

Szkoła Główna Handlowa w Warszawie

Kolegium Analiz Ekonomicznych

mgr Kamil Gala

Autoreferat rozprawy doktorskiej pt.

**Przestrzenne aspekty ryzyka ubezpieczeniowego
w ubezpieczeniach komunikacyjnych**

Praca doktorska

napisana pod kierunkiem naukowym

dra hab. Wojciecha Bijaka, prof. SGH

Warszawa 2020

1. Przedmiot, cel i zakres pracy

Przedmiotem rozprawy doktorskiej jest przestrzenne zróżnicowanie ryzyka ubezpieczeniowego na polskim rynku ubezpieczeń komunikacyjnych obejmującym ubezpieczenia odpowiedzialności cywilnej posiadaczy pojazdów mechanicznych (OC p.p.m.) oraz ubezpieczenia *Autocasco* (AC). Wobec wieloznaczności oraz złożoności pojęcia ryzyka ubezpieczeniowego (Ronka-Chmielowiec, 2013) w niniejszej pracy jest ono rozumiane jako ryzyko związane z zawartymi umowami ubezpieczenia, realizujące się w postaci wypadków ubezpieczeniowych objętych ochroną oraz związanych z nimi odszkodowań i świadczeń. Definicja ta kładzie nacisk na prawdopodobieństwo wystąpienia wypadku ubezpieczeniowego (jednego lub wielu) oraz wysokość wynikającej z niego szkody.

Literatura przedmiotu (np. Taylor, 2001; Brouhns i in., 2002; Gala, 2017) wskazuje, że ryzyko ubezpieczeniowe (mierzone np. za pomocą wskaźnika częstości szkód) wykazuje zauważalne zróżnicowanie przestrzenne. Głównym celem rozprawy jest ocena istotności wymiaru geograficznego jako czynnika pozwalającego na lepszą ocenę ryzyka ubezpieczeniowego w ubezpieczeniach komunikacyjnych. Do tego celu zostały opracowane i wykorzystane narzędzia łączące ze sobą doświadczenia ekonometrii przestrzennej oraz aktuarialnej taryfikacji *a priori*.

Taryfikacja *a priori* jest elementem procesu oceny ryzyka ubezpieczeniowego, w którym składka ustalana jest w momencie zawarcia umowy na podstawie obserwowalnych cech ubezpieczonego oraz przedmiotu ubezpieczenia (np. pojazdu lub nieruchomości). Ubezpieczenia komunikacyjne, przede wszystkim ubezpieczenie OC p.p.m. oraz ubezpieczenie AC, stanowią znaczną część polskiego rynku ubezpieczeń majątkowych, co wynika z dużej liczby pojazdów mechanicznych w Polsce oraz obowiązkowości ubezpieczenia OC p.p.m. Masowy charakter tych ubezpieczeń sprawia, że podstawowym narzędziem służącym do ustalania składki są metody statystyczne.

Na przestrzeni lat ukształtowały się zarówno specyficzne metody aktuarialne (biorące pod uwagę np. asymetryczne rozkłady zmiennych objaśnianych), jak i zbiór czynników uważanych za dobre predyktory powstawania szkód ubezpieczeniowych, takie jak wiek ubezpieczonego, pojemność silnika pojazdu czy też region jego użytkowania. Równolegle, w dużej mierze bez powiązania ze statystyką aktuarialną, rozwijają się metody ekonometrii przestrzennej. Metody te służą do badania zjawisk umiejscowionych w przestrzeni geograficznej, najczęściej charakteryzujących się przestrzennym zróżnicowaniem oraz powiązaniem między procesami zachodzącymi w sąsiadujących ze sobą obszarach (tzw.

autokorelacja przestrzenna). Można wskazać (por. np. Lee i Pace, 2009) wiele występujących w rzeczywistości sytuacji, które prowadzą do zaobserwowania autokorelacji przestrzennej – zarówno w wyniku rzeczywistych oddziaływań między regionami (tzw. efekt zarażania), jak i w wyniku błędów pomiaru lub pominięcia istotnych zmiennych o charakterze przestrzennym. Szczególnie ten drugi efekt może być istotny w ubezpieczeniach komunikacyjnych, biorąc pod uwagę, że granice jednostek podziału administracyjnego nie odzwierciedlają specyfiki ruchu drogowego oraz zjawisk pogodowych, które to czynniki wpływają na ryzyko powstania szkód komunikacyjnych. Niezależnie jednak od faktycznego źródła autokorelacji, pominięcie jej podczas budowy modelu może spowodować, że otrzymane estymatory parametrów nie będą posiadały pożądaných własności, takich jak zgodność, nieobciążoność czy efektywność.

Rozprawa obejmuje swoim zakresem teoretyczną oraz empiryczną analizę modeli statystycznych stosowanych do taryfikacji *a priori* uwzględniających przestrzenny charakter danych ubezpieczeniowych. Pierwszy rozdział jest poświęcony taryfikacji *a priori*. W rozdziale drugim przedstawiono zagadnienia związane z danymi przestrzennymi oraz opisano podstawowe modele stosowane w ekonometrii przestrzennej. W rozdziale trzecim sformułowano przestrzenny uogólniony model liniowy definiujący pewną rodzinę modeli, a następnie przedstawiono należące do tej rodziny modele z czynnikiem wielopoziomowym, a także omówiono metody segmentacji przestrzennej. W ostatnim rozdziale rozważania teoretyczne zostały uzupełnione analizą empiryczną danych dotyczących polskiego rynku ubezpieczeń komunikacyjnych, w ramach której dokonano opisu przestrzennego zróżnicowania ryzyka ubezpieczeniowego oraz estymacji parametrów modeli predykcyjnych i oceny skuteczności omawianych metod i modeli.

2. Przyczyny podjęcia tematu badawczego – jego teoretyczne i praktyczne znaczenie

Zgodnie z art. 33 ustawy z dnia 11 września 2015 r. o działalności ubezpieczeniowej i reasekuracyjnej (Dz.U. 2020 poz. 895) zakład ubezpieczeń jest zobowiązany do ustalenia wysokości składki ubezpieczeniowej po dokonaniu oceny ryzyka ubezpieczeniowego (ust. 1). Składka ta powinna być ustalona w wysokości zapewniającej co najmniej wykonanie wszystkich zobowiązań z umów ubezpieczenia i pokrycie kosztów wykonywania działalności ubezpieczeniowej zakładu ubezpieczeń (ust. 2), a zakład gromadzi odpowiednie dane statystyczne do ustalenia składki na ich podstawie (ust. 3).

