

Andrzej Stryjek

Szkoła Główna Handlowa w Warszawie

Porównanie estymacji parametru n -wymiarowej kopuli dla $2 \leq n \leq 4$ przy zastosowaniu kanonicznej metody największej wiarygodności i metody minimalnej odległości*

Streszczenie. W artykule dokonano porównania estymacji parametru kopuli, stosując kanoniczną metodę największej wiarygodności i metodę minimalnej odległości. W tej drugiej metodzie rozważano trzy warianty funkcji odległości: odległość Cramera von Mises, odległość Kołmogorowa–Smirnova i odległość z przestrzeni L^1 . Na podstawie przeprowadzonych symulacji badano obciążenie i błąd średniokwadratowy wymienionych estymatorów dla n -wymiarowych kopuli Claytona, Franka i Farlie–Gumbel–Morgensterna, gdzie $n=2$, $n=3$ lub $n=4$. Najczęściej najmniejszym obciążeniem i błędem średniokwadratowym charakteryzował się estymator otrzymany kanoniczną metodą największej wiarygodności.

Słowa kluczowe: kopula, estymacja, estymacja KMNW, estymacja metodą minimalnej odległości, kopula Claytona, kopula Franka, kopula Farlie–Gumbela–Morgensterna

Wstęp

W analizie zależności między zmiennymi losowymi, które są stosowane w modelach matematycznych i statystycznych, jednym z najpopularniejszych wskaźników jest współczynnik korelacji liniowej Pearsona. Miarą tą można posługiwać się w celu uwzględnienia w modelu liniowych zależności między badanymi zmiennymi.

* Prezentowane badanie symulacyjne zostało opracowane w ramach badania statutowego nr 03/S/0021/12 w Instytucie Ekonometrii Szkoły Głównej Handlowej w Warszawie (por. A. Stryjek, *Porównanie wybranych metod estymacji parametru rodziny n -wymiarowej kopuli dla $n \geq 2$* , Szkoła Główna Handlowa w Warszawie, Warszawa 2013).

W przypadku gdy analizowane zależności nie są liniowe, jako miarę zależności można wykorzystać współczynnik τ Kendalla lub ρ Spearmana. Każdy z tych wskaźników jest wartością liczbową, zatem w bardzo syntetyczny sposób wyraża stopień zależności między zmiennymi losowymi. Innym sposobem uwzględniania tych zależności jest wykorzystanie kopuli. Kopula to taka funkcja, która podaje zależność między dystrybuantą wielowymiarowej zmiennej losowej a dystrybuantami rozkładów brzegowych tej zmiennej. W literaturze można odnaleźć wiele przykładów kopuli¹, a dokładniej przykładów rodzin kopuli, w większości przypadków bowiem formuła analityczna kopuli zawiera parametr lub kilka parametrów. Parametry te są używane, aby za pomocą jednej rodziny kopuli wyrażać różne typy zależności między rozkładami brzegowymi.

Przykłady zastosowania kopuli w finansach i ubezpieczeniach zainteresowany czytelnik znajdzie w pracach Domana, Heilperna, Stryjka². Podstawowy problem, jaki pojawia się przy zastosowaniu kopuli w analizach empirycznych, to prawidłowe oszacowanie parametru lub parametrów rodziny kopuli na podstawie posiadanej próby. W wymienionej wyżej literaturze podstawową metodą estymacji parametru kopuli jest metoda największej wiarygodności oraz jej modyfikacje, tj. metoda wnioskowania dla rozkładów brzegowych oraz kanoniczna metoda największej wiarygodności. Innym sposobem wyznaczenia oszacowań parametrów kopuli jest zastosowanie metody minimalnej odległości. Jednak skuteczność tej metody nie jest szeroko dyskutowana w literaturze. Kwestia ta dotychczas została podjęta w zasadzie w dwóch artykułach³. Bieżący artykuł ma na celu włączenie się w dyskusję i próbę uzupełnienia przeprowadzonych analiz dotyczących estymacji parametru kopuli metodą minimalnej odległości.

1. Kopula. Metody estymacji parametrów rodziny kopuli

Omówienie rodzin kopuli oraz metod estymacji ich parametrów poprzedzone jest definicją pojęcia kopuli. n -wymiarową kopułą C , gdzie $n \geq 2$, jest każda

¹ U. Cherubini, E. Luciano, W. Vecchiato, *Copula method in finance*, John Wiley and Sons, Chichester 2004; S. Heilpern, *Funkcje łączące*, Wyd. AE im. Oskara Langego, Wrocław 2007; H. Joe, *Multivariate Models and Dependence Concepts*, Chappman & Hall/CRC, New York 1997; R.B. Nelsen, *An introduction to copulas*, Springer, New York 2006.

² R. Doman, *Zastosowania kopuli w modelowaniu dynamiki zależności na rynkach finansowych*, Wyd. UE w Poznaniu, Poznań 2011; S. Heilpern, op. cit.; A. Stryjek, *Porównanie wybranych metod estymacji parametru rodziny kopuli*, Szkoła Główna Handlowa, Warszawa 2011; A. Stryjek, *O pewnym problemie wyboru metody szacowania wartości zagrożonej*, Instytut Ekonometrii Szkoły Główniej Handlowej, Warszawa 2012.

³ H. Tsukahara, *Semiparametric estimation in copula models*, „The Canadian Journal of Statistics” 2005, vol. 33, z. 3; G.N.F. Weiss, *Copula parameter estimation: numerical considerations and implications for risk management*, „The Journal of Risk” 2010, t. 13, nr 1.

dystrybuanta n -wymiarowej zmiennej losowej $(X_1, X_2, \dots, X_n)$, której dziedziną została obcięta do kostki $[0,1]^n$, przy czym wszystkie rozkłady brzegowe, tzn. rozkłady zmiennych $X_1, X_2, \dots, X_n$, są jednostajne na odcinku $[0,1]$.

Jak już zostało wspomniane, kopula jest łącznikiem między rozkładem łącznym wielowymiarowej zmiennej losowej a rozkładami brzegowymi tej zmiennej. Własność ta jest konsekwencją podstawowego twierdzenia w teorii kopuli, czyli twierdzenia Sklara. Sformułowanie tego twierdzenia podajemy za Nelsenem⁴.

Niech H będzie dystrybuantą n -wymiarowej zmiennej losowej o rozkładach brzegowych określonych przez dystrybuanty $F_1, F_2, \dots, F_n$ oraz $\mathbb{R} = \mathbb{R} \cup \{-\infty, +\infty\}$. Wówczas istnieje n -wymiarowa kopula C taka, że dla każdego $x \in \mathbb{R}^n$ zachodzi

$$H(x_1, x_2, \dots, x_n) = C(F_1(x_1), F_2(x_2), \dots, F_n(x_n)). \quad (1)$$

Jeżeli funkcje $F_1, F_2, \dots, F_n$ są ciągłe, to C jest wyznaczona jednoznacznie. W przeciwnym przypadku jest ona wyznaczona jednoznacznie tylko na zbiorze $F_1(\mathbb{R}) \times F_2(\mathbb{R}) \times \dots \times F_n(\mathbb{R})$. Jeżeli zaś C jest n -wymiarową kopulą i $F_1, F_2, \dots, F_n$ są dystrybuantami rozkładów jednowymiarowych zmiennych losowych, to funkcja H zdefiniowana wzorem (1) jest dystrybuantą n -wymiarowej zmiennej losowej, której dystrybuanty rozkładów brzegowych to $F_1, F_2, \dots, F_n$.

