



Stanisław Heilpern

Uniwersytet Ekonomiczny we Wrocławiu
Wydział Zarządzania, Informatyki i Finansów
Katedra Statystyki
stanislaw.heilpern@ue.wroc.pl

ANALIZA WIELOWYMIAROWEJ STRUKTURY ZALEŻNOŚCI – ZASTOSOWANIE W RODZINNYCH UBEZPIECZENIACH NA ŻYCIE¹

Streszczenie: Praca dotyczy modelowania wielowymiarowej struktury zależności, gdy liczba wymiarów jest większa niż dwa. Do modelowania wykorzystano hierarchiczne archimedesowe funkcje łączące (HAC): całkowicie i częściowo zagnieżdżone. Jest to proste uogólnienie klasycznych archimedesowych funkcji łączących, które nie sprawdzają się w przypadku większej liczby wymiarów. Omówiono metody dopasowania funkcji łączących do danych i metody estymacji ich parametrów oraz przedstawiono przykładową konstrukcję HAC. Ostatnia część pracy dotyczy rozszerzonej wersji renty wdowiej. Do wyznaczenia aktuarialnej wartości renty wykorzystano HAC.

Słowa kluczowe: renta, wielowymiarowa zależność, funkcja łącząca.

JEL Classification: C30, G22, J10.

Wprowadzenie

Wielowymiarowa struktura zależności, w przypadku gdy liczba wymiarów jest większa niż dwa, jest istotnie różna od dwuwymiarowej struktury [Heilpern, 2014]. Archimedesowe funkcje łączące dość dobrze modelują dwuwymiarową strukturę zależności. Ze względu na dostatecznie prostą postać często są wykorzystywane w praktycznych zastosowaniach. Dla większej liczby wymiarów ich

¹ Praca naukowa finansowana ze środków na naukę w latach 2013-2016 jako projekt badawczy nr DEC-2013/09/B/HS4/00490.

zastosowanie jest jednak znacznie ograniczone. Głównie przez fakt, że modelują one jedynie przypadek jednakowo skorelowanych par zmiennych brzegowych.

W pracy przedstawiono rozszerzenie klasycznych archimedesowych funkcji łączących do ich hierarchicznej wersji (HAC). Dzięki temu został istotnie poszerzony zakres zastosowań tego typu funkcji łączących. Omówione zostały całkowicie i częściowo zagnieżdżone archimedesowe funkcje łączące. Przedstawiono metody dopasowania funkcji łączących do danych i estymacji ich parametrów oraz przykładową konstrukcją HAC.

Ostatnia część pracy dotyczy rozszerzonej wersji renty wdowiej, szczególnego przypadku ubezpieczenia grupowego, uwzględniającej opiekę nad dzieckiem. Struktura zależności została opisana trójwymiarową HAC oraz wykorzystana do wyznaczenia wartości aktuarialnej rozpatrywanej renty.

1. Wielowymiarowe funkcje łączące

Funkcja łącząca (*copula*) C jest łącznikiem między dystrybuantą rozkładu łącznego F a dystrybuantami brzegowymi F_i wektora losowego $\mathbf{X} = (X_1, \dots, X_d)$ [Nelsen, 2006; Heilpern, 2007]:

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)).$$

Gdy dystrybuanty są ciągłe, to funkcja łącząca jest wyznaczona jednoznacznie. W podobny sposób określamy funkcję łączącą przeżycia C^* :

$$S(x_1, \dots, x_d) = C^*(S_1(x_1), \dots, S_d(x_d)),$$

gdzie:

S – jest łączną,

$S_i(x) = P(X_i > x)$ – są brzegowymi funkcjami przeżycia.

Funkcja łącząca przeżycia ma te same własności jak zwykła funkcja łącząca.

W przypadku dwóch wymiarów, znając funkcję łączącą, można w prosty sposób wyznaczyć funkcję łączącą przeżycia:

$$C^*(u_1, u_2) = u_1 + u_2 - 1 + C(1 - u_1, 1 - u_2).$$

Dla większej liczby wymiarów wzór jest już bardziej skomplikowany.

W zagadnieniach aktuarialnych, zwłaszcza dotyczących ubezpieczeń na życie, częściej wykorzystuje się funkcje przeżycia, a nie dystrybuanty. Funkcje łączące przeżycia znajdują wtedy szersze zastosowanie.

