Uniwersytet Ekonomiczny w Krakowie
Dziedzina nauki: Nauki społeczne
Dyscyplina naukowa: Ekonomia i finanse

mgr Krystian Szczęsny

Agregacja ryzyka i ocena efektu dywersyfikacji w procesie wyznaczania kapitałowych wymogów wypłacalności w Solvency II

Rozprawa doktorska

Promotor: dr hab. Stanisław Wanat, prof. UEK
Promotor pomocniczy: dr Anna Denkowska

Kraków, 2025

Podziękowania

Składam serdeczne podziękowania Panu dr hab. Stanisławowi Wanatowi, prof. UEK, za niezwykle cenne wskazówki, merytoryczne wsparcie oraz zaangażowanie w powstawanie niniejszej pracy. Pana wiedza, doświadczenie i cierpliwość były dla mnie nieocenioną pomocą na każdym etapie badań. Chciałbym również podziękować Pani dr Annie Denkowskiej za cenne uwagi oraz inspirujące rozmowy, które przyczyniły się do powstania niniejszej pracy. Pragnę Państwu podziękować także za umożliwienie mi udziału w konferencjach naukowych oraz za wprowadzenie mnie w świat badań i pracy naukowej.

Nie ma słów, którymi mógłbym w pełni wyrazić wdzięczność mojej żonie Katarzynie oraz moim rodzicom Joannie i Mieczysławowi za ich nieustanne wsparcie. Z najgłębszą wdzięcznością Wam właśnie dedykuję tę pracę.

Serdecznie dziękuję wszystkim, którzy zawsze byli dla mnie oparciem - mojej Siostrze Aleksandrze, Dziadkom, całej Rodzinie Magnesów i Rokitów, a także wszystkim osobom, które dodawały mi motywacji podczas pisania tej pracy.

Spis treści

Wykaz oznaczeń

AM - Bezwzględna Miara Ryzyka Modelowego

ARA - Adaptive Rearrangement Algorithm

ART1 - Teoria rezonansu adaptacyjnego 1

ART2 - Teoria rezonansu adaptacyjnego 2

BSCR - Basic Solvency Capital Requirement

C-vine - C-kaskada kopul

CLB - Dolna granica ryzyka z uwzględnieniem wiarygodności założeń

CRM - Credibility-based Relative Measure

CUB - Górna granica ryzyka z uwzględnieniem wiarygodności założeń

DNN - Głębokie sieci neuronowe

DQL - Głębokie uczenie Q

D-vine - D-kaskada kopul

ED - Efekt dywersyfikacji

EIOPA - European Insurance and Occupational Pensions Authority

FCM - Metoda rozmytych c-średnich

FS - Formuła Standardowa

GTM - Generatywne odwzorowanie topograficzne

HC - Grupowanie hierarchiczne

KHM - Metoda k-harmonicznych średnich

KKM - Jądrowa metoda k-średnich

K-Means - Metoda k-średnich

K-Means++ - Ulepszona metoda k-średnich

K-Medoids - Metoda k-medoidów

k-NN - Metoda k-najbliższych sąsiadów

LDA - Liniowa analiza dyskryminacyjna

LES$_p(X_i)$ - Dolny oczekiwany shortfall (Lower Expected Shortfall) na poziomie ufności p

Logit - Regresja logistyczna

LRL - Liniowa regresja

MCM - Metoda Monte Carlo

MM - Modele mieszanin

MoRC - Model Risk Capital

NBC - Naiwny klasyfikator Bayesa

Perceptron - Perceptron

QLDA - Kwadratowa analiza dyskryminacyjna

Q-Learning - Uczenie Q

RM - Względna Miara Ryzyka Modelowego

SARSA - Algorytm SARSA

SCR - Kapitałowy Wymóg Wypłacalności

SFCR - Solvency and Financial Condition Report

SKM - Sekwencyjna metoda k-średnich

SOM - Samoorganizująca się mapa (mapa Kohonena)

SVM - Maszyna wektorów nośnych

TDM - Metoda różnic czasowych

$ES_p(X_i)$ - Oczekiwana strata (Expected Shortfall) na poziomie ufności p

$VaR_p(X_i)$ - Wartość zagrożona (Value at Risk) na poziomie ufności p

ϱ - Miara ryzyka

$\|\cdot\|_2$ - Norma euklidesowa

Wstęp

W 2016 r. weszła w życie Dyrektywa Solvency II, będąca najważniejszą zmianą regulacyjną na unijnym rynku ubezpieczeniowym od ponad 30 lat. Wprowadza nowe, jednolite dla wszystkich krajów UE wymogi w zakresie zarządzania ryzykiem i określania wypłacalności. Celem jest zwiększenie bezpieczeństwa funkcjonowania zakładów ubezpieczeń oraz ochrona ubezpieczonych przed ich niewypłacalnością. Realizację tych założeń ma gwarantować spełnienie przez firmy ubezpieczeniowe dwóch wymogów kapitałowych: minimalnego wymogu wypłacalności (ang. Minimum Capital Requirement - MCR) oraz wymogu kapitałowego wypłacalności (ang. Solvency Capital Requirement - SCR). SCR definiowany jest jako kapitał ekonomiczny (ang. economic capital), który powinien zapewnić bezpieczeństwo ubezpieczonym w przypadku wystąpienia nieprzewidzianych strat. Powinien z prawdopodobieństwem 0.995 gwarantować, że ubezpieczyciel w ciągu 12 miesięcy będzie w stanie wypełniać swoje zobowiązania. Jest obliczany przynajmniej raz w roku oraz gdy nastąpi istotna zmiana profilu ryzyka ubezpieczyciela. Dyrektywa podaje metody wyznaczania MCR oraz SCR w postaci tzw. formuły standardowej (FS). Zgodnie z tą formułą, SCR zakładu ubezpieczeń ustala się w drodze agregacji wymogów kapitałowych dla poszczególnych modułów i podmodułów ryzyka. Ze względu na fakt, że nie wszystkie czynniki ryzyka realizują się w tym samym czasie, wyznaczony w wyniku agregacji SCR jest na ogół nie większy niż suma wymogów kapitałowych dla poszczególnych modułów i podmodułów. W takiej sytuacji dywersyfikacja ryzyka oraz prawidłowo oszacowany efekt dywersyfikacji (ED) stanowią kluczowy element procesu zarządzania ryzykiem w zakładach ubezpieczeń. Mechanizm ten umożliwia agregację jednostek narażonych na identyczne lub zbliżone kategorie ryzyka w ramach portfeli, co prowadzi do ograniczenia potencjalnych negatywnych konsekwencji finansowych na poziomie poszczególnych podmiotów. Ponadto, efekt dywersyfikacji stwarza zakładom ubezpieczeń możliwość efektywnego zarządzania zróżnicowanymi kategoriami ryzyka oraz ich natężeniem, na które narażone są zarówno portfele ubezpieczeniowe, jak i poszczególne linie biznesowe oraz portfele inwestycyjne.

Z metodologicznego punktu widzenia oszacowanie na adekwatnym poziomie ED wymaga uwzględnienia w procesie agregacji właściwej struktury zależności między agregowanymi ryzykami. W ramach formuły standardowej wymogi kapitałowe agreguje się z wykorzystaniem metody wariancji–kowariancji, w której macierz korelacji została podana i ustalona w dyrektywie. Zatem każdy z zakładów ubezpieczeń w Unii Europejskiej (UE), stosując FS, musi korzystać z dokładnie takiej samej macierzy korelacji. Stosując tę metodę, otrzymuje się oszacowanie ED na adekwatnym poziomie, pod warunkiem że agregowane ryzyka ubezpieczyciela podlegają wielowymiarowemu rozkładowi normalnemu.

W praktyce jednak ryzyka często mają rozkłady różne od wielowymiarowego rozkładu normalnego, co sprawia, że współczynnik korelacji liniowej jest niewystarczający w określeniu zależności, ponieważ nie wychwytuje relacji innych niż liniowe. Twórcy dyrektywy podali także wartości korelacji równe 0. Należy podkreślić, że wartość korelacji równa zero nie oznacza braku zależności. Oznacza jedynie brak związku liniowego, co w praktyce może prowadzić do sprzecznych wniosków. Co najistotniejsze, miara korelacji liniowej jest również bardzo wrażliwa na obserwacje odstające, przez co pojedyncze ekstremalne zdarzenia mogą w istotny sposób zniekształcić wyniki tej miary. W konsekwencji efekt dywersyfikacji jest szacowany przy użyciu metod, które mogą niewłaściwie przypisywać relacje między czynnikami ryzyka. Prowadzi to do niedoszacowania lub przeszacowania SCR i w istotny sposób wpływa na funkcjonowanie zakładu ubezpieczeń oraz na jego wypłacalność. Z niedoskonałości FS zdają sobie

sprawę sami autorzy Dyrektywy, zalecając poszukiwanie i stosowanie modeli wewnętrznych, które w lepszym stopniu odzwierciedlają profil ryzyka ubezpieczyciela niż wykorzystanie FS. Każdy model wewnętrzny musi zostać zaakceptowany przez organ nadzoru - w Polsce jest to Komisja Nadzoru Finansowego. Wątpliwości ma także Europejski Urząd Nadzoru Ubezpieczeń i Pracowniczych Programów Emerytalnych (ang. European Insurance and Occupational Pensions Authority - EIOPA), który w październiku 2020 r. rozpoczął ogólnoeuropejskie badanie porównawcze dotyczące dywersyfikacji w modelach wewnętrznych. Cel badania EIOPA jest trojaki:

1. Uzyskanie przeglądu obecnych podejść na rynku w zakresie określania ED oraz przeanalizowanie i porównanie ich poziomu.

2. Lepsze zrozumienie związku między sposobami modelowania zależności a agregacją ryzyka oraz wynikającymi z nich korzyściami z dywersyfikacji.

3. Poprawa jakości i konwergencji nadzoru nad dywersyfikacją w modelach wewnętrznych.

W dalszych rozważaniach centralne miejsce zajmuje model agregacji, rozumiany jako sposób przejścia od indywidualnych zmiennych ryzyka do zmiennej zagregowanego ryzyka, na podstawie której wyznacza się kapitałowy wymóg wypłacalności. Formalnie, niech $\boldsymbol{X} = (X_1, \ldots, X_d)$ oznacza wektor indywidualnych (agregowanych) ryzyk o rozkładach brzegowych $F_1, \ldots, F_d$.

Model agregacji obejmuje trzy elementy: (i) specyfikację rozkładów brzegowych F_i, (ii) opis struktury zależności między składowymi $\boldsymbol{X}$ (np. poprzez macierz korelacji, kopulę lub inną reprezentację współzależności, w tym zależności ogonowe), oraz (iii) regułę agregacji $\psi(\cdot)$ prowadzącą do zmiennej zagregowanej $Y = \psi(\boldsymbol{X})$ (w praktyce najczęściej $Y = \sum_{i=1}^{d} X_i$). Na podstawie Y wyznacza się kapitałowy wymóg wypłacalności SCR.

W formule standardowej Solvency II rolę modelu agregacji pełni metoda wariancji-kowariancji z ustaloną macierzą korelacji, która zastępuje szczegółowy opis współzależności między dwiema agregowanymi ryzykami jedną liczbą (współczynnikiem korelacji liniowej). Założenie to jest adekwatne przy wielowymiarowym rozkładzie normalnym i słabych zależnościach ogonowych, lecz w praktyce może odbiegać od profilu ryzyka zakładu. W efekcie zarówno SCR, jak i szacowany efekt dywersyfikacji stają się funkcją niepewności co do struktury zależności, a więc są na ryzyko modelu. Ocena tej niepewności oraz jej wpływu na stabilność (odporność) wyników stanowi bezpośrednie uzasadnienie dla rozważania alternatywnych modeli agregacji, w których zależności mogą być odwzorowane dokładniej (np. poprzez kopule, kaskady kopul lub komponenty oparte na uczeniu maszynowym).

Poniżej sprecyzowano zamierzenia badawcze oraz hipotezy dotyczące: (i) ilościowej oceny ryzyka modelu agregacji wynikającego z niepewnej struktury zależności, (ii) wykorzystania tej oceny do badania odporności oszacowań SCR i efektu dywersyfikacji na scenariusze skrajnie niekorzystnych współzależności oraz (iii) porównania formuły standardowej z metodami agregacji opartymi na kopulach i uczeniu maszynowym, które mogą być zastosowane w modelach wewnętrznych.

Za cel główny przyjęto ilościową ocenę ryzyka modelu standardowej formuły agregacji wymogów kapitałowych w Solvency II, wynikającego z niepewności struktury zależności między rodzajami ryzyka, oraz opracowanie i porównanie alternatywnych metod agregacji opartych na kopulach i uczeniu maszynowym, możliwych do zastosowania w modelach wewnętrznych.

Uzyskane wyniki zostaną wykorzystane do oceny odporności SCR oraz efektu dywersyfikacji na skrajne scenariusze współzależności.

Cel ten doprecyzowano ośmioma celami szczegółowymi:

C1: Scharakteryzowanie formuły standardowej stosowanej w Solvency II do wyznaczania kapitałowych wymogów wypłacalności, w tym założeń metody wariancji–kowariancji oraz roli macierzy korelacji.

C2: Omówienie zastosowań uczenia maszynowego w działalności ubezpieczycieli ze szczególnym uwzględnieniem wypłacalności i wsparcia procesów agregacji ryzyka.

C3: Zaprezentowanie sposobów agregacji ryzyka przy braku, częściowej oraz pełnej informacji o strukturze zależności między rodzajami ryzyka.

C4: Scharakteryzowanie ryzyka modelu w kontekście Solvency II oraz opracowanie i zastosowanie podejścia ilościowego do oceny ryzyka modelu agregacji wynikającego z niepewnej struktury zależności między agregowanymi ryzykami.

C5: Określenie wpływu niepewności współczynnika korelacji oraz zależności w ogonach rozkładu na SCR i korzyści z dywersyfikacji.

C6: Wskazanie i rozwinięcie metod modelowania zależności łączących kopule i techniki uczenia maszynowego, które zastosowane w procesie agregacji pozwolą wiarygodnie uchwycić rzeczywistą strukturę współzależności między ryzykami, a tym samym umożliwią określenie kapitałowego wymogu wypłacalności oraz efektu dywersyfikacji na adekwatnym poziomie.

C7: Opracowanie narzędzi umożliwiających wyznaczenie dolnego i górnego ograniczenia wartości zagrożonej (VaR) dla zagregowanego ryzka dostosowanych do różnego zakresu informacji o zależnościach.

C8: Zbadanie ryzyka modelu formuły standardowej (FS) z wykorzystaniem dolnego i górnego ograniczenia VaR oraz wskazanie sytuacji potencjalnego niedoszacowania lub przeszacowania SCR.

W odniesieniu do przedstawionych celów badawczych sformułowano następującą hipotezę główną:

HG: Niepewność struktury zależności między rodzajami ryzyka generuje istotne ryzyko modelu w formule standardowej Solvency II, przez co wyniki SCR i efekt dywersyfikacji uzyskane z FS nie są odporne na skrajne scenariusze współzależności, natomiast metody agregacji oparte na kopulach i uczeniu maszynowym, lepiej odwzorowujące rzeczywiste zależności, dostarczają stabilniejszych i bardziej adekwatnych oszacowań.

Opis zamierzeń badawczych domyka zestaw hipotez szczegółowych, które rozwijają i precyzują hipotezę główną:

H1: Jednowymiarowe rozkłady agregowanych zmiennych ryzyka estymowane przez sztuczne sieci neuronowe lepiej odzwierciedlają dane poza próbą niż rozkłady parametryczne.

H2: Modele wykorzystujące uczenie maszynowe lepiej identyfikują nieliniowe struktury zależności między ryzykami niż modele parametryczne z góry narzuconą strukturą zależności, dzięki czemu dostarczają dokładniejszych prognoz kwantyli straty i bardziej adekwatnych poziomów SCR.

H3: Ilościowa ocena ryzyka modelu agregacji umożliwia wskazanie, kiedy formuła standardowa zaniża lub zawyża kapitałowe wymogi wypłacalności, przez co dostarcza podstaw do ustalania buforów kapitałowych i parametrów reasekuracji, redukując ryzyko błędnej oceny SCR przy niepewnej strukturze zależności.

H4: Poszerzanie wiarygodnej informacji o strukturze zależności między ryzykami, od braku wiedzy po pełniejszą specyfikację, systematycznie zawęża pasmo niepewności oszacowań SCR i stabilizuje ocenę efektu dywersyfikacji.

Ryzyko modelu, rozumiane jako możliwość uzyskania zniekształconych wyników wskutek błędnych założeń, uproszczeń lub niepewności związanej z estymacją, ma szczególne znaczenie w modelach agregacji ryzyka. Modele te przekształcają informacje o rozkładach brzegowych poszczególnych rodzajów ryzyka oraz o ich współzależnościach w miary ryzyka portfelowego (np. VaR, SCR). Niepewność co do kształtu rozkładów, parametrów i struktury zależności (w tym w ogonach) przekłada się bezpośrednio na niepewność wielkości kapitałowych oraz na ocenę efektu dywersyfikacji. W tym ujęciu odporność stanowi lustrzane odbicie ryzyka modelu: im większa wrażliwość wyników na skrajne konfiguracje zależności, tym wyższe ryzyko modelu i tym słabsza odporność otrzymanych oszacowań. Innymi słowy, jeżeli miary i procedury ilościowej oceny ryzyka modelu potwierdzają stabilność wyników wobec skrajnie niekorzystnych scenariuszy współzależności, model spełnia kryterium odporności.

Praca ma charakter teoretyczno-empiryczny. Teoretyczna część pracy porządkuje zagadnienia wokół kilku kluczowych problemów. Po pierwsze, przedstawia ramy regulacyjne Solvency II oraz rolę agregacji w wyznaczaniu SCR, wskazując ograniczenia formuły standardowej opartej na metodzie wariancji–kowariancji oraz stałej macierzy korelacji (m.in. założenie wielowymiarowej normalności oraz nieuwzględnianie zależności ogonowych). Centralnym wątkiem jest tu odporność wyników na skrajne, zwłaszcza niekorzystne scenariusze współzależności.

Po drugie, zarysowuje aparat matematyczny: miary ryzyka, pojęcie efektu dywersyfikacji oraz definicję „modelu agregacji" jako kombinacji rozkładów brzegowych, struktury zależności i reguły łączenia ryzyk. Omawia metody opisu zależności z wykorzystaniem kopul w podejściu parametrycznym i nieparametrycznym. Omawia, jak różny zakres informacji o zależnościach (pełna, częściowa, brak) przekłada się na graniczne oszacowania zagregowanego ryzyka.

Po trzecie, przedstawia rozwój uczenia maszynowego jako uzupełnienie lub alternatywę dla klasycznych podejść, podkreślając ich zdolność do uchwycenia złożonych, nieliniowych struktur zależności. Akcentuje synergię z teorią kopul. Uczenie maszynowe wspiera estymację i aproksymację złożonych struktur zależności (w tym identyfikację postaci kopuli), zaś kopule służą do redukcji wymiaru oraz do generowania syntetycznych danych uczących w sytuacjach z lukami lub ograniczeniami poufności. W kontekście agregacji narzędzia te mogą podnosić adekwatność i odporność oszacowań względem formuły standardowej.

Po czwarte, porządkuje pojęcie ryzyka modelu, wskazując jego podstawowe kategorie (koncepcyjne, implementacyjne, związane z jakością danych oraz nadmiernym dopasowaniem) oraz źródła (ludzkie, technologiczne, środowiskowe), a następnie przechodzi do jego oceny w ujęciu jakościowym i ilościowym w realiach Solvency II.

W efekcie ta część dostarcza spójnych podstaw teoretycznych do badania najważniejszego problemu pracy: na ile wyniki modeli agregacji - zwłaszcza formuły standardowej - pozostają odporne na ekstremalne scenariusze zależności oraz w jaki sposób alternatywne opisy współzależności (kopule, kaskady kopul, metody uczenia maszynowego) mogą tę odporność wzmacniać.

W części empirycznej prowadzone badania obejmują dwa zasadnicze obszary. Pierwszy dotyczy modelowania zależności pomiędzy ryzykami z wykorzystaniem metod uczenia maszynowego, które mogą znaleźć zastosowanie w budowie modeli wewnętrznych. Drugi koncentruje się na badaniu wrażliwości miary VaR na przyjętą strukturę zależności oraz na ilościowej ocenie ryzyka modelu w formule standardowej Solvency II. Ocena ta służy bezpośrednio do badania odporności wyników SCR oraz efektu dywersyfikacji na skrajnie niekorzystne scenariusze współzależności.

W pierwszym etapie badania wykorzystano dane modułu ubezpieczeń majątkowych, pochodzące ze sprawozdań o wypłacalności i kondycji finansowej (ang. Solvency and Financial Condition Report – SFCR) ubezpieczycieli z Niemiec. Kapitałowy wymóg wypłacalności wyznaczano w oparciu o wielowymiarowy rozkład skonstruowany zgodnie z twierdzeniem Sklara, poprzez dekompozycję na rozkłady brzegowe poszczególnych ryzyk oraz kopulę opisującą strukturę zależności między nimi. Zastosowano dwa podejścia: parametryczne i nieparametryczne. W podejściu parametrycznym dopasowywano rozkłady brzegowe, a strukturę zależności modelowano kaskadą kopul. W podejściu nieparametrycznym konstrukcja rozkładu wielowymiarowego opierała się na architekturze sieci neuronowych: dla każdego ryzyka budowano sieć, estymującą rozkład brzegowy, z funkcją straty opartą na log-wiarygodności oraz ograniczeniami, zapewniającymi zgodność z własnościami dystrybuanty jednowymiarowej. Odrębna sieć odwzorowywała strukturę zależności przy nałożonych warunkach wynikających z teorii kopuli. Po oszacowaniu rozkładów brzegowych i zależności konstruowano rozkład wielowymiarowy, a następnie wyznaczano SCR oraz efekt dywersyfikacji. Uzyskane wyniki porównano z wartościami otrzymanymi w formule standardowej, co pozwoliło ocenić różnice wynikające z wyboru podejścia do modelowania zależności. Walidację przeprowadzono dwuetapowo: (i) porównano dopasowanie rozkładów brzegowych, (ii) oceniono dopasowanie rozkładów wielowymiarowych z użyciem miar jakości prognoz probabilistycznych, w tym odległości energetycznej. W tym ujęciu lepsza identyfikacja rozkładów brzegowych i współzależności sprzyja adekwatniejszym oszacowaniom SCR.

W drugim etapie opracowano cztery algorytmy do badania wpływu struktury zależności na VaR i w konsekwencji do ilościowej oceny ryzyka modelu FS. Pierwszy algorytm zakłada wielowymiarową normalność oraz znajomość macierzy korelacji i bada wrażliwość VaR na zmiany współczynnika korelacji ρ. Drugi i trzeci algorytm skupiają się na niejednoznaczności opisu zależności przez korelację liniową: przy danych rozkładach brzegowych (a) wyznaczano parametry kopuli Archimedesa odpowiadające zadanemu współczynnikowi korelacji ρ metodą siecznych, (b) poszukiwano dla ustalonego ρ struktury zależności maksymalizującej VaR z wykorzystaniem wielokryterialnej optymalizacji Pareto. Czwarty algorytm służy ocenie wpływu niepewności w ogonach rozkładu: przy znanych rozkładach brzegowych zależność opisywano zniekształconą mieszanką kopul, z rozróżnieniem części centralnej i ogonów. Wykorzystując symulowane wyżarzanie, poszukiwano takiej kombinacji kopul w ogonach, która prowadzi do najwyższej wartości VaR. Uzyskane wyniki posłużyły do oceny ryzyka modelu agregacji stosowanego w formule standardowej Solvency II przy wyznaczaniu SCR, opartego na metodzie wariancji–kowariancji.

Struktura rozprawy odzwierciedla powyższe etapy badań. W rozdziale pierwszym przedstawiono zarys dyrektywy Solvency II, jej główne założenia oraz znaczenie dla systemu zarządzania ryzykiem w sektorze ubezpieczeniowym. Zaprezentowano formułę standardową wraz ze sposobem agregacji ryzyka w modułach i podmodułach oraz omówiono wątpliwości EIOPA wynikające z badań dotyczących stosowania FS w Unii Europejskiej, a także rolę modeli wewnętrznych i wymogi formalne związane z ich wdrażaniem.

Rozdział drugi wprowadza matematyczne podstawy miar ryzyka oraz pojęcie agregacji ryzyka. Na podstawie literatury dokonano podziału metod agregacji na techniki łączące kapitały ekonomiczne lub miary ryzyka bez opisu rzeczywistej struktury zależności oraz podejścia oparte na rozkładach wielowymiarowych i jawnej strukturze współzależności, przy czym uwaga koncentruje się na tej drugiej klasie. Zdefiniowano efekt dywersyfikacji i omówiono jego znaczenie w ubezpieczeniach, a następnie przedstawiono metody agregacji, bazujące na kopulach w ujęciu parametrycznym i nieparametrycznym. Zaprezentowano także przegląd podejść badających wpływ informacji o zależnościach na zagregowane ryzyko i efekt dywersyfikacji, porządkując je według zakresu wiedzy: pełna znajomość rozkładu wielowymiarowego, znane rozkłady brzegowe z częściową informacją o zależnościach oraz wyłącznie rozkłady brzegowe.

Rozdział trzeci zawiera syntetyczną charakterystykę wybranych metod uczenia maszynowego oraz przykłady ich zastosowań w ubezpieczeniach, m.in. w ocenie ryzyka, wycenie polis i detekcji nadużyć. Na podstawie przeglądu literatury pokazano sposoby łączenia uczenia maszynowego z kopulami, od estymacji i identyfikacji struktur zależności po wykorzystanie kopul do redukcji wymiaru i generowania syntetycznych danych uczących.

Rozdział czwarty porządkuje zagadnienie ryzyka modelu. Definiuje pojęcia modelu i ryzyka modelu, ilustruje ich znaczenie przykładami historycznymi z sektora finansowego Zaproponowano w nim autorską klasyfikację ryzyka modelu według rodzajów i źródeł. Całość osadzono w realiach Solvency II, ze szczególnym uwzględnieniem ryzyk wynikających ze stosowania FS oraz potencjalnych zagrożeń związanych z modelami wewnętrznymi. Przedstawiono ramy oceny ryzyka modelu: elementy zarządzania modelem, podejście jakościowe oraz komponent ilościowy oparty na wskaźnikach mierzących wpływ założeń modelowych na łączny poziom ryzyka oraz niepewności wyników.

W rozdziale piątym przedstawiono autorskie nieparametryczne podejście do wyznaczania SCR oraz efektu dywersyfikacji, oparte na kopulach i uczeniu maszynowym. Bazuje ono na architekturze dwóch głębokich sieci neuronowych, pierwszej estymującej rozkłady brzegowe poszczególnych ryzyk, drugiej estymującej strukturę zależności. Zaproponowaną metodę porównano z podejściem parametrycznym, bazującym na kaskadach kopul. W tym celu, wykorzystujac oba podejścia, agregowano ryzyko czterech segmentów ubezpieczyciela; dokładniej, agregowanymi ryzykami były wskaźniki zespolone wyznaczone dla tych segmentów na podstawie danych pochodzących ze sprawozdań SFCR dla rynku niemieckiego. Dopasowanie modelu w podejściu parametrycznym oraz uczenie w podejściu nieparametrycznym przeprowadzono na zbiorze treningowym, a weryfikację na zbiorze testowym z użyciem log-wiarygodności oraz odległości energetycznej. Uzyskane wyniki wskazują na przewagę podejścia bazującego na głębokich sieciach neuronowych.

W rozdziale szóstym oceniono wpływ struktury zależności między agregowanymi ryzykami na SCR, wykorzystując cztery autorskie algorytmy. Pierwszy bada wrażliwość VaR na zmiany współczynników korelacji przy założeniu wielowymiarowego rozkładu normalnego. Dwa kolejne pokazują niewystarczalność korelacji liniowej w szacowaniu VaR dla zagregowanego ryzyka; ta sama wartość ρ może odpowiadać odmiennym strukturom zależności i prowadzić

do różnych wyników. Ostatni algorytm uwzględnia niepewność w ogonach rozkładów, co jest kluczowe przy wyznaczaniu SCR w reżimie Solvency II. Opracowane procedury posłużyły do oceny ryzyka modelu agregacji stosowanego w formule standardowej Solvency II przy wyznaczaniu SCR, opartego na metodzie wariancji–kowariancji, z użyciem miar opisanych w rozdziale czwartym. Uzyskane wyniki podkreślają znaczenie ryzyka modelu w kontekście modeli wewnętrznych oraz stosowania FS, w której zależności między ryzykami są opisane wyłącznie macierzą korelacji. Pokazują, że ta sama wartość korelacji może prowadzić do różnych struktur zależności i odmiennych wyników VaR, a nieuwzględnienie tej niepewności, zwłaszcza w ogonach rozkładów, może skutkować istotnym błędem w oszacowaniu wymogów kapitałowych.

1. Szacowanie kapitałowych wymogów wypłacalności w Solvency II

1.1 Ogólna charakterystyka Solvency II

Pierwszym formalnym działaniem w kierunku koordynacji przepisów ustawowych, administracyjnych i wykonawczych odnoszących się do podejmowania i prowadzenia działalności w dziedzinie ubezpieczeń bezpośrednich innych niż ubezpieczenia na życie w Europie było wprowadzenie w 1973 r. pierwszej dyrektywy 73/239/EWG a następnie w 1979 r. pierwszej dyrektywy dotyczącej ubezpieczeń na życie - dyrektywy 79/267/EWG. Skutkiem tych działań było ujednolicenie wymogów wypłacalności w krajach członkowskich Unii Europejskiej. Dyrektywa nakładała na zakłady ubezpieczeniowe wymóg posiadania marginesu wypłacalności, skąd pochodzi nazwa "Wypłacalność I", czyli "Solvency I". W kolejnych latach udoskonalano te dyrektywy poprzez dyrektywy 88/357/EWG, 90/619/EWG, 92/49/EWG i 92/96/EWG. W połowie lat 90. XX wieku podjęto na poziomie Europejskim dalsze działania mające na celu poprawę i nadanie nowego znaczenia istniejącym przepisom dotyczącym wypłacalności. Przeprowadzono wówczas badania porównawcze, z których wynikało, że obowiązujące rezerwy dotyczące funduszy własnych nie uwzględniały odpowiednio wszystkich ryzyk, na jakie narażony jest ubezpieczyciel. Zdecydowano zatem o kompleksowej reformie przepisów dotyczących nadzoru ubezpieczeniowego. Ze względu na złożoność tematu w międzyczasie wprowadzano jedynie pilniejsze zmiany poprzez dyrektywę dotyczącą ubezpieczeń na życie (dyrektywa 2002/83/WE) oraz dyrektywę dotycząca wymogów ubezpieczeń innych niż na życie (dyrektywa 2002/13/WE).

Prace nad dyrektywą Solvency II Komisja Europejska rozpoczęła w 2001 r. Ich celem było kontynuowanie prac nad wymogami wypłacalności zapoczątkowanymi w dyrektywie Solvency I. Dyrektywa Solvency I miała na celu przegląd i aktualizację unijnego systemu wypłacalności, ale nie radziła sobie z różnorodnością profili ryzyka firm ubezpieczeniowych; dlatego dyrektywa Solvency II stanowi fundamentalną reformę przepisów dotyczących wypłacalności ubezpieczycieli, która miała umożliwić m.in. ujednolicenie w całej UE technik wyceny aktywów i pasywów. Zasady Solvency I nadal mają zastosowanie do zakładów, do których Solvency II nie ma zastosowania (małe zakłady ubezpieczeń, instytucje pracownicze programów emerytalnych oraz fundusze odpraw pośmiertnych).

W dyrektywie Solvency II (2009/138/WE) przyjęto kilka założeń. Pierwsze, organy nadzoru powinny posiadać narzędzia do oceny wypłacalności zakładów ubezpieczeń. Narzędzia te mają umożliwić zarówno ocenę ilościową, jak i jakościową działalności zakładów. Drugim powodem jest to, że ubezpieczyciele będą zachęcani do odpowiedniego zarządzania ryzykami, na które będą narażeni. Kolejne założenie dotyczy ilościowego określenia wymogów kapitałowych, które powinny być konstruowane dwupoziomowo jako wymogi kapitału docelowego i wymogi kapitału minimalnego, co dałoby organom nadzoru niezbędny czas na uruchomienie systemu naprawczego w zakładzie. W dyrektywie przyjęto również założenie spójności rozwiązań w całym sektorze finansowym, co oznacza, że zapisy dyrektywy Solvency II mają być zgodne z zasadami przyjętymi w sektorze bankowym. Ze względu na globalne skutki podejmowanych przez ubezpieczycieli decyzji, nowy system powinien w szczególności obejmować konglomeraty i ubezpieczeniowe grupy kapitałowe. Przyjęto również, że w całej Unii należy przyjąć wspólne standardy nadzorcze, rachunkowe i sprawozdawcze.

W Rozporządzeniu Delegowanym Unii Europejskiej (ang. European Union - UE) 2015/35 z dnia 10 października 2014 r., uzupełniającym dyrektywę Parlamentu Europejskiego i Rady 2009/138/WE w sprawie podejmowania i prowadzenia działalności ubezpieczeniowej i reasekuracyjnej (Solvency II), wyróżnia się trzy główne filary: wymogi dotyczące wyceny oraz wymogi kapitałowe oparte na ryzyku (Filar I), poprawę zarządzania (Filar II) oraz zwiększenie przejrzystości (Filar III). Konstrukcja pierwszego filaru w szczególności przypomina rozwiązania przyjęte w regulacjach bankowych z 2004 r., zawartych w dokumentach Bazylea II, i opiera się na trzech powiązanych elementach: SCR, MCR oraz zasadach wyceny. Każdemu filarowi przypisano odrębne kategorie ryzyka, które odzwierciedlają specyfikę działalności zakładów ubezpieczeń.

Filar I uwzględnia mierzalne ilościowo rodzaje ryzyka związane z działalnością zakładu ubezpieczeń. Filar II obejmuje wymogi jakościowe dotyczące zarządzania ryzykiem oraz systemu nadzoru zakładu ubezpieczeń, w tym proces własnej oceny ryzyka i wypłacalności (ang. Own Risk and Solvency Assessment - ORSA). Filar III obejmuje narzędzia do ujawniania informacji i zapewnienia przejrzystości, w szczególności Raport o wypłacalności i sytuacji finansowej zakładu (ang. Solvency and Financial Condition Report - SFCR), publikowany dla opinii publicznej, zawierający informacje o kapitale, ryzykach i zarządzaniu zakładem, oraz Regularny raport nadzorczy (ang. Regular Supervisory Report - RSR), przekazywany organowi nadzoru, zawierający szczegółowe informacje finansowe i dotyczące ryzyka, w tym wyniki ORSA.

W Tabeli 1.1 przedstawiamy strukturę dyrektywy Solvency II z podziałem na filary, wraz z ich opisem.

Filar	**Zakres**	**Opis**	**Artykuły Rozporządzenia i Dyrektywy**
I	Wymogi ilościowe dotyczące wyceny i kapitału	SCR, MCR, zasady wyceny aktywów i zobowiązań	Rozporządzenie 2015/35: Rozdz. II; Dyrektywa 2009/138/WE: Art. 75–101
II	Wymogi jakościowe i nadzorcze	System zarządzania ryzykiem, system nadzoru zakładów ubezpieczeń, proces ORSA	Rozporządzenie 2015/35: Rozdz. III–V; Dyrektywa 2009/138/WE: Art. 40–44
III	Ujawnianie informacji i przejrzystość	Raporty SFCR i RSR, obowiązki sprawozdawcze wobec organów nadzoru i opinii publicznej	Rozporządzenie 2015/35: Rozdz. VI; Dyrektywa 2009/138/WE: Art. 51–53

Tabela 1.1: Filarowa struktura Solvency II.
Źródło: Opracowanie własne.

Filar I określa wymagania ilościowe, koncentrując się na bilansie ekonomicznym oraz na wynikających z niego funduszach własnych. W pierwszej części Rozporządzenia Delegowanego z 2014 r. przedstawiono w kolejnych rozdziałach zasady dotyczące wyceny aktywów i pasywów, rezerw techniczno-ubezpieczeniowych, środków własnych, formuły standardowej kapitałowego wymogu wypłacalności, kapitałowego wymogu wypłacalności - pełnych i częściowych modeli

wewnętrznych, a także minimalnego wymogu kapitałowego, w tym dotyczącego pozycji sekurytyzacyjnych.

Aktywa wycenia się w kwocie, za jaką w warunkach rynkowych mogłyby zostać sprzedane, natomiast pasywa - w kwocie, za jaką mogłyby zostać przeniesione lub rozliczone w warunkach rynkowych[1].

W rozporządzeniu Delegowanym z 2014 r. (Art. 9-16) zamieszczono uszczegółowiony opis zakresu oraz metod wyceny, wraz z zasadami ogólnymi i hierarchią wyceny aktywów oraz zobowiązań. Uwzględniono także zobowiązania warunkowe, metody wyceny dla jednostek powiązanych oraz dla poszczególnych kategorii zobowiązań. Określono składniki aktywów, którym przypisuje się wartość zerową (np. wartość firmy oraz wartości niematerialne i prawne), a także doprecyzowano zasady dotyczące odroczonych podatków dochodowych.

Zgodnie z Solvency II zakłady ubezpieczeń zobowiązane są do tworzenia rezerw techniczno ubezpieczeniowych w celu pokrycia zobowiązań ubezpieczeniowych lub reasekuracyjnych wobec ubezpieczonych i beneficjentów. Wartość tych rezerw odpowiada kwocie, jaką zakład ubezpieczeń musiałby posiadać na dzień bilansowy w celu uregulowania wszystkich istniejących zobowiązań wobec ubezpieczonych. W niektórych przypadkach obowiązki te mogą wykraczać w przyszłość, tak jak ma to miejsce w przypadku ubezpieczeń emerytalnych. Do obliczenia rezerw techniczno-ubezpieczeniowych wykorzystuje się dane dostępne na rynkach finansowych oraz ogólnodostępne dane dotyczące ryzyk ubezpieczeniowych[2]. Wysokość rezerw techniczno-ubezpieczeniowych wyznacza się jako sumę[3]:

- najlepszego oszacowania będącego bieżącą wartością oczekiwanych przyszłych przepływów pieniężnych, przy zastosowaniu odpowiedniej struktury terminowej stopy procentowej wolnej od ryzyka. Najlepsze oszacowanie wyznacza się w oparciu o aktualne i wiarygodne informacje, stosując metody aktuarialne i statystyczne;

- margines ryzyka mający na celu zwiększenie wartości rezerw techniczno ubezpieczeniowych do kwoty, jaką zakład ubezpieczeń lub reasekuracji zażądałby za przejęcie zobowiązań ubezpieczeniowych lub reasekuracyjnych i wywiązania się z nich.

Dodatkowo, wyznaczając rezerwy techniczno-ubezpieczeniowe, zakłady ubezpieczeń muszą uwzględnić wszystkie wydatki poniesione w związku z obsługą zobowiązań, inflację oraz wszystkie płatności na rzecz ubezpieczających oraz beneficjentów[4].

W dyrektywie Solvency II wprowadzono nową definicję pojęcia funduszy własnych (ang. own funds). W polskiej nomenklaturze fundusze własne określane są także jako środki własne. Środki własne to aktywa wolne od wszelkich przewidywalnych zobowiązań, dostępne na pokrycie strat wynikających z niesprzyjających wahań w obszarze prowadzonej działalności, zarówno przy założeniu kontynuacji działalności, jak i przy jej likwidacji[5]. Zakład ubezpieczeń określa dostępne dla niego środki własne, dzieląc je na dwie grupy:

- podstawowe środki własne to pozycje bilansowe oraz inne środki zakładu ubezpieczeń, które są dostępne do pokrycia strat i stanowią nadwyżkę aktywów nad zobowiązaniami.

[1] Art. 75 dyrektywy Solvency II.
[2] Art. 76 dyrektywy Solvency II.
[3] Art. 77 dyrektywy Solvency II.
[4] Art. 78 dyrektywy Solvency II.
[5] Ustęp 48 Preambuły do dyrektywy Solvency II.

Mają być one wolne od zobowiązań i trwale dostępne do pokrycia strat zarówno w przypadku kontynuacji działalności, jak i jej likwidacji. Przykładami podstawowych środków własnych są: opłacony kapitał zakładowy, kapitał zapasowy,kapitał pojednawczy oraz zyski zatrzymane, dostępne natychmiastowo do pokrycia strat;

- uzupełniające środki własne to pozycje, które mogą być wykorzystane do pokrycia strat, ale nie spełniają kryteriów podstawowych środków własnych lub mają ograniczenia czasowe lub warunkowe. Do uzupełniających środków własnych zalicza się instrumenty finansowe lub należności, które mogą zostać wykorzystane przez zakład ubezpieczeń do zwiększenia dostępnych środków w przypadku strat. Są to m.in. akredytywy, gwarancje, nieopłacony kapitał zakładowy lub założycielski[6], który po opłaceniu staje się częścią podstawowych środków własnych.. Jeśli w przyszłości dojdzie do opłacenia uzupełniających środków własnych, wówczas staną się one pozycjami podstawowych środków własnych.

Po określeniu dostępnych środków własnych zakład ubezpieczeń dokonuje ich klasyfikacji, dzieląc je na trzy kategorie (poziomy). W klasyfikacji uwzględnia się pozycje środków własnych podstawowych lub uzupełniających oraz stopień ich dostępności i podporządkowania[7]. Środki z poziomu pierwszego są kapitałem najwyższej klasy, np. wpłacony kapitał zakładowy. Uzupełniające środki własne nie są kwalifikowane do poziomu pierwszego, mogą jednak zostać przyporządkowane do poziomu drugiego i trzeciego. Klasyfikacja na poziomy jest istotna dla określenia dopuszczonych środków własnych (ang. eligible own funds), czyli tych środków własnych, które kwalifikują się do pokrycia wymogów kapitałowych[8]. Zakład ubezpieczeń jest wypłacalny, jeśli wartość dopuszczonych środków własnych jest większa lub równa sumie wymogów kapitałowych.

Zakłady ubezpieczeń są zobowiązane do spełnienia dwóch regulacyjnych wymogów wypłacalności, określonych w ramach dyrektywy Solvency II, a są to:

- Kapitałowy Wymóg Wypłacalności (zgodnie z wcześniejszym oznaczeniem - SCR), czyli poziom dopuszczonych środków własnych, który powinien zapewnić ubezpieczonym, że w razie wystąpienia nieprzewidzianych strat zakład ubezpieczeń wywiązuje się ze swoich zobowiązań i wypłaca odszkodowania. SCR wyznaczany jest z uwzględnieniem wszystkich mierzalnych rodzajów ryzyka, na które narażony jest ubezpieczyciel. Może zostać obliczony za pomocą formuły standardowej, danej przez Twórców dyrektywy, modeli wewnętrznych lub parametrów własnych przygotowanych przez zakłady ubezpieczeń. Obliczany jest przynajmniej raz w roku lub gdy nastąpiła istotna zmiana profilu ubezpieczyciela[9].

- Minimalny wymóg wypłacalności (zgodnie z wcześniejszym oznaczeniem - MCR) pokrywany jest przez taką wielkość dopuszczonych podstawowych środków własnych, poniżej której, przy założeniu kontynuacji prowadzenia działalności przez zakłady ubezpieczeń, ubezpieczający i beneficjenci byliby narażeni na niedopuszczalny poziom ryzyka. MCR obliczany jest jako funkcja liniowa zbioru lub podzbioru następujących zmiennych: rezerwy techniczno-ubezpieczeniowe zakładu, składka przypisana, podatki odroczone oraz

[6] Art. 89 dyrektywy Solvency II.
[7] Art. 93-97 dyrektywy Solvency II.
[8] Art. 98 dyrektywy Solvency II.
[9] Art. 100-105 dyrektywy Solvency II.

wydatki administracyjne. Zakłady ubezpieczeń obliczają MCR co najmniej raz na kwartał i informują organ nadzoru o wyniku obliczeń. MCR określa minimalny poziom środków własnych, poniżej którego ich wartość nie powinna spaść. W dyrektywie określono także dolny próg dla MCR oraz ustalono, że ma on stanowić od 25% do 45% SCR[10].

Filar II określa wymagania jakościowe dotyczące zarządzania ryzykiem i systemu nadzoru zakładów ubezpieczeń, w tym proces własnej oceny ryzyka i wypłacalności (ORSA). Z tego powodu zadania Filaru II obejmuja trzy obszary; nadzór nad strategią i procesami zakładów ubezpieczeń, wdrożenie systemu zarządzania oraz prowadzenie własnej oceny ryzyka i wypłacalności.

- Organy nadzoru ubezpieczeniowego dokonują przeglądu i oceny strategii oraz procesów sprawozdawczych zakładów ubezpieczeń i reasekuracji w celu zapewnienia zgodności z przepisami dyrektywy Solvency II. Wspomniany przegląd i ocena obejmują ocenę wymogów jakościowych związanych z systemem zarządzania, ocenę rodzajów ryzyka, na które narażony jest zakład ubezpieczeń, oraz zdolność zakładów ubezpieczeń do szacowania tych ryzyk. Organy nadzoru badają kondycję finansową oraz adekwatność metod i praktyk stosowanych przez zakłady ubezpieczeń w celu określenia zdarzeń lub przyszłych zmian warunków gospodarczych, które mogą mieć niekorzystny wpływ na kondycję finansową zakładu[11]. Jeśli zakład ubezpieczeń ustaliłby swoje wymogi kapitałowe na nieadekwatnym poziomie, to organ nadzoru zobowiązuje zakład do narzutu kapitałowego, który te wymogi podwyższy[12].

- Zgodnie z systemem zarządzania, zakłady ubezpieczeń i zakłady reasekuracji muszą posiadać pisemne zasady dotyczące zarządzania ryzykiem, kontroli wewnętrznej, audytu wewnętrznego oraz outsourcingu[13]. Sporządzone zasady muszą zostać zatwierdzone przez organ administrujący, zarządzający lub nadzorczy. Zakłady ubezpieczeń lub reasekuracji zapewniają spełnienie wymagań dotyczących kwalifikacji zawodowych oraz dobrej reputacji przez osoby sprawujące kluczowe funkcje w tych zakładach[14].

- Własna ocena ryzyka i wypłacalności (wcześniej oznaczona jako ORSA) to procesy i procedury stosowane w celu identyfikacji, oceny, monitorowania, zarządzania i raportowania krótkoterminowego oraz długoterminowego ryzyka, a także określenia funduszy własnych niezbędnych do zapewnienia, że ogólne potrzeby przedsiębiorstwa w zakresie wypłacalności są przez cały czas spełnione. ORSA łączy wymogi kapitałowe z Filaru I oraz oceny ryzyka z Filaru II, tworząc rzeczywisty poziom potrzeb w zakresie wypłacalności zakładu ubezpieczeń i uwzględniając także ryzyka, które nie zostały wzięte pod uwagę w ramach Filaru I, m. in. utrata reputacji. Firma musi również określić ilościowo dalszą zdolność do spełnienia MCR oraz SCR w horyzoncie planowania biznesowego (od trzech do pięciu lat), z uwzględnieniem nowych transakcji. Te informacje nie są jawne i są przekazywane jedynie organowi nadzoru. ORSA jest jednym z elementów branych pod uwagę przez nadzorcę przy ustalaniu, czy wymagany jest dodatkowy narzut kapitałowy.

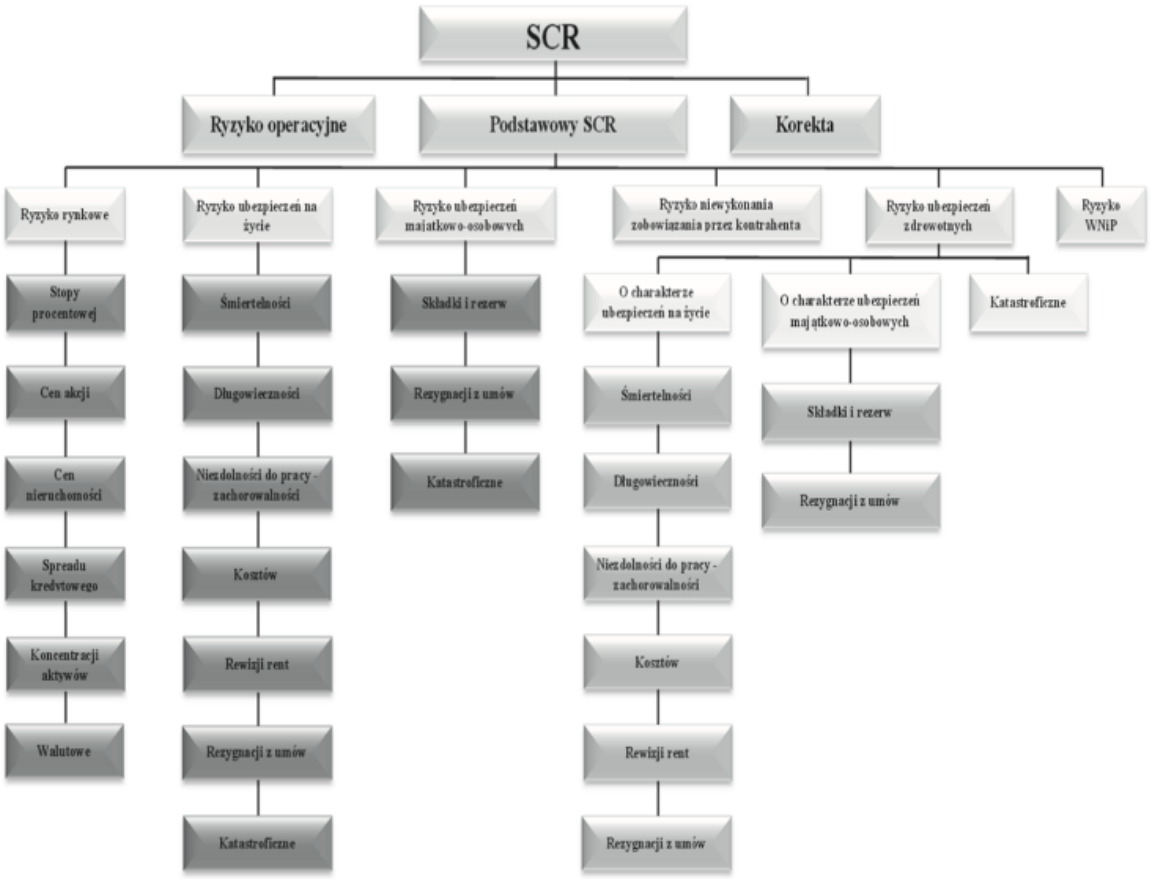
[10] Art. 129 dyrektywy Solvency II.
[11] Art. 36 dyrektywy Solvency II
[12] Art. 37 dyrektywy Solvency II
[13] (Art. 44, 46, 47, 49 dyrektywy Solvency II).
[14] Art. 42 i 43 dyrektywy Solvency II.

Filar III Solvency II obejmuje wymogi dotyczące publicznego ujawniania informacji oraz sprawozdawczości nadzorczej, zarówno na poziomie indywidualnym, jak i grupowym. Zakład ubezpieczeń jest zobowiązany do przygotowania następujących raportów:

- Raport o wypłacalności i kondycji finansowej (jw. SFCR) opisujący działalność, wyniki, operacje, profil ubezpieczyciela, zasady wyceny aktywów, rezerwy techniczno-ubezpieczeniowe, wymogi kapitałowe oraz inne zobowiązania zakładu ubezpieczeń. SFCR jest raportem publicznym, i każdy może go odnaleźć na stronie internetowej ubezpieczyciela.
- Sprawozdania nadzorcze (jw. RSR) są podobne do SFCR, ale zawierają informacje, które uznawane są za zbyt szczegółowe lub zbyt poufne do ujawnienia publicznego.
- Sprawozdania ilościowe (ang. Quantitative Reporting Templates - QRT) obejmują formularze oraz szablony sprawozdawcze, zawierające informacje ilościowe na temat zakładów ubezpieczeń. Rozróżnia się sprawozdania kwartalne oraz roczne, przy czym te pierwsze mają bardziej ograniczony zakres.

1.2 Formuła standardowa

W Filarze I wprowadzono wymóg wypłacalności SCR, który określa poziom kapitału potrzebnego, aby na poziomie ufności 99,5% zakład ubezpieczeń mógł wywiązać się ze zobowiązań wobec ubezpieczonych w przypadku wystąpienia nieprzewidzianych strat. Formuła standardowa określa sposób obliczania całkowitego SCR poprzez agregację wymogów kapitałowych poszczególnych modułów i podmodułów ryzyka. Podział na poszczególne moduły i podmoduły prezentuje poniższy schemat na Rysunku 1.1:



Rysunek 1.1: Struktura modułowa Solvency II.
Źródło: Opracowanie własne.

Zgodnie z zaprezentowanym schematem struktury modułowej (rys. 1.1), obliczany SCR stanowi sumę pozycji z poziomu pierwszego, czyli:
- podstawowego SCR (ang. Basic Solvency Capital Requirement - BSCR),
- wymogu wypłacalności z tytułu ryzyka operacyjnego,
- dostosowania z tytułu zdolności rezerw techniczno-ubezpieczeniowych i podatków odroczonych do pokrywania strat[15].

W szczególnosci dla ustalenia BSCR zakłady ubezpieczeń dokonują identyfikacji poszczególnych rodzajów ryzyka, na jakie są narażone[16], a następnie wyznaczają wymóg kapitałowy dla każdego rodzaju ryzyka i na koniec agregują wyznaczone wymogi kapitałowe, wykorzystując metodę wariancji–kowariancji. Zależności między agregowanymi ryzykami opisuje macierz korelacji liniowej, która została ustalona i podana w dyrektywie [17].

1.2.1 Metoda wariancji–kowariancji w agregacji wymogów kapitałowych

W celu ustalenia BSCR, zakłady ubezpieczeniowe agregują wyznaczone wymogi kapitałowe

[15] Art. 103 Dyrektywy
[16] (Pkt 1 zał IV Dyrektywy)
[17] Załącznik IV „Dyrektywy Parlamentu Europejskiego i Rady 2009/138/WE z dnia 25 listopada 2009 r. w sprawie podejmowania i prowadzenia działalności ubezpieczeniowej oraz reasekuracyjnej"

ustalone dla modułów zidentyfikowanych ryzyk z drugiego poziomu zgodnie ze wzorem:

$$BSCR = \sqrt{WRW^T} \tag{1.1}$$

gdzie:

- $W = [SCR_{Market}, SCR_{Counterp}, SCR_{Life}, SCR_{Health}, SCR_{Non-Life}]$
  oznacza wektor składający się z wymogów kapitałowych dla następujących modułów ryzyka:
  - SCR_{Market} dla modułu ryzyka rynkowego
  - $SCR_{Counterp}$ dla modułu ryzyka niewywiązania się ze zobowiązania przez kontrahenta (ang. Counterparty risk - Counterp.)
  - SCR_{Life} dla modułu ryzyka ubezpieczeniowego w ubezpieczeniach na życie
  - SCR_{Health} dla modułu ryzyka ubezpieczeniowego w ubezpieczeniach zdrowotnych
  - $SCR_{Non-Life}$ dla modułu ryzyka ubezpieczeniowego w ubezpieczeniach majątkowo osobowych
- R jest macierzą współczynników korelacji liniowej między modułami ryzyka, przedstawioną poniżej jako Tabela 1.2.

Moduł ryzyka Ubezpieczyciela	**Rynkowe**	**Niewykonania zobowiązania**	**Na życie**	**Na zdrowie**	**Majątkowo osobowe**
Rynkowe	1	0,25	0,25	0,25	0,25
Niewykonania zobowiązania	0,25	1	0,25	0,25	0,5
Na życie	0,25	0,25	1	0,25	0
Na zdrowie	0,25	0,25	0,25	1	0
Majątkowo osobowe	0,25	0,5	0	0	1

Tabela 1.2: Wartości korelacji między modułami ryzyka Ubezpieczyciela.
Źródło: Opracowanie własne.

Natomiast każdy z wymienionych powyżej wymogów kapitałowych dla modułów na drugim poziomie zakład ubezpieczeniowy ustala dla ryzyk określonych w następujących podmodułach:

1.2.1.1 Ryzyko rynkowe

Moduł „ryzyko rynkowe" odzwierciedla ryzyko wynikające z poziomu lub zmienności rynkowych cen instrumentów finansowych, mających wpływ na wartość aktywów i zobowiązań zakładu ubezpieczeń.

SCR dla modułu ryzyka rynkowego wyznaczamy zgodnie ze wzorem:

$$SCR_{Market} = \sqrt{WRW^T} \tag{1.2}$$

gdzie:

- $W = [SCR_{IR}, SCR_{Spread}, SCR_{Concentration}, SCR_{Currency}, SCR_{Equity}, SCR_{Property}]$ oznacza wektor składający się z wymogów kapitałowych dla następujących podmodułów ryzyka:
  - SCR_{IR} - podmoduł stopy procentowej
  - SCR_{Spread} - podmoduł ryzyka spreadu
  - $SCR_{Concentration}$ - podmoduł koncentracji ryzyka rynkowego
  - $SCR_{Currency}$ - podmoduł ryzyka wynikającego z wymiany walut
  - SCR_{Equity} - podmoduł ryzyka związanego z inwestowaniem w akcje
  - $SCR_{Property}$ - podmoduł ryzyka związanego z inwestowaniem w nieruchomości
- R jest macierzą współczynników korelacji liniowej między podmodułami ryzyka, przedstawioną w Tabeli 1.3.

	Stóp procentowych	**Spreadu**	**Koncentracji**	**Walutowe**	**Cen akcji**	**Cen nieruchomości**
Stóp procentowych	1	↑ 0/↓ 0.5	0	0.25	↑ 0/↓ 0.5	↑ 0/↓ 0.5
Spreadu	↑ 0/↓ 0.5	1	0	0.25	0.75	0.5
Koncentracji	0	0	1	0	0	0
Walutowe	0.25	0.25	0	1	0.25	0.25
Cen akcji	↑ 0/↓ 0.5	0.75	0	0.25	1	0.75
Cen nieruchomości	↑ 0/↓ 0.5	0.5	0	0.25	0.75	1

Tabela 1.3: Wartości korelacji między podmodułami w ryzyku rynkowym.
Źródło: Opracowanie własne.

Wartość korelacji dla stopy procentowej wynosi 0, jeśli wymóg kapitałowy dla ryzyka stopy procentowej stanowi wymóg kapitałowy dla ryzyka wzrostu struktury terminowej stopy procentowej wolnej od ryzyka, z uwzględnieniem zdolności rezerw techniczno ubezpieczeniowych do pokrywania strat. W pozostałych przypadkach wartość współczynnika korelacji wynosi 0,5. Wartości korelacji liniowej w macierzy R zostały ustalone i podane w dyrektywie.

1.2.1.2 Ryzyko niewykonania zobowiązania przez kontrahenta

Moduł tego ryzyka odzwierciedla straty wynikające z nieoczekiwanego niewykonania zobowiązań przez kontrahentów i dłużników zakładu ubezpieczeń. Rozróżnia się dwa rodzaje ekspozycji: typ 1 i typ 2. Typ 1 obejmuje ekspozycje, które nie mogą zostać zdefiniowane, a kiedy istnieje prawdopodobieństwo, że kontrahent posiada rating. Zawierają się w nim ekspozycje z tytułu umów ograniczania ryzyka, środków pieniężnych na rachunku bankowym oraz należności depozytowych od cedentów. Typ 2 obejmuje ekspozycje, które są dywersyfikowane, gdy istnieje prawdopodobieństwo, że kontrahent nie posiada ratingu. Ekspozycje typu 2 obejmują ekspozycje z tytułu należności od pośredników i ubezpieczonych, pożyczki zabezpieczone hipotecznie, a także nieopłacone zobowiązania i należności od cedentów. $SCR_{Counterp.}$ dla modułu ryzyka niewykonania zobowiązania przez kontrahenta wyznacza się zgodnie ze wzorem:

$$SCR_{Counterp.} = \sqrt{SCR^2_{Counterp.,1} + 1,5 \cdot SCR_{Counterp.,1} \cdot SCR_{Counterp.,2} + SCR^2_{Counterp.,2}} \tag{1.3}$$

gdzie:

- $SCR_{Counterp.,1}$ oznacza wymóg kapitałowy dla ryzyka niewykonania zobowiązania przez kontrahenta dla ekspozycji typu 1,
- $SCR_{Counterp.,2}$ oznacza wymóg kapitałowy dla ryzyka niewykonania zobowiązania przez kontrahenta dla ekspozycji typu 2,
- wielkość 1.5 opisuje wartość efektu dywersyfikacji.

1.2.1.3 Ryzyko w ubezpieczeniach na życie

Moduł tego ryzyka obejmuje ryzyko wynikające z działalności ubezpieczeniowej w zakresie ubezpieczeń na życie, związane zarówno z inherentnymi zagrożeniami, jak i z procesami zachodzącymi w trakcie prowadzenia działalności. Obliczenia wymogów kapitałowych dla podmodułów modułu ryzyka w ubezpieczeniach na życie opierają się na określonych scenariuszach opisanych w dyrektywie. Wymóg kapitałowy dla modułu ryzyka ubezpieczeniowego w ubezpieczeniach na życie wyznaczamy zgodnie ze wzorem:

$$SCR_{Life} = \sqrt{WRW^T} \tag{1.4}$$

gdzie:

- $W = [SCR_{Mort}, SCR_{Long}, SCR_{Dis}, SCR_{Exp}, SCR_{Rev}, SCR_{Lapse}, SCR_{CAT}]$ oznacza wektor składający się z następujących podmodułów ryzyka:
  - SCR_{Mort} - podmoduł ryzyka śmiertelności - SCR_{Long} - podmoduł ryzyka długowieczności
  - SCR_{Dis} - podmoduł ryzyka niezdolności do pracy – zachorowalności
  - SCR_{Exp} - podmoduł ryzyka związanego z wysokością ponoszonych kosztów
  - SCR_{Rev} - podmoduł ryzyka korekty
  - SCR_{Lapse} - podmoduł ryzyka związanego z rezygnacjami z umów
  - SCR_{CAT} - podmoduł ryzyka związanego z katastrofami w ubezpieczeniach na życie

- R jest macierzą współczynników korelacji liniowej między podmodułami ryzyka, przedstawioną w Tabeli 1.4.

Podmoduł ryzyka w ubezpieczeniach na życie	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
(1)	1	-0.25	0.25	0.25	0	0	0.25
(2)	-0.25	1	0	0.25	0.25	0.25	0
(3)	0.25	0	1	0.5	0	0	0.25
(4)	0.25	0.25	0.5	1	0.5	0.5	0.25
(5)	0	0.25	0	0.5	1	0	0
(6)	0	0.25	0	0.5	0	1	0.25
(7)	0.25	0	0.25	0.25	0	0.25	1

Objaśnienia: (1) śmiertelność; (2) długowieczność; (3) zachorowalność; (4) koszty; (5) wysokość rent; (6) rezygnacja; (7) katastrofy.

Tabela 1.4: Wartości korelacji między podmodułami ryzyka w ubezpieczeniach na życie.
Źródło: Opracowanie własne.

1.2.1.4 Ryzyko w ubezpieczeniach zdrowotnych

Moduł ten odzwierciedla ryzyko wynikające ze zobowiązań ubezpieczeniowych i reasekuracyjnych w ubezpieczeniach zdrowotnych w odniesieniu do objętych nimi zagrożeń oraz procesów odbywających się przy prowadzeniu działalności. Wymóg kapitałowy dla modułu ryzyka ubezpieczeniowego w ubezpieczeniach na zdrowie wyznaczamy zgodnie ze wzorem:

$$SCR_{Health} = \sqrt{WRW^T} \tag{1.5}$$

gdzie:

- $W = [SCR_{SLT\ Health}, SCR_{NSLT\ Health}, SCR_{Health\ CAT}]$ jest wektorem składającym się z wymogów kapitałowych dla podmodułów ryzyka:
  - $SCR_{SLT\ Health}$ - wymóg kapitałowy dla podmodułu ryzyka ubezpieczeniowego w ubezpieczeniach zdrowotnych o charakterze ubezpieczeń na życie - $SCR_{NSLT\ Health}$ - wymóg kapitałowy dla podmodułu ryzyka ubezpieczeniowego w ubezpieczeniach zdrowotnych o charakterze ubezpieczeń majątkowo-osobowych - $SCR_{Health\ CAT}$ - wymóg kapitałowy z tytułu ryzyka katastroficznego w ubezpieczeniach zdrowotnych
- R jest macierzą współczynników korelacji liniowej między podmodułami ryzyka w ubezpieczeniach zdrowotnych przedstawioną w Tabeli 1.5.

Podmoduł ryzyka w ubezpieczeniach zdrowotnych	SLT Health	NSLT Health	Health CAT
SLT Health	1	0.75	0
NSLT Health	0.75	1	0.25
Health CAT	0.25	0.25	1

Tabela 1.5: Wartości korelacji między podmodułami ryzyka w ubezpieczeniach zdrowotnych.
Źródło: Opracowanie własne.

1.2.1.5 Ryzyko w ubezpieczeniach majątkowo-osobowych

Ryzyko w tych ubezpieczeniach wynika ze zobowiązań majątkowych i osobowych w odniesieniu zarówno do rodzajów zagrożeń objętych umowami ubezpieczenia, jak i do procesów odbywających się w trakcie prowadzenia działalności. Ryzyko w ubezpieczeniach majątkowo-osobowych obejmuje również ryzyko wynikające z niepewności związanej z założeniami dotyczącymi skorzystania przez ubezpieczających z opcji, takich jak możliwość odnowienia lub rozwiązania umowy. SCR dla modułu ryzyka w ubezpieczeniach majątkowo-osobowych wyznaczamy zgodnie z poniższą formułą:

$$SCR_{Non\_Life} = \sqrt{WRW^T} \tag{1.6}$$

gdzie:

- $W = [NL_{pr}, NL_{Lapse}, NL_{CAT}]$ oznacza wektor złożony z wymogów kapitałowych:
  - NL_{CAT} - wymóg kapitałowy dla ryzyka katastroficznego - NL_{pr} - wymóg kapitałowy z tytułu składki i rezerw - NL_{Lapse} - ryzyko wynikające z rezygnacji z umów
- R jest macierzą współczynników korelacji liniowej między podmodułami ryzyka w ubezpieczeniach majątkowo-osobowych, przedstawioną w Tabeli 1.6.

Podmoduł ryzyka w ubezpieczeniach majątkowo-osobowych	NL_{pr}	NL_{Lapse}	NL_{CAT}
NL_{pr}	1	0.25	0
NL_{Lapse}	0.25	1	0
NL_{CAT}	0	0	1

Tabela 1.6: Wartości korelacji między podmodułami ryzyka w ubezpieczeniach majątkowo-osobowych.
Źródło: Opracowanie własne.

Wartość NL_{CAT} szacuje się na podstawie testów stresu. NL_{Lapse} powinien być równy stracie podstawowych środków własnych zakładów, która mogłaby wynikać z połączenia dwóch szoków: braku kontynuacji 40% polis ubezpieczeniowych, co skutkowałoby wzrostem rezerw tech-

niczno ubezpieczeniowych bez marginesu ryzyka oraz spadku o 40% liczby przyszłych bazowych umów ubezpieczenia lub reasekuracji objętych umowami reasekuracji czynnej i uwzględnianych w obliczeniach rezerw techniczno ubezpieczeniowych. NL_{pr} dla ryzyka składki i rezerw wyznacza się, postępując zgodnie z następującym wzorem:

$$NL_{pr} = \rho(\sigma) \cdot V \tag{1.7}$$

gdzie:

- V - oznacza miarę działalności ubezpieczyciela
- $\rho(\sigma)$ - jest funkcją odchylenia standardowego związaną z rozkładem zmiennej modelującej ryzyko składki i rezerw.

Początkowo w dyrektywie przyjęto rozkład logarytmiczno-normalny dla zmiennej modelującej ryzyko składki i rezerw, a funkcję $\rho(\sigma)$ określono w następujący sposób:

$$\rho(\sigma) = \frac{\exp\left(N_{0.995} \cdot \sqrt{\log(\sigma^2 + 1)}\right)}{\sqrt{\sigma^2 + 1}} - 1 \tag{1.8}$$

gdzie:

$N_{0.995}$ - oznacza kwantyl rzędu 0.995 rozkładu normalnego.

Ostatecznie jednak autorzy dyrektywy założyli rozkład normalny zmiennej modelującej ryzyko i przybliżyli funkcję $\rho(\sigma)$ formułą:

$$\rho(\sigma) \approx 3 \cdot \sigma \tag{1.9}$$

Aby wyznaczyć wartości miary działalności ubezpieczyciela V oraz odchylenia standardowego σ, ryzyko składki i rezerw dzieli się na dwanaście segmentów. Wartość V jest sumą miar działalności dla i – tego segmentu V_i, natomiast miara działalności i - tego segmentu jest sumą miar ryzyka składki $V_{(prem,i)}$ oraz miary ryzyka rezerw $V_{(res,i)}$, korygowaną o współczynnik dywersyfikacji geograficznej:

$$V_i = \left(V_{(prem,i)} + V_{(res,i)}\right) \cdot (0.75 + 0.25 \cdot DIV_i) \tag{1.10}$$

gdzie:

DIV_i - współczynnik Herfindahla dywersyfikacji składki w podziale na 18 regionów świata.

Miara ryzyka składki $V_{(prem,i)}$ dla i-tego segmentu liczona jest jako suma maksimum z szacowanej zarobionej w ciągu kolejnych 12 miesięcy składki (P_i) i zarobionej w ciągu ostatnich 12 miesięcy składki $(P_{(last,i)})$, obecnej szacowanej składki zarobionej z istniejących kontraktów po okresie kolejnych 12 miesięcy $((FP_{(existing,i)})$ oraz obecnej szacowanej składki dla kontraktów rozpoczynających się w okresie kolejnych 12 miesięcy $((FP_{(future,i)})$:

$$V_{(prem,i)} = \max(P_i; P_{(last,i)}) + FP_{(existing,i)} + FP_{(future,i)} \tag{1.11}$$

Miara ryzyka rezerw $V_{(res,i)}$ liczona jest jako wielkość najlepszego oszacowania rezerw techniczno-ubezpieczeniowych na niewypłacone odszkodowania i świadczenia.

Odchylenie standardowe σ wyznacza się jako:

$$\sigma = \sqrt{\frac{1}{V^2} \cdot \Big(\sum_{i=1}^{12} \sum_{j=1}^{12} w_{i,j} \cdot \sigma_i \cdot \sigma_j \cdot V_i \cdot V_j \Big)} \tag{1.12}$$

gdzie:

- V - miara działalności ubezpieczyciela
- V_i, V_j - miary działalności dla $i-tego$ i $j-tego$ segmentu
- σ_i, σ_j - odchylenia standardowe mierzące ryzyko składki i rezerw dla poszczególnych segmentów
- $w_{i,j}$ - współczynniki korelacji między $i-tym$ a $j-tym$ segmentem przedstawione w Tabeli 1.7

Segmenty	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**	**(11)**	**(12)**
(1)	1	0.5	0.5	0.25	0.5	0.25	0.5	0.25	0.5	0.25	0.25	0.25
(2)	0.5	1	0.25	0.25	0.25	0.25	0.5	0.5	0.5	0.25	0.25	0.25
(3)	0.5	0.25	1	0.25	0.25	0.25	0.25	0.5	0.5	0.25	0.25	0.5
(4)	0.25	0.25	0.25	1	0.25	0.25	0.25	0.5	0.5	0.5	0.25	0.5
(5)	0.5	0.25	0.25	0.25	1	0.5	0.5	0.25	0.25	0.25	0.5	0.25
(6)	0.25	0.25	0.25	0.25	0.5	1	0.5	0.25	0.5	0.25	0.5	0.25
(7)	0.5	0.5	0.25	0.25	0.5	0.5	1	0.25	0.5	0.25	0.5	0.25
(8)	0.25	0.5	0.5	0.5	0.25	0.25	0.25	1	0.5	0.5	0.25	0.25
(9)	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	1	0.25	0.25	0.5
(10)	0.25	0.25	0.25	0.5	0.25	0.25	0.25	0.5	0.25	1	0.25	0.25
(11)	0.25	0.25	0.25	0.25	0.5	0.5	0.5	0.25	0.25	0.25	1	0.25
(12)	0.25	0.25	0.5	0.5	0.25	0.25	0.25	0.25	0.5	0.25	0.25	1

Objaśnienia: (1) komunikacyjne, odpowiedzialność cywilna; (2) komunikacyjne pozostałe; (3) morskie, lotnicze i transportowe (MLT); (4) od ognia i innych szkód rzeczowych; (5) odpowiedzialność cywilna; (6) kredyty i gwarancje; (7) ochrona prawna; (8) świadczenie pomocy; (9) pozostałe nie na życie; (10) nieproporcjonalna reasekuracja - majątkowe; (11) nieproporcjonalna reasekuracja - wypadkowe; (12) nieproporcjonalna reasekuracja - MLT.

Tabela 1.7: Wartości korelacji między podmodułami ryzyka majątkowego.
Źródło: Opracowanie własne.

Odchylenie standardowe σ_i, dla $i-tego$ segmentu wyznacza się na drodze agregacji ryzyka

składki $\sigma_{i,pr}$ oraz ryzyka rezerw $\sigma_{i,res}$ zgodnie z poniższym wzorem:

$$\sigma_i = \frac{\sqrt{(\sigma_{i,pr}V_{i,pr}^2 + 2\alpha\sigma_{i,res}V_{i,pr}V_{i,res} + (\sigma_{i,res}V_{i,res})^2}}{V_i} \tag{1.13}$$

gdzie:

- α - współczynnik korelacji liniowej między rodzajami ryzyka,
- $\sigma_{i,pr}$ - wartości ryzyka składki,
- $\sigma_{i,res}$ - wartości ryzyka rezerw.

Wartości $\sigma_{i,pr}$ i $\sigma_{i,res}$ ustalone i podane w dyrektywie przedstawiamy w Tabeli 1.8:

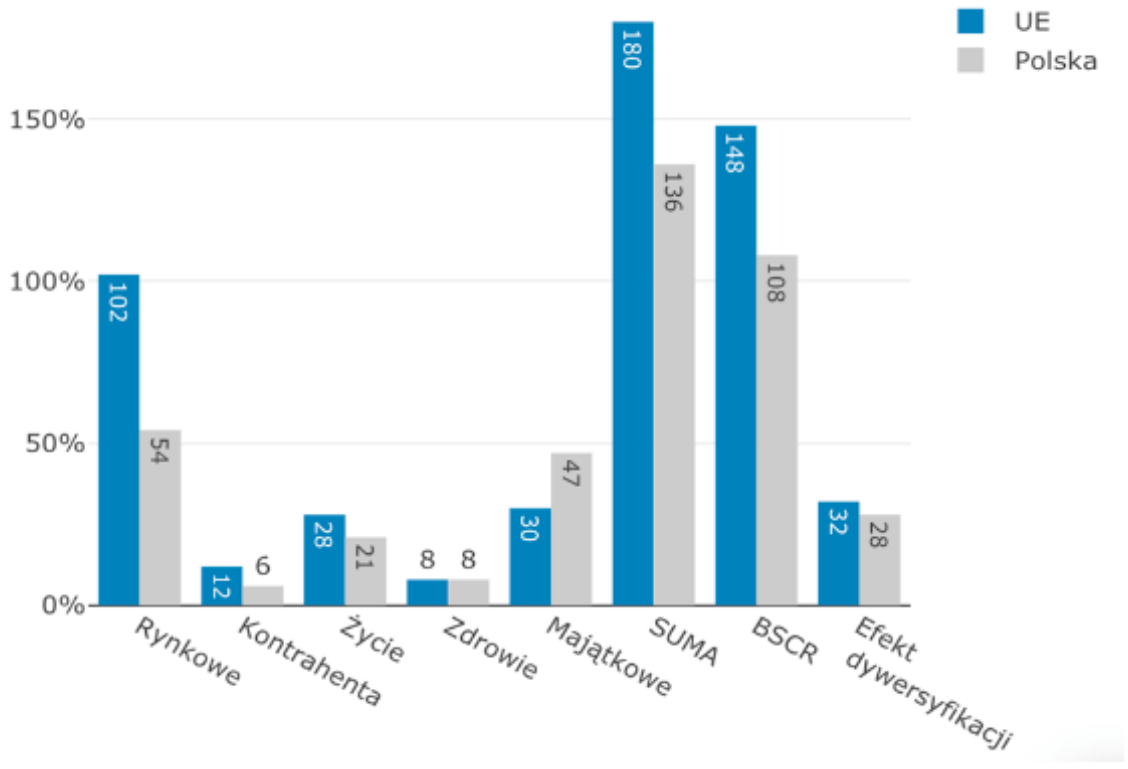
i	**Segment**	$\sigma_{i,pr}$	$\sigma_{i,res}$
1	Komunikacyjne, odpowiedzialność cywilna	0.10	0.09
2	Komunikacyjne pozostałe	0.08	0.08
3	Morskie, lotnicze i transportowe	0.15	0.11
4	Od ognia i innych szkód rzeczowych	0.08	0.10
5	Odpowiedzialność cywilna	0.14	0.11
6	Kredyty i gwarancje	0.12	0.19
7	Ochrona prawna	0.07	0.12
8	Świadczenie pomocy	0.09	0.20
9	Pozostałe nie na życie	0.13	0.20
10	Nieproporcjonalna reasekuracja – majątkowe	0.17	0.20
11	Nieproporcjonalna reasekuracja – wypadkowe	0.17	0.20
12	Nieproporcjonalna reasekuracja – MLT	0.17	0.20

Tabela 1.8: Wartości odchyleń dla ryzyka majątkowego.
Źródło: Opracowanie własne.

1.3 Wątpliwości dotyczące poprawności formuły standardowej

Wymogi kapitałowe dla modułów i podmodułów ryzyka wyznacza się na drodze agregacji rodzajów ryzyka. Te z kolei, nie wszystkie występują równocześnie, dlatego kapitał określony w wyniku takiej agregacji jest na ogół nie większy od kapitału potrzebnego do zabezpieczenia się przed każdym rodzajem ryzyka z osobna. Otrzymana różnica nazywana jest korzyścią z dywersyfikacji lub efektem dywersyfikacji i pełni ważną rolę w zakładach ubezpieczeń, ponieważ dzięki niemu ubezpieczyciel może łączyć jednostki narażone na to samo ryzyko w portfele,

redukując ewentualne niekorzystne skutki finansowe. Dodatkowo, daje to ubezpieczycielom możliwość zarządzania różnymi rodzajami ryzyka, na które narażone są portfele ubezpieczeń, segmenty oraz portfele inwestycyjne. Wykorzystanie metody wariancji–kowariancji z ustalonymi wartościami korelacji między modułami a podmodułami ryzyka zostało ocenione w piątym badaniu ilościowym (QIS5) przeprowadzonym przez Europejski Urząd Nadzoru Ubezpieczeń i Pracowniczych Programów Emerytalnych (j.w. EIOPA). Innymi słowy, badanie dotyczyło oceny efektów dywersyfikacji uzyskanych w wyniku zastosowania danej metody w FS. W badaniu wzięło udział blisko 70% zakładów ubezpieczeń i reasekuracji, które zostaną objęte dyrektywą Solvency II. Uczestniczyło w nim 50 ubezpieczycieli z Polski (24 zakłady ubezpieczeń na życie – 89% udziału w rynku oraz 26 zakładów ubezpieczeń Działu II – 89% udziału w rynku). Z raportu wynika, że w rezultacie zastosowania standardowego podejścia suma wymogów kapitałowych (łącznego ryzyka) dla wszystkich nośników ryzyka zakładów biorących udział w badaniu wynosi 1328 mld euro, przy czym efekt dywersyfikacji oszacowano na poziomie 466 mld euro (35,1 proc.), a zdolność pokrywania strat z rezerw techniczno-ubezpieczeniowych oraz podatków odroczonych wynosi 314 mld euro (23,7 proc.). Dało to oszacowanie kapitałowego wymogu wypłacalności dla badanych zakładów na poziomie 547 mld euro, co stanowiło około 41,2%. Dla ubezpieczycieli Unii Europejskiej solo (bez grup kapitałowych) efekt dywersyfikacji wynosił 32%; oznacza to, że łączne wymogi kapitałowe dla tych modułów są o 32% niższe od sumy ich wymogów. Dla ubezpieczycieli z Polski efekt dywersyfikacji oszacowano na poziomie 28%. Wyniki tych badań przedstawiono na Rysunku 1.2.



Rysunek 1.2: Efekt dywersyfikacji ubezpieczycieli z Polski.
Źródło: Opracowanie własne.

Można zauważyć, że otrzymany efekt dywersyfikacji z wykorzystaniem metody warian-

cji–kowariancji jest znaczący. Ustalenie efektu dywersyfikacji zależy od wielu czynników, takich jak poziom korelacji, zależności w ogonach rozkładów, liczba czynników ryzyka, kształt rozkładów podstawowych oraz struktura zależności między nimi. Badania QIS5 potwierdzają, że efekt dywersyfikacji odgrywa istotną rolę w zarządzaniu ryzykiem ubezpieczyciela, ale metoda agregacji ryzyka wskazana w dyrektywie - metoda wariancji–kowariancji - jest metodologicznie poprawna jedynie wtedy, gdy rozkłady agregowanych czynników ryzyka są normalne, a struktura zależności między nimi może być opisana za pomocą współczynnika korelacji liniowej. Z niedoskonałości tej metody zdają sobie sprawę sami jej twórcy, którzy zezwalają, a nawet zachęcają zakłady ubezpieczeń do budowania modeli wewnętrznych, które w lepszym stopniu oddadzą profil działalności zakładu i ustalą ED na właściwym poziomie. Doświadczenia praktyków w budowaniu modeli wewnętrznych skłoniły EIOPA w październiku 2020 r. do wszczęcia ogólnoeuropejskich badań porównawczych dotyczących dywersyfikacji w modelach wewnętrznych; „Study on Diversification in Internal Models". Ustalono trzy cele badania:

1. uzyskanie przeglądu obecnych podejść na rynku w zakresie określania ED oraz przeanalizowanie i porównanie jego poziomu,

2. lepsze zrozumienie związku między sposobami modelowania zależności a agregacją ryzyka oraz wynikającymi z tego korzyściami z dywersyfikacji,

3. poprawa jakości i konwergencji nadzoru nad dywersyfikacją w modelach wewnętrznych.

Badanie zostało zaplanowane w dwóch etapach, aby zrównoważyć złożoność oraz kompletność analiz. Pierwsza faza, rozpoczęta na początku października 2020 r., koncentrowała się na zależnościach między głównymi modułami ryzyka, tj. ryzykiem rynkowym, kredytowym, na życie, majątkowym, zdrowotnym oraz operacyjnym (zob. tab. 1.2). Etap ten zakończył się w styczniu 2021 r. i dostarczył podstawowych informacji na temat struktury powiązań między modułami ryzyka. Aby uzyskać pełniejsze zrozumienie efektów dywersyfikacji, uwzględniono również profil ryzyka poszczególnych zakładów ubezpieczeń.

Druga faza badania, uruchomiona w październiku 2021 r. i zakończona w styczniu 2022 r., została poświęcona ocenie zależności między podmodułami ryzyka, które składają się na poszczególne moduły. Dla przykładu, w odniesieniu do modułu ubezpieczeń innych niż na życie, zebrano dane dotyczące całkowitego ryzyka oraz jego struktury na poziomie europejskim. Ponadto, przeprowadzono szczegółowe analizy kilku segmentów działalności w celu identyfikacji dominujących czynników ryzyka - w szczególności w obszarach odpowiedzialności cywilnej, ubezpieczeń komunikacyjnych, pożarowych i ogólnych, a także w zakresie ryzyka kredytów i poręczeń. Badanie umożliwiło zakładom ubezpieczeń przedstawienie własnych opinii na temat profilu ryzyka przy wykorzystaniu technicznie ugruntowanego punktu odniesienia, jakim są ramy regulacyjne Solvency II. Ostateczny raport podsumowujący (tzw. „Comparative Study on diversification in internal models") został opublikowany 24 stycznia 2024. Przy czym EIOPA zaznaczyłą, że będzie kontynuować nadzór i monitoring wyników wdrożeń.

1.4 Modele wewnętrzne

Zakłady ubezpieczeń i reasekuracji mogą obecnie wyznaczać SCR, stosując FS daną przez twórców dyrektywy lub za pomocą częściowych bądź pełnych modeli wewnętrznych. Modele wewnętrzne umożliwiają lepsze dopasowanie SCR do profilu prowadzonej działalności ubezpieczeniowej. Zakłady ubezpieczeń i reasekuracji mogą stosować częściowe modele wewnętrzne

w celu obliczania jednego lub kilku modułów i podmodułów podstawowego kapitałowego wymogu wypłacalności, ryzyka operacyjnego, dostosowania SCR z tytułu zdolności rezerw techniczno-ubezpieczeniowych oraz podatków odroczonych na pokrycie strat[18]. Przygotowany model wewnętrzny musi zostać zatwierdzony przez organ nadzoru i ma spełniać szereg wymogów[19]. Do najważniejszych należą:

- Test użyteczności, w którym zakłady ubezpieczeń i reasekuracji wykazują, że model wewnętrzny jest powszechnie stosowany i odgrywa ważną rolę w ich systemie zarządzania ryzykiem, procesach decyzyjnych oraz w procesach oceny i alokacji kapitału gospodarczego oraz kapitału zabezpieczającego wypłacalność.