Konkurencja na polskim rynku ubezpieczeń komunikacyjnych, wzrost świadomości klientów oraz regulacje prawne wymuszają na zakładach ubezpieczeń udoskonalanie stosowanych metod oceny ryzyka. W konsekwencji można zaobserwować rosnące zainteresowanie nowymi danymi o posiadaczu pojazdu, takimi jak jego sytuacja kredytowa (Gala i Kolak, 2015), punkty karne i wykroczenia (Pinquet i in., 2011), a także danymi dotyczącymi liczby pokonywanych rocznie kilometrów oraz stylu jazdy (Lemaire i in., 2016). Należy jednak zwrócić uwagę, że skorzystanie z tych możliwości może być ograniczone ze względu na koszty, krajowe przepisy prawa czy też względy praktyczne (np. konieczność uzyskania zgody klienta na pozyskanie danych z zewnętrznego źródła). W tej sytuacji alternatywą może być bardziej efektywne wykorzystanie danych, które już są w posiadaniu zakładu ubezpieczeń. Identyfikacja właściwych zmiennych taryfowych i określenie ich siły oddziaływania pozwala nie tylko na lepsze dostosowanie składki do ryzyka związanego z umową ubezpieczenia, ale również na aktywne ukierunkowanie działań sprzedażowych zakładu ubezpieczeń na grupy klientów o niskiej szkodowości. Warto również zwrócić uwagę na to, że w szerszym ujęciu polityka cenowa zakładów ubezpieczeń może stanowić finansową zachętę do określonych zachowań mających wpływ na bezpieczeństwo ruchu drogowego (np. wybór przez kierowców samochodów o mniejszej mocy silnika).

Znaczenie aspektów przestrzennych w badaniu zjawisk ekonomicznych podkreślone zostało w pracy Anselina (1988), który jako główne zagadnienia wskazuje przestrzenną niejednorodność oraz autokorelację przestrzenną. Zwraca on jednocześnie uwagę na konieczność uwzględnienia efektów przestrzennych w specyfikacji modelu w celu właściwej estymacji jego parametrów oraz weryfikacji za pomocą testów statystycznych. Autor wymienia dwa główne źródła autokorelacji przestrzennej – błędy związane z pomiarem zjawiska oraz agregacją danych (terytorialne jednostki obserwacji nie odzwierciedlają przestrzennego rozkładu zjawisk), a także rzeczywiste procesy związane z rozprzestrzenianiem się zjawisk oraz interakcjami między różnymi obszarami w przestrzeni (np. migracjami ludności lub transportem towarów).

Oba z wymienionych źródeł autokorelacji przestrzennej można znaleźć w ubezpieczeniach komunikacyjnych. Po pierwsze, tradycyjna taryfikacja na poziomie jednostki geograficznej (kodu pocztowego, gminy, powiatu) może być obarczona błędem wynikającym z faktu, że granice tych jednostek często nie pokrywają się z występowaniem dużych aglomeracji czy ciągów komunikacyjnych (np. autostrad). Po drugie, mogą występować interakcje między różnymi obszarami, np. w postaci intensywnego ruchu drogowego między obszarami (np. dojazdy do pracy). W obu przypadkach na poziomie

danych przekrojowych zostanie zaobserwowana autokorelacja przestrzenna, którą należy uwzględnić w specyfikacji modelu taryfikacyjnego. Warto również zwrócić uwagę na fakt, że o ile zwiększenie szczegółowości stosowanego podziału na obszary mogłoby pozwolić na lepsze odzwierciedlenie przestrzennego charakteru badanego zjawiska, to konsekwencją tego podejścia jest mała liczba obserwacji w obrębie jednostek terytorialnych lub brak obserwacji dla niektórych z nich. W takiej sytuacji uzyskanie wiarygodnych wyników jest utrudnione lub wręcz niemożliwe.

Istnieje bogata literatura naukowa zarówno w obszarze aktuarialnej oceny ryzyka w ubezpieczeniach komunikacyjnych, jak i w obszarze analizy danych przestrzennych. Zagadnienia związane z aktuarialną oceną ryzyka w ubezpieczeniach komunikacyjnych, zarówno *a priori*, jak i *a posteriori*, zostały szeroko opisane m.in. w pracach Denuit i in. (2007) oraz Lemaire'a (1995). Z kolei zagadnienia związane z ekonometrią przestrzenną rozwijane były m.in. w pracach Anselina (1988) oraz LeSage i Pace (2001), natomiast w polskiej literaturze można wskazać pracę pod redakcją Suheckiego (2010). Należy jednak zwrócić uwagę na fakt, że w literaturze aktuarialnej istnieje stosunkowo niewiele prac, w których szczegółowo analizowane są przestrzenne aspekty ryzyka ubezpieczeniowego w ubezpieczeniach komunikacyjnych, przy czym żadna z tych prac nie obejmuje kompleksowo omawianego zagadnienia. W pracy Boskova i Verralla (1994) autorzy przyjmują założenia dotyczące przestrzennej struktury efektów losowych, natomiast nie podają metody ich wyodrębnienia ze składnika systematycznego modelu. W pracy Brouhns i in. (2002) szerzej przedstawiono wykorzystanie modelu Boskova i Verralla do analizy danych z rynku belgijskiego i opisano w pełni procedurę estymacji jego parametrów, natomiast pominięta została kwestia oceny zdolności predykcyjnych modelu. Z kolei w pracach Taylora (1989; 2001) wykorzystywana jest metoda przestrzennej interpolacji efektów geograficznych wyznaczanych jako reszty z uogólnionego modelu liniowego, natomiast nie są wprost przyjmowane żadne założenia dotyczące autokorelacji przestrzennej bądź przestrzennego zróżnicowania badanego zjawiska, które uzasadniałyby takie postępowanie. Znaczenie przestrzennego zróżnicowania ryzyka ubezpieczeniowego było wskazywane również w opracowaniach Raichle (1997), Brubakera i Byllsa (1997) oraz Lee i Lee (2016), aczkolwiek większa uwaga była skupiona na ryzyku związanym z ubezpieczeniem mienia (w tym pożaru, powodzi i katastrof naturalnych), a problematyka ubezpieczeń komunikacyjnych nie została szerzej podjęta. Prace te wskazują jednak na praktyczne znaczenie analiz geograficznych w ubezpieczeniach.