Funkcja przeżycia, jak i funkcja łącząca przeżycia, jest dystrybuantą zmiennej losowej skupionej na kostce $[0, 1]^d$, o jednostajnych rozkładach brzegowych. Zachodzi bowiem zależność:

$$C(\mathbf{u}) = P(U_1 \leq u_1, \dots, U_d \leq u_d),$$

gdzie:

$$\mathbf{u} = (u_1, \dots, u_d), \text{ a } U_i = F_i(X_i).$$

W przypadku gdy zmienne losowe są niezależne, odpowiadająca im funkcja łącząca przyjmuje postać:

$$\Pi(\mathbf{u}) = u_1 \cdot \dots \cdot u_d.$$

Każda funkcja łącząca C jest ograniczona z dołu, jak i z góry przez tzw. granice Frecheta-Hoeffdinga:

$$W(\mathbf{u}) \leq C(\mathbf{u}) \leq M(\mathbf{u}).$$

Górna granica $M(\mathbf{u}) = \min(u_1, \dots, u_d)$ jest funkcją łączącą opisującą strukturę zależności indukowaną przez ściśle zależne dodatnio zmienne losowe, czyli zmienne współmonotoniczne. Natomiast dolna granica $W(\mathbf{u}) = \max(u_1 + \dots + u_d - n + 1, 0)$ nie jest dla wymiaru $d > 2$ funkcją łączącą. W przypadku dwóch wymiarów W jest funkcją łączącą odpowiadającą ściśle zależnym zmiennym losowym, które są ujemnie skorelowane, czyli przeciwnotoniczne. Już dla trzech wymiarów nie istnieją jednak trzy zmienne losowe wzajemnie przeciwnotoniczne.

Innym przykładem funkcji łączącej jest eliptyczna funkcja łącząca. Spełnia ona zależność:

$$C(\mathbf{u}) = \Phi(\Phi_1^{-1}(u_1), \dots, \Phi_d^{-1}(u_d)),$$

gdzie:

Φ – jest dystrybuantą d -wymiarowego rozkładu eliptycznego, np. normalnego, t -Studenta, czy logistycznego,

Φ_i – jednowymiarową dystrybuantą eliptyczną.

W przeciwieństwie do eliptycznych funkcji łączących, archimedesowe funkcje łączące (AC) można przedstawić jawnym, w miarę prostym wzorem:

$$C(\mathbf{u}) = \varphi^1(\varphi(u_1) + \dots + \varphi(u_d)),$$

gdzie:

generator $\varphi: [0, 1] \rightarrow [0, \infty]$ spełnia następujące warunki: $\varphi(0) = \infty$, $\varphi(1) = 0$, φ^{-1} jest funkcją całkowicie monotoniczną, czyli dla każdego $k = 0, 1, \dots$, zachodzi:

$$(-1)^k \frac{d^k}{dt^k} \varphi^{-1}(t) \geq 0.$$

W przypadku dwóch zmiennych wystarczy, że φ jest funkcją ciągłą, malejącą i wypukłą.

Znając postać generatora $\varphi_C(u)$ archimedesowej funkcji łączącej C , możemy w prosty sposób obliczyć miarę zależności, np. wartość współczynnika korelacji rang Kendalla między parami brzegowych zmiennych losowych:

$$\tau(C) = 4 \int_0^1 \frac{\varphi_C(t)}{\varphi_C'(t)} dt + 1.$$

Niestety dla każdej pary zmiennych losowych otrzymujemy tę samą wartość współczynnika. Zmienne te są jednakowo skorelowane, zawiąza to znacznie pole praktycznego zastosowania AC w przypadku większej niż 2 liczby wymiarów. Można je ponadto stosować jedynie dla zmiennych losowych dodatnio skorelowanych, wynika to z warunku dotyczącego całkowitej monotoniczności funkcji odwrotnej do generatora [Nelsen, 2006; Heilpern, 2007].

W praktyce wykorzystuje się rodziny archimedesowych funkcji łączących indeksowanych parametrami. W tabeli 1 podane są najczęściej wykorzystywane rodziny tych funkcji. Przedstawiono ich wzór, zakres wartości parametrów oraz wartość odpowiadającego współczynnika korelacji Kendalla.

Tabela 1. Wybrane rodziny archimedesowych funkcji łączących

Rodzina	Wzór	Parametr	Współczynnik Kendalla
Claytona	$(u_1^{-\theta} + \dots + u_d^{-\theta} - d + 1)^{-1/\theta}$	$\theta > 0$	$\tau = \frac{\theta}{\theta + 2}$
Gumbela	$\exp(-((-\ln u_1)^\theta + \dots + (-\ln u_d)^\theta)^{1/\theta})$	$\theta \geq 1$	$\tau = 1 - \frac{1}{\theta}$
Franka	$-\frac{1}{\theta} \ln \left(1 + \frac{(e^{-\theta u_1} - 1) \dots (e^{-\theta u_d} - 1)}{(e^{-\theta} - 1)^{d-1}} \right)$	$\theta > 0$	$\tau = 1 - \frac{4}{\theta} \left(1 - \frac{1}{\theta} \int_0^\theta \frac{t}{e^t - 1} dt \right)$
AMH	$\frac{\prod_{j=1}^d u_j}{1 - \theta \prod_{j=1}^d (1 - u_j)}$	$0 \leq \theta \leq 1$	$\tau = \frac{3\theta - 2}{3\theta} - \frac{2(1 - \theta)^2}{3\theta^2} \ln(1 - \theta)$

Źródło: Opracowanie własne.

