

**Załącznik nr 3**

do wniosku o wszczęcie postępowania habilitacyjnego  
dr Alicji Wolny-Dominiak

**Autoreferat na temat dorobku i osiągnięć  
w pracy naukowo-badawczej w języku polskim**

**Katowice, styczeń 2017**



## Autoreferat

### 1. Imię i nazwisko

Alicja Wolny-Dominiak

### 2. Posiadane dyplomy i stopnie naukowe

- Magister matematyki o specjalności: informatyczna  
Uniwersytet Śląski w Katowicach, Wydział Matematyki 1996  
Promotor: Prof. dr hab. Maciej Sablik
- Doktor nauk ekonomicznych w zakresie ekonomii,  
Akademia Ekonomiczna w Katowicach, Wydział Ekonomii, 2002  
Tytuł rozprawy:  
*Metody statystycznej analizy poziomu rezerwy szkodowej w zakładzie ubezpieczeń*  
Promotor: Prof. dr hab. Włodzimierz Szkutnik  
Recenzenci: Prof. dr hab. Wanda Ronka-Chmielowiec, Prof. dr hab. Jerzy Mika

### 3. Informacje o dotychczasowym zatrudnieniu w jednostkach naukowych

- W latach 1997-2000 byłam zatrudniona na stanowisku asystenta w Katedrze Ekonometrii na Wydziale Zarządzania Akademii Ekonomicznej w Katowicach.
- W latach 2000-2003 byłam zatrudniona na stanowisku asystenta w Katedrze Matematyki na Wydziale Zarządzania Akademii Ekonomicznej w Katowicach.
- W latach 2003-2005 byłam zatrudniona na stanowisku adiunkta w Katedrze Matematyki na Wydziale Zarządzania Akademii Ekonomicznej w Katowicach.
- Od roku 2005 jestem zatrudniona na stanowisku adiunkta w Katedrze Metod Statystyczno-Matematycznych na Wydziale Ekonomii Uniwersytetu Ekonomicznego w Katowicach (do roku 2010 Akademii Ekonomicznej w Katowicach).

#### 4. Wskazanie osiągnięcia naukowego

##### a) tytuł osiągnięcia naukowego

Jako osiągnięcie naukowe w rozumieniu art. 16 ust. 2 ustawy z dnia 14 marca 2003 o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz. U. nr 65, poz. 595 ze zm.) przedstawiam monografię pod tytułem:

***Taryfikacja w ubezpieczeniach majątkowych z wykorzystaniem modeli mieszanych***

##### b)

Autor: Alicja Wolny-Dominiak

Rok wydania: 2014

Nazwa wydawnictwa: Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach

Recenzent wydawniczy: dr hab. Stanisław Wanat

Publikacja została dofinansowana ze środków Narodowego Centrum Nauki w ramach projektu nr NN 111461540.

##### c) omówienie celu naukowego ww. pracy i osiągniętych wyników wraz z omówieniem ich ewentualnego wykorzystania

Celem głównym monografii jest zastosowanie modeli mieszanych z efektami stałymi i losowymi w procesie taryfikacji (odpowiedniki używane w ang. *ratemaking*, *experience pricing*) istotnie rozszerzające stosowane obecnie przez ubezpieczycieli modele z efektami stałymi. Podjęte w monografii zagadnienie dotyczy zatem taryfikacji w portfelu ubezpieczeń majątkowych, dla którego ubezpieczyciel dysponuje licznym (masowym) zbiorem polis ubezpieczeniowych oraz czas od momentu zajścia zdarzenia do momentu zgłoszenia szkody jest na tyle krótki, że może być pominięty. Przy tych założeniach, rozpatrywany proces taryfikacji jest najczęściej stosowany w takich ubezpieczeniach jak ubezpieczenia komunikacyjne OC/AC, ubezpieczenia nieruchomości czy ubezpieczenia turystyczne.

Wraz z rozwojem technik komputerowych i bazodanowych, ubezpieczyciele odeszli od stosowania metod deterministycznych bazujących na procencie przechodząc do modeli stochastycznych klasy GLM (generalized linear model). Dzięki specjalistycznemu oprogramowaniu dedykowanemu do taryfikacji (np. Emblem), te popularne w wielu dziedzinach modele stały się powszechnie używane w codziennej praktyce zakładów ubezpieczeń. W literaturze z zakresu taryfikacji można znaleźć wiele rozszerzeń klasycznych

 3

modeli GLM, które uwzględniają różnorodne efekty występujące w portfelach ubezpieczeniowych. W monografii zaproponowano autorskie modyfikacje modeli GLM na potrzeby procesu taryfikacji. Dodatkowo zaproponowano wykorzystanie szeroko stosowanej w genetyce klasy hierarchicznych modeli GLM (HGLM) pozwalających doskonale rozwiązać pięć zidentyfikowanych problemów występujących w procesie taryfikacji. Do poprawnego zrozumienia tych zagadnień niezbędne jest zdefiniowanie trzech podstawowych pojęć: ryzyka, składki czystej (odpowiedniki używane w ang. *pure risk premium*, *risk premium*, *pure premium*) oraz taryfikacji.

#### *Zdefiniowanie ryzyka i składki czystej*

W taryfikacji ryzyko utożsamiane jest z pojedynczą polisą i definiuje się je jako zmienną losową odpowiadającą łącznej wartości szkód dla tego ryzyka. Istotą w działalności ubezpieczyciela jest wyznaczenie składki ryzyka dla pojedynczego ryzyka. Składkę tą definiuje się jako wartość oczekiwaną łącznej wartości szkód. Oczywiście ubezpieczyciel nie zna wartości tej charakterystyki, a zatem wymaga ona szacowania. Obecnie do wyznaczenia wartości składki czystej stosuje się modele statystyczne GLM pozwalające na estymację wartości oczekiwanej łącznej wartości szkód dla pojedynczego ryzyka poprzez estymację parametrów tych modeli.

