowiednich procesów przestrzennych, $\beta_{0i}, \beta_{1i}, \beta_{2i}$ stanowią parametry trendu przestrzennego, $\beta_{0i,j}$ jest parametrem odpowiadającym za poziom wartości oczekiwanej procesu, s_1, s_2 to koordynaty przestrzenne, a $\boldsymbol{\varepsilon}_i, \boldsymbol{\varepsilon}_{i,j}$ są wektorami wartości białych szumów przestrzennych. Natomiast realizacje procesów $X_{1,1}, X_{2,1}$ wyznaczone zostały na podstawie zależności liniowej zgodnie z równaniem (6.1).

Dodatkowo założona została zależność regresyjna dla procesu X_1 względem procesu X_2 określona równaniem (5.6). Po ustaleniu specyfikacji równań dla kolejnych procesów przestrzennych oraz specyfikacji modelowej uwzględniającej zależność między procesami, problem skali został ponownie rozważony na podstawie wyników analizy symulacyjnej. W pierwszym kroku procedury generowane były realizacje procesów przestrzennych $X_1, X_2, X_{1,2}, X_{2,2}$ na poziomie agregacji NUTS 5. Dla procesów przestrzennych X_1, X_2 założono własność systematycznej niejednorodności, opisaną za pomocą modelu przestrzennego trendu pierwszego stopnia zgodnego z równaniem (6.5), a dla procesów $X_{1,1}, X_{2,2}$ założono własność jednorodności przestrzennej zgodnie z równaniem (6.6)⁸. Dla wszystkich procesów założono brak autokorelacji przestrzennej procesu resztowego.

Przyjęte wartości parametrów w modelach generujących dane przedstawione zostały w tabeli 6.9. Ponownie dla modeli trendu przestrzennego ustalono wielkość zmienności procesu resztowego zapewniającą stopień dopasowania modelu do danych na poziomie zbliżonym do 80%.

⁸ Procesy $X_{1,2}, X_{2,2}$ jako szumy przestrzenne charakteryzują się brakiem autokorelacji przestrzennej.

Tabela 6.9.

Własności procesów przestrzennych $X_1, X_2, X_{1,2}, X_{2,2}$

Parametry przestrzennego trendu pierwszego stopnia							
Proces X_1				Proces X_2			
β_{01}	β_{11}	β_{21}	R^2	β_{02}	β_{12}	β_{22}	R^2
10,0	1,5	3,0	0,8	20,0	0,5	1,0	0,8
Brak autokorelacji przestrzennej				Brak autokorelacji przestrzennej			
Statystyki opisowe szumu przestrzennego							
Proces $X_{1,2}$			Proces $X_{2,2}$				
$E(X_{1,2})$	$S(X_{1,2})$	$V(X_{1,2})$	$E(X_{2,2})$	$S(X_{2,2})$	$V(X_{2,2})$		
30,0	1,5	5,0	40,0	2,0	5,0		
Brak autokorelacji przestrzennej			Brak autokorelacji przestrzennej				

Źródło: opracowanie własne.

Zbiór realizacji procesów przestrzennych generowany był w trzech wariantach, ze względu na odmienne zależności korelacyjne między procesami. W wariantcie pierwszym procedury symulacyjnej założono dla procesów $X_1, X_2, X_{1,2}, X_{2,2}$ zależności korelacyjne przedstawione w tabeli 6.10. W wariantcie tym założono silną, dodatnią zależność korelacyjną między procesami X_1, X_2 oraz dodatnie zależności korelacyjne o mniejszej sile między parami procesów $(X_{1,2}, X_{2,2}), (X_1, X_{1,2}), (X_1, X_{2,2}), (X_2, X_{1,2})$ i $(X_2, X_{2,2})$. W wariantcie drugim założono brak korelacji między procesami przestrzennymi $X_1, X_2, X_{1,2}, X_{2,2}$. Natomiast w wariantcie trzecim założono silną, ujemną zależność między procesami X_1, X_2 oraz ujemne zależności korelacyjne o mniejszej sile dla par procesów $(X_1, X_{2,2})$ i $(X_2, X_{1,2})$. Z kolei dodatnie zależności korelacyjne założono między procesami $X_{1,2}, X_{2,2}, X_1, X_{1,2}$ oraz $X_2, X_{2,2}$.

W kolejnym kroku dokonano osobno agregacji realizacji procesów $X_{1,1}, X_{2,1}, X_{1,2}, X_{2,2}$ zgodnie z równaniem (1.3). W wyniku procesu agregacji otrzymano dane przestrzenne odniesione do układu na wyższym poziomie agregacji NUTS 4. W ostatnim kroku, na podstawie równania (6.1) wyznaczone zostały zagregowane dane przestrzenne dla procesów X_1, X_2 . W wyniku wykonania dwóch kolejnych kroków procedury symulacyjnej otrzymano dla każdego z trzech wariantów 10000 układów danych przestrzennych na poziomie agregacji NUTS 5 oraz 10000 układów danych przestrzennych na poziomie agregacji NUTS 4.

W kolejnym kroku dokonano osobno agregacji realizacji procesów $X_{1,1}, X_{2,1}, X_{1,2}, X_{2,2}$ zgodnie z równaniem (1.3). W wyniku procesu agregacji otrzymano dane przestrzenne odniesione do układu na wyższym poziomie agregacji NUTS 4. W ostatnim kroku, na podstawie równania (6.1) wyznaczone zostały zagregowane dane przestrzenne dla procesów X_1, X_2 . W wyniku wykonania

dwóch kolejnych kroków procedury symulacyjnej otrzymano dla każdego z trzech wariantów 10000 układów danych przestrzennych na poziomie agregacji NUTS 5 oraz 10000 układów danych przestrzennych na poziomie agregacji NUTS 4.

Tabela 6.10.

Założona struktura korelacyjna dla procesów $X_1, X_2, X_{1,2}, X_{2,2}$

Wariant pierwszy				
	X_1	X_2	$X_{1,2}$	$X_{2,2}$
X_1	1,0	0,9	0,5	0,4
X_2		1,0	0,4	0,5
$X_{1,2}$			1,0	0,6
$X_{2,2}$				1,0
Wariant drugi				
	X_1	X_2	$X_{1,2}$	$X_{2,2}$
X_1	1,0	0,0	0,0	0,0
X_2		1,0	0,0	0,0
$X_{1,2}$			1,0	0,0
$X_{2,2}$				1,0
Wariant trzeci				
	X_1	X_2	$X_{1,2}$	$X_{2,2}$
X_1	1,0	-0,9	0,5	-0,4
X_2		1,0	-0,4	0,5
$X_{1,2}$			1,0	0,6
$X_{2,2}$				1,0

Źródło: opracowanie własne.

W kolejnym kroku dokonano osobno agregacji realizacji procesów $X_{1,1'}, X_{2,1'}, X_{1,2'}, X_{2,2'}$ zgodnie z równaniem (1.3). W wyniku procesu agregacji otrzymano dane przestrzenne odniesione do układu na wyższym poziomie agregacji NUTS 4. W ostatnim kroku, na podstawie równania (6.1) wyznaczone zostały zagregowane dane przestrzenne dla procesów $X_{1'}, X_{2'}$. W wyniku wykonania dwóch kolejnych kroków procedury symulacyjnej otrzymano dla każdego z trzech wariantów 10000 układów danych przestrzennych na poziomie agregacji NUTS 5 oraz 10000 układów danych przestrzennych na poziomie agregacji NUTS 4.

Następnie przeprowadzono ocenę własności procesów przestrzennych $X_{1'}, X_{2'}, X_{1,1'}, X_{2,1'}, X_{1,2'}, X_{2,2'}$ na obydwu poziomach agregacji. Dla procesów przestrzennych $X_{1'}, X_{2'}, X_{1,1'}, X_{2,1'}$ dokonano estymacji parametrów modeli przestrzennego

trendu stopnia pierwszego, a dla procesów przestrzennych $X_{1,2}$, $X_{2,2}$ wyznaczono podstawowe statystyki opisowe, zgodnie z równaniami (6.5) i (6.6). Następnie zgodnie z przyjętą specyfikacją modelu regresji zapisaną za pomocą równania (5.6) oszacowano parametry modelu oraz wykonano test Morana dla procesu resztowego. Na podstawie zbiorów otrzymanych ocen parametrów policzono statystyki opisowe w postaci wartości średniej, odchylenia standardowego oraz współczynnika zmienności. Wyniki otrzymane dla kolejnych wariantów struktury zależności korelacyjnych między procesami przestrzennymi przedstawiono w tabelach 6.11–6.19.

Tabela 6.11.

Analiza własności procesów oraz zależności między nimi przed agregacją: wariant pierwszy

Układ NUTS 5										
Analiza trendu przestrzennego dla procesów X_1 oraz $X_{1,1}$										
Statystyki opisowe	$\beta_1(X_1)$	p	$\beta_2(X_1)$	p	R^2	$\beta_1(X_2)$	p	$\beta_2(X_2)$	p	R^2
Średnia	1,50	~0,00	3,00	~0,00	0,81	44,96	~0,00	90,55	~0,00	0,48
Odch. std.	0,03	~0,00	0,03	~0,00	0,01	1,77	~0,00	2,01	~0,00	0,01
Wsp. zm. (%)	1,77	–	0,96	–	0,66	3,93	–	2,22	–	2,74
Analiza trendu przestrzennego dla procesów X_2 oraz $X_{2,1}$										
Statystyki opisowe	$\beta_1(X_1)$	p	$\beta_2(X_1)$	p	R^2	$\beta_1(X_2)$	p	$\beta_2(X_2)$	p	R^2
Średnia	0,50	~0,00	1,00	~0,00	0,81	20,04	~0,00	40,10	~0,00	0,48
Odch. std.	0,01	~0,00	0,01	~0,00	0,01	0,84	~0,00	0,79	~0,00	0,01
Wsp. zm. (%)	1,68	–	0,90	–	0,62	4,18	–	1,97	–	2,43
Analiza własności procesów $X_{1,2}$ oraz $X_{2,2}$										
Statystyki opisowe	$E(X_{1,2})$	$E(X_{2,2})$	$S(X_{1,2})$	$S(X_{2,2})$	$I(X_{1,2})$	p	$I(X_{2,2})$	p	–	–
Średnia	29,99	40,00	1,50	2,00	~0,00	0,52	0,00	0,49	–	–
Odch. std.	0,03	0,04	0,02	0,02	0,01	0,28	0,01	0,29	–	–
Wsp. zm. (%)	0,11	0,11	1,54	1,23	–	–	–	–	–	–
Analiza regresji										
Statystyki opisowe	α_1	p	β_1	p	β_2	p	R^2	V	$I(r)$	p
Średnia	2,72	~0,00	0,14	~0,00	0,28	~0,00	0,96	0,01	~0,00	0,52
Odch. std.	0,03	~0,00	0,02	~0,00	0,03	~0,00	0,00	0,00	0,01	0,30
Wsp. zm. (%)	1,11	–	15,36	–	11,53	–	0,11	0,73	–	–

Źródło: opracowanie własne.

Tabela 6.12.

Analiza własności procesów oraz zależności między nimi po wykonaniu agregacji danych: wariant pierwszy

Układ NUTS 4										
Analiza trendu przestrzennego dla procesów X_1 oraz $X_{1,1}$										
Statystyki opisowe	$\beta_1(X_1)$	p	$\beta_2(X_1)$	p	R^2	$\beta_1(X_{1,1})$	p	$\beta_2(X_{1,1})$	p	R^2
Średnia	1,50	~0,00	3,00	~0,00	0,94	1093,87	~0,00	452,56	~0,00	0,05
Odch. std.	0,04	~0,00	0,05	~0,00	0,01	11,92	~0,00	12,56	~0,00	0,00
Wsp. zm. (%)	2,52	-	1,58	-	0,67	1,09	-	2,78	-	2,22
Analiza trendu przestrzennego dla procesów X_2 oraz $X_{2,1}$										
Statystyki opisowe	$\beta_1(X_2)$	p	$\beta_2(X_2)$	p	R^2	$\beta_1(X_{2,1})$	p	$\beta_2(X_{2,1})$	p	R^2
Średnia	0,50	~0,00	1,00	~0,00	0,94	484,48	~0,00	200,49	~0,00	0,05
Odch. std.	0,01	~0,00	0,02	~0,00	0,01	5,79	~0,00	4,95	~0,00	0,00
Wsp. zm. (%)	2,48	-	1,60	-	0,71	1,20	-	2,47	-	2,39
Analiza własności procesów $X_{1,2}$ oraz $X_{2,2}$										
Statystyki opisowe	$E(X_{1,2})$	$E(X_{2,2})$	$S(X_{1,2})$	$S(X_{2,2})$	$I(X_{1,2})$	p	$I(X_{2,2})$	p	-	-
Średnia	196,11	261,54	112,31	149,74	-0,01	0,62	-0,01	0,62	-	-
Odch. std.	0,21	0,29	0,20	0,28	0,00	0,02	0,00	0,02	-	-
Wsp. zm. (%)	0,11	0,11	0,18	0,19	-	-	-	-	-	-
Analiza regresji										
Statystyki opisowe	α_1	p	β_1	p	β_2	p	R^2	V	$I(r)$	p
Średnia	2,73	~0,00	0,13	~0,00	0,27	~0,00	0,99	0,01	-0,01	0,53
Odch. std.	0,09	~0,00	0,05	~0,00	0,10	~0,00	0,00	0,00	0,03	0,25
Wsp. zm. (%)	3,46	-	36,30	-	37,21	-	0,14	2,95	-	-

Źródło: opracowanie własne.

W tabelach 6.11–6.12, 6.14–6.15, 6.17–6.18 przedstawiono obliczone na podstawie zbiorów ocen kolejnych parametrów wartości średnich tych ocen, odchyłeń standardowych oraz współczynników zmienności. Dla wszystkich trzech wariantów struktury zależności korelacyjnych między procesami X_1 , X_2 , $X_{1,2}$, $X_{2,2}$ otrzymano podobne wyniki w sensie oceny występowania problemu skali. Ponownie wnioski podzielone zostały na analizę zmian dla procesów X_1 i X_2 oraz na analizę zmian dla procesów $X_{1,1}$, $X_{2,1}$, $X_{1,2}$, $X_{2,2}$ jako składowych procesów X_1 i X_2 .

Tabela 6.13.

Analiza korelacyjna dla procesów $X_1, X_2, X_{1,2}, X_{2,2}$: wariant pierwszy

NUTS 5						
	X_1	X_2	$X_{1,2}$	$X_{2,2}$	$X_{1,1}$	$X_{2,1}$
X_1	1,00	0,90 (0,38)	0,50 (3,46)	0,40 (4,13)	0,80 (0,95)	0,69 (1,44)
X_2		1,00	0,40 (4,41)	0,50 (2,94)	0,69 (1,45)	0,81 (0,84)
$X_{1,2}$			1,00	0,60 (2,26)	0,91 (0,42)	0,60 (2,43)
$X_{2,2}$				1,00	0,60 (2,24)	0,91 (0,34)
$X_{1,1}$					1,00	0,73 (1,33)
$X_{2,1}$						1,00
NUTS 4						
	X_1	X_2	$X_{1,2}$	$X_{2,2}$	$X_{1,1}$	$X_{2,1}$
X_1	1,00	0,9 (1,61)	0,02 (365,23)	0,02 (419,68)	0,04 (186,92)	0,03 (211,31)
X_2		1,00	0,00 –	0,00 –	0,02 (322,54)	0,03 (274,47)
$X_{1,2}$			1,00	0,99 (0,01)	0,98 (0,03)	0,98 (0,04)
$X_{2,2}$				1,00	0,98 (0,02)	0,98 (0,03)
$X_{1,1}$					1,00	0,99 (0,01)
$X_{2,1}$						1,00

Źródło: opracowanie własne.