Na podstawie próby $\mathbf{X}_i = (X_{i1}, \dots, X_{id})$, $1 \leq i \leq n$, możemy skonstruować empiryczną funkcję łączącą \hat{C}_n , opisaną wzorem:

$$\hat{C}_n(\mathbf{u}) = \frac{1}{n} \sum_{i=1}^n \prod_{j=1}^d \mathbf{1}_{\{\hat{U}_{ij,n} \leq u_j\}},$$

gdzie:

pseudoobserwacje $\hat{U}_{ij} = \hat{F}_j(X_{ij}) = \frac{1}{n+1}(R_{ij})$, $\hat{F}_j(X_{ij})$ są empirycznymi dystrybuantami brzegowymi,

R_{ij} są rangami X_{ij} w X_{1j}, \dots, X_{nj} .

2. Hierarchiczne archimedesowe funkcje łączące

Definicja hierarchicznych archimedesowych funkcji łączących (HAC) jest indukcyjna. Są to archimedesowe funkcje łączące o argumentach będących wartościami innych HAC [Hofert, 2011]. W odróżnieniu od klasycznych archimedesowych funkcji łączących mają one $d - 1$ parametrów.

Spośród HAC możemy wyróżnić całkowicie zagnieżdżone AC (FNAC), określone formułą:

$$\begin{aligned} C_{d-1}(\mathbf{u}) &= \varphi_{d-1}^{-1} \left(\varphi_{d-1}(u_d) + C_{d-2}(u_1 + \dots + u_{d-1}) \right) \quad (1) \\ &= \varphi_{d-1}^{-1} \left(\varphi_{d-1}(u_d) + \varphi_{d-1} \left(\varphi_{d-2}^{-1} \left(\dots \varphi_1^{-1} \left(\varphi_1(u_1) + \varphi_1(u_2) \right) \dots \right) \right) \right), \end{aligned}$$

gdzie:

C_i – jest FNCA z i parametrami,

φ_i – są generatorami dwuwymiarowej AC.

Funkcje φ_i nie mogą być dowolnymi generatorami AC. Można pokazać, że aby (1) była funkcją łączącą, to złożenie $\varphi_{i+1} \circ \varphi_i^{-1}$ musi być funkcją całkowicie monotoniczną [McNeil, 2008]. Fakt ten ma głębokie konsekwencje. Na ogół wszystkie archimedesowe funkcje łączące C_i tworzące FNCA muszą należeć do tej samej rodziny. Rzadko kiedy można łączyć różne rodziny. Przykładowo, generatory φ_{i+1} , φ_i należące do rodzin Clayтона i Gumbela nigdy nie będą spełniać wspomnianego warunku. Ale dla rodziny AMH i Clayтона, gdy parametr θ rodziny Clayтона spełnia nierówność $\theta \geq 1$, to warunek ten jest spełniony. Więcej informacji na ten temat czytelnik może znaleźć w pracach [Hofert, 2008; McNeil, 2008].

Ponadto współczynniki korelacji rang Kendalla τ_i , odpowiadające archimedesowemu funkcją łączącym C_i , muszą spełniać warunek:

$$\tau_1 \geq \tau_2 \geq \dots \geq \tau_{d-1}.$$

Zwykle większa wartość parametru AC odpowiada większemu stopniu zależności, tak że muszą zachodzić wtedy zależności:

$$\theta_1 \geq \theta_2 \geq \dots \geq \theta_{d-1}.$$

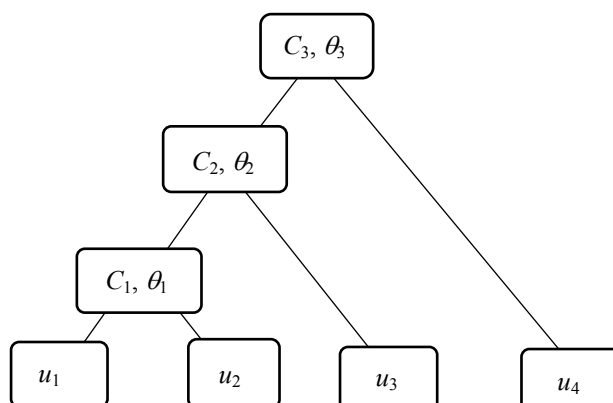
Pewną modyfikacją FNAC są częściowo zagnieżdżone AC (PNAC) [McNeil, Frey i Embrechts, 2006]. Jest to złożenie klasycznej AC z FNAC. Przykładowo dla czterech wymiarów przyjmuje ona postać:

