

**Justyna Majewska**

Uniwersytet Ekonomiczny w Katowicach

# WYBRANE METODY ESTYMACJI PARAMETRÓW FUNKCJI ŁĄCZĄCYCH

## Wprowadzenie

Funkcje łączące<sup>1</sup> dystrybuantę rozkładu  $d$ -wymiarowego z dystrybuantami jednowymiarowych rozkładów brzegowych pozwalają częściowo rozwiązać problem polegający na nieznaności postaci analitycznej empirycznego łącznego rozkładu stóp zwrotu.

Funkcje łączące są bardzo popularnym narzędziem służącym opisywaniu nieliniowych i asymetrycznych zależności między zmiennymi opisującymi rynki finansowe<sup>2</sup>. Możliwość wydzielenia z dowolnej dystrybuanty wielowymiarowej struktury zależności<sup>3</sup> jest wykorzystywana również m.in. w analizie ryzyka portfeli inwestycyjnych czy modelowaniu zdolności kredytowej<sup>4</sup>. Do fundamentalnych prac z zakresu teorii funkcji łączących należą prace Joego i Nelsena<sup>5</sup> dostarczające szczegółowej wiedzy na temat funkcji łączących, ich własności

---

<sup>1</sup> Termin przyjęty za S. Heilpernem, w literaturze przedmiotu *copula* tłumaczy się również: kopuła, funkcja kopuli, funkcja powiązań. Zob. R. Doman: *Zastosowania funkcji łączących w modelowaniu dynamiki zależności na rynkach finansowych*. Wydawnictwo Uniwersytetu Ekonomicznego, Poznań 2011, s. 7.

<sup>2</sup> Zob. np. J. Rodriguez: *Measuring Financial Contagion: A Copula Approach*. „Journal of Empirical Finance” 2007, 41, 401-423; S. Chen, S. Poon: *Modelling International Stock Market Contagion Using Copula and Risk Appetite*. Working Paper 2007.

<sup>3</sup> Jest to jedno z wielu podejść do badania zależności. Współczynnik korelacji liniowej, szeroko stosowany w finansach jako miara zależności, jest odpowiednim narzędziem do mierzenia zależności między stopami zwrotu instrumentów finansowych tylko w przypadku, gdy łączny rozkład tych zwrotów jest eliptyczny. Zatem jeśli dane empiryczne przeczą przyjęciu takiego założenia, do badania zależności powinno się stosować miary dające się określić jedynie w terminach funkcji łączących rozważane zmienne. R. Doman: Op. cit.

<sup>4</sup> Zob. np. Y. Malevergne, D. Sornette: *Testing the Gaussian Copula Hypothesis for Financial Assets Dependence*. „Quantitative Finance” 2003, 3, 231-250; D. Li: *On Default Correlation: A Copula Approach*. „Journal of Fixed Income” 2000, 9, 43-54.

<sup>5</sup> H. Joe: *Multivariate Models and Dependence Concepts*. Chapman & Hall, London 1997; R. Nelsen: *An Introduction to Copulas*. 2nd ed. Springer, New York 2006.

i modelowania zależności. Szybki rozwój metod modelowania zależności z wykorzystaniem funkcji łączących i wzrost obszarów ich aplikacji znajduje odzwierciedlenie w bogatej literaturze. McNeil i in. oraz Denuit i in.<sup>6</sup> rozważają funkcje łączące w kontekście zarządzania ryzykiem. Cherubini i in.<sup>7</sup> przedstawili funkcje łączące językiem matematyki finansowej. Choros i in.<sup>8</sup> dokonali przeglądu parametrycznych i nieparametrycznych metod estymacji parametrów funkcji łączących zarówno dla zmiennych losowych, jak i szeregów czasowych. Genest i Favre<sup>9</sup> zaprezentowali semiparametryczną metodę estymacji dla zmiennych niezależnych i pochodzących z jednakowego rozkładu, popierając teorię szczegółowymi przykładami empirycznymi. Natomiast Patton<sup>10</sup> dokonał przeglądu modeli opartych na funkcjach łączących, wykorzystując je do prognozowania struktury zależności szeregów finansowych i ekonomicznych.

