Uniwersytet
Ekonomiczny
w Katowicach

Studia Ekonomiczne. Zeszyty Naukowe
Uniwersytetu Ekonomicznego w Katowicach
ISSN 2083-8611 Nr 297 · 2016

Andrzej Stryjek

Szkoła Główna Handlowa w Warszawie
Instytut Ekonometrii
astryj@sgh.waw.pl

PORÓWNANIE METOD ESTYMACJI PARAMETRU W KLASIE WYBRANYCH DWUWYMIAROWYCH KOPULI ARCHIMEDESOWYCH

Streszczenie: Celem artykułu jest porównanie dokładności estymacji parametru kopuli za pomocą kanonicznej metody największej wiarygodności, metody kalibracji, metody estymacji przedziałowej i metody minimalnej odległości. Analiza polegała na komputerowej symulacji 250-elementowych prób z rozkładu o funkcji łączącej Claytona, Gumbela, Yagera i Farlie-Gumbel-Morgensterna dla 20 losowo wybranych parametrów tych kopuli. Kryterium porównania metod stanowiło obciążenie i błąd średniokwadratowy estymatorów.

Słowa kluczowe: kopula, kanoniczna metoda największej wiarygodności, metoda minimalnej odległości, τ-Kendalla, estymacja przedziałowa.

Wprowadzenie

W ostatnich latach, oprócz typowych wskaźników określających stopień zależności między zmiennymi losowymi (współczynnik korelacji Pearsona, τ-Kendalla lub ρ-Spearmana), coraz częściej są stosowane funkcje łączące (kopule). Wykorzystanie kopuli w praktyce, tj. dla danych empirycznych, wymaga oszacowania parametru lub parametrów funkcji łączącej. Często wymienianym w literaturze sposobem estymacji jest metoda największej wiarygodności. Oprócz niej są stosowane dwie metody: wnioskowania dla rozkładów brzegowych i kalibracji [por. Cherubini i in., 2004; Heilpern, 2007; Doman, 2011]. Rzadziej wymienia się metodę estymacji przedziałowej [Heilpern, 2007] oraz metodę minimalnej odległości. Metoda minimalnej odległości została zaproponowana w artykule Wolfowitza [1953] jako nowy sposób szacowania parametrów dystrybuanty rozkładu. Ponieważ każda kopula jest dystrybuantą obciętą do zbioru $[0, 1]^2$, więc metodę minimalnej odległości można użyć do estymacji parametrów funkcji łączących. W pracach Tsukahara [2005] i Weiss [2010] podjęto dyskusję o skuteczności tego sposobu estymacji w porównaniu z estymacją metodą największej wiarygodności.

Niniejszy artykuł włącza się w tę dyskusję przez analizę innych metod estymacji parametru funkcji łączącej, tj. kanoniczną metodę największej wiarygodności, metodę kalibracji za pomocą współczynnika τ-Kendalla, metodę estymacji przedziałowej oraz metodę minimalnej odległości z zastosowaniem odległości Kołmogorowa-Smirnowa, odległości Cramera von Misesa i odległości indukowanej przez normę z przestrzeni L_1. Podjęta analiza obejmuje trzy rodziny kopuli: rodzinę Claytona, Gumbela i Yagera, które zalicza się do archimedesowych funkcji łączących. Dodatkowo analizowano kopulę Farlie-Gumbel-Morgensterna (FGM), która nie jest kopulą archimedesową.

1. Kopula. Kopule archimedesowe

Kopulą (funkcją łączącą) nazywamy każdą dwuwymiarową dystrybuantę, która jest obcięta do zbioru $[0, 1] \times [0, 1]$, taką że dystrybuanty rozkładów brzegowych są dystrybuantami rozkładów jednostajnych na odcinku $[0, 1]$.

Popularność kopuli w modelach opisujących zależności między zmiennymi losowymi stosowanymi do opisu procesów zachodzących na rynkach finansowych wynika z faktu, iż umożliwiają w większym stopniu niż tradycyjne miary, takie jak współczynnik korelacji liniowej Pearsona, τ-Kendalla i ρ-Spearmana, opis struktury powiązań między tymi zmiennymi. Zauważmy, że wymienione mierniki to wartość liczbowa, a kopula to funkcja.