- Standardy statystyczne, zgodnie z którymi zakład ubezpieczeń udowadnia, że metody stosowane do wyznaczania prognozy rozkładu prawdopodobieństwa opierają się na odpowiednich metodach aktuarialnych i technikach statystycznych oraz są zgodne z metodami stosowanymi do ustalenia wartości rezerw techniczno ubezpieczeniowych. Stosowane metody opierają się na wiarygodnych informacjach i realistycznych założeniach, a wykorzystane dane są dokładne, kompletne i adekwatne.

- Standardy kalibracji, w których zakłady ubezpieczeń wskazują, że zastosowany horyzont czasowy oraz miara ryzyka w modelu wewnętrznym zapewniają ubezpieczonym i beneficjentom poziom ochrony równoważny poziomowi określonemu w dyrektywie. Jeśli jest to możliwe, SCR obliczany jest na podstawie prognozy rozkładu prawdopodobieństwa; natomiast jeśli nie jest to możliwe, organy nadzoru mogą zezwolić na zastosowanie przybliżeń, pod warunkiem że ubezpieczyciele wykażą, iż poziom ochrony ubezpieczających jest równoważny z poziomem określonym w dyrektywie. Ponadto, organy nadzoru mogą zażądać weryfikacji modelu na wzorcowych portfelach umów oraz zewnętrznych danych.

- Przypisanie zysków i strat, w których zakłady ubezpieczeń dokonują przeglądu przyczyn i źródeł powstania zysków i strat każdego głównego obszaru działalności zakładu. Wykazują, w jaki sposób wybrana w modelu wewnętrznym kategoryzacja ryzyka wyjaśnia przyczyny i źródła powstawania zysków oraz strat.

- Standardy walidacji, w których zakłady ubezpieczeń dokonują analizy stabilności modelu wewnętrznego, badają wrażliwość uzyskanych wyników na zmiany podstawowych założeń leżących u jego podstaw. Obejmuje on również ocenę dokładności, kompletności i adekwatności danych stosowanych w modelu wewnętrznym.

- Standardy dokumentacji, w których zakłady ubezpieczeń sporządzają dokumentację zawierającą szczegółowy opis teorii, założeń oraz podstaw matematycznych i empirycznych leżących u podstaw modelu wewnętrznego. Dodatkowo w dokumentacji określone są przypadki, w których model wewnętrzny nie działa prawidłowo.

Gdy model wewnętrzny spełnia przytoczone w Dyrektywie opisane powyżej wymogi, zakład ubezpieczeń może przystąpić do procesu jego zatwierdzenia przez organ nadzoru. Przed złożeniem związanego z tym oficjalnego wniosku zakłady ubezpieczeń mogą (choć nie muszą) wziąć udział w procesie przedadaptacyjnym, którego celem jest zapoznanie organu nadzoru z

[18] Art. 112 dyrektywy Solvency II.
[19] Art. 120-125 dyrektywy Solvency II.

przygotowanym modelem wewnętrznym. W momencie udziału zakładu ubezpieczeń w procesie przedadaptacyjnym nie jest konieczne posiadanie pełnej dokumentacji ani przygotowanego całego modelu wewnętrznego. Niemniej jednak brak odpowiednich dokumentów może utrudnić lub uniemożliwić współpracę organu nadzoru z ubezpieczycielem. Organ nadzoru może wymagać od zakładu ubezpieczeń uzupełnienia informacji dotyczących modelu. Zakłady ubezpieczeń podczas procesu przeadaptacyjnego mogą wykryć elementy modelu wewnętrznego, które wymagają poprawy. Udział zakładu ubezpieczeń w procesie przedadaptacyjnym nie wpływa na to, czy złożony później wniosek o zatwierdzenie modelu wewnętrznego zostanie zatwierdzony, czy też nie. Wniosek o wykorzystanie modelu wewnętrznego zakład ubezpieczeń składa w formie pisma przewodniego wraz z załącznikami, w skład których wchodzą m.in. informacje na temat polityki zmian w modelu wewnętrznym, zakres modelu wewnętrznego, wyniki badań ORSA, plany dotyczące ulepszenia modelu wewnętrznego oraz porównanie SCR otrzymanego z wykorzystaniem Formuły Standardowej i modelu wewnętrznego. Wszystkie dokumenty powinny zostać dostarczone w języku narodowym kraju, w którym znajduje się organ nadzoru. Od momentu otrzymania kompletnych dokumentów od zakładu ubezpieczeń, organ nadzoru w przeciągu 6 miesięcy podejmuje decyzję o zatwierdzeniu lub odrzuceniu modelu wewnętrznego. Podczas pierwszego procesu zatwierdzenia modelu wewnętrznego organy nadzoru wyrażają zgodę na politykę zmian, zgodnie z którą zakład ubezpieczeń będzie mógł wprowadzać modyfikacje w modelu wewnętrznym. Organ nadzoru może zatwierdzić model wewnętrzny lub zatwierdzić go warunkowo, co oznacza, że będzie na bieżąco monitorować implementację przez zakład ubezpieczeń warunków narzuconych na model wewnętrzny, albo może odrzucić wniosek o zatwierdzenie modelu wewnętrznego. Przy odrzuceniu wniosku organ nadzoru wydaje uzasadnienie swojej decyzji. Po zatwierdzeniu wniosku przez organ nadzoru, zakład ubezpieczeń może wprowadzać w nim zmiany zgodnie z zatwierdzoną polityką wprowadzania zmian. Została ona podzielona na dwa rodzaje wprowadzanych zmian: główne zmiany w modelu i zmiany drugorzędne[20]. Główne zmiany muszą zostać zatwierdzone przez organ nadzoru, a do momentu ich zatwierdzenia zakład ubezpieczeń musi korzystać z modelu wewnętrznego, który nie zawiera zmian. Drugorzędne zmiany modelu wewnętrznego nie muszą być zatwierdzane przez organ nadzoru, jeśli są zgodne z polityką zmian. Natomiast zakład ubezpieczeń ma obowiązek sporządzać raporty z wprowadzonych drugorzędnych zmian do modelu wewnętrznego i przedstawiać je organowi nadzoru. Ten z kolei może zakwestionować wprowadzone zmiany, uznając, że są to główne zmiany w modelu lub nie spełniają one standardów Solvency II.

[20] Art. 115 dyrektywy Solvency II.

2. Modelowanie zależności w procesie agregacji ryzyka a efekt dywersyfikacji

Agregacja ryzyka stanowi powszechnie występującą koncepcję w obszarze zarządzania ryzykiem zakładów ubezpieczeń. W ujęciu ogólnym odnosi się ona do procesu łączenia oraz analizy powiązań pomiędzy poszczególnymi kategoriami ryzyka w celu uzyskania całościowego obrazu profilu ryzyka przedsiębiorstwa. Zgodnie z definicją przedstawioną przez ISO (2009), agregacja ryzyka oznacza połączenie różnych rodzajów ryzyka w jedno, co umożliwia pełniejsze zrozumienie ogólnego poziomu ryzyka. Natomiast, Bazylejski Komitet Nadzoru Bankowego (BCBS, 2010) definiuje ją jako proces integrowania mniej złożonych miar ryzyka w ramach organizacji w celu uzyskania bardziej kompleksowej miary całościowej ekspozycji na ryzyko. W ramach systemu Solvency II agregacji podlegają wymogi kapitałowe, wyrażone poprzez kapitał ekonomiczny określany jako różnica między wartością zagrożoną (VaR) a wartością oczekiwaną. W szczególnym przypadku, gdy wymóg kapitałowy wyznaczany jest bezpośrednio na podstawie miary ryzyka (np. (VaR)), procesowi agregacji podlegają właśnie te miary ryzyka. W innych sytuacjach agregacji mogą nie podlegać same kapitały ekonomiczne, określone na podstawie miar ryzyka, lecz informacje opisujące zróżnicowanie zmiennej modelującej ryzyko, takie jak wariancja czy odchylenie standardowe.

Metody agregacji ryzyka, stosowane w celu wyznaczenia kapitału ekonomicznego odpowiadającego łącznej ekspozycji na ryzyko, można podzielić na dwie główne grupy. Pierwsza grupa obejmuje metody, które agregują kapitały ekonomiczne lub miary ryzyka dla poszczególnych rodzajów ryzyka, bez uwzględniania rzeczywistej struktury zależności pomiędzy nimi. Wybór konkretnej metody agregacji implikuje tym samym określoną strukturę zależności (najczęściej komonotoniczną). Druga grupa metod, w odróżnieniu od pierwszej, koncentruje się na modelowaniu wielowymiarowego rozkładu ryzyka, co pozwala na bardziej precyzyjne odwzorowanie zależności pomiędzy poszczególnymi kategoriami ryzyka. W niniejszej pracy uwaga skupiona jest na metodach należących do drugiej grupy, w których kluczową rolę odgrywa struktura zależności między agregowanymi rodzajami ryzyka.

2.1 Miary ryzyka

Miary ryzyka stanowią kluczowe narzędzie umożliwiające ubezpieczycielom określenie wielkości kapitału niezbędnego do zabezpieczenia się przed potencjalnymi stratami wynikającymi z prowadzonej działalności. Umożliwiają one ilościowe ujęcie ekspozycji na ryzyko oraz skuteczne zarządzanie jego poziomem poprzez wyrażenie go w postaci jednej liczby rzeczywistej. W Solvency II, dysponując zmienną losową opisującą straty dla danego rodzaju ryzyka, ubezpieczyciel dąży do wyznaczenia konkretnej wartości wymogu kapitałowego, która zapewni mu utrzymanie wypłacalności. W tym celu stosowane są miary ryzyka, które umożliwiają określenie poziomu kapitału koniecznego do pokrycia potencjalnych strat. W dalszej części podrozdziału przedstawiono koncepcję koherentnych miar ryzyka, stanowiących podstawę teoretyczną procesu agregacji ryzyka w ramach nowoczesnych systemów zarządzania kapitałem.

Rozważmy przestrzeń probabilistyczną $(\Omega, \mathcal{A}, \mathbb{P})$ oraz zbiór wszystkich zmiennych losowych o wartościach rzeczywistych w tej przestrzeni $L^0 := L^0(\Omega, \mathcal{A}, \mathbb{P})$. Rozkłady zmiennych losowych $X_1, X_2, X_3 \ldots$ należących do przestrzeni L^0 są dane za pomocą dystrybuant

$F_1, F_2, F_3 \ldots$. W dalszej części pracy przyjęto, że modelują one indywidualne rodzaje ryzyka ubezpieczyciela. Miara ryzyka to funkcjonał ϱ odwzorowujący L^0 w zbiór liczb rzeczywistych $\varrho : L^0 \to \mathbb{R} \cup \{+\infty\}$. Oczywistym wyzwaniem jest sprowadzenie informacji o całym rozkładzie zmiennej losowej do jednej liczby, opisującej jej ryzyko. Z tego względu kwestia odpowiedniego zaprojektowania miary ryzyka nabiera istotnego znaczenia praktycznego i ma kluczowe znaczenie zarówno dla sektora finansowego, jak i ubezpieczeniowego, oraz dla organów nadzoru regulacyjnego. Postać przyjmowanej miary ryzyka wynika z przyjętych aksjomatów narzuconych na funkcjonał ϱ, które determinują jego własności teoretyczne oraz sposób interpretacji w kontekście oceny ryzyka. W literaturze można znaleźć wiele konstrukcji miar ryzyka (Pedersen & E., 1998; Artzner, Delbaen, Eber, & Heath, 1999; Rockafellar, 2002; Venter, 2009; Dhaene, Vanduffel, Goovaerts, Kaas, & Vyncke, 2006; Byrne, 2004; Föllmer, 2011). W niniejszej pracy rozważać będziemy wprowadzone w (Artzner et al., 1999) koherentne miary ryzyka ϱ.

Miara ryzyka ϱ jest określana jako koherentna, jeżeli spełnia następujące cztery aksjomaty:

1. Monotoniczność:

   Dla dowolnych X_i, X_j z L^0, jeśli $X_i \leq_{st} X_j$, to $\varrho(X_i) \leq \varrho(X_j)$, gdzie $\leq_{st}$ oznacza porządek stochastyczny[21].

   Aksjomat ten gwarantuje racjonalność oceny ryzyka. Jeżeli strata X_i jest mniejsza lub równa stracie X_j w każdym scenariuszu, to ryzyko przypisane do X_i nie może być większe niż ryzyko przypisane do X_j.

2. Translacyjna ekwiwariantność: [22]

   Dla dowolnych $a \in \mathbb{R}$ i X_i z L^0 zachodzi $\varrho(X_i + a) = \varrho(X_i) + a$

   Dodanie pewnej deterministycznej kwoty c do pozycji X_i powoduje wzrost miary ryzyka o dokładnie tę samą wartość. Aksjomat ten umożliwia interpretację $\varrho(X_i)$ jako minimalnego wymaganego kapitału zabezpieczającego.

3. Dodatnia jednorodność:

   Dla dowolnych $\lambda \geq 0$ i X_i z L^0 zachodzi $\varrho(\lambda X_i) = \lambda \varrho(X_i)$ Powiększenie pozycji portfela λ-krotnie skutkuje proporcjonalnym wzrostem wymaganego kapitału. Warunek ten wprowadza liniową skalowalność ryzyka.

4. Subaddytywność:

   Dla dowolnych X_i, X_j z L^0 zachodzi $\varrho(X_i + X_j) \leq \varrho(X_i) + \varrho(X_j)$

   Łączne ryzyko dwóch pozycji nie może być większe od sumy ryzyk indywidualnych. Własność ta odzwierciedla efekt dywersyfikacji. Połączenie niezależnych źródeł ryzyka prowadzi do redukcji całkowitego ryzyka portfela.

Aksjomaty te zapewniają wewnętrzną spójność oraz intuicyjną interpretowalność miary ryzyka w kontekście zarządzania kapitałem. Zapewniają, że jest zgodna z podstawowymi zasadami zarządzania ryzykiem:

[21] Więcej informacji na temat tego porządku można znaleźć w (Muller i Stoyan, 2002).

[22] W literaturze, a szczególnie w aktuarialnej i finansowej (zwłaszcza w kontekście Solvency II, Basel III), bardziej rozpowszechniona jest nazwa „translacyjna niezmienniczość" lub „niezmienniczość względem przesunięcia".

- Monotoniczność gwarantuje racjonalność porównań ryzyk,
- Translacyjna ekwiwariantność pozwala interpretować miarę jako wymóg kapitałowy,
- Dodatnia jednorodność zapewnia liniową zależność między ekspozycją a ryzykiem,
- Subaddytywność formalnie opisuje korzyści wynikające z dywersyfikacji.

Nie wszystkie miary ryzyka, popularne zarówno wśród organów regulacyjnych, jak i zarządzających ryzykiem, spełniają powyższe warunki. Przykładem może być miara wartości zagrożonej (ang. Value at Risk – VaR), która, mimo szerokiego zastosowania w praktyce, nie jest subaddytywna, co może prowadzić do przeszacowania ryzyka w portfelach zdywersyfikowanych. Z kolei, oczekiwana strata warunkowa (ang. Expected Shortfall – ES), zwana również warunkową wartością zagrożoną (ang. Conditional Value at Risk – $CVaR$), spełnia wszystkie cztery aksjomaty i stanowi przykład w pełni koherentnej miary ryzyka.

Wartość zagrożoną na poziomie ufności p definiuje się jako:

$$VaR_p(X_i) = inf\{x \in \mathbb{R} : \mathbb{P}(X_i \leq x) \geq p\}, p \in (0,1), X_i \in L^0 \tag{2.1}$$

natomiast ES jako:

$$ES_p(X_i) = \frac{1}{1-p}\int_p^1 VaR_q(X_i)dq, p \in (0,1), X_i \in L^0 \tag{2.2}$$

Ponadto, oczekiwana wartość w dolnym ogonie rozkładu (ang. Lower Expected Shortfall – LES) jest dana wzorem:

$$LES_p(X_i) = \frac{1}{p}\int_0^p VaR_q(X_i)dq, p \in (0,1), X_i \in L^0 \tag{2.3}$$

W kontekście agregacji ryzyka w Solvency II miara ryzyka pełni kluczową rolę, ponieważ na jej podstawie wyznaczane są wymogi kapitałowe dla pojedynczych rodzajów ryzyka oraz wymóg kapitałowy dla zagregowanego ryzyka. Zgodnie z dyrektywą, wymogi kapitałowe wyznaczane są w oparciu o VaR, a zatem o miarę, która nie jest koherentna, ponieważ nie spełnia aksjomatu subaddytywności. Brak subaddytywności jest sprzeczny z poglądem, że z agregacją ryzyka powinna wiązać się korzyść z dywersyfikacji. Zatem nie ma pewności, że wymóg kapitałowy dla łącznego ryzyka (agregując dwa rodzaje ryzyka, jest to $VaR_p(X_i+X_j)$) będzie mniejszy od sumy wymogów kapitałowych dla poszczególnych rodzajów ryzyka ($VaR_p(X_i)+VaR_p(X_j)$. W związku z tym kwestia spełnienia aksjomatu subaddytywności w kontekście agregacji ryzyka ma fundamentalne znaczenie zarówno teoretyczne, jak i praktyczne. W literaturze szeroko dyskutowano problem addytywności miary VaR oraz konsekwencje wynikające z niespełnienia tego warunku, m.in. w pracach (Danielsson et al., 2013; P. Embrechts, McNeil, & Straumann, 2002; P. G. R. L. Embrechts P., 2009; P. Embrechts, Puccetti, & Rüschendorf, 2013) Co więcej, VaR na poziomie p nie daje informacji o dotkliwości strat z ogona, które występują z prawdopodobieństwem mniejszym niż $1-p$, w przeciwieństwie do ES na tym samym poziomie ufności, który opiera się na średniej strat powyżej p (por. (Artzner et al., 1999; Dhaene et al., 2006)).

2.2 Agregacja ryzyka i efekt dywersyfikacji

Ubezpieczyciele w swojej działalności podlegają ekspozycji na liczne i zróżnicowane rodzaje ryzyka, oznaczane dalej jako $X_1, \ldots, X_d$. Poszczególne kategorie ryzyka mogą mieć

odmienne źródła oraz charakter, obejmując między innymi ryzyko ubezpieczeniowe, rynkowe, kredytowe i operacyjne; jednak ich łączne oddziaływanie determinuje całkowitą ekspozycję zakładu ubezpieczeń na ryzyko.

Zasadniczym elementem procesu zarządzania ryzykiem jest wyznaczenie poziomu kapitału $\kappa(Y)$, gdzie Y oznacza zagregowaną zmienną ryzyka wynikającą z połączenia składników $X_1, \ldots, X_d$. Zmienna Y może być zdefiniowana w sposób ogólny jako:

$$Y = \psi(X_1, X_2, \ldots, X_d), \tag{2.4}$$

gdzie $\psi(\cdot)$ jest funkcją (metodą) agregacji, opisującą sposób łączenia ryzyk cząstkowych w całkowite ryzyko portfela. Kapitał $\kappa(Y)$ określa poziom zabezpieczenia niezbędny do pokrycia potencjalnych strat wynikających z realizacji zdarzeń losowych wpływających na działalność zakładu ubezpieczeń. Proces ten stanowi podstawę koncepcji agregacji ryzyka, której nieodłącznym elementem jest efekt dywersyfikacji, wynikający z istnienia zależności pomiędzy poszczególnymi źródłami ryzyka. Uwzględnienie tych zależności w modelu ma kluczowe znaczenie dla racjonalnego określenia wymogu kapitałowego oraz efektywnego zarządzania portfelem ryzyk, ponieważ umożliwia bardziej realistyczne odwzorowanie zachowania całkowitej ekspozycji w warunkach niepewności.

W praktyce można wyróżnić dwa zasadnicze podejścia do agregacji ryzyka, różniące się sposobem traktowania struktury zależności pomiędzy składnikami $X_1, \ldots, X_d$.

Pierwsza grupa metod obejmuje podejścia, w których struktura zależności pomiędzy zmiennymi losowymi nie jest jawnie modelowana. W tym ujęciu proces agregacji koncentruje się wyłącznie na łączeniu poszczególnych składników ryzyka w jedną zmienną zagregowaną, bez szczegółowego opisu powiązań między nimi. Nie oznacza to jednak przyjęcia założenia o ich statystycznej niezależności, lecz raczej braku explicite określonego modelu współzależności. Zastosowanie konkretnej metody agregacji z tej grupy (na przykład sumy składników lub agregacji opartej na wartości ekstremalnej) implikuje określoną, często skrajną strukturę zależności, taką jak struktura komonotoniczna, w której wszystkie składniki osiągają swoje wartości ekstremalne jednocześnie. W efekcie podejścia nieuwzględniające zależności mogą prowadzić do istotnych różnic w ocenie całkowitego ryzyka portfela, w zależności od przyjętego założenia o relacjach pomiędzy składnikami X_i.

Druga grupa metod obejmuje podejścia, które w sposób jawny modelują strukturę zależności pomiędzy składnikami ryzyka. Ujęcie to pozwala na bardziej realistyczne odwzorowanie efektu dywersyfikacji, ponieważ umożliwia uwzględnienie współzależności pomiędzy poszczególnymi źródłami ryzyka. Do tej grupy zalicza się między innymi podejścia oparte na założeniu niezależności składników ryzyka, a także bardziej zaawansowane metody, w których estymowany jest pełny rozkład wielowymiarowy portfela z wykorzystaniem funkcji kopuli. Modelowanie zależności za pomocą kopuli pozwala oddzielić rozkłady brzegowe od struktury współzależności, dzięki czemu możliwe jest elastyczne odwzorowanie zarówno liniowych, jak i nieliniowych zależności występujących pomiędzy ryzykami.

W dalszej części rozdziału omówione zostaną szczegółowo metody należące do obu grup, ze szczególnym uwzględnieniem podejść wykorzystujących modelowanie struktury zależności zarówno w ujęciu parametrycznym, jak i z zastosowaniem sieci neuronowych.

2.2.1 Agregacja bez uwzględniania struktury zależności

W pierwszej grupie metod agregacji ryzyka ze wspomnianych na wstępie niniejszego rozdziału wyróżnia się podejścia służące do wyznaczania kapitału $\kappa(Y)$ w oparciu o indywidualne

rodzaje ryzyka, bez identyfikacji rzeczywistej struktury zależności pomiędzy nimi. Do najprostszych metod należą w tym przypadku metody ważone. Potrzebny kapitał $\kappa(Y)$ wyznaczany jest w oparciu o znajomość wielkości straty oraz prawdopodobieństwa jej wystąpienia. Te informacje można uzyskać zarówno z rozkładów dyskretnych, jak i ciągłych. Wówczas kapitał jest dany jako:

$$\kappa(Y) = \sum_{i=1}^{d} c_{X_i} p_{X_i} \tag{2.5}$$

gdzie c_{X_i} to wielkość straty, a p_{X_i} to prawdopodobieństwo wystąpienia straty dla ryzyka X_i.

Bardzo podobną do powyższej metody jest metoda, w której nadaje się udziały (wagi) poszczególnym rodzajom ryzyka. Dla przykładu, agregując segmenty w podmodule ryzyka i składki w ubezpieczeniach majątkowych, udziały poszczególnych segmentów w portfelu można określić na podstawie składki zarobionej. Wówczas kapitał wyznacza się jako:

$$\kappa(Y) = \sum_{i=1}^{d} c_{X_i} w_{X_i} \tag{2.6}$$

gdzie c_{X_i} to strata, a w_{X_i} to udział ryzyka X_i w portfelu.
Kolejne metody z tej grupy opierają się na sumowaniu wyznaczonych miar ryzyka. Jeśli rodzaje ryzyka są przez ciągłe zmienne losowe X_i, to potrzebny kapitał $\kappa(Y)$ można wyznaczyć jako:

- Średnią wartość strat wynikających z poszczególnych rodzajów ryzyk X_i:

$$\kappa(Y) = \sum_{i=1}^{d} \mathbb{E}(X_i) \tag{2.7}$$

  gdzie $E(X_i)$ jest wartością oczekiwaną.

- Graniczny poziom straty poprzez wyznaczenie wartości zagrożonej VaR_{X_i} dla każdego z ryzyk X_i i ich zsumowania:

$$\kappa(Y) = \sum_{i=1}^{d} VaR_p(X_i) \tag{2.8}$$

- Średnią wartość strat większych niż graniczny poziom straty wyznacza się poprzez określenie warunkowej wartości zagrożonej ES_{X_i} dla każdego z ryzyk X_i i ich zsumowanie:

$$\kappa(Y) = \sum_{i=1}^{d} ES_p(X_i) \tag{2.9}$$

- Suma kapitałów ekonomicznych (podejście stosowane w Solvency II) danych jako różnica granicznego poziomu straty i średniej wartości straty dla każdego z ryzyk X_i:

$$\kappa(Y) = \sum_{i=1}^{d} \left(VaR_p(X_i) - \mathbb{E}(X_i)\right) \tag{2.10}$$

Przedstawiona powyżej grupa metod agregacji w celu wyznaczenia potrzebnego kapitału koncentruje uwagę na indywidualnych rodzajach ryzyka $X_1, \ldots, X_d$, nie uwzględniając pełnego obrazu wielowymiarowego ryzyka $(X_1, \ldots, X_d)$. Kluczową kwestią w wyznaczeniu wielowymiarowego rozkładu jest identyfikacja zależności pomiędzy agregowanymi rodzajami ryzyka. W rzeczywistości ryzyka ubezpieczyciela często charakteryzują się zależnością, której przejawem jest np. ich współwystępowanie w czasie. Nowe uwzględnienie zależności między agregowanymi rodzajami ryzyka ubezpieczyciela może prowadzić do negatywnych skutków dla zakładu. Aby to zilustrować, rozważmy skrajny scenariusz, w którym dwa rodzaje ryzyka są przeciwnie zależne w tym sensie, że jeśli jedno ryzyko maleje, to drugie rośnie. Patrząc na sumę ryzyk, ich tendencje doskonale się równoważą, tak że ryzyka zapewniają ubezpieczycielowi naturalne zabezpieczenie. Jako inny klasyczny przykład rozważmy d niezależnych ryzyk o tym samym rozkładzie i ustalonych odchyleniach standardowych wykorzystanych do pomiaru niepewności portfela. Proste obliczenia pokazują, że odchylenie standardowe portfela jest proporcjonalne do pierwiastka liczby agregowanych ryzyk $(\sqrt{d})$. Można to interpretować w ten sposób, że niepewność zmniejsza się w stosunku do średniej portfela, czyli jest proporcjonalna do d, wraz ze wzrostem liczby ryzyk d.

2.2.2 Agregacja z uwzględnieniem struktury zależności

Druga grupa metod agregacji ryzyka opiera się na uwzględnieniu struktury zależności pomiędzy agregowanymi rodzajami ryzyka, czyli na modelowaniu pełnego rozkładu wielowymiarowego wektora $(X_1, \ldots, X_d)$. Wymaga to zatem określenia nie tylko rozkładów brzegowych poszczególnych składników, lecz także sposobu ich powiązania w ramach wspólnego rozkładu.

Jednym z klasycznych i szeroko stosowanych podejść jest metoda wariancji–kowariancji, oparta na założeniu istnienia liniowej zależności pomiędzy poszczególnymi rodzajami ryzyka. W tym ujęciu struktura zależności modelowana jest za pomocą macierzy korelacji, która opisuje siłę i kierunek związku między każdą parą zmiennych losowych. Dla dwóch rodzajów ryzyka X_i, X_j współczynnik korelacji $\rho(X_i, X_j)$ przyjmuje wartości z przedziału $-1 \leq \rho(X_i, X_j) \leq 1$, określając tym samym stopień współzależności pomiędzy nimi. Macierz korelacji w sposób jednoznaczny determinuje strukturę zależności jedynie w przypadku rozkładów należących do klasy rozkładów eliptycznych, takich jak rozkład normalny czy t-Studenta. W pozostałych przypadkach korelacja liniowa nie wystarcza do pełnego opisu współzależności, co istotnie ogranicza zakres zastosowania tej metody w praktyce modelowania ryzyka.

Metoda wariancji–kowariancji jest często stosowana ze względu na prostotę implementacji i łatwość interpretacji wyników. Jej procedura obejmuje dwa zasadnicze etapy. W pierwszym z nich wyznaczane są indywidualne kapitały $\kappa(X_i)$ dla poszczególnych składników portfela. W drugim etapie następuje ich agregacja z uwzględnieniem współczynników korelacji liniowej między składnikami ryzyka. Formuła agregacji przyjmuje postać:

$$\kappa(Y) = \sqrt{\begin{pmatrix} \kappa(X_1) \\ \kappa(X_2) \\ \vdots \\ \kappa(X_d) \end{pmatrix}^T \begin{pmatrix} 1 & \rho_{12} & \ldots & \rho_{1d} \\ \rho_{21} & 1 & \vdots & \rho_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{d1} & \rho_{d2} & \ldots & 1 \end{pmatrix} \begin{pmatrix} \kappa(X_1) \\ \kappa(X_2) \\ \vdots \\ \kappa(X_d) \end{pmatrix}}. \tag{2.11}$$

Z metodologicznego punktu widzenia metoda wariancji–kowariancji jest poprawna jedynie w przypadku, gdy analizowane zmienne $X_1, \ldots, X_d$ mają wspólny rozkład eliptyczny, a

struktura ich zależności może zostać w pełni opisana przez macierz współczynników korelacji liniowej. W rzeczywistych danych ubezpieczeniowych i finansowych założenia te są jednak rzadko spełnione, gdyż zależności między rodzajami ryzyka są często nieliniowe, asymetryczne lub związane z zachowaniami ekstremalnymi w ogonach rozkładów. Ogranicza to wiarygodność wyników uzyskiwanych przy użyciu tej metody.

W konsekwencji zastosowanie metody wariancji–kowariancji w sytuacjach, w których zależności między składnikami portfela mają charakter silnych współzależności ogonowych, może prowadzić do niedoszacowania całkowitego ryzyka oraz wymaganego kapitału. Oznacza to, że model nie doszacowuje prawdopodobieństwa wspólnego wystąpienia ekstremalnych strat w wielu kategoriach ryzyka jednocześnie, co w praktyce skutkuje przeszacowaniem efektu dywersyfikacji. Z tego względu w analizach, w których istotne są zależności nieliniowe lub asymetryczne, zaleca się stosowanie bardziej elastycznych narzędzi, takich jak kopule[23], umożliwiających oddzielne modelowanie rozkładów brzegowych oraz struktury zależności między zmiennymi losowymi. Zastosowanie kopuli pozwala na elastyczne odwzorowanie zarówno liniowych, jak i nieliniowych relacji pomiędzy różnymi rodzajami ryzyka, niezależnie od ich rozkładów brzegowych, co czyni je szczególnie użytecznym narzędziem w procesie agregacji ryzyka.

Procedura modelowania oparta na kopulach obejmuje następujące etapy:

- identyfikację rozkładów brzegowych zmiennych losowych $(X_1, \ldots, X_d)$,
- estymację struktury zależności pomiędzy nimi poprzez dopasowanie odpowiedniej kopuli $C(u_1, \ldots, u_d)$,
- wyznaczenie rozkładu zmiennej zagregowanej $Y = \psi(X_1, \ldots, X_d)$, gdzie ψ oznacza funkcję agregującą,
- obliczenie wymaganego kapitału $\kappa(Y)$ w oparciu o przyjętą miarę ryzyka $\varrho(\cdot)$.

2.2.3 Efekt dywersyfikacji

Agregacja licznych rodzajów ryzyka, na które narażony jest zakład ubezpieczeń, przy czym nie wszystkie z nich realizują się równocześnie, powoduje, że całkowity kapitał niezbędny do zabezpieczenia się przed tymi ryzykami jest z reguły mniejszy niż suma kapitałów wymaganych dla poszczególnych rodzajów ryzyka rozpatrywanych oddzielnie. Zjawisko to, określane mianem efektu dywersyfikacji, stanowi jeden z kluczowych elementów procesu zarządzania ryzykiem w zakładach ubezpieczeń. Efekt dywersyfikacji pozostaje ściśle powiązany z modelowaniem zależności pomiędzy ryzykami (Wanat, 2014), gdyż to właśnie struktura współzależności decyduje o stopniu redukcji łącznego wymogu kapitałowego. Znaczenie prawidłowego modelowania zależności szczególnie uwidoczniło się w kontekście kryzysu finansowego w latach 2008–2009, kiedy to okazało się, że błędne założenia dotyczące współzależności między aktywami a ryzykami doprowadziły do znacznego niedoszacowania całkowitej ekspozycji na ryzyko. W literaturze oraz praktyce pojawiały się komentarze wskazujące, że przyczyną problemów było m.in. niewłaściwe stosowanie kopuli Gaussa, która nie odzwierciedla zależności w ogonach rozkładu oraz ogólne niedoszacowanie korelacji. W skrajnych opiniach podkreślano

[23] Istotę kopuli oraz ich zastosowanie w modelowaniu wielowymiarowych rozkładów ryzyka w kontekście wyznaczania wymogów kapitałowych omówiono w dalszej części niniejszego rozdziału, tj. w podrozdziałach 2.3–2.6.

wręcz, że „wszystko, co opiera się na korelacji, jest szarlatanerią", co uwidacznia wagę i złożoność zagadnienia modelowania współzależności w ocenie ryzyka. W obszarze ubezpieczeń dywersyfikacja pełni dwie zasadnicze funkcje. Po pierwsze, umożliwia zakładowi realizację jego podstawowej roli, jaką jest tworzenie wspólnoty ryzyka, czyli łączenie jednostek narażonych na podobne zagrożenia w portfele pozwalające na redukcję finansowych skutków indywidualnych strat. To właśnie dzięki efektowi dywersyfikacji możliwe jest oferowanie ochrony ubezpieczeniowej po ekonomicznie uzasadnionej cenie. Po drugie, dywersyfikacja stanowi narzędzie zarządzania portfelem ryzyk, umożliwiające kontrolę i optymalizację ekspozycji na różne rodzaje ryzyka w ramach działalności ubezpieczeniowej, inwestycyjnej i operacyjnej. Można zatem stwierdzić, że prawidłowa ocena i wykorzystanie efektu dywersyfikacji stanowi fundament efektywnego funkcjonowania zakładów ubezpieczeń oraz kluczowy element procesu wyznaczania wymogów kapitałowych w ramach współczesnych systemów regulacyjnych.

Do oceny efektu dywersyfikacji stosuje się tzw. współczynnik dywersyfikacji ED, który stanowi miarę ilościową korzyści wynikających z łączenia różnych rodzajów ryzyka w ramach jednego portfela ryzyk. W niniejszym opracowaniu został on zdefiniowany w następujący sposób:

$$ED = 1 - \frac{\kappa(Y)}{\sum_{i=1}^{d} \kappa(X_i)}, \tag{2.12}$$

gdzie $\kappa(Y)$ oznacza wymóg kapitałowy dla całkowitego portfela, a $\kappa(X_i)$ – wymóg kapitałowy odpowiadający poszczególnym rodzajom ryzyka X_i.

Wartość współczynnika ED informuje o relacji pomiędzy wymogiem kapitałowym dla portfela zagregowanego a sumą kapitałów dla jego składowych. Wartość $ED = 0$ wskazuje na brak efektu dywersyfikacji i odpowiada sytuacji, w której ryzyka są w pełni dodatnio skorelowane (komonotoniczne), co powoduje, że całkowity wymóg kapitałowy jest równy sumie kapitałów cząstkowych. Wartości dodatnie ($ED > 0$) oznaczają występowanie pozytywnego efektu dywersyfikacji, czyli redukcję wymaganego kapitału dzięki częściowej kompensacji pomiędzy składnikami portfela. Im wyższa wartość współczynnika, tym większe korzyści z dywersyfikacji. Z kolei, wartości ujemne ($ED < 0$) wskazują na efekt odwrotny. Wzrost całkowitego wymogu kapitałowego po agregacji ryzyk może mieć miejsce w przypadku silnych, dodatnich lub nieliniowych zależności pomiędzy zmiennymi, zwłaszcza w obszarach ogonowych rozkładów.

Właściwe oszacowanie współczynnika ED wymaga dokładnego odwzorowania struktury zależności pomiędzy agregowanymi rodzajami ryzyka. To właśnie sposób modelowania tych zależności decyduje o poprawności wnioskowania na temat korzyści z dywersyfikacji oraz o realności oceny kapitału ekonomicznego zakładu ubezpieczeń. W dalszej części pracy szczegółowo omówiono dwie odmienne koncepcje modelowania struktury zależności: podejście parametryczne, oparte na analitycznych postaciach funkcji kopuli, oraz podejście nieparametryczne, wykorzystujące sztuczne sieci neuronowe do estymacji złożonych, nieliniowych relacji między zmiennymi ryzyka. Porównanie wyników uzyskanych w obu podejściach stanowi kluczowy element analizy prowadzonej w niniejszym rozdziale.

2.3 Kopula jako narzędzie modelowania zależności

W celu bardziej elastycznego i realistycznego odwzorowania rzeczywistej struktury zależności pomiędzy agregowanymi rodzajami ryzyka stosuje się kopule. Formalnie d-wymiarową

kopulę definiuje się następująco:

$$C(u_1, \ldots, u_d) = \mathbb{P}(U_1 \leq u_1, \ldots, U_d \leq u_d), \tag{2.13}$$

gdzie $U_1, \ldots, U_d$ są zmiennymi losowymi o rozkładzie jednostajnym na przedziale $[0, 1]$. Kopula jest zatem dystrybuantą rozkładu wielowymiarowego o brzegach jednostajnych.

Kopula umożliwia połączenie dowolnych jednowymiarowych rozkładów brzegowych w rozkład wielowymiarowy. Zależność tę formalizuje twierdzenie Sklara.

Twierdzenie Sklara. Niech F oznacza d-wymiarową dystrybuantę łączną zmiennej losowej $(X_1, \ldots, X_d)$ z rozkładami brzegowymi $F_1, \ldots, F_d$.
Wówczas istnieje taka kopula $C : [0, 1]^d \to [0, 1]$, że dla każdego $(x_1, \ldots, x_d) \in \mathbb{R}^d$:

$$F(x_1, \ldots, x_d) = C\big(F_1(x_1), \ldots, F_d(x_d)\big). \tag{2.14}$$

Odwrotnie, jeśli C jest kopulą, a $F_1, \ldots, F_d$ są dystrybuantami jednowymiarowymi, to funkcja F zdefiniowana powyżej jest dystrybuantą łączną o brzegach $F_1, \ldots, F_d$. Jeżeli $F_1, \ldots, F_d$ są ciągłe, to kopula C jest wyznaczona jednoznacznie.

Jeżeli $F_1, \ldots, F_d$ są funkcjami ściśle rosnącymi (różnowartościowymi), to istnieją ich funkcje odwrotne F_i^{-1} $(i = 1, \ldots, d)$ i zachodzi

$$C(u_1, \ldots, u_d) = F\big(F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)\big), \qquad (u_1, \ldots, u_d) \in [0, 1]^d. \tag{2.15}$$

Zatem, przy znanych rozkładach brzegowych F_i, cała informacja o zależnościach pomiędzy $X_1, \ldots, X_d$ zawarta jest w kopuli C.

Zakładając ciągłość brzegów, gęstość wspólna f może być zapisana jako

$$f(x_1, \ldots, x_d) = c(u_1, \ldots, u_d) \prod_{i=1}^{d} f_i(x_i), \qquad u_i = F_i(x_i), \tag{2.16}$$

gdzie

$$c(u_1, \ldots, u_d) = \frac{\partial^d C(u_1, \ldots, u_d)}{\partial u_1 \cdots \partial u_d} \tag{2.17}$$

jest gęstością kopuli. Każda kopula $C \in \Xi_d$ (gdzie Ξ_d oznacza zbiór wszystkich d-wymiarowych kopul) spełnia tzw. ograniczenia Frécheta–Hoeffdinga:

$$\max\left\{\textstyle\sum_{i=1}^{d} u_i - d + 1,\ 0\right\} \ \leq\ C(u_1, \ldots, u_d) \ \leq\ \min\{u_1, \ldots, u_d\}, \qquad (u_1, \ldots, u_d) \in [0, 1]^d. \tag{2.18}$$

Fakt mówiący o tym, że rzeczywista struktura zależności może zostać opisana kopulą, skłonił zarówno środowisko akademickie, jak i praktyków do konstrukcji różnych kopul oddających przeróżne struktury zależności. Do podstawowych kopul wykorzystywanych w sektorze ubezpieczeniowym należą kopule eliptyczne i kopule Archimedesa.

Kopule eliptyczne powiązane są z rozkładami eliptycznymi; najczęściej używa się kopuli Gaussa i kopuli t-Studenta. Kopula Gaussa ma postać

$$C_R^{\mathrm{Ga}}(u_1, \ldots, u_d) = \Phi_R\big(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)\big), \tag{2.19}$$

gdzie Φ_R oznacza dystrybuantę wielowymiarowego rozkładu normalnego standardowego z macierzą korelacji R, a Φ - dystrybuantę jednowymiarowego rozkładu normalnego standardowego. Kopula t-Studenta definiowana jest przez

$$C^{\mathrm{St}}_{R,\nu}(u_1, \ldots, u_d) = t_{R,\nu}\big(t_\nu^{-1}(u_1), \ldots, t_\nu^{-1}(u_d)\big), \tag{2.20}$$

gdzie $t_{R,\nu}$ jest dystrybuantą wielowymiarowego rozkładu t-Studenta o ν stopniach swobody i macierzy korelacji R, zaś t_ν jest dystrybuantą jednowymiarową. Parametr ν umożliwia modelowanie zależności w ogonach rozkładu.

Ograniczeniem kopul eliptycznych jest jednak fakt, że umożliwiają one modelowanie jedynie zależności symetrycznych, co ogranicza ich zastosowanie w sytuacjach, gdy obserwowane są asymetrie między ryzykami. Problem ten przezwycięża klasa kopul Archimedesa, która pozwala na odwzorowanie zarówno asymetrycznych zależności, jak i zależności w ogonach rozkładu (Hofert, 2008). Kopule z tej klasy definiuje się za pomocą tzw. generatora kopuli φ (McNeil & Nešlehová, 2009):

$$C(u_1, \ldots, u_d) = \phi^{-1}\big(\varphi(u_1) + \cdots + \varphi(u_d)\big), \tag{2.21}$$

gdzie $\varphi : [0, 1] \to [0, \infty]$ jest funkcją ściśle malejącą i wypukłą, spełniającą $\varphi(1) = 0$, $\varphi(0) = \infty$, a φ^{-1} oznacza funkcję odwrotną (w ujęciu uogólnionym w razie potrzeby).

W literaturze przedmiotu wskazuje się cztery najczęściej stosowane kopule Archimedesa: Gumbela, Joe, Claytona oraz Franka (F. A.-C. Genest C., 2007). Kopule Gumbela (Gumbel, 1960) oraz Joe (Joe, 1996) należą do kopul asymetrycznych, charakteryzujących się większym prawdopodobieństwem skupionym w górnym ogonie rozkładu, natomiast kopula Claytona (Clayton, 1978) koncentruje większe prawdopodobieństwo w dolnym ogonie. Kopula Franka (Frank, 1979), mająca charakter symetryczny, zwykle wykazuje mniejszą gęstość w ogonach w porównaniu z kopulami Gumbela, Joe czy Claytona; jednak pozwala na modelowanie umiarkowanych zależności.

Szczególne znaczenie mają dwuparametryczne kopule Archimedesa, które umożliwiają jednoczesne modelowanie zależności w obu ogonach rozkładu. Przykładowo, kopula Claytona opisuje zależności w lewym (dolnym) ogonie, natomiast kopula Joe — w prawym (górnym) ogonie. Połączenie tych dwóch funkcji prowadzi do powstania kopuli Joe-Claytona, umożliwiającej opisanie struktury zależności zarówno w dolnym, jak i w górnym ogonie. W literaturze przedstawiono Ponadto, inne połączenia jednoparametrycznych kopul Archimedesa, takie jak kopule Claytona-Gumbela, Joe-Gumbela, Joe-Claytona oraz Joe-Franka. Należy zauważyć, że kopula Claytona nie nadaje się do modelowania zjawisk charakteryzujących się zależnością w górnym ogonie, natomiast znajduje zastosowanie w przypadku zależności w dolnym ogonie. W celu rozszerzenia zakresu możliwych do odwzorowania struktur zależności w ramach danej rodziny kopul definiuje się tzw. kopule obrócone (rotated copulas), które pozwalają na modelowanie zależności w przeciwnych ogonach rozkładu (Brechmann, 2013).

Kopule stanowią efektywne narzędzie umożliwiające konstrukcję rozkładów wielowymiarowych oraz generowanie obserwacji z tych rozkładów. Znajduje to szczególne zastosowanie w analizie portfeli zawierających wiele rodzajów ryzyka. Konstrukcja ta ma kluczowe znaczenie w zarządzaniu ryzykiem, zwłaszcza w kontekście oceny zależności między ryzykami, które często nie mogą być ujęte w ramach klasycznych modeli liniowej korelacji.

W celu skonstruowania rozkładu wielowymiarowego należy:

1. dopasować rozkłady brzegowe $F_1, \ldots, F_d$,

2. oszacować kopulę C, opisującą strukturę zależności między składnikami,

3. zdefiniować wspólną dystrybuantę $F(x_1, \ldots, x_d) = C(F_1(x_1), \ldots, F_d(x_d))$.

Symulacja obserwacji z tak określonego rozkładu przebiega według następującej procedury:

1. generuje się realizację $(U_1, \ldots, U_d) \sim C$, gdzie C to wcześniej dopasowana kopula,

2. oblicza się $X_i = F_i^{-1}(U_i)$ dla $i = 1, \ldots, d$,

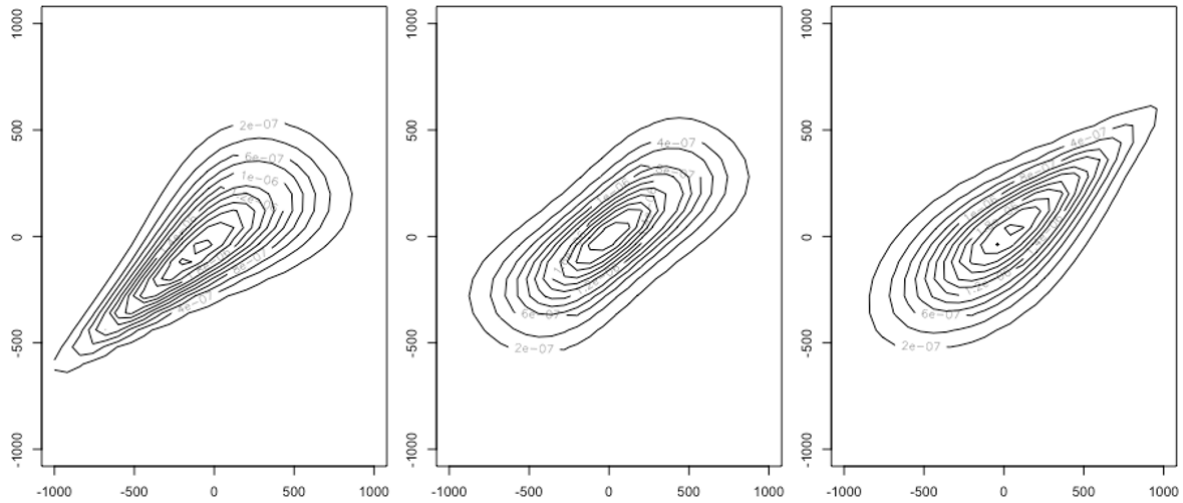
3. wektor $(X_1, \ldots, X_d)$ ma rozkład wspólny F z brzegami F_i oraz zależnością opisaną przez C.

Zastosowanie kopul w zarządzaniu ryzykiem finansowym umożliwia realistyczne modelowanie współzależności pomiędzy czynnikami ryzyka, w tym zwłaszcza zależności w ogonach rozkładu. Cecha ta pozwala na uwzględnienie podwyższonego prawdopodobieństwa jednoczesnego wystąpienia ekstremalnych strat w wielu pozycjach portfela, co nie jest możliwe w przypadku modeli opartych wyłącznie na korelacjach liniowych, takich jak modele oparte na wspólnym rozkładzie normalnym.

W praktyce finansowej kopule są szeroko wykorzystywane m.in. do modelowania ryzyka skrajnego oraz analizy portfelowej, gdzie kluczowe znaczenie ma dokładne ujęcie współzależności pomiędzy składnikami portfela. W kontekście regulacyjnym (zarówno w ramach wymogów kapitałowych zgodnych z Bazyleą II/III, jak i Solvency II) odpowiednie odwzorowanie struktury zależności stanowi podstawę do wyznaczenia miar ryzyka, takich jak wartość zagrożona (VaR) oraz oczekiwany niedobór (ES).

Zastosowanie kopuli umożliwia również ocenę efektu dywersyfikacji. W zależności od przyjętego typu kopuli, wspólna dystrybuanta strat portfela może wykazywać różny poziom zależności skrajnych, co wpływa na oszacowanie redukcji ryzyka wynikającej z rozproszenia ekspozycji. Badania empiryczne wskazują, że wybór konkretnej kopuli może istotnie wpływać na wyniki agregacji ryzyka. Dla identycznych rozkładów brzegowych różne kopule prowadzą do odmiennych szacunków wymaganego kapitału oraz korzyści z dywersyfikacji. Niedoszacowanie zależności, np. poprzez zastosowanie kopuli niedostatecznie uwzględniającej współwystępowanie skrajnych zdarzeń, prowadzi do zbyt optymistycznej oceny ryzyka oraz zaniżenia wymogów kapitałowych. Z kolei, zastosowanie kopuli modelującej silną zależność ogonową (np. kopuli t-Studenta lub niektórych kopuli Archimedesowych) skutkuje niższym efektem dywersyfikacyjnym i wyższymi wymogami kapitałowymi, co stanowi bardziej konserwatywne i bezpieczne podejście z punktu widzenia instytucji finansowej.

Aby zilustrować znaczenie struktury zależności w agregacji ryzyka, rozważono przykład z rozkładami brzegowymi $X_1 \sim \mathcal{N}(0, 392^2)$ oraz $X_2 \sim \mathcal{N}(0, 248^2)$. Strukturę zależności opisano kolejno za pomocą kopul Archimedesa: Claytona, Franka i Gumbela, tak aby współczynnik korelacji liniowej między (X_1, X_2) wynosił 0.25. Na Rysunku 2.1 przedstawiono wykresy konturowe uzyskanych rozkładów dwuwymiarowych (od lewej: Claytona, Franka, Gumbela).



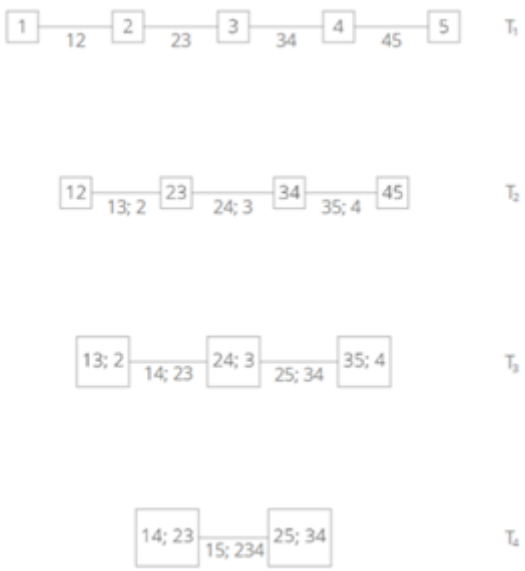
Rysunek 2.1: Wykresy konturowe dla kopul Claytona, Franka i Gumbela.
Źródło: Opracowanie własne.

Kopule eliptyczne są elastyczne w modelowaniu zależności dla większej liczby zmiennych, lecz z natury są symetryczne. Kopule Archimedesa dobrze odwzorowują asymetrie w przypadku dwóch zmiennych, jednak przy większym wymiarze pojedynczy parametr zależności narzuca tę samą siłę związku dla wszystkich par, co ogranicza elastyczność. Z tego względu w zastosowaniach ubezpieczeniowych stosuje się także kaskady kopuli (vine copulas), które pozwalają realistyczniej opisać złożone struktury współzależności w wyższych wymiarach.

2.4 Kaskady kopul

W przypadku modelowania zależności wielowymiarowych przydatnym narzędziem są kaskady kopuli. W odróżnieniu od wielowymiarowych kopuli Archimedesa, gdzie parametr zależności dla każdej pary zmiennych losowych jest taki sam, kaskady kopuli uwzględniają zależności między każdą parą zmiennych losowych. Wprowadzone przez (Bedford & Cooke, 2002) kaskady kopul to elastyczna klasa modeli, w których wielowymiarowy rozkład przedstawiany jest za pomocą kopul dwuwymiarowych. Pozwala to na elastyczność w modelowaniu zależności między większą liczbą zmiennych losowych, uwzględniając przy tym asymetrię rozkładów oraz zależności w ogonach. W odróżnieniu od wielowymiarowych kopul Archimedesa, gdzie parametr zależności dla każdej pary zmiennych losowych jest taki sam, kaskady kopul uwzględniają zależności między każdą parą zmiennych losowych.

Stosowane są dwa modele kaskad kopul (Czado, 2019): D-kaskada kopul (D-vine) oraz C-kaskada kopul (C-vine). Sposób ich konstrukcji najczęściej przedstawiany jest w sposób graficzny. Przykładową pięciowymiarową dekompozycję D-kaskady kopul przedstawia Rysunek 2.2.



Rysunek 2.2: Pięciowymiarowa dekompozycja D-vine kopul.
Źródło: Opracowanie własne.

W wierzchołkach drzewa T_1 umieszcza się indeksy zmiennych losowych, które łączy się krawędziami, a następnie wyznacza się ich etykiety. Krawędź łączącą wierzchołek 1 z wierzchołkiem 2 etykietuje się jako 12, wierzchołek 2 z wierzchołkiem 2 jako 23 itd. Do wierzchołków drzewa T_2 trafiają etykiety z drzewa T_1, następnie wierzchołki łączy się krawędziami i nadaje im się etykiety. Krawędź łączącą wierzchołek 12 z wierzchołkiem 23 etykietuje się jako 13;2, gdzie po średniku umieszcza się indeks zmiennej, występującej zarówno w wierzchołku 12, jak i w wierzchołku 23, a przed średnikiem umieszcza się indeksy pozostałych zmiennych w kolejności rosnącej. Zmienne przed średnikiem to zmienne warunkowane, a po średniku to zmienne warunkujące. Kolejne krawędzie drzewa T_2 wyznacza się analogicznie. W drzewie T_3 w wierzchołkach umieszcza się etykiety z drzewa T_2. Wierzchołki drzewa łączy się krawędziami, które etykietuje się, powtarzając rozumowanie z drzewa T_2. Dla przykładu etykietą dla krawędzi łączącej wierzchołek 13;2 z wierzchołkiem 24;3 jest 14;23. Po średniku umieszcza się zmienne występujące zarówno w wierzchołku 13;2, jak i w wierzchołku 24;3 w kolejności rosnącej, a przed średnikiem pozostałe zmienne w kolejności rosnącej. W ostatnim drzewie T_4 do wierzchołków grafu trafiają etykiety z drzewa T_4. Etykietę dla krawędzi w tym drzewie wybiera się analogicznie do etykiet w drzewach wcześniejszych.

Zatem, wyznaczając D-kaskadę kopul dla d zmiennych losowych, otrzymuje się d-1 drzew. Pierwsze drzewo składa się z d wierzchołków i $d-1$ krawędzi. Okazuje się, że wybór porządku w drzewie T_1 nie jest jednoznaczny i determinuje porządek w kolejnych drzewach. Dla d wierzchołków w drzewie T_1, zauważając, że krawędzie w drzewie T_1 są nieskierowane, istnieje $\frac{d!}{2}$ możliwość wyboru dekompozycji D-kaskady kopul.

Mając ustalony porządek w drzewie T_1, przystępuje się do wyboru kopuli i jej parametrów dla każdej krawędzi drzewa T_1. Jeśli drzewo T_1 składa się z d wierzchołków, to zostaje dopasowanych $d-1$ kopul. Każda kopula jest indeksowana etykietą krawędzi, dla której została dobrana. Następnie, z wybranych kopul losuje się obserwacje. Otrzymuje się $d-1$ macierzy obserwacji o wymiarach $n \times 2$, (n oznacza liczebność próby), które trafiają do wierzchołków drzewa T_2. Dalej, dla każdej krawędzi drzewa T_2 dobierana jest rodzina kopul warunkowych oraz ich

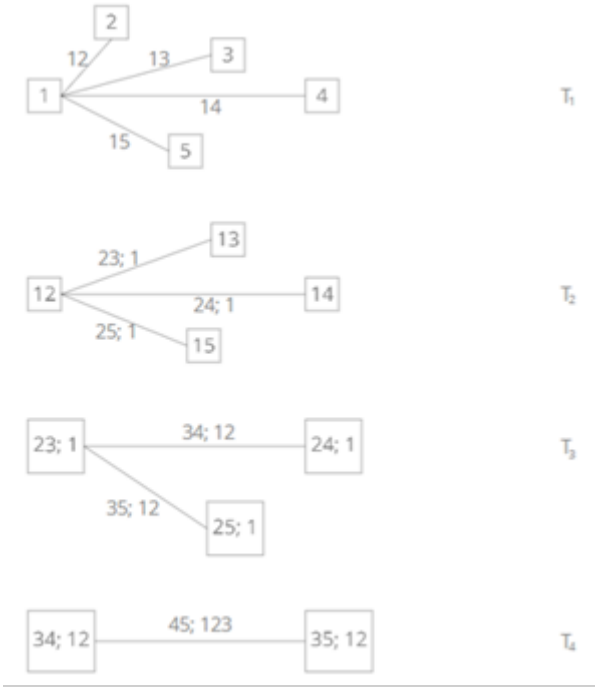
parametry. Wybranych kopul jest $d-2$, a każda kopula indeksowana jest etykietą krawędzi, do której została dobrana. Procedurę tę powtarza się aż do momentu wyznaczenia dwuwymiarowej kopuli warunkowej dla dwóch ostatnich wierzchołków drzewa.

Zgodnie z (Czado, 2019), gęstość d-wymiarowego rozkładu dla konstrukcji D-vine może być zapisana jako:

$$f(\boldsymbol{x}) = \prod_{k=1}^{d} f_k(x_k) \times \prod_{\ell=1}^{d-1} \prod_{i=1}^{d-\ell} c_{i,i+\ell|i+1:i+\ell-1} \left(u_{i|i+1:i+\ell-1},\, u_{i+\ell|i+1:i+\ell-1}\right), \tag{2.22}$$

gdzie $c_{i,j|D}$ oznacza warunkową gęstość kopuli dla zmiennych X_i i X_j, warunkując względem zbioru D, a $u_{k|D}$ oznacza warunkową dystrybuantę zmiennej X_k daną warunkiem D.

Inaczej wygląda sytuacja w przypadku, gdy zadaniem jest wyznaczenie dekompozycji C-kaskady kopul. Poniższa grafika (rys. 2.3) ilustruje przykład dekompozycji dla pięciu zmiennych losowych.



Rysunek 2.3: Pięciowymiarowa dekompozycja C-vine kopul.
Źródło: Opracowanie własne.

W odróżnieniu od dekompozycji D, w dekompozycji C-kaskady kopul wybierana jest zmienna, która jest korzeniem drzewa. Po ustaleniu korzenia dla drzewa T_1 łączy się go z pozostałymi wierzchołkami drzewa T_1, a następnie krawędzie drzewa T_1 etykietuje się w taki sam sposób, jak krawędzie drzewa T_1 w dekompozycji D-vine. Krawędzie drzewa T_1 stają się wierzchołkami drzewa T_2, a z nich wybiera się korzeń dla drzewa T_2. Korzeń łączy się krawędziami z pozostałymi wierzchołkami drzewa T_2 i nadaje im etykiety. Etykietą dla krawędzi łączącej korzeń 12 i wierzchołek 13 to 23;1, korzeń 12 i wierzchołek 14 to 24;1, korzeń 12 i wierzchołek 15 to 25;1. Zatem zmienną warunkującą w drzewie T_2 jest zmienna, która była korzeniem drzewa T_1. Kontynuując to rozumowanie, dla drzewa T_3 zmiennymi warunkującymi są zmienne, które znajdują się w korzeniu drzewa T_2. Porządek w drzewie T_1 nie jest jednoznaczny

i nie determinuje porządku w kolejnych drzewach. Występuje $\frac{d!}{2}$ możliwych dekompozycji C-vine (Aas, Czado, Frigessi, & Bakken, 2009).

Gdy został już wybrany korzeń drzewa T_1, dla każdej krawędzi drzewa T_1 dobierana jest kopula, którą indeksuje się etykietą danej krawędzi drzewa T_1. Po dobraniu wszystkich kopuli w drzewie T_1 losowane są z nich obserwacje. Otrzymuje się $d-1$ macierzy obserwacji o wymiarach $n \times 2$ (n oznacza liczebność próby), z których najpierw wyznacza się korzeń drzewa T_2. Podobnie jak przy wyborze korzenia dla drzewa T_1, wyznacza się macierz wartości bezwzględnych współczynnika τ Kendalla dla $d-1$ zmiennych i wybiera się tę, dla której suma w kolumnie macierzy jest największa. Zmienna ta zostaje korzeniem drzewa T_2, a pozostałe zmienne umieszcza się na wierzchołkach i łączy z korzeniem krawędziami. Mając wyznaczone drzewo T_2, dla każdej krawędzi dobiera się kopule warunkowe indeksowane etykietami krawędzi drzewa T_2. Procedurę tę powtarza się, aż zostanie wyznaczona warunkowa kopula dwuwymiarowa dla dwóch ostatnich wierzchołków.
W przypadku dekompozycji C-vine funkcja gęstości przyjmuje postać: Dla konstrukcji $C-vine$ wzór na gęstość przyjmuje postać (Czado, 2019):

$$f(\boldsymbol{x}) = \prod_{k=1}^{d} f_k(x_k) \times \prod_{j=1}^{d-1} \prod_{i=j+1}^{d} c_{j,i|1:j-1}\left(u_{j|1:j-1},\, u_{i|1:j-1}\right), \tag{2.23}$$

gdzie interpretacja oznaczeń jest analogiczna jak w przypadku konstrukcji $D-vine$.

Struktura C-kaskady kopul stosowana jest, jeśli wśród zmiennych losowych, z których konstruowany jest rozkład wielowymiarowy, jedna ze zmiennych losowych wykazuje dużą zależność z pozostałymi zmiennymi losowymi. Natomiast, jeśli taka zmienna losowa nie istnieje, wykorzystuje się dekompozycję D-kaskady kopul.

Szczegółowe omówienie konstrukcji $D-vine$ oraz $C-vine$, wraz z ich teoretycznymi podstawami i implementacją w języku R, przedstawiono w monografii (Czado, 2019). Obie konstrukcje stanowią elastyczne podejście do modelowania złożonych zależności pomiędzy wieloma zmiennymi losowymi, umożliwiając dekompozycję wielowymiarowych kopuli na szereg kopuli dwuwymiarowych z odpowiednimi warunkowaniami.

W kontekście zarządzania ryzykiem w zakładach ubezpieczeń konstrukcje kaskady kopul są szczególnie przydatne, ponieważ pozwalają na realistyczne odwzorowanie struktury współzależności między różnymi źródłami ryzyka, także w ogonach rozkładów. Takie podejście jest zgodne z wymogami regulacyjnymi Solvency II, które nakładają na ubezpieczycieli obowiązek wyznaczania kapitałowych wymogów wypłacalności w oparciu o pełne rozkłady ryzyka, przy uwzględnieniu ich wzajemnych powiązań. Dzięki kaskadom kopul możliwe jest lepsze uchwycenie efektu dywersyfikacji i bardziej precyzyjne oszacowanie miar takich jak wartość zagrożona czy oczekiwany niedobór, co wpływa na bezpieczeństwo finansowe instytucji i zgodność z regulacjami.

2.5 Kopule Nieparametryczne

Ograniczeniem kaskad kopuli jest fakt, że dekomponują one rozkład wielowymiarowy na rozkład opisywany za pomocą kopul dwuwymiarowych, które dla zadanych parametrów opisują z góry określony kształt zależności. Dodatkowo, w praktyce często przyjmuje się założenie upraszczające, tzn. że dwuwymiarowe kopule zależą od zmiennych warunkujących tylko pośrednio, poprzez warunkowe rozkłady brzegowe. Oznacza to, że kopula $c_{13|2}\left(U_1, U_3\right)$ pod

warunkiem $U_2 = u_2$ jest taka sama dla wszystkich wartości $u_2 \in (0, 1)$. Podejście oparte na kopulach nieparametrycznych polega na braku założenia apriorii dotyczącego kształtu zależności. W badaniach empirycznych wykorzystano dwie kopule nieparametryczne: kopulę szachownicy kopulę odcinkowo liniową.

Kopula szachownicy

Idea kopuli szachownicy (ang. checkerboard copula) polega na wykorzystaniu równomiernego podziału przestrzeni jednostkowej w celu konstrukcji funkcji schodkowej, przybliżającej rzeczywistą strukturę zależności.

Niech dana będzie macierz obserwacji o wymiarach $n \times d$, a $m \in \mathbb{N}$ oznacza liczbę podziałów przedziału jednostkowego w każdej współrzędnej (m jest dzielnikiem n). Wówczas przestrzeń $[0, 1]^d$ dzieli się na m^d równych kostek, indeksowanych przez zbiór

$$\mathcal{I} = \left\{ \mathbf{i} = (i_1, \ldots, i_d) \in \{1, \ldots, m\}^d \right\}. \tag{2.24}$$

Każdej kostce odpowiada podobszar przestrzeni:

$$I_{\mathbf{i},m} = \prod_{j=1}^{d} \left(\frac{i_j - 1}{m}, \frac{i_j}{m} \right]. \tag{2.25}$$

Niech μ oznacza miarę (rzeczywistą lub empiryczną) odpowiadającą rozkładowi kopuli, a λ miarę Lebesgue'a w $\mathbb{R}^d$. Kopulę szachownicy definiuje się jako:

$$C(\mathbf{x}) = \sum_{\mathbf{i} \in \mathcal{I}} m^d \mu(I_{\mathbf{i},m}) \cdot \lambda \left([0, \mathbf{x}] \cap I_{\mathbf{i},m} \right), \quad \mathbf{x} \in [0, 1]^d, \tag{2.26}$$

gdzie $[0, \mathbf{x}] = \prod_{j=1}^{d} [0, x_j]$.

Z punktu widzenia generowania obserwacji, konstrukcja (2.26) oznacza wybór kostki $I_{\mathbf{i},m}$ z prawdopodobieństwem $\mu(I_{\mathbf{i},m})$, a następnie losowanie punktu z rozkładu jednostajnego na tej kostce.

Alternatywnie, można zastosować monotoniczną wersję kopuli szachownicy:

$$C(\mathbf{x}) = \sum_{\mathbf{i} \in \mathcal{I}} m^d \mu(I_{\mathbf{i},m}) \cdot \min \left\{ x_1 - \tfrac{i_1 - 1}{m}, \ldots, x_d - \tfrac{i_d - 1}{m}, \tfrac{1}{m} \right\}. \tag{2.27}$$

W zastosowaniach empirycznych zastosowano wersję, w której miara μ jest zastępowana miarą empiryczną $\hat{\mu}$, określoną jako:

$$\hat{\mu}([0, \mathbf{x}]) = \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_{\{R_1^{(k)} \le x_1,\, R_2^{(k)} \le x_2,\, \ldots,\, R_d^{(k)} \le x_d\}}, \tag{2.28}$$

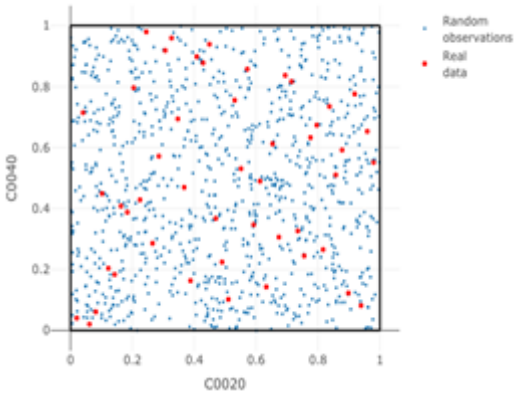
gdzie $R_j^{(k)}$ oznacza rangę k-tej obserwacji w j-tej zmiennej marginesowej (przeskalowanej do przedziału $[0, 1]$).

Dla $d = 2$, wartość m oznacza liczbę podziałów przestrzeni $[0, 1]^2$ na m^2 kwadratów o bokach długości $\frac{1}{m}$, które stanowią dyskretną siatkę wykorzystywaną do aproksymacji ciągłej kopuli.

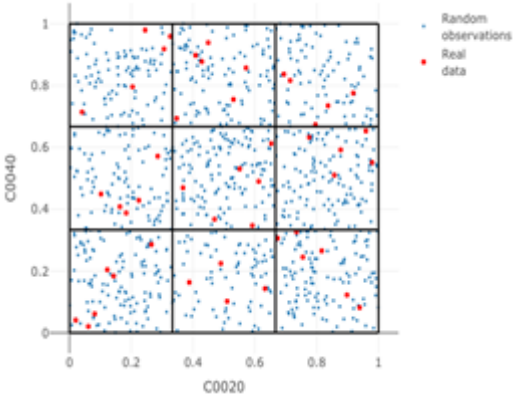
Przykład

Wyznaczając struktury zależności między dwoma segmentami ubezpieczyciela, dla każdego

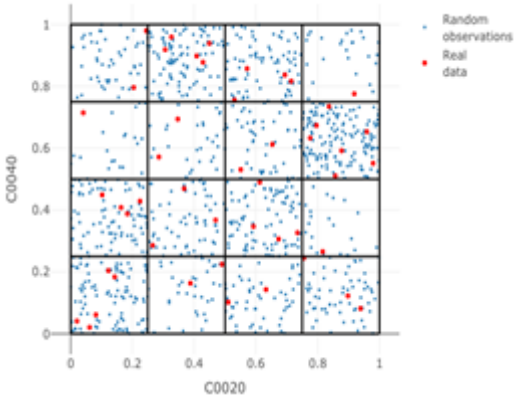
segmentu wzięto 48 obserwacji. Na poniższych Rysunkach 2.4 - 2.9 przedstawiono dopasowanie kopuli szachownicy do danych rzeczywistych. Czerwone punkty oznaczają obserwacje rzeczywiste, natomiast czarne to realizacje z kopuli.



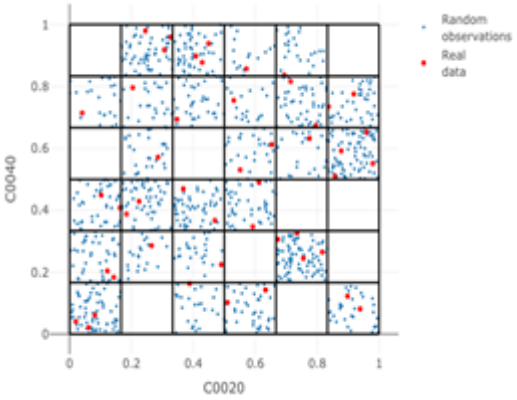
Rysunek 2.4: Kopula szachownicy ($m = 1$).
Źródło: Opracowanie własne.



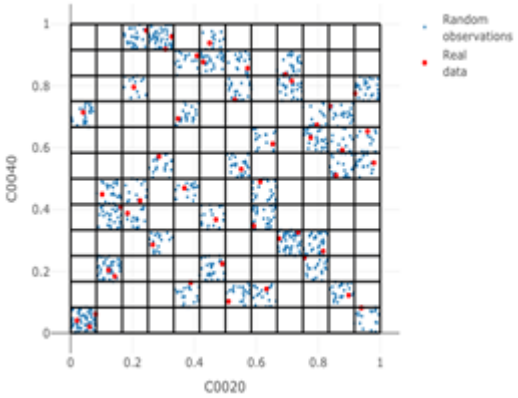
Rysunek 2.5: Kopula szachownicy ($m = 3$).
Źródło: Opracowanie własne.



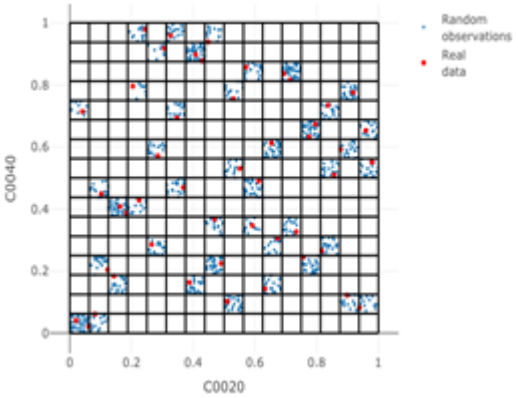
Rysunek 2.6: Kopula szachownicy ($m = 4$).
Źródło: Opracowanie własne.



Rysunek 2.7: Kopula szachownicy ($m = 6$).
Źródło: Opracowanie własne.



Rysunek 2.8: Kopula szachownicy ($m = 12$).
Źródło: Opracowanie własne.



Rysunek 2.9: Kopula szachownicy ($m = 16$).
Źródło: Opracowanie własne.

Kopula odcinkowo liniowa

Kopula odcinkowo-liniowa (ang. piecewise linear copula) stanowi elastyczną metodę przybliżania struktury zależności pomiędzy zmiennymi losowymi. Konstrukcja ta opiera się na podziale przestrzeni jednostkowego hiperześcianu $[0, 1]^d$ na skończoną liczbę nieskładających się na siebie regionów (płatów), z których każdy może mieć różny kształt i rozmiar.

Niech $\mathcal{L} = \{\ell_1, \ldots, \ell_k\}$ oznacza taki podział, a $p_1, \ldots, p_k$ będą wagami spełniającymi $p_i > 0$ oraz $\sum_{i=1}^{k} p_i = 1$. Wówczas kopula odcinkowo-liniowa określona na $\mathcal{L}$ z wagami $p = (p_1, \ldots, p_k)$ ma postać:

$$C_{p,\mathcal{L}}(\mathbf{x}) = \sum_{i=1}^{k} p_i \, \lambda_{\ell_i}(\mathbf{x}), \tag{2.29}$$

gdzie $\mathbf{x} = (x_1, \ldots, x_d) \in [0, 1]^d$, a

$$\lambda_{\ell_i}(\mathbf{x}) = \frac{1}{\lambda(\ell_i)} \lambda\big([0, \mathbf{x}] \cap \ell_i\big), \tag{2.30}$$

zaś λ oznacza standardową d-wymiarową miarę Lebesgue'a. Wartość $\lambda_{\ell_i}(\mathbf{x})$ opisuje proporcję objętości części płata ℓ_i zawartej w prostopadłościanie $[0, \mathbf{x}]$.

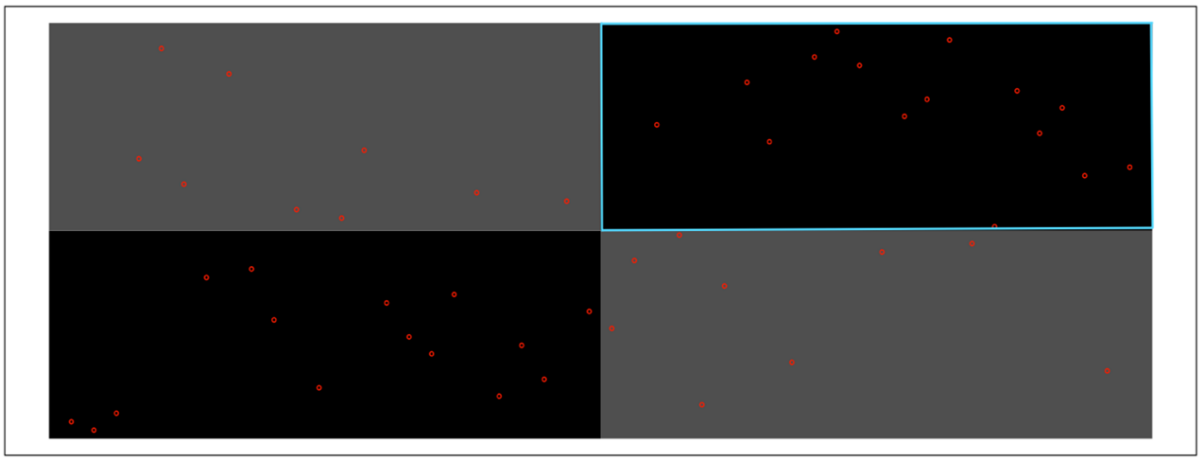
Konstrukcja ta umożliwia dopasowanie do bardziej złożonych struktur zależności w danych empirycznych. W przeciwieństwie do kopuli szachownicy, gdzie przestrzeń $[0, 1]^d$ dzielona jest na regularną siatkę kostek tej samej wielkości, tutaj dopuszcza się zmienność rozmiarów i kształtów płatów, co przekłada się na większą elastyczność odwzorowania lokalnych struktur zależności.

Szczególnym przypadkiem kopuli odcinkowo-liniowej jest sytuacja, w której $p_i = \lambda(\ell_i)$ dla każdego $i = 1, \ldots, k$. Wówczas funkcja $C_{p,\mathcal{L}}$ spełnia warunki kopuli, czyli stanowi dystrybuantę rozkładu wielowymiarowego z rozkładami brzegowymi jednostajnymi na $[0, 1]$.

Kopule odcinkowo-liniowe stosowane są w szczególności do estymacji kopuli empirycznej, modelowania złożonych zależności w danych finansowych, a także w analizie ryzyka i modelowaniu zależności w ogonach rozkładów (Durante & Sempi, 2015; C. Genest, Rémillard, & Beaudoin, 2009).

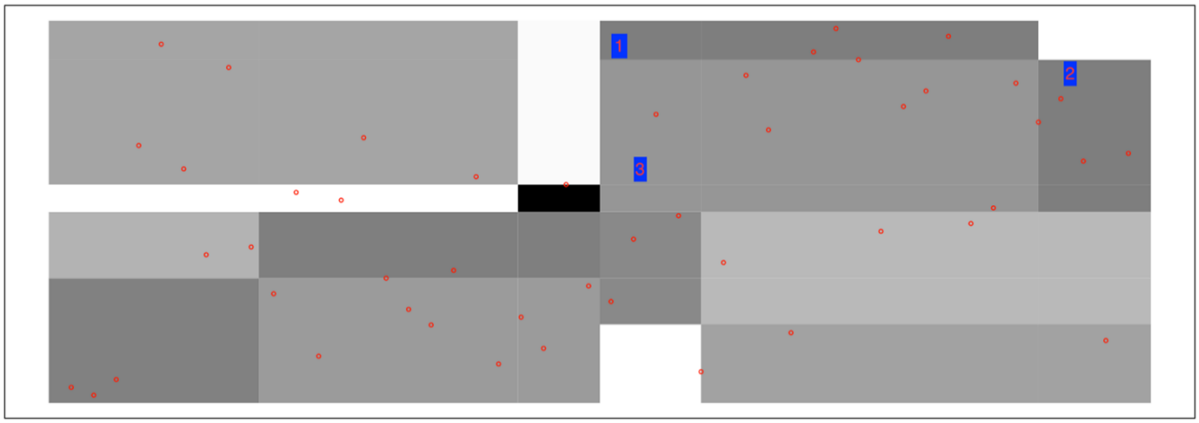
Przykład

Odcinkowo liniowa kopula w odróżnieniu od kopuli szachownicy, w której kostki dzielące przestrzeń miały równe długości boków, odcinkowo liniowa kopula dzieli przestrzeń na tak zwane liście o różnych wymiarach. Określa się minimalną liczbę obserwacji w liściach, poniżej której nie następuje dalszy podział. W analizowanym przykładzie ustalono, że algorytm nie powinien dzielić liścia, jeśli zawiera on 3 obserwacje. Podczas pierwszej iteracji (rys. 2.10) algorytm podzielił przestrzeń na cztery identyczne liście:



Rysunek 2.10: Odcinkowo liniowa kopula – pierwsza iteracja.
Źródło: Opracowanie własne.

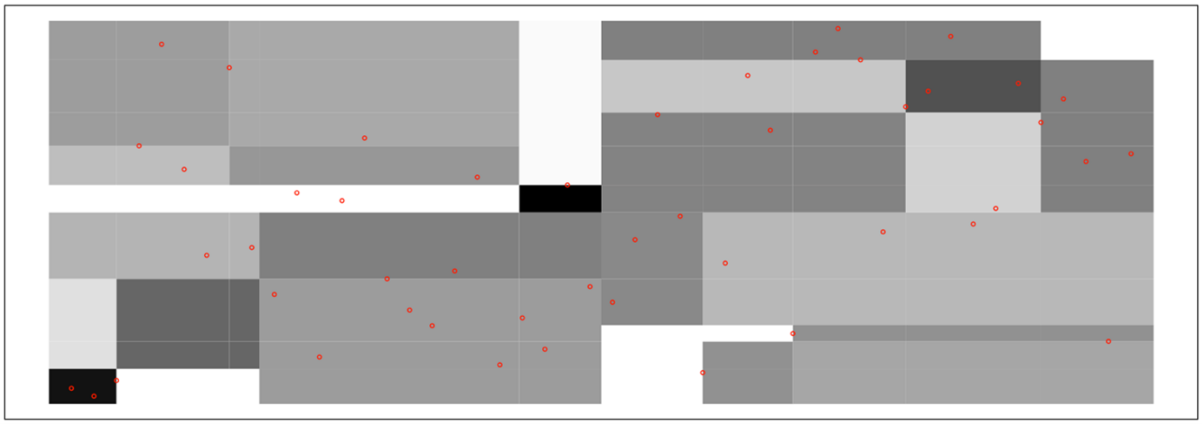
Liść zaznaczony na niebiesko został podzielony na trzy kolejne liście. Na Rysunku 2.11 pokazano, że w liściach 1 i 2 znajdują się trzy obserwacje, więc te dwa liście nie zostaną podzielone w następnej iteracji, natomiast trzeci liść zostanie podzielony.



Rysunek 2.11: Odcinkowo liniowa kopula – druga iteracja.
Źródło: Opracowanie własne.

Na Rysunku 2.11 pokazano, że trzeci liść został podzielony na cztery liście, a w każdym z nich znajdują się dwie lub trzy obserwacje. Dlatego w tej iteracji (rys. 2.12) algorytm kończy

dzielenie tego liścia.



Rysunek 2.12: Odcinkowo liniowa kopula – trzecia iteracja.
Źródło: Opracowanie własne.

Jak zaprezentowano w powyższych przykładach, kopule nieparametryczne dopasowują się do danych rzeczywistych, dzieląc przestrzeń na kostki czy liście. Jest to ich główną przewagą nad kopulami z rodziny Archimedesa czy kaskadami kopul, gdzie struktura zależności jest określona dla danego parametru.

2.6 Wpływ informacji o strukturze zależności na ocenę zagregowanego ryzyka i efekt dywersyfikacji

W niniejszym podrozdziale analizowany jest wpływ zakresu informacji o strukturze zależności pomiędzy składowymi ryzyka na kształt rozkładu zmiennej $Y = X_1 + \cdots + X_d$, modelującej zagregowane ryzyko portfela. Na podstawie rozkładu zmiennej Y wyznaczane są miary ryzyka, które Z kolei, służą do określania wymogów kapitałowych, a także do oceny efektu dywersyfikacji.

Wyznaczenie rozkładu zagregowanego ryzyka Y stanowi zadanie złożone, zależne zarówno od wiedzy o rozkładach brzegowych zmiennych $X_1, \ldots, X_d$, jak i od wiedzy na temat ich współzależności. W praktyce zazwyczaj zakłada się, że funkcja agregująca ψ przyjmuje postać sumy składowych, czyli $\psi(X_1, \ldots, X_d) = X_1 + \cdots + X_d$.

Można wyróżnić dwa główne przypadki, zależnie od dostępnych informacji. W pierwszym przypadku dostępna jest pełna wiedza o rozkładzie wielowymiarowym wektora $(X_1, \ldots, X_d)$, co pozwala na bezpośrednie wyznaczenie rozkładu zmiennej Y. W drugim przypadku wiedza jest niepełna — najczęściej dostępne są jedynie dobrze oszacowane rozkłady brzegowe, natomiast struktura zależności pozostaje częściowo lub całkowicie nieznana.

W sytuacji ograniczonej wiedzy bezpośrednie wyznaczenie rozkładu zmiennej Y nie jest możliwe. Wówczas stosuje się podejścia polegające na wyznaczeniu ograniczeń dolnych i górnych dla dystrybuanty tej zmiennej. W kontekście regulacyjnym, np. w ramach systemu Solvency II, ograniczenia te mają istotne znaczenie praktyczne, gdyż umożliwiają wyznaczenie przedziałów niepewności dla miar ryzyka, w szczególności dla wartości zagrożonej $\mathrm{VaR}_\alpha(Y)$.

W dalszej części rozdziału omówione zostaną różne poziomy wiedzy o strukturze zależności - od jej pełnej znajomości po całkowity brak wiedzy - oraz ich wpływ na rozkład agregatu ryzyka i ocenę efektu dywersyfikacji.

2.6.1 Pełna wiedza o rozkładzie wielowymiarowym

W tym przypadku możliwe jest dokładne wyznaczenie dystrybuanty zmiennej Y oraz miar ryzyka, takich jak $\mathrm{VaR}_\alpha(Y)$ czy $\mathrm{ES}_\alpha(Y)$, co pozwala na precyzyjne określenie wymogów kapitałowych oraz efektu dywersyfikacji.