Przykłady rodzin kopuli, które można odnaleźć w literaturze przedmiotu, można podzielić na trzy grupy: rodzina kopuli eliptycznych, rodzina kopuli archimedesowych i pozostałe rodziny kopuli.

Do rodziny kopuli eliptycznych zalicza się kopulę Gaussa, kopulę t Studenta i kopulę wykładniczo-potęgową⁵. Jednak ponieważ w klasie rozkładów eliptycznych adekwatną miarą zależności między zmiennymi losowymi jest współczynnik korelacji liniowej Pearsona, więc wykorzystanie tej grupy kopuli w wielu modelach statystycznych nie ma należytego uzasadnienia.

Ważną grupę rodzin kopuli, które wykorzystywane są w modelowaniu zależności między zmiennymi losowymi, stanowią kopule archimedesowe. Ich zaletą jest fakt, że każda rodzina jest parametryzowana tylko jednym parametrem. Ponadto są one opisane relatywnie nieskomplikowanymi formułami analitycznymi. Dla $n > 2$ można wskazać pięć rodzin kopuli archimedesowych:

$$C_\alpha(\mathbf{u}) = -\frac{1}{\alpha} \cdot \ln \left(1 + \frac{(e^{-\alpha u_1} - 1) \cdot \dots \cdot (e^{-\alpha u_n} - 1)}{(e^{-\alpha} - 1)^{n-1}} \right) \quad \text{dla } \alpha > 0, \quad (2)$$

$$C_\alpha(\mathbf{u}) = (u_1^{-\alpha} + u_2^{-\alpha} + \dots + u_n^{-\alpha} - n + 1)^{-\frac{1}{\alpha}} \quad \text{dla } \alpha > 0, \quad (3)$$

⁴ R.B. Nelsen, op. cit.

⁵ Por. S. Heilpern, op. cit.

$$C_{\alpha}(\mathbf{u}) = \exp\left(-\left((-\ln u_1)^{\alpha} + \dots + (-\ln u_n)^{\alpha}\right)^{\frac{1}{\alpha}}\right) \quad \text{dla } \alpha \geq 0, \quad (4)$$

$$C_{\alpha}(\mathbf{u}) = \frac{(1-\alpha) \cdot u_1 u_2 \cdot \dots \cdot u_n}{\prod_{i=1}^n (1-\alpha(1-u_i)) - \alpha \cdot u_1 u_2 \cdot \dots \cdot u_n} \quad \text{dla } \alpha \in [0,1], \quad (5)$$

$$C_{\alpha}(\mathbf{u}) = \exp\left(\frac{1}{\alpha} \cdot \left(1 - \prod_{i=1}^n (1 - \alpha \ln u_i)\right)\right) \quad \text{dla } \alpha \in (0,1], \quad (6)$$

gdzie $\mathbf{u} = (u_1, u_2, \dots, u_n)$.

Wzory (2)-(6) określają, odpowiednio, rodzinę kopuli Franka, Claytona, Gumbela, Ali–Mikhail–Haq i Gumbela–Barnetta.

W przypadku $n=2$ można wskazać wiele przykładów kopuli, nie tylko kopuli eliptycznych i archimedesowych⁶. Jednak dla $n>2$ nie ma ich tak szerokiego spektrum. Relatywnie prostym przykładem rodziny kopuli spoza dwóch grup opisanych wyżej jest rodzina Farlie–Gumbel–Morgensterna (FGM)

$$C_{\alpha}(\mathbf{u}) = \left(\prod_{i=1}^n u_i\right) \cdot \left(1 + \sum_{i_1 < i_2} \alpha_{i_1 i_2} \cdot (1-u_{i_1})(1-u_{i_2}) + \right. \\ \left. + \sum_{i_1 < i_2 < i_3} \alpha_{i_1 i_2 i_3} \cdot (1-u_{i_1})(1-u_{i_2})(1-u_{i_3}) + \alpha_{12\dots n} \prod_{i=1}^n (1-u_i)\right), \quad (7)$$

gdzie wszystkie parametry są z przedziału $[-1,1]$. Zauważmy, że dla $n=3$ rodzina (7) parametryzowana jest przez cztery, a dla $n=4$ aż przez jedenaście parametrów.

Następnie przyjmujemy, że $(x'_1, x'_2, \dots, x'_n)$, gdzie $n \geq 2$ i $t=1, 2, \dots, T$, oznacza próbę zawierającą T realizacji zmiennej losowej $(X_1, X_2, \dots, X_n)$. Poniżej omówimy dwa sposoby estymacji parametru α kopuli C_{α} na podstawie próby $(x'_1, x'_2, \dots, x'_n)$.

Pierwsza metoda estymacji parametru kopuli to metoda największej wiarygodności (MNW) oraz jej modyfikacja – kanoniczna metoda największej wiarygodności (KMNW). Obie metody są przykładem standardowych metod estymacji parametru kopuli zalecanych w literaturze⁷.

MNW wymaga przyjęcia założenia, że rodzina kopuli C_{α} jest parametryzowana przez wektor $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_k)$, gdzie $\alpha \in I_C$, a rozkłady brzegowe zmiennej $(X_1, X_2, \dots, X_n)$ są znane i określone przez dystrybuanty $F_1(x_1; \beta_1), F_2(x_2; \beta_2), \dots, F_n(x_n; \beta_n)$. Ponadto każda dystrybuanta $F_i(x_i; \beta_i)$, $1 \leq i \leq n$, jest zależna od parametrów, które tworzą wektor $\beta_i = (\beta_{i1}, \beta_{i2}, \dots, \beta_{ik_i})$, gdzie $k_i \geq 1$. Wówczas gęstość rozkładu łącznego zmiennej losowej $(X_1, X_2, \dots, X_n)$ przyjmuje postać

⁶ Por. R.B. Nelsen, op. cit.

⁷ Por. S. Heilpern, op. cit.; U. Cherubini, E. Luciano, W. Vecchiato, op. cit.; R. Doman, op. cit.

$$f(x_1, x_2, \dots, x_n; \alpha; \beta_1, \beta_2, \dots, \beta_n) = c_\alpha(F_1(x_1; \beta_1), F_2(x_2; \beta_2), \dots, F_n(x_n; \beta_n)) \cdot f_1(x_1; \beta_1) \cdot f_2(x_2; \beta_2) \cdot \dots \cdot f_n(x_n; \beta_n), \quad (8)$$

gdzie $c_\alpha(u_1, u_2, \dots, u_n) = \frac{\partial^n C_\alpha}{\partial u_1 \partial u_2 \dots \partial u_n}(u_1, u_2, \dots, u_n)$ jest gęstością kopuli $C_\alpha(u_1, u_2, \dots, u_n)$, a $f_1(x_1; \beta_1), f_2(x_2; \beta_2), \dots, f_n(x_n; \beta_n)$ to gęstości rozkładów brzegowych.