Funkcja łącząca stanowi połączenie między dystrybuantami rozkładów brzegowych dwuwymiarowej zmiennej losowej (X_1, X_2) a dystrybuantą rozkładu łącznego. Własność tę formułuje twierdzenie Sklara [zob. Cherubini i in., 2004, s. 56]. *Niech H będzie dystrybuantą dwuwymiarowej zmiennej losowej (X_1, X_2) o rozkładach brzegowych określonych przez dystrybuanty F_1 i F_2 oraz $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$. Wówczas istnieje dwuwymiarowa kopula C taka, że dla każdego $(x_1, x_2) \in \overline{\mathbb{R}}^2$ zachodzi:*

$$H(x_1, x_2) = C\big(F_1(x_1), F_2(x_2)\big). \tag{1}$$

Jeżeli funkcje F_1 i F_2 są ciągłe, to C jest wyznaczona jednoznacznie. W przeciwnym przypadku jest ona wyznaczona jednoznacznie tylko na zbiorze $F_1(\overline{\mathbb{R}}) \times F_2(\overline{\mathbb{R}})$. Jeżeli zaś C jest dwuwymiarową kopulą oraz F_1 i F_2 są dystry-

buantami rozkładów pewnych jednowymiarowych zmiennych losowych, to funkcja H zdefiniowana wzorem (1) *jest dystrybuantą dwuwymiarowej zmiennej losowej, której dystrybuanty rozkładów brzegowych to funkcje F_1 i F_2.*

Każda kopula C jest funkcją ograniczoną z góry przez funkcję $W^2(x_1, x_2) = \max\{x_1 + x_2 - 1\}$, z dołu zaś przez $M^2(x_1, x_2) = \min\{x_1, x_2\}$. Innymi słowy dla każdego $(x_1, x_2) \in [0,1]^2$ zachodzi nierówność:

$$M^2(x_1, x_2) \leq C(x_1, x_2) \leq W^2(x_1, x_2).$$

Powyższa nierówność w literaturze z zakresu kopuli nosi nazwę nierówności Frécheta-Hoeffdinga. Dowodzi się, że funkcje M^2 i W^2 są kopulami.

Często funkcja łącząca C jest zależna od parametru (parametrów) $\alpha \in I$, gdzie I jest zbiorem określającym zakres parametru α, dla którego C jest kopulą. Wówczas stosujemy oznaczenie C_α. W szczególnych przypadkach wartości parametru α kopula C_α może być równa funkcji łączącej M^2, W^2 lub Π. Przy czym $\Pi(x_1, x_2) = x_1 \cdot x_2$ jest kopulą niezależności i opisuje rozkład dwuwymiarowej zmiennej losowej, dla której rozkłady brzegowe są niezależne.

W dwuwymiarowym przypadku można podać wiele przykładów funkcji, które są kopulami [por. Nelsen, 2006]. Wśród nich ważną klasę funkcji łączących stanowią kopule archimedesowe. Kopula archimedesowa to funkcja łącząca, która ma postać:

$$C_\varphi(x_1, x_2) = \varphi^{[-1]}\big(\varphi(x_1) + \varphi(x_2)\big), \tag{2}$$

przy czym $\varphi: [0,1] \to [0, +\infty]$ jest ciągłą, ściśle malejącą i wypukłą funkcją, która spełnia warunek $\varphi(1) = 0$. Funkcję φ nazywamy generatorem kopuli archimedesowej C_φ. Natomiast funkcję $\varphi^{[-1]}$ nazywamy pseudoodwrotnością generatora φ. Funkcja ta jest zdefiniowana jako:

$$\varphi^{[-1]}(t) = \begin{cases} \varphi^{-1}(t) \text{ dla } 0 \leq t \leq \varphi(0) \\ 0 \text{ dla } \varphi(0) \leq t \leq +\infty \end{cases}.$$

Kopule archimedesowe charakteryzują się stosunkowo prostą konstrukcją oraz relatywnie do innych rodzajów funkcji łączących nieskomplikowaną formą analityczną i parametryzacją. Ponadto współczynnik τ-Kendalla wyraża się jako [por. Heilpern, 2007, s. 85]:

$$\tau = 1 + 4 \cdot \int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt, \tag{3}$$

a więc generator kopuli archimedesowej φ determinuje τ.