#### *Zdefiniowanie taryfikacji*

Modele GLM znajdują tak szerokie zastosowanie w praktyce, ponieważ podstawową własnością portfela ryzyk jest jego **niejednorodność**. Oznacza to, że ryzyka generują różne łączne wartości szkód. Powoduje to, iż przypisanie jednej składki ryzyka każdemu ryzyku nie jest sprawiedliwe. Praktyką zakładów ubezpieczeń jest więc przeprowadzanie **taryfikacji definiowanej jako grupowanie portfela ryzyk na grupy, w których każdemu ryzyku przypisuje się jednakową składkę czystą**. Grupowanie to odbywa się na podstawie czynników powodujących niejednorodność portfela.

Czynniki te dzieli się na:

- czynniki obserwowalne (w momencie zawierania umowy ubezpieczeniowej) - czynniki opisujące osobę ubezpieczającą się, czynniki opisujące przedmiot ubezpieczenia oraz zmienna przestrzenia (w sensie regionu geograficznego),
- czynniki nieobserwowalne - umiejętności kierowcy, bezpieczeństwo dzielnicy, w której znajduje się nieruchomość itp.

Czynniki obserwowalne dzielą portfel na grupy ryzyk (np. ryzyka podzielone pod względem wieku ubezpieczającego się w ubezpieczeniach komunikacyjnych). Czynniki nieobserwowalne dzielą się następująco:

- czynniki nieobserwowalne, które można sprowadzić do czynników obserwowalnych w momencie zawierania umowy ubezpieczeniowej - region użytkowania pojazdu (nieobserwowalny) -> region rejestracji pojazdu (obserwowalny),
- czynnik nieobserwowalny odpowiadający przykładowo wewnętrznym cechom osoby ubezpieczającej się - czynnik indywidualny dla każdego ryzyka traktowany jako zmienna losowa o określonym rozkładzie.

Uwzględniając czynniki ryzyka, proces taryfikacji jest przeprowadzany w dwóch etapach determinowanych przez te czynniki:

- taryfikacja a priori - grupowanie portfela ryzyk na grupy ryzyk jednorodne pod względem czynników obserwowalnych i wyznaczenie dla każdej grupy ryzyk jednakowej składki ryzyka a priori,
- taryfikacja a posteriori - korekta składek a priori uwzględniająca indywidualny czynnik nieobserwowalny.

### *Problemy występujące w procesie taryfikacji*

Proces taryfikacji jest ciągle rozwijany zarówno w literaturze przedmiotu jak i w codziennej praktyce ubezpieczycieli. Wychodząc od praktyki ubezpieczeniowej, a następnie analizując problemy poruszane w opracowaniach naukowych **zauważyłam istotne problemy** związane z wykorzystywanymi modelami GLM.

### *Problem 1. Założenie postaci rozkładu w modelu*

Modele GLM ograniczają się jedynie do rozkładów pochodzących z dyspersyjnej rodziny rozkładów wykładniczych. Często modelowane zmienne losowe nie spełniają tego założenia. np. rozkład liczby szkód.

### *Problem 2. Efekt nadmiernej liczby zer oraz efekt nadmiernej dyspersji*

Dla liczby szkód zakłada się klasycznie rozkład Poissona. Jednak w portfelach ubezpieczeniowych występuje zazwyczaj znaczna liczba ryzyk, dla których nie wystąpiła żadna szkoda. Powstaje wtedy efekt nadmiernej liczby zer. Dodatkowo efekt ten powoduje efekt nadmiernej dyspersji.

**Problem 3.** Założenie niezależności ryzyk w portfelu

Oznacza to, iż zakłada się niezależność pomiędzy zmiennymi losowymi (ryzykami). Założenie to determinuje klasę modeli, którą można stosować w procesie taryfikacji. W rzeczywistości można zaobserwować, iż często założenie to nie jest w portfelu spełnione.

**Problem 4.** Czynniki nieobserwowalne i składka a posteriori

Czynniki nieobserwowalne powodują najczęściej dodatkowy wzrost nadmiernej dyspersji. Zakłada się wtedy rozkład ujemny-dwumianowy, a jeśli oba efekty występują jednocześnie - rozkład ZINB (ang. *zero-inflated negative-binomial*). Oznacza to, iż zmienna losowa reprezentująca czynnik nieobserwowalny przyjmuje rozkład Gamma. Jednak przyjęcie innego rozkładu dla czynnika może dawać lepsze wyniki.

**Problem 5.** Wykorzystywanie zgromadzonych danych

Przy obecnym zaawansowaniu systemów bazodanowych ubezpieczyciele od lat gromadzą duże ilości danych, których jednak nie wykorzystują w pełni w taryfikacji. Można zauważyć, że portfele ubezpieczeniowe mają raczej charakter przekrojowo-czasowy a nie jedynie przekrojowy. Klienci zakładów najczęściej odnawiają swoje polisy w kolejnych latach.

*Zdefiniowanie szczegółowych celów osiągnięcia naukowego*

W oparciu o powyższe spostrzeżenia uzasadnione jest **rozszerzenie modeli GLM** stosowanych w zakładach ubezpieczeń oraz **modyfikację modeli mieszanych** proponowanych w literaturze problemu a nie wdrożonych jeszcze w praktyce. W związku z powyższym w monografii postawiono następujące cele szczegółowe:

**Cel pierwszy:** zaproponowanie modyfikacji modeli GLM głównie pod względem założeń odnośnie przyjmowanych rozkładów oraz konstrukcja stosowanych rozkładów – dotyczy problemów 1-2.