Estymator wyznacza się przez taki dobór wartości $(\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_n)$, aby zmaksymalizować logarytm funkcji wiarygodności

$$l^{\text{MKNW}}(\alpha, \beta_1, \beta_2, \dots, \beta_n) = \sum_{t=1}^T \ln \left(c_\alpha \left(F_1(x_t^t; \beta_1), F_2(x_t^t; \beta_2), \dots, F_n(x_t^t; \beta_n) \right) \right) + \sum_{t=1}^T \left(\ln f_1(x_t^t; \beta_1) + \ln f_2(x_t^t; \beta_2) + \dots + \ln f_n(x_t^t; \beta_n) \right). \quad (9)$$

Jeżeli przyjmiemy, że rozkłady brzegowe zmiennej $(X_1, X_2, \dots, X_n)$ są reprezentowane przez dystrybuanty empiryczne $F_i^T(x_i) = \frac{1}{T+1} \cdot \sum_{t=1}^T \mathbb{1}_{\{x_i^t: x_i^t \leq x_i\}}$ dla $i \in \{1, 2, \dots, n\}$, przy czym symbol $\mathbb{1}_A$ oznacza funkcję charakterystyczną⁸ zbioru A , to wzór (9) przyjmuje postać:

$$l^{\text{KMNW}}(\alpha) = \sum_{t=1}^T \ln \left(c_\alpha \left(\frac{r_1^t}{T+1}, \frac{r_2^t}{T+1}, \dots, \frac{r_n^t}{T+1} \right) \right), \quad (10)$$

gdzie r_i^t oznacza rangę obserwacji x_i^t dla $i = 1, 2, \dots, n$ oraz $t = 1, 2, \dots, T$. Estymator $\hat{\alpha}$, który maksymalizuje funkcję (10), nazywa się estymatorem otrzymanym kanoniczną metodą największej wiarygodności (KMNW). Jak podaje Heilpern⁹, estymatory MNW i KMNW są asymptotycznie normalne. Estymator KMNW jest estymatorem zgodnym¹⁰.

Metoda minimalnej odległości pojawiła się w statystyce w 1953. J. Wolfowitz w swojej pracy¹¹ zaproponował, aby odległość $\delta(X, Y)$ między zmiennymi losowymi X i Y mierzyć jako supremum modułu różnicy między wartościami dystrybuant obu zmiennych w takich samych punktach po wszystkich argumentach rzeczywistych, czyli

$$\delta(X, Y) = \sup_{x \in \mathbb{R}} |F_X(x) - F_Y(x)|. \quad (11)$$

⁸ Funkcja charakterystyczna zbioru A to funkcja $\mathbb{1}_A: A \rightarrow \{0, 1\}$ określona warunkami: $\mathbb{1}_A(t) = 1$, gdy $t \in A$ oraz $\mathbb{1}_A(t) = 0$, gdy $t \notin A$.

⁹ S. Heilpern, op. cit., s. 134-136.

¹⁰ C. Genest, K. Ghoudi, L. -P. Rivest, *A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions*, „Biometrika” 1995.

¹¹ J. Wolfowitz, *Estimation by the minimum distance method*, „Annals of the Institute of Statistical Mathematics” 1953, t. 5.

Miarę określoną wzorem (11) można uogólnić przez zastosowanie różnych funkcji odległości d , tj.

$$\delta(X, Y) = d(F_X, F_Y). \quad (12)$$

Jeżeli we wzorze (12) przyjmiemy jedną z dystrybuant jako dystrybuantę empiryczną wyznaczoną na podstawie próby, to estymacja parametrów drugiej dystrybuanty w przypadku zastosowania metody minimalnej odległości polega na takim wyborze wartości jej parametrów, aby odległość $\delta(X, Y)$ była minimalna.

Metodę minimalnej odległości można zastosować do estymacji parametrów rodziny kopuli, gdyż kopule są szczególnym przypadkiem dystrybuant. Analogicznie do pojęcia dystrybuanty empirycznej definiuje się pojęcie kopuli empirycznej¹². Jest to funk-

cja określona na zbiorze $\left\{ \left(\frac{i_1}{T}, \frac{i_2}{T}, \dots, \frac{i_n}{T} \right) : i_j \in \{0, 1, 2, \dots, T\} \text{ dla } j \in \{1, 2, \dots, n\} \right\}$, zdefiniowana jako

$$C_T \left(\frac{i_1}{T}, \frac{i_2}{T}, \dots, \frac{i_n}{T} \right) = \frac{1}{T} \cdot \sum_{t=1}^T \mathbb{1}_{\{r_t^1: r_t^1 \leq i_1\}} \cdot \mathbb{1}_{\{r_t^2: r_t^2 \leq i_2\}} \cdot \dots \cdot \mathbb{1}_{\{r_t^n: r_t^n \leq i_n\}}, \quad (13)$$

gdzie r_t^i oznacza rangę obserwacji x_t^i dla $i = 1, 2, \dots, n$ oraz $t = 1, 2, \dots, T$.

W badaniu symulacyjnym zaprezentowanym w dalszej części artykułu jako miary odległości między kopulami zostały użyte funkcje odległości: odległość Cramera von Mises (δ^1), odległość Kołmogorowa–Smirnova (δ^2) i odległość indukowana przez normę z przestrzeni L^1 (δ^3). Są one zdefiniowane wzorami (14)–(16).

$$\delta^1(C_1, C_2) = \int_{[0,1]^n} (C_1(\mathbf{u}) - C_2(\mathbf{u}))^2 d\mathbf{u}, \quad (14)$$

$$\delta^2(C_1, C_2) = \sup_{\mathbf{u} \in [0,1]^n} |C_1(\mathbf{u}) - C_2(\mathbf{u})|, \quad (15)$$

$$\delta^3(C_1, C_2) = \int_{[0,1]^n} |C_1(\mathbf{u}) - C_2(\mathbf{u})| d\mathbf{u}. \quad (16)$$

W pracy Tsukahary¹³ wykazano, że estymator metodą minimalnej odległości dla funkcji odległości (14) jest asymptotycznie normalny.

Na zakończenie zauważmy, że w przypadku oszacowania parametrów rodziny kopuli C_{α}^A i innej rodziny C_{ϕ}^B na podstawie próby $(x_1^t, x_2^t, \dots, x_n^t)$, gdzie $n \geq 2$ i $t = 1, 2, \dots, T$, za pomocą metody minimalnej odległości, w naturalny sposób pojawia się też kryterium wyboru najlepiej dopasowanej do tej próby kopuli spośród

¹² U. Cherubini, E. Luciano, W. Vecchiato, op. cit., s. 161.

¹³ H. Tsukahara, *Semiparametric estimation in copula models*, „The Canadian Journal of Statistics” 2005, vol. 33, z. 3.

obu oszacowanych. Optymalnym wyborem jest ten estymator, dla którego otrzymujemy najmniejszą wartość przyjętej funkcji odległości δ . Dla kanonicznej metody największej wiarygodności takiego wniosku nie można sformułować bez dodatkowych kryteriów i obliczeń.

2. Opis przeprowadzonych symulacji

Przeprowadzone badanie symulacyjne polegało na tym, że dla ustalonej rodziny kopuli i przyjętych z góry dla tej rodziny wartości jej parametrów, wygenerowano 1000 prób z rozkładu opisanego tą kopulą. Następnie na podstawie tych prób oszacowano parametry rozpatrywanej rodziny za pomocą kanonicznej metody największej wiarygodności i metody minimalnej odległości. Celem badania jest ustalenie, która z metod estymacji generuje oszacowania bliższe prawdziwym wartościom parametrów kopuli.

W badaniu opisanym w bieżącym artykule rozważane są trzy rodziny kopuli: kopula Clayтона (3), kopula Franka (2) i kopula FGM (7). Oprócz przypadku kopuli dwuwymiarowej ($n=2$), analizie podlegały kopule dla $n=3$ oraz dla $n=4$.

Analiza skuteczności kanonicznej metody największej wiarygodności i minimalnej odległości (z funkcją odległości Cramera von Mises) dla $n=2$ była przedmiotem rozważań w pracy Stryjka¹⁴. Wybór wszystkich wartości parametrów kopuli wynikał tam z próby takiego doboru tych wielkości, aby dla dwuwymiarowej rodziny kopuli opisywały one różne typy zależności (dodatnia lub ujemna korelacja) między zmiennymi. W bieżącym badaniu przyjęto wartości parametrów kopuli tak samo jak w innej pracy wymienionego autora¹⁵, z wyjątkiem parametru $\alpha=-5$ dla kopuli Franka, wartość ta bowiem jest poza zakresem zmienności parametru rodziny Franka w wielowymiarowym przypadku ($n \geq 2$).