2. Metody estymacji parametru kopuli

W celu zastosowania funkcji łączącej jako opisu zależności między badanymi zmiennymi losowymi zachodzi najczęściej potrzeba wyspecyfikowania wartości parametru α funkcji łączącej C_α. Jak już zostało wspomniane, w literaturze są wymieniane następujące metody szacowania tego parametru:

- metoda największej wiarygodności,
- metoda wnioskowania dla rozkładów brzegowych,
- kanoniczna metoda największej wiarygodności,
- metoda kalibracji na podstawie współczynnika τ-Kendalla lub ρ-Spearmana,
- estymacja przedziałowa,
- metoda minimalnej odległości.

Dwie pierwsze metody są trudniejsze do zastosowania w praktyce, ponieważ wymagają sformułowania *a priori*, jakie są rozkłady brzegowe zmiennej (X_1, X_2). Zatem oprócz estymacji parametrów kopuli dodatkowym problemem jest estymacja parametrów przyjętych rozkładów. W prezentowanym badaniu były więc rozpatrywane tylko cztery pozostałe z wymienionych metod. Ich syntetyczny opis stanowi przedmiot dalszej części tego punktu artykułu.

2.1. Metoda kalibracji

Szacowanie wartości parametru α kopuli C_α metodą kalibracji polega na wyznaczeniu takiej wartości $\hat{\alpha}$, która daje identyczną wartość miary τ-Kendalla lub ρ-Spearmana, co wartość ich odpowiedniego estymatora wyznaczonego na podstawie próby (x_1^t, x_2^t) dla $t = 1, 2, \dots, T$. Jak podaje Cherubini i in. [2004, s. 99-102], nieobciążonym i zgodnym estymatorem τ-Kendalla jest $\hat{\tau} = \frac{2}{T(T-1)} \cdot \sum_{t=1}^{T} \sum_{s=1}^{T} \operatorname{sgn}\left((x_1^t - x_1^s) \cdot (x_2^t - x_2^s)\right)$, natomiast ρ-Spearmana $\hat{\rho} = \frac{12}{T(T^2-1)} \cdot \sum_{t=1}^{T} (r_1^t - \bar{r}_1)(r_2^t - \bar{r}_2)$, gdzie r_i^t to rangi obserwacji x_i^t dla $i = 1, 2$, a $\bar{r}_i$ oznacza średnią z wartości r_i^t.

Mając wyznaczoną wartość $\hat{\tau}$ (lub $\hat{\rho}$), wystarczy przyrównać tę wartość do wzoru na współczynnik τ (lub ρ) i wyznaczyć $\hat{\alpha}$ jako rozwiązanie tego równania z niewiadomą α. Niestety dla wielu kopuli wzór wyrażający współczynnik τ-Kendalla lub ρ-Spearmana ma skomplikowaną postać analityczną, taką że rozwiązanie wspomnianego równania jest kłopotliwe lub niemożliwe (także w sposób numeryczny).

2.2. Estymacja przedziałowa

Przedział ufności dla parametru kopuli C_α na poziomie $1-\beta \in (0,1)$ ma postać:

$$\hat{\alpha}_T - z_{\frac{\beta}{2}} \cdot \frac{4S|g'(\tau_T)|}{\sqrt{T}} \leq \alpha \leq \hat{\alpha}_T + z_{\frac{\beta}{2}} \cdot \frac{4S|g'(\tau_T)|}{\sqrt{T}}, \tag{4}$$

gdzie:

- T jest liczebnością próby,
- τ_T jest estymatorem z próby współczynnika τ-Kendalla,
- funkcja g jest analitycznym zapisem zależności parametru kopuli α od wartości τ-Kendalla, tj. $\alpha = g(\tau)$,
- $\hat{\alpha}_T = g(\tau_T)$,
- z_β jest kwantylem rzędu $1-\beta$ standardowego rozkładu normalnego,
- $S = \sqrt{\frac{1}{T} \cdot \sum_{i=1}^{T} \left(W_i + \widetilde{W}_i - 2\overline{W}\right)^2}$,
- macierz $\boldsymbol{I}_{ij} = \begin{cases} 1 \text{ gdy } x_i < x_j \text{ i } y_i < y_j \\ 0 \text{ przeciwnym przypadku} \end{cases}$ dla $i \neq j$ oraz $\boldsymbol{I}_{ii} = 1$,
- $W_i = \frac{1}{T} \cdot \sum_{i=1}^{T} \boldsymbol{I}_{ij}$,
- $\overline{W} = \frac{1}{T} \cdot \sum_{i=1}^{T} W_i$,
- $\widetilde{W}_i = \frac{1}{T} \cdot \sum_{i=1}^{T} \boldsymbol{I}_{ji}$,
- $\boldsymbol{I}_{ji} = \begin{cases} 1 \text{ gdy } x_i \leq x_j \text{ i } y_i \leq y_j \\ 0 \text{ przeciwnym przypadku} \end{cases}$.