**Cel drugi:** wprowadzenie efektów losowych do modeli GLM i uzasadnienie przejścia od modeli GLM do modeli mieszanych – dotyczy problemów 3-4.

**Cel trzeci:** zaproponowanie modeli mieszanych klasy HGLM oraz NLMM (ang. *non-linear mixed model*) do taryfikacji uwzględniając specyfikę i rozmiar gromadzonych danych – dotyczy problemów 3-5.

**Cel czwarty:** opracowanie procedur estymacji w zaproponowanych modelach i pełna ich implementacja w programie R pozwalająca na bezpośrednią aplikację przez ubezpieczyciela – dotyczy problemów 3-5.

Monografia wskazana jako osiągnięcie naukowe wychodzi na przeciw tak określonym celom. Zasadniczo powyższe cele mają charakter głównie naukowy przyczyniając się moim zdaniem do rozwoju modelowania statystycznego w analizowanym obszarze. Jednak rezultaty mają bezpośrednie przełożenie na praktykę ubezpieczycieli. Realizacja tych celów stanowi **autorski wkład w rozwój naukowy oraz praktyczny procesu taryfikacji.**

#### *Struktura monografii*

Monografia składa się z 5 rozdziałów a jej struktura wynika bezpośrednio z procedury przeprowadzania taryfikacji.

**Rozdział pierwszy** przedstawia ogólną koncepcję rozszerzenia obecnego podejścia modeli GLM w taryfikacji do stosowania modeli mieszanych. Aby to uzyskać opisane zostały dwa zagadnienia poruszane w monografii: taryfikacja oraz wyznaczanie wartości składki czystej dla pojedynczego ryzyka. Uwagę skupiono na zdefiniowaniu pojęć ubezpieczeniowych oraz odpowiadających im pojęć statystycznych w celu ujednoczenia procesu taryfikacji proponowanego w pracy, gdyż w literaturze przedmiotu autorzy stosują różną nomenklaturę. Szczegółowo scharakteryzowano pojęcie ryzyka jako zmiennej losowej oraz pojęcie wartości składki jako estymator wartości oczekiwanej zmiennej losowej. Przedstawiono dwa możliwe podejścia w taryfikacji:

- podejście dwumodelowe – stosuje się model wartości szkody oraz model częstości szkód,
- podejście jednomodelowe – stosuje się model łącznej wartości szkód.

W obu podejściach opisano wykorzystywane w praktyce modele statystyczne poruszając następujące zagadnienia:

- czynniki ryzyka występujące w modelach oraz związane z tym etapy a priori i a posteriori w taryfikacji,
- niezależność ryzyk w portfelu,
- charakterystyka danych,
- estymacja parametrów modeli,
- efekt nadmiernej dyspersji oraz nadmiernej liczby zer.

5  
7

Ponadto podjęto temat celowości zastosowania modeli mieszanych w procesie taryfikacji oraz rozważono problem danych występujących w portfelach ubezpieczeniowych.

Ten fragment pracy można podsumować wyodrębnieniem problemów 1-5 opisanych powyżej. Pozwoliło to dalej na realizację celów 1-4 poprzez **zapropozowanie i skonstruowanie** modeli aplikowalnych w taryfikacji.

## 1. Modele z efektami stałymi

- 1.1. Nieliniowy model NLM (*non-linear model*) – ZIGP (*zero-inflated generalised Poisson*)/ZINB częstości szkód z efektami stałymi uwzględniający jednocześnie efekt nadmiernej liczby ryzyk bezszkodowych w portfelu
- 1.2. Model GLM-ZA wartości szkody z dodanymi ryzykami bezszkodowymi
- 1.3. Model GLM-CPOIS (*compound Poisson*) łącznej wartości szkód rozszerzający złożony rozkład Tweedie (Poisson-Gamma) o modelowanie ryzyk bezszkodowych

## 2. Modele mieszane

- 2.1. zależności występujące dla klastrów (dane przekrojowe)
  - 2.1.1. Hierarchiczny model GLM (HGLM) w taryfikacji a priori
  - 2.1.2. Nieliniowy model mieszany NLMM-ZIGP – rozszerzenie modelu z pkt. 1.1
  - 2.1.3. Model HGLM-ZA – rozszerzenie modelu z pkt. 1.2
  - 2.1.4. Model HGLM-CPOIS – rozszerzenie modelu z pkt. 1.3
- 2.2. zależności występujące w czasie (dane przekrojowe-czasowe)
  - 2.2.1. Hierarchiczny model GLM (HGLM) w taryfikacji a posteriori – uwzględnienie indywidualnego czynnika nieobserwowalnego mierzonego przy pomocy liczby szkód w czasie

**Rozdział drugi** zawiera opis rozkładów prawdopodobieństwa zmiennych losowych mających zastosowanie w procesie taryfikacji, zarówno w podejściu dwumodelowym, jak i w podejściu jednomodelowym. W pierwszej kolejności omówiono dyspersyjną rodzinę rozkładów wykładniczych (EDM) skupiając się na eksploatowanej w praktyce podrodzinie rozkładów Tweedie. Dalej scharakteryzowano zaawansowane rozkłady liczby szkód, które nie należą do rodziny EDM, ale uwzględniają efekty występujące w danych wykorzystywanych w taryfikacji, opisane w rozdziale 1:

- efekt nadwyżki wartości zerowych - rozkład ZIP (*zero-inflated Poisson*),
- efekt nadmiernej dyspersji – rozkład NB,



- efekt nadwyżki wartości zerowych i efekt nadmiernej dyspersji – ZIGP, ZINB.