Dla kopuli Clayтона rozpatrywano następujące wartości parametru: $\alpha=1$, $\alpha=5$ i $\alpha=15$, dla kopuli Franka $\alpha=-5$ i $\alpha=3$ w przypadku $n=2$ oraz $\alpha=3$ dla $n>2$. Natomiast dla rodziny FGM przyjęto $\alpha=-0,9$, $\alpha=0,1$ i $\alpha=0,9$ w przypadku dwuwymiarowym. Ponieważ dla $n=3$ ta rodzina jest parametryzowana przez cztery parametry, a dla $n=4$ przez jedenaście parametrów, więc przyjęto w obu tych wariantach, że wszystkie parametry będą jednocześnie równe $-0,9$ lub $0,1$, lub $0,9$.

Ponieważ wraz ze wzrostem wymiaru kopuli n , istotnie wydłużał się czas obliczeń, więc w badaniu przyjęto w zależności od n kilka wariantów liczebności generowanych prób. Dla kopuli dwuwymiarowej rozpatrywano: $T=20$, $T=40$

¹⁴ A. Stryjek, *Porównanie wybranych metod estymacji parametru rodziny n -wymiarowej kopuli dla $n \geq 2$* , Szkoła Główna Handlowa, Warszawa 2013.

¹⁵ A. Stryjek, *Użyteczność kopuli w finansach i ubezpieczeniach*, „Zeszyty Naukowe WSB we Wrocławiu” 2011, nr 20.

i $T=120$, dla $n=3$ rozpatrywano: $T=20$, $T=40$ i $T=80$, a dla $n=4$ dwie wersje liczebności¹⁶: $T=20$ i $T=40$.

Dla ustalonej rodziny kopuli C_α , ustalonej wartości parametru α oraz ustalonej długości próby T symulacje przeprowadzano według następującego schematu:

1. Wygenerowanie n -wymiarowej próby o liczebności T z rozkładu opisanego kopulą C_α .

2. Wyznaczenie na podstawie otrzymanej próby wartości estymatorów parametru α czterema metodami: kanoniczną metodą największej wiarygodności oraz metodą minimalnej odległości z odległością Cramera von Mises (CvM), odległością Kołmogorowa–Smirnova (KS) i odległością z przestrzeni L^1 .

3. Powtórzenie pierwszego i drugiego etapu 1000 razy.

Jako rezultat powyższego algorytmu otrzymywano szereg zawierający 1000 wartości będących oszacowaniami parametru α kopuli C_α (przy ustalonej długości próby i metodzie estymacji). Szereg ten w dalszej części artykułu będzie nazywany empirycznym rozkładem estymatora parametru α . Przy czym dla kopuli FGM otrzymano dla $n=3$ – cztery, a dla $n=4$ – jedenaście 1000-elementowych szeregów. Zatem, podobnie jak dla pozostałych rodzin kopuli, dla kopuli FGM również dla każdego jej parametru otrzymano empiryczny rozkład estymatora.

Generowanie jednego n -wymiarowego elementu próby dla zadanej rodziny kopuli oraz jej parametru przeprowadzono metodą próbkowania warunkowego (*conditional sampling*). Algorytm podajemy za Heilpernem¹⁷:

1) wygenerować n liczb losowych: $v_1, v_2, \dots, v_n$ z rozkładu jednostajnego na przedziale $[0,1]$,

2) przyjąć $u_1 = v_1$,

3) wyznaczyć kolejno wartości $u_2, u_3, \dots$ jako rozwiązania odpowiednich równań postaci:

$$v_2 = \frac{\frac{\partial}{\partial u_1} C(u_1, u_2, 1, \dots, 1)}{\frac{\partial}{\partial u_1} C(u_1, 1, \dots, 1)},$$

$$v_3 = \frac{\frac{\partial^2}{\partial u_1 \partial u_2} C(u_1, u_2, u_3, 1, \dots, 1)}{\frac{\partial^2}{\partial u_1 \partial u_2} C(u_1, u_2, 1, \dots, 1)},$$

.....

¹⁶ Dla rodziny FGM przeprowadzono obliczenia wyłącznie dla $T=20$.

¹⁷ S. Heilpern, op. cit., s. 100.

$$v_k = \frac{\frac{\partial^{k-1}}{\partial u_1 \partial u_2 \dots \partial u_{k-1}} C(u_1, u_2, \dots, u_{k-1}, u_k, 1, \dots, 1)}{\frac{\partial^{k-1}}{\partial u_1 \partial u_2 \dots \partial u_{k-1}} C(u_1, u_2, \dots, u_{k-1}, 1, 1, \dots, 1)},$$

dla $k=2, 3, \dots, n$.

Dla $n > 2$ wyznaczenie wartości $u_2, u_3, \dots, u_n$ sprowadza się do rozwiązania odpowiednich równań określonych w punkcie 3) powyższego algorytmu. Ich postać analityczna jest na ogół skomplikowana, zatem szukane rozwiązanie należy wyznaczyć w sposób numeryczny. W prezentowanym badaniu wykorzystano funkcję `uniroot.all` dostępną w pakiecie R-CRAN 3.1.1. Formuły analityczne równań 3) dla kopuli Claytona, Franka i FGM zostały wyznaczone w programie Wolfram Mathematica 9, a następnie zostały zaadaptowane do pakietu R-CRAN 3.1.1. Ze względu na ich stopień skomplikowania nie są zamieszczone w artykule, lecz mogą być zainteresowanym czytelnikom udostępnione przez autora w formie elektronicznej. Maksymalizację logarytmu funkcji wiarygodności i minimalizację każdej ze stosowanych funkcji odległości przeprowadzono z wykorzystaniem procedury `nmlinb`¹⁸ dostępnej w pakiecie R-CRAN 3.1.1.

Podsumowując, należy stwierdzić, że badanie symulacyjne polegało na tym, iż dla każdego ustalonego przypadku rodziny kopuli (Claytona, Franka lub FGM), wartości jej parametru (parametrów) oraz wybranej metody estymacji (kanonicznej metody największej wiarygodności, metody minimalnej odległości z odległością Cramera von Mises, Kołmogorowa–Sminova lub z przestrzeni L^1) otrzymano 1000 oszacowań znanego *a priori* parametru.

3. Otrzymane wyniki

Analizie poddano 56 różnych przypadków otrzymanych jako wynik rozważania różnych liczebności symulowanych prób (T), trzech rodzin n -wymiarowych kopuli, $n=2, 3, 4$ oraz kilku wariantów wartości parametrów α dla tych rodzin kopuli. Ponieważ trójwymiarowa ($n=3$) rodzina kopuli FGM jest zależna od czterech parametrów, a czterowymiarowa ($n=4$) od jedenastu parametrów, więc ostatecznie dla każdej metody szacowania parametru kopuli (KMNW, L^1 , CvM i KS) otrzymano 113 różnych empirycznych rozkładów estymatorów. Dlatego też w tabelach podsumowujących wyniki analizy (tab. 4 i tab. 5) dla kopuli FGM ($n=3$ i $n=4$) zamieszczona jest dodatkowa, pierwsza kolumna, która zawiera numer

¹⁸ Opis procedury dostępny na stronie <http://netlib.bell-labs.com/cm/cs/cstr/153.pdf> [1.10.2013].

szacowanego parametru tej rodziny, tzn. dla $n=3$ są to wartości od 1 do 4, a dla $n=4$ od 1 do 11.