Przedział (4) jest skonstruowany na podstawie statystyki $\sqrt{T} \cdot \frac{\tau_T - \tau}{4S}$, która ma w granicy standardowy rozkład normalny [Heilpern, 2007, s. 131]. W przeciwieństwie do estymatora metodą kalibracji, ten jest estymatorem przedziałowym, co stanowi jego zaletę.

2.3. Kanoniczna metoda największej wiarygodności (*KMNW*)

W tej metodzie estymacji przyjmuje się, że rozkłady brzegowe zmiennej (X_1, X_2) są opisane za pomocą dystrybuant empirycznych $F_1^T(x_1) = \frac{1}{T+1} \cdot \sum_{t=1}^{T} \mathbb{1}_{\{x_1^t:\, x_1^t \leq x_1\}}$ i $F_2^T(x_2) = \frac{1}{T+1} \cdot \sum_{t=1}^{T} \mathbb{1}_{\{x_2^t:\, x_2^t \leq x_2\}}$, przy czym $\mathbb{1}_A$ jest funkcją charakterystyczną zbioru A, zaś (x_1^t, x_2^t) to T-elementowa próba zawierająca realizacje zmiennej (X_1, X_2).

Logarytm funkcji wiarygodności przy powyższych założeniach ma postać:

$$l(\alpha) = \sum_{t=1}^{T} \ln(c_\alpha(u_1, u_2)), \tag{5}$$

gdzie $u_1 = F_1(x_1)$ i $u_2 = F_2(x_2)$ oraz $c_\alpha(u_1, u_2) = \frac{\partial \mathcal{C}}{\partial u_1 \partial u_2}(u_1,\ u_2)$ jest gęstością funkcji łączącej $\mathcal{C}_\alpha$. Jak podaje Heilpern [2007, s. 135], wzór (5) można równoważnie zapisać za pomocą rang r_i^t obserwacji x_i^t dla $i = 1, 2$ oraz $t = 1, 2, \ldots, T$ w postaci:

$$l(\alpha) = \sum_{t=1}^{T} \ln\left(c_\alpha\left(\frac{r_1^t}{T+1}, \frac{r_2^t}{T+1}\right)\right). \tag{6}$$

Estymatorem parametru α jest taka wartość $\hat{\alpha}$, która maksymalizuje funkcję (6). Estymator otrzymany kanoniczną metodą największej wiarygodności jest asymptotycznie normalny [Heilpern, 2007, s. 136].

2.4. Metoda minimalnej odległości

W pracy Wolfowitza [1953] zaproponowano, aby odległość δ między zmiennymi losowymi X i Y mierzyć jako:

$$\delta(X, Y) = \sup_{x \in \mathbb{R}} |F_X(x) - F_Y(x)|,$$

gdzie F_X i F_Y są dystrybuantami zmiennych X i Y. Metodę tę można uogólnić przez zastąpienie normy supremum inną funkcją odległości d, czyli $\delta(X, Y) = d(F_X(x), F_Y(x))$. Jeżeli przyjmiemy, że dystrybuanta F_X jest dystrybuantą empiryczną wyznaczoną na podstawie próby, to oszacowania parametrów drugiej dystrybuanty F_Y można wyznaczyć tak, aby odległość $\delta(X, Y)$ była minimalna.

Odpowiednikiem dystrybuanty empirycznej dla funkcji łączącej jest kopula empiryczna. Kopula empiryczna [por. Cherubini i in., 2004, s. 161] to funkcja określona na zbiorze:

$$\left\{\left(\frac{i_1}{T}, \frac{i_2}{T}\right) : i_j \in \{0, 1, 2, \ldots, T\} \text{ dla } j \in \{1, 2\}\right\}$$

jako:

$$\mathcal{C}_T\left(\frac{i_1}{T}, \frac{i_2}{T}\right) = \frac{1}{T} \cdot \sum_{t=1}^{T} \left(\mathbb{1}_{\{r_1^t :\, r_1^t \le i_1\}} \cdot \mathbb{1}_{\{r_2^t :\, r_2^t \le i_2\}}\right),$$

gdzie r_i^t oznacza rangę obserwacji x_i^t dla $i = 1, 2$ oraz $t = 1, 2, \ldots, T$.