$$C(u_1, u_2, u_3, u_4) = C_3(C_1(u_1, u_2), C_1(u_3, u_4)).$$

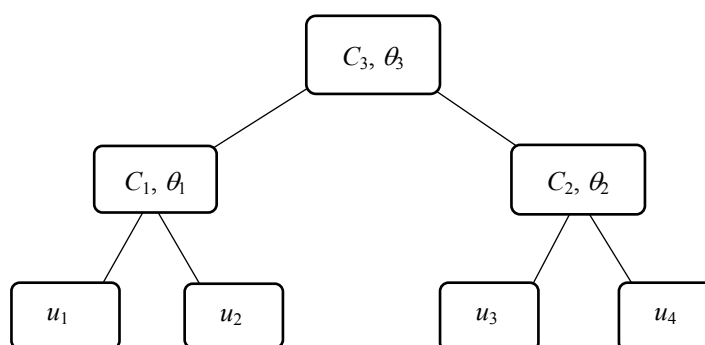
Struktury tak określonej PNAC oraz FNAC danej wzorem:

$$C(u_1, u_2, u_3, u_4) = C_3(C_2(C_1(u_1, u_2), u_3) u_4)$$

są przedstawione graficznie na rys. 1. W tym przypadku składowe, dwuwymiarowe archimedesowe funkcje łączące C_i , również muszą spełniać własności takie jak FNAC.



a) całkowicie zagnieżdżona AC



b) częściowo zagnieżdżona AC

Rys. 1. Hierarchiczne archimedesowe funkcje łączące

Źródło: Aas i Berg [2009].

3. Wybór funkcji łączącej

Ważnym etapem analizy struktury zależności jest wybór odpowiedniej funkcji łączącej. Powinna ona być jak najlepiej „dopasowana” do danych empirycznych. W niniejszej pracy zaproponowano następującą procedurę wyboru takiej funkcji łączącej:

- ustalenie potencjalnych rodzin funkcji łączących,
- estymacja parametrów tych rodzin,
- wybór najlepszej rodziny według ustalonego kryterium.

3.1. Metody estymacji parametrów

Spośród wszystkich metod estymacji parametrów funkcji łączącej można wyróżnić dwie najważniejsze: metodę największej wiarygodności oraz momentów, opartą na miarach zależności. Metoda największej wiarygodności polega na wyborze wektora parametrów $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$ maksymalizującego funkcję wiarygodności. Są różne warianty postaci tej funkcji [Okhrin, 2009]. Jedna z nich przyjmuje następującą postać:

$$l(\boldsymbol{\theta}) = \sum_{i=1}^n \log c(\hat{F}_1(X_{i1}), \dots, \hat{F}_d(X_{id}), \boldsymbol{\theta}),$$

gdzie:

c – jest funkcją gęstości funkcji łączącej,

\hat{F}_j – brzegową dystrybuantą empiryczną.

W niniejszej pracy wybrano metodę momentów. Jest to prosta metoda, polegająca na wyborze takiego parametru, czy wektora parametrów, dla którego odpowiadająca jemu miara zależności jest równa empirycznej wartości tej miary. Na przykład, dla rodziny Claytona estymator parametru θ powinien być równy:

$$\hat{\theta} = \frac{2\hat{\tau}}{1 - \hat{\tau}},$$

gdzie: $\hat{\tau}$ jest empiryczną wersją współczynnika korelacji Kendalla (patrz tabela 1). Parametry eliptycznych funkcji łączących są wprost współczynnikami korelacji Pearsona.

3.2. Kryteria wyboru rodziny

Jednym z kryteriów wyboru „najlepiej” dopasowanej do danych empirycznych funkcji łączącej jest maksymalizacja funkcji wiarygodności [Okhrin i Ristig, 2014]. Dla każdej rodziny funkcji łączących estymujemy parametry, a następnie wybieramy tę rodzinę, dla której dla wyestymowanego parametru funkcja wiarygodności przyjmuje największą wartość.

Można też zastosować kryterium oparte na statystyce Cramera-von Misesa, preferujące funkcję łączącą „najlepiej” dopasowaną do jej empirycznej wersji [Aas i Berg, 2009]:

$$\begin{aligned} S_n &= n \int_{[0,1]^d} (\hat{C}_n(\mathbf{u}) - C_\theta(\mathbf{u}))^2 dC_n(\mathbf{u}) \\ &= \sum_{i=1}^n (\hat{C}_n(\mathbf{u}_i) - C_\theta(\mathbf{u}_i))^2, \end{aligned}$$

gdzie:

$\mathbf{u}_i = \frac{n}{n+1}(\hat{F}_1(x_{i1}), \dots, \hat{F}_d(x_{id})) = \frac{1}{n+1}(R_{i1}, \dots, R_{id})$ są pseudoobserwacjami,

\hat{C}_n – empiryczną funkcją łączącą.