Jako pierwsze zostały poddane ocenie potencjalne zmiany w kształtowaniu się trendu przestrzennego na skutek procesu agregacji danych przestrzennych, osobno dla procesu X_1 oraz procesu X_2 . Obliczone średnie ocen parametrów $\beta_1(X_1)$, $\beta_2(X_1)$ dla procesu X_1 oraz średnie ocen parametrów $\beta_1(X_2)$, $\beta_2(X_2)$ dla procesu X_2 nie ulegały zmianie na wyższym poziomie agregacji NUTS 4. Oznacza to, że nie występowały systematyczne zmiany w otrzymywanych ocenach dla modelu trendu przestrzennego na skutek procesu agregacji. Dla obydwu modeli trendu przestrzennego wartości współczynnika determinacji R^2 uległy podwyższeniu o około 15% w wyniku procesu agregacji oraz nie stwierdzono na podstawie testu Morana występowania własności autokorelacji przestrzennej procesu resztowego przed i po wykonaniu agregacji danych.

Tabela 6.14.

Analiza własności procesów oraz zależności między nimi przed agregacją: wariant drugi

Układ NUTS 5										
Analiza trendu przestrzennego dla procesów X_1 oraz $X_{1,1}$										
Statystyki opisowe	$\beta_1(X_1)$	p	$\beta_2(X_1)$	p	R^2	$\beta_1(X_{1,1})$	p	$\beta_2(X_{1,1})$	p	R^2
Średnia	1,50	-0,00	2,99	-0,00	0,80	44,93	-0,00	89,88	-0,00	0,57
Odch. std.	0,03	-0,00	0,03	-0,00	0,01	1,34	-0,00	1,52	-0,00	0,01
Wsp. zm. (%)	1,81	-	0,97	-	0,65	2,98	-	1,69	-	1,70
Analiza trendu przestrzennego dla procesów X_2 oraz $X_{2,1}$										
Statystyki opisowe	$\beta_1(X_2)$	p	$\beta_2(X_2)$	p	R^2	$\beta_1(X_{2,1})$	p	$\beta_2(X_{2,1})$	p	R^2
Średnia	0,50	-0,00	1,00	-0,00	0,81	20,17	-0,00	39,99	-0,00	0,57
Odch. std.	0,01	-0,00	0,01	-0,00	0,01	0,70	-0,00	0,80	-0,00	0,01
Wsp. zm. (%)	1,70	-	0,95	-	0,67	3,47	-	1,99	-	2,06
Analiza własności procesów $X_{1,2}$ oraz $X_{2,2}$										
Statystyki opisowe	$E(X_{1,2})$	$E(X_{2,2})$	$S(X_{1,2})$	$S(X_{2,2})$	$I(X_{1,2})$	p	$I(X_{2,2})$	p	-	-
Średnia	30,00	40,00	1,50	2,01	-0,01	0,61	0,00	0,45	-	-
Odch. std.	0,03	0,04	0,02	0,03	0,01	0,26	0,01	0,29	-	-
Wsp. zm. (%)	0,10	0,11	1,48	1,27	-	-	-	-	-	-
Analiza regresji										
Statystyki opisowe	α_1	p	β_1	p	β_2	p	R^2	V	$I(r)$	p
Średnia	0,00	0,54	1,50	-0,00	2,99	-0,00	0,80	0,02	0,00	0,49
Odch. std.	0,05	0,28	0,04	-0,00	0,06	-0,00	0,01	0,00	0,01	0,31
Wsp. zm. (%)	-	-	2,39	-	2,12	-	0,65	0,69	-	-

Źródło: opracowanie własne.

Zupełnie inaczej kształtowały się zmiany trendu przestrzennego w wyniku agregacji danych dla procesów $X_{1,1}$ oraz procesu $X_{2,1}$. Na poziomie agregacji NUTS 4 wartości średnich ocen parametrów $\beta_1(X_{1,1})$, $\beta_2(X_{1,1})$ dla procesu $X_{1,1}$ oraz średnich ocen parametrów $\beta_1(X_{2,1})$, $\beta_2(X_{2,1})$ dla procesu $X_{2,1}$ były znacznie wyższe w porównaniu z wartościami otrzymanymi na poziomie agregacji NUTS 5. Również w przypadku agregacji danych przestrzennych dla procesów $X_{1,2}$ i $X_{2,1}$ obliczone średnie z ocen dotyczących wartości oczekiwanych $E(X_{1,2})$, $E(X_{2,2})$ oraz odchyłeń standardowych $S(X_{1,2})$, $S(X_{2,2})$ uległy znacznemu podwyższeniu. Fakt ten ponownie stanowi konsekwencję zastosowanego sposobu agregacji danych przestrzennych polegającego na wykonaniu operacji sumowania wybranych wartości danych przestrzennych.

Tabela 6.15.

Analiza własności procesów oraz zależności między nimi po wykonaniu agregacji danych: wariant drugi

Układ NUTS 4										
Analiza trendu przestrzennego dla procesów X_1 oraz $X_{1,1}$										
Statystyki opisowe	$\beta_1(X_1)$	p	$\beta_2(X_1)$	p	R^2	$\beta_1(X_{1,1})$	p	$\beta_2(X_{1,1})$	p	R^2
Średnia	1,50	-0,00	3,00	-0,00	0,94	1093,46	-0,00	448,36	-0,00	0,05
Odch. std.	0,03	-0,00	0,04	-0,00	0,01	9,04	-0,00	9,52	-0,00	0,00
Wsp. zm. (%)	2,30	-	1,27	-	0,69	0,83	-	2,12	-	1,74
Analiza trendu przestrzennego dla procesów X_2 oraz $X_{2,1}$										
Statystyki opisowe	$\beta_1(X_2)$	p	$\beta_2(X_2)$	p	R^2	$\beta_1(X_{2,1})$	p	$\beta_2(X_{2,1})$	p	R^2
Średnia	0,50	-0,00	1,00	-0,00	0,94	485,11	-0,00	199,84	-0,00	0,05
Odch. std.	0,01	-0,00	0,01	-0,00	0,01	4,76	-0,00	4,92	-0,00	0,00
Wsp. zm. (%)	2,66	-	1,45	-	0,77	0,98	-	2,46	-	2,05
Analiza własności procesów $X_{1,2}$ oraz $X_{2,2}$										
Statystyki opisowe	$E(X_{1,2})$	$E(X_{2,2})$	$S(X_{1,2})$	$S(X_{2,2})$	$I(X_{1,2})$	p	$I(X_{2,2})$	p	-	-
Średnia	196,17	261,52	112,31	149,74	-0,01	0,62	-0,01	0,61	-	-
Odch. std.	0,20	0,29	0,22	0,28	0,00	0,02	0,00	0,03	-	-
Wsp. zm. (%)	0,10	0,11	0,19	0,19	-	-	-	-	-	-
Analiza regresji										
Statystyki opisowe	α_1	p	β_1	p	β_2	p	R^2	V	$I(r)$	p
Średnia	0,00	0,38	1,49	-0,00	2,99	-0,00	0,94	0,02	-0,01	0,52
Odch. std.	0,23	0,31	0,12	-0,00	0,23	-0,00	0,01	0,00	0,03	0,29
Wsp. zm. (%)	-	-	7,92	-	7,61	-	0,68	2,63	-	-

Źródło: opracowanie własne.

Otrzymane wyniki pozwoliły również na analizę zmian w otrzymywanych ocenach parametrów modelu regresji liniowej dla procesu X_1 względem procesu X_2 określonego za pomocą równania (5.6), przy uwzględnieniu składnika przestrzennego trendu pierwszego stopnia dla procesu objaśnianego. Zarówno średnia ocen współczynnika regresji α_1 , jak i średnie ocen parametrów trendu β_1 , β_2 nie uległy zmianie na skutek procesu agregacji danych. Również w przypadku modelu regresji nie stwierdzono własności autokorelacji przestrzennej procesu resztowego na wyższym poziomie agregacji. Oznacza to, że nie występują systematyczne zmiany w otrzymywanych wynikach badań na skutek procesu agregacji danych również dla modelu regresji poszerzonego o składnik trendu przestrzennego. W przypadku wszystkich rozpatrywanych modeli można stwierdzić również, że proces agregacji danych przestrzennych nie

przyczynia się do powstawania autokorelacji przestrzennej analizowanych procesów (zob. tabele 6.11–6.12, 6.14–6.15, 6.17–6.18).

Tabela 6.16.

Analiza korelacyjna dla procesów $X_1, X_2, X_{1,2}, X_{2,2}$; wariant drugi

NUTS 5						
	X_1	X_2	$X_{1,2}$	$X_{2,2}$	$X_{1,1}$	$X_{2,1}$
X_1	1,00	0,00 –	0,00 –	0,00 –	0,56 (2,40)	0,00 –
X_2		1,00	0,00 –	0,00 –	0,00 –	0,56 (2,49)
$X_{1,2}$			1,00	0,00 –	0,82 (0,93)	0,00 –
$X_{2,2}$				1,00	0,00 –	0,82 (0,75)
$X_{1,1}$					1,00	0,00 –
$X_{2,1}$						1,00
NUTS 4						
	X_1	X_2	$X_{1,2}$	$X_{2,2}$	$X_{1,1}$	$X_{2,1}$
X_1	1,00	0,00 –	0,00 –	0,00 –	0,02 (258,66)	0,00 –
X_2		1,00	0,00 –	0,00 –	0,00 –	0,00 –
$X_{1,2}$			1,00	0,98 (0,01)	0,98 (0,03)	0,98 (0,04)
$X_{2,2}$				1,00	0,98 (0,02)	0,97 (0,03)
$X_{1,1}$					1,00	0,99 (0,01)
$X_{2,1}$						1,00

Źródło: opracowanie własne.

Wyniki analizy zależności korelacyjnych między procesami dla kolejnych wariantów przedstawione zostały w tabelach 6.13, 6.16 oraz 6.19. Dla każdego z wariantów wyznaczone wartości średnie z ocen współczynnika korelacji liniowej Pearsona dla procesów X_1, X_2 nie ulegały zmianom pod wpływem procesu agregacji. Ponownie jednak wystąpił wzrost wartości współczynnika zmienności wyznaczanych ocen. Oznacza to, że rozkład współczynnika korelacji liniowej Pearsona dla procesów X_1, X_2 może charakteryzować się większym rozstępem na wyższym poziomie agregacji.

Tabela 6.17.

Analiza własności procesów oraz zależności między nimi przed agregacją: wariant trzeci

Układ NUTS 5										
Analiza trendu przestrzennego dla procesów X_1 oraz $X_{1,1}$										
Statystyki opisowe	$\beta_1(X_1)$	p	$\beta_2(X_1)$	p	R^2	$\beta_1(X_{1,1})$	p	$\beta_2(X_{1,1})$	p	R^2
Średnia	1,50	-0,00	3,00	-0,00	0,80	45,19	-0,00	89,91	-0,00	0,47
Odch. std.	0,03	-0,00	0,03	-0,00	0,01	2,12	-0,00	2,14	-0,00	0,01
Wsp. zm. (%)	2,02	-	1,11	-	0,62	4,69	-	2,38	-	2,64
Analiza trendu przestrzennego dla procesów X_2 oraz $X_{2,1}$										
Statystyki opisowe	$\beta_1(X_2)$	p	$\beta_2(X_2)$	p	R^2	$\beta_1(X_{2,1})$	p	$\beta_2(X_{2,1})$	p	R^2
Średnia	0,50	-0,00	1,00	-0,00	0,81	20,02	-0,00	40,09	-0,00	0,48
Odch. std.	0,01	-0,00	0,01	-0,00	0,01	0,91	-0,00	0,97	-0,00	0,01
Wsp. zm. (%)	2,03	-	1,14	-	0,69	4,53	-	2,42	-	2,92
Analiza własności procesów $X_{1,2}$ oraz $X_{2,2}$										
Statystyki opisowe	$E(X_{1,2})$	$E(X_{2,2})$	$S(X_{1,2})$	$S(X_{2,2})$	$I(X_{1,2})$	p	$I(X_{2,2})$	p	-	-
Średnia	29,99	40,00	1,50	2,00	0,00	0,52	0,00	0,49	-	-
Odch. std.	0,03	0,04	0,02	0,02	0,01	0,28	0,01	0,29	-	-
Wsp. zm. (%)	0,11	0,11	1,54	1,23	-	-	-	-	-	-
Analiza regresji										
Statystyki opisowe	α_1	p	β_1	p	β_2	p	R^2	V	$I(r)$	p
Średnia	-2,99	-0,00	3,00	-0,00	5,99	-0,00	0,99	0,01	0,00	0,50
Odch. std.	0,01	-0,00	0,01	-0,00	0,01	-0,00	0,00	0,00	0,02	0,33
Wsp. zm. (%)	-0,25	-	0,20	-	0,13	-	0,01	0,74	-	-

Źródło: opracowanie własne.

Natomiast w przypadku wszystkich wariantów dla par procesów, gdzie pierwszy proces należał do zbioru X_1, X_2 , a drugi proces należał do zbioru $X_{1,1}, X_{2,1}, X_{1,2}, X_{2,2}$ zależności korelacyjne okazały się statystycznie nieistotne na wyższym poziomie agregacji NUTS 4. Z kolei siła zależności korelacyjnych dla par procesów ze zbioru $X_{1,1}, X_{2,1}, X_{1,2}, X_{2,2}$ uległa znacznemu podwyższeniu w wyniku procesu agregacji i była bliska jedności.

Tabela 6.18.