W związku z powyższym, niniejsza rozprawa doktorska ma na celu wypełnienie luki istniejącej na styku statystyki aktuarialnej oraz ekonometrii przestrzennej, które to dyscypliny do tej pory rozwijały się w dużej mierze niezależnie od siebie. Głównym przedmiotem zainteresowania jest tutaj zwiększenie efektywności procesu oceny ryzyka ubezpieczeniowego w ubezpieczeniach komunikacyjnych (rozumianej jako jak najlepsze dopasowanie składki do rzeczywistego ryzyka związanego z umową ubezpieczenia) poprzez uwzględnienie przestrzennych aspektów tego rodzaju ryzyka. Wiąże się to ze sformułowaniem i empiryczną weryfikacją modeli predykcyjnych, które oprócz standardowych zmiennych taryfowych (np. cech pojazdu lub jego posiadacza) biorą pod uwagę przestrzenną strukturę danych oraz potencjalne interakcje między różnymi lokalizacjami. Warto zwrócić przy tym uwagę, że przyjęte podejście daje możliwość optymalizacji procesu oceny ryzyka poprzez lepsze wykorzystanie danych będących w posiadaniu zakładu ubezpieczeń oraz danych regionalnych, które są dostępne publicznie.

Ważnym elementem pracy jest analiza empiryczna danych indywidualnych dotyczących całego rynku ubezpieczeń komunikacyjnych w Polsce, dotychczas niespotykana w polskiej literaturze aktuarialnej. Wykorzystanie danych rynkowych pozwala na przyjęcie poziomu szczegółowości niemożliwego do osiągnięcia dla pojedynczego zakładu ubezpieczeń, który dla części obszarów może nie dysponować żadnymi danymi lub posiada dla nich zbyt mało obserwacji.

Podsumowując, od strony teoretycznej praca może przyczynić się do usystematyzowania i rozwoju metod taryfikacji *a priori*, które łączą ze sobą narzędzia klasycznej statystyki aktuarialnej oraz ekonometrii przestrzennej. Korzyścią od strony praktycznej będzie lepsze poznanie procesów zachodzących na polskim rynku ubezpieczeń komunikacyjnych oraz identyfikacja podejścia do taryfikacji odpowiedniego do specyfiki tego rynku.

3. Określenie podstawowych pytań badawczych, celu badań, tezy głównej i hipotez szczegółowych

Celem głównym niniejszej rozprawy jest ocena istotności przestrzennych aspektów ryzyka ubezpieczeniowego w ubezpieczeniach komunikacyjnych w procesie ustalania składki ubezpieczeniowej oraz wskazanie najlepszej metody wykorzystania przestrzennego charakteru danych ubezpieczeniowych. Należy podkreślić, że jednym z kryteriów oceny metod taryfikacji *a priori* jest również ich złożoność oraz możliwość zastosowania w praktyce. Obok celu głównego zostały sformułowane również **cele szczegółowe**, które można podzielić na trzy kategorie:

1. Cele poznawcze:

- C1. przedstawienie znaczenia danych przestrzennych w analizach związanych z ryzykiem ubezpieczeniowym w ubezpieczeniach komunikacyjnych oraz wskazanie kierunków rozwoju modeli aktuarialnych w tym obszarze,
- C2. identyfikacja i opis przestrzennego zróżnicowania oraz autokorelacji przestrzennej badanego zjawiska na podstawie rzeczywistych danych.

2. Cele metodyczne:

- C3. stworzenie wspólnych ram koncepcyjnych pozwalających na analizę różnych wariantów modeli przestrzennych,
- C4. zdefiniowanie różnych modeli pozwalających na analizę efektów przestrzennych.

3. Cel praktyczny:

- C5. estymacja parametrów rozważanych modeli przestrzennych, ocena ich efektywności taryfikacyjnej oraz identyfikacja najlepszego podejścia z praktycznego punktu widzenia.

W pracy zostały postawione następujące **hipotezy badawcze**:

H1. Ryzyko ubezpieczeniowe w ubezpieczeniach komunikacyjnych jest zróżnicowane przestrzennie oraz występuje korelacja między poziomem ryzyka w różnych obszarach. Obszary położone blisko siebie są bardziej podobne do siebie pod względem ryzyka ubezpieczeniowego niż obszary położone daleko od siebie.

H2. Precyzja wnioskowania statystycznego na temat ryzyka ubezpieczeniowego w badanym obszarze może być zwiększona dzięki wykorzystaniu danych dotyczących pozostałych obszarów.

H3. Uwzględnienie wymiaru przestrzennego w modelu taryfikacji *a priori* pozwala na zwiększenie jego możliwości predykcyjnych i efektywności procesu taryfikacji.

H4. Przestrzenne aspekty ryzyka ubezpieczeniowego mają istotne znaczenie z punktu widzenia jego właściwej oceny i ustalenia składki odpowiedniej dla ryzyka ponoszonego przez zakład ubezpieczeń.

4. Metodyka badania

Sformułowane w rozprawie cele oraz hipotezy badawcze uzasadniają podział metod badawczych na dwie grupy – teoretyczne oraz empiryczne.

Metody teoretyczne obejmują opracowanie modeli taryfikacyjnych uwzględniających przestrzenny wymiar danych oraz analizę ich własności. W tej części rozprawy został sformułowany przestrzenny uogólniony model liniowy, natomiast pogłębionej analizie został poddany jego szczególny przypadek – model z czynnikiem wielopoziomowym (ang. *multi-level factor model*, *MLF* – por. Ohlsson i Johansson, 2010). Oprócz standardowego modelu *MLF* opisano również jego hierarchiczną wersję (*HMLF*), a także przestrzenny model *MLF* (*SMLF*). Model *SMLF* bazuje na autorskim uogólnieniu klasycznego modelu Bühlmann-Strauba polegającym na uchyleniu założenia o niezależności nieobserwowanych parametrów ryzyka. Przedstawiono również metody segmentacji przestrzennej, które mogą zostać wykorzystane do konstrukcji zmiennych taryfowych o charakterze geograficznym, stanowiąc w ten sposób alternatywę dla bardziej skomplikowanych modeli przestrzennych, w tym autorską metodę segmentacji opartą na algorytmie wykrywania społeczności w grafie nieskierowanym.