Wybieramy wtedy rodzinę minimalizującą statystykę S_n .

W kryterium S_n suma przebiega po wszystkich pseudoobserwacjach \mathbf{u}_i . Kryterium to można rozszerzyć sumując po kracie wartości z całej kostki $[0, 1]^d$ [Nelsen, 2006]:

$$L = \left\{ \left(\frac{i_1}{m}, \dots, \frac{i_d}{m} : i_k = 0, 1, \dots, m \right) \right\}.$$

Dla większej liczby wymiarów możemy jednak wtedy natrafić na trudności natury numerycznej.

3.3. Przykład konstrukcji hierarchicznej archimedesowej funkcji łączącej

Poniżej przedstawiono przykład pokazujący konstrukcję hierarchicznej archimedesowej funkcji łączącej. Konstrukcja ta jest oparta na propozycji Góreckiego i Hoferta [2014]. Przykład jest czysto akademicki, wykorzystujący rzeczywiste dane dotyczące notowań pięciu spółek: BZWBK, Pekao, Handlowy, Ciech i Budimex w okresie od 2.01.2015 r. do 14.07.2015 r.

Konstrukcja HAC jest oparta na macierzy korelacji Kendalla. Macierz tę przedstawiono w tabeli 2.

Tabela 2. Macierz korelacji Kendalla

τ	BZWBK	Pekao	Handlowy	Ciech	Budimex
BZWBK	1	0,26520	0,20201	0,16145	0,14400
Pekao		1	0,23073	0,09080	0,07700
Handlowy			1	0,11316	0,12591
Ciech				1	0,21938
Budimex					1

Źródło: Opracowanie własne.

Największy współczynnik korelacji Kendalla, równy 0,2652, jest między spółkami BZWBK i Pekao. Wybieramy więc jako C_1 archimedesową funkcję łączącą opisującą zależność między tymi spółkami. Następnie wybieramy największy zagregowany współczynnik Kendalla między spółkami Handlowy, Ciech, Budimex oraz zbiorem $BP = \{BZWBK, Pekao\}$. Zagregowany współczynnik korelacji Kendalla między zbiorami $A = \{u_{i_1}, \dots, u_{i_l}\}$ i $B = \{u_{j_1}, u_{j_q}\}$ jest wyznaczony formułą:

$$\tau(A, B) = h\left(\tau(u_{i_1}, u_{j_1}), \tau(u_{i_1}, u_{j_2}), \dots, \tau(u_{i_q}, u_{j_q})\right),$$

gdzie:

funkcja agregująca h jest niezmiennicza ze względu na permutację argumentów, $h(u, \dots, u) = u$.

W naszym przypadku wybraliśmy średnią arytmetyczną. Przykładowo, $\tau(\text{Handlowy, BP}) = (0,20201 + 0,23073)/2 = 0,21637$. Największy współczynnik Kendalla jest jednak między spółkami Ciech i Budimex i wynosi 0,21938. W następnym kroku porównujemy zagregowane współczynniki Kendalla między Handlowy, BP oraz CB, gdzie $CB = \{\text{Ciech, Budimex}\}$. W rezultacie otrzymujemy następującą PNAC:

$$C(u_1, u_2, u_3, u_4, u_5) = C_4(C_3(C_1(u_1, u_2), u_3), C_2(u_4, u_5)).$$

Stosując kryterium oparte na minimalizacji funkcji S_n , najmniejszą wartość – równą 0,0290 – otrzymamy dla PNAC Claytona. Parametry poszczególnych składowych funkcji łączących C_i wynoszą odpowiednio:

$$\theta_1 = 0,7218; \quad \theta_2 = 0,5621; \quad \theta_3 = 0,5522; \quad \theta_4 = 0,2694.$$

Wartość funkcji wiarygodności jest natomiast równa $l = 34,7390$.

Tabela 3. Optymalne wartości kryteriów dla różnych rodzin PNAC i AC

	S_n	l	AC S_n
Clayton	0,0230	35,5577	0,05939
	0,0232	35,5637	
Gumbel	0,0351	37,0104	0,04403
	0,0361	37,6197	
Frank	0,0312	30,0013	0,05210
	0,0319	30,0180	
AMH	0,0275	32,2448	0,15837
	0,0279	32,2856	

Źródło: Opracowanie własne.

Powyższe dopasowanie funkcji łączącej do danych można trochę poprawić, stosując zamiast średniej arytmetycznej jako funkcji agregującej h , kombinację liniową funkcji max i min. Wtedy dla PNAC Claytona i parametrów $\theta_3 = 0,5063$ oraz $\theta_4 = 0,1832$ wartość kryterium $S_n = 0,0230$ jest trochę mniejsza, a funkcja wiarygodności $l = 35,5577$ nieco większa. Dla parametrów $\theta_3 = 0,5063$ oraz $\theta_4 = 0,1966$ otrzymamy z kolei największą wartość funkcji wiarygodności $l = 35,5637$, ale za to większą wartość kryterium $l = 0,0232$. W tabeli 3 podano wartości optymalnych wartości kryteriów dla różnych rodzin PNAC i AC.