Analiza własności procesów oraz zależności między nimi po wykonaniu agregacji danych: wariant trzeci

Układ NUTS 4										
Analiza trendu przestrzennego dla procesów X_1 oraz $X_{1,1}$										
Statystyki opisowe	$\beta_1(X_1)$	p	$\beta_2(X_1)$	p	R^2	$\beta_1(X_{1,1})$	p	$\beta_2(X_{1,1})$	p	R^2
Średnia	1,51	~0,00	3,00	~0,00	0,93	1095,77	~0,00	448,68	~0,00	0,05
Odch. std.	0,03	~0,00	0,04	~0,00	0,01	14,51	~0,00	13,22	~0,00	0,00
Wsp. zm. (%)	2,32	-	1,48	-	0,62	1,32	-	2,95	-	2,46
Analiza trendu przestrzennego dla procesów X_2 oraz $X_{2,1}$										
Statystyki opisowe	$\beta_1(X_2)$	p	$\beta_2(X_2)$	p	R^2	$\beta_1(X_{2,1})$	p	$\beta_2(X_{2,1})$	p	R^2
Średnia	0,50	~0,00	1,00	~0,00	0,94	484,50	~0,00	200,47	~0,00	0,05
Odch. std.	0,01	~0,00	0,01	~0,00	0,01	5,94	~0,00	6,00	~0,00	0,00
Wsp. zm. (%)	2,31	-	1,49	-	0,68	1,23	-	2,99	-	2,68
Analiza własności procesów $X_{1,2}$ oraz $X_{2,2}$										
Statystyki opisowe	$E(X_{1,2})$	$E(X_{2,2})$	$S(X_{1,2})$	$S(X_{2,2})$	$I(X_{1,2})$	p	$I(X_{2,2})$	p	-	-
Średnia	196,12	261,54	112,28	149,76	-0,01	0,62	-0,01	0,61	-	-
Odch. std.	0,17	0,23	0,22	0,27	0,00	0,03	0,00	0,02	-	-
Wsp. zm. (%)	0,09	0,09	0,20	0,18	-	-	-	-	-	-
Analiza regresji										
Statystyki opisowe	α_1	p	β_1	p	β_2	p	R^2	V	$I(r)$	p
Średnia	-2,98	~0,00	2,99	~0,00	5,98	~0,00	1,00	0,01	0,03	0,30
Odch. std.	0,04	~0,00	0,02	~0,00	0,04	~0,00	0,00	0,00	0,04	0,28
Wsp. zm. (%)	-1,23	-	0,63	-	0,63	-	0,02	2,25	-	-

Źródło: opracowanie własne.

Na podstawie uzyskanych wyników przeprowadzonych analiz stwierdzono różnice w kształtowaniu się zmian ocen parametrów pod wpływem procesu agregacji, między procesami X_1 , X_2 oraz procesami $X_{1,1}$, $X_{2,1}$, $X_{1,2}$, $X_{2,2}$. Dla procesów X_1 , X_2 nie zidentyfikowano problemu skali. Oznacza to, że analiza trendowo-regresyjna na podstawie procesów, których realizacje wyrażone są za pomocą ważonych danych przestrzennych powinna pozwalać na otrzymywanie zbliżonych wyników na różnych poziomach agregacji.

Tabela 6.19.

Analiza korelacyjna dla procesów $X_1, X_2, X_{1,2}, X_{2,2}$; wariant trzeci

NUTS 5						
	X_1	X_2	$X_{1,2}$	$X_{2,2}$	$X_{1,1}$	$X_{2,1}$
X_1	1,00	-0,90 (0,04)	0,50 (2,78)	-0,40 (4,95)	0,80 (0,74)	-0,73 (1,5)
X_2		1,00	-0,40 (3,67)	0,49 (3,41)	-0,74 (1,07)	0,80 (1,02)
$X_{1,2}$			1,00	0,60 (2,25)	0,91 (0,41)	0,22 (7,92)
$X_{2,2}$				1,00	0,22 (9,63)	0,91 (0,33)
$X_{1,1}$					1,00	-0,19 (10,44)
$X_{2,1}$						1,00

NUTS 4						
	X_1	X_2	$X_{1,2}$	$X_{2,2}$	$X_{1,1}$	$X_{2,1}$
X_1	1,00	-0,90 (0,17)	0,02 (291,00)	0,00 -	0,04 (170,25)	-0,01 (491,02)
X_2		1,00	0,00 -	0,02 (303,49)	-0,02 (409,87)	0,04 (175,35)
$X_{1,2}$			1,00	0,98 (0,01)	0,97 (0,03)	0,97 (0,04)
$X_{2,2}$				1,00	0,98 (0,02)	0,97 (0,03)
$X_{1,1}$					1,00	0,98 (0,02)
$X_{2,1}$						1,00

Źródło: opracowanie własne.

Natomiast w przypadku procesów $X_{1,1}, X_{2,1}, X_{1,2}, X_{2,2}$ ustalono występowanie systematycznych zmian na skutek procesu agregacji. Dla procesów $X_{1,1}$ oraz $X_{2,1}$ stwierdzono istotne zmiany w ocenach parametrów trendu przestrzennego przy zmianie poziomu agregacji na wyższy. Z kolei procesy $X_{1,2}$ i $X_{2,2}$ charakteryzowały się podobnym poziomem przyrostu procentowego wartości średniej oraz odchylenia standardowego w wyniku agregacji danych. Otrzymane rezultaty świadczą o konieczności unikania w przestrzennych badaniach ekonomicznych analiz dotyczących procesów, których charakter zgodny jest z procesami $X_{1,1}, X_{2,1}, X_{1,2}, X_{2,2}$.

6.3. Analiza symulacyjna dla procesów o własności autokorelacji przestrzennej

Procesy przestrzenne rozpatrywane w poprzednich podrozdziałach charakteryzowały się stałym poziomem wartości oczekiwanej i wariancji oraz brakiem autokorelacji przestrzennej, a następnie systematycznymi zmianami przestrzennymi w wartości oczekiwanej. W niniejszym podrozdziale rozważony został problem skali dla procesów przestrzennych charakteryzujących się autokorelacją przestrzenną. W tym celu przeprowadzone zostało badanie, w ramach którego generowane były ponownie ważne dane przestrzenne stanowiące realizacje procesów X_1, X_2 o własnościach autozależności przestrzennych. Przeprowadzona analiza symulacyjna pozwoliła na określenie zmian w otrzymywanych ocenach parametrów zachodzących na skutek procesu agregacji danych dla tego typu procesów.

W związku z tym w przyjętej procedurze symulacyjnej, podobnie jak w podrozdziale 6.1 generowane były realizacje pięciu procesów przestrzennych $X_1, X_2, X_{1,2}, X_{1,1}, X_{2,1}$, dla których zależności funkcyjne określone zostały w równaniu (6.1)⁹. Ponownie generowano realizacje trzech procesów $X_1, X_2, X_{1,2}$, natomiast realizacje procesów $X_{1,1}, X_{2,1}$ wyznaczone zostały na podstawie założonych zależności funkcyjnych. W rozdziale piątym stwierdzono, że w przypadku procesów przestrzennych o własności autokorelacji przestrzennej, generowanie zbioru realizacji na podstawie modelu SAR nie gwarantuje odzwierciedlenia struktury przestrzennej o tym samym charakterze. Podkreślone zostało, że estymacja modelu SAR pozwala na identyfikację własności autokorelacji przestrzennej dla wybranych procesów, jednak wykorzystanie empirycznej postaci modelu SAR nie pozwala na generowanie danych przestrzennych, których przestrzenna zmienność zgadzałaby się z rzeczywistą strukturą zależności przestrzennych analizowanego obszaru. W związku z tym w rozdziale piątym zaproponowana została autorska procedura symulacyjna polegająca na wykorzystaniu empirycznych rozkładów procesów w ramach przyjętego obszaru analizy. W niniejszym podrozdziale dane odniesione do układu jednostek terytorialnych NUTS 5 generowane były również zgodnie z rozwiązaniem opartym na empirycznym rozkładzie procesów, dzięki czemu odwzorowane zostały zależności między procesami oraz własność autokorelacji przestrzennej.

W związku z tym przyjęto trzy procesy przestrzenne na poziomie agregacji NUTS 5, dochody własne gmin na 1 mieszkańca (proces X_1), liczbę podmiotów gospodarczych przypadających na 10000 mieszkańców (proces X_2) oraz liczbę ludności (proces $X_{1,2}$)¹⁰, gdzie pomiędzy pierwszymi dwoma procesami zachodzi zależność ekonomiczna. W tabelach 6.20–6.22 przedstawiono sta-

⁹ Tak jak w podrozdziale 6.1 założono, że $X_{1,2}=X_{2,2}$, co oznacza tożsamość mianowników wyrażeń definiujących procesy X_1, X_2 .

¹⁰ Dane przestrzenne dla roku 2017 pozyskane zostały z Banku Danych Lokalnych Głównego Urzędu Statystycznego.

tystyki opisowe rozkładów rozważanych procesów. Wszystkie procesy charakteryzowały się wysokim udziałem odchylenia standardowego w średniej, w przypadku liczby ludności przekraczającym 100%. Dla każdego z rozkładów zidentyfikowano silną, dodatnią skośność. Uzyskane statystyki opisowe wskazały na znaczne odstępstwa empirycznych rozkładów od rozkładu normalnego. W związku z tym wykonano dwa testy na zgodność z rozkładem normalnym oraz rozkładem gamma. W przypadku wszystkich procesów stwierdzono brak zgodności z rozkładem normalnym oraz z rozkładem gamma. Wyznaczone statystyki testów wskazują jednak na wyższą zgodność rozkładów empirycznych z rozkładem gamma. Przeprowadzony został również test Morana. Obliczone statystyki testu pozwalają na stwierdzenie występowania dodatniej autokorelacji przestrzennej dla każdego z procesów.

Tabela 6.20.

Analiza rozkładu dochodów własnych gmin na 1 mieszkańca

Statystyki opisowe			
Średnia	1689,21	Wsp. zm. (%)	76,29
Odch. std.	1288,78	Współczynnik skośności	17,19
Wyniki testów statystycznych			
Rodzaj testu	Statystyka testu		Wartość p
Test Morana	0,37		~0,00
Test na zgodność rozkładu normalnego	103670,49		~0,00
Test na zgodność rozkładu gamma	10,55		~0,00

Źródło: opracowanie własne.

Tabela 6.21.

Analiza rozkładu liczby podmiotów gospodarczych przypadających na 10000 mieszkańców

Statystyki opisowe			
Średnia	797,52	Wsp. zm. (%)	44,12
Odch. std.	351,89	Współczynnik skośności	5,39
Wyniki testów statystycznych			
Rodzaj testu	Statystyka testu		Wartość p
Test Morana	0,38		~0,00
Test na zgodność rozkładu normalnego	2068,36		~0,00
Test na zgodność rozkładu gamma	8,90		~0,00

Źródło: opracowanie własne.

Tabela 6.22.
Analiza rozkładu ludności

Statystyki opisowe			
Średnia	12125,33	Wsp. zm. (%)	127,66
Odch. std.	15479,00	Współczynnik skośności	4,76
Wyniki testów statystycznych			
Rodzaj testu	Statystyka testu		Wartość p
Test Morana	0,18		-0,00
Test na zgodność rozkładu normalnego	15223,00		-0,00
Test na zgodność rozkładu gamma	18,97		-0,00

Źródło: opracowanie własne.

Zgodnie z przyjętą procedurą symulacyjną w pierwszym kroku wygenerowano po 10000 realizacji trzech procesów przestrzennych $X_{1'}$, $X_{2'}$, $X_{1,2'}$ a następnie wyznaczono realizacje procesów $X_{1,1'}$, $X_{2,2'}$ na podstawie równania (6.1). Podobnie jak w podrozdziale 5.4, na podstawie empirycznych wartości procesów wyznaczono kolejne centyle: centyl dziesiąty, centyl trzydziesty, centyl pięćdziesiąty, centyl siedemdziesiąty i centyl dziewięćdziesiąty oraz przyporządkowano poszczególne jednostki terytorialne do jednej z sześciu klas według przynależności do odpowiedniego przedziału wartości centyli. W ten sposób otrzymane zostały dyskretne rozkłady jednostek terytorialnych z układu NUTS 5 dla wszystkich trzech procesów. Ustalone rozkłady jednostek terytorialnych wykorzystane zostały następnie przy generowaniu realizacji procesów, gdzie wartości danych przestrzennych losowane były z odpowiedniego przedziału dystrybuanty rozkładu gamma. Odpowiednie postaci rozkładu gamma oszacowano na podstawie empirycznych rozkładów przyjętych procesów¹¹. Zgodnie z przeprowadzonymi rozważaniami w rozdziale piątym, określona w ten sposób procedura symulacyjna pozwoliła na generowanie realizacji procesów o wybranych własnościach przy jednoczesnym uwzględnieniu struktury przestrzennej, z której wynikało przestrzenne zróżnicowanie tych procesów.

W drugim kroku procedury symulacyjnej dokonano agregacji generowanych danych przestrzennych odniesionych do procesów $X_{1,1'}$, $X_{2,1'}$, $X_{1,2}$ na podstawie równania (1.3). Następnie zgodnie z równaniem (6.1) wyznaczono wartości realizacji procesów $X_{1'}$, $X_{2'}$ na wyższym poziomie agregacji (NUTS 4). W wyniku wykonania obydwu kroków, otrzymano dla pięciu procesów zbiory danych przestrzennych odniesionych do układów jednostek terytorialnych: NUTS 5 oraz NUTS 4.

¹¹ W celu obniżenia wartości średniej oraz odchylenia standardowego odniesionych do oszacowanych rozkładów gamma, w przypadku wszystkich rozkładów empirycznych, wartości danych podzielono przez 100.

Wygenerowane zbiory realizacji procesów $X_1, X_2, X_{1,1}, X_{2,1}, X_{1,2}$ pozwoliły na obliczenie wartości średniej i odchylenia standardowego oraz oszacowanie parametrów modeli autoregresji przestrzennej SAR osobno dla każdego z procesów zgodnie z równaniami (6.7) i (6.8). Dodatkowo oszacowano parametry modelu SAR uwzględniającego zależność regresyjną dla procesu X_1 względem procesu X_2 na podstawie równania (6.9) oraz wyznaczono miary średniego oddziaływania A_D, A_I, A_R, A_T . W przypadku analizy regresji dla procesów X_1, X_2 wykonano również test Morana na podstawie reszt. Kolejne specyfikacje modelu autoregresji przestrzennej SAR zapisane zostały za pomocą równań:

$$\mathbf{X}_i = \alpha_{0i} + \mathbf{q}_i \mathbf{W} \mathbf{X}_i + \boldsymbol{\varepsilon}_i, \quad (6.7)$$

$$\mathbf{X}_{i,j} = \alpha_{0i,j} + \mathbf{q}_{i,j} \mathbf{W} \mathbf{X}_{i,j} + \boldsymbol{\varepsilon}_{i,j}, \quad (6.8)$$

$$\mathbf{X}_1 = \alpha_0 + \alpha_1 \mathbf{X}_2 + \mathbf{q} \mathbf{W} \mathbf{X}_2 + \boldsymbol{\varepsilon}, \quad (6.9)$$

gdzie $\mathbf{X}_i, \mathbf{X}_{i,j}$ są wektorami wartości odpowiednich procesów przestrzennych, $\alpha_{0i}, \alpha_{0i,j}, \alpha_0$ oznaczają parametry odpowiadające za poziom wartości oczekiwanej procesów, α_1 jest parametrem strukturalnym modelu regresji, natomiast $\mathbf{q}_i, \mathbf{q}_{i,j}, \mathbf{q}$ są to parametry autoregresji przestrzennej w modelach dla odpowiednich procesów, \mathbf{W} jest standaryzowaną macierzą sąsiedztwa pierwszego rzędu, a $\boldsymbol{\varepsilon}_i, \boldsymbol{\varepsilon}_{i,j}, \boldsymbol{\varepsilon}$ są wektorami wartości białych szumów przestrzennych.

