

Joanna Małgorzata Landmesser

Szkoła Główna Gospodarstwa Wiejskiego w Warszawie

**ZASTOSOWANIE FUNKCJI KOPULA W MODELOWANIU
WIELOWYMIAROWYCH CZASÓW TRWANIA
NA PRZYKŁADZIE DECYZJI PROKREACYJNYCH
AKTYWNYCH ZAWODOWO KOBIEC**

1. Wstęp

Funkcje kopula są użytecznym narzędziem pomocnym w modelowaniu rozkładów łącznych, gdy znane są rozkłady brzegowe; w szczególności, gdy rozkłady brzegowe nie są normalne. Słowa „kopula” po raz pierwszy w statystycznym sensie użył Sklar w roku 1959 [Sklar 1959]. W ostatnich dwudziestu latach doszło do intensywnego rozwoju zarówno teorii dotyczących kopuli, jak i ich zastosowań. Powstało kilka monografii poświęconych tej tematyce, np. [Joe 1997; Nelsen 1999]. Frees i Valdez [Frees, Valdez 1998] wskazują na przydatność kopuli w naukach aktuarialnych. Cherubini i in. [Cherubini, Luciano, Vecchiato 2004] skupiają się na zastosowaniach w analizach finansowych. Liczne badania dotyczą zależności między zmiennymi dyskretnymi, np. [Cameron, Trivedi, Zimmer 2004]. Ostatnio modele kopul są coraz częściej stosowane w modelowaniu wielowymiarowych czasów trwania. Georges i in. [Georges i in. 2001] oraz Cameron z Trivedim [Cameron, Trivedi 2005] prezentują możliwości i zastosowania kopuli w ramach wielowymiarowej analizy przeżycia.

Celem artykułu jest wykorzystanie funkcji kopula w kontekście wielowymiarowych rozkładów czasów trwania tak, aby możliwe było modelowanie wielowymiarowych funkcji przeżycia. Konstruując empiryczny przykład, na podstawie danych z Niemieckiego Panelu Socjoekonomicznego GSOEP, modelujemy strukturę zależności między długością czasu oczekiwania na pierwsze dziecko a czasem trwania aktualnego zatrudnienia po zawarciu związku małżeńskiego. Podjęta zostanie próba weryfikacji hipotezy o związku pomiędzy opóźnieniem decyzji o prokreacji a aktywnością zawodową kobiet.

2. Definicja i przykłady funkcji kopula

Kopula jest funkcją, która łączy jednowymiarowe rozkłady brzegowe w rozkłady łączne. Poprzez wielowymiarową kopulę rozumiemy następującą funkcję:

$$C(u_1, u_2, \dots, u_n) = \Pr[U_1 \leq u_1, U_2 \leq u_2, \dots, U_n \leq u_n],$$

gdzie U_1, U_2, \dots, U_n są jednowymiarowymi zmiennymi losowymi. W niniejszej pracy zastosowanie znajdują dwuwymiarowe kopule $C(u_1, u_2) = \Pr[U_1 \leq u_1, U_2 \leq u_2]$.

Formalnie:

Definicja 1. Dwuwymiarowa kopula jest funkcją $C: [0, 1]^2 \rightarrow [0, 1]$ spełniającą:

1. $C(u, 0) = C(0, u) = 0$ dla dowolnego $u \in [0, 1]$.
2. $C(u, 1) = C(1, u) = u$ dla dowolnego $u \in [0, 1]$.
3. Dla wszystkich $0 \leq u_{11} \leq u_{12} \leq 1$ oraz $0 \leq u_{21} \leq u_{22} \leq 1$

$$C([u_{11}, u_{21}] \times [u_{12}, u_{22}]) = C(u_{12}, u_{22}) - C(u_{11}, u_{22}) - C(u_{12}, u_{21}) + C(u_{11}, u_{21}) \geq 0.$$

