



AKADEMIA GÓRNICZO-HUTNICZA
im. Stanisława Staszica w Krakowie



WYDZIAŁ ZARZĄDZANIA

STUDIA STACJONARNE

KIERUNEK: Informatyka i Ekonometria

PRACA DYPLOMOWA

licencjacka

Jakub Ignatik

Analiza przestrzenna przestępczości w Polsce

Spatial analysis of crime in Poland

Promotor: dr Jacek Wolak

Zatwierdzam do rejestracji i dopuszczam do obrony

.....
Data i podpis promotora

Kraków, 2019

Oświadczenie studenta

Upředzony(-a) o odpowiedzialności karnej na podstawie art. 115 ust. 1 i 2 ustawy z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych (t.j. Dz.U. z 2018 r. poz. 1191 z późn. zm.): „Kto przywłaszczu sobie autorstwo albo wprowadzu w błąd co do autorstwa całości lub części cudzego utworu albo artystycznego wykonania, podlega grzywnie, karze ograniczenia wolności albo pozbawienia wolności do lat 3. Tej samej karze podlega, kto rozpowszechnia bez podania nazwiska lub pseudonimu twórcy cudzy utwór w wersji oryginalnej albo w postaci opracowania, artystyczne wykonanie albo publicznie zniekształca taki utwór, artystyczne wykonanie, fonogram, wideogram lub nadanie.”, a także upředzony(-a) o odpowiedzialności dyscyplinarnej na podstawie art. 307 ust. 1 ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (Dz. U. z 2018 r. poz. 1668 z późn. zm.) „Student podlega odpowiedzialności dyscyplinarnej za naruszenie przepisów obowiązujących w uczelni oraz za czyn uchybiający godności studenta.”, oświadczam, że niniejszą pracę dyplomową wykonałem(-am) osobiście i samodzielnie i nie korzystałem(-am) ze źródeł innych niż wymienione w pracy.

Jednocześnie Uczelnia informuje, że zgodnie z art. 15a ww. ustawy o prawie autorskim i prawach pokrewnych Uczelni przysługuje pierwszeństwo w opublikowaniu pracy dyplomowej studenta. Jeżeli Uczelnia nie opublikowała pracy dyplomowej w terminie 6 miesięcy od dnia jej obrony, autor może ją opublikować, chyba że praca jest częścią utworu zbiorowego. Ponadto Uczelnia jako podmiot, o którym mowa w art. 7 ust. 1 pkt 1 ustawy z dnia 20 lipca 2018 r. – Prawo o szkolnictwie wyższym i nauce (Dz. U. z 2018 r. poz. 1668 z późn. zm.), może korzystać bez wynagrodzenia i bez konieczności uzyskania zgody autora z utworu stworzonego przez studenta w wyniku wykonywania obowiązków związanych z odbywaniem studiów, udostępniać utwór ministrowi właściwemu do spraw szkolnictwa wyższego i nauki oraz korzystać z utworów znajdujących się w prowadzonych przez niego bazach danych, w celu sprawdzania z wykorzystaniem systemu antyplagiatowego. Minister właściwy do spraw szkolnictwa wyższego i nauki może korzystać z prac dyplomowych znajdujących się w prowadzonych przez niego bazach danych w zakresie niezbędnym do zapewnienia prawidłowego utrzymania i rozwoju tych baz oraz współpracujących z nimi systemów informatycznych.

.....
(czytelny podpis studenta)

Spis treści

Streszczenie.....	7
Abstract	8
Wstęp	9
1. Charakterystyka przestępczości i przestępstw	10
1.1 Wstęp	10
1.2 Definicja przestępczości	10
1.3 Metody pomiaru przestępczości	11
Statystyki kryminalne.....	11
Badania sondażowe	12
1.4. Podział przestępstw	13
1.5. Analiza zjawiska przestępczości podejmowana w pracach badawczych	13
1.6. Podsumowanie	20
2. Wybrane metody i modele analizy danych przestrzennych.....	21
2.1. Wstęp	21
2.2. Charakterystyka danych przestrzennych	21
2.3. Eksploracyjna analiza danych przestrzennych	23
Macierze wag	24
Wariogram.....	26
Korelogram.....	28
Globalna autokorelacja przestrzenna	28
Lokalna autokorelacja przestrzenna	32
2.4. Modele przestrzenne	33
Charakterystyka i estymacja modeli przestrzennych	33
Strategia wyboru modelu	35
Weryfikacja statystyczna modelu.....	38

2.5. Podsumowanie.....	39
3. Analiza przestrzennego zróżnicowania przestępczości w Polsce	41
3.1 Wstęp	41
3.2 Analiza przestępczości w Polsce	41
Analiza graficzna.....	44
3.3 Analiza przestrzenna przestępczości	47
Wariogramy.....	47
Korelogramy.....	48
Testy globalne	50
Testy lokalne	56
3.4. Wybór zmiennych.....	57
3.5. Podsumowanie.....	62
4. Modelowanie przestrzenne przestępczości w Polsce.....	63
4.1. Wstęp.....	63
4.2. Dobór odpowiedniej postaci modelu	63
4.3. Model regresji liniowej dla przestępstw gospodarczych	67
Interpretacja modelu liniowego.....	70
4.4. Model SEM dla przestępstw kryminalnych.....	71
Interpretacja modelu SEM	74
4.5. Model SAR dla przestępstw drogowych	76
Interpretacja modelu SAR.....	79
4.6. Podsumowanie.....	81
Zakończenie	82
Bibliografia	85
Spis rysunków	89
Spis tabel.....	90

Streszczenie

Celem niniejszej pracy jest przedstawienie zjawiska przestępczości w Polsce z wykorzystaniem narzędzi analizy przestrzennej, w rozbiciu na trzy główne kategorie przestępczości: przestępczość kryminalną, gospodarczą i drogową. Wybór tego narzędzia zdaniem autora wydaje się być słuszny, gdyż na rozmiar przestępczości może mieć wpływ sytuacja w sąsiednich jednostkach administracyjnych. Podejście przestrzenne modelowania przestępczości zostało już zaprezentowane w literaturze, jednak niniejsza praca różni się pod względem zakresu i metodologii badań (podział przestępstw na trzy kategorie, dane z roku 2017), a także wyciągniętych wniosków. Praca została opracowana z wykorzystaniem programu R oraz arkusza kalkulacyjnego MS Excel.

W celu identyfikacji autokorelacji przestrzennej, zastosowano narzędzia eksploracyjnej analizy danych (ESDA). Narysowano wariogramy oraz korelogramy, które potwierdziły stacjonarność modeli, istotność opóźnień przestrzennych i pokazały, że liczbę przestępstw można tłumaczyć także lokalizacją. Zasadność podejścia przestrzennego pokazały też testy globalne Morana, Geary'ego C i join-count oraz test lokalny Morana, a także wykonane kartogramy.

Końcowym wynikiem pracy jest przedstawienie zweryfikowanych statystycznie (za pomocą testu Walda, LR, autokorelacji reszt LM, Breuscha-Pagana i Shapiro-Wilka) modeli: modelu liniowego dla przestępstw gospodarczych, modelu błędu przestrzennego (SEM) dla przestępstw kryminalnych oraz modelu opóźnienia przestrzennego (SAR) dla przestępstw drogowych. Wykorzystanie w badaniu modeli przestrzennych pozwoliło na pokazanie unikalnego charakteru każdej z kategorii przestępstw. Stosowane podejście potwierdziło też występowanie zależności przestrzennych między powiatami dla przestępstw kryminalnych oraz drogowych.

Słowa kluczowe

przestępczość, analiza przestrzenna, eksploracyjna analiza danych przestrzennych, modele przestrzenne

Abstract

The aim of this paper is to present the phenomenon of crime in Poland with the use of spatial analysis tools, divided into three main categories of crime: criminal, economic and road crime. In the author's opinion, the choice of this tool seems to be correct, as the size of crime may be affected by the situation in neighbouring administrative units. The spatial approach to crime modelling has already been presented in the literature, however, this paper differs in terms of the scope and methodology of research (division of crimes into three categories, data from year 2017) and conclusions drawn. The paper was prepared with the use of the R program and MS Excel spreadsheet.

In order to identify spatial autocorrelation, exploratory data analysis tools (ESDA) were used. Drawings of variograms and correlograms confirmed the stationary nature of the models, the significance of spatial delays and showed that the number of crimes can also be explained by localization. The validity of the spatial approach was also demonstrated by the Moran, Geary C and join-count global tests, the Moran local test and cartograms.

The final result of the study is the presentation of statistically verified (using Wald, LR, LM autocorrelation, Breusch-Pagan and Shapiro-Wilk tests) models: linear model for economic crime, spatial error model (SEM) for criminal crime and spatial lag model (SAR) for road crime. The use of spatial models in the study allowed to show the unique character of each category of crime. The applied approach also confirmed the existence of spatial relationships between districts for criminal and road traffic offences.

Keywords

crime, spatial analysis, exploratory spatial data analysis, spatial models

Wstęp

Hierarchia potrzeb Masłowa plasuje potrzebę bezpieczeństwa na drugim stopniu, tym samym umiejscawiając ją w gronie podstawowych wymogów, bez zapewnienia których, człowiek ma ograniczone możliwości rozwoju. Bezpieczeństwo, mierzone np. za pomocą liczby przestępstw, stanowi ciekawy element badań zarówno na polu socjologicznym, jak i ekonomicznym.

Celem niniejszej pracy jest przedstawienie zjawiska przestępczości w Polsce z wykorzystaniem narzędzi analizy przestrzennej. Dla każdej z trzech głównych kategorii przestępczości zbadano powiązania przestrzenne między powiatami oraz postanowiono zbadać, jakie czynniki mają na nie istotny wpływ. W tym celu zbudowano modele objaśniające poziom tych zjawisk w roku 2017 oraz poddano je weryfikacji statystycznej.

Metody zastosowane w niniejszej pracy zalicza się do analizy przestrzennej. Wybór tego narzędzia zdaniem autora wydaje się być słuszny, gdyż na rozmiar przestępczości może mieć wpływ sytuacja w sąsiednich jednostkach administracyjnych. W celu identyfikacji autokorelacji przestrzennej zastosowano narzędzia eksploracyjnej analizy danych, które potwierdziły obecność procesów przestrzennych. Ostatnią część pracy stanowi stworzenie odpowiednich modeli wyjaśniających przestępczość.

Pierwszy rozdział stanowi teoretyczny wstęp do zagadnienia przestępczości oraz przegląd literatury, który pozwoli na wyłonienie determinant przestępczości. W drugim rozdziale zaprezentowano metodologię badań oraz przedstawiono niezbędne w dalszej części pracy narzędzia eksploracyjnej analizy danych przestrzennych. Tematem trzeciego rozdziału jest wstępna analiza badanego zjawiska oraz opisanie zbioru zmiennych stosowanych w modelach. W ostatnim rozdziale dla każdej z kategorii przestępczości utworzony został model, który za pomocą zmiennych omówionych w rozdziale trzecim wyjaśnia poziom przestępczości w Polsce.

1. Charakterystyka przestępczości i przestępstw

1.1 Wstęp

Przestępczość jest zjawiskiem szczególnie niebezpiecznym. Stanowi ona ośrodek badań wielu dziedzin, gdyż jej skutki można odczuć w wielu aspektach życia - takich jak sfera prywatna (zachowanie, efektywność w nauce i pracy, aktywność w czasie wolnym) albo społeczna (relacje z drugim człowiekiem, poczucie bezpieczeństwa, otwartość na społeczne inicjatywy). Z punktu widzenia ekonomii najważniejszy jest jednak wymiar ekonomiczny. Obejmuje on koszty przestępstw (ukradzione mienie, naprawa zniszczeń), ich zapobiegania (funkcjonowanie policji, montaż systemów antywłamaniowych) oraz egzekwowania (wymiar sprawiedliwości, więziennictwo). Oprócz powyższych, na wymiar ekonomiczny może też mieć wpływ zachowanie jednostek, np. wspomniana wcześniej efektywność w pracy.

Niniejszy rozdział będzie poświęcony teoretycznemu wstępowi do zagadnienia przestępczości. Zostaną w nim przedstawione niezbędne w dalszej części pracy podstawowe definicje oraz podział przestępstw. Rozdział zakończy się przeglądem literatury, która wskaże na różnorodność możliwych analiz ilościowych w temacie przestępczości.

1.2 Definicja przestępczości

Źródłem, w którym można doszukiwać się definicji przestępczości, jest kodeks karny, gdyż to w nim podane są rodzaje przestępstw oraz odpowiadające im możliwe kary pozbawienia wolności. W powyższym dokumencie brak jednak wyprowadzonej wprost definicji, można za to wyprowadzić ją na podstawie wybranych artykułów. Według L. Gardeckiego, który na podstawie kodeksu karnego wyprowadził pojęcie przestępczości, przestępstwo to „*czyn (działanie lub zaniechanie) człowieka, zabroniony przez ustawę pod groźbą kary jako zbrodnia lub występki, zawiniony umyślnie lub nieumyślnie oraz społecznie szkodliwy w stopniu wyższym niż znikomy*”¹.

¹ Gardocki L., 2013, *Pojęcie przestępstwa i podziały przestępstw w polskim prawie karnym*, Annales Universitatis Mariae Curie-Skłodowska. Sectio G, nr 60, Lublin, s.30

1.3 Metody pomiaru przestępczości

Pomiar przestępczości koncentruje się wokół statystyk kryminalnych oraz badań ankietowych.

Statystyki kryminalne

Wśród statystyk kryminalnych można wyróżnić statystyki policyjne oraz statystyki sądowe. Statystyki policyjne mogą przyjmować następującą formę²:

- **przestępczość rejestrowana**, która dotyczy zdarzeń wstępnie zakwalifikowanych przez organy ścigania jako przestępstwa. Wskaźnik ten rzadko stosuje się do określenia rozmiaru badanego zjawiska, gdyż niekoniecznie liczba przestępstw zarejestrowanych odzwierciedla liczbę osób, którym winę udowodniono. Poza tym, dane dotyczące tej formy przestępczości są dostępne publicznie wyłącznie bez kwalifikacji prawnej czynu,
- **przestępczość stwierdzona**, czyli liczba zdarzeń, co do których w zakończonym postępowaniu przygotowawczym potwierdzono, że są przestępstwami. Jest to najbardziej precyzyjna miara, ponieważ jednostką obliczeniową nie jest postępowanie, a czyn, dzięki czemu wzrasta prawdopodobieństwo poprawnej klasyfikacji zdarzenia w porównaniu z przestępczością rejestrowaną. Ponadto, na tym poziomie dostępne są dane dla poszczególnych kategorii przestępczości,
- **przestępczość wykrywana**, dotycząca osób wstępnie zakwalifikowanych przez organy ścigania jako sprawców przestępstw. Miara ta przypomina przestępczość rejestrowaną, jednak jej jednostką nie jest postępowanie, lecz osoba. Dane te pozwalają na sprofilowanie sprawców przestępstw, a także na wstępne ocenienie, za pomocą innych zestawów danych, skuteczności pracy policji (wskaźnik wykrywalności sprawców przestępstw).

Statystyki sądowe dzielą się na następujące grupy³:

- **przestępczość osądzana**, która określa liczbę orzeczeń sądowych przypadających na jedną osobę oraz zbiór osób wstępnie uznanych za przestępców, które pozostawiono sądowi do oceny. Dane te dotyczą niestety wyłącznie orzeczeń sądów pierwszej instancji, czyli również wyroków nieprawomocnych. Poza tym

² Błachut J., 2007, *Problemy związane z pomiarem przestępczości*, Łódź, Wolters Kluwer Polska, s.166-198

³ Błachut J., 2007, *Problemy związane z pomiarem przestępczości*, Łódź, Wolters Kluwer Polska, s.223-244

osoba, która popełniła wiele przestępstw i stawia się przed sądem, będzie odnotowana wyłącznie według czynu o największej przewidzianej karze,

- **przestępczość w ujęciu jurydycznym**, która bierze pod uwagę liczbę wyroków prawomocnych. Z racji jednak, że utrudniony jest dostęp do prawomocnych wyroków z oskarżenia prywatnego oraz tych z sądów drugiej i trzeciej instancji, jest to zbiór danych dość kłopotliwy w zastosowaniu.

Jedną z wad statystyk kryminalnych jest tzw. „ciemna liczba przestępstw”. Zgodnie ze słownikiem PWN⁴, jest to „*liczba przestępstw faktycznie popełnionych, lecz nie objętych przez statystykę kryminalną wskutek nieujawnienia ich przez organy ścigania*”. Dotyczy ona zarówno przestępstw tuszowanych (np. ze względów politycznych), przestępstw niewykrytych, jak i przestępstw, które nie zostały zgłoszone (dotyczy to takich przestępstw, jak: przemoc domowa, przestępstwa na tle seksualnym, przemoc psychiczna i wynika m.in. ze strachu lub wstydu ofiary przed ujawnieniem czynu).

Drugą wadą jest ciągła zmienność prawa, która utrudnia porównywanie przestępstw między latami. Może to powodować sytuację, kiedy coś, co nie było czynem zabronionym, staje się w roku następnym przestępstwem i powoduje duży przyrost przestępczości.

Badania sondażowe

Wśród badań sondażowych można wyróżnić⁵:

- **badania typu *self-report* (sondaże sprawców)**, które, jak wskazuje nazwa, polegają na przeprowadzeniu sondażu, w którym ludzie pytani są o popełnione przestępstwa w całym okresie swojego życia lub w wybranym horyzoncie czasowym. Wyzwaniem jest określenie pojęć „przestępstwo” i „przestępczość”, aby odróżnić je od zachowań dewiacyjnych, oraz wybranie odpowiedniej próby reprezentatywnej,
- **badania wiktymizacyjne (sondaże ofiar)**, które miały stanowić sposób na pozbycie się ciemnej liczby przestępstw. W tego typu sondażach reprezentatywna grupa ludzi, którą mogą być np. pojedyncze osoby lub gospodarstwa domowe, pytana jest o zdarzenia, w których było się ofiarą przestępstwa.

⁴ <https://encyklopedia.pwn.pl/haslo/ciemna-liczba-przestepstw;3886513.html> (dostęp: 07.06.2019)

⁵ Błachut J., 2007, *Problemy związane z pomiarem przestępczości*, Łódź, Wolters Kluwer Polska, s.256-315

Badania sondażowe nie zyskały miana alternatywy dla statystyk kryminalnych, pozostając danymi uzupełniającymi⁶. Wynika to z problemów metodologicznych, takich jak różnica w określeniu obiektów badań, zadawanych pytaniach, czy różne horyzonty czasowe. Mimo tego, stanowią one dobry wskaźnik dla określenia ciemnej liczby przestępstw, zwłaszcza na podstawie badań wiktymizacyjnych.

1.4. Podział przestępstw

Przestępstwa można dzielić według takich czynników, jak: podmiot przestępstwa, skutek, formę uczestnictwa, czy wysokość sankcji. Zgodnie jednak z podziałem zastosowanym w statystykach policyjnych, a w konsekwencji również i w Banku Danych Lokalnych, wśród przestępstw można wyróżnić (kolejność malejąca pod względem liczby przestępstw dla kategorii):

- **przestępstwa kryminalne**, w których największą grupę stanowią przestępstwa przeciwko mieniu (art. 278-295 KK)⁷. Oprócz tego są to m.in. przestępstwa przeciwko życiu i zdrowiu (art. 148-162 KK), bezpieczeństwu powszechnemu (art. 163-172 KK), wolności (art. 189-193 KK), rodzinie i zdrowiu (art. 206-211 KK),
- **przestępstwa gospodarcze**, czyli przestępstwa przeciwko ochronie informacji (art. 265-269 KK), wiarygodności dokumentów (art. 270-277 KK), obrotowi gospodarczemu (art. 296-309 KK), obrotowi pieniędzmi i papierami wartościowymi (art. 310-315 KK),
- **przestępstwa drogowe**, czyli przestępstwa przeciwko bezpieczeństwu w komunikacji (art. 173-180 KK)

1.5. Analiza zjawiska przestępczości podejmowana w pracach badawczych

Problem ekonometrycznej analizy poziomu przestępczości był podejmowany w wielu pracach badawczych⁸. Kluczowym zagadnieniem przy wszystkich metodach statystycznych jest określenie zbioru zmiennych objaśniających, które będą najlepiej tłumaczyć badane zjawisko. Do tej pory jednak nie został wyłoniony zbiór uniwersalnych

⁶ Jak wyżej

⁷ W roku 2017 stanowiły one 80% przestępstw kryminalnych. Jest to również najliczniejsza grupa, jeśli brać pod uwagę ogół przestępstw (49% ogółu przestępstw).

⁸ Patrz: omawiane poniżej prace

zmiennych, które znalazłyby zastosowanie w każdej z prac naukowych. Różnorodność badań spowodowana jest mnogością istniejących teorii, z których każda zwraca uwagę na odmienne przyczyny szerzącego się bezprawia.

Jedną z pierwszych prób analizy podjął G.S. Becker⁹, twórca ekonomicznej teorii przestępczości, który założył racjonalność postępowania przestępców. Według Beckera, złamanie prawa determinuje wykonany przez nich rachunek zysków i strat, który z kolei zależy od takich czynników, jak: wiek, płeć, rasa, czy poziom zamożności przestępcy.

Za początek badań ekonometrycznych w tym zakresie można uznać artykuł I. Ehrlicha¹⁰, który zbudował model regresji wielorakiej, który opisywał przestępczość w Stanach Zjednoczonych w latach 40., 50. oraz 60. ubiegłego wieku. W badaniu tym liczba przestępstw na osobę jest wyjaśniana przez takie zmienne objaśniające, jak średni czas odsiadki przestępców w więzieniu, średni dochód rodzin, odsetek ludności o innym, niż biały kolorze skóry, współczynnik maskulinizacji¹¹, stopa bezrobocia wśród mężczyzn z miast w wieku 14-24 i 35-39 lat, średnia liczba lat nauki dla osób powyżej 25 roku życia, czy też zmienna binarna oznaczająca przynależność do południa Stanów Zjednoczonych, co można uznać za pewien początek analizy przestrzennej przy badaniu przestępczości. Jak można zauważyć, Ehrlich wykorzystał w swoim badaniu wszystkie te czynniki, o których wspominał Becker (pomijając rachunek zysków oraz strat). Główne zależności, które zostały wychwycone przez model, są następujące:

- prawdopodobieństwo zatrzymania i ukarania karą pozbawienia wolności oraz średni czas odbywania kary w więzieniach mają ujemny wpływ na wielkość przestępczości praktycznie każdego rodzaju,
- nierówności dochodowe i średni dochód mają duży, dodatni wpływ na przestępstwa przeciwko mieniu, natomiast mniejszy na przestępstwa kryminalne,
- zwiększenie odsetka ludzi o innym niż biały kolorze skóry powoduje, dla wszystkich przestępstw, wzrost zmiennej objaśnianej

Mocno zaakcentowana w literaturze jest teoria dezorganizacji społecznej, do której rozpropagowania przyczynili się głównie C.R. Shaw oraz H.D. McKay¹²,

⁹ Becker G. S., 1968, *Crime and Punishment: An Economic Approach*, Journal of Political Economy nr 76, s. 169-217

¹⁰ Ehrlich I., 1973, *Participation in Illegitimate Activities: A Theoretical and Empirical Investigation*, The Journal of Political Economy nr 81, s. 521-565

¹¹ Współczynnik maskulinizacji to liczba mężczyzn przypadająca na sto kobiet.

¹² Shaw C., McKay H., 1942, *Juvenile Delinquency and Urban Areas*, University of Chicago Press, Chicago

analizując przestępczość wśród nieletnich w Chicago za pomocą map, w pierwszych dekadach ubiegłego wieku. Zgodnie z ich koncepcją, badane zjawisko nie jest determinowane przez zachowanie pojedynczych jednostek, lecz przez warunki społeczne, w których dana osoba egzystuje: warunki ekonomiczne najbliższego otoczenia, zróżnicowanie etniczne oraz rotację ludności¹³. Oprócz wyodrębnienia zmiennych, które mogą pełnić rolę zmiennych objaśniających dla modelu, Shaw i McKay wskazali też przestrzenny wzór przestępczości, stwierdzając, że spada ona wraz z oddalaniem się od centrum miasta. Wzór ten nie zanikał jednak pod wpływem czasu oraz zmieniania się struktury etnicznej miasta. Stąd wniosek, że na bezprawie wśród nieletnich, spośród proponowanych zmiennych, wpływ ma wyłącznie sytuacja w sąsiedztwie.

Test teorii dezorganizacji społecznej wykonali w swoim opracowaniu R.J. Sampson i W.B. Groves¹⁴. W latach 80. ubiegłego wieku zbudowali oni model regresji wielorakiej na podstawie badania ankietowego na terenie Anglii i Walii wśród mieszkańców tamtych terenów. Oprócz determinant przestępczości wskazanych przez Shawa i McKaya¹⁵, Sampson i Groves wskazali także na urbanizację oraz rozbięcie rodziny, przy czym wyłącznie pierwszy z tych czynników okazał się być statystycznie istotnym i razem z niskim statusem społeczno-ekonomicznym i mobilnością mieszkaniową wpływać dodatnio na poziom przestępczości¹⁶.

Do interesujących opracowań, które skupiły się na teorii dezorganizacji społecznej, można z pewnością zaliczyć artykuł R. Martineza i in.¹⁷, w którym przeanalizowano zależność między liczbą zabójstw a napływem imigrantów w mieście San Diego. W badaniu, do którego użyto regresji ujemnej dwumianowej o stałych efektach, za zmienne objaśniające przyjęto liczbę dorosłych przypadającą na jedno dziecko, procent osób urodzonych za granicą, wskaźnik niekorzyści ekonomicznej, stabilności mieszkaniowej oraz odsetek zatrudnienia zawodowego wśród dorosłych. Opracowanie powyższe pokazało, że wysoka mobilność mieszkaniowa może, wbrew teorii dezorganizacji społecznej, przyczynić się do spadku przestępczości. Jak sugerują

¹³ Wong C., 2002, *Clifford R. Shaw and Henry D. McKay: The Social Disorganization Theory*. CSISS Classics, UC Santa Barbara: Center for Spatially Integrated Social Science

¹⁴ Groves W.B., Sampson R.J., 1989, *Community Structure and Crime: Testing Social-Disorganization Theory*, *American Journal of Sociology*, nr 94, s. 774-802.

¹⁵ Shaw C., McKay H., 1942, *Juvenile Delinquency and Urban Areas*, University of Chicago Press, Chicago

¹⁶ Bogacka E., 2012, *Zastosowanie teorii społecznej dezorganizacji do badania zróżnicowania przestrzennego przestępczości*, *Wiadomości statystyczne*, nr 1, s. 43-51

¹⁷ Martinez R., Stowell J.I., Lee M.T., 2010, *Immigration and crime in an era of transformation: a longitudinal analysis of homicides in San Diego neighborhoods, 1980-2000*, *CRIMINOLOGY*, nr 3, s. 797-829

autorzy, w większych miastach społeczna dezorganizacja może być skutkiem nie zróżnicowanej struktury etnicznej, a ekonomicznych barier.

Kolejną teorią, która, obok teorii dezorganizacji społecznej, jest często wykorzystywana w pracach badawczych, jest teoria działań rutynowych. Została ona opracowana przez L.E. Cohena oraz M. Felsona¹⁸, którzy opierając się na ekologii społecznej Amosa Hawleya wyodrębnili trzy czynniki, które zbiegając się w tym samym czasie oraz przestrzeni umożliwiają popełnienie przestępstwa. Powyższe przyczyny to: potencjalny przestępca, odpowiedni cel i brak osoby, która mogłaby zapobiec złamaniu prawa. Cohen i Felson nie podali zbioru zmiennych niezależnych, które byłyby istotne dla ich teorii, jednak nie brakuje opracowań, które się tego podejmują. Jednym z najnowszych jest praca E. Bogackiej¹⁹, w której zbadano przestępstwa związane z kradzieżą. Autorka artykułu do każdego z czynników wymienionego przez Cohena i Felsona przypisała zmienne, które mogą wpływać na obserwowane zjawisko. W modelu regresji finalnie znalazły się dwie zmienne: liczba mieszkań na 100 km² (odpowiedni cel) oraz gęstość zaludnienia (brak osoby mogącej zapobiec przestępstwu), jednak istotna statystycznie okazała się być wyłącznie pierwsza z nich, mając pozytywny związek z przestępczością.