Szereg istniejących modeli funkcji łączących, od najprostszych statycznych po coraz bardziej złożone modele z dynamiką sterowaną, wymaga świadomego wyboru metody estymacji parametrów. Parametry funkcji mogą być estymowane różnymi metodami. W praktyce najczęściej są stosowane następujące metody estymacji: parametryczne (*fully parametric*, ML), stopniowe parametryczne (*stepwise parametric*, IFM) oraz semiparametryczne oparte na funkcji największej wiarygodności. Metody parametryczne i semiparametryczne zostały porównane w badaniu symulacyjnym przez Kim i in.<sup>11</sup> Podejście semiparametryczne oparte na funkcji największej wiarygodności, w którym rozkłady brzegowe są wyznaczone na podstawie ich empirycznych odpowiedników z parametrami funkcji łączących oszacowanymi metodą największej wiarygodności, wykazuje największą efektywność esymacji i najmniej obciążone oszacowania parametrów funkcji łączących.

---

<sup>6</sup> A.J. McNeil, R. Frey, P. Embrechts: *Quantitative Risk Management*. Princeton University Press, Princeton 2005; M. Denuit, J. Dhaene, M. Goovaerts, R. Kaas: *Actuarial Theory for Dependent Risks: Measures, Orders and Models*. Wiley, Chichester 2005.

<sup>7</sup> U. Cherubini, E. Luciano, W. Vecchiato: *Copula Methods in Finance*. Wiley, 2004.

<sup>8</sup> B. Choros, R. Ibragimov, E. Permiakova: *Copula Estimation*. W: *Copula Theory and Its Applications*. Eds. P. Jaworski, F. Durante, W. Härdle, T. Rychlik. Springer, Dordrecht 2010 (Netherlands), 77-92.

<sup>9</sup> C. Genest, A.C. Favre: *Everything You Always Wanted to Know about Copula but You Were Afraid to Ask*. „J Hydrol Eng” 2007, 12, 347-368.

<sup>10</sup> A. Patton: *A Review of Copula Models for Economic Time Series*. „J Multivariate Anal” 2012, 110, 4-18.

<sup>11</sup> G. Kim, M. Silvapulle, P. Silvapulle: *Comparison of Semiparametric and Parametric Methods for Estimating Copulas*. „Computational Statistics & Data Analysis” 2007, 51, 2836-2850.

W niniejszym artykule dokonano oceny odporności podstawowych i zarazem najpopularniejszych metod estymacji parametrów funkcji łączących, przy czym metody te będą rozważane w kontekście występowania obserwacji nietypowych<sup>12</sup>.

W pierwszej części przedstawiono elementarne podstawy teorii funkcji łączących niezbędne dla dalszej części artykułu. Druga część stanowi ogólny przegląd metod estymacji, ze szczegółowym opisem metod, które są analizowane w artykule. W ostatniej części zaprezentowano wyniki eksperymentu symulacji stochastycznych pozwalających na porównanie prezentowanych metod pod względem odporności rozważanych metod.

## 1. Wybrane elementy teorii funkcji łączących

Fundamentalne twierdzenie Sklara<sup>13</sup> pozwala na dekompozycję dystrybucji rozkładu  $d$ -wymiarowego na  $d$  jednowymiarowych dystrybuant brzegowych oraz  $d$ -wymiarową funkcję łączącą: jeśli  $\mathbf{Y} \equiv [Y_1, \dots, Y_d]^T \sim F$ , przy czym  $Y_i \sim F_i$ , wtedy  $\exists C : [0, 1]^d \rightarrow [0, 1]$  taka, że<sup>14</sup>:

$$F(\mathbf{y}) = C(F_1(y_1), \dots, F_d(y_d)) \quad \forall \mathbf{y} \in \mathbb{R}^d \quad (1)$$

Odwrotna implikacja jest również prawdziwa i czyni wzór (1) szczególnie użytecznym dla praktyków. To znaczy dla funkcji łączącej  $C$  i jednowymiarowych dystrybuant  $F_1, \dots, F_d$  można na podstawie (1) wyznaczyć taką  $d$ -wymiarową dystrybuantę  $F$ , dla której  $F_1, \dots, F_d$  są dystrybuantami brzegowymi.