Podsumowując, PNAC Clayтона o parametrach:

$$\theta_1 = 0,7218; \quad \theta_2 = 0,5621; \quad \theta_3 = 0,5063; \quad \theta_4 = 0,1832$$

minimalizuje wartość kryterium $S_i = 0,0230$, natomiast największą funkcję wiarygodności $l = 37,6197$ daje PNAC Gumbela z parametrami:

$$\theta_1 = 1,3609; \quad \theta_2 = 1,2810; \quad \theta_3 = 1,2999; \quad \theta_4 = 1,0834.$$

Na pytanie, którą funkcję łączącą wybrać, nie ma w tym przypadku jednoznacznej odpowiedzi. Wybór zależy od subiektywnych preferencji prowadzącego badania. Należy też zauważyć, że klasyczne AC są zdecydowanie gorsze, dają znacznie większą wartość funkcji S_n .

4. Rozszerzona renta wdowia

Rozpatrzmy następującą rentę [Rakkolainen, 2011], będącą rozszerzeniem renty wdowiej. Załóżmy, że mąż ma x lat, żona y , a syn jest w wieku z . Przyjmijmy, że po śmierci męża wdowa dostaje 1 jednostkę pieniężną, a syn dostaje 0,2 jednostki. Gdy wdowa umrze, to syn dostaje 0,6 jednostki. Ponadto syn dostaje rentę jedynie do wieku $z + n$, np. na czas nauki. Przedstawiają one pewną propozycję umowy między firmą ubezpieczeniową a klientem.

Niech X będzie czasem życia męża, Y czasem życia żony, a Z czasem życia syna. Wtedy

$$S(x, y, z) = P(X > x, Y > y, Z > z)$$

jest łączną funkcją przeżycia, a

$$S_X(x) = P(X > x) = {}_x p_0^M, \quad S_Y(y) = P(Y > y) = {}_y p_0^K, \quad S_Z(z) = P(Z > z) = {}_z p_0^M$$

brzegowymi funkcjami przeżycia, które możemy wyznaczyć na podstawie tablic trwania życia. Wtedy:

$$S(x, y, z) = C(S_X(x), S_Y(y), S_Z(z)),$$

gdzie C jest funkcją łączącą przeżycia.

Wartość aktuarialna naszej renty wynosi:

$$a_{x|yz} = \sum_{k=1}^n v^k (1,2 {}_k \rho_{011} + {}_k \rho_{010} + 0,6 {}_k \rho_{001}) + \sum_{k=n+1}^w v^k {}_k \rho_{01},$$

gdzie:

$$v = 1/(1 + \xi), \quad \xi \text{ jest techniczną stopą zwrotu oraz}$$

$${}_k \rho_{011} = P(X \leq x + k, Y > y + k, Z > z + k | X > x, Y > y, Z > z),$$

$${}_k \rho_{010} = P(X \leq x + k, Y > y + k, Z \leq z + k | X > x, Y > y, Z > z),$$

$${}_k \rho_{001} = P(X \leq x + k, Y \leq y + k, Z > z + k | X > x, Y > y, Z > z),$$

$${}_k \rho_{01} = P(X \leq x + k, Y > y + k | X > x, Y > y, Z > z).$$

Prawdopodobieństwo ${}_k\rho_{011}$ jest równe:

$$\begin{aligned} {}_k\rho_{011} &= P(X \leq x+k, Y > y+k, Z > z+k | X > x, Y > y, Z > z) \\ &= \frac{P(x < X \leq x+k, Y > y+k, Z > z+k)}{P(X > x, Y > y, Z > z)} \\ &= \frac{P(X > x, Y > y+k, Z > z+k) - P(X > x+k, Y > y+k, Z > z+k)}{P(X > x, Y > y, Z > z) - P(X > x+k, Y > y, Z > z+k)} \\ &= \frac{S(x, y+k, z+k) - S(x+k, y+k, z+k)}{S(x, y, z)}. \end{aligned}$$

W podobny sposób wyznaczamy pozostałe prawdopodobieństwa:

$$\begin{aligned} {}_k\rho_{010} &= \frac{S(x, y+k, z) - S(x+k, y+k, z) - S(x, y+k, z+k)}{S(x, y, z)} \\ &\quad + \frac{S(x+k, y+k, z+k)}{S(x, y, z)}, \\ {}_k\rho_{001} &= \frac{S(x, y, z+k) - S(x+k, y, z+k) - S(x, y+k, z+k)}{S(x, y, z)} \\ &\quad + \frac{S(x+k, y+k, z+k)}{S(x, y, z)}, \\ {}_k\rho_{01} &= \frac{S(x, y+k, z) - S(x+k, y, z, z)}{S(x, y, z)}. \end{aligned}$$