Część autorów przy doborze zmiennych do badania decyduje się wykorzystać więcej niż jedną teorię. Do takich opracowań należy np. artykuł Kingi Kądziołki²⁰, w którym można odszukać zmienne pochodzące z wszystkich wcześniej omawianych teorii. Autorka badania w celu analizy przestępczości dla powiatów w Polsce wykorzystowała drzewo regresyjne. Pozwoliło to na przydzielenie każdego z powiatu do jednej z ośmiu grup, co ukazało obszary zbliżone do siebie względem statystyk. Wyłoniono też w ten sposób zmienne najbardziej istotnie wpływające na poziom bezprawia: wskaźnik urbanizacji, odsetek gospodarstw jednoosobowych, natężenie przestępstw w powiatach sąsiednich, współczynnik rozwodów oraz udzielone noclegi na 1000 ludności. Aby ustalić, który z tych czynników ma największy wpływ na zmienną zależną, Kądziołka zbudowała drzewa losowe. Pokazały one, że na przestępczość najmocniej oddziałuje wskaźnik urbanizacji.

¹⁸ Cohen, L.E., Felson M., 1979, *Social Change and Crime Rate Trends: A Routine Activities Approach*, American Sociological Review, nr 44, s. 588-608

¹⁹ Bogacka E., 2016, *Czynniki kradzieży rzeczy cudzej w świetle teorii działań rutynowych. Przykład województwa wielkopolskiego*, Rozwój Regionalny i Polityka Regionalna, nr 36, s. 199-207

²⁰ Kądziołka K., 2016, *Determinanty przestępczości w Polsce. Analiza zależności z wykorzystaniem drzew regresyjnych*, Ekonomia. Rynek, gospodarka, społeczeństwo, nr 45, s. 53-81

Wykorzystania wielu teorii podejmuje się też w swojej książce W. Florczak²¹. W celu przeanalizowania przestępczości dla Polski w latach 1970-2008 wyłoniono zbiór 58 zmiennych objaśniających, które posłużyły do interpretacji zmienności czterech grup przestępstw: przeciwko życiu i zdrowiu, przeciwko mieniu, prowadzenie pojazdu przez osobę nietrzeźwą lub pod wpływem środków odurzających oraz pozostałych. Przy doborze zmiennych do modelu kierowano się takimi teoriami przestępczości jak np. teoria konfliktu, teoria racjonalnego wyboru, teoria anomii, czy wspomniane już wcześniej: teoria dezorganizacji społecznej i teoria działań rutynowych. Wykonane równania regresji oraz analiza mnożnikowa pozwoliły ustalić, która grupa czynników ma największy wpływ na daną kategorię przestępstw oraz jak zmiana jednej cechy wpłynie na poziom badanego zjawiska w kraju. Okazało się m.in., że na poziom przestępczości w Polsce, najbardziej wpływa pogorszenie się wskaźników rozwoju społecznego oraz złagodzenie polityki karnej, szczególnie w warunkach narastających nierówności ekonomicznych. Z kolei najbardziej efektywnym sposobem na obniżenie się poziomu przestępczości, jeśli chodzi o administrację państwową, jest zwiększenie dotacji na system sądowy.

Kolejnym opracowaniem, gdzie uwaga autora nie jest skupiona na wyłącznie jednej teorii jest artykuł P. Bieńka i in.²², w którym za pomocą klasycznej metody najmniejszych kwadratów objaśniana jest dla każdego powiatu liczba przestępstw przypadająca na osobę. W badaniu wykorzystano interakcje zmiennych ze zmienną binarną odpowiedzialną za wskazanie dużych miast, stąd możliwość wychwycenia różnic między wpływem czynnika na miejscowości o różnej liczbie ludności. Do determinant mających istotny wpływ na liczbę przestępstw na jedną osobę mają wpływ takie zmienne jak zużycie energii (destymulanta), wydatki na alkohol (destymulanta), czy średni dochód gospodarstwa domowego (stymulanta). Okazało się też być istotnym, czy powiat położony jest blisko granicy z Niemcami (odległość do 125 km dawała wartość '1').

Odnosząc się do wskazanej w artykule Bieńka zmiennej dotyczącej odległości od granicy z zachodnim sąsiadem Polski, warto odnieść się do książki E. Bogackiej²³ na temat struktury przestrzennej i czynników przestępczości na wspomnianym wyżej obszarze. Do tego terytorium zaliczone zostało województwo zachodniopomorskie,

²¹ Florczak W., 2013, *Co wywołuje przestępczość i jak ją ograniczać? Wielowymiarowa analiza makroekonomiczna*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź

²² Bieńek P., Cichocki S., Szczepaniec M., 2012, *Czynniki ekonomiczne a poziom przestępczości – badanie ekonometryczne*, Zeszyty Prawnicze, nr 12.1, s. 147–172

²³ Bogacka E., 2012: *Struktura przestrzenna i czynniki przestępczości na obszarze nadgranicznym Polski z Niemcami*, Bogucki Wydawnictwo Naukowe, Poznań

lubuskie oraz dolnośląskie. W swoim opracowaniu Bogacka skupiła się na dwóch teoriach przestępczości: teorii społecznej dezorganizacji oraz teorii działań rutynowych. Dla każdej z wymienionych koncepcji zastosowano metodę najmniejszych kwadratów. Stworzono w ten sposób dwa modele liniowe, z których każdy objaśnia liczbę wszystkich przestępstw na 10000 ludności. Z modelu regresji dla teorii społecznej dezorganizacji istotną determinantą okazało się być wyłącznie saldo migracji wewnętrznych (destymulanta), z kolei dla drugiej teorii były to: procent kobiet w wieku 15-24 lata w liczbie ludności ogółem (stymulanta), liczba targowisk na 100 km² (stymulanta), przeciętne miesięczne wynagrodzenie na osobę (stymulanta) oraz współczynnik feminizacji (symulanta). Ciekawy wniosek płynie ze skutków otwarcia się granicy polsko-niemieckiej oraz przystąpienia Polski do Unii Europejskiej (badanie prowadzono w latach 2000-2009). Mianowicie, wydarzenia te podziały stymulująco na przestępstwa powiązane z granicą, czyli przemylem i nielegalnym przekroczeniem granicy, jednak nie miały żadnego wpływu na pozostałą formę przestępczości. Bogacka przeprowadziła również testy na obecność autokorelacji przestrzennej zmiennej objaśnianej. Nie została jednak ona wykryta, więc postanowiono poprzestać na modelach liniowych.

Istnieją jednak badania, w których testy wskazują na konieczność zastosowania modeli uwzględniających autokorelację przestrzenną. Jedną z takich prac jest artykuł M. F. Cracolici oraz T. E. Uberti²⁴. Dotyczy on wyłonienia determinant przestępczości w prowincjach Włoch na lata 1999 i 2003. Zastosowano jeden model dla czterech zmiennych objaśnianych: liczby morderstw, liczby kradzieży, liczby oszustw finansowych oraz liczby wymuszeń pieniędzy. Kartogramy sugerowały, że w przypadku niektórych z nich, mogą tworzyć się klastry o podwyższonym poziomie przestępczości. W celu weryfikacji tej hipotezy wyznaczono najpierw dla każdej zmiennej objaśnianej trzy macierze odległości standaryzowane wierszami. Przeprowadzono dla każdej z nich testy Morana I i Geary'ego C, a także graficznie przedstawiono wykres rozrzutu Morana. Następnie dokonano estymacji modelu najmniejszych kwadratów, którego celem było wyjaśnienie przestępczości w prowincjach Włoch. Wśród zmiennych objaśniających można wymienić: średni czas spędzony w więzieniu, odsetek przestępstw z anonimowym sprawcą, udział przedsiębiorstw z sektora przemysłowego w całkowitej liczbie przedsiębiorstw, bezrobocie wśród mężczyzn w wieku 25-29 lat oraz odsetek obcokrajowców. Rozważano użycie zarówno modelu opóźnienia przestrzennego jak

²⁴ Cracolici M.F., Uberti T.E., 2008, *Geographical Distribution of Crime in Italian Provinces: A Spatial Econometric Analysis*, Jahrbuch für Regionalwissenschaft, nr 29, s. 1-28

i błędu przestrzennego. Ostatecznie, znacznie częściej testy wskazywały na zasadność budowania dla zmiennej objaśnianej model opóźnienia przestrzennego. Otrzymane wyniki wskazały na zaniknięcie zależności przestrzennych z biegiem lat dla wszystkich przestępstw poza kradzieżą, natomiast istotność zmiennych objaśniających była różna w zależności od rozpatrywanego typu przestępstwa.

Warto również wspomnieć o pracy M. A. Andresena²⁵, w której oprócz klasycznej, posłużono się alternatywną metodą mierzenia przestępczości z użyciem przemocy, a mianowicie wyliczono zagrożenie przestępczością nie w danej dzielnicy miasta, lecz w przylegających obszarach. Wspomniane wyżej dzielnice dotyczą miasta Vancouver na rok 2001. Andresen wykorzystał dwa statystyczne podziały miasta: na tzw. *census tracts* (obszary spisu powszechnego, 110 dzielnic) oraz na 990 *dissemination areas* (990 jednostek). Zmienne objaśniające zostały podzielone na 3 kategorie: opisujące populację (np. odsetek mężczyzn w wieku 15-24 lata, zróżnicowanie etniczne, rodzice samotnie wychowujący dzieci), socjalno-ekonomiczne warunki (np. średni dochód domostwa, bezrobocie) i domostwa (ilość budynków zbudowanych przed 1961 rokiem, średnia wartość domostwa itp.). Zarówno test I Morana jak i kartogram przynależności do ćwiartek Morana wskazały na autokorelację przestrzenną dla obu podziałów miasta, jak i obu rodzajów metod obliczania przestępczości. Ostatecznie stworzone zostały cztery modele błędu przestrzennego. Wyniki pokazały, że zależności między obszarami są lepiej widoczne dla mniejszych jednostek statystycznych, a alternatywna metoda mierzenia nie jest metodą najlepszą, jednak ze względu na znikomy koszt pozyskania danych oraz odmienne od klasycznej metody wyniki powinna być brana pod uwagę przy badaniu przestępczości z użyciem przemocy, gdzie ma skłonności do bycia lepszą metodą od tej klasycznej.

Interesującym opracowaniem jest też artykuł X. Ye oraz L. Wu²⁶, którzy badali wskaźnik zabójstw dla miasta Chicago. W ich pracy została wykorzystana eksploracyjna analiza danych przestrzennych (ESDA), której wyniki sugerowały obecność autokorelacji przestrzennej zmiennej objaśnianej dla wszystkich punktów w czasie (badanie zostało przeprowadzone dla lat 1965-1990 w odstępach pięcioletnich). Zbudowane modele liniowe pokazały jednak, że autokorelacja zachodzi jedynie dla roku 1965. Ye i Wu nie zgodzili się jednak co do wyniku regresji, kierując się wynikami ESDA, wysokimi

²⁵ Andresen M.A., 2011, *The Ambient Population and Crime Analysis*, The Professional Geographer, nr 2, s. 193-212

²⁶ Ye X., Wu L., 2011, *Analyzing the dynamics of homicide patterns in Chicago: ESDA and spatial panel approaches*, Applied Geography, nr 31, s. 800-807

skorygowanymi współczynnikami determinacji R^2 oraz niejasnościami co do interpretacji zmiennych zależnych. Zbudowali modele panelowe z efektami ustalonymi, do pierwszego z nich włączając opóźnienie przestrzenne, natomiast do drugiego błąd przestrzenny. Model z włączonym błędem przestrzennym okazał się być lepszym z modeli i dla niego wyprowadzono końcowe wnioski. Wszystkie z rozważanych zmiennych objaśnianych (odsetek ludności poniżej granicy ubóstwa, odsetek ludności pozostającej w tym samym miejscu zamieszkania co 5 lat temu oraz odsetek cudzoziemców), bazujących na teorii społecznej dezorganizacji, okazały się być istotne w wyjaśnianiu liczby zabójstw. Odsetek ludności poniżej granicy ubóstwa zaprezentował się jako stymulanta przestępczości, natomiast pozostałe dwie zmienne to destymulanty.

W literaturze krajowej również można znaleźć prace dotyczące badania przestępczości i wykorzystującą analizę przestrzenną. Do takich opracowań należy artykuł K. Kądziołki²⁷. Głównym celem badania jest zbadanie przestępczości ogółem na poziomie powiatów w roku 2014. Macierz wag bazowała na macierzy sąsiedztwa według wspólnej granicy. Aby ograniczyć nadmierną korelację między zmiennymi objaśniającymi, za pomocą metody Warda wyłoniono determinanty reprezentatywne dla każdej z utworzonych w dendrogramie grup. Zgodnie ze wskazaniem testu LM utworzono model opóźnienia przestrzennego. Okazało się, że na poziom przestępczości w danym powiecie wpływa przestępczość w powiatach sąsiednich, średnie zarobki brutto, wskaźnik urbanizacji, procent gospodarstw jednoosobowych, liczba rozwodów na 1000 ludności, współczynnik migracji na 1000 osób oraz przyznane zakwaterowanie na 1000 osób. Wszystkie zmienne objaśniające zaprezentowały się jako stymulanty.

1.6. Podsumowanie

W niniejszym rozdziale zaprezentowano podstawowe pojęcia dotyczące przestępczości oraz metody jej pomiaru i podziału, co pozwoli na zrozumienie przeprowadzonych w dalszej części pracy badań. W tym celu utworzono także przegląd literatury, który zwrócił uwagę na ciekawsze opracowania oraz przedstawił różnorodność stosowanych w analizie przestępczości metod.

²⁷ Kądziołka K., 2016, *Analysis of the crime rate in Poland in spatial and temporal terms*, Central and Eastern European Journal of Management and Economics, nr 1, s. 81-96

2. Wybrane metody i modele analizy danych przestrzennych

2.1. Wstęp

W tytule pracy N. Becka²⁸ możemy przeczytać: „*Przestrzeń to więcej niż geografia*”. Cytat ten doskonale odzwierciedla rolę analizy przestrzennej w otaczającym nas środowisku. To właśnie dzięki jej metodom możemy badać wpływ pojedynczej jednostki na sąsiednie regiony, na co nie pozwalają zwykłe metody ekonometryczne. Ekonometria przestrzenna jest stosunkowo nową dziedziną, jednak wykorzystuje się ją już w takich obszarach naukowych, jak: medycyna, politologia, biologia i socjologia.

2.2. Charakterystyka danych przestrzennych

Twórca pojęcia „ekonometria przestrzenna”, J.H.P. Paelinck, wraz z L.H. Klaassenem²⁹ sformułował pięć reguł budowania przestrzennych modeli ekonometrycznych. Zgodnie z nimi, przy badaniu należy uwzględnić:

- współzależność przestrzenną zmiennych objaśniających,
- asymetrię i heterogeniczność (dotyczy zróżnicowania relacji ekonomicznych między dwoma jednostkami),
- allotopię (na wartość dla danego obszaru wpływają wartości w jednostkach sąsiednich),
- odmienne interakcje *ex ante* i *ex post*,
- zróżnicowanie charakterystyk przestrzeni (nie można patrzeć na dane ekonomiczne jak na zwykłe punkty).

Model przestrzenny jest tworzony, gdy ma miejsce któryś z efektów przestrzennych powodujących niedotrzymanie założeń MNK. Wśród nich można wyróżnić heterogeniczność przestrzenną oraz autokorelację przestrzenną.

Przestrzenna heterogeniczność ma miejsce, gdy zachodzi przestrzenna, monotoniczna lub skokowa, zmiana rozmieszczenia badanej zmiennej³⁰. Dochodzi wtedy do zmiany średniej, wariancji lub kowariancji w przestrzeni, co spowodowane jest brakiem zachowania stacjonarności przestrzennej. Narzędziami, za pomocą których

²⁸ Beck N., 2006, *Space Is More than Geography: Using Spatial Econometrics in the Study of Political Economy*, International Studies Quarterly, nr 50, s. 27-44

²⁹ Paelinck J.H.P., Klaassen L.H., 1983, *Ekonometria przestrzenna*, PWN, Warszawa

³⁰ Suhecki B. (red.), 2010, *Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych*, Wydawnictwo C.H.Beck, Warszawa, s.102

można badać przestrzenną heterogeniczność, są kartogramy. Z metod matematycznych przydatne są miary położenia (średnia, mediana przestrzenna) i rozproszenia (elipsa odchylenia standardowego, standardowa odległość).

Zjawisko autokorelacji przestrzennej może być skojarzone z pierwszym prawem geograficznym Toblera³¹, które mówi o tym, że „wszystko jest ze sobą powiązane, ale obiekty bliższe są ze sobą bardziej powiązane niż dalsze.” Zdarzenie to zachodzi bowiem, gdy wystąpienie cechy w danej jednostce zwiększa lub zmniejsza szansę na wystąpienie tej cechy w jednostkach sąsiednich³². Przyczyną tego może być powiązanie zjawiska z jednostkami terytorialnymi (dane zgrupowane według wartości mogą zostać rozdzielone przez granicę administracyjną, przez co efekt będzie widoczny w obu lokalizacjach) lub aspekt społeczno-ekonomiczny człowieka, który formowany jest przez odległość i lokalizację.

Wyszczególnić można dwa rodzaje przestrzennej autokorelacji: dodatnią oraz ujemną³³. **Dodatnia autokorelacja** oznacza, że wartości obserwowanych zmiennych grupują się w klastry wysokich lub niskich wartości. Odwrotnym zjawiskiem jest **autokorelacja ujemna**, której występowanie może być przedstawione w formie graficznej jako szachownica, gdyż wartości wysokie przeplatają się z niskimi. W obu przypadkach wyróżnia się także tzw. wyspy (ang.: *hot spots*), które stanowią obserwacje odstające. W przypadku braku autokorelacji rozmieszczenie wartości jest losowe i nie można mówić o tworzeniu się wzorców przestrzennych. Powyższe warianty zostały pokazane na poniższym rysunku.

Rysunek 2.1. Rodzaje autokorelacji przestrzennej



Źródło: Opracowanie własne.

³¹ Tobler W.R., 1970, *A computer model simulating urban growth in Detroit region*, Economic geography, nr 2, s. 236

³² Woźniak A., Sikora J., 2007, *Autokorelacja przestrzenna wskaźników infrastruktury wodno-ściekowej woj. małopolskiego*, Infrastruktura i ekologia terenów wiejskich, nr 2, s.315-329

³³ Suchecki B. (red.), 2010, *Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych*, Wydawnictwo C.H.Beck, Warszawa, s.103

Zależność przestrzenna może się pojawić w dwóch miejscach modelu: na poziomie błędu (błąd przestrzenny) oraz jako przestrzennie skorelowane obserwacje zmiennej objaśnianej (opóźnienie przestrzenne)³⁴. W przypadku pierwszej zależności regiony z wysokim/niskim błędem regresji graniczą z obszarami o także wysokim/niskim poziomie błędu. Z kolei obecność opóźnienia przestrzennego oznacza, że region z wysokim/niskim natężeniem badanej zmiennej otoczony jest przez regiony z podobnymi wartościami tej zmiennej.

Omówione wyżej dane przestrzenne podzielić można na cztery rodzaje³⁵:

- dane punktowe, które pokazują konkretne punkty w przestrzeni (np. rozmieszczenie budynków),
- dane liniowe (np. położenie dróg),
- dane obszarowe, o zmienności skokowej, dotyczące zjawisk i procesów ekonomicznych. Zmienne są obserwowane dla obiektów w postaci fragmentów powierzchni posiadających granice (np. gminy, powiaty, województwa),
- dane powierzchniowe, o ciągłej zmienności, dotyczące najczęściej zasobów naturalnych

Wszystkie z wymienionych danych przestrzennych mogą być badane nie tylko w pewnej przestrzeni geograficznej, ale też w zadanych okresach czasu. Dane te przyjmują wtedy charakter danych przestrzenno-czasowych³⁶.

2.3. Eksploracyjna analiza danych przestrzennych

Zbiorem narzędzi, które umożliwiają zbadanie przestrzennych danych, jest eksploracyjna analiza danych przestrzennych (*ang. ESDA - exploratory spatial data analysis*). Głównym jej zadaniem jest zbadanie autokorelacji przestrzennej. Pozwala na zidentyfikowanie odstających obserwacji, klastrów podobnych wartości, przestrzennych reżimów a także przestrzennej niestacjonarności³⁷.

³⁴ Kopczewska K., 2006, *Ekonometria i statystyka przestrzenna z wykorzystaniem programu R CRAN*, CeDeWu, Warszawa, s. 14

³⁵ Haining R., 2003, *Spatial Data Analysis Theory and Practice*, Cambridge University Press, s.44, s.52

³⁶ Suhecki B. (red.), 2010, *Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych*, Wydawnictwo C.H.Beck, Warszawa s.39

³⁷ Anselin L., 1998, *Interactive Techniques and Exploratory Spatial Data Analysis* w Longley P. i in., *Geographical Information Systems: Principles, Techniques, Management and Application*, New York: John Wiley & Sons, s. 258

Macierze wag

Kluczowym elementem analizy danych przestrzennych jest określenie macierzy wag, która definiuje powiązania między jednostkami. Sposób ich tworzenia ma znaczący wpływ na wyniki przeprowadzonych badań, gdyż wagi decydują o tym, która z otaczających nas jednostek oddziałuje silniej na badany region.

Macierz wag, definiowana zwykle literą W , jest macierzą kwadratową o wymiarach $n \times n$, gdzie n to liczba badanych obiektów. i -ty wiersz macierzy określa siłę, z jaką inne regiony wpływają na i -ty region. Jednostka nie może wpływać bezpośrednio na samą siebie, dlatego macierz jest w postaci diagonalnej.

W literaturze obecne są dwa rodzaje macierzy wag: macierze sąsiedztwa oraz odległości. Pierwsze z nich są macierzami binarnymi, których wartości obrazuje poniższy wzór:

$$W_{ij} = \begin{cases} 1, & \text{gdy obiekt } i \text{ sąsiaduje z obiektem } j \\ 0, & \text{gdy obiekt } i \text{ nie sąsiaduje z obiektem } j \text{ lub } i = j \end{cases} \quad (2.1)$$

Ważne jest określenie, co rozumie się poprzez „sąsiedztwo”. Podstawowe kryterium wykorzystuje macierz sąsiedztwa według wspólnej granicy. Wartość w macierzy wag przyjmuje wartość 1, gdy regiony ze sobą graniczą, czyli są sąsiadami pierwszego rzędu. Możliwe jest jednak określenie sąsiedztwa według drugiego, trzeciego itd. rzędów. Oznacza to, że sąsiedzi naszych sąsiadów są także naszymi sąsiadami i dla nich również przewidziana jest wartość 1. Regiony nie muszą graniczyć krawędzią. Możliwe jest ustalenie sąsiedztwa zgodnie z tzw. kryterium królowej, czyli liczy się także graniczenie samymi wierzchołkami jednostek³⁸.

Innym rozwiązaniem jest macierz oparta o kryterium odległości. Obiekt i jest wtedy sąsiadem obiektu j , gdy leży on w odległości nie większej niż d km. Odległość ta mierzona jest między geograficznymi środkami obiektów.

Sąsiedztwo może być także definiowane poprzez macierz k najbliższych sąsiadów. Środek każdego z badanych obszarów łączy się z k najbliższymi obszarami, co pozwala na ustalenie dla każdej jednostki k sąsiadów. Wadą tego rozwiązania jest niesymetryczność powstałej macierzy, przez co wymaga ona „usymetrycznienia”, aby nie zaburzać wyników testów na autokorelację przestrzenną³⁹.

³⁸ Anselin L, 1988, *Spatial Econometrics: Methods and models*, Springer Science, s.18

³⁹ Kopczewska K., 2006, *Ekonometria i statystyka przestrzenna z wykorzystaniem programu R CRAN*, CeDeWu, Warszawa, s. 59

Macierze odległości bazują na odległości między środkami geograficznymi lub innymi, charakterystycznymi punktami (np. stolice państw). Są one niebinarne, a ich wartości wylicza się na podstawie następującego wzoru⁴⁰:

$$W_{ij} = \begin{cases} \frac{1}{d_{ij}^\alpha}, & \text{gdy } d_{ij} \leq d^* \\ 0, & \text{gdy } d_{ij} > d^* \end{cases} \quad (2.2)$$

Wartość α to z góry ustalony parametr, d_{ij} to odległość między dwoma regionami, natomiast d^* to przyjęta wartość, po przekroczeniu której nie uznaje się wpływu zmiennej z obszaru i na zmienną z obszaru j . Inny wzór na obliczanie elementów macierzy prezentuje się następująco⁴¹:

$$W_{ij} = \begin{cases} e^{-\alpha d_{ij}}, & \text{gdy } d_{ij} \leq d^* \\ 0, & \text{gdy } d_{ij} > d^* \end{cases} \quad (2.3)$$

W powyższym wzorze do obliczenia elementów macierzy wykorzystano liczbę Eulera, jednak jest to wciąż funkcja wykorzystująca jedynie odległość. Czasem stosuje się funkcję, która bierze pod uwagę nie tylko odległość między regionami, ale też wielkość charakteryzującą stopień powiązania między regionami i oraz j , oznaczaną jako v_{ij} . Wartością tą może być np. stosunek długości granicy między regionami i , j do obwodu danego regionu⁴². Funkcja, która wykorzystuje powyższą miarę, ma następującą formę:

$$w_{ij} = d_{ij}^{-\alpha} * v_{ij}^\beta \quad (2.4)$$

gdzie: $i, j = 1, 2, \dots, n$, $d_{ij} > 0$, $i \neq j$ oraz α, β to ustalone parametry

Jeśli parametry α i β będą dodatnie, największe wartości w macierzy odległości zostaną uzyskane, gdy odległość między regionami będzie jak najmniejsza, a ich wspólna granica jak najdłuższa⁴³.

Czasem autorzy badań decydują się na macierze odległości ekonomicznej. Przykładem takiego działania może być macierz bazująca na środkach komunikacji między regionami, biorąca pod uwagę ich relatywną ważność⁴⁴. Ten typ macierzy jednak

⁴⁰ Suchecki B. (red.), 2010, *Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych*, Wydawnictwo C.H.Beck, Warszawa, s.105

⁴¹ Ibidem

⁴² Zeliaś A., 1991, *Ekonometria przestrzenna*, Państwowe Wydawnictwo Ekonomiczne, Warszawa, s. 100

⁴³ Ibidem

⁴⁴ Bodson P., Peters D., 1975. *Estimation of the Coefficients of a Linear Regression in the Presence of Spatial Autocorrelation: An Application to a Belgium Labor Demand Function*, Environment and Planning, nr 7, s. 455-472

może powodować trudności przy interpretacji, gdyż wagi powinny być egzogeniczne względem modelu⁴⁵.

Aby umożliwić porównanie parametrów dla różnych modeli przestrzennych oraz dla łatwiejszej interpretacji autokorelacji i autoregresji, rzędy macierzy są zazwyczaj standaryzowane wierszami zgodnie z poniższym wzorem, dzięki czemu wagi przybierają wartości z zakresu $[0,1]$, sumując się rzędami do 1⁴⁶:

$$w_{ij}^* = \frac{w_{ij}}{\sum_j w_{ij}} \quad (2.5)$$

gdzie w_{ij} to waga przestrzenna określająca powiązanie między regionami i oraz j .

W praktyce, najczęściej stosowanymi macierzami są macierze sąsiedztwa według wspólnej granicy⁴⁷, co wydaje się być intuicyjnym wyborem i takiego wyboru dokonano w badaniu empirycznym w niniejszej pracy.