Estymator parametru α kopuli $\mathcal{C}_\alpha$ jest wartością, która minimalizuje odległość $d(\mathcal{C}_\alpha, \mathcal{C}_T)$.

W tym miejscu warto zauważyć, że własności tego estymatora są dotychczas w małym stopniu opisane w literaturze. W pracy Tsukahara [2005] autor wykorzystał odległość Cramera von Misesa określoną jako:

$$d(\mathcal{C}_1, \mathcal{C}_2) = \int_{(0,1)^2} \left(\mathcal{C}_1(u, v) - \mathcal{C}_2(u, v)\right)^2 du dv$$

i wykazał, że dla tej funkcji odległości estymator jest asymptotycznie normalny. Rozważania na temat dokładności estymacji parametru funkcji łączącej metodą minimalnej odległości są zawarte również w pracach Mendes [2007] i Weiss [2010].

2.5. Opis badania

Celem zaprezentowanego w tym punkcie badania symulacyjnego jest porównanie dokładności opisanych wyżej metod estymacji parametru kopuli.

Analiza jest przeprowadzona dla czterech funkcji łączących: kopuli Claytona, kopuli Gumbela, kopuli Yagera i kopuli Farlie-Gumbel-Morgensterna. Trzy pierwsze funkcje to kopule archimedesowe.

Kopula Claytona jest określona wzorem:

$$C_\alpha(u,v) = (u^{-\alpha} + v^{-\alpha} - 1)^{-\frac{1}{\alpha}}, \tag{7}$$

gdzie $\alpha \in (-1, 0) \cup (0, +\infty)$. Generatorem kopuli (7) jest funkcja $\varphi(t) = \frac{1}{\alpha}(t^{-\alpha} - 1)$. Zatem współczynnik τ-Kendalla wynosi $\tau = \frac{\alpha}{\alpha+2}$.

Funkcja łącząca Gumbela jest zdefiniowana jako:

$$C_\alpha(u,v) = \exp\left(-((-\ln(u))^\alpha + (-\ln(v))^\alpha)^{\frac{1}{\alpha}}\right), \tag{8}$$

gdzie $\alpha \in [1, +\infty)$. Generatorem kopuli (8) jest funkcja $\varphi(t) = \left(-\ln(t)\right)^\alpha$. Współczynnik τ-Kendalla wynosi $\tau = \frac{\alpha-1}{\alpha}$.

Kopula Yagera to funkcja określona formułą:

$$C_\alpha(u,v) = \frac{1}{1+((u^{-1}-1)^\alpha+(v^{-1}-1)^\alpha)^{\frac{1}{\alpha}}}, \tag{9}$$

gdzie $\alpha \in [1, +\infty)$. Generator kopuli (9) to funkcja $\varphi(t) = \left(\frac{1}{t} - 1\right)^\alpha$. Współczynnik τ-Kendalla jest równy $\tau = \frac{\alpha-2}{\alpha}$.

Kopula Farlie-Gumbel-Morgensterna jest to funkcja łącząca opisana wzorem:

$$C(u,v) = uv + \alpha uv(1-u)(1-v), \tag{10}$$

gdzie $\alpha \in [-1, 1]$. Współczynnik τ-Kendalla wynosi zaś $\tau = \frac{2\alpha}{9}$. Kopula FGM może więc być stosowana do modelowania słabych zależności między zmiennymi losowymi, gdyż $-\frac{2}{9} \le \tau \le \frac{2}{9}$.

Dla każdej kopuli C_α (spośród czterech wymienionych wyżej) wylosowano dwadzieścia wartości jej parametru α_i $(i = 1, 2, \dots, 20)$. Przy czym w przypadku funkcji łączących Claytona, Gumbela i Yagera górne ograniczenie przedziału dla parametru zostało ustalone arbitralnie jako wartość 30. Okazuje się bowiem, że dla wartości przekraczających ten próg są generowane próby o bardzo zbliżonych rozkładach, co zaobserwowano na podstawie wykresów punktowych utworzonych dla wielu prób. W następnym etapie symulacji dla każdej wartości α_i wygenerowano próbę o liczebności 250 z rozkładu opisanego kopulą C_{α_i}.