Dodatkowo skonstruowano rozkład (*ZA zero-adjustment*) uwzględniający wartości zerowe, mający zastosowanie w modelowaniu wartości szkód dla ryzyka, np. w przypadku gdy nie jest znana liczba szkód. W ostatniej części omówiono przypadek złożonego rozkładu Poissona łącznej wartości szkód, który jest wykorzystywany w jednomodelowej taryfikacji. Przyjmuje się wtedy założenie rozkładu Gamma pojedynczej wartości szkody uzyskując przypadek rozkładu Tweedie dla  $1 < p < 2$  (ang. *Tweedie compound Poisson distribution*).

W rozdziale 3 podjęto tematykę dotyczącą samej metodologii związanej z modelami mieszanymi. Dane przekrojowo-czasowe, dane o strukturze klastrowej lub niekompletne generują zależności, które wpływają bezpośrednio na konstrukcje modelu. Zignorowanie tego faktu może prowadzić do zafałszowanych wyników. Dobrym rozwiązaniem jest wykorzystanie modeli mieszanych, w miejsce modeli z efektami stałymi. Uwagę skupiono na modelach nieliniowych, gdyż de facto tylko taka klasa modeli znajduje zastosowanie w poruszanych zagadnieniach ubezpieczeniowych. Popularne modele liniowe nie są prawidłowe, chociażby ze względu na multiplikatywny charakter procesu taryfikacji.

Przedstawiono w kolejności poszczególne klasy modeli, **jedynie te, które dalej zaproponowano do wykorzystania w taryfikacji**. Rozważania rozpoczęto od dobrze znanych modeli:

1. Nieliniowy model z efektami stałymi (NLM),
2. Uogólniony model liniowy z efektami stałymi (GLM),
3. Uogólniony model liniowy z efektami stałymi i modelowanym parametrem dyspersji (DGLM).

Scharakteryzowano krótko najważniejsze z punktu widzenia taryfikacji elementy modelowania: macierz wariancji-kowariancji  $V$  dla ryzyk (por. wzór (3.7)), funkcję wiążącą (*link function*), offset, modelowanie parametru dyspersji. Opisano krótko metody estymacji parametrów modeli i omówiono programistyczne możliwości estymacji w programie R.

Dalej rozważania kontynuowano podejmując tematykę modeli mieszanych. Omówiono następujące modele:

4. Nieliniowy model z efektami stałymi i losowymi (NLMM),
5. Uogólniony model liniowy z efektami stałymi i losowymi (GLMM),
6. Hierarchiczny uogólniony model liniowy z efektami stałymi (HGLM).

Rozważania rozszerzono (w porównaniu z GLM) o wyznaczenie brzegowej wartości oczekiwanej i wariancji dla pojedynczego ryzyka oraz macierzy wariancji-kowariancji  $V$  (por.

wzory począwszy od (3.41)). Rozdział kończy obszerna część dotycząca metod estymacji parametrów modelu mieszanego: ogólnej metody brzegowej wiarygodności oraz hierarchicznego algorytmu IWSL. Opisano również problemy związane z implementacją komputerową metod w programie R oraz, w koniecznych przypadkach, w programie SAS.

Rozdział 4 poświęcono w całości modelowaniu w procesie taryfikacji. Punktem wyjścia była prezentacja obecnej praktyki ubezpieczeniowej i dalej wyznaczenie wzoru na składkę czystą. Przedstawiono w pierwszej kolejności **dwumodelowe** podejście stosujące modele GLM z efektami stałymi: wartości szkody oraz częstości szkód. Rozważania zakończono przykładem empirycznym, w którym estymowano parametry modeli: GLM-IG (*inverse-Gaussian*) – model wartości szkody i GLM-POIS – model częstości szkód. W obliczeniach wykorzystano **rzeczywiste dane z polskiego rynku ubezpieczeniowego**.

Dalej podjęto tematykę uwzględniania w taryfikacji ryzyk bezszkodowych (nie wystąpiła żadna szkoda) istotnie wpływających na wartość składki czystej w grupach. W tym celu **skonstruowano model częstości szkód NLM-ZIGP** będący rozszerzeniem popularnego modelu ZIP uwzględniającego efekt nadmiernej liczby wartości zerowych o parametr modelujący dyspersję. Model ten nie jest formalnie modelem GLM, dlatego też dodatkowo wprowadzono funkcję wiarygodności będącą podstawą do implementacji komputerowej modelu. Następnie podjęto tematykę modelowania wartości zerowych dla ryzyk bezszkodowych w modelu wartości szkody, w którym **złożono skonstruowany w rozdziale 2 rozkład ZA**. Wyprowadzono wartość oczekiwaną oraz wariancję dla pojedynczego ryzyka, a następnie – postać funkcji wiarygodności. Ze wzoru (4.29) widać, że estymacja parametrów modelu sprowadza się do szacowania dwóch modeli – dla części zerowej i części niezerowej. Stąd implementacja komputerowa w R sprowadza się do szacowania dwóch modeli GLM. Na końcu zaprezentowano skonstruowane modele na przykładzie empirycznym.