Niech $F_{\boldsymbol{X}}$ oznacza dystrybuantę łączną wektora $(X_1, \ldots, X_d)$. Wówczas dystrybuanta zmiennej Y wyraża się wzorem:

$$F_Y(y) = \mathbb{P}(Y \leq y) = \int_{I(y)} dF_{\boldsymbol{X}}(x_1, \ldots, x_d), \tag{2.31}$$

gdzie

$$I(y) = \left\{ (x_1, \ldots, x_d) \in \mathbb{R}^d : \sum_{i=1}^{d} x_i \leq y \right\}. \tag{2.32}$$

Wyznaczenie powyższej całki może być numerycznie wymagające (Arbenz, Embrechts, & Puccetti, 2011a, 2011b). W celu uproszczenia analizy często rozważa się dwie ekstremalne struktury zależności: niezależność i komonotoniczność.

Przypadek niezależności.
Niech $(X_1^\perp, \ldots, X_d^\perp)$ oznacza wektor losowy o tych samych rozkładach brzegowych co $(X_1, \ldots, X_d)$, ale z założeniem niezależności między zmiennymi. Wówczas dystrybuanta wspólna ma postać:

$$F_{X^\perp}(x_1, \ldots, x_d) = \prod_{i=1}^{d} F_{X_i}(x_i). \tag{2.33}$$

Rozkład zmiennej $Y^\perp = X_1^\perp + \cdots + X_d^\perp$ może zostać wyznaczony za pomocą splotu rozkładów brzegowych, np. z wykorzystaniem transformaty Fouriera (Duhamel & Vetterli, 1990) lub metod rekurencyjnych (np. formuły Panjera (Panjer, 1981)). Takie podejście może prowadzić do niedoszacowania całkowitego ryzyka. Jeśli Y i $Y^\perp$ należą do tej samej klasy Frécheta, wówczas mają tę samą wartość oczekiwaną: $\mathbb{E}[Y] = \mathbb{E}[Y^\perp]$.

Przypadek komonotoniczności
Niech $(X_1^c, \ldots, X_d^c)$ oznacza wektor losowy o komonotonicznej strukturze zależności i tych samych brzegach co $(X_1, \ldots, X_d)$, tj.:

$$F_{X^c}(x_1, \ldots, x_d) = \min\{F_{X_1}(x_1), \ldots, F_{X_d}(x_d)\}, \tag{2.34}$$

co jest równoważne zapisowi:

$$(X_1^c, \ldots, X_d^c) \overset{d}{=} \left(F_{X_1}^{-1}(U), \ldots, F_{X_d}^{-1}(U) \right), \quad U \sim \mathcal{U}(0,1). \tag{2.35}$$

Rozkład $Y^c = X_1^c + \cdots + X_d^c$ może być wyznaczony przez sumę kwantyli, co zazwyczaj prowadzi do przeszacowania ryzyka. Zgodnie z wynikami Dhaene, Denuit, Goovaerts, Kaas, and Vyncke (2002), dla każdej wypukłej funkcji v, dla której istnieje wartość oczekiwana, zachodzi:

$$\mathbb{E}[v(Y)] \leq \mathbb{E}[v(Y^c)]. \tag{2.36}$$

Oznacza to, że Y^c dominuje Y w sensie porządku wypukłego (convex order), co przekłada się m.in. na relację $\mathrm{VaR}(Y^c) \geq \mathrm{VaR}(Y)$.

Rozkłady z zależnością liniową
W praktyce często przyjmuje się założenie, że zmienne mają strukturę liniowej zależności, np. rozkład wielowymiarowy normalny. Wówczas:

$$Y \sim \mathcal{N}\left(\mu = \sum_{i=1}^{d} \mu_i,\ \sigma^2 = \sum_{i=1}^{d}\sum_{j=1}^{d} \sigma_{ij}\right), \tag{2.37}$$

co umożliwia łatwe obliczenie wartości $\mathrm{VaR}_\alpha(Y)$ i $\mathrm{ES}_\alpha(Y)$.

Podejście numeryczne - algorytm AEP
W przypadku braku analitycznego opisu dystrybuanty łącznej $F_{\boldsymbol{X}}$ można zastosować algorytm AEP (ang. Additive Extension Procedure) zaproponowany przez Arbenz et al. (2011a, 2011b), który umożliwia numeryczne przybliżenie rozkładu zmiennej Y bez konieczności stosowania symulacji Monte Carlo lub całkowania gęstości. Metoda ta opiera się na geometrycznej aproksymacji zbioru poziomicowego funkcji $\sum x_i \leq y$, co umożliwia szybką i dokładną estymację rozkładu agregatu w przypadku znanych rozkładów brzegowych i kopuli.

2.6.2 Znane rozkłady brzegowe i częściowa informacja o strukturze zależności

Miary zależności odgrywają kluczową rolę w analizie zależności pomiędzy zmiennymi losowymi, umożliwiając ilościowe ujęcie stopnia zależności. Każda z miar, takich jak τ-Kendalla, ρ-Spearmana, współczynnik Blomqvista czy gamma Giniego, charakteryzuje się inną wrażliwością na zmiany w rozkładzie obserwacji i różnym sposobem ujęcia zależności pomiędzy zmiennymi. Współczynnik τ-Kendalla opiera się na liczbie par obserwacji zgodnych i niezgodnych, odzwierciedlając ogólną monotoniczność zależności pomiędzy zmiennymi. Współczynnik ρ-Spearmana mierzy siłę zależności w oparciu o korelację rang, co czyni go mniej wrażliwym na wartości skrajne, lecz bardziej podatnym na nieliniowości w środkowych zakresach rozkładu. Miara Blomqvista koncentruje się na zależności pomiędzy medianami rozkładów brzegowych, przez co jest szczególnie użyteczna w analizie symetrycznych struktur zależności. Natomiast, gamma Giniego, będąca uogólnieniem współczynnika Giniego na przypadek dwuwymiarowy, w większym stopniu akcentuje zależność w centralnych częściach rozkładu, dzięki czemu stanowi alternatywę dla klasycznych współczynników korelacji w sytuacjach, gdy zależność pomiędzy zmiennymi jest nieliniowa lub asymetryczna.

Znajomość wartości wybranej miary pozwala zdefiniować zbiór wszystkich kopul dwuwymiarowych, dla których dana miara przyjmuje ustaloną wartość, co umożliwia konstrukcję odpowiadających im górnych i dolnych ograniczeń Frécheta–Hoeffdinga. W dalszej części podrozdziału przedstawiono wybrane miary zależności oraz wynikające z nich analityczne postaci ograniczeń dla kopul dwuwymiarowych.

Rozważając przypadek dwuwymiarowy, niech $\kappa : \Xi_2 \to [-1, 1]$ będzie daną miarą zgodności, a $k \in [-1, 1]$. Wówczas zbiór

$$\mathcal{F}_\kappa := \{C \in \Xi_2 \mid \kappa(C) = k\} \tag{2.38}$$

gdzie Ξ_2 jest klasą wszystkich kopul dwuwymiarowych, a $\mathcal{F}$ odpowiada takim kopulom dwuwymiarowym, dla których wartość miary κ wynosi k. Więcej na temat miar zgodności

można znaleźć w (Liebscher, 2014; Fuchs & Schmidt, 2014). W literaturze przedstawiono wyznaczanie ograniczeń kopuli (dystrybuanty) przy założeniu znajomości wybranych miar zależności, takich jak: współczynnik τ Kendalla (τ), współczynnik ρ Pearsona, współczynnik Blomqvista (β), współczynnik rang Spearmana (ϕ) oraz gamma Giniego (γ). Dla miar zależności τ, ρ oraz β najlepsze dolne punktowe ograniczenia zostały podane w pracach (Liebscher, 2014; Fuchs & Schmidt, 2014). Odpowiednie wzory mają postać:

- w przypadku, gdy znany jest współczynnik korelacji τ-Kendalla

$$\underline{F}_{\tau}(u_1, u_2) = \max\left\{0; u_1 + u_2 - 1; \frac{1}{2}\left(u_1 + u_2 - \sqrt{(u_1 - u_2)^2 + 1 - \tau}\right)\right\}, \tag{2.39}$$

- w przypadku, gdy znany jest współczynnik korelacji ρ-Spearmana

$$\underline{F}_{\rho}(u_1, u_2) = \max\left\{0; u_1 + u_2 - 1; \frac{1}{2}\left(u_1 + u_2 - \varphi(u_1, u_2, \rho_S)\right)\right\}, \tag{2.40}$$

  gdzie:

$$\varphi(u_1, u_2, \rho_s) = \begin{pmatrix} (9(1-\rho_s) + 3\sqrt{9(1-\rho_s)^2 - 3(u_1 - u_2)^6)^3}^{\frac{1}{3}} \\ +(9(1-\rho_s) - 3\sqrt{9(1-\rho_s)^2 - 3(u_1 - u_2)^6)^3}^{\frac{1}{3}} \end{pmatrix} \tag{2.41}$$

- w przypadku, gdy znany jest współczynnik korelacji Blomqvista

$$\underline{F}_{\beta}(u_1, u_2) = \max\left\{0; u_1 + u_2 - 1; \frac{\beta + 1}{4} - \left(\frac{1}{2} - u_1\right)_+ - \left(\frac{1}{2} - u_2\right)_+\right\}, \tag{2.42}$$

  gdzie $x_+ = max(0; x)$.

W pracy (Kokol Bukovšek, Mazo, Mesiar, & Omladič, 2021) opracowano odpowiednio górne i dolne granice dla rang Spearmana oraz gamma Giniego. Postać analityczna tych ograniczeń jest bardziej złożona niż dla miar zależności prezentowanych powyżej. W dalszej części podrozdziału przedstawiono ideę oraz uzyskane ograniczenia.

W celu wyznaczenia lokalnych ograniczeń dla zbioru kopul o ustalonej wartości współczynnika Spearmana rozważa się rodzinę wszystkich kopul dwuwymiarowych, dla których ta miara przyjmuje określoną wartość. Niech

$$\mathcal{F}_{\phi} := \{\, C \in \Xi_2 \mid \phi(C) = \phi \,\}, \tag{2.43}$$

gdzie Ξ_2 oznacza klasę wszystkich kopul dwuwymiarowych, a F_ϕ oznacza zbiór wszystkich kopul dwuwymiarowych, dla których współczynnik Spearmana przyjmuje ustaloną wartość $\phi \in \left[-\frac{1}{2}, 1\right]$.

Wśród wszystkich kopul Ξ_2 wyróżniają się tak zwane granice Frécheta-Hoeffdinga, które określają skrajne przypadki możliwej zależności pomiędzy zmiennymi losowymi. Granice te definiuje się następująco:

$$W(a, b) = \max\{0,\ a + b - 1\}, \tag{2.44}$$

$$M(a, b) = \min\{a,\ b\}, \tag{2.45}$$

gdzie W stanowi dolną granicę Frécheta-Hoeffdinga, odpowiadającą przypadkowi ujemnej zależności, natomiast M jest górną granicą, odpowiadającą przypadkowi dodatniej zależności. Dla wszystkich par punktów $(a, b) \in \mathbb{I}^2$ otrzymuje się dolne ograniczenie w postaci:

$$\underline{F}_\phi(a,b) = \begin{cases} \frac{1}{2}\left(a+b-\sqrt{\frac{2}{3}(1-\phi)+(b-a)^2}\right), & \text{jeśli } b \notin \{0,1\} \text{ i } \frac{1-\phi}{6b} \le a \le 1-\frac{1-\phi}{6(1-b)}, \\ W(a,b), & \text{w przeciwnym razie.} \end{cases} \tag{2.46}$$

Funkcja $\underline{F}_\phi$ jest kopulą dla wszystkich $\phi \in \left[-\frac{1}{2}, 1\right]$, przy czym $\underline{F}_{-1/2} = W$ oraz $\underline{F}_1 = M$, a ponadto, spełnia warunek symetrii $\underline{F}_\phi(a,b) = \underline{F}_\phi(b,a)$.

Aby znaleźć $\bar{F}_\phi$, obliczenia są znacznie trudniejsze, a ich ideę przybliżono poniżej. Kwadrat jednostkowy $\mathbb{I}^2$ podzielono na siedem obszarów w zależności od ϕ:

$$\Delta_\phi^1 = \left\{ \begin{array}{l} a \le \frac{1}{2}\left(1-\frac{\sqrt{3}}{3}\sqrt{1+2\phi}\right), \\ b \ge \frac{1}{2}\left(1+\frac{\sqrt{3}}{3}\sqrt{1+2\phi}\right),\ b \le a+\frac{1}{2}\left(1-\frac{\sqrt{3}}{3}\sqrt{1+2\phi}\right) \end{array} \right\}. \tag{2.47}$$

$$\Delta_\phi^2 = \left\{ \begin{array}{l} b \le \frac{1}{2}\left(1+\frac{\sqrt{3}}{3}\sqrt{1+2\phi}\right), \\ \frac{1}{3}\left(2b-1+\sqrt{(2b-1)^2+1+2\phi}\right) \le a \le \frac{1}{3}\left(b+1-\sqrt{(2b-1)^2+1+2\phi}\right) \end{array} \right\}. \tag{2.48}$$

$$\Delta_\phi^3 = \left\{ \begin{array}{l} a \ge \frac{1}{2}\left(1-\frac{\sqrt{3}}{3}\sqrt{1+2\phi}\right), \\ \frac{1}{3}\left(a+1+\sqrt{(2a-1)^2+1+2\phi}\right) \le b \le \frac{1}{3}\left(2a+2-\sqrt{(2a-1)^2+1+2\phi}\right) \end{array} \right\}. \tag{2.49}$$

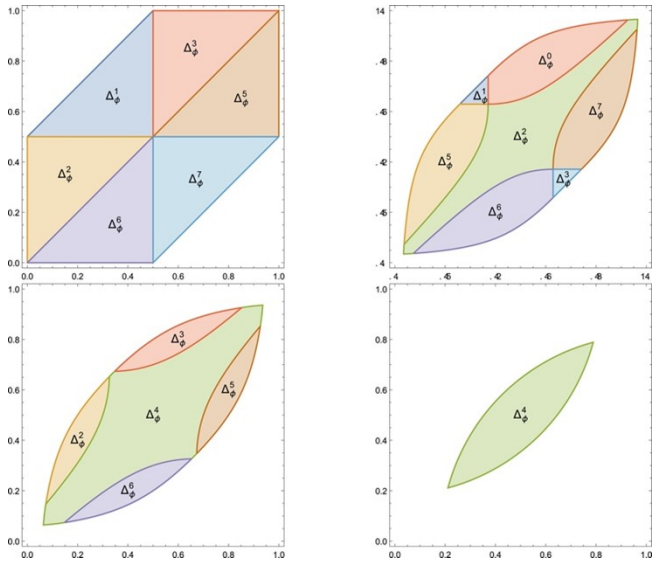
$$\Delta_\phi^4 = \left\{ \begin{array}{l} (a,b) \in \mathbb{I}^2, \\ \frac{1}{3}\left(b+1-\sqrt{(2b-1)^2+1+2\phi}\right) \le a \le \frac{1}{3}\left(b+1+\sqrt{(2b-1)^2+1+2\phi}\right), \\ \frac{1}{3}\left(a+1-\sqrt{(2a-1)^2+1+2\phi}\right) \le b \le \frac{1}{3}\left(a+1+\sqrt{(2a-1)^2+1+2\phi}\right) \end{array} \right\}. \tag{2.50}$$

$$\Delta_\phi^5 = \left\{ \begin{array}{l} b \ge \frac{1}{2}\left(1-\frac{\sqrt{3}}{3}\sqrt{1+2\phi}\right), \\ \frac{1}{3}\left(b+1+\sqrt{(2b-1)^2+1+2\phi}\right) \le a \le \frac{1}{3}\left(2b+2-\sqrt{(2b-1)^2+1+2\phi}\right) \end{array} \right\}. \tag{2.51}$$

$$\Delta_\phi^6 = \left\{ \begin{array}{l} a \le \frac{1}{2}\left(1+\frac{\sqrt{3}}{3}\sqrt{1+2\phi}\right), \\ \frac{1}{3}\left(2a-1+\sqrt{(2a-1)^2+1+2\phi}\right) \le b \le \frac{1}{3}\left(a+1-\sqrt{(2a-1)^2+1+2\phi}\right) \end{array} \right\}. \tag{2.52}$$

$$\Delta_\phi^7 = \left\{ \begin{array}{l} b \le \frac{1}{2}\left(1-\frac{\sqrt{3}}{3}\sqrt{1+2\phi}\right), \\ a \ge \frac{1}{2}\left(1+\frac{\sqrt{3}}{3}\sqrt{1+2\phi}\right), \\ b \ge a-\frac{1}{2}\left(1-\frac{\sqrt{3}}{3}\sqrt{1+2\phi}\right) \end{array} \right\}. \tag{2.53}$$

natomiast geometryczną interpretację na Rysunku 2.13, na których przedstawiono otrzymane obszary $\Delta_\phi^1, \ldots, \Delta_\phi^7$ dla wartości $\phi = -\frac{1}{2}, -\frac{2}{5}, -\frac{32}{100}, 0$.



Rysunek 2.13: Wykres obszarów dla współczynnika rang Spearmana.
Źródło: Wykresy pochodzą z (Kokol Bukovšek et al., 2021).

Dla wartości parametru $\phi \in \left[-\frac{1}{2}, -\frac{1}{3}\right]$ wszystkie obszary $\Delta_\phi^1, \Delta_\phi^2, \Delta_\phi^3, \Delta_\phi^4, \Delta_\phi^5, \Delta_\phi^6$ oraz Δ_ϕ^7 są niepuste. Wraz ze wzrostem wartości ϕ niektóre z tych obszarów stopniowo ulegają redukcji, a część z nich staje się pusta. Dla $\phi = -\frac{1}{2}$ obszar Δ_ϕ^4 redukuje się do przekątnej jednostkowego kwadratu. W przedziale $\phi \in \left(-\frac{1}{3}, -\frac{1}{5}\right]$ obszary Δ_ϕ^1 oraz Δ_ϕ^7 są puste, natomiast dla $\phi \in \left(-\frac{1}{5}, \frac{1}{4}\right]$ obszar Δ_ϕ^4 pozostaje niepusty. Przy dalszym wzroście wartości współczynnika, tj. dla $\phi \in \left(\frac{1}{4}, 1\right]$, wszystkie obszary $\Delta_\phi^1, \Delta_\phi^2, \Delta_\phi^3, \Delta_\phi^4, \Delta_\phi^5, \Delta_\phi^6$ oraz Δ_ϕ^7 stają się puste, co odpowiada sytuacji, w której górne ograniczenie pokrywa się z kopulą M. Dla każdego punktu $(a, b) \in \mathbb{I}^2$ zdefiniowano supremum $\bar{F}_\phi(a, b)$ zbioru $\mathcal{F}_\phi$ dla wszystkich $\phi \in \left[-\frac{1}{2}, 1\right]$ w postaci:

$$\bar{F}_\phi(a,b) = \begin{cases} \frac{1}{2}\left(2b - 1 + \frac{\sqrt{3}}{3}\sqrt{1+2\phi}\right), & (a,b) \in \Delta_\phi^1, \\ \frac{1}{3}\left(2b - 1 + \sqrt{(1-2b)^2 + 1 + 2\phi}\right), & (a,b) \in \Delta_\phi^2, \\ \frac{1}{3}\left(a + 3b - 2 + \sqrt{(1-2a)^2 + 1 + 2\phi}\right), & (a,b) \in \Delta_\phi^3, \\ \frac{1}{2}\left(a + b - 1 + \sqrt{3(b-a)^2 + (1-2a)(1-2b) + \frac{2}{3}(1+2\phi)}\right), & (a,b) \in \Delta_\phi^4, \\ \frac{1}{3}\left(3a + b - 2 + \sqrt{(1-2b)^2 + 1 + 2\phi}\right), & (a,b) \in \Delta_\phi^5, \\ \frac{1}{3}\left(2a - 1 + \sqrt{(1-2a)^2 + 1 + 2\phi}\right), & (a,b) \in \Delta_\phi^6, \\ \frac{1}{2}\left(2a - 1 + \frac{\sqrt{3}}{3}\sqrt{1+2\phi}\right), & (a,b) \in \Delta_\phi^7, \\ M(a,b), & \text{w przeciwnym razie.} \end{cases} \tag{2.54}$$

Wartość $\bar{F}_{-\frac{1}{2}}$ stanowi kombinację funkcji M, wynikającą z granic Frécheta–Hoeffdinga, tj. $\bar{F}_{-\frac{1}{2}} = M\Big(2, \big\{\big[0, \frac{1}{2}\big], \big[\frac{1}{2}, 1\big]\big\}, (2,1), 1\Big)$, natomiast $\bar{F}_\phi = M$ dla $\phi \in \big[\frac{1}{4}, 1\big]$. Dla $\phi \in \Big(-\frac{1}{2}, \frac{1}{4}\Big)$ ograniczenie $\bar{F}_\phi$ jest właściwą quasi-kopulą. Ponadto, dla wszystkich $\phi \in \big[-\frac{1}{2}, 1\big]$ funkcja $\bar{F}_\phi$ spełnia własność symetryczności $\bar{F}_\phi(a,b) = \bar{F}_\phi(b,a)$.

Kolejną miarą, dla której wyznaczono lokalne ograniczenia, jest gamma Giniego. Analogicznie jak w przypadku współczynnika rang Spearmana, definiuje się zbiór

$$\mathcal{G}_\gamma := \{\, C \in \Xi_2 \mid \gamma(C) = \gamma \,\}, \tag{2.55}$$

dla każdej wartości $\gamma \in [-1,1]$. Zbiór $\mathcal{G}_\gamma$ obejmuje wszystkie kopule dwuwymiarowe, dla których współczynnik gamma Giniego przyjmuje ustaloną wartość w przedziale $\gamma \in [-1,1]$. Podobnie jak w przypadku wyznaczania ograniczeń dla współczynnika Spearmana, w celu konstrukcji granic definiuje się odpowiednie obszary, które dzielą kwadrat jednostkowy $\mathbb{I}^2$. Dla wszystkich punktów $(a,b) \in \mathbb{I}^2$ przedziały wyglądają następująco:

$$\Omega^1_\gamma = \left\{ \begin{array}{l} a \le \frac{1}{2}, \\ \frac{1}{2}\Big(1 + \frac{1+\gamma}{1-2a}\Big) \le b \le 1 - \frac{1+\gamma}{4a} \end{array} \right\}. \tag{2.56}$$

$$\Omega^2_\gamma = \left\{ \begin{array}{l} b \le \frac{1}{2}\Big(1 + \frac{1+\gamma}{1-2a}\Big), \\ (1+2a-2b)^2 + 4a(1-b) \ge 1+\gamma, \\ b \ge \frac{1}{3}\Big(a + 1 + \frac{1}{2}\sqrt{(2a-1)^2 + 3(1+\gamma)}\Big), \\ b \ge \frac{1}{4}\Big(6a - 1 + \frac{1+\gamma}{1-2a}\Big) \end{array} \right\}. \tag{2.57}$$

$$\Omega^3_\gamma = \left\{ \begin{array}{l} b \le \frac{1}{3}\Big(a + 1 + \frac{1}{2}\sqrt{(2a-1)^2 + 3(1+\gamma)}\Big), \\ b \le \frac{1}{8}\Big(3a + 6 - \frac{1+\gamma}{a}\Big), \\ a \le \frac{1}{11}\Big(3 + 5b - \sqrt{9(2b-1)^2 + 11(1+\gamma)}\Big) \end{array} \right\}. \tag{2.58}$$

$$\Omega^4_\gamma = \left\{ \begin{array}{l} a \ge \frac{1}{3}\Big(b + 1 - \frac{1}{2}\sqrt{(2b-1)^2 + 3(1+\gamma)}\Big), \\ a \ge \frac{1}{8}\Big(3b - 1 + \frac{1+\gamma}{1-b}\Big), \\ b \ge \frac{1}{11}\Big(3 + 5a + \sqrt{9(2a-1)^2 + 11(1+\gamma)}\Big) \end{array} \right\}. \tag{2.59}$$

$$\Omega^5_\gamma = \left\{ \begin{array}{l} \frac{1}{11}\Big(3 + 5b - \sqrt{9(2b-1)^2 + 11(1+\gamma)}\Big) \le a \le \frac{1}{11}\Big(3 + 5b + \sqrt{9(2b-1)^2 + 11(1+\gamma)}\Big), \\ \frac{1}{11}\Big(3 + 5a - \sqrt{9(2a-1)^2 + 11(1+\gamma)}\Big) \le b \le \frac{1}{11}\Big(3 + 5a + \sqrt{9(2a-1)^2 + 11(1+\gamma)}\Big), \\ b \le -2a + \sqrt{3a(a+2) - (1+\gamma)}, \qquad a \le -2b + \sqrt{3b(b+2) - (1+\gamma)} \end{array} \right\}. \tag{2.60}$$

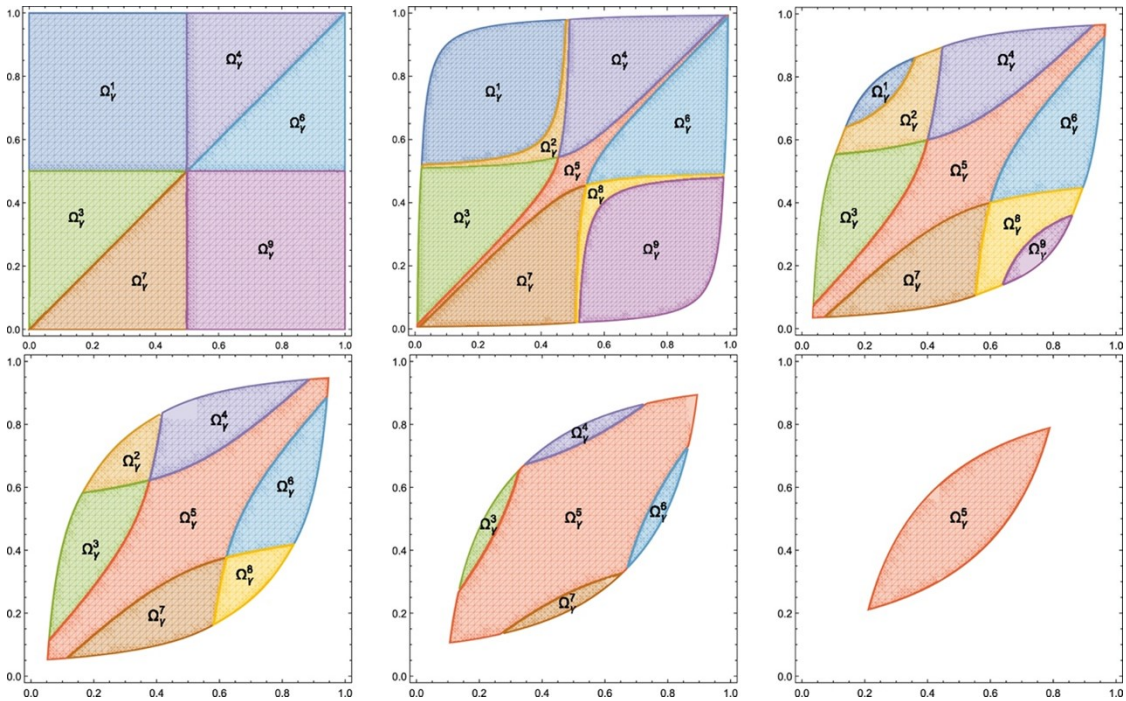
$$\Omega^6_\gamma = \left\{ \begin{array}{l} b \ge \frac{1}{3}\Big(a + 1 - \frac{1}{2}\sqrt{(2a-1)^2 + 3(1+\gamma)}\Big), \\ b \ge \frac{1}{8}\Big(3a - 1 + \frac{1+\gamma}{1-a}\Big), \\ a \ge \frac{1}{11}\Big(3 + 5b + \sqrt{9(2b-1)^2 + 11(1+\gamma)}\Big) \end{array} \right\}. \tag{2.61}$$

$$\Omega_\gamma^7 = \left\{ \begin{array}{l} a \le \frac{1}{3}\left(b + 1 + \frac{1}{2}\sqrt{(2b-1)^2 + 3(1+\gamma)}\right), \\ a \le \frac{1}{8}\left(3b + 6 - \frac{1+\gamma}{b}\right), \\ b \le \frac{1}{11}\left(3 + 5a - \sqrt{9(2a-1)^2 + 11(1+\gamma)}\right) \end{array} \right\}. \tag{2.62}$$

$$\Omega_\gamma^8 = \left\{ \begin{array}{l} a \le \frac{1}{2}\left(1 + \frac{1+\gamma}{1-2b}\right), \\ (1-2a+2b)^2 + 4b(1-a) \ge 1+\gamma, \\ a \ge \frac{1}{3}\left(b + 1 + \frac{1}{2}\sqrt{(2b-1)^2 + 3(1+\gamma)}\right), \\ a \ge \frac{1}{4}\left(6b - 1 + \frac{1+\gamma}{1-2b}\right) \end{array} \right\}. \tag{2.63}$$

$$\Omega_\gamma^9 = \left\{ \begin{array}{l} b \le \frac{1}{2}, \\ \frac{1}{2}\left(1 + \frac{1+\gamma}{1-2b}\right) \le a \le 1 - \frac{1+\gamma}{4b} \end{array} \right\}. \tag{2.64}$$

Na poniższym rysunku (rys. 2.14) przedstawiono graficzną interpretację otrzymanych obszarów $\Omega_\gamma^1, \ldots, \Omega_\gamma^9$ dla wybranych wartości parametru $\gamma = \{-1, -\frac{24}{25}, -\frac{4}{5}, -\frac{7}{10}, -\frac{43}{100}\}$ oraz 0.



Rysunek 2.14: Wykres obszarów dla współczynnika gamma Giniego.
Źródło: wykresy pochodzą z (Kokol Bukovšek et al., 2021)

Dla wszystkich wartości $\gamma \in \left[-1, -\frac{3}{4}\right]$ wszystkie obszary Ω_γ^i są niepuste. W przypadku $\gamma = -1$ obszar Ω_γ^5 redukuje się do przekątnej jednostkowego kwadratu, natomiast obszary Ω_γ^2 i Ω_γ^8 sprowadzają się do dwóch prostych prostopadłych $a = \frac{1}{2}$ oraz $b = \frac{1}{2}$.

Dla $\gamma \in \left(-\frac{3}{4}, -\frac{4}{9}\right]$ obszary Ω_γ^1 i Ω_γ^9 stają się puste, natomiast dla $\gamma \in \left(-\frac{4}{9}, -\frac{4}{13}\right]$ puste są obszary $\Omega_\gamma^1,\ \Omega_\gamma^2,\ \Omega_\gamma^8$ oraz Ω_γ^9.

W przedziale $\gamma \in \left(-\frac{4}{13}, \frac{1}{2}\right]$ niepusty pozostaje jedynie obszar Ω_γ^5, natomiast dla $\gamma \in \left(\frac{1}{2}, 1\right]$ wszystkie obszary Ω_γ^i są już puste.

Supremum $\bar{G}_\gamma(a,b)$ w każdym punkcie $(a,b) \in \mathbb{I}^2$ dla dowolnej wartości parametru $\gamma \in [-1,1]$ ma postać funkcji częściowej:

$$\bar{G}_\gamma(a,b) = \begin{cases} \frac{1}{2}\left(a+b-1+\sqrt{(a+b-1)^2+1+\gamma}\right), & (a,b) \in \Omega_\gamma^1, \\ \frac{1}{4}\left(a+3b-2+\sqrt{(a+b-1)^2+(1-2a)(1-2b)+2(1+\gamma)}\right), & (a,b) \in \Omega_\gamma^2, \\ \frac{1}{7}\left(2a+4b-3+\sqrt{(2a+4b-3)^2+7(1+\gamma)}\right), & (a,b) \in \Omega_\gamma^3, \\ \frac{1}{7}\left(3a+5b-4+\sqrt{(4a+2b-3)^2+7(1+\gamma)}\right), & (a,b) \in \Omega_\gamma^4, \\ \frac{1}{2}\left(a+b-1+\frac{\sqrt{3}}{3}\sqrt{5(a+b-1)^2-2(1-2a)(1-2b)+2(1+\gamma)}\right), & (a,b) \in \Omega_\gamma^5, \\ \frac{1}{7}\left(5a+3b-4+\sqrt{(2a+4b-3)^2+7(1+\gamma)}\right), & (a,b) \in \Omega_\gamma^6, \\ \frac{1}{7}\left(4a+2b-3+\sqrt{(4a+2b-3)^2+7(1+\gamma)}\right), & (a,b) \in \Omega_\gamma^7, \\ \frac{1}{4}\left(3a+b-2+\sqrt{(a+b-1)^2+(1-2a)(1-2b)+2(1+\gamma)}\right), & (a,b) \in \Omega_\gamma^8. \end{cases} \tag{2.65}$$

2.6.3 Znajomość tylko rozkładów brzegowych

W praktyce znacznie prostsze jest dopasowanie rozkładów brzegowych X_i niż pełnego rozkładu wielowymiarowego wektora $(X_1, \ldots, X_d)$. W konsekwencji rozkład Y może być elementem należącym do dopuszczalnej klasy Frécheta, obejmującej wszystkie rozkłady Y zgodne z ustalonymi rozkładami X_i i możliwymi strukturami zależności pomiędzy nimi:

$$\mathfrak{S}(F_1, \ldots, F_d) = \{Y = X_1 + \ldots + X_d : X_i \sim F_i, i = 1, \ldots, d\} \tag{2.66}$$

Dalej oznaczono górną granicę klasy przez $\bar{e}_\tau$ i odpowiednio dolną granicę przez $\underline{e}_\tau$. Poszukiwanie tych granic było poruszane w literaturze od lat 80. W (Tchen, 1980) pokazano, że komonotoniczna struktura zależności prowadzi do elementu maksymalnego (L^c) ze względu na porządek wypukły $(\preccurlyeq_{cx})$ w klasie dopuszczalnej $\mathfrak{S}$:

Zgodnie z klasycznym podejściem Frécheta–Hoeffdinga–Tchena, dla każdej zagregowanej zmiennej Y należącej do klasy Frécheta zachodzi:

$$\forall\, Y \in \mathfrak{S}: \quad Y \preccurlyeq_{cx} Y^c := \sum_{i=1}^{d} F_i^{-1}(U), \tag{2.67}$$

gdzie $U \sim \text{Uniform}(0,1)$. Zmienna Y^c odpowiada komonotonicznej strukturze zależności między składnikami X_i i stanowi element maksymalny w sensie porządku wypukłego $(\preccurlyeq_{cx})$ w

klasie $\mathfrak{S}$. Na tej podstawie definiuje się górne ograniczenie miary ryzyka w postaci:

$$\bar{e}_\tau = e_\tau\left(Y^c\right), \tag{2.68}$$

gdzie funkcjonał $e_\tau(X)$ oznacza ekspektyl poziomu τ. Określa asymetryczną wartość oczekiwaną (Newey & Powell, 1987; Taylor, 2008; Bellini, Klar, Müller, & Rosazza Gianin, 2014). Zgodnie z definicją przedstawioną przez (Newey & Powell, 1987), expectyl wyraża punkt minimalizujący ważoną wartość oczekiwaną błędu kwadratowego:

$$e_\tau(X) = \underset{e\in\mathbb{R}}{\operatorname{argmin}}\, E\left[\left(\tau\mathbf{1}_{\{X>e\}} + (1-\tau)\mathbf{1}_{\{X<e\}}\right)(X-e)^2\right]. \tag{2.69}$$

Funkcjonał $e_\tau(X)$ stanowi uogólnienie klasycznej wartości oczekiwanej dla $\tau = 0.5$ oraz przyjmuje coraz bardziej konserwatywny charakter wraz ze wzrostem poziomu τ, kładąc większy nacisk na wartości z prawego ogona rozkładu. Jak wykazali (Bellini et al., 2014), expectyl $e_\tau(X)$ jest spójną miarą ryzyka dla $\tau \in [\frac{1}{2}, 1)$, co wynika z jego monotoniczności, subaddytywności i dodatniej jednorodności. W przypadku identycznych rozkładów brzegowych $F_i = F_1$ dla $i = 2, \ldots, d$, z własności dodatniej jednorodności i komonotonicznej addytywności expectyla wynika zależność:

$$\bar{e}_\tau = e_\tau(dX_1) = d\, e_\tau(X_1) = \sum_{i=1}^{d} e_\tau(X_i), \tag{2.70}$$

co oznacza, że w przypadku pełnej komonotoniczności expectyl zagregowanej zmiennej Y jest równy sumie expectyli jej składników X_i (Dhaene et al., 2002; Bellini et al., 2014).

Znalezienie dolnego ograniczenia jest bardziej złożone. W pracy (Bellini & Di Bernardino, 2015) pokazano, że

$$\underline{e}_\tau \;\geq\; E(Y) \;=\; \sum_{i=1}^{d} E(X_i). \tag{2.71}$$

Jeśli dopuszczalna klasa $\mathfrak{S}$ zawiera przypadek, w którym zmienna Y przyjmuje stałą wartość równą swojej średniej $E(Y)$, to $E(Y)$ stanowi najmniejszy element w sensie porządku wypukłego, a nierówność jest ostra. Oznacza to, że w zbiorze dopuszczalnych rozkładów istnieje sytuacja całkowitej pewności, co odpowiada przypadkowi deterministycznemu i wyznacza dolną granicę w porządku wypukłym. Metodę numerycznego wyznaczania ograniczeń $\underline{e}_\tau$ i $\bar{e}_\tau$ zaproponowano w (Jakobsons & Vanduffel, 2015), natomiast (Jakobsons, Han, & Wang, 2016) przedstawili podejście oparte na równaniach całkowych dla przypadku identycznych rozkładów brzegowych. W (P. Embrechts et al., 2013; Puccetti & Rüschendorf, 2012, 2015) autorzy zaproponowali algorytm przegrupowania (*Rearrangement Algorithm*, RA), który zapewnia lepsze przybliżenia najlepszych możliwych dolnych i górnych granic w przypadku VaR. Wadą tego algorytmu jest to, że wraz ze wzrostem liczby agregowanych zmiennych X_i jego złożoność obliczeniowa istotnie wzrasta, co ogranicza efektywność obliczeń w przypadku dużych wymiarów. Natomiast, w pracy (Hofert, Memartoluie, Saunders, & Wirjanto, 2017) opracowano ulepszenie algorytmu RA w postaci *Adaptive Rearrangement Algorithm* (ARA). Ideę tego algorytmu przedstawiono poniżej.

Dla ustalonego poziomu ufności $\alpha \in (0, 1)$ oraz znanych rozkładów brzegowych $F_1, \ldots, F_d$ określa się odpowiadające im funkcje kwantylowe $F_1^{-1}, \ldots, F_d^{-1}$. Liczba punktów dyskretyzacji oznaczana jest przez N i przyjmuje wartości będące kolejnymi potęgami liczby 2, tzn. $N = 2^k$.

Ponadto, przyjmuje się wektor tolerancji zbieżności $\varepsilon = (\varepsilon_1, \varepsilon_2)$, gdzie: ε_1 oznacza indywidualną tolerancję zbieżności w ramach danej iteracji (dla ustalonego N), a ε_2 oznacza łączną tolerancję zbieżności pomiędzy wynikami uzyskanymi dla kolejnych wartości N. Dla każdej wartości $N = 2^k$ oraz ustalonej maksymalnej liczby iteracji algorytm ARA przebiega według poniższych kroków:

1. W pierwszym etapie algorytmu konstruuje się dwie macierze wartości kwantyli, które stanowią dyskretne przybliżenie rozkładów brzegowych zmiennych losowych $X_1, \ldots, X_d$. Macierz $\underline{X}^{\alpha}$ odpowiada dolnemu ograniczeniu VaR dla zmiennej zagregowanej $Y = X_1 + \ldots + X_d$, natomiast macierz $\bar{X}^{\alpha}$ służy do wyznaczenia ograniczenia górnego. Każda kolumna tych macierzy zawiera wartości kwantyli jednego z rozkładów brzegowych, natomiast każdy wiersz odpowiada temu samemu poziomowi prawdopodobieństwa $p \in [\alpha, 1)$. Wartości te stanowią punkt wyjścia do kolejnych etapów algorytmu, w których następuje permutacja kolumn w celu uzyskania ekstremalnych zależności między zmiennymi. Dla $i \in \{1, \ldots, N\}$ oraz $j \in \{1, \ldots, d\}$ definiuje się macierz dla ograniczenia dolnego $\underline{X}^{\alpha} = (\underline{x}_{ij}^{\alpha})$, gdzie

$$\underline{x}_{ij}^{\alpha} = F_j^{-1}\left(\alpha + \frac{(1-\alpha)(i-1)}{N}\right). \tag{2.72}$$

   Analogicznie definiuje się macierz dla ograniczenia górnego $\bar{X}^{\alpha} = (\bar{x}_{ij}^{\alpha})$, gdzie

$$\bar{x}_{ij}^{\alpha} = F_j^{-1}\left(\alpha + \frac{(1-\alpha)i}{N}\right). \tag{2.73}$$

   Jeżeli dla $i = N$ oraz $j \in \{1, \ldots, d\}$ zachodzi $F_j^{-1}(1) = \infty$, to w celu uniknięcia wartości nieskończonych przyjmuje się

$$F_j^{-1}\left(\alpha + \frac{(1-\alpha)(N-\frac{1}{2})}{N}\right). \tag{2.74}$$

2. W kolejnym kroku algorytmu dokonuje się losowego permutowania elementów w każdej kolumnie macierzy $\underline{X}^{\alpha}$ oraz $\bar{X}^{\alpha}$. Celem tego etapu jest uniknięcie przypadkowych zależności pomiędzy kolumnami, które mogłyby wynikać z uporządkowania kwantyli w poprzednim kroku.

3. W kolejnym etapie algorytmu dokonuje się iteracyjnego przestawiania kolumn dla $j \in \{1, 2, \ldots, d, 1, 2, \ldots, d, \ldots\}$ w macierzach $\underline{X}^{\alpha}$ oraz $\bar{X}^{\alpha}$. Każda kolumna jest ustawiana przeciwnie do sumy wszystkich pozostałych kolumn, co pozwala uzyskać układ wartości odpowiadający skrajnemu (najmniej lub najbardziej korzystnemu) współzależnemu zachowaniu zmiennych losowych. Proces ten jest powtarzany aż do osiągnięcia maksymalnej liczby przestawień kolumn lub spełnienia warunku zbieżności:

$$\left|\frac{s(\underline{Y}^{\alpha}) - s(\underline{X}^{\alpha})}{s(\underline{X}^{\alpha})}\right| \leq \varepsilon_1, \tag{2.75}$$

$$\left|\frac{s(\bar{Y}^{\alpha}) - s(\bar{X}^{\alpha})}{s(\bar{X}^{\alpha})}\right| \leq \varepsilon_1, \tag{2.76}$$

gdzie funkcja $s(X)$ oznacza minimalną sumę wiersza macierzy $X = (x_{ij})$, czyli

$$s(X) = \min_{1 \le i \le N} \sum_{j=1}^{d} x_{ij}. \tag{2.77}$$

Wartość $s(X)$ reprezentuje najmniejszą możliwą sumę składników portfela dla danego układu kwantyli i stanowi przybliżenie VaR dla sumy zmiennych losowych przy poziomie ufności α. Po każdej iteracji macierze $\underline{Y}^{\alpha}$ i $\bar{Y}^{\alpha}$ przedstawiają układ wartości po dokonaniu kolejnego przestawienia kolumn, natomiast $\underline{X}^{\alpha}$ i $\bar{X}^{\alpha}$ odnoszą się do stanu sprzed tej operacji, czyli do macierzy uzyskanych przed rozpoczęciem bieżącego cyklu przestawień. Po zakończeniu iteracji ostateczne wartości ograniczeń wyznacza się jako minimalne sumy wierszy uzyskanych macierzy:

$$\underline{s}_N = s\left(\underline{X}^{\alpha}\right), \qquad \bar{s}_N = s\left(\bar{X}^{\alpha}\right). \tag{2.78}$$

4. Po zakończeniu iteracji dla danego N sprawdza się, czy wyniki uzyskane dla ograniczenia dolnego i górnego są wystarczająco zbieżne. Ocena ta opiera się na porównaniu względnej różnicy pomiędzy wartościami $\underline{s}_N$ i $\bar{s}_N$, zgodnie z warunkiem wspólnej tolerancji zbieżności:

$$\left| \frac{\bar{s}_N - \underline{s}_N}{\bar{s}_N} \right| \le \varepsilon_2. \tag{2.79}$$

Spełnienie powyższego kryterium oznacza, że wyniki dla obu ograniczeń można uznać za wystarczająco dokładne, a algorytm kończy swoje działanie. W przeciwnym przypadku proces obliczeniowy jest kontynuowany dla większej liczby punktów dyskretyzacji $N = 2^k$, aż do osiągnięcia zbieżności lub przekroczenia maksymalnej liczby iteracji.

5. Jeżeli zarówno indywidualne, jak i wspólne warunki zbieżności zostały spełnione, algorytm kończy działanie i zwraca ostateczne wyniki w postaci pary wartości $\left(\underline{s}_N, \bar{s}_N\right)$, odpowiadających dolnemu i górnemu ograniczeniu VaR.

Dotychczasowe badania numeryczne wykazały, że ARA prezentuje bardzo dobrą dokładność.

W dalszej części rozważa się przypadki, w których analizowane są granice VaR, przy założeniu znajomości rozkładów brzegowych opisujących poszczególne rodzaje ryzyka oraz dodatkowego ograniczenia w postaci znanej granicy wariancji sumy. Założenie to ma istotne znaczenie praktyczne, ponieważ w wielu rzeczywistych sytuacjach odpowiada maksymalnemu zakresowi dostępnych informacji podczas szacowania wartości VaR dla portfela, w tym również w kontekście agregacji ryzyka oraz oceny wypłacalności zakładów ubezpieczeń w ramach regulacji takich jak Bazylea III czy Solvency II.

Wariancja sumy portfela może być zazwyczaj oszacowana z odpowiednią dokładnością na podstawie danych historycznych bądź też implikowana z wykorzystaniem dostępnych informacji o korelacjach między ryzykami. Z punktu widzenia interpretacyjnego, ponieważ wariancja odzwierciedla średni rozrzut zagregowanej straty portfela wokół wartości oczekiwanej, jej znajomość ma istotny wpływ na określenie maksymalnej możliwej wartości VaR. Uwzględnienie górnej granicy wariancji zamiast jej dokładnej wartości pozwala dodatkowo odzwierciedlić niepewność statystyczną związaną z estymacją drugiego momentu rozkładu.

W pracy (Bernard, Rüschendorf, & Vanduffel, 2017a) przedstawiono granice VaR dla zmiennej losowej Y, będącej sumą czynników ryzyka o znanych rozkładach brzegowych. Niech $\alpha \in (0,1)$ oraz $X_j \sim F_j$ dla $j = 1, \ldots, d$, a agregowana strata portfela wynosi $Y = \sum_{j=1}^{d} X_j$. Wówczas zachodzą następujące relacje:

$$VaR_\alpha(Y) \leq \overline{VaR}_\alpha^m := ES_\alpha(Y^c) = \sum_{j=1}^{d} ES_\alpha(X_j), \tag{2.80}$$

oraz

$$VaR_\alpha(Y) \geq \underline{VaR}_\alpha^m := LES_\alpha(Y^c) = \sum_{j=1}^{d} LES_\alpha(X_j), \tag{2.81}$$

gdzie Y^c oznacza komonotoniczną strukturę zależności między X_i, czyli sytuację pełnej dodatniej zależności pomiędzy nimi. Granice $\underline{VaR}_\alpha^m$ i $\overline{VaR}_\alpha^m$ opisują zatem skrajne wartości możliwej wartości zagrożonej portfela Y przy znanych rozkładach brzegowych, lecz nieznanej strukturze zależności.

W pracy (Bernard et al., 2017a) przedstawiono również uogólnione granice VaR dla przypadku, w którym oprócz znajomości rozkładów brzegowych poszczególnych czynników ryzyka $X_j \sim F_j$ $(j = 1, \ldots, d)$ dostępna jest również informacja o maksymalnej wariancji zagregowanej straty $Y = \sum_{j=1}^{d} X_j$, dana jako $\mathrm{VaR}(Y) \leq s^2$. Dla poziomu istotności $\alpha \in (0,1)$ granice te przyjmują postać:

$$VaR_\alpha(Y) \leq \overline{VaR}_\alpha^{m,s} := \min\left(\mu + s\sqrt{\frac{\alpha}{1-\alpha}}, \overline{VaR}_\alpha^m\right), \tag{2.82}$$

oraz

$$VaR_\alpha(Y) \geq \underline{VaR}_\alpha^{m,s} := \max\left(\mu - s\sqrt{\frac{1-\alpha}{\alpha}}, \underline{VaR}_\alpha^m\right), \tag{2.83}$$

gdzie $\mu = \mathbb{E}(X_i, ..., X_d)$ oznacza wartość oczekiwaną portfela, a s^2 stanowi górne ograniczenie jego wariancji. Wyznaczając takie ograniczenie dla agregacji ryzyka zgodnie z dyrektywą Solvency II, wariancję s^2 można oszacować w oparciu o macierz korelacji pomiędzy modułami ryzyka. Dla znanych odchyleń standardowych poszczególnych składników σ_j oraz współczynników korelacji ρ_{ij} wartość ta może zostać wyznaczona według wzoru:

$$s^2 = \sum_{j=1}^{d} \sigma_j^2 + 2\sum_{i<j} \rho_{ij}\sigma_i\sigma_j, \tag{2.84}$$

czyli w sposób zgodny z agregacją ryzyk stosowaną w FS. W prezentowanych dalej podejściach wariancja sumy może być wyznaczana dokładnie w ten sam sposób.

W wielu praktycznych zastosowaniach przyjmuje się założenie o znajomości dwóch pierwszych momentów poszczególnych rodzajów ryzyka oraz zakresu możliwych wartości straty portfela. Ujęcie to odzwierciedla realistyczne ograniczenia informacyjne występujące w procesie szacowania ryzyka w praktyce. Zastąpienie dokładnych wartości momentów ich maksymalnymi dopuszczalnymi wartościami pozwala uwzględnić niepewność statystyczną związaną z estymacją parametrów rozkładów. Wykorzystanie tych informacji prowadzi do wyznaczenia granic określonych przez (Bernard, Kazzi, & Vanduffel, 2019). Niech $Y \in [a, b]$, $E(Y) = \mu$

oraz $E(Y^k) \leq d_k$ będą dla $k = 2, 3, \ldots, r$. Dodatkowo, zakłada się, że $E(X_i) = \mu_i$ oraz $\mathrm{VaR}(X_i) = v_i^2$. Wówczas zachodzą ograniczenia:

$$VaR_\alpha(Y) \leq \overline{VaR}_\alpha^{ab,\mu,d,im} := \min\left(\mu + \psi,\ \mu + \sum_{i=1}^{n} v_i \sqrt{\frac{\alpha}{1-\alpha}}\right), \tag{2.85}$$

$$VaR_\alpha(Y) \geq \underline{VaR}_\alpha^{ab,\mu,d,im} := \max\left(\mu - \psi\frac{1-\alpha}{\alpha},\ \mu - \sum_{i=1}^{n} v_i \sqrt{\frac{1-\alpha}{\alpha}}\right), \tag{2.86}$$

gdzie ψ jest maksymalną wartością w przedziale

$$\left[0,\ \min\left\{ b - \mu,\ (\mu - a)\frac{\alpha}{1-\alpha}\right\}\right], \tag{2.87}$$

taką, że spełnione są nierówności:

$$(\mu + \psi)^k (1-\alpha) + \left(\mu - \psi\frac{1-\alpha}{\alpha}\right)^k \alpha \leq d_k, \qquad k = 1, \ldots, r. \tag{2.88}$$

Większość rozkładów wykorzystywanych w modelowaniu ryzyka finansowego i ubezpieczeniowego charakteryzuje się jednomodalnością, co wynika z faktu, że cecha ta jest powszechnie obserwowana w rzeczywistych zbiorach danych ubezpieczeniowych. W pracy (Bernard et al., 2019) wyprowadzono granice wartości zagrożonej dla rozkładów jednomodalnych o zadanej wartości oczekiwanej i wariancji, co umożliwia bardziej realistyczne ujęcie struktury ryzyka w porównaniu z założeniem pełnej dowolności rozkładu. Granice te przyjmują następującą postać:

$$VaR_\alpha(Y) \leq \overline{VaR}_\alpha^{u,\mu,s} := \begin{cases} \mu + s\sqrt{\dfrac{4}{9(1-\alpha)} - 1}, & \text{dla } \alpha \in [5/6, 1), \\ \mu + s\sqrt{\dfrac{3\alpha}{4-3\alpha}}, & \text{dla } \alpha \in [0, 5/6), \end{cases} \tag{2.89}$$

$$VaR_\alpha(Y) \geq \underline{VaR}_\alpha^{u,\mu,s} := \begin{cases} \mu - s\sqrt{\dfrac{1-\alpha}{1/3+\alpha}}, & \text{dla } \alpha \in (1/6, 1), \\ \mu - s\sqrt{\dfrac{4}{9\alpha} - 1}, & \text{dla } \alpha \in (0, 1/6], \end{cases} \tag{2.90}$$

gdzie $\mu = E(Y)$ oznacza wartość oczekiwaną portfela, a $s^2 = \mathrm{VaR}(Y)$ stanowi jego wariancję. Włączenie założenia jednomodalności prowadzi do węższych i bardziej realistycznych przedziałów niepewności dla wartości zagrożonej w porównaniu z przypadkami, w których rozkład ryzyka pozostaje dowolny.

3. Metody uczenia maszynowego w szacowaniu SCR

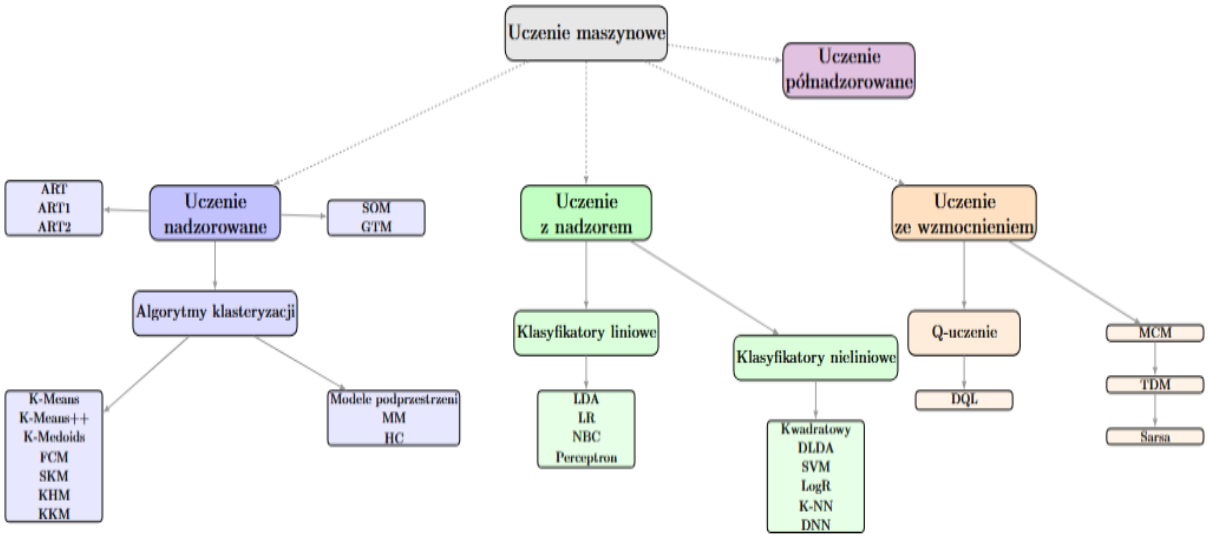
Rozwój metod analityki predykcyjnej i uczenia maszynowego w ostatnich latach otworzył nowe możliwości w obszarze modelowania ryzyka ubezpieczeniowego, w tym wyznaczania SCR-u. Wykorzystanie tradycyjnych metod aktuarialnych opartych na modelach liniowych oraz na symulacjach monte Carlo jest uzupełniane lub zastępowane metodami opartymi na uczeniu maszynowym. Metody te pozwalają na efektywniejsze uchwycenie złożonych, nieliniowych zależności pomiędzy zmiennymi ryzyka oraz lepsze odwzorowanie struktury portfela ubezpieczeniowego. Techniki uczenia maszynowego znajdują zastosowanie zarówno w modelowaniu częstotliwości i rozmiaru szkód, jak i w estymacji zależności pomiędzy ryzykami, analizie scenariuszy oraz w procesach agregacji ryzyka. Atutem metod uczenia maszynowego jest ich zdolność do pracy na dużych i złożonych zbiorach danych pochodzących z różnych źródeł: danych polisowych, szkodowych i finansowych. Pozwala to na budowę bardziej realistycznych i kompleksowych modeli ryzyka, które lepiej odzwierciedlają rzeczywiste zależności występujące w portfelu ubezpieczeniowym.

W ostatnich latach w literaturze można zauważyć prace, w których łączona jest teoria kopul z metodami uczenia maszynowego. Wynika to z faktu, iż kopule pozwalają na modelowanie struktur zależności, natomiast uczenie maszynowe pozwala na znacznie dokładniejsze estymowanie i aproksymowanie złożonych funkcji wielowymiarowych. Z uwagi na fakt, iż kopule opisują strukturę zależności między danymi, wykorzystuje się je do redukcji wymiaru danych wejściowych do algorytmów uczenia maszynowego. Dzięki temu, model uczenia maszynowego uczy się na danych, które wnoszą istotne informacje do modelu. Kolejną kwestią jest wykorzystanie uczenia maszynowego w identyfikacji postaci kopuli. Klasyczne metody numeryczne okazują się być bardzo kosztowne z uwagi na czas obliczeń oraz pamięć ram. Przykładowo, dla funkcji generatora kopul Archimedesa budowana jest sieć neuronowa, która zachowuje postać oraz teoretyczne warunki tej funkcji. Innym przykładem jest wykorzystanie kopul w generowaniu danych dla procesu uczenia maszynowego. Często dane są poufne lub w danych występują luki. Kopule uczą się struktury zależności danych, a następnie generują dane syntetyczne, które są wykorzystywane w procesie uczenia algorytmów uczenia maszynowego.

W dalszej części rozdziału opisano wybrane metody uczenia maszynowego, takie jak sieci neuronowe, lasy losowe oraz maszyny wektorów nośnych. Następnie przeprowadzono studium literaturowe, ukazujące wykorzystanie algorytmów uczenia maszynowego w ubezpieczeniach, ze szczególnym uwzględnieniem aktuariatu. Ostatnia część rozdziału poświęcona została łączeniu kopul z algorytmami uczenia maszynowego, zarówno w kontekście wykorzystania kopul w redukcji wymiaru, jak również w generowaniu danych syntetycznych, a także w wykorzystaniu uczenia maszynowego w estymacji i konstrukcji kopul.

3.1 Charakterystyka wybranych algorytmów uczenia maszynowego

Algorytmy uczenia maszynowego dzielimy na cztery podstawowe klasy: uczenie bez nadzoru (ang. Unsupervised Learning, UL), uczenie nadzorowane (ang. Supervised Learning, SL), uczenie przez wzmacnianie (ang. Reinforcement Learning, RL) oraz uczenie częściowo nadzorowane (ang. Semi-Supervised Learning, SSL). W pracy (Awad, 2012) dokonano jednej z możliwych taksonomii algorytmów uczenia maszynowego i przedstawiono ją na poniższym rysunku (rys. 3.1).



Rysunek 3.1: Taksonomia algorytmów uczenia maszynowego.
Źródło: Opracowanie własne.

W sektorze ubezpieczeń najczęściej stosowane są algorytmy uczenia nadzorowanego, stanowiące podstawę wielu nowoczesnych rozwiązań analitycznych, wykorzystywanych m.in. do taryfikacji, oceny ryzyka oraz wykrywania nadużyć. Celem tego rodzaju algorytmów jest zbudowanie estymatora zdolnego do przewidywania wartości zmiennej zależnej (objaśnianej) na podstawie zestawu zmiennych niezależnych (objaśniających), które opisują dany obiekt, np. klienta, polisę czy zdarzenie ubezpieczeniowe. W kontekście uczenia maszynowego wartości zmiennej zależnej określane są często mianem etykiet (ang. labels).

W procesie uczenia algorytm otrzymuje zestaw danych treningowych składających się z par: wektor wartości zmiennych objaśniajacych oraz odpowiadająca mu wartość zmiennej objaśnianej (etykieta, wartość docelowa), przypisanych poszczególnym obiektom (jednostkom, obserwacjom). Na tej podstawie model uczy się zależności między zmiennymi, porównując przewidywane wyniki z rzeczywistymi i minimalizując powstały błąd. W efekcie uzyskiwany jest model, który potrafi uogólniać wiedzę zdobytą na danych treningowych i dokonywać trafnych prognoz dla nowych, nieznanych obserwacji.

Algorytmy uczenia nadzorowanego znajdują szerokie i zróżnicowane zastosowania w działalności ubezpieczeniowej. Wykorzystywane są między innymi do modelowania prawdopodobieństwa wystąpienia szkody i prognozowania jej wysokości, a także do oceny ryzyka portfela ubezpieczeń oraz segmentacji klientów według ich profilu ryzyka. Istotną rolę odgrywają również w procesach wykrywania nadużyć i anomalii w danych operacyjnych oraz w optymalizacji procesów likwidacji szkód i zarządzania rezerwami techniczno-ubezpieczeniowymi. W dalszej części omówiono najczęściej stosowane w praktyce ubezpieczeniowej algorytmy uczenia maszynowego.

3.1.1 Algorytm maszyny wektorów nośnych (SVM)

Maszyna wektorów nośnych (ang. Support Vector Machine – SVM) jest jednym z najważniejszych algorytmów uczenia nadzorowanego, stosowanym zarówno w klasyfikacji, jak i regresji. Jej celem jest znalezienie hiperpłaszczyzny decyzyjnej, która w sposób optymalny rozdziela obserwacje należące do różnych klas lub najlepiej aproksymuje zależność pomiędzy

zmiennymi w przypadku regresji.

Dla problemu klasyfikacji binarnej przyjmuje się, że dane treningowe mają postać:

$$\{(x_i, y_i)\}_{i=1}^{n}, \quad x_i \in \mathbb{R}^m, \quad y_i \in \{-1, 1\} \tag{3.1}$$

gdzie x_i oznacza wektor cech (zmienne objaśniające), a y_i - zmienną objaśnianą (etykietę klasy).

W przypadku danych liniowo separowalnych, SVM poszukuje hiperpłaszczyzny postaci:

$$w^\top x + b = 0 \tag{3.2}$$

która maksymalizuje margines separacji między klasami. Warunek poprawnej klasyfikacji można zapisać jako:

$$w^\top x + b = 0, \tag{3.3}$$

$$y_i(w^\top x_i + b) \geq 1, \quad \text{dla } i = 1, \ldots, n. \tag{3.4}$$

Zadanie optymalizacyjne sprowadza się do minimalizacji normy wektora wag w:

$$\min_{w,b} \frac{1}{2}\|w\|^2 \quad \text{przy warunku} \quad y_i(w^\top x_i + b) \geq 1. \tag{3.5}$$

Punkty danych, które leżą najbliżej hiperpłaszczyzny, określane są mianem wektorów nośnych (ang. support vectors), i to one wyznaczają granicę decyzyjną modelu.

W rzeczywistych zastosowaniach dane często nie są idealnie separowalne, dlatego wprowadza się zmienne miękkie $\xi_i \geq 0$, które dopuszczają pewne błędy klasyfikacji. Wówczas funkcja celu przyjmuje postać:

$$\min_{w,b,\xi} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \xi_i, \tag{3.6}$$

$$\text{przy warunku} \quad y_i(w^\top x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \tag{3.7}$$

gdzie parametr $C > 0$ kontroluje kompromis między szerokością marginesu a liczbą dopuszczalnych błędów.

W sytuacjach, gdy dane nie są separowalne liniowo, SVM wykorzystuje tzw. trik jądrowy (ang. kernel trick). Polega on na przekształceniu danych do przestrzeni o wyższym wymiarze za pomocą funkcji $\Phi(x)$, w której stają się one separowalne liniowo. Zamiast jawnego przekształcenia wykorzystuje się funkcję jądra $K(x_i, x_j)$, która oblicza iloczyn skalarny:

$$K(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle. \tag{3.8}$$

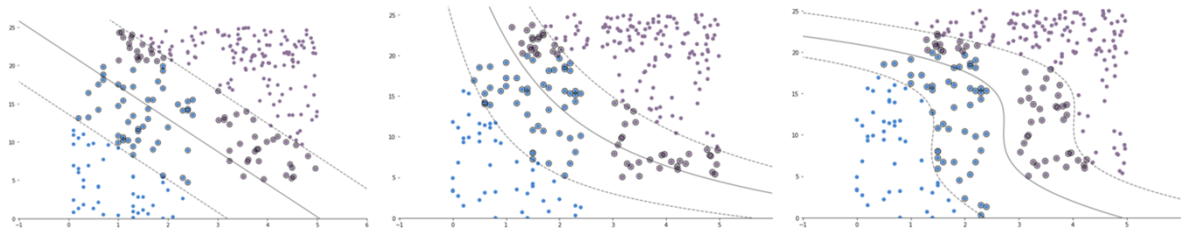
w przestrzeni przekształconych cech. Operator $\langle \cdot, \cdot \rangle$ oznacza iloczyn skalarny dwóch wektorów i w przestrzeni euklidesowej odpowiada sumie iloczynów ich współrzędnych:

$$\langle x, z \rangle = \sum_{k=1}^{m} x_k z_k. \tag{3.9}$$

Dzięki zastosowaniu funkcji jądra, SVM może efektywnie modelować nieliniowe zależności między zmiennymi bez konieczności bezpośredniego odwzorowania danych w przestrzeń o wyższym wymiarze. W praktyce najczęściej wykorzystuje się jądra: liniowe, wielomianowe, sigmoidalne oraz radialne (ang. Radial Basis Function - RBF).

Poniżej na Rysunku 3.2 przedstawiono przykłady klasyfikacji odpowiednio dla jądra:

- liniowego $K\left(\mathbf{x}_i, \mathbf{x}_j\right) = \left\langle \mathbf{x}_i, \mathbf{x}_j \right\rangle$,
- wielomianowego $K\left(\mathbf{x}_i, \mathbf{x}_j\right) = \left(\left\langle \mathbf{x}_i, \mathbf{x}_j \right\rangle + c\right)^2$,
- radialnego $K\left(\mathbf{x}_i, \mathbf{x}_j\right) = \exp\left(-\frac{\left\|\mathbf{x}_i - \mathbf{x}_j\right\|^2}{2\sigma^2}\right)$.



Rysunek 3.2: Klasyfikator SVM dla trzech funkcji jądra.
Źródło: Opracowanie własne.

W wersji regresyjnej algorytm działa analogicznie, minimalizując błędy większe niż zadana tolerancja ε, co pozwala uzyskać gładką funkcję przybliżającą dane.

Zaletą SVM jest stabilność i odporność na przeuczenie, szczególnie w przypadkach, gdy liczba zmiennych jest duża w stosunku do liczby obserwacji. W ubezpieczeniach algorytm ten znajduje zastosowanie m.in. w:

- klasyfikacji klientów według poziomu ryzyka,
- prognozowaniu prawdopodobieństwa szkody,
- wykrywaniu nadużyć i anomalii w danych,
- modelowaniu wartości odszkodowań.

Dzięki solidnym podstawom teoretycznym oraz zdolności do modelowania złożonych, nieliniowych zależności, maszyna wektorów nośnych stanowi jedno z najbardziej uniwersalnych i efektywnych narzędzi analizy predykcyjnej w zastosowaniach ubezpieczeniowych.

3.1.2 Głębokie sieci neuronowe

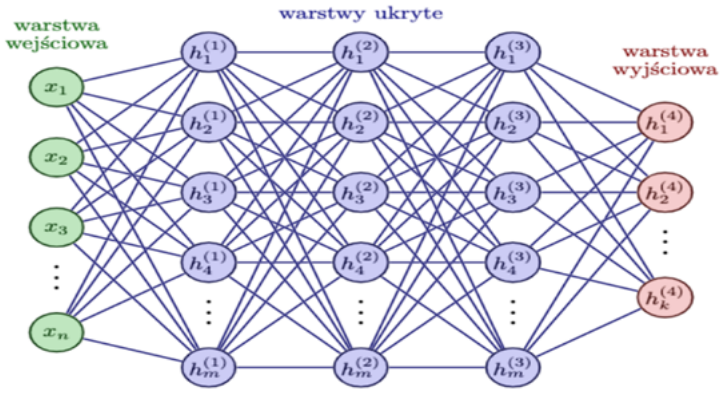
Początki koncepcji sztucznych sieci neuronowych (ang. Artificial Neural Networks – ANN) sięgają lat czterdziestych XX wieku. McCulloch and Pitts (1943) przedstawili pierwszą formalną koncepcję modelu neuronowego, stanowiącą próbę odwzorowania sposobu działania ludzkiego mózgu w formie matematycznej. Jednak, jak podkreśla Tadeusiewicz (2009), współczesne sieci neuronowe stanowią nadal znacznie uproszczoną reprezentację biologicznych pierwowzorów.

Architektura głębokiej sieci neuronowej składa się zazwyczaj z trzech podstawowych warstw:

- warstwy wejściowej – odpowiedzialnej za odbieranie danych zewnętrznych i przekazywanie ich do kolejnych warstw w celu uczenia modelu,
- warstw ukrytych – przetwarzających sygnały otrzymane z warstw poprzednich i przekazujących wyniki do kolejnych neuronów,

- warstwy wyjściowej – generującej ostateczny wynik obliczeń na podstawie sygnałów pochodzących z warstwy poprzedzającej.

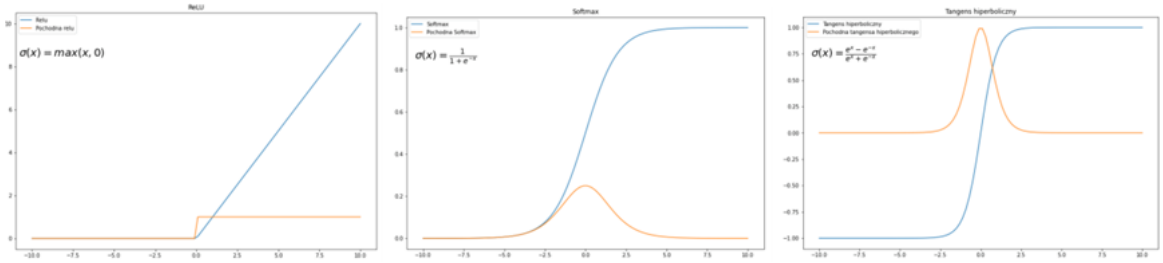
Tak skonstruowany model umożliwia odwzorowanie złożonych, nieliniowych zależności pomiędzy zmiennymi, co stanowi podstawę jego szerokiego zastosowania w uczeniu maszynowym i analizie danych. Na Rysunku 3.3 przedstawiono sieć neuronową z n neuronami wejściowymi, trzema warstwami ukrytymi, z których każda ma m neuronów, oraz warstwą wyjściową z k neuronami.



Rysunek 3.3: Sztuczna sieć neuronowa.
Źródło: Opracowanie własne.

W dotychczasowej literaturze opisano wiele architektur sieci neuronowych. W niniejszym podrozdziale zostanie opisana jedna z sieci, która będzie wykorzystywana w badaniach empirycznych – jednokierunkowa sieć neuronowa. Wartości przewidywane przez sieć neuronową zależą od kilku kluczowych elementów jej architektury. Należą do nich: wagi sieci neuronowej w, liczba warstw ukrytych L_s, liczba neuronów w poszczególnych warstwach ukrytych m oraz funkcje aktywacji wykorzystywane w tych warstwach. Właściwy dobór wymienionych parametrów ma istotny wpływ na zdolność sieci do odwzorowania złożonych zależności pomiędzy zmiennymi wejściowymi a prognozowaną wartością zmiennej wyjściowej.

Funkcje aktywacji σ wykorzystywane są po to, aby sieć neuronowa mogła uczyć się nieliniowych zależności. Najczęściej wykorzystywane funkcje aktywacji przedstawiono na Rysunku 3.4. Kolejno są to funkcje ReLu, Softmax i tangens hiperboliczny.



Rysunek 3.4: Funkcje aktywacji.
Źródło: Opracowanie własne.

Procedura uczenia głębokiej sieci neuronowej rozpoczyna się od przygotowania danych, które najczęściej dzieli się na zbiór uczący i testowy, a niekiedy również na zbiór walidacyjny.

Aby rozpocząć proces uczenia, do warstwy wejściowej sieci ($l = 0$) wprowadza się wektor danych wejściowych:

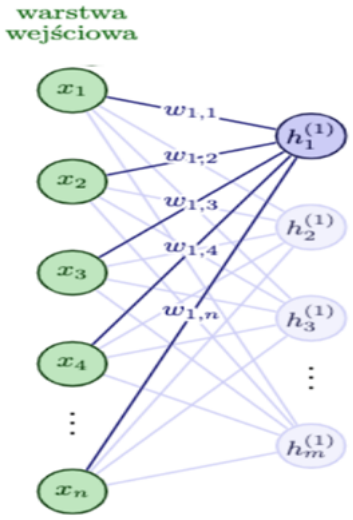
$$h^{(0)}(x) = x. \tag{3.10}$$

Następnie, między każdą parą warstw wyznaczane są wagi połączeń, gdzie $w_{jk}^{(l)}$ oznacza wagę łączącą k-ty neuron w $(l-1)$-szej warstwie z j-tym neuronem w l-tej warstwie. Dodatkowo, wprowadzany jest bias $b_j^{(l)}$, który umożliwia przesunięcie funkcji aktywacji względem osi X, co pozwala na lepsze odwzorowanie nieliniowych zależności w danych.

Aktywacja j-tego neuronu w l-tej warstwie wyraża się wzorem:

$$h_j^{(l)} = \sigma\left(\sum_k w_{jk}^{(l)} h_k^{(l-1)} + b_j^{(l)}\right), \tag{3.11}$$

gdzie $\sigma(\cdot)$ oznacza funkcję aktywacji, a $h_k^{(l-1)}$ – wyjście z k-tego neuronu w poprzedniej warstwie. Rysunek 3.5 przedstawia połączenie każdego neuronu z warstwy wejściowej z pierwszym neuronem w warstwie ukrytej.



Rysunek 3.5: Połączenie warstwy wejściowej z pierwszą warstwą ukrytą.
Źródło: Opracowanie własne.

Jako wartość wyjściową głębokiej sieci neuronowej uzyskuje się funkcję wyjściową sieci w ostatniej warstwie ($l = L_s + 1$) z zastosowaniem funkcji aktywacji σ:

$$h^{(L_s+1)}(x) = \sigma\left(b^{(L_s+1)} + w^{(L_s+1)} h^{(L_s)}(x)\right) := f(x, \theta), \tag{3.12}$$

gdzie θ oznacza zbiór wszystkich parametrów modelu, tj. wag i biasów sieci.

Dla wartości wyjściowej sieci neuronowej $f(x, \theta)$ obliczana jest funkcja błędu (funkcja kosztu), stanowiąca miarę różnicy pomiędzy wartościami przewidywanymi przez model a rzeczywistymi wartościami $y(x)$, które sieć ma aproksymować. Istnieje wiele postaci funkcji błędu, a wybór odpowiedniej zależy od celu uczenia oraz charakteru problemu, na przykład klasyfikacyjnego lub regresyjnego.

Jedną z najczęściej stosowanych funkcji błędu w zadaniach regresyjnych jest błąd średniokwadratowy (ang. Mean Squared Error – MSE). Przy założeniu, że sieć posiada jedno wyjście, funkcja ta przyjmuje postać:

$$L(x, \theta) = \frac{1}{n} \sum_{x} \|y(x) - f(x, \theta)\|^2 , \tag{3.13}$$

gdzie n oznacza liczbę obserwacji w zbiorze uczącym.

W celu ograniczenia ryzyka nadmiernego dopasowania modelu do danych uczących stosuje się regularyzację, polegającą na dodaniu do funkcji błędu dodatkowego składnika kary. Techniki regularizacji odgrywają kluczową rolę w procesie uczenia modeli, zwłaszcza w przypadku złożonych struktur, takich jak sieci neuronowe, które są szczególnie podatne na zjawisko przeuczenia. Dwie najpopularniejsze regularyzacje to l_1 i l_2:

1. regularyzacja l_1:

$$L(x, \theta) = \frac{1}{n} \sum_{x} \|y(x) - f(x, \theta)\|^2 + \gamma \sum_{i} \sum_{j} |w_{ij}| \tag{3.14}$$

2. regularyzacja l_2:

$$L(x, \theta) = \frac{1}{2n} \sum_{x} \|y(x) - f(x, \theta)\|^2 + \gamma \sum_{i} \sum_{j} w_{ij}^2 \tag{3.15}$$

Zastosowanie regularyzacji l_1 i l_2 ma wpływ na nadmierne dopasowanie modelu, ponieważ wagi niektórych ukrytych warstw zbliżają się do (lub są równe) 0. W konsekwencji ich wpływ jest mniejszy, a ostateczna sieć jest prostsza, ponieważ jest bliższa mniejszej sieci. Również dla mniejszych wartości wag zmniejsza się wartość wyjścia z funkcji aktywacji neuronu ukrytego. Dla wartości bliskich 0 wiele funkcji aktywacji zachowuje się liniowo.

Aby zbudować model sieci neuronowej, najpierw musimy zainicjalizować wagi. W praktyce najczęściej wagi inicjalizuje się losowo.[24] W uczeniu modelu sieci neuronowej stosuje się algorytm propagacji wstecznej (Rumelhart, Hinton, & Williams, 1986), który ma na celu zminimalizowanie funkcji błędu poprzez dostosowanie wag i odchyleń sieci. Algorytm ten opiera się na gradiencie funkcji błędu względem wartości wag. Intuicyjnie, jeśli gradient wyznacza kierunek najszybszego wzrostu błędu w przestrzeni wag, to kierunek przeciwny powinien gwarantować szybki spadek błędu.Wagę $w_{ij}^{(l)}$ aktualizowaną w l-tej warstwie oznacza się jako $w_{ij}^{(l)}(n)$, natomiast jej wartość po dokonaniu aktualizacji zapisuje się w postaci $w_{ij}^{(l)}(n+1)$. Przyjmując, że η oznacza współczynnik uczenia, proces aktualizacji wag można wyrazić następująco:

$$w_{ij}^{(l)}(n+1) = w_{ij}^{(l)}(n) + \Delta w_{ij}^{(l)}(n) = w_{ij}^{(l)}(n) - \eta \cdot \frac{\partial L(n)}{\partial w_{ij}^{(l)}(n)} \tag{3.16}$$

Przedstawiona metoda aktualizacji wag nie zawsze zapewnia efektywną optymalizację procesu uczenia sieci neuronowej. W praktyce stosowanych jest wiele alternatywnych metod, które różnią się sposobem modyfikacji kroku uczenia oraz uwzględnieniem informacji o wcześniejszych iteracjach. Do najczęściej wykorzystywanych podejść należą:

[24] Porównanie metod inicjalizacji wag można znaleźć w (Thimm & Fiesler, 1996).

- metoda momentu (Polyak, 1964),
- algorytm Nesterova (Nesterov, 1983),
- metoda AdaGrad (Duchi, Hazan, & Singer, 2011),
- metoda RMSProp (Tieleman & Hinton, 2012),
- optymalizatory Adam oraz Nadam (Dozat, 2016).

W klasycznej metodzie spadku gradientu zmiana wag $\Delta w_{ij}^{(l)}(n)$ zależy wyłącznie od bieżącego gradientu funkcji kosztu $C(n)$:

$$\Delta w_{ij}^{(l)}(n) = -\eta \, \frac{\partial L(n)}{\partial w_{ij}^{(l)}(n)}, \tag{3.17}$$

gdzie η oznacza współczynnik uczenia. W praktyce metoda ta często prowadzi do oscylacji w kierunku spadku gradientu oraz spowolnienia zbieżności. W celu ograniczenia tego zjawiska wprowadzono tzw. *momentum* (Polyak, 1964), które uwzględnia kierunek poprzedniej aktualizacji wag:

$$\Delta w_{ij}^{(l)}(n) = \mu \, \Delta w_{ij}^{(l)}(n-1) - \eta \cdot \frac{\partial L(n)}{\partial w_{ij}^{(l)}(n)}, \tag{3.18}$$

gdzie $\mu \in [0, 1)$ jest współczynnikiem momentu określającym wpływ poprzedniego kroku aktualizacji. Nowe wartości wag oblicza się zgodnie z relacją:

$$w_{ij}^{(l)}(n+1) = w_{ij}^{(l)}(n) + \Delta w_{ij}^{(l)}(n). \tag{3.19}$$

Algorytm Nesterova (Nesterov, 1983) stanowi modyfikację metody momentu. Gradient obliczany jest nie w bieżącym punkcie, lecz w punkcie przewidywanym przez składnik pędu:

$$\Delta w_{ij}^{(l)}(n) = \mu \, \Delta w_{ij}^{(l)}(n-1) - \eta \cdot \frac{\partial L}{\partial w_{ij}^{(l)}} \Big(w_{ij}^{(l)}(n) + \mu \, \Delta w_{ij}^{(l)}(n-1) \Big). \tag{3.20}$$

Rozwiązanie to pozwala lepiej przewidywać kierunek spadku gradientu i ogranicza oscylacje, szczególnie w obszarach o złożonym kształcie funkcji kosztu.

Algorytm Adam (ang. Adaptive Moment Estimation) (Kingma & Ba, 2014) łączy zalety metod momentu oraz adaptacyjnych procedur dostosowania kroku uczenia, takich jak RMSProp. W tym podejściu wykorzystywane są dwie średnie ruchome: pierwszego rzędu (momentum) oraz drugiego rzędu (kwadratów gradientu):

$$g_{ij}^{(l)}(n) = \frac{\partial L(n)}{\partial w_{ij}^{(l)}(n)}, \tag{3.21}$$

$$m_{ij}^{(l)}(n) = \beta_1 \, m_{ij}^{(l)}(n-1) + (1-\beta_1) \, g_{ij}^{(l)}(n), \tag{3.22}$$

$$v_{ij}^{(l)}(n) = \beta_2 \, v_{ij}^{(l)}(n-1) + (1-\beta_2) \, (g_{ij}^{(l)}(n))^2, \tag{3.23}$$

gdzie $\beta_1, \beta_2 \in [0, 1)$ są współczynnikami wygładzania średnich ruchomych. Aby zminimalizować obciążenie estymatorów w początkowych iteracjach, stosuje się poprawki biasu:

$$\hat{m}_{ij}^{(l)}(n) = \frac{m_{ij}^{(l)}(n)}{1 - \beta_1^n}, \quad \hat{v}_{ij}^{(l)}(n) = \frac{v_{ij}^{(l)}(n)}{1 - \beta_2^n}. \tag{3.24}$$

Aktualizacja wag w metodzie Adam ma postać:

$$w_{ij}^{(l)}(n+1) = w_{ij}^{(l)}(n) - \eta \cdot \frac{\hat{m}_{ij}^{(l)}(n)}{\sqrt{\hat{v}_{ij}^{(l)}(n)} + \epsilon}, \tag{3.25}$$

gdzie ϵ jest niewielką stałą numeryczną, zapobiegającą dzieleniu przez zero.

Algorytm Nadam (ang. Nesterov-accelerated Adaptive Moment Estimation) (Dozat, 2016) łączy zalety algorytmu Adam z mechanizmem przyspieszenia Nesterova. Pozwala to na adaptacyjne dostosowanie kroku uczenia oraz lepsze przewidywanie kierunku spadku gradientu. Aktualizacja wag w tym przypadku przyjmuje postać:

$$w_{ij}^{(l)}(n+1) = w_{ij}^{(l)}(n) - \eta \cdot \frac{\beta_1 \, \hat{m}_{ij}^{(l)}(n) + \frac{(1-\beta_1)\, g_{ij}^{(l)}(n)}{(1-\beta_1^n)}}{\sqrt{\hat{v}_{ij}^{(l)}(n)} + \epsilon}. \tag{3.26}$$

Wprowadzenie składnika Nesterova pozwala na szybszą reakcję na zmiany kierunku gradientu, co prowadzi do przyspieszenia zbieżności w początkowych etapach uczenia oraz zwiększenia stabilności optymalizacji w porównaniu z klasycznym algorytmem Adam. Z tego względu w części empirycznej niniejszej pracy zastosowano optymalizator Nadam jako procedurę uczenia modeli sieci neuronowych.

3.1.3 Drzewa decyzyjne

Drzewo decyzyjne stanowi hierarchiczną strukturę umożliwiającą podział przestrzeni danych na grupy o podobnych charakterystykach w celu dokonania klasyfikacji lub regresji. Budowa drzewa opiera się na iteracyjnym procesie dzielenia przestrzeni zmiennych objaśniających na rozłączne regiony, które w sposób możliwie jednorodny grupują obserwacje pod względem zmiennych objaśniających. Metoda ta jest jedną z najczęściej stosowanych technik w uczeniu nadzorowanym ze względu na intuicyjność interpretacji oraz zdolność odwzorowania nieliniowych zależności pomiędzy zmiennymi.