Metody empiryczne obejmują statystyczną analizę danych pochodzących z ogólnopolskiej bazy Ośrodka Informacji Ubezpieczeniowego Funduszu Gwarancyjnego (OI UFG). Baza ta zawiera informacje o zawartych umowach ubezpieczenia OC p.p.m. oraz AC, zdarzeniach ubezpieczeniowych oraz o wypłaconych lub odmówionych odszkodowaniach i świadczeniach. Zasilanie bazy OI UFG przez zakłady ubezpieczeń jest obowiązkowe, natomiast art. 102 ustawy z dnia 22 maja 2003 r. o ubezpieczeniach obowiązkowych, UFG i PBUK (Dz.U. 2019 poz. 2214) definiuje zakres gromadzonych w niej danych. Z kolei art. 102 ust. 7 wskazuje, że UFG może przetwarzać dane dla celów innych niż określone w ustawie, po modyfikacji, która nie pozwoli na ustalenie tożsamości osoby, której dane dotyczą. Pozwala to Ubezpieczeniowemu Funduszowi Gwarancyjnemu na pełnienie funkcji centrum zaawansowanych analiz w zakresie procesów zachodzących na polskim rynku

ubezpieczeń komunikacyjnych. W związku z tym niniejsza rozprawa realizuje nie tylko określone cele badawcze, ale również dodatkowy cel przetwarzania danych, jakim jest ocena potencjału analitycznego Ośrodka Informacji UFG oraz wsparcie (teoretycznie i empiryczne) rozwoju usług związanych z zaawansowaną analityką świadczonych przez UFG dla rynku ubezpieczeń.

Analizie zostały poddane umowy OC p.p.m. oraz AC zawarte w 2017 r. Zbiór danych dot. umów OC p.p.m. zawierał ok. 16,5 mln rekordów, natomiast w przypadku AC było to 2,8 mln rekordów. Oba zbiory zostały podzielone losowo na zbiór uczący (70% obserwacji) oraz zbiór walidacyjny (30% obserwacji). Zakres wykorzystanych danych obejmował podstawowe dane dot. pojazdu (rodzaj, marka) oraz dane dot. posiadaczy pojazdu (wiek, płeć, miejsce zamieszkania), a także dane dot. przebiegu ubezpieczenia i historii szkodowej pojazdu i jego posiadaczy.

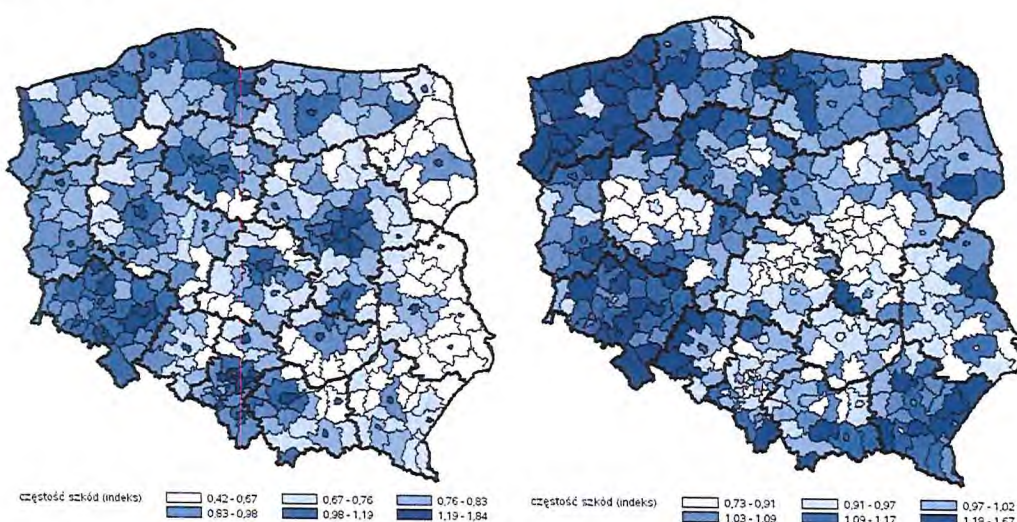
Z przyjętych celów rozprawy oraz hipotez badawczych wynika potrzeba przeprowadzenia analizy empirycznej zgodnie z metodyką obejmującą pełen proces budowy modelu predykcyjnego, obejmujący zarówno opis badanego zjawiska (cel C2), jak i budowę modeli oraz ocenę ich efektywności (cel C5). Jest to konieczne, aby właściwie ocenić istotność danych przestrzennych w procesie taryfikacji i korzyści płynące z ich wykorzystania (cel główny). Analiza empiryczna została więc podzielona na trzy części, w ramach których wyodrębniono etapy. Każdy z etapów mógł zawierać jeden lub więcej kroków.

- **Część 1: analiza eksploracyjna:**
 - **etap 1:** przygotowanie zbiorów uczących i zbiorów walidacyjnych,
 - **etap 2:** statystyczny opis rozkładu przestrzennego badanych statystyk szkodowych,
 - **etap 3:** wizualizacja statystyk szkodowych za pomocą kartogramów,
 - **etap 4:** obliczenie miar opisujących przestrzenny rozkład ryzyka ubezpieczeniowego, w tym współczynników autokorelacji przestrzennej Morana, Geary'ego i Getisa-Orda,
 - **etap 5:** wstępna selekcja zmiennych objaśniających na podstawie ich związku ze zmienną objaśnianą oraz budowa jak najlepszego modelu niezawierającego zmiennych geograficznych (nazywanego dalej „modelem bazowym”).
- **Część 2: estymacja parametrów modeli przestrzennych:**
 - **etap 1:** analiza wyników estymacji parametrów modelu bazowego,

- **etap 2:** estymacja parametrów modeli MLF przy różnych definicjach regionów,
- **etap 3:** estymacja parametrów modeli HMLF przy różnych definicjach regionów,
- **etap 4:** estymacja parametrów modeli SMLF,
- **etap 5:** przeprowadzenie segmentacji przestrzennej i analiza modeli uwzględniających uzyskane w ten sposób zmienne taryfowe.
- **Część 3: porównanie efektywności modeli z efektami przestrzennymi:**
 - **etap 1:** ocena efektywności modeli z czynnikiem wielopoziomowych za pomocą przyjętych kryteriów,
 - **etap 2:** ocena efektywności modeli uwzględniających zmienne taryfowe uzyskane w wyniku przeprowadzenia segmentacji przestrzennej.