Wariogram

Wariogram jest narzędziem eksploracyjnej analizy danych, który pozwala na zbadanie stacjonarności zmiennej. Funkcja ta, przedstawiana na wykresie, odpowiada za pokazanie zależności między wariancją cechy i jej położeniem. W praktyce stosowany jest semiwariogram (nazywany także wariogramem), który stanowi połowę wariogramu i wyrażony jest następującym wzorem⁴⁸:

$$\gamma(h) = \frac{1}{2} \text{var}[Z(x) - Z(x + h)] = \frac{1}{2} E[\{Z(x) - Z(x + h)\}^2] \quad (2.6)$$

gdzie $Z(x)$ to wartość Z w punkcie x , natomiast h to wektor odległości między dwoma punktami.

Wyróżnia się także wariogram empiryczny, którego postać estymowana jest za pomocą metody momentów, na podstawie bezpośrednich wyników i ma poniższy wzór⁴⁹:

$$\hat{\gamma}(h) = \frac{1}{2m(h)} \sum_{j=1}^{m(h)} \{z(x_j) - z(x_j + h)\}^2 \quad (2.7)$$

gdzie $i = 1, 2, \dots, N$, a $m(h)$ to zbiór par punktów o odległości h .

⁴⁵ Kopczewska K., 2006, *Ekonometria i statystyka przestrzenna z wykorzystaniem programu R CRAN*, CeDeWu, Warszawa, s. 56

⁴⁶ Suhecki B. (red.), 2010, *Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych*, Wydawnictwo C.H.Beck, Warszawa, s.107

⁴⁷ Kopczewska K., 2006, *Ekonometria i statystyka przestrzenna z wykorzystaniem programu R CRAN*, CeDeWu, Warszawa, s. 57

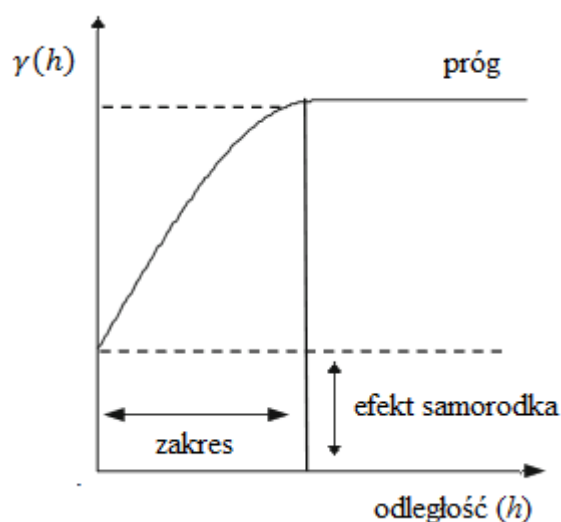
⁴⁸ Oliver M.A., Webster R., 2014, *A tutorial guide to geostatistics: Computing and modelling variograms and kriging*, Catena, s. 56-69

⁴⁹ Ibidem

Istnieje także możliwość wyestymowania wariogramu teoretycznego, w tym celu wybierając jeden z czterech kształtów wariogramów: nugget effect, sferyczny, wykładniczy oraz gaussowski. Wybrany wykres musi jak najlepiej opisywać wariogram eksperymentalny. Przy napotkaniu trudności w dopasowywaniu, można posłużyć się kombinacją kilku podstawowych kształtów.

Dla interpretacji graficznej wariogramu można wyróżnić trzy charakterystyczne parametry, widoczne na poniższym rysunku.

Rysunek 2.2. Charakterystyka wariogramu



Źródło: opracowanie własne na podstawie książki M.A. Oliver i R. Webstera⁵⁰.

Znaczenie zaznaczonych części semiwariogramu jest następujące⁵¹:

- **próg** (*sill*) - maksymalna wariancja semiwariogramu; obecność takiego punktu świadczy o stacjonarności modelu (semiwariogram rosnący do nieskończoności oznacza niestacjonarność),
- **efekt samorodka** (*nugget variance*) - punkt przecięcia z osią OY, początkowa wartość semiwariogramu dla wartości $h = 0$; jego wysoka wartość może świadczyć o znacząco różnych wartościach dla bliskich sobie lokalizacji,
- **zakres** (*range*) - odległość h dla osiągniętego progu, jest to największy dystans dla którego zachodzi autokorelacja.

⁵⁰ Oliver M.A., Webster R., 2015, *Basic Steps in Geostatistics: The Variogram and Kriging*, s. 31

⁵¹ Suchecki B. (red.), 2010, *Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych*, Wydawnictwo C.H.Beck, Warszawa s.78-80

Korelogram

Narzędziem, które pozwoli na zbadanie występowania przestrzennej autokorelacji, jest korelogram przestrzenny. Jego zbudowanie ma na celu zbadanie, do ilu opóźnień przestrzennych bądź do jakiej odległości między jednostkami zachodzi autokorelacja przestrzenna. Wykres może być sporządzony z wykorzystaniem metody korelacji lub metody Morana⁵². Jeśli chociaż jedna wartość jest istotna, można mówić o istotności całego korelogramu. Liczba badanych opóźnień przestrzennych wynosi zazwyczaj 5-6⁵³. Jeśli między nimi pojawi się przerwa w istotności, może to świadczyć o niestacjonarności modelu i należy uznać za istotne jedynie opóźnienia przestrzenne, które występują przed powstałą luką. Przy badaniu istotności opóźnień przestrzennych należy stosować korektę Bonferroniego, która polega na podwojeniu wartości p-value lub stosowaniu poziomu istotności dzielonego przez liczbę opóźnień⁵⁴. Ważnym aspektem korelogramu jest też nachylenie powstałych wartości. Jeżeli można zaobserwować wzrost współczynnika korelacji lub statystyki Morana, świadczy to o istnieniu makro-regionów⁵⁵. Odwrotny trend oznacza obecność kilku wiodących regionów, co może być skutkiem takiego zjawiska jak podwyższone koszty transportu.

Globalna autokorelacja przestrzenna

Globalne miary pozwalają zbadać obecność autokorelacji przestrzennej na terenie całego obszaru. W odróżnieniu od miar lokalnych, jest to jednak miara uśredniona i nie pozwala na znalezienie konkretnych lokalizacji, które wraz z sąsiadującymi regionami przyczyniają się do istnienia badanego zjawiska⁵⁶. Aby przeprowadzić rzetelne testy z wykorzystaniem statystyk przestrzennych, należy spełnić następujące założenia⁵⁷:

- **stacjonarność zmiennej** - wystarczy słaba stacjonarność, czyli sytuacja, kiedy średnia i wariancja są stałe, a autokorelacja zależy wyłącznie od dzielącej obiekty odległości,
- **izotropizm** - wzorce przestrzenne nie mogą zależeć od kierunku badania.

⁵² Kopczewska K., 2006, *Ekonometria i statystyka przestrzenna z wykorzystaniem programu R CRAN*, CeDeWu, Warszawa, s. 119

⁵³ Ibidem

⁵⁴ Kopczewska K., 2006, *Ekonometria i statystyka przestrzenna z wykorzystaniem programu R CRAN*, CeDeWu, Warszawa, s. 120

⁵⁵ Ibidem

⁵⁶ Kopczewska K., 2006, *Ekonometria i statystyka przestrzenna z wykorzystaniem programu R CRAN*, CeDeWu, Warszawa, s. 69

⁵⁷ Kopczewska K., 2006, *Ekonometria i statystyka przestrzenna z wykorzystaniem programu R CRAN*, CeDeWu, Warszawa, s. 70

Najczęściej stosowanymi globalnymi miarami autokorelacji przestrzennej są: statystyka Morana I, Geary'ego C oraz statystyki join-count.

Statystyka globalna Morana I ma postać⁵⁸:

$$I = \frac{1}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} * \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (2.8)$$

gdzie:

- w_{ij} - elementy macierzy wag W
- n - liczba wszystkich regionów
- x_i - wartość badanej cechy w regionie i

Statystyka globalna Morana I testuje następujący zbiór hipotez⁵⁹:

H₀: Brak autokorelacji przestrzennej (wartości zmiennej rozmieszczone są losowo)

H₁: Występuje autokorelacja przestrzenna

Weryfikacja hipotez odbywa się poprzez momenty teoretyczne albo podejście permutacyjne⁶⁰. Wartość oczekiwana statystyki, obliczana poniższym wzorem, jest bliska zeru, co oznacza losowość:

$$E(I) = -\frac{1}{n-1} \quad (2.9)$$

Statystyka Morana I przyjmuje zazwyczaj wartości w zakresie $[-1,1]$ i można je interpretować jako współczynnik korelacji⁶¹. Pierwiastek z tej liczby pokazuje, jaki procent zmienności cechy można tłumaczyć lokalizacją (w taki sposób można odczytywać również wartości korelogramu zbudowanego za pomocą metody Morana). Dodatnie wartości oznaczają autokorelację dodatnią, czyli klastry o podobnych wartościach, natomiast ujemne wyniki świadczą o ujemnej autokorelacji, czyli istnieniu tzw. *hot spots*⁶².

Graficzną reprezentacją statystyki Morana I jest wykres punktowy Morana. Pokazuje on relację między liczbą odchyłeń wartości zmiennej od średniej na osi OX

⁵⁸ Suchecki B. (red.), 2010, *Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych*, Wydawnictwo C.H.Beck, Warszawa s.112

⁵⁹ Suchecki B. (red.), 2010, *Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych*, Wydawnictwo C.H.Beck, Warszawa s.114

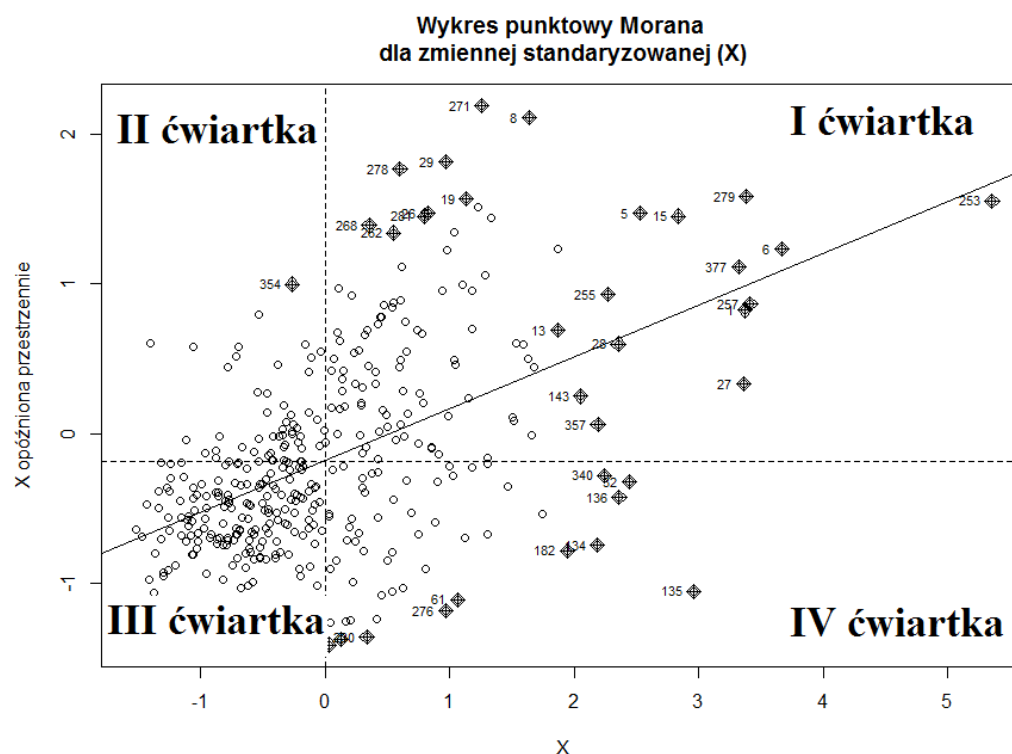
⁶⁰ Kopczewska K., 2006, *Ekonometria i statystyka przestrzenna z wykorzystaniem programu R CRAN*, CeDeWu, Warszawa, s. 72

⁶¹ Kopczewska K., 2006, *Ekonometria i statystyka przestrzenna z wykorzystaniem programu R CRAN*, CeDeWu, Warszawa, s. 73

⁶² Ibidem

i średnią wartością cechy w sąsiednich regionach na osi OY⁶³. Wartość statystyki można odczytać z współczynnika kierunkowego linii regresji, która dopasowywana jest względem naniesionych punktów. Wykres stosować można dla każdej zmiennej, możliwe też jest odczytanie z niego, zgodnie z regułą trzech sigm⁶⁴, odstających obserwacji. Poniżej znajduje się przykład omawianego wykresu oraz jego interpretacja.

Rysunek 2.3. Przykład wykresu punkowego Morana



Źródło: opracowanie własne z wykorzystaniem programu R.

Wykres jest podzielony na cztery ćwiartki względem punktu (0,0). To, do których ćwiartek przynależą w przeważającej liczbie punkty, determinuje obecny typ autokorelacji przestrzennej. Przynależność do I lub III ćwiartki (odpowiednio: wysokie wartości zmiennej i jej sąsiadów; niskie wartości zmiennej i jej sąsiadów) świadczy o pozytywnej autokorelacji przestrzennej, czyli tworzeniu się klastrów podobnych wartości⁶⁵. Obecność punktu w II albo IV ćwiartce (odpowiednio: wysokie wartości zmiennej, niskie wartości w sąsiedztwie; niskie wartości zmiennej, wysokie wartości w sąsiedztwie) oznacza z kolei ujemną autokorelację przestrzenną. Jeśli obserwacje są

⁶³ Le Sage J.P., 2008, *An Introduction to Spatial Econometrics*, Revue d'économie industrielle, nr 123, s.35-36

⁶⁴ Kopczewska K., 2006, *Ekonometria i statystyka przestrzenna z wykorzystaniem programu R CRAN*, CeDeWu, Warszawa, s. 75

⁶⁵ Le Sage J.P., 2008, *An Introduction to Spatial Econometrics*, Revue d'économie industrielle, nr 123, s.36-37

dość równomiernie rozłożone we wszystkich ćwiartkach wykresu, można mówić o braku autokorelacji⁶⁶. W powyższym przypadku można mówić o autokorelacji dodatniej, gdyż większość punktów należy do nieparzystych ćwiartek. Należy też zwrócić uwagę na znaczną liczbę obserwacji odstających, które oznaczone zostały przekreślonymi rombami.

Kolejną miarą globalnej autokorelacji przestrzennej jest **statystyka Geary'ego C**, zadana wzorem⁶⁷:

$$C = \frac{n-1}{2 \sum_i \sum_j w_{ij}} * \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - x_j)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (2.10)$$

Statystyka Geary'ego C przyjmuje ten sam zestaw hipotez, co statystyka I Morana. Hipotezę zerową o braku autokorelacji przestrzennej można zweryfikować za pomocą liczby wyznaczonej z powyższego wzoru. Zawiera się ona w przedziale [0,2], a jej wartość oczekiwana wynosi 1, co oznacza brak autokorelacji przestrzennej⁶⁸. Wartości większe niż jeden oznaczają autokorelację ujemną, natomiast wartości poniżej tej liczby wskazują na autokorelację dodatnią. Statystyka C Geary'ego jest mocno skorelowana ze statystyką I Morana⁶⁹, więc w niektórych przypadkach może istnieć potrzeba wykorzystania odmiennej miary globalnej autokorelacji.

Kolejną z globalnych miar autokorelacji jest **statystyka join-count** (joint-count), której idea bazuje na nanoszeniu kolorów, odpowiadających danej grupie wartości, na mapę. Brana pod uwagę jest liczba połączeń pól. Z racji przekształcania zmiennych na zmienne binarne, można na mapie wyróżnić trzy grupy: WW (ang. *white-white*, biało-białe), BB (ang. *black-black*, czarno-czarne) wskazujące na autokorelację dodatnią oraz BW (ang. *black-white*, czarno-białe). Są one opisane są następującymi wzorami⁷⁰:

$$BB = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n w_{ij} x_i x_j \quad (2.11)$$

$$WW = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n w_{ij} (1 - x_i)(1 - x_j) \quad (2.12)$$

⁶⁶ Kopczewska K., 2006, *Ekonometria i statystyka przestrzenna z wykorzystaniem programu R CRAN*, CeDeWu, Warszawa, s. 74

⁶⁷ Suhecki B. (red.), 2010, *Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych*, Wydawnictwo C.H.Beck, Warszawa s.114

⁶⁸ Kopczewska K., 2006, *Ekonometria i statystyka przestrzenna z wykorzystaniem programu R CRAN*, CeDeWu, Warszawa, s. 81

⁶⁹ Suhecki B. (red.), 2010, *Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych*, Wydawnictwo C.H.Beck, Warszawa s.115

⁷⁰ Suhecki B. (red.), 2010, *Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych*, Wydawnictwo C.H.Beck, Warszawa s.110

$$BW = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - x_j)^2 \quad (2.13)$$

Hipotezę zerową o braku autokorelacji przestrzennej testuje się za pomocą wartości BW, porównując jej wartość z jej rozkładem przy założeniu losowym⁷¹. Może odbywać się to na dwa sposoby: ze zwracaniem lub bez.

Lokalna autokorelacja przestrzenna

Statystyki LISA (ang. *Local Indicators of Spatial Association*, lokalne wskaźniki związków przestrzennych) należą do lokalnych miar autokorelacji przestrzennej. Pozwalają na zbadanie autokorelacji na bardziej szczegółowym stopniu, niż robią to miary globalne. Może okazać się, że pomimo braku globalnej autokorelacji można na mapie doszukać się klastrów podobnych wartości bądź obserwacji odstających. Żeby poprawnie wykorzystać statystykę LISA, konieczne jest spełnienie następujących założeń⁷²:

- wskaźnik dla każdej obserwacji powinien wskazywać zakres istotnego przestrzennego grupowania podobnych wartości wokół danej obserwacji
- suma wskaźnika dla wszystkich obserwacji jest proporcjonalna do globalnego wskaźnika asocjacji przestrzennej.

Pierwszą statystyką, która wchodzi w skład lokalnych wskaźników związków przestrzennych, jest **lokalna statystyka Morana I**, wyrażona wzorem⁷³:

$$I_i = z_i \sum_j w_{ij} z_j \quad (2.14)$$

gdzie z_i, z_j to odchylenia od średniej (odpowiednio: badana jednostka, regiony sąsiadujące), j to liczba sąsiadów natomiast w_{ij} to macierz wag. Ujemna wartość oznacza autokorelację ujemną, a dodatnia: autokorelację dodatnią. Powyższa statystyka jest testowana z wykorzystaniem hipotezy zerowej o braku autokorelacji przestrzennej. Testy mogą być wykonane przy użyciu podejścia randomizacji lub permutacji⁷⁴. Przy podejściu permutacji należy rozważyć użycie korekty Bonferoniego, która obniża poziom istotności

⁷¹ Kopczewska K., 2006, *Ekonometria i statystyka przestrzenna z wykorzystaniem programu R CRAN*, CeDeWu, Warszawa, s. 83

⁷² Anselin L., 1995, *Local indicators of spatial association*, *Geographical Analysis*, nr 27, s. 93–115

⁷³ Ibidem

⁷⁴ Ibidem

odwrotnie proporcjonalnie do liczby powtórzeń, aby uniknąć popełnienia błędu I rodzaju⁷⁵.

Drugą miarą lokalną, która wchodzi w skład statystyk LISA, jest lokalny odpowiednik globalnego testu Geary'ego C.

Lokalna statystyka Geary'ego wyraża się wzorem:

$$c_i = \sum_j w_{ij} (z_i - z_j)^2 \quad (2.15)$$

gdzie oznaczenia powielają się ze statystyką lokalną Morana I.

Podobnie jest z interpretacją, która jest identyczna jak w przypadku poprzedniej miary. Należy dodać, że w obu przypadkach wartości p-value poniżej poziomu istotności świadczą o dodatniej autokorelacji, natomiast wartości powyżej liczby $1 - \alpha$ świadczą o ujemnej autokorelacji⁷⁶.

2.4. Modele przestrzenne

W sytuacji wykrytej autokorelacji przestrzennej, nie można stosować zwykłej regresji liniowej. Mogłoby to zaburzyć parametry modelu liniowego i spowodować przekłamanie w wartościach parametrów modelu, gdyż taki model nie uwzględnia istotnego czynnika przestrzennego. W tym celu stosuje się modele, które uwzględniają macierz wag przestrzennych oraz współczynnik autokorelacji przestrzennej. Interakcja przestrzenna może przybierać wiele form i odnosić się do jednego z trzech parametrów modelu: zmiennej objaśnianej, zmiennych objaśniających lub składnika losowego⁷⁷. Podstawowe modele uwzględniają tylko jeden z wymienionych parametrów, natomiast metody zaawansowane skupiają się na kilku spośród wymienionych czynników.

Charakterystyka i estymacja modeli przestrzennych

Pierwszym z podstawowych modeli przestrzennych jest **model opóźnienia przestrzennego SLM** (ang. *Spatial Lag Model*), zwany również **modelem autoregresyjnym SAR** (ang. *Spatial Autoregressive Model*), który wykorzystuje

⁷⁵ Suhecki B. (red.), 2010, *Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych*, Wydawnictwo C.H.Beck, Warszawa s.125

⁷⁶ Kopczewska K., 2006, *Ekonometria i statystyka przestrzenna z wykorzystaniem programu R CRAN*, CeDeWu, Warszawa, s. 91

⁷⁷ Suhecki B. (red.), 2010, *Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych*, Wydawnictwo C.H.Beck, Warszawa s.239

opóźnione przestrzennie zmienne objaśniane regionów sąsiednich. Przyjmuje następującą postać⁷⁸:

$$Y = \rho W y + X \beta + \epsilon$$

$$\epsilon \sim N(0, \sigma^2 I)$$
(2.16)

gdzie:

- W to macierz wag przestrzennych,
- ρ to parametr autoregresyjny,
- Y to zmienna objaśniana,
- X to zmienne objaśniające,
- β to współczynniki modelu,
- ϵ to błąd modelu.

W modelu SLM testowana jest istotność parametru autoregresyjnego ρ . Jeśli równa się on zero, to nie można mówić o występowaniu autokorelacji przestrzennej. Model przyjmuje wtedy takie samo równanie jak zwykła regresja liniowa. Interpretacja istotnego parametru nie zawsze jednak jest tak intuicyjna, oznaczając większy wpływ jednostek sąsiednich przy wyższej wartości parametru autoregresyjnego. Istotny parametr może wskazywać na jeden z poniższych scenariuszów⁷⁹:

- występuje rzeczywiste przenikanie się zmiennych objaśnianych, ponieważ rozpatrywane determinanty odpowiadają badanym jednostkom, więc „rozlewanie się” wartości na sąsiednie regiony stanowi wynik modelu teoretycznego,
- ma miejsce niedopasowanie między skalą zjawiska a podziałem terytorialnym i włączenie do modelu opóźnionej zmiennej przestrzennej zapobiega wtedy utracie informacji, która wynika ze sztucznego podziału administracyjnego

Możliwe jest ustalenie, jak bardzo zmiana zmiennych objaśniających w regionach sąsiednich wpływa na zmienną objaśnianą w naszym regionie. Jest to tzw. **efekt ekwilibrium** (*ang. equilibrium effect*), który bazuje na wyliczeniu z równania SLM wartości oczekiwanej zmiennej endogenicznej. Wzór ten wygląda następująco⁸⁰:

$$E(y) = (I - \rho W)^{-1} X \beta$$
(2.17)

⁷⁸ Ward M., Gleditsch K., 2007, *An Introduction to Spatial Regression Models in the Social Sciences*, Michael D. Ward & Kristian Skrede Gleditsch

⁷⁹ Anselin L., Bera A., 1998, *Spatial Dependence in linear Regression Models with an Introduction to Spatial Econometrics*, Handbook of Applied Economic Statistics, s. 237-289

⁸⁰ Ward M., Gleditsch K., 2007, *An Introduction to Spatial Regression Models in the Social Sciences*, Michael D. Ward & Kristian Skrede Gleditsch

O sile interakcji przestrzennej informuje parametr przy X , czyli $(I - \rho W)^{-1}$. Jeśli nie zachodzi autokorelacja przestrzenna, wpływ zmiennej objaśnianej pozostaje zerowy.

Kolejnym z podstawowych modeli przestrzennych jest **model błędu przestrzennego SEM** (ang. *Spatial Error Model*), który wykorzystuje autokorelację przestrzenną składnika losowego.

Przyjmuje on następującą postać⁸¹:

$$\begin{aligned} y &= X\beta + \xi \\ \xi &= \lambda W\xi + \varepsilon \end{aligned} \quad (2.18)$$

W modelu SEM zmienne objaśniane w regionach sąsiednich nie wywierają bezpośredniego wpływu na daną jednostkę. Model regresji liniowej jest jednak obciążony błędem, gdyż interakcja przestrzenna zawarta jest w błędzie modelu. O sile tej zależności świadczy parametr λ , którego istotność oznacza obecność autokorelacji przestrzennej w modelu. Może to wynikać z pominięcia niemierzalnych zmiennych, które są przestrzennie skorelowane⁸² (np. układ rzek, podłoże historyczne). Inną możliwą interpretacją tego zjawiska jest pokazanie błędu przestrzennego jako zakłócenia, które powoduje autokorelację przestrzenną dla błędów pomiarowych bądź zmiennych, które są inaczej zbędne dla modelu⁸³. Spowodowany tym problem szacunku może być rozwiązany poprzez zwiększenie próby albo spójne szacowanie parametru.

Do szacowania parametrów obu modeli można wykorzystać kilka metod, jednak nie wystarczy do tego klasyczna metoda najmniejszych kwadratów (MNK). Najbardziej popularną z nich jest metoda największej wiarygodności (MNW), której wadą jest potrzeba skomplikowanego programowania⁸⁴. Do innych używanych metod można zaliczyć podwójną metodę najmniejszych kwadratów (2MNK), metodę zmiennych instrumentalnych (MZI), metodę momentów, czy też coding method⁸⁵.

Strategia wyboru modelu

Po wyborze macierzy wag oraz analizie ESDA, która potwierdziła możliwość istnienia autokorelacji przestrzennej, należy zbudować model, który będzie zawierał interakcję

⁸¹ Suhecki B. (red.), 2010, *Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych*, Wydawnictwo C.H.Beck, Warszawa s.250

⁸² Kopczewska K., 2006, *Ekonometria i statystyka przestrzenna z wykorzystaniem programu R CRAN*, CeDeWu, Warszawa, s. 133

⁸³ Anselin L., Bera A., 1998, *Spatial Dependence in linear Regression Models with an Introduction to Spatial Econometrics*, Handbook of Applied Economic Statistics, s. 249

⁸⁴ Suhecki B. (red.), 2010, *Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych*, Wydawnictwo C.H.Beck, Warszawa, s.267

⁸⁵ Anselin, L., *Spatial Econometrics*, Chapter Fourteen, *Spatial Econometrics* Luc Anselin, *A Companion to Theoretical Econometrics* Edited by Badi H. Baltagi 2001, 2003, Blackwell Publishing Ltd, s. 319-323

przestrzenną. Wyróżnić można dwie główne metody, które wskazują na kolejność przy budowie modelu.

Pierwszą, bardziej intuicyjną, techniką jest metoda klasyczna (od szczególnego do ogólnego, *ang. specific to general*), bazująca na trzech krokach⁸⁶:

- zbudowanie i estymacja wstępnego modelu regresji liniowej,
- testowanie zjawiska autokorelacji przestrzennej,
- w przypadku braku wykrycia powyższego należy podjąć procedury naprawcze (np. transformacja zmiennej, uwzględnienie dodatkowych zależności).

Kroki drugi i trzeci należy powtarzać aż do osiągnięcia ostatecznego, poprawnie wyestymowanego modelu z autokorelacją przestrzenną. Podejście powyższe nie zawsze przynosi najlepsze rezultaty ze względu na możliwość zastosowania niewłaściwych testów diagnostycznych, jednak w literaturze⁸⁷ jest to zalecana technika budowy modelu.

Drugą techniką jest podejście Hendry'ego (od ogólnego do szczególnego, *ang. general to specific*), która przyjmuje odwrotny porządek. Punkt wyjścia stanowi model z wszystkimi możliwymi efektami przestrzennymi, który jest następnie zawężany za pomocą przeprowadzanych testów istotności⁸⁸. Zostało udowodnione, że podejście to daje wyniki gorsze od podejścia klasycznego⁸⁹, więc dalsza część pracy skupi się na pierwszej z wymienionych technik.