Oznaczmy symbolami r_{XY} , r_{XZ} oraz r_{YZ} współczynniki korelacji Kendalla między poszczególnymi zmiennymi. Ponadto niech:

$$r_Z = \frac{r_{XZ} + r_{YZ}}{2}, r_X = \frac{r_{XY} + r_{XZ}}{2}, r_Y = \frac{r_{XY} + r_{YZ}}{2}.$$

Rozpatrzmy dwa przypadki struktur zależności, gdy największy jest współczynnik r_{XY} oraz r_{XZ} . Pierwszy odpowiada sytuacji, gdy jest duża więź między małżonkami, żyją w podobnych warunkach, czy narażeni są na te same ryzyka. Drugi natomiast zachodzi, gdy syn ma podobne geny jak ojciec, może mieć te same dziedziczne choroby, a mniejsza jest w tym przypadku zależność genetyczna między matką a synem.

Gdy największy jest współczynnik korelacji Kendalla r_{XY} , to struktura zależności przyjmuje postać:

$$S(x, y, z) = C_2(C_1(xp_0^M, yp_0^K), zp_0^M).$$

W drugim przypadku otrzymujemy natomiast:

$$S(x, y, z) = C_2(C_1(xp_0^M, zp_0^M), yp_0^K).$$

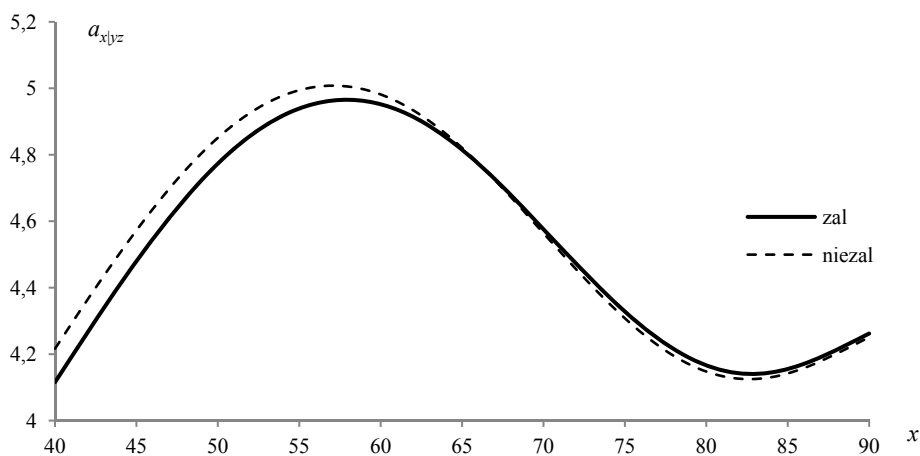
Tabela 4. Wartości renty $a_{x|yz}$ dla różnych r_{XY}

r_{XY}	$a_{x yz}$	r_{XY}	$a_{x yz}$
0	4,15434	0,12	4,06151
0,02	4,13886	0,14	4,04604
0,04	4,12338	0,16	3,9498
0,06	4,10791	0,18	3,91685
0,08	4,09244	0,2	3,88389
0,1	4,07698		

Źródło: Opracowanie własne.

Założmy teraz, że małżonkowie mają po 40 lat, a ich syn 15 lat i będzie otrzymywać ewentualną rentę do 25 roku życia, tzn. $x = y = 40$ oraz $z = 15$ i $n = 10$. Przyjmujemy ponadto, że techniczna stopa zwrotu jest stała i wynosi $\xi = 0,03$, czyli $v = 0,97087$, oraz że jest większa zależność między długością życia męża i syna niż żony i syna: $r_{XZ} = 0,15$ i $r_{YZ} = 0,05$. W tabeli 4 znajdują się wartości renty $a_{x|yz}$ dla różnych stopni zależności między długością życia małżonków, czyli współczynnika korelacji Kendalla r_{XY} , wyrażone w umownych jednostkach pieniężnych. Widzimy, że wraz ze wzrostem stopnia zależności długości życia małżonków r_{XY} maleje wartość renty $a_{x|yz}$.

Wartości renty $a_{x|yz}$ w zależności od wieku małżonków są przedstawione na rys. 2, gdzie założono, że $x = y$, $r_{XZ} = 0,15$, $r_{YZ} = 0,05$, $r_{XY} = 0,1$, $z = 15$ oraz $n = 10$. Ponadto rozpatrzono osobno przypadek dotyczący niezależności długości życia wszystkich trzech osób.