Podstawowy algorytm budowy drzew decyzyjnych został sformalizowany w ramach koncepcji klasyfikacji i regresji (ang. Classification and Regression Trees – CART), zaproponowanej przez Breiman, Friedman, Olshen, and Stone (1984). Drzewa CART wykorzystują podziały rekurencyjne przestrzeni zmiennych objaśniających do tworzenia struktury drzewa, która może być stosowana zarówno w problemach klasyfikacyjnych, jak i regresyjnych. W pierwszym przypadku model służy do przypisania obserwacji do jednej z określonych kategorii, natomiast w drugim – do estymacji wartości zmiennej ilościowej.

W analizie regresji przyjmuje się, że drzewo decyzyjne można przedstawić jako funkcję $T(\mathbf{x}, \theta)$, która dzieli przestrzeń zmiennych objaśniających na M rozłącznych regionów

$R_1, R_2, \ldots, R_M$, z których każdy jest opisany przypisaną wartością c_m dla R_m ($m = 1, 2, \ldots, M$). Dla danej obserwacji $\mathbf{x}_i$ wartość przewidywana przez model wyrażana jest równaniem:

$$f(\mathbf{x}_i \mid \Theta) = \sum_{m=1}^{M} c_m \mathbf{1}_{R_m}(\mathbf{x}_i), \tag{3.27}$$

gdzie $\Theta = \{R_m, c_m\}_{m=1}^{M}$ oznacza komplet parametrów modelu, a $\mathbf{1}_{R_m}(\mathbf{x}_i)$ jest funkcją wskaźnikową, przyjmującą wartość 1 dla $\mathbf{x}_i \in R_m$ i 0 w przeciwnym przypadku.

Wartości c_m wyznaczane są w taki sposób, aby minimalizować funkcję straty, zdefiniowaną jako:

$$L(y_i, \hat{y}_i) = \sum_{i=1}^{N} \left(y_i - \sum_{m=1}^{M} \hat{c}_m \mathbf{1}_{R_m}(\mathbf{x}_i) \right)^2, \tag{3.28}$$

gdzie $\hat{y}_i$ oznacza wartość przewidywaną przez model dla obserwacji $\mathbf{x}_i$, a N jest liczbą obserwacji w zbiorze uczącym.

Regiony $R_1, R_2, \ldots, R_M$ są wyznaczane zgodnie z algorytmem rekurencyjnego podziału binarnego. W pierwszym kroku wybierana jest zmienna objaśniająca $X_{\cdot j}$ oraz wartość progowa s, które dzielą przestrzeń na dwa regiony:

$$R_1(j, s) = \{\mathbf{x}_i \mid X_{\cdot j} < s\}, \quad R_2(j, s) = \{\mathbf{x}_i \mid X_{\cdot j} \geq s\}. \tag{3.29}$$

Podział ten dobierany jest w taki sposób, aby minimalizował błąd wewnątrzgrupowy, co można zapisać jako:

$$\underset{j,s}{\operatorname{argmin}} \left[\sum_{i:\mathbf{x}_i \in R_1(j,s)} (y_i - \hat{c}_{R_1(j,s)})^2 + \sum_{i:\mathbf{x}_i \in R_2(j,s)} (y_i - \hat{c}_{R_2(j,s)})^2 \right]. \tag{3.30}$$

Następnie, algorytm rekurencyjnie poszukuje kolejnych podziałów w obrębie powstałych regionów, stosując tę samą zasadę minimalizacji błędu. Proces ten jest kontynuowany do momentu osiągnięcia określonego kryterium zatrzymania, którym może być minimalna liczba obserwacji w regionie końcowym, maksymalna głębokość drzewa lub próg poprawy jakości podziału. Wynikiem procedury jest struktura przypominająca drzewo, w której każdy węzeł wewnętrzny reprezentuje test warunkowy, a każdy liść (węzeł końcowy) - końcowy region R_m z przypisaną wartością predykcji c_m.

Zaletą drzew decyzyjnych jest ich przejrzystość i interpretowalność. Model można przedstawić graficznie w formie struktury drzewa, co ułatwia analizę czynników decydujących o przypisaniu obserwacji do danej klasy lub wartości prognozy. Drzewa decyzyjne nie wymagają również założeń dotyczących rozkładu zmiennych i dobrze radzą sobie z danymi o mieszanych typach (ilościowych i jakościowych).

Jednocześnie należy zauważyć, że pojedyncze drzewa decyzyjne są wrażliwe na niewielkie zmiany w danych uczących, co może prowadzić do niestabilności modelu. W celu przezwyciężenia tego ograniczenia opracowano techniki uczenia zespołowego (ang. ensemble methods), takie jak lasy losowe (ang. Random Forests) czy wzmacnianie gradientowe (ang. Gradient Boosting), które łączą wiele drzew decyzyjnych w celu poprawy stabilności i jakości predykcji.

3.1.4 Lasy Losowe

Las losowy (ang.Random Forest) został po raz pierwszy zaproponowany przez Breiman (2001) jako metoda zespołowa oparta na drzewach decyzyjnych. Odnosi się on do zbioru nieprzycinanych drzew regresyjnych

$$\{T(\mathbf{x}, \Theta_b),\ b = 1, 2, \ldots, B\}, \tag{3.31}$$

które są generowane na podstawie próbkowania bootstrapowego z oryginalnego zbioru danych treningowych. Oznacza to, że dla każdego drzewa b losowany jest zbiór uczący poprzez dobór próbek z powtórzeniami, co zapewnia różnorodność struktury poszczególnych drzew. W rezultacie można traktować $\{\Theta_b\}$ jako niezależne wektory losowe o identycznym rozkładzie.

Przyjmując B jako całkowitą liczbę drzew w lesie losowym, model dla zmiennej zależnej definiuje się jako średnią wartości uzyskanych z poszczególnych drzew regresyjnych:

$$f_B(\mathbf{x} \mid \Theta) = \frac{1}{B} \sum_{b=1}^{B} f(\mathbf{x} \mid \Theta_b). \tag{3.32}$$

Zgodnie z prawem wielkich liczb, gdy $B \to \infty$, spełniony jest warunek:

$$E_{\mathbf{x},y}\left[(y - f_B(\mathbf{x} \mid \Theta))^2\right] \longrightarrow E_{\mathbf{x},y}\left[(y - E_\Theta f(\mathbf{x} \mid \Theta))^2\right]. \tag{3.33}$$

W konstrukcji lasu losowego istotną rolę odgrywa dodatkowe losowanie podzbioru zmiennych objaśniających na każdym węźle drzewa, spośród których wybierana jest zmienna generująca najlepszy podział. Procedura ta ogranicza korelację pomiędzy drzewami i tym samym zmniejsza wariancję modelu zespołowego, zachowując przy tym jego niskie obciążenie.

Dzięki swojej strukturze las losowy charakteryzuje się wysoką dokładnością predykcji, odpornością na przeuczenie oraz możliwością efektywnego modelowania złożonych, nieliniowych zależności pomiędzy zmiennymi. Ponadto, umożliwia on ocenę istotności poszczególnych zmiennych, co czyni go użytecznym narzędziem w analizie danych, w tym w zastosowaniach ubezpieczeniowych i finansowych.

3.1.5 Wzmacnianie gradientowe

Algorytm wzmacniania gradientowego (ang. Gradient Boosting), opracowany przez Friedman (2001), polega na sekwencyjnym budowaniu wielu drzew regresyjnych. W każdej iteracji procesu konstrukcji modelu generowane jest nowe drzewo regresji z wykorzystaniem informacji pochodzących z wcześniej utworzonych drzew. Innymi słowy, każde kolejne drzewo uczone jest na podstawie błędów (reszt) uzyskanych w poprzednich iteracjach, co umożliwia stopniową poprawę dokładności modelu.

W wyniku tej procedury powstaje zbiór drzew

$$T_s(\mathbf{x}; \Theta_s), \quad s = 1, \ldots, S, \tag{3.34}$$

gdzie S oznacza liczbę iteracji (kroków wzmacniania). Model wzmacniania gradientowego można przedstawić jako sumę wyników poszczególnych drzew:

$$F_S(\mathbf{x} \mid \Theta) = \sum_{s=1}^{S} f_s(\mathbf{x} \mid \Theta_s), \tag{3.35}$$

gdzie $\Theta_s = \{R_{ms}, c_{ms}\}_{m=1}^{M_s}$, a $f_s(\mathbf{x} \mid \Theta_s)$ reprezentuje model utworzony przez drzewo $T_s(\mathbf{x}; \Theta_s)$.

Dla każdej iteracji s wyznaczane są optymalne parametry Θ_s poprzez rozwiązanie problemu minimalizacji funkcji straty:

$$\hat{\Theta}_s = \underset{\Theta_s}{\operatorname{argmin}} \sum_{i=1}^{N} L\left(y_i,\ F_{s-1}(\mathbf{x}_i \mid \Theta) + f_s(\mathbf{x}_i \mid \Theta_s)\right), \tag{3.36}$$

gdzie $L(y_i, \cdot)$ oznacza funkcję straty mierzącą różnicę pomiędzy wartością rzeczywistą a prognozowaną.

W kolejnych iteracjach model jest aktualizowany według zależności:

$$F_s(\mathbf{x}_i \mid \hat{\Theta}) = F_{s-1}(\mathbf{x}_i \mid \hat{\Theta}) + \sum_{m=1}^{M_s} \hat{c}_{ms} \mathbf{1}_{R_{ms}}(\mathbf{x}_i), \tag{3.37}$$

gdzie $\hat{c}_{ms}$ określane jest jako wartość minimalizująca funkcję straty w regionie R_{ms}:

$$\hat{c}_{ms} = \underset{c}{\operatorname{argmin}} \sum_{\mathbf{x}_i \in R_{ms}} L\left(y_i,\ F_{s-1}(\mathbf{x}_i \mid \Theta) + c\right). \tag{3.38}$$

Procedura ta prowadzi do iteracyjnego przybliżania funkcji celu poprzez kolejne dopasowywanie modeli do gradientu błędu funkcji straty, co stanowi istotę wzmacniania gradientowego. Każde kolejne drzewo minimalizuje błąd predykcji pozostały po poprzednich iteracjach, dzięki czemu model końcowy uzyskuje wysoką dokładność i zdolność generalizacji.

Algorytm wzmacniania gradientowego można interpretować jako iteracyjny proces przybliżania nieznanej funkcji zależności pomiędzy zmiennymi objaśniającymi a zmienną objaśnianą poprzez kolejne kroki w kierunku przeciwnym do gradientu funkcji straty. W każdej iteracji model uczony jest na tzw. pseudoresztach, czyli ujemnych wartościach pochodnych funkcji straty względem bieżących prognoz. W ten sposób każda kolejna iteracja koryguje błędy popełnione przez poprzednie drzewa, dążąc do minimalizacji globalnego błędu predykcji.

Z teoretycznego punktu widzenia metoda ta stanowi uogólnienie klasycznych technik gradientowych na przypadek uczenia modeli zespołowych, w których poszczególne drzewa regresji pełnią rolę prostych, lokalnych aproksymatorów. Dzięki temu, wzmacnianie gradientowe umożliwia modelowanie złożonych, nieliniowych zależności przy zachowaniu wysokiej stabilności i odporności na przeuczenie. Ponadto, metoda ta jest elastyczna pod względem wyboru funkcji straty, co pozwala na jej zastosowanie zarówno w problemach regresyjnych, jak i klasyfikacyjnych. W praktyce stanowi jedno z najskuteczniejszych narzędzi predykcyjnych wykorzystywanych w uczeniu maszynowym, w tym również w modelowaniu ryzyka ubezpieczeniowego.

3.2 Wykorzystanie uczenia maszynowego w ubezpieczeniach

Stosowanie metod uczenia maszynowego w sektorze ubezpieczeniowym sięga lat 80 XX wieku. Początkowo wykorzystywano klasyczne metody regresyjne, w tym regresję liniową oraz uogólnione modele liniowe (GLM), które stanowiły podstawę analizy aktuarialnej. W pracy (Dugas, Bengio, & Chapados, 2003) przedstawiono jedno z pierwszych systematycznych studiów nad zastosowaniem metod uczenia maszynowego w naukach aktuarialnych, obejmujących modele regresyjne, drzewa decyzyjne, sieci neuronowe oraz maszyny wektorów nośnych.

Dynamiczny rozwój uczenia maszynowego w ubezpieczeniach nastąpił jednak dopiero w ostatniej dekadzie, co wynikało z połączenia trzech kluczowych czynników:

- wprowadzenia bardziej zaawansowanych funkcji aktywacji w sieciach neuronowych,
- gwałtownego wzrostu dostępności dużych, różnorodnych zbiorów danych,
- rosnącej mocy obliczeniowej, w szczególności dzięki wykorzystaniu procesorów graficznych (ang. Graphics Processing Unit - GPU).

Obecnie uczenie maszynowe stanowi istotny element procesów analitycznych w zakładach ubezpieczeń, umożliwiając modelowanie złożonych zależności pomiędzy zmiennymi ryzyka, a także usprawnienie wyceny, taryfikacji, detekcji oszustw oraz szacowania rezerw techniczno-ubezpieczeniowych.

Jednym z głównych obszarów zastosowań uczenia maszynowego jest wycena a priori, której celem jest prognozowanie oczekiwanych kosztów związanych z nową umową ubezpieczeniową przy braku informacji o historii szkodowej klienta. W pracy (Sakthivel & Rajitha, 2017) zastosowano sieci neuronowe do przewidywania częstotliwości szkód a posteriori, wykorzystując dane o wcześniejszej częstotliwości roszczeń oraz współczynniki wiarygodności obliczone z wykorzystaniem teorii Bayesowskiej. Wynikiem modelu była estymacja rocznej częstotliwości szkód w kolejnym roku polisowym. Z kolei, Wüthrich (2019) oraz Wüthrich and Merz (2019) zaproponowali model Combined Actuarial Neural Network (CANN), który łączy klasyczny model GLM z siecią neuronową w celu uchwycenia nieliniowych zależności nieuwzględnionych w prostszych modelach parametrycznych.

Kolejnym istotnym kierunkiem rozwoju metod uczenia maszynowego w ubezpieczeniach jest wykorzystanie nieustrukturyzowanych danych telematycznych w procesie taryfikacji ubezpieczeń komunikacyjnych. Dane te, rejestrowane w wysokiej rozdzielczości (np. sekundowo), obejmują informacje o prędkości, przyspieszeniu, lokalizacji i stylu jazdy kierowcy. Boucher, Pérez-Marín, and Santolino (2013) wykazali, że relacja pomiędzy przebiegiem rocznym a częstotliwością szkód nie jest liniowa. Wraz z doświadczeniem kierowcy ryzyko wypadku maleje. W kolejnych pracach (Weidner & Wüthrich, 2017; Wüthrich, 2018b) zastosowano analizę skupień oraz metody redukcji wymiaru (ang. Principal Component Analysis – PCA) do grupowania kierowców na podstawie cech jazdy, takich jak rozkład prędkości i przyspieszeń. Gao and Wüthrich (2019) wykorzystali konwolucyjne sieci neuronowe (ang. Convolutional Neural Network – CNN) do klasyfikacji kierowców na podstawie krótkich fragmentów podróży, wykorzystując dane dotyczące prędkości, kąta skrętu i przyspieszenia. Pomimo ograniczonej liczby obserwacji uzyskano satysfakcjonujące wyniki klasyfikacji, potwierdzające potencjał metod głębokiego uczenia. W kolejnych badaniach (Gao, Meng, & Wüthrich, 2020) dowiedziono, że wykorzystanie map ciepła przyspieszeń i prędkości w konwolucyjnych sieciach neuronowych prowadzi do poprawy dokładności wyceny składek.

Zastosowanie metod uczenia maszynowego obejmuje również proces wyznaczania rezerw techniczno-ubezpieczeniowych, w którym tradycyjne modele często zawodzą ze względu na brak kompletności danych oraz ich nieustrukturyzowany charakter. (Wüthrich, 2018) zaproponował wykorzystanie płytkich sieci neuronowych do modelowania czynników rozwojowych w wyznaczaniu rezerw szkodowych. (Gabrielli, 2019) rozszerzyli klasyczny model Poissona na model uogólniony o strukturze sieci neuronowej, co umożliwiło wspólne modelowanie trójkątów szkód dla różnych linii biznesowych. W późniejszych pracach (Gabrielli, 2020; Lindholm, Lindskog, & Wahl, 2020). zastosowali bardziej elastyczne metody, takie jak sieci neuronowe i maszyny wzmacniania gradientowego (ang. Gradient Boosting Machine – GBM), do modelowania liczby i wielkości roszczeń. W modelach tych rezerwy ustala się w sposób dynamiczny i istnieje możliwość uwzględnienia współzależności między częstością a wysokością szkód.

W kontekście modelowania indywidualnych rezerw, Gabrielli and Wüthrich (2018) wykorzystali sieci neuronowe do prognozowania indywidualnych ścieżek rozwoju szkód w oparciu o historię portfela ubezpieczeniowego. Wüthrich (2018a) zastosował drzewa regresji do modelowania płatności, wykorzystując dane dotyczące zgłoszenia i przebiegu roszczenia. Kuo (2020) zaproponował model oparty na rekurencyjnych sieciach neuronowych (ang. Recurrent Neural Network – RNN) dla roszczeń zgłoszonych, lecz nieuregulowanych typu RBNS (ang. Reported But Not Settled - RBNS), w którym zastosowano architekturę kodera-dekodera (ang. Long Short-Term Memory) do generowania rozkładu przyszłych wypłat. Użycie bayesowskich sieci neuronowych w tym kontekście umożliwiło dodatkowo kwantyfikację niepewności prognoz.

Algorytmy uczenia maszynowego są również stosowane w modelach predykcyjnych dotyczących analizy ryzyka kredytowego, przewidywania częstotliwości szkód czy oceny wartości pojazdów. Quan and Valdez (2018) porównali różne miary jakości dopasowania, takie jak współczynnik determinacji R^2, indeks Giniego, błąd średniokwadratowy (MSE) oraz błąd procentowy (MAPE) dla modeli opartych na drzewach decyzyjnych. Bärtl and Krummaker (2020) wykorzystali natomiast drzewa decyzyjne, lasy losowe oraz sieci neuronowe do prognozowania wartości roszczeń w kontekście finansowania kredytów.

Znaczący wkład w popularyzację metod uczenia maszynowego w naukach aktuarialnych wniosły także instytucje branżowe. Szwajcarskie Stowarzyszenie Aktuariuszy (Schelldorfer & Wüthrich, 2019; Ferrario & Hämmerli, 2019) opublikowało serię samouczków wprowadzających do zastosowań metod analizy danych w kontekście ubezpieczeniowym. Z kolei Towarzystwo Aktuariuszy (ang. Society of Actuaries - SOA) oraz Kanadyjski Instytut Aktuariuszy (Society of Actuaries & Canadian Institute of Actuaries, 2019) prowadziły badania dotyczące predykcyjnego modelowania w ubezpieczeniach na życie, wskazując na rosnące znaczenie technik uczenia maszynowego w procesach taryfikacji i zarządzania ryzykiem.

Inne wykorzystanie algorytmów ML w sektorze ubezpieczeniowym przedstawia Tabela 3.1.

Pozycja	Zakres wykorzystania	Algorytm
(Chang & Lai, 2021)	Sieci ANN wykorzystywane do przewidywania skłonności konsumentów do zakupu polisy ubezpieczeniowej	ANN
(Fang et al., 2016)	Prognozowanie rentowności klienta ubezpieczeniowego	RF
(Desik & Behera, 2012)	Tworzenie reguł biznesowych na podstawie danych klientów	DT
(Neumann et al., 2019)	Przewidywanie decyzji klientów po wypadku	DT
(Dhieb et al., 2019)	Wykrywanie fałszywych roszczeń komunikacyjnych	XGBoost
(Grize et al., 2020)	Modele retencji i dynamiczna wycena online	CART, NN, XGBoost
(Corlosquet-Habart & Janssen, 2018)	Wykorzystanie big data w ubezpieczeniach	NN, RF, GBM, SVM
(Shen et al., 2021)	Oceny ryzyka kredytowego	ANN, SVM, CART
(Arora & Kaur, 2020)	Wybór cech w ocenie ryzyka kredytowego	SVM, DT, kNN
(Panigrahi et al., 2018)	Wykrywanie oszustw komunikacyjnych	DT, RF
(Rukhsar et al., 2022), (Kirlidog & Asuk, 2012), (Sundarkumar & Siddeshwar, 2015)	ML w oszustwach ubezpieczeniowych (zawyżone roszczenia, fikcyjne polisy itd.)	SVM, RF, DT, ANN, SVR
(Gramegna & Giudici, 2020)	Dlaczego klient kupuje/odwołuje polisę non-life	ANN

Tabela 3.1: Algorytmy uczenia maszynowego w sektorze ubezpieczeniowym.

Źródło: Opracowanie własne.

Podsumowując, rozwój metod uczenia maszynowego znacząco poszerzył możliwości analityczne w sektorze ubezpieczeniowym. Pozwolił na uwzględnienie danych heterogenicznych, nieliniowych zależności oraz efektów czasowych, których klasyczne modele aktuarialne nie były w stanie uchwycić. Zastosowanie tych metod przyczyniło się do poprawy dokładności wyceny, skuteczniejszej detekcji oszustw i bardziej precyzyjnego szacowania rezerw, stanowiąc tym samym istotny krok w kierunku nowoczesnego, opartego na danych zarządzania ryzykiem.

3.3 Kopule a uczenie maszynowe - wzajemne zastosowania

W prezentowanym rozdziale omówiono przykłady wykorzystania kopul w procesie redukcji wymiaru, zastosowania algorytmów uczenia maszynowego w estymacji kopul oraz użycia kopul do generowania danych syntetycznych przeznaczonych do trenowania modeli uczenia maszynowego. Przedstawione podejścia pozwalają lepiej uchwycić złożoną strukturę ryzyka oraz budować modele o wyższej dokładności predykcyjnej.

3.3.1 Kopule w redukcji wymiaru

Wraz z rozwojem nauki i technologii obserwuje się dynamiczny wzrost ilości danych, zarówno pod względem ich objętości, jak i wymiaru. Wysokowymiarowe zbiory danych zawierają dużą liczbę zmiennych, z których część może być wzajemnie zależna. W takich przypadkach zastosowanie znajduje redukcja wymiaru, rozumiana jako proces przekształcania zbioru danych w przestrzeń o niższym wymiarze przy minimalnej utracie informacji oraz przy zachowaniu istotnych właściwości strukturalnych danych.

Redukcja wymiaru umożliwia uproszczenie modeli, poprawę ich interpretowalności oraz skrócenie czasu uczenia algorytmów, a Ponadto, ogranicza ryzyko przeuczenia. W praktyce proces ten często polega na identyfikacji i eliminacji cech zależnych lub mało informatywnych, które nie wnoszą istotnego wkładu w model predykcyjny. W ostatnich latach wybór zmiennych objaśniających stał się jednym z kluczowych etapów procesu uczenia maszynowego, zarówno w fazie eksploracji danych, jak i w przygotowaniu danych do modelowania.

Do najczęściej stosowanych metod redukcji wymiaru zalicza się analizę głównych składowych (ang. Principal Component Analysis, PCA) wprowadzoną przez Karla Pearsona w 1901 r. (Pearson, 1901), sekwencyjną selekcję wsteczną i do przodu (Marill & Green, 1963; Whitney, 1971), regresję najmniejszych kątów (ang. Least Angle Regression – LARS) (Efron, Hastie, Johnstone, & Tibshirani, 2004), a także techniki oparte na algorytmach ewolucyjnych, w tym algorytmach genetycznych (Singh, Balamurugan, & Jebamalar Leavline, 2016). Szeroki przegląd metod i zastosowań redukcji wymiaru w kontekście współczesnych technik uczenia maszynowego przedstawiono w pracy (Rani, Khurana, Kumar, & Kumar, 2022), gdzie omówiono zarówno klasyczne podejścia liniowe, jak i nowoczesne metody nieliniowe wykorzystywane w analizie dużych zbiorów danych.

Z metodologicznego punktu widzenia techniki redukcji wymiaru można podzielić na dwie główne grupy: metody selekcji zmiennych objaśniających i metody ekstrakcji zmiennych objaśniających. W pierwszej grupie redukcja wymiaru polega na wyborze podzbioru istniejących zmiennych, które najlepiej reprezentują strukturę danych lub zmienną objaśnianą. Do metod tego typu zalicza się m.in. selekcję wsteczną, selekcję do przodu oraz algorytmy oparte na ocenie znaczenia zmiennych objaśniających (ang. feature importance). Druga grupa obejmuje metody ekstrakcji zmiennych objaśniających, w których nowe zmienne tworzone są jako kombinacje

liniowe lub nieliniowe wyjściowych zmiennych objaśniających, np. w analizie głównych składowych lub w metodach nieliniowych, takich jak t-SNE (ang. t-distributed Stochastic Neighbor Embedding) czy UMAP (ang. Uniform Manifold Approximation and Projection). Podział ten ma kluczowe znaczenie praktyczne, ponieważ wybór odpowiedniej techniki redukcji wymiaru zależy od charakteru danych, celu analizy oraz przyjętego modelu predykcyjnego.

W niniejszym podrozdziale uwagę skupiono na redukcji wymiaru z wykorzystaniem kopuli, której celem jest identyfikacja struktury zależności pomiędzy zmiennymi objaśniającymi. W artykule (Femmam & Femmam, 2022) przedstawiono metodę redukcji wymiaru z wykorzystaniem kopuli dwuwymiarowych. Jej celem jest identyfikacja i eliminacja silnych zależności pomiędzy zmiennymi objaśniającymi, tak aby uzyskać zredukowany zbiór danych, pozbawiony współzależności między zmiennymi. Metoda opiera się na estymacji kopuli Gaussa dla wszystkich par zmiennych objaśniających oraz na ocenie siły ich współzależności na podstawie współczynnika τ Kendalla.

Dane wejściowe tworzą macierz $X_{n\times d}$, w której n oznacza liczbę obserwacji, a d liczbę zmiennych objaśniających. Każdą kolumnę macierzy, odpowiadającą zmiennej objaśniającej X_j, przekształca się do postaci pseudo-obserwacji, wykorzystując rangi elementów:

$$U_{ij} = \frac{r_{ij}}{n+1}, \tag{3.39}$$

gdzie r_{ij} oznacza rangę obserwacji x_{ij} w kolumnie j, wyznaczaną jako numer porządkowy danej wartości po posortowaniu wszystkich obserwacji tej zmiennej w porządku rosnącym.

Takie przekształcenie zapewnia jednostajny rozkład wartości na przedziale $[0, 1]$ i pozwala modelować zależności między zmiennymi niezależnie od ich rozkładów brzegowych. Dla każdej pary zmiennych (X_i, X_j), gdzie $i, j \in \{1, 2, \ldots, d\}$ oraz $i \neq j$, budowana jest empiryczna kopula opisująca ich strukturę zależności:

$$\widehat{C}_n(u, v) = \frac{1}{n} \sum_{k=1}^{n} 1(U_{ki} \leq u,\ U_{kj} \leq v), \tag{3.40}$$

Na podstawie kopuli empirycznej obliczany jest współczynnik τ_{ij} Kendalla, a następnie odpowiadający mu teoretyczny parametr kopuli Gaussa α_{ij} zgodnie z:

$$\tau_{ij} = \frac{2}{\pi} \arcsin(\alpha_{ij}). \tag{3.41}$$

Wartość $|\alpha_{ij}|$ określa siłę zależności pomiędzy zmiennymi objaśniającymi X_i i X_j. Jeżeli $|\alpha_{ij}|$ przekracza ustalony próg (w pracy przyjęto wartość 0.5), jedna z tych zmiennych zostaje usunięta z macierzy danych X. Proces ten jest powtarzany iteracyjnie dla wszystkich par zmiennych, aż do momentu, gdy w zbiorze nie pozostaną zmienne silnie skorelowane. Ostatecznie uzyskuje się zredukowany zbiór danych $X_{n\times(d-p)}$, w którym kolumny reprezentują zmienne objaśniające niezależne lub słabo zależne, przy czym p oznacza liczbę usuniętych zmiennych. Zbiór ten zachowuje istotne informacje zawarte w danych wejściowych, eliminując jednocześnie nadmiarowość wynikającą z wysokiej zależności między zmiennymi. Metoda charakteryzuje się prostotą obliczeniową i skutecznością w redukcji silnych zależności, a jej jedynym ograniczeniem jest losowy wybór zmiennej do usunięcia w przypadku wystąpienia silnej korelacji pomiędzy dwiema zmiennymi objaśniającymi.

W artykule (Femmam, Brahimi, & Femmam, 2023) zaproponowano rozszerzenie opisanej powyżej metody, polegające na optymalizacji procesu selekcji zmiennych. W tym podejściu

wyeliminowano element losowości w wyborze usuwanej zmiennej poprzez wprowadzenie mechanizmu grupowania zmiennych objaśniających o zbliżonej strukturze zależności. Podobnie jak w poprzedniej metodzie, punktem wyjścia jest macierz danych $X_{n\times d}$. Dla każdej pary zmiennych (X_i, X_j), przy czym $i, j \in \{1, 2, \ldots, d\}$ oraz $i \neq j$, wyznacza się teoretyczną kopulę dwuwymiarową, najczęściej kopulę Gaussa, opisującą ich zależności. Następnie, obliczany jest parametr α_{ij}, jak w powyższej metodzie. Dla każdej zmiennej X_k, gdzie $k \in \{1, 2, \ldots, d\}$, tworzy się zbiór zmiennych silnie z nią skorelowanych:

$$G_k = \{X_j : |\alpha_{kj}| \geq 0.5,\ j \neq k\}, \tag{3.42}$$

gdzie G_k reprezentuje grupę zmiennych zależnych od X_k, a j indeksuje pozostałe zmienne porównywane z X_k. Wszystkie grupy tworzą rodzinę $\mathcal{G} = G_1, G_2, \ldots, G_d$. Spośród wszystkich utworzonych grup wybierana jest grupa mająca najwięcej elementów i oznaczana jest jako G_l. Wszystkie zmienne należące do tej grupy są usuwane z macierzy danych X oraz ze zbioru grup $\mathcal{G}$. Następnie, algorytm ponownie analizuje pozostałe zależności i kontynuuje proces eliminacji, aż w danych nie pozostaną żadne pary zmiennych, dla których $|\alpha_{ij}| \geq 0.5$. W wyniku działania algorytmu uzyskuje się nową macierz danych $X_{n\times(d-p)}$, w której kolumny stanowią zbiór zmiennych objaśniających o niskim poziomie współzależności, a p oznacza liczbę usuniętych zmiennych. Zredukowany zbiór danych zachowuje istotne informacje zawarte w danych wejściowych, eliminując jednocześnie nadmiarowość wynikającą z silnych korelacji między zmiennymi. Metoda charakteryzuje się prostą implementacją i niewielką złożonością obliczeniową, a dzięki zastosowaniu mechanizmu grupowania zapewnia większą stabilność i deterministyczność procesu selekcji cech w porównaniu z wcześniejszymi podejściami.

W pracy (Lall, Chowdhury, & Ghosh, 2021) zaproponowano podejście do selekcji zmiennych objaśniających z uwzględnieniem zmiennej objaśnianej y. Algorytm jednocześnie minimalizuje wzajemną informację pomiędzy wybranymi zmiennymi objaśniającymi oraz maksymalizuje zależność pomiędzy zmienną objaśnianą y a każdą ze zmiennych objaśniających X_i, gdzie $i \in \{1, 2, \ldots, d\}$. Dzięki temu, możliwe jest usunięcie zmiennych zależnych w zbiorze X i zachowanie jedynie tych, które wnoszą unikalną informację o zmiennej objaśnianej. Wzajemną informację pomiędzy dwiema zmiennymi losowymi X_i i y zdefiniowano na podstawie entropii kopuli:

$$I_c(X_i, y) = -H(C(u, v)), \tag{3.43}$$

gdzie $C(u, v)$ oznacza kopulę opisującą strukturę zależności pomiędzy zmiennymi X_i i y. Zależność ta wskazuje, że wzajemna informacja może być interpretowana jako ujemna entropia odpowiadającej jej kopuli - im mniejsza entropia, tym silniejsza zależność pomiędzy zmiennymi.

Dla zbioru zmiennych objaśniających $X = X_1, X_2, \ldots, X_d$ zadaniem jest wybór takiego podzbioru $S \subseteq X$ o liczności k, który maksymalizuje informacje względem zmiennej objaśnianej y, przy jednoczesnej minimalizacji redundancji pomiędzy zmiennymi w S. Zmienna objaśniająca X_i jest lepsza względem y niż zmienna X_j, jeśli:

$$I_c(X_i, Y) > I_c(X_j, Y). \tag{3.44}$$

Kolejnym etapem algorytmu jest iteracyjny proces selekcji zmiennych objaśniających, który ma na celu uzyskanie podzbioru zmiennych objaśniających o maksymalnej wartości informacyjnej względem zmiennej objaśnianej y oraz minimalnym poziomie wzajemnej redundancji. Przez s_j oznaczamy indeks zmiennej objaśniającej X_i, która została wybrana w j-tej iteracji algorytmu, natomiast zbiór $S_t = \{X_{s_1}, X_{s_2}, \ldots, X_{s_t}\}$ reprezentuje podzbiór zmiennych wybranych

po t iteracjach. W tym celu w każdej iteracji dokonywany jest wybór zmiennej X_i, która w największym stopniu zwiększa ilość informacji o y, przy jednoczesnym ograniczeniu zależności z już wybranymi zmiennymi. Algorytm rozpoczyna działanie od wyboru zmiennej X_{s_1}, dla której wzajemna informacja $I_c(X_i, y)$ przyjmuje największą wartość spośród wszystkich zmiennych objaśniających:

$$X_{s_1} = \arg\max_{X_i \in X} I_c(X_i, y). \tag{3.45}$$

Następnie, w kolejnych krokach wybierana jest zmienna $X_{s_{t+1}}$, która maksymalizuje różnicę pomiędzy istotnością względem zmiennej objaśnianej y a redundancją względem już wybranych zmiennych objaśniających:

$$X_{s_{t+1}} = \arg\max_{X_i \in (X \setminus S_t)} \Big[I_c(X_i, y) - I_c(X_i; X_{s_1}, X_{s_2}, \ldots, X_{s_t}) \Big], \tag{3.46}$$

gdzie $S_t = \{X_{s_1}, X_{s_2}, \ldots, X_{s_t}\}$ oznacza zbiór zmiennych wybranych po t iteracjach. Proces selekcji jest kontynuowany do momentu, aż liczność zbioru S_t osiągnie wartość k, co oznacza, że wybrano optymalny zestaw zmiennych:

$$S = \{X_{s_1}, X_{s_2}, \ldots, X_{s_k}\}. \tag{3.47}$$

3.3.2 Algorytmy uczenia maszynowego w estymacji kopul

W ostatnich latach w literaturze obserwuje się rosnące zainteresowanie wykorzystaniem algorytmów uczenia maszynowego w analizie zależności między zmiennymi losowymi. Metody te stanowią naturalne rozszerzenie klasycznych podejść statystycznych, umożliwiając modelowanie złożonych, nieliniowych zależności bez konieczności przyjmowania silnych założeń parametrycznych. Jak opisano w rozdziale drugim, do opisu struktury zależności wykorzystywane są kopule, w szczególności kopule Archimedesa, które pozwalają na modelowanie zależności dwuwymiarowych. W przypadku rozkładów wielowymiarowych stosuje się kaskady kopul, które dekomponują rozkład wielowymiarowy na zestaw kopul dwuwymiarowych. Zarówno kopuli Archimedesa, jak i kaskady kopul charakteryzują się ograniczoną elastycznością. Kształt zależności jest z góry określony przez wybraną rodzinę kopul oraz jej parametry. W efekcie modele te mogą nie być w stanie uchwycić bardziej złożonych struktur zależności obserwowanych w danych rzeczywistych. Zastosowanie algorytmów uczenia maszynowego pozwala przezwyciężyć to ograniczenie. Modele te umożliwiają bezpośrednie odwzorowanie struktury zależności na podstawie danych, bez konieczności definiowania postaci kopuli parametrycznej. Dzięki temu, możliwe jest uchwycenie nieregularnych, asymetrycznych zależności, które wykraczają poza możliwości klasycznych kopul parametrycznych, ponieważ dana rodzina i dany parametr wybranej kopuli nie są w stanie opisać określonej struktury zależności. Właściwości modeli wynikające z własności kopul łączą się z elastycznością modeli uczenia maszynowego. Metody te są konstruowane w taki sposób, aby odzwierciedlały własności kopuli poprzez odpowiednią architekturę modelu lub właściwą konstrukcję funkcji kosztu. W rezultacie otrzymuje się model, który zachowuje interpretowalność dla teorii kopul, a zarazem wykorzystuje zdolność algorytmów uczenia maszynowego do aproksymacji i estymacji dowolnych nieliniowych struktur zależności. Jednym z kierunków łączenia teorii kopul z metodami uczenia maszynowego jest wykorzystanie głębokich sieci neuronowych do przybliżenia generatora kopuli Archimedesa φ. Pozwala to ująć proces uczenia jako aproksymację funkcji φ spełniającej określone warunki monotoniczności i dodatniości, co zapewnia zachowanie interpretowalnej struktury kopuli przy

jednoczesnym wykorzystaniu elastyczności sieci neuronowych. Parametryzacja funkcji generatora dla kopul Archimedesa została ograniczona do prostych postaci, ponieważ skomplikowane funkcje generatora prowadzą do trudności w obliczeniu gęstości kopuli - składnika niezbędnego do oszacowania największej wiarygodności. Kopule Archimedesa buduje się z funkcji generatora φ,, które są funkcjami monotonicznymi. Wprost z twierdzenia Bernsteina (Duchon, 2011) wynika, iż φ jest monotoniczna wtedy i tylko wtedy, gdy φ jest transformatą Laplace'a dodatniej zmiennej losowej M:

$$\varphi(t) = \int_0^\infty e^{-ts} dF_M(s) = \mathbb{E}_M\left(\exp\left(-tM\right)\right) \tag{3.48}$$

Twierdzenie Bernsteina wskazuje, że funkcje generujące φ można aproksymować skończoną sumą funkcji wykładniczych, co stanowi praktyczne przybliżenie kopul Archimedesa. Jeśli wektor $U = (U_1, \ldots, U_d)$ ma rozkład opisany przez kopułę C generowaną przez funkcję φ, będącą transformatą Laplace'a nieujemnej zmiennej losowej M, wówczas każdą U_i można zapisać jako

$$U_i = \varphi\left(\frac{E_i}{M}\right), \tag{3.49}$$

gdzie

$$E_i \sim \mathrm{Exp}(1). \tag{3.50}$$

W tym ujęciu M stanowi wspólny czynnik losowy, odpowiedzialny za zależność między zmiennymi U_i, natomiast E_i są niezależnymi zmiennymi wykładniczymi, które wprowadzają indywidualną losowość każdej U_i.

Ideę tę rozwinięto w pracy (Ling, Fang, & Kolter, 2020), w której zaproponowano sieć neuronową Archimedean Copula Network (ACNet), która będzie oznaczana jako φ^{nn}. Jest to sieć neuronowa reprezentująca funkcję generatora kopuli Archimedesa, przy użyciu dużej, lecz skończonej mieszanki funkcji wykładniczych. Architektura sieci φ^{nn} wygląda podobnie do standardowej sieci typu feedforward, z tą różnicą, że dane wyjściowe każdej warstwy stanowią wypukłą kombinację skończonej liczby ujemnych funkcji wykładniczych. Sieć została zaprojektowana w taki sposób, aby zachowywać kluczowe własności funkcji generatora kopuli Archimedesa, w szczególności jej monotoniczność. Każda warstwa sieci realizuje dwa etapy obliczeń: najpierw tworzy wypukłą kombinację wyników z warstwy poprzedniej, a następnie mnoży uzyskany wynik przez czynnik wykładniczy $\exp(-B_i^{(l)} t)$, gdzie $B_i^{(l)} > 0$ jest parametrem decydującym o tempie zanikania składnika wykładniczego. Wagi $A_{ij}^{(l)}$ są nieujemne i znormalizowane tak, że $\sum_j A_{ij}^{(l)} = 1$, co gwarantuje, że każda warstwa generuje poprawną wypukłą kombinację. Struktura sieci może być opisana rekurencyjnie zależnością:

$$\varphi_i^{nn(l)}(t) = e^{-B_i^{(l)}(t)} \sum_{j=1}^{H_{l-1}} A_{ij}^{(l)} \, \varphi_j^{nn(l-1)}(t), \qquad l = 1, \ldots, L, \tag{3.51}$$

gdzie $\varphi_1^{nn(0)}(t) = 1$ oraz H_l to liczba neuronów w warstwie l. Wyjście całej sieci obliczane jest zgodnie ze wzorem:

$$\varphi^{nn}(t) = \sum_{j=1}^{H_L} A_{1j}^{(L+1)} \, \varphi_j^{nn(L)}(t). \tag{3.52}$$

Parametry $A_{ij}^{(l)}$ i $B_i^{(l)}$ nie są klasycznymi wagami połączeń neuronowych $w_{ij}^{(l)}$. Współczynniki $A_{ij}^{(l)}$ określają wagi, ale wewnątrz wypukłej kombinacji, które muszą być nieujemne oraz sumować się do jedności. natomiast $B_i^{(l)}$ kontrolują tempo zanikania poszczególnych funkcji wykładniczych w warstwie l. Takie ograniczenia zapewniają, że sieć zachowuje własności monotoniczności wymagane dla generatora kopuli Archimedesa. Dzięki takiej konstrukcji cała sieć definiuje skończoną mieszaninę funkcji wykładniczych, która z definicji jest monotoniczna i tym samym spełnia warunki bycia poprawnym generatorem kopuli Archimedesa. Wartość kopuli $C(u_1, \ldots, u_d)$ obliczana jest zgodnie ze wzorem:

$$C(u) = \varphi^{-1}\left(\sum_i \varphi(u_i)\right), \tag{3.53}$$

przy czym odwrotność funkcji φ^{-1} dla generatora neuronowego φ^{nn} nie ma postaci analitycznej i jest wyznaczana numerycznie z wykorzystaniem metody Newtona.

W kolejnym podejściu do modelowania kopul Archimedesa Ng, Hasan, Elkhalil, and Tarokh (2021) zaproponowali koncepcję generatywnych kopul Archimedejskich (ang. Generative Archimedean Copulas – GAC). W odróżnieniu od sieci ACNet, w której bezpośrednio aproksymowano funkcję generatora φ przy pomocy mieszaniny funkcji wykładniczych, model GAC parametryzuje rozkład zmiennej losowej M, której transformata Laplace'a definiuje generator kopuli. Idea metody podobnie opiera się na wykorzystaniu twierdzenia Bernsteina, zgodnie z którym każda monotoniczna funkcja φ może być przedstawiona jako transformata Laplace'a pewnej nieujemnej zmiennej losowej M, zgodnie ze wzorem 3.37. Zatem zamiast bezpośrednio przybliżać φ, autorzy proponują uczenie sieci neuronowej (dalej oznaczanej jako $G(\varepsilon; \theta)$), która generuje próbki M z rozkładu F_M. Dzięki temu, zachowane zostają teoretyczne własności funkcji generatora, a model zyskuje dużą elastyczność w odwzorowywaniu różnorodnych zależności nieliniowych pomiędzy zmiennymi. Sieć $G(\varepsilon; \theta)$ pełni rolę modelu generatywnego. Jej zadaniem jest przekształcenie rozkładu $\varepsilon \sim \mathcal{N}(0, 1)$ (lub $\varepsilon \sim U(0, 1)$) w rozkład dodatniej zmiennej $M = G(\varepsilon; \theta)$. Aby zapewnić nieujemność generowanej zmiennej M, na wyjściu sieci zastosowano funkcję aktywacji wykładniczej:

$$M = \exp(h_\theta(\varepsilon)), \tag{3.54}$$

gdzie $h_\theta(\varepsilon)$ oznacza wyjście ostatniej warstwy ukrytej. Dzięki temu, każda próbka M jest dodatnia, co jest warunkiem koniecznym dla poprawnej definicji transformaty Laplace'a. Taka architektura jest znacznie prostsza niż konstrukcja ACNet, lecz jej rola jest fundamentalna. Rozkład zmiennej M, a nie sama funkcja φ, jest bezpośrednio modelowany przez sieć. Oznacza to, że sieć $G(\varepsilon; \theta)$ nie musi spełniać żadnych dodatkowych ograniczeń dotyczących monotoniczności czy wypukłości, ponieważ te własności są automatycznie zapewnione przez fakt, że φ jest transformatą Laplace'a nieujemnej zmiennej losowej. Parametry sieci θ są uczone w taki sposób, aby rozkład $M = G(\varepsilon; \theta)$ prowadził do generatora φ najlepiej dopasowanego do danych empirycznych. Autorzy rozważają trzy metody treningu sieci: metodę największej wiarygodności (ang. Maximum Likelihood Estimation – MLE), polegająca na maksymalizacji logarytmu gęstości kopuli, dopasowanie Craméra–von Misesa oraz uczenie kontradyktoryczne. Model GAC można rozszerzyć na przypadek hierarchicznych kopul Archimedesa (HAC), w których różne grupy zmiennych opisują różne poziomy zależności.

Jak opisano w podrozdziale 2.1, kaskady kopul są elastyczniejszym narzędziem w modelowaniu struktur zależności niż kopule Archimedesa. W pracy (Sun, Cuesta-Infante, & Veerama-

chaneni, 2019) zaproponowano podejście do automatycznego uczenia struktury kaskad kopul z wykorzystaniem metod uczenia maszynowego. Ideą metody jest konstrukcja kaskad kopul jako sekwencyjnego problemu decyzyjnego, który można sformułować w ramach uczenia przez wzmacnianie. W takim ujęciu proces uczenia agenta przypomina stopniowe „rozrastanie się" drzewa kopul, w którym każda decyzja o połączeniu wierzchołków wpływa na przyszłe możliwe wybory oraz końcową jakość dopasowania modelu. Dzięki temu, możliwe jest jednoczesne poznanie struktury drzewa kaskad kopul oraz wybór dwuwymiarowych kopul składowych w sposób zautomatyzowany i globalnie optymalny. Kaskada kopul może być opisana jako sekwencja zagnieżdżonych drzew, w których krawędzie odpowiadają dwuwymiarowym kopulom warunkowym. W zaproponowanym modelu zamiast tego proponuje się sformułowanie procesu budowy drzewa jako sekwencyjnego problemu decyzyjnego, w którym wybierane są kolejne krawędzie (i, j), maksymalizując globalną funkcję celu.

Proces konstrukcji kaskady kopul można sprowadzić do sekwencyjnych decyzji, w których w kolejnych krokach wybierane są połączenia między zmiennymi losowymi. Stan układu w chwili t opisuje aktualną konfigurację drzewa:

$$s_t = (E_t, \mathcal{N}_t^L, \mathcal{N}_t^R), \tag{3.55}$$

gdzie E_t oznacza zbiór istniejących krawędzi, $\mathcal{N}_t^L$ węzły już połączone, a $\mathcal{N}_t^R$ węzły jeszcze niepołączone. Na każdym etapie wybierana jest para wierzchołków:

$$a_t = (i, j), \quad \text{gdzie } i \in \mathcal{N}_t^L,\ j \in \mathcal{N}_t^R, \tag{3.56}$$

która tworzy nową krawędź w drzewie. Po dodaniu połączenia zbiór krawędzi i węzłów jest aktualizowany zgodnie z zasadą:

$$E_{t+1} = E_t \cup \{(i, j)\}, \quad \mathcal{N}_{t+1}^L = \mathcal{N}_t^L \cup \{j\}, \quad \mathcal{N}_{t+1}^R = \mathcal{N}_t^R \setminus \{j\}. \tag{3.57}$$

W ten sposób struktura drzewa rozwija się iteracyjnie, aż do momentu, gdy wszystkie zmienne zostaną połączone w kompletną konfigurację kaskady kopul. W każdym kroku ocenia się wpływ dodania nowej krawędzi na jakość dopasowania modelu. W tym celu definiuje się funkcję nagrody, mierzącą przyrost logarytmu wiarygodności danych przy przejściu ze stanu s_{t-1} do stanu s_t:

$$R_t = L(x, s_t) - L(x, s_{t-1}), \tag{3.58}$$

gdzie $L(x, s_t)$ oznacza wartość funkcji log-wiarygodności dla obserwacji x przy aktualnej strukturze drzewa s_t. Wartość nagrody jest tym większa, im bardziej nowo dodane połączenie poprawia dopasowanie modelu do danych empirycznych. W celu ograniczenia przeuczenia, całkowita nagroda może dodatkowo uwzględniać składnik kary za zbyt złożoną lub gęstą strukturę:

$$R = R_{\text{likelihood}} + \lambda R_{\text{penalty}}, \tag{3.59}$$

gdzie współczynnik λ kontroluje równowagę między jakością dopasowania a prostotą modelu. Tak zdefiniowana nagroda kieruje procesem konstrukcji w stronę modeli dobrze dopasowanych do danych, a jednocześnie o ograniczonej złożoności. Celem procesu uczenia jest maksymalizacja oczekiwanej sumy nagród uzyskanych w trakcie konstrukcji całego drzewa kaskad kopul. Funkcję celu zapisuje się jako:

$$J(\phi) = \mathbb{E}_{\tau \sim \pi_\phi} \left[\sum_{t=1}^{T} R_t \right], \tag{3.60}$$

gdzie:

- ϕ oznacza wektor parametrów (wag) sieci neuronowej,
- $\pi_\phi(a_t|s_t)$ jest rozkładem prawdopodobieństwa wyboru akcji a_t w stanie s_t, określanym przez bieżące parametry ϕ,
- $\tau = \{(s_t, a_t)\}_{t=1}^{T}$ reprezentuje całą sekwencję decyzji podjętych w procesie budowy drzewa,
- R_t to nagroda uzyskana w kroku t, zdefiniowana wcześniej jako przyrost log-wiarygodności modelu po dodaniu nowej krawędzi,
- T oznacza łączną liczbę kroków niezbędnych do utworzenia pełnego drzewa kopul.

Rozkład π_ϕ jest modelowany przez sieć neuronową, która uczy się wybierać kolejne połączenia między zmiennymi w sposób maksymalizujący funkcję celu $J(\phi)$. Parametry sieci są aktualizowane zgodnie z zasadą, która pozwala obliczyć gradient funkcji celu względem ϕ:

$$\nabla_\phi J(\phi) = \mathbb{E}_{\tau\sim\pi_\phi}\left[\sum_{t=1}^{T} R_t \nabla_\phi \log \pi_\phi(a_t|s_t)\right]. \tag{3.61}$$

W ten sposób model stopniowo zwiększa prawdopodobieństwo wyboru tych połączeń (i, j), które w największym stopniu poprawiają dopasowanie struktury kopuli do danych empirycznych. Rozkład prawdopodobieństwa wyboru akcji $\pi_\phi(a_t|s_t)$ jest modelowany za pomocą sieci neuronowej, której zadaniem jest wyznaczanie kolejnych połączeń (i, j) w procesie budowy drzewa. Sieć ta zawiera kryterium wyboru i przyjmuje na wejściu wektor reprezentujący aktualny stan s_t, a następnie generuje rozkład prawdopodobieństwa możliwych akcji a_t. Na tej podstawie wybierane jest połączenie, które maksymalizuje oczekiwaną nagrodę zdefiniowaną w funkcji celu $J(\phi)$. Ponieważ decyzje podejmowane na wczesnych etapach konstrukcji drzewa wpływają na dostępne połączenia w kolejnych krokach, proces ten nie spełnia własności Markowa. Oznacza to, że bieżący stan s_t zależy nie tylko od poprzedniego stanu s_{t-1}, lecz również od całej sekwencji wcześniejszych decyzji $\tau = \{(s_1, a_1), (s_2, a_2), \ldots, (s_t, a_t)\}$. Aby uchwycić tę długookresową zależność, stosuje się sieć rekurencyjną z długą i krótką pamięcią (ang. Long Short-Term Memory – LSTM). W każdym kroku t wektor ukryty sieci LSTM, oznaczany jako h_t, jest aktualizowany zgodnie z zależnością:

$$h_t = \mathrm{LSTM}(s_t, h_{t-1}), \tag{3.62}$$

gdzie h_{t-1} reprezentuje stan pamięci z poprzedniego kroku, a s_t jest wektorem cech opisujących bieżącą konfigurację drzewa kopuli. Na podstawie wektora h_t wyznaczany jest rozkład prawdopodobieństwa wyboru akcji:

$$\pi_\phi(a_t|s_t) = \mathrm{softmax}(Wh_t + b), \tag{3.63}$$

gdzie W i b oznaczają odpowiednio macierz wag oraz wektor przesunięcia warstwy wyjściowej.

Zastosowanie architektury LSTM umożliwia sieci zapamiętywanie kontekstu wcześniejszych decyzji i uwzględnianie ich wpływu przy wyborze kolejnych połączeń. Dzięki temu, budowa drzewa przebiega w sposób spójny, a wyuczona struktura lepiej odwzorowuje złożone zależności między zmiennymi losowymi.

Po wyznaczeniu pełnej struktury drzewa kaskady kopul, w kolejnym etapie dobierane są odpowiednie kopule dwuwymiarowe dla każdej krawędzi $(i, j) \in E_T$. Celem tego etapu jest oszacowanie parametrów θ_{ij} oraz wybór rodziny kopuli C_{ij}, która najlepiej opisuje zależność pomiędzy zmiennymi X_i i X_j, ewentualnie warunkowo względem zbioru zmiennych $D(i, j)$, połączonych na wcześniejszych poziomach drzewa. Dla każdej krawędzi (i, j) maksymalizowana jest lokalna funkcja log-wiarygodności:

$$\hat{C}_{ij}(\hat{\theta}_{ij}) = \arg\max_{\theta_{ij}} \log L_{ij}(x_i, x_j; \theta_{ij}), \tag{3.64}$$

gdzie $L_{ij}(x_i, x_j; \theta_{ij})$ oznacza wartość funkcji wiarygodności dla obserwacji (x_i, x_j), a θ_{ij} jest wektorem parametrów odpowiadającym danej rodzinie kopul. Rodzina kopuli C_{ij} wybierana jest spośród najczęściej stosowanych, takich jak kopula Claytona, Gumbela, Franka czy t-Studenta, w zależności od kierunku i siły zależności między zmiennymi oraz charakteru zależności ogonowych.

W artykule (Janke & Steinke, 2021) przedstawiono metodę niejawnych kopuł generatywnych (ang. Implicit Generative Copulas – IGC), umożliwiającą elastyczne i nieparametryczne modelowanie struktury zależności w danych poprzez uczenie rozkładu gęstości kopuli. Kluczowym założeniem metody jest rozdzielenie modelowania zależności wielowymiarowych od modelowania rozkładów brzegowych.

Pierwszym krokiem metody IGC jest nauczenie tzw. wewnętrznego (latentnego) rozkładu Q_θ, który odzwierciedla strukturę zależności występującą w danych, lecz bez uwzględnienia rozkładów brzegowych. Oznacza to, że na tym etapie model koncentruje się wyłącznie na relacjach pomiędzy zmiennymi losowymi, pomijając ich indywidualne rozkłady brzegowe. W tym celu wykorzystuje się generatywną sieć neuronową g_θ, której zadaniem jest przekształcenie prostego, znanego rozkładu bazowego w bardziej złożony rozkład w przestrzeni latentnej. Jako rozkład bazowy przyjmuje się standardowy rozkład normalny:

$$z \sim \mathcal{N}(0, I), \tag{3.65}$$

gdzie $z \in \mathbb{R}^K$ jest wektorem losowym o niezależnych składowych, a I oznacza macierz jednostkową. Następnie, próbki z są przekształcane przez funkcję:

$$y = g_\theta(z), \tag{3.66}$$

gdzie $g_\theta : \mathbb{R}^K \to \mathbb{R}^d$ to parametryzowana sieć neuronowa o wagach θ, a $y \in \mathbb{R}^d$ stanowi próbkę z rozkładu Q_θ. Sieć g_θ uczy się w ten sposób nieliniowej transformacji, która odwzorowuje proste próbki z rozkładu normalnego w próbki o złożonej strukturze zależności, podobnej do tej, która występuje w danych rzeczywistych. Rozkład Q_θ można interpretować jako wewnętrzną reprezentację danych, w której uchwycona jest struktura zależności między zmiennymi, lecz bez wymuszania jednostajnych rozkładów brzegowych. Etap ten stanowi fundament IGC, ponieważ uczy on sieć sposobu, w jaki zmienne są ze sobą powiązane w przestrzeni latentnej, zanim nałożony zostanie warunek jednostajnych rozkładów w kolejnym kroku prowadzącym do konstrukcji kopuli Q_θ^C.

Drugim krokiem jest przekształcenie wewnętrznego rozkładu Q_θ w taki sposób, aby otrzymać jednostajne rozkłady brzegowe na przedziale $[0, 1]$. Dzięki temu, uzyskany rozkład można interpretować jako kopulę, czyli model opisujący strukturę zależności między zmiennymi, niezależnie od rozkładów brzegowych. Osiąga się to poprzez zastosowanie transformacji (ang.

Probability Integral Transform – PIT), polegającej na użyciu dystrybuanty empirycznej do każdej składowej y_i wektora $y = g_\theta(z)$:

$$v_i = F_{Y_i}(y_i), \quad i = 1, \ldots, d, \tag{3.67}$$

gdzie F_{Y_i} oznacza dystrybuantę empiryczną zmiennej Y_i, a $v_i \in [0, 1]$. W wyniku tego przekształcenia powstaje wektor:

$$v = (v_1, \ldots, v_d), \tag{3.68}$$

którego każda składowa ma jednostajny rozkład brzegowy, natomiast zależności między nimi pozostają niezmienione. Wektor v reprezentuje zatem próbkę pochodzącą z rozkładu kopuli Q_θ^C, określonego przez wewnętrzny model g_θ oraz transformację F_Y:

$$v = F_Y(g_\theta(z)), \quad v \sim Q_\theta^C. \tag{3.69}$$

Kopula Q_θ^C opisuje więc jedynie zależności pomiędzy zmiennymi, przy zachowaniu jednostajnych rozkładów brzegowych. Celem uczenia modelu jest takie dostrojenie parametrów θ, aby generatywna kopula Q_θ^C jak najdokładniej odwzorowywała rzeczywistą strukturę zależności obecną w danych, opisaną przez empiryczną kopulę P^C. Sieć g_θ uczy się odwzorowania, które zachowuje zależności pomiędzy zmiennymi, a transformacja PIT sprawia, że te zależności zostają przedstawione w przestrzeni kopuli, w której każda zmienna ma rozkład jednostajny. Proces uczenia IGC polega na dopasowaniu rozkładu kopuli generowanej przez model Q_θ^C do kopuli empirycznej P^C, opisującej rzeczywistą strukturę zależności w danych. Kopula empiryczna P^C jest uzyskiwana na podstawie rzeczywistych danych po transformacji PIT, dzięki czemu wszystkie jej rozkłady brzegowe są jednostajne na przedziale $[0, 1]$. Ponieważ celem jest odwzorowanie jedynie zależności między zmiennymi, uczenie przebiega w przestrzeni $[0, 1]^d$ po transformacji PIT, w której wszystkie rozkłady brzegowe są jednostajne. Do pomiaru różnicy pomiędzy rozkładami Q_θ^C i P^C stosowana jest odległość energetyczna, będącą nieparametryczną miarą rozbieżności między rozkładami prawdopodobieństwa. W przypadku dwóch rozkładów P i Q oznaczono ją jako $D_E^2(P, Q)$. Miara energetyczna została opisana w podrozdziale 5.1.3. Wartość $D_E(P, Q)$ jest równa zeru wtedy i tylko wtedy, gdy rozkłady P i Q są identyczne; dlatego minimalizacja tej odległości prowadzi do dopasowania modelu generatywnego do danych. W praktyce, podczas uczenia modelu IGC minimalizuje się empiryczną postać tej miary:

$$\mathcal{L}_{\text{energy}}(\theta) = D_E^2(P^C, Q_\theta^C), \tag{3.70}$$

co odpowiada minimalizacji średniej odległości pomiędzy próbkami generowanymi przez model a próbkami pochodzącymi z danych rzeczywistych. Kluczowym elementem skutecznego uczenia metodami gradientowymi jest zapewnienie różniczkowalności całego procesu przekształceń. Klasyczna transformacja PIT wymaga operacji sortowania, która nie jest różniczkowalna, co uniemożliwia propagację gradientów przez model. Aby rozwiązać ten problem, autorzy wprowadzili warstwę soft-rank, będącą gładką aproksymacją sortowania. Warstwa ta przypisuje elementom ich rangi w sposób ciągły, umożliwiając obliczanie przybliżonych rang w formie różniczkowalnej. Zastosowanie warstwy soft-rank pozwala na propagację gradientów przez cały model, dzięki czemu parametry sieci g_θ mogą być skutecznie aktualizowane metodami gradientowymi. W rezultacie możliwe jest optymalizowanie funkcji celu $\mathcal{L}_{\text{energy}}(\theta)$ w sposób stabilny i spójny z zasadami uczenia głębokiego.

Po zakończeniu procesu uczenia model IGC może być wykorzystywany do generowania nowych próbek, które zachowują strukturę zależności charakterystyczną dla danych rzeczywistych. Generowanie odbywa się w kilku etapach, odpowiadających kolejnym transformacjom modelu:

1. Losowana jest próbka $z \sim \mathcal{N}(0, I)$ z rozkładu bazowego o niezależnych składowych.

2. Wektor z jest przekształcany przez wyuczoną sieć neuronową g_θ, co daje wektor $y = g_\theta(z)$, będący próbką z wewnętrznego rozkładu Q_θ.

3. Na każdą składową y_i nakładana jest empiryczna dystrybuanta $\widehat{F}_{Y_i}$, aby uzyskać jednostajny rozkład brzegowy:
$$v_i = \widehat{F}_{Y_i}(y_i), \quad i = 1, \ldots, d. \tag{3.71}$$
Funkcja $\widehat{F}_{Y_i}$ oznacza empiryczną dystrybuantę oszacowaną na podstawie próbek y_i generowanych przez model g_θ, a więc stanowi przybliżenie nieznanej dystrybuanty prawdziwego rozkładu zmiennej Y_i. W wyniku tego powstaje wektor $v = (v_1, \ldots, v_d)$, który leży w jednostkowej kostce $[0, 1]^d$ i stanowi próbkę z wyuczonej kopuli Q_θ^C.

Proces ten można zapisać jako odwzorowanie:

$$v = T_\theta(z) = F_Y(g_\theta(z)), \tag{3.72}$$

gdzie T_θ jest różniczkowalną transformacją przekształcającą próbki z przestrzeni rozkładu normalnego w przestrzeń kopuli, a $F_Y = (F_{Y_1}, \ldots, F_{Y_d})$ oznacza wektor empirycznych dystrybuant zastosowanych do poszczególnych składowych. Transformacja T_θ reprezentuje złożenie funkcji generatywnej g_θ oraz transformacji PIT za pomocą wektora dystrybuant F_Y.

Jeżeli celem jest uzyskanie próbek w oryginalnej przestrzeni danych, wówczas na każdą składową można zastosować odwrotność dystrybuanty empirycznej $F_{X_i}^{-1}$, otrzymując:

$$x_i = F_{X_i}^{-1}(v_i), \quad i = 1, \ldots, d. \tag{3.73}$$

Dzięki temu, generowane dane $x = (x_1, \ldots, x_d)$ posiadają takie same rozkłady brzegowe jak dane rzeczywiste, przy jednoczesnym zachowaniu ich wzajemnych zależności opisanych przez wyuczoną kopulę Q_θ^C.

3.3.3 Kopule w generowaniu danych dla algorytmów uczenia maszynowego

Kopule są również wykorzystywane do generowania syntetycznych danych w technikach uczenia maszynowego. Wraz z rozwojem metod analizy opartych na danych pojawia się problem ich słabej dostępności, ograniczonej jakości oraz ochrony prywatności. Dane syntetyczne częściowo łagodzą te wyzwania, umożliwiając skalowanie zbiorów danych treningowych bez utraty ważnych właściwości statystycznych danych rzeczywistych. Jest to szczególnie ważne w trzech przypadkach:

- kiedy instytucje finansowe dysponują niewielką ilością danych historycznych ze względu na wysokie koszty pozyskania lub niekompletne zapisy z przeszłości,

- występują znaczne różnice między dostępnymi danymi, co powoduje, że metody uczenia maszynowego nie dokonują dokładnych prognoz,

- jeśli instytucje nie mają możliwości bezpośredniej wymiany swoich danych, informacje są poufne lub wrażliwe.

Problemy nasilają się w badaniach ubezpieczeniowych i aktuarialnych. Firmy ubezpieczeniowe mogą mieć ograniczone dane dotyczące niektórych produktów lub klientów, co utrudnia konstruowanie wiarygodnych modeli prognostycznych. W innych przypadkach dane historyczne są częściowe, ponieważ stare systemy informatyczne są zastępowane przez nowe, firmy dokonują połączeń lub występują zmiany regulacyjne. Ponadto, zarówno zbiory danych dotyczące roszczeń, jak i polis mogą być sporadycznie "zaszumione"lub przestarzałe, co pogarsza ich jakość. Większość danych związanych z ubezpieczeniami, takich jak dane dotyczące stanu zdrowia, dochodów oraz historia roszczeń, jest prywatna, a ich wykorzystanie jest ściśle regulowane surowymi przepisami (np. RODO), które ograniczają naszą zdolność do swobodnego ujawniania danych do celów badawczych. W takiej sytuacji wytwarzanie danych syntetycznych przy użyciu kopul stanowi odpowiednie rozwiązanie, ponieważ realistyczne zależności między zmiennymi można odtworzyć, aby uzyskać dane, które są bezpieczne do pracy w modelowaniu ryzyka, procesach wyceny czy przewidywaniu roszczeń.

W (Sei, Onesimu, & Ohsuga, 2022) przedstawiono metodę wstępnego przetwarzania, która generuje syntetyczny zbiór danych na bazie kopul, równocześnie redukując szum. Eksperymenty na danych syntetycznych i rzeczywistych z użyciem drzew decyzyjnych, głębokich sieci neuronowych, k-najbliższych sąsiadów oraz maszyn wektorów nośnych wykazały istotny wzrost dokładności modeli. Wykorzystanie kopul w tym samym kontekście można znaleźć w (Meyer, Schefzik, Thorarinsdottir, & Gneiting, 2021; Benali, Notton, Fouilloy, & Voyant, 2021; Jeong, Lee, & Lee, 2016). W (Nejad, Erdogan, & Cirillo, 2021) kopula Archimedesowa, w szczególności kopula Claytona, została wykorzystana do odwzorowania zależności między zmiennymi socjodemograficznymi podczas generowania populacji syntetycznej. Na podstawie danych pochodzących z różnych źródeł autorzy estymowali rozkłady brzegowe poszczególnych obserwacji oraz macierz zależności pomiędzy nimi. Kopula posłużyła do połączenia tych rozkładów w jeden model wielowymiarowy, umożliwiając generowanie realistycznych obserwacji reprezentujących mieszkańców małych obszarów geograficznych. Dzięki temu, możliwe było utworzenie populacji syntetycznej, zachowującej strukturę zależności obserwowaną w danych rzeczywistych, co pozwoliło następnie na wiarygodne modelowanie i analizę ewakuacji osób pozbawionych samochodów. Zgodnie z podejściem przedstawionym w (Meyer et al., 2021), kopule Gaussa lub kaskady kopul umożliwiają tworzenie syntetycznych obserwacji, które zachowują zarówno indywidualne właściwości zmiennych, jak i ich zależności. W tym ujęciu macierz korelacji estymowana z danych rzeczywistych stanowi podstawę budowy kopuli, a z niej generowane są nowe próbki o spójnej strukturze statystycznej. Takie dane syntetyczne mogą być następnie wykorzystywane do trenowania modeli uczenia maszynowego bez ujawniania wrażliwych danych, przy jednoczesnym zachowaniu ich wartości informacyjnej. W (Chu, Ip, Lam, & So, 2022) zaproponowano zastosowanie kaskad kopul do generowania syntetycznych danych, mając na uwadze, iż ze względu na wrażliwe informacje w danych, dane te nie mogą być publikowane. Model umożliwia dokładne odwzorowanie zależności pomiędzy zmiennymi w danych mieszanych. Dane syntetyczne generowane są poprzez próbkowanie warunkowe z dopasowanego modelu kaskad kopuli, co pozwala zachować strukturę zależności przy jednoczesnym ograniczeniu ryzyka ujawnienia wartości rzeczywistych. Podejście to wykorzystuje klasyczną dekompozycję $C-vine$ i $D-vine$, w której każda para zmiennych jest modelowana odrębną kopulą, a typ kopuli dobierany jest na podstawie miary dopasowania. Z kolei, w (Griesbauer, Czado, Frigessi, & Haff, 2025) przedstawiono rozszerzenie klasycznego podejścia $C-vine$, nazwane uciętymi

$C-vine$, które służy do generowania danych syntetycznych przy zachowaniu kompromisu pomiędzy użytecznością a prywatnością. Autorzy przycinają strukturę kaskad kopul powyżej określonego poziomu drzewa, co skutkuje odrzuceniem słabszych lub potencjalnie wrażliwych zależności między zmiennymi. Dzięki temu, możliwe jest ograniczenie ujawniania informacji przy zachowaniu najistotniejszych zależności. Metoda ta szczególnie dobrze sprawdza się w generowaniu tabelarycznych danych syntetycznych, które mają być wykorzystywane w uczeniu maszynowym. Oba podejścia opierają się na klasycznej teorii kaskad kopul, jednak różnią się celem: pierwsze koncentruje się na danych mieszanych i ochronie prywatności poprzez modelowanie zależności, drugie natomiast wykorzystuje strukturalne przycinanie zależności, aby zapewnić kompromis między jakością statystyczną a bezpieczeństwem danych syntetycznych. Na szczególną uwagę zasługują podejścia wykorzystujące kopule do nieparametrycznego generowania danych syntetycznych, które pozwalają odwzorować złożone zależności pomiędzy zmiennymi bez konieczności przyjmowania założeń o ich rozkładach.Jednym z takich rozwiązań jest metoda zaproponowana w (Restrepo, Arango, Trujillo, & Zapata, 2023), w której przedstawiono w pełni nieparametryczne podejście do generowania danych syntetycznych z wykorzystaniem kopul. Podejście to zakłada, że zarówno kopula C, opisująca zależności między zmiennymi, jak i rozkłady brzegowe $F_i \sim X_i$ są nieznane. Znamy jedynie dane rzeczywiste X_{nd}, gdzie n oznacza liczbę obserwacji, a d liczbę zmiennych. Celem metody jest odtworzenie struktury zależności pomiędzy zmiennymi na podstawie empirycznej kopuli skonstruowanej bezpośrednio z danych, zdefiniowanej jako:

$$\hat{U}_{ij} = \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{X_{kj}\leq X_{ij}} \quad \forall j \in \{1,\ldots,d\},\ i \in \{1,\ldots,n\}, \tag{3.74}$$

Dla każdej zmiennej X_j rozważa się przedział podzielony na t_j podprzedziałów:

$$\min(X_i) = a_{0i} < a_{1i} < \cdots < a_{t_i i} = \max(X_i), \tag{3.75}$$

przy czym liczba podziałów t_i może być wybrana w zależności od oczekiwanej dokładności estymacji rozkładu brzegowego. Każdy podprzedział definiuje się jako:

$$B_{si} = \begin{cases} [a_{(s-1)i},\ a_{si}], & \text{dla } s = 1, \\ (a_{(s-1)i},\ a_{si}], & \text{dla } s > 1, \end{cases} \qquad s \in \{1,\ldots,t_i\}, \tag{3.76}$$

gdzie B_{si} oznacza s-ty podprzedział dla zmiennej X_i. Na tej podstawie definiuje się funkcję:

$$R(B_{sj}) = \frac{1}{n}\sum_{r=1}^{s}\sum_{k=1}^{n} \mathbf{1}_{\{X_{kj}\in B_{rj}\}}, \quad \forall s \in \{1,\ldots,t_j\}, \tag{3.77}$$

która stanowi empiryczny estymator dystrybuanty zmiennej X_i obliczany w kolejnych podprzedziałach. Wektor:

$$\left[R(B_{s1}),\ldots,R(B_{sp})\right], \tag{3.78}$$

jest estymatorem nieobciążonym rzeczywistych wartości dystrybuant brzegowych:

$$\left[F_1(a_{s1}),\ldots,F_p(a_{sp})\right]. \tag{3.79}$$

Ponieważ postać kopuli C jest nieznana, losowo wybierany jest indeks $m \in \{1, \ldots, n\}$ i przyjmowany jest m-ty wiersz macierzy $\hat{U}$ jako reprezentatywny wektor wartości empirycznej kopuli.

Następnie, dla każdego wymiaru $i \in \{1, \ldots, d\}$:

1. losowane są obserwacje ze zmiennej $U \sim \mathcal{U}(0, 1)$,
2. wyznacza najmniejsze s spełniające warunek:

$$R(B_{si}) \geq \hat{U}_{mi}, \tag{3.80}$$

3. oblicza się wartość syntetyczną:

$$\hat{X}_i = a_{(s-1)i} + \left(a_{si} - a_{(s-1)i}\right) U. \tag{3.81}$$

Wygenerowany w ten sposób wektor $(\hat{X}_1, \ldots, \hat{X}_d)$ zachowuje zarówno strukturę rozkładów brzegowych, jak i strukturę zależności pomiędzy zmiennymi zaobserwowaną w danych rzeczywistych. Cały proces przebiega bez jawnego modelowania postaci kopuli ani przyjmowania jakichkolwiek założeń parametrycznych, co czyni metodę w pełni nieparametryczną i elastyczną w zastosowaniu do różnych typów danych.

4. Ryzyko modelu

Modelowanie ma charakter wielodyscyplinarny; obejmuje wiedzę na temat biznesu, teorii finansowych, modelowania matematycznego, implementacji oprogramowania komputerowego oraz projektowania interfejsów użytkownika. Model stanowi uproszczony opis rzeczywistości, stworzony w celu zrozumienia, analizy lub prognozowania zjawisk. Według klasycznej definicji podanej w (Silver, 2012), model to narzędzie służące do przekształcania złożoności świata rzeczywistego w logiczną strukturę, która może być zrozumiała i użyteczna do analizy. Modelowanie pozwala na badanie kluczowych zależności, identyfikowanie wzorców oraz testowanie hipotez w kontrolowany sposób. W literaturze naukowej modele są definiowane jako abstrakcyjne reprezentacje wybranych procesów. Na przykład Derman (1996) uważa, że model pełni funkcję pośrednika między teorią a rzeczywistością, umożliwiając testowanie założeń teoretycznych w kontekście danych empirycznych. W praktyce modele przyjmują różne formy, od prostych równań matematycznych po zaawansowane algorytmy oparte na sztucznej inteligencji. Model jest narzędziem służącym do sporządzania ograniczonego opisu rzeczywistości, który służy do podejmowania określonych decyzji. Modele nie są pełnym odwzorowaniem rzeczywistości, lecz skupiają się na kluczowych elementach istotnych dla danego zagadnienia. Jak zauważył Box (1979), „wszystkie modele są błędne, ale niektóre są użyteczne". Najważniejszą kwestią w ocenie modelu jest jego zdolność do spełnienia zamierzonych celów, takich jak przewidywanie zdarzeń czy wspieranie decyzji. Pomimo swojej użyteczności, modele mają swoje ograniczenia, które wynikają z konieczności upraszczania rzeczywistości. Upraszczanie, choć pozwala na analizę złożonych systemów, niesie ze sobą ryzyko błędnych założeń, które są niewystarczające do reprezentacji rzeczywistych zależności. Modele są wdrażane do życia codziennego, zatem działają w dynamicznym środowisku, gdzie zmieniające się warunki mogą sprawić, że ich założenia i wyniki tracą na aktualności. W takich sytuacjach pojawia się pojęcie ryzyka modelu, związanego z konsekwencjami użycia modeli, które mogą być niewłaściwe, niedostosowane do sytuacji lub błędnie interpretowane. Derman (1996) opisuje to zjawisko jako ryzyko wynikające z błędów w modelu teoretycznym. Z kolei, Gatarek (2002) definiuje je jako stratę powstałą w wyniku zastosowania niewłaściwego modelu do wyceny instrumentu finansowego lub pomiaru ryzyka. Według Dowd (2005) problem ten dotyczy błędów w szacowaniu miar ryzyka spowodowanych niedoskonałościami modelu. Kuziak (2005) przedstawia tę kwestię szerzej, wskazując, że sytuacja ta występuje, gdy formalny i poprawny opis fragmentu rzeczywistości odbiega od rzeczywistego stanu, co może prowadzić do strat. Ryzyko modelu jest szczególnie istotne w sektorach takich jak bankowość i ubezpieczenia, a także w działalności związanej z zarządzaniem inwestycjami, gdzie modele matematyczne i statystyczne odgrywają kluczową rolę w analizie ryzyka, wycenie instrumentów finansowych i ocenie wypłacalności. Ryzyko modelu może wynikać z różnych źródeł, takich jak niepełne lub niewiarygodne dane wejściowe, niewłaściwe założenia teoretyczne, błędy programistyczne czy ograniczenia samego modelu. Przykładowo, użycie niewłaściwego modelu w wycenie portfela kredytowego może skutkować niedoszacowaniem ryzyka niewypłacalności klientów. Co więcej, ryzyko modelu może być potęgowane przez zmieniające się warunki rynkowe, które sprawiają, że modele, choć dokładne w przeszłości, stają się mniej adekwatne w nowych realiach. Zarządzanie ryzykiem modelu stanowi kluczowy element zarządzania ryzykiem w organizacjach. Wymaga regularnej walidacji i weryfikacji stosowanych modeli, dokładnej dokumentacji oraz transparentności procesów decyzyjnych. W erze szybko rozwijających się technologii, w tym sztucznej inteligencji i uczenia maszynowego, wyzwania związane z ryzykiem modelu stają

się coraz bardziej złożone, co podkreśla konieczność inwestowania w odpowiednie zasoby i kompetencje. W niniejszym rozdziale przedstawiona będzie istota ryzyka modelu, jego główne przyczyny oraz strategie zarządzania, które pozwolą na minimalizację jego skutków.

W pierwszej części rozdziału zostaną omówione rodzaje ryzyka modelu, w tym m.in. ryzyko koncepcyjne, implementacyjne, jakość danych oraz problem nadmiernego dopasowania. Następnie, zostaną przedstawione źródła ryzyka, które mogą wynikać z czynników ludzkich, technologicznych oraz środowiskowych. Ostatnia część rozdziału dotyczy oceny ryzyka modelu, obejmującej zarówno podejścia jakościowe, jak i ilościowe.

4.1 Znaczenie ryzyka modelu w instytucjach finansowych

Instytucje finansowe wykorzystują modele w celu zarządzania ryzykiem, wyceny instrumentów finansowych oraz podejmowania decyzji operacyjnych i strategicznych. W związku z tym dla tych instytucji kluczowe znaczenie odgrywa ocena ryzyka, jakie niesie zastosowanie określonego modelu. Literatura przedmiotu podkreśla znaczenie ryzyka modelu w następujących obszarach:

- Stabilność finansowa - modele finansowe wykorzystuje się do oceny ryzyk wynikających m.in. z ryzyka kredytowego, rynkowego oraz operacyjnego. Niedokładne lub błędnie zbudowane modele mogą prowadzić do niewłaściwej oceny ekspozycji na ryzyko, co przekłada sie na nieadekwatną alokację kapitału. Niepoprawne matematycznie oraz niezgodne z regulacjami modele mogą wpływać na stabilność instytucji finansowych, przyczyniając się do ryzyka systemowego.

- Zgodność z regulacjami - instytucje finansowe, stosując modele, muszą dostosować się do regulacji, np. Bazylea III, Solvency I czy Solvency II. Niedokładne modele mogą prowadzić do niezgodności z przepisami, co może skutkować sankcjami ze strony organów nadzoru. Zarządzanie ryzykiem w instytucjach finansowych jest istotne zarówno dla stabilności systemu finansowego, jak i dla generowania ryzyka systemowego.

- Podejmowanie decyzji operacyjnych i strategicznych - modele mają na celu wspieranie podejmowania decyzji w instytucjach finansowych. Błędne modele mogą prowadzić do niewłaściwych decyzji, przez co instytucje ponoszą straty. Odpowiednie zarządzanie ryzykiem operacyjnym, w tym jakością modeli, jest kluczowe dla efektywnego funkcjonowania banków.

- Zapobieganie ryzyku systemowemu - ryzyko modelu, wynikające z błędnych założeń i nadmiernej wiary w modele kwantyfikujące ryzyko, może samo w sobie generować ryzyko systemowe, gdyż błędne modele są stosowane jednocześnie przez wiele instytucji; (Danielsson, 2016) antidotum stanowi większy nacisk na testy odporności modeli oraz uwzględnianie niepewności modelowej w zarządzaniu ryzykiem. Z kolei, błędne modelowanie sieci powiązań finansowych może prowadzić do niedoszacowania efektów kaskadowych i zwiększać podatność systemu na szoki. Aby temu zapobiec, Acemoglu i in. (2015) (Acemoglu, Ozdaglar, & Tahbaz-Salehi, 2015) zalecają projektowanie bardziej odpornych struktur sieciowych oraz włączenie heterogeniczności i niepewności do modeli systemowych. Z kolei, niewłaściwe modelowanie współzależności instytucji finansowych (np. w ramach CoVaR), co może zaniżać miary ryzyka systemowego, Acharya i in.

(2017) (Acharya, Pedersen, Philippon, & Richardson, 2017) zalecają kalibrację modeli z wykorzystaniem danych stresowych oraz dynamicznych korelacji, odzwierciedlających rzeczywistą współzależność w okresach kryzysowych.

W przeszłości można odnaleźć wiele negatywnych zdarzeń, które miały niepożądane skutki i wynikały z niewłaściwego wykorzystania modeli:

- Kryzys finansowy w latach 2007-2008 - Przed kryzysem finansowym wiele instytucji finansowych opierało swoje modele oceny ryzyka kredytowego na założeniach ignorujących możliwość spadku cen nieruchomości i korelacje między aktywami (Commission, 2011). Modele te niedoszacowywały ryzyka, prowadząc do błędnej wyceny ryzykownych aktywów. Rajan (2010) podkreśla, że uproszczone modele nie odzwierciedlały rzeczywistości i przyczyniły się do nadmiernego zadłużenia oraz destabilizacji systemu finansowego. Stosowanie tych samych modeli w różnych instytucjach stworzyło efekt domina, uwypuklając znaczenie ryzyka modelu w kontekście ryzyka systemowego.

- Upadek Long-Term Capital Management (LTCM) w 1998 r. uwypuklił ryzyko związane z modelami finansowymi. Fundusz wykorzystywał zaawansowane modele matematyczne do strategii arbitrażowych, zakładając niskie prawdopodobieństwo jednoczesnych kryzysów na różnych rynkach. Założenia te okazały się błędne, prowadząc do ogromnych strat i konieczności interwencji banku centralnego w celu uniknięcia destabilizacji systemu finansowego. Analizy LTCM (Lopez & Saidenberg, 2001; Shirreff, 1999) wskazują, że upadek funduszu wynikał z zastosowania modeli opartych na nierealistycznych założeniach. Przykład ten podkreśla znaczenie ryzyka modelu w decyzjach operacyjnych i strategicznych oraz w zapobieganiu ryzyku systemowemu - błędne modele mogą wymusić interwencję instytucji publicznych, aby powstrzymać efekt domina w systemie finansowym.

- Afera Enronu na początku lat 2000 ujawniła znaczenie ryzyka modelu w zarządzaniu finansami i zgodności z regulacjami. Firma wykorzystywała skomplikowane modele finansowe do wyceny kontraktów opartych na nierealistycznych założeniach, co doprowadziło do jednego z największych skandali finansowych i upadku spółki. Analizy (McLean & Elkind, 2003; Fox, 2003) podkreślają rolę błędnych modeli w kryzysie, ukazując ich wpływ na decyzje finansowe i stabilność sektora. Przykład Enronu uwypukla znaczenie ryzyka modelu zarówno w kontekście zgodności z regulacjami - afera przyczyniła się do reform, takich jak Sarbanes-Oxley Act - jak i w kontekście stabilności finansowej, ponieważ upadek spółki negatywnie wpłynął na inwestorów i sektor energetyczny.

- Afera „Londyńskiego Wieloryba" w 2012 r. jest klasycznym przykładem ryzyka modelu w praktyce bankowej. Trader JPMorgan Chase dokonywał ogromnych transakcji na instrumentach pochodnych w ramach Synthetic Credit Portfolio (SCP), opierając się na modelach ryzyka, które znacząco niedoszacowywały potencjalne straty. Błędy w konstrukcji modelu Value at Risk, a także jego celowe modyfikacje, spowodowały zaniżone raportowanie ryzyka i nadmierne zwiększenie ekspozycji na ryzykowne instrumenty (Zeisberger & Chen, 2014). W rezultacie straty przekroczyły 6 miliardów dolarów, podważając reputację banku i zaufanie inwestorów. Raport Senatu USA (on Investigations, 2013) wykazał, że problem nie wynikał wyłącznie z samego modelu, lecz także z niedostatecznego zarządzania ryzykiem - słabego nadzoru, nieadekwatnych procedur kontrolnych

i niewystarczającego raportowania. Analizy (Zeissler & Metrick, 2019) podkreślają, że modele ilościowe nie uwzględniały skrajnych scenariuszy ani korelacji między instrumentami, a decyzje strategiczne opierały się na błędnych założeniach. Przypadek ten uwidocznił ograniczenia modeli finansowych, ich wpływ na stabilność finansową oraz konieczność zgodności narzędzi modelowych z regulacjami i procedurami decyzyjnymi w instytucjach.