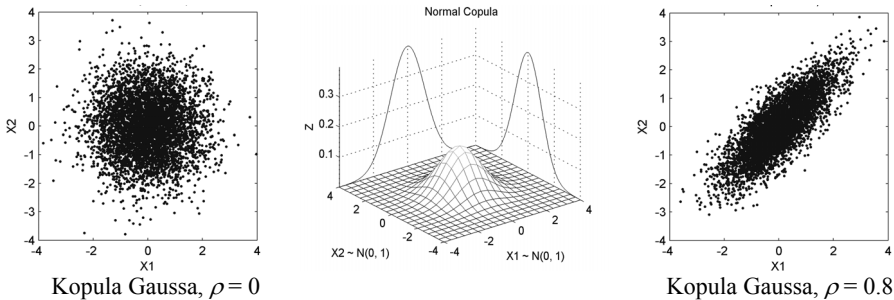
Znaczenie funkcji kopula w statystyce wyraża twierdzenie Sklara [Sklar 1959].

Twierdzenie Sklara. Niech $F(x_1, x_2, \dots, x_n)$ oznacza n -wymiarową funkcję rozkładu z funkcjami rozkładów brzegowych $F_1(x_1), F_2(x_2), \dots, F_n(x_n)$. Wtedy istnieje n -kopula C taka, że dla wszystkich rzeczywistych (x_1, x_2, \dots, x_n)

$$F(x_1, x_2, \dots, x_n) = C(F_1(x_1), F_2(x_2), \dots, F_n(x_n)).$$

Jeśli wszystkie rozkłady brzegowe są ciągłe, to kopula jest wyznaczona jednoznacznie.

Mnogość funkcji kopula występujących w literaturze wiąże się z tym, że każda z nich wyraża odmienną strukturę zależności w danych. Najprostsza kopula to kopula produktowa o postaci $C(u_1, u_2) = u_1 \cdot u_2$, adekwatna w wypadku niezależności zmiennych.



Rys. 1. Kopula gaussowska

Źródło: opracowanie własne.

Kopula gaussowska zadana jest formułą $C(u_1, u_2; \theta) = \Phi_G(\Phi^{-1}(u_1), \Phi^{-1}(u_2); \theta)$, gdzie Φ_G to funkcja rozkładu dla 2-wymiarowego standaryzowanego rozkładu normalnego, natomiast Φ^{-1} to odwrotność funkcji rozkładu dla 1-wymiarowego standa-

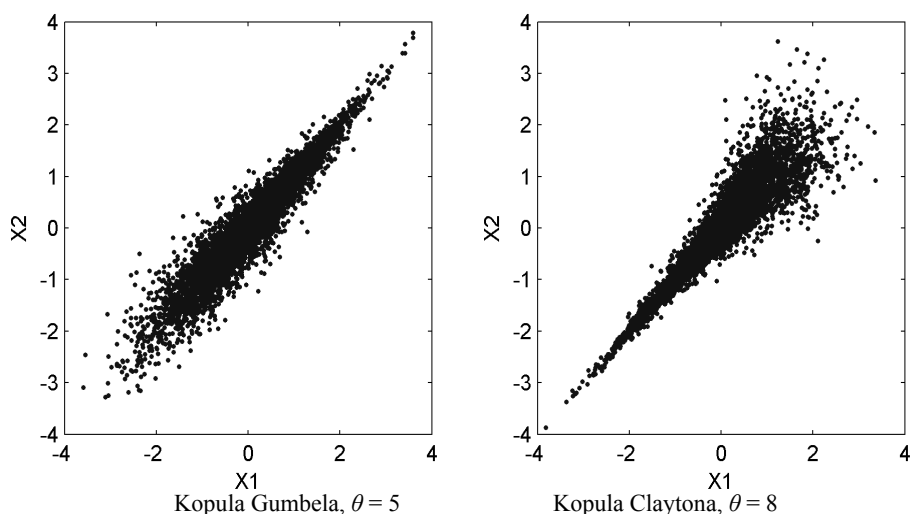
ryzowanego rozkładu normalnego. Kopula ta nadaje się do opisu zarówno dodatnich, jak i ujemnych zależności. Siłę zależności wyraża wartość parametru $\theta \in (-1, 1)$.