5. Wyniki i wnioski z przeprowadzonego badania

Wyniki analizy opisowej prowadzą do wniosku, że ryzyko ubezpieczeniowego jest zróżnicowane geograficznie, o czym świadczy przestrzenny rozkład badanych statystyk szkodowych. Rozkład częstości szkód, będącej głównym przedmiotem zainteresowania niniejszej pracy, został przedstawiony na Rys. 1. Dane prezentowane są według adresu najstarszego posiadacza pojazdu, a ponadto zostały przedstawione w formie indeksu, tzn. oryginalna wartość statystyki dla regionu została podzielona przez wartość statystyki dla całego rynku.



Rys. 1. Częstość szkód (indeks) w ujęciu powiatowym w ubezpieczeniach OC p.p.m. (po lewej) i AC (po prawej).

W przypadku OC p.p.m. daje się zauważyć duży wpływ ośrodków miejskich na kształtowanie się ryzyka ubezpieczeniowego, natomiast w przypadku AC wyróżniają się w Polsce województwa zachodnie. Ryzyko ubezpieczeniowe cechuje się również przestrzenną autokorelacją, o czym świadczą wysokie wartości statystyki Morana (ponad 0,4 w przypadku częstości szkód). Warto również podkreślić, że na wysokim poziomie szczegółowości podziału geograficznego (gminy, kody pocztowe) zauważalny staje się problem związany z tym, że większość zakładów ubezpieczeń nie dysponuje danymi obejmującymi wszystkie regiony, a także niektóre regiony cechują się małą liczebnością próby. Stanowi to uzasadnienie do rozważania zastosowania metod bayesowskich i modeli z efektami losowymi.

Wyniki modelowania predykcyjnego wskazują, że model bazowy uwzględnia szeroki zakres danych, obejmujący zarówno podstawowe dane pojazdu i jego posiadacza, jak i dane dotyczące przebiegu ubezpieczenia, w tym długość okresu bezszkodowego oraz liczbę i częstość szkód. Model bazowy nie jest jednak w stanie objaśnić w pełni obserwowanego przestrzennego zróżnicowania danych, o czym świadczą rozkłady efektów losowych oszacowanych za pomocą modeli MLF na poziomie powiatu, gminy i kodu pocztowego. Z kolei modele HMLF i SMLF wskazują, że korelacja między liczbą szkód z tytułu zawartych umów występuje nie tylko w ramach regionu (co wynika z istotności regionu jako zmiennej taryfowej), ale również między regionami. Modele HMLF wskazują na istotny związek między ryzykiem ubezpieczeniowym w regionach wchodzących w skład tego samego regionu nadrzędnego (w szczególności powiatów będących w obrębie tego samego województwa), natomiast model SMLF wskazuje na dodatnią autokorelację między efektami losowymi na poziomie powiatu. Wreszcie, wyniki segmentacji przestrzennej wskazują na możliwość konstrukcji zmiennych taryfowych, które włączone do modelu bazowego poprawiają jego dopasowanie do danych, natomiast jakość takiego modelu jest porównywalna do modelu uzyskanego poprzez włączenie do modelu bazowego województwa jako zmiennej taryfowej.

Efektywność taryfikacyjna rozważanych modeli została zbadana na trzech poziomach – portfela ubezpieczeń, podportfeli wyznaczonych przez wybrane zmienne taryfowe oraz na poziomie indywidualnym. Do tego celu wykorzystano odpowiednio względny błąd prognozy, pierwiastek błędu średniokwadratowego oraz statystyki *lift*. Zastosowane kryteria efektywności pozwalają stwierdzić, że w większości przypadków uwzględnienie efektów przestrzennych pozwala na lepszą ocenę ryzyka w porównaniu do modelu bazowego. Należy jednak podkreślić, że uzyskane wyniki nie pozwalają na wybór modelu, który byłby jednoznacznie najlepszy ze względu na każde kryterium. Wzrost efektywności uzyskany

dzięki uwzględnieniu efektów przestrzennych z wykorzystaniem rozważanych modeli różni się istotnie w zależności od rodzaju umowy oraz analizowanego kryterium. Jeżeli chodzi o modele z czynnikiem wielopoziomowym, w przypadku umów AC dobre wyniki daje model HMLF z grupowaniem kodów pocztowych na podstawie pierwszych dwóch cyfr (poziom portfela oraz podportfeli), a także modele MLF i HMLF z gminą jako zmienną taryfową (poziom indywidualny). W przypadku OC p.p.m. warto wyróżnić modele na poziomie gminy – HMLF na poziomie podportfeli oraz MLF na poziomie indywidualnym. Porównywalne wyniki daje również model MLF dla kodów pocztowych. Z kolei najbardziej skomplikowany pod względem struktury stochastycznej model SMLF okazał się najlepszy jedynie w kontekście prognozowania liczby szkód dla całego portfela, natomiast przy bardziej szczegółowych kryteriach oceny korzyści płynące z jego stosowania nie były oczywiste. W przypadku segmentacji przestrzennej wszystkie rozważane metody cechują się podobną skutecznością. Zarówno rozszerzony model bazowy, jak i metoda grafowa oraz metoda *k*-średnich dają zbliżone wyniki, jeżeli chodzi o kryteria oceny modeli przyjęte w niniejszej pracy. Ich zastosowanie pozwala na poprawę trafności prognozy w ujęciu geograficznym, natomiast nie ma istotnego wpływu na prognozę w przekrojach związanych z pojazdem i jego posiadaczami.