Pierwszym etapem w metodzie klasycznej jest test Morana na autokorelację przestrzenną reszt modelu liniowego, który wyraża się wzorem⁹⁰:

$$I = \frac{n}{S_0} \cdot \frac{u' W u}{u' u} \quad (2.19)$$

w którym u jest wektorem $[n \times 1]$ reszt z modelu MNK.

Jeśli zostanie odrzucona hipoteza zerowa o braku autokorelacji, warto wykonać powyższy test także dla reszt z modelu szacowanego na logarytmach zmiennych⁹¹. Gdy

⁸⁶ Florax R. i in., 2003, *Specification Searches in Spatial Econometrics: The Relevance of Hendry's Methodology*, Regional Science and Urban Economics, nr 33, s. 2-3;5

⁸⁷ Florax R. i in., 2003, *Specification Searches in Spatial Econometrics: The Relevance of Hendry's Methodology*, Regional Science and Urban Economics, nr 33, s. 3;16-17

⁸⁸ Florax R, Nijkamp P., 2003, *Misspecification in Linear Spatial Regression Models*, Tinbergen Institute Discussion Paper, nr 3, s.18

⁸⁹ Ibidem

⁹⁰ Florax R, Nijkamp P., 2003, *Misspecification in Linear Spatial Regression Models*, Tinbergen Institute Discussion Paper, nr 3, s.15

⁹¹ Kopczewska K., 2006, *Ekometria i statystyka przestrzenna z wykorzystaniem programu R CRAN*, CeDeWu, Warszawa, s. 123

nie zostanie wykryta przestrzenna interakcja, to brak odrzucenia hipotezy zerowej mógł wynikać z powodu estymacji związku nieliniowego modelem liniowym.

Gdy autokorelacja przestrzenna mimo to występuje, należy wykonać testy mnożnika Lagrange'a (LM, *ang. Lagrange Multiplier*). Wyniki tego testu pozwolą na ustalenie, który z modeli, SAR czy SEM, będzie bardziej odpowiedni dla analizowanych danych. Oba testy w hipotezie zerowej zakładają nieistotność parametru odpowiedzialnego za autokorelację przestrzenną. Test LM dla błędu przestrzennego ma następującą formę⁹²:

$$LM_{SEM} = \frac{1}{T_1} \cdot \left(\frac{e^T W e}{\hat{\sigma}^2} \right)^2 \quad (2.20)$$

$$T_1 = tr[(W^T + W)W]$$

Drugi z kolei test LM wykonuje się dla opóźnienia przestrzennego, a wyraża go poniższy wzór⁹³:

$$LM_{SAR} = \frac{1}{T_2} \cdot \left(\frac{e^T W y}{\hat{\sigma}^2} \right)^2 \quad (2.21)$$

$$T_2 = T_1 + \frac{(WX\hat{\beta})^T M_x (WX\hat{\beta})}{\hat{\sigma}^2}$$

Oba testy mają rozkład χ^2 z jednym stopniem swobody. Zgodnie z podejściem klasycznym, bardziej istotny test wskazuje na model, który należy użyć (w przypadku, gdy żaden z testów nie jest istotny, należy pozostać na modelu liniowym)⁹⁴.

Istnieje jednak podejście odporne, które zamiast zwykłych testów mnożnika Lagrange'a proponuje użycie testów odpornych (RLM, *ang. Robust Lagrange Multiplier*). Są to testy alternatywne, które badają obecność jednego zjawiska, podczas gdy stwierdzona została obecność drugiego, a opisane są one następującymi wzorami⁹⁵:

$$RLM_{SEM} = \frac{1 - T_2 T_1}{T_1} \cdot \left(\frac{e^T W y}{\hat{\sigma}^2} - \frac{T_1 e^T W e}{T_2 \hat{\sigma}^2} \right)^2 \quad (2.22)$$

$$RLM_{SAR} = \frac{1}{T_2 - T_1} \cdot \left(\frac{e^T W y}{\hat{\sigma}^2} - \frac{e^T W e}{\hat{\sigma}^2} \right)^2 \quad (2.23)$$

⁹² Suhecki B. (red.), 2010, *Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych*, Wydawnictwo C.H.Beck, Warszawa s.296

⁹³ Suhecki B. (red.), 2010, *Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych*, Wydawnictwo C.H.Beck, Warszawa s.295

⁹⁴ Florax R. i in., 2003, *Specification Searches in Spatial Econometrics: The Relevance of Hendry's Methodology*, Regional Science and Urban Economics, nr 33, s. 5-6

⁹⁵ Suhecki B. (red.), 2010, *Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych*, Wydawnictwo C.H.Beck, Warszawa s.297

Najczęściej stosowanym podejściem jest jednak metoda hybrydowa, która łączy ze sobą obie metody. W przypadku, gdy oba testy zwykle nie odrzucają hipotezy zerowej, o zastosowanym modelu decyduje bardziej istotna wersja testu odpornego LM⁹⁶.

Weryfikacja statystyczna modelu

Następnym krokiem po zbudowaniu modelu będzie jego statystyczna weryfikacja. Metody poniżej przedstawione będą służyły do oceny modelu estymowanego metodą MNW, ponieważ, jak wspomniano wcześniej, jest to najbardziej popularna metoda estymacji.

Konieczne jest zweryfikowanie założenia o normalności reszt rozkładu, zarówno w modelu liniowym, jak i ostatecznym modelu przestrzennym. Dla regresji liniowej nie ma jednak określonego testu, który dominuje nad innymi. Z tego powodu można wykorzystać takie standardowe testy, jak: test Shapiro-Wilka, Jarque-Bera, czy Kołmogorowa⁹⁷. W modelu przestrzennym prócz wyżej wymienionych stosuje się także test LM, który bada, czy autokorelacja przestrzenna została w całości wyjaśniona przez model i nie występuje już w resztach. Test LM stanowi nieco zmienioną wersję testu mnożników Lagrange'a, który pomaga przy wyborze odpowiedniego modelu.

Do zbadania dopasowania modelu nie można wykorzystać klasycznego współczynnika determinacji, gdyż przez obecność procesów przestrzennych traci on swoje własności⁹⁸. Odpowiadającą mu miarę stanowi współczynnik pseudo- R^2 , który pokazuje udział objaśnionej wariancji w całkowitej wariancji i wyraża się następującym wzorem⁹⁹:

$$pseudo - R^2 = \frac{S^2(\hat{y})}{S^2(y)} = corr(\hat{y}, y)^2 \quad (2.24)$$

Pozostałymi miarami, które pozwalają na porównanie modeli przestrzennych, są: logarytm wiarygodności (L), kryterium informacyjne Akaike'a (AIC), kryterium Bayesowskie (BIC) oraz kryterium informacyjne Schwartz'a (SC). Im mniejsze są podane wartości, tym lepiej dopasowany jest model.

⁹⁶ Florax R. i in., *Specification Searches in Spatial Econometrics: The Relevance of Hendry's Methodology*, *Regional Science and Urban Economics*, nr 33, 2003, s. 6-7

⁹⁷ Opis testów normalności znajduje się m.in. w opracowaniu Biecka:

Biecek P., 2013, *Wybrane testy normalności*, Materiały Fundacji SmarterPoland.pl

⁹⁸ Kopczewska K., 2006, *Ekonometria i statystyka przestrzenna z wykorzystaniem programu R CRAN*, CeDeWu, Warszawa, s. 124

⁹⁹ Suhecki B. (red.), 2010, *Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych*, Wydawnictwo C.H.Beck, Warszawa s.293-294

Do diagnostyki modeli zalicza się też test istotności parametrów, które odpowiadają za interakcję przestrzenną. Powyższą rolę spełnia test Walda, który opisywany jest następującymi równaniami¹⁰⁰:

$$Wald(\rho) = \left[\frac{\hat{p}}{S(\hat{p})} \right]^2 \quad (2.25)$$

$$Wald(\lambda) = \left[\frac{\hat{\lambda}}{S(\hat{\lambda})} \right]^2 \quad (2.26)$$

Jakość modelu można sprawdzić testem LR (*ang. Likelihood Ratio*, iloraz wiarygodności), który porównuje logarytmy wiarygodności dla obu testowanych modeli (modelu SAR i modelu SEM). Wyższa wartość wskazuje na lepiej dopasowany model. Wzór na test LR prezentuje się następująco¹⁰¹:

$$LR_{SAR} = 2(\log L_{MNV(\lambda)} - \log L_{MNK}) \quad (2.27)$$

$$LR_{SEM} = 2(\log L_{MNV(\lambda)} - \log L_{MNK}) \quad (2.28)$$

Aby sprawdzić prawidłową specyfikację modelu, należy też ustalić, czy spełniana jest następująca nierówność¹⁰²:

$$Wald \geq LR \geq LM \quad (2.29)$$

W diagnostyce ważne też jest zbadanie heteroskedastyczności rozkładu reszt, zarówno w modelu liniowym, jak i w przestrzennym. Do tego celu stosuje się test Breuscha-Pagana (BP)¹⁰³. Odrzucenie hipotezy zerowej o homoskedastyczności składnika losowego może wynikać z błędnej specyfikacji modelu lub niestabilności parametrów¹⁰⁴. Prawidłowo specyfikowany model przestrzenny nie powinien cechować się heteroskedastycznością reszt.

2.5. Podsumowanie

W niniejszym rozdziale przedstawiono charakterystykę danych przestrzennych, gdzie szerzej omówiono autokorelację przestrzenną oraz wspomniano o heterogeniczności przestrzennej. Następnie pokazano narzędzia eksploracyjnej analizy danych

¹⁰⁰ Suhecki B. (red.), 2010, *Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych*, Wydawnictwo C.H.Beck, Warszawa s.298

¹⁰¹ Ibidem

¹⁰² Ibidem

¹⁰³ Opis testu Breuscha Pagana znajduje się m.in. w: Zaman A., 2000, *The Inconsistency of the Breusch-Pagan Test*, Journal of Economic and Social Research, nr 2, s.1-5

¹⁰⁴ Kopczevska K., 2006, *Ekonometria i statystyka przestrzenna z wykorzystaniem programu R CRAN*, CeDeWu, Warszawa, s. 124

przestrzennych, które pozwalają na wstępną analizę interakcji przestrzennych dla danych. Wśród nich można wyróżnić wariogram (badanie stacjonarności zmiennej), korelogram (istotność opóźnień przestrzennych), a także testy autokorelacji przestrzennej: globalnej oraz lokalnej. W kolejnej części pracy przedstawiono modele przestrzenne oraz omówiono ich estymację oraz strategię wyboru. Na koniec pokazano metody weryfikacji statystycznej zaprezentowanych modeli.

3. Analiza przestrzennego zróżnicowania przestępczości w Polsce

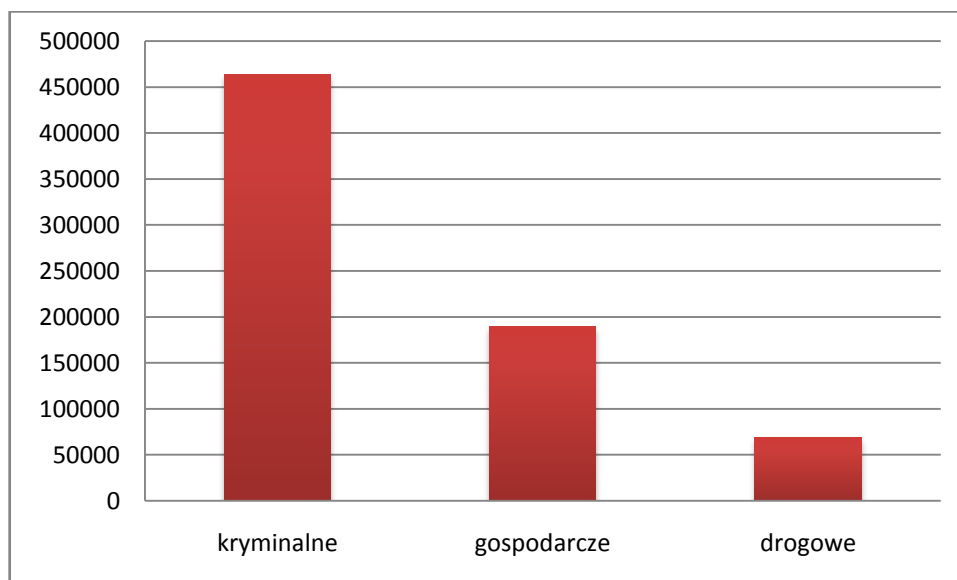
3.1 Wstęp

W niniejszym rozdziale zaprezentowana zostanie wstępna analiza przestępczości w Polsce na poziomie powiatów, w rozbiciu na trzy główne kategorie: przestępstwa na tle kryminalnym, przestępstwa gospodarcze oraz przestępstwa drogowe. Po omówieniu podstawowych statystyk opisowych i przedstawieniu badanych zmiennych na kartogramach, zostaną zastosowane metody eksploracyjnej analizy danych (ESDA) oraz wyłonione zmienne objaśniające do budowanych w dalszej części pracy modeli.

3.2 Analiza przestępczości w Polsce

Na początku warto zauważyć dysproporcję w liczbie przestępstw dla poszczególnych kategorii, co prezentuje rysunek 3.1.

Rysunek 3.1. Liczba przestępstw stwierdzonych przez policję w zakończonych postępowaniach przygotowawczych w roku 2017



Źródło: Opracowanie własne z wykorzystaniem arkusza kalkulacyjnego Excel.

Z powyższego wykresu wynika, że przestępstw kryminalnych jest prawie dwa razy więcej niż zsumowanych przestępstw z pozostałych kategorii. Tak jak było to

wspomniane w rozdziale 1.4., ma na to wpływ powszechność przestępstw przeciwko mieniu, które stanowią ponad połowę przypadków w swojej klasie.

Jeszcze lepszą prezentację badanego zjawiska stanowią statystyki opisowe dla 380 polskich powiatów, wykonane z osobna dla każdej z kategorii przestępstw. Zostały one stworzone na podstawie danych z Banku Danych Lokalnych GUS, dotyczących przestępczości stwierdzonej przez policję w zakończonych postępowaniach przygotowawczych¹⁰⁵. Aby umożliwić porównanie między małymi oraz dużymi jednostkami, wszystkie przestępstwa przeliczane są na 1000 ludności. Statystyki opisowe zaprezentowane są w poniższej tabeli.

Tabela 3.1. Wybrane statystyki opisowe dla przestępstw w roku 2017

Przestępstwa	Statystyki opisowe								
	Min.	Q1	Mediana	Średnia	Q3	Max.	Odch. st.	Kurtoza	Skośność
Kryminalne	3,35	7,23	9,48	10,44	12,63	35,49	4,68	3,25	1,43
Gospodarcze	0,57	1,81	2,68	4,58	4,79	139,29	8,66	160,5	11,17
Drogowe	0,96	1,58	1,93	1,97	2,28	4,24	0,52	0,85	0,72

Źródło: Opracowanie własne z wykorzystaniem arkusza kalkulacyjnego Excel.

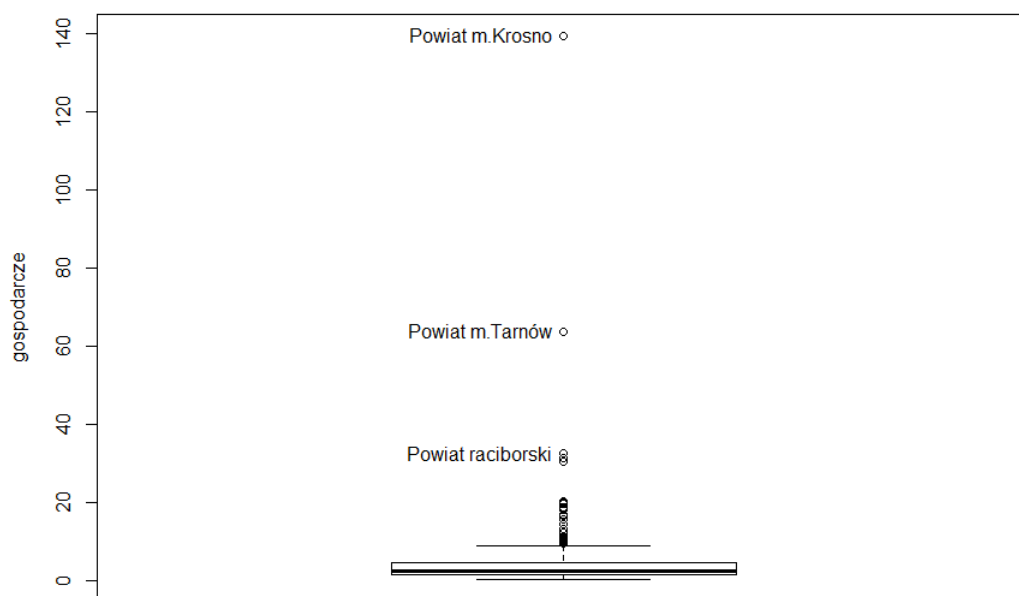
Średnia potwierdza poprzednio wykonany wykres. Przestępstwa kryminalne znacząco odstają od średniej dla pozostałych kategorii.

Mediana wskazuje na równomierny rozkład cechy w dwóch przypadkach. Dla przestępstw gospodarczych odstaje ona sporo od średniej, co może oznaczać problem z obserwacjami odstającymi, czemu należałoby się przyjrzeć.

Powyższe przypuszczenie potwierdzają pozostałe statystyki. Odchylenie standardowe w przypadku przestępstw gospodarczych przewyższa ich średnią. Kurtoza, która jest miarą koncentracji wyników, przyjmuje tym większe wartości, im większe skupienie cechy wokół średniej. Dla wszystkich rozpatrywanych kategorii przyjmuje ona wartości dodatnie, co świadczy o tendencji do gromadzenia się obserwacji przy średniej. Przestępstwa gospodarcze mają bardzo wysoką wartość, co spowodowane jest dużą liczbą obserwacji odstających. Poniżej znajduje się wykres pudełkowy, który pozwoli na wyodrębnienie powiatów o bardzo wysokim poziomie przestępczości gospodarczej.

¹⁰⁵ Dokładny opis przestępczości stwierdzonej został zaprezentowany w rozdziale 1.3.

Rysunek 3.2. Wykres pudełkowy dla przestępstw gospodarczych w roku 2017



Źródło: Opracowanie własne z wykorzystaniem programu R.

Zauważyć można, że najbardziej odstającym powiatem jest miasto Krosno (przy średniej 4,58 osiąga ono wartość 139,29, co stanowi ponad 30-krotne przekroczenie średniej). Nie jest to jednak błąd w danych, a odchylenie od normy z poprzednich lat, gdzie liczba nie przekraczała ośmiu przestępstw na tysiąc mieszkańców. Wynika to z prowadzenia dwóch dużych postępowań, które przez kilka lat pozostawały niezauważone¹⁰⁶. Powiat miasta Tarnów również zdecydowanie odstaje od reszty obserwacji, w tym przypadku jest to jednak wartość, która pozostaje zbliżona do ostatnich obserwacji i nie jest spowodowana żadnym nagłym zdarzeniem (jak to miało miejsce w przypadku Krosna). W celu uniknięcia problemów ze stacjonarnością danych oraz poprawności tworzonych modeli, w pracy zdecydowano się na usunięcie obu wspomnianych wyżej powiatów ze zbioru dla przestępczości gospodarczej.

Jeśli chodzi o pozostałe statystyki opisowe, skośność, będąca miarą asymetrii obserwowanych wyników, dla wszystkich kategorii przyjmuje wartości większe od zera. Świadczy to o prawostronnej asymetrii rozkładu, zatem większej niż dla rozkładu normalnego liczbie powiatów o podwyższonym poziomie przestępczości.

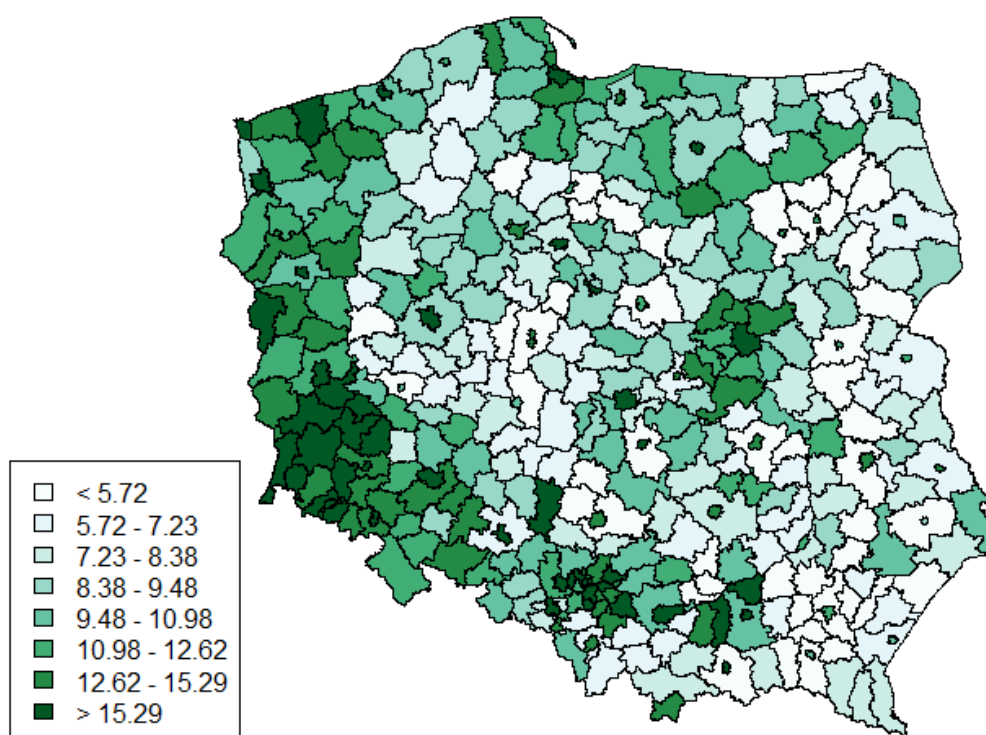
Na podstawie powyższych statystyk opisowych można wysnuć przypuszczenie o problemie z normalnością rozkładu dla wszystkich z badanych kategorii, zwłaszcza dla przestępstw gospodarczych.

¹⁰⁶<https://krosno112.pl/aktualnosci/inne/item/5285-najniebezpieczniejsze-miasto-w-polsce-to-krosno> (dostęp: 17.06.2019)

Analiza graficzna

Poza podstawowymi statystykami opisowymi dla przestępczości, ciekawych wniosków może też dostarczyć analiza kartogramów. Dzięki nim zbadać można rozkład badanej cechy w ujęciu przestrzennym, co pozwoli na wstępne określenie możliwej autokorelacji przestrzennej. Na poniższych rysunkach zauważyć można trzy kartogramy, po jednej dla każdej z kategorii przestępczości. Wartości zostały naniesione na mapę konturową powiatów w Polsce, która została pobrana ze strony GIS¹⁰⁷.

Rysunek 3.3. Kartogram liczby przestępstw kryminalnych na 1000 mieszkańców dla powiatów Polski w roku 2017



Źródło: Opracowanie własne z wykorzystaniem programu R.

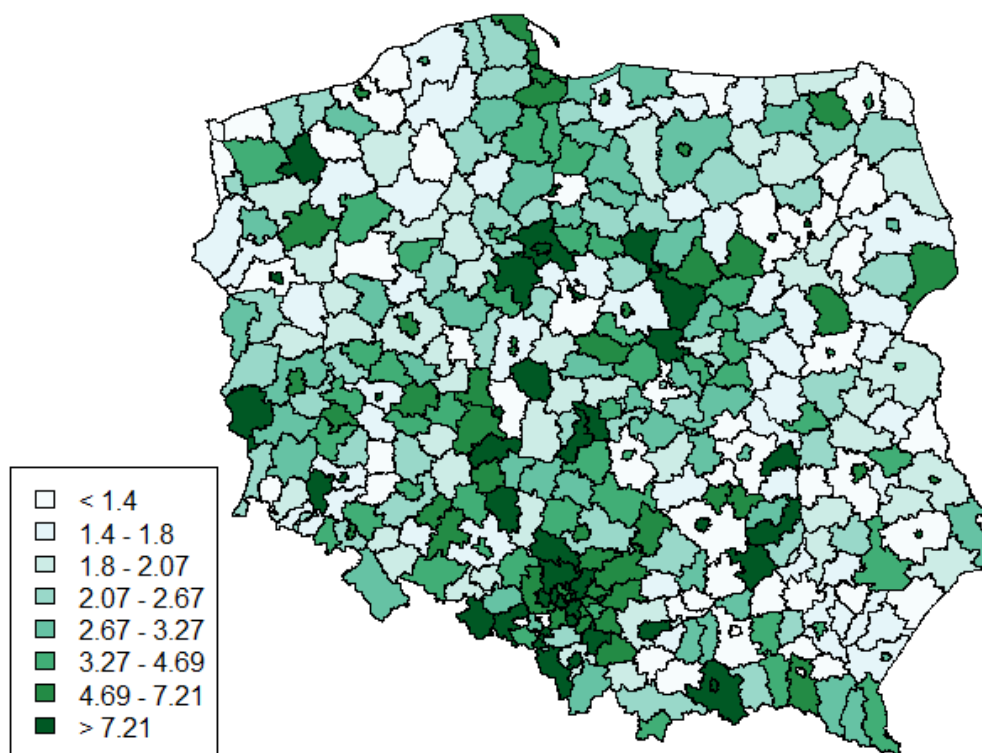
Największy poziom przestępczości odnotowano dla powiatów położonych niedaleko granicy polsko-niemieckiej (obszary leżące w granicach województw: dolnośląskiego, lubuskiego oraz zachodniopomorskiego). Poza tym, tworzenie się pewnych obszarów o podwyższonym poziomie niebezpieczeństwa można zauważyć na terenie Górnośląskiego Okręgu Przemysłowego, a także wokół Warszawy oraz na terenie województwa warmińsko-mazurskiego. Czołówkę powiatów o największej liczbie przestępstw kryminalnych na 1000 osób stanowią miasta na prawach powiatu położone

¹⁰⁷ <https://gis-support.pl/granice-administracyjne/> (dostęp: 20.03.2019)

na Górnym i Dolnym Śląsku: Siemianowice Śląskie (35,49), Jelenia Góra (27,6), Tychy (26,41), Katowice (26,27) oraz Legnica (26,22). Najbardziej bezpieczne miejsca to z kolei następujące powiaty, leżące głównie w województwie podkarpackim oraz wielkopolskim: powiat kaliski (3,35), brzozowski (3,44), słupecki (3,56), strzyżowski (3,58) i grudziądzki (3,74). Miasto na prawach powiatu, które jest najmniej zagrożone przestępczością, to Zamość (9,03).

Przyczyn istnienia klastrów przestępczości kryminalnej przy zachodniej granicy Polski można doszukiwać się na kilku płaszczyznach. Czapliński¹⁰⁸ wskazuje na brak stabilności struktury społecznej, która jest pokłosiem przesiedleń po II wojny światowej. Nieco inaczej zjawisko tłumaczy Rau¹⁰⁹ oraz Bogacka¹¹⁰, którzy jako przyczynę podają ruch tranzytowy, ułatwiający sprawcy anonimowość.

Rysunek 3.4. Kartogram liczby przestępstw gospodarczych na 1000 mieszkańców dla powiatów Polski w roku 2017



Źródło: Opracowanie własne z wykorzystaniem programu R.