Rozważmy rodzinę funkcji $\varphi: [0, 1] \rightarrow [0, \infty)$ z ciągłymi pochodnymi na odcinku $(0, 1)$ o własnościach: $\varphi(1) = 0$, $\varphi(0) = \infty$, $\varphi'(1) < 0$ (funkcje malejące) oraz $\varphi''(1) > 0$ (wypukłe). Wtedy przez dwuwymiarową kopulę archimedesowską rozumiemy kopulę o kształcie $C(u_1, u_2; \theta) = \varphi^{[-1]}(\varphi(u_1) + \varphi(u_2))$. Funkcja φ zwana jest generatorem kopuli, natomiast $\varphi^{[-1]}$ to quasi-odwrotność generatora. Główną cechą kopuli archimedesowskich jest to, iż całkowita informacja o strukturze zależności zmiennych jest zawarta w jednowymiarowym generatorze φ .

Przykładem kopuli archimedesowskiej jest kopula Gumbela:

$$(C(u_1, u_2); \theta) = \exp \left\{ - \left((-\log(u_1))^\theta + (-\log(u_2))^\theta \right)^{\frac{1}{\theta}} \right\}.$$

Parametr zależności θ przyjmuje tu wartości z przedziału $[1, \infty)$; $\theta = 1$ oznacza niezależność, a $\theta \rightarrow \infty$ całkowitą zależność. Kopula Gumbela nie jest w stanie oddać ujemnej zależności, pokazuje zaś silną współzależność w prawym ogonie i słabą w lewym.



Rys. 2. Kopule Gumbela i Claytona

Źródło: opracowanie własne.

Z kolei popularna kopula Claytona ma postać $(C(u_1, u_2); \theta) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-\frac{1}{\theta}}$, gdzie $\theta \in (0, \infty)$, przy czym $\theta \rightarrow 0$ oznacza niezależność. Kopula Claytona nie odda-

je ujemnej zależności, pokazuje silną współzależność w lewym ogonie i słabą w prawym.

3. Miary zależności zmiennych losowych

Współczynnik korelacji Pearsona ρ_P często stosowany do pomiaru siły związku między dwiema zmiennymi ma wady. Mierzy on siłę zależności liniowej i w ogólnym przypadku analiza danych na podstawie tego współczynnika może prowadzić do błędnych wniosków. Jego wartość można wyznaczyć tylko dla zmiennych mających skończone wariancje. Niezależność zmiennych zawsze implikuje zerową korelację, ale jest to prawda tylko dla zmiennych o rozkładzie normalnym. Słaba korelacja nie implikuje słabej zależności. Dodatkowo korelacja nie jest miarą niezmienniczą.

Alternatywnym podejściem może być rozpatrywanie współczynnika mierzącego stopień uporządkowania zgodnego i niezgodnego: współczynnika Tau Kendalla. Nelsen [Nelsen 1999] pokazał, że Tau Kendalla może być wyrażone za pomocą kopuli:

$$\tau_K = 4 \int_0^1 \int_0^1 C(u_1, u_2) dC(u_1, u_2) - 1.$$

Dla kopuli Gaussa $\tau_K = \frac{2}{\pi} \arcsin(\theta)$, dla Claytona $\tau_K = \frac{\theta}{\theta + 2}$, a dla Gumbela $\tau_K = 1 - \frac{1}{\theta}$.

Również współczynnik korelacji rangowej Rho Spearmana ρ_S , będący miarą korelacji porządkowej dla dwóch zmiennych losowych, można wyrazić funkcją kopuli. Generalnie te dwa ostatnie współczynniki mierzą stopień zależności funkcyjnej w przeciwieństwie do współczynnika Pearsona, który mierzy tylko siłę zależności liniowej.