W wyniku przeprowadzonej procedury symulacyjnej, otrzymano zbiory 10000 ocen wartości oczekiwanej i odchylenia standardowego procesów $X_1, X_2, X_{1,1}, X_{2,1}, X_{1,2}$ oraz ocen parametrów modeli SAR. Następnie wyznaczono podstawowe statystyki na podstawie ocen w postaci średniej, odchylenia standardowego oraz współczynnika zmienności. Otrzymane wyniki na poziomie agregacji NUTS 5 przedstawione zostały w tabeli 6.23, a na poziomie NUTS 4 w tabeli 6.24. Dodatkowo dla obydwu poziomów agregacji przeprowadzono analizę korelacyjną, której wyniki zamieszczono w tabeli 6.27.

Przedstawione w tabelach 6.23–6.24 wartości średniej, odchylenia standardowego oraz współczynnika zmienności pozwoliły na ustalenie wpływu procesu agregacji danych na otrzymywane oceny parametrów analizowanych procesów. Ponownie wnioski dotyczące własności procesów podzielone zostały na osobną analizę procesów X_1 i X_2 oraz analizę procesów $X_{1,1}, X_{2,1}, X_{1,2}$.

W przypadku wartości oczekiwanej procesów X_1 i X_2 , obliczone wartości średnich wzrosły, odpowiednio z poziomu 16,99 do 23,21 dla procesu X_1 oraz z poziomu 8,02 do poziomu 10,07 dla procesu X_2 . Zmiany procentowe wyniosły 36,61% dla procesu X_1 oraz 25,62% dla procesu X_2 . Zmiany wartości oczekiwanych $E(X_1), E(X_2)$, na skutek procesu agregacji danych przestrzennych świadczą o występowaniu problemu skali. Odnotowane zmiany wartości oczekiwanych

procesów nie wykazują jednak prawdziwości i najprawdopodobniej wynikają z rzeczywistej struktury przestrzennej obszaru. Otrzymane średnie dla zbioru ocen odchylenia standardowego $S(X_1)$, $S(X_2)$ wskazują, że jego poziom uległ obniżeniu przy zmianie poziomu agregacji, co stanowi naturalny efekt łągodzenia czynników przypadkowych w wyniku procesu agregacji. Należy jednak podkreślić, że zmiany wartości odchylenia standardowego są znaczne i wynoszą $-20,86\%$ w przypadku procesu X_1 oraz $-27,97\%$ w przypadku procesu X_2 (zob. tabele 6.23–6.24).

Tabela 6.23.

Analiza własności procesów oraz zależności między nimi przed agregacją danych

Układ NUTS 5								
Podstawowe własności procesów oraz miary oddziaływania przestrzennego								
Statystyki opisowe	$E(X_1)$	$E(X_2)$	$E(X_{1,2})$	$E(X_{1,1})$	$E(X_{2,1})$	$S(X_1)$	$S(X_2)$	$S(X_{1,2})$
Średnia	16,99	8,02	122,41	2 938,02	1 275,82	14,19	3,86	155,36
Odch. std.	0,06	0,02	0,93	37,56	11,99	0,20	0,03	3,68
Wsp. zm. (%)	0,35	0,18	0,68	1,15	0,87	1,36	0,89	2,33
Statystyki opisowe	$S(X_{1,1})$	$S(X_{2,1})$	$q(X_1)$	p	$q(X_2)$	p	$q(X_{1,2})$	p
Średnia	5 806,02	2 087,84	0,71	$\sim 0,00$	0,74	$\sim 0,00$	0,31	$\sim 0,00$
Odch. std.	209,72	62,29	0,01	$\sim 0,00$	0	$\sim 0,00$	0,02	$\sim 0,00$
Wsp. zm. (%)	3,61	2,85	1,63	–	1,22	–	5,38	–
Statystyki opisowe	$q(X_{1,1})$	p	$q(X_{2,1})$	p	A_0	A_1	A_R	A_T
Średnia	0,35	$\sim 0,00$	0,32	$\sim 0,00$	2,01	0,17	$\sim 0,00$	2,17
Odch. std.	0,02	$\sim 0,00$	0,02	$\sim 0,00$	0,05	0,00	$\sim 0,00$	0,05
Wsp. zm. (%)	5,62	–	6,61	–	2,26	2,26	2,72	2,16
Analiza regresji								
Statystyki opisowe	α_1	p	q	p	R^2	$V(r)$	$I(r)$	p
Średnia	1,95	$\sim 0,00$	0,35	$\sim 0,00$	0,49	0,54	$\sim 0,00$	0,43
Odch. std.	0,04	$\sim 0,00$	0,01	$\sim 0,00$	0,01	0,01	$\sim 0,00$	0,15
Wsp. zm. (%)	1,83	–	3,11	–	2,34	1,93	–	–

Źródło: opracowanie własne.

W przypadku procesów X_1 , $X_{2,1}$, $X_{1,2}$ obliczone wartości średnie dla ocen dotyczących wartości oczekiwanej $\bar{E}(X_{1,1})$, $E(X_{2,1})$, $E(X_{1,2})$ uległy znacznemu podwyższeniu na skutek zmiany poziomu agregacji z poziomu NUTS 5 na poziom NUTS 4. W wyniku agregacji danych przestrzennych wzrosły również wartości średnich ocen odchylenia standardowego $S(X_{1,1})$, $S(X_{2,1})$, $S(X_{1,2})$. Tak, jak zostało wskazane w poprzednich podrozdziałach, wynik ten jest rezultatem zastoso-

wanego sposobu agregacji danych przestrzennych dla procesów $X_{1,1}$, $X_{2,1}$, $X_{1,2}$, gdzie wartości danych uzyskane zostały na podstawie operacji sumowania. (zob. tabele 6.23–6.24).

Tabela 6.24.

Analiza własności procesów oraz zależności między nimi po wykonaniu agregacji danych

Układ NUTS 4								
Podstawowe własności procesów oraz miary oddziaływania przestrzennego								
Statystyki opisowe	$E(X_1)$	$E(X_2)$	$E(X_{1,2})$	$E(X_{1,1})$	$E(X_{2,1})$	$S(X_1)$	$S(X_2)$	$S(X_{1,2})$
Średnia	23,21	10,07	800,46	18 732,35	8 215,52	11,23	2,78	511,17
Odch. std.	0,12	0,03	5,45	215,70	71,28	0,36	0,08	10,12
Wsp. zm. (%)	0,51	0,31	0,68	1,15	0,87	3,16	2,94	1,98
Statystyki opisowe	$S(X_{1,1})$	$S(X_{2,1})$	$q(X_1)$	p	$q(X_2)$	p	$q(X_{1,2})$	p
Średnia	16 094,08	6 560,56	0,55	0	0,55	0	0,11	0,06
Odch. std.	602,51	198,81	0,02	0	0,02	0	0,02	0,05
Wsp. zm. (%)	3,74	3,03	4,21	–	3,66	–	16,34	–
Statystyki opisowe	$q(X_{1,1})$	p	$q(X_{2,1})$	p	A_D	A_I	A_R	A_T
Średnia	0,26	0,00	0,18	0,00	2,58	0,21	~0,00	2,78
Odch. std.	0,03	0,00	0,02	0,00	0,28	0,03	~0,00	0,28
Wsp. zm. (%)	12,40	–	12,37	–	10,91	13,37	11,79	10,11
Analiza regresji								
Statystyki opisowe	α_1	p	q	p	R^2	$V(r)$	$I(r)$	p
Średnia	2,55	~0,00	0,32	~0,00	0,57	0,32	0,05	0,07
Odch. std.	0,11	~0,00	0,03	~0,00	0,03	0,02	0,02	0,05
Wsp. zm. (%)	4,29	–	10,43	–	4,54	5,25	–	–

Źródło: opracowanie własne.

W kolejnym kroku dokonano oceny wyników uzyskanych na podstawie osobnej estymacji parametrów modelu SAR dla każdego z procesów X_1 , X_2 , $X_{1,1}$, $X_{2,1}$, $X_{1,2}$, których specyfikacje wyrażone zostały za pomocą równań (6.7) i (6.8). W tabeli 6.25 przedstawiono zmiany procentowe dla uzyskiwanych wartości średnich z ocen parametrów autoregresji, przy przejściu na wyższy poziom agregacji. W przypadku wszystkich procesów wartości średnich z ocen parametrów autoregresji uległy obniżeniu. W wyniku procesu agregacji wystąpiły następujące spadki średnich z ocen: dla parametru autoregresji $q(X_1)$ miała miejsce zmiana wartości z 0,71 do 0,55, dla parametru $q(X_2)$ zmiana z wartości 0,74 do 0,55, dla parametru $q(X_{1,2})$ zmiana z wartości 0,31 do 0,11, dla parametru $q(X_{1,1})$ zmiana wartości z 0,35 do 0,26 oraz dla parametru $q(X_{2,1})$ zmiana

wartości z 0,32 do 0,18. Dla parametrów $q(X_1)$, $q(X_2)$ oraz $q(X_{1,1})$ zmiany procentowe przekraczały 25%, dla parametru $q(X_{2,1})$ zmiana przekroczyła 40%, a w przypadku parametru $q(X_{1,2})$ zmiana procentowa była większa od 60% (zob. tabele 6.23–6.25).

Tabela 6.25.

Zmiany procentowe dla średnich ocen parametrów autoregresji modeli SAR

Poziom agregacji	$q(X_1)$	$q(X_2)$	$q(X_{1,2})$	$q(X_{1,1})$	$q(X_{2,1})$
NUTS 5	0,71	0,74	0,31	0,35	0,32
NUTS 4	0,55	0,55	0,11	0,26	0,18
Zmiana procentowa (%)	-22,54	-25,68	-64,52	-25,71	-43,75

Źródło: opracowanie własne.

W ostatnim kroku przeprowadzono ocenę wyników otrzymanych na podstawie estymacji parametrów modelu regresji określonego równaniem (6.9). Należy podkreślić, że wszystkie wartości średnich ocen parametrów modelu regresji uległy zmianie w wyniku agregacji generowanych danych przestrzennych. Dla parametru regresji α_1 wystąpił wzrost wartości średniej ocen z poziomu 1,95 do poziomu 2,55, a dla parametru autoregresji q spadek wartości średniej ocen z poziomu 0,35 do 0,32. Znaczny wzrost średniej ocen dla parametru regresji α_1 przy niższym spadku wartości średniej ocen dla parametru autoregresji q przyczynił się do wzrostu wartości średnich dla wszystkich miar oddziaływania przestrzennego. Średnia dla miary oddziaływania bezpośredniego A_D wzrosła z poziomu 2,01 do poziomu 2,58, średnia dla miary oddziaływania pośredniego A_I z poziomu 0,17 do 0,21 oraz średnia dla miary oddziaływania całkowitego A_T z poziomu 2,17 do 2,78 (zob. tabela 6.23–6.24, 6.26). W tabeli 6.26 zaprezentowane zostały wielkości zmian procentowych średnich ocen parametrów, które wystąpiły na skutek agregacji danych przestrzennych. W przypadku parametru autoregresji zmiana procentowa wyniosła -8,57%, a w przypadku parametru regresji α_1 oraz miar średniego oddziaływania przestrzennego A_D , A_I , A_T zmiana procentowa przekroczyła 20%.

Tabela 6.26.

Zmiany procentowe średnich ocen parametrów modelu regresji oraz miar oddziaływania przestrzennego

Poziom agregacji	α_1	q	A_D	A_I	A_T
NUTS 5	1,95	0,35	2,01	0,17	2,17
NUTS 4	2,55	0,32	2,58	0,21	2,78
Zmiana procentowa (%)	30,77	-8,57	28,36	23,53	28,11

Źródło: opracowanie własne.

Dodatkowo zbadano zależności korelacyjne między procesami X_1 , X_2 , $X_{1,1}$, $X_{2,1}$, $X_{1,2}$ na obydwu poziomach agregacji. Wyniki przedstawione zostały w tabeli 6.27. W przypadku zależności korelacyjnej między procesami X_1 , X_2 odnotowano wzrost wartości współczynnika korelacji liniowej Pearsona z poziomu 0,65 do poziomu 0,72. Otrzymany wynik jest ważny, ponieważ w poprzednich podrozdziałach nie doszło do zmiany zależności korelacyjnej między procesami X_1 , X_2 w wyniku procesu agregacji.

Tabela 6.27.

Analiza korelacyjna dla generowanych danych przestrzennych przed i po agregacji

NUTS 5					
	X_1	X_2	$X_{1,2}$	$X_{1,1}$	$X_{2,1}$
X_1	1,00	0,65 (1,33)	0,40 (2,23)	0,58 (1,61)	0,46 (2,36)
X_2		1,00	0,50 (2,03)	0,54 (2,52)	0,63 (1,70)
$X_{1,2}$			1,00	0,89 (0,94)	0,95 (0,35)
$X_{1,1}$				1,00	0,93 (0,81)
$X_{2,1}$					1,00
NUTS 4					
	X_1	X_2	$X_{1,2}$	$X_{1,1}$	$X_{2,1}$
	1,00	0,72 (2,65)	0,03 (69,54)	0,55 (3,16)	0,26 (8,34)
		1,00	0,11 (16,28)	0,47 (4,22)	0,42 (3,89)
			1,00	0,79 (1,39)	0,91 (0,56)
				1,00	0,91 (0,88)
					1,00

Źródło: opracowanie własne.

Na podstawie wyników zamieszczonych w tabeli 6.27 można stwierdzić również różnice w wynikach uzyskanych na różnych poziomach agregacji, w porównaniu do wyników otrzymanych w podrozdziałach 6.1 i 6.2, dotyczących zależności korelacyjnych dla pozostałych par procesów. Największe spadki wartości współczynnika korelacji liniowej odnotowano dla par procesów (X_1 , $X_{1,2}$), (X_2 , $X_{1,2}$), (X_1 , $X_{2,1}$), ($X_{1,2}$, $X_{2,1}$). Dla pozostałych par procesów zmiany współczynnika korelacji na skutek procesu agregacji były znacznie mniejsze. Należy

zauważyć, że zmiany o których mowa nie są tak znaczne, jak w procedurze symulacyjnej dla procesów szumów przestrzennych oraz procesów charakteryzujących się trendem przestrzennym, gdzie na wyższym poziomie agregacji NUTS 4 stwierdzono wzrost współczynnika korelacji liniowej Pearsona między procesami $X_{1,1}$, $X_{2,2}$, $X_{1,2}$ do jedności oraz brak korelacji między pozostałymi parami procesów. Ustalone zmiany wartości współczynnika korelacji liniowej pozwalają na stwierdzenie, że również dla zależności korelacyjnych między procesami dokonano identyfikacji problemu skali. Przeprowadzona analiza symulacyjna dotyczyła procesów charakteryzujących się własnością autokorelacji przestrzennej, gdzie generowane były ważne dane przestrzenne. Otrzymane wyniki zastosowanej procedury symulacyjnej pozwalają na stwierdzenie występowania zmian w wyznaczanych ocenach parametrów modeli w wyniku procesu agregacji danych, co świadczy o identyfikacji problemu skali. Oznacza to, że w przypadku ustalenia własności autokorelacji przestrzennej procesów w ramach przestrzennych badań ekonomicznych należy założyć możliwość znacznych różnic w otrzymywanych ocenach analizowanych parametrów na różnych poziomach agregacji. Znajomość potencjalnych zmian w ocenach parametrów może ułatwić oraz uwiarygodnić interpretację wyników badań, gdzie otrzymane różnice w ocenach parametrów można uznać za dopuszczalne w ramach występującego problemu skali. Należy podkreślić, że podobnie, jak w rozdziale piątym nie stwierdzono występowania problemu skali dla procesów szumów przestrzennych oraz procesów charakteryzujących się przestrzennym trendem pierwszego stopnia.