¹⁰⁸ <https://wiadomosci.onet.pl/kraj/rzeczpospolita-wzrasta-przestepczosc-najgorzej-na-zachodzie-polski/rc7rj> (dostęp: 16.06.2019)

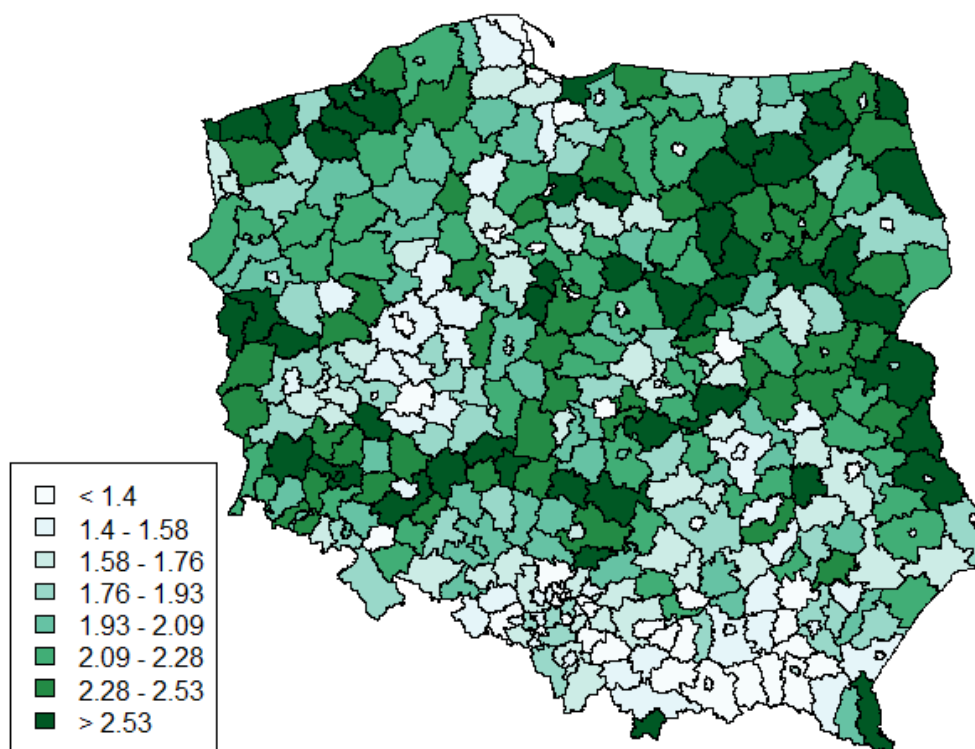
¹⁰⁹ Ibidem

¹¹⁰ Bogacka E., 2012: *Struktura przestrzenna i czynniki przestępczości na obszarze nadgranicznym Polski z Niemcami*. Bogucki Wydawnictwo Naukowe, Poznań, s.47

Dla przestępstw gospodarczych trudno jest o wyróżnienie skupiska powiatów o podwyższonym poziomie zagrożenia. Można wskazać jedynie teren Górnośląskiego Okręgu Przemysłowego i jego okolice, jednak powiaty o największym poziomie badanego zjawiska niekoniecznie leżą na Górnym Śląsku. Powiaty o największej liczbie przestępstw gospodarczych na tysiąc osób, po opisanym w poprzedniej części rozdziału usunięciu Krosna i Tarnowa, to: powiat raciborski (32,58), kluczborski (31,40), opatowski (30,67), miasta Tychy (20,41) i złotoryjski (20,04). Po przeciwnej stronie rankingu stoją powiaty: piotrkowski (0,57), kolbuszowski (0,67), grudziądzki (0,67), skierniewicki (0,71) i proszowicki (0,71).

Wysoki poziom przestępczości gospodarczej na terenie GOP można tłumaczyć wysokim uprzemysłowieniem tamtego rejonu. W rezultacie istnieje wiele firm, wśród których, co przyznaje sama policja¹¹¹, znacznie łatwiej jest o anonimowość.

Rysunek 3.5. Kartogram liczby przestępstw drogowych na 1000 mieszkańców dla powiatów Polski w roku 2017



Źródło: Opracowanie własne z wykorzystaniem programu R.

Jeśli chodzi o przestępczość drogową, na mapie wyraźnie zaznacza się podział między terenami wyżinnymi i kotlinami a terenami nizinnymi. Obszary o wysokim

¹¹¹<https://dziennikzachodni.pl/przestepczosc-w-woj-slaskim-maleje-ale-korupcja-to-nasza-specjalnosc-statystyki-policji/ar/3723630> (dostęp: 17.06.2019)

zagrożeniu, gdzie odnaleźć można większe skupiska powiatów o podwyższonym poziomie niebezpieczeństwa, to głównie tereny wiejskie w północno-wschodniej Polsce oraz poziomy pas bliżej południowej części kraju. Na czele rankingu stoją powiaty: grójecki (4,24), świdwiński (3,74), ślubicki (3,49), miasta Legnica (3,43) i ostrzeszowski (3,39). Po drugiej stronie są powiaty miast: Bytom (0,96), Bydgoszcz (0,98), Toruń (1,03), Poznań (1,05), Gdynia (1,06).

Duże znaczenie w wyjaśnianiu wysokiej przestępczości drogowej może mieć stan dróg. Według mapy nawierzchni polskich dróg¹¹², północno-wschodnia Polska charakteryzuje się większym niż w reszcie kraju udziałem dróg o pogorszonej jakości. Dodatkowo, w 2017 roku, 82% wszystkich przestępstw drogowych polegało na prowadzeniu pojazdu w stanie nietrzeźwości lub pod wpływem środka odurzającego¹¹³. Na terenach wiejskich trudniej jest skontrolować trzeźwość kierowcy, więc niewykluczone, że także tam należy szukać przyczyny.

3.3 Analiza przestrzenna przestępczości

Wariogramy

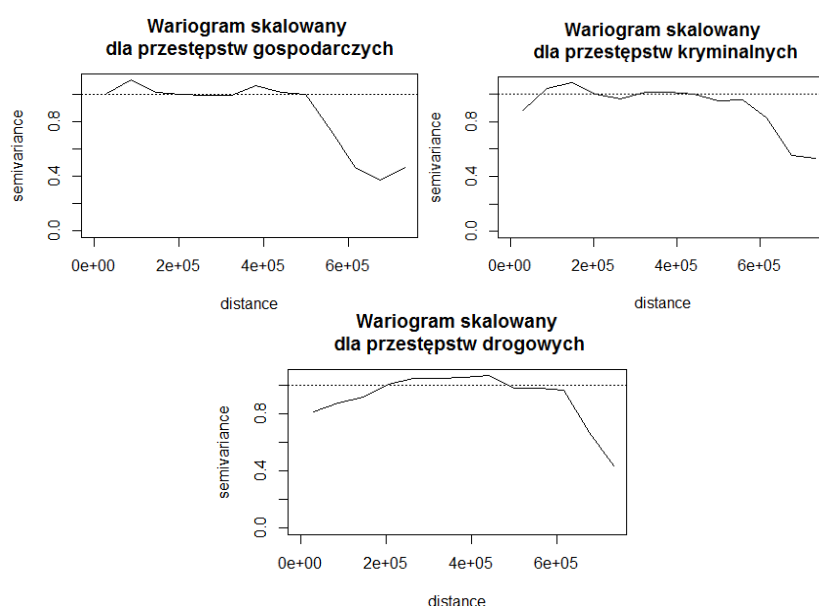
Pierwszym narzędziem eksploracyjnej analizy danych, które zostanie zastosowane w celu analizy przestrzennej przestępczości, jest wariogram. Pozwoli on na weryfikację założenia o stacjonarności rozpatrywanych zmiennych, w celu przeprowadzenia rzetelnych testów na autokorelację przestrzenną¹¹⁴. Wykonane wariogramy widoczne są na rysunku 3.6. Zostały one przeskalowane, żeby umożliwić ich porównanie między sobą.

¹¹² <http://ssc.siskom.waw.pl/mapa-nawierzchni/mapa-nawierzchnia.png> (dostęp: 17.06.2019)

¹¹³ Obliczenia własne na podstawie <http://www.statystyka.policja.pl/st/kodeks-karny/przestepstwa-przeciwko-2> (dostęp: 17.06.2019)

¹¹⁴ Wariogram został szczegółowo opisany w rozdziale 2.3.

Rysunek 3.6. Wykresy skalowanych wariogramów dla przestępstw w roku 2017



Źródło: Opracowanie własne z wykorzystaniem programu R.

Dla wszystkich kategorii przestępstw zauważyć można obecność punktu o maksymalnej wartości funkcji semiwariancji (w przypadku przestępstw gospodarczych semiwariogram zaczyna na końcu rosnać, jednak następuje to znacznie poniżej progu, więc przyjąć można stacjonarność zmiennej). Oznacza to, że do pewnej odległości (widocznej na osi OX) występuje autokorelacja przestrzenna. W tym przypadku ma ona znaczniejszy zasięg dla przestępstw drogowych, gdzie semiwariogram cechuje się większym zakresem. We wszystkich kategoriach nie można mówić o wysokim efekcie samorodka, który świadczyłby o znacząco różnych wartościach dla bliskich sobie lokalizacji. Punkt początkowy wykresu znajduje się poniżej poziomu wariancji próbkowej, oznaczonej przerywaną linią, więc wartość osiągnięta przez punkt jest relatywnie mała.

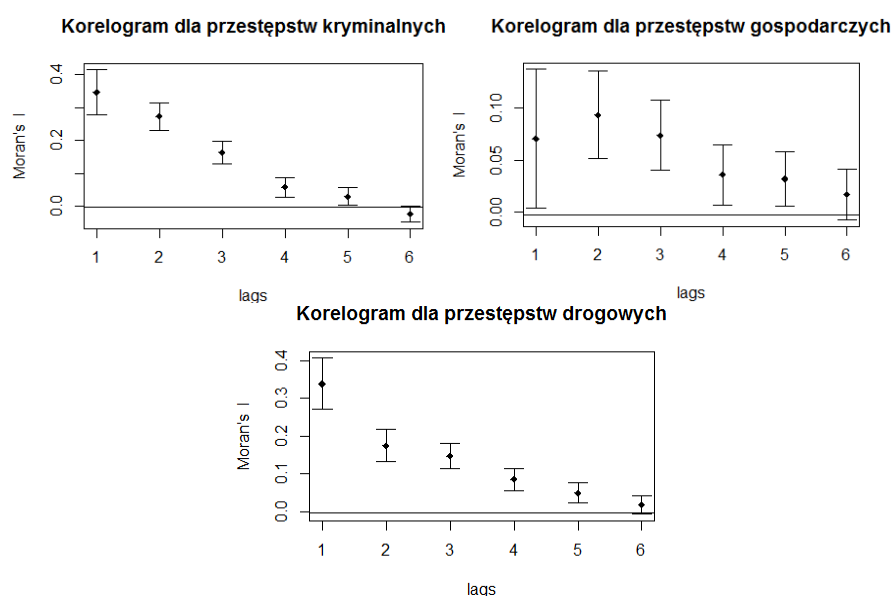
Korelogramy

Aby dokładniej przyjrzeć się autokorelacji przestrzennej, której obecność zasygnalizowano przy badaniu wariogramów, należy posłużyć się korelogramem. Pozwoli on na określenie, ile opóźnień przestrzennych jest istotnych dla każdej z kategorii przestępstw oraz jak silna jest ta zależność¹¹⁵. Widoczne na rysunku 3.7. wykresy biorą pod uwagę 6 opóźnień przestrzennych, co stanowi standardową liczbę, braną pod uwagę przy wykonywaniu korelogramu. Ważnym parametrem przy konstrukcji

¹¹⁵ Korelogram został szczegółowo opisany w rozdziale 2.3.

korelogramu jest także macierz sąsiedztwa. W programie R domyślnie stosuje się macierz pierwszego rzędu standaryzowaną rzędami, czyli sąsiadami danego regionu są bezpośrednio przylegające do niego jednostki¹¹⁶. Postanowiono pozostać przy tym wyborze, gdyż jest on najbardziej intuicyjny dla polskich powiatów. Zostanie on również zastosowany w dalszej części pracy, ponieważ macierze polegające na odległości w promieniu zadanej liczby kilometrów bądź liczbie k-sąsiadów nie wydają się być odpowiednie dla nieregularnych i gęsto rozlokowanych kształtów powiatów.

Rysunek 3.7. Wykresy korelogramów dla przestępstw w roku 2017



Źródło: Opracowanie własne z wykorzystaniem programu R.

W przypadku przestępstw kryminalnych oraz gospodarczych można zauważyć spadek wartości statystyki I Morana wraz z kolejnymi opóźnieniami przestrzennymi. Oznacza to, że istnieje kilka wiodących regionów, które wraz ze zwiększaniem się od nich odległości tracą swoją mocną pozycję. Wartości są na dość wysokim poziomie, co sugeruje słuszność brania pod uwagę powiatów sąsiednich. W przypadku przestępstw kryminalnych statystyka przyjmuje na końcu wartość ujemną, co oznacza, że przy tak dużym opóźnieniu przestrzennym występuje autokorelacja ujemna i powiaty tworzą „szachownicę”. Jest to jednak niewielki efekt, gdyż jego wartość oscyluje wokół zera (podobnie, jak dla wszystkich badanych opóźnień przestrzennych dla przestępczości gospodarczej). W tej kategorii wartość statystyki początkowo rośnie, a potem zaczyna maleć. Nie można tu jednak mówić o tworzeniu makroregionów, gdyż wzrost powinien

¹¹⁶ Macierze wag zostały szczegółowo opisane w rozdziale 2.3.

trwać chociaż do drugiego lub trzeciego opóźnienia przestrzennego¹¹⁷. Aby dokładniej zbadać tą sytuację oraz upewnić się co do interpretacji dla pozostałych kategorii przestępstw, należy przeprowadzić test istotności opóźnień przestrzennych. Zostanie on wykonany z użyciem korekty Bonferroniego, przy wariancie podwojonej wartości p-value.

Tabela 3.2. Test istotności opóźnień przestrzennych dla korelogramów przestępstw w roku 2017

Opóźnienie przestrzenne	Kryminalne		Gospodarcze		Drogowe	
	Moran I	p-value	Moran I	p-value	Moran I	p-value
1	0,35	0,00	0,07	0,05	0,34	0,00
2	0,27	0,00	0,09	0,00	0,17	0,00
3	0,16	0,00	0,07	0,00	0,15	0,00
4	0,06	0,00	0,04	0,01	0,09	0,00
5	0,03	0,03	0,03	0,01	0,05	0,00
6	-0,02	0,20	0,02	0,22	0,02	0,10

Źródło: Opracowanie własne z wykorzystaniem programu R oraz arkusza kalkulacyjnego Excel.

Zarówno dla przestępstw kryminalnych jak i drogowych istotne jest pierwsze pięć opóźnień przestrzennych. Nie można zatem mówić o autokorelacji ujemnej dla szóstego opóźnienia dla pierwszej z kategorii. W przypadku przestępstw gospodarczych występuje luka w istotności, co mogłoby wskazywać na niestacjonarność badanej zmiennej, jednak taka możliwość została wcześniej odrzucona.

Testy globalne

Pomimo wcześniejszych oznak istnienia autokorelacji przestrzennej, decydującymi testami na jej obecność będą testy globalne oraz lokalne.

Pierwszym krokiem są globalne miary, czyli statystyka Morana I oraz Geary'ego C¹¹⁸. Wyniki testów przedstawia tabela 3.3.

¹¹⁷ Kopczewska K., 2006, *Ekonometria i statystyka przestrzenna z wykorzystaniem programu R CRAN*, CeDeWu, Warszawa, s. 119

¹¹⁸ Testy globalne zostały szczegółowo omówione w rozdziale 2.3.

Tabela 3.3. Globalne statystyki autokorelacji przestrzennej dla przestępstw w roku 2017

Przestępstwa	Moran I		Geary C	
	statystyka	p-value	statystyka	p-value
Kryminalne	0,35	0,00	0,57	0,00
Gospodarcze	0,07	0,01	0,90	0,06
Drogowe	0,34	0,00	0,65	0,00

Źródło: Opracowanie własne z wykorzystaniem programu R oraz arkusza kalkulacyjnego Excel.

Oba testy globalne są jednoznaczne dla przestępstw kryminalnych oraz drogowych, wskazując na ich dodatnią autokorelację. Z racji tego, że statystykę Morana I można interpretować jako współczynnik korelacji, dla wskazanych wyżej kategorii można mówić o przeciętnej korelacji wartości cechy w powiecie z wartościami w powiatach sąsiednich. Zmienność obu zmiennych w 12% ($0,35^2 \approx 0,34^2 \approx 0,12$) można tłumaczyć lokalizacją¹¹⁹. Z kolei w przypadku przestępstw gospodarczych jedynie test Morana I wykazał istotność dodatniej autokorelacji. Zastanawiająca jest duża różnica między wartościami statystyk, chociaż należy pamiętać o odmiennych zakresach przyjmowanych przez obie statystyki: [-1,1] dla Morana I oraz [0,2] dla Geary'ego C.

Aby zdecydować o obecności autokorelacji (lub jej braku) dla przestępstw gospodarczych, wykonano test join-count, który w nieco innym kontekście mierzy autokorelację przestrzenną. Statystyka ta wymaga, żeby podzielić zbiór danych na parę grup wartości. W niniejszej pracy zostało to wykonane poprzez podział zmiennej za pomocą kwartyli na cztery równoliczne grupy, które otrzymały nazwy: „mało”, „średnio”, „dużo”, „b.dużo”. Przyjęto poziom istotności na poziomie 0,05, więc wartość statystyki z jest istotna, gdy jej wartość bezwzględna przekracza liczbę 1,64. Wyniki testu join-count przedstawia tabela 3.4.

¹¹⁹ Kopczewska K., 2006, *Ekonometria i statystyka przestrzenna z wykorzystaniem programu R CRAN*, CeDeWu, Warszawa, s. 122

Tabela 3.4. Statystyki join-count dla przestępstw gospodarczych w roku 2017

Join-count	Przestępstwa gospodarcze	
	statystyka	z-value
mało:mało	17,34	4,12
średnio:średnio	12,79	0,71
dużo:dużo	10,22	-0,86
b.dużo:b.dużo	16,32	3,35
średnio:mało	29,17	2,72
dużo:mało	25,12	0,88
dużo:średnio	25,55	1,11
b.dużo:mało	20,32	-1,88
b.dużo:średnio	16,28	-3,98
b.dużo:dużo	15,90	-3,95
Jtot	132,34	-3,52

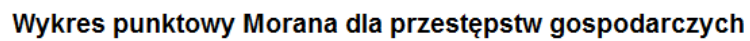
Źródło: Opracowanie własne z wykorzystaniem programu R oraz arkusza kalkulacyjnego Excel.

Z testu join-count wynika, że test Geary'ego C mógł zostać odrzucony przez brak autokorelacji przestrzennej wewnątrz dwóch grup kolorystycznych (średnio:średnio i dużo:dużo). Widać jednak, że wewnątrz pozostałych jednolitych grup, a także w znaczącej części między nimi (Jtot określa istotność wszystkich połączeń międzygrupowych), można zaobserwować autokorelację przestrzenną. Istotne jest to zwłaszcza dla grupy „b.duże”, gdyż w każdym przypadku statystyka z cechuje się istotnością. Z powyższych powodów zdecydowano się przychylić do testu Morana I, który wskazał na obecność autokorelacji przestrzennej dla przestępstw gospodarczych.

Graficzną reprezentacją statystyki Morana I jest wykres punktowy Morana. Dzięki niemu możliwe będzie zaobserwowanie odstających obserwacji¹²⁰. Wykres dla każdej z kategorii zostanie zaprezentowany na rysunku 3.8.

¹²⁰ Wykres punktowy Morana został szczegółowo omówiony w rozdziale 2.3.

Wykres punktowy Morana dla przestępstw kryminalnych



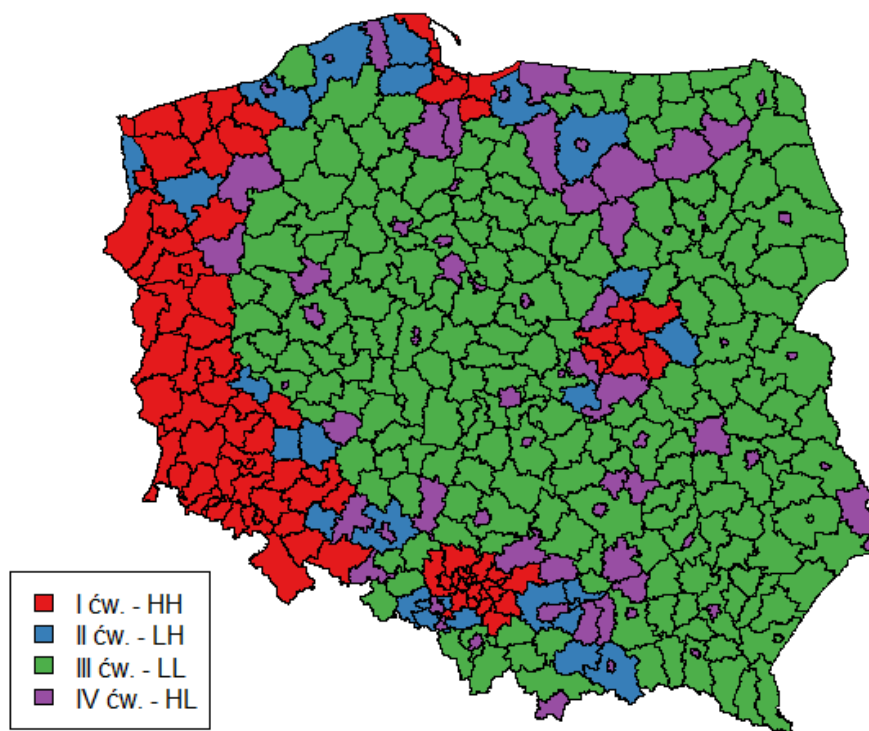
We wszystkich kategoriach przestępstw pojawia się podobna liczba obserwacji odstających, oznaczonych na wykresie przekreślonymi kwadratami. Z powodu udzielenia przez testy pozytywnych wyników oraz sporej liczby outlierów, zdecydowano się nie usuwać żadnego z powiatów.

Można też zauważyć, że zgodnie z globalnymi testami na autokorelację większość obserwacji położonych jest w pierwszej i trzeciej ćwiartce. Zależność ta wydaje się zachodzić najslabiej dla przestępstw gospodarczych, gdzie punktów położonych w pozostałych dwóch ćwiartkach jest relatywnie dużo.

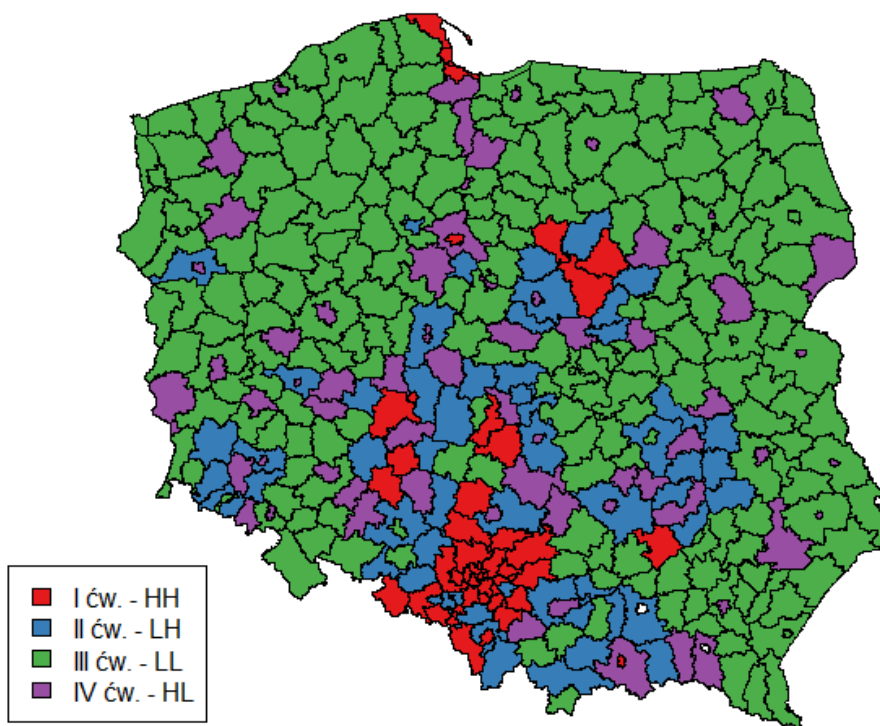
Ciekawym narzędziem analizy jest też utworzenie kartogramów na podstawie przynależności punktów do ćwiartek wykresu Morana. Powinny one przypominać zwykłe kartogramy, jednak dzięki zmniejszeniu liczby kolorów będą one bardziej czytelne i pozwolą na łatwiejsze uchwycenie tworzących się klastrów podobnych wartości.

Rysunek 3.9. Kartogramy przynależności powiatów do ćwiartek wykresu punkowego Morana dla przestępstw w roku 2017

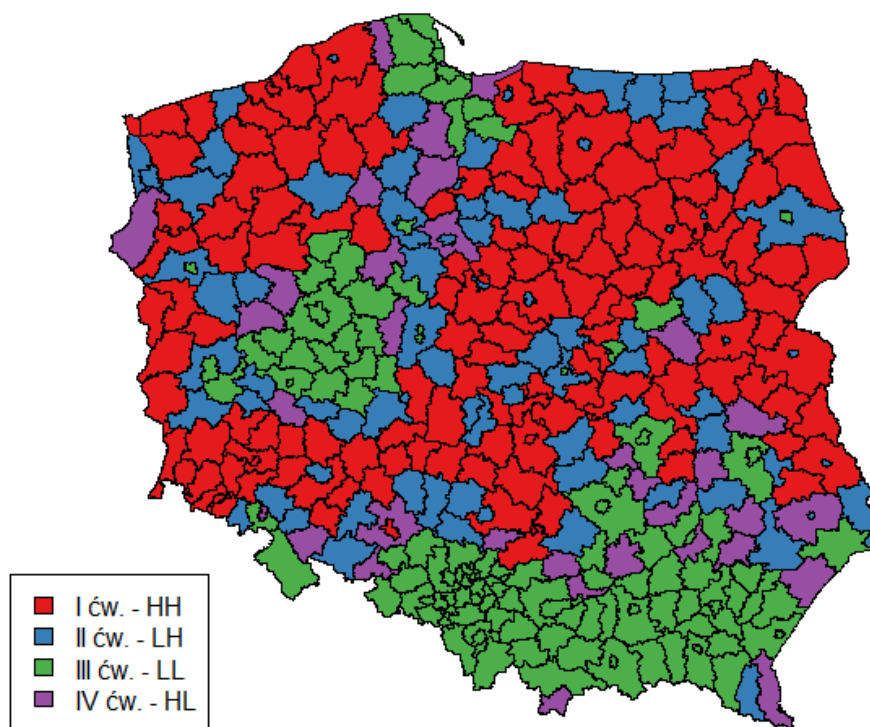
Przynależność przestępstw kryminalnych do ćwiartek wykresu Morana



Przynależność przestępstw gospodarczych do ćwiartek wykresu Morana



Przynależność przestępstw drogowych do ćwiartek wykresu Morana



Źródło: Opracowanie własne z wykorzystaniem programu R.

Wyniki pokrywają się z przedstawionymi wcześniej kartogramami na danych liczbowych. Są jednak przypadki, gdy regiony posądzone o tworzenie się klastrów wysokich wartości ich nie tworzą, a także bywają odwrotne sytuacje. Dla przestępstw kryminalnych na terenie województwa warmińsko-mazurskiego nie tworzą się jednak skupiska o wysokiej przestępczości, natomiast widać je w okolicach Trójmiasta. Region ten okazuje się mieć także znaczenie przy przestępstwach gospodarczych, co nie było widoczne początkowo na kartogramie. W przypadku przestępstw drogowych zaskoczeniem okazuje się być skala tego zjawiska. Jest ona znacznie większa niż wcześniej sugerował to kartogram. Większa część terenów nizinnych tworzy klastry, co może okazać się ciekawe przy dalszej analizie.

Testy lokalne

Na koniec zaprezentowane zostaną wyniki testu lokalnego Morana, który to bardziej szczegółowo przygląda się autokorelacji, będąc w stanie wskazać pojedyncze powiaty odpowiedzialne za odrzucenie hipotezy zerowej dla testów globalnych.