Generowanie pary liczb z rozkładu opisanego kopulą C_α odbywało się na podstawie algorytmu opartego na pojęciu kopuli warunkowej. Szczegóły można znaleźć w pracy Cherubini i in. [2004]. Tutaj podano algorytm za Heilpern [2007, s. 100]:

- wygenerować wartości v_1 i v_2 z rozkładu jednostajnego na przedziale $[0, 1]$,
- przyjąć $u_1 = v_1$ i u_2 jako rozwiązanie równania $v_2 = \frac{\frac{\partial C_\alpha(u_1,u_2)}{\partial u_1}}{\frac{\partial C_\alpha(u_1,1)}{\partial u_1}}$.

Na podstawie każdej 250-elementowej próby wyznaczono wartość estymatora $\hat{\alpha}_i$ z wykorzystaniem metod opisanych w drugim punkcie artykułu. Procedurę generowania próby z rozkładu opisanego funkcją łączącą C_{α_i} i wyznaczania wartości $\hat{\alpha}_i$ iterowano 1000 razy.

Łącznie dla wszystkich kopuli (Claytona, Gumbela, Yagera i FGM) oraz dla każdej metody estymacji otrzymano 1000 oszacowań osiemdziesięciu wylosowanych parametrów. W dalszej części artykułu zestaw tysiąca oszacowań jednego parametru będziemy nazywać rozkładem estymatora na podstawie symulacji.

Do estymacji parametru kopuli metodą minimalnej odległości zastosowano trzy warianty funkcji odległości:

- odległość Cramera von Misesa (CvM):

$$d(C_T, C_\alpha) = \sum_{i=1}^{T}\sum_{j=1}^{T}\left(C_T\left(\frac{i}{T},\frac{j}{T}\right) - C_\alpha\left(\frac{i}{T},\frac{j}{T}\right)\right)^2,$$

- odległość Kołmogorowa-Smirnowa (KS):

$$d(C_T, C_\alpha) = \max_{\left(\frac{i}{T},\frac{j}{T}\right)\in\mathcal{L}}\left|C_T\left(\frac{i}{T},\frac{j}{T}\right) - C_\alpha\left(\frac{i}{T},\frac{j}{T}\right)\right|,$$

- odległość indukowaną przez normę z przestrzeni L_1:

$$d(C_T, C_\alpha) = \sum_{i=1}^{T}\sum_{j=1}^{T}\left|C_T\left(\frac{i}{T},\frac{j}{T}\right) - C_\alpha\left(\frac{i}{T},\frac{j}{T}\right)\right|.$$

Obliczenia zostały przeprowadzone w pakiecie R.3.1.1. Maksymalizacja lub minimalizacja wartości funkcji została wykonana za pomocą procedury *nlminb*. Natomiast numeryczne rozwiązywanie równań nieliniowych przeprowadzono procedurą *uniroot.all*. Niezbędne wzory w algorytmie symulowania próby z rozkładu określonego kopulą, jak również formuły poszczególnych funkcji wiarygodności zostały wyznaczone w programie Mathematica 9.

3. Otrzymane wyniki

Podstawowe kryteria służące do porównania metod szacowania wartości parametru kopuli na podstawie otrzymanych symulacyjnych rozkładów estymatorów to: obciążenie estymatora oraz błąd średniokwadratowy.

Obciążeniem estymatora $\hat{\theta}$ parametru θ nazywamy wielkość $\mathbb{E}(\hat{\theta} - \theta)$. W badaniu symulacyjnym za obciążenie estymatora $\hat{\theta}_i$ przyjęto $\frac{1}{T} \cdot \sum_{n=1}^{T}\left(\hat{\theta}_i^n - \theta_i\right)$, gdzie $T = 1000$, θ_i jest parametrem kopuli $(i = 1, 2, \dots, 20)$, $\hat{\theta}_i^n$ jest wartością estymatora wyznaczoną na podstawie próby wygenerowanej w n-tej iteracji $(n = 1, 2, \dots, 1000)$.

Błędem średniokwadratowym estymatora $\hat{\theta}$ parametru θ nazywamy wielkość $\mathbb{E}\left(\left(\hat{\theta} - \theta\right)^2\right)$. Błąd średniokwadratowy parametru kopuli θ_i $(i = 1, \dots, 20)$ wyznaczano jako $\frac{1}{T}\sum_{n=1}^{T}\left(\hat{\theta}_i^n - \theta_i\right)^2$.

W tabeli 1 są zamieszczone wyniki symulacji dotyczące obciążenia (po uśrednieniu wyników względem 20 wylosowanych parametrów) poszczególnych estymatorów w zależności od poszczególnych funkcji łączących.