4.2 Rodzaje ryzyka modelu

Ryzyko modelu może przyjmować różne formy i występować w wielu obszarach zastosowania modeli. W literaturze spotyka się liczne klasyfikacje tego zjawiska, różniące się zakresem i kryteriami podziału. W dalszej części przedstawiona zostanie przyjęta w tej pracy klasyfikacja, której celem jest uporządkowanie najważniejszych rodzajów ryzyka modelu z perspektywy ich wpływu na funkcjonowanie instytucji finansowych oraz stabilność systemu finansowego. Przyjęta klasyfikacja nawiązuje do podejścia stosowanego w ramach regulacji Solvency II, w której szczególną uwagę zwraca się na identyfikację, pomiar i kontrolę ryzyka modelu jako elementu systemu zarządzania ryzykiem.

4.2.1 Ryzyko koncepcyjne

Ryzyko koncepcyjne, czyli błędna specyfikacja modelu, odnosi się do sytuacji, w której model nieodpowiednio odwzorowuje rzeczywistość, prowadząc do niepoprawnych wyników i błędnych decyzji. Wynika to z niewłaściwych założeń teoretycznych, uproszczeń w konstrukcji modelu oraz błędów w specyfikacji statystycznej. W kontekście agregacji ryzyka w Solvency II można zauważyć następujące zagrożenia:

- Rozkłady normalne i korelacje liniowe różnią się od macierzy stosowanej w FS. W FS przyjęto uproszczone założenie, że zmienne agregowane mają wielowymiarowy rozkład normalny z odgórnie ustalonymi wartościami macierzy korelacji. Stosowanie stałych korelacji liniowych może prowadzić do błędnej oceny wymogów kapitałowych, skutkując ich przeszacowaniem lub niedoszacowaniem.

- Rozkłady brzegowe różnią się od rozkładu normalnego. W FS przyjmuje się, że rozkłady ryzyka w portfelu ubezpieczyciela są normalne. W praktyce jednak często wykazują one asymetrię i „grube ogony", co oznacza większe prawdopodobieństwo wystąpienia zdarzeń ekstremalnych niż przewiduje rozkład normalny. Stosowanie założenia o normalności może prowadzić do błędnej oceny prawdopodobieństwa wysokiego ryzyka. Przykładowo, w ubezpieczeniach majątkowych straty katastroficzne mogą znacznie odbiegać od prognoz opartych na rozkładzie normalnym.

- Struktura zależności między agregowanymi czynnikami ryzyka nie zawsze jest liniowa. FS zakłada liniowe zależności między czynnikami ryzyka, podczas gdy w rzeczywistości zależności te mogą być nieliniowe. W ubezpieczeniach majątkowych ekstremalne zdarzenia, takie jak huragany czy powodzie, powodują straty w różnych segmentach jednocześnie. Huragany niszczą budynki i infrastrukturę, opóźniając naprawy, a rosnące ceny materiałów budowlanych oraz usług remontowych dodatkowo zwiększają wypłaty z polis. W ubezpieczeniach komunikacyjnych ekstremalne warunki pogodowe zwiększają

liczbę wypadków i uszkodzeń pojazdów, szczególnie w obszarach miejskich. Przerwy w działalności firm i instytucji finansowych generują dodatkowe roszczenia, co w krótkim czasie kumuluje wiele szkód.

Takie sytuacje prowadzą do nieliniowych zależności między ryzykami, w których zdarzenia wzajemnie się wzmacniają. Tradycyjne modele liniowe, zakładające proporcjonalne i stałe zależności, nie uchwycą tej dynamiki. Do opisu skumulowanych ryzyk konieczne jest stosowanie bardziej zaawansowanych narzędzi, takich jak kopule, pozwalających dokładnie modelować nieliniowe zależności.

4.2.2 Ryzyko implementacji algorytmów

Ryzyko związane z implementacją algorytmów może wynikać z wielu przyczyn, począwszy od problemów z samymi algorytmami po trudności w integracji z systemem. Największym ryzykiem jest to, że algorytm został niepoprawnie zaprogramowany. Niewielkie różnice między matematycznym zapisem algorytmu a zaprogramowanym algorytmem mogą prowadzić do nieprawidłowych wyników. Błędy w zaprogramowanym algorytmie mogą wynikać z niedokładnego zrozumienia algorytmu, pominięcia istotnych założeń lub błędnego odwzorowania równań matematycznych na język programowania, co powoduje ograniczenia modelu lub brak zamierzonych rezultatów. Istotną kwestią są również błędy numeryczne. W przypadku obliczeń wieloetapowych na każdym etapie mogą występować błędy, które będą kumulować się z etapu na etap. Implementacja algorytmu może nie zawierać obsługi wyjątków, takich jak dzielenie przez zero lub błędne formaty danych. Pominięcie w obsłudze takich wyjątków może prowadzić do błędnych wyników lub zatrzymania systemu. Problem integracji stanowi kolejne wyzwanie, zwłaszcza gdy modele muszą współpracować z różnorodnymi systemami. Przykładem może być zapis znaków, takich jak UTF-8 i ASCII, lub zapis dat, które mogą powodować błędy w przesyłaniu danych pomiędzy różnymi środowiskami. Nieścisłości mogą też wynikać z odmiennych standardów numeracji indeksów w różnych językach programowania. Python i C++ zaczynają numerowanie od zera, a R od jedynki, co wymaga szczególnej uwagi podczas implementacji i korzystania z różnych języków programowania równocześnie. Problemem dużych modeli jest ich wydajność; dlatego stosuje się obliczenia równoległe, które umożliwiają przyspieszenie obliczeń. Jednak obliczenia równoległe niosą ze sobą pewne ryzyko, m.in. problemy z synchronizacją danych oraz kolizje dostępu do wspólnej pamięci. Dodatkowo, nierównomierne rozłożenie zadań między wątki a procesy może obniżać efektywność obliczeń. Tworząc model stochastyczny, istotnym elementem jest właściwy dobór generatorów liczb pseudolosowych. Nieodpowiedni generator może prowadzić do nieprzewidywalności wyników i trudności w ich weryfikacji. W (Culver, Heitmann, & Weiß, 2018) przedstawiono znaczenie wyboru ziarna dla wyniku w procesach oceny ryzyka.

4.2.3 Ryzyko wynikające z jakości danych

W dzisiejszym świecie, w którym decyzje finansowe, operacyjne i strategiczne coraz bardziej opierają się na danych, ryzyko wynikające z ich niskiej jakości staje się poważnym problemem dla wielu sektorów, w tym dla ubezpieczeń. Dane o niedostatecznej jakości mogą prowadzić do błędnych analiz, niewłaściwego modelowania ryzyka, a w konsekwencji do nieoptymalnych lub wręcz sprzecznych decyzji. Problemy z danymi mogą przybierać różne formy: od błędów wprowadzania, przez braki w kluczowych informacjach, aż po dane nieaktualne lub

nieadekwatne względem celu ich wykorzystania. W kontekście aktuariatu, gdzie dane stanowią fundament kalkulacji rezerw techniczno-ubezpieczeniowych i wymogów kapitałowych, ich jakość odgrywa kluczową rolę. Rozporządzenie delegowane Komisji (UE) 2015/35 (European Commission, 2015) w ramach Solvency II nakłada szczególne wymogi dotyczące jakości danych, zwracając uwagę na:

- Dokładność danych, które powinny być wolne od błędów i odpowiadać rzeczywistości. Przykładem może być nieprawidłowe wprowadzenie danych do bazy, np. kwot szkód czy błędów w datach zdarzeń. Błędy te mogą wynikać również z wykorzystania przestarzałych tabel śmiertelności lub nieuwzględniania różnic walutowych w kalkulacji odszkodowań, co prowadzi do istotnych zniekształceń wyników analizy ryzyka.

- Kompletność danych, czyli fakt, że wszystkie istotne informacje powinny być dostępne. Brak ważnych informacji w danych, takich jak szczegółowe informacje o przyczynach szkód, charakterystyka polis czy informacje o reasekuracji, ogranicza możliwości przeprowadzenia dokładnej analizy ryzyka. Braki w danych podczas procesu raportowania utrudniają walidację i porównywanie informacji. Na przykład, brak szczegółów dotyczących rodzaju aktywów w portfelu inwestycyjnym może prowadzić do niewłaściwej oceny ryzyka związanego z koncentracją portfela.

- Adekwatność danych, które muszą być odpowiednie do celu, w jakim są wykorzystywane. Na przykład dane historyczne z rynków zagranicznych, które nie uwzględniają specyfiki lokalnych przepisów lub różnic demograficznych, mogą być nieadekwatne do analizy ryzyka na konkretnym rynku. Brak reprezentatywności wykorzystywanych danych, szczególnie w przypadku szkód katastroficznych, może prowadzić do błędów w modelach aktuarialnych i skutkować niedoszacowaniem wymogów kapitałowych. EIOPA (2022) zwraca uwagę na konieczność wykorzystania unikalnych identyfikatorów, takich jak LEI, aby zapewnić precyzyjne połączenie danych z różnych źródeł i umożliwić dokładne analizy powiązań oraz koncentracji ryzyk. Przykładem adekwatnego użycia danych może być analiza szkód spowodowanych przez huragany, która wymaga danych historycznych uwzględniających zmienność klimatyczną w określonym regionie.

Zaniedbanie tych kryteriów może prowadzić do poważnych konsekwencji, w tym niedoszacowania rezerw, błędnej oceny ryzyk i naruszenia regulacyjnych wymogów wypłacalności.

4.2.4 Ryzyko nadmiernego dopasowania

Nadmierne dopasowanie pojawia się, gdy model matematyczny uczy się zbyt szczegółowo danych treningowych. Zamiast uogólniać wzorce, model „zapamiętuje" szumy i anomalie obecne w danych, co prowadzi do błędnych predykcji na danych rzeczywistych. Problem ten jest szczególnie istotny w uczeniu maszynowym oraz aktuariacie, w których modele są używane do prognozowania ryzyka finansowego, strat ubezpieczeniowych i rezerw techniczno-ubezpieczeniowych. Nawet niewielkie błędy mogą mieć poważne konsekwencje, zarówno finansowe, jak i operacyjne.

W uczeniu maszynowym nadmierne dopasowanie występuje najczęściej, gdy modele są bardzo złożone, m.in. w przypadku głębokich sieci neuronowych czy bardzo głębokich drzew decyzyjnych, które mają tendencję do zapamiętywania szczegółów danych zamiast uczenia się

ogólnych wzorców. Problemem jest także zbyt długi czas treningu. Dla przykładu, zbyt wiele epok powoduje, że model dostosowuje się do najdrobniejszych szczegółów danych treningowych. Dodatkowo, ograniczony zbiór danych sprawia, że model nie ma wystarczającej liczby przykładów, aby nauczyć się wszystkich istotnych zależności.

W modelach aktuarialnych nadmierne dopasowanie modeli do danych stanowi istotne ryzyko, ponieważ modele są wykorzystywane do prognozowania kluczowych wskaźników, takich jak prawdopodobieństwo wystąpienia szkód, rezerwy techniczne czy ryzyko rezygnacji klientów. Przyczyną jest złożoność modelu, która przewyższa rzeczywiste potrzeby. Dla przykładu, firma ubezpieczeniowa buduje model przewidujący prawdopodobieństwo wystąpienia szkód w ubezpieczeniach komunikacyjnych, korzystając z sieci neuronowej z wieloma warstwami. Ten sam problem może być opisany za pomocą regresji logistycznej. Złożony model zapamiętuje szczegóły w danych treningowych, zamiast uczyć się ogólnych zależności. W efekcie model doskonale dopasowuje się do danych historycznych, lecz posiada słabą zdolność prognozowania przyszłych szkód. W konsekwencji model dokonuje nierealistycznych predykcji.

Innym powodem nadmiernego dopasowania jest niereprezentatywność danych historycznych. Model prognozujący straty w ubezpieczeniach rolniczych może opierać się na danych z ostatnich trzech lat, które obejmują wyjątkowo suche lata. W efekcie model przewiduje wysokie ryzyko strat związanych z suszą, ignorując występowanie mokrych sezonów w przyszłości. Takie podejście prowadzi do zawyżonych składek ubezpieczeniowych, co może zniechęcać rolników do wykupywania polis i obniżać konkurencyjność firmy. Z kolei, przeszacowanie rezerw technicznych zamraża kapitał, który mógłby zostać wykorzystany w innych obszarach działalności firmy.

Przykładem nadmiernego dopasowania jest także uwzględnianie zbyt szczegółowych zmiennych w modelach. Model wyceny składki w ubezpieczeniach komunikacyjnych może przypisać wysokie ryzyko kierowcom w wieku 35 lat, mieszkającym w małej miejscowości, na podstawie jednego przypadku poważnego wypadku. Tego typu szczegółowe zmienne, które nie są reprezentatywne, prowadzą do zawyżania składek i zniechęcania klientów. Podobnie, modele opisujące ryzyko w ubezpieczeniach rolniczych mogą uwzględniać zmienne, takie jak liczba hektarów ziemi uprawnej sąsiadów czy odległość od najbliższego miasta, które nie mają rzeczywistego wpływu na szkody. Model przypisuje takim zmiennym zbyt dużą wagę, co skutkuje nierealistycznymi prognozami.

Choć w praktyce wykorzystywane modele w aktuariacie, takie jak uogólnione modele liniowe (ang. Generalized Linear Models - GLM), Chain Ladder czy kopula, są mniej podatne na nadmierne dopasowanie niż zaawansowane algorytmy uczenia maszynowego, mogą również ulegać temu problemowi. W przypadku GLM nadmierne dopasowanie może wynikać z uwzględnienia zbyt wielu zmiennych niezależnych. Na przykład uwzględnienie szczegółowych zmiennych demograficznych, takich jak zawód w podziale na wiele kategorii, może prowadzić do dopasowywania do niereprezentatywnych danych, co skutkuje błędną wyceną ryzyka w małych grupach.

W metodzie Chain Ladder przeuczenie może wynikać z dopasowania modelu do zdarzeń, które wystąpiły pojedynczo w danych historycznych, takich jak jednorazowy wzrost liczby szkód spowodowany zmianą regulacji. W takich przypadkach model traktuje te anomalie jako trwałe trendy, co prowadzi do przeszacowania rezerw technicznych. Problemem jest również założenie liniowości wzorców, które ignoruje zmienność wynikającą z cykli gospodarczych lub zmian w polityce likwidacji szkód. W rezultacie model może generować zbyt konserwatywne prognozy, które blokują kapitał firmy. Problem nadmiernego dopasowania nie może być jednak

całkowicie wyeliminowany, ale można go ograniczyć poprzez stosowanie regularyzacji, odpowiednią optymalizację hiperparametrów oraz weryfikację wyników modeli na danych testowych i rzeczywistych.

4.3 Źródła ryzyka modelu

Skuteczne zarządzanie ryzykiem modelu wymaga nie tylko samej kontroli wyników, ale przede wszystkim rozpoznania i analizy głównych źródeł, które mogą prowadzić do błędów czy nieścisłości. Ryzyko to może mieć różne podłoże i oddziaływać na odmienne etapy pracy z modelem, od jego konstrukcji, przez praktyczne funkcjonowanie i interpretację rezultatów, aż po etap wdrożenia.

W niniejszym rozdziale omówiono trzy zasadnicze obszary, z których mogą wynikać zagrożenia dla poprawności modeli: czynnik ludzki, technologiczny oraz środowiskowy.

4.3.1 Czynnik ludzki

Czynnik ludzki należy do podstawowych źródeł ryzyka modelu. Popełnienie błędu może wystąpić na każdym etapie, od projektowania, przez implementację i walidację, aż po interpretację wyników. Najmniejsza pomyłka na każdym z etapów może istotnie obniżyć jakość wyników i skuteczność zastosowanego modelu. Niewystarczająca wiedza osób odpowiedzialnych za budowę modeli zwiększa ryzyko koncepcyjne i implementacyjne, ponieważ przyjęte założenia mogą być błędne, a zastosowane modele niewłaściwe. Ważnym obszarem jest także jakość danych. Pominięcie braków danych, błędy w procesie czyszczenia oraz niewłaściwe przygotowanie zbioru danych mogą sprzyjać nadmiernemu dopasowaniu, czyli sytuacji, w której model dobrze odtwarza dane historyczne, natomiast ma problem z prognozowaniem przyszłości. Ryzyko może pojawić się również w trakcie testów walidacyjnych. Jeśli scenariusze testowe nie uwzględniają zdarzeń skrajnych, model pozostaje niewystarczająco odporny na potencjalne wstrząsy w rzeczywistości. Brak kompleksowej weryfikacji założeń w ekstremalnych warunkach wzmacnia ryzyko koncepcyjne. Dodatkowym zagrożeniem są błędy interpretacyjne. Brak intuicji w odczycie wyników przez analityków może skutkować zmianami w modelu, które zamiast poprawy obniżają jego efektywność w dłuższym horyzoncie.

4.3.2 Czynnik technologiczny

Wiele algorytmów powstaje w oparciu o założenie, że dane są spójne i kompletne. W praktyce bywa jednak inaczej; zbiory mogą być niejednorodne, zawierać błędy lub być puste. W takich sytuacjach wyniki analiz łatwo stają się niereprezentatywne. Jeśli przyjmie się błędne założenia dotyczące danych, pojawia się ryzyko koncepcyjne, ponieważ model wówczas nie odzwierciedla rzeczywistych zależności. Kolejnym problemem jest skalowalność. Gdy baza danych rośnie, a obliczenia stają się coraz bardziej złożone, nie każdy algorytm potrafi sobie z tym poradzić. Dotyczy to zwłaszcza metod opartych na sztucznej inteligencji i głębokim uczeniu, które wykorzystują ogromne ilości pamięci oraz mocy obliczeniowej. Jeżeli model nie jest przygotowany do pracy na dużych zbiorach danych, wykorzystuje całą pamięć i przerywa działanie. Takie sytuacje świadczą o ryzyku implementacyjnym, ponieważ algorytm, zamiast wspierać proces analizy, staje się jego przeszkodą.

4.3.3 Czynniki środowiskowe

Kryzysy gospodarcze w oczywisty sposób wpływają na stabilność finansową zakładów ubezpieczeń. Dobrym przykładem jest globalny kryzys z lat 2007-2008, który doprowadził do gwałtownego spadku wartości aktywów w portfelach inwestycyjnych wielu firm. W efekcie ubezpieczyciele musieli zaostrzyć politykę inwestycyjną i prowadzić działalność z większą ostrożnością. Silny spadek wartości akcji oraz nieruchomości osłabił zdolność niektórych instytucji do pokrywania przyszłych zobowiązań wobec klientów. Na działalność ubezpieczycieli wpływają także zmiany inflacji i stóp procentowych. Wyższa inflacja podnosi koszty wypłat, co jest szczególnie mocno odczuwalne w ubezpieczeniach majątkowych. Z kolei, stopy procentowe decydują o rentowności portfeli inwestycyjnych.

Zmiany regulacyjne, takie jak wprowadzona dyrektywa Solvency II w Unii Europejskiej, mają kluczowe znaczenie dla działalności firm ubezpieczeniowych. Dyrektywa wymusza na firmach ubezpieczeniowych większe zabezpieczenie kapitałowe, tak aby zakłady były w stanie pokryć swoje zobowiązania. Dodatkowo, zmiany w przepisach dotyczących ochrony danych osobowych, takich jak RODO, wymuszają na firmach ubezpieczeniowych konieczność dostosowania procedur oraz wprowadzenia nowych rozwiązań technologicznych, co wiąże się z dodatkowymi kosztami.

Ważnym czynnikiem kształtującym działalność ubezpieczycieli są także preferencje i potrzeby klientów. W ostatnich latach coraz większym zainteresowaniem cieszą się ubezpieczenia zdrowotne, co zmusza firmy do dostosowywania oferty do nowych oczekiwań. Na sytuację zakładów wpływa również rosnąca konkurencja rynkowa. Obniżanie cen czy wprowadzanie innowacyjnych produktów przez konkurentów bezpośrednio oddziałuje na rentowność i wymaga szybkiej reakcji ze strony innych podmiotów.

Zmiany klimatyczne coraz silniej oddziałują na działalność ubezpieczycieli. Coraz częstsze i gwałtowniejsze zjawiska pogodowe, takie jak powodzie, huragany czy pożary, prowadzą do znaczącego wzrostu liczby roszczeń. Dobrym przykładem jest huragan Katrina z 2005 r., który spowodował ogromne straty w Stanach Zjednoczonych i zmusił firmy ubezpieczeniowe do wypłaty rekordowych odszkodowań. Podobne doświadczenia miała również Polska. Powódź z 1997 r., określana mianem „powodzi tysiąclecia", wywołała olbrzymie szkody materialne oraz tysiące zgłoszeń od poszkodowanych, a kolejne powodzie w 2010 r. ponownie wystawiły sektor ubezpieczeniowy na ciężką próbę. Takie wydarzenia pokazują, że zmiany klimatyczne wymuszają modyfikację modeli ryzyka oraz dostosowanie składek do rosnącego prawdopodobieństwa wystąpienia katastrof naturalnych.

4.4 Ocena ryzyka modelu

Aby ocenić ryzyko modelu, stosuje się dwa podejścia: ilościowe (m.in. testowanie wsteczne, testy warunków skrajnych, analiza wrażliwości i stabilności wyników) oraz jakościowe (ocena poprawności koncepcyjnej, adekwatności założeń, jakości danych, dokumentacji i nadzoru nad cyklem życia modelu). W praktyce ryzyko modelu jest także kwalifikowane jako szczególna kategoria ryzyka operacyjnego. W ocenie ryzyka modelu kluczowe znaczenie mają dwie formy weryfikacji: testowanie wsteczne oraz testy warunków skrajnych. Ideą testowania wstecznego jest sprawdzenie, czy model może być zastosowany w praktyce. Ocena jest przeprowadzana na podstawie danych historycznych. Dla przykładu, jeśli firma ubezpieczeniowa przygotowuje model prognozujący straty w 2024 r., to model ten ocenia się na podstawie danych z 2022

r., porównując je ze stratami, które wystąpiły w 2023 r. Taka weryfikacja potwierdza przydatność modelu do szacowania przyszłych strat. Testy warunków skrajnych oceniają wpływ rzadkich, lecz potencjalnie bardzo dotkliwych zdarzeń. Choć prawdopodobieństwo ich wystąpienia jest niewielkie, konsekwencje mogą istotnie oddziaływać na funkcjonowanie firmy. Testy te odpowiadają zatem na pytanie: „Jak duża może być strata?", a nie „Jak bardzo jest ona prawdopodobna?". Poniżej przedstawiono podstawowe metody stosowane w ramach testów warunków skrajnych:

- analiza scenariuszowa obejmuje wiele czynników ryzyka oraz bada wpływ zdarzeń katastroficznych na ryzyko przedsiębiorstw. Bazuje na analizie scenariuszy historycznych, zawierających informacje o rynkach finansowych w przeszłości, oraz na scenariuszach hipotetycznych dotyczących wydarzeń, które mogą wystąpić;
- metoda wartości ekstremalnych służy do modelowania rzadko występujących zdarzeń, które powodują ekstremalne straty. Estymuje straty w ogonach, a testy mogą być oparte na analizie skośności i grubości oraz na analizie innych cech ogonów rozkładów;
- metoda maksymalnej straty określa, dla jakich kombinacji zmian czynników ryzyka otrzymujemy największe straty;
- analiza wrażliwości często powiązana jest z analizą niepewności, która dotyczy analizy danych wejściowych modelu. Niepewność modelu może wynikać z wielu źródeł, takich jak błędy danych wejściowych lub błędy oszacowania parametrów modelu. Analiza wrażliwości jest narzędziem, które bada, jak różne niepewności modelu wpływają na jego wyniki. Analiza wrażliwości może przybierać różne formy:
  - analiza pojedynczej wrażliwości polega na zmianie jednego parametru w modelu, za każdym razem innego;
  - wielokrotna analiza wrażliwości przeprowadzana jest w celu oceny jednoczesnych zmian dwóch lub więcej parametrów modelu;
  - analiza progowa może pomóc w określeniu progów dla pewnych zmiennych, które wywołują wyniki będące przedmiotem zainteresowania;
  - analiza wrażliwości probabilistycznej polega na jednoczesnym pseudolosowym wyborze wielu parametrów analizy modelu. W analizie powinny zostać uwzględnione wszystkie parametry, które są obarczone błędem oszacowania.

Celem analizy wrażliwości jest określenie zakresu wartości, dla których badany model jest poprawny, oraz zwiększenie niezawodności modelu w przypadku zmieniających się danych wejściowych. Analizy wrażliwości mogą oceniać zmiany nie tylko w parametrach modelu, ale także w specyfikacji modelu. Prawidłowo zaprojektowana analiza wrażliwości jest narzędziem wykorzystywanym w modelowaniu, ponieważ dostarcza informacji na temat poprawności oraz stabilności prognoz modelu. Dostarcza informacji na temat istotności parametrów w modelu oraz może przyczynić się do specyfikacji modelu poprzez ocenę indywidualnego wkładu danej zmiennej oraz potrzeby uwzględnienia jej w modelu lub nie.

4.4.1 Zarządzanie modelem i ocena jakościowa ryzyka

Zagadnienia związane z oceną jakościową ryzyka modelu są szeroko dyskutowane w literaturze, zarówno w kontekście ubezpieczeń, jak i szerszego zarządzania ryzykiem finansowym (Kozmenko, Bieliaieva, & Ivashyna, 2015; Idris, 2024; Bernard, 2023; Varadarajan & Others, 2024). Ocena jakościowa ryzyka to narzędzie, które pomaga firmom lepiej zrozumieć działanie modelu oraz sprawniej radzić sobie z różnymi rodzajami zagrożeń. Obok analiz opartych wyłącznie na danych liczbowych uwzględnia się także sposób, w jaki model został zbudowany, jak również to, jak jest wykorzystywany. Pozwala to uchwycić czynniki, które mogą decydować o skuteczności modelu w praktyce. Ocenę jakościową wykorzystuje się do:

- zapewnienia zgodności modeli z sytuacją rzeczywistą - ocena jakościowa pozwala upewnić się, że modele matematyczne i statystyczne odzwierciedlają rzeczywiste warunki rynkowe oraz są w stanie przewidzieć potencjalne zagrożenia;
- monitorowania i zarządzania ryzykiem - ocena ta umożliwia ciągłe monitorowanie ryzyk, identyfikowanie nowych zagrożeń oraz aktualizację strategii zarządzania ryzykiem;
- spełnienia wymogów regulacyjnych - Solvency II wymaga, aby modele były regularnie przeglądane i walidowane. Ocena jakościowa jest kluczowym elementem spełniania tych wymogów.

W ramach oceny jakościowej analizuje się, czy dane wejściowe do modelu są poprawne, sprawdza się, czy metodologia modelowania jest prawidłowa, oraz czy oczekiwane wyniki są adekwatne do rzeczywistego ryzyka. Modele ryzyka oparte są często na założeniach o rozkładach statystycznych zmiennych losowych. Dla przykładu, w ryzyku kredytowym zakłada się, że straty mają rozkład normalny. Wówczas ocena jakościowa uwzględnia weryfikację, czy to założenie jest spełnione w świecie rzeczywistym, jak również możliwe konsekwencje odstępstw od niego. W (American Academy of Actuaries, 2019a) zaleca się w praktyce zarządzania ryzykiem modelu sprawdzenie tych założeń w kontekście aktualnej sytuacji gospodarczej oraz zmian na rynku kredytowym. W sytuacji, w której założenia nie są spełnione, konieczna jest zmiana modelu. Dalej ważną kwestią są dane wejściowe, ponieważ na ich podstawie budujemy model. Dane muszą być kompletne, dokładne i wolne od błędów. Ocena jakości danych obejmuje sprawdzenie ich kompletności, dokładności oraz aktualności. Przykładem z raportu EIOPA jest analiza danych dotyczących liczby i wieku ubezpieczonych, które są używane do modelowania ryzyka śmiertelności. Należy ocenić, czy dane, które wykorzystujemy, są regularnie aktualizowane i czy są reprezentatywne, tzn. czy opisują wszystkie przypadki osób objętych ubezpieczeniem. Kolejną kwestią jest to, iż modele muszą być zgodne z wymogami regulacyjnymi, takimi jak Solvency II. Ocena ta obejmuje sprawdzenie modelu pod kątem wymagań dotyczących przejrzystości, sprawozdawczości oraz adekwatności kapitałowej. Na przykład, weryfikacja, czy model uwzględnia wszystkie wymagane scenariusze ryzyka. Przykładem może być analiza tego, czy model właściwie uwzględnia ryzyko inflacji w swoich prognozach kapitałowych. Ocena wykorzystania modelu koncentruje się na analizie osób odpowiedzialnych za model oraz na weryfikacji ich kwalifikacji. Przykładem może być ocena decyzji dotyczących rezerw technicznych pod kątem interpretacji wyników modelu, jak również poprawności udokumentowania wyników. American Academy of Actuaries (2019b) sugeruje, aby regularnie przeprowadzać szkolenia dla osób odpowiedzialnych za model oraz wykonywać audyty wewnętrzne, które zapewnią, że model jest używany zgodnie z jego przeznaczeniem. Firmy

ubezpieczeniowe mają obowiązek formalnie zarządzać ryzykiem modelu poprzez ustanowienie polityk i procedur, które są dostosowane do działalności firmy. Poprzez wdrożenie polityki, która określa role i obowiązki związane z zarządzaniem modelami, firma może skuteczniej identyfikować, oceniać, monitorować i raportować ryzyko modeli. Polityka ta powinna być regularnie sprawdzana przez zarząd w celu zapewnienia jej aktualności oraz dostosowania do bieżących wymagań rynkowych i regulacyjnych. Inną kwestią jest proces inwentaryzacji modeli, polegający na zestawieniu wszystkich modeli używanych w organizacji, ich wersji, właścicieli, celów i ograniczeń. Regularne aktualizowanie inwentaryzacji modeli jest kluczowe dla skutecznego zarządzania ryzykiem modeli. Podczas przeglądu inwentaryzacji firma może zauważyć brak aktualizacji modeli od wielu lat. Zatem inwentaryzacja wskaże, które modele wymagają przeglądu lub aktualizacji. Zapewni to ich aktualność i dokładność. Dokumentacja modeli jest niezbędna, aby umożliwić odtworzenie modelu, a także podczas wizyt kontrolnych wytłumaczenie działania modelu. Dokumentacja powinna zawierać wszystkie informacje na temat modeli, poczynając od założeń modelu, przez opis i charakterystykę danych wejściowych, opis teoretyczny modeli, aż po interpretację wyników oraz ograniczenia modelu. Regularne przeglądy i aktualizacje dokumentacji są kluczowe dla utrzymania dokładności i aktualności modeli. Zarządzanie ryzykiem modeli i ich jakość w firmach ubezpieczeniowych są kluczowymi elementami zapewniającymi stabilność finansową oraz zgodność z regulacjami Solvency II. Bardzo ważną kwestią jest stworzenie solidnej dokumentacji modelu prognozowania rezerw technicznych, która może być aktualizowana po wprowadzeniu nowych danych. Niezbędnym elementem kontroli procesu modelowania jest zarządzanie integralnością danych wejściowych, obliczeń i wyników. Przez co minimalizuje się ryzyko błędów. Wdrożenie skutecznych procedur kontroli zmian w modelach zapewnia, że każda modyfikacja jest odpowiednio udokumentowana i zatwierdzona.

Nieodłącznym elementem jest walidacja, będąca niezależnym przeglądem i testowaniem modelu. Podczas walidacji sprawdza się scenariusze i założenia, jak również ocenia zgodność z wymogami regulacyjnymi. Systematyczny przegląd modeli wraz z ich walidacją pozwala zapewnić ich wiarygodność i adekwatność w zmieniających się warunkach rynku. Na przykład, podczas walidacji modelu kapitałowego można zidentyfikować założenie dotyczące inflacji jako potencjalne źródło ryzyka. Jeśli pojawią się niezgodności, będzie można przetestować model na scenariuszach uwzględniających różne poziomy inflacji.

Jeśli model wewnętrzny Solvency II zostanie wdrożony, może być uruchamiany co miesiąc na nowych danych. Przez to zakład cały czas będzie kontrolował poprawność modelu, jak również będzie miał wiedzę na temat aktualności danych. Dzięki temu łatwiej wychwycić błędy. W praktyce pozwala to na szybkie reagowanie i dopasowywanie modelu do aktualnej sytuacji rynkowej. Analiza różnicy między przewidywanymi a rzeczywistymi wynikami finansowymi pokazuje, czy model dobrze spełnia swoje zadanie, czy może wymaga zmian.

Podsumowując, zarządzanie oraz ocena jakości ryzyka modelu są fundamentem stabilności finansowej zakładów ubezpieczeń. Aby skutecznie zarządzać ryzykiem modeli i zapewnić zgodność z regulacjami Solvency II, ubezpieczyciele powinni wdrożyć podejście obejmujące działania takie jak:

- Stworzenie ram zarządzania Ryzykiem Modelu - opracowanie kompleksowych ram zarządzania ryzykiem modelu, które obejmują wszystkie aspekty, od identyfikacji i oceny ryzyka, po monitorowanie i raportowanie.

- Wprowadzenie standardów modelowania - ustanowienie standardów dotyczących jakości

danych, metodologii modelowania, dokumentacji oraz walidacji modeli.

- Szkolenia i edukacja - zapewnienie odpowiednich szkoleń dla pracowników odpowiedzialnych za tworzenie, walidację i używanie modeli.
- Audyt i przeglądy Wewnętrzne - regularne audyty i przeglądy wewnętrzne modeli oraz procesów zarządzania ryzykiem modelu pozwalają na wykrycie i korektę potencjalnych problemów.
- Współpraca z organami nadzoru - regularna komunikacja i współpraca z organami nadzoru w celu zapewnienia zgodności z wymaganiami Solvency II oraz uwzględnienia najnowszych wytycznych i najlepszych praktyk w zarządzaniu ryzykiem modelu.

4.4.2 Podejście ilościowe do oceny ryzyka modelu

Ocena jakościowa ryzyka modelu jest chętnie stosowana przez praktyków ze względu na swoją prostotę i łatwość wdrożenia; najczęściej opiera się na systemie punktowym, w którym eksperci przyznają punkty za kryteria takie jak jakość danych, stopień skomplikowania modelu oraz liczba przyjętych założeń. Następnie, suma punktów pozwala przypisać modelowi poziom ryzyka (niski, średni, wysoki). Metoda ta daje szybki obraz sytuacji, ale jak zauważają Black, Bonsall, Godfrey, Khalaf, and Rees (2017), duża rola subiektywnej oceny specjalistów znacząco utrudnia pełną powtarzalność wyników. Dodatkowo, sama redukcja złożoności modelu, która w takim systemie obniża jego ocenę ryzyka, wcale nie musi oznaczać rzeczywistego spadku ryzyka błędnych wyników.

Właśnie dlatego w praktyce poszukuje się także innych sposobów oceny. Jednym z nich jest traktowanie ryzyka modeli jako części ryzyka operacyjnego. W takim podejściu błędy w działaniu modeli traktuje się jak incydenty operacyjne, które można opisać pod względem częstotliwości i skali skutków. Na podstawie zebranych danych historycznych da się oszacować potencjalne straty, np. takie, które mogą wystąpić raz na 10 lub 100 lat. Niestety Prudential Regulation Authority (2023) potwierdza, że problemem jest to, iż danych o faktycznych stratach spowodowanych przez modele wciąż jest niewiele. W odpowiedzi na ograniczenia ocen jakościowych oraz traktowania ryzyka modelu jako ryzyka operacyjnego, rozwijane są ilościowe metody oceny ryzyka modelu, w szczególności oparte na analizie niepewności i porównaniu modeli alternatywnych. Black et al. (2017) oraz Institute and Faculty of Actuaries (IFoA) w (Black, Bonsall, Godfrey, Khalaf, & Rees, 2018) wskazują, że skutecznym podejściem do oceny ryzyka modelowego jest porównywanie wyników modeli opartych na różnych założeniach. Im większe rozbieżności lub szerszy przedział wyników, tym wyższe ryzyko modelowe. Podobne podejście rozwijane jest w ramach koncepcji nazywanych przedziałami niepewności. W tym podejściu, zamiast jednej wartości miary, takiej jak Value at Risk, wyznacza się jej przedział dla różnych akceptowalnych modeli (Bernard, Rüschendorf, & Vanduffel, 2017b). Dzięki temu można uchwycić skalę potencjalnych odchyleń wyników i lepiej ocenić wiarygodność samej miary. Inną propozycją jest wykorzystanie entropii względnej do określenia odległości pomiędzy modelem bazowym a modelami alternatywnymi. W ten sposób Glasserman and Xu (2014) i Breuer and Csiszár (2016) wyznaczają maksymalną wartość miary ryzyka. W literaturze pojawiają się również prace wskazujące na konieczność badania wrażliwości modelu na poszczególne założenia (Jacobs, 2015). Takie podejście umożliwia zidentyfikowanie tych elementów konstrukcji, które w największym stopniu poszerzają przedział niepewności, a tym samym

generują największe ryzyko. Dzięki temu, ilościowa ocena ryzyka modelu staje się nie tylko sposobem na wyznaczenie bardziej konserwatywnych miar, ale także praktycznym narzędziem wspierającym proces walidacji i poprawy samych modeli. W (Bernard, Kazzi, & Vanduffel, 2023) zaproponowano szereg narzędzi ilościowych służących do oceny ryzyka modelu. Jednym z nich jest miara warunkowego wkładu w redukcję ryzyka modelu (ang. Conditional Model Risk Contribution Measure). Jej celem jest dekompozycja całkowitego ryzyka modelowanego zjawiska na części pochodzące z poszczególnych grup założeń. Miara ta opiera się na analizie szerokości przedziału niepewności danej miary ryzyka (np. VaR). Dla wybranej podgrupy założeń porównuje się krańcowe wartości tej miary w sytuacji, gdy założenia te są uwzględnione, z wartościami odpowiadającymi przypadkowi ich pominięcia. Otrzymany wskaźnik określa względne zawężenie przedziału niepewności, a tym samym udział rozpatrywanych założeń w całkowitym ryzyku modelowym.

Miara ryzyka ϱ definiowana jest jako odwzorowanie:

$$\varrho : \mathcal{M} \to \mathbb{R}, \tag{4.1}$$

gdzie $\mathcal{M}$ oznacza zbiór wszystkich rzeczywistych lub wielowymiarowych zmiennych losowych określonych na ustalonej przestrzeni probabilistycznej $(\Omega, \mathcal{F}, \mathbb{P})$:

$$\mathcal{M} = \{\, X : \Omega \to \mathbb{R}^d \mid X \text{ jest } \mathcal{F}\text{-mierzalna} \,\}. \tag{4.2}$$

Zbiór $\mathcal{M}$ stanowi zatem przestrzeń wszystkich dopuszczalnych modeli,na których możliwe jest określenie miary ryzyka ϱ. Dalej rozważany jest zbiór n założeń $\{a_i\}_{i=1}^n$, które definiują dany model. Dla wybranej podgrupy założeń analizowany jest ich wpływ na zakres niepewności wartości miary ryzyka. W tym celu wprowadza się zstępujący ciąg zbiorów:

$$A_k = \{\, Y \in \mathcal{M} : Y \text{ spełnia założenia } \{a_i\}_{i \le k} \,\}. \tag{4.3}$$

Zbiór A_k opisuje klasę wszystkich modeli zgodnych z przyjętymi założeniami do poziomu k. Przy założeniu, że obowiązują jedynie elementy A_k, można określić odpowiadające im górne i dolne granice miary ryzyka ϱ. Umożliwia to ilościową ocenę wpływu kolejnych założeń $\{a_i\}$ na przedział niepewności tej Dla danej miary ryzyka ϱ definiujemy:

$$C(\rho, \mathcal{A}_k, \mathcal{A}_l) = 1 - \frac{\bar{\varrho}_l - \underline{\varrho}_l}{\bar{\varrho}_k - \underline{\varrho}_k} \tag{4.4}$$

gdzie:

$$\bar{\varrho}_k = \sup_{X \in \mathcal{A}_k} \varrho(X), \quad \underline{\varrho}_k = \inf_{X \in \mathcal{A}_k} \varrho(X),$$
$$\bar{\varrho}_l = \sup_{X \in \mathcal{A}_l} \varrho(X), \quad \underline{\varrho}_l = \inf_{X \in \mathcal{A}_l} \varrho(X),$$

a $\mathcal{A}_k$, $\mathcal{A}_l$ oznaczają zbiory modeli zgodnych odpowiednio z mniejszym i większym zestawem znanych założeń.

Wartość C mieści się w przedziale $[0, 1]$ i informuje, jak duży wpływ ma dana grupa założeń na całkowitą niepewność modelu. Im wyższa wartość C, tym większa istotność badanych założeń. Bardzo często nie wszystkie założenia można jednoznacznie uznać za poprawne lub stwierdzić, że są falszywe. Założenia weryfikuje się na podstawie danych empirycznych, testów

statystycznych lub opinii ekspertów, przez co nie wszystkie są istotne w ten sam sposób. Miarą, która uwzględnia różne wagi dla założeń, jest miara oparta na wiarygodności tych założeń.

Niech $z_j \in [0, 1]$ oznacza stopień wiarygodności przypisany założeniu a_j, przy założeniu że wcześniejsze założenia $\{a_i\}_{i<j}$ są znane i zaakceptowane. Oznaczmy przez $\bar{\varrho}_j$ oraz $\underline{\varrho}_j$ odpowiednio górną i dolną granicę ryzyka wyznaczoną w zbiorze modeli spełniających założenia do j-tego poziomu.

Wówczas górna granica (ang. Credibility-based Upper Bound - CUB) i dolna granica (ang. Credibility-based Lower Bound - CLB) ryzyka, z uwzględnieniem wiarygodności założeń, wyrażają się wzorami:

$$\text{CUB}\left(\varrho, \{A_m\}_m, \{z_j\}_j\right) = \bar{\varrho}_k + \sum_{m=r+1}^{n} \left(\prod_{j=r+1}^{m} z_j \right) (\bar{\varrho}_m - \bar{\varrho}_{m-1}) \tag{4.5}$$

$$\text{CLB}\left(\varrho, \{A_m\}_m, \{z_j\}_j\right) = \underline{\varrho}_k + \sum_{m=r+1}^{n} \left(\prod_{j=r+1}^{m} z_j \right) \left(\underline{\varrho}_m - \underline{\varrho}_{m-1}\right) \tag{4.6}$$

Im większa wartość z_j, tym większe zaufanie do danego założenia, a tym samym węższy przedział niepewności. W przypadku gdy $z_j = 1$ dla wszystkich j, otrzymujemy klasyczny scenariusz pełnego zaufania do założeń. Natomiast, gdy $z_j = 0$ dane założenie nie wpływa na ograniczenie przedziału, model pozostaje bardziej konserwatywny. Z praktycznego punktu widzenia takie podejście pozwala na bardziej elastyczne i realistyczne modelowanie ryzyka modelu. Umożliwia uwzględnienie wiedzy statystycznej lub eksperckiej w ocenie ryzyka modelu.

Kolejnym podejściem do ilościowej oceny ryzyka modelu są miary porównujące wartość miary ryzyka wyznaczonej na podstawie danego modelu z wartością referencyjną, obliczoną dla modelu bazowego lub obserwacji empirycznych. Wyróżniamy tutaj dwie kategorie:

Bezwzględna Miara Ryzyka Modelowego (ang. Absolute Measure – AM) mierzy, o ile ryzyko oszacowane przez model odbiega od wartości referencyjnej w skali procentowej. Wyraża się wzorem:

$$\text{AM}(\varrho, A_k, Y^*) = \frac{\bar{\varrho}_k - \varrho(Y^*)}{\varrho(Y^*)} \tag{4.7}$$

gdzie $\bar{\varrho}_k$ to górna granica miary ryzyka zgodnie z zestawem założeń A_k, zaś $\varrho(Y^*)$ oznacza wartość tej miary wyliczoną dla referencyjnego modelu Y^*. Miara ta interpretuje ryzyko modelu jako względny błąd oszacowania. Dodatnia wartość oznacza przeszacowanie ryzyka, a ujemna - jego niedoszacowanie.

Względna Miara Ryzyka Modelowego (ang. Relative Measure – RM) porównuje rozstęp pomiędzy ryzykiem oszacowanym przez model a wartością referencyjną względem całkowitego rozstępu niepewności wyznaczonego przez granice modelu. Jej postać to:

$$\text{RM}(\varrho, A_k, Y^*) = \frac{\bar{\varrho}_k - \varrho(Y^*)}{\bar{\varrho}_k - \underline{\varrho}_k} \tag{4.8}$$

Miara RM przyjmuje wartości w przedziale $[0, 1]$ i informuje, jak blisko (procentowo) znajduje się wartość referencyjna względem „najgorszego przypadku" przewidzianego przez model. Im wyższa wartość RM, tym bardziej ryzyko referencyjne odbiega od modelu, jednocześnie wskazując na potencjalnie wysokie ryzyko modelu.

Zarówno AM, jak i RM umożliwiają ocenę ryzyka modelu z punktu widzenia odchyleń względem wartości odniesienia, przy czym RM dodatkowo uwzględnia pełny zakres niepewności w ocenie.

Na podstawie CLB i CUB definiuje się dwie nowe miary: miarę absolutną opartą na wiarygodności (ang. Credibility-based Absolute Measure - CAM) oraz miarę względną opartą na wiarygodności (ang. Credibility-based Relative Measure - CRM). Ich celem jest rozszerzenie klasycznych miar AM i RM o element wiarygodności przypisanej poszczególnym założeniom modelu. Dzięki temu, pozwalają uzyskać bardziej informacyjną i praktyczną ocenę ryzyka modelowego, zwłaszcza w sytuacjach, gdy klasyczne granice niepewności są zbyt szerokie i mało użyteczne w interpretacji.

CAM mierzy relatywną różnicę między wartością miary ryzyka oszacowaną przez model a górną granicą przedziału CUB:

$$\mathrm{CAM}(\varrho, CUB, CLB, X^*) = \frac{CUB - \varrho(X^*)}{\varrho(X^*)}. \tag{4.9}$$

Miara ta pokazuje, o ile wyższe może być ryzyko w scenariuszu najbardziej niekorzystnym (przy danych wiarygodnościach założeń) w stosunku do wartości wynikającej z modelu.

CRM natomiast ocenia położenie wartości modelowej względem całego przedziału niepewności:

$$\mathrm{CRM}(\varrho, CUB, CLB, X^*) = \frac{CUB - \varrho(X^*)}{CUB - CLB}. \tag{4.10}$$

Wartości bliskie 0 wskazują, że wynik modelu znajduje się blisko górnej granicy, a ryzyko niedoszacowania jest niewielkie; natomiast wartości wysokie oznaczają, że wynik leży bliżej dolnej granicy, co sugeruje większą podatność modelu na ryzyko błędnych założeń. W praktyce CAM i CRM umożliwiają bardziej elastyczną i realistyczną ocenę ryzyka modelowego niż klasyczne miary AM i RM, ponieważ pozwalają uwzględnić różną wiarygodność poszczególnych założeń, opartą na testach statystycznych lub opiniach ekspertów.

W kolejnych przykładach pokażemy praktyczne zastosowanie tych miar i wprowadzimy nową miarę, która pozwala przeliczyć ocenę ryzyka modelowego na wymaganą rezerwę kapitałową (ang. Model Risk Capital – MoRC).

Przykład 1.

W formule standardowej Solvency II SCR dla modułów ryzyka *Life* i *Health* wyznaczany jest na drodze agregacji, wykorzystując metodę wariancji–kowariancji, z ustaloną dla wszystkich zakładów ubezpieczeń Unii Europejskiej wartością współczynnika korelacji $\rho_{\text{Life, Health}} = 0.25$. Aby przeanalizować ryzyko modelu wynikające z tego założenia, przyjmujemy, że znane są tylko rozkłady brzegowe tych modułów $N(0, 392)$, $N(0, 248)$.

Znając rozkłady brzegowe, rozważamy sześć alternatywnych opisów zależności: pięć kopul Gaussa z korelacjami $\rho \in \{-1, 0, 0.25, 0.50, 1\}$ oraz kopulę t-Studenta z $df = 4$ i $\rho = 0.25$, która modeluje zależność w ogonie rozkładu. Wyniki symulacji przedstawia Tabela 4.1.

Gdy nie znamy w ogóle struktury zależności, dopuszczalny zbiór modeli $\mathcal{A}_k$ prowadzi do granic $\underline{\varrho}_k = 370.92$ (ujemna korelacja) i $\bar{\varrho}_k = 1648.53$ (komonotoniczność), a pełny przedział niepewności wynosi $\Delta_k = 1277.61$. Jeśli mamy informacje o tym, że korelacje są w przedziale $\rho \in [0, 0.5]$ (zbiór $\mathcal{A}_l$), wówczas przedział zawęża się do $\underline{\varrho}_l = 1194.83$ i $\bar{\varrho}_l = 1439.67$, czyli $\Delta_l = 244.84$.

Scenariusz	**Kopula**	ϱ	$\mathbf{VaR_{0.995}}$
Ujemna korelacja	Gauss	-1	370.92
Brak korelacji	Gauss	0	1194.83
Formuła Standardowa	**Gauss**	**0.25**	**1322.92**
Silniejsza korelacja	Gauss	0.50	1439.67
Komonotoniczna	Gauss	1	1648.53
t-Student copula	t-Student ($df = 4$)	0.25	1406.87

Tabela 4.1: VaR dla różnych struktur zależności w kontekście ryzyka modelu.
Źródło: Opracowanie własne.

Na tej podstawie obliczamy wartości miar ryzyka modelu wprowadzone powyżej. Warunkowa miara kontrybucji (C przyjmuje wartość

$$C = C(\varrho, \mathcal{A}_k, \mathcal{A}_l) = 1 - \frac{\bar{\varrho}_l - \underline{\varrho}_l}{\bar{\varrho}_k - \underline{\varrho}_k} = 1 - \frac{244.84}{1277.61} \approx 0.81,$$

co oznacza, że ograniczenie korelacji do przedziału $[0, 0.5]$ redukuje niepewność wynikającą z zależności o blisko 19 %. Bezwzględna miara AM, odniesiona do wartości referencyjnej $\rho = 0.25$, wynosi

$$\text{AM} = \frac{\bar{\varrho}_k - \varrho(Y^*)}{\varrho(Y^*)} = \frac{1648.53 - 1322.92}{1322.92} \approx 0.25,$$

czyli najgorszy przypadek VaR jest o 25% wyższy niż wynik formuły standardowej. Względna miara RM przyjmuje identyczną wartość 0.25, co umieszcza scenariusz referencyjny w pierwszej ćwiartce najgorszego przypadku. Przy poziomie wiarygodności $z = 0.7$ przedziały CUB/CLB dają odpowiednio

$$\text{CLB} = 947.66, \qquad \text{CUB} = 1502.32,$$

a zatem praktyczny przedział $[947.66,\ 1502.32]$ jest o 26% węższy od pełnego $[370.92,\ 1648.53]$.

Na podstawie granic CLB i CUB można także obliczyć miary CAM oraz CRM. Dla wartości $\varrho(Y^*) = 1322.92$ (formuła standardowa) otrzymujemy:

$$\text{CAM} = \frac{\text{CUB} - \varrho(Y^*)}{\varrho(Y^*)} = \frac{1502.32 - 1322.92}{1322.92} \approx 0.14,$$

co oznacza, że w scenariuszu najbardziej niekorzystnym ryzyko może być o około 14% wyższe niż wartość wskazana przez formułę standardową. Z kolei

$$\text{CRM} = \frac{\text{CUB} - \varrho(Y^*)}{\text{CUB} - \text{CLB}} = \frac{1502.32 - 1322.92}{1502.32 - 947.66} \approx 0.32,$$

czyli wartość leży w około jednej trzeciej odległości od górnej granicy praktycznego przedziału. Wskazuje to, że mimo redukcji niepewności poprzez uwzględnienie wiarygodności założeń nadal istnieje zauważalne ryzyko niedoszacowania, ale jest ono mniejsze niż sugerowałyby miary AM i RM. Ustalona w dyrektywie korelacja 0.25 daje SCR (1323) stosunkowo blisko środka

częściowego przedziału [1195, 1440], ale jednocześnie jedynie 25% odległości od najgorszego scenariusza, co w świetle miary RM wskazuje na umiarkowanie wysokie ryzyko modelu. Już ogólne stwierdzenie, że korelacja nie przekracza 0.5, zmniejsza niepewność zależności o jedną piątą. Ponadto, kopula t-Studenta przy tej samej korelacji wykazuje większy o około 6% VaR niż kopula normalna, co może wskazywać na to, iż kopula normalna nie oddaje zależności w ogonie. Przykład ten pokazuje, jak miary C, CUB, CLB, AM i RM stanowią spójne narzędzia ilościowej oceny ryzyka modelu, wynikającego z niepewności struktury zależności w formule standardowej Solvency II.

Ostatnim omawianym wskaźnikiem służącym do oceny ryzyka modelowego jest MoRC. Stanowi on miarę określającą dodatkową rezerwę kapitałową, którą instytucja powinna utrzymywać w celu zabezpieczenia się przed niepewnością wynikającą z zastosowanego modelu. MoRC łączy ocenę względnego ryzyka modelowego (CRM) z rozpiętością granic wiarygodności ryzyka (CUB oraz CLB) i formalnie definiuje się go jako:

$$MoRC(CRM, CUB, CLB, f) = f(CRM) \cdot (CUB - CLB), \tag{4.11}$$

gdzie $f : [0, 1] \to [0, 1]$ jest funkcją rosnącą, spełniającą warunki $f(0) = 0$ oraz $f(1) = 1$. Wartość $f(CRM)$ interpretuje się jako udział maksymalnego wymaganego kapitału. Przyjmuje wartośc od 0% przy braku ryzyka do 100% w przypadku jego maksymalnego poziomu.

Przykład 2.

Na podstawie wyników z Przykładu 1 można obliczyć dodatkową rezerwę kapitałową związaną z ryzykiem modelu. Do wyznaczenia MoRC wykorzystujemy otrzymane granice wiarygodności CLB = 947.66 oraz CUB = 1502.32, a także względną miarę ryzyka modelu CRM $\approx$ 0.32. Inspirując się ramami Solvency II, przyjmujemy funkcję $f(x) = \sqrt{x}$, która odzwierciedla konserwatywne podejście. Nawet umiarkowane wartości CRM prowadzą do stosunkowo wysokiego bufora kapitałowego. Podstawiając wartości z Przykładu 1, otrzymujemy:

$$MoRC = \sqrt{0.32} \times (1502.32 - 947.66) \approx 314.0.$$

Otrzymany wynik oznacza, że wymagany bufor kapitałowy stanowi około 314 jednostek, czyli 23.7% wartości referencyjnego $SCR = 1322.92$.

Przykład ten pokazuje, że MoRC pozwala na praktyczne przełożenie oceny ryzyka modelu (CRM wraz z granicami CLB i CUB) na wielkość dodatkowej rezerwy kapitałowej. Ponadto, wybór funkcji f umożliwia dopasowanie poziomu bezpieczeństwa do wymogów regulacyjnych oraz polityki zarządzania ryzykiem. W tym przypadku zastosowanie funkcji $\sqrt{x}$ odpowiada ostrożnemu podejściu zgodnemu z dyrektywą Solvency II, w której szczególną uwagę poświęca się wypłacalności zakładu ubezpieczeń.

5. Wyznaczanie efektu dywersyfikacji w ramach modeli wielowymiarowych opartych na kopulach i sieciach neuronowych

Efekt dywersyfikacji stanowi jeden z najistotniejszych elementów w ocenie wypłacalności zakładów ubezpieczeń w ramach dyrektywy Solvency II. Portfel ubezpieczyciela obejmuje wiele rodzajów ryzyk, które nie występują w tym samym czasie. Zatem całkowite ryzyko portfela ubezpieczyciela jest zazwyczaj niższe niż suma ryzyk, które się w nim znajdują. Wyznaczenie efektu dywersyfikacji na właściwym poziomie pozwala na określenie poziomu wymaganego kapitału, zapewniającego jednocześnie adekwatny poziom bezpieczeństwa finansowego zakładu. W badaniach przeprowadzonych przez EIOPA wskazano jednoznacznie, iż sposób modelowania zależności jest jednym z najważniejszych elementów wyznaczania efektu dywersyfikacji. Dlatego kluczową kwestią w procesie modelowania ryzyka jest odpowiedni dobór metody, pozwalającej na jak najlepsze opisanie struktury zależności pomiędzy agregowanymi czynnikami ryzyka. Tylko wówczas możliwe jest uzyskanie efektu dywersyfikacji na poziomie odpowiadającym rzeczywistym korzyściom z agregacji ryzyka, a tym samym zapewnienie spójności pomiędzy modelem matematycznym a faktycznym profilem ryzyka ubezpieczyciela. Dla danych rzeczywistych niemieckich ubezpieczycieli pozyskanych z raportu SFCR, w oparciu o twierdzenie Sklara, zbudowano wielowymiarowe rozkłady w dwóch podejściach:

1. Parametrycznym - postać wielowymiarowego rozkładu opiera się na z góry ustalonej postaci analitycznej rozkładów brzegowych i kopuli. Strukturę zależności wyznaczono z wykorzystaniem kaskad kopul C (C-vine) i D (D-vine).

2. Nieparametryczny - postać wielowymiarowego rozkładu nie zakłada z góry konkretnej postaci rozkładów brzegowych ani struktury zależności. Zarówno rozkłady brzegowe, jak i struktura zależności między nimi zostały wyznaczone przy użyciu modelu opartego na sieciach neuronowych. Pozwala to na uchwycenie nieliniowych i złożonych zależności, które często występują w rzeczywistych danych ubezpieczeniowych.

W dalszej części pracy zrealizowano cel badawczy C6, polegający na wskazaniu i rozwinięciu metod modelowania zależności łączących kopule oraz techniki uczenia maszynowego, które zastosowane w procesie agregacji pozwalają wiarygodnie uchwycić rzeczywistą strukturę współzależności między ryzykami. Tym samym umożliwiają określenie kapitałowego wymogu wypłacalności oraz efektu dywersyfikacji na adekwatnym poziomie.

Badania prowadzone są na danych rzeczywistych pozyskanych z raportów SFCR dotyczących niemieckich ubezpieczycieli, obejmujących cztery segmenty działalności w zakresie ubezpieczeń majątkowych i osobowych. W dalszej części rozdziału szczegółowo przedstawiono zastosowanie sieci neuronowych w identyfikacji wielowymiarowego rozkładu, opisano przyjętą metodykę badań empirycznych oraz charakterystykę danych wykorzystanych w badaniu. Następnie zaprezentowano sposób wyznaczania wielowymiarowego rozkładu oraz wyznaczono efekt dywersyfikacji.

5.1 Metodyka badania

5.1.1 Modelowane zmienne ryzyka

Z uwagi na brak ogólnodostępnych danych na potrzeby niniejszego badania, ryzyko składki i rezerw dla poszczególnych segmentów modelowane jest za pomocą współczynnika zespolonego wprowadzonego w (Schubert & Grießmann, 2007), danego poniższym wzorem:

$$X_i = \frac{NCE_i + OE_i}{NP_i} \tag{5.1}$$

gdzie dla i-tego segmentu NCE_i oznacza odszkodowania i świadczenia netto (bezpośrednia działalność ubezpieczeniowa), OE_i to koszty poniesione, natomiast NP_i to składki zarobione netto (bezpośrednia działalność ubezpieczeniowa). Zmienną losową L modelującą zagregowane ryzyko segmentów otrzymuje się, agregując współczynniki zespolone X_i za pomocą następującej formuły:

$$Y = \sum_{i=1}^{d} X_i \tag{5.2}$$

Wymogi kapitałowe dla poszczególnych segmentów $\kappa(L_i)$ oraz zagregowanego ryzyka $\kappa(L)$ wyznaczono w następujący sposób:

$$\kappa(X_i) = VaR_{0.995}(X_i) - \mathbb{E}(X_i) \tag{5.3}$$

$$\kappa(Y) = VaR_{0.995}(Y) - \mathbb{E}(Y) \tag{5.4}$$

efekt dywersyfikacji zgodnie ze wzorem:

$$ED = 1 - \frac{\kappa(Y)}{\sum_{i=1}^{d} \kappa(X_i)} \tag{5.5}$$

Dane rzeczywiste zostały podzielone na dwa zbiory: treningowy, służący do dopasowania modeli, oraz testowy, wykorzystywany do oceny jakości dopasowania. Na zbiorze treningowym wyznaczane są zarówno rozkłady brzegowe o charakterze parametrycznym, jak i rozkłady modelowane za pomocą sieci neuronowych. W celu oceny jakości dopasowania wyznaczonych rozkładów do danych rzeczywistych przeprowadza się porównanie z rozkładem empirycznym z wykorzystaniem testu Kołmogorowa–Smirnowa. Strukturę zależności między zmiennymi X_i, kluczową z punktu widzenia analizy efektu dywersyfikacji, modeluje się przy użyciu dwóch metod:

1. Podobnie jak w przypadku wyznaczania rozkładów brzegowych, stosowane jest podejście parametryczne, w którym strukturę zależności modeluje się za pomocą kaskady kopuli typu C-vine oraz D-vine.

2. W podejściu nieparametrycznym zależności są modelowane z wykorzystaniem sieci neuronowych.

W kolejnym etapie, zgodnie z twierdzeniem Sklara, konstruowane są rozkłady wielowymiarowe, stanowiące połączenie rozkładów brzegowych oraz funkcji kopuli, opisującej zależności między zmiennymi. W podejściu parametrycznym rozkłady brzegowe łączone są z kopulami

typu C-vine i D-vine, co umożliwia uzyskanie parametrycznego rozkładu wielowymiarowego. Natomiast, wykorzystanie rozkładów brzegowych oraz kopuli identyfikowanych przy użyciu sieci neuronowych pozwala na uzyskanie nieparametrycznego rozkładu wielowymiarowego

5.1.2 Dopasowanie rozkładu wielowymiarowego z wykorzystaniem sieci neuronowych

Modele sieci neuronowych można postrzegać jako zdefiniowaną funkcję, która pobiera dane wejściowe (argumenty funkcji) i generuje wynik (wartości funkcji). Budując odpowiednią architekturę sieci neuronowej i określając sposób jej uczenia, sieć może zostać wykorzystana do opisu różnych problemów. W niniejszej pracy sieć neuronowa traktowana jest jako estymator rozkładu wielowymiarowego. Celem jest opracowanie modelu sieci neuronowej zdolnego do odwzorowania rozkładu danych empirycznych oraz generowania nowych realizacji zgodnych z oszacowanym modelem. Przyjęte podejście opiera się na twierdzeniu Sklara, które umożliwia reprezentację rozkładu wielowymiarowego jako połączenie rozkładów brzegowych za pomocą kopuli:

$$F\left(x_1, x_2, \ldots, x_d\right) = C\left(F_1\left(x_1\right), F_2\left(x_2\right), \ldots, F_d\left(x_d\right)\right), \tag{5.6}$$

gdzie $F\left(x_1, x_2, \ldots, x_d\right)$ to dystrybuanta wielowymiarowego rozkładu, natomiast $F_j\left(x_j\right)$ to dystrybuanty rozkładów brzegowych, a C jest kopulą opisującą strukturę zależności między segmentami. Dla każdego segmentu $j \in \{1, 2, \ldots, d\}$ na podstawie danych rzeczywistych uczona jest sieć neuronowa, estymująca rozkład prawdopodobieństwa opisany przez $F_j(x_j)$. Każda zmienna ma swój własny rozkład prawdopodobieństwa, zatem trenowanych jest d niezależnych sieci neuronowych. Na każdą z rozważanych sieci neuronowych nałożono ograniczenia wynikające z założeń teorii prawdopodobieństwa, tak aby ich architektura gwarantowała, że generowane wartości wyjściowe spełniają własności dystrybuanty jednowymiarowej. Przez parametr $\theta = (\theta_1, \ldots, \theta_d)$ oznaczany jest zbiór wag dla każdej wyznaczonej sieci neuronowej dla każdego z rozkładów. Dla każdego F_j przypisany jest oddzielny zestaw parametrów θ_j, dostrajany w procesie uczenia, tak aby $\hat{F}_j\left(x_j, \theta_j\right)$ jak najlepiej odpowiadała rzeczywistemu rozkładowi. Po wyznaczeniu rozkładów brzegowych $\hat{F}_j(x_j, \theta_j)$ budowana jest osobna sieć neuronowa estymująca kopulę $C(u)$, gdzie $u = (u_1, \ldots, u_d)$, a $u_j = \hat{F}_j(x_j, \theta_j)$. Konstrukcji podlega jedna architektura sieci, która następnie jest trenowana, a jej zadaniem jest uchwycenie zależności między agregowanymi zmiennymi losowymi. Podobnie jak w przypadku rozkładów brzegowych, na sieć nakładane są ograniczenia wynikające z teorii kopuli. Przez parametry θ_c oznaczany jest zestaw wag i biasów sieci neuronowej, estymującej kopulę. Parametry te dostrajane są w procesie uczenia tak, aby $\hat{C}(u, \theta_c)$ jak najlepiej odzwierciedlała rzeczywistą strukturę zależności pomiędzy zmiennymi. Po wyestymowaniu rozkładów brzegowych oraz kopuli, zgodnie z twierdzeniem Sklara, dystrybuantę wielowymiarowego rozkładu można zapisać jako funkcję argumentu $\mathbf{x} = (x_1, x_2, \ldots, x_d)$, reprezentującego wartości poszczególnych zmiennych losowych w przestrzeni $\mathbb{R}^d$:

$$\hat{F}\left(\mathbf{x}; \theta, \theta_c\right) = \hat{C}\left(\hat{F}_1\left(x_1; \theta_1\right), \hat{F}_2\left(x_2; \theta_2\right), \ldots, \hat{F}_d\left(x_d; \theta_d\right); \theta_c\right), \tag{5.7}$$

Ponieważ zarówno rozkłady brzegowe, jak i kopula są różniczkowalne, można wyznaczyć gęstość rozkładu wielowymiarowego jako kombinację gęstości rozkładów brzegowych oraz gęstości kopuli:

$$\hat{f}\left(x;\theta,\theta_c\right)=\hat{c}\left(\hat{F}_1\left(x_1;\theta_1\right),\hat{F}_2\left(x_2;\theta_2\right),\ldots,\hat{F}_d\left(x_d;\theta_d\right);\theta_c\right)\cdot\prod_{j=1}^{d}\hat{f}_j\left(x_j;\theta_j\right), \tag{5.8}$$

gdzie $\hat{f}_j\left(x_j;\theta_j\right)$ to wyestymowane gęstości rozkładów brzegowych, a $\hat{c}\left(u;\theta_c\right)$ to wyestymowana gęstość kopuli opisująca strukturę zależności.

W praktyce dane wykorzystywane do uczenia sieci neuronowej mają postać macierzy obserwacji:

$$\mathbf{X}=\begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,d} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,d} \end{bmatrix}, \tag{5.9}$$

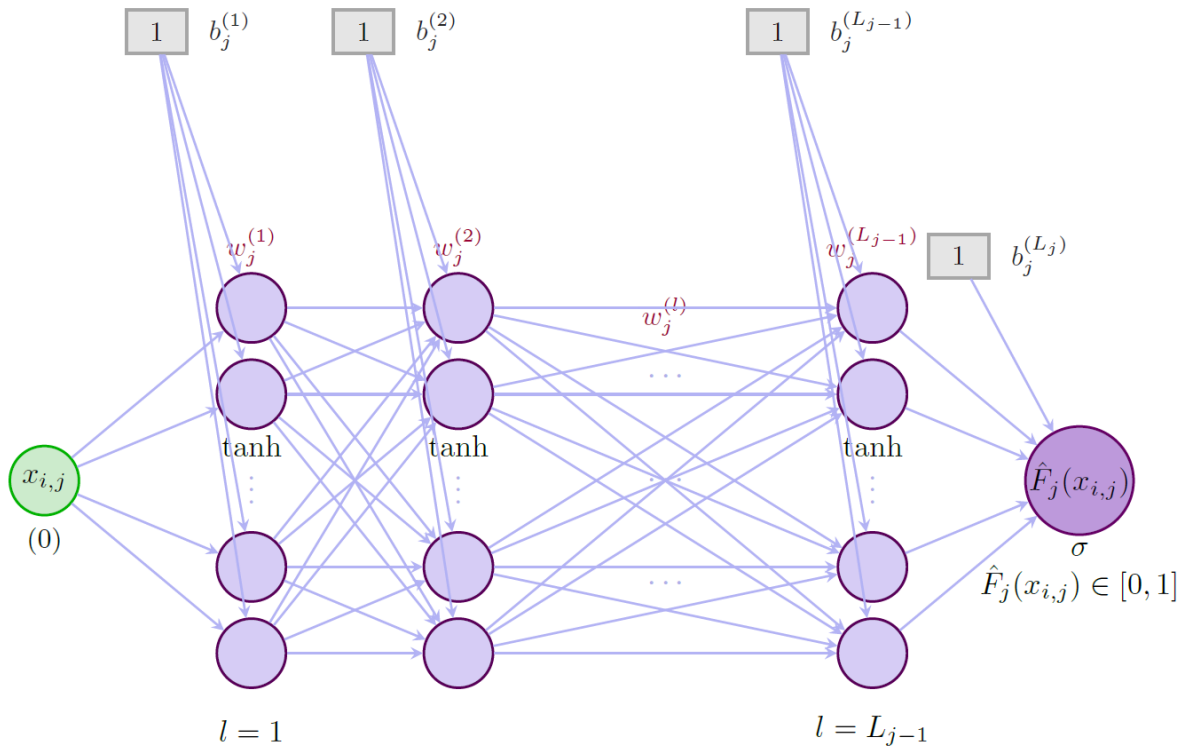
gdzie $x_{i,j}$ oznacza wartość i-tej obserwacji zmiennej losowej X_j. Każdy wiersz macierzy $\mathbf{X}$, zapisany jako wektor $\mathbf{x}_i=(x_{i,1},x_{i,2},\ldots,x_{i,d})$, reprezentuje pojedynczą obserwację wektora losowego $\mathbf{X}=(X_1,X_2,\ldots,X_d)$, natomiast każda kolumna zawiera wszystkie realizacje jednej zmiennej X_j. Taki zapis danych pozwala na bezpośrednie wykorzystanie ich w procesie uczenia sieci neuronowej, zarówno w modelach estymujących rozkłady brzegowe, jak i w modelu kopuli neuronowej opisującym strukturę zależności pomiędzy zmiennymi.

5.1.2.1 Dopasowanie rozkładów brzegowych

Wyznaczenie rozkładów brzegowych stanowi pierwszy etap budowy wielowymiarowego rozkładu opartego na sieciach neuronowych. Na tym etapie szacowane są dystrybuanty oraz odpowiadające im funkcje gęstości dla każdej zmiennej wejściowej, które następnie stanowią podstawę do wyznaczenia kopuli opisującej zależności między zmiennymi. Poniżej kolejno opisana została architektura sieci oraz proces jej uczenia dla rozkładów brzegowych.

Architektura sieci

Schemat architektury sieci neuronowej wykorzystanej do estymacji jednowymiarowej dystrybuanty rozkładu prawdopodobieństwa dla j-tej zmiennej losowej przedstawiono na poniższym rysunku (rys. 5.1). Sieć neuronowa składa się z trzech warstw: wejściowej, ukrytych oraz wyjściowej.



Rysunek 5.1: Architektura sieci neuronowej dla rozkładów brzegowych.
Źródło: Opracowanie własne.

Warstwa wejściowa sieci jest jednoargumentowa $x_{i,j} \in \Omega \subset \mathbb{R}$:

$$h^{(0)} = x_{i,j} \in \Omega \subset \mathbb{R}. \tag{5.10}$$

Każda z kolejnych warstw ukrytych $l \in \{0, \ldots, L_j - 1\}$ dokonuje liniowego przekształcenia wektora wyjściowego z warstwy poprzedniej poprzez zastosowanie wag $w_j^{(l)}$ oraz dodanie wektora przesunięcia $b_j^{(l)}$ (bias). Następnie, aby wprowadzić nieliniowość, na wynik działania warstwy liniowej nakładana jest funkcja aktywacji. W zastosowanej sieci funkcją aktywacji jest tangens hiperboliczny, który zapewnia gładkość odwzorowania, stabilizując wartości wyjściowe w przedziale $(-1, 1)$. Formalnie proces zachodzący w $l - tej$ warstwie ukrytej można opisać równaniem:

$$h_j^{(l+1)} = \tanh\left(w_j^{(l)} h_j^{(l)} + b_j^{(l)}\right). \tag{5.11}$$

W warstwie wyjściowej stosowana jest funkcja aktywacji sigmoid, która zapewnia, że wynik sieci przyjmuje wartości w przedziale $[0, 1]$, co jest zgodne z wartościami dystrybuanty:

$$\hat{F}_j = \text{sigmoid}\left(w_j^{(L_j)} h_j^{(L_j)} + b_j^{(L_j)}\right) \in [0, 1]. \tag{5.12}$$

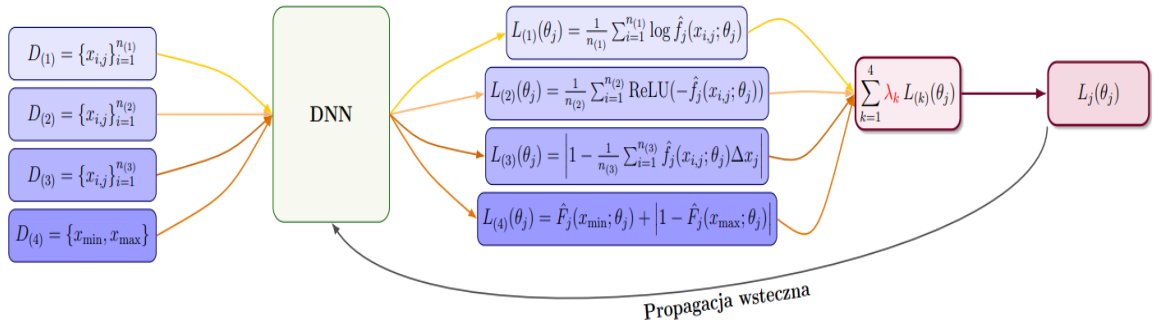
Po wyznaczeniu dystrybuanty rozkładu prawdopodobieństwa $\hat{F}_j(x_{i,j}, \theta_j)$, gdzie parametry $\theta_j = \{w_j^{(l)}, b_j^{(l)}\}$ odpowiadają wagom i biasom sieci, gęstość prawdopodobieństwa może zostać wyznaczona jako pochodna dystrybuanty względem $x_{i,j}$:

$$\hat{f}_j(x_{i,j}, \theta_j) = \frac{d\hat{F}_j(x_{i,j}, \theta_j)}{dx_{i,j}}. \tag{5.13}$$

Dzięki temu, że sieć zbudowana jest z funkcji różniczkowalnych (tanh w warstwach ukrytych oraz sigmoid w wyjściu), pochodna $\hat{F}_j$ może zostać wyznaczona wprost za pomocą mechanizmu automatycznego różniczkowania Baydin, Pearlmutter, Radul, and Siskind (2018) dostępnego w bibliotekach uczenia maszynowego (np. TensorFlow lub PyTorch). W ten sposób sieć dla rozkładów brzegowych dostarcza zarówno dystrybuantę, jak i gęstość, bez konieczności wprowadzania dodatkowych założeń co do postaci rozkładu.

Proces uczenia sieci

Proces uczenia sieci dla rozkładów brzegowych przedstawiono na poniższym schemacie (rys. 5.2). Dane treningowe sieci można podzielić na cztery zbiory, które kolejno podawane są do sieci, a na ich podstawie obliczane są poszczególne składowe funkcji kosztu odpowiadające teoretycznym założeniom rozkładu prawdopodobieństwa. Wartości nakładanych na sieć kar są następnie sumowane w jedną łączną funkcję straty, na podstawie której parametry sieci są aktualizowane metodą propagacji wstecznej. Poniżej szczegółowo opisano proces uczenia.



Rysunek 5.2: Procedura nauki sieci neuronowej dla rozkładów brzegowych.
Źródło: Opracowanie własne.

Pierwszy zbiór treningowy oznaczany jest jako

$$D_{(1)} = \{ x_{i,j} \}_{i=1}^{n_{(1)}}, \tag{5.14}$$

gdzie $n_{(1)}$ oznacza liczbę obserwacji w tym zbiorze. Dane stanowią rzeczywiste obserwacje, do których dopasowywany jest rozkład prawdopodobieństwa. Na podstawie zbioru $D_{(1)}$ definiowany jest następujący pierwszy składnik funkcji straty, odpowiadający logarytmowi wiarygodności:

$$L_{(1)}(\theta_j) = -\frac{1}{n_{(1)}} \sum_{i=1}^{n_{(1)}} \log \hat{f}_j\left(x_{i,j}\,; \theta_j\right). \tag{5.15}$$

Intuicyjnie ten warunek wymusza na sieci neuronowej, aby przypisywała wysokie wartości gęstości prawdopodobieństwa dla obserwacji ze zbioru $D_{(1)}$. Innymi słowy, sieć neuronowa powinna jak najlepiej dopasować rozkład brzegowy $\hat{f}_j(x; \theta_j)$ do danych rzeczywistych.

Drugi zbiór treningowy jest oznaczany jako

$$D_{(2)} = \{x_{i,j}\}_{i=1}^{n_{(2)}}, \tag{5.16}$$

gdzie $n_{(2)}$ oznacza liczbę elementów w tym zbiorze. Elementy $x_{i,j}$ w tym przypadku nie muszą odpowiadać obserwacjom rzeczywistym, lecz są punktami wybranymi z dziedziny Ω_j zmiennej losowej X_j. Zbiór $D_{(2)}$ służy do sprawdzenia, czy estymowana gęstość prawdopodobieństwa $\hat{f}_j(x_j;\theta_j)$ przyjmuje wartości nieujemne, co jest warunkiem koniecznym dla każdej gęstości prawdopodobieństwa. Aby ten warunek został spełniony, wprowadzany jest drugi składnik funkcji straty:

$$L_{(2)}(\theta_j) = \frac{1}{n_{(2)}} \sum_{i=1}^{n_{(2)}} \left| \mathrm{ReLu}\left(-\hat{f}_j(x_{i,j};\theta_j)\right) \right|, \tag{5.17}$$

Dzięki takiej postaci ujemne wartości gęstości prawdopodobieństwa powodują dodanie dodatniej kary do funkcji straty. Jeżeli $\hat{f}_j(x_{i,j};\theta_j) \geq 0$ dla wszystkich punktów $x_{i,j} \in D_{(2)}$, to składnik ten przyjmuje wartość zerową. Oczekiwanym efektem procesu uczenia jest zatem:

$$L_{(2)}(\theta_j) \to 0, \tag{5.17}$$

co oznacza, że model generuje nieujemne wartości gęstości prawdopodobieństwa w całej rozpatrywanej dziedzinie Ω_j.

Trzeci zbiór treningowy oznaczany jest jako

$$D_{(3)} = \{x_{i,j}\}_{i=1}^{n_{(3)}}, \tag{5.18}$$

gdzie $n_{(3)}$ oznacza jego liczebność. Elementy $x_{i,j}$ są wybierane z dziedziny Ω_j w taki sposób, jak w $D_{(2)}$ Ponieważ każda gęstość prawdopodobieństwa musi spełniać warunek normalizacji:

$$\int_{\Omega} \hat{f}_j(x_j;\theta_j)\, dx_j = 1, \tag{5.19}$$

wprowadzany jest składnik kary, kontrolujący jego spełnienie. Wykorzystując dyskretną aproksymację całki, składnik ten przyjmuje postać:

$$L_{(3)}(\theta_j) = \left| 1 - \frac{1}{n_{(3)}} \sum_{i=1}^{n_{(3)}} \hat{f}_j(x_{i,j};\theta_j)\, \Delta x_j \right|, \tag{5.20}$$

gdzie Δx_j oznacza krok dyskretyzacji. Jeśli gęstość $\hat{f}_j(x_j;\theta_j)$ została poprawnie wyestymowana, wartość całki powinna być bliska jedności, a składnik $L_{(3)}(\theta_j)$ powinien przyjmować wartości bliskie zeru. Dlatego oczekiwanym efektem procesu uczenia jest:

$$L_{(3)}(\theta_j) \to 0, \tag{5.21}$$

co oznacza, że estymowana gęstość spełnia warunek normalizacji.

Czwarty zbiór treningowy oznaczany jest jako

$$D_{(4)} = \{x_{\min}, x_{\max}\}, \tag{5.22}$$

gdzie $x_{\min}$ oraz $x_{\max}$ wyznaczają odpowiednio dolną i górną granicę rozpatrywanego przedziału Ω_j dla zmiennej losowej X_j. W praktyce wartości te obliczane są jako minimum i maksimum dostępnych obserwacji danej zmiennej, tj.:

$$x_{\min} = \min(x_{1,j}, x_{2,j}, \ldots, x_{n,j}), \qquad x_{\max} = \max(x_{1,j}, x_{2,j}, \ldots, x_{n,j}), \tag{5.23}$$

Punkty te służą do kontroli poprawności wartości dystrybuanty na brzegach dziedziny. Z własności dystrybuanty wynika, że:

$$\hat{F}_j(x_{\min};\theta_j) = 0, \qquad \hat{F}_j(x_{\max};\theta_j) = 1. \tag{5.24}$$

Aby wymusić spełnienie tych warunków, wprowadzany jest czwarty składnik kary:

$$L_{(4)}(\theta_j) = \left|\hat{F}_j(x_{\min};\theta_j) - 0\right| + \left|1 - \hat{F}_j(x_{\max};\theta_j)\right|. \tag{5.25}$$

Jeśli sieć wyznacza poprawnie dystrybuantę, składnik ten powinien przyjmować wartości bliskie zeru. Oczekujemy zatem, że w procesie uczenia:

$$L_{(4)}(\theta_j) \to 0, \tag{5.26}$$

co oznacza, że dystrybuanta $\hat{F}_j$ przyjmuje wartości zgodne z warunkami brzegowymi.

Po zdefiniowaniu wszystkich składników funkcji celu, proces uczenia sieci polega na iteracyjnej minimalizacji łącznej straty:

$$L(\theta_j) = \sum_{k=1}^{4} \lambda_k L_{(k)}(\theta_j), \tag{5.27}$$

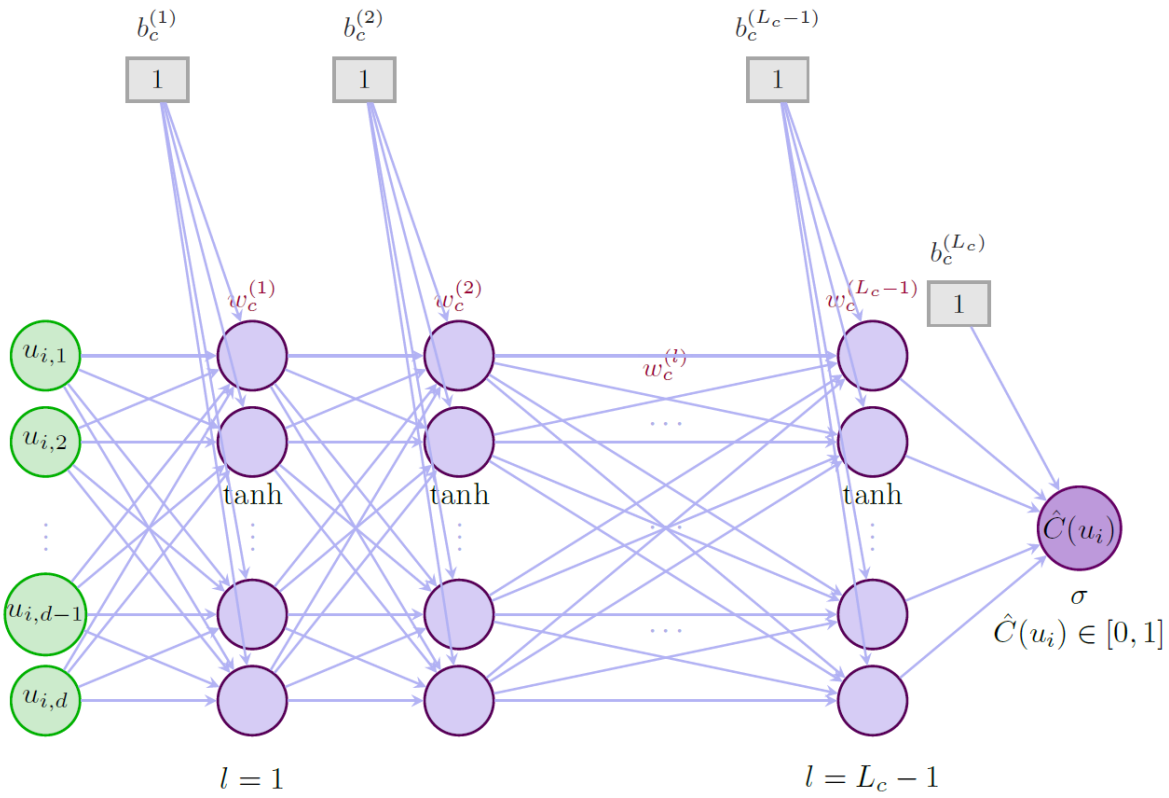
gdzie parametry λ_k odpowiadają udziałowi w całkowitej funkcji straty $k-tej$ funkcji straty. W każdym kroku uczenia dane z odpowiednich zbiorów $D_{(1)}$, $D_{(2)}$, $D_{(3)}$ i $D_{(4)}$ przechodzą przez sieć neuronową, która wyznacza wartości dystrybuanty $\hat{F}_j(x_j;\theta_j)$. Następnie, dzięki zastosowaniu mechanizmu automatycznego różniczkowania, obliczana jest również odpowiadająca jej gęstość $\hat{f}_j(x_j;\theta_j)$. Uzyskane wartości pozwalają na wyznaczenie wszystkich częściowych składników kosztu $L_{(k)}(\theta_j)$. Wyznaczane są gradienty względem $\theta_j = \{w_j^{(l)}, b_j^{(l)}\}$ i za pomocą propagacji wstecznej aktualizowane są wagi. W pracy zastosowano optymalizator typu Nadam, który łączy metodę Adama z momentem Nesterova, co zapewnia stabilność oraz przyspiesza proces uczenia. Aby zapobiec przeuczeniu, do funkcji celu wprowadzono element regularyzacji l_2, który dodatkowo karze funkcję celu za zbyt duże wartości wag. Dzięki temu, kontrolowana jest złożoność modelu i ograniczona tendencja do nadmiernego dopasowania sieci do danych treningowych. Proces uczenia jest powtarzany iteracyjnie w kolejnych epokach, aż do momentu, gdy wartości składników $L_{(k)}(\theta_j)$ osiągają zamierzone wartości.