Rys. 2. Wartości renty $a_{x|yz}$ w zależności od wieku małżonków

Źródło: Opracowanie własne.

Można zauważyć pewną cykliczność wykresu wartości renty $a_{x|yz}$ w zależności od wieku małżonków. Naprzód wartość renty rośnie wraz ze wzrostem ich wieku, następnie po 58 roku maleje. Na uwagę zasługuje kolejny wzrost wartości renty po 82 latach. Jest to wiek czysto hipotetyczny w przypadku rodziców, ale może on przykładowo dotyczyć dziadków, którzy opiekują się wnukiem. Również zmienna jest zależność wielkości renty $a_{x|yz}$ od wieku małżonków i przypadku niezależności. Do 67 roku dla zależnych długości życia wartość renty jest mniejsza, a dla starszych osób jest większa niż dla niezależnych długości życia.

Podsumowanie

Praca dotyczyła analizy wielowymiarowej struktury zależności. Zwrócono głównie uwagę na przypadek większej niż dwa liczby wymiarów. Ma to szczególne znaczenie, gdy struktura ta jest opisana archimedesową funkcją łączącą, która ze względu na prostotę często jest wykorzystywana w praktyce. W przypadku większej liczby wymiarów jest jednak wtedy wymagana jednakowa korelacja między parami zmiennych losowych, co jest rzadko spotykane w praktyce.

W tym celu zaproponowano zastosowanie pewnego uogólnienia tych funkcji łączących: hierarchicznych funkcji łączących (HAC), całkowicie i częściowo zagnieżdżonych. Nie mają one tych wad, co klasyczne archimedesowe funkcje łączące, i są lepiej dopasowane do empirycznych danych. W pracy przedstawiono przykładową konstrukcję częściowo zagnieżdżonej HAC. Przeprowadzono też analizę rozszerzonej wersji renty wdowiej. Do wyznaczenia aktuarialnej wartości tej renty wykorzystano HAC. Zbadano też wpływ stopnia zależności długości życia osób oraz ich wieku na wartość tej renty.

Literatura

- Aas K., Berg D. (2009), *Models for Construction of Multivariate Dependence – A Comparison Study*, „The European Journal of Finance”, Vol. 15(7).
- Górecki J., Holena M. (2014), *Structure Determination and Estimation of Hierarchical Archimedean Copulas Based on Kendall Correlation Matrix* [in:] A. Appice, M. Ceci, C. Loglisci, G. Manco, E. Masciari, Z.W. Ras (eds.), *New Frontiers in Mining Complex Patterns*, Lecture Notes in Computer Science, Springer, New York.
- Heilpern S. (2007), *Funkcje łączące*, Wydawnictwo AE, Wrocław.
- Heilpern S. (2014), *Multivariate Measures of Dependence Based on Copulas*, „Mathematical Economics”, No. 10(17).

- Hofert M. (2008), *Sampling Archimedean copulas*, „Computational Statistics & Data Analysis”, Vol. 52(12).
- Hofert M. (2011), *Efficiently Sampling Nested Archimedean Copulas*, „Computational Statistics and Data Analysis”, Vol. 55(1).
- McNeil A.J. (2008), *Sampling Nested Archimedean Copulas*, „Journal Statistical Computation and Simulation”, Vol. 78(6).
- McNeil A.J., Frey R., Embrechts P. (2006), *Quantitative Risk Management: Concepts, Techniques and Tools*, Priceton University Press, Priceton.
- Nelsen R.B. (2006), *An Introduction to Copulas* (II ed.), Springer, New York.
- Okhrin O. (2009), *Fitting high-dimensional. Copulae to Data*, <http://edoc.hu-berlin.de/series/sfb-649-papers/2010-22/PDF/22.pdf> (dostęp: 27.11.2015).
- Okhrin R., Ristig A. (2014), *Hierarchical Archimedean Copulae: The HAC Package*, „Journal Statistical Software”, Vol. 58(4).
- Rakkolainen T. (2011), *Insurance Mathematics*, <http://web.abo.fi/fak/mmf/mate/kurser/forsa-kring/insurancemath.pdf> (dostęp: 27.11.2015).

ANALYSIS OF THE MULTIVARIATE DEPENDENCE STRUCTURE – APPLICATION IN THE FAMILY LIFE INSURANCE

Summary: The paper is devoted to the modeling of the multivariate dependent structure, where the number of dimension is greater than two. The hierarchical Archimedean copulas (HAC) are using to this end: the fully and partially nested. This is the simple generalization of the classical Archimedean copulas, which do not work in the multi-dimensional case. The methods of fitting the copulas to the empirical data, the methods of estimation of the parameters of copulas and the exemplary constructions of HAC are investigated. The last part of the paper is devoted to the expanded version of a widow's pension. The HAC is used to derive actuarial value of such pension.

Keywords: actuarial pension, multivariate dependence, copula.