4. Kopuła dla wielowymiarowej funkcji przeżycia

Niech T będzie zmienną losową o funkcji rozkładu F i gęstości f oznaczającą czas trwania. Jednowymiarową funkcję przeżycia definiujemy jako $S(t) = P[T > t] = 1 - F(t)$, natomiast jednowymiarową funkcję hazardu jako

$$h(t) = \lim_{dt \rightarrow 0} \frac{P(t \leq T < t + dt \mid T \geq t)}{dt} = \frac{f(t)}{S(t)}.$$

Szacuje ona bezpośrednio ryzyko tego, że wydarzenie nastąpi w przedziale czasowym pomiędzy t i $t + dt$, pod warunkiem że do momentu t zajście jego nie nastąpiło. Często hazard $h(t)$ jest modelowany za pomocą zmiennych objaśniających, wtedy na przykład $h(t) = \exp(X\beta)h_0(t)$, gdzie $h_0(t)$ to tak zwany hazard bazowy (PH model).

Powyższe definicje można rozszerzyć na przypadek wielowymiarowy. Wtedy wielowymiarowa funkcja przeżycia $S(t_1, \dots, t_n) = \Pr[T_1 > t_1, \dots, T_n > t_n]$, gdzie T_1, \dots, T_n

oznaczają n czasów trwania o jednowymiarowych funkcjach przeżycia $S_j(t_j)$. Za-
chodzi $S_j(t_j) = S(0, \dots, 0, t_j, 0, \dots, 0)$, ale $S(t_1, \dots, t_n) \neq 1 - F(t_1, \dots, t_n)$.

Wielowymiarowa funkcja gęstości zadana jest wzorem
 $f(t_1, \dots, t_n) = \partial_{1, \dots, n} F(t_1, \dots, t_n) = (-1)^n S(t_1, \dots, t_n)$, zaś wielowymiarowa funkcja ha-
zardu przyjmuje postać

$$h(t_1, \dots, t_n) = \lim_{\max dt_j \rightarrow 0} \frac{P(t_1 \leq T_1 \leq t_1 + dt_1, \dots, t_n \leq T_n \leq t_n + dt_n | T_1 \geq t_1, \dots, T_n \geq t_n)}{dt_1 \dots dt_n} = \frac{f(t_1, \dots, t_n)}{S(t_1, \dots, t_n)}.$$

Kopule są naturalnym narzędziem do wyprowadzania wielowymiarowych funk-
cji przeżycia z jednowymiarowych funkcji brzegowych. Wielowymiarowa funkcja
przeżycia S ma następującą reprezentację: $S(t_1, \dots, t_n) = \tilde{C}(S_1(t_1), \dots, S_n(t_n))$, gdzie \tilde{C}
to tzw. kopuła przeżycia (twierdzenie Sklara dla funkcji przeżycia). Kopuła \tilde{C} łączy
brzegowe funkcje przeżycia w łączną funkcję przeżycia poniekąd analogicznie do
sposobu, w jaki zwykła kopuła C łączy brzegowe funkcje rozkładu w rozkład łączny.

Niech $n = 2$. Wtedy łączna funkcja przeżycia:

$$\begin{aligned} S(t_1, t_2) &= P[T_1 > t_1, T_2 > t_2] = \\ &= 1 - F(t_1) + 1 - F(t_2) - P[T_1 > t_1] - P[T_2 > t_2] + P[T_1 > t_1, T_2 > t_2] = \\ &= 1 - F(t_1) - F(t_2) + F(t_1, t_2) = S_1(t_1) + S_2(t_2) - 1 + C(F_1(t_1), F_2(t_2)) = \\ &= S_1(t_1) + S_2(t_2) - 1 + C(1 - S_1(t_1), 1 - S_2(t_2)) = \tilde{C}(S_1(t_1), S_2(t_2)). \end{aligned}$$

Kopulę przeżycia \tilde{C} definiujemy więc jako $\tilde{C}(u, v) = u + v - 1 + C(1 - u, 1 - v)$.

W ten sposób kopuła \tilde{C} wyraża łączne ryzyko porażki [Cameron, Trivedi 2005].