5.1.2.2 Dopasowanie kopuli

Wyznaczenie kopuli stanowi drugi etap procesu estymacji rozkładu wielowymiarowego z wykorzystaniem sieci neuronowych. Na tym etapie modelowana jest struktura zależności pomiędzy zmiennymi losowymi o uprzednio oszacowanych rozkładach brzegowych. Zadaniem sieci neuronowej jest zatem aproksymacja kopuli. W dalszej części rozdziału przedstawiono szczegółową architekturę sieci neuronowej przeznaczonej do modelowania kopuli oraz opisano proces jej uczenia, wraz z warunkami zapewniającymi zgodność teoretyczną i stabilność numeryczną uzyskanego modelu.

Architektura sieci

Architekturę sieci neuronowej dla kopuli przedstawiono na poniższym Rysunku 5.3.



Rysunek 5.3: Architektura sieci neuronowej dla kopuli.
Źródło: Opracowanie własne.

Podobnie jak w przypadku sieci neuronowej zbudowanej dla rozkładów brzegowych, konstruowana jest sieć neuronowa dla kopuli $\hat{C}((u_1, u_2, \ldots, u_d); \theta_c)$, której celem jest uchwycenie nieliniowych zależności między zmiennymi, przy jednoczesnym zachowaniu własności teoretycznych kopuli. Dane wejściowe do sieci stanowi wektor:

$$u_i = (u_{i,1}, u_{i,2}, \ldots, u_{i,d}) \in [0,1]^d, \tag{5.28}$$

gdzie każdy element $u_{i,j}$ jest wartością dystrybuanty rozkładu brzegowego dla j-tej zmiennej i i-tej obserwacji rzeczywistej:

$$u_{i,j} = \hat{F}_j(x_{i,j}; \theta_j), \quad j = 1, 2, \ldots, d, \quad i = 1, 2, \ldots, n. \tag{5.29}$$

Każdy $u_{i,j}$ należy do przedziału $[0, 1]$. Architektura sieci obejmuje warstwę wejściową, warstwy ukryte oraz warstwę wyjściową. Argumentem warstwy wejściowej dla danej obserwacji i jest zatem wektor:

$$h_c^{(0)} = u_i = (u_{i,1}, u_{i,2}, \ldots, u_{i,d}) \in [0,1]^d. \tag{5.30}$$

Rozkłady brzegowe zostały wyznaczone za pomocą sieci opisanych w podrozdziale 5.1.1. Każda kolejna warstwa ukryta $l \in \{0, \ldots, L_c - 1\}$ wykonuje liniową transformację wyjścia poprzedniej warstwy z użyciem wag $w_c^{(l)}$ i wektora przesunięcia $b_c^{(l)}$ (bias), a następnie wprowadza nieliniowość przez funkcję aktywacji tanh:

$$h_c^{(l+1)} = \tanh\left(w_c^{(l)} h_c^{(l)} + b_c^{(l)}\right). \tag{5.31}$$

W celu zapewnienia spełnienia warunku definiującego, że wartości kopuli mieszczą się w przedziale $[0, 1]$, w warstwie wyjściowej zastosowano sigmoidalną funkcję aktywacji:

$$\hat{C}((u_1, u_2, \ldots, u_d); \theta_c) = \text{sigmoid}\left(w_c^{(L_c)} h_c^{(L_c)} + b_c^{(L_c)}\right) \in [0, 1]. \tag{5.32}$$

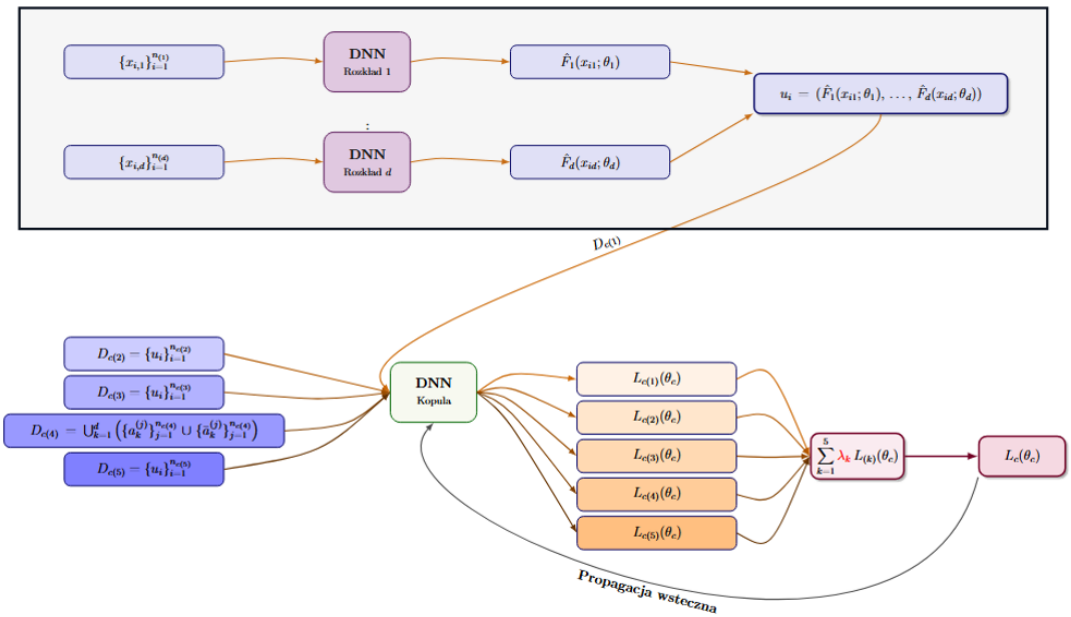
Mając estymator dystrybuanty kopuli $\hat{C}((u_1, u_2, \ldots, u_d); \theta_c)$, gdzie $\theta_c = \{w_c^{(l)}, b_c^{(l)}\}$, wyznaczana jest gęstość kopuli $\hat{c}((u_1, u_2, \ldots, u_d); \theta_c)$ jako pochodna cząstkowa rzędu d względem wszystkich składowych u_i:

$$\hat{c}((u_1, u_2, \ldots, u_d); \theta_c) = \frac{\partial^d \hat{C}((u_1, u_2, \ldots, u_d); \theta_c)}{\partial u_1 \, \partial u_2 \cdots \partial u_d}. \tag{5.33}$$

Ponieważ użyte funkcje (tanh w warstwach ukrytych i sigmoid na wyjściu) są różniczkowalne, pochodna $\hat{C}((u_1, u_2, \ldots, u_d), \theta_c)$ może zostać obliczona analogicznie do modelu rozkładów brzegowych. Dzięki temu otrzymywany jest jednocześnie estymator dystrybuanty oraz odpowiadającej jej gęstości, co umożliwia odwzorowanie struktury zależności między zmiennymi brzegowymi.

Proces uczenia sieci

Proces uczenia sieci dla modelu kopuli zilustrowano na poniższym schemacie (rys. 5.4). Dane wejściowe, po przekształceniu przez modele brzegowe, trafiają do modelu kopuli, który estymuje wielowymiarową dystrybuantę. Sieci neuronowe dla rozkładów zostały wcześniej wytrenowane, a ich wagi zamrożone. Na tej podstawie obliczane są poszczególne składowe funkcji kosztu, odpowiadające zarówno poprawności estymacji, jak i spełnieniu własności teoretycznych kopuli. Wartości tych składników są następnie sumowane w jedną łączną funkcję strat.



Rysunek 5.4: Procedura nauki sieci neuronowej dla kopuli.
Źródło: Opracowanie własne.

Zbiór treningowy $D_{c(1)}$ zawiera pseudo-obserwacje dla danych rzeczywistych wyznaczone z wykorzystaniem uprzednio skonstruowanych sieci neuronowych dla rozkładów brzegowych:

$$u_i = \big(\hat{F}_1(x_{i1};\theta_1), \hat{F}_2(x_{i2};\theta_2), \ldots, \hat{F}_d(x_{id};\theta_d)\big), \tag{5.34}$$

gdzie $u_i \in [0,1]^d$ oznacza wektor pseudo-obserwacji odpowiadający obserwacji x_i. Wartości parametrów $\theta = (\theta_1, \ldots, \theta_d)$ zostały ustalone podczas wcześniejszego etapu uczenia sieci dla rozkładów brzegowych. Na tej podstawie definiowany jest pierwszy składnik funkcji straty w postaci log-wiarygodności:

$$L_{c(1)}(\theta_c) \;=\; \frac{1}{n_{c(1)}} \sum_{i=1}^{n_{c(1)}} \log \hat{f}_c(u_i;\, \theta_c)\,. \tag{5.35}$$

Zbiór $D_{c(2)}$ stanowi zestaw punktów pomocniczych wykorzystywanych w procesie uczenia modelu kopuli. Punkty te są równomiernie rozmieszczone w jednostkowej kostce $[0,1]^d$ i mogą być interpretowane jako siatka próbek z przestrzeni kopuli. Każdy punkt ma postać

$$u_i = (u_{i1}, u_{i2}, \ldots, u_{id}) \in [0,1]^d, \tag{5.36}$$

a cały zbiór zapisujemy jako:

$$D_{c(2)} = \{u_i\}_{i=1}^{n_{c(2)}}. \tag{5.37}$$

Zbiór ten służy do zapewnienia zgodności z założeniem teoretycznym, zgodnie z którym estymowana gęstość $\hat{f}_c$ nie może przyjmować wartości ujemnych. Na jego podstawie definiowana jest funkcja straty w postaci:

$$L_{c(2)}(\theta_c) \;\approx\; \frac{1}{n_{c(2)}} \sum_{i=1}^{n_{c(2)}} \mathrm{ReLu}\Big(-\hat{f}_c(u_i;\, \theta_c)\Big), \tag{5.38}$$

której minimalizacja wymusza nieujemność estymowanej gęstości. Oczekiwanym efektem procesu uczenia jest zatem:

$$L_{c(2)}(\theta_c) \to 0. \tag{5.39}$$

Zbiór $D_{c(3)}$ odpowiada za normalizację gęstości kopuli. W praktyce wykorzystuje się ten sam równomierny zbiór punktów w przestrzeni $[0,1]^d$, co w $D_{c(2)}$. Na podstawie wartości $\hat{f}_c$ w tych punktach definiuje się składnik funkcji kosztu:

$$L_{c(3)}(\theta_c) \;\approx\; \left| 1 - \frac{1}{n_{c(3)}} \sum_{i=1}^{n_{c(3)}} \hat{f}_c(u_i;\, \theta_c) \prod_{j=1}^{d} \Delta_j \right|, \qquad u_i \in D_{c(3)} \subset [0,1]^d. \tag{5.40}$$

gdzie Δ_j jest krokiem siatki w j-tym wymiarze. Oczekiwanym efektem procesu uczenia jest zatem

$$L_{c(3)}(\theta_c) \to 0. \tag{5.41}$$

Zbiór $D_{c(4)}$ służy do sprawdzenia warunków brzegowych, które musi spełniać kopula. Zgodnie z definicją kopuli, dla każdego wymiaru $k = 1, \ldots, d$ zachodzą zależności:

$$\hat{C}(t_1, \ldots, t_{k-1}, 0, t_{k+1}, \ldots, t_d;\, \theta_c) = 0, \qquad t_i \in [0,1], \tag{5.42}$$

oraz

$$\hat{C}(1,\ldots,1,\,t_k,\,1,\ldots,1;\,\theta_c) = t_k, \qquad t_k \in [0,1]. \tag{5.43}$$

W celu sprawdzenia spełnienia powyższych warunków definiuje się punkty testowe na brzegach jednostkowej kostki $[0,1]^d$:

$$a_k^{(j)} = (t_1^{(j)},\ldots,t_{k-1}^{(j)},0,t_{k+1}^{(j)},\ldots,t_d^{(j)}), \qquad \bar{a}_k^{(j)} = (1,\ldots,1,\,t_k^{(j)},\,1,\ldots,1), \tag{5.44}$$

gdzie $t^{(j)} \in [0,1]^d$, $j = 1,\ldots,n_{c(4)}$. Na tej podstawie składnik funkcji kosztu przyjmuje postać:

$$L_{c(4)}(\theta_c) \;=\; \sum_{k=1}^{d}\sum_{j=1}^{n_{c(4)}} \hat{C}(a_k^{(j)};\,\theta_c) \;+\; \sum_{k=1}^{d}\sum_{j=1}^{n_{c(4)}} \left|\hat{C}(\bar{a}_k^{(j)};\,\theta_c) - t_k^{(j)}\right|. \tag{5.45}$$

Zbiór $D_{c(4)}$ można zatem zapisać jako:

$$D_{c(4)} = \bigcup_{k=1}^{d} \left(\{a_k^{(j)}\}_{j=1}^{n_{c(4)}} \cup \{\bar{a}_k^{(j)}\}_{j=1}^{n_{c(4)}}\right). \tag{5.46}$$

Oczekiwanym efektem procesu uczenia jest zatem

$$L_{c(4)}(\theta_c) \to 0, \tag{5.47}$$

co zapewnia spełnienie warunków brzegowych przez $\hat{C}$ na krawędziach jednostkowej kostki $[0,1]^d$.

W celu lepszego odwzorowania dystrybuanty wprowadza się dodatkowy punktowy składnik uczenia oparty na porównaniu wartości $\hat{C}$ z jej odpowiednikiem empirycznym. Dla losowo wybranych punktów $u \in [0,1]^d$ wyznacza się empiryczną dystrybuantę:

$$\widehat{F}_{\text{emp}}(u) = \frac{1}{n_{c(1)}} \sum_{v \in D_{c(1)}} \text{flag}(u,v), \tag{5.48}$$

gdzie

$$\text{flag}(u,v) = \begin{cases} 1, & \text{gdy } v_j < u_j \;\; \forall j \le d, \\ 0, & \text{w przeciwnym razie.} \end{cases} \tag{5.49}$$

Wielkość ta stanowi estymator prawdopodobieństwa $P(V \le u)$. Następnie wyznacza się wartość $\hat{F}_c(u;\theta)$ w tych punktach:

$$\hat{F}_c(u;\theta) = \hat{C}\big(\hat{F}_1(u_1;\theta_1),\;\ldots,\;\hat{F}_d(u_d;\theta_d);\theta_c\big). \tag{5.50}$$

Funkcja celu dla opisanego warunku liczona jest na zbiorze $D_{c(5)} = \{u_i\}_{i=1}^{n_{c(5)}}$, gdzie u_i to losowo rozmieszczone punkty w przestrzeni $[0,1]^d$, a funkcja ograniczająca, która jest nakładana na sieć, ma postać $L_{c(5)}$:

$$L_{c(5)}(\theta) = \frac{1}{n_{c(5)}} \sum_{i=1}^{n_{c(5)}} \left|\hat{F}_c(u_i;\theta) - \widehat{F}_{\text{emp}}(u_i)\right|. \tag{5.51}$$

Oczekuje się, że

$$L_{c(5)}(\theta) \to 0. \tag{5.52}$$

Całkowita funkcja kosztu wykorzystywana podczas uczenia modelu kopuli jest definiowana jako ważona suma pięciu składników opisanych we wcześniejszych sekcjach:

$$L_c(\theta) = \lambda_1 L_{c(1)}(\theta_c) + \lambda_2 L_{c(2)}(\theta_c) + \lambda_3 L_{c(3)}(\theta_c) + \lambda_4 L_{c(4)}(\theta_c) + \lambda_5 L_{c(5)}(\theta_c), \tag{5.53}$$

gdzie $\lambda_1, \ldots, \lambda_5 > 0$ stanowią współczynniki wagowe regulujące wpływ poszczególnych składników na końcową funkcję celu. W każdym kroku uczenia korzysta się z odpowiednich zbiorów danych $D_{c(1)}$–$D_{c(5)}$. Gradienty funkcji celu obliczane są względem parametrów $\theta_c = \{w_c^{(l)}, b_c^{(l)}\}_{l=1}^{L_c}$, natomiast parametry sieci opisujących rozkłady brzegowe $\theta = (\theta_1, \ldots, \theta_d)$ są zamrożone. Następnie wagi sieci kopuli $\theta_c = \{w_c^{(l)}, b_c^{(l)}\}_{l=1}^{L_c}$ są aktualizowane metodą propagacji wstecznej. W implementacji wykorzystano optymalizator Nadam oraz regularyzację l_2 na wagach, co stabilizuje i przyspiesza konwergencję oraz przeciwdziała przeuczeniu. Proces uczenia wykonywany jest epoka po epoce, aż do osiągnięcia pożądanych wartości $L_{c(k)}(\theta)$, co zapewnia spełnienie własności teoretycznych kopuli oraz dobre dopasowanie modelu do danych.

5.1.2.3 Generowanie realizacji z wielowymiarowego rozkładu

Po wyznaczeniu i ustaleniu parametrów sieci estymujących rozkłady brzegowe $\hat{F}_j(x_j; \theta_j)$ oraz $\hat{f}_j(x_j; \theta_j)$ dla $j = 1, \ldots, d$, a także sieci neuronowej estymującej kopulę $\hat{c}(u; \theta_c)$, gdzie $u = (\hat{F}_1(x_1; \theta_1), \ldots, \hat{F}_d(x_d; \theta_d))$ przystępuje się do generowania realizacji $\mathbf{x}^{(\text{gen})} = (x_1^{(\text{gen})}, \ldots, x_d^{(\text{gen})})$ z wielowymiarowego rozkładu opartego na wyznaczonych modelach.

Dla każdej zmiennej losowej X_j określa się zbiór m punktów w przedziale $[0, 1]$, ponieważ dane rzeczywiste zostały znormalizowane do tej przestrzeni:

$$\mathcal{X}_j = \left\{0, \frac{1}{m-1}, \frac{2}{m-1}, \ldots, 1\right\}, \qquad j = 1, \ldots, d. \tag{5.54}$$

Iloczyn kartezjański tych zbiorów tworzy siatkę punktów na przestrzeni dziedziny wielowymiarowej:

$$\mathcal{G} = \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_d \subset [0, 1]^d. \tag{5.55}$$

Dla każdego punktu $\mathbf{x}^{(\text{g})} = (x_1^{(\text{g})}, x_2^{(\text{g})}, \ldots, x_d^{(\text{g})}) \in \mathcal{G}$ wyznaczane są wartości dystrybuant brzegowych $\hat{F}_j(x_j^{(\text{g})}; \theta_j)$ oraz odpowiadające im gęstości $\hat{f}_j(x_j^{(\text{g})}; \theta_j)$, dla $j = 1, \ldots, d$. Następnie obliczana jest gęstość kopuli $\hat{c}(\mathbf{u}; \theta_c)$, gdzie

$$u = \left(\hat{F}_1(x_1^{(\text{g})}; \theta_1), \hat{F}_2(x_2^{(\text{g})}; \theta_2), \ldots, \hat{F}_d(x_d^{(\text{g})}; \theta_d)\right). \tag{5.56}$$

Zgodnie z twierdzeniem Sklara, wielowymiarową gęstość można zapisać jako kombinację gęstości brzegowych oraz gęstości kopuli:

$$\hat{f}(\mathbf{x}^{(\text{g})}; \theta, \theta_c) = \hat{c}\left(\hat{F}_1(x_1^{(\text{g})}; \theta_1), \ldots, \hat{F}_d(x_d^{(\text{g})}; \theta_d); \theta_c\right) \cdot \prod_{j=1}^{d} \hat{f}_j(x_j^{(\text{g})}; \theta_j). \tag{5.57}$$

Następnie normalizowane są wartości $\hat{f}(\mathbf{x}^{(\mathrm{g})};\theta,\theta_c)$ na siatce $\mathcal{G}$ w celu uzyskania dyskretnego rozkładu prawdopodobieństwa:

$$p(\mathbf{x}^{(\mathrm{g})}) = \frac{\hat{f}(\mathbf{x}^{(\mathrm{g})};\theta,\theta_c)}{\sum_{\mathbf{z}^{(\mathrm{g})}\in\mathcal{G}} \hat{f}(\mathbf{z}^{(\mathrm{g})};\theta,\theta_c)}, \qquad \sum_{\mathbf{x}^{(\mathrm{g})}\in\mathcal{G}} p(\mathbf{x}^{(\mathrm{g})}) = 1. \tag{5.58}$$

W celu wygenerowania N realizacji z rozkładu p, zdefiniowanego na siatce $\mathcal{G}$, proces generowania polega na N-krotnym losowaniu indeksu

$$j \in \{1, 2, \ldots, |\mathcal{G}|\}, \tag{5.59}$$

zgodnie z dyskretnym rozkładem prawdopodobieństwa określonym następująco:

$$\{\, p(\mathbf{x}^{(1)}),\ p(\mathbf{x}^{(2)}),\ \ldots,\ p(\mathbf{x}^{(|\mathcal{G}|)}) \,\}. \tag{5.60}$$

Następnie, dla każdego wylosowanego indeksu j, wybierany jest odpowiadający mu punkt siatki:

$$\mathbf{x}_i^{(\mathrm{gen})} = \mathbf{x}^{(j)} \in \mathcal{G}, \qquad i = 1, \ldots, N. \tag{5.61}$$

Punkty $\mathbf{x}_i^{(\mathrm{gen})}$ są generowane zgodnie z wyznaczonym wielowymiarowym rozkładem $\hat{f}$.

5.1.3 Ocena dopasowania modelu

W procesie modelowania rozkładów wielowymiarowych istotnym etapem jest ocena jakości uzyskanych wyników oraz weryfikacja, w jakim stopniu skonstruowany rozkład odzwierciedla rzeczywistą strukturę danych. Celem takiej analizy jest nie tylko sprawdzenie, czy model dobrze dopasował się do obserwacji rzeczywistych, lecz także ocena jego zdolności generowania realizacji. W ramach przeprowadzonej analizy zastosowano dwa komplementarne podejścia oceny dopasowania modelu: pierwsze oparte na logarytmie wiarygodności sprawdzające w jakim stopniu model odwzorowuje rzeczywistą strukturę danych testowcyh, oraz drugie bazujące na odległości energetycznej, umożliwiające porównanie realizacji wygenerowanych z wielowymiarowego rozkładu z danymi rzeczywistymi.

Pierwsze z zastosowanych podejść polega na ocenie jakości modelu wielowymiarowego na podstawie danych testowych, niezależnych od procesu uczenia. Metoda ta koncentruje się na wyznaczeniu średniej wartości logarytmu funkcji największej wiarygodności, co pozwala określić, w jakim stopniu model odwzorowuje strukturę prawdopodobieństwa obserwowaną w danych rzeczywistych. Zgodnie z twierdzeniem Sklara, wartość tej miary wyznaczana jest według wzoru:

$$\ell = \frac{1}{N}\sum_{i=1}^{N}\left[\log c(U_i) + \sum_{j=1}^{d}\log f_j(x_{ij})\right], \tag{5.62}$$

gdzie $c(U_i)$ oznacza gęstość kopuli w punkcie $U_i = (F_1(x_{i1}), \ldots, F_d(x_{id}))$, natomiast $f_j(x_{ij})$ to gęstości rozkładów brzegowych poszczególnych zmiennych. Ponieważ obliczenia oparte są na gęstościach rozkładów, miara ta umożliwia ilościową ocenę zgodności modelu z rzeczywistym rozkładem danych. Wyższa wartość średniego logarytmu funkcji wiarygodności oznacza, że model przypisuje większe prawdopodobieństwo obserwacjom testowym, a tym samym lepiej odzwierciedla zależności pomiędzy zmiennymi i charakteryzuje się wyższą zdolnością generalizacji.

Drugie podejście oceny wyznaczonych rozkładów wielowymiarowych opiera się na analizie zgodności realizacji wygenerowanych z modelu z rzeczywistymi obserwacjami danych empirycznych. Ocena ta wykorzystuje narzędzia stosowane w analizie jakości prognoz probabilistycznych. Pierwszym jest tzw. *odległość energetyczna* (ang. *Energy Distance*), wprowadzoną przez (Gneiting & Raftery, 2007). Miara ta jest szeroko stosowana do oceny prognoz probabilistycznych w przestrzeniach wielowymiarowych i znajduje coraz szersze zastosowanie w modelach generatywnych, w tym w modelach opartych na kopulach i sieciach neuronowych (Ziel & Berg, 2019). Odległość energetyczna pozwala ocenić, w jakim stopniu rozkład estymowany przez model odzwierciedla rzeczywistą strukturę prawdopodobieństwa obserwowanych danych. W przeciwieństwie do klasycznych testów dopasowania, takich jak test Kolmogorowa–Smirnowa czy test Craméra–von Misesa, nie wymaga znajomości gęstości rozkładu, a jedynie próbek z rozkładu modelowego i rzeczywistego. Dzięki temu może być stosowana w ocenie rozkładów o dowolnym wymiarze, w tym także wysokowymiarowych modeli bazujących na kopulach lub sieciach neuronowych, estymujących rozkład wielowymiarowy.

Niech $\mathbf{X} \in \mathbb{R}^d$ będzie wektorem losowym o rozkładzie modelowym F_X, a $\mathbf{y} \in \mathbb{R}^d$ obserwacją rzeczywistą. Miara oparta na odległości energetycznej jest definiowana jako:

$$\Upsilon_\beta(F_X, \mathbf{y}) = \underbrace{\mathbb{E}\Big(\|\mathbf{X} - \mathbf{y}\|_2^\beta\Big)}_{\text{trafność (bliskość do y)}} - \frac{1}{2} \underbrace{\mathbb{E}\Big(\|\mathbf{X} - \tilde{\mathbf{X}}\|_2^\beta\Big)}_{\text{rozproszenie prognoz}}, \tag{5.63}$$

gdzie $\tilde{\mathbf{X}}$ jest niezależną kopią zmiennej losowej $\mathbf{X}$, $\|\cdot\|_2$ oznacza normę euklidesową, a $\beta \in (0, 2)$ jest parametrem określającym czułość miary na wartości odstające oraz różnice w wariancji.

Pierwszy składnik miary Υ_β mierzy przeciętną odległość pomiędzy realizacjami wygenerowanymi przez model a rzeczywistymi obserwacjami, odzwierciedlając trafność modelu w odwzorowaniu danych empirycznych. Drugi składnik opisuje rozproszenie próbek pochodzących z rozkładu modelowego, czyli stopień niepewności generowanych realizacji. Zrównoważenie tych dwóch składników pozwala jednocześnie ocenić zarówno precyzję modelu, jak i jego wiarygodność probabilistyczną.

Miara ta jest bezpośrednio powiązana z koncepcją odległości między rozkładami, zdefiniowaną jako:

$$\mathcal{D}(P, Q) = 2\,\mathbb{E}\|\mathbf{X} - \mathbf{Y}\|_2^\beta - \mathbb{E}\|\mathbf{X} - \mathbf{X}'\|_2^\beta - \mathbb{E}\|\mathbf{Y} - \mathbf{Y}'\|_2^\beta, \tag{5.64}$$

gdzie $\mathbf{X}, \mathbf{X}' \sim P$ to niezależne realizacje z rozkładu rzeczywistego, a $\mathbf{Y}, \mathbf{Y}' \sim Q$ to realizacje z rozkładu modelowego. Odległość energetyczna $\mathcal{D}(P, Q)$ przyjmuje wartości nieujemne i równa się zeru wtedy i tylko wtedy, gdy $P = Q$, co oznacza, że stanowi rzeczywistą metrykę między rozkładami.

Z praktycznego punktu widzenia, w kontekście modeli bazujących na kopulach i sieciach neuronowych estymujących rozkład wielowymiarowy, odległość energetyczna pozwala na porównanie jakości dopasowania rozkładów modelowych do rozkładu empirycznego danych testowych bez konieczności estymowania gęstości. Obliczenia opierają się na próbkach Monte Carlo, a estymator postaci próbnej przyjmuje postać:

$$\widehat{\mathcal{D}}(P, Q) = \frac{2}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \|\mathbf{x}_i - \mathbf{y}_j\|_2^\beta - \frac{1}{n^2} \sum_{i,k=1}^{n} \|\mathbf{x}_i - \mathbf{x}_k\|_2^\beta - \frac{1}{m^2} \sum_{j,l=1}^{m} \|\mathbf{y}_j - \mathbf{y}_l\|_2^\beta, \tag{5.65}$$

gdzie $\mathbf{x}_i$ to obserwacje empiryczne, a $\mathbf{y}_j$ – realizacje wygenerowane z modelu. Niższa wartość $\widehat{\mathcal{D}}(P, Q)$ oznacza lepsze dopasowanie modelu do danych rzeczywistych.

Niech $\widehat{\Upsilon}_\beta(F_X, \mathbf{y}_t)$ oznacza wartość miary obliczoną dla t-tej obserwacji ($t = 1, \ldots, N$). W celu zwiększenia odporności na wartości odstające, ostateczną wartość miary wyznaczono jako medianę wszystkich ocen:

$$\widetilde{\Upsilon}_\beta = \text{median}\left\{\widehat{\Upsilon}_\beta(F_X, \mathbf{y}_1), \ldots, \widehat{\Upsilon}_\beta(F_X, \mathbf{y}_N)\right\}. \tag{5.66}$$

Wartość miary $\widetilde{\Upsilon}_\beta$ umożliwia bezpośrednie porównanie jakości dopasowania różnych modeli rozkładu wielowymiarowego - zarówno parametrycznych (np. opartych na kopulach C-vine lub D-vine), jak i nieparametrycznych (np. z wykorzystaniem sieci neuronowych). Niewątpliwą zaletą tej metody jest jej niezależność od przyjętej postaci analitycznej gęstości oraz zdolność do uogólnienia na przestrzenie o wysokim wymiarze. W badaniach empirycznych wykazano, że odległość energetyczna jest stabilna numerycznie, dobrze różnicuje jakość modeli probabilistycznych i może być z powodzeniem stosowana jako alternatywa lub uzupełnienie miar opartych na log-wiarygodności (Ziel & Berg, 2019; Székely & Rizzo, 2013; Efron & Hastie, 2020).

Drugą miarą stosowaną do oceny jakości prognoz wielowymiarowych jest *Variogram Score* (VS), zaproponowana przez (Scheuerer & Hamill, 2015). Miara ta stanowi rozwinięcie koncepcji wykorzystywanej w analizie przestrzennej i opiera się na porównaniu struktur wariogramów, czyli oczekiwanych różnic pomiędzy parami komponentów wektora losowego. W przeciwieństwie do odległości energetycznej, która koncentruje się na globalnej zgodności rozkładów, *Variogram Score* skupia się na ocenie struktury współzależności między poszczególnymi elementami wektora prognozy. Dzięki temu, lepiej odzwierciedla zgodność modelu z rzeczywistą strukturą korelacji i wariancji, co czyni go szczególnie użytecznym w analizach danych o wyraźnej strukturze przestrzennej lub czasowo-przestrzennej.

Niech $X \in \mathbb{R}^d$ będzie d-wymiarową zmienną losową o rozkładzie F_X, a $y \in \mathbb{R}^d$ obserwacją rzeczywistą. Miarę VS jakości prognozy rzędu $p > 0$ definiuje się jako:

$$\text{VS}_p(F_X, y) = \sum_{i=1}^{d}\sum_{j=1}^{d} w_{ij}\left(|y_i - y_j|^p - \mathbb{E}|X_i - X_j|^p\right)^2, \tag{5.67}$$

gdzie $w_{ij} \geq 0$ oznaczają wagi poszczególnych par komponentów (zazwyczaj przyjmuje się $w_{ij} = 1$).

Pierwszy składnik wewnątrz nawiasu odzwierciedla rzeczywiste odległości pomiędzy komponentami obserwacji, natomiast drugi odnosi się do średnich odległości pomiędzy tymi samymi komponentami w próbkach pochodzących z rozkładu modelowego. Miara ta ocenia więc, na ile model poprawnie odwzorowuje współzależności pomiędzy zmiennymi - im mniejsza wartość VS, tym bliższa jest struktura zależności prognozy do tej obserwowanej w danych rzeczywistych.

Podobnie jak w przypadku odległości energetycznej, ocena globalna dla całego zbioru danych uzyskiwana jest przez agregację wartości cząstkowych dla każdej obserwacji. W praktyce, w celu zwiększenia odporności na wartości odstające, ostateczną wartość miary wyznacza się jako medianę wszystkich wyników:

$$\widetilde{\text{VS}}_p = \text{median}\left\{\widehat{\text{VS}}_p(F_X, y_1), \ldots, \widehat{\text{VS}}_p(F_X, y_N)\right\}. \tag{5.68}$$

Zastosowanie *Variogram Score* pozwala zatem na uzupełnienie informacji uzyskiwanych z miar globalnych, takich jak odległość energetyczna, poprzez analizę lokalnych relacji między

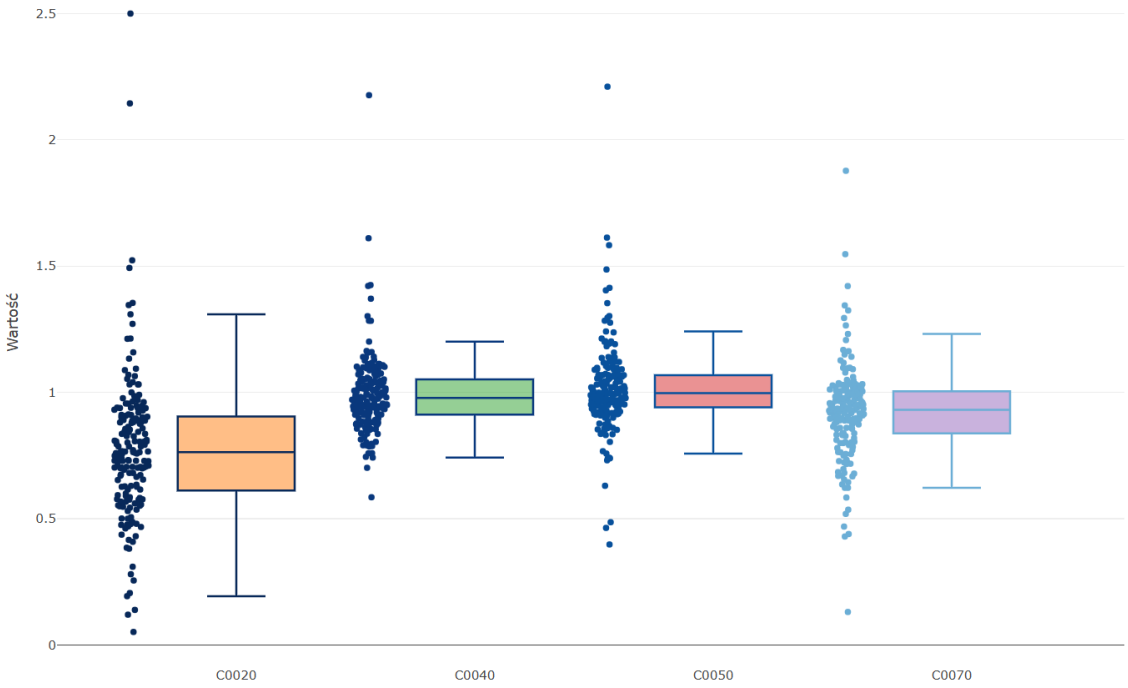
komponentami rozkładu. W kontekście oceny dopasowania rozkładów wielowymiarowych, a zwłaszcza modeli opartych na kopulach lub sieciach neuronowych, VS jest szczególnie wartościowa, gdy istotne są wzorce współzależności pomiędzy zmiennymi, np. w modelowaniu ryzyka, portfeli ubezpieczeniowych lub struktur korelacji przestrzennej. W badaniach empirycznych potwierdzono, że VS stanowi stabilną i komplementarną miarę jakości prognoz wielowymiarowych (Scheuerer & Hamill, 2015; Ziel & Berg, 2019), zwłaszcza w zastosowaniach, w których kluczową rolę odgrywa precyzyjne uchwycenie struktury zależności.

5.2 Charakterystyka danych

Do analizy wybrano dane dotyczące niemieckich ubezpieczycieli majątkowych (non-life), przede wszystkim z uwagi na dużą liczbę aktywnie działających podmiotów w tym sektorze oraz szeroką dostępność raportów o wypłacalności i kondycji finansowej SFCR. Z uwagi na fakt, że ubezpieczyciele w Niemczech nie prowadzą działalności we wszystkich 12 segmentach wyszczególnionych w dyrektywie Solvency II, w badaniach analizowane są wskaźniki zespolone X_i dla czterech wybranych segmentów działalności ubezpieczeniowej, wyznaczonych na podstawie danych pozyskanych z raportów SFCR z lat 2017-2022.
Wybrane segmenty to:

- C0020 – ubezpieczenia na wypadek utraty dochodów;
- C0040 – ubezpieczenia odpowiedzialności cywilnej z tytułu użytkowania pojazdów mechanicznych;
- C0050 – pozostałe ubezpieczenia pojazdów;
- C0070 – ubezpieczenia od ognia i innych szkód rzeczowych.

Na poniższym Rysunku 5.5 za pomocą wykresów pudełkowych przedstawiamy rozkład wskaźników zespolonych dla analizowanych segmentów ubezpieczycieli, natomiast w poniższej Tabeli 5.1 prezentujemy ich podstawowe statystyki opisowe.

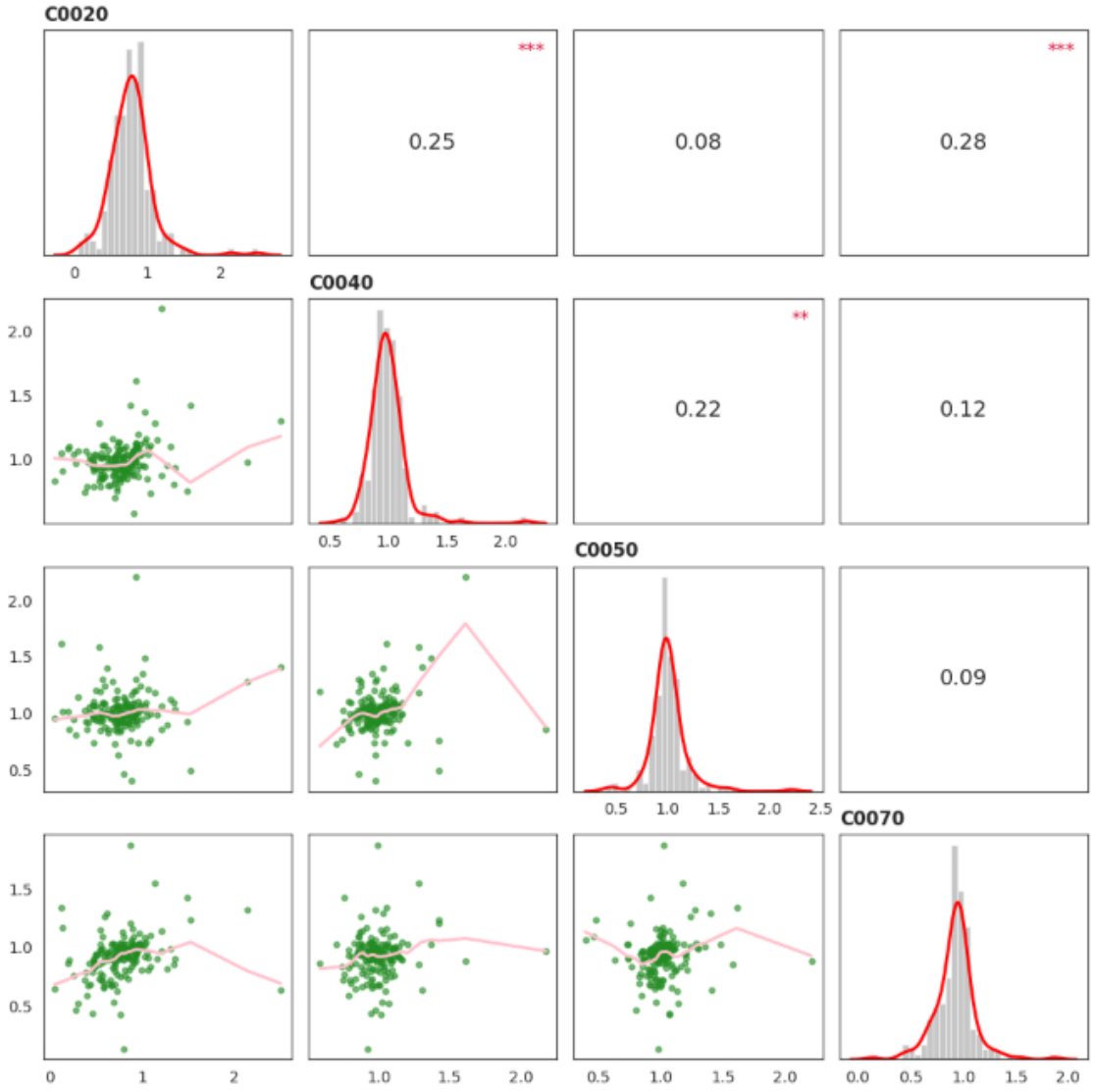


Rysunek 5.5: Rozkład wskaźników zespolonych dla segmentów ubezpieczycieli.
Źródło: Opracowanie własne.

Statystyka	**C0020**	**C0040**	**C0050**	**C0070**
Min	0,0516	0,5850	0,3979	0,1308
Max	2,5000	2,1769	2,2105	1,8775
Q1 25%	0,6115	0,9122	0,9385	0,8954
Mediana	0,7660	0,9768	0,9963	0,9330
Q3 75%	0,9046	1,0493	1,0678	1,0957
Średnia	0,7776	0,9919	1,0121	0,9232
Wariancja	0,0862	0,0253	0,0334	0,0355
Odchylenie std.	0,2937	0,1590	0,1828	0,1885
Rozstęp	2,4484	1,5919	1,8126	1,7467
CV	0,3776	0,1603	0,1806	0,2041
MAD	0,1984	0,1011	0,1122	0,1240
Skośność	1,6438	3,8579	1,6815	0,3167
Kurtoza	8,6248	18,2941	11,9286	5,2685
LCL	0,7342	0,9684	0,9851	0,8954
UCL	0,8210	1,0154	1,0391	0,9510
N	176	176	176	176

Tabela 5.1: Statystyki dla współczynników zespolonych.
Źródło: Opracowanie własne.

Segment C0020 charakteryzuje się największą rozpiętością danych oraz relatywnie wysokim współczynnikiem zmienności. Wysoka skośność świadczy o asymetrii, natomiast podwyższona kurtoza potwierdza obecność obserwacji odstających oraz większą koncentrację wartości wokół mediany. Segment C0040 cechuje się jeszcze asymetrią rozkładu oraz bardzo wysoką kurtozą. Dane w tym segmencie są w dużym stopniu skupione wokół wartości w centrum rozkładu, lecz jednocześnie występują wartości ekstremalne, które istotnie wpływają na kształt rozkładu. W przypadku segmentu C0050 obserwuje się większą równowagę. Średnia i mediana przyjmują zbliżone wartości, a współczynnik zmienności pozostaje na stosunkowo niskim poziomie. Wartość kurtozy wskazuje jednak, że również w tym segmencie pojawiają się obserwacje odstające. Segment C0070 wyróżnia się najmniejszą zmiennością oraz najniższym odchyleniem standardowym spośród analizowanych wskaźników. Rozkład cechuje się wysoką symetrią, a umiarkowana kurtoza sugeruje ograniczoną liczbę wartości odstających.



Rysunek 5.6: Charakterystyka analizowanych danych.
Źródło: Opracowanie własne.

Na Rysunku 5.6 przedstawiono analizę graficzną zmiennych C0020, C0040, C0050 i C0070.

Pozwala ona ocenić charakter zależności oraz intuicyjnie określić rozkład zmiennych. Rozkłady pierwszych trzech zmiennych charakteryzują się wyraźną prawostronną asymetrią oraz wysoką kurtozą, co wskazuje na obecność wartości odstających i znaczne odstępstwa od rozkładu normalnego. Takie właściwości mogą prowadzić do zaniżenia wartości współczynnika korelacji Pearsona. Wykresy rozrzutu sugerują, że niektóre zależności mogą mieć charakter nieliniowy. Dla pary C0020–C0070, mimo umiarkowanie niskiej korelacji liniowej wynoszącej 0.28, układ punktów oraz przebieg krzywej wygładzającej wskazują na bardziej złożoną strukturę zależności. Podobny efekt widoczny jest w przypadku zmiennych C0020–C0040 oraz C0040–C0050, gdzie współczynniki korelacji (0.25 i 0.22) są istotne statystycznie, ale rozrzut obserwacji nie wskazuje na liniową zależność. Zmienna C0070 wyróżnia się największą stabilnością i symetrią rozkładu, co ogranicza wpływ wartości odstających; jednak w jej przypadku także nie obserwuje się jednoznacznych zależności liniowych z pozostałymi zmiennymi. Na podstawie przeprowadzonej analizy stwierdzono, że zależności między zmiennymi mogą mieć charakter nieliniowy. Rozkłady charakteryzują się wysoką skośnością oraz kurtozą, co wskazuje na obecność wartości odstających i odchylenia od normalności. W konsekwencji proste miary korelacji liniowej mogą okazać się niewystarczające, natomiast pełniejsze odwzorowanie struktury zależności wymaga zastosowania metod odpornych na nieliniowości oraz asymetrię.

5.3 Wyniki estymacji i oceny dopasowania modeli wielowymiarowych opartych na kaskadach kopul i sieciach neuronowych

W niniejszym podrozdziale przedstawiono wyniki estymacji oraz oceny modeli wielowymiarowych rozkładów prawdopodobieństwa. W pierwszej części opisano sposób wyznaczenia rozkładów brzegowych, które stanowią punkt wyjścia do konstrukcji rozkładu wielowymiarowego. Zastosowano dwa podejścia estymacyjne: nieparametryczne, oparte na uczeniu sieci neuronowych odwzorowujących dystrybuanty i gęstości przy zachowaniu ich własności teoretycznych, oraz parametryczne, w którym dopasowano rozkłady o z góry określonej postaci analitycznej.

Druga część podrozdziału poświęcona została modelowaniu struktury zależności pomiędzy zmiennymi losowymi. W tym celu skonstruowano trzy alternatywne modele kopul: neuronową oraz klasyczne kaskadowe konstrukcje typu C-vine i D-vine.

W ostatniej części podrozdziału przeprowadzono ocenę jakości dopasowania uzyskanych rozkładów wielowymiarowych. Analizie poddano ich zgodność z danymi empirycznymi oraz zdolność do odtwarzania struktury współzależności między zmiennymi. Weryfikacji dokonano na podstawie wartości logarytmu funkcji wiarygodności oraz miar jakości dopasowania, takich jak Energy Score i Variogram Score. Uzyskane wyniki umożliwiają kompleksową ocenę jakości modeli i stanowią podstawę do dalszej analizy efektu dywersyfikacji w kontekście zastosowań ubezpieczeniowych.

Dane wejściowe zostały podzielone na dwa zbiory: treningowy i testowy. Do zbioru treningowego losowo przydzielono 66% danych, natomiast pozostałe 33% utworzyły zbiór testowy.

5.3.1 Estymacja rozkładów brzegowych

W procesie estymacji rozkładów brzegowych wykorzystano dwa podejścia: oparte na sieciach neuronowych oraz na klasycznych rozkładach parametrycznych. W pierwszym etapie, na zbiorze treningowym, przeprowadzono proces dopasowania rozkładów. Sieci neuronowe

uczono w taki sposób, aby estymowane funkcje dystrybuant i gęstości spełniały teoretyczne warunki rozkładu prawdopodobieństwa. Na wykresach poniżej przedstawiono przebieg procesu uczenia. Równolegle, dla tego samego zbioru danych treningowych, przeprowadzono estymację rozkładów w podejściu parametrycznym, a następnie dla każdego segmentu wybrano rozkłady, które w największym stopniu odzwierciedlały dane rzeczywiste. W drugim etapie, na zbiorze testowym, wykonując test Kołmogorowa-Smirnowa, dokonano niezależnej oceny dopasowanych rozkładów. Przeprowadzenie testu na danych, których modele nie widziały podczas uczenia, pozwoliło zweryfikować ich zdolność do uogólniania oraz porównać skuteczność podejścia parametrycznego i opartego na sieciach neuronowych.

Estymacja z wykorzystaniem sieci neuronowych

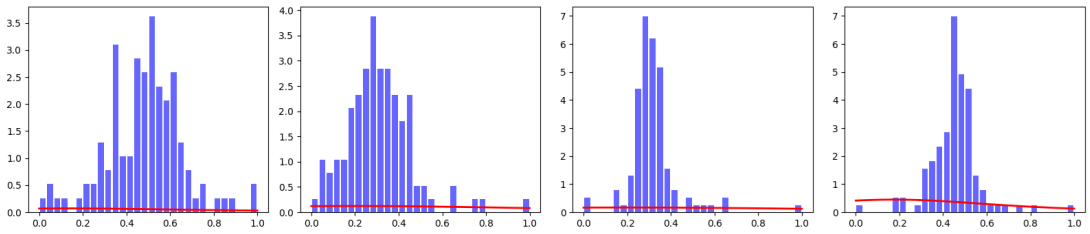
W estymacji rozkładów brzegowych dla poszczególnych segmentów zastosowano sieci neuronowe, których zadaniem było odwzorowanie funkcji dystrybuant $\hat{F}_j(x_j)$ oraz odpowiadających im gęstości $\hat{f}_j(x_j)$, gdzie $j = \{1, 2, 3, 4\}$. Dla każdego segmentu C0020, C0040, C0050, C0070 skonstruowano odrębną sieć neuronową, uczoną niezależnie. Jak opisano w podrozdziale 5.1, aby DNN mogły być interpretowane jako poprawne rozkłady brzegowe, nałożono na nie ograniczenia wynikające z teoretycznych właściwości rozkładów prawdopodobieństwa.

Proces uczenia modeli brzegowych realizowany był w oparciu o minimalizację funkcji kosztu $L(\theta_j)$, stanowiącej sumę ważoną czterech składników $L_{(1)}$–$L_{(4)}$:

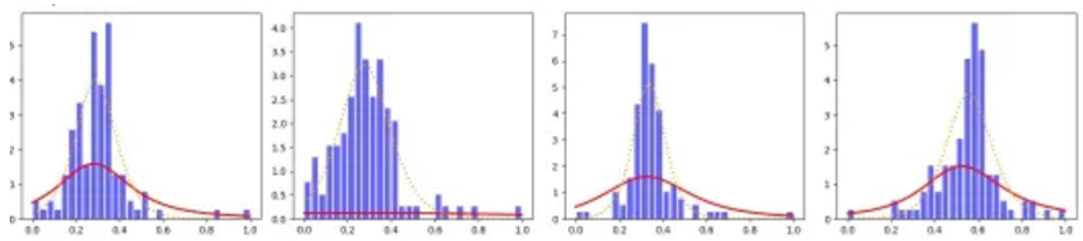
- $L_{(1)}$ odpowiada za minimalizację ujemnego log-wiarygodności zgodnie ze wzorem (5.15),
- $L_{(2)}$ wymusza nieujemność estymowanej gęstości $\hat{f}_j$, zgodnie ze wzorem (5.17),
- $L_{(3)}$ zapewnia normalizację gęstości zgodnie ze wzorem (5.20),
- $L_{(4)}$ kontroluje spełnienie warunków brzegowych dystrybuanty, wymuszając przyjmowanie przez funkcję $\hat{F}_j$ wartości 0 i 1 odpowiednio dla $x_{\min}$ i $x_{\max}$, zgodnie ze wzorem (5.25).

Ostateczna funkcja straty $L(\theta_j)$ wyznaczana jest zgodnie ze wzorem 5.27. Dla każdego z czterech segmentów skonstruowano niezależną sieć neuronową. Architektura każdej sieci obejmowała pięć warstw ukrytych, a w każdej z nich 10 neuronów. We wszystkich warstwach ukrytych zastosowano funkcję aktywacji *tanh (tangens hiperboliczny)*, natomiast w warstwie wyjściowej użyto funkcji *sigmoid*, która ogranicza wartości wyjściowe estymowanej dystrybuanty do przedziału $[0, 1]$, zapewniając zgodność z własnościami dystrybuanty rozkładu prawdopodobieństwa. Proces uczenia przeprowadzono z wykorzystaniem optymalizatora *Nadam* z początkowym współczynnikiem uczenia równym $5 \cdot 10^{-4}$. Dodatkowo zastosowano adaptacyjną regulację tempa uczenia za pomocą mechanizmu *ReduceLROnPlateau* (Chollet et al., 2015), który monitorował wartość funkcji kosztu $L(\theta_j)$ i automatycznie zmniejszał współczynnik uczenia o połowę w przypadku braku poprawy przez 1000 epok. W każdej warstwie sieci zastosowano regularizację typu l_2 z parametrem $\lambda = 0.001$, która ograniczała wartości wag neuronów, redukując ryzyko przeuczenia. Dodatkowym czynnikiem stabilizującym proces uczenia były $L_{(2)}$ - $L_{(4)}$, wymuszające spełnienie własności rozkładu prawdopodobieństwa. Połączenie klasycznej regularizacji wag typu l_2, funkcji kar wynikających z własności prawdopodobieństwa oraz adaptacyjnej kontroli tempa uczenia pozwoliło uzyskać modele charakteryzujące się dobrą zdolnością generalizacji.

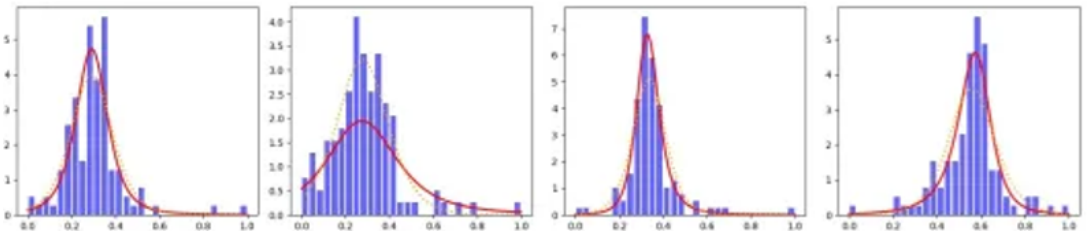
Na Rysunkach 5.7–5.11 przedstawiono proces dopasowywania się sieci neuronowych do poszczególnych segmentów C0020, C0040, C0050 oraz C0070 w kolejnych epokach uczenia.



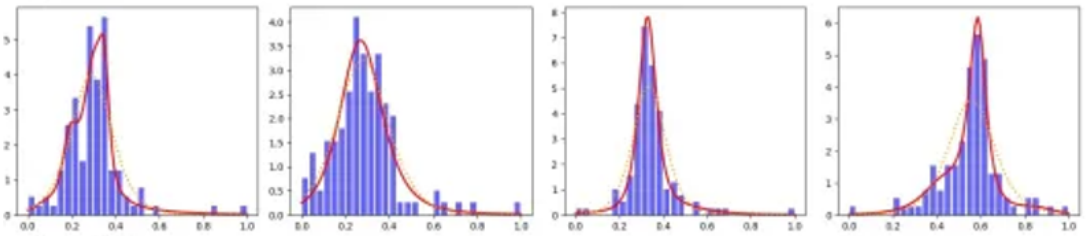
Rysunek 5.7: Dopasowanie sieci neuronowych do segmentów po 10 epokach.
Źródło: Opracowanie własne.



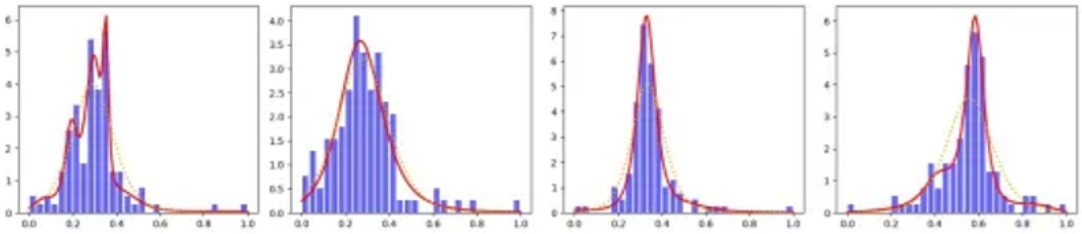
Rysunek 5.8: Dopasowanie sieci neuronowych do segmentów po 3000 epokach.
Źródło: Opracowanie własne.



Rysunek 5.9: Dopasowanie sieci neuronowych do segmentów po 6000 epokach.
Źródło: Opracowanie własne.



Rysunek 5.10: Dopasowanie sieci neuronowych do segmentów po 8000 epokach.
Źródło: Opracowanie własne.



Rysunek 5.11: Dopasowanie sieci neuronowych do segmentów po 10000 epokach.
Źródło: Opracowanie własne.

Estymacja parametryczna

Do wskaźników zespolonych w poszczególnych segmentach dopasowywano parametryczne rozkłady prawdopodobieństwa z wykorzystaniem metody największej wiarygodności. Dopasowano następujące rozkłady: normalny, logarytmiczno-normalny, logistyczny, weibulla oraz gamma.

Rozkład *normalny* opisuje zmienną losową o gęstości:

$$f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \tag{5.69}$$

gdzie μ oznacza wartość oczekiwaną, a σ odchylenie standardowe.

Rozkład *logarytmiczno-normalny* znajduje zastosowanie w przypadku zmiennych dodatnich o silnej asymetrii. Jego gęstość wyraża się wzorem:

$$f(x;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x-\mu)^2}{2\sigma^2}\right), \quad x > 0. \tag{5.70}$$

Rozkład *logistyczny*, będący symetryczną alternatywą wobec rozkładu normalnego, ma funkcję gęstości:

$$f(x;\mu,s) = \frac{\exp\left(-\frac{x-\mu}{s}\right)}{s\left(1+\exp\left(-\frac{x-\mu}{s}\right)\right)^2}, \tag{5.71}$$

gdzie μ to parametr położenia, a $s > 0$ to parametr skali.

Rozkład *Weibulla*, często stosowany w analizie ryzyka i niezawodności, ma gęstość postaci:

$$f(x;k,\lambda) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} \exp\left[-\left(\frac{x}{\lambda}\right)^k\right], \quad x > 0, \tag{5.72}$$

gdzie $k > 0$ oznacza parametr kształtu, a $\lambda > 0$ parametr skali.

Rozkład *gamma*, stosowany do modelowania nieujemnych wartości losowych, definiowany jest jako:

$$f(x;\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x > 0, \tag{5.73}$$

gdzie $\alpha > 0$ to parametr kształtu, a $\beta > 0$ to parametr skali.

Wybór rozkładu następuje na podstawie bayesowskiego kryterium informacyjnego BIC (Schwarz, 1978)), które równoważy jakość dopasowania z liczbą parametrów modelu. Rozkład o najmniejszej wartości BIC jest wykorzystywany w dalszych analizach. W każdym z

czterech segmentów wybór pada na rozkład logistyczny, charakteryzujący się symetrycznym kształtem oraz cięższymi ogonami w porównaniu z rozkładem normalnym. Dopasowane rozkłady prawdopodobieństwa wraz z oszacowanymi parametrami dla poszczególnych segmentów przedstawiono w Tabeli 5.2.

Segment	**Rozkład**
C0020	Logis (0.76; 0.15)
C0040	Logis (0.98; 0.07)
C0050	Logis (1.01; 0.09)
C0070	Logis (0.92; 0.10)

Tabela 5.2: Dopasowane rozkłady prawdopodobieństwa i ich parametry.
Źródło: Opracowanie własne.

Ocena dopasowania rozkładów brzegowych

W dalszej części przeprowadzono ocenę jakości dopasowania modeli rozkładów brzegowych na zbiorze testowym. Modele te stanowią podstawę do konstrukcji rozkładów wielowymiarowych. Rozkłady brzegowe wyznaczone przy użyciu sieci neuronowych zostaną następnie wykorzystane do estymacji rozkładu wielowymiarowego z zastosowaniem drugiej architektury sieci neuronowej, przeznaczonej do modelowania struktury zależności między zmiennymi. Z kolei rozkłady uzyskane w podejściu parametrycznym posłużą do budowy parametrycznych modeli wielowymiarowych, w których zależności pomiędzy segmentami ubezpieczeń opisano za pomocą kaskad kopul C-vine oraz D-vine.

W konsekwencji rozkłady brzegowe wyznaczone na etapie analizy jednowymiarowej stanowią punkt wyjścia dla dwóch alternatywnych podejść do konstrukcji rozkładu wielowymiarowego:

1. podejścia parametrycznego, opartego na rozkładach o określonej postaci analitycznej,
2. podejścia nieparametrycznego, wykorzystującego modele sieci neuronowych, umożliwiającego elastyczne odwzorowanie zależności nieliniowych.

W celu zapewnienia obiektywnego porównania jakości dopasowania rozkładów brzegowych estymowanych z wykorzystaniem podejścia nieparametrycznego i parametrycznego przeprowadzono analizę na zbiorze testowym, niewykorzystywanym w procesie uczenia modeli. Wykorzystanie danych testowych gwarantuje niezależność próby weryfikacyjnej od próby uczącej, ogranicza ryzyko przeszacowania jakości dopasowania oraz umożliwia ocenę zdolności modeli do generalizacji poza dane treningowe. Dzięki temu uzyskane wyniki lepiej odzwierciedlają rzeczywistą jakość modeli, a nie efekt ewentualnego przeuczenia.

Punktem wyjścia dla oceny jakości dopasowania jest własność transformacji PIT (*Probability Integral Transform*). Dla ciągłej dystrybuanty F oraz zmiennej losowej $X \sim F$ zachodzi zależność:

$$F(X) \sim U(0,1), \tag{5.74}$$

co oznacza, że transformowane wartości zmiennej losowej powinny mieć rozkład jednostajny na przedziale $[0,1]$. Ocenę jakości dopasowania modeli brzegowych oparto na weryfikacji tej własności na zbiorze testowym.

Niech $\{x_{i,s}\}_{i=1}^{N_s}$ oznacza obserwację badanego wskaźnika w segmencie s, pochodzącą ze zbioru testowego, gdzie N_s jest liczbą obserwacji w tym zbiorze. Dla każdej metody $m \in \{\text{logistyczny}, \text{DNN}\}$ oraz każdego $x_{i,s}$ obliczono wartości

$$u_{i,s}^m = \hat{F}_s^m\left(x_{i,s}\right), \tag{5.75}$$

gdzie $\hat{F}_s^m$ oznacza oszacowaną dystrybuantę rozkładu brzegowego uzyskaną odpowiednio z modelu logistycznego lub z sieci neuronowej dla s-tego segmentu. Zgodność uzyskanych wartości $\{u_{i,s}^m\}_{i=1}^{n_s}$ z rozkładem jednostajnym na $[0, 1]$ weryfikowano testem Kołmogorowa–Smirnowa, formułując hipotezę zerową

$$H_0 : \; u_{i,s}^m \sim U(0, 1). \tag{5.76}$$

Wyniki oceny dopasowania przedstawiono w Tabeli 5.3.

Segment	**Opis rozkładu**	**Statystyka D**	**p-value**
C0020	Parametryczny	0.088982	0.7049
	Sieć neuronowa	0.0631	0.9613
C0040	Parametryczny	0.092183	0.6635
	Sieć neuronowa	0.0890	0.6879
C0050	Parametryczny	0.16509	0.07128
	Sieć neuronowa	0.1125	0.4142
C0070	Parametryczny	0.14656	0.1433
	Sieć neuronowa	0.0836	0.7729

Tabela 5.3: Ocena dopasowania rozkładów (test Kołmogorowa–Smirnowa) na zbiorze testowym.
Źródło: Opracowanie własne.

We wszystkich segmentach hipoteza $H_0 : u_{i,s}^m \sim U(0, 1)$ nie została odrzucona przy $\alpha = 0.05$ (wynik estymacji parametrycznej dla C0050 jest graniczny, $p = 0.07128$). Na podstawie uzyskanych wyników można stwierdzić, że zarówno podejście parametryczne, jak i oparte na sieciach neuronowych zapewniają poprawnie skalibrowane rozkłady brzegowe dla danych niewykorzystywanych w procesie uczenia, co potwierdza ich zdolność do generalizacji. Jednocześnie we wszystkich analizowanych segmentach model oparty na sieciach neuronowych charakteryzował się niższą wartością statystyki D oraz wyższymi wartościami p w porównaniu z modelem parametrycznym, co wskazuje na lepszą kalibrację rozkładów prawdopodobieństwa uzyskanych z wykorzystaniem sieci neuronowych. Lepsze dopasowanie modeli neuronowych może wynikać z ich zdolności do odwzorowania złożonych, nieliniowych zależności w danych ubezpieczeniowych, których modele parametryczne nie są w stanie w pełni uchwycić przy założeniu określonej postaci analitycznej rozkładu.

5.3.2 Estymacja struktury zależności

W dalszej części dokonano analizy struktury zależności pomiędzy analizowanymi segmentami ubezpieczeń w dwóch podejściach: nieparametrycznym oraz parametrycznym. W podejściu nieparametrycznym zastosowano sieć neuronową odwzorowującą nieliniowe zależności pomiędzy zmiennymi przy zachowaniu własności teoretycznych kopuli. Natomiast w podejściu

parametrycznym wykorzystano kaskady kopul typu C-vine oraz D-vine, umożliwiające dekompozycję rozkładu wielowymiarowego na kopule dwuwymiarowe.

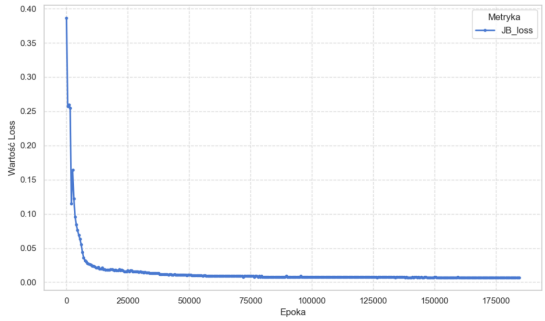
Estymacja z wykorzystaniem głębokich sieci neuronowych

Podobnie jak w przypadku sieci dla rozkładów brzegowych, do estymacji kopuli $\hat{C}(u;\theta_c)$ wykorzystano DNN. Jej zadaniem jest odwzorowanie czterowymiarowej dystrybuanty kopuli oraz odpowiadającej jej gęstości $\hat{f}_c(u;\theta_c)$. Na wejściu model otrzymuje wektor pseudo-obserwacji $u = (u_1, u_2, u_3, u_4) \in [0,1]^4$, uzyskanych z wcześniej wytrenowanych sieci dla rozkładów brzegowych. Wagi sieci dla rozkładów brzegowych zostały zamrożone, a ich parametry nie były aktualizowane w trakcie uczenia kopuli. We wszystkich warstwach ukrytych zastosowano funkcję aktywacji *tanh*, natomiast w warstwie wyjściowej wykorzystano funkcję aktywacji *sigmoid*, która ogranicza wartości wyjściowe estymowanej dystrybuanty do przedziału $[0,1]$.

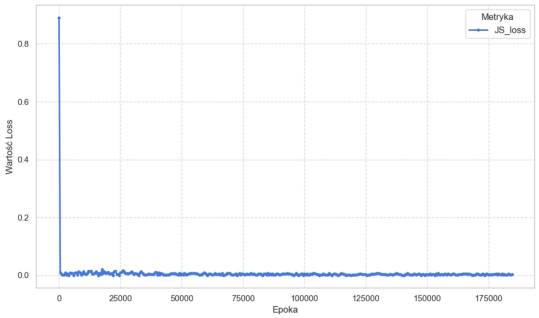
Proces uczenia modelu kopuli realizowany był w oparciu o minimalizację funkcji kosztu $L_c(\theta_c)$ opisanej we wzorze (5.53), stanowiącej ważoną sumę pięciu składników $L_{c(1)}$–$L_{c(5)}$:

- $L_{c(1)}$ odpowiada za minimalizację ujemnego log-wiarygodności (tj. maksymalizację funkcji wiarygodności, odpowiadającej dopasowaniu modelu do danych), zgodnie ze wzorem (5.35),
- $L_{c(2)}$ wymusza nieujemność estymowanej gęstości $\hat{f}_c$ poprzez zastosowanie funkcji *ReLu* na ujemnych wartościach gęstości, zgodnie ze wzorem (5.38),
- $L_{c(3)}$ zapewnia normalizację gęstości w przestrzeni $[0,1]^d$, minimalizując odchylenie całki z estymowanej gęstości od jedności, zgodnie ze wzorem (5.40),
- $L_{c(4)}$ kontroluje spełnienie warunków brzegowych kopuli zgodnie ze wzorem (5.45),
- $L_{c(5)}$ minimalizuje odchylenie pomiędzy wartościami modelowej dystrybuanty $\hat{F}_c(u_i;\theta)$ a jej empirycznym odpowiednikiem $\widehat{F}_{\text{emp}}(u_i)$, zgodnie ze wzorem (5.51).

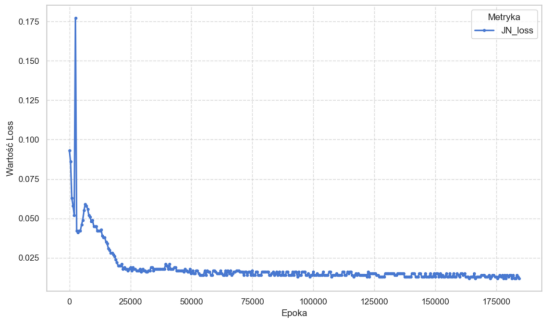
Uczenie przeprowadzono przy użyciu optymalizatora Nadam z początkowym współczynnikiem uczenia $5 \cdot 10^{-4}$, a jego wartość była adaptacyjnie regulowana zgodnie z zasadą *ReduceLROnPlateau* (Chollet et al., 2015), która automatycznie zmniejszała tempo uczenia o połowę w przypadku braku poprawy funkcji celu przez określone 1000 epok. Architektura sieci obejmuje cztery warstwy ukryte, a w każdej z nich znajduje się 10 neuronów. W każdej warstwie ukrytej kopuli wykorzystano regularizację typu l_2 z parametrem $\lambda = 0.001$. Dodatkowo, składniki funkcji kosztu $L_{c(2)}$–$L_{c(4)}$ pełniły funkcję regularizacyjną, podobnie jak w modelu sieci rozkładów brzegowych.



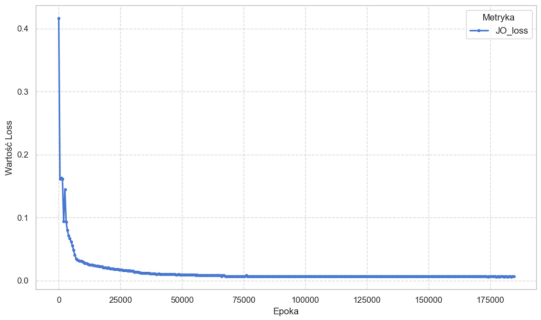
(a) $L_{c(4)}$ – warunki brzegowe kopuli



(b) $L_{c(3)}$ – normalizacja gęstości



(c) $L_{c(2)}$ – zgodność z dystrybuantą empiryczną



(d) $L_{c(2)}$ – nieujemność gęstości

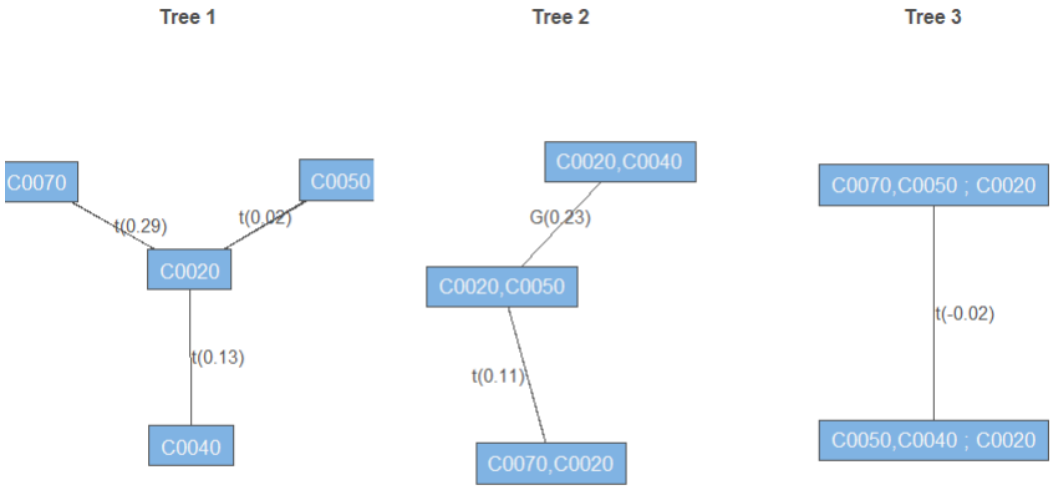
Rysunek 5.12: Funkcje ograniczające $L_{c(2)}$–$L_{c(5)}$ dla sieci neuronowej kopuli.
Źródło: Opracowanie własne.

Na Rysunku 5.12 przedstawiono przebiegi poszczególnych składników funkcji ograniczających $L_{c(2)}$–$L_{c(5)}$, odpowiadających za własności teoretyczne kopuli w trakcie uczenia sieci neuronowej kopuli $\hat{C}(u;\theta_c)$. Wszystkie krzywe strat maleją do zera, co potwierdza skuteczność procesu optymalizacji oraz stabilne dopasowanie modelu do danych. We wszystkich przypadkach zauważalny jest charakterystyczny dla sieci neuronowych etap gwałtownej redukcji błędu w początkowej fazie uczenia, po którym następuje stopniowa stabilizacja wartości w pobliżu zera. Takie zachowanie świadczy o prawidłowej zbieżności procesu uczenia. Model stopniowo spełnia wszystkie teoretyczne warunki kopuli. Brak oscylacji lub wzrostów strat w końcowych etapach uczenia wskazuje, że proces uczenia przebiegał stabilnie numerycznie, a gradienty nie ulegały ani zanikaniu, ani eksplozji. Wartości wszystkich funkcji ograniczających nałożonych na sieć po osiągnięciu stabilnego minimum utrzymują się na niskim poziomie przez resztę procesu uczenia. Oznacza to, że model osiągnął punkt optymalny, w którym równocześnie spełnione są warunki definicyjne kopuli oraz zapewnione jest właściwe dopasowanie do rozkładu empirycznego. Stabilny przebieg strat potwierdza również skuteczność zastosowanej regularizacji l_2 oraz funkcji ograniczających, które zapobiegły nadmiernemu dopasowaniu do danych treningowych. Brak wzrostu błędu po osiągnięciu minimum wskazuje, że model nie ulega przeuczeniu. Ostatecznie uzyskany model można uznać za zbieżny, stabilny i zgodny z teoretycznymi założeniami kopuli.

Estymacja parametryczna

Struktura C-vine

Struktura C-vine stanowi hierarchiczną reprezentację struktury zależności pomiędzy zmiennymi, w której na kolejnych poziomach (drzewach) modelowane są zarówno zależności między zmiennymi, jak i zależności warunkowe. Konstrukcja opiera się na wyborze zmiennej centralnej w pierwszym drzewie oraz dopasowaniu kopul dwuwymiarowych do par utworzonych z tym węzłem. W kolejnych drzewach analizowane są zależności pomiędzy pozostałymi zmiennymi, warunkowane przez zmienne już uwzględnione. Dzięki temu C-vine umożliwia elastyczne opisanie zarówno prostych, jak i bardziej złożonych zależności wielowymiarowych. W procesie estymacji zastosowano porządkowanie oparte na współczynniku τ Kendalla, który umożliwia identyfikację zmiennej o najsilniejszych powiązaniach z pozostałymi zmiennymi. Wybrana zmienna staje się korzeniem pierwszego drzewa. Wybór odpowiednich rodzin kopul dla poszczególnych krawędzi dokonany został przy użyciu kryterium informacyjnego AIC, co zapewnia kompromis między dopasowaniem a złożonością modelu. W oparciu o tę konstrukcję dopasowano strukturę C-vine, opisującą współzależności pomiędzy segmentami C0020, C0040, C0050 oraz C0070. Na Rysunku 5.13 przedstawiono wyrysowane drzewa dla kopuli C-vine:



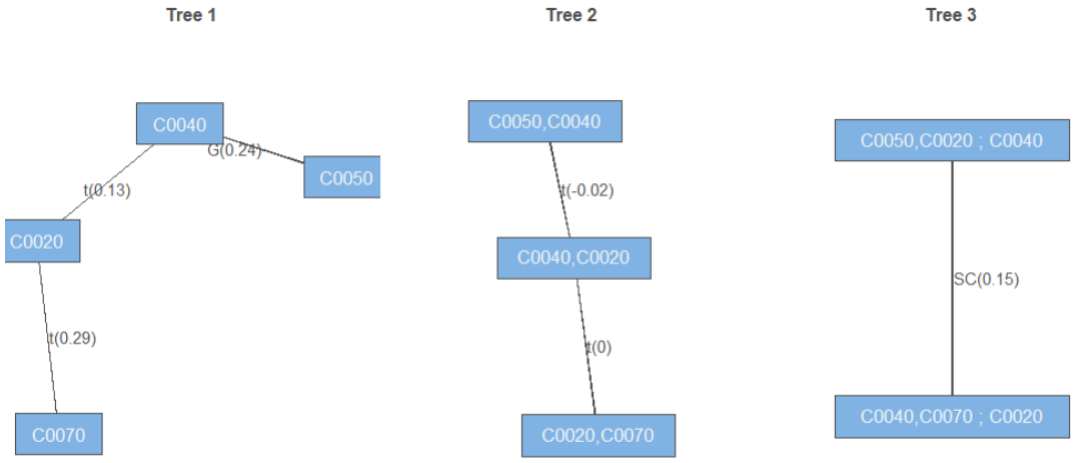
Rysunek 5.13: Struktura drzew kopuli C-vine dla segmentów C0020-C0070.
Źródło: Opracowanie własne.

Na krawędziach drzew przedstawiono dopasowane kopule dwuwymiarowe, a w nawiasach podano wartości współczynnika Kendalla τ dla każdej dobranej kopuli. W drzewie T_1 centralną zmienną jest C0020, która tworzy połączenia z pozostałymi segmentami. Najsilniejsza zależność o współczynniku $\tau = 0.29$ występuje pomiędzy C0020 a C0070 i została opisana kopulą t-Studenta. Zmienna C0020 wykazuje również umiarkowaną zależność z C0040, modelowaną kopulą t-Studenta ($\tau = 0.13$), natomiast z C0050 wiąże ją bardzo słaba zależność opisana kopulą t-Studenta ($\tau = 0.02$). W drzewie T_2 analizowane są zależności warunkowe. Najsilniejsza z nich to połączenie pomiędzy (C0020,C0040) a (C0020,C0050), opisane kopulą Gumbela z $\tau = 0.23$. Nieco słabsza zależność ($\tau = 0.11$) występuje między (C0020,C0050) a (C0070,C0020), również modelowana kopulą t-Studenta. W drzewie T_3 uwzględniono zależności wyższego

rzędu. Zależność pomiędzy (C0070,C0050 ; C0020) a (C0050,C0040 ; C0020) została opisana kopulą t-Studenta z ujemnym współczynnikiem $\tau = -0.02$, co wskazuje na bardzo słabą odwrotną korelację między tymi zmiennymi warunkowymi.

Struktura D-vine

W przeciwieństwie do struktury C-vine, w której jedna zmienna pełni rolę centralną i tworzy połączenia z pozostałymi zmiennymi, struktura D-vine zakłada uporządkowanie zmiennych w formie sekwencji. Każda z nich jest powiązana jedynie z najbliższymi sąsiadami, co sprawia, że zależności pomiędzy zmiennymi przyjmują postać łańcucha. W pierwszym drzewie analizowane są bezpośrednie zależności pomiędzy zmiennymi sąsiadującymi, natomiast kolejne drzewa opisują zależności pomiędzy parami bardziej oddalonymi, z uwzględnieniem zmiennych warunkujących. Takie ujęcie pozwala lepiej przedstawić lokalny charakter współzależności oraz uporządkowaną strukturę powiązań w danych. Budowę modelu D-vine rozpoczyna się od ustalenia optymalnego porządku zmiennych, a następnie dopasowuje się kopule dwuwymiarowe dla poszczególnych krawędzi i stopniowo modeluje zależności warunkowe w kolejnych etapach konstrukcji. Problem wyboru porządku w drzewie T_1 jest sprowadzony do problemu komiwojażera. Buduje się graf pełny ważony, którego wierzchołkami są zmienne losowe X_1, X_2, X_3, X_4 opisujące segmenty. Następnie do każdej krawędzi łączącej wierzchołki przypisuje się wagi, obliczając $1 - |\tau_{i,j}|$ dla $i, j \in \{1, 2, 3, 4\}$. Wyznacza się najkrótszą drogę, zaczynającą się z każdego wierzchołka, a następnie spośród najkrótszych dróg z każdego wierzchołka wybierana jest ta najkrótszą. Po wybraniu porządku w pierwszym drzewie modelowana jest struktura D-vine, która jest przedstawiona na Rysunku 5.14.



Rysunek 5.14: Struktura drzew kopuli D-vine dla segmentów C0020-C0070.
Źródło: Opracowanie własne.

Na krawędziach drzew zapisano dopasowane kopule dwuwymiarowe, a w nawiasach okrągłych podano wartości współczynnika τ Kendalla dla każdej dobranej kopuli.

W drzewie T_1 centralną zmienną jest $C0020$, która tworzy połączenia z pozostałymi segmentami. Najsilniejsza zależność o współczynniku $\tau = 0.29$ występuje pomiędzy C0020 a C0070 i została opisana kopulą t-Studenta. Zmienna C0020 wykazuje również umiarkowaną zależność z C0040, modelowaną kopulą t-Studenta ($\tau = 0.13$); natomiast zależność pomiędzy C0040 a C0050 została opisana kopulą Gumbela z $\tau = 0.24$.

W drzewie T_2 analizowane są zależności warunkowe. Najsłabsza z nich, o niewielkiej ujemnej wartości $\tau = -0.02$, występuje pomiędzy (C0050,C0040) a (C0040,C0020) i została opisana kopulą t-Studenta. Zależność między (C0040,C0020) a (C0020,C0070) ma współczynnik $\tau = 0$ i została również opisana kopulą t-Studenta.

W drzewie T_3 uwzględniono zależności wyższego rzędu. Zależność pomiędzy (C0050, C0020;C0040) a (C0040,C0070;C0020) została opisana kopulą Claytona o współczynniku $\tau = 0.15$, co wskazuje na umiarkowaną dodatnią zależność między tymi zmiennymi warunkowymi.

5.3.3 Ocena dopasowania rozkładu wielowymiarowego

Po wyznaczeniu rozkładów brzegowych oraz kopul opisujących strukturę zależności pomiędzy zmiennymi, zgodnie z twierdzeniem Sklara, skonstruowano wielowymiarowe rozkłady na podstawie zbioru danych treningowych. W analizie zastosowano dwa odmienne podejścia modelowania: parametryczne oraz nieparametryczne. Pierwsze z nich ma charakter parametryczny i opiera się na estymacji rozkładu wielowymiarowego poprzez wykorzystanie rozkładów brzegowych o postaci analitycznej oraz kopul, opisujących strukturę zależności z rodzin kaskad C-vine i D-vine. W tym celu, na podstawie przeprowadzonych analiz dopasowania, dla każdej zmiennej przyjęto rozkład logistyczny, który następnie połączono z wykorzystaniem kaskad kopul C-vine oraz D-vine, odwzorowujących zależności pomiędzy komponentami wektora losowego. Drugie podejście ma charakter nieparametryczny i wykorzystuje modelowanie oparte na sieciach neuronowych. W tym przypadku zarówno rozkłady brzegowe, jak i struktura zależności pomiędzy zmiennymi są estymowane bezpośrednio przez sieć neuronową, co umożliwia większą elastyczność modelu i pozwala uchwycić nieliniowe relacje między zmiennymi.