Najgorszy wynik jest w przypadku przestępstw gospodarczych: 14 powiatów wykazuje istotną autokorelację ujemną, a 16 autokorelację dodatnią. Lepszą sytuację

można zaobserwować dla przestępstw kryminalnych: 67 powiatów o dodatniej autokorelacji i 6 o ujemnej autokorelacji. Najlepszymi wynikami cechuje się jednak przestępczość drogowa: 80 powiatów o dodatniej autokorelacji i 2 o ujemnej autokorelacji. Uzyskane wyniki potwierdzają zaprezentowane wcześniej testy globalne oraz kartogramy przynależności do ćwiartek wykresu punktowego Morana.

3.4. Wybór zmiennych

Kluczowym zagadnieniem przy budowie modeli jest ustalenie zbioru zmiennych objaśniających. We wszystkich modelach badany czynnikiem jest lokalizacja, jednak oprócz tego potrzeba wyłonienia jak największej liczby pozostałych zmiennych, które mogą okazać się istotne, oraz są ogólnie dostępne w skali powiatów. Kierując się tymi oboma kryteriami, za pomocą przeglądu literatury dokonano we wcześniejszej części pracy oraz subiektywnej oceny autora wybrano dla każdej z trzech badanych zmiennych objaśnianych odpowiedni zbiór zmiennych.

Zmienne objaśniane, przedstawiane już w poprzedniej części pracy, to:

- **przestępstwa kryminalne stwierdzone przez policję w zakończonych postępowaniach przygotowawczych w roku 2017 na 1000 osób (*Krym*),**
- **przestępstwa gospodarcze stwierdzone przez policję w zakończonych postępowaniach przygotowawczych w roku 2017 na 1000 osób (*Gosp*),**
- **przestępstwa drogowe stwierdzone przez policję w zakończonych postępowaniach przygotowawczych w roku 2017 na 1000 osób (*Drog*).**

Dla wszystkich z powyższych zmiennych jest brane pod uwagę 12 wspólnych zmiennych objaśniających, które zostały wybrane przy pomocy przeglądu literatury. Mając na uwadze teorię działań rutynowych, według której do popełnienia przestępstwa wymagany jest m.in. brak osoby, która mogłaby zapobiec przestępstwu, wzięto pod uwagę **gęstość zaludnienia** (*Gestosc_zaludnienia*). Jej wysoki poziom, zgodnie z powyższą teorią, powinien przyczynić się do spadku przestępczości, ponieważ wzrosłaby liczba świadków. Z kolei kwalifikując wskazaną zmienną wg teorii działań rutynowych jako odpowiedni cel, można spodziewać się wzrostu badanego zjawiska, bo większe zagęszczenie ludności to większa liczba potencjalnych ofiar. Silnie powiązana z gęstością zaludnienia jest inna zmienna, a mianowicie **współczynnik urbanizacji** (*Urbanizacja*), który wskazuje na odsetek ludzi mieszkających w miastach. Większy poziom tego współczynnika przekłada się na większe zagęszczenie ludności. W pracy

Kądziołki¹²¹ wykazano, że urbanizacja najmocniej ze wszystkich czynników oddziałuje na przestępczość.

Wspomniani wyżej odpowiedni cel mogą, zwłaszcza dla złodziei, stanowić słabo strzeżone mieszkania. Nadzór nad własną wartością jest lepszy, gdy dom zamieszkuje więcej osób. Przestępczość powinna być wtedy na niższym poziomie. Stąd obecność zmiennej dotyczącej **przeciętnej liczby mieszkańców na jedno mieszkanie** (*Mieszkania*). Podobny wniosek wysnuła w swoim artykule Kądziołka¹²², która wykazała procent gospodarstw jednoosobowych jako stymulantę dla ogółu przestępstw.

Analizując prace z zakresu teorii społecznej dezorganizacji można zauważyć znaczącą rolę migracji. W swojej książce, tłumaczącej przestępczość na zachodzie kraju, Bogacka¹²³ wykazała jej destymulujący charakter, natomiast Kądziołka¹²⁴ udowodniła jej stymulujący charakter biorąc pod uwagę ogólną przestępczość na terenie całego kraju. Można przypuszczać, że zwiększony poziom migracji przyczynia się do zawirowań w strukturze społecznej, a napływ nowej, nieobeznanej lokalnie ludności to dobry znak dla przestępców, zwłaszcza tych specjalizujących się w kradzieżach domostw. W tym celu w modelu zawarta została zmienna dotycząca **współczynnika migracji** (*Wspolczynnik_migracji*), która jest saldem migracji (różnica między liczbą zameldowań i wymeldowań) na pobyty stały ludności na 10 tys. osób. Podobne zawirowania może również tworzyć ruch turystyczny, który powoduje sporą rotację ludności. Jak pokazali Shaw i McKay¹²⁵, jest to jeden z czynników mogących determinować przestępczość. Przy dużej rotacji ludności łatwiej jest o anonimowość sprawcy przestępstwa, więc może mieć ona działanie stymulujące. Aby to zbadać, w modelu zawarto **miejsca noclegowe na 1000 osób** (*Nocleg*).

Istotną cechą może okazać się także inna zmienna wywodząca się z teorii społecznej dezorganizacji, a mianowicie **liczba rozwodów na 1000 ludności** (*Rozwody_na_1000*), zaproponowana początkowo przez Sampsona i Grovesa¹²⁶.

¹²¹ Kądziołka K., 2016, *Determinanty przestępczości w Polsce. Analiza zależności z wykorzystaniem drzew regresyjnych*, *Ekonomia. Rynek, gospodarka, społeczeństwo*, nr 45, s. 53-81

¹²² Kądziołka K., 2016, *Analysis of the crime rate in Poland in spatial and temporal terms*, *Central and Eastern European Journal of Management and Economics*, nr 1, s. 81-96

¹²³ Bogacka E., 2012: *Struktura przestrzenna i czynniki przestępczości na obszarze nadgranicznym Polski z Niemcami*, Bogucki Wydawnictwo Naukowe, Poznań

¹²⁴ Kądziołka K., 2016, *Analysis of the crime rate in Poland in spatial and temporal terms*, *Central and Eastern European Journal of Management and Economics*, nr 1, s. 81-96

¹²⁵ Shaw C., McKay H., 1942, *Juvenile Delinquency and Urban Areas*, University of Chicago Press, Chicago

¹²⁶ Groves W.B., Sampson R.J., 1989, *Community Structure and Crime: Testing Social-Disorganization Theory*, *American Journal of Sociology*, nr 94, s. 774-802.

Podobnie, jak współczynnik migracji, może przyczyniać się do osłabienia więzów społeczności, co jest szczególnie odczuwalne przez dzieci pozostające pod opieką wyłącznie jednego z rodziców.

Kolejną zmienną braną pod uwagę jest **procent rodzin korzystających z pomocy społecznej** (*Pomoc_spoeczna*). Zmienna ta była brana pod uwagę w pracy Kądziołki¹²⁷, jednak nie znalazła się w ostatecznym modelu przestrzennym. Wybór tej zmiennej podyktowany jest ekonomiczną teorią przestępczości Beckera, według której przestępca posługuje się rachunkiem zysków i strat, aby zdecydować o złamaniu prawa. Procent rodzin korzystających z pomocy społecznej może okazać się dobrym odzwierciedleniem sytuacji materialnej ludzi zamieszkujących dany powiat, gdyż w pracach naukowych wykazywano stymulujący charakter odsetka ludności żyjącej poniżej granicy ubóstwa. Inną ciekawą zmienną, która z pewnością wpływa na wspomniany rachunek zysków i strat, to **przeciętne miesięczne wynagrodzenie brutto** (*Wynagrodzenie*). W swojej książce Bogacka¹²⁸ wykazała jej stymulujący charakter dla zachodnich powiatów Polski. Poza tym, jest to też zmienna bardzo intuicyjna, ponieważ pieniądze to częsty czynnik, przez który dochodzi do popełnienia przestępstwa. Ekonomiczna teoria przestępczości podyktowała też obecność zmiennej dotyczącej bezrobocia. Brak pracy oznacza brak zarobków, co może budzić frustrację, a także skłaniać bezrobotnego do poszukiwania nielegalnych źródeł zarobku. W pracach badawczych stosuje się różne zmienne dotyczące bezrobocia, np. w badaniu dla włoskich prowincji¹²⁹ było to bezrobocie wśród mężczyzn w wieku 25-29 lat. W niniejszej pracy wykorzystano **odsetek bezrobotnych długookresowo**¹³⁰ **wśród liczby bezrobotnych ogółem** (*Bezrobocie*). Uznano, że krótkookresowe bezrobocie nie musi mieć negatywnych skutków, gdyż może być związane z przejściową sytuacją dotyczącą zmiany pracy lub miejsca zamieszkania.

Kierując się pracą Bieńka¹³¹, postanowiono dołączyć zmienną dotyczącą alkoholu. Jak zostało wcześniej pokazane, 82% przestępstw drogowych dotyczyło jazdy pod wpływem alkoholu lub środków odurzających. Spożycie może się także przyczyniać do zwiększenia agresji i częstszych bójek oraz innych przestępstw powiązanych z przemocą.

¹²⁷ Kądziołka K., 2016, *Determinanty przestępczości w Polsce. Analiza zależności z wykorzystaniem drzew regresyjnych*, Ekonomia. Rynek, gospodarka, społeczeństwo, nr 45, s. 53-81

¹²⁸ Bogacka E., 2012: *Struktura przestrzenna i czynniki przestępczości na obszarze nadgranicznym Polski z Niemcami*, Bogucki Wydawnictwo Naukowe, Poznań

¹²⁹ Cracolici M.F., Uberti T.E., 2008, *Geographical Distribution of Crime in Italian Provinces: A Spatial Econometric Analysis*, Jahrbuch fur Regionalwissenschaft, nr 29, s. 1-28.

¹³⁰ Czyli powyżej jednego roku.

¹³¹ Bieńek P., Cichocki S., Szczepaniec M., 2012, *Czynniki ekonomiczne a poziom przestępczości – badanie ekonometryczne*, Zeszyty Prawnicze, nr 12.1, s. 147-172

Z racji braku danych dotyczących średniego spożycia alkoholu dla powiatów, wykorzystano **wydatki z budżetu powiatu na przeciwdziałanie alkoholowi na osobę (Alkohol)**. Postanowiono założyć, że wydatki te dobrze oddają poziom problemów spowodowanych nadmiernym spożyciem alkoholu.

Według danych Służby Więziennej, w 2017 roku kobiety stanowiły zaledwie niecałe 4% liczby osób tymczasowo aresztowanych i skazanych¹³². Świadczy to o tym, że zdecydowanie częściej to mężczyźni popełniają przestępstwa. Stąd postanowiono w modelu uwzględnić to zjawisko, za zmienną uznając **współczynnik feminizacji (Feminizacja)**, czyli liczbę kobiet przypadającą na 100 mężczyzn.

Ostatnią ze zmiennych wspólnych dla wszystkich przestępstw będzie **wskaźnik wykrywalności sprawców przestępstw stwierdzonych przez policję (Wykrywalność przestępstw)**, który informuje o odsetku przestępstw, dla których odnaleziono sprawcę¹³³. Wyższy współczynnik powinien działać destymulująco, gdyż świadomość tego, jak łatwo zostać wykrytym, przekładać się będzie na mniejszą skłonność do popełnienia przestępstwa. Zmienna ta była obecna m.in. w pracy Kądziołki¹³⁴, dotyczącej przestępczości ogółem. W niniejszej pracy wskaźnik nie będzie brany dla roku 2017, lecz będzie jego średnią z ostatnich 4 lat, gdyż osoba zamierzająca popełnić przestępstwo rozpatruje szanse na bycie złapanym poprzez pryzmat wydarzeń z przeszłości.

Przestępstwa na tle kryminalnym bazować będą wyłącznie na wymienionych wyżej dwunastu zmiennych. Z kolei dla przestępstw gospodarczych oraz drogowych w skład modeli wchodzi również po trzy zmienne wybrane subiektywnie przez autora. Dla przestępstw gospodarczych są to:

- **podmioty wpisane do rejestru REGON na 10 000 ludności (Regon)** - większa liczba podmiotów gospodarczych sprzyja pozostaniu anonimowym w tłumie i szansie na niewykryte przestępstwo, co widać na przykładzie Górnośląskiego Okręgu Przemysłowego,
- **nakłady inwestycyjne w przedsiębiorstwach na 1 mieszkańca (Inwestycje)** - większa wartość będzie oznaczała dysponowanie przez podmioty większym kapitałem, co sprzyja przestępczości gospodarczej o większej skali, a także większym pieniądзом na ewentualne procesy sądowe,

¹³² <https://sw.gov.pl/strona/statystyka-roczna> (data dostępu: 20.06.2019)

¹³³ Wskaźnik ten jest jednak liczony oddzielnie dla każdej z kategorii przestępstw.

¹³⁴ Kądziołka K., 2016, *Analysis of the crime rate in Poland in spatial and temporal terms*, Central and Eastern European Journal of Management and Economics, nr 1, s. 81-96

- **wartość umów/decyzji o dofinansowanie ze środków Unii Europejskiej na 1 mieszkańca (*Dotacje*)** - podobnie jak dla poprzedniej zmiennej, i w tym przypadku większe kwoty oznaczają większe możliwości manipulacji. Wartości zostały przeliczone nie na 1 umowę, lecz na 1 mieszkańca z powodu braku danych.

Dodatkowe zmienne dla przestępstw drogowych to:

- **udział wydatków na drogi publiczne w wydatkach budżetu ogółem (*Wydatki*)** - zmniejszone wydatki na drogi publiczne mogą oznaczać zły stan dróg, więc większa wartość będzie oznaczać, że poprawiana jest jakość nawierzchni,
- **natężenie dróg gminnych i powiatowych o twardej lub gruntowej nawierzchni na 100 km² (*Drogi*)** - większe natężenie dróg oznacza utrudniony ruch samochodowy oraz zwiększoną szansę na wypadek, więc zmienna jest stymulantą,
- **zmienna binarna rozróżniająca powiaty grodzkie oraz ziemskie (*Miasto*)**: w mieście panują inne warunki jazdy oraz ograniczenia prędkości, więc należy rozdzielić tereny miejskie od terenów pozamiejskich.

W danych brakowało wyłącznie jednej wartości, tj. miejsc noclegowych na 1000 osób (*Nocleg*) dla powiatu miasta Mysłowice. Wartość postanowiono zastąpić średnią z pięciu ostatnich lat, dla których była ona podana (lata 2011-2015). Aby dokładniej przyjrzeć się zmiennym objaśniającym, wykonano tabelę z podstawowymi statystykami opisowymi.

Tabela 3.5. Wybrane statystyki opisowe dla zmiennych objaśniających w roku 2017

Przestępstwa	Statystyki opisowe								
	Min.	Q1	Mediana	Średnia	Q3	Max.	Odch. st.	Kurtoza	Skośność
Gestosc_zaludnienia	19,00	61,00	91,00	371,26	188,75	3785,00	658,90	5,64	2,45
Wspolczynnik_migracji	-184,40	-73,53	-33,05	-14,56	14,55	635,30	99,49	7,81	2,21
Pomoc_spoleczna	1,70	4,58	6,40	6,71	8,50	18,80	2,78	0,65	0,70
Rozwody_na_1000	0,60	1,30	1,60	1,59	1,90	2,70	0,41	-0,54	-0,03
Bezrobocie	13,20	33,68	39,70	38,92	45,00	56,40	8,04	-0,16	-0,38
Feminizacja	96,00	102,00	104,00	104,71	107,00	119,00	3,84	0,80	1,01
Urbanizacja	0,00	30,78	45,91	50,63	63,76	100,00	27,43	-0,55	0,60
Wykrywalnosc_krym	33,93	55,15	61,27	61,22	67,80	90,23	9,40	0,20	-0,04
Wykrywalnosc_gosp	48,70	77,28	82,99	82,51	88,32	98,23	8,22	1,00	-0,68
Wykrywalnosc_drog	96,58	98,68	99,13	99,02	99,49	100,00	0,64	1,72	-1,11
Wynagrodzenie	2960,24	3561,35	3747,29	3874,45	4048,41	7515,99	518,26	9,65	2,41
Mieszkania	1,85	2,66	2,91	2,89	3,15	4,02	0,38	-0,06	-0,16
Alkohol	4,55	13,89	16,22	16,99	18,76	57,79	5,53	13,69	2,65
Nocleg	0,57	5,14	8,56	24,02	15,91	413,15	57,73	24,40	4,85
Regon	499,00	747,75	900,50	947,90	1050,25	2503,00	286,89	4,93	1,77
Inwestycje	250,00	1369,50	2246,50	3126,48	3881,00	23687,00	2950,71	14,00	3,15
Dotacje	636,18	2060,66	3046,83	4044,02	4710,21	24115,04	3302,08	10,70	2,85
Wydatki	1,00	7,38	12,60	13,69	17,83	59,10	7,99	2,98	1,24
Drogi	35,40	97,90	123,40	157,51	181,95	619,30	96,61	2,99	1,75

Źródło: Opracowanie własne z wykorzystaniem arkusza kalkulacyjnego Excel.

Powyższe statystyki są bardzo zróżnicowane, jednak można doszukać się kilku wspólnych własności. Dane w większości posiadają dodatnią kurtozę, co świadczy o skupianiu się obserwacji wokół średniej. Podobna sytuacja występuje dla skośności, gdzie zdecydowana większość zmiennych posiada ją większą od zera, zatem dominują obserwacje poniżej średniej. Brak jest obserwacji, które powodowałyby tak duże zaburzenia statystyk, jak wcześniej przedstawione miasto Krosno.

3.5. Podsumowanie

W niniejszym rozdziale zajęto się analizą zmiennych wchodzących w skład modeli. Dla zmiennych objaśnianych zbadano ich autokorelację oraz wykryto dwie obserwacje odstające, które zostały usunięte z dalszych badań dotyczących przestępczości gospodarczej. Wyznaczono również zbiór zmiennych objaśniających oraz przedstawiono ich podstawowe statystyki opisowe. W świetle dokonanych badań konieczne jest utworzenie modeli, które będą uwzględniać autokorelację przestrzenną dla każdej z kategorii przestępczości.

4. Modelowanie przestrzenne przestępczości w Polsce

4.1. Wstęp

Niniejszy rozdział stanowi ciąg dalszy empirycznych badań z rozdziału trzeciego, gdzie wykazano obecność autokorelacji dla wszystkich kategorii przestępczości oraz wyznaczono zmienne objaśniane do modeli, które będą stanowić temat tej części pracy. Zostaną one wyestymowane zgodnie z podejściem „od szczególnego do ogólnego”, czyli tzw. metodą klasyczną. Sposób postępowania został szczegółowo omówiony w rozdziale 2.4, jednak poszczególne etapy będą przypomniane w trakcie całego rozdziału.

4.2. Dobór odpowiedniej postaci modelu

Pierwszy etap stanowi estymacja modelu regresji liniowej na zmiennych opisanych w rozdziale trzecim. Dla przestępstw kryminalnych model wyestymowany został na wszystkich 12 bazowych zmiennych. Dla pozostałych kategorii przestępczości zdecydowano się jednak ograniczyć podstawowy zbiór zmiennych, z racji na niemerytoryczność niektórych z nich w odniesieniu do przestępstw innych niż kryminalne. Poniżej znajduje się ostateczna lista zmiennych branych pod uwagę, w rozbiciu na trzy omawiane kategorie.

Przestępstwa kryminalne (*Krym*):

1. *Gestosc_zaludnienia* - gęstość zaludnienia,
2. *Wspolczynnik_migracji* - współczynnik migracji,
3. *Pomoc_spoleczna* - procent rodzin korzystających z pomocy społecznej,
4. *Rozwody_na_1000* - liczba rozwodów na 1000 ludności,
5. *Bezrobocie* - odsetek bezrobotnych długookresowo wśród liczby bezrobotnych ogółem,
6. *Feminizacja* - współczynnik feminizacji,
7. *Urbanizacja* - współczynnik urbanizacji,
8. *Wykrywalnosc_krym* - wskaźnik wykrywalności sprawców przestępstw kryminalnych stwierdzonych przez policję,
9. *Wynagrodzenie* - przeciętne miesięczne wynagrodzenie brutto,
10. *Mieszkania* - przeciętna liczba mieszkańców na jedno mieszkanie,
11. *Alkohol* - wydatki z budżetu powiatu na przeciwdziałanie alkoholowi na osobę,
12. *Nocleg* - miejsca noclegowe na 1000 osób.

Przestępstwa gospodarcze (*Gosp*):

1. *Gestosc_zaludnienia* - gęstość zaludnienia,
2. *Wspolczynnik_migracji* - współczynnik migracji,
3. *Pomoc_spoleczna* - procent rodzin korzystających z pomocy społecznej,
4. *Bezrobocie* - odsetek bezrobotnych długookresowo wśród liczby bezrobotnych ogółem,
5. *Urbanizacja* - współczynnik urbanizacji,
6. *Wykrywalnosc_gosp* - wskaźnik wykrywalności sprawców przestępstw gospodarczych stwierdzonych przez policję,
7. *Wynagrodzenie* - przeciętne miesięczne wynagrodzenie brutto,
8. *Regon* - podmioty wpisane do rejestru REGON na 10 000 ludności,
9. *Inwestycje* - nakłady inwestycyjne w przedsiębiorstwach na 1 mieszkańca,
10. *Dotacje* - wartość umów/decyzji o dofinansowanie ze środków Unii Europejskiej na 1 mieszkańca.

Przestępstwa drogowe (*Drog*):

1. *Gestosc_zaludnienia* - gęstość zaludnienia,
2. *Wspolczynnik_migracji* - współczynnik migracji,
3. *Pomoc_spoleczna* - procent rodzin korzystających z pomocy społecznej,
4. *Bezrobocie* - odsetek bezrobotnych długookresowo wśród liczby bezrobotnych ogółem,
5. *Feminizacja* - współczynnik feminizacji,
6. *Urbanizacja* - współczynnik urbanizacji,
7. *Wykrywalnosc_drog* - wskaźnik wykrywalności sprawców przestępstw drogowych stwierdzonych przez policję,
8. *Wynagrodzenie* - przeciętne miesięczne wynagrodzenie brutto,
9. *Alkohol* - wydatki z budżetu powiatu na przeciwdziałanie alkoholizmowi na osobę,
10. *Nocleg* - miejsca noclegowe na 1000 osób,
11. *Wydatki* - udział wydatków na drogi publiczne w wydatkach budżetu ogółem,
12. *Drogi* - natężenie dróg gminnych i powiatowych o twardej lub gruntowej nawierzchni na 100 km²,
13. *Miasto* - zmienna binarna rozróżniająca powiaty grodzkie oraz ziemskie.

Ponieważ model liniowy niekoniecznie musi być ostatecznym celem, zdecydowano się nie przedstawiać jego parametrów, a jedynie zaprezentować wyniki następujących testów: Shapiro-Wilka (normalność reszt rozkładu), Breuscha-Pagana (heteroskedastyczność) oraz najważniejszy, Morana (autokorelacja przestrzenna reszt). Wyniki prezentuje tabela 4.1.

Tabela 4.1. Wyniki testów Shapiro-Wilka, Breuscha-Pagana oraz Morana dla modeli regresji liniowej

Przestępstwa	p-value dla testów		
	Shapiro-Wilka	Breuscha-Pagana	Morana
Kryminalne	0,00	0,01	0,00
Gospodarcze	0,00	0,29	0,98
Drogowe	0,00	0,47	0,00

Źródło: Opracowanie własne z wykorzystaniem programu R oraz arkusza kalkulacyjnego Excel.

Z przeprowadzonych testów wynika, że dla przestępstw gospodarczych autokorelacja przestrzenna została wyjaśniona modelem liniowym i nie ma potrzeby stosowania bardziej skomplikowanych modeli. Dalsza estymacja dla przestępstw gospodarczych będzie zatem polegać na poprawie modelu regresji. Może okazać się, że po usunięciu nieistotnych zmiennych i ewentualnych obserwacji odstających diagnostyka poprawi się.

Przestępstwa kryminalne oraz drogowe w dalszym ciągu wykazują autokorelację przestrzenną dla reszt, więc może istnieć potrzeba wykorzystania któregoś z omawianych wcześniej modeli przestrzennych. Żeby jednak potwierdzić powyższe przypuszczenie, należy wyestymować model na logarytmach. Jeśli nie zostanie wykryta przestrzenna interakcja, to obecność autokorelacji reszt dla modelu liniowego mogła być spowodowana estymacją związku nieliniowego modelem liniowym. Wyniki dla zlogarytmizowanych zmiennych przedstawia poniższa tabela¹³⁵.

Tabela 4.2. Wyniki testów Shapiro-Wilka, Breuscha-Pagana oraz Morana dla modeli regresji liniowej ze zlogarytmizowanymi zmiennymi

Przestępstwa	p-value dla testów		
	Shapiro-Wilka	Breuscha-Pagana	Morana
Kryminalne	0,00	0,33	0,00
Drogowe	0,00	0,79	0,00

Źródło: Opracowanie własne z wykorzystaniem programu R oraz arkusza kalkulacyjnego Excel.

¹³⁵ Zmienne *Urbanizacja*, *Współczynnik migracji* i *Miasto* nie zostały zlogarytmizowane, gdyż posiadają wartości niedodatnie.

W dalszym ciągu utrzymuje się autokorelacja przestrzenna, zatem potwierdzona została słusność zamiaru przejścia do bardziej skomplikowanych modeli. W wyniku zlogarytmowania zmiennych w obu modelach poprawił się wynik testu Breuscha-Pagana, jednak w dalszym ciągu reszty nie są zgodne z rozkładem normalnym. Powyższy model wydaje się być lepszym od zwykłego modelu liniowego, jednak zlogarytmizowanie wszystkich zmiennych mogłoby spowodować zbyt duże uproszczenia. Postanowiono sprawdzić, czy po zlogarytmowaniu wyłącznie zmiennej objaśnianej wyniki testów diagnostycznych będą podobne. Podobne przekształcenie zdecydowano się sprawdzić także dla przestępstw gospodarczych.

Tabela 4.3. Wyniki testów Shapiro-Wilka, Breuscha-Pagana oraz Morana dla modeli regresji liniowej ze zlogarytmizowaną zmienną objaśnianą

Przestępstwa	p-value dla testów		
	Shapiro-Wilka	Breuscha-Pagana	Morana
Kryminalne	0,00	0,32	0,00
Gospodarcze	0,00	0,35	0,93
Drogowe	0,02	0,34	0,00

Źródło: Opracowanie własne z wykorzystaniem programu R oraz arkusza kalkulacyjnego Excel.

Model dla przestępstw gospodarczych posiada wyższą wartość p-value dla testu Breuscha-Pagana, a także, chociaż jest to niewidoczne przez zaokrąglenie, dla testu Shapiro-Wilka. Test na normalność rozkładu reszt wyszedł też lepiej dla pozostałych kategorii przestępstw, chociaż pogorszył się wynik testu Breuscha-Pagana w porównaniu z logarytmizacją wszystkich zmiennych. Mimo to, w dalszej części pracy zdecydowano się na stosowanie zlogarytmizowanej zmiennej objaśnianej dla wszystkich kategorii przestępstw, gdyż poprawa p-value dla testu Shapiro-Wilka oznacza, że konieczne będzie usuwanie mniejszej ilości obserwacji odstających, a wszystkie modele wciąż zachowują homoskedastyczność.

Kolejnym krokiem, gdy model liniowy nie wyeliminował autokorelacji przestrzennej powiatów, jest przeprowadzenie testów mnożnika Lagrange’a. Służą one zbadaniu, który z modeli przestrzennych, SLM czy SEM, będzie bardziej odpowiedni dla zmiennej¹³⁶. Ich wyniki przedstawia tabela 4.4.

¹³⁶ Testy mnożnika Lagrange’a zostały bardziej szczegółowo omówione w rozdziale 2.4.

Tabela 4.4. Wyniki testów mnożnika Lagrange’a dla reszt modeli regresji liniowej

Przestępstwa	p-value dla testów			
	LMerr	LMlag	RLMerr	RLMlag
Kryminalne	0,00	0,00	0,01	0,25
Drogowe	0,00	0,00	0,46	0,00

Źródło: Opracowanie własne z wykorzystaniem programu R oraz arkusza kalkulacyjnego Excel.