Wielowymiarową funkcję przeżycia można określać jako $S(t_1, \dots, t_n)$
 $= \tilde{C}(S_1(t_1), \dots, S_n(t_n))$, ale też jako $S(t_1, \dots, t_n) = C(S_1(t_1), \dots, S_n(t_n))$ dla pewnej kopu-
li rozkładów C . Oba podejścia nie dają jednak w efekcie tych samych funkcji prze-
życia z wyjątkiem kilku przypadków. Praca z kopulą rozkładów oraz kopulą prze-
życia jest ekwiwalentna dla kopuli eliptycznych (wtedy $\tilde{C} = C$).

5. Estymacja kopuli

Funkcji kopuła można używać do modelowania zależności, gdy rozkłady brze-
gowe są warunkowe albo też nie są warunkowe. Celem badania może być estymacja
parametrów regresji dla rozkładów brzegowych i tylko dodatkowo oszacowanie para-
metru zależności, ale również zbadanie natury zależności w modelach, które nie zawie-
rają zmiennych zależnych. Popularne metody estymacji modeli kopuli są następujące:

- symultaniczna estymacja wszystkich parametrów metodą *full maximum likelihood*,
- sekwencyjna 2-krokowa metoda największej wiarygodności (TSML), w której
najpierw są estymowane rozkłady brzegowe, a później dopiero parametr zależ-
ności θ (wykorzystując funkcję kopuła z podstawionymi do niej wynikami kro-
ku pierwszego),

- uogólniona metoda momentów (GMM).

6. Przykład empiryczny

Zastosowanie funkcji kopula zaprezentujemy na przykładzie modelowania związków między decyzjami prokreacyjnymi zameżnych kobiet a ich aktywnością zawodową.

Kształtowanie się postaw prokreacyjnych wobec kariery zawodowej stanowi często przedmiot zainteresowań socjologów. Po wojnie w większości krajów europejskich aktywność zawodowa kobiet była niska, co wynikało z tradycyjnego modelu kulturowego: praca kobiet była podporządkowana funkcjom rodzinnym. Po zakończeniu edukacji kobiety podejmowały zatrudnienie, ale trwale je przerywały po wyjściu za mąż lub po urodzeniu dziecka. Masowe wejście kobiet na rynek pracy w latach 60. spowodowane było oddziaływaniem ruchów feministycznych oraz wzrostem świadomości kobiet [Slany 2006].

W ostatnim ćwierćwieczu odnotowano, że zwiększona aktywność zawodowa kobiet wraz z presją rynku pracy powodują spadek dzietności ([*Family policy and family life in Europe. Volume I...* 1997; 2002]). Długie i nieprzerwane prokreacją uczestnictwo kobiet na rynku pracy możliwe jest dzięki rozwojowi środków kontroli urodzeń pozwalających na określanie czasu pojawienia się pożądanego potomstwa. Zanotowano podwyższanie się wieku kobiet rodzących dzieci. Parametry „starego i nowego reżimu demograficznego”, odnoszące się do sytuacji kobiet, przedstawił Livi-Bacci [Livi-Bacci 2004].

Wśród małżeństw dochodzi do racjonalizacji zachowań prokreacyjnych na skutek kalkulacji liczby dzieci, tak aby zmaksymalizować ich „oczekiwaną użyteczność” [Sah 1991]. „Koszt dziecka” to zwłaszcza czas matki poświęcony na jego wychowanie. Kobiety, które dobrze zarabiają, pozostają aktywne zawodowo i opóźniają prokreację. Mimo iż większość kobiet stara się połączyć rodzenie dzieci z pracą zawodową, to jednak nasilają się ujemne związki między płodnością a aktywnością zawodową kobiet. Opóźnianie prokreacji lub bezdzietność są mocno podkreślane w teorii „drugiego przejścia demograficznego” [Van de Kaa 1987]. Zgodnie z nią na czoło wysuwa się para, która inwestuje przede wszystkim w siebie, co określane jest jako DINKS (*Double Income No Kids*).