Ocena wyznaczonych wielowymiarowych rozkładów została przeprowadzona w dwóch ujęciach. W pierwszym z nich modele oceniono na podstawie danych testowych, które nie były wykorzystywane w procesie uczenia. Takie podejście pozwala na obiektywną weryfikację jakości dopasowania modelu oraz ocenę jego zdolności do generalizacji poza zbiór treningowy. W ramach tej analizy wyznaczano średnią wartość logarytmu funkcji wiarygodności, zgodnie ze wzorem (5.62), która informuje o stopniu zgodności pomiędzy rozkładem generowanym przez model a empirycznym rozkładem danych. Wysoka wartość tej miary świadczy o tym, że model trafnie odwzorowuje strukturę zależności pomiędzy zmiennymi oraz poprawnie opisuje ich wspólną zmienność w danych testowych.

Model	**Średnia log-wiarygodność**
C-vine	1.429865
D-vine	1.34221
Sieć neuronowa	3.116601

Tabela 5.4: Średnia logarytmiczna wiarygodność modeli na zbiorze testowym.
Źródło: Opracowanie własne.

Wyniki w Tabeli 5.4 pokazują wyraźną różnicę między podejściem parametrycznym a nieparametrycznym. Miara logarytmu wiarygodności dla modelu sieci neuronowej jest dwukrotnie wyższa niż dla modeli parametrycznych. Pokazuje to, że model oparty na sieciach neuronowych znacznie lepiej opisuje zmienność i zależności występujące w danych testowych. Wartości miary dla modeli parametrycznych są istotnie niższe i zbliżone do siebie. Podane wartości miary

potwierdzają, iż podejście oparte na sieciach neuronowych jest bardziej elastyczne i ma lepsze zdolności w modelowaniu złożonych struktur zależności niż podejście parametryczne. Tym samym model oparty na sieci neuronowej w lepszym stopniu odzwierciedla empiryczny rozkład danych. W procesie generowania realizacji będzie on tworzył próbki bardziej zbliżone do rzeczywistych obserwacji niż modele parametryczne. Drugie podejście oceny dopasowania modeli opiera się na analizie jakości realizacji generowanych z wyznaczonych rozkładów w porównaniu z danymi rzeczywistymi. W tym celu z każdego modelu wylosowano 10000 realizacji, które następnie porównano z obserwacjami rzeczywistymi przy użyciu miar *Energy Score* oraz *Variogram Score*, obliczonych zgodnie ze wzorami (5.66) i (5.68). Miary te umożliwiają ilościową ocenę podobieństwa pomiędzy rozkładem generowanym przez model a rozkładem empirycznym, bez konieczności analitycznego wyznaczania gęstości. Niższe wartości obu miar wskazują na mniejszą odległość pomiędzy realizacjami generowanymi przez model a danymi rzeczywistymi, co oznacza, że wygenerowane realizacje są „bliżej"danych rzeczywistych.

Model	**Energy Score**	**Variogram Score**
C-vine	0.1783883	0.3448577
D-vine	0.1808768	0.3343024
Sieć neuronowa	0.177288	0.318990

Tabela 5.5: Wartości miar *Energy Score* i *Variogram Score* dla analizowanych modeli.
Źródło: Opracowanie własne.

Wyniki w Tabeli 5.5 uzyskane dla miar *Energy Score* oraz *Variogram Score* wskazują, że model oparty na sieciach neuronowych osiągnął najniższe wartości obu wskaźników. Oznacza to, że generowane przez niego realizacje w największym stopniu odzwierciedlają dane rzeczywiste. Z kolei modele parametryczne, oparte na kopulach C-vine i D-vine, uzyskały wyższe wartości tych miar, co sugeruje, że ich zdolność do uchwycenia złożonych i nieliniowych zależności pomiędzy zmiennymi jest w pewnym stopniu ograniczona. Uzyskane rezultaty są spójne z wcześniejszą analizą logarytmu funkcji wiarygodności, potwierdzając, że podejście wykorzystujące sieci neuronowe lepiej odwzorowuje rzeczywisty rozkład danych oraz wykazuje większą efektywność w generowaniu realistycznych realizacji.

5.4 Wyznaczanie efektu dywersyfikacji

Efekt dywersyfikacji mierzy korzyści wynikające z agregacji ryzyka. Wartość efektu dywersyfikacji jest zdeterminowana sposobem modelowania struktury zależności. Opisanie struktury zależności na rzeczywistym poziomie jest zatem kluczową kwestią. W niniejszym podrozdziale efekt ten został wyznaczony w oparciu o dwa podejścia:

1. **Podejście parametryczne:**
   (a) Za pomocą formuły standardowej:
      - zakładając macierz korelacji daną przez twórców dyrektywy,
      - zakładając niezależność liniową między zmiennymi losowymi opisującymi moduły ryzyka (wartości korelacji między wszystkimi modułami ryzyka są równe 0),

- zakładając pełną zależność liniową między zmiennymi losowymi opisującymi moduły ryzyka (wartości korelacji pomiędzy wszystkimi modułami ryzyka są równe 1).

(b) Przy założeniu braku informacji na temat zależności – w tym przypadku wyznaczono przedziały zgodnie z algorytmem ARA.

(c) Wyznaczając strukturę zależności za pomocą kopuli typu C-vine oraz D-vine, przy rozkładach brzegowych dopasowanych parametrycznie.

2. **Podejście nieparametryczne:**

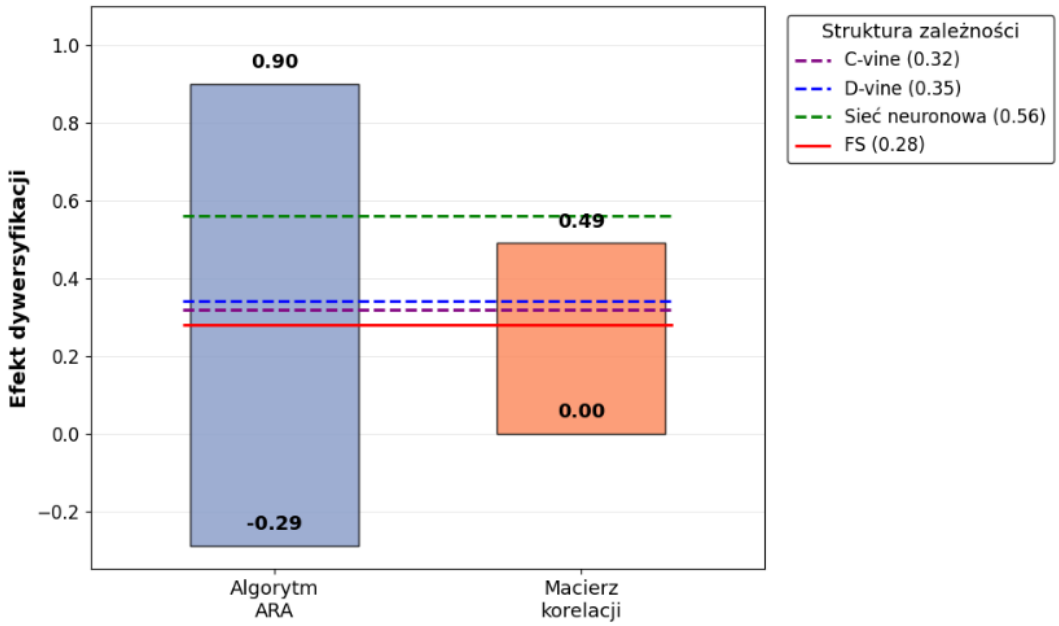
(a) Wyznaczając pełny rozkład wielowymiarowy, w którym zarówno rozkłady brzegowe, jak i struktura zależności zostały oszacowane przy wykorzystaniu modelu opartego na sieciach neuronowych.

Wyniki przeprowadzonych analiz przedstawiono w Tabeli 5.6, która syntetycznie zestawia wartości efektu dywersyfikacji uzyskane dla poszczególnych modeli:

Model	**Wartość dolna**	**Wartość górna**
Algorytm ARA	–0.29	0.90
Macierz korelacji	0.00	0.49
C-vine	0.32	
D-vine	0.35	
Sieć neuronowa	0.56	
FS	0.28	

Tabela 5.6: Wartości graniczne i odniesienia dla modeli zależności.
Źródło: Opracowanie własne.

Dla pełniejszego zobrazowania uzyskanych wyników, wartości te zostały przedstawione graficznie na Rysunku 5.15, co umożliwia łatwiejszą ocenę różnic pomiędzy poszczególnymi podejściami:



Rysunek 5.15: Otrzymane efekty dywersyfikacji.
Źródło: Opracowanie własne.

W przypadku zastosowania algorytmu ARA uzyskano najszerszy zakres możliwych wartości efektu dywersyfikacji, mieszczący się w przedziale od −0.29 do 0.90. Wynika to z faktu, iż w tej metodzie zakłada się jedynie znajomość rozkładów brzegowych, bez informacji o strukturze zależności. Przedział ten można interpretować jako pełny zakres możliwych wartości efektu dywersyfikacji, zależnych od przyjętej struktury zależności pomiędzy zmiennymi.

Następnie zakres niepewności wynikający z przyjętej struktury zależności został ograniczony do przypadku, w którym założono, że zależności między agregowanymi segmentami mają charakter liniowy. Przeprowadzono analizę wpływu wartości współczynnika korelacji na efekt dywersyfikacji, uzyskując wyniki w przedziale od 0.00 do 0.49.

Ograniczenie się wyłącznie do liniowej struktury zależności powoduje zatem zmniejszenie przedziału niepewności. Zgodnie z FS, efekt dywersyfikacji dla macierzy korelacji podanej przez twórców dyrektywy wynosi 0.28.

W dalszej części analizy efekt dywersyfikacji został określony z wykorzystaniem dwóch podejść: parametrycznego oraz nieparametrycznego. Oba podejścia opierają się na modelowaniu zależności pomiędzy segmentami przy użyciu kopuli, co pozwala na elastyczne i precyzyjne odwzorowanie zależności występujących w danych. W podejściu parametrycznym zależności pomiędzy zmiennymi modelowane są z wykorzystaniem kaskad kopul C-vine oraz D-vine, które umożliwiają dekompozycję wielowymiarowej struktury zależności na kopule dwuwymiarowe. Takie rozwiązanie pozwala uchwycić złożone, często nieliniowe zależności między zmiennymi oraz dokładniej odzwierciedlić charakter zależności obserwowanych w rzeczywistych danych. Porównanie wyników uzyskanych w podejściu parametrycznym z wartością referencyjną FS wskazuje na wyższy efekt dywersyfikacji. Dla modelu C-vine wartość efektu dywersyfikacji wynosi 0.32, natomiast dla modelu D-vine 0.35. W podejściu nieparametrycznym, w którym zarówno rozkłady brzegowe, jak i struktura zależności zostały oszacowane z wykorzystaniem sieci neuronowych, uzyskano jeszcze wyższą wartość efektu dywersyfikacji równą 0.56. Model nieparametryczny oparty na sieci neuronowej uzyskał najwyższą wartość średniej logarytmu

wiarygodności na zbiorze testowym, co potwierdza jego lepsze dopasowanie do rzeczywistej struktury zależności między segmentami w porównaniu z modelami parametrycznymi. Analiza miar energetycznych wskazuje, że generowane przez niego realizacje cechują się mniejszym błędem dopasowania, co świadczy o wyższej jakości modelu. Lepsze dopasowanie rozkładu wielowymiarowego przekłada się na oszacowanie efektu dywersyfikacji, które jest bliższe wartości odpowiadającej rzeczywistej strukturze współzależności. Uzyskana w podejściu z sieciami neuronowymi wartość 0.56 znacznie przewyższa wyniki dla kaskad kopul C-vine i D-vine, co potwierdza skuteczność rozwiązania nieparametrycznego.

6. Ocena ryzyka modelu wynikająca z niepewności struktury zależności

W czwartym rozdziale zwrócono uwagę na znaczenie ryzyka modelu w procesie oceny wypłacalności zakładów ubezpieczeń. Odnosząc ten problem do FS, można go sformułować następująco: w jakim stopniu metoda agregacji kapitałowych wymogów wypłacalności przyjęta w FS jest odporna na błędną lub niepełną specyfikację struktury zależności między rodzajami ryzyka ubezpieczyciela oraz w jakim stopniu odporność tej formuły wpływa na wartość wymogu kapitałowego SCR. Identyfikacja właściwej struktury zależności ma kluczowe znaczenie, ponieważ jej błędne lub nadmiernie uproszczone określenie może prowadzić do istotnych konsekwencji finansowych, co ilustrują przykłady przedstawione w podrozdziale 4.1.

W rozdziale 2 przedstawiono algorytmy umożliwiające ocenę niepewności wartości zagrożonej (VaR), wynikającej z niepewności dotyczącej struktury zależności pomiędzy agregowanymi czynnikami ryzyka. Istotne znaczenie ma tutaj analiza wrażliwości miary ryzyka na niepewność struktury zależności w procesie agregacji ryzyka. Zaprezentowano m.in. algorytmy, które nie wymagają żadnej informacji o strukturze zależności. W niniejszym rozdziale rozwijana jest ta tematyka. Przedstawiono cztery algorytmy umożliwiające ocenę niepewności wyznaczania VaR przy różnym zakresie informacji o zależnościach.

W pierwszym z zaprezentowanych algorytmów (*Algorytm 1*) przyjęto założenie o istnieniu wielowymiarowego rozkładu normalnego, natomiast brakowało pewności co do wartości współczynników korelacji liniowej. Algorytm ten pozwala badać, w jaki sposób zmiany w macierzy korelacji pomiędzy agregowanymi zmiennymi losowymi (ryzykami ubezpieczyciela) wpływają na wartość VaR. Umożliwia to ocenę odporności efektu dywersyfikacji obliczonego z wykorzystaniem FS, w której ryzyka agregowane są metodą wariancji–kowariancji, na niepewność wynikającą z przyjętej macierzy korelacji.

Korelacja Pearsona, w odróżnieniu od miary tau Kendalla, zależy nie tylko od struktury zależności (kopuli), lecz także od rozkładów brzegowych. Dla kopul z rodziny Archimedesa nie istnieje prosty wzór łączący parametr kopuli z korelacją Pearsona, ponieważ zawsze zależy on od wyboru rozkładów brzegowych. W oparciu o metodę siecznych opracowano algorytm (*Algorytm 3*) wyznaczający parametr zależności dla kopul Archimedesa dla zadanych rozkładów brzegowych oraz zadanej korelacji liniowej. Opisywany algorytm odnosi się do zależności opisanych za pomocą kopul z rodziny Archimedesa. Idąc krok dalej, rozważono problem znalezienia dla ustalonej wartości korelacji liniowej takiej struktury zależności, która maksymalizuje wartość VaR. Mając informacje o rozkładach brzegowych i stosując optymalizację wielokryterialną w sensie Pareto, opracowano algorytm (*Algorytm 2*), który rozwiązuje to zagadnienie. Oba zaprezentowane algorytmy odnoszą się zatem do analizy wrażliwości efektu dywersyfikacji na struktury zależności charakteryzujące się identycznymi współczynnikami korelacji liniowej pomiędzy agregowanymi rodzajami ryzyka.

Zbudowane trzy pierwsze algorytmy odnoszą się do analizy wrażliwości na korelację liniową. Czwarty algorytm uwzględnia szerszy zakres informacji o strukturze zależności. W praktyce stosunkowo prosto można oszacować strukturę zależności w części centralnej rozkładu, natomiast problematyczne jest określenie zależności w ogonach rozkładów. W kontekście Solvency II ma to szczególne znaczenie, ponieważ przy wyznaczaniu $VaR_{0.995}$ dominują zależności ogonowe. W odpowiedzi na pytanie, jak zależności w ogonach wpływają na wartość VaR, opracowano algorytm (*Algorytm 4*) oparty na zniekształconej mieszance kopul oraz

algorytmie symulowanego wyżarzania.

Opracowane autorskie algorytmy przyczyniły się do realizacji celów badawczych: C4, C5, C7, C8. Wszystkie zaproponowane algorytmy można sprowadzić do wspólnego zagadnienia, jakim jest analiza wrażliwości wartości VaR na niepewność struktury zależności w procesie agregacji ryzyka. Realizując cele C4 i C8, opracowane algorytmy w połączeniu z miarami ilościowymi pozwalają na stosowanie podejścia ilościowego do oceny ryzyka modelu.

6.1 Analiza wpływu niepewności macierzy korelacji liniowej na wartość VaR w wielowymiarowym modelu normalnym

Metoda agregacji wykorzystana w FS zakłada, że agregowane rodzaje ryzyka mają wielowymiarowy rozkład normalny, tzn. rozkłady brzegowe są normalne, a zależności pomiędzy nimi opisuje macierz korelacji. W tym ujęciu istotne staje się pytanie o wrażliwość oszacowań wartości VaR na parametry tej macierzy. Innymi słowy, należy zbadać, w jaki sposób zmiany poszczególnych elementów macierzy korelacji przekładają się na wartość wyznaczonego VaR. Zgodnie z postanowieniami dyrektywy, wartości współczynników korelacji są jednoznacznie określone i stanowią element obligatoryjny FS. Niezależnie od tego, że zakład ubezpieczeń może, w oparciu o własne dane empiryczne, wyznaczyć macierz korelacji odbiegającą od tej wskazanej w regulacji, jest zobligowany do stosowania macierzy określonej w dyrektywie.

Algorytm umożliwia analizę wpływu zmian zarówno pojedynczej, jak i wielu wartości korelacji między ryzykami na wartość VaR. Opracowana metoda umożliwia ilościową ocenę wpływu niepewności macierzy korelacji na wynik agregacji ryzyk w ramach FS. Umożliwia ona analizę, w jakim stopniu przyjęte w FS wartości współczynników korelacji, które są jedynie pojedynczymi liczbami opisującymi złożone zależności liniowe, badają wpływ zmian wartości korelacji na wartość zagregowanego VaR oraz efekt dywersyfikacji. W praktyce rzeczywiste zależności między ryzykami mogą mieścić się w szerszym przedziale. dlatego metoda ta pozwala badać stabilność wyników również w sytuacji, gdy korelacje odbiegają od wartości bazowych określonych w regulacjach, jak również od korelacji wyznaczonej przez sam zakład. Dzięki temu stanowi ona praktyczne narzędzie do oceny wrażliwości i odporności modelu na niepewność w estymacji zależności między ryzykami, a tym samym identyfikuje ograniczenia ryzyka modelu w ramach FS. Algorytm ma następujący przebieg:

Algorytm 1

1. Określa macierz korelacji między analizowanymi rodzajami ryzyka, przyjmując jej wyjściową postać zgodną z założeniami FS.
2. Wybiera jeden lub kilka elementów macierzy korelacji, które będą podlegały modyfikacji.
3. Dokonuje zmian wartości wybranych współczynników korelacji o zadaną wielkość, symetrycznie względem przekątnej macierzy.
4. Na podstawie zmodyfikowanej macierzy ponownie oblicza zagregowaną wartość VaR.
5. Porównuje uzyskane wyniki z wartością bazową, oceniając, jak zmiany poszczególnych korelacji wpływają na wynik agregacji oraz efekt dywersyfikacji.

6. Analizuje wrażliwość modelu na zmiany korelacji, umożliwiając identyfikację przedziału istotności, w którym modyfikacje tych parametrów nie prowadzą do statystycznie znaczących różnic w wartości VaR.

W analizie wrażliwości jedno- i wieloczynnikowej ocenia się wpływ jednoczesnych zmian jednego, dwóch lub większej liczby parametrów modelu na zagregowaną wartość VaR. Szczegóły algorytmu przedstawione są poniżej. Rozważania prowadzi się dla SCR, zgodnie z FS, gdzie SCR definiuje się jako różnicę $\text{VaR}_{0.995}$ i średniej, natomiast $W = [SCR_1, \ldots, SCR_d]$ oznacza wektor wymogów kapitałowych poszczególnych rodzajów ryzyka. Gdy średnia jest równa 0, zachodzi tożsamość $SCR = \text{VaR}$. Przyjmuje się, że w procesie agregacji uczestniczy d ryzyk, w związku z czym obliczenia prowadzi się dla macierzy wymiaru $d \times d$.

Definiuje się macierz R^*, otrzymaną z macierzy R poprzez modyfikację wybranych elementów o ε_h, gdzie indeks h identyfikuje zastosowaną perturbację ε_h. Ponieważ macierz korelacji jest symetryczną macierzą kwadratową wymiaru $d \times d$, zmiany wartości korelacji wprowadza się symetrycznie względem przekątnej.

Zatem liczba możliwych zmian odpowiada liczbie wyrazów pod przekątną, czyli $\frac{d(d-1)}{2}$. Jeśli dokonuje się $2k$ zmian wyrazów macierzy R (k pod przekątną i k symetrycznie nad przekątną), to istnieje

$$\binom{\frac{d(d-1)}{2}}{k} \tag{6.1}$$

możliwości wyboru tych zmian.

Dalej definiuje się macierz $S_{i_h j_h}$ wymiaru $d \times d$, złożoną z zer, z wyjątkiem jedynki występującej wyłącznie na pozycji (i_h, j_h). Formalnie można ją zapisać jako:

$$S_{i_h j_h} = e_{i_h} e_{j_h}^{\top}, \tag{6.2}$$

gdzie e_{i_h} oznacza wektor bazowy z jedynką na pozycji i_h i zerami w pozostałych miejscach.

Innymi słowy:

$$S_{i_h, j_h} = \begin{cases} 1, & \text{na pozycji } (i_h, j_h), \\ 0, & \text{w pozostałych miejscach.} \end{cases} \tag{6.3}$$

Przykładowa postać takiej macierzy wygląda następująco:

$$S_{i_h j_h} = \begin{bmatrix} 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & 1 & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \end{bmatrix}, \tag{6.4}$$

gdzie jedynka znajduje się w $i_h - tym$ wierszu i $j_h - tej$ kolumnie.

Taka konstrukcja pozwala w prosty sposób opisać dowolną modyfikację wyrazów macierzy korelacji R. Jeśli wybrany wyraz macierzy (i_h, j_h) oraz symetrycznie (j_h, i_h) są zwiększone o ε_h, to nową macierz korelacji R^* można zapisać jako:

$$R^* = R + \sum_{h=1}^{k} \left(\varepsilon_h S_{i_h j_h} + \varepsilon_h S_{j_h i_h} \right). \tag{6.5}$$

Korzystając z powyższej postaci macierzy R^*, wyznacza się $WR^*W^\top$:

$$\begin{aligned} WR^*W^\top &= W\Big(R + \sum_{h=1}^{k} \left(\varepsilon_h S_{i_h j_h} + \varepsilon_h S_{j_h i_h}\right)\Big)W^\top \\ &= WRW^\top + \sum_{h=1}^{k} \varepsilon_h \left(WS_{i_h j_h}W^\top + WS_{j_h i_h}W^\top\right). \end{aligned} \tag{6.6}$$

Ponieważ dla ustalonych i_h i j_h:

$$\varepsilon_h WS_{i_h j_h}W^\top = \varepsilon_h SCR_{i_h} SCR_{j_h}, \tag{6.7}$$

oraz

$$\varepsilon_h WS_{j_h i_h}W^\top = \varepsilon_h SCR_{j_h} SCR_{i_h}, \tag{6.8}$$

otrzymuje się:

$$WR^*W^\top = WRW^\top + 2\sum_{h=1}^{k} \varepsilon_h SCR_{i_h} SCR_{j_h}. \tag{6.9}$$

Powyższe obliczenia prowadzą do następującego ogólnego wniosku. Niech R będzie symetryczną macierzą korelacji, a (i_h, j_h) oznacza numer wiersza i kolumny pod przekątną macierzy R, do której dodano ε_h dla $h \in \{1, 2, \ldots, k\},\ k \leq \frac{d(d-1)}{2}$ (analogicznie, (i_h, j_h) oznacza numer wiersza i kolumny nad przekątną macierzy R, do której dodano ε_h). Przekształconą macierz zapisano jako R^*. Wtedy dla dowolnego W otrzymano:

$$WR^*W^\top = WRW^\top + 2\sum_{h=1}^{k} \varepsilon_h SCR_{i_h} SCR_{j_h}. \tag{6.10}$$

Z powyższego wniosku wynika, że jeśli dokonuje się k zmian wartości korelacji w macierzy R, a każdą korelację zmieniamy odpowiednio o $\varepsilon_1, \ldots, \varepsilon_k$, to $WR^*W^\top$ jest równy sumie $WRW^\top$ (gdzie macierz R jest macierzą, w której zostały wykonane zmiany) oraz podwojonych iloczynów wymogów kapitałowych, między którymi zmieniamy wartość korelacji i wartości zmiany ε_h.

6.2 Maksymalizacja wartości VaR przy ustalonej korelacji liniowej

Wartość korelacji liniowej opisuje w sposób jednoznaczny jedynie zależności liniowe. Zatem może istnieć wiele różnych struktur zależności dla tej samej wartośći korelacji liniowej. Przyjmując zadane rozkłady brzegowe oraz ustaloną macierz korelacji, poszukiwany jest taki rozkład wielowymiarowy, który, przy zachowaniu zadanej korelacji liniowej, prowadzi do maksymalnej wartości VaR. Pokazuje, że korelacja liniowa, będąca jedynie liczbową miarą zależności, nie odzwierciedla w pełni rzeczywistej zależności między zmiennymi, które w praktyce mogą mieć charakter nieliniowy lub asymetryczny. W efekcie, nawet przy ustalonej macierzy korelacji można otrzymać wiele różnych struktur zależności. W kontekście FS metoda ta pozwala ocenić, na ile przyjęte w regulacji założenia dotyczące zależności między modułami ryzyka są wystarczające do dokładnej identyfikacji rzeczywistej struktury zależności między rodzajami ryzyka ubezpieczyciela. Proponowany algorytm stanowi tym samym praktyczne narzędzie do badania

odporności i stabilności wyników agregacji SCR oraz do identyfikacji ryzyka modelu wynikającego z uproszczonego, liniowego opisu zależności pomiędzy ryzykami. Algorytm przebiega w następujący sposób:

Algorytm 2

1. Zadawane są rozkłady brzegowe, a także ustalana jest macierz korelacji liniowej między zmiennymi losowymi.

2. Generowany jest zbiór możliwych przekształceń danych poprzez permutacje wierszy w danej kolumnie macierzy, co pozwala uzyskać różne struktury zależności przy zachowaniu tych samych rozkładów brzegowych.

3. Dla każdej permutacji obliczana jest wartość VaR oraz różnica uzyskanej macierzy korelacji i zadanej macierzy w punkcie 1.

4. Problem formułowany jest jako zadanie wielokryterialne, w którym dąży się do jednoczesnej maksymalizacji wartości VaR oraz minimalizacji różnicy między otrzymaną a zadaną macierzą korelacji.

5. W kolejnych iteracjach oceniane są wygenerowane rozwiązania w sensie dominacji Pareto, a zbiór najlepszych, niezdominowanych rozwiązań jest na bieżąco aktualizowany.

6. Proces powtarzany jest do momentu, aż zbiór Pareto ustabilizuje się, tworząc front reprezentujący kompromisy pomiędzy maksymalizacją wartości VaR a dokładnością odwzorowania zadanej struktury korelacyjnej.

Szczegóły algorytmu przedstawione są poniżej. Niech $X = \{X_{i,j}\}_{i=1,\ldots,n}^{j=1,\ldots,d}$ oznacza macierz danych o wymiarach $n \times d$, w której j-ta kolumna odpowiada zmiennej losowej $X^{(j)}$ o rozkładzie F_j. Przez $P^{(j)}$ oznacza się permutację zbioru indeksów $\{1, 2, \ldots, n\}$ w sobie dla j-tej kolumny. Macierz $X_{P^{(1)},\ldots,P^{(d)}}$ otrzymuje się poprzez zastosowanie niezależnych permutacji $P^{(j)}$ do wierszy w danej kolumnie macierzy X:

$$X_{P^{(1)},\ldots,P^{(d)}} = \{X_{P^{(j)}(i),j}\}_{i=1,\ldots,n}^{j=1,\ldots,d}. \tag{6.11}$$

Należy zauważyć, że zastosowanie permutacji nie wpływa na właściwości statystyczne zmiennych losowych $X^{(j)}$, natomiast prowadzi do zmiany struktury zależności pomiędzy tymi rozkładami. Zbiór wszystkich możliwych przekształceń X można zapisać jako

$$\Omega = \{X_{P^{(1)},\ldots,P^{(d)}} : P^{(j)} \in \mathcal{S}_n \text{ dla każdego } j = 1, \ldots, d\}, \tag{6.12}$$

gdzie $\mathcal{S}_n$ oznacza wszystkie permutacje zbioru $\{1, 2, \ldots, n\}$.

Problem sprowadza się do wyznaczenia takich permutacji $P^{(1)}, \ldots, P^{(d)}$, dla których macierz $X_{P^{(1)},\ldots,P^{(d)}}$ równocześnie maksymalizuje VaR i zachowuje zadaną macierz korelacji. Postawiony problem można ująć jako problem maksymalizacji wielokryterialnej:

$$\max_{P^{(1)},\ldots,P^{(d)} \in \mathcal{S}_n} \left(VaR(X_{P^{(1)},\ldots,P^{(d)}}),\ -CE(X_{P^{(1)},\ldots,P^{(d)}}) \right). \tag{6.13}$$

Pierwsze kryterium jest tożsame z maksymalizacją wartości zagrożonej:

$$VaR(X_{P^{(1)},\ldots,P^{(d)}}) = \text{quantile}_{\alpha}\left(\sum_{j=1}^{d} X_{P^{(j)},j}\right), \tag{6.14}$$

gdzie quantile_{α} oznacza kwantyl rzędu α, natomiast $P^{(j)}$ stanowi permutację elementów w j-tej kolumnie macierzy danych.

Drugie kryterium związane jest z minimalizacją błędu dopasowania korelacji wyznaczonej dla $X_{P^{(1)},\ldots,P^{(d)}}$ z zadaną korelacją:

$$CE(X_{P^{(1)},\ldots,P^{(d)}}) = \sum_{k=1}^{d}\sum_{\ell=1}^{d} (C_{\text{obecna}}[k,\ell] - C_{\text{docelowa}}[k,\ell])^2, \tag{6.15}$$

gdzie C_{obecna} oznacza macierz korelacji obliczoną na podstawie $X_{P^{(1)},\ldots,P^{(d)}}$, natomiast C_{docelowa} jest zadaną macierzą docelową. W celu rozwiązania postawionego problemu zastosowano podejście oparte na analizie Pareto (Deb, 2001; Ehrgott, 2005). Podejście to umożliwia wyznaczenie zbioru wszystkich niezdominowanych rozwiązań, czyli takich, dla których nie istnieje inne rozwiązanie lepsze jednocześnie w obu kryteriach. W wyniku permutacji kolumn zmieniają się struktury zależności, a we froncie Pareto znajdują się wielowymiarowe rozkłady o takich strukturach zależności, które stanowią najlepsze kompromisy pomiędzy maksymalizacją VaR a minimalizacją błędu różnicy między zadaną macierzą korelacji a macierzą uzyskaną dla wyznaczonego wielowymiarowego rozkładu. W celu porównania rozwiązań definiuje się poniżej pojęcie dominacji Pareto, odpowiednie dla rozważanego problemu. Niech X_P oraz $X_{P'}$ oznaczają dwie macierze powstałe po zastosowaniu różnych permutacji $P^{(1)},\ldots,P^{(d)}$ oraz $P'^{(1)},\ldots,P'^{(d)}$. Mówi się, że rozwiązanie X_P dominuje rozwiązanie $X_{P'}$, co zapisuje się jako $X_P \succ X_{P'}$, jeżeli spełnione są warunki:

$$VaR(X_P) \geq VaR(X_{P'}) \quad \text{oraz} \quad CE(X_P) \leq CE(X_{P'}), \tag{6.16}$$

przy czym co najmniej jedna z powyższych nierówności jest ostra:

$$VaR(X_P) > VaR(X_{P'}) \quad \text{lub} \quad CE(X_P) < CE(X_{P'}). \tag{6.17}$$

Oznacza to, że rozwiązanie X_P nie jest gorsze od $X_{P'}$ w żadnym z kryteriów i jest lepsze w co najmniej jednym z nich.

Zbiór wszystkich rozwiązań, które pozostają niezdominowane w sensie Pareto, można zapisać jako

$$S^* = \{X_P \in \Omega \mid \nexists X_{P'} \in \Omega \text{ takie, że } X_{P'} \succ X_P\}. \tag{6.18}$$

Zatem zbiór S^* składa się z permutacji macierzy X, dla których nie istnieje inne rozwiązanie lepsze równocześnie w obu kryteriach. W dalszej analizie front Pareto będzie stanowił punkt odniesienia przy wyznaczaniu kompromisów między maksymalizacją VaR a minimalizacją błędu odwzorowania macierzy korelacji. Niech k oznacza numer iteracji, gdzie $k = 1, 2, \ldots, K$, a K jest całkowitą liczbą iteracji.

W k-tej iteracji losowana jest permutacja $P_k^{(1)},\ldots,P_k^{(d)}$, która następnie jest wykorzystana do zbudowania macierzy:

$$X^{(k)} = X_{P_k^{(1)},\ldots,P_k^{(d)}}. \tag{6.19}$$

Dla macierzy $X^{(k)}$ obliczane są wartości kryteriów:

$$VaR^{(k)} = VaR(X^{(k)}), \qquad CE^{(k)} = CE(X^{(k)}). \tag{6.20}$$

Następnie, rozwiązanie $X^{(k)}$ jest porównywane z rozwiązaniami znajdującymi się już w wcześniejszych iteracjach w zbiorze Pareto S^*. Algorytm może napotkać trzy przypadki:

- jeśli rozwiązanie $X^{(k)}$ jest zdominowane przez rozwiązania należące do zbioru Pareto S^*, to wówczas nie zostaje ono zaakceptowane,
- jeśli rozwiązanie $X^{(k)}$ nie jest zdominowane, to wówczas rozwiązanie $X^{(k)}$ zostaje dodane do zbioru S^*,
- jeśli wszystkie rozwiązania, które znajdują się w zbiorze Pareto S^*, są zdominowane przez rozwiązanie $X^{(k)}$, to zostaną usunięte.

Zatem aktualizacja zbioru Pareto ma postać:

$$S^* \leftarrow S^* \cup \{X^{(k)}\} \setminus \{X_P \in S^* \mid X^{(k)} \succ X_P\}. \tag{6.21}$$

W ten sposób w kolejnych iteracjach $k = 1, 2, \ldots, K$ zbiór Pareto stopniowo się rozszerza, a proces ten prowadzi do przybliżenia optymalnego frontu reprezentującego kompromisy pomiędzy maksymalizacją VaR a minimalizacją błędu odwzorowania macierzy korelacji.

6.3 Wyznaczanie parametru zależności dla kopul Archimedesa przy zadanej korelacji liniowej

Niniejszy podrozdział porusza dokładnie tę samą tematykę, co podrozdział 6.2, zawężając strukturę zależności do kopul Archimedesa. Proponowany algorytm wyznacza parametr zależności θ dla kopuli Archimedesa w taki sposób, aby odpowiadał zadanej wartości korelacji liniowej między zmiennymi losowymi o zadanych rozkładach brzegowych. W kontekście ryzyka modelu algorytm pozwala szczegółowo zbadać, jakie struktury zależności pomiędzy ryzykami opisane kopulami z rodziny Archimedesa odpowiadają tej samej, zadanej wartości korelacji liniowej oraz jakie wartości zagregowanego VaR są z nimi związane. Umożliwia tym samym identyfikację sytuacji, w których modele o jednakowej macierzy korelacji prowadzą do odmiennych wyników agregacji, co wskazuje na potencjalną wrażliwość modelu na przyjętą postać zależności. W kontekście FS metoda ta pozwala ocenić, w jakim stopniu założenie o liniowym charakterze zależności może prowadzić do zaniżenia lub przeszacowania efektu dywersyfikacji, a tym samym do niedokładnego wyznaczenia SCR. Dzięki temu algorytm stanowi istotne narzędzie w analizie odporności modelu na uproszczenia dotyczące struktury zależności i umożliwia lepsze zrozumienie potencjalnych źródeł ryzyka modelu wynikających z przyjętych założeń regulacyjnych. Przebieg algorytmu jest następujący:

Algorytm 3

1. W pierwszym kroku zadawane są rozkłady brzegowe oraz parametr korelacji liniowej, dla którego ma zostać znaleziony parametr zależności kopuli oraz kopula z rodziny Archimedesa.

2. Definiowana jest funkcja celu, opisująca różnicę między korelacją uzyskaną dla dwuwymiarowego rozkładu opisanego kopulą Archimedesa oraz rozkładami brzegowymi a zadaną wartością korelacji.

3. W kolejnych iteracjach, wykorzystując metodę siecznych, wyznaczane są kolejne przybliżenia parametru, a ich akceptacja zależy od spełnienia warunków zapewniających stabilność obliczeń i odpowiednie tempo zbieżności.

4. Jeżeli warunki te nie są spełnione, stosowana jest metoda bisekcji, gwarantująca zachowanie zbieżności procesu numerycznego.

5. Iteracje kontynuowane są do momentu osiągnięcia założonej dokładności, odnosząc się do otrzymanego przedziału potencjalnych parametrów kopuli jak i wartośći funkcji celu.

6. Ostatecznie za parametr zależności przyjmowana jest wartość, dla której uzyskana korelacja liniowa jest najbliższa zadanej wartośći współczynnika korelacji.

Szczegóły algorytmu przedstawiono poniżej. Rozważany jest dwuwymiarowy wektor losowy $X = (X_1, X_2)$ z ustalonymi rozkładami brzegowymi F_1 i F_2. Struktura zależności jest opisana kopulą C_θ z parametrem zależności θ. Przez Θ oznaczmy zbiór wszystkich dopuszczalnych parametrów dla kopuli C_θ. Zgodnie z twierdzeniem Sklar'a, wielowymiarowy rozkład można zapisać:

$$H_\theta(x_1, x_2) \;=\; C_\theta\big(F_1(x_1),\, F_2(x_2)\big). \tag{6.22}$$

Dla ustalonego $\varrho^\star \in (-1, 1)$ i dla ustalonej kopuli C_θ szukany jest parametr zależności θ, aby korelacja (X_1, X_2) miała wartość $\rho^\star$. Wprowadzamy funkcję parametru korelacji dla rozkładu H_θ:

$$\rho(\theta) = \mathrm{Corr}_{H_\theta}(X_1, X_2). \tag{6.23}$$

Zadanie sprowadza się do wyznaczenia takiego parametru zależności kopuli $\hat{\theta} \in \Theta$, dla którego spełniony jest warunek:

$$\rho(\hat{\theta}) = \rho^\star. \tag{6.24}$$

Algorytm opiera się na symulacjach oraz na numerycznym przeszukiwaniu parametrów Θ. Funkcja celu ma postać:

$$f(\theta) = \widehat{\rho}(\theta) - \rho^\star. \tag{6.25}$$

W oparciu o metodę opisaną przez Brenta, (Brent, 1973) budowany jest algorytm numeryczny służący do znalezienia takiego parametru $\hat{\theta} \in \Theta$, że:

$$f(\hat{\theta}) \;\to\; 0, \tag{6.26}$$

co jest tożsame z osiągnięciem zadanej korelacji liniowej przez rozkład H_θ.

Zakładamy ciągłość funkcji $\theta \mapsto f(\theta)$ na pewnym przedziale $[\theta_\mathrm{L}, \theta_\mathrm{U}] \subset \Theta$. Wybieramy dopuszczalny przedział parametrów dla danej kopuli $[\theta_\mathrm{L}, \theta_\mathrm{U}]$ i obliczamy $f(\theta_\mathrm{L})$ oraz $f(\theta_\mathrm{U})$. Jeżeli spełniony jest warunek:

$$f(\theta_\mathrm{L}) \cdot\; f(\theta_\mathrm{U}) \le 0, \tag{6.27}$$

Wtedy, zgodnie z twierdzeniem o wartości pośredniej (Apostol, 1974; Rudin, 1976), istnieje $\theta^\star \in [\theta_\mathrm{L}, \theta_\mathrm{U}]$ takie, że $f(\theta^\star) = 0$.

W k-tej iteracji na przedziale $[\theta_{\mathrm{L}}^{(k)}, \theta_{\mathrm{U}}^{(k)}]$ wyznaczany jest kandydat $\hat{\theta}$ jako $s^{(k)} \in (\theta_{\mathrm{L}}^{(k)}, \theta_{\mathrm{U}}^{(k)})$, wykorzystując regułę siecznych (Ortega & Rheinboldt, 1970; Stoer & Bulirsch, 2002; Traub, 1964):

$$s^{(k)} \;=\; \theta_{\mathrm{U}}^{(k)} - f_n(\theta_{\mathrm{U}}^{(k)})\,\frac{\theta_{\mathrm{U}}^{(k)} - \theta_{\mathrm{L}}^{(k)}}{f(\theta_{\mathrm{U}}^{(k)}) - f(\theta_{\mathrm{L}}^{(k)})}. \tag{6.28}$$

Jeśli kandydat $s^{(k)}$ nie spełnia kryteriów akceptacji:

- $s^{(k)} \in (\theta_{\mathrm{L}}^{(k)}, \theta_{\mathrm{U}}^{(k)})$ ($s^{(k)}$ należy do dziedziny),
- niech $L^{(k)} = \theta_{\mathrm{U}}^{(k)} - \theta_{\mathrm{L}}^{(k)}$. Akceptacja nowego kandydata $s^{(k)}$ jest możliwa tylko wtedy, gdy po wyznaczeniu nowej długości przedziału

  $$L_{\mathrm{new}}^{(k)} = \max\{\, s^{(k)} - \theta_{\mathrm{L}}^{(k)},\; \theta_{\mathrm{U}}^{(k)} - s^{(k)} \,\} \tag{6.29}$$

  spełnia

  $$L_{\mathrm{new}}^{(k)} \;\le\; \kappa\, L^{(k)}, \qquad \kappa \le \frac{1}{2},$$

  co zapewnia skrócenie przedziału możliwych rozwiązań nie gorsze niż w metodzie bisekcji (Burden & Faires, 2011; Brent, 1973),
- konstrukcja $s^{(k)}$ jest stabilna numerycznie, jeśli spełnia jednocześnie warunki:
  - mianownik jest wartością niezerową $|f(\theta_{\mathrm{U}}^{(k)}) - f(\theta_{\mathrm{L}}^{(k)})| > \varepsilon_f$,
  - parametr $s^{(k)}$ nie jest równy krańcom dziedziny. Wprowadzany jest stały parametr $\eta \in (0, \frac{1}{2})$ i sprawdza się, czy punkt $s^{(k)}$ jest w odległości η od krańców dziedziny:

    $$s^{(k)} \in \left(\theta_{\mathrm{L}}^{(k)} + \eta L^{(k)},\; \theta_{\mathrm{U}}^{(k)} - \eta L^{(k)}\right) \quad\Longleftrightarrow\quad \min\{\, s^{(k)} - \theta_{\mathrm{L}}^{(k)},\; \theta_{\mathrm{U}}^{(k)} - s^{(k)} \,\} \;\ge\; \eta\, L^{(k)},$$

    gdzie $L^{(k)} = \theta_{\mathrm{U}}^{(k)} - \theta_{\mathrm{L}}^{(k)}$. Zapewnia to, że $s^{(k)}$ leży wewnątrz przedziału, dzięki czemu algorytm jest stabilniejszy oraz ma efektywniejszą kontrakcję.
  - warunek kroku - wymagamy, aby odległość kandydata $s^{(k)}$ od bliższego krańca przedziału nie była zbyt mała:

    $$\min\{\, |s^{(k)} - \theta_{\mathrm{L}}^{(k)}|,\; |\theta_{\mathrm{U}}^{(k)} - s^{(k)}| \,\} \;\ge\; \Delta_{\min}, \tag{6.30}$$

    $$\Delta_{\min} \;=\; \max\left\{\varepsilon_x,\; \varepsilon_{\mathrm{rel}} \cdot \max\left(|\theta_{\mathrm{L}}^{(k)}|,\; |\theta_{\mathrm{U}}^{(k)}|\right)\right\}, \tag{6.31}$$

    $$\varepsilon_x > 0, \qquad \varepsilon_{\mathrm{rel}} > 0. \tag{6.32}$$

    Parametry ε_x oraz $\varepsilon_{\mathrm{rel}}$ pełnią rolę progów numerycznych, kontrolujących minimalny krok zmiany wartości parametru θ w kolejnych iteracjach i są ustalane przed uruchomieniem algorytmu. Parametr ε_x oznacza bezwzględną dokładność θ, czyli najmniejszą dopuszczalną zmianę parametru. Z kolei, $\varepsilon_{\mathrm{rel}}$ stanowi względną dokładność, określającą minimalną zmianę wyrażoną w proporcji do bieżącej wielkości przedziału poszukiwań. Oba te parametry razem definiują próg, poniżej którego algorytm uznaje, że zmiana jest zbyt mała i nie prowadzi do poprawy obliczeń.

Warunek eliminuje ruchy tak małe, że nie przynoszą już kontrakcji przedziału. W odróżnieniu od poprzedniego warunku (który kontroluje położenie $s^{(k)}$ odnosząc się do krańcowych punktów przedziału w oparciu o $L^{(k)}$), pełni on rolę progu numerycznego dla wielkości kroku. Zapobiega wykonywaniu ruchów tak małych, że są praktycznie nierozróżnialne i nie powodują kontrakcji, niezależnie od długości aktualnego przedziału.

Jeśli przynajmniej jedno z powyższych kryteriów akceptacji nie jest spełnione, kandydata wyznaczamy zgodnie z metodą bisekcji:

$$s^{(k)} = \frac{\theta_{\mathrm{L}}^{(k)} + \theta_{\mathrm{U}}^{(k)}}{2}. \tag{6.33}$$

Następnie oblicza się $f\big(s^{(k)}\big)$ i aktualizuje bieżący przedział dziedziny tak, aby na jego końcach wciąż występowały przeciwne znaki:

$$(\theta_{\mathrm{L}}^{(k+1)},\, \theta_{\mathrm{U}}^{(k+1)}) \;=\; \begin{cases} \big(\theta_{\mathrm{L}}^{(k)},\, s^{(k)}\big), & \text{jeśli } f\big(\theta_{\mathrm{L}}^{(k)}\big)\, f\big(s^{(k)}\big) \le 0, \\ \big(s^{(k)},\, \theta_{\mathrm{U}}^{(k)}\big), & \text{w przeciwnym razie.} \end{cases} \tag{6.34}$$

Iteracja kończy się, gdy spełniony jest przynajmniej jeden z warunków:

$$\theta_{\mathrm{U}}^{(k)} - \theta_{\mathrm{L}}^{(k)} \;\le\; \tau_x, \tag{6.35}$$

$$\min_{\vartheta \in \{\theta_{\mathrm{L}}^{(k)},\, s^{(k)},\, \theta_{\mathrm{U}}^{(k)}\}} |f(\vartheta)| \;\le\; \tau_f, \tag{6.36}$$

gdzie

$$\tau_x > 0, \qquad \tau_f > 0 \tag{6.37}$$

Parametry τ_x oraz τ_f określają kryteria zakończenia iteracji i są ustalane przed uruchomieniem algorytmu. Parametr τ_x oznacza dopuszczalną dokładność w przestrzeni argumentów, czyli maksymalną szerokość przedziału parametrów kopuli, przy której algorytm uznaje, że rozwiązanie zostało wystarczająco dokładnie określone. Natomiast, τ_f stanowi dopuszczalną dokładność w przestrzeni wartości funkcji celu, określającą, jak blisko zera musi znajdować się wartość funkcji $f(\theta)$, aby uznać, że osiągnięto wymaganą zgodność między uzyskaną a zadaną korelacją. W praktyce oba te parametry kontrolują moment zatrzymania obliczeń. Iteracje kończą się, gdy dalsze zmniejszanie przedziału parametrów kopuli lub wartości funkcji celu nie prowadzi już do istotnego zwiększenia dokładności rozwiązania.

Jako estymator parametru $\hat{\theta}$ przyjmuje się punkt z najmniejszą wartością $|f|$ w ostatniej iteracji:

$$\hat{\theta} \;=\; \arg \min_{\vartheta \in \{\theta_{\mathrm{L}}^{(k)},\, s^{(k)},\, \theta_{\mathrm{U}}^{(k)}\}} |f(\vartheta)|. \tag{6.38}$$

6.4 Analiza wpływu zależności ogonowych na wartość VaR przy wykorzystaniu zniekształconych mieszanek kopul i symulowanego wyżarzania

Jak wcześniej zasygnalizowano w niniejszym opracowaniu, kopule służą do opisu struktury zależności między zmiennymi losowymi. Umożliwiają one modelowanie zależności pomiędzy rozkładami brzegowymi niezależnie od ich kształtu, dzięki czemu mogą być stosowane

w szerokim zakresie problemów dotyczących analizy ryzyka. W praktyce jednak szczególne trudności pojawiają się w opisie zależności w ogonach rozkładów, gdzie liczba obserwacji jest niewielka, a zależności pomiędzy zmiennymi stają się niestabilne. W takich obszarach standardowe kopule, często nie są w stanie właściwie odwzorować zależności występujących w sytuacjach skrajnych, w ogonach rozkładów. Do modelowania zależności zarówno w części centralnej, gdzie dane są liczne i stabilne, jak i w ogonach rozkładów, gdzie zależności są bardziej zmienne i trudniejsze do uchwycenia, można wykorzystać zniekształconą mieszankę kopul (ang. Distorted Mix Copula – DMC) zaproponowaną przez Li, Yuen, and Yang (2014). Łącząc model DMC z algorytmem symulowanego wyżarzania, możliwe jest wyznaczenie takiej struktury zależności w ogonach rozkładów, która, przy założeniu znanej struktury w części centralnej, prowadzi do maksymalnej wartości VaR. Pozwala to na identyfikację najbardziej niekorzystnej struktury zależności między ryzykami w obszarach odpowiadających sytuacjom skrajnym. Z punktu widzenia ryzyka modelu ma to kluczowe znaczenie, ponieważ umożliwia ilościową ocenę wpływu niepewności dotyczącej zależności w ogonach rozkładów na wyniki modelu i miary ryzyka, takie jak wartość VaR. W obszarach ogonowych, odpowiadających sytuacjom skrajnym, współwystępowanie niekorzystnych zdarzeń pomiędzy różnymi rodzajami ryzyka może znacząco wzrosnąć, co prowadzi do kumulacji strat. Standardowe podejścia oparte na korelacji liniowej lub klasycznych kopulach często nie są w stanie właściwie uchwycić tego efektu, zakładając zbyt słabe zależności w ekstremach rozkładu. Niewłaściwe odwzorowanie tych zależności może prowadzić do istotnego zaniżenia wartości VaR, a w konsekwencji do błędnej oceny wymaganego kapitału oraz niedoszacowania ekspozycji na ryzyko w warunkach stresowych. Z punktu widzenia ryzyka modelu ma to kluczowe znaczenie, ponieważ w praktyce dane dotyczące zależności w obszarze centralnym rozkładów są stabilne, natomiast w obszarach ogonowych obserwacje są nieliczne i obarczone dużą niepewnością. Jest to częsty problem w analizie ryzyka, gdyż w tych skrajnych sytuacjach zależności pomiędzy ryzykami mogą się nasilać, prowadząc do kumulacji strat. Opracowany algorytm pozwala zatem zbadać, w jaki sposób te niepewne i potencjalnie silniejsze zależności w ogonach rozkładów wpływają na wartość VaR, stanowiąc istotny element oceny odporności i wiarygodności modelu w warunkach ekstremalnych. Algorytm przebiega następująco:

Algorytm 4

1. Określana jest kopula, opisująca zależności w centralnej części rozkładu oraz zbiór kopul, reprezentujących możliwe struktury zależności w ogonach.

2. Tworzona jest zniekształcona mieszanka kopul DMC, która umożliwia łączne modelowanie zależności w części centralnej i w ogonach rozkładu.

3. Wykorzystując metodę symulowanego wyżarzania,wyznaczana jest taka kombinacja wag (udziałów) poszczególnych kopul w ogonie rozkładu, by maksymalizować VaR.

4. Interacyjnie identyfikowana jest kombinacja kopul w ogonie, dająca najbardziej niekorzystną strukturę zależności, czyli taka, która prowadzi do największej wartości VaR.

Szczegóły algorytmu przedstawiono poniżej. Obszar centralny zależności, dla której dostępna jest większa pewność, reprezentowany jest przez jedną kopulę, natomiast obszary rozkładu charakteryzujące się większą niepewnością opisywane są za pomocą kolejnych kopul.

Kopula DMC definiowana jest jako kombinacja wypukła kopul składowych. Niech $m \geq 2$ oznacza liczbę kopul C_i, gdzie $i = 0, \ldots, m$ oraz $\alpha_i > 0$ oznaczają udziały poszczególnych kopul, spełniające warunek:

$$\sum_{i=0}^{m} \alpha_i = 1. \tag{6.39}$$

Funkcje zniekształcające definiowane są jako odwzorowania

$$D_{ij} : [0,1] \to [0,1], \qquad i = 0, \ldots, m, \ j = 1, \ldots, d, \tag{6.40}$$

które są ciągłe, rosnące i spełniają warunki brzegowe

$$D_{ij}(0) = 0, \qquad D_{ij}(1) = 1, \tag{6.41}$$

oraz relację

$$\sum_{i=0}^{m} \alpha_i D_{ij}(v) = v, \qquad v \in [0,1]. \tag{6.42}$$

Dzięki funkcjom zniekształcającym każda z kopul w DMC odpowiada za określony obszar rozkładu, zachowując właściwości teoretyczne kopuli. Innymi słowy, funkcje zniekształcające pełnią rolę operatorów modyfikujących kopule, określając sposób odwzorowania struktury zależności. W szczególności umożliwiają one regulowanie intensywności zależności ogonowej przy zachowaniu jednostajności rozkładów brzegowych, co oznacza, że otrzymana kombinacja kopul pozostaje rozkładem jednostajnym. Wówczas dowolna zniekształcona mieszanka kopul $C : [0,1]^d \to [0,1]$ przyjmuje postać:

$$C(u_1, \ldots, u_d) = \sum_{i=0}^{m} \alpha_i \, C_i\big(D_{i1}(u_1), \ldots, D_{id}(u_d)\big). \tag{6.43}$$

Aby wygenerować realizacje z kopuli DMC, stosujemy następujący algorytm:

1. Losowana jest kopula, z odpowiedniej części rozkładu i, zgodnie z rozkładem dyskretnym $P(Z = i) = \alpha_i$.

2. Losowany jest wektor $V = (V_1, \ldots, V_d)$ z kopuli C_i.

3. Wylosowane realizacje przekształcane są za pomocą funkcji zniekształcającej $U_j = D_{ij}^{-1}(V_j)$ dla każdej składowej j.

W przypadku, gdy znane są rozkłady brzegowe $L_1, L_2, \ldots, L_d$ oraz kopula C_0 opisująca zależność w części centralnej rozkładu, natomiast struktura zależności w ogonach nie jest znana, zależności w ogonach rozkładów można opisać za pomocą zbioru kopul:

$$\mathfrak{C} = \{C_1, C_2, \ldots, C_K\}, \tag{6.44}$$

gdzie K oznacza liczbę rozważanych kopul. Dla tak określonego przypadku kopula DMC przyjmuje postać:

$$\hat{C}^{\gamma^1,\ldots,\gamma^m}(u_1,\ldots,u_d) = \alpha_0 C_0(D_{01}(u_1),\ldots,D_{0d}(u_d)) + \sum_{i=1}^{m} \alpha_i \tilde{C}^{\gamma^i}(D_{i1}(u_1),\ldots,D_{id}(u_d)) \tag{6.45}$$

gdzie:

$$\widetilde{C}^{\gamma_i} = \sum_{j=1}^{K} \gamma_j C_j \tag{6.46}$$

Kopula DMC jest reprezentowana przez kombinację kopul ze zbioru $\mathfrak{C}$. Zadaniem jest wyznaczenie takich wartości parametrów γ, dla których uzyskiwana jest maksymalna wartość miary ryzyka Value at Risk. Poszukiwana jest zatem najbardziej niekorzystna kombinacja zależności w ogonach, prowadząca do najwyższej możliwej wartości VaR. Wektor wag γ przyjmuje wartości opisane jako sympleks $\triangle^{K-1}$. Zbiór wszystkich dopuszczalnych kombinacji kopul ogonowych można przedstawić w postaci:

$$\gamma \in \triangle^{K-1} = \left\{ \gamma = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_K \end{pmatrix} \in \mathbb{R}^K \;\middle|\; \sum_{j=1}^{K} \gamma_j = 1,\ \gamma_j \geq 0 \text{ dla wszystkich } j \right\}. \tag{6.47}$$

Pojawia się problem optymalizacyjny, którego celem jest wyznaczenie takiej kombinacji wag $\bar{\gamma}$, dla której wartość miary ryzyka VaR osiąga wartość maksymalną. Problem ten można sformułować w postaci:

$$\max_{\bar{\gamma}=(\gamma^1,\ldots,\gamma^m)\in(\triangle^{K-1})^m} VaR\left(L^{\bar{\gamma}}\right), \tag{6.48}$$

gdzie $L^{\bar{\gamma}}$ oznacza łączną stratę określoną jako

$$L^{\bar{\gamma}} = \Psi(L_1,\ldots,L_d), \tag{6.49}$$

dla wektora $(L_1,\ldots,L_d)$ o kopuli $\hat{C}^{\gamma^1,\ldots,\gamma^m}$ oraz zadanych rozkładach brzegowych L_i. W ten sposób zadanie sprowadza się do identyfikacji najbardziej niekorzystnej struktury zależności pomiędzy składnikami ryzyka, prowadzącej do maksymalizacji wartości VaR przy zachowaniu ustalonych rozkładów brzegowych.

Do wyznaczenia kombinacji kopul w ogonie, maksymalizujących wartość VaR, zastosowano metodę symulowanego wyżarzania (Simulated Annealing – SA). Jest to heurystyczny algorytm optymalizacyjny, przeszukujący przestrzeń dopuszczalnych rozwiązań, inspirowany procesem fizycznego wyżarzania metali (Kirkpatrick, Gelatt, & Vecchi, 1983; Černý, 1985). Działanie algorytmu opiera się na probabilistycznej akceptacji nowych rozwiązań. Kluczowym parametrem metody jest temperatura, odpowiedzialna za akceptację rozwiązań, które nie poprawiają wartości funkcji celu. Na początkowym etapie temperatura przyjmuje wysokie wartości, co umożliwia akceptację rozwiązań gorszych pod względem funkcji celu, pozwalając na uniknięcie zbieżności do minimum lokalnego. W kolejnych iteracjach, wraz z obniżaniem temperatury, prawdopodobieństwo akceptacji mniej korzystnych rozwiązań maleje, co powoduje koncentrację procesu poszukiwania w obszarze najlepszego dotychczas rozwiązania.

Algorytm przebiega następująco. W każdej iteracji k losowany jest wektor wag, odpowiadający udziałom poszczególnych kopul w ogonie poprzez dodanie zakłócenia losowego $\xi^{(k)} \sim \mathcal{N}(0, \sigma^2 I)$ do poprzedniego rozwiązania oraz rzutowanie na zbiór rozwiązań dopuszczalnych:

$$\theta^{\text{cand}} \;=\; \Pi_{\Theta}\Big(\theta^{(k-1)} + \xi^{(k)} \Big). \tag{6.50}$$

Funkcja Π_{Θ} rzutuje nowego kandydata θ^{cand} na $\Theta = (\triangle^{K-1})^m$. Dzięki niej wszystkie wylosowane elementy wektora θ^{cand} są nieujemne oraz sumują sie do jedności. Następnie, obliczamy wartość funkcji celu dla θ^{cand}, oraz dla poprzedniego rozwiązania: $f\big(\theta^{\text{cand}}\big)$ i $f\big(\theta^{(k-1)}\big)$, i wyznaczamy różnicę pomiędzy tymi wartościami:

$$\Delta f \;=\; f\big(\theta^{\text{cand}}\big) \;-\; f\big(\theta^{(k-1)}\big). \tag{6.51}$$

Kolejnym krokiem jest zdefiniowanie funkcji temperatury T_k, odpowiedzialnej za akceptowanie rozwiązań. Zgodnie z (Geman & Geman, 1984) i (Hájek, 1988), jeśli funkcja temperatury spełnia warunki:

$$\sum_{k=1}^{\infty} T_k = \infty \qquad \text{oraz} \qquad \sum_{k=1}^{\infty} \frac{T_k^2}{\ln k} < \infty. \tag{6.52}$$

to zapewnia teoretyczną zbieżność do optimum globalnego. Nowe rozwiązanie zostaje zaakceptowane, jeśli:

$$\theta^{(k)} \;=\; \begin{cases} \theta^{\text{cand}}, & \text{gdy } \Delta f \le 0, \\ \theta^{\text{cand}}, & \text{z prawdopodobieństwem } \exp\!\left(-\dfrac{\Delta f}{T_k}\right), \quad \Delta f > 0, \\ \theta^{(k-1)}, & \text{w przeciwnym przypadku.} \end{cases} \tag{6.53}$$

Dalej algorytm przechodzi do kolejnej iteracji, zwiększając indeks k i powtarzając opisane kroki.

6.5 Wpływ specyfikacji zależności na agregację ryzyka i SCR

W niniejszym podrozdziale zaprezentowano analizę wpływu specyfikacji struktury zależności na agregację ryzyka oraz wartość SCR na przykładzie dwóch modułów: aktuarialnego w ubezpieczeniach na życie (X_1) oraz aktuarialnego w ubezpieczeniach zdrowotnych (X_2). Przyjęto zgodnie z (P. Embrechts, Wang, & Wang, 2015), że zmienne ryzyka dla obu modułów miały rozkłady normalne o parametrach podanych w Tabeli 6.1. Wyznaczano wartość zagrożoną dla $Y = X_1 + X_2$ na poziomie bezpieczeństwa $\alpha = 0.995$ ($\text{VaR}_{0.995}(Y)$). Ponieważ $\mathbb{E}(Y) = 0$ ($\mathbb{E}(X_1) = 0$, $\mathbb{E}(X_2) = 0$), zachodziła równość $\text{VaR}_{0.995}(Y) = SCR$; w dalszym ciągu, dla uproszczenia zapisu, stosowano oznaczenie $\text{VaR}_{0.995}$. Analizę prowadzono z wykorzystaniem algorytmów przedstawionych w poprzednich podrozdziałach.

Moduł	Rozkład	Średnia	Odchylenie standardowe
Ryzyko aktuarialne w ubezpieczeniach na życie	Normalny	0	392
Ryzyko aktuarialne w ubezpieczeniach zdrowotnych	Normalny	0	248

Tabela 6.1: Parametry rozkładów dla poszczególnych segmentów.
Źródło: Opracowanie własne.

Pomiędzy rozkładami narzucono strukturę zależności odpowiadającą współczynnikowi korelacji liniowej $\rho = 0.25$. Celem analizy było wykazanie, że mimo normalności rozkładów brzegowych zagregowanych ryzyk rzeczywista struktura współzależności mogła odbiegać od liniowej, co w konsekwencji prowadziło do zróżnicowanych wartości SCR nawet przy ustalonej wartości ρ. Zagadnienie to pozostawało ściśle związane z oceną odporności FS na błędną specyfikację zależności oraz z badaniem wrażliwości efektu dywersyfikacji i miar ryzyka na przyjętą wartość korelacji.

Odwołując się do podrozdziału 4.4.2, przyjęto, że rozkłady brzegowe były znane (zob. tab. 6.1) oraz zdefiniowano zbiór $A_5 = \{a_1, a_2, a_3, a_4, a_5\}$, porządkujący dostępne informacje o strukturze zależności pomiędzy rozkładami. Każdy element zbioru odpowiadał innemu poziomowi wiedzy - od całkowitego braku informacji po coraz bardziej szczegółowe specyfikacje współzależności. Zestawienie tych poziomów przedstawiono w Tabeli 6.2.

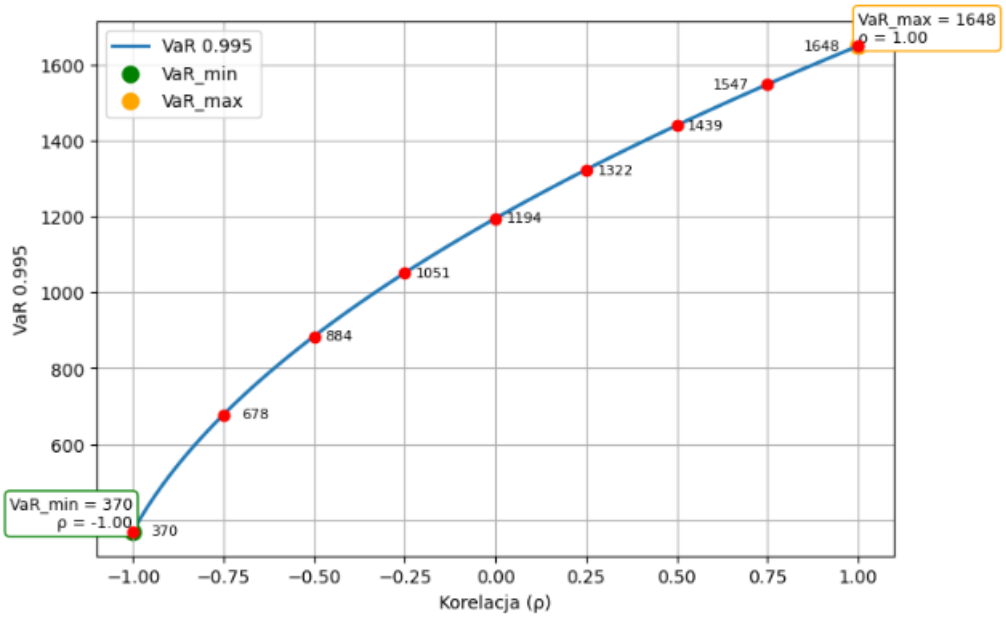
	Informacje o zależności	Metoda
a_1	Brak wiedzy na temat struktury zależności	*Algorytm ARA*
a_2	Nieznany współczynnik korelacji liniowej ρ, dwuwymiarowy rozkład normalny	*Algorytm 1*
a_3	Znany współczynnik korelacji liniowej ρ, nieznana kopula	*Algorytm 2*
a_4	Znany współczynnik korelacji liniowej ρ, kopula Archimedesa	*Algorytm 3*
a_5	Znana struktura zależności w części centralnej	*Algorytm 4*

Tabela 6.2: Informacje o strukturze zależności.
Źródło: Opracowanie własne.

Nieznany współczynnik korelacji liniowej, dwuwymiarowy rozkład normalny (Algorytm 1)

Przy założeniu wielowymiarowego rozkładu normalnego współczynnik korelacji liniowej ρ stanowi właściwe narzędzie do opisu struktury zależności. Z uwagi na niepewność co do wartości tego współczynnika możliwe było jedynie wyznaczenie dolnego i górnego ograniczenia $\text{VaR}_{0.995}$ dla zagregowanego ryzyka. Ograniczenia te oszacowano z wykorzystaniem Algorytmu 1, który umożliwiał również analizę wrażliwości $\text{VaR}_{0.995}$ na zmiany ρ wynikające m.in.

z błędu estymacji (np. ograniczona liczebność próby, niestacjonarność, zmiany reżimu) oraz z celowych przesunięć stosowanych w analizie scenariuszowej i testach warunków skrajnych. Wyniki zaprezentowano na Rysunku 6.1.



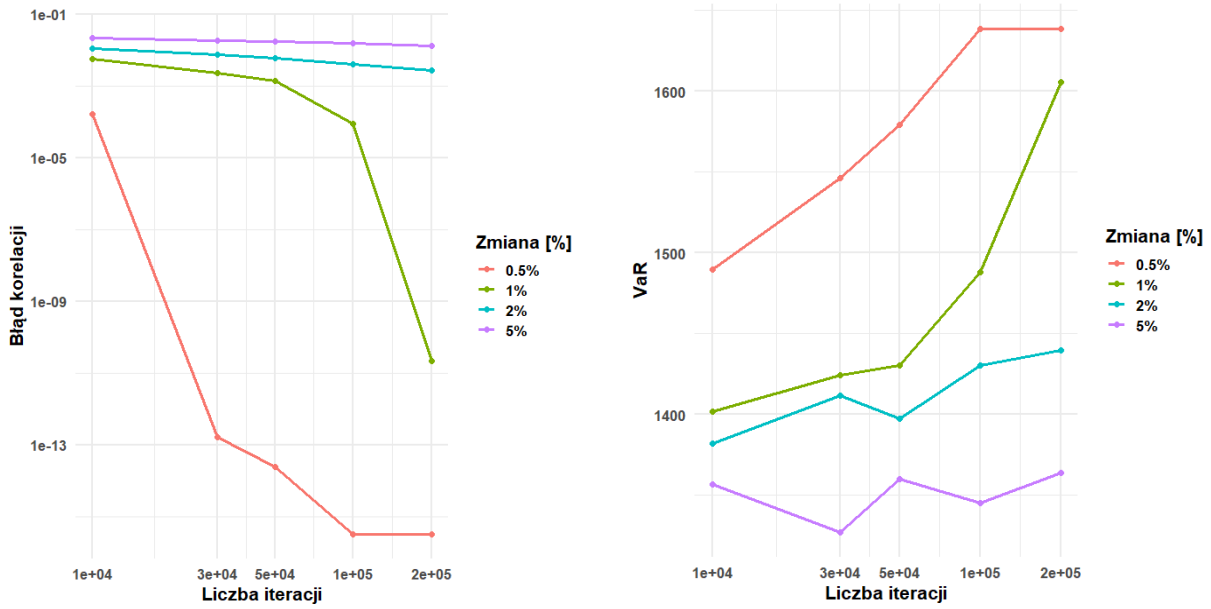
Rysunek 6.1: Zależność agregowanego $VaR(0.995)$ od współczynnika korelacji.
Źródło: Opracowanie własne.

Uzyskane wyniki wskazywały, że przy założeniu wielowymiarowego rozkładu normalnego niepewność estymacji ρ istotnie wpływała na $\text{VaR}_{0.995}$ dla zagregowanego ryzyka dwóch modułów (aktuarialnego w ubezpieczeniach na życie oraz aktuarialnego w ubezpieczeniach zdrowotnych). Wartość zagrożona rosła monotonicznie wraz ze wzrostem ρ i wykazywała wyraźną wrażliwość na jego zmiany, co potwierdzało małą odporność na błędy w ρ. W konsekwencji zastosowania metody wariancji–kowariancji w formule standardowej uwidaczniało się ryzyko modelu: występowało zarówno niedoszacowanie, jak i przeszacowanie SCR, co może skutkować ryzykiem naruszenia wymogów kapitałowych bądź nadmiernym zaangażowaniem kapitału. Może to także utrudniać planowanie kapitałowe i projektowanie reasekuracji, co w konsekwencji uzasadnia prowadzenie rozszerzonych testów wrażliwości oraz systematyczne dokumentowanie własnej oceny ryzyka i wypłacalności (Own Risk and Solvency Assessment, ORSA).

Znany współczynnik korelacji liniowej ρ, nieznana kopula (Algorytm 2)

Korelacja liniowa dobrze opisuje zależności pomiędzy zmiennymi w przypadku stosowania wielowymiarowego rozkładu normalnego. Jednak w ogólnym przypadku te same wartości współczynników korelacji liniowej mogą odpowiadać różnym strukturom zależności. Oznacza to, że sama macierz korelacji nie determinuje jednoznacznie pełnej struktury zależności pomiędzy zmiennymi. W podrozdziale 6.2 został opracowany algorytm pozwalający na wyznaczenie takiej struktury zależności, która dla zadanego współczynnika korelacji liniowej $\rho = 0.25$ prowadzi do największej wartości $\text{VaR}_{0.995}$. Odnosi się to do założenia a_3. Na Ry-

sunku 6.2 przedstawiono zależność pomiędzy liczbą iteracji algorytmu a jakością odwzorowania macierzy korelacji oraz odpowiadającą jej wartością $\text{VaR}_{0.995}$. Na lewym wykresie pokazano minimalny błąd między oczekiwaną korelacją $\rho = 0.25$ a korelacjami otrzymanymi za pomocą algorytmu w zależności od liczby iteracji i poziomu zmian. Przy czym zmiana oznacza procent modyfikowanych elementów w każdej kolumnie w trakcie przekształceń. Krzywe wskazują, że dla 0.5% modyfikowanych obserwacji błąd korelacji maleje wraz z liczbą iteracji, natomiast w pozostałych przypadkach (1%, 2%, 5%) błąd stabilizuje się na poziomie 10^{-2}–10^{-3}, co oznacza brak zgodności z wartością docelową 0.25. Na prawym wykresie zaprezentowano wartości $\text{VaR}_{0.995}$ uzyskane dla składowych frontu Pareto charakteryzujących się minimalnym błędem dopasowania korelacji. Można zauważyć, że wraz ze wzrostem iteracji wartość $\text{VaR}_{0.995}$ rośnie i stabilizuje się na najwyższym poziomie przy 0.5% zmieniających się obserwacjach w danej kolumnie, podczas gdy dla 1%, 2%, 3% wartości VaR pozostają istotnie niższe.



(a) Błąd dopasowania korelacji

(b) Wartości VaR

Rysunek 6.2: Wartości błądu dopasowania korelacji oraz wartości VaR w kolejnych iteracjach. Źródło: Opracowanie własne.

Uzyskane wyniki jednoznacznie potwierdzają, że najlepszy kompromis pomiędzy maksymalizacją wartości $\text{VaR}_{0.995}$ a zgodnością z zadaną macierzą korelacji osiągany jest przy poziomie modyfikacji elementów równym $0,5\%$ w każdej kolumnie oraz przy liczbie iteracji nie mniejszej niż 10^5. W Tabeli 6.3 przedstawiono front Pareto obejmujący rozwiązania, dla których błąd macierzy korelacji był mniejszy niż 10^{-10}.

$\mathrm{VaR}_{0.995}$	**Błąd przybliżenia współczynnika korelacji**
1557.915	1.017252×10^{-14}
1568.437	3.020109×10^{-14}
1570.490	3.686309×10^{-14}
1592.357	3.104877×10^{-13}
1632.457	9.696069×10^{-14}
1634.256	3.378994×10^{-12}
1634.573	2.093017×10^{-11}
1635.805	3.035560×10^{-10}

Tabela 6.3: Wartości frontu Pareto.
Źródło: Opracowanie własne.

Spośród uzyskanych wyników wybrano rozwiązanie o najlepszym dopasowaniu do zadanej wartości współczynnika korelacji 0.25, dla którego wartość VaR wyniosła 1557.915, przy błędzie rzędu 1.02×10^{-14}. Uzyskane rezultaty potwierdzają, że nawet przy identycznych współczynnikach korelacji liniowej między agregowanymi rodzajami ryzyka VaR pozostaje silnie uzależniona od przyjętej postaci struktury zależności.

Znany współczynnik korelacji liniowej ρ, kopula Archimedesa (Algorytm 3)

W założeniu a_3 przyjęto, że znane są rozkłady brzegowe oraz wartość współczynnika korelacji liniowej ρ, natomiast brak jest informacji o strukturze zależności pomiędzy zmiennymi (może ona być zupełnie dowolna). Założenie a_4 stanowi zawężenie przypadku a_3, ograniczając strukturę zależności do rodziny kopul Archimedesa. Wykorzystując algorytm opisany w podrozdziale 6.3, wyznaczono parametr zależności θ, odpowiadający zadanej wartości korelacji liniowej $\rho = 0,25$, dla następujących kopul Archimedesa: Clayton, Frank, Gumbel, Joe Pokazano tym samym, że dla takiej samej wartości korelacji istnieje wiele kopul z rodziny Archimedesa. Na tabeli poniżej (tab. 6.4) przedstawiono uzyskane wartości $\hat{\theta}$ oraz odpowiadające im korelacje empiryczne ρ_{emp}, wraz z błędami dopasowania do przyjętej wartości współczynnika korelacji $\rho = 0.25$. Wartości błędów potwierdzają wysoką dokładność przybliżenia zadanej korelacji.

Kopula	$\hat{\theta}$	$\rho_{\mathbf{emp}}$	**Błąd dopasowania**
Clayton	0.3719	0.2500002	2.31×10^{-7}
Frank	1.6266	0.2500001	6.60×10^{-8}
Gumbel	1.1850	0.2500001	9.73×10^{-8}
Joe	1.3173	0.2499994	-6.02×10^{-7}

Tabela 6.4: Parametry zależności kopul Archimedesa dla ustalonej korelacji $\rho = 0.25$.
Źródło: Opracowanie własne.

Otrzymane wartośći VaR dla wyznaczonych kopul przedstawiono poniżej (tab. 6.5).

Kopula	$\mathrm{VaR}_{0.995}$
Clayton	1234.91
Frank	1275.12
Gumbel	1436.42
Joe	1501.83

Tabela 6.5: Wartości VaR dla kopul Archimedesa dla ustalonej korelacji $\rho = 0.25$.
Źródło: Opracowanie własne.

Przykład ten pokazuje, że dla tej samej wartości współczynnika korelacji liniowej można uzyskać różne kopule Archimedesa. W konsekwencji zauważamy wrażliwość VaR na przyjętą strukturę zależności, nawet gdy współczynniki korelacji liniowej pomiędzy agregowanymi rodzajami ryzyka pozostają identyczne.

Znana struktura zależności w części centralnej (Algorytm 4)

W standardowym podejściu Solvency II przyjmuje się, że zależności pomiędzy ryzykami można opisać za pomocą korelacji liniowej, czyli kopuli Gaussa. Kopula Gaussa nie uwzględnia jednak zależności w ogonach rozkładów, gdzie znajduje się informacja kluczowa dla wyznaczenia $\mathrm{VaR}_{0.995}$. W analizowanym przykładzie przyjęto, że w części centralnej zależności odpowiadają kopuli Gaussa, natomiast w ogonach występuje niepewność co do ich struktury. Podany przykład odnosi się do założenia a_5. Celem analizy jest zbadanie wpływu tej niepewności w ogonach na wartość $\mathrm{VaR}_{0.995}$, z wykorzystaniem algorytmu opisanego w podrozdziale 6.4.

W celu połączenia obszaru centralnego i obszaru w ogonie wprowadzono spójną strukturę kopuli z wykorzystaniem następujących funkcji zniekształcających:

$$D_2(x;\alpha_2) = \frac{x - \alpha_2 x^2}{\alpha_2 + (1-2\alpha_2)x}, \qquad D_3(x;\alpha_3) = \frac{\alpha_3 x^2}{\alpha_3 + (1-2\alpha_3)(1-x)}, \tag{6.54}$$

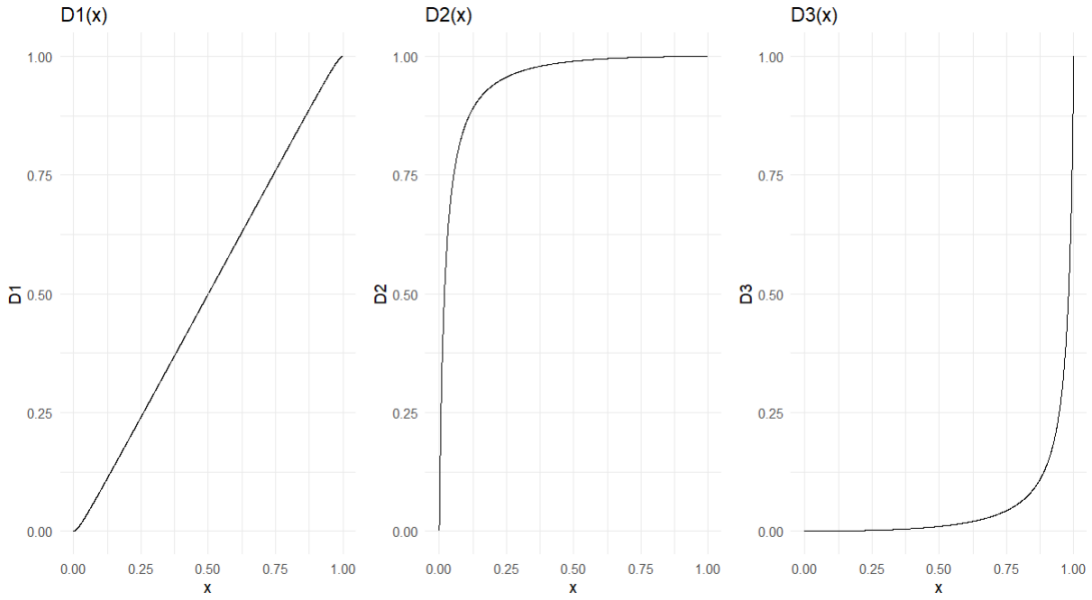
oraz

$$D_1(x;\alpha_1,\alpha_2,\alpha_3) = \frac{x - \alpha_2 D_2(x;\alpha_2) - \alpha_3 D_3(x;\alpha_3)}{\alpha_1}. \tag{6.55}$$

Parametry $\alpha_1, \alpha_2, \alpha_3$ są nieujemne i spełniają warunek normalizacji:

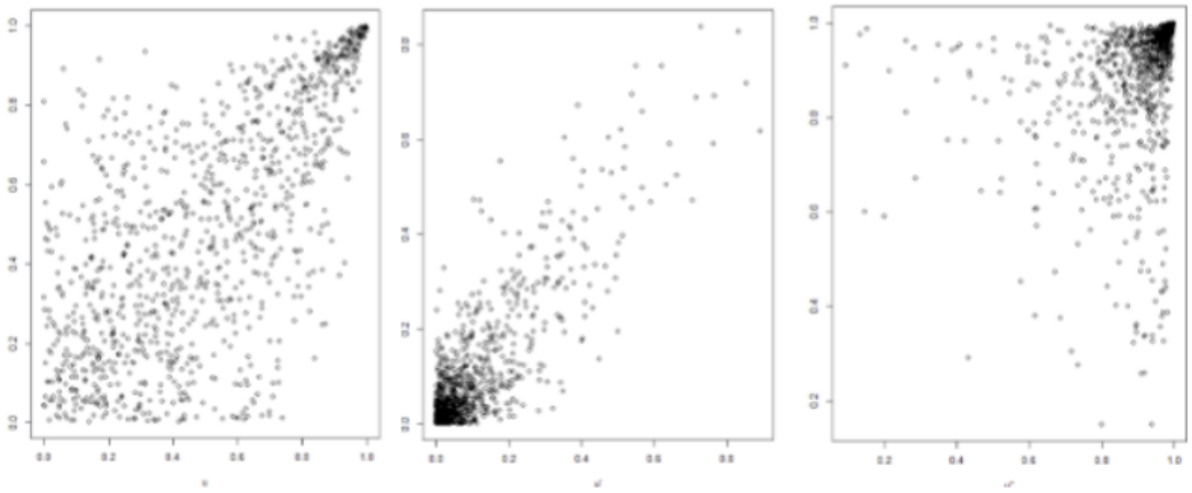
$$\alpha_1 + \alpha_2 + \alpha_3 = 1. \tag{6.56}$$

Oznacza to, że α_1 kontroluje udział obszaru centralnego rozkładu, natomiast α_2 i α_3 określają udział dolnego i górnego ogona. Na Rysunku 6.3 przedstawiono kształt funkcji zniekształcających $D_1(x)$, $D_2(x)$ oraz $D_3(x)$ dla $\alpha = 0.02$. Wartość parametru $\alpha = 0.02$ oznacza, że niepewność w strukturze zależności obejmuje po 2% obserwacji w lewym i prawym ogonie rozkładu, natomiast pozostałe 96% odpowiada obszarowi centralnemu, w którym zależności są zgodne z kopulą Gaussa. Funkcja $D_1(x)$ przebiega niemal liniowo w całym przedziale $[0, 1]$, co oznacza, że odpowiada za obszar centralny rozkładu i zachowuje proporcjonalne odwzorowanie prawdopodobieństwa. Funkcja $D_2(x)$ rośnie bardzo szybko w pobliżu zera, a następnie spłaszcza się w okolicach wartości jeden, co prowadzi do koncentracji masy prawdopodobieństwa w dolnym ogonie i pozwala na opisanie zależności występujących przy skrajnych stratach. Odwrotne zachowanie obserwuje się dla funkcji $D_3(x)$, która przez większą część przedziału pozostaje bliska zeru, a gwałtownie rośnie dopiero przy wartościach bliskich jedności. Oznacza to, że jej wpływ koncentruje się w górnym ogonie, a więc w obszarze skrajnych strat. Zastosowanie tych trzech funkcji umożliwia rozdzielenie wpływu kopul na obszar centralny oraz na oba ogony rozkładu.