Tabela 1. Obciążenie estymatorów (liczebność próby 250)

Kopula	Obciążenie estymatora				
	kalibracja	$KMNW$	L_1	CvM	KS
Claytona	0,1647	**0,0461**	0,9747	-0,3043	-1,1514
Gumbela	**0,1487**	-16,7100	2,2071	-0,9309	-3,0859
Yagera	**0,1358**	-0,2080	3,5916	-1,3500	-4,5178
FGM	**-0,0013**	-0,0014	-0,0026	-0,0030	-0,0035

Źródło: Opracowanie własne.

Najczęściej obciążenie estymatora najbardziej zbliżone do wartości zero występowało w przypadku metody kalibracji. Tylko dla kopuli Claytona takie obciążenie wystąpiło dla metody $KMNW$. Metoda $KMNW$ charakteryzowała się bardzo dużym obciążeniem dla funkcji łączącej Gumbela. Najprawdopodobniej wynik ten powstał na skutek kłopotów numerycznych przy maksymalizacji funkcji wiarygodności. Warto zauważyć, że przy estymacji metodą minimalnej odległości najlepsze wyniki otrzymywano dla funkcji odległości CvM, a najsłabsze dla odległości KS.

Zauważmy także, iż $KMNW$ i metoda minimalnej odległości na ogół szacowały wartość parametru z niedoborem, natomiast metoda kalibracji z nadmiarem.

W tabeli 2 zawarto wyniki dotyczące błędu średniokwadratowego dla wszystkich kopuli.

Podobnie jak przy poprzednim kryterium, można stwierdzić, że na podstawie błędu średniokwadratowego najlepsza metoda estymacji to kalibracja. Tym razem najmniejszy błąd otrzymano dla tej metody w przypadku kopuli Gumbela i Yagera, a dla metody $KMNW$ − kopuli Claytona i FGM. Jednakże dla kopuli FGM różnica w błędzie średniokwadratowym jest nieznaczna dla wszystkich metod estymacji.

Tabela 2. Błąd średniokwadratowy estymatorów

Kopula	Błąd średniokwadratowy estymatora				
	kalibracja	$KMNW$	L_1	CvM	KS
Claytona	4,1412	**1,2046**	8,6022	5,6184	8,0119
Gumbela	**3,1238**	325,6885	13,1600	4,4505	17,2912
Yagera	**3,3847**	3,4840	29,2406	5,6580	31,8366
FGM	0,0341	**0,0315**	0,0334	0,0333	0,0370

Źródło: Opracowanie własne.

Na zakończenie zostanie omówiona estymacja przedziałowa. Wyniki dla tej metody zebrano w tabeli 3. W badaniu symulacyjnym przyjęto, że poziom ufności wynosi $0{,}95$. W kolumnie „parametr" jest podany odsetek przypadków, gdy „prawdziwa" wartość parametru zawierała się w wyznaczonym przedziale ufności. W następnych kolumnach tabeli 3 podano dodatkowo, jaki procent spośród oszacowań otrzymanych pozostałymi metodami (kalibracji, KMNW i metodą minimalnej odległości) zawierał się w obliczonych przedziałach ufności.

Tabela 3. Wyniki badań dla metody estymacji przedziałowej

Kopula	Udział procentowy otrzymanych wartości w przedziale ufności					
	parametr	kalibracja	$KMNW$	L_1	CvM	KS
Claytona	95,18	100,00	98,13	98,75	100,00	97,14
Gumbela	95,88	100,00	0,21	89,83	100,00	62,35
Yagera	96,29	100,00	97,20	67,16	100,00	38,04
FGM	94,81	100,00	99,98	99,98	99,98	99,98

Źródło: Opracowanie własne.

Jak widać, poziom ufności nie został osiągnięty tylko dla funkcji łączącej FGM. Natomiast analiza pozostałych metod estymacji na podstawie przedziału ufności potwierdza, że najskuteczniejszą metodą estymacji jest kalibracja, następnie w kolejności jest $KMNW$ i metoda minimalnej odległości z funkcją odległości CvM.