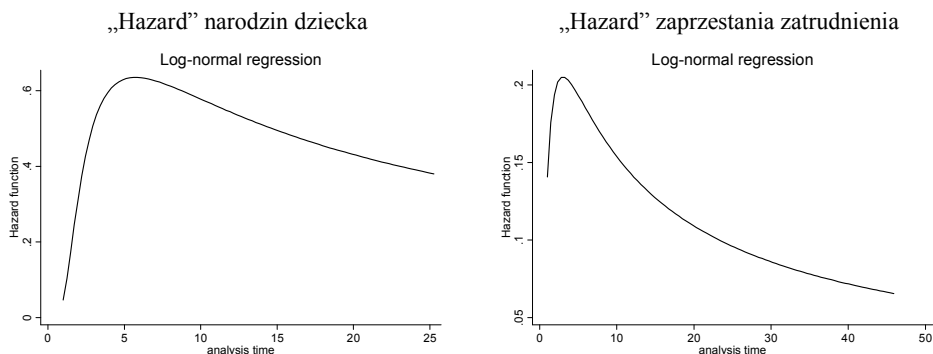
Takie negatywne tendencje odnotowano w wielu krajach europejskich. W niniejszej analizie będzie opisany problem na podstawie danych z Niemieckiego Panelu Socjo-Ekonomicznego GSOEP. Wyczerpująca analiza lokalnych uwarunkowań zawarta została w pracach: [*Family policy and family life in Europe. Volume I...* 1997; *Family policy and family life in Europe. Volume II...* 2002].

Wykorzystując dane z panelu GSOEP, konstruujemy przykład empiryczny dotyczący zastosowań kopul w modelowaniu wielowymiarowych czasów trwania. Dane dotyczą 5535 Niemek. Kobiety te wstąpiły w związek małżeński w latach 1946-1995 i trwały w nim przez co najmniej jeden rok. Na początku małżeństwa pracowały, ale później z pewnych powodów przerywały zatrudnienie. W trakcie

trwania związku urodziły dziecko. Budujemy dwie zmienne oznaczające czasy trwania: *doczdz* – długość czasu oczekiwania na pierwsze dziecko (od ślubu do narodzin w latach), *dzatr* – czas trwania zatrudnienia (od ślubu do zaprzestania pracy w latach). Obserwacje dla tych zmiennych nie są cenzurowane. Zachodzi statystycznie istotna zależność między tymi zmiennymi ($\rho_P = 0,1049$, $\tau_K = 0,2922$). Celem uniknięcia problemów ze zbieżnością w trakcie estymacji modeli przekształcamy oryginalne dane w dane ciągłe. Zmienną objaśniającą długość czasu oczekiwania na pierwsze dziecko jest *wiekmatki* – wiek matki w chwili narodzin pierwszego dziecka, natomiast zmienną objaśniającą – czas trwania zatrudnienia jest *wiekpanny* – wiek kobiety w chwili zawierania związku małżeńskiego.

Zaletą wykorzystania podejścia kopuli do określenia struktury zależności między zmiennymi jest możliwość opisu tej zależności na podstawie modeli dla rozkładów brzegowych. W przykładzie parametryczne modele brzegowe są szacowane i testowane z użyciem jednowymiarowych modeli przeżycia. Wyniki estymacji modeli hazardu PH i AFT przy założonych rozkładach brzegowych Weibulla, Gomperta, lognormalnym oraz log-logistycznym zaprezentowano w tab. 1. Zarówno w wypadku modelowania ryzyka (szansy) narodzin dziecka, jak i modelowania ryzyka zaprzestania zatrudnienia najlepszy jest model AFT log-normalny, który zakłada monotoniczne stopy hazardu: najpierw rosnące, a potem malejące. Model ten jest każdorazowo preferowany z powodu najwyższych wartości logarytmu funkcji wiarygodności.

W drugim kroku sekwencyjnej metody największej wiarygodności (TSML) szacujemy parametr zależności θ dla kopul przeżycia Gumbela, Claytona i Gaussa, podstawiając w jednowymiarowych funkcjach przeżycia szacunki parametrów uzyskane w kroku poprzednim. Wyniki przedstawiono w tab. 2. Najlepsze dopasowanie wykazuje kopula gaussowska, natomiast w rodzinie kopul archimedesowskich lepsza jest kopula Claytona oddająca silniejszą zależność w lewych ogonach.