Uzyskane wyniki pozwalają na potwierdzenie postawionych w pracy hipotez badawczych. **Hipoteza H1** została potwierdzona poprzez analizę geograficznego rozkładu statystyk szkodowych (częstości szkód, średniej szkody, składki czystej), w tym współczynników autokorelacji przestrzennej. **Hipoteza H2** została potwierdzona poprzez analizę właściwości modeli MLF, HMLF i SMLF. Wyniki estymacji wskazują, że w większości przypadków wiarygodność danych dotyczących poszczególnych regionów jest zbyt niska i wskazana jest estymacja efektów przestrzennych z uwzględnieniem danych dotyczących całego rynku. **Hipoteza H3** została potwierdzona poprzez walidację modeli na zbiorze danych, który nie został wykorzystany do estymacji parametrów tych modeli. O ile nie jest możliwe wskazanie jednego modelu, który byłby najlepszy ze względu na wszystkie rozważane kryteria, tak większość przedstawionych modeli przestrzennych pozwala na uzyskanie wyników lepszych niż te uzyskiwane przy wykorzystaniu modelu bazowego. Z kolei **hipoteza H4** została potwierdzona poprzez pozytywną weryfikację hipotez H1 – H3. Wyniki estymacji parametrów modeli potwierdzają występowanie zależności przestrzennych, m.in. poprzez istotny wpływ hierarchii w modelach HMLF czy też dodatnią autokorelację przestrzenną między ukrytymi czynnikami ryzyka w modelu SMLF. Z kolei wyniki walidacji modeli wskazują na zalety uwzględnienia wymiaru geograficznego w analizie, które są

związane z poprawą zdolności predykcyjnej modelu taryfikacyjnego i mogą zostać przez zakład ubezpieczeń przekształcone w wymierne korzyści finansowe. Mając z kolei na uwadze wyniki przeprowadzonych analiz teoretycznych i empirycznych należy również uznać, że cel główny oraz cele szczegółowe określone w niniejszej rozprawie zostały osiągnięte.

Wkład niniejszej pracy w rozwój nauki jest dwojaki. Po pierwsze, walor metodyczny pracy wynika z ogólnego zdefiniowania przestrzennego uogólnionego modelu liniowego, który może pełnić rolę wspólnych ram koncepcyjnych dla szerokiej rodziny modeli przestrzennych. Dokonano również uogólnienia klasycznego modelu Bühlmana-Strauba w celu uwzględnienia korelacji między ukrytymi czynnikami ryzyka, w niniejszej pracy interpretowanej jako autokorelacja przestrzenna, a także zbadano własności tego modelu i zaproponowano podejście do estymacji jego parametrów. Według wiedzy autora taki model nie został jeszcze opisany w literaturze aktuarialnej w kontekście oceny ryzyka ubezpieczeniowego, podobnie zaproponowana metoda segmentacji przestrzennej za pomocą algorytmu grafowego. Warto podkreślić, że prezentowane modele przestrzenne mogą być wykorzystane do analizy dowolnych danych obszarowych, nie tylko z dziedziny ubezpieczeń. Po drugie, walor poznawczy wynika z przedstawienia wyników analizy danych Ubezpieczeniowego Funduszu Gwarancyjnego, który jako jedyny podmiot posiada dane dotyczące całego polskiego rynku ubezpieczeń komunikacyjnych. Duży wolumen danych przekłada się na większą wiarygodność wyników, jednocześnie uzasadniając potrzebę stosowania metod bayesowskich, gdyż nawet w skali całego rynku dane dotyczące niektórych regionów nie mogą być uznane za wystarczająco wiarygodne ze statystycznego punktu widzenia.

Mimo swojego szerokiego zakresu niniejsza rozprawa nie wyczerpuje potencjału związanego z wykorzystaniem danych przestrzennych w analizach aktuarialnych i możliwe jest wskazanie potencjalnych kierunków dalszych badań. W zakresie metodycznym można wskazać rozważenie wariantów przestrzennego GLM opartych na przestrzennym procesie Markowa (por. Boskov i Verrall, 1994) – estymacja tych modeli jest bardziej skomplikowana w porównaniu do modeli z rodziny MLF ze względu na konieczność stosowania metod symulacyjnych. Co więcej, model SMLF jest otwarty na różne specyfikacje macierzy wariancji-kowariancji nieobserwowanych czynników ryzyka oraz alternatywne metody kalibracji jego parametrów. Szerokie pole do dalszej analizy dają również metody segmentacji przestrzennej, które przy tworzeniu segmentów mogą uwzględniać różne cechy regionów oraz łączących ich relacji. Interesującym kierunkiem może być też analiza danych punktowych dotyczących miejsca wystąpienia zdarzeń komunikacyjnych. Z kolei w zakresie

przedmiotowym możliwe jest rozszerzenie analizy na inne statystyki szkodowe, w szczególności średnią szkodę oraz składkę czystą – w pierwszym przypadku jako rozkład zmiennej objaśnianej w uogólnionym modelu liniowym można przyjąć rozkład gamma lub odwrotny gaussowski, natomiast w drugim – modele Tweediego (por. Kaas i in., 2001). O ile ogólny sposób postępowania byłby zbliżony do modelowania częstości szkód, to sam algorytm selekcji zmiennych i kryteria oceny efektywności modeli musiałyby zostać zmodyfikowane w celu uwzględnienia ciągłego (w przypadku średniej szkody) lub dyskretno-ciągłego (w przypadku składki czystej) charakteru zmiennej objaśnianej. Przeprowadzenie takiej analizy byłoby zgodne z praktycznym podejściem zakładającym równoległe modelowanie liczby oraz wartości szkód, a następnie połączenie wyników w jeden model taryfikacyjny. Znacznie rozszerzyłoby to jednak zakres niniejszej rozprawy, nie zmieniając jednocześnie głównych wniosków płynących z przeprowadzonej analizy.