Zakończenie

W niniejszej pracy omówiona została potrzeba nowego spojrzenia na obecny w literaturze statystycznej i ekonometrycznej problem zmiennej jednostki odniesienia. W ramach tego problemu rozpatrzono dwa osobne zagadnienia, problem sposobu podziału przestrzeni oraz problem skali. W literaturze wskazywane jest, że istotę obydwu problemów stanowi możliwość uzyskania odmiennych wyników badań na skutek zmiany układu jednostek terytorialnych lub w wyniku wykonania procesu agregacji danych. Zgadając się co do istoty rozważanych problemów dyskusji podlega sposób ich zdefiniowania i podejścia do rozwiązania.

W rozdziale pierwszym przeprowadzono teoretyczne rozważania na temat agregacji danych, ponieważ proces agregacji jest kluczowym elementem problemu zmiennej jednostki odniesienia. Omówienie kolejnych zagadnień pozwoliło na zaproponowanie metodyki budowy systemu agregacji danych, na który składa się zbiór reguł dotyczących gromadzenia danych pierwotnych, procesu agregacji danych oraz udostępniania danych potencjalnym odbiorcom, przy ustalonych układach agregacji (układzie odbiorców, układzie podmiotowym agregacji, układzie przedmiotowym agregacji oraz układzie czasowym). Poprawnie opracowany system agregacji powinien pozwolić na gromadzenie istotnych danych z punktu widzenia odbiorcy oraz zapewnić możliwość odpowiedniego ich udostępniania.

W treści rozdziału poruszona została także kwestia zapewnienia porównywalności obszarów w ramach wykonywanych analiz, gdzie wyróżniono dwa sposoby wyrażenia wartości danych. Zgodnie ze sposobem pierwszym dane wyrażane są w wartościach średnich, natomiast w podejściu drugim dane przestrzenne są ważone w stosunku do odpowiednio rozumianej wielkości obszaru. W związku z tym w rozdziale pierwszym autor przedstawił procedurę prawidłowej agregacji danych przestrzennych w zależności od sposobu ich wyrażenia,

co było pomijane we wcześniejszych pracach dotyczących problemu zmiennej jednostki odniesienia.

W związku z możliwością uzyskania odmiennych wyników badań spowodowanych problemem zmiennej jednostki odniesienia, w rozdziale drugim omówiono zagadnienie poprawności przestrzennych badań ekonomicznych. W ramach przeprowadzonych rozważań przedstawiono propozycje kolejnych autorskich pojęć: własności jednorodności przyczynowej danych przestrzennych, jednorodnego zbioru układów, błędu identyfikacji jednorodnego zbioru układów, błędu wyznaczenia obszaru wniosków, mikrozależności oraz makrozależności przestrzennych.

Pojęcie własności jednorodności przyczynowej jest istotne dla przestrzennych badań ekonomicznych, ponieważ jedynie dane charakteryzujące się tą własnością poprawnie odzwierciedlają oddziaływanie tzw. splotu przyczyn głównych. Analizy przeprowadzone na podstawie jednorodnych przyczynowo danych przestrzennych powinny pozwolić na prawidłową identyfikację przestrzennych własności badanych procesów oraz ekonomicznych zależności między procesami.

Identyfikacja własności jednorodności przyczynowej danych przestrzennych odniesionych do układów jednostek terytorialnych pozwala na wybór właściwych w konkretnym badaniu układów. W związku z tym autor określił jednorodny zbiór układów, jako zbiór układów jednostek terytorialnych na różnych poziomach agregacji, gdzie odniesione do tych układów dane przestrzenne posiadają własność jednorodności przyczynowej. Wykorzystanie układów jednostek terytorialnych należących do jednorodnego zbioru układów powinno zapewnić poprawność przeprowadzanych badań.

W ramach autorskiej propozycji potencjalnych błędów badawczych, jako pierwszy wyróżniono błąd identyfikacji jednorodnego zbioru układów. Błąd ten dotyczy kwestii identyfikacji jednorodnego zbioru układów i polega na wyborze układu jednostek terytorialnych, w przypadku którego dane przestrzenne nie posiadają własności jednorodności przyczynowej. Błąd ten jest istotny dla poprawności badań, ponieważ niejednorodne przyczynowo dane przestrzenne nie odzwierciedlają oddziaływania splotów przyczyn głównych. Wskazane zostało, że ustalenie w ramach badań właściwego zbioru układów jednostek terytorialnych pozwala na uniknięcie błędu identyfikacji jednorodnego zbioru układów.

Jako drugi, wyróżniono błąd wyznaczenia obszaru wniosków, którego popełnienie polega na niewłaściwym powiązaniu stosowanego narzędzia badawczego z ustalonym obszarem wniosków, do którego odniesione zostaną wyniki badań. Istota błędu polega na tym, że w przypadku nie spełnienia wymaganych własności procesów w granicach ustalonego obszaru wniosków, zastosowane narzędzie badawcze traci swoją wartość poznawczą. Autor wskazuje na konieczność identyfikacji tego błędu i ewentualne zmniejszenie granic obszaru

wniosków w celu uzyskania odpowiednich własności procesów przestrzennych lub przyjęcie innych, odpowiednich narzędzi badawczych.

Jako ostatnie rozpatrzone zostało zagadnienie powiązań między funkcjonującymi w literaturze pojęciami mikrozależności oraz makrozależności ekonomicznych. Rozważenie problematyki przestrzennej w tematyce wzajemnej relacji między mikroekonomią a makroekonomią pozwoliło autorowi na propozycję nowych pojęć mikrozależności oraz makrozależności przestrzennych. Do określenia różnicy między mikrozależnościami a makrozależnościami przestrzennymi wykorzystane zostało kryterium oparte na możliwości zmiany charakteru analizowanej zależności ekonomicznej, rozumianej jako istotnej zmiany siły lub postaci funkcyjnej badanej zależności przy przejściu między wybranymi jednostkami terytorialnymi tworzącymi obszar wniosków. Mikrozależności przestrzenne określone zostały jako zależności między analizowanymi procesami, które posiadają stały charakter dla wszystkich jednostek terytorialnych tworzących obszar wniosków. Natomiast makrozależności przestrzenne określone zostały jako zależności, w których dochodzi do zmiany ich charakteru w ramach przyjętego obszaru wniosków. Autor ponownie podkreślił niebezpieczeństwo popełnienia błędu wyznaczenia obszaru wniosków w przypadku identyfikacji makrozależności przestrzennych, ponieważ analizowane w ramach ustalonego obszaru procesy mogą nie posiadać odpowiednich własności dla stosowanego narzędzia badawczego.

W rozdziale trzecim dokonano krytycznego przeglądu literatury dotyczącej problemu zmiennej jednostki odniesienia. Wskazane zostało, że przeprowadzone tamże analizy oparte były na idei losowego wyznaczenia granic układów jednostek terytorialnych w ramach *zoning system* lub *grouping system*. Zgodnie z wnioskami sformułowanymi w rozdziale drugim, w ramach podjętego problemu badawczego powinien zostać utworzony jednorodny zbiór układów składający się z układów danych przestrzennych o własności jednorodności przyczynowej. W sytuacji wykorzystania *zoning system* lub *grouping system*, układy jednostek terytorialnych wyznaczane są w losowy sposób, a odniesione do nich dane nie będą posiadały własności jednorodności przyczynowej, ze względu na pomieszanie oddziaływania różnych splotów przyczyn. W związku z tym, analiza przeprowadzona na podstawie tego typu danych przestrzennych nie pozwoli na poprawną identyfikację własności badanych procesów oraz zależności przyczynowych.

Wykorzystanie *zoning system* oraz *grouping system* stwarza potencjalne niebezpieczeństwo otrzymania niepoprawnych wniosków podczas rozwiązywania postawionego problemu badawczego. Przełożyło się to na wyniki prezentowanych w literaturze badań, gdzie na skutek stosowania systemu losowego generowania układów jednostek terytorialnych otrzymywano przedział wyznaczanych ocen o szerokim rozstępie. W związku z tym, w prezentowanych w literaturze badaniach stwierdzano istotne zmiany w otrzymywanych wynikach w zależności od wyboru układu jednostek terytorialnych. Skutkowało

to każdorazowo identyfikacją problemu zmiennej jednostki odniesienia, którą autor pracy określił jako identyfikację pozorną.

Krytyczna ocena prac z literatury przedmiotu skonfrontowana z autorskimi rozważaniami na temat poprawności przestrzennych badań ekonomicznych wskazała na potrzebę redefinicji problemu sposobu podziału przestrzeni oraz problemu skali. W związku z tym w rozdziale trzecim zaprezentowana została propozycja redefinicji obydwu problemów. Zdefiniowanie i opracowanie w nowy sposób zagadnienia zmiennej jednostki odniesienia stanowi przyczynek do poszerzenia obszaru wiedzy w zakresie ekonometrii przestrzennej i jednocześnie wzbogaca metodykę przestrzennych badań ekonomicznych.

W ramach zaproponowanej redefinicji, problem sposobu podziału przestrzeni został określony jako problem utworzenia pojedynczego układu jednostek terytorialnych na ustalonym poziomie agregacji w taki sposób, by w ramach podjętego problemu badawczego należał on do jednorodnego zbioru układów. Natomiast problem skali określony został jako problem istotnych różnic w otrzymywanych ocenach własności procesów przestrzennych oraz łączących je zależności przyczynowych, gdy analiza wykonywana jest na podstawie układów jednostek terytorialnych przyporządkowanych w ramach podjętego problemu badawczego do jednorodnego zbioru układów.

Należy podkreślić, że rozwiązanie problemu sposobu podziału przestrzeni ma fundamentalne znaczenie dla przestrzennych badań ekonomicznych, ponieważ każda analiza na podstawie danych przestrzennych wymaga ich odniesienia do konkretnego układu jednostek terytorialnych. Wybór właściwego układu jednostek terytorialnych w ramach postawionego problemu badawczego determinuje powodzenie dalszych badań. Decyzja badacza o wyborze układu jednostek terytorialnych, gdzie odniesione do niego dane przestrzenne nie będą posiadały własności jednorodności przyczynowej skutkowałą będzie niemożliwością rozwiązania postawionego problemu badawczego lub może doprowadzić do niepoprawnych wniosków i pozornego rozwiązania tego problemu.

Również analizy mające na celu rozwiązanie problemu skali mają duże znaczenie dla przestrzennych badań ekonomicznych, ponieważ dane przestrzenne najczęściej nie są upubliczniane dla wszystkich poziomów agregacji. W większości przypadków dane są publikowane dla wyższych poziomów agregacji niż w rzeczywistości zostały zebrane. Jeśli w przypadku takiej sytuacji, wykorzystanie danych przestrzennych na wyższym poziomie agregacji przyczyni się do uzyskania innych ocen parametrów, to należy mieć świadomość wpływu tego faktu na wyniki prowadzonych badań. W sytuacji dostępu do danych przestrzennych na różnych poziomach agregacji, badacz powinien znać potencjalne kierunki zmian w otrzymywanych rezultatach.

Osiągnięty cel praktyczny pracy stanowi wypracowanie podejścia metodycznego pozwalającego na rozwiązanie problemu sposobu podziału przestrzeni na drodze przykładu empirycznego oraz analizę problemu skali w ramach pro-

cedur symulacyjnych, które przedstawione zostały w trzech kolejnych rozdziałach. Autor wskazał, że rozwiązanie problemu sposobu podziału przestrzeni polega na przyjęciu właściwego układu jednostek terytorialnych, który może zostać przyporządkowany do jednorodnego zbioru układów. Rozwiązanie tego problemu przedstawione zostało w rozdziale czwartym na przykładzie przestrzennych badań ekonomicznych dotyczących rozwoju rolnictwa w Polsce na poziomie makroregionów. Wyznaczony problem badawczy polegał na ustaleniu przebiegu granic nowego układu jednostek terytorialnych na poziomie makroregionów, gdzie odniesione do niego dane przestrzenne poprawnie odzwierciedlały oddziaływanie przyczyn wpływających na zachodzące zmiany w rolnictwie. Granice nowego układu czterech makroregionów rolniczych wyznaczone zostały w wyniku przeprowadzenia analizy przestrzennego zróżnicowania struktury agrarnej, co stanowiło rozwiązanie podjętego problemu sposobu podziału przestrzeni.

W rozdziale czwartym rozważony został także w ujęciu przestrzennym, podjęty przez Pawłowskiego w odniesieniu do szeregów czasowych problem ustalenia relacji między makroparametrem a mikroparametrami. Przeprowadzone zostało badanie ekonomiczne w układzie makroregionów rolniczych, w wyniku którego dokonano identyfikacji mikrozależności przestrzennych dotyczących wpływu wielkości gospodarstwa rolnego na poziom zużycia nawozów azotowych. Następnie wyznaczona została ocena makroparametru na podstawie kombinacji liniowej ocen mikroparametrów, co pozwoliło na ustalenie relacji między mikroparametrami dla makroregionów rolniczych a makroparametrem dla obszaru Polski. Wykonane w pracy badanie poszerzyło postawiony przez Pawłowskiego problem o aspekty przestrzenne, gdzie do ustalenia mikrozależności dla całego obszaru makroregionów wykorzystano autorskie miary średniego oddziaływania przestrzennego.

W ostatnich dwóch rozdziałach przeprowadzono analizę problemu skali, która oparta została na zastosowaniu procedur symulacyjnych. W ramach redefinicji problemu skali przedstawione zostało autorskie podejście do analizy symulacyjnej w konfrontacji z podejściem reprezentowanym w literaturze przedmiotu. W procesie generowania danych przestrzennych oraz ich agregacji wykorzystano wyłącznie układy jednostek terytorialnych, które przyporządkowane zostały do jednorodnego zbioru układów. W wyniku tak określonej procedury symulacyjnej otrzymano rozkłady ocen badanych parametrów na każdym z przyjętych poziomów agregacji, co pozwoliło na ustalenie, w jakim stopniu problem skali wpływał na konkretne wyniki badań.

W związku z przedstawionymi dwoma sposobami zapewnienia porównywalności obszarów w ramach wykonywanych badań, przeprowadzona w rozdziale piątym analiza symulacyjna oparta została na generowanych danych przestrzennych, które wyrażone zostały w wartościach średnich. Natomiast w rozdziale szóstym w analizie symulacyjnej generowane były dane przestrzenne ważone wielkością obszaru. W obydwu przypadkach sprawdzono wpływ agregacji danych przestrzennych na otrzymywane wyniki badań, gdzie

w trzech scenariuszach założone zostały wybrane własności procesów przestrzennych. W scenariuszu pierwszym generowano realizacje procesów szumów przestrzennych o stałym poziomie wartości oczekiwanej i wariancji oraz braku autokorelacji przestrzennej. W scenariuszu drugim założono dla analizowanych procesów występowanie przestrzennego trendu stopnia pierwszego, a w scenariuszu trzecim założono własność autokorelacji przestrzennej.