Podsumowanie

Jak wiadomo, zależności między zmiennymi losowymi, które stosuje się do opisu działania rynków finansowych, mają skomplikowany charakter. Często modelowanie za pomocą rozkładu normalnego (lub ogólniej: eliptycznego) nie jest podejściem poprawnym. Dla rozkładów, które nie są eliptyczne, stosowanie współczynnika korelacji liniowej jest niewłaściwe wówczas, gdy chcemy wyrazić zależność między rozkładami brzegowymi a rozkładem łącznym wielowymiarowej zmiennej losowej. W tej sytuacji alternatywne podejście stanowią funkcje łączące.

Opisane w artykule badanie symulacyjne to początek większego i bardziej złożonego zadania, które polega na porównaniu dokładności popularnych metod estymacji parametru kopuli z estymacją minimalnej odległości. Uzyskane wyniki wskazują, że prostą i skuteczną metodą szacowania jest metoda kalibracji na podstawie współczynnika τ-Kendalla. Jednakże metoda ta dla większości kopuli nie jest możliwa do zastosowania ze względu na trudności obliczeniowe zasygnalizowane w drugim punkcie artykułu.

Jak pokazuje przeprowadzona analiza, zamiast kalibracji, z dość dobrym skutkiem, można zastosować kanoniczną metodę największej wiarygodności. W stosunku do kalibracji jest obarczona na ogół większym obciążeniem i błędem średniokwadratowym estymatora. Tylko w przypadku kopuli Gumbela otrzymane wyniki dyskwalifikują ten sposób estymacji.

Obie te metody estymacji mają pewną słabość w zastosowaniu praktycznym, tj. gdy dla posiadanej próby chcemy wybrać spośród wielu różnych funkcji łączących tę, która najlepiej odzwierciedla zależności w próbie. W przypadku kalibracji i $KMNW$ literatura zaleca, aby najpierw oszacować parametry wszystkich funkcji łączących na podstawie próby, a następnie na podstawie odrębnego kryterium wybrać tę najlepiej dopasowaną do danych. Heilpern [2007, s. 138-147] wskazuje jako potencjalne kryteria wyboru: analizę graficzną, testy zgodności lub porównanie z kopulą empiryczną. A zatem stosując metodę kalibracji lub $KMNW$, ostatecznego wyboru funkcji łączącej dla próby należy dokonać w dwóch etapach.

W przypadku metody minimalnej odległości drugi etap nie jest potrzebny, gdyż szacowanie parametru odbywa się przez minimalizację funkcji odległości między kopulą teoretyczną i kopulą empiryczną. Przy ustalonej funkcji odległości jako kryterium wyboru kopuli najlepiej dopasowanej do próby można użyć kryterium minimalnej wartości funkcji odległości.

Literatura

Cherubini U., Luciano E., Vecchiato W. (2004), *Copula Method in Finance*, John Wiley and Sons, Chichester.

Doman R. (2011), *Zastosowania kopuli w modelowaniu dynamiki zależności na rynkach finansowych*, Wydawnictwo Uniwersystetu Ekonomicznego, Poznań.

Heilpern S. (2007), *Funkcje łączące*, Wydawnictwo Akademii Ekonomicznej im. Oskara Langego, Wrocław.

Mendes B., de Melo E., Nelsen R. (2007), *Robust Fits for Copula Models*, Communications in Statistics − Simulation and Computation, s. 997-1017.

Nelsen R.B. (2006), *An Introduction to Copulas*, Springer, New York.

Tsukahara H. (2005), *Semiparametric Estimation in Copula Models*, "The Canadian Journal of Statistics", s. 357-375.

Weiss G.N.F. (2010), *Copula Parameter Estimation: Numerical Considerations and Implications for Risk Management*, "The Journal of Risk", s. 17-53.

Wolfowitz J. (1953), *Estimation by the Minimum Distance Method*, "Annals of the Institute of Statistical Mathematics", Vol. 5, s. 9-23.

THE COMPARISON OF PARAMETER ESTIMATION METHODS IN THE CLASS OF SOME TWO-DIMENSIONAL ARCHIMEDEAN COPULAS

Summary: The aim of the paper is to compare the accuracy of copula parameter estimation methods: canonical maximum likelihood method, method of calibration using Kendall's τ, interval estimation method and minimum distance method. The analysis was based on computer simulations of samples (every sample consisted of 250 elements) generated from distributions described by Clayton, Gumbel, Yager and Farlie-Gumbel-Morgenstern copulas and 20 randomly selected parameters for them. The methods were compared by the measures: bias and mean squared error of the estimator.

Keywords: copula, canonical maximum likelihood method, minimum distance method, Kendall's τ, interval estimation.