6. Bibliografia

1. Anselin L., 1988, *Spatial Econometrics: Methods and Models*, Kluwer Academic Publishers, Dordrecht.
2. Anselin L., 1995, *Local Indicators of Spatial Association – LISA*, "Geographical Analysis", Vol. 27, No. 2.
3. Anselin, L., 1996, *The Moran Scatterplot as an ESDA Tool to Assess Local Instability in Spatial Association*, [w:] "Fischer, M., Scholten, H. and Unwin, D., Eds., *Spatial Analytical Perspectives on GIS*", Taylor and Francis, London, s. 111-125.
4. Bailey R.A., Simon L., 1960, *Two studies in Automobile Insurance Ratemaking*, „ASTIN Bulletin”, Vol. 1, s. 192-217.
5. Bailey, R.A., 1963, *Insurance rates with minimum bias*, „Proceedings of the Casualty Actuarial Society” L, s. 4–13.
6. Blondel V.D, Guillaume J.L, Lambiotte R., Lefebvre E., 2008, *Fast unfolding of communities in large networks*, „Journal of Statistical Mechanics: Theory and Experiment”, nr 10.
7. Boskov, M., Verrall, R.J., 1994, *Premium rating by geographical area using spatial models*, „ASTIN Bulletin” 24, s. 131-143.
8. Brouhns, N., Denuit, M., Masuy, B., Verrall, R., 2002, *Ratemaking by geographical area: A case study using the Boskov and Verrall model*, Discussion paper 0202, Publications of the Institut de statistique, Louvain-la-Neuve, s. 1-26.
9. Brubaker R., Bylls R., 1997, *Geocoding: Description and Uses*, Health Care Issues for Property/Casualty Insurers 1997 Discussion Paper Program, Casualty Actuarial Society, s. 153-164, online: <https://www.casact.org/pubs/dpp/dpp97/97dpp.pdf> [dostęp: 31.05.2020].
10. Bühlmann H., Gisler A., 2005, *A Course in Credibility Theory and its Applications*, Springer-Verlag Berlin Heidelberg.
11. Cieślak B., 2017, *Zastosowania telematyki w ubezpieczeniach komunikacyjnych*, „Roczniki Kolegium Analiz Ekonomicznych”, zeszyt 45/2017, s. 79-92.
12. Denuit M., Dhaene J., Goovaerts M., Kaas R., 2005, *Actuarial Theory for Dependent Risks: Measures, Orders and Models*, John Wiley & Sons Ltd.
13. Denuit M., Maréchal X., Pitrebois S., Walhin J., 2007, *Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems*, Wiley, New York.
14. Dionne G., Vanasse Ch., 1989, *A Generalization of Automobile Insurance Rating*

- Models: The Negative Binomial Distribution with a Regression Component*, „ASTIN Bulletin”, Vol. 19, No. 2, s. 199-212.
15. Finger, R.J., 2001, *Risk Classification*, [w:] “Foundations of Casualty Actuarial Science” 4th ed., Arlington, VA: Casualty Actuarial Society.
 16. Francis L., 2014, *Unsupervised Learning*, [w:] „Predictive Modeling Applications in Actuarial Science, Volume I: Predictive Modeling Techniques” (red. Frees E., Derrig R., Meyers G.), Cambridge University Press, New York., s. 280-314.
 17. Frątczak, E. (red.), 2012, *Zaawansowane metody analiz statystycznych*, Szkoła Główna Handlowa, Warszawa.
 18. Gala K., Kolak K., 2015, *Wykorzystanie informacji kredytowej w procesie oceny ryzyka ubezpieczeniowego w ubezpieczeniach komunikacyjnych*, „Roczniki Kolegium Analiz Ekonomicznych”, zeszyt 37/2015, s. 73-87.
 19. Gala K., 2016, *Lojalność klientów w ubezpieczeniach komunikacyjnych w świetle danych UFG*, „Studia Ekonomiczne, Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach”, nr 301, red. E. Dziwok, K. Sawicz, A. Sroczyńska-Baron, s. 43-62.
 20. Gala K., 2017, *Taryfikacja a priori z uwzględnieniem efektów przestrzennych*, „Śląski Przegląd Statystyczny”, nr 15/2017, s. 99-114.
 21. Gala K., Bobrowski M., 2017, *Lojalność klientów w ubezpieczeniach autocasco*, „Śląski Przegląd Statystyczny”, nr 15/2017, s. 81-98.
 22. Gala K., 2018, *Taryfikacja a priori w ubezpieczeniach komunikacyjnych z uwzględnieniem zależności przestrzennej*, „Roczniki Kolegium Analiz Ekonomicznych”, zeszyt 51/2018, s. 147-168.
 23. Geary R.C., 1954, *The Contiguity Ratio and Statistical Mapping*, „The Incorporated Statistician”, Vol. 5, No. 3, s. 115-146.
 24. Getis A., Ord J.K, 1992, *The analysis of spatial association by distance statistics*, „Geographical Analysis”, nr 24.
 25. Griffith D.A., 1983, *The Boundary value problem in spatial statistical analysis*, „Journal of Regional Science”, vol. 23, no. 3, s. 377–387, doi: <https://doi.org/10.1111/j.1467-9787.1983.tb00996.x>.
 26. Griffith D.A., 1992, *What is spatial autocorrelation? Reflections on the past 25 years of spatial statistics*. „Espace géographique”, tome 21, no. 3, s. 265-280; doi: <https://doi.org/10.3406/spgeo.1992.3091>
 27. Hastie, T., Tibshirani R. Friedman J., 2009, *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, New York, NY: Springer.

28. Jewell W.S., 1975, *The use of collateral data in credibility theory: a hierarchical model*, „Giornale dell’Istituto Italiano degli Attuari”, no. 38, s. 1-16.
29. Kaas R., Goovaerts M.J., Dhaene J., Denuit M., 2001, *Modern Actuarial Risk Theory*, Kluwer Academic Publishers, Dordrecht.
30. Larose D., 2005, *Discovering Knowledge in Data: An Introduction to Data Mining*, Wiley, Hoboken, NJ.
31. Lee J.S., Lee I.S., 2016, *The study of insurance premium rate GIS mapping considering the storm and flood hazard risks*, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLI-B8, 2016 XXIII ISPRS Congress, 12–19 July 2016, Prague, Czech Republic, online: <https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLI-B8/95/2016/isprs-archives-XLI-B8-95-2016.pdf> [dostęp: 31.05.2020].
32. Leja F., 1976, *Rachunek różniczkowy i całkowity*. PWN, Warszawa.
33. Lemaire J., 1995, *Bonus-Malus Systems in Automobile Insurance*. Kluwer Academic Publishers, Boston-Dordrecht-London, ISBN 0-7923-9545-X.
34. Lemaire J., Park S.C., Wang K.C., 2016, *The use of annual mileage as a rating variable*, „ASTIN Bulletin”, vol. 46, iss. 1, s. 39-69.
35. LeSage J., 1999, *The Theory and Practice of Spatial Econometrics*, Department of Economics, University of Toledo, Toledo.
36. LeSage J., Pace R. K., 2009, *Introduction to Spatial Econometrics*, BocaRaton, FL: CRC Press.
37. Loimaranta K., 1972, *Some asymptotic properties of bonus systems*, „ASTIN Bulletin”, Vol. VI. Part 3, s. 233-245.
38. Maddala G.S., 2001, *Introduction to Econometrics*, John Wiley & Sons, Ltd.
39. Mildenhall S., 1999, *A Systematic Relationship between Minimum Bias and Generalized Linear Models*, „Proceedings of the Casualty Actuarial Society” LXXXVI, pp. 393-486.
40. Moran P.A.P., 1950, *Notes on Continuous Stochastic Phenomena*, „Biometrika”, Vol. 37, No. 1/2, s. 17-23.
41. Nelder J.A., Wedderburn R.W.M., 1972, *Generalized Linear Models*, "Journal of the Royal Statistical Society", Series A (General), Vol. 135, No. 3 (1972), s.370-384.
42. Ohlsson, E., 2005, *Simplified estimation of structure parameters in hierarchical credibility*, The 36th ASTIN Colloquium, Zürich 2005.
43. Ohlsson E., Johansson B., 2010, *Non-Life Insurance Pricing with Generalized Linear*