W przypadku scenariusza trzeciego podkreślona została konieczność wzięcia pod uwagę struktury przestrzennej analizowanego obszaru, którą autor określa, jako ogół istniejących zależności przestrzennych na ustalonym obszarze, w tym tzw. autozależności oraz zależności pomiędzy procesami. W związku z tym, procedura generowania kolejnych realizacji procesów, które byłyby odzwierciedleniem tej samej struktury przestrzennej poszerzona została o wykorzystanie empirycznego rozkładu procesów.

W przypadku scenariusza pierwszego i drugiego, zarówno na podstawie generowanych danych wyrażonych w wartościach średnich, jak i ważonych danych przestrzennych nie stwierdzono istotnych różnic w ocenach własności procesów przestrzennych oraz łączących je zależnościach przyczynowych, które byłyby wynikiem procesu agregacji. Oznacza to, że zastosowanie analizy regresji dla procesów szumów przestrzennych lub procesów charakteryzujących się płaszczyznowym trendem przestrzennym powinna gwarantować uzyskanie podobnych rezultatów, bez względu na wybrany poziom agregacji danych przestrzennych.

Natomiast występowanie problemu skali stwierdzone zostało w scenariuszu trzecim, gdzie analizowane procesy charakteryzowały się dodatnią autokorelacją przestrzenną. Otrzymane wyniki wskazały, że w przypadku identyfikacji własności autokorelacji przestrzennej zachodzą istotne zmiany w otrzymywanych rezultatach badań na skutek procesu agregacji danych. Interpretacja problemu skali nie jest łatwa, ponieważ ustalono, że w zależności od wartości parametru autoregresji zmiany ocen badanych parametrów na skutek procesu agregacji przebiegają w różny sposób. Wobec tego autor wskazuje na fakt, zgodnie z którym za charakter zmian w otrzymywanych ocenach parametrów odpowiada rzeczywista struktura przestrzenna. Dlatego tak istotnym jest prawidłowe odzwierciedlenie przez generowane dane rzeczywistej struktury przestrzennej, wyrażającej zależności między badanymi procesami.

Motywacją do podjęcia problematyki w przedstawionej pracy była występująca w literaturze przedmiotu luka metodyczna i implementacyjna w zakresie problemu zmiennej jednostki odniesienia, zarówno w aspekcie problemu podziału przestrzeni, jak i problemu skali. Należy podkreślić, że zaproponowane w pracy nowe definicje problemu sposobu podziału przestrzeni oraz problemu skali, a także przedstawione przykłady empiryczne oraz analizy symulacyjne poszerzyły stan badań dotyczących problemu zmiennej jednostki odniesienia. Zaproponowana metodyka badań powinna zostać zastosowana w kolejnych badaniach przestrzennych w celu potwierdzenia jej przydatności na polu statystyki i ekonometrii przestrzennej.

Literatura

- Adler, R.J. (1983). *The geometry of random fields*. New York: John Wiley.
- Amrhein, C.G. (1995). Searching for the elusive aggregation effect: evidence from statistical simulations. *Environment and Planning A*, 27, s. 105–119.
- Anselin, L. (1988). *Spatial econometrics: method and models*. Boston: Kluwer Academic Press.
- Anselin, L. (1990). Spatial dependence and spatial structural instability in applied regression analysis. *Journal of Regional Science*, 30, s. 185–207.
- Anselin, L. (2002). Under the hood. Issues in the specification and interpretation of spatial regression models. *Agricultural Economics*, 27, s. 247–267.
- Anselin, L. (2003). Spatial externalities, spatial multipliers, and spatial econometrics. *International Regional Science Review*, 26, s. 153–166.
- Anselin, L., Florax, R.J.G.M., i Rey, S.J. (red.) (2004). *Advances in spatial econometrics. Methodology, tools and applications*. Berlin: Springer.
- Anselin, L., i Florax, R.J.G.M. (red.) (1995). *New direction in spatial econometrics*. Berlin: Springer.
- Anselin, L., i Griffith, D. (1988). Do spatial effects really matter in regression analysis? *Papers of the Regional Science Association*, 65, s. 11–34.
- Anselin, L., i Rey, S.J. (1991). Properties of tests for spatial dependence in linear regression models. *Geographical Analysis*, 23, s. 112–131.
- Anselin, L., i Rey, S.J. (1997). Introduction to the special issue on spatial econometrics. *International Regional Science Review*, 20, s. 1–7.
- Antczak, E. (2013). Przestrzenny taksonomiczny miernik rozwoju. *Wiadomości Statystyczne*, 7, s. 37–53.
- Arbia, G. (1988). *Spatial data configuration in statistical analysis of regional economics and related problems*. Dordrecht: Kluwer Academic Press.

- Arbia, G. (1990). On second order non-stationarity in two-dimensional lattice processes. *Computational Statistics and Data Analysis*, 9, s. 147–60.
- Arbia, G. (2001). The role of spatial effects in the empirical analysis of regional concentration. *Journal of Geographical Systems*, 3, s. 271–281.
- Arbia, G. (2006). *Spatial econometrics, statistical foundations and applications to regional convergence*. Heidelberg: Springer.
- Arbia, G., Bee, M., i Espa, G. (2013). Testing isotropy in spatial econometric models. *Spatial Economic Analysis*, 8(3), s. 228–40.
- Arbia, G., i Lafratta, G. (2005). Exploring non-linear dependence in the tails of bivariate marginals of a random field. *Geographical Analysis*, 37(4), s. 423–437.
- Bal-Domańska, B., i Pietrzak, M.B. (2014). Modelowanie wzrostu gospodarczego na podstawie rozszerzonego modelu Solowa-Swana z uwzględnieniem aspektu przestrzennego. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu. Problemy rozwoju regionalnego i lokalnego*, 331, s. 11–18.
- Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society B*, 36, s. 192–236.
- Bivand, R. (1980). Autokorelacja przestrzenna a metody analizy statystycznej w geografii. W: Z. Chojnicki (red.), *Analiza regresji w geografii*. Poznań: PWN, s. 23–38.
- Bivand, R. (1981). *Modelowanie geograficznych układów czasoprzestrzennych*. Warszawa–Poznań: PWN.
- Bivand, R. (1984). Regression modeling with spatial dependence: an application of some class selection and estimation methods. *Geographical Analysis*, 16(1), s. 25–37.
- Bivand, R., Pebesma, E.J., i Gómez-Rubio, V. (2008). *Applied spatial data analysis with R*. New York: Springer.
- Blalock, H. (1964). *Causal inferences in nonexperimental research*. Chapel Hill: University of North Carolina Press.
- Bołt, T.W., Krauze, K., i Kulawczuk, T. (1985). *Agregacja modeli ekonometrycznych*. Warszawa: PWE.
- Box, G.E.P., i Jenkins, G.M. (1983). *Analiza szeregów czasowych*. Warszawa: PWN.
- Bukraba-Rylska, I. (2008). *Socjologia wsi polskiej*. Warszawa: PWN.
- Ceriani, L., i Verme, P. (2012). The origins of the Gini index: extracts from *Variabilità e Mutabilità* (1912) by Corrado Gini. *Journal of Economic Inequality*, 10(3), s. 421–443.
- Chojnicki, Z. (1966). *Zastosowanie modeli grawitacji i potencjału w badaniach przestrzenno-ekonomicznych*. Warszawa: PWN.
- Chojnicki, Z., i Czyż, T. (1973). *Metody taksonomii numerycznej w regionalizacji geograficznej*. Warszawa: PWN.

- Christakos, G. (1992). *Random field modelling in the earth sciences*. San Diego: Academic Press.
- Christakos, G. (2000). *Modern spatiotemporal geostatistics*. Oxford: Oxford University Press.
- Clif, A., i Ord, J.K. (1973). *Spatial autocorrelation*. London: Pion.
- Clif, A., i Ord, J.K. (1981). *Spatial processes, models and applications*. London: Pion.
- Conley, T.G., i Topa, G. (2002). Socio-economic distance and spatial pattern of unemployment. *Journal of Applied Econometrics*, 17, s. 303–327.
- Cressie, N.A.C. (1993). *Statistics for spatial data*. New York: John Wiley & Sons.
- Czyż, T. (1978). *Metody generalizacji układów przestrzennych*. Warszawa: PWN.
- Dark, S.J., i Bram, D. (2007). The modifiable areal unit problem (MAUP) in physical geography. *Progress in Physical Geography*, 31(5), s. 471–479.
- Dudley, G. (1991). Scale, aggregation, and the modifiable area unit problem. *The Operational Geographer*, 9(3), s. 28–33.
- Elhorst, J.P. (2001). Dynamic models in space and time. *Geographical Analysis*, 33, s. 119–140.
- Eurostat. (2015). *Eurostat nomenclature of territorial units for statistics: NUTS guidelines 2015*. Pobrane 30.02.2018 z <https://ec.europa.eu/eurostat/web/ess/-/eurostat-nomenclature-of-territorial-units-for-statistics-nuts-guidelines-2015>.
- Florax, R., i Folmer, H. (1992). Specification and estimation of spatial linear regression models. *Regional Science and Urban Economics*, 22, s. 405–432.
- Flowerdew, R. (2011). How serious is the modifiable areal unit problem for analysis of english census data? *Population Trends*, 145, s. 1–13.
- Fotheringharn, A.S., i Wong, D.W.S. (1991). The modifiable areal unit problem in multivariate analysis. *Environment and Planning A*, 23, s. 1025–1044.
- Gajowski, J., i Przybycin, Z. (1999). *Statystyka przestrzenna. Wybrane zagadnienia*. Katowice: Wydawnictwo AE w Katowicach.
- Geary, R. (1954). The contiguity ratio and statistical mapping. *The Incorporated Statistician*, 5, s. 115–145.
- Gehlke, C.E., i Biehl, K. (1934). Certain effects of grouping upon the size of the correlation coefficient in census tract material. *Journal of the American Statistical Association*, 29, s. 169–170.
- Getis, A., Mur, J., i Zoller, H. (2004). *Spatial econometrics and spatial statistics*. Basingstoke: Palgrave Macmillan.
- Getis, A., i Ord, J.K. (1992). The analysis of spatial association by use of distance statistics. *Geographical Analysis*, 24(3), s. 189–206.

- Goraj, L., Cholewa, I., Osuch, D., i Płonka, R. (2010). *Analiza skutków zmian we Wspólnotowej Typologii Gospodarstw Rolnych*. Warszawa: IERiGŻ-PIB.
- Granger, C.W.J. (1974). Spatial data and time series analysis. W: A. Scott (red.), *Studies in Regional Science*. London: Pion.
- Granger, C.W.J. (1975). Aspects of the analysis and interpretation of temporal and spatial data. *The Statistician*, 24, s. 197–210.
- Griffith, D.A. (1988). *Advanced spatial statistics*. Dordrecht: Kluwer Academic Press.
- Griffith, D.A. (2000). Eigenfunction properties and approximations of selected incidence matrices employed in spatial analysis. *Linear Algebra and its Applications*, 321, s. 95–112.
- Griffith, D.A., i Haining, R.P. (red.), (1986). *Transformations through space and time*. Leiden: Martin Nijhoff Publishers.
- Haining, R.P. (1990). *Spatial data analysis in the social and environmental sciences*. Cambridge: University Press, Cambridge.
- Haining, R.P. (1994). Diagnostics for regression model in spatial econometrics. *Journal of Regional Science*, 34, s. 325–341.
- Haining, R.P. (2005). *Spatial data analysis. Theory and practice*. Cambridge: Cambridge University Press.
- Haining, R.P. (1979). Statistical tests and process generators for random field models. *Geographer Analysis*, 11, s. 45–46.
- Hellwig, Z. (1972). *Elementy rachunku prawdopodobieństwa i statystyki matematycznej*. Warszawa: PWN.
- Hordijk, L. (1979). Problems in estimating econometric relations in space. *Papers in Regional Science Association*, 42, s. 99–115.
- Jankowski, R., i Walukiewicz, H. (1997). Modeling of two-dimensional random fields. *Probabilistic Engineering Mechanics*, 12(2), s. 115–121.
- Jenks, G.F. (1967). The data model concept in statistical mapping. *International Yearbook of Cartography*, 7, s. 186–190.
- Kelejian, H. i Robinson, D. (1995). Spatial correlation: a suggested alternative to the autoregressive model. W: L. Anselin, i R. Florax (red.). *New direction in spatial econometrics*. Berlin: Springer.
- Klaassen, J.H.P., Paelinck, L.H., i Wagenaar, s. (1982). *Systemy przestrzenne*. Warszawa: PWN.
- Klein, L.R. (1946). Macroeconomics and the theory of rational behavior. *Econometrica*, 14(2), s. 93–108.
- Kopczewska, K. (2006). *Ekonometria i statystyka przestrzenna: z wykorzystaniem programu R CRAN*. Warszawa: CeDeWu.
- Lange, O. (1961). *Wstęp do ekonometrii*. Warszawa: PWN.
- Lange, O. (1978). *Ekonomia polityczna*. Tom 1 i 2. Warszawa: PWN.

- Lange, O., i Banasiński, A. (1968). *Teoria statystyki*. Warszawa: PWE.
- Łaszkiwicz, E. (2016). *Ekonometria przestrzenna III. Modele wielopoziomowe: teoria i zastosowanie*. Warszawa: C.H. Beck.
- Leonenko, N.N., i Ivanov A.V. (1989). *Statistical analysis of random fields*. Dordrecht: Kluwer Academic Press.
- LeSage, J.P. (1997). Regression analysis of spatial data. *Journal of Regional Analysis and Policy*, 27, s. 83–94.
- LeSage, J.P. i Pace, R.K. (2007). A matrix exponential spatial specification. *Journal of Econometrics*, 140(1), s. 190–214.
- LeSage, J.P., i Pace, R.K. (2009). *Introduction to spatial econometrics*. Boca Raton: Wiley.
- LeSage, J.P., i Pace, R.K. (red.). (2004). *Advances in econometrics: spatial and spatiotemporal econometrics*. Amsterdam: Elsevier.
- Manley, D., Flowerdew, R.T.N., i Steel, D. (2006). Scales, levels and processes: studying spatial patterns of British census variables. *Computers, Environment and Urban Systems*, 30(2), s. 143–160.
- May, K. (1946). The aggregation problem for a one-industry model. *Econometrica*, 14(4), s. 285–298.
- Michna, W. (2007). *Sterowane i samoistne przemiany struktury agrarnej w różnych regionach kraju*. Warszawa: IERiGŻ–PIB.
- Moran, P.A.P. (1948). The interpretation of statistical maps. *Journal of the Royal Statistical Society, Series B*, 10, s. 243–251.
- Moran, P.A.P. (1950). Notes on continuous stochastic phenomena. *Biometrika*, 37, s. 17–23.
- Moran, P.A.P. (1973). A Gaussian Markovian process on a square lattice. *Applied Probability*, 10, s. 54–62.
- Müller-Frączek, I., i Pietrzak, M.B. (2011a). Space-time modelling of the unemployment rate in Polish poviats. *Dynamic Econometric Models*, 11, s. 203–213.
- Müller-Frączek, I., i Pietrzak, M.B. (2011b). Analiza stopy bezrobocia w Polsce z wykorzystaniem przestrzennego modelu MESS. *Acta Universitatis Lodziensis. Folia Oeconomica*, 253, s. 215–223.
- Müller-Frączek, I., i Pietrzak, M.B. (2012). Analiza stopy bezrobocia w Polsce w ujęciu przestrzenno-czasowym. *Oeconomia Copernicana*, 2, s. 43–55.
- Müller-Frączek, I., i Pietrzak, M.B. (2013). Zastosowanie modelu MESS w przestrzenno-czasowej analizie stopy bezrobocia w Polsce. *Acta Universitatis Lodziensis. Folia Oeconomica*, 293, s. 101–110.
- Oleński, J. (2001). *Ekonomika informacji*. PWE.
- Openshaw, S. (1977a). A geographical solution to scale and aggregation problems in region-building, partitioning and spatial modelling. *Transactions of the Institute of British Geographers. New Series*, 2(2), s. 459–472.