Z twierdzenia Sklara wynikają zastosowania funkcji łączących. Znając rozkłady brzegowe i łączny rozkład prawdopodobieństwa wektora zmiennych losowych, można dopasować odpowiednią funkcję łączącą. Funkcja łącząca może być zastosowana do wyznaczania miar zależności między zmiennymi np.  $\tau$  Kendalla,  $\rho$  Spearmana. Wtedy miarę  $\tau$  Kendalla zapisujemy jako (przy założeniu, że dystrybucje brzegowe  $(Y_1, Y_2)$  są ciągłe,  $C$  funkcją łączącą):

<sup>12</sup> Odporność można rozumieć w sensie ogólnym, uwzględniającym wszystkie rodzaje odstępstw od założeń (*robust*) lub jako odporność tylko na wyróżnione odstępstwo (*resistant*). Wówczas rozróżnia się przykładowo metody odporne na występowanie obserwacji nietypowych, na inny niż założony w modelu rozkład oraz na brak spełnienia warunku niezależności. Zob. S. Heilpern: *Modele odporne*. W: *Statystyczne metody analizy danych*. Red. W. Ostasiewicz. Wydawnictwo Akademii Ekonomicznej im. O. Langego, Wrocław 1998, s. 235-236.

<sup>13</sup> A. Sklar: *Fonctions de repartition a n dimensions et leurs margers*. Publications de l'Institut Statistique de l'Universite de Paris 1959, 8, 229-231.

<sup>14</sup> W przypadku gdy wielowymiarowa dystrybuanta jest ciągła, funkcja łącząca jest wyznaczona jednoznacznie.

$$\tau(Y_1, Y_2) = 4 \iint_{[0,1]^2} C(u_1, u_2) dC(u_1, u_2) - 1$$

a  $\rho$  Spearmana jako:

$$\rho(Y_1, Y_2) = 12 \iint_{[0,1]^2} C(u_1, u_2) du_1 du_2 - 3$$

Funkcje łączące służą również wyznaczaniu współczynników zależności w ogonach rozkładu. Współczynnik zależności w górnym (dolnym) ogonie  $\lambda_U$  ( $\lambda_L$ ) zapisujemy odpowiednio jako dla  $u \in [0, 1]$ :

$$\lambda_U = \lim_{u \rightarrow 1^-} P(Y_2 > F_{Y_2}^{-1}(u) | Y_1 > F_{Y_1}^{-1}(u)) = \lim_{u \rightarrow 1^-} \frac{1 - 2u + C(u, u)}{1 - u}$$

$$\lambda_L = \lim_{u \rightarrow 0^+} P(Y_2 < F_{Y_2}^{-1}(u) | Y_1 < F_{Y_1}^{-1}(u)) = \lim_{u \rightarrow 0^+} \frac{C(u, u)}{u}$$

$$\lambda_U, (\lambda_L) \in [0, 1]$$

Gdy  $\lambda_U, (\lambda_L) \in (0, 1]$  – zmienne  $Y_1$  i  $Y_2$  są zależne, gdy  $\lambda_U, (\lambda_L) = 0$  – niezależne. Ponadto funkcja łącząca umożliwia dokonanie symulacji łącznego rozkładu prawdopodobieństwa wektora zmiennych przy zadanych rozkładach brzegowych. W praktycznych zagadnieniach etap symulacji musi być poprzedzony oszacowaniem nieznanymi parametrów funkcji łączących na podstawie danych empirycznych.

Bardzo ważną kwestią jest wybór odpowiednich funkcji łączących. Zazwyczaj dla danych rozkładów brzegowych eksperymentuje się z różnymi funkcjami w celu zbadania wrażliwości interesującej wielkości na wybór odpowiedniej funkcji. Pominięto w artykule funkcje łączące typu Gaussa, które nie są w stanie wychwycić zależności pomiędzy zdarzeniami ekstremalnymi. Poniżej przedstawiono jawne postaci wybranych funkcji łączących, należących do klasy kopul archimedesowych, wykorzystywane w dalszej części artykułu:

a) funkcja łącząca Clayтона<sup>15</sup>:

$$C_{\theta}^{Cl}(u_1, u_2) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta}, \theta \in R \setminus \{0\}$$

b) funkcja łącząca Franka<sup>16</sup>:

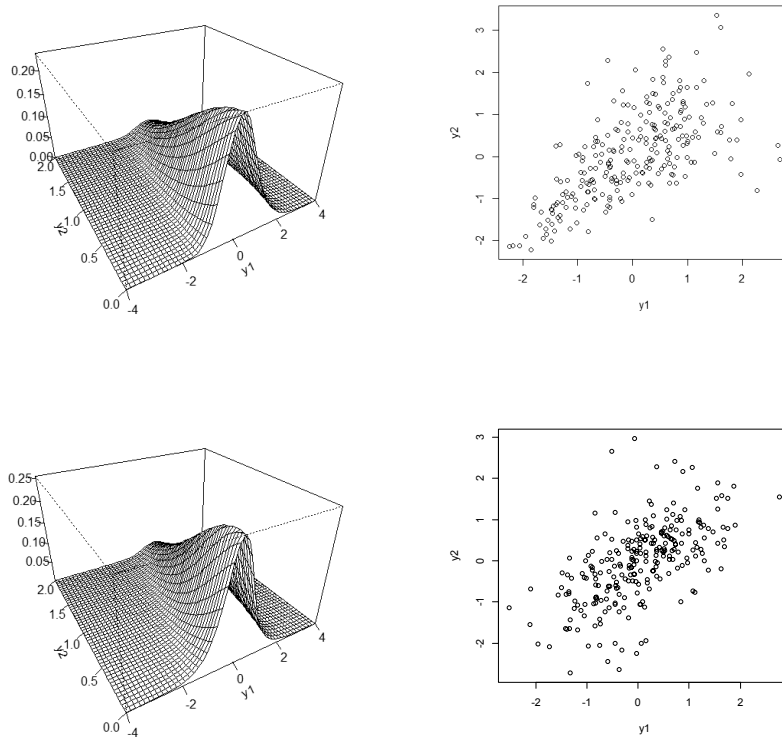
$$C_{\theta}^{Fr}(u_1, u_2) = \begin{cases} -\frac{1}{\theta} \ln \left( 1 + \frac{(\exp(-\theta u_1) - 1)(\exp(-\theta u_2) - 1)}{\exp(-\theta) - 1} \right); & \theta \in R \setminus \{0\} \\ u_1 u_2; & \theta = 0 \end{cases}$$

<sup>15</sup> D.G. Clayton: *A Model for Association in Bivariate Life Tables and Its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence*. „Biometrika” 1978, 65, 141-151.

c) funkcja łącząca Gumbela<sup>17</sup>:

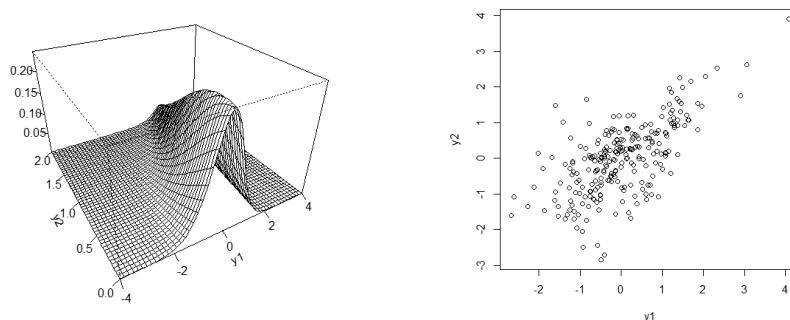
$$C_{\theta}^{GH}(u_1, u_2) = \exp(-((-\ln u_1)^{\theta} + (-\ln u_2)^{\theta})^{1/\theta}), \theta \in \mathbb{R} \setminus \{0\}$$

Rysunek 1 przedstawia rozkłady dwuwymiarowe dla zadanych rozkładów brzegowych – rozkładów normalnych – i ustalonego współczynnika korelacji  $\tau = 0,5$  otrzymane za pomocą wybranych funkcji łączących.



<sup>16</sup> M.J. Frank: *On the Simultaneous Associativity of  $F(x,y)$  and  $xyF(x,y)$* . „Aequationes Mathematicae” 1979, 19, 194-226.

<sup>17</sup> E.J. Gumbel: *Bivariate Exponential Distributions*. „Journal of the American Statistical Association” 1960, 55, 698-707.