Dla czterech zastosowanych wariantów szacowania parametrów otrzymano, że empiryczne rozkłady estymatorów nie są rozkładami normalnymi. Nie ma też zgodności między tymi rozkładami, a w znacznej części przypadków średnia w otrzymanych rozkładach symulacyjnych nie jest równa prawdziwej wartości szacowanego parametru.

Brak podstaw do odrzucenia hipotezy o tym, że otrzymane w symulacji wartości pochodzą z rozkładu normalnego na podstawie testu Shapiro–Wilka, występuje w zaledwie kilku przypadkach. Po jednym przypadku na 113 badanych dla estymatora KMNW i minimalnej odległości z funkcją odległości z przestrzeni L^1 , dwa przypadki dla odległości CvM i trzy dla odległości KS.

Ponadto dla empirycznych rozkładów estymatorów na ogół brak jest zgodności między rozkładem estymatora KMNW a estymatorem z metody minimalnej odległości dla wszystkich trzech funkcji odległości (L^1 , CvM i KS). Fakt ten weryfikowany był na podstawie testu Kołmogorowa–Smirnova. Największy odsetek zbadanych przypadków (55%), dla których nie ma podstaw do odrzucenia hipotezy, że rozkłady estymatorów są zgodne, występuje dla estymatora z odległością L^1 i CvM. Dla pozostałych par estymatorów zgodność występuje w 25–30% przypadków. Najmniej takich przypadków zarejestrowano dla KMNW oraz KS (zaledwie 13%).

Za podstawowe wskaźniki, które należy uznać jako decydujące o ocenie skuteczności badanych metod, trzeba przyjąć obciążenie estymatora oraz jego błąd średniokwadratowy (MSE).

Przypomnijmy, że obciążenie estymatora to różnica między średnią w rozkładzie empirycznym estymatora a prawdziwą wartością parametru. W tabeli 1 podano obciążenie dla każdej z czterech metod estymacji w podziale na trzy rodziny kopuli. Na podstawie 23 badanych przypadków dla rodziny kopuli Claytona, 11 przypadków dla rodziny kopuli Franka i 79 przypadków dla rodziny kopuli FGM, estymator otrzymany KMNW charakteryzuje się najmniejszym obciążeniem dla pierwszej i trzeciej z rozpatrywanych rodzin kopuli. Dla kopuli Claytona stosunkowo małe obciążenie jest także dla estymatora z odległością CvM. Najmniejsze obciążenie dla kopuli Franka otrzymano dla metody minimalnej odległości z odległością z przestrzeni L^1 .

W tabeli 2 podano obciążenie estymatorów, uwzględniając dodatkowo wymiar kopuli n . Dla dwuwymiarowej kopuli Claytona najmniejsze obciążenie jest dla metody minimalnej odległości CvM, a w przypadku trójwymiarowej kopuli Franka i dwuwymiarowej kopuli FGM dla metody minimalnej odległości z odległością KS. W pozostałych przypadkach najmniejsze obciążenie mają estymatory KMNW.

Z analizy obciążenia estymatorów dla wszystkich 113 przypadków (por. tab. 4) wynika, że KMNW najczęściej charakteryzowała się najmniejszym obciążeniem.

Tabela 1. Porównanie obciążenia badanych estymatorów

Kopula	KMNW	L^1	CvM	KS
	obciążenie			
Claytona	0,0971	17,7991	-0,1547	-1,1886
Franka	-0,0283	0,0270	0,0721	0,0800
FGM	0,0021	0,0577	0,0625	0,0706

Źródło: opracowanie własne na podstawie przeprowadzonych symulacji.

Tabela 2. Porównanie obciążenia badanych estymatorów w zależności od wymiaru kopuli n

Kopula	n	KMNW	L^1	CvM	KS
		obciążenie			
Claytona	2	0,1642	13,0076	0,1442	-1,2235
	3	0,0585	14,5629	-0,2890	-1,1414
	4	0,0544	29,8407	-0,4018	-1,2071
Franka	2	0,0380	-0,0550	0,0520	0,0935
	3	-0,3354	0,0844	0,0436	0,0134
	4	0,2333	0,1872	0,1752	0,1391
FGM	2	0,0166	0,0121	0,0110	0,0098
	3	-0,0111	0,0212	0,0185	0,0292
	4	0,0126	0,1100	0,1245	0,1324

Źródło: opracowanie własne na podstawie przeprowadzonych symulacji.

Własność ta wystąpiła w 42% przypadków. Dla metody minimalnej odległości oraz funkcji odległości z przestrzeni L^1 , CvM oraz KS udział takich przypadków wynosił, odpowiednio, 4%, 25% i 30%. Odsetek przypadków o najmniejszym obciążeniu dla KMNW zmniejsza się wraz ze wzrostem wymiaru kopuli (por. tab. 3). Dzieje się tak głównie na rzecz metody minimalnej odległości z funkcją odległości CvM i KS.

Tabela 3. Odsetek przypadków o najmniejszym obciążeniu estymatora w zależności od wymiaru kopuli n (%)

n	KMNW	L^1	CvM	KS
2	67	0	17	17
3	38	6	25	31
4	32	2	29	37

Źródło: opracowanie własne na podstawie przeprowadzonych symulacji.

Tabela 4. Porównanie obciążenia badanych estymatorów

Numer parametru	Kopula	n	Wartość parametru	T	KMNW	L^1	CvM	KS
					obciążenie			
1	Clayton	2	1	20	-0,0171	0,3265	0,1469	0,0543
1		2	1	40	-0,0277	0,1189	0,0664	0,0369
1		2	1	120	-0,0121	0,0543	0,0405	0,0283
1		2	5	20	0,1965	6,0934	0,2889	-0,6077
1		2	5	40	0,1608	1,0643	0,2057	-0,2130
1		2	5	120	0,0222	0,2452	0,0562	-0,0646
1		2	15	20	0,6956	94,8658	0,9771	-5,7421
1		2	15	40	0,3358	12,7199	-0,3292	-3,2951
1		2	15	120	0,1236	1,5798	-0,1544	-1,2088
1		3	1	20	-0,0555	0,0931	0,0268	0,0022
1		3	1	40	-0,0406	0,0584	0,0361	0,0371
1		3	1	80	-0,0289	0,0066	-0,0041	-0,0100
1		3	5	20	0,0859	2,4089	0,0197	-0,5738
1		3	5	40	0,0729	0,5195	-0,0473	-0,2897
1		3	5	80	0,0341	0,2492	0,0200	-0,0795
1		3	15	20	0,2491	116,6635	-0,9180	-4,6986
1		3	15	40	0,1498	9,2009	-0,9897	-2,9763
1		3	15	80	0,0597	1,8664	-0,7444	-1,6843
1		4	1	20	-0,0326	0,0953	0,0538	0,0734
1		4	1	40	-0,0406	0,0584	0,0361	0,0371
1		4	5	20	0,0906	3,5061	0,0624	-0,2867
1		4	5	40	0,0224	0,5480	0,0195	-0,1187
1		4	15	20	0,1971	159,2171	-1,1770	-4,2328
1		4	15	40	0,0896	15,6196	-1,4055	-2,7151
1	FGM	2	-0,9	20	0,1923	0,2085	0,2200	0,2305
1		2	-0,9	40	0,1432	0,1595	0,1702	0,1879
1		2	-0,9	120	0,0551	0,0781	0,0818	0,0895
1		2	0,1	20	-0,0259	-0,0478	-0,0526	-0,0581
1		2	0,1	40	0,0315	0,0192	0,0183	0,0227
1		2	0,1	120	0,0457	0,0434	0,0419	0,0382
1		2	0,9	20	-0,1661	-0,1966	-0,2134	-0,2391
1		2	0,9	40	-0,0991	-0,1177	-0,1264	-0,1394
1	Frank	2	0,9	120	-0,0277	-0,0375	-0,0411	-0,0441
1		2	-5	20	-0,0869	-0,7816	-0,0091	0,2871
1		2	-5	40	-0,0983	-0,2420	0,0082	0,1926
1		2	-5	120	-0,0075	-0,0140	0,0450	0,0797
1		2	3	20	0,1715	0,4045	0,0904	-0,0799
1		2	3	40	0,1830	0,2069	0,1027	0,0268
1		2	3	120	0,0663	0,0961	0,0750	0,0547
1		3	3	20	-0,0830	0,1918	0,1110	0,0678
1		3	3	40	-0,4905	0,0268	-0,0008	-0,0270
1		3	3	80	-0,4327	0,0346	0,0205	-0,0005
1		4	3	20	0,0245	0,0052	-0,0037	-0,0413
1		4	3	40	0,4421	0,3691	0,3540	0,3195