- Openshaw, s. (1977b). Algorithm 3: a procedure to generate pseudo-random aggregations of N zones into M zones, where M is less than N'. *Environment and Planning A*, 9, s. 1423–1428.
- Openshaw, s. (1977c). Optimal zoning systems for spatial interaction models. *Environment and Planning A*, 9, s. 169–184.
- Openshaw, s. (1978a). An optimal zoning approach to the study of spatially aggregated data. W: I. Masser, P. J. B. Brown (red.), *Spatial representation and spatial interaction*. Leiden: Martinus Nijhoff Social Sciences Division, s. 95–113.
- Openshaw, s. (1978b). An empirical study of some zone-design criteria. *Environment and Planning A*, 10(7), s. 781–794.
- Openshaw, s. (1984a). Ecological fallacies and the analysis of areal census data. *Environment and Planning A*, 16, s. 17–31.
- Openshaw, s. (1984b). The modifiable areal unit problem. *Concepts and Techniques in Modern Geography*, 38, s. 1–40.
- Openshaw, S., i Taylor, P.J. (1979). A million or so correlation coefficients: three experiments on the modifiable areal unit problem. W: N. Wrigley (red.), *Statistical methods in the spatial sciences*. London: Pion, s. 127–144.
- Ord, J.K. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association*, 70, s. 120–126.
- Ord, J.K. (1975). Estimation methods for spatial interaction. *Journal of the American Statistical Association*, 70, s. 120–126.
- Paelinck, J.H.P. (2000). On aggregation in spatial econometric modelling. *Journal of Geographical Systems*, 2(2), s. 157–165.
- Paelinck, J.H.P. (2001). A multiple gap approach to spatial econometrics. *The Annals of Regional Science*, 36(2), s. 219–228.
- Paelinck, J.H.P. i Klaassen, L.H. (1979). *Spatial econometrics*. Farnborough: Saxon House.
- Paelinck, J.H.P. i Nijkamp, P. (1975). *Operational theory and methods in regional analysis*. Farnborough: Saxon House.
- Paelinck, J.H.P., i Klaassen, L.H. (1983). *Ekonometria przestrzenna*. Warszawa: PWN.
- Pawłowski, Z. (1969). *Ekonometria*. Warszawa: PWN.
- Pietrzak, M. B. (2014c). The modifiable areal unit problem: analysis of correlation and regression. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 9(4), s. 113–131.
- Pietrzak, M.B. (2010a). Problem identyfikacji struktury danych przestrzennych. *Acta Universitatis Nicolai Copernici. Oeconomia*, 41, s. 83–98.
- Pietrzak, M.B. (2010b). Analiza danych przestrzennych a jakość informacji. W: T. Trzaskalik (red.), *Modelowanie Preferencji a Ryzyko '09*, Katowice: Wydawnictwo AE w Katowicach, s. 323–338.

- Pietrzak, M.B. (2010c). Dwuetapowa procedura budowy przestrzennej macierzy wag z uwzględnieniem odległości ekonomicznej. *Oeconomia Copernicana*, 1, s. 65–78.
- Pietrzak, M.B. (2010d). Wykorzystanie odległości ekonomicznej w przestrzennej analizie stopy bezrobocia dla Polski. *Oeconomia Copernicana*, 1, s. 79–98.
- Pietrzak, M.B. (2011). Ocena siły oddziaływania procesów objaśniających dla modeli przestrzennych. *Prace i Materiały Wydziału Zarządzania Uniwersytetu Gdańskiego, Modelowanie i prognozowanie gospodarki narodowej*, 4/8, s. 458–498.
- Pietrzak, M.B. (2011). Wykorzystanie przestrzennego modelu regresji przelącznikowej w analizie stopy bezrobocia dla Polski. *Prace i Materiały Wydziału Zarządzania Uniwersytetu Gdańskiego, Modelowanie i prognozowanie gospodarki narodowej*, 4/8, s. 453–466.
- Pietrzak, M.B. (2012). Interpretation of the structural parameters of the SDM model based on the example of the analysis of average gross wages and salaries in Poland. W: J. Pocięcha (red.), *Methods and models for analysing and forecasting economic processes*. Cracow: University of Economics Press, s. 44–55.
- Pietrzak, M.B. (2012). Wykorzystanie odległości ekonomicznej w przestrzennych analizach procesów ekonomicznych. W: B. Pawełek (red.), *Modelowanie i prognozowanie zjawisk społeczno-gospodarczych. Aktualny stan i perspektywy rozwoju*. Kraków: Wydawnictwo UE w Krakowie, s. 96–106.
- Pietrzak, M.B. (2012). Wykorzystanie przestrzennego modelu regresji przelącznikowej w analizie regionalnej konwergencji w Polsce. *Ekonomia i Prawo*, 11(4), s. 167–185.
- Pietrzak, M.B. (2013). Interpretation of structural parameters for models with spatial autoregression. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 8(2), s. 129–155.
- Pietrzak, M.B. (2014a). Redefining the modifiable areal unit problem within spatial econometrics: the case of the scale problem. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 9(2), s. 111–132.
- Pietrzak, M.B. (2014b). Redefining the modifiable areal unit problem within spatial econometrics, the case of the aggregation problem. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 9(3), s. 131–151.
- Pietrzak, M.B. (2014d). Taksonomiczny miernik rozwoju (TMR) z uwzględnieniem zależności przestrzennych. *Przegląd Statystyczny*, 61(2), s. 181–201.
- Pietrzak, M.B. (2016). The problem of the inclusion of spatial dependence within the TOPSIS Method. *Montenegrin Journal of Economics*, 12(3), s. 69–86.

- Pietrzak, M.B., i Balcerzak, A.P. (2016). A spatial SAR model in evaluating influence of entrepreneurship and investments on unemployment in Poland. W: M. Reiff, P. Gezik (red.), *Proceedings of the international scientific conference quantitative methods in economics multiple criteria decision making XVIII*. Vratna: Letra Interactive, s. 303–308.
- Pietrzak, M.B., i Rutkowska, E. (2011). Analiza czasu przejazdu do centrum miasta z wykorzystaniem modelu przestrzennego SAR. *Acta Universitatis Lodziensis. Folia Oeconomica*, 253, s. 269–282.
- Pietrzak, M.B., i Walczak, D. (2012). The evaluation of the agrarian structure in the Pomerania and Kujawy regions in the years 1921 and 2002. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 14(6), s. 211–215.
- Pietrzak, M.B., i Walczak, D. (2013). Changes in the agrarian structure in Poland in the years 1921–2002 based on the example of selected provinces from three annexed territories. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, 15(4), s. 333–336.
- Pietrzak, M.B., i Walczak, D. (2014). The analysis of the agrarian structure in Poland with the special consideration of the years 1921 and 2002. *Bulgarian Journal of Agricultural Science*, 20(5), s. 1018–1039.
- Pietrzak, M.B., i Wilk, J. (2014). Odległość ekonomiczna w modelowaniu zjawisk przestrzennych z wykorzystaniem modelu grawitacji. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu. Taksonomia 22: Klasyfikacja i analiza danych: teoria i zastosowania*, 327, s. 177–185.
- Pietrzak, M.B., i Ziemkiewicz, B. (2017). The use of random fields in the modifiable areal unit problem. W: M. Papież, s. Śmiech (red.), *The 11th professor Aleksander Zelias international conference on modelling and forecasting of socio-economic phenomena. Conference proceedings*. Cracow: Foundation of the Cracow University of Economics, s. 319–328.
- Pietrzak, M.B., Wilk, J., Bivand, R., i Kossowski, T. (2014). The application of Local Indicators for Categorical Data (LICD) in identifying of spatial dependences in the analysis of socio-economic development. *Comparative Economic Research. Central and Eastern Europe*, 17 (4), s. 203–220.
- Pietrzak, M.B., Wilk, J., Kossowski, T., i Bivand, R. (2014). the identification of spatial dependence in the analysis of regional economic development: join-count test application. W: M. Papież, s. Śmiech (red.), *Proceedings of the 8th professor Aleksander Zelias international conference on modelling and forecasting of socio-economic phenomena*. Cracow: Foundation of the Cracow University of Economics, s. 135–144.
- Pietrzak, M.B.. (2016). Zastosowanie przestrzennego taksonomicznego mierznika rozwoju (pTMR) w analizie rynku pracy w Polsce. *Studia Ekonomiczne: Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach*, 291, s. 47–58.

- Przybycin, Z. (1992). *Zastosowanie pól losowych w ekonomicznych modelach przestrzennych*. Katowice: Wydawnictwo AE w Katowicach.
- Przybycin, Z. (2004). *Metody i modele statystyki przestrzennej*. Katowice: Wydawnictwo AE w Katowicach.
- Ramm, A.G. (2005). *Random fields estimation*. New Jersey: World Scientific.
- Reiff, M., Surmanová, K., Balcerzak, A.P., i Pietrzak, M.B. (2016). Multiple criteria analysis of European Union agriculture. *Journal of International Studies*, 9(3), s. 62–74.
- Reynolds, H.D. (1998). *The modifiable area unit problem: empirical analysis by statistical simulation*. Doctoral thesis, graduate Department of Geography, University of Toronto.
- Ripley, B.D. (1988). *Statistical inference for spatial processes*. Cambridge: Cambridge University Press.
- Ripley, B.D. (2004). *Spatial statistics*. New York: John Wiley & Sons.
- Robinson, W.S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, 15(3), s. 351–357.
- Rozanow, J.A. (1974). *Wstęp do teorii procesów stochastycznych*. Warszawa: PWN.
- Rozporządzenie Rady Ministrów z dnia 19 grudnia 2017 r. w sprawie programu badań statystycznych statystyki publicznej na rok 2018 (Dz. U. poz. 2471).
- Schabenberger, O. i Gotway, C.A. (2002). *Statistical methods for spatial data analysis*. Boca Raton: Chapman & Hall/CRC.
- Schabenberger, O., i Gotway, C. A. (2005). *Statistical methods for spatial data analysis, texts in statistical science*. Boca Raton: Chapman & Hall/CRC.
- Sherman, M. (1996). Variance estimation for statistics computed from spatial lattice data. *Journal of the Royal Statistical Society. Series B*, 58, s. 509–523.
- Skarżyńska, A., Goraj, L., i Ziętek, I. (2005). *Metodologia SGM „2002” dla typologii gospodarstw rolnych w Polsce*. Warszawa: IERiGŻ–PIB.
- Suhecka, J. (red.), (2014). *Statystyka przestrzenna. Metody analizy struktur przestrzennych*. Warszawa: C.H. Beck.
- Suhecki, B. (red.), (2010). *Ekonometria przestrzenna, metody i modele analizy danych przestrzennych*. Warszawa: C.H. Beck.
- Suhecki, B. (red.), (2012). *Ekonometria przestrzenna II, modele zaawansowane*. Warszawa: C.H. Beck.
- Swieszniak, A.A. (1965). *Podstawowe metody funkcji losowych*. Warszawa: PWN.
- Szczepankiewicz, E. (1985). *Zastosowania pól losowych*. Warszawa: PWN.
- Szulc, E. (2001). Ekonometryczna analiza przestrzenno-czasowego procesu bezrobocia. W: A. Zeliaś (red.), *Przestrzenno-czasowe modelowanie i prognozowanie zjawisk gospodarczych*. Kraków: Wydawnictwo UE w Krakowie, s. 179–190.

- Szulc, E. (2003a). Identyfikacja odstępów czasowych realizacji zależności w przestrzenno-czasowych modelach ekonometrycznych. W: A. Zeliaś (red.), *Przestrzenno-czasowe modelowanie i prognozowanie zjawisk gospodarczych*. Kraków: Wydawnictwo UE w Krakowie, s. 357–366.
- Szulc, E. (2003b). Odległość ekonomiczna w badaniu struktury zależności w procesach gospodarczych. W: Z. Zieliński (red.), *Dynamiczne modele ekonometryczne*. Toruń: Wydawnictwo UMK w Toruniu, s. 311–318.
- Szulc, E. (2007a). Modelowanie zależności między przestrzenno-czasowymi procesami ekonomicznymi. W: Z. Zieliński (red.), *Dynamic econometric models*, Toruń: Wydawnictwo UMK w Toruniu, s. 161–170.
- Szulc, E. (2007b). *Ekonometryczna analiza wielowymiarowych procesów gospodarczych*. Toruń: Wydawnictwo UMK w Toruniu.
- Tate, N., i Atkinson, P.M. (red.), (2001). *Modelling scale in geographical information science*. Chichester: John Wiley & Sons.
- Tate, N.P. Atkinson, P.M. (red.), *Modelling scale in geographical information science*. Chichester: John Wiley & Sons, s. 105–122.
- Taylor, P.J. (1977). *Quantitative methods in geography: an introduction to spatial analysis*. Boston: Houghton Mifflin.
- Taylor, P.J., i Johnston, R.J. (1979). *Geography of elections*. New York: Holmes & Meier Publishers.
- Theil, H. (1965). *Linear aggregation of economic relations*. Amsterdam: North-Holland Publishing Company.
- Tinbergen, J. (1939). *Business cycles in the United States of America 1919–1932*. Genewa: League of Nations.
- Tjostheim, D. (1978). Statistical spatial series modelling. *Advances in Applied Probability*, 10, s. 130–154.
- Tobler, W.R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46, s. 234–240.
- Tobler, W.R. (1979). Lattice tuning. *Geographical Analysis*, 11, s. 36–44.
- Tobler, W.R. (1989). Frame independent spatial analysis. W: M.F. Goodchild, i s. Gopal (red.), *The accuracy of spatial databases*. London: Taylor & Francis, s. 115–122.
- Tranmer, M., i Steel, D. (2001). Using local census data to investigate scale effects. W: N.J. Tate i P.M. Atkinson (red.), *Modelling scale in geographical information science*. Chichester: John Wiley and Sons, s. 105–122.
- Ustawa z dnia 29 czerwca 1995 r. o statystyce publicznej (Dz. U. z 2016 r. poz. 1068, z późn. zm.).
- Ustawa z dnia 29 listopada 2000 r. o zbieraniu i wykorzystywaniu danych rachunkowych z gospodarstw rolnych (Dz.U. 2001 nr 3 poz. 20).