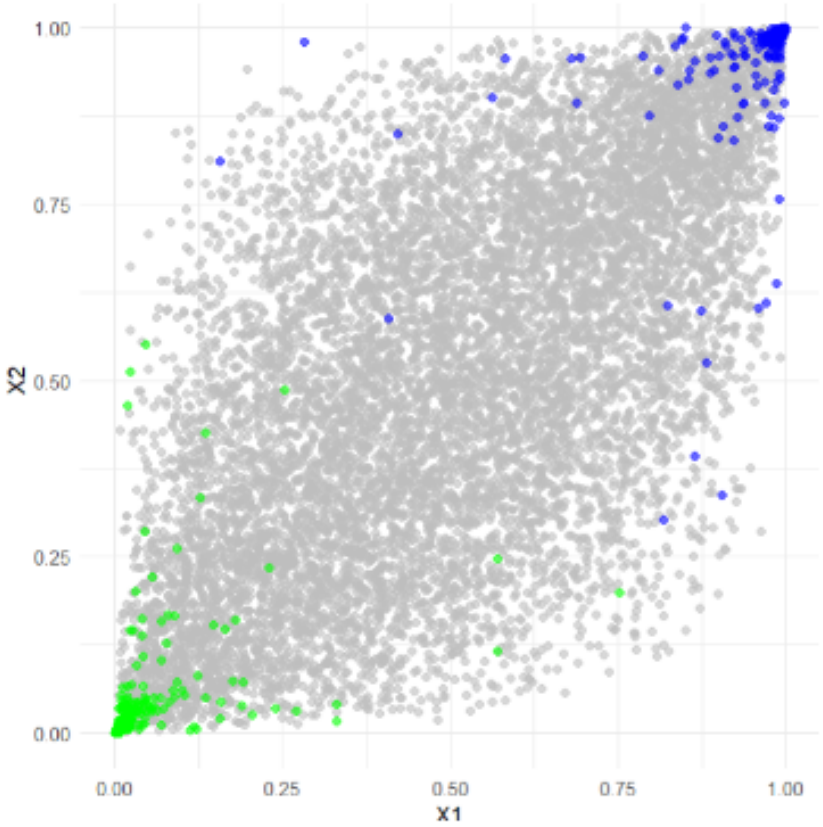
Rysunek 6.3: Przykładowe funkcje sydorsji.
Źródło: Opracowanie własne.

Na Rysunku 6.4 przedstawiono wykorzystanie funkcji zniekształcających. Na lewym wykresie znajdują się realizacje wygenerowane z kopuli Joe. Na środkowym panelu pokazano te same realizacje po transformacji funkcją D_2, która koncentruje masę prawdopodobieństwa w dolnym ogonie. Z kolei, prawy panel przedstawia wyniki po zastosowaniu funkcji D_3, co odpowiada przekształceniu zależności do górnego ogona rozkładu.



Rysunek 6.4: Próbka wygenerowana z kopuli Joe i przekształcona przez funkcje D_1, D_2 oraz D_3.
Źródło: Opracowanie własne.

Z kolei, na Rysunku 6.5 przedstawiono realizacje kopuli DMM, w której obszar centralny rozkładu opisany został kopulą Gaussa (szare punkty), natomiast punkty w obszarze lewego (zielone punkty) i prawego ogona (niebieskie punkty) zostały wygenerowane na podstawie kopuli Joe oraz przekształcone przy użyciu funkcji zniekształcających.



Rysunek 6.5: Próbka wygenerowana z kopuli Joe i przekształcona przez funkcje D_1, D_2 oraz D_3.
Źródło: Opracowanie własne.

Pojawia się zatem naturalne pytanie, dla jakich kopul w ogonach wartość $\text{VaR}_{0.995}$ osiąga największy poziom. W dalszej części zastosowano algorytm symulowanego wyżarzania (SA), za pomocą którego wyznaczono kombinację kopul Franka, Claytona, Gumbela oraz Joe w ogonach, prowadzącą do maksymalizacji wartości $\text{VaR}_{0.995}$. W niniejszym przykładzie założono, że dane zostały podzielone na trzy obszary: obszar centralny, lewy ogon oraz prawy ogon rozkładu. Podział ten opiera się na ustalonych prawdopodobieństwach α_1, α_2 oraz α_3, gdzie α_1 odpowiada obszarowi centralnemu, natomiast α_2 i α_3 – odpowiednio obszarowi dolnego i górnego ogona. Przyjęto wartości $\alpha_2 = \alpha_3 = 0.02$ i wykorzystując algorytmu symulowanego wyżarzania, poszukiwano wag maksymalizujących wartość miary $\text{VaR}_{0.995}$. Dokładniej, celem było ustalenie wag γ_1^{dolny}, γ_2^{dolny}, γ_3^{dolny}, γ_4^{dolny} oraz $\gamma_1^{\text{górny}}$, $\gamma_2^{\text{górny}}$, $\gamma_3^{\text{górny}}$, $\gamma_4^{\text{górny}}$, które odpowiadają odpowiednio za udział kopul Franka, Claytona, Gumbela oraz Joe w dolnym ogonie i górnym ogonie.

Algorytm SA w kolejnych iteracjach modyfikował wektory γ i oceniał odpowiadającą im wartość $\text{VaR}_{0.995}$. W ten sposób wyznaczono kombinację wag maksymalizującą $\text{VaR}_{0.995}$. Wagi te odpowiadały scenariuszowi najbardziej niekorzystnej zależności w ogonach. W wyniku procedury otrzymano $\text{VaR}_{0.995} = 1632$, podczas gdy wartość oszacowana zgodnie z formułą standardową Solvency II wynosiła 1322. Różnica 310 (ok. 23.4‰) ujawniła znaczną wrażliwość wyniku na specyfikację zależności ogonowej i potencjalne niedoszacowanie ryzyka przez metodę wariancji–kowariancji w formule standardowej. Otrzymany scenariusz „worst–case" wskazywał na istotne ryzyko modelu: niewielkie zmiany w parametryzacji współzależności mogły prowadzić do materialnie wyższych wartości VaR/SCR . W ujęciu praktycznym uzasadnia to stosowanie rozszerzonych testów wrażliwości i strestestów dla parametrów zależności oraz stosowanie konserwatywnych buforów kapitałowych oraz weryfikację dopasowania programów reasekuracyjnych.

6.6 Wpływ skrajnych scenariuszy współzależności na VaR: implikacje dla ryzyka modelu FS

Wyniki przedstawione w poprzednim podrozdziale wskazywały, że algorytmy zaprezentowane w podrozdziałach 6.1–6.4 umożliwiają ocenę wpływu struktury zależności między rodzajami ryzyka na rozkład zagregowanej zmiennej Y oraz na wartość $\text{VaR}_{0.995}$. W niniejszym podrozdziale wyniki te uzupełniono o górne oszacowanie $\text{VaR}_{0.995}$ w warunkach braku informacji o strukturze zależności (założenie a_1), a następnie wykorzystano je do analizy wrażliwości $\text{VaR}_{0.995}$ na przyjętą strukturę zależności oraz do oceny ryzyka stosowania metody wariancji–kowariancji w formule standardowej Solvency II przy wyznaczaniu SCR (dalej: ryzyka FS).

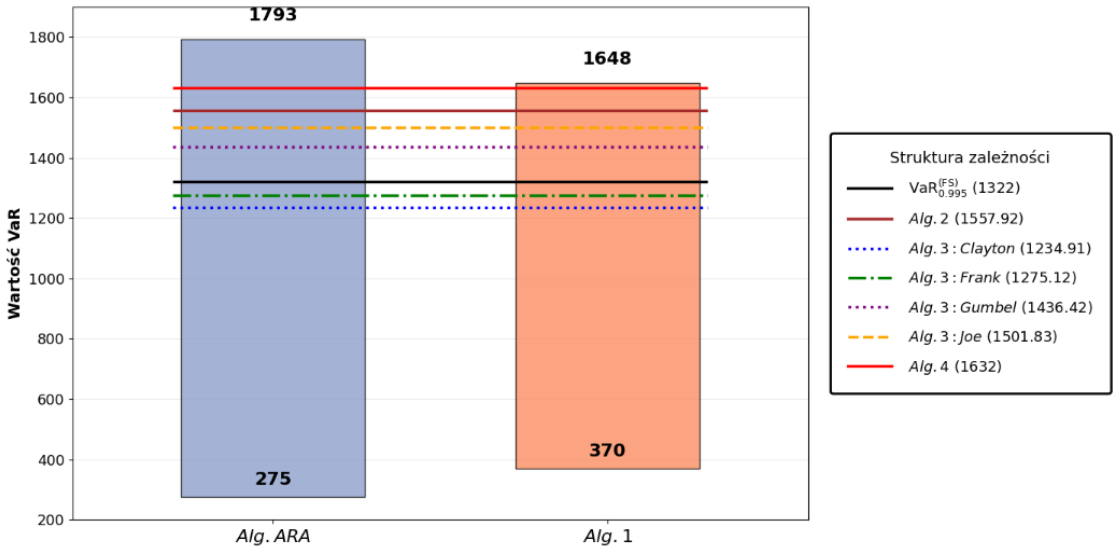
Górne oszacowania $\text{VaR}_{0.995}$ przy założonych informacjach o strukturze zależności zestawiono w Tabeli 6.6 oraz zilustrowano na Rysunku 6.6. Na Rysunku 6.6 zaznaczono również wartość odniesienia wyznaczoną metodą wariancji–kowariancji stosowaną w formule standardowej Solvency II, tj. $\text{VaR}_{0.995}^{(\text{FS})} = 1322$.

W warunkach braku informacji o strukturze zależności (znane jedynie rozkłady brzegowe; założenie a_1) górne ograniczenie $\text{VaR}_{0.995}$ wyznaczono algorytmem ARA (podrozdział 2.6.1). Następnie, zgodnie z założeniem a_2, przyjęto liniowy charakter współzależności i z wykorzystaniem algorytmu z podrozdziału 6.1 określono przedział wartości $\text{VaR}_{0.995}$ odpowiadających wszystkim dopuszczalnym współczynnikom korelacji liniowej. W założeniu a_3 posłużono się informacją o wartości $\rho = 0.25$ jako jedynym parametrem opisującym zależność. W opar-

ciu o algorytm z podrozdziału 6.2 wyznaczono strukturę współzależności maksymalizującą $\mathrm{VaR}_{0.995}$ przy ustalonym ρ. Kolejno, dla założenia a_4, przyjęto że znana jest kopula z rodziny Archimedesa opisująca strukturę zależności; stosując algorytm z podrozdziału 6.3 oszacowano jej parametry, a następnie obliczono wartości $\mathrm{VaR}_{0.995}$. Wreszcie, w założeniu a_5 przyjęto, że struktura zależności jest znana w części centralnej rozkładu, natomiast nieznana w ogonach; wykorzystując algorytm z podrozdziału 6.4, uzyskano odpowiadające temu oszacowania $\mathrm{VaR}_{0.995}$.

	Informacje o zależności	**Metoda**	**Górne oszacowanie**
a_1	Brak wiedzy na temat struktury zależności	*Alg. ARA*	1793
a_2	Nieznany współczynnik korelacji liniowej ρ, dwuwymiarowy rozkład normalny	*Alg. 1*	1648
a_3	Znany współczynnik korelacji liniowej ρ, nieznana kopula	*Alg. 2*	1557.92
a_4	Znany współczynnik korelacji liniowej ρ:	*Alg. 3*	
	dla kopuli Archimedesa:		
	Claytona		1234.92
	Franka		1275.12
	Gumbela		1436.42
	Joe		1501.83
a_5	Znana struktura zależności w części centralnej	*Alg. 4*	1632

Tabela 6.6: Górne oszacowania $\mathrm{VaR}_{0.995}$ przy zakładanych informacjach o strukturze zależności.
Źródło: Opracowanie własne.



Rysunek 6.6: Górne oszacowania $\mathrm{VaR}_{0.995}$ przy zakładanych informacja o strukturze zależności, $\mathrm{VaR}^{(\mathrm{FS})}_{0.995}$ oraz możliwe wartości $\mathrm{VaR}_{0.995}$ dla założeń a_1 - a_5.
Źródło: Opracowanie własne.

Zestawione wyniki powyżej (Tab. 6.6 i Rys. 6.6) ilustrują wpływ kolejnych założeń dotyczących struktury zależności pomiędzy X_1 i X_2 na wartość górnego oszacowania $\mathrm{VaR}_{0.995}$. Na wykresie dodatkowo wskazano $\mathrm{VaR}_{0.995}^{(\mathrm{FS})}$ oraz możliwe wartości $\mathrm{VaR}_{0.995}$ dla założeń a_1 i a_2. W scenariuszu a_1, w którym brakuje informacji o strukturze zależności, otrzymano najwyższą wartość górnego oszacowania $\mathrm{VaR}_{0.995}^{(\mathrm{FS})}$ równą 1793 (dalej $\overline{VaR}_{0.995} = 1793$), co odzwierciedla największe ryzyko modelu. Wraz z wprowadzaniem dodatkowych założeń, najpierw dotyczących zależności liniowej (a_2), następnie znanej korelacji (a_3) oraz określonego typu kopuli Archimedesa (a_4), uzyskiwane wartości $\mathrm{VaR}_{0.995}$ stopniowo maleją. Redukuje to ryzyko modelu poprzez uwzględnienie coraz bardziej szczegółowej informacji o strukturze zależności. Ostatnie założenie (a_5), w którym znana jest struktura zależności w części centralnej rozkładu, a nieznana struktura zależności w ogonach rozkładów, prowadzi do pośredniego oszacowania (1632), co wskazuje na istotny wpływ informacji o strukturze zależności w ogonach rozkładów na $\mathrm{VaR}_{0.995}$.

W celu ilościowej oceny ryzyka FS zastosowano zmodyfikowaną w następujacy sposób wersję miary AM (4.7):

$$AM = \frac{\mathrm{VaR}_{0.995}^{(\mathrm{a_i})} - \mathrm{VaR}_{0.995}^{(\mathrm{FS})}}{\mathrm{VaR}_{0.995}^{(\mathrm{FS})}} \times 100\%, \tag{6.57}$$

gdzie $\mathrm{VaR}_{0.995}^{(\mathrm{a_i})}$ oznacza górne oszacowanie wartości zagrożonej, przy założeniu że są dostępne informacje a_i, natomiast $\mathrm{VaR}_{0.995}^{(\mathrm{FS})}$ odpowiada wartości referencyjnej otrzymanej na podstawie FS. Zastosowano również przekształconą miarę ryzyka modelu RM (4.8), która różni się od RM tym, iż odchylenie względem wartości referencyjnej odnosi się do całkowitego przedziału niepewności:

$$RM = \frac{\mathrm{VaR}_{0.995}^{(\mathrm{a_i})} - \mathrm{VaR}_{0.995}^{(\mathrm{FS})}}{\overline{VaR}_{0.995} - \underline{VaR}_{0.995}}, \tag{6.58}$$

gdzie $\overline{VaR}_{0.995}$ oraz $\underline{VaR}_{0.995}$ oznaczają odpowiednio górne i dolne oszacowanie $\mathrm{VaR}_{0.995}$ otrzymane przy braku jakichkolwiek informacji o strukturze zależności (założenie a_1). Miara RM określa, jak duża część potencjalnego przedziału niepewności realizowana jest przez dane odchylenie względem FS. Wartości dodatnie wskazują na przypadki, w których FS oszacowała ryzyko poniżej wartości uzyskanej w modelu alternatywnym, natomiast wartości ujemne pokazują sytuację odwrotną.

Uzyskane wartości tych miar przedstawiono poniżej, w Tabeli 6.7.

	Informacje o zależności	Metoda	Górne oszac.	AM	RM
a_1	Brak wiedzy na temat struktury zależności	*Alg. ARA*	1793	35.63%	31.04%
a_2	Nieznany współczynnik korelacji liniowej ρ, dwuwymiarowy rozkład normalny	*Alg. 1*	1648	24.66%	21.48%
a_3	Znany współczynnik korelacji liniowej ρ, nieznana kopula	*Alg. 2*	1557.92	17.85%	15.5%
a_4	Znany współczynnik korelacji:	*Alg. 3*			
	liniowej ρ dla kopuli Archimedesa:				
	Claytona		1234.91	-6.59%	-5.7%
	Franka		1275.12	8.66%	7.5%
	Gumbela		1436.42	13.60%	11.8%
	Joe		1501.83	-3.55%	-3.1%
a_5	Znana struktura zależności w części centralnej	*Alg. 4*	1632	23.45%	20.4%

Tabela 6.7: Miary ryzyka modelu AM i RM w zależności od posiadanej informacji strukturze zależności. Źródło: Opracowanie własne.

Wartości dodatnie miar AM i RM wskazują, że $\mathrm{VaR}_{0.995}^{(a_i)}$ przewyższa wartość referencyjną $\mathrm{VaR}_{0.995}^{(\mathrm{FS})}$, co oznacza, że FS zaniżała ryzyko wobec modelu alternatywnego; wartości ujemne oznaczają sytuację odwrotną, w której FS zawyżała ryzyko. W Tabeli 6.7 obserwuje się wyraźny trend: wraz z doprecyzowaniem informacji o zależności od a_1 do a_3 miary maleją (z 35.63% i 31.04% do 17.85% i 15.5%), co świadczy o zawężaniu pasma niepewności i spadku ryzyka modelu; dodatni znak utrzymujący się dla a_1–a_3 potwierdza, że w tych scenariuszach FS zaniżała poziom ryzyka. Dla a_4 istotne znaczenie ma wybór rodziny kopuli przy stałej korelacji liniowej: kopule Gumbela i Joe prowadzą do dodatnich AM i RM, natomiast kopule Claytona i Franka prowadzą do wartości ujemnych, co ujawnia wrażliwość FS na specyfikację ogonową przy tej samej wartości ρ. W przypadku a_5, gdy zależność w części centralnej jest znana, a ogony pozostają nieokreślone, wartości $AM \approx 23.45\%$ i $RM \approx 20.4\%$ rosną względem a_3, co wskazuje, że niepewność ogonowa stanowi kluczowe źródło ryzyka modelu i ograniczonej odporności FS. Normalizacja RM przez pełny przedział niepewności $[\underline{VaR}_{0.995}, \overline{VaR}_{0.995}]$ pozwala ocenić względną skalę odchylenia: przykładowo $RM = 31.04\%$ przy a_1 oznacza, że błąd względem FS realizuje około jednej trzeciej dostępnego przedziału, podczas gdy wartości rzędu 7–12% obserwowane dla niektórych kopul Archimedesa sygnalizują mniejszą, choć nadal istotną, część tego pasma. Całościowo wyniki potwierdzają ograniczoną odporność FS na błędną lub niekompletną specyfikację zależności: przy braku informacji lub nieokreślonych ogonach ryzyko bywa zaniżane, a przy wybranych rodzinach kopuli może być także zawyżane. Z perspektywy zarządczej uzasadnia to prowadzenie analiz wrażliwości na klasę zależności, raportowanie pasm wartości VaR, włączanie ustaleń do ORSA oraz weryfikację i dostrajanie programów reasekuracyjnych.

W przeprowadzonej powyżej ocenie ryzyka modelu przyjmuje się, że każde założenie jest równie istotne. Podejście opisane w podrozdziale 4.4.2 rozszerza tę koncepcję, wprowadzając

czynniki z_j, które przypisują każdemu z założeń określony stopień istotności w modelu. Zatem dodajemy kolejne informacje na temat struktury zależności, określając ich istotność z_j: od znajomości rozkładów brzegowych i braku informacji o strukturze zależności (a_1), przez opis zależności za pomocą macierzy korelacji (a_2), aż po kalibrację kopul Archimedesa dla zadanej wartości korelacji liniowej (a_3).

k	$(\underline{VaR}_{0.995}^{(a_k)}, \overline{VaR}_{0.995}^{(a_k)})$	z_k	$\prod_{j=2}^{k} z_j$	$(\Delta\underline{VaR}_{0.995}^{(a_k)}, \Delta\overline{VaR}_{0.995}^{(a_k)})$	A_k, B_k
1	(275, 1793)	–	–	–	–
2	(370, 1648)	0.7	0.7	(95, -145)	(341.5, 1691.5)
3	(1234.91, 1501.83)	0.7	0.49	(864.91, -146.17)	(765.3, 1619.9)

Tabela 6.8: Granice wiarygodności.
Źródło: Opracowanie własne.

Wyniki przedstawione w Tabeli 6.8 wskazują, że wprowadzanie kolejnych założeń dotyczących struktury zależności prowadzi do zawężania przedziałów niepewności modelu (A_k, B_k). Zastosowanie współczynnika istotności $z_j = 0.7$ oznacza, że każdemu nowemu założeniu przypisano 70% jego wiarygodności. Oznacza to częściowe, lecz niepełne ograniczenie niepewności wynikającej z braku pełnej informacji o strukturze zależności między czynnikami ryzyka. Ostateczny przedział ryzyka redukuje się do $(A_3, B_3) = (765.3, 1619.9)$, odzwierciedlając skumulowany wpływ wprowadzanych założeń o strukturze zależności. Dla uzyskanego przedziału wiarygodności (A_3, B_3), odpowiadającego wartościom (CLB, CUB), obliczono miary CAM i CRM zgodnie z podrozdziałem 4.4.2. Miary te umożliwiają ilościową ocenę ryzyka modelu, uwzględniając stopień wiarygodności przyjętych założeń oraz pozwalają na relatywne porównanie odporności modelu FS na niepewność w opisie struktury zależności.

Miara CAM określa różnicę pomiędzy górną granicą przedziału wiarygodności a wartością referencyjną $\mathrm{VaR}_{0.995}^{(\mathrm{FS})}$:

$$CAM = \frac{CUB - \mathrm{VaR}_{0.995}^{(\mathrm{FS})}}{\mathrm{VaR}_{0.995}^{(\mathrm{FS})}} \times 100\%. \tag{6.59}$$

Miara CRM wyraża położenie wartości referencyjnej $\mathrm{VaR}_{0.995}^{(\mathrm{FS})}$ wewnątrz przedziału niepewności $[CLB, CUB]$ i odnosi to położenie do całkowitej szerokości przedziału:

$$CRM = \frac{CUB - \mathrm{VaR}_{0.995}^{(\mathrm{FS})}}{CUB - CLB} \times 100\%. \tag{6.60}$$

Uzyskane wartości tych miar przedstawiono poniżej (tab. 6.9):

Miara	Wartość
CAM	22.53%
CRM	34.86%

Tabela 6.9: Wartości miar CAM i CRM dla przedziału wiarygodności $(CLB,\ CUB)$.
Źródło: Opracowanie własne.

Wartość $CAM = 22.53\%$ wskazuje, że zastosowanie korelacji liniowej w ramach metody FS prowadzi do niedoszacowania wymaganego kapitału o 22.53% względem wartości wynikającej z najbardziej niekorzystnej, dopuszczalnej struktury zależności. Z kolei, $CRM = 34.86\%$ odzwierciedla względny poziom ryzyka modelu, uwzględniający szerokość całego przedziału niepewności. Oznacza to, że około 34.86% potencjalnego przedziału niepewności struktury zależności można przypisać ryzyku wynikającemu z przyjęcia uproszczonego założenia o liniowej korelacji w metodzie FS. Przy zastosowanym współczynniku wiarygodności $z_j = 0.7$, reprezentującym 70% poziom istotności do kolejnych założeń dotyczących struktury zależności, uzyskane wartości CAM i CRM odzwierciedlają częściową redukcję niepewności modelu, wynikającą z niepełnej informacji o zależnościach między czynnikami ryzyka.

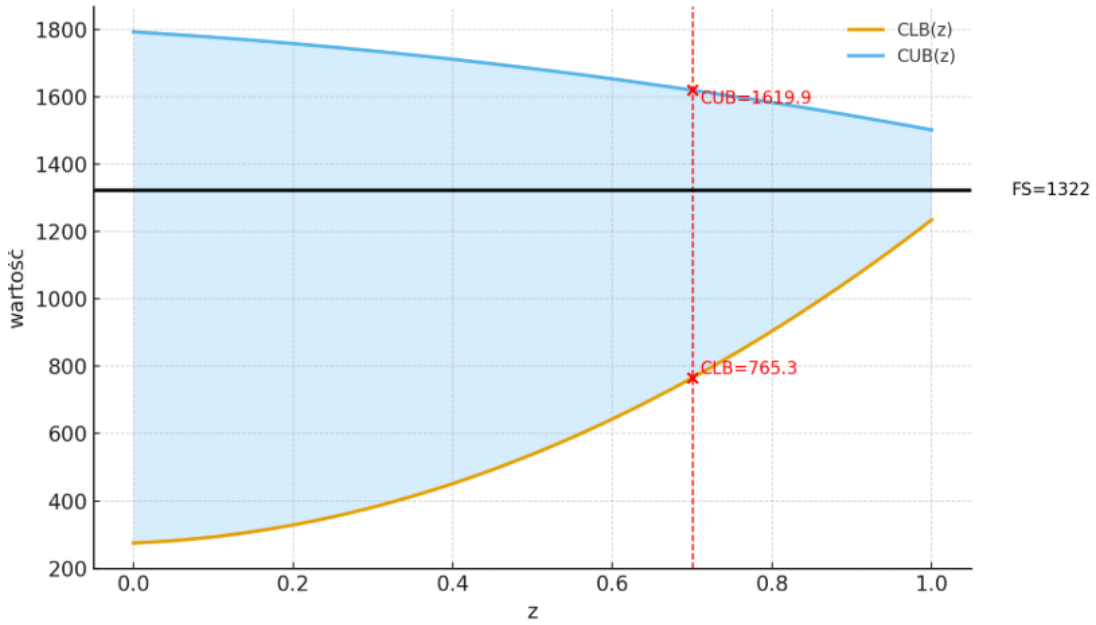
Pojawia się zatem pytanie, jaką dodatkową kwotę kapitału powinien utrzymywać ubezpieczyciel stosujący metodę FS, aby zabezpieczyć się przed ryzykiem modelu wynikającym z niejednoznaczności korelacji liniowej przy opisie struktury zależności. Odpowiedzi dostarcza miara $MoRC$, opisana w rozdziale 4.4.2, która łączy względną miarę ryzyka modelu (CRM) z rozpiętością granic wiarygodności ryzyka (CUB oraz CLB) i formalnie definiowana jest następująco:

$$MoRC(CRM, CUB, CLB, f) = f(CRM) \cdot (CUB - CLB), \tag{6.61}$$

Wartość $f(CRM)$ interpretuje się jako udział maksymalnego wymaganego kapitału.

Dla funkcji $f(x) = x^2$ otrzymano wartość $MoRC = 103.8$. Oznacza ona dodatkową kwotę kapitału, jaką ubezpieczyciel powinien utrzymywać w celu zabezpieczenia się przed potencjalnym błędem wynikającym z przyjętej struktury zależności. Wartość ta odzwierciedla wielkość rezerwy kapitałowej odpowiadającej niepewności modelowej, stanowiąc miarę wrażliwości wyniku metody FS na założenia dotyczące zależności między rodzajami ryzyka.

Współczynniki z_j w wyznaczaniu CLB i CUB można interpretować jako udział kolejnych założeń dotyczących struktury zależności w procesie oceny ryzyka modelu. Określają one stopień istotności, z jakim dane założenie jest uwzględnione w budowie przedziału wiarygodności: przy $z_j = 1$ wpływ danego założenia jest w pełni odzwierciedlony, natomiast przy $z_j = 0$ nie jest on w ogóle brany pod uwagę. Wartości pośrednie $0 < z_j < 1$ odpowiadają częściowej wiarygodności informacji o zależności i umożliwiają modelowanie stopniowej redukcji niepewności w miarę wprowadzania dodatkowych założeń. W kontekście metody FS, w której odporność na błędną specyfikację struktury zależności między ryzykami ubezpieczyciela ma kluczowe znaczenie dla adekwatności wymaganego kapitału SCR. Poniżej (rys. 6.7) przedstawiono wrażliwość granic przedziału $[CLB(z),\ CUB(z)]$ w funkcji wartości współczynnika z, ilustrującą wpływ poziomu wiarygodności założeń na szerokość przedziału niepewności.



Rysunek 6.7: Przedział $[\mathrm{CLB}(z), \mathrm{CUB}(z)]$ w funkcji z.
Źródło: Opracowanie własne.

Na Rysunku 6.7 przedstawiono zależność przedziału niepewności $[CLB(z), CUB(z)]$ od parametru z, który odzwierciedla stopień uwzględnienia kolejnych założeń dotyczących struktury zależności. Widoczny jest asymetryczny charakter zmian: dolna granica $CLB(z)$ rośnie w tempie szybszym niż maleje górna granica $CUB(z)$. Zjawisko to wskazuje na asymetrię wrażliwości metody FS na błędną specyfikację macierzy korelacji, co oznacza, że ryzyko niedoszacowania wymaganego kapitału jest istotnie większe niż ryzyko jego przeszacowania. Dodatkowo, wartość $\mathrm{VaR}_{0.995}$ uzyskana w ramach metody FS znajduje się relatywnie wysoko względem możliwego przedziału wartości, co sugeruje, że przy alternatywnych strukturach zależności poziom wymaganego kapitału mógłby być istotnie niższy.

Podsumowanie

W dyrektywie Solvency II kluczowym elementem procesu wyznaczania wymogów kapitałowych jest sposób agregacji ryzyka poszczególnych modułów i podmodułów. Agregacja ta jest bezpośrednio determinowana przez przyjętą strukturę zależności pomiędzy rodzajami ryzyka. Formuła standardowa przyjęta w dyrektywie zakłada, że zależności te mają charakter liniowy, a ich wartości zostały określone jednolicie dla wszystkich zakładów ubezpieczeń w Unii Europejskiej. Takie ujednolicenie struktury zależności upraszcza proces wyznaczania wymogu kapitałowego, jednak może prowadzić do jego niedoszacowania lub przeszacowania w przypadku występowania nieliniowych lub asymetrycznych współzależności pomiędzy czynnikami ryzyka. Twórcy dyrektywy są świadomi ograniczeń tego podejścia, dlatego dopuszczają możliwość budowy modeli wewnętrznych, które lepiej odzwierciedlają profil działalności ubezpieczyciela.

W pracy skoncentrowano się na analizie wpływu błędnej specyfikacji struktury zależności na oszacowanie kapitałowych wymogów wypłacalności. Punktem ciężkości badań była ilościowa ocena ryzyka modelu formuły standardowej, wynikającego z niepewności współzależności między rodzajami ryzyka, ze szczególnym uwzględnieniem wrażliwości wyników na specyfikację macierzy korelacji oraz niepewności w ogonach rozkładów. Równie ważnym etapem było autorskie opracowanie i zastosowanie nieparametrycznego podejścia, łączącego kopule ze sztucznymi sieciami neuronowymi, przeznaczonego do agregacji ryzyka w sposób dokładniej odzwierciedlający rzeczywistą strukturę zależności.

Za cel główny badań przyjęto **ilościową ocenę ryzyka modelu standardowej formuły agregacji wymogów kapitałowych w ramach Solvency II, wynikającego z niepewności struktury zależności między rodzajami ryzyka, oraz opracowanie i porównanie alternatywnych metod agregacji opartych na kopulach i uczeniu maszynowym, możliwych do zastosowania w modelach wewnętrznych. Uzyskane wyniki posłużyły do oceny odporności miary SCR oraz efektu dywersyfikacji na skrajne scenariusze współzależności.**

Podsumowując przeprowadzone działania, które pozwoliły na osiągniecie postawionego celu, w trakcie analizy:

- zbadano wrażliwość oszacowań SCR oraz efektu dywersyfikacji na niepewność elementów macierzy korelacji przyjętej w FS (identycznej dla wszystkich zakładów w UE),
- sprawdzono konsekwencje niepewności zależności ogonowych, których korelacja liniowa nie wychwytuje w ocenie wymogów kapitałowych,
- zweryfikowano wpływ wyboru klasy kopuli i jej parametrów na wyniki agregacji,
- opracowano i porównano alternatywne procedury agregacji ryzyka – parametryczne (kaskady kopul) oraz nieparametryczne, integrujące kopule z głębokimi sieciami neuronowymi, które pozwalają wierniej odwzorować nieliniowe i asymetryczne współzależności obserwowane w danych,
- oszacowano miary ryzyka modelu FS i na ich podstawie zidentyfikowano przypadki niedoszacowania lub przeszacowania SCR przy stosowaniu FS,
- sformułowano wnioski dotyczące odporności FS.

Realizację głównego celu pracy rozpoczęto od zaprezentowanej w rozdziale pierwszym charakterystyki systemu Solvency II oraz analizy sposobu agregacji ryzyka w poszczególnych modułach i podmodułach. W pierwszej kolejności przedstawiono podstawowe założenia FS oraz metody wyznaczania wymogu kapitałowego. Następnie przeprowadzono analizę literatury przedmiotu, obejmującą zarówno ograniczenia formuły standardowej, jak i dotychczasowe badania nad konstrukcją oraz zastosowaniem modeli wewnętrznych.

W drugim rozdziale pracy zaprezentowano analizę literatury poświęconą metodom agregacji ryzyka, ze szczególnym uwzględnieniem modeli opartych na teorii kopuli oraz ich zastosowaniu w sektorze ubezpieczeniowym. Na podstawie analizy literatury dokonano klasyfikacji podejść badawczych ze względu na zakres dostępnej informacji o strukturze zależności między ryzykami, wyróżniając przypadki pełnej wiedzy o rozkładzie wielowymiarowym, częściowej znajomości struktury zależności oraz znajomości jedynie rozkładów brzegowych.

W trzecim rozdziale przedstawiono charakterystykę metod uczenia maszynowego. Wskazano na niezwykle ważne aspekty dla prowadzonych badań dotyczących możliwości łączenia uczenia maszynowego z kopulami, w szczególności do estymacji, identyfikacji struktur zależności oraz do redukcji wymiaru i generowania syntetycznych danych uczących.

W rozdziale czwartym szczególną uwagę poświęcono zagadnieniu ryzyka modelu. Przedstawiono autorską klasyfikację ryzyka modelu według rodzajów i źródeł. W szczególności uwagę skupiono na ryzykach wynikających ze stosowania FS oraz potencjalnych zagrożeniach związanych z modelami wewnętrznymi. Przedstawiono ramy oceny ryzyka modelu.

W trakcie badań opracowano i wytrenowano autorską architekturę sieci neuronowej, dostosowaną do specyfiki danych ubezpieczeniowych oraz charakterystyki analizowanych zmiennych. Opracowane modele agregacji ryzyka stanowią fundament przeprowadzonych analiz, gdyż na ich podstawie dokonano oceny jakości odwzorowania struktury zależności między rodzajami ryzyka oraz dokładności oszacowania wymogu kapitałowego. W celu zbadania jakości modeli zastosowano zestaw miar dopasowania, obejmujący logarytm wiarygodności, odległość energetyczną (Energy Score) oraz wariogramową (Variogram Score). Miary te umożliwiają ilościową ocenę zgodności rozkładów generowanych przez model z danymi rzeczywistymi, a także porównanie precyzji modeli parametrycznych i nieparametrycznych. W części dotyczącej modelowania zależności wykorzystano dane pochodzące ze sprawozdań SFCR niemieckich ubezpieczycieli dla ubezpieczeń majątkowych. Wyniki estymacji przedstawiono w rozdziale piątym.

W dalszej części pracy, w rozdziale szóstym, opracowano i zaimplementowano autorski zestaw metod służących do ilościowej oceny ryzyka modelu wynikającego z błędnej specyfikacji struktury zależności pomiędzy rodzajami ryzyka. Zaprojektowane algorytmy pozwalają analizować wpływ poziomu informacji o zależnościach na wynik agregacji oraz na wartość zagregowanej miary ryzyka VaR. W ramach przeprowadzonych badań rozpatrzono przypadki wskazujące na brak jednoznaczności w opisie struktury zależności przez korelację liniową oraz zbadano, jaki wpływ mają zależności w ogonach rozkładów.

Opracowane metody umożliwiają oszacowanie miar ryzyka modelu FS i na ich podstawie zidentyfikowanie przypadków niedoszacowania lub przeszacowania SCR przy stosowaniu FS. Wyniki uzyskane w tej części badań stanowią podstawę do oceny odporności formuły standardowej oraz wskazują kierunki możliwej modyfikacji metodyki agregacji ryzyka w ramach modeli wewnętrznych.

Główny cel pracy zdekomponowano na cele szczegółowe, z których pierwsze dotyczą ujęcia poznawczego i teoretycznego, natomiast kolejne odnoszą się do części empirycznej. W tabe-

lach poniżej przedstawiono zestawienie celów wraz z przypisanymi im rozdziałami (etapami realizacji) oraz opisem podejścia badawczego. W Tabelach 6.10 - 6.12 zaprezentowano schemat prezentacji zrealizowanych celów wraz z przypisanymi rozdziałami, w których opisano podejście badawcze, podane równiez w wersji skróconej. Pierwsza Tabela 6.10 obejmuje cele o charakterze poznawczym, odnoszące się do przeglądu literatury, analizy formuły standardowej oraz opracowania podstaw metodologicznych. W Tabelach 6.11 - 6.12 zawarto informacje o potwierdzonych hipotezach badawczych wraz z informacjami o wynikach empirycznych, które je potwierdzają.

Dla czytelnej prezentacji tabel wprowadzamy skróty dla Rozdziałów - Rozdz., natomiast przez H - hipotezy.

Tabela 6.10: Zestawienie realizacji celów badawczych w rozdziałach pracy

Cel	Rozdz.	Opis podejścia badawczego
C2	1.1	Przedstawiono tło historyczne poprzedzające wdrożenie dyrektywy Solvency II. Zaprezentowano główne założenia trzech filarów dyrektywy oraz omówiono ich znaczenie dla funkcjonowania i nadzoru zakładów ubezpieczeń.
	1.2	Opisano FS zawartą w dyrektywie Solvency II, a także przedstawiono metodę agregacji kapitałowych wymogów wypłacalności z wykorzystaniem podejścia wariancji–kowariancji. Scharakteryzowano sposób, w jaki agregowane są poszczególne moduły i podmoduły ryzyka zdefiniowane w ramach FS.
	1.3	Omówiono zainicjowane przez EIOPA badania ilościowe (Quantitative Impact Studies-QIS) wraz z przedstawieniem płynących z nich wniosków oraz wskazaniem zidentyfikowanych wątpliwości dotyczących praktycznego stosowania FS.
	1.4	Przeanalizowano wymagania, jakie musi spełnić zakład ubezpieczeń ubiegający się o zgodę na wdrożenie modelu wewnętrznego. Wskazano kluczowe kryteria, procedury oraz znaczenie nadzoru w procesie zatwierdzania modelu wewnętrznego.
	2.1	Zdefiniowano pojęcie miary ryzyka oraz przedstawiono własności koherentnej miary ryzyka. Wprowadzono definicje miar Value at Risk oraz Expected Shortfall.

Kontynuacja na następnej stronie.

Tabela 6.10: (cd.) Zestawienie realizacji celów badawczych w rozdziałach pracy

Cel	Rozdz.	Opis podejścia badawczego
	2.2	Przedstawiono grupy metod agregacji ryzyka w zależności od sposobu modelowania zależności między czynnikami ryzyka. Omówiono podejścia nieuwzględniające zależności oraz metody, w których zależności są modelowane z wykorzystaniem funkcji kopuli. Z przeprowadzonego studium literaturowego wynika, że kluczową kwestią w agregacji ryzyka i wyznaczaniu efektu dywersyfikacji jest identyfikacja struktury zależności między agregowanymi czynnikami ryzyka.
C3	2.3, 2.4, 2.5	Zdefiniowano w sposób formalny kopulę jako narzędzie modelowania zależności między zmiennymi losowymi. Omówiono główne rodziny kopul wykorzystywanych w analizie ryzyka, w tym kopule eliptyczne, kopule Archimedesa, kaskady kopul oraz kopule nieparametryczne. Przedstawiono ich właściwości, ze szczególnym uwzględnieniem sposobu opisu zależności w ogonach rozkładów oraz możliwości dopasowania do danych empirycznych.
	2.6	Dokonano przeglądu literatury dotyczącej agregacji ryzyka w zależności od dostępnej informacji na temat struktury zależności między czynnikami ryzyka. Zaproponowano autorski podział metod na trzy grupy: obejmujące przypadki znanych rozkładów brzegowych przy braku informacji o strukturze zależności, częściowej informacji o strukturze zależności oraz znanego wielowymiarowego rozkładu agregowanych czynników ryzyka.
C1	3.1	Omówiono najczęściej stosowane w praktyce ubezpieczeniowej algorytmy uczenia maszynowego. Przedstawiono ich budowę, zasadę działania oraz podstawowe różnice wynikające z przyjętych założeń matematycznych. Zaprezentowano takie algorytmy jak: maszyna wektorów nośnych, głębokie sieci neuronowe, drzewa decyzyjne, lasy losowe oraz metody wzmacniania gradientowego.
	3.2	Przeprowadzono przegląd badań dotyczących wykorzystania metod uczenia maszynowego w aktuariacie i ubezpieczeniach, wskazując główne obszary ich zastosowań. W szczególności omówiono użycie algorytmów w procesach taryfikacji i wyceny składek, prognozowania rentowności klientów, detekcji oszustw ubezpieczeniowych, modelowaniu rezerw techniczno-ubezpieczeniowych, analizie ryzyka kredytowego, ocenie prawdopodobieństwa szkód oraz wykorzystaniu danych telematycznych i big data w ubezpieczeniach.

Kontynuacja na następnej stronie.

Tabela 6.10: (cd.) Zestawienie realizacji celów badawczych w rozdziałach pracy

Cel	Rozdz.	Opis podejścia badawczego
	3.3	Dokonano przeglądu literatury dotyczącej wzajemnych zastosowań kopul i algorytmów uczenia maszynowego, co pozwoliło na opracowanie autorskiego podziału metod na trzy grupy. Pierwsza z nich obejmuje wykorzystanie kopul w procesie redukcji wymiaru, w tym selekcję zmiennych na podstawie zależności między zmiennymi objaśniającymi oraz analizę zależności pomiędzy zmiennymi objaśniającymi a zmienną objaśnianą. Druga grupa dotyczy zastosowania algorytmów uczenia maszynowego w estymacji kopul, obejmując wykorzystanie sieci neuronowych do aproksymacji generatora kopul Archimedesa, ustalania porządku drzewa w kaskadach kopul oraz użycia generatywnych sieci neuronowych do tworzenia realizacji zgodnych z własnościami kopul. Trzecia grupa odnosi się do generowania danych syntetycznych przeznaczonych do trenowania modeli uczenia maszynowego, co ma szczególne znaczenie w kontekście ograniczonej dostępności i wrażliwości danych rzeczywistych. Przedstawiono nieparametryczny algorytm odwzorowujący dane rzeczywiste, z którego można wygenerować realizacje i wykorzystując je do uczenia algorytmów uczenia maszynowego.
C4	4.1	Przedstawiono znaczenie ryzyka modelu w działalności instytucji finansowych, ze szczególnym uwzględnieniem jego wpływu na stabilność systemu finansowego, zgodność regulacyjną oraz procesy decyzyjne. Omówiono główne obszary, w których ryzyko modelowe odgrywa kluczową rolę oraz zaprezentowano przykłady historycznych zdarzeń, które pokazują konsekwencje wykorzystania modelu z błędnymi założeniami.
	4.2	Przedstawiono klasyfikację rodzajów ryzyka modelu, wskazując ich występowanie i znaczenie w ramach Solvency II. Omówiono cztery główne obszary: ryzyko koncepcyjne, ryzyko implementacji algorytmów, ryzyko wynikające z jakości danych oraz ryzyko nadmiernego dopasowania modeli. Opisano etapy modelu, na których te ryzyka mogą występować, wskazując ich wpływ na dokładność obliczeń SCR.
	4.3	Przedstawiono główne źródła ryzyka modelu, wskazując ich znaczenie w kontekście Solvency II. Omówiono trzy zasadnicze grupy czynników: ludzki, technologiczny i środowiskowy, które mogą prowadzić do błędów w konstrukcji, implementacji i wykorzystaniu modeli. Podkreślono, że w ramach Solvency II istotne jest identyfikowanie tych źródeł na każdym etapie cyklu życia modelu.

Kontynuacja na następnej stronie.

Tabela 6.10: (cd.) Zestawienie realizacji celów badawczych w rozdziałach pracy

Cel	Rozdz.	Opis podejścia badawczego
	4.4	W oparciu o studium literaturowe opracowano metody służące do identyfikacji ryzyka modelu w dwóch podejściach: jakościowym oraz ilościowym. Podejście jakościowe opracowano w kontekście wymogów Solvency II, koncentrując się na inwentaryzacji i dokumentacji modelu. Dodatkowo, zwrócono uwagę na jakość danych, niezależność walidacji oraz regularne przeglądy modeli, które zapewniają przejrzystość i zgodność z regulacjami. W podejściu ilościowym skupiono się na analizie niepewności oraz porównywaniu modeli alternatywnych. Wprowadzono mierniki wpływu założeń w ryzyko modelu, służące ocenie, w jakim stopniu poszczególne założenia wpływają na całkowitą niepewność danego modelu. Zdefiniowano granice wiarygodności ryzyka (CLB, CUB), które określają przedział niepewności dodając wiarygodność założeń modelu. Kolejne bezwzględne (AM) i względne (RM) wskaźniki pozwalają porównać przyjęty model z modelem referencyjnym. Ich rozszerzeniem są CAM i CRM, uwzględniające wiarygodność poszczególnych założeń, umożliwiające bardziej realistyczną ocenę wpływu błędnych lub niepewnych założeń. Dodatkowo, wprowadzono miernik MoRC, który przekształca ocenę ryzyka modelowego na wymaganą dodatkową rezerwę kapitałową. Zestaw zaproponowanych mierników pozwala na ilościową, przejrzystą i zgodną z regulacjami ocenę wpływu przyjętych założeń na wyniki modelu oraz zakres niepewności związany z ich stosowaniem.

Źródło: opracowanie własne.

Tabela 6.11: Podsumowanie wyników badań i weryfikacja hipotez - rozdział 6

Cel	Rozdz.	Opis podejścia badawczego	H	Wnioski
	6.2	Opracowano *Algorytm 3* wyznaczający parametr zależności dla kopul Archimedesa (Claytona, Gumbela, Franka i Joe), odpowiadający zadanej korelacji liniowej. Algorytm łączy metodę siecznych i bisekcji, co zapewnia wysoką dokładność oraz stabilność obliczeń.		Uzyskane wyniki dla algorytmu ARA oraz czterech autorskich Algorytmów 1–4, zaprezentowane w Tabelach 6.7–6.9 w rozdziale szóstym, potwierdzają, że struktura zależności między ryzykami ma kluczowe znaczenie dla poziomu ryzyka modelu. Przeprowadzone analizy wykazały, że przy takiej samej wartości współczynnika korelacji liniowej kształt zależności jest różny, co prowadzi do znacząco odmiennych wyników agregacji ryzyka. Oznacza to, że korelacja liniowa, stosowana w FS, nie oddaje w pełni charakteru zależności między ryzykami, szczególnie w obszarach ekstremalnych wartości. Wyniki badań wskazują, że ogony rozkładów odpowiadające sytuacjom skrajnym mają istotny udział w kształtowaniu całkowitego ryzyka portfela, co uzasadnia potrzebę ich jednoznacznego modelowania. Wykorzystanie ilościowych mierników ryzyka modelowego umożliwiło Ponadto, identyfikację sytuacji, w których FS prowadzi do przeszacowania lub niedoszacowania wymogów kapitałowych. Opracowane algorytmy pozwalają ocenić zakres niepewności w estymacji wymaganego kapitału.
	6.3	Wykorzystując podejście wielokryterialne oparte na dominacji Pareto, opracowano *Algorytm 2* umożliwiający wyznaczenie takiej struktury zależności między ryzykami, która przy zadanej korelacji liniowej maksymalizuje wartość VaR.		
C7				
C8	6.1	Opracowano *Algorytm 1* umożliwiający analizę wrażliwości wartości VaR na zmiany współczynników korelacji między rodzajami ryzyka. Algorytm modyfikuje wybrane wartości w macierzy korelacji, wyznaczając przedział możliwych VaR. Dodatkowo, algorytm może zostać wykorzystany do określenia przedziału istotności zmian w macierzy korelacji.	H3	
C5			H4	
	6.4	Opracowano *Algorytm 4* wykorzystujący zniekształconą mieszankę kopul (Distorted Mix Copula, DMC) w połączeniu z metodą symulowanego wyżarzania badający wpływ założeń w ogonie na VaR. Znana jest struktura zależności w obszarze centralnym, natomiast niepewna jest zależność w ogonach rozkładów.		

Źródło: opracowanie własne.

Tabela 6.12: Podsumowanie wyników badań i weryfikacja hipotez - rozdział 5

Cel	Rozdz.	Opis podejścia badawczego	H	Wnioski
C8	5.1	Przedstawiono metodę dopasowania wielowymiarowego rozkładu z użyciem sieci neuronowych, obejmującą opis architektury i sposobu uczenia opisując funkcje celu zapewniające spełnienie warunków teoretycznych dla rozkładów brzegowych i kopuli. Omówiono narzędzia oceny jakości dopasowania, w tym miary takie jak odległość energetyczna, służące do weryfikacji zgodności z danymi rzeczywistymi.	**H1** **H2**	Przeprowadzone badania pokazują, że zastosowanie sieci neuronowych do jednoczesnej estymacji rozkładów brzegowych i struktury zależności umożliwia elastyczne modelowanie wielowymiarowego rozkładu ryzyk bez przyjmowania ograniczających założeń parametrycznych. Model nieparametryczny znacznie lepiej opisał dane rzeczywiste, co potwierdzają wyższe wartości log-wiarygodności oraz niższe wartości miar Energy Score i Variogram Score. Lepsze odwzorowanie struktury zależności między ryzykami przełożyło się bezpośrednio na wyższy efekt dywersyfikacji – 0.56 w porównaniu do 0.28 dla FS i około 0.35 dla modeli parametrycznych. Wyniki te wskazują, że stosowanie uproszczonych zależności liniowych może prowadzić do przeszacowania ryzyka i niedoszacowania dywersyfikacyjnego, natomiast opracowana metoda pozwala wyznaczyć rzeczywistą strukturę zależności.
	5.2	Przedstawiono charakterystykę danych wykorzystanych w badaniach.		
	5.3	Porównano podejście parametryczne i nieparametryczne do estymacji wielowymiarowych modeli. W podejściu parametrycznym rozkłady brzegowe dopasowano metodą MLE z wyborem według BIC (najlepszy rozkład logistyczny) oraz modelowano zależności kopulami C-vine i D-vine. W podejściu nieparametrycznym użyto sieci neuronowych do estymacji rozkładów brzegowych i kopuli. Porównanie modeli przeprowadzono dwutorowo: porównano rozkłady brzegowe parametryczne i nieparametryczne na zbiorze testowym, wykorzystując transformację PIT i test Kołmogorowa–Smirnowa, natomiast jakość generowanych realizacji z rozkładów wielowymiarowych oceniono na danych rzeczywistych z wykorzystaniem miar energetycznych Energy Score i Variogram Score.		
	5.4	Wyznaczono efekt dywersyfikacji w oparciu o podejścia parametryczne FS i kopulę C-vine i D-vine oraz podejście nieparametryczne wykorzystując sieci neuronowe. Otrzymane wyniki zestawiono z przypadkiem braku informacji na temat struktury zależnośći wykorzystując algorytm ARA.		

Z przeprowadzonych badań wynika, że model estymacji wielowymiarowego rozkładu oparty na sieciach neuronowych pozwala znacząco poprawić odwzorowanie struktury zależności między poszczególnymi rodzajami ryzyk. Model ten wyróżnia się większą elastycznością oraz zdolnością do uchwycenia złożonych, nieliniowych i asymetrycznych zalezności w porównaniu z klasycznymi podejściami parametrycznymi, takimi jak kaskady kopul typu C-vine i D-vine.

Dzięki zbudowanym algorytmom i miarom oceny ryzyka modelowego możliwa była szczegółowa analiza wpływu błędnej specyfikacji zależności na poziom kapitałowego wymogu wypłacalności, wyznaczanego zgodnie z formułą standardową dyrektywy Solvency II. Wyniki badań jednoznacznie wskazują, że agregacja oparta na założeniu o stałej liniowej korelacji może prowadzić zarówno do przeszacowania, jak i niedoszacowania wymaganego kapitału.

W związku z powyższym potwierdzona została prawdziwość hipotezy głównej pracy, zgodnie z którą **niepewność struktury zależności między rodzajami ryzyka generuje istotne ryzyko modelu w formule standardowej Solvency II, przez co wyniki SCR i efekt dywersyfikacji uzyskane z FS nie są odporne na skrajne scenariusze współzależności, natomiast metody agregacji oparte na kopulach i uczeniu maszynowym, lepiej odwzorowujące rzeczywiste zależności, dostarczają stabilniejszych i bardziej adekwatnych oszacowań.**

Podkreśla to kluczowe znaczenie stosowania zaawansowanych metod, umożliwiających dokładniejsze modelowanie struktury zależności, szczególnie w kontekście oceny wypłacalności zakładów ubezpieczeń.

Na podstawie uzyskanych wyników można zakładom ubezpieczeń rekomendować włączenie ilościowej oceny ryzyka modelu do stałych procesów zarządczych, w szczególności do ORSA i planowania kapitałowego. Praktycznie oznacza to regularne wyznaczanie przedziałów niepewności dla VaR, SCR i efektu dywersyfikacji oraz przekładanie tych przedziałów na bufor kapitałowy, na przykład z użyciem wskaźników AM i RM oraz ich wersji ważonych wiarygodnością CAM i CRM, a także miary MoRC, która pozwala wyrazić ryzyko modelu w jednostkach kapitału. W raportach wewnętrznych i materiałach dla komitetów ryzyka powinny systematycznie pojawiać się zarówno wartości punktowe, jak i ich pasma wraz z krótką interpretacją konsekwencji decyzyjnych.

Drugim filarem rekomendacji może być dyscyplina w badaniu wrażliwości. Zakłady powinny rutynowo analizować, jak zmiany elementów macierzy korelacji przyjętej w formule standardowej wpływają na kwantyle strat, oraz konfrontować tę samą wartość korelacji z różnymi strukturami zależności opisanymi przez kopule. Osobne miejsce trzeba poświęcić zależnościom ogonowym, które najmocniej kształtują wysokie kwantyle i mogą odwrócić wnioski co do wielkości buforów. Taki zestaw scenariuszy skrajnych współzależności powinien być powtarzany cyklicznie, tak aby uchwycić ewentualny dryf w czasie.

Można zalecić także równoległe utrzymywanie rozwiązania opartego na formule standardowej oraz alternatywnych modeli agregacji: parametrycznych z wykorzystaniem kopul oraz nieparametrycznych, które łącząc kopule z głębokimi sieciami neuronowymi umożliwiają elastyczną estymacji rozkładów brzegowych i struktury zależności. Modele te należy porównywać nie tylko na podstawie wyników SCR, lecz także na podstawie jakości dopasowania do danych poza próbą, wykorzystując transformacje PIT, test Kołmogorowa–Smirnowa dla rozkładów brzegowych oraz miary jakości prognoz probabilistycznych dla rozkładów wielowymiarowych, takie jak Energy Score i Variogram Score. W momencie gdy alternatywny model systematycznie dostarcza węższych, stabilniejszych pasm niepewności i lepszego dopasowania, powinien on informować decyzje kapitałowe i parametry reasekuracji, nawet jeśli FS pozostaje punktem odniesienia regulacyjnego.

Z perspektywy nadzorczej i regulacyjnej wnioski prowadzą do kilku spójnych postulatów. Pożądane jest, aby w sprawozdawczości nadzorczej pojawił się obowiązek ujawniania przedziałów niepewności dla SCR i efektu dywersyfikacji wraz z opisem metody ich wyznaczania pod różnymi założeniami o zależnościach. Minimalny zestaw scenariuszy ekstremalnych, obejmujący warianty silnych zależności w ogonach i alternatywne struktury kopul przy tej samej korelacji liniowej, powinien zostać zdefiniowany i stosowany sektorowo. W modelach wewnętrznych warto wyraźnie dopuścić rozwiązania oparte na kopulach i metodach nieparametrycznych pod warunkiem spełnienia standardów walidacyjnych, a wyniki jakości dopasowania powinny być prezentowane w ujednoliconych szablonach.

W świetle uzyskanych wyników i zidentyfikowanego potencjału badawczego naturalnym kierunkiem dalszych prac jest rozwinięcie oraz integracja opracowanych metod w ramach bardziej złożonych struktur modelowania zależności. W szczególności planowane jest połączenie algorytmów zaprezentowanych w rozdziale trzecim, w którym konstrukcja drzewa kaskad kopul zostanie wyznaczona z wykorzystaniem jednej sieci neuronowej, natomiast estymacja dwuwymiarowych kopul Archimedesa będzie realizowana przy użyciu drugiej sieci neuronowej, odpowiadającej za estymację generatora kopuli Archimedesa. Z kolei, konstrukcja kopuli DMC, w której parametrycznie modelowany jest zarówno obszar centralny, jak i ogonowy zależności, zostanie oparta na algorytmie przedstawionym w rozdziale piątym. Zwracamy również uwagę, że opracowany w niniejszej pracy model estymacji rozkładów brzegowych oraz wielowymiarowego rozkładu ryzyka z wykorzystaniem sieci neuronowych stanowi uniwersalne i elastyczne narzędzie analityczne, które może znaleźć zastosowanie w różnych obszarach analizy aktuarialnej, modelowania ryzyka oraz zarządzania kapitałem. Poniżej przedstawiono rekomendowane kierunki praktycznego wykorzystania opracowanego podejścia w kontekście ubezpieczeń i analiz aktuarialnych:

- modelowanie wysokości szkody (severity) - estymacja rozkładu wypłaty pojedynczej szkody na podstawie cech klienta, rodzaju ubezpieczenia lub charakterystyki zdarzenia szkodowego,
- modelowanie częstości szkód (frequency) - określanie rozkładu liczby szkód w zadanym okresie dla danego ubezpieczonego lub grupy, bez konieczności zakładania klasycznego rozkładu dyskretnego (np. rozkładu Poissona),
- wyznaczanie brzegowych rozkładów wypłat w rezerwach szkodowych - estymacja rozkładu płatności w kolejnych latach rozwoju szkody, wykorzystywana w modelach IBNR oraz RBNS,
- modelowanie zależności między częstością a wysokością szkód - wspólne modelowanie częstości i severity jako zmiennych współzależnych (np. klient o wysokiej częstości szkód ma też tendencję do wysokich wypłat),
- rezerwy techniczno-ubezpieczeniowe - modelowanie zależności pomiędzy liniami biznesowymi, rozwojem szkody a wypłatami w różnych okresach rozliczeniowych,
- wyceny ryzyka - wspólna analiza prawdopodobieństw wypłaty w wielu liniach ubezpieczeń w celu wyceny ryzyka portfela lub pakietu produktów,

- ryzyko katastroficzne - modelowanie współwystępowania szkód (np. burze, powodzie) w różnych lokalizacjach lub segmentach portfela, gdzie zależności są silne, nieliniowe i asymetryczne,
- ubezpieczenia zdrowotne i na życie - modelowanie zależności między długością życia, wiekiem zachorowania, czasem leczenia, kosztami leczenia itd. w ramach produktów długoterminowych.

Spis tabel

Spis rysunków

Literatura

Aas, K., Czado, C., Frigessi, A., & Bakken, H. (2009). Pair-copula constructions of multiple dependence. Insurance: Mathematics and Economics, 44(2), 182–198.

Acemoglu, D., Ozdaglar, A., & Tahbaz-Salehi, A. (2015). Systemic risk and stability in financial networks. American Economic Review, 105(2), 564-608.

Acharya, V. V., Pedersen, L. H., Philippon, T., & Richardson, M. (2017). Measuring systemic risk. The Review of Financial Studies, 30(1), 2-47.

American Academy of Actuaries. (2019a). Model risk management practice note. (Practice note providing guidance on model risk management)

American Academy of Actuaries. (2019b). Model risk management practice note. Practice Note of the American Academy of Actuaries.

Apostol, T. M. (1974). Mathematical analysis (2nd ed.). Reading, MA: Addison–Wesley.

Arbenz, P., Embrechts, P., & Puccetti, G. (2011a). The aep algorithm for the fast computation of the distribution of the sum of dependent random variables. Bernoulli, 17(2), 562–591. doi: 10.3150/10-BEJ287

Arbenz, P., Embrechts, P., & Puccetti, G. (2011b). A numerically stable algorithm for the computation of the distribution of the sum of dependent risks. Journal of Computational and Applied Mathematics, 236(4), 524–540. doi: 10.1016/j.cam.2011.08.027

Arora, N., & Kaur, P. D. (2020). A bolasso based consistent feature selection enabled random forest classification algorithm: An application to credit risk assessment. Applied Soft Computing, 86.

Artzner, P., Delbaen, F., Eber, J., & Heath, D. (1999). Coherent measures of risk. Mathematical Finance, 9(3), 203–228.

Awad, H. (2012). Taxonomy of machine learning algorithms to classify real-time interactive applications. Unpublished manuscript.

Bärtl, M., & Krummaker, S. (2020). Prediction of claims in export credit finance: A comparison of four machine learning techniques. Risks, 8(1), 22.

Baydin, A. G., Pearlmutter, B. A., Radul, A. A., & Siskind, J. M. (2018). Automatic differentiation in machine learning: a survey. Journal of Machine Learning Research, 18(153), 1–43. Retrieved from `http://jmlr.org/papers/v18/17-468.html`

Bedford, T., & Cooke, R. M. (2002). Vines: A new graphical model for dependent random variables. Annals of Statistics, 30(4), 1031–1068.

Bellini, F., & Di Bernardino, E. (2015). Risk measures based on generalized quantiles. European Journal of Operational Research, 245(1), 291–300.

Bellini, F., Klar, B., Müller, A., & Rosazza Gianin, E. (2014). Generalized quantiles as risk measures. Insurance: Mathematics and Economics, 54, 41–48.

Benali, A., Notton, G., Fouilloy, A., & Voyant, C. (2021). Copula-based modeling and machine learning for multivariate dependence in renewable energy data. Energy and AI, 5, 100092. doi: 10.1016/j.egyai.2021.100092

Bernard, C. (2023). A practical approach to quantitative model risk assessment. Variance: Advancing the Science of Risk, 17(1), 55–77.

Bernard, C., Kazzi, R., & Vanduffel, S. (2019). Risk bounds for partially specified unimodal distributions.

Bernard, C., Kazzi, R., & Vanduffel, S. (2023). A practical approach to quantitative model risk assessment. Variance, 16(1), 1–29.

Bernard, C., Rüschendorf, L., & Vanduffel, S. (2017a). Value-at-risk bounds with given marginals and maximum variance. Journal of Multivariate Analysis, 155, 33–45. doi: 10.1016/j.jmva.2017.01.008

Bernard, C., Rüschendorf, L., & Vanduffel, S. (2017b). Value-at-risk bounds with variance constraints. Journal of Risk and Insurance, 84(3), 923–959.

Black, R., Bonsall, J., Godfrey, D., Khalaf, T., & Rees, B. (2017). Model risk: Illuminating the black box. British Actuarial Journal, 22(e8), 1–35.

Black, R., Bonsall, J., Godfrey, D., Khalaf, T., & Rees, B. (2018). Model risk: Illuminating the black box. British Actuarial Journal, 23(e2), 1–58. (Institute and Faculty of Actuaries, Model Risk Working Party)

Boucher, J.-P., Pérez-Marín, A. M., & Santolino, M. (2013). Pay-as-you-drive insurance: the effect of the kilometers on the risk of accident. Anales del Instituto de Actuarios Españoles, 19, 135–154.

Box, G. E. P. (1979). Robustness in the strategy of scientific model building. In R. L. Launer & G. N. Wilkinson (Eds.), Robustness in statistics (pp. 201–236). Academic Press.

Brechmann, S. U., E. C. (2013). Modeling dependence with c- and d-vine copulas: The r package vinecopula. Journal of Statistical Software, 52(3), 1–27.

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. doi: 10.1023/A:1010933404324

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and regression trees. Belmont, CA: Wadsworth International Group.

Brent, R. P. (1973). Algorithms for minimization without derivatives. Englewood Cliffs, NJ: Prentice-Hall.

Breuer, T., & Csiszár, I. (2016). Measuring distribution model risk. Mathematical Finance, 26(2), 395–411.

Burden, R. L., & Faires, J. D. (2011). Numerical analysis (9th ed.). Boston, MA: Brooks/Cole, Cengage Learning.

Byrne, L. S., P. (2004). Different risk measures: Different portfolio compositions? (Working Paper). University of Reading.

Černý, V. (1985). Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm. Journal of Optimization Theory and Applications, 45(1), 41–51. doi: 10.1007/BF00940812

Chang, W. T., & Lai, K. H. (2021). A neural network-based approach in predicting consumers' intentions of purchasing insurance policies. Acta Informatica Pragensia, 10(2), 138–154.

Chollet, F., et al. (2015). Keras.

Chu, A., Ip, W., Lam, K., & So, M. (2022). Vine copula statistical disclosure control for mixed-type data. Computational Statistics Data Analysis, 173, 107544.

Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. Biometrika, 65(2), 141–151.

Commission, F. C. I. (2011). The financial crisis inquiry report: Final report of the national commission on the causes of the financial and economic crisis in the united states. Washington, D.C.: U.S. Government Printing Office.

Corlosquet-Habart, M., & Janssen, J. (2018). Big data for insurance companies. Hoboken, NJ: John Wiley Sons.

Culver, Q., Heitmann, D., & Weiß, C. (2018). The influence of seed selection

on the solvency ii ratio. arXiv preprint arXiv:1801.05409. Retrieved from https://arxiv.org/abs/1801.05409

Czado, C. (2019). Analyzing dependent data with vine copulas: A practical guide with r (Vol. 222). Cham: Springer.

Danielsson, J. (2016). Model risk of risk models. Journal of Financial Stability, 22, 1-10.

Danielsson, J., Embrechts, P., Goodhart, C., Keating, C., Muennich, F., Renault, O., & Shin, H. S. (2013). Fat tails, value-at-risk and subadditivity. Journal of Econometrics, 172(2), 283–291.

Deb, K. (2001). Multi-objective optimization using evolutionary algorithms. Chichester, UK: John Wiley Sons.

Derman, E. (1996). Model risk. Risk, 9(5), 139–145.

Desik, P. H. A., & Behera, S. (2012). Acquiring insurance customer: The chaid way. SSRN Electronic Journal. doi: 10.2139/ssrn.2169897

Dhaene, J., Denuit, M., Goovaerts, M., Kaas, R., & Vyncke, D. (2002). The concept of comonotonicity in actuarial science and finance: Theory. Insurance: Mathematics and Economics, 31(1), 3–33. doi: 10.1016/S0167-6687(02)00147-5

Dhaene, J., Vanduffel, S., Goovaerts, M. J., Kaas, R., & Vyncke, D. (2006). Risk measures and comonotonicity: A review. North American Actuarial Journal, 10(4), 107–122.

Dhieb, N., Ghazzai, H., Besbes, H., & Massoud, Y. (2019). Extreme gradient boosting machine learning algorithm for safe auto insurance operations. International Journal of Intelligent Systems Technologies and Applications, 18(2), 142–157.

Dowd, K. (2005). Measuring market risk (2nd ed.). Chichester: Wiley.

Dozat, T. (2016). Incorporating nesterov momentum into adam. In Proceedings of the 4th international conference on learning representations, workshop track (pp. 1–4).

Duchi, J., Hazan, E., & Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. Journal of Machine Learning Research, 12, 2121–2159.

Duchon, M. (2011). A generalized bernstein approximation theorem. Tatra Mountain Mathematical Publications, 49, 99–109.

Dugas, C., Bengio, Y., & Chapados, N. (2003). Insurance risk modeling using neural networks and generalized linear models. In Casualty actuarial society forum (pp. 177–217).

Duhamel, P., & Vetterli, M. (1990). Fast fourier transforms: A tutorial review and a state of the art. Signal Processing, 19(4), 259–299. doi: 10.1016/0165-1684(90)90095-Q

Durante, F., & Sempi, C. (2015). Principles of copula theory. CRC Press.

Efron, B., & Hastie, T. (2020). Computer age statistical inference: Algorithms, evidence, and data science. Cambridge: Cambridge University Press.

Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. The Annals of Statistics, 32(2), 407–499. doi: 10.1214/009053604000000067

Ehrgott, M. (2005). Multicriteria optimization (Vol. 491). Berlin, Heidelberg: Springer.

Embrechts, P., McNeil, A., & Straumann, D. (2002). Correlation and dependence in risk management: Properties and pitfalls. Risk Management: Value at Risk and Beyond, 176–223.

Embrechts, P., Puccetti, G., & Rüschendorf, L. (2013). Model uncertainty and var aggregation. Journal of Banking Finance, 37(8), 2750–2764.

Embrechts, P., Wang, B., & Wang, R. (2015). Aggregation-robustness and model uncertainty of regulatory risk measures. Finance and Stochastics, 19(4), 763-790.

Embrechts, P. G. R. L., P. (2009). Additivity properties for value-at-risk under archimedean dependence and heavy-tailedness. Insurance: Mathematics and Economics, 44(2), 164–169.

European Commission. (2015). Commission delegated regulation (eu) 2015/35 of 10 october 2014 supplementing directive 2009/138/ec of the european parliament and of the council on the taking-up and pursuit of the business of insurance and reinsurance (solvency ii). Official Journal of the European Union, 1-797.

Fang, K., Jiang, Y., & Song, M. (2016). Customer profitability forecasting using big data analytics: A case study in the insurance industry. Computers Industrial Engineering, 101, 554–564.

Femmam, K., Brahimi, B., & Femmam, S. (2023). An optimized feature selection technique based on bivariate copulas "gbcfs". Journal of Combinatorial Optimization, 45(2), 478–493.

Femmam, K., & Femmam, S. (2022). Fast and efficient feature selection method using bivariate copulas. Journal of Advances in Information Technology, 13(3), 301-305.

Ferrario, A., & Hämmerli, N. (2019). Introduction to machine learning for actuaries. Swiss Association of Actuaries Tutorial, 1–28.

Föllmer, S. A., H. (2011). Stochastic finance: An introduction in discrete time (3rd ed.). Berlin: De Gruyter.

Fox, L. (2003). Enron: The rise and fall. Hoboken, NJ: Wiley.

Frank, M. J. (1979). On the simultaneous associativity of f(x,y) and x + y – f(x,y). Aequationes Mathematicae, 19, 194–226.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. The Annals of Statistics, 29(5), 1189–1232. doi: 10.1214/aos/1013203451

Fuchs, S., & Schmidt, K. D. (2014). Bivariate copulas: Transformations, asymmetry and measures of concordance. Kybernetika, 50(1), 109–125.

Gabrielli, A. (2019). A neural network boosted double over–dispersed poisson claims reserving model. SSRN Electronic Journal.

Gabrielli, A. (2020). A neural network boosted double overdispersed poisson claims reserving model. ASTIN Bulletin: The Journal of the International Actuarial Association, 50(1), 25–60.

Gabrielli, A., & Wüthrich, M. V. (2018). Neural network embedding of the over-dispersed poisson reserving model. Scandinavian Actuarial Journal, 2018(9), 794–817.

Gao, G., Meng, S., & Wüthrich, M. V. (2020). Improving car insurance pricing using telematics data with convolutional neural networks. Scandinavian Actuarial Journal, 2020(9), 818–845.

Gao, G., & Wüthrich, M. V. (2019). Convolutional neural network classification of telematics car driving data. Risks, 7(1), 6.

Gatarek, D. (2002). Ryzyko modelu. Rynek Terminowy(4), 24–31.

Geman, S., & Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-6(6), 721–741. doi: 10.1109/TPAMI.1984.4767596

Genest, C., Rémillard, B., & Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. Insurance: Mathematics and Economics, 44(2), 199–213.

Genest, F. A.-C., C. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. Journal of Hydrologic Engineering, 12(4), 347–368.

Glasserman, P., & Xu, X. (2014). Robust risk measurement and model risk. Quantitative Finance, 14(1), 29–58.
Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association, 102(477), 359–378.
Gramegna, A., & Giudici, P. (2020). Why to buy insurance? an explainable artificial intelligence approach. Risks, 8(4), 137.
Griesbauer, E., Czado, C., Frigessi, A., & Haff, I. H. (2025). Tvinesynth: A truncated c-vine copula generator of synthetic tabular data to balance privacy and utility. arXiv preprint arXiv:2503.15972.
Grize, Y., Fischer, W., & Lützelschwab, C. (2020). Machine learning applications in nonlife insurance. Applied Stochastic Models in Business and Industry, 36(4), 523–537. doi: 10.1002/asmb.2543
Gumbel, E. J. (1960). Bivariate exponential distributions. Journal of the American Statistical Association, 55(292), 698–707.
Hájek, B. (1988). Cooling schedules for optimal annealing (Vol. 13) (No. 2). Springer. doi: 10.1287/moor.13.2.311
Hofert, M. (2008). Sampling archimedean copulas. Computational Statistics and Data Analysis, 52(12), 5163–5174.
Hofert, M., Memartoluie, A., Saunders, D., & Wirjanto, T. (2017). Improved algorithms for computing worst value-at-risk: Numerical challenges and the adaptive rearrangement algorithm. The Journal of Risk, 19(2), 1–38.
Idris, H. (2024). Exploring financial risk management: A qualitative study on risk identification, evaluation and mitigation in banking, insurance and corporate finance. ResearchGate Preprint.
Jacobs, M. (2015). The quantification and aggregation of model risk: Perspectives on potential approaches. International Journal of Financial Engineering and Risk Management, 2(2), 124–150.
Jakobsons, E., Han, X., & Wang, R. (2016). Bounds for expectiles given marginals and correlations. Insurance: Mathematics and Economics, 71, 394–406.
Jakobsons, E., & Vanduffel, S. (2015). Bounds for risk measures given marginals and covariances. Insurance: Mathematics and Economics, 64, 1–10.
Janke, D., & Steinke, F. (2021). Implicit generative copulas. arXiv preprint arXiv:2109.14567.
Jeong, E.-H., Lee, S.-Y., & Lee, Y.-S. (2016). Copula-based data generation for privacy-preserving data mining. Journal of Information Science and Engineering, 32(3), 627–646.
Joe, H. (1996). Families of m-variate distributions with given margins and m(m-1)/2 bivariate dependence parameters (Vol. 28). Institute of Mathematical Statistics.
Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. Science, 220(4598), 671–680. doi: 10.1126/science.220.4598.671
Kirlidog, M., & Asuk, C. (2012). A fraud detection approach with data mining in health insurance. Procedia - Social and Behavioral Sciences, 62, 989–994.
Kokol Bukovšek, D., Mazo, G., Mesiar, R., & Omladič, M. (2021). Bounds on bivariate distribution functions with given spearman's rho or gini's gamma. Journal of Computational and Applied Mathematics, 391.

Kozmenko, O., Bieliaieva, N., & Ivashyna, K. (2015). Statistical model of risk assessment of insurance companies. Investment Management and Financial Innovations, 12(2), 120–128.

Kuo, K. (2020). Individual claims forecasting with bayesian mixture density networks. Casualty Actuarial Society E-Forum, Winter 2020, 1–22.

Kuziak, K. (2005). Ryzyko modeli w szacowaniu instrumentów finansowych. Zeszyty Naukowe Uniwersytetu Ekonomicznego(690), 45–58.

Lall, S., Chowdhury, S. R., & Ghosh, S. (2021). Copula-based mutual information for feature selection in machine learning. Pattern Recognition Letters, 152, 164–171.

Li, L., Yuen, K. C., & Yang, J. (2014). Distorted mix method for constructing copulas with tail dependence. Insurance: Mathematics and Economics, 57(3), 77–89. doi: 10.1016/j.insmatheco.2014.05.002

Liebscher, E. (2014). Copula-based dependence measures. Dependence Modeling, 2(1), 49–64.

Lindholm, M., Lindskog, F., & Wahl, F. (2020). Estimation of conditional mean squared error of prediction for claims reserving. Annals of Actuarial Science, 14(1), 93-128.

Ling, Z., Fang, X., & Kolter, J. Z. (2020). Deep archimedean copulas. Advances in Neural Information Processing Systems (NeurIPS), 33, 16083–16093.

Lopez, J. A., & Saidenberg, M. R. (2001). The development of internal models approaches to bank regulation & supervision: Lessons from the market risk amendment. Federal Reserve Bank of San Francisco Economic Review.

Marill, T., & Green, D. M. (1963). On the effectiveness of receptors in recognition systems. IEEE Transactions on Information Theory, 9(1), 11–17. doi: 10.1109/TIT.1963.1057810

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. The Bulletin of Mathematical Biophysics, 5(4), 115–133. doi: 10.1007/BF02478259

McLean, B., & Elkind, P. (2003). The smartest guys in the room: The amazing rise and scandalous fall of enron. New York: Portfolio.

McNeil, A. J., & Nešlehová, J. (2009). Multivariate archimedean copulas, d-monotone functions and ℓ_1-norm symmetric distributions. Annals of Statistics, 37(5B), 3059–3097.

Meyer, D., Schefzik, R., Thorarinsdottir, T. L., & Gneiting, T. (2021). Copula-based synthetic data augmentation for machine-learning emulators. Geoscientific Model Development, 14(8), 5205–5223. doi: 10.5194/gmd-14-5205-2021

Nejad, M. M., Erdogan, S., & Cirillo, C. (2021). A statistical approach to small area synthetic population generation as a basis for carless evacuation planning. Journal of Transport Geography, 90, 102902.

Nesterov, Y. (1983). A method for solving the convex programming problem with convergence rate $o(1/k^2)$. In Sssr.

Neumann, L., Nowak, R., Okuniewski, R., & Wawrzynski, P. (2019). Machine learning-based predictions of customers' decisions in car insurance. Applied Artificial Intelligence, 33(12), 1046–1063.

Newey, W. K., & Powell, J. L. (1987). Asymmetric least squares estimation and testing. Econometrica, 55(4), 819–847.

Ng, Y., Hasan, A., Elkhalil, K., & Tarokh, V. (2021). Generative archimedean copulas. arXiv preprint arXiv:2102.11351.

on Investigations, U. S. P. S. (2013). Jpmorgan chase whale trades: A case history of derivatives risks abuses (Tech. Rep.). United States Senate.

Ortega, J. M., & Rheinboldt, W. C. (1970). Iterative solution of nonlinear equations in several variables. New York: Academic Press.

Panigrahi, R. K., Palkar, R., et al. (2018). Comparative analysis on classification algorithms of auto-insurance fraud detection based on feature selection algorithms. International Journal of Innovative Technology and Exploring Engineering, 8(2S2), 470–475.

Panjer, H. H. (1981). Recursive evaluation of a family of compound distributions. ASTIN Bulletin, 12(1), 22–26. doi: 10.1017/S0515036100006796

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 2(11), 559–572. doi: 10.1080/14786440109462720

Pedersen, C., & E., S. S. (1998). An extended family of financial risk measures. The Geneva Papers on Risk and Insurance Theory, 23(2), 89–117.

Polyak, B. T. (1964). Some methods of speeding up the convergence of iteration methods. Zh. Vychisl. Mat. Mat. Fiz., 4(5), 791–803. Retrieved from `https://www.mathnet.ru/eng/zvmmf7713` (USSR Comput. Math. Math. Phys. translation)

Prudential Regulation Authority. (2023). Model risk management principles for banks (supervisory statement ss1/23). Bank of England, Supervisory Statement, 1–38.

Puccetti, G., & Rüschendorf, L. (2012). Computation of sharp bounds on the distribution of a function of dependent risks. Journal of Computational and Applied Mathematics, 236(7), 1833–1840.

Puccetti, G., & Rüschendorf, L. (2015). Computation of sharp bounds on the expected value of a supermodular function of risks with given marginals. Communications in Statistics - Simulation and Computation, 44(3), 705–718.

Quan, Z., & Valdez, E. A. (2018). Predictive analytics of insurance claims using multivariate decision trees. Dependence Modeling, 6(1), 377–407.

Rajan, R. G. (2010). Fault lines: How hidden fractures still threaten the world economy. Princeton, NJ: Princeton University Press.

Rani, R., Khurana, M., Kumar, A., & Kumar, N. (2022). Big data dimensionality reduction techniques in iot: Review, applications and open research challenges. Cluster Computing, 25, 4027–4058. doi: 10.1007/s10586-022-03634-y

Restrepo, N., Arango, J. C., Trujillo, L., & Zapata, C. M. (2023). Nonparametric generation of synthetic data using copulas. Electronics, 12(7), 1601.

Rockafellar, U. S. Z. M., R. T. (2002). Deviation measures in risk analysis and optimization. Finance and Stochastics, 6(2), 191–203.

Rudin, W. (1976). Principles of mathematical analysis (3rd ed.). New York: McGraw–Hill.

Rukhsar, L., Bangyal, W. H., Nisar, K., & Nisar, S. (2022). Prediction of insurance fraud detection using machine learning algorithms. Mehran University Research Journal of Engineering Technology, 41(1), 33–40.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. Nature, 323(6088), 533–536. doi: 10.1038/323533a0

Sakthivel, K. M., & Rajitha, C. S. (2017). Artificial intelligence for estimation of future claim frequency in non-life insurance. Global Journal of Pure and Applied Mathematics, 13(6), 1701–1710.

Schelldorfer, J., & Wüthrich, M. V. (2019). Nesting classical actuarial models into neural networks. Swiss Association of Actuaries Tutorial, 1–35.

Scheuerer, M., & Hamill, T. M. (2015). Variogram-based proper scoring rules for probabilistic forecasts of multivariate quantities. Monthly Weather Review, 143(4), 1321–1334.
Schubert, T., & Grießmann, G. (2007). German proposal for a standard approach for solvency ii. The Geneva Papers on Risk and Insurance - Issues and Practice, 32(1), 133–150.
Schwarz, G. (1978). Estimating the dimension of a model. The Annals of Statistics, 6(2), 461–464. doi: 10.1214/aos/1176344136
Sei, Y., Onesimu, J. A., & Ohsuga, A. (2022). Machine learning model generation with copula-based synthetic dataset for local differentially private numerical data. IEEE Access, 10, 101656–101671. doi: 10.1109/ACCESS.2022.3208715
Shen, F., Zhao, X., Kou, G., & Alsaadi, F. E. (2021). A new deep learning ensemble credit risk evaluation model with an improved synthetic minority oversampling technique. Applied Soft Computing, 98.
Shirreff, D. (1999). Dealing with financial risk. London: Risk Books.
Silver, N. (2012). The signal and the noise: Why so many predictions fail – but some don't. New York: Penguin Press.
Singh, D., Balamurugan, S., & Jebamalar Leavline, E. (2016). Literature review on feature selection methods for high-dimensional data. International Journal of Computer Applications, 136(1). doi: 10.5120/ijca2016908317
Society of Actuaries, & Canadian Institute of Actuaries. (2019). The use of predictive analytics in the canadian life insurance industry (Tech. Rep.). Society of Actuaries and Canadian Institute of Actuaries.
Stoer, J., & Bulirsch, R. (2002). Introduction to numerical analysis (3rd ed.). New York: Springer.
Sun, Y., Cuesta-Infante, A., & Veeramachaneni, K. (2019). Learning vine copula models for synthetic data generation. arXiv preprint arXiv:1812.01226.
Sundarkumar, R., & Siddeshwar, V. (2015). One-class support vector machine based undersampling: Application to churn prediction and insurance fraud detection.
Székely, G. J., & Rizzo, M. L. (2013). Energy statistics: A class of statistics based on distances. Journal of Statistical Planning and Inference, 143(8), 1249-1272.
Tadeusiewicz, R. (2009). Sieci neuronowe. Kraków: Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie.
Taylor, J. W. (2008). Estimating value at risk and expected shortfall using expectiles. Journal of Financial Econometrics, 6(2), 231–252.
Tchen, A. H. (1980). Inequalities for distributions with given marginals. Annals of Probability, 8(4), 814–827.
Thimm, G., & Fiesler, E. (1996). High-order and multilayer perceptron initialization. In Proceedings of the international conference on neural networks (pp. 331–336). IEEE.
Tieleman, T., & Hinton, G. (2012). Lecture 6.5: Rmsprop, divide the gradient by a running average of its recent magnitude. Coursera Lecture Notes / Technical report.
Traub, J. F. (1964). Iterative methods for the solution of equations. Englewood Cliffs, NJ: Prentice-Hall.
Varadarajan, A., & Others. (2024). Evaluation of risk level assessment: A review. International Journal of Engineering Research Technology, 13(2), 45–53.
Venter, G. G. (2009). Risk measures and capital allocation. North American Actuarial Journal, 13(3), 1–25.

Wanat, S. (2014). Efekt dywersyfikacji ryzyka w solvency ii w świetle wyników ilościowego badania wpływu qis5. Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu(371), 320–330.

Weidner, A., & Wüthrich, M. V. (2017). Telematics car driving behavior data in insurance: clustering and dimension reduction. ASTIN Bulletin, 47(2), 453–485.

Whitney, A. W. (1971). A direct method of nonparametric measurement selection. IEEE Transactions on Computers, 20(9), 1100–1103. doi: 10.1109/T-C.1971.223410

Wüthrich, M. V. (2018a). Neural networks applied to chain-ladder reserving. European Actuarial Journal, 8(2), 407–436.

Wüthrich, M. V. (2018b). Neural networks applied to telematics car driving behavior data. European Actuarial Journal, 8(2), 407–429.

Wüthrich, M. V. (2019). Yes, we cann! Available at SSRN.

Wüthrich, M. V., & Merz, M. (2019). Nesting classical actuarial models into neural networks. Available at SSRN.

Wüthrich, M. V. (2018). Neural networks applied to chain–ladder reserving. European Actuarial Journal, 8(2), 407–436.

Zeisberger, C., & Chen, A. (2014). Jpmorgan and the london whale. (Case ID: 6003, INSEAD)

Zeissler, A. G., & Metrick, A. (2019). Jpmorgan chase london whale d: Risk-management practices. Journal of Financial Crises, 2(1), 1–27. (Case study analysis of the London Whale loss)

Ziel, F., & Berg, K. (2019). Probabilistic forecasting of real-time electricity prices: The case of the german intraday market. Applied Energy, 238, 1687–1701.