cd. tab. 4

Numer parametru	Kopula	n	Wartość parametru	T	KMNW	L^1	CvM	KS
					obciążenie			
1	FGM	3	-0,9	20	0,0269	0,1253	0,1562	0,1737
1		3	-0,9	40	-0,0675	0,0433	0,0652	0,0705
1		3	-0,9	80	-0,0903	-0,0066	0,0183	0,0301
1		3	0,1	20	-0,0053	-0,0163	-0,0136	0,0067
1		3	0,1	40	-0,0090	-0,0005	-0,0055	0,0027
1		3	0,1	80	0,0167	0,0236	0,0184	0,0112
1		3	0,9	20	-0,1690	-0,1614	-0,1660	-0,1381
1		3	0,9	40	-0,0929	-0,1004	-0,1047	-0,0779
1		3	0,9	80	-0,0334	-0,0602	-0,0613	-0,0317
2		3	-0,9	20	0,1134	0,2599	0,2752	0,2243
2		3	-0,9	40	0,0037	0,1942	0,2112	0,1564
2		3	-0,9	80	-0,0478	0,1390	0,1622	0,1108
2		3	0,1	20	-0,0123	-0,0189	-0,0197	-0,0010
2		3	0,1	40	0,0047	0,0158	0,0096	0,0124
2		3	0,1	80	0,0027	-0,0043	-0,0094	0,0080
2		3	0,9	20	-0,2335	-0,1899	-0,1953	-0,1594
2		3	0,9	40	-0,0937	-0,0781	-0,0828	-0,0578
2		3	0,9	80	-0,0330	-0,0510	-0,0541	-0,0366
3		3	-0,9	20	0,0522	0,1699	0,1927	0,1869
3		3	-0,9	40	0,0010	0,1917	0,2087	0,1765
3		3	-0,9	80	-0,0388	0,1298	0,1502	0,0951
3	FGM	3	0,1	20	0,0097	-0,0094	-0,0053	0,0269
3		3	0,1	40	0,0054	-0,0054	-0,0068	0,0141
3		3	0,1	80	-0,0110	-0,0149	-0,0171	-0,0079
3		3	0,9	20	-0,1734	-0,1588	-0,1612	-0,1101
3		3	0,9	40	-0,0801	-0,0888	-0,0925	-0,0591
3		3	0,9	80	-0,0292	-0,0419	-0,0439	-0,0218
4		3	-0,9	20	0,5228	0,2663	0,1587	0,3762
4		3	-0,9	40	0,4007	0,5143	0,2648	0,3243
4		3	-0,9	80	0,1902	0,7562	0,3937	0,3010
4		3	0,1	20	-0,0319	-0,1908	-0,0209	0,0434
4		3	0,1	40	-0,0384	-0,1487	-0,0363	0,0039
4		3	0,1	80	-0,0217	-0,0819	-0,0175	-0,0190
4		3	0,9	20	-0,3310	-0,2622	-0,1986	-0,2098
4		3	0,9	40	-0,1154	-0,2082	-0,1671	-0,1860
4		3	0,9	80	0,0088	-0,1656	-0,1400	-0,1861
1	FGM	4	-0,9	20	-0,0114	0,0670	0,1413	0,1991
1		4	0,1	20	-0,0263	0,0006	-0,0153	-0,0114
1		4	0,9	20	-0,2185	-0,1478	-0,1486	-0,0858
2		4	-0,9	20	0,0754	0,1418	0,2229	0,2241
2		4	0,1	20	-0,0257	0,0215	0,0078	0,0382
2		4	0,9	20	-0,1783	-0,1275	-0,1300	-0,0857
3		4	-0,9	20	0,3196	0,3460	0,4561	0,4503
3		4	0,1	20	-0,0181	0,0067	-0,0065	-0,0105
3		4	0,9	20	-0,2192	-0,1176	-0,1204	-0,0780

Numer parametru	Kopula	n	Wartość parametru	T	KMNW	L^1	CvM	KS
					obciążenie			
4	FGM	4	-0,9	20	0,0421	0,1303	0,2074	0,2010
4		4	0,1	20	0,0189	0,0207	0,0103	0,0400
4		4	0,9	20	-0,1988	-0,1262	-0,1294	-0,0930
5		4	-0,9	20	0,2928	0,3357	0,4474	0,4378
5		4	0,1	20	-0,0056	-0,0248	-0,0307	0,0087
5		4	0,9	20	-0,2226	-0,1286	-0,1320	-0,0744
6		4	-0,9	20	0,2327	0,2945	0,3890	0,3923
6		4	0,1	20	0,0081	-0,0082	-0,0194	0,0338
6		4	0,9	20	-0,2343	-0,1243	-0,1267	-0,0779
7		4	-0,9	20	0,4821	0,9318	0,3394	0,5176
7		4	0,1	20	-0,0494	-0,3053	-0,0427	0,0779
7		4	0,9	20	-0,3363	-0,2775	-0,2156	-0,1795
8		4	-0,9	20	0,7406	1,1753	0,8651	0,7296
8		4	0,1	20	-0,1177	-0,3062	-0,0669	0,0555
8		4	0,9	20	-0,4463	-0,2983	-0,2274	-0,1736
9		4	-0,9	20	0,6947	1,1692	0,8817	0,7342
9		4	0,1	20	-0,0527	-0,3163	-0,0332	0,0833
9		4	0,9	20	-0,3770	-0,2897	-0,2102	-0,1755
10		4	-0,9	20	0,6625	1,1679	0,8780	0,7033
10		4	0,1	20	-0,1042	-0,3241	-0,0559	0,0882
10		4	0,9	20	-0,3941	-0,2500	-0,1764	-0,1667
11		4	-0,9	20	1,0254	1,8599	1,5697	0,8811
11		4	0,1	20	-0,0416	-0,4288	-0,1359	0,0378
11		4	0,9	20	-0,9023	-0,4383	-0,2845	-0,3523

Źródło: opracowanie własne na podstawie przeprowadzonych symulacji.