Rys. 1. Przebieg krzywych hazardu dla narodzin dziecka i zaprzestania zatrudnienia

Źródło: obliczenia własne.

Tabela 1. Wyniki estymacji parametrycznych modeli jednowymiarowych

Ryzyko (szansa) narodzin dziecka								
Wyszczególnienie	Weibull		Gompertz		lognormalny		log-logistyczny	
	ocena parametru	z	ocena parametru	z	ocena parametru	z	ocena parametru	z
wiekmatki	0,8128582	-49,46	0,8273195	-43,84	0,0644619	44,40	0,0682023	43,34
const	-	-	-	-	-0,4876656	-13,03	-0,5722859	-14,48
p	2,391829	-	-	-	-	-	-	-
gamma	-	-	0,2947558	53,45	-	-	0,2689445	-
sigma	-	-	-	-	0,4696189	-	-	-
ln L	-3680,3054		-4772,5544		-3616,5732		-3674,5685	
Ryzyko zaprzestania zatrudnienia								
Wyszczególnienie	Weibull		Gompertz		lognormalny		log-logistyczny	
	ocena parametru	z	ocena parametru	z	ocena parametru	z	ocena parametru	z
wiekpanny	1,033184	8,95	1,028506	7,70	-0,0270096	-7,81	-0,0276607	-8,14
const	-	-	-	-	2,076698	25,57	2,009215	25,00
p	0,973373	-	-	-	-	-	-	-
gamma	-	-	-0,0261551	-15,59	-	-	0,5396752	-
sigma	-	-	-	-	0,9453674	-	-	-
ln L	-8261,6913		-8129,257		-7432,476		-7490,485	

Źródło: obliczenia własne.

Tabela 2. Wyniki oszacowań kopul przeżycia dla log-normalnych rozkładów brzegowych

Wyszczególnienie	Gumbel		Clayton		Gauss	
	ocena parametru	z	ocena parametru	z	ocena parametru	z
theta	1,091187	108,49	0,3053369	14,05	0,1909393	21,07
ln L	51,833405		115,84009		202,55611	

Źródło: obliczenia własne.

Kopule mogą być przydatne celem rozpoznania zmiany w czasie dla zależności pomiędzy czasem oczekiwania na dziecko i czasem trwania zatrudnienia. Zbiór danych podzielono na dwie części: pierwszą zawierającą informacje o małżeństwach z lat 1946-1970 oraz drugą z lat 1971-1995 (dwa równe 25-letnie podokresy). W pierwszym okresie zmienna *doczdz* była umiarkowanie skorelowana ze zmienną *dzatr*, natomiast w drugim okresie ta korelacja okazała się silniejsza. W porównaniu z latami 1946-1970 w latach 1971-1995 ryzyko narodzin dziecka stało się niższe, natomiast ryzyko przerwania zatrudnienia wyższe. W obu podokresach oszacowano kopule przeżycia dla log-normalnych rozkładów brzegowych. Przy estymacji kopuli gaussowskiej wystąpiły problemy ze zbieżnością algorytmu. W kopulach archimedesowskich zanotowano zaś znaczne różnice w szacunkach parametrów. W okresie 1971-1996 wartości parametrów θ były wyższe niż dla lat 1946-1970 oraz dodatkowo kopuła Gumbela, wskazująca na silniejszą zależność w

prawych ogonach, okazała się nieznacznie lepsza. Sugeruje to większą zależność między momentem macierzyństwa oraz przerwania pracy zawodowej w późniejszych latach trwania związku małżeńskiego. Decyzje prokreacyjne są więc odradzane. W latach 1945-1970 kobiety stosunkowo wcześniej zostawały niepracującymi matkami. Natomiast „nowoczesne” kobiety, zaangażowane w swoją karierę zawodową, odkładają macierzyństwo na później.