W ostatniej części rozdziału rozważono taryfikację jednomodelową, w której skonstruowano model łącznej wartości szkody w dwóch przypadkach. Pierwszym przypadkiem jest model GLM wykorzystujący rozkład Tweedie jako przypadek złożonego rozkładu Poissona (Poisson-Gamma), natomiast w drugim – model, w którym dodano nadwyżkę wartości zerowych (ryzyk bezszkodowych) do rozkładu Poissona. Zdefiniowano składkę czystą oraz przedstawiono działanie modelu na przykładzie wskazując na możliwości praktycznego wdrożenia tego modelu.

W rozdziale 5 wprowadzono modele mieszane jako rozszerzenie modeli GLM. Istotą stosowania tych modeli jest uwzględnienie efektów występującego w portfelach ryzyk, jakim są różnego rodzaju **zależności**. Założenie niezależności ryzyk określa się w literaturze mianem założenia fundamentalnego w taryfikacji a priori jednak w praktyce łatwo znaleźć sytuacje, w których ryzyka będą zależne (zob. opis i rysunki w roz. 1). Można powiedzieć wtedy, że portfel ryzyk ma strukturę klastrową, gdzie wewnątrz klastra ryzyka są zależne. Przykładowym klastrem jest zbiór ryzyk, którym przypisano ten sam region geograficzny. Jednym z narzędzi służących do modelowania tych zależności są **modele mieszane**. W monografii **zaproponowano** zastosowanie modeli mieszanych dla portfeli o strukturze klastrowej, wprowadzając efekty losowe do modeli z efektami stałymi. W tabeli 1 zawarto charakterystykę efektów losowych w kontekście taryfikacji.

Tabela 1. Charakterystyka losowego czynnika ryzyka w taryfikacji

Przypadek	Charakterystyka losowego czynnika ryzyka	Przykład
<i>Taryfikacja a priori</i>		
Przypadek 1	Losowy czynnik ryzyka uwzględniający zależności w grupach danych według kategorii zmiennej	Obszar geograficzny – ryzyka, którym jest przypisany ten sam obszar są zależne, a w przeciwnym przypadku są niezależne
Przypadek 2	Czynnik losowy uwzględniający zależności w grupach danych według polis w portfelu	W portfelu znajduje się ryzyko (lub więcej ryzyk), dla którego wystąpiła więcej niż jedna szkoda i w danych zarejestrowano kilka pojedynczych szkód dla jednego ryzyka
Przypadek 3	Czynnik losowy to czas (kategorie to okresy trwania polisy dla ryzyka)	W portfelu, który ma charakter przekrojowo-czasowy grupa składa się z pojedynczego ryzyka w kolejnych okresach
Przypadek 4	Losowy czynnik ryzyka, dla którego obserwuje się jedynie część kategorii	Marka pojazdu – w danych nie rejestruje się wszystkich możliwych marek samochodów, zmienna posiada bardzo wiele (kilka tysięcy) poziomów
<i>Taryfikacja a posteriori</i>		
	Czynnik losowy to identyfikator ryzyka w danych przekrojowo-czasowych	Szacuje się liczbę szkód dla pojedynczego ryzyka w kolejnym okresie na podstawie historii szkodowości, co służy pomiarowi czynnika nieobserwowalnego, np. skłonności kierowcy do szybkiej jazdy

Istotnym problemem w wykorzystywaniu modeli mieszanych jest estymacja parametrów. Dla klasycznych modeli taryfikacyjnych GLM bezpośrednio zastosowanie znajduje metoda największej wiarygodności. Jednak w sytuacji, gdy konstruuje się model mieszany nie można wskazać jednoznacznie metody estymacji. W literaturze przedmiotu zaproponowano wiele różnorodnych metod, zarówno w podejściu klasycznym jak i w podejściu bayesowskim. Specyfika zagadnienia taryfikacji, jak również obecna praktyka zakładów ubezpieczeń stosujących modele GLM skłania do tego, aby w przypadku portfeli o strukturze klastrowej stosować modele HGLM wykorzystując w estymacji rozszerzoną wiarygodność (ang. *extended likelihood*). Ponadto model HGLM pozwala na modelowanie nieobserwowalnych czynników ryzyka istotnych np. w modelach wiarygodności. Niniejsza praca proponuje nowatorskie i twórcze zastosowanie modelu HGLM w taryfikacji.

W pierwszej części rozdziału skonstruowano model HGLM dla portfeli o strukturze klastrowej przy założeniu rozkładu Tweedie. W zależności od wartości parametru  $p$  może on służyć do modelowania liczby szkód, średniej wartości szkody czy łącznej wartości szkody dla pojedynczego ryzyka:

- $p=1$  - model HGLM z warunkowym rozkładem Poissona - mieszany model częstości szkód dla ryzyka mający zastosowanie w dwumodelowej taryfikacji a priori,
- $1 < p < 2$  - model HGLM z warunkowym rozkładem Tweedie (w sensie złożonego rozkładu Poissona) - mieszany model łącznej wartości szkód dla ryzyka mający zastosowanie w jednomodelowej taryfikacji a priori,
- $p=2$  - model HGLM z warunkowym rozkładem gamma - mieszany model wartości szkody dla ryzyka mający zastosowanie w dwumodelowej taryfikacji a priori,
- $p=3$  - model HGLM z warunkowym rozkładem odwróconym-normalnym - mieszany model wartości szkody dla ryzyka mający zastosowanie w dwumodelowej taryfikacji a priori.

Ponadto przedstawiono model GLMM-Tweedie zaczerpnięty z literatury przedmiotu, jednak rozpatrywany jako specyficzny przypadek modelu HGLM.