- Models*, Springer-Verlag Berlin Heidelberg.
44. Ostasiewicz W. (red.), 2004, *Składki i ryzyko ubezpieczeniowe. Modelowanie stochastyczne*, Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu, Wrocław.
 45. Paelinck J, Klaassen L., 1979, *Spatial econometrics*. Saxon House, Farnborough.
 46. Pinquet J., Guillén M., Bolancé C., 2001, *Allowance for the age of claims in bonus-malus systems*, „ASTIN Bulletin”, 31(2), s. 337-348.
 47. Pinquet J., Dionne G., Vanasse C., Mathieu M., 2011, *Incentive Mechanisms for Safe Driving: A Comparative Analysis with Dynamic Data*. „The Review of Economics and Statistics”, 93 (1), s. 218-227.
 48. Raichle W. M., 1997, *Insurance Geographics*, Health Care Issues for Property/Casualty Insurers 1997 Discussion Paper Program, Casualty Actuarial Society, s. 141-152, online: <https://www.casact.org/pubs/dpp/dpp97/97dpp.pdf> [dostęp: 31.05.2020].
 49. Ronka-Chmielowiec W. (red.), 2013, *Metody aktuarialne. Zastosowania matematyki w ubezpieczeniach*, Wydawnictwo Naukowe PWN, Warszawa.
 50. Siddiqi N., 2006, *Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring*, John Wile & Sons, Inc, Hoboken, New Jersey.
 51. Sobiecki D., 2013, Dwustopniowe modelowanie składki za ubezpieczenie komunikacyjne OC, „Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu”, nr 312, 2013, s. 116 – 134.
 52. Suhecki B. (red.), 2010, *Ekonometria przestrzenna. Metody i modele analizy danych przestrzennych*, Wydawnictwo C.H. Beck, Warszawa.
 53. Szymańska A., 2014, *Statystyczna analiza systemów bonus-malus w ubezpieczeniach komunikacyjnych*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź.
 54. Śliwiński A., 2012, *Przestrzenne zróżnicowanie ryzyka ubezpieczeniowego a efektywność ubezpieczeń na życie*, Oficyna Wydawnicza Szkoły Głównej Handlowej w Warszawie, Warszawa.
 55. Taylor, G., 1989, *Use of spline functions for premium rating by geographic area*. „ASTIN Bulletin”, 19, s. 91–122.
 56. Taylor, G., 2001, *Geographic Premium Rating by Whittaker Spatial Smoothing*. „ASTIN Bulletin”, 31(1), s. 147–160. doi:10.2143/AST.31.1.999
 57. Tobler W., 1970, *A computer model simulating urban growth in the Detroit region*, „Economic Geography”, nr 46 (2).

58. Weibel E., Walsh J., 2008, *Territory Analysis with Mixed Models and Clustering*, Casualty Actuarial Society Discussion Paper Program.
59. Werner, G., Modlin, C., 2016, *Basic Ratemaking*, Casualty Actuarial Society, Fifth Edition, online:
https://www.casact.org/library/studynotes/Werner_Modlin_Ratemaking.pdf [dostęp: 31.05.2020]
60. Wolthuis, H., 2003, *Life Insurance Mathematics (The Markovian Model)*, IAE, Universiteit van Amsterdam.
61. Wolny-Dominiak A., Wanat S., Sobiecki D., 2018, *Selected Statistical Models in Non-Life Insurance with R*, LAP LAMBERT Academic Publishing.
62. Yao J., 2016, *Clustering in General Insurance Pricing*, [w:] „Predictive Modeling Applications in Actuarial Science, Volume II: Case Studies in Insurance” (red. Frees E., Derrig R., Meyers G.), Cambridge University Press, New York., s. 159-179.

Źródła sieciowe i akty prawne:

1. *Biuletyn roczny. Rynek ubezpieczeń 2018*, Komisja Nadzoru Finansowego, https://www.knf.gov.pl/?articleId=67227&p_id=18 (dostęp: 31.05.2020).
2. *Stanowisko Komisji Nadzoru Finansowego z dnia 28 września 2015 r. w sprawie kalkulacji wysokości składki ubezpieczeniowej*, Komisja Nadzoru Finansowego, https://www.knf.gov.pl/knf/pl/komponenty/img/kalkulacja_wysokosci_skladki_ubezpieczeniowej_28-09-15_42937.pdf (dostęp: 31.05.2020)
3. *Słownik języka polskiego*, Polskie Wydawnictwo Naukowe, <https://sjp.pwn.pl/sjp/efektywny;2556022.html> (dostęp: 31.05.2020)
4. Ustawa z dnia 23 kwietnia 1964 r. - Kodeks cywilny, Dz.U. 2019 poz. 1145 z późn. zm.
5. Ustawa z dnia 11 września 2015 r. o działalności ubezpieczeniowej i reasekuracyjnej, Dz.U. 2020 poz. 895 z późn. zm.
6. Ustawa z dnia 22 maja 2003 r. o ubezpieczeniach obowiązkowych, Ubezpieczeniowym Funduszu Gwarancyjnym i Polskim Biurze Ubezpieczycieli Komunikacyjnych, Dz.U. 2019 poz. 2214 z późn. zm.