7. Podsumowanie

Celem badania było wykorzystanie idei funkcji kopula w kontekście łącznych rozkładów czasów trwania dla modelowania wielowymiarowych funkcji przeżycia. W artykule zawarto praktyczną implementację podejścia kopuli. Wydaje się, że funkcje kopula są stosownym narzędziem do analizowania struktury zależności pomiędzy czasami trwania, w szczególności w związku z tym, że rozkłady symetryczne nie są w tym wypadku naturalnymi kandydatami.

Przytoczony przykład empiryczny pozwala ująć związek aktywności zawodowej kobiet z decyzjami związanymi z rozrodczością. Szacując oddzielne modele kopuli dla różnych przedziałów czasowych, można zaobserwować ewaluację w czasie kształtowania się postaw prokreacyjnych wobec kariery zawodowej.

Literatura

- Cameron A.C., Li T., Trivedi P.K., Zimmer D.M. (2004), *Modeling the differences in counted outcomes using bivariate copula models: with application to mismeasured counts*, "Econometrics Journal", 7, s. 566-584.
- Cameron A.C., Trivedi P.K. (2005), *Microeconometrics. Methods and applications*, Cambridge University Press, New York.
- Cherubini U., Luciano E., Vecchiato W. (2004), *Copula methods in finance*, John Wiley, New York.
- Family policy and family life in Europe. Volume I. Structures and trends in the 1980s* (1997), red. F.-X. Kaufmann, A. Kuijsten, H.-J. Schulze, K.P. Strohmeier, Clarendon Press, Oxford.
- Family policy and family life in Europe. Volume II. Problems and Issues in Comparative Perspective*, (2002), red. F.-X. Kaufmann, A. Kuijsten, H.-J. Schulze, K.P. Strohmeier, Clarendon Press, Oxford.
- Frees E.W., Valdez E.A. (1998), *Understanding relationships using copulas*, "North American Actuarial Journal", 2(1), s. 1-25.
- Georges P., Lamy A.-G., Nicolas A., Quibel G., Roncalli T. (2001), *Multivariate survival modeling: A unified approach with copulas, unpublished paper*, France: Groupe de Recherche Opérationnelle Crédit Lyonnais.
- Hougaard P. (1989), *Fitting a multivariate failure time distribution*, "IEEE Trans Reliability", 38, s. 444-448.
- Joe H. (1997), *Multivariate models and dependence concepts*, Chapman & Hall, London.
- Kaufmann F.-X. (1997), *Herausforderungen des Sozialstaates, Suhrkamp*, Frankfurt am Main.
- Livi-Bacci M. (2004), *The narrow path of policies*, "Genus", 40(1), s. 207-231.
- Nelsen R.B. (1999), *An introduction to copulas*, Springer, New York.
- Sah R. (1991), *The effect of child mortality changes on fertility choice and parental welfare*, "Journal of Political Economy", 99, s. 582-606.
- Sklar A. (1959), *Fonctions de répartition à n dimensions et leurs marges*, Publ. Inst. Statist. Univ. Paris, 8, s. 229-231.

Slany K. (2006), *Alternatywne formy życia małżeńsko-rodzinnego w ponowoczesnym świecie*, Zakład Wydawniczy Nomos, Kraków.

Van de Kaa D. (1987), *Europe's second demographic transition*, "Population Bulletin", 42(1), s. 3-24.

THE APPLICATION OF COPULA FUNCTIONS TO MODELING OF MULTIVARIATE DURATIONS IN THE CASE OF PROCREATION DECISIONS AMONG PROFESSIONALLY ACTIVE WOMEN

Summary

Copulas are a useful method for deriving joint distributions given the marginals especially when we work with non-normal distributions. The aim of our paper is to use this idea in the context of joint duration distributions for the modeling of multivariate survival functions. In our paper, a numerical example is considered, which is based on the data from the German Socio-Economic Panel Study GSOEP. We model the dependence between the duration of time a woman remains childless and the duration of employment among German married women. Marginal models are fitted and tested with the use of standard univariate survival models, and the dependence parameter is estimated in a sequential second-stage procedure.