W dalszej części skonstruowano modele częstości szkód NLMM-ZIGP oraz wartości szkody HGLM-ZA z uwzględnieniem nadwyżki ryzyk bezszkodowych wykorzystując analogiczne modele z efektami stałymi omawiane w rozdziale 4. Zaimplementowaną metodę estymacji modelu NLMM w programie SAS. Dalej rozważono model mieszany HGLM-CPOIS do taryfikacji jednomodelowej wraz z implementacją w programie R.

Istotną częścią taryfikacji jest wyznaczenie składki czystej rozumianej jako pewna wartość oczekiwana. Dlatego też **dla każdego proponowanego modelu mieszanego wyprowadzono**: wartość oczekiwaną warunkową oraz wartość oczekiwaną brzegową dla ryzyka jak również macierz wariancji-kowariancji obrazującą zależności występujące portfelu. Szczegółowe obliczenia dla różnych przypadków zawarto w Dodatku.

W ostatniej części rozdziału skupiono się na eksploatowanym w aktuariacie **zagadnieniu związanemu z teorią wiarygodności**, jakim jest taryfikacja a posteriori i składka a posteriori. Jak wspomniano w rozdziale 1, na wartość składki czystej dla pojedynczego ryzyka, poza czynnikami obserwowalnymi, wpływa także indywidualny czynnik nieobserwowalny. Przykładem jest ubezpieczenie komunikacyjne, w którym czynnik nieobserwowalny jest utożsamiany z indywidualnymi cechami kierowcy (osoby ubezpieczającej się) wpływającymi na szkodowość danego ryzyka. Jako, iż czynnik ten jest nieobserwowalny, jest traktowany jako realizacja zmiennej losowej o z góry zadany rozkładzie prawdopodobieństwa. Tradycyjnie w taryfikacji indywidualny czynnik nieobserwowalny jest uwzględniany w modelowaniu liczby szkód. Na wstępie rozważono szeroko omawiane w literaturze zagadnienie **mieszanej Poisson rozkładów** (ang. *mixed Poisson*) pozwalającej na uwzględnienie czynnika nieobserwowalnego w portfelu z **danymi przekrojowymi**. Wychodząc od tego podejścia zaproponowano modele HGLM i NLMM bazujące na **danych przekrojowo-czasowych**. W modelach tych **indywidualny czynnik nieobserwowalny jest mierzony** poprzez obserwację liczby szkód w kolejnych okresach trwania polisy. Zauważono, że modele te można potraktować jako pewien specyficzny przypadek modelu mieszanego dla danych o strukturze klastrowej (por. Tab. 5.1). Podobnie jak wyżej rozpatrywano efekty ZI oraz nadmiernej dyspersji dla liczby szkód przechodząc do rozkładu ZIP/ZIGP. Podobnie jak w przypadku modeli wcześniejszych, wyprowadzono wzory na brzegową wartość oczekiwaną, wariancję oraz kowariancję ryzyk. Wskazano metodę estymacji parametrów modelu oraz sposób implementacji komputerowej modelu w programie **R** lub w przypadku, gdy nie było to możliwe, w programie **SAS**. Wyznaczono składkę czystą w podejściu dwumodelowym oraz jednomodelowym. W celu zobrazowania działania modeli pokazano przykłady empiryczne na rzeczywistych danych.

## **Dodatek A - wyprowadzenia**

Zawiera wyprowadzenia wzorów oraz uwagi techniczne.

## **Dodatek B – opis portfeli ubezpieczeniowych**

W przykładach wykorzystano dwa portfele polis pochodzące z polskiego rynku ubezpieczeniowego, oznaczane jednolicie w całej pracy jako danePL oraz danePLpanel. Jeden portfel zawiera dane w układzie przekrojowym, natomiast drugi dane w układzie przekrojowo-czasowym. Dodatek zawiera szczegółowy opis danych.

Kod programu R dostępny na stronie internetowej <http://veb.ue.katowice.pl/wobli/> w celu lepszego zrozumienia konstruowanych modeli oraz popularyzację uzyskanych wyników.

W estymacji parametrów większości modeli zastosowano program komputerowy R. Wyjątkiem jest model klasy NLMM z założeniem rozkładu ZIGP, dla którego wykorzystano program SAS. W większości przypadków zostały użyte funkcje z różnorodnych pakietów programu R. jednak pewne nowe zastosowania wymagały przeprowadzenia implementacji komputerowej metody estymacji opartej na h-wiarygodności (por. Lee and Nelder, 2006, str. 180-220).

### *Uzyskane rezultaty*

W monografii zrealizowałam postawione cele 1-4 a co za tym idzie w dużym stopniu rozwiązałam zidentyfikowane problemy 1-5. Poza głównymi celami, w pracy podjęłam wiele mniejszych zagadnień, które wyniknęły w trakcie prac zarówno merytorycznych jak również programistycznych. Podsumowując, jako najważniejsze rezultaty uzyskane w monografii oraz stanowiące mój wkład w rozwój nauk ekonomicznych w dyscyplinie finanse wskazuję:

- propozycję wykorzystania modeli mieszanych w procesie taryfikacji i szacowaniu składki czystej w ubezpieczeniach nieżyciowych wraz z obszernym uzasadnieniem,
- konstrukcję modeli klasy HGLM oraz NLMM i opracowanie metod estymacji parametrów dla portfeli o strukturze klasowej,
- konstrukcję modeli klasy HGLM oraz NLMM i opracowanie metod estymacji parametrów z uwzględnieniem indywidualnego nieobserwowalnego czynnika ryzyka,
- implementację komputerową powyższych metod estymacji oraz opracowanie przykładów empirycznych działania skonstruowanych modeli dla portfeli ubezpieczeniowych pochodzących od ubezpieczyciela działającego na rynku polskim,
- liczne wyprowadzenia analityczne na potrzeby wyznaczania składki czystej takie jak postać rozkładu ZA czy odpowiednie charakterystyki zmiennej losowej,
- systematyzację procesu taryfikacji w oparciu o:

- o rozwiązania stosowane w zakładach ubezpieczeń,
- o **własne propozycje odnośnie zdefiniowanych problemów 1-5,**
- o literaturę światową z wielu dziedzin nauki, w których obserwowane są podobne zjawiska (np. modelowanie zależności).