Drugim kryterium porównania metod estymacji parametru kopuli jest wielkość błędu średniokwadratowego (MSE). Błąd ten to średnia z kwadratów odchyleń otrzymanych oszacowań od prawdziwej wartości parametru. O ile kryterium obciążenia estymatora klasyfikowało KMNW jako lepszą od pozostałych, o tyle według kryterium błędu MSE takie stwierdzenie jest mniej prawomocne. W 49% przypadków KMNW ma najmniejszy błąd średniokwadratowy. Metoda minimalnej odległości (KS) ma ten udział mniejszy o 7 punktów procentowych, czyli 42%, natomiast w przypadku odległości z przestrzeni L^1 i CvM udziały są równe, odpowiednio 4% i 6%.

Wyniki błędu MSE dla wszystkich empirycznych rozkładów estymatorów z badania symulacyjnego są zamieszczone w tabeli 5. Niski błąd średniokwadratowy występuje dla metody minimalnej odległości z funkcją odległości KS na ogół dla kopuli FGM. Skuteczność tej metody w świetle MSE zwiększa się wraz ze wzrostem wymiaru kopuli n . Fakt ten ma także potwierdzenie w porównaniu udziału przypadków o najniższych wartościach błędu MSE wymiarem kopuli (por. tab. 6). Udział takich przypadków wraz ze wzrostem gwałtownie maleje dla KMNW na korzyść metody z odległością KS.

Tabela 5. Porównanie błędu średniokwadratowego (MSE) otrzymanych symulacyjnych rozkładów badanych estymatorów

Numer parametru	Kopula	n	α	T	KMNW	L^1	CvM	KS
					MSE			
	Claytona	2	1	20	0,3859	1,2926	0,6785	0,5570
			1	40	0,2235	0,2959	0,2426	0,2512
			1	120	0,0841	0,0793	0,0740	0,0741
			5	20	1,5535	544,6747	7,0511	4,0616
			5	40	0,8159	7,1865	3,0828	2,4373
			5	120	0,2347	1,0334	0,8397	0,8844
			15	20	11,3690	18875,5398	134,0357	39,8416
			15	40	5,0697	699,8849	22,8199	22,4241
			15	120	1,6862	12,0904	6,1537	7,7062
		3	1	20	0,2218	0,3638	0,2768	0,3128
			1	40	0,1420	0,1701	0,1574	0,1951
			1	80	0,0655	0,0694	0,0674	0,0823
			5	20	0,6947	82,6577	4,1683	3,2131
			5	40	0,3658	3,4576	1,9902	1,9816
			5	80	0,1671	1,0690	0,8443	0,9624
			15	20	4,8042	24512,3914	48,8800	35,6166
			15	40	2,4416	308,2991	13,6975	17,9669
			15	80	1,1025	15,7669	6,6692	9,4160
		4	1	20	0,1323	0,3254	0,2611	0,3367
			1	40	0,1420	0,1701	0,1574	0,1951
			5	20	0,4992	345,5859	3,4911	3,2051
			5	40	0,2270	2,6065	1,4445	1,6915
			15	20	3,2322	34549,5585	27,0495	30,2671
			15	40	1,7446	1394,0146	12,4052	16,8687
	FGM	2	-0,9	20	0,2117	0,2230	0,2254	0,2369
			-0,9	40	0,1128	0,1257	0,1300	0,1468
			-0,9	120	0,0339	0,0440	0,0446	0,0498
			0,1	20	0,4264	0,4458	0,4198	0,4209
			0,1	40	0,2270	0,2277	0,2207	0,2402
			0,1	120	0,0799	0,0803	0,0784	0,0866
			0,9	20	0,1825	0,2164	0,2209	0,2485
			0,9	40	0,0856	0,0999	0,1025	0,1138
			0,9	120	0,0283	0,0309	0,0318	0,0359
1	Frank	2	-5	20	3,1177	9,7600	4,4830	4,1895
1		2	-5	40	1,5051	2,5390	1,9116	1,7819
1		2	-5	120	0,4130	0,5906	0,5517	0,6121
1		2	3	20	2,7950	5,1392	3,3918	3,0577
1		2	3	40	1,2033	1,6039	1,4030	1,4582
1		2	3	120	0,3395	0,4569	0,4331	0,4492
1		3	3	20	2,0480	2,0761	1,7431	1,7070
1		3	3	40	2,4828	0,8818	0,8330	0,8622
1		3	3	80	1,8526	0,4111	0,3984	0,4366
1		4	3	20	0,5848	1,4834	1,3452	1,3929
1		4	3	40	1,2108	1,5061	1,4046	1,3917

Numer parametru	Kopula	n	α	T	KMNW	L^1	CvM	KS
					MSE			
1		3	-0,9	20	0,1176	0,1679	0,1859	0,1896
1		3	-0,9	40	0,0198	0,0722	0,0810	0,0885
1		3	-0,9	80	0,0114	0,0300	0,0372	0,0447
1		3	0,1	20	0,4907	0,4169	0,4007	0,3919
1		3	0,1	40	0,2921	0,2856	0,2836	0,2756
1		3	0,1	80	0,1318	0,1569	0,1543	0,1625
1		3	0,9	20	0,2546	0,2126	0,2118	0,1945
1		3	0,9	40	0,1172	0,1069	0,1076	0,0977
1		3	0,9	80	0,0489	0,0624	0,0623	0,0534
2		3	-0,9	20	0,1992	0,3038	0,3112	0,2373
2		3	-0,9	40	0,0509	0,1794	0,1836	0,1481
2		3	-0,9	80	0,0196	0,0965	0,1041	0,0796
2		3	0,1	20	0,4997	0,4339	0,4187	0,4041
2		3	0,1	40	0,2911	0,2871	0,2802	0,2721
2		3	0,1	80	0,1437	0,1674	0,1640	0,1637
2		3	0,9	20	0,3331	0,2391	0,2400	0,2080
2		3	0,9	40	0,1091	0,0931	0,0938	0,0835
2		3	0,9	80	0,0443	0,0567	0,0580	0,0547
3		3	-0,9	20	0,1298	0,2065	0,2264	0,2193
3		3	-0,9	40	0,0533	0,1707	0,1777	0,1533
3		3	-0,9	80	0,0206	0,0960	0,1025	0,0734
3	FGM	3	0,1	20	0,5039	0,4557	0,4403	0,3983
3		3	0,1	40	0,2913	0,2743	0,2682	0,2637
3		3	0,1	80	0,1385	0,1710	0,1667	0,1674
3		3	0,9	20	0,2620	0,2177	0,2117	0,1612
3		3	0,9	40	0,0968	0,1027	0,1026	0,0842
3		3	0,9	80	0,0408	0,0507	0,0506	0,0454
4		3	-0,9	20	0,9353	0,3127	0,2585	0,3508
4		3	-0,9	40	0,6140	0,5833	0,3501	0,2789
4		3	-0,9	80	0,2387	0,8445	0,3879	0,2278
4		3	0,1	20	0,7733	0,8448	0,7690	0,2992
4		3	0,1	40	0,6100	0,7267	0,6822	0,2037
4		3	0,1	80	0,3608	0,6302	0,5989	0,1562
4		3	0,9	20	0,6519	0,5388	0,4253	0,2692
4		3	0,9	40	0,2459	0,4038	0,3387	0,1927
4		3	0,9	80	0,0596	0,3003	0,2617	0,1784
1			-0,9	20	0,1012	0,1369	0,1883	0,2261
1			0,1	20	0,6238	0,4723	0,4475	0,4002
1			0,9	20	0,3668	0,2241	0,2139	0,1260
2		4	-0,9	20	0,2195	0,2121	0,2829	0,2452
2			0,1	20	0,6419	0,5121	0,4902	0,4043
2			0,9	20	0,3147	0,1996	0,1985	0,1438
3			-0,9	20	0,5365	0,4577	0,5587	0,5030
3			0,1	20	0,6495	0,4707	0,4467	0,3880
3			0,9	20	0,3545	0,2023	0,1974	0,1369