- Walczak, D., i Pietrzak, M.B. (2016). Analysis of agrarian structure in Poland in 1921 and 2002 based on the example of selected districts. W: M.H. Bilgin, H. Danis, E. Demir, i U. Can (red.), *Business challenges in the changing economic landscape. Proceedings of the 14th Eurasia Business and Economics Society conference*. Cham: Springer, s. 461–472.
- Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, 41, s. 434–449.
- Whittle, P. (1962). Stochastic processes in several dimensions. *Bulletin of the International Statistical Institute*, 49, s. 974–993.
- Wilk, J., Pietrzak, M.B., Bivand, R., i Kossowski, T. (2015). Wpływ wyboru metody klasyfikacji na identyfikację zależności przestrzennych — zastosowanie testu join-count. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu. Taksonomia 24: Klasyfikacja i analiza danych: teoria i zastosowania*, 384, s. 296–304.
- Wilson, A.G. i Bennet, R.J. (1985). *Mathematical methods in human geography and planning*. New York: John Wiley.
- Wiśniewski J. W., i Zieliński, Z. (1998). *Elementy ekonometrii*. Toruń: Wydawnictwo UMK w Toruniu.
- Woś, A. (red.). (1998). *Encyklopedia agrobiznesu*. Warszawa: Fundacja Innowacja.
- Wrigley, N. (1979). *Statistical methods in the spatial sciences*. Pion: London
- Yule, G.U., i Kendall, M.G. (1950). *An introduction to the theory of statistics*. London: Griffin.
- Yule, G.U., i Kendall, M.G. (1966). *Wstęp do teorii statystyki*. Warszawa: PWE.
- Zadora, K. (1974). Modelowanie i predykcja pól losowych. *Przegląd Statystyczny*, 3, s. 435–449.
- Zeliaś, A. (red.), (1991). *Ekonometria przestrzenna*. Warszawa: PWE.
- Zieliński, Z. (1990). *Ekonometryczne modele pól losowych: postawienie problemu, podstawowe pojęcia i określenia, wytyczne kierunków badań*. Opracowanie w ramach CPBP, Toruń.
- Zieliński, Z. (1991). *Liniowe modele ekonometryczne jako narzędzie opisu i analizy przyczynowych zależności zjawisk ekonomicznych*. Toruń: Wydawnictwo Adam Marszałek.
- Zieliński, Z. (2002). *Analiza ekonomicznych procesów stochastycznych. Pisma wybrane*. Toruń: Wydawnictwo UMK.
- Zieliński, Z., i Szulc, E. (1991). Ekonomiczne pola losowe jako przedmiot analizy ekonometrycznej. W: J.W. Wiśniewski (red.), *Dynamiczne modele ekonometryczne*, Toruń: Wydawnictwo UMK w Toruniu, s. 31–38.

Spis tabel

Tabela 2.1. Etapy przestrzennych badań ekonomicznych	38
Tabela 2.2. Etapy agregacji danych przestrzennych	40
Tabela 2.3. Etapy przestrzennych badań ekonomicznych: rozwinięcie	42
Tabela 2.4. Wyniki testu Morana dla lat 2004–2007	56
Tabela 3.1. Stopień wykonania kolejnych etapów przestrzennych badań ekonomicznych w pracy Gehlke i Biehl (1934)	87
Tabela 3.2. Etapy badań przestrzennych zgodnie z nowym paradygmatem nauk empirycznych	100
Tabela 4.1. Wyniki testu Morana dla zużycia nawozów mineralnych oraz średniej powierzchni gospodarstw rolnych	126
Tabela 4.2. Wyniki estymacji parametrów modeli końcowych dla makroregionów rolniczych	128
Tabela 4.3. Wyniki estymacji parametrów modelu końcowego dla obszaru Polski	129
Tabela 4.4. Ocena mikrozależności dla makroregionów rolniczych	137
Tabela 4.5. Ocena makrozależności dla obszaru Polski	139
Tabela 5.1. Parametry generowanych szumów przestrzennych	147
Tabela 5.2. Analiza problemu skali dla szumów przestrzennych: wariant pierwszy	148
Tabela 5.3. Analiza problemu skali dla szumów przestrzennych: wariant drugi	149
Tabela 5.4. Analiza problemu skali dla szumów przestrzennych: wariant trzeci	150
Tabela 5.5. Analiza problemu skali dla szumów przestrzennych: wariant czwarty	151
Tabela 5.6. Analiza problemu skali dla szumów przestrzennych: wariant piąty	152
Tabela 5.7. Procentowy wzrost wartości odchyłeń standardowych procesów X_1 oraz X_2 (%)	153
Tabela 5.8. Własności procesów przestrzennych	156
Tabela 5.9. Analiza problemu skali dla procesów X_1 i X_2 : wariant pierwszy	157
Tabela 5.10. Analiza problemu skali dla procesów X_1 i X_2 : wariant drugi	158
Tabela 5.11. Analiza problemu skali dla procesów X_1 i X_2 : wariant trzeci	159

Tabela 5.12. Analiza problemu skali dla procesów X_1 i X_2 : wariant czwarty	160
Tabela 5.13. Analiza problemu skali dla procesów X_1 i X_2 : wariant piąty	161
Tabela 5.14. Analiza rozkładu zużycia nawozów azotowych	166
Tabela 5.15. Analiza rozkładu powierzchni gospodarstw rolnych	166
Tabela 5.16. Wartości współczynnika korelacji liniowej dla wybranych wariantów systemu wag	167
Tabela 5.17. Wyniki procedury symulacyjnej: scenariusz pierwszy, wariant pierwszy	169
Tabela 5.18. Wyniki procedury symulacyjnej: scenariusz pierwszy, wariant drugi	170
Tabela 5.19. Wyniki procedury symulacyjnej: scenariusz pierwszy, wariant trzeci	171
Tabela 5.20. Wyniki procedury symulacyjnej: scenariusz drugi, wariant pierwszy	173
Tabela 5.21. Wyniki procedury symulacyjnej: scenariusz drugi, wariant drugi	174
Tabela 5.22. Wyniki procedury symulacyjnej: scenariusz drugi, wariant trzeci	175
Tabela 5.23. Zmiany procentowe dla średnich ocen parametrów w wariancie trzecim scenariusza drugiego	176
Tabela 5.24. Wariant pierwszy: podwyższony poziom autokorelacji przestrzennej procesów X_1^* , X_2^*	178
Tabela 5.25. Wariant drugi: obniżony poziom autokorelacji przestrzennej procesów X_1^* , X_2^*	179
Tabela 5.26. Wariant trzeci: obniżony poziom autokorelacji przestrzennej procesu X_1^* oraz podwyższony poziom autokorelacji przestrzennej procesu X_2^*	180
Tabela 5.27. Wariant czwarty: podwyższony poziom autokorelacji przestrzennej procesu X_1^* oraz obniżony poziom autokorelacji przestrzennej procesu X_2^*	181
Tabela 6.1. Własności procesów szumów przestrzennych	185
Tabela 6.2. Założona struktura korelacyjna procesów $X_{1'}$, $X_{2'}$, $X_{1,2}$	186
Tabela 6.3. Analiza problemu skali dla szumów przestrzennych: wariant pierwszy	187
Tabela 6.4. Analiza korelacyjna dla procesów $X_{1'}$, $X_{2'}$, $X_{1,2}$: wariant pierwszy	188
Tabela 6.5. Analiza problemu skali dla szumów przestrzennych: wariant drugi	189
Tabela 6.6. Analiza korelacyjna dla procesów $X_{1'}$, $X_{2'}$, $X_{1,2}$: wariant drugi	190
Tabela 6.7. Analiza problemu skali dla szumów przestrzennych: wariant trzeci	191
Tabela 6.8. Analiza korelacyjna dla procesów $X_{1'}$, $X_{2'}$, $X_{1,2}$: wariant trzeci	192
Tabela 6.9. Własności procesów przestrzennych $X_{1'}$, $X_{2'}$, $X_{1,2'}$, $X_{2,2}$	195
Tabela 6.10. Założona struktura korelacyjna dla procesów $X_{1'}$, $X_{2'}$, $X_{1,2'}$, $X_{2,2}$	196
Tabela 6.11. Analiza własności procesów oraz zależności między nimi przed agregacją: wariant pierwszy	197
Tabela 6.12. Analiza własności procesów oraz zależności między nimi po wykonaniu agregacji danych: wariant pierwszy	198
Tabela 6.13. Analiza korelacyjna dla procesów $X_{1'}$, $X_{2'}$, $X_{1,2'}$, $X_{2,2}$: wariant pierwszy	199
Tabela 6.14. Analiza własności procesów oraz zależności między nimi przed agregacją: wariant drugi	200
Tabela 6.15. Analiza własności procesów oraz zależności między nimi po wykonaniu agregacji danych: wariant drugi	201

Tabela 6.16. Analiza korelacyjna dla procesów $X_{1,t}, X_{2,t}, X_{1,2,t}, X_{2,2,t}$: wariant drugi	202
Tabela 6.17. Analiza własności procesów oraz zależności między nimi przed agregacją: wariant trzeci	203
Tabela 6.18. Analiza własności procesów oraz zależności między nimi po wykonaniu agregacji danych: wariant trzeci	204
Tabela 6.19. Analiza korelacyjna dla procesów $X_{1,t}, X_{2,t}, X_{1,2,t}, X_{2,2,t}$: wariant trzeci	205
Tabela 6.20. Analiza rozkładu dochodów własnych gmin na 1 mieszkańca	207
Tabela 6.21. Analiza rozkładu liczby podmiotów gospodarczych przypadających na 10000 mieszkańców	207
Tabela 6.22. Analiza rozkładu ludności	208
Tabela 6.23. Analiza własności procesów oraz zależności między nimi przed agregacją danych	210
Tabela 6.24. Analiza własności procesów oraz zależności między nimi po wykonaniu agregacji danych	211
Tabela 6.25. Zmiany procentowe dla średnich ocen parametrów autoregresji modeli SAR	212
Tabela 6.26. Zmiany procentowe średnich ocen parametrów modelu regresji oraz miar oddziaływania przestrzennego	212
Tabela 6.27. Analiza korelacyjna dla generowanych danych przestrzennych przed i po agregacji	213

Spis rysunków

Rysunek 1.1. Hierarchia przestrzenna oparta na układach NUTS 3 oraz NUTS 2	33
Rysunek 2.1. Podział terytorium Polski na makroregiony rolnicze SGM	52
Rysunek 2.2. Porównanie układu SGM z układem NUTS 1	53
Rysunek 2.3. Przestrzenne zróżnicowanie stopy bezrobocia rejestrowanego	55
Rysunek 2.4. Zróżnicowanie przestrzenne struktury agrarnej w Polsce	65
Rysunek 2.5. Mapa granic zaborów w Polsce	67
Rysunek 2.6. Przykładowy podział Polski na trzy makroregiony	69
Rysunek 3.1. Układ jednostek terytorialnych nr 1	91
Rysunek 3.2. Układ jednostek terytorialnych nr 2	92
Rysunek 3.3. Układ jednostek terytorialnych wygenerowany w ramach <i>grouping system</i>	93
Rysunek 4.1. Przestrzenne zróżnicowanie struktury agrarnej w układzie jednostek terytorialnych NUTS 3	110
Rysunek 4.2. Przestrzenne zróżnicowanie struktury agrarnej w układzie jednostek terytorialnych NUTS 2	111
Rysunek 4.3. Przestrzenne zróżnicowanie struktury agrarnej w układzie jednostek terytorialnych NUTS 1	112
Rysunek 4.4. Koncentracja użytków rolnych w Polsce: podział na klasy oparty na wartościach kwartyli	114
Rysunek 4.5. Koncentracja użytków rolnych w Polsce: podział na klasy oparty na metodzie podziału naturalnego	115
Rysunek 4.6. Podział Polski na dwa makroregiony rolnicze: makroregion wschodni oraz makroregion zachodni	116
Rysunek 4.7. Wstępny układ makroregionów rolniczych na tle przestrzennego zróżnicowania struktury agrarnej	117
Rysunek 4.8. Układ makroregionów rolniczych M1, M2, M3, M4	120

Rysunek 4.9. Średnie zużycie nawozów azotowych oraz średnia powierzchnia gospodarstw rolnych w układzie makroregionów rolniczych SGM	124
Rysunek 4.10. Średnie zużycie nawozów azotowych oraz średnia powierzchnia gospodarstw rolnych w układzie województw	125
Rysunek 5.1. Układy jednostek terytorialnych wykorzystane w analizie symulacyjnej	145
Rysunek 5.2. Hipotetyczny jednorodny zbiór układów	146

„Recenzowana praca jest dość obszernym studium teoretyczno-empirycznym z wykorzystaniem eksperymentów symulacyjnych, poświęconym zagadnieniu zmiennej jednostki odniesienia w przestrzennych badaniach zjawisk i procesów ekonomicznych. Stanowi ona zwięźczone wieloletnich intensywnych, rzetelnych badań prowadzonych przez Autora. Wnosi ważny wkład do teorii i praktyki ekonometrycznej z zakresu ekonometrii przestrzennej, jednego z najszybciej rozwijających nurtów tej dyscypliny naukowej. (...) Wnosi twórczy wkład do metodyki badań przestrzennych, a opracowanie w nowy sposób zagadnienia zmiennej jednostki odniesienia stanowi wzbogacenie i poszerzenie obszaru wiedzy w zakresie ekonomii i ekonometrii przestrzennej”.

prof. zw. dr hab. JAN ZAWADZKI

„Przedłożona mi do recenzji książka stanowi nowatorskie i odkrywczym opracowanie jednego z podstawowych problemów statystyki i ekonometrii przestrzennej – tytułowego problemu zmiennej jednostki odniesienia, w literaturze zagranicznej znanego jako *Modifiable Areal Unit Problem* (MAUP), co można tłumaczyć także jako problem modyfikowalności jednostki przestrzennej. (...) Podjęta w pracy problematyka jest bardzo ważna nie tylko dla rozważań teoretycznych prowadzących do nowych rozwiązań metodycznych i metodologicznych, ale przede wszystkim dla badań rzeczywistych zjawisk ekonomicznych zlokalizowanych przestrzennie. Badania takich zjawisk prowadzi się zawsze na podstawie danych odnoszących się do określonych układów jednostek (regionów) przestrzennych. Konsekwencje wyboru określonego układu przestrzennego, o których m. in. traktuje recenzowana praca stanowią przedmiot szczególnego zainteresowania analityków gospodarczych zajmujących się badaniami regionalnymi. (...) Praca stanowi interesujące studium oparte na krytycznym przeglądzie literatury prezentującej dotychczasowe podejścia do będących przedmiotem zainteresowania problemów, rzeczowej dyskusji uzasadniającej potrzebę redefinicji tych problemów, przejrzystej ilustracji empirycznej oraz szeroko zakrojonych analizach symulacyjnych.

dr hab. ELŻBIETA SZULC, prof. UMK

ISBN 978-83-62049-32-5



9 788362 049325



POLSKIE
TOWARZYSTWO
EKONOMICZNE