Część rezultatów pracy nad monografią zawarliśmy w pakiecie programu R zaakceptowanym i umieszczonym w repozytorium CRAN pod nazwą {insuranceData} (por. <https://cran.r-project.org/package=insuranceData>). Pakiet zawiera zbiór portfeli ubezpieczeniowych analizowanych w monografii oraz portfele udostępnione przez autorów w innych pracach z obszaru taryfikacji. Dzięki temu rezultaty znajdują bezpośrednią aplikację w praktyce ubezpieczeniowej, jako rozwinięcie modeli stosowanych obecnie.

## 5. Omówienie pozostałych osiągnięć naukowo – badawczych

W trakcie studiów matematycznych zaczęłam interesować się nowatorskimi w latach dziewięćdziesiątych różnorodnymi zagadnieniami z zakresu metod aktuarialnych. Będąc słuchaczem Letniej Szkoły Nauk Aktuarialnych przy WNE Uniwersytetu Warszawskiego poznałam podstawy ubezpieczeń życiowych, ubezpieczeń niezyciowych oraz matematyki finansowej. Szczególne moje zainteresowanie wzbudziło zagadnienie z dziedziny finansów ubezpieczycieli, jakim są **rezerwy techniczno-ubezpieczeniowe**. W efekcie zgłębiania mojej wiedzy moją uwagę skierowałam na problem wyznaczania tzw. rezerwy szkodowej. W literaturze przedmiotu zaproponowano wiele metod statystycznych pozwalających na szacowanie jej poziomu, jednak z praktycznego punktu widzenia nieznana jest „optymalna” metoda, którą mógłby stosować ubezpieczyciel. Zatem po zatrudnieniu się na Akademii Ekonomicznej w Katowicach rozpoczęłam pracę nad doktoratem o tej tematyce uwzględniając **istotny wpływ poziomu tej rezerwy na gospodarkę finansową zakładu ubezpieczeń**. W trakcie trwania moich badań opracowałam pewien model wielokryterialny pozwalający na wybór wartości rezerwy szkodowej z przedziału wyznaczonego na podstawie estymatorów rezerwy otrzymywanych różnymi metodami statystycznymi. Ostateczną decyzję o wartości rezerwy szkodowej decydent podejmował w oparciu o system wskaźników finansowych zakładów ubezpieczeń<sup>1</sup>. Wyniki moich badań znalazły uznanie i wzięłam udział w grantie pł.

<sup>1</sup> W okresie moich prac nad modelem wyznaczania rezerwy szkodowej polska wykładnia prawa pozwalała zakładom ubezpieczeń na stosowanie jedynie jednej określonej metody kalkulacji rezerwy szkodowej (np. deterministycznie). Obecnie jednak, w wyniku uregulowań *Solvency II*, stosuje się zasadę *Best Estimate*, której idea bazuje na wyborze najlepszego estymatora rezerwy.

„Metody wielokryterialne na polskim rynku finansowym” finansowanym przez Ministerstwo Nauki i Szkolnictwa Wyższego w latach 2002-2005.

Po doktoracie moją uwagę zwrócił kolejny problem z dziedziny finansów ubezpieczycieli, jakim jest wyznaczanie składki dla pojedynczego ryzyka. Badania zawęziłam do problemu szacowania (w sensie statystycznym) składki czystej dla pojedynczego ryzyka w masowym portfelu ubezpieczeń niezyciowych oraz związanego z tym bezpośrednio procesu taryfikacji. Dlatego też moje prace podążyły z jednej strony w kierunku analizy procesu taryfikacji, a z drugiej strony w kierunku zaawansowanego modelowania statystycznego ogólnie (nie jedynie w zakresie taryfikacji). W wyniku analiz zauważyłam, iż **wiele innych modeli**, niż tylko stosowanych w praktyce modeli GLM, może być użytecznych w tym zagadnieniu oraz, że **część tych modeli znajduje zastosowanie w szacowaniu rezerwy szkodowej**. Moje propozycje związane z modelowaniem zależności i wykorzystaniem modeli mieszanych klasy HGLM w taryfikacji zyskały pozytywnie przyjęte przez ekspertów i uzyskałam na rozwój tej tematyki **indywidualny projekt badawczy** pt. „Proces kalkulacji stóp taryf z zastosowaniem hierarchicznych uogólnionych modeli liniowych w zakładach ubezpieczeń majątkowych” finansowany przez Narodowe Centrum Nauki w latach 2011-2014. Rezultaty moich badań w tym temacie zawarłam w monografii wskazanej jako osiągnięcie naukowe. Ponadto w trakcie moich, wciąż trwających, prac badawczych po doktoracie powstał szereg artykułów, które można sklasyfikować według kryterium poruszanej tematyki ubezpieczeniowej oraz modelowania statystycznego. Poniższe zestawienie zawiera najważniejsze pozycje.