cd. tab. 5

Numer parametru	Kopula	n	α	T	KMNW	L^1	CvM	KS
					MSE			
4	FGM	4	-0,9	20	0,1672	0,1922	0,2480	0,2211
4			0,1	20	0,6690	0,4785	0,4572	0,4050
4			0,9	20	0,3428	0,2089	0,2062	0,1526
5			-0,9	20	0,4567	0,4170	0,5265	0,4937
5			0,1	20	0,6516	0,4758	0,4504	0,3983
5			0,9	20	0,3669	0,1995	0,1931	0,1258
6			-0,9	20	0,3937	0,3642	0,4509	0,4280
6			0,1	20	0,6205	0,4769	0,4548	0,3944
6			0,9	20	0,4003	0,1990	0,1918	0,1344
7			-0,9	20	0,8973	1,3437	0,5110	0,4781
7			0,1	20	0,8530	0,8552	0,7768	0,1852
7			0,9	20	0,6751	0,6055	0,4776	0,1894
8			-0,9	20	1,3664	1,9029	1,4445	0,7546
8			0,1	20	0,8480	0,8532	0,7886	0,1713
8			0,9	20	0,8882	0,6280	0,5071	0,1597
9			-0,9	20	1,2534	1,8950	1,4740	0,7736
9			0,1	20	0,8450	0,8386	0,7516	0,1842
9			0,9	20	0,7620	0,6178	0,4725	0,1651
10			-0,9	20	1,2133	1,8979	1,4861	0,7067
10			0,1	20	0,8590	0,8475	0,7715	0,1956
10			0,9	20	0,7982	0,5357	0,4065	0,1653
11			-0,9	20	1,9658	3,5183	2,9277	0,8512
11			0,1	20	0,9381	0,9749	0,9115	0,0450
11			0,9	20	1,7552	0,9301	0,6604	0,2788

Źródło: opracowanie własne na podstawie przeprowadzonych symulacji.

Tabela 6. Udział przypadków o najniższym błędzie średniokwadratowym estymatora w zależności od wymiaru kopuli (%)

n	KMNW	L^1	CvM	KS
2	83	0	17	0
3	52	0	6	42
4	26	10	0	64

Źródło: opracowanie własne na podstawie przeprowadzonych symulacji.

Podsumowanie

Na podstawie przeprowadzonych symulacji nie można jednoznacznie wskazać, która z metod estymacji parametru kopuli jest najlepsza. Na podstawie analizy obciążenia i błędu średniokwadratowego estymatorów i po uśrednieniu

wszystkich wyników względem wymiarów kopuli, różnych wariantów parametrów i liczebności generowanych prób najlepiej oceniona jest KMNW. Ponadto dodatkową zaletą KMNW jest krótszy czas obliczeń niż dla metody minimalnej odległości.

Należy jednak zauważyć, że pod względem obciążenia estymatora uzyskano dobre rezultaty także dla kopuli Clayтона przy metodzie minimalnej odległości (CvM), dla kopuli FGM (L^1 , CvM i KS) oraz dla kopuli Franka (L^1).

Porównując błąd średniokwadratowy estymatorów, należy zwrócić uwagę, iż niskie wartości MSE otrzymano dla metody minimalnej odległości z funkcją odległości KS zastosowaną do kopuli FGM. Zaobserwowano również, że wraz ze wzrostem wymiaru kopuli n wielkość MSE zmniejsza się.

Metoda minimalnej odległości ma tę zaletę w opozycji do KMNW, że jej wyniki można bezpośrednio wykorzystać do wyboru optymalnej kopuli dla zgromadzonych danych empirycznych, wówczas gdy chcemy wybrać jedną rodzinę kopuli spośród kilku różnych rodzin. Należy wtedy wskazać jako najlepszą rodzinę kopuli tę, dla której otrzymujemy najmniejszą wartość minimalizowanej funkcji odległości. Dla KMNW takiego wyboru należałoby dokonać na podstawie dodatkowych obliczeń i kryteriów.

Metodę minimalnej odległości można uznać za skuteczny sposób estymacji parametru rodziny kopuli, jednak na ogół generującą trochę słabsze wyniki niż KMNW. Najsłabsze wyniki generuje dla odległości z przestrzeni L^1 , lepsze natomiast otrzymywane są wtedy, gdy zastosowane będą funkcje odległości Kołmogorowa–Smirnova i Cramera von Mises.

Literatura

- Cherubini U., Luciano E., Vecchiato W., *Copula method in finance*, John Wiley and Sons, Chichester 2004.
- Doman R., *Zastosowania kopuli w modelowaniu dynamiki zależności na rynkach finansowych*, Wyd. UE w Poznaniu, Poznań 2011.
- Genest C., Ghoudi K., Rivest L.-P., *A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions*, „Biometrika” 1995.
- Heilpern S., *Funkcje łączące*, Wyd. AE im. Oskara Langego, Wrocław 2007.
- Joe H., *Multivariate Models and Dependence Concepts*, Chapman & Hall/CRC, New York 1997.
- Nelsen R.B., *An introduction to copulas*, Springer, New York 2006.
- Stryjek A., *Porównanie wybranych metod estymacji parametru rodziny kopuli*, Szkoła Główna Handlowa, Warszawa 2011.
- Stryjek A., *Użyteczność kopuli w finansach i ubezpieczeniach*, „Zeszyty Naukowe WSB we Wrocławiu” 2011, nr 20.
- Stryjek A., *O pewnym problemie wyboru metody szacowania wartości zagrożonej*, Instytut Ekonometrii Szkoły Głównej Handlowej, Warszawa 2012.

- Stryjek A., *Modeling the dependence of risk variables using copulas. The comparison of copula parameter estimation methods*, w: *Zastosowanie metod ilościowych w zarządzaniu ryzykiem w działalności inwestycyjnej*, red. A.S. Barczak, P. Tworek, Wyd. Uniwersytetu Ekonomicznego w Katowicach, Katowice 2013.
- Stryjek A., *Porównanie wybranych metod estymacji parametru rodziny n -wymiarowej kopuli dla $n \geq 2$* , Szkoła Główna Handlowa, Warszawa 2013.
- Tsukahara H., *Semiparametric estimation in copula models*, „The Canadian Journal of Statistics” 2005, vol. 33, z. 3.
- Weiss G.N.F., *Copula parameter estimation: numerical considerations and implications for risk management*, „The Journal of Risk” 2010, t. 13, nr 1.
- Wolfowitz J., *Estimation by the minimum distance method*, „Annals of the Institute of Statistical Mathematics” 1953, t. 5.

A comparison of n -dimensional copula parameter estimation for $2 \leq n \leq 4$ using canonical maximum likelihood method and minimum distance method

Summary. The estimation using canonical maximum likelihood method (CMLE) and minimum distance method (MDE) is compared in the article. In the MDE method three types of distance functions are considered: Cramer von Mises distance, Kolmogorov–Smirnov distance and the distance metrics from $L1$ space. The bias and the mean squared error were analyzed on the grounds of the simulations for the n -dimensional Clayton copula, Frank copula and Farlie–Gumbel–Morgenstern copula, where $n = 2$, $n = 3$, and $n = 4$. In most considered cases, the smallest bias and mean square error were received for MLE method.

Key words: copula, estimation, CMLE, minimum distance estimation, Clayton copula, Frank copula, Farlie–Gumbel–Morgenstern copula