Testy w wersji zwykłej nie wskazują, który model będzie lepszy, gdyż w obu p-value utrzymuje się na poziomie zerowym. Decydujące są zatem odporne wersje testów. Dla przestępstw kryminalnych został wskazany model błędu przestrzennego, natomiast dla przestępstw drogowych jest to model opóźnienia przestrzennego.

W dalszej części pracy zostaną wyestymowane trzy modele: model liniowy dla przestępstw gospodarczych, model błędu przestrzennego dla przestępstw kryminalnych oraz model opóźnienia przestrzennego dla przestępstw drogowych.

4.3. Model regresji liniowej dla przestępstw gospodarczych

Poniżej zostały przedstawione wstępne wyniki estymacji modelu liniowego dla przestępstw gospodarczych.

Tabela 4.5. Wyniki estymacji modelu liniowego dla przestępstw gospodarczych

Model liniowy	Współczynnik	p-value
(Intercept)	-1,24	0,02
Gestosc_zaludnienia	0,00021	0,00
Wspolczynnik_migracji	0,00052	0,22
Pomoc_spoleczna	-0,02	0,11
Bezrobocie	0,0016	0,7
Urbanizacja	0,0086	0,00
Wykrywalnosc_gosp	0,02	0,00
Wynagrodzenie	0,000028	0,73
Inwestycje	0,0000014	0,91
Regon	-0,00012	0,45
Dotacje	-0,000015	0,12

Źródło: Opracowanie własne z wykorzystaniem programu R oraz arkusza kalkulacyjnego Excel.

W celu zawarcia w modelu wyłącznie statystycznie istotnych zmiennych, zdecydowano się wybrać za pomocą metody Hellwiga¹³⁷ oraz regresji krokowej wstecz

¹³⁷ Opis metody znajduje się m.in. w: Dziechciarz J., 2003, *Ekonometria. Metody, przykłady, zadania*, Wydawnictwo Akademii Ekonomicznej, Wrocław, s. 51-53

wybrać optymalny zbiór zmiennych objaśnianych. Obie metody wskazały na zbiór zmiennych, dla którego estymacja została zaprezentowana poniżej.

Tabela 4.6. Wyniki estymacji modelu liniowego dla przestępstw gospodarczych

Model liniowy	Współczynnik	p-value
(Intercept)	-1,08	0,00
Gestosc_zaludnienia	0,00020	0,01
Pomoc_spoleczna	-0,03	0,02
Urbanizacja	0,0076	0,00
Wykrywalnosc_gosp	0,02	0,00

Źródło: Opracowanie własne z wykorzystaniem programu R oraz arkusza kalkulacyjnego Excel.

Wszystkie z proponowanych zmiennych okazały się być istotne. W poniższej tabeli zaprezentowano wyniki testów oraz kryteria informacyjne.

Tabela 4.7. Wyniki testu Shapiro-Wilka, Breuscha-Pagana oraz kryteria informacyjne dla modelu liniowego przestępstw gospodarczych

p-value dla testów			AIC	R ²	Skorygowane R ²
Shapiro-Wilka	Breuscha-Pagana	F			
0,00	0,05	0,00	691,83	0,35	0,33

Źródło: Opracowanie własne z wykorzystaniem programu R oraz arkusza kalkulacyjnego Excel.

Problem modelu stanowi brak spełnienia założenia o normalności rozkładu reszt. Można też mówić o dość niskim współczynniku R², jednak uzyskanie wysokiej wartości jest trudne do osiągnięcia, gdyż zbiór cech wpływających na decyzję o przestępstwie jest dość obszerny.

W celu poprawy normalności rozkładu reszt, zdecydowano się na eliminację obserwacji odstających. Ze zbioru danych usunięto około 6% obserwacji (21 powiatów). Były to głównie powiaty o najwyższej stopie przestępczości, jednak nie stanowiły one jednolitego ośrodka. Były to powiaty z takich województw jak: województwo śląskie, dolnośląskie, opolskie, świętokrzyskie, małopolskie, a nawet podlaskie. Wyniki modelu po powyższej operacji przedstawia tabela 4.8.

Tabela 4.8. Wyniki estymacji modelu liniowego dla przestępstw gospodarczych po usunięciu obserwacji odstających

Model liniowy	Współczynnik	p-value
(Intercept)	-0,86	0,00
Gestosc_zaludnienia	0,00017	0,00
Pomoc_spoeczna	-0,03	0,00
Urbanizacja	0,01	0,00
Wykrywalnosc_gosp	0,02	0,00

Źródło: Opracowanie własne z wykorzystaniem programu R oraz arkusza kalkulacyjnego Excel.

Wyniki zbliżone są do poprzedniego modelu, jednak poprawiła się istotność niektórych parametrów. Najważniejsza jest jednak poniższa diagnostyka.

Tabela 4.9. Wyniki testu Shapiro-Wilka, Breuscha-Pagana oraz kryteria informacyjne dla modelu liniowego przestępstw gospodarczych po usunięciu obserwacji odstających

p-value dla testów			AIC	R ²	Skorygowane R ²
Shapiro-Wilka	Breuscha-Pagana	F			
0,01	0,86	0,00	470,96	0,51	0,50

Źródło: Opracowanie własne z wykorzystaniem programu R oraz arkusza kalkulacyjnego Excel.

W wyniku usunięcia 6% obserwacji odstających model poprawił swoje dopasowanie, co wyraża się we wzroście współczynnika R² oraz spadku wartości kryterium AIC. Na poziomie istotności 0,01 można też przyjąć normalność rozkładu reszt, zwłaszcza, że inne testy wskazują na poziom istotności 0,05, co widać w poniższej tabeli.

Tabela 4.10. Wyniki testów na normalność dla modelu liniowego przestępstw gospodarczych po usunięciu obserwacji odstających

p-value dla testów			
Doornika-Hansena	Shapiro-Wilka	Lillieforsa	Jarque'a-Bera
0,05	0,01	0,04	0,09

Źródło: Opracowanie własne z wykorzystaniem programu Gretl oraz arkusza kalkulacyjnego Excel.

Z racji poprawy diagnostyki modelu po zmniejszeniu liczby obserwacji, postanowiono pozostać na zredukowanym modelu, którego interpretacja zostanie przedstawiona w kolejnym podrozdziale.

Interpretacja modelu liniowego

Tabela 4.11. Wyniki estymacji modelu liniowego dla przestępstw gospodarczych

Model liniowy	Współczynnik	p-value
(Intercept)	-0,86	0,00
Gestosc_zaludnienia	0,00017	0,00
Pomoc_spoleczna	-0,03	0,00
Urbanizacja	0,01	0,00
Wykrywalnosc_gosp	0,02	0,00

Źródło: Opracowanie własne z wykorzystaniem programu R oraz arkusza kalkulacyjnego Excel.

Zarówno gęstość zaludnienia (*Gestosc_zaludnienia*) jak i współczynnik urbanizacji (*Urbanizacja*) przyczyniają się do wzrostu przestępczości gospodarczej. Ma to swoje uzasadnienie w korzyściach aglomeracji miejskich, które to odznaczają się większą niż tereny wiejskie gęstością zaludnienia oraz urbanizacją. Wspomniane korzyści aglomeracji, powodujące lokowanie się firm w obszarach silnie zurbanizowanych i o dużej gęstości zaludnienia, to m.in. tworzenie dużego rynku zbytu, dostęp do wykwalifikowanej siły roboczej i rozbudowanej infrastruktury miejskiej¹³⁸. Duże zagęszczenie firm tworzy jednak sprzyjające warunki do pozostania anonimowym, przez co wzrasta przestępczość gospodarcza.

Destymulujący wpływ odsetka rodzin korzystających z pomocy społecznej (*Pomoc_spoleczna*) może wynikać z charakteru sprawców przestępstw gospodarczych. W największej części jest to tzw. „przestępczość białych kołnierzyków”, gdzie sprawcami przestępstw są osoby o wysokim statusie społecznym¹³⁹. Z kolei z pomocy społecznej korzystają w głównej mierze osoby ubogie, które nie pełnią kierowniczych stanowisk w firmach. Stąd prawdziwa może być zależność, że im większy udział osób ubogich w społeczeństwie, tym mniej jest popełnianych przestępstw gospodarczych.

Dodatni współczynnik przy wskaźniku wykrywalności przestępstw gospodarczych (*Wykrywalnosc_gosp*) wydaje się być niezgodny z intuicją, która podpowiada odwrotny skutek dobrej pracy policji. Przyczyną może być stosowanie w przypadku tej kategorii na tyle różnych sposobów popełnienia przestępstwa, że informacja o zatrzymaniu „białych kołnierzyków” nie zatrzymuje potencjalnych przestępców przed złamaniem prawa.

¹³⁸ Harasimowicz A., 2015, *Efekty aglomeracji - czynnik czy bariera rozwoju miast? Analiza wybranych aspektów w świetle literatury przedmiotu*, Studia Miejskie, nr 20, s. 24-27

¹³⁹ Chorbot P., 2009, *Przestępczość białych kołnierzyków w świetle amerykańskich teorii kryminologicznych. Analogie do polskich realiów*, Białostockie studia prawnicze, nr 6, s. 268-281

4.4. Model SEM dla przestępstw kryminalnych

Zgodnie z wskazaniem testów LM, dla przestępstw kryminalnych zostanie wykonany model błędu przestrzennego. Wyniki estymacji przedstawia poniższa tabela.

Tabela 4.12. Wyniki estymacji modelu SEM dla przestępstw kryminalnych

Model SEM	Współczynnik	p-value
Lambda	0,33	0,00
(Intercept)	4,16	0,00
Gestosc_zaludnienia	0,0000044	0,90
Wspolczynnik_migracji	0,00021	0,24
Pomoc_spoleczna	0,0022	0,77
Rozwody_na_1000	0,06	0,21
Bezrobocie	-0,0043	0,02
Feminizacja	-0,02	0,06
Urbanizacja	0,0071	0,00
Wykrywalnosc_krym	-0,0027	0,17
Wynagrodzenie	0,000047	0,15
Mieszkania	-0,29	0,00
Alkohol	0,01	0,00
Nocleg	-0,00035	0,34

Źródło: Opracowanie własne z wykorzystaniem programu R oraz arkusza kalkulacyjnego Excel.

Parametr lambda, odpowiedzialny za autokorelację przestrzenną reszt modelu, okazał się istotny. Potwierdza to słuszność wyboru modelu błędu przestrzennego dla przestępstw kryminalnych. Większość zmiennych jest jednak nieistotna statystycznie, dlatego w następnym kroku zaprezentowany zostanie zbiór zmiennych ograniczony przy pomocy regresji krokowej wstecz.

Tabela 4.13. Wyniki estymacji modelu SEM dla przestępstw kryminalnych

Model SEM	Współczynnik	p-value
Lambda	0,34	0,00
(Intercept)	4,33	0,00
Wspolczynnik_migracji	0,00038	0,01
Bezrobocie	-0,0041	0,02
Feminizacja	-0,01	0,05
Urbanizacja	0,0082	0,00
Mieszkania	-0,34	0,00
Alkohol	0,0098	0,00

Źródło: Opracowanie własne z wykorzystaniem programu R oraz arkusza kalkulacyjnego Excel.

Poniżej przedstawiono wyniki niektórych testów oraz kryteria informacyjne dla powyższego modelu.

Tabela 4.14. Wyniki testu Shapiro-Wilka, Breuscha-Pagana oraz kryteria informacyjne dla modelu SEM przestępstw kryminalnych

p-value dla testów		Logarytm wiarygodności	AIC	pseudo-R ²
Shapiro-Wilka	Breuscha-Pagana			
0,00	0,75	-38,68	95,37	0,61

Źródło: Opracowanie własne z wykorzystaniem programu R oraz arkusza kalkulacyjnego Excel.

Współczynnik pseudo-R² nie jest wysoki, ale biorąc pod uwagę skomplikowaną naturę przestępstw wartość ta jest jak najbardziej akceptowalna. Na podstawie powyższej tabeli zauważyć można, że model posiada problem wyłącznie z normalnością rozkładu reszt.

Aby rozwiązać powyższe zagadnienie, zgodnie z wskazaniem wykresu kwantyl-kwantyl zdecydowano się na usunięcie dwóch powiatów o największych resztach modelu, a mianowicie powiatu brzeskiego oraz dąbrowskiego. Są to powiaty leżące w województwie małopolskim, nie sąsiadujące ze sobą będące sąsiadami drugiego rzędu. Posiadają dość wysoką liczbę przestępstw kryminalnych w odróżnieniu od swoich sąsiadów, którzy mają ją znacznie mniejszą, i dlatego znacznie odstają od stworzonego modelu. Wyniki nowego modelu przedstawia tabela 4.15.

Tabela 4.15. Wyniki estymacji modelu SEM dla przestępstw kryminalnych po usunięciu obserwacji odstających

Model SEM	Współczynnik	p-value
Lambda	0,34	0,00
(Intercept)	4,46	0,00
Wspolczynnik_migracji	0,00037	0,01
Bezrobocie	-0,0048	0,01
Feminizacja	-0,015	0,03
Urbanizacja	0,0082	0,00
Mieszkania	-0,35	0,00
Alkohol	0,01	0,00

Źródło: Opracowanie własne z wykorzystaniem programu R oraz arkusza kalkulacyjnego Excel.

Nie widać dużych różnic w porównaniu z poprzednim modelem, jednak ważne okazały się ponownie przeprowadzone testy diagnostyczne.

Tabela 4.16. Wyniki testu Shapiro-Wilka, Breuscha-Pagana oraz kryteria informacyjne dla modelu SEM przestępstw kryminalnych po usunięciu obserwacji odstających

p-value dla testów		Logarytm wiarygodności	AIC	pseudo-R ²
Shapiro-Wilka	Breuscha-Pagana			
0,35	0,68	-22,57	63,13	0,64

Źródło: Opracowanie własne z wykorzystaniem programu R oraz arkusza kalkulacyjnego Excel.

Znacznie poprawił się wynik testu Shapiro-Wilka, a także kryteria informacyjne. W wyniku powyższych rezultatów zdecydowano się na dalsze stosowanie modelu bez powiatu brzeskiego oraz dąbrowskiego.

Końcowym etapem weryfikacji modelu jest wykonanie testu Walda, LR oraz LM¹⁴⁰, które powinny spełnić nierówność daną wzorem 2.29. Wyniki dla finalnego modelu przedstawia poniższa tabela.

Tabela 4.17. Statystyka Walda, LR i LM dla modelu SEM przestępstw kryminalnych po usunięciu obserwacji odstających

test Walda		test LR		test LM	
statystyka	p-value	statystyka	p-value	statystyka	p-value
26,07	0,00	27,45	0,00	0,54	0,46

Źródło: Opracowanie własne z wykorzystaniem programu R oraz arkusza kalkulacyjnego Excel.

Same wyniki testów wskazują na dobrą specyfikację modelu, gdyż p-value przekroczyło poziom istotności tylko dla testu LM, jednak jest to efekt pożądany. Świadczy to o tym, że model SEM wyjaśnił autokorelację przestrzenną i nie występuje ona w resztach modelu. Nie zachodzi jednak nierówność z wzoru 2.29, gdyż statystyka LR jest większa od statystyki Walda. Może to wynikać z niespełnienia założenia o normalności rozkładu składnika losowego modelu, istnienia nieliniowej zależności między zmienną objaśnianą a zmiennymi objaśniającymi, złego doboru zmiennych do modeli lub złego wyboru macierzy wag przestrzennych¹⁴¹. Ponieważ udowodniono normalność reszt, postanowiono wybrać inną macierz odległości, a mianowicie macierz binarną (macierz sąsiedztwa pierwszego rzędu bez standaryzacji wierszami). Statystyki przedstawia tabela 4.18.

¹⁴⁰ Szczegółowe omówienie testów w rozdziale 2.4.

¹⁴¹ Suhecki B. (red.), 2010, *Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych*, Wydawnictwo C.H.Beck, Warszawa s.298-299

Tabela 4.18. Statystyka Walda, LR i LM dla modelu SEM przestępstw kryminalnych po usunięciu obserwacji odstających i zastosowaniu macierzy binarnej

test Walda		test LR		test LM	
statystyka	p-value	statystyka	p-value	statystyka	p-value
30,64	0,00	30,22	0,00	0,54	0,46

Źródło: Opracowanie własne z wykorzystaniem programu R oraz arkusza kalkulacyjnego Excel.

Zmiana macierzy wag przyczyniła się do poprawnej specyfikacji modelu, więc postanowiono pozostać przy macierzy binarnej. Polepszyły się także kryteria informacyjne, co przedstawia tabela 4.19.

Tabela 4.19. Wyniki testu Shapiro-Wilka, Breuscha-Pagana oraz kryteria informacyjne dla modelu SEM przestępstw kryminalnych po usunięciu obserwacji odstających i zastosowaniu macierzy binarnej

p-value dla testów		Logarytm wiarygodności	AIC	pseudo-R ²
Shapiro-Wilka	Breuscha-Pagana			
0,43	0,52	-21,18	60,36	0,64

Źródło: Opracowanie własne z wykorzystaniem programu R oraz arkusza kalkulacyjnego Excel.

Wykorzystanie macierzy binarnej nie zmieniło wyników dla eksploracyjnej analizy danych, lecz spowodowało niewielkie zmiany we współczynnikach modelu. Dlatego przed interpretacją modelu zostanie on pokazany w swojej ostatecznej formie.

Interpretacja modelu SEM

Tabela 4.20. Wyniki estymacji modelu SEM dla przestępstw kryminalnych

Model SEM	Współczynnik	p-value
Lambda	0,07	0,00
(Intercept)	4,42	0,00
Wspolczynn timer migracji	0,00040	0,00
Bezrobocie	-0,0050	0,00
Feminizacja	-0,01	0,04
Urbanizacja	0,0080	0,00
Mieszkania	-0,36	0,00
Alkohol	0,01	0,00

Źródło: Opracowanie własne z wykorzystaniem programu R oraz arkusza kalkulacyjnego Excel.

Pierwszy istotny parametr to λ , który świadczy o istnieniu autokorelacji przestrzennej błędu modelu. Oznacza to, że przestępczość kryminalna w danym powiecie

nie wpływa na przestępczość w powiatach sąsiednich. Jeśli jednak w powiecie wystąpi nagle zmiana w resztach modelu, to samo stanie się w powiatach sąsiednich. Może to powodować zmienna, której nie uwzględniono w zbiorze zmiennych objaśniających. Z racji, że wysokie wartości grupują się przy zachodniej granicy Polski, może to też być wspomnianym efektem braku stabilności społecznej lub ruchu tranzytowego, jednak te dane są albo niemierzalne (współczynnik migracji oddaje tylko niedawne zmiany ludności, nie sięga wystarczająco daleko w przeszłość), albo dane nie są, przynajmniej publicznie, dostępne (ruch tranzytowy dla powiatów).

Kolejnym istotnym parametrem jest, wspomniany przed chwilą, współczynnik migracji (*Wspolczynnik_migracji*). Dodatni znak odpowiada interpretacji zawartej w rozdziale trzecim i jest zgodny z literaturą. Zawierania w strukturze społecznej, napływ nowej ludności stwarzają dogodne warunki przestępcom. Współczynnik ma niewielką wartość, jednak zmienna może przyjmować duże wartości.

Okazuje się, że odsetek bezrobotnych długookresowo wśród liczby bezrobotnych ogółem (*Bezrobocie*) wpływa ujemnie na przestępczość kryminalną. Nie odpowiada to intuicji, która powiązuje wyższe zagrożenie przestępczością z wyższym bezrobociem. Ujemny znak może jednak mieć swoje uzasadnienie w teorii działań rutynowych, która została przedstawiona w rozdziale 1.5. W pracy Cohena i Felsona¹⁴² podjęta została próba wytłumaczenia rosnącej przestępczości w Stanach Zjednoczonych latach 1960-1975, kiedy poprawiały się warunki życia ludności: spadała stopa ubóstwa, rosła skolaryzacja, a także spadało bezrobocie. Rosnąca przestępczość przy spadku bezrobocia tłumaczona była zmniejszonym nadzorem nad domostwami. Większa liczba osób udających się do pracy oznaczała zwiększoną liczbę pustych, niestrzeżonych mieszkań.

Z wspomnianymi przed chwilą pustymi mieszkaniem ma też związek zmienna dotycząca przeciętnej liczby mieszkańców na jedno mieszkanie (*Mieszkania*). Ujemny znak wskazuje na destymulujący charakter zmiennej. Większa liczba osób przypadających na jedno mieszkanie powoduje z pewnością spadek przestępstw polegających na kradzieży mienia, gdyż rośnie liczba potencjalnych świadków.

Zgodny z przewidywaniami jest ujemny znak przy zmiennej *Feminizacja*, wskazującej na feminizację społeczeństwa. Większa liczba kobiet przypadających na 100 mężczyzn przyczynia się do spadku przestępczości kryminalnej, która, zgodnie ze statystykami Służby Więziennej, jest domeną mężczyzn.

¹⁴² Cohen, L.E., Felson M., 1979, *Social Change and Crime Rate Trends: A Routine Activities Approach*, American Sociological Review, nr 44, s. 588-608

Dodatnia wartość współczynnika dla współczynnika urbanizacji (*Urbanizacja*) odpowiada teorii działań rutynowych, która przypisuje urbanizacji kategorię odpowiedniego celu. Większe skupiska ludności sprzyjają przestępcy, gdyż w tłumie trudniej jest o wykrycie takich przestępstw jak np. kradzież. Tym samym miasta charakteryzują się większą przestępczością, gdyż powiaty miejskie posiadają wartość współczynnika równą 1.

Dodatni parametr dla zmiennej *Alkohol*, mówiącej o wydatkach z budżetu powiatu na przeciwdziałanie alkoholowi na osobę, sugeruje, że wydatki te nie obniżają, lecz podwyższają poziom przestępczości. Nie są one z pewnością nastawione bezpośrednio na poczucie bezpieczeństwa obywateli, lecz na obniżenie konsumpcji alkoholu. Trudno zatem mówić w tym przypadku o odwrotnym działaniu niż zamierzone, jednak należy zwrócić uwagę na skutek zwiększania wydatków na przeciwdziałanie alkoholizmowi.

4.5. Model SAR dla przestępstw drogowych

Ostatni z modeli, zgodnie z wskazaniem testów LM, będzie modelem opóźnienia przestrzennego. Wyniki estymacji dla przestępstw drogowych przedstawia poniższa tabela.

Tabela 4.21. Wyniki estymacji modelu SAR dla przestępstw drogowych

Model SAR	Współczynnik	p-value
Rho	0,30	0,00
(Intercept)	-4,13	0,05
Gestosc_zaludnienia	-0,00016	0,00
Wspolczynnik_migracji	-0,00015	0,26
Pomoc_spoleczna	0,02	0,00
Bezrobocie	0,00071	0,62
Feminizacja	0,0028	0,61
Urbanizacja	-0,00046	0,63
Wykrywalnosc_drog	0,04	0,04
Wynagrodzenie	0,000032	0,20
Alkohol	0,01	0,01
Nocleg	0,000072	0,80
Wydatki	0,0035	0,04
Drogi	0,000044	0,84
Miasto	0,01	0,93

Źródło: Opracowanie własne z wykorzystaniem programu R oraz arkusza kalkulacyjnego Excel.

Parametr rho, odpowiedzialny za autokorelację przestrzenną, okazał się istotny. Potwierdza to słuszność wyboru modelu opóźnienia przestrzennego dla przestępstw drogowych. Większość zmiennych jest jednak nieistotna statystycznie, dlatego w następnym kroku zaprezentowany zostanie zbiór zmiennych ograniczony przy pomocy regresji krokowej wstecz.

Tabela 4.22. Wyniki estymacji modelu SAR dla przestępstw drogowych

Model SAR	Współczynnik	p-value
Rho	0,30	0,00
(Intercept)	-3,61	0,05
Gestosc_zaludnienia	-0,00015	0,00
Pomoc_spoleczna	0,02	0,00
Wykrywalnosc_drog	0,04	0,04
Alkohol	0,01	0,00
Wydatki	0,0033	0,03

Źródło: Opracowanie własne z wykorzystaniem programu R oraz arkusza kalkulacyjnego Excel.

Nie udało się odzyskać istotności dla żadnej ze statystycznie nieistotnych zmiennych z pełnego modelu. Kolejnym krokiem jest wykonanie pierwszych testów oraz wyznaczenie kryteriów informacyjnych.

Tabela 4.23. Wyniki testu Shapiro-Wilka, Breuscha-Pagana oraz kryteria informacyjne dla modelu SAR przestępstw drogowych

p-value dla testów		Logarytm wiarygodności	AIC	pseudo-R ²
Shapiro-Wilka	Breuscha-Pagana			
0,02	0,60	57,75	-99,50	0,37

Źródło: Opracowanie własne z wykorzystaniem programu R oraz arkusza kalkulacyjnego Excel.

Współczynnik pseudo-R² nie jest duży, jednak należy mieć na uwadze skomplikowaną strukturę przestępstw drogowych. Nie zawsze też sytuacja w danym powiecie ma wpływ na przestępstwo drogowe, które dzieje się na jego obszarze. Dla niewielkiej jednostki, jaką jest powiat, niemałą część uczestników ruchu stanowią osoby spoza powiatu. W pracy przyjęto jednak, że mimo wszystko powiat ma znaczenie dla ruchu drogowego na swoim obszarze, m.in. poprzez wpływanie na lokalnych mieszkańców oraz uczestników ruchu.

Występuje problem z normalnością rozkładu reszt. Część testów nie odrzuca hipotezy zerowej o normalności na poziomie 0,05, jednak z wykresu kwantyl-kwantyl można zauważyć, że dwa powiaty zdecydowanie odstają od reszty. Są to: powiat grójecki

oraz powiat miasta Legnica. Są one otoczone powiatami o również wysokich wartościach, jednak mimo wszystko znacznie od nich odstają, gdyż należą do pierwszych pięciu powiatów o największym zagrożeniu drogową przestępczością. Wyniki nowego modelu przedstawia poniższa tabela.

Tabela 4.24. Wyniki estymacji modelu SAR dla przestępstw drogowych po usunięciu obserwacji odstających

Model SAR	Współczynnik	p-value
Rho	0,26	0,00
(Intercept)	-3,83	0,03
Gestosc_zaludnienia	-0,00015	0,00
Pomoc_spoleczna	0,02	0,00
Wykrywalnosc_drog	0,04	0,02
Alkohol	0,01	0,00
Wydatki	0,0037	0,01

Źródło: Opracowanie własne z wykorzystaniem programu R oraz arkusza kalkulacyjnego Excel.

Parametry modelu zmieniły się nieznacznie, jednak najważniejsze jest sprawdzenie, czy poprawił się wynik testu Shapiro-Wilka.

Tabela 4.25. Wyniki testu Shapiro-Wilka, Breuscha-Pagana oraz kryteria informacyjne dla modelu SAR przestępstw drogowych po usunięciu obserwacji odstających

p-value dla testów		Logarytm wiarygodności	AIC	pseudo-R ²
Shapiro-Wilka	Breuscha-Pagana			
0,95	0,29	70,07	-124,14	0,39

Źródło: Opracowanie własne z wykorzystaniem programu R oraz arkusza kalkulacyjnego Excel.

Poprawiły się wszystkie kryteria informacyjne modelu oraz, co najważniejsze, wynik testu Shapiro-Wilka. Pogorszyła się jedynie homoskedastyczność modelu, jednak wciąż p-value dla testu Breuscha-Pagana ma wysoką wartość. W wyniku powyższych testów zdecydowano się na dalsze stosowanie modelu bez powiatu grójeckiego oraz m. Legnica.

Podobnie jak przy przestępstwach kryminalnych, końcowym etapem weryfikacji modelu jest wykonanie testu Walda, LR oraz LM, które powinny spełnić nierówność daną wzorem 2.29. Wyniki przedstawia następująca tabela.

Tabela 4.26. Statystyka Walda, LR i LM dla modelu SAR przestępstw drogowych po usunięciu obserwacji odstających

test Walda		test LR		test LM	
statystyka	p-value	statystyka	p-value	statystyka	p-value
18,99	0,00	18,50	0,00	0,42	0,52

Źródło: Opracowanie własne z wykorzystaniem programu R oraz arkusza kalkulacyjnego Excel.