#### 1. Zagadnienia związane z taryfikacją oraz szacowaniem składki czystej:

1.1. modelowanie zależności pomiędzy liczbą szkód a średnią wartością szkody dla ryzyka z wykorzystaniem dwuwymiarowej kopuli, modelowanie zależności pomiędzy LOB-ami (*line-of-business*) z wykorzystaniem kopuli, symulacyjne wyznaczanie składki czystej, symulacyjne wyznaczanie składki kwantylowej

- Wolny-Dominiak A. (2016) The copula-based total claim amount regression model with an unobserved risk factor, *Przeгляд Статистичны* LXIII (3), 309-327.
- Wolny-Dominiak A. Wąsat S. (2016) Taryfikacja a priori z wykorzystaniem kopuli, *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu* 415, 258-265.
- Wolny-Dominiak A. Wąsat S. (2015) Total claims amount model using longitudinal data, *Proceedings of the 10th Actuarial Science in Theory and In Practice conference*, University of Economics in Bratislava, 191-200.

1.2. regresja kwantylowa w taryfikacji



- Wolny-Dominiak A. Trzpiot G. (2013) GLM and quantile regression models in a priori ratemaking, *Insurance Review* 4, 50-57.
- Wolny-Dominiak A. and Studnik J. (2013). Estimation of claim counts quantiles, *Proceedings of the 7th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena*, Foundation of the Cracow University of Economics, 206-214.

### 1.3. efekty ZI, ZA w modelach nieliniowych z efektami stałymi

- Wolny-Dominiak A. (2013) Zero-inflated claim count modeling and testing – a case study, *Ekonometria* 1 (39), 144-151.
- Wolny-Dominiak, A. (2013) A Score Test for Zero-adjusted Effect in Claim Severity Modeling, *Proceedings of the 28<sup>th</sup> International Workshop on Statistical Modelling*, Springer volume 2, Università di Palermo, 819-822.

### 1.4. estymacja modelu mieszanego Poissona metodą EM (Expectation-Maximization)

- Trzęsiok M. Wolny-Dominiak A. (2015) Złożony mieszany rozkład Poissona – zastosowania ubezpieczeniowe, *Studia Ekonomiczne* 227/15, 85-95.
- Wolny-Dominiak A. Trzęsiok M. (2016). MixedPoisson: Mixed Poisson Models. R package version 2.0. <https://CRAN.R-project.org/package=MixedPoisson>.

### 1.5. modelowanie w taryfikacji

- Wolny-Dominiak A. (2014) Jednomodelowa taryfikacja a priori w krótkoterminowych ubezpieczeniach majątkowych, *Ekonometria* 46, 34-42.
- Wolny-Dominiak A. (2013) Modeling claim severity via h-likelihood, *Proceedings of 31<sup>th</sup> International Conference Mathematical Methods in Economics*, College of Polytechnics Jihlava, Czech Republic, 1028-1035.
- Wolny-Dominiak A. (2011) Szacowanie stóp taryf w ubezpieczeniach majątkowych z wykorzystaniem modelu HGLM, *Insurance Study, Research Paper of Poznan University of Economics* 182, 318-328.
- Wolny-Dominiak A. (2011) Analiza porównawcza modeli mieszanych szacowania stóp taryf w ubezpieczeniach majątkowych z wykorzystaniem krosvalidacji, *Actuarial Issues – Theory and Practice, Research Papers of Wrocław University of Economics* 207, 229-238.
- Wolny-Dominiak A. (2012) Ranking and classification of automobile insurance policies according to the number of claims, *Mathematical Economics* 8(15), 147-158.

## 2. Zagadnienia związane z szacowaniem rezerwy szkodowej

- Wolny-Dominiak A. (2016) Bootstrap Mean Squared Error of Prediction in Loss Reserving, *Contemporary Trends and Challenges in Finance – Proceedings from the 2nd International Conference in Finance*, Wrocław, Series: Springer Proceedings in Business and Economics, /in print/.



- Wolny-Dominiak A. (2014) Loss reserving using growth curve modeling, *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu* 342, „Ubezpieczenia wobec wyzwań XXI wieku”, red. Wanda Ronka-Chmielowiec, 331-338.
- Wolny-Dominiak A. (2014) Zmodyfikowana regresja logarytmiczno-normalna w szacowaniu rezerwy szkodowej, *Studia Ekonomiczne* 181, 220-234.

### 3. Zagadnienia związane z rozwiązaniami informatycznymi w programie R

- 3.1. pakiet {MixedPoisson} (por. <https://cran.r-project.org/package=MixedPoisson>) zawiera funkcje służące do estymacji parametrów mieszanego modelu Poissona za pomocą metody EM wykorzystując całkowanie numeryczne kwadraturą Gaussa-Laguerre
- 3.2. zastosowania programu komputerowego R w analizach przestrzennych.

**Podsumowując**, poza monografią wskazaną w osiągnięciu naukowym, do swojego dorobku po uzyskaniu stopnia doktora zaliczam w szczególności:

- samodzielne autorstwo w ponad 21 artykułach w czasopismach oraz 6 rozdziałów w monografiach i materiałach konferencyjnych.
- współautorstwo 10 artykułów w czasopismach oraz 7 rozdziałów w materiałach konferencyjnych, w których byłam wiodącym autorem pod względem koncepcyjnym jak również edycyjnym,
- współautorstwo 2 pakietów statystycznych programu komputerowego R, w których byłam wiodącym autorem pod względem koncepcyjnym, merytorycznym, technicznym i edycyjnym.

Ponadto realizowałam po doktoracie:

- 10 podtematów w ramach badań statutowych
- 4 tematy w ramach badań własnych
- dwa granty naukowe: jako kierownik grantu (jeden), wykonawca (jeden)

Brałam udział jako referent w 20 konferencjach międzynarodowych (referaty głównie w języku angielskim) oraz 20 ogólnokrajowych (referaty w języku polskim).