Każdy z testów indywidualnie przedstawia pożądane wartości, a także spełniona jest nierówność zadana wzorem 2.29. Oznacza to, że model jest dobrze wyspecyfikowany oraz że można przejść do interpretacji modelu opóźnienia przestrzennego dla przestępstw drogowych.

Interpretacja modelu SAR

Tabela 4.27. Wyniki estymacji modelu SAR dla przestępstw drogowych

Model SAR	Współczynnik	p-value
Rho	0,26	0,00
(Intercept)	-3,83	0,03
Gestosc_zaludnienia	-0,00015	0,00
Pomoc_spoeczna	0,02	0,00
Wykrywalnosc_drog	0,04	0,02
Alkohol	0,01	0,00
Wydatki	0,0037	0,01

Źródło: Opracowanie własne z wykorzystaniem programu R oraz arkusza kalkulacyjnego Excel.

Pierwszy istotny parametr to ρ , który świadczy o istotności opóźnienia przestrzennego zmiennej objaśnianej. Oznacza to, że przestępczość drogowa w danym powiecie wpływa na przestępczość drogową w powiatach sąsiednich. Można zatem tłumaczyć skalę przestępczości drogowej w badanym powiecie również za pomocą wartości zmiennej objaśnianej (czyli także zmiennych objaśnianych) w powiatach, które zgodnie z macierzą wag są jego sąsiadami. Wniosek ten pokrywa się z omawianą wcześniej cechą przestępczości drogowej, a mianowicie jej przenikalnością. Dla tak niewielkich obszarów jak powiaty sporą część ruchu drogowego generują osoby z innych powiatów, które np. dojeżdżają do pracy lub szkoły, albo podróżują w bardziej odległe miejsca i na terenie danego powiatu przebywają przez krótki odcinek czasu.

W modelu istotna okazała się być gęstość zaludnienia (*Gestosc_zaludnienia*), która przyczynia się do obniżenia przestępczości drogowej. Może to być skutek zwiększonej kontroli policji w gęściej zaludnionych obszarach, a także bardziej

restrykcyjnych norm na obszarach zabudowanych, które to charakteryzują się większą gęstością zaludnienia. Podobne wnioski zawarte zostały w pracy Grimma i Treibich¹⁴³, gdzie zauważono spadek śmiertelności pieszych na skutek wypadków samochodowych przy podwyższeniu gęstości zaludnienia, dla niezmiennego się wskaźnika urbanizacji. Grimm i Treibich sugerują, że może być to spowodowane powiązaniem większej gęstości zaludnienia z niższą średnią prędkością pojazdów.

Pierwszą stymulantą przestępczości drogowej jest odsetek rodzin korzystających z pomocy społecznej (*Pomoc_spoeczna*). Pomoc społeczna przeznaczona jest dla ubogich rodzin, które nie mogą sobie pozwolić na nowe, bardziej bezpieczne auta. Część rodzin boryka się także z problemem alkoholizmu, co może sprzyjać prowadzeniu samochodu pod wpływem alkoholu. Poza tym, większy odsetek ludzi korzystających z pomocy społecznej oznacza mniej zamożne społeczeństwo. To z kolei koreluje z wyższą śmiertelnością w wypadkach drogowych, gdyż dla społeczeństw o średnim i wysokim stopniu rozwoju, wraz ze wzrostem zamożności spada ryzyko śmiertelnego wypadku¹⁴⁴.

Kolejną cechą powodującą wzrost przestępczości jest wykrywalność przestępstw drogowych (*Wykrywalnosc_drog*). Jest to niezgodne z intuicją, która podpowiada odwrotny skutek lepszej pracy policji. Zmienna ta nie ma jednak tak dużego wpływu biorąc pod uwagę, jakie wartości może przyjmować. Statystyki opisowe w rozdziale trzecim pokazały, że wykrywalność przestępstw drogowych mieści się w przedziale [96,58;100], więc dla każdego powiatu można mówić o dużej wykrywalności i o dobrej pracy policji w każdej z badanych jednostek.

Dodatni parametr dla zmiennej *Alkohol*, mówiącej o wydatkach z budżetu powiatu na przeciwdziałanie alkoholowi na osobę, sugeruje, że wydatki te nie obniżają, lecz podwyższają poziom przestępczości drogowej. Jak wspomniano w rozdziale 3.2. przy analizie graficznej, 82% przestępstw drogowych w roku 2017 dotyczyło jazdy pod wpływem alkoholu lub środków odurzających. Zwiększone wydatki powiatu na przeciwdziałanie alkoholizmowi mogą wskazywać na duży problem mieszkańców z nadmiernym spożywaniem alkoholu, a w rezultacie zwiększoną szansę na prowadzenie pojazdu w stanie nietrzeźwości.

¹⁴³ Grimm M., Treibich C., 2013, *Determinants of road traffic crash fatalities across Indian states*, Health Economics, nr 22, s. 915-930

¹⁴⁴ Kopits E., Cropper M., 2005, *Traffic Fatalities and Economic Growth*, Accident Analysis and Prevention, nr 37, s. 172-173

Ostatnią zmienną istotną dla modelu jest udział wydatków na drogi publiczne w wydatkach budżetu ogółem (*Wydatki*), będący stymulantą przestępczości drogowej. Podobnie jak dla zmiennej *Alkohol*, zwiększony poziom zmiennej może w rezultacie prowadzić do poprawy warunków na drodze, jednak w krótszym odstępie czasu efekt jest odwrotny. Duże wydatki na infrastrukturę drogową oznaczają, że drogi są obecnie w złym stanie technicznym, a pojawiające się remonty utrudniają ruch zarówno kierowcom, jak i pieszym.

4.6. Podsumowanie

Niniejszy rozdział rozpoczął się od przedstawienia zmiennych objaśniających dla każdej z trzech kategorii przestępczości. Następnie za pomocą testu Morana oraz testów mnożnika Lagrange’a wybrano odpowiedni model dla każdej ze zmiennych objaśnianych. Dla przestępstw gospodarczych, z racji nieodrżucenia hipotezy zerowej testu Morana, wybrano model liniowy. Dla pozostałych kategorii przestępstw zostały utworzone modele przestrzenne: model SAR dla przestępstw drogowych oraz model SEM dla przestępstw kryminalnych.

Do każdego z utworzonych modeli zastosowano regresję krokową wstecz (dla modelu liniowego również metodę Hellwiga), aby wyznaczyć zmienne istotnie statystycznie. Następnie zbadano modele pod kątem spełnienia założeń dotyczących normalności rozkładu reszt oraz homoskedastyczności. W każdym przypadku wystąpił problem z normalnością, więc zastosowano redukcję zbioru obserwacji. Dla przestępstw kryminalnych oraz drogowych usunięto po dwie obserwacje, natomiast dla przestępstw gospodarczych 21 (łącznie z usuniętymi wcześniej powiatami m. Krosno i m. Tarnów, model bazuje na 367 powiatach). Modele przestrzenne zostały również przetestowane pod kątem statystyk Walda, LR i LM. W modelu SEM konieczna była zmiana macierzy wag ze standaryzowanej wierszami na binarną.

Zmienne, które nie znalazły się w żadnym z finalnych modeli, to: miejsca noclegowe na 1000 osób (*Nocleg*), przeciętne miesięczne wynagrodzenie brutto (*Wynagrodzenie*) oraz liczba rozwodów na 1000 ludności (*Rozwody_na_1000*). Ze zbioru zmiennych subiektywnie wybranych przez autora statystycznie istotna okazała się wyłącznie zmienna dotycząca udziału wydatków na drogi publiczne w wydatkach budżetu ogółem (*Wydatki*). Żadna ze zmiennych nie pojawiła się we wszystkich modelach, można więc mówić o unikalnym charakterze każdego typu przestępstw.

Zakończenie

Celem niniejszej pracy było przedstawienie zjawiska przestępczości w Polsce za pomocą analizy przestrzennej. W badaniu posłużono się danymi dla powiatów z roku 2017, pochodzącymi z Banku Danych Lokalnych. Wyodrębniono trzy główne kategorie przestępczości, które poddano dalszej analizie: przestępczość kryminalna, gospodarcza oraz drogowa, które przeliczono na 1000 mieszkańców. Dla wszystkich z nich przyjęto macierz odległości będącą macierzą sąsiedztwa pierwszego rzędu, standaryzowaną wierszami.

Za pomocą eksploracyjnej analizy danych ustalono, że każda kategoria przestępczości odznacza się występowaniem autokorelacji przestrzennej. Ze względu na możliwą niestacjonarność modelu, dla przestępstw gospodarczych konieczne było odrzucenie dwóch obserwacji odstających: powiatu miasta Krosno oraz powiatu miasta Tarnów.

Kolejnym krokiem było wyodrębnienie zmiennych objaśniających dla każdej z kategorii przestępczości. Mając na uwadze przeanalizowaną literaturę, wybrano zbiór 12 zmiennych, które mogły pojawić się w każdym z modeli. Dla przestępstw gospodarczych oraz drogowych dokonano także subiektywnego wyboru zmiennych, wybierając po trzy zmienne zgodne z intuicją autora pracy.

Test Morana pokazał, że w przypadku przestępstw gospodarczych model liniowy całkowicie wytłumaczył autokorelację przestrzenną. Dla przestępstw kryminalnych oraz gospodarczych konieczne było utworzenie modeli przestrzennych. Testy mnożnika Lagrange'a wskazały na model opóźnienia przestrzennego dla przestępstw drogowych oraz model błędu przestrzennego dla przestępstw kryminalnych. Wszystkie ze zmiennych objaśnianych zostały zlogarytmizowane, gdyż występował problem z normalnością rozkładu reszt. W każdym z utworzonych modeli przeprowadzono regresję krokową wstecz (dla przestępstw gospodarczych zastosowano też metodę Hellwiga), która wyłoniła zmienne istotne statystycznie, a także usunięto obserwacje odstające, zaburzające normalność rozkładu reszt. Finalnie modele przestrzenne bazowały na 378 powiatach, natomiast model liniowy na 367. Model błędu przestrzennego wymagał zmiany macierzy wag ze standaryzowanej wierszami na binarną, co potwierdziły wyniki testu Walda, LR oraz LM.

Dla objaśnienia przestępczości gospodarczej istotnie pozytywne okazały się zmienne dotyczące przestrzeni, a mianowicie gęstość zaludnienia oraz wskaźnik

urbanizacji. Można to tłumaczyć korzyściami płynącymi z lokowania firm w aglomeracjach miejskich, który to proces sprzyja anonimowości sprawców. Stymulantą okazała się też wykrywalność przestępstw, co ma uzasadnienie w coraz bardziej skomplikowanych metodach przestępczości gospodarczej, a w rezultacie nieodstraszaniem potencjalnych sprawców. Jediną istotnie negatywną zmienną jest odsetek rodzin korzystających z pomocy społecznej, co powiązane jest ze zdominowania przestępczości gospodarczej przez osoby wykształcone na kierowniczych stanowiskach.

Model błędu przestrzennego dla przestępstw kryminalnych pokazał istotność parametru λ , który świadczyć może o występowaniu nieuwzględnionego w modelu czynnika odpowiedzialnego za różnice w wartościach przestępstw kryminalnych dla powiatów. Stymulantami okazały się: współczynnik migracji (zawierania w strukturze społecznej i napływ nowej ludności sprzyjają przestępcom), współczynnik urbanizacji (w tłumie trudniej wykryć takie przestępstwa jak np. kradzież, łatwiej też o anonimowość) i wydatki powiatu na przeciwdziałanie alkoholizmowi (zwiększone wydatki sugerują większe spożycie alkoholu, który z kolei zwiększa agresję i bywa przyczyną przemocy). Poziom przestępczości obniżają: zwiększony odsetek bezrobotnych długookresowo wśród liczby bezrobotnych ogółem (zwiększony nadzór nad domostwami), zwiększony współczynnik feminizacji (mężczyźni częściej popełniają przestępstwa kryminalne) i zwiększona liczba mieszkańców na jedno mieszkanie (większy nadzór nad domostwem).

Model opóźnienia przestrzennego dla przestępstw drogowych pokazał istotność parametru ρ , czyli liczbę przestępstw drogowych w powiecie można tłumaczyć poziomem przestępczości tej kategorii w powiatach sąsiednich. ma to intuicyjne wyjaśnienie, gdyż dla tak niewielkich obszarów jak powiaty sporą część ruchu drogowego generują osoby z innych powiatów. Jediną destymulantą okazała się być gęstość zaludnienia, co może być skutkiem zmniejszenia dozwolonej prędkości w obszarach gęsto zaludnionych oraz wzmożonej na tych terenach kontroli policji. Przestępczość drogową stymulują następujące zmienne: odsetek rodzin korzystających z pomocy społecznej (wzrost zamożności przyczynia się do spadku liczby wypadków, natomiast większy poziom zmiennej oznaczają wzrost ubóstwa), wykrywalność przestępstw (może to być spowodowane niewielkimi różnicami zmiennej dla powiatów, graniczącej z poziomem 100%), a także udział wydatków na drogi publiczne w wydatkach budżetu ogółem i wydatki z budżetu powiatu na przeciwdziałanie

alkoholowi na osobę, które to zmienne przyjmując większe wartości sugerują problem powiatu z alkoholizmem i stanem dróg.

Podsumowując, wnioski wyciągnięte z badań są zgodne z przeprowadzanymi już badaniami, polegającymi na poszukiwaniu przyczyn przestępczości. W niniejszej pracy zwrócono jednak uwagę na aspekt przestrzenności badanego zjawiska, który powinien być uwzględniany przy próbie wyjaśniania przestępczości.

Bibliografia

1. Andresen M.A., 2011, *The Ambient Population and Crime Analysis*, The Professional Geographer, nr 2, s. 193–212
2. Anselin L., 1988, *Spatial Econometrics: Methods and models*, Springer Science
3. Anselin L., 1995, *Local indicators of spatial association*, Geographical Analysis, nr 27
4. Anselin L., 1998, *Interactive Techniques and Exploratory Spatial Data Analysis* w Longley P. i in., *Geographical Information Systems: Principles, Techniques, Management and Application*, New York: John Wiley & Sons, s. 253-266
5. Anselin L., Bera A., 1998, *Spatial Dependence in linear Regression Models with an Introduction to Spatial Econometrics*, Handbook of Applied Economic Statistics
6. Anselin, L., Spatial Econometrics, Chapter Fourteen, Spatial Econometrics Luc Anselin, *A Companion to Theoretical Econometrics* Edited by Badi H. Baltagi 2001, 2003, Blackwell Publishing Ltd, s. 310-330
7. Beck N., 2006, *Space Is More than Geography: Using Spatial Econometrics in the Study of Political Economy*, International Studies Quarterly, nr 50, s. 27-44
8. Becker G. S., 1968, *Crime and Punishment: An Economic Approach*, Journal of Political Economy nr 76, s. 169-217
9. Biecek P., 2013, *Wybrane testy normalności*, Materiały Fundacji SmarterPoland.pl
10. Bieniek P., Cichocki S., Szczepaniec M., 2012, *Czynniki ekonomiczne a poziom przestępczości – badanie ekonometryczne*, Zeszyty Prawnicze, nr 12.1, s. 147–172
11. Błachut J., 2007, *Problemy związane z pomiarem przestępczości*, Łódź, Wolters Kluwer Polska
12. Bodson P., Peters D., 1975. *Estimation of the Coefficients of a Linear Regression in the Presence of Spatial Autocorrelation: An Application to a Belgium Labor Demand Function*, Environment and Planning, nr 7, s. 455-472
13. Bogacka E., 2012, *Zastosowanie teorii społecznej dezorganizacji do badania zróżnicowania przestrzennego przestępczości*, Wiadomości statystyczne, nr 1, s. 43-51

14. Bogacka E., 2012: *Struktura przestrzenna i czynniki przestępczości na obszarze nadgranicznym Polski z Niemcami*, Bogucki Wydawnictwo Naukowe, Poznań
15. Bogacka E., 2016, *Czynniki kradzieży rzeczy cudzej w świetle teorii działań rutynowych. Przykład województwa wielkopolskiego*, *Rozwój Regionalny i Polityka Regionalna*, nr 36, s. 199-207
16. Chorbot P., 2009, *Przestępczość białych kołnierzyków w świetle amerykańskich teorii kryminologicznych. Analogie do polskich realiów*, *Białostockie studia prawnicze*, nr 6, s. 268-281
17. Cohen, L.E., Felson M., 1979, *Social Change and Crime Rate Trends: A Routine Activities Approach*, *American Sociological Review*, nr 44, s. 588-608
18. Cracolici M.F., Uberti T.E., 2008, *Geographical Distribution of Crime in Italian Provinces: A Spatial Econometric Analysis*, *Jahrbuch für Regionalwissenschaft*, nr 29, s. 1-28
19. Dziechciarz J., 2003, *Ekonometria. Metody, przykłady, zadania*, Wydawnictwo Akademii Ekonomicznej, Wrocław
20. Ehrlich I., 1973, *Participation in Illegitimate Activities: A Theoretical and Empirical Investigation*, *The Journal of Political Economy* nr 81, s. 521-565
21. Florax R, Nijkamp P., 2003, *Misspecification in Linear Spatial Regression Models*, *Tinbergen Institute Discussion Paper*, nr 3, s. 1-28
22. Florax R. i in., 2003, *Specification Searches in Spatial Econometrics: The Relevance of Hendry's Methodology*, *Regional Science and Urban Economics*, nr 33, s. 1-30
23. Florczak W., 2013, *Co wywołuje przestępczość i jak ją ograniczać? Wielowymiarowa analiza makroekonomiczna*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź
24. Gardocki L., 2013, *Pojęcie przestępstwa i podziały przestępstw w polskim prawie karnym*, *Annales Universitatis Mariae Curie-Skłodowska. Sectio G*, nr 60, Lublin, s. 29-41
25. Grimm M., Treibich C., 2013, *Determinants of road traffic crash fatalities across Indian states*, *Health Economics*, nr 22, s. 915-930
26. Groves W.B., Sampson R.J., 1989, *Community Structure and Crime: Testing Social-Disorganization Theory*, *American Journal of Sociology*, nr 94, s. 774-802

27. Haining R., 2003, *Spatial Data Analysis Theory and Practice*, Cambridge University Press
28. Harasimowicz A., 2015, *Efekty aglomeracji - czynnik czy bariera rozwoju miast? Analiza wybranych aspektów w świetle literatury przedmiotu*, Studia Miejskie, nr 20, s. 23-34
29. Kądziołka K., 2016, *Analysis of the crime rate in Poland in spatial and temporal terms*, Central and Eastern European Journal of Management and Economics, nr 1, s. 81-96
30. Kądziołka K., 2016, *Determinanty przestępczości w Polsce. Analiza zależności z wykorzystaniem drzew regresyjnych*, *Ekonomia. Rynek, gospodarka, społeczeństwo*, nr 45, s. 53-81
31. Kopczewska K., 2006, *Ekonometria i statystyka przestrzenna z wykorzystaniem programu R CRAN*, CeDeWu, Warszawa
32. Kopits E., Cropper M., 2005, *Traffic Fatalities and Economic Growth*, *Accident Analysis and Prevention*, nr 37, s. 169-178
33. Le Sage J.P., 2008, *An Introduction to Spatial Econometrics*, *Revue d'économie industrielle*, nr 123
34. Martinez R., Stowell J.I., Lee M.T., 2010, *Immigration and crime in an era of transformation: a longitudinal analysis of homicides in San Diego neighborhoods, 1980—2000*, *CRIMINOLOGY*, nr 3, s. 797-829
35. Oliver M.A., Webster R., 2014, *A tutorial guide to geostatistics: Computing and modelling variograms and kriging*, *Catena*, s. 56-69
36. Oliver M.A., Webster R., 2015, *Basic Steps in Geostatistics: The Variogram and Kriging*
37. Paelinck J.H.P, Klaassen L.H., 1983, *Ekonometria przestrzenna*, PWN, Warszawa
38. Shaw C., McKay H., 1942, *Juvenile Delinquency and Urban Areas*, University of Chicago Press, Chicago
39. Suhecki B. (red.), 2010, *Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych*, Wydawnictwo C.H.Beck, Warszawa
40. Tobler W.R., 1970, *A computer model simulating urban growth in Detroit region*, *Economic geography*, nr 2, s. 234-240

41. Ward M., Gleditsch K., 2007, *An Introduction to Spatial Regression Models in the Social Sciences*, Michael D. Ward & Kristian Skrede Gleditsch
42. Wong C., 2002, *Clifford R. Shaw and Henry D. McKay: The Social Disorganization Theory*. CSISS Classics, UC Santa Barbara: Center for Spatially Integrated Social Science
43. Woźniak A., Sikora J., 2007, *Autokorelacja przestrzenna wskaźników infrastruktury wodno-ściekowej woj. małopolskiego*, Infrastruktura i ekologia terenów wiejskich, nr 2, s.315-329
44. Ye X., Wu L., 2011, *Analyzing the dynamics of homicide patterns in Chicago: ESDA and spatial panel approaches*, Applied Geography, nr 31, s. 800-807
45. Zaman A., 2000, *The Inconsistency of the Breusch-Pagan Test*, Journal of Economic and Social Research, nr 2, s.1-5
46. Zeliaś A., 1991, *Ekonometria przestrzenna*, Państwowe Wydawnictwo Ekonomiczne, Warszawa

Źródła internetowe:

1. <http://ssc.siskom.waw.pl/mapa-nawierzchni/mapa-nawierzchnia.png>
(dostęp: 17.06.2019)
2. <http://www.statystyka.policja.pl/st/kodeks-karny/przestepstwa-przeciwko-2>
(dostęp: 17.06.2019)
3. <https://dziennikzachodni.pl/przestepczosc-w-woj-slaskim-maleje-ale-korupcja-to-nasza-specjalnosc-statystyki-policji/ar/3723630> (dostęp: 17.06.2019)
4. <https://encyklopedia.pwn.pl/haslo/ciemna-liczba-przestepstw;3886513.html>
(dostęp: 07.06.2019)
5. <https://krosno112.pl/aktualnosci/inne/item/5285-najniebezpieczniejsze-miasto-w-polsce-to-krosno> (dostęp: 17.06.2019)
6. <https://wiadomosci.onet.pl/kraj/rzeczpospolita-wzrasta-przestepczosc-najgorzej-na-zachodzie-polski/rc7rj> (dostęp: 16.06.2019)

Spis rysunków

Rysunek 2.1. Rodzaje autokorelacji przestrzennej	22
Rysunek 2.2. Charakterystyka wariogramu	27
Rysunek 2.3. Przykład wykresu punktowego Morana	30
Rysunek 3.1. Liczba przestępstw stwierdzonych przez policję w zakończonych postępowaniach przygotowawczych w roku 2017	41
Rysunek 3.2. Wykres pudełkowy dla przestępstw gospodarczych w roku 2017	43
Rysunek 3.3. Kartogram liczby przestępstw kryminalnych na 1000 mieszkańców dla powiatów Polski w roku 2017	44
Rysunek 3.4. Kartogram liczby przestępstw gospodarczych na 1000 mieszkańców dla powiatów Polski w roku 2017	45
Rysunek 3.5. Kartogram liczby przestępstw drogowych na 1000 mieszkańców dla powiatów Polski w roku 2017	46
Rysunek 3.6. Wykresy skalowanych wariogramów dla przestępstw w roku 2017	48
Rysunek 3.7. Wykresy korelogramów dla przestępstw w roku 2017	49
Rysunek 3.8. Wykresy punktowe Morana dla przestępstw w roku 2017	53
Rysunek 3.9. Kartogramy przynależności powiatów do ćwiartek wykresu punktowego Morana dla przestępstw w roku 2017	55

Spis tabel

Tabela 3.1. Wybrane statystyki opisowe dla przestępstw w roku 2017	42
Tabela 3.2. Test istotności opóźnień przestrzennych dla korelogramów przestępstw w roku 2017	50
Tabela 3.3. Globalne statystyki autokorelacji przestrzennej dla przestępstw w roku 2017..	51
Tabela 3.4. Statystyki join-count dla przestępstw gospodarczych w roku 2017	52
Tabela 3.5. Wybrane statystyki opisowe dla zmiennych objaśniających w roku 2017 ..	62
Tabela 4.1. Wyniki testów Shapiro-Wilka, Breuscha-Pagana oraz Morana dla modeli regresji liniowej	65
Tabela 4.2. Wyniki testów Shapiro-Wilka, Breuscha-Pagana oraz Morana dla modeli regresji liniowej ze zlogarytmizowanymi zmiennymi	65
Tabela 4.3. Wyniki testów Shapiro-Wilka, Breuscha-Pagana oraz Morana dla modeli regresji liniowej ze zlogarytmizowaną zmienną objaśnianą	66
Tabela 4.4. Wyniki testów mnożnika Lagrange’a dla reszt modeli regresji liniowej	67
Tabela 4.5. Wyniki estymacji modelu liniowego dla przestępstw gospodarczych	67
Tabela 4.6. Wyniki estymacji modelu liniowego dla przestępstw gospodarczych	68
Tabela 4.7. Wyniki testu Shapiro-Wilka, Breuscha-Pagana oraz kryteria informacyjne dla modelu liniowego przestępstw gospodarczych	68
Tabela 4.8. Wyniki estymacji modelu liniowego dla przestępstw gospodarczych po usunięciu obserwacji odstających	69
Tabela 4.9. Wyniki testu Shapiro-Wilka, Breuscha-Pagana oraz kryteria informacyjne dla modelu liniowego przestępstw gospodarczych po usunięciu obserwacji odstających	69
Tabela 4.10. Wyniki testów na normalność dla modelu liniowego przestępstw gospodarczych po usunięciu obserwacji odstających	69
Tabela 4.11. Wyniki estymacji modelu liniowego dla przestępstw gospodarczych	70
Tabela 4.12. Wyniki estymacji modelu SEM dla przestępstw kryminalnych	71
Tabela 4.13. Wyniki estymacji modelu SEM dla przestępstw kryminalnych	71
Tabela 4.14. Wyniki testu Shapiro-Wilka, Breuscha-Pagana oraz kryteria informacyjne dla modelu SEM przestępstw kryminalnych	72
Tabela 4.15. Wyniki estymacji modelu SEM dla przestępstw kryminalnych po usunięciu obserwacji odstających	72
Tabela 4.16. Wyniki testu Shapiro-Wilka, Breuscha-Pagana oraz kryteria informacyjne dla modelu SEM przestępstw kryminalnych po usunięciu obserwacji odstających	73

Tabela 4.17. Statystyka Walda, LR i LM dla modelu SEM przestępstw kryminalnych po usunięciu obserwacji odstających	73
Tabela 4.18. Statystyka Walda, LR i LM dla modelu SEM przestępstw kryminalnych po usunięciu obserwacji odstających i zastosowaniu macierzy binarnej	74
Tabela 4.19. Wyniki testu Shapiro-Wilka, Breuscha-Pagana oraz kryteria informacyjne dla modelu SEM przestępstw kryminalnych po usunięciu obserwacji odstających i zastosowaniu macierzy binarnej	74
Tabela 4.20. Wyniki estymacji modelu SEM dla przestępstw kryminalnych	74
Tabela 4.21. Wyniki estymacji modelu SAR dla przestępstw drogowych	76
Tabela 4.22. Wyniki estymacji modelu SAR dla przestępstw drogowych	77
Tabela 4.23. Wyniki testu Shapiro-Wilka, Breuscha-Pagana oraz kryteria informacyjne dla modelu SAR przestępstw drogowych	77
Tabela 4.24. Wyniki estymacji modelu SAR dla przestępstw drogowych po usunięciu obserwacji odstających	78
Tabela 4.25. Wyniki testu Shapiro-Wilka, Breuscha-Pagana oraz kryteria informacyjne dla modelu SAR przestępstw drogowych po usunięciu obserwacji odstających	78
Tabela 4.26. Statystyka Walda, LR i LM dla modelu SAR przestępstw drogowych po usunięciu obserwacji odstających	79