Rys. 1. Rozkłady dwuwymiarowe dla zadanych rozkładów brzegowych i ustalonego współczynnika korelacji  $\tau = 0,5$  otrzymane za pomocą funkcji łączących (od góry): Claytona, Franka i Gumbela

Wybór tych funkcji jest arbitralną decyzją autora, jednak wynikającą z wcześniejszych jego badań w obszarze analiz na rynku finansowym<sup>18</sup>. Przypomnijmy, funkcje łączące archimedesowe (Claytona, Gumbela) mogą się cechować asymetrycznymi wzorcami zależności w ogonach rozkładu. Funkcja łącząca Claytona ma zależność w ogonie dolnym, funkcja łącząca Gumbela – w ogonie górnym. Z kolei funkcja łącząca Franka nie ma zależności w ogonach. W przypadku funkcji łączących archimedesowych zaletą jest jej nieskomplikowana postać analityczna (są funkcjami łączącymi jednoparametrowymi).

## 2. Estymacja parametrów funkcji łączących

Parametry funkcji łączących mogą być estymowane różnymi metodami. Przy założeniu, że prawdziwa funkcja łącząca należy do rodziny parametrycznej  $\{C_\theta : \theta \in \Theta\}$ , zgodne i asymptotycznie normalne oszacowania wektora parametrów można uzyskać metodami największej wiarygodności<sup>19</sup>.

Na potrzeby dalszych rozważań przytoczono definicję gęstości funkcji połączeń. Jeśli gęstość funkcji łączącej  $C$  istnieje, to jest ona dana wzorem:

$$c(u_1, \dots, u_d) = \frac{\partial^d C(u_1, \dots, u_d)}{\partial u_1 \dots \partial u_d}$$

<sup>18</sup> Zob. np. G. Trzpiot, J. Majewska: *Testing for Tail Independence in Extreme Value Models Application on Polish Stock Exchange*. „Acta Universitatis Lodzianis. Folia Economica” 2011, 137-146.

<sup>19</sup> Szczegółowy opis estymacji metodą największej wiarygodności zob. w podrozdziale 10.1 pracy: H. Joe: Op. cit.

Ponadto jeśli  $C$  jest funkcją łączącą dystrybuanty  $F$  mającej dystrybuanty brzegowe  $F_1, \dots, F_d$  i gęstości brzegowe  $f_1, \dots, f_d$ , to zachodzi następujący związek:

$$f(y_1, \dots, y_d) = c(F_1(y_1), \dots, F_d(y_d)) \prod_{i=1}^d f_i(y_i) \quad (2)$$

Z (2) wynika postać funkcji największej wiarygodności  $L = \sum_{j=1}^n \log f(y_1^{(j)}, \dots, y_d^{(j)})$  określonej na  $d$ -wymiarowej próbie  $y^{(j)} = (y_1^{(j)}, \dots, y_d^{(j)})$ ,  $j = 1, 2, \dots, n$ , z gęstością  $f$ :

$$L = L_C + \sum_{i=1}^d L_i \quad (3)$$

gdzie:  $L_C = \sum_{j=1}^n \log c(F_1(y_1^{(j)}), \dots, F_d(y_d^{(j)}))$  jest związany ze strukturą zależności reprezentowaną przez funkcję łączącą  $C$ ,  $L_i = \sum_{j=1}^n \log f_i(y_i^{(j)})$ ,  $i = 1, 2, \dots, d$  zawiera informacje z każdej z dystrybuant brzegowych.

Najpopularniejsze jest stosowanie dwustopniowej metody największej wiarygodności. Zatem rozważamy funkcję łączącą  $C$  z rodziny funkcji łączących zależną od wektora parametrów  $\theta$ :  $C = C(u_1, \dots, u_d; \theta)$  i jednowymiarowych dystrybuant brzegowych  $F_i$  z gęstościami  $f_i$ , zależnych od wektora parametrów  $\alpha_i$ , co zapisujemy:  $F_i = F_i(y_i; \alpha_i)$ ,  $f_i = f_i(y_i; \alpha_i)$ .

Estymator największej wiarygodności  $(\hat{\alpha}_1^{MLE}, \dots, \hat{\alpha}_d^{MLE}, \hat{\theta}^{MLE})$  wektora parametrów  $(\alpha_1, \dots, \alpha_d, \theta)$  ma postać:

$$\begin{aligned} (\hat{\alpha}_1^{MLE}, \dots, \hat{\alpha}_d^{MLE}, \hat{\theta}^{MLE}) &= \arg \max_{\alpha_1, \dots, \alpha_d, \theta} L(\alpha_1, \dots, \alpha_d, \theta) = \\ &= \arg \max_{\alpha_1, \dots, \alpha_d, \theta} (L_C(\alpha_1, \dots, \alpha_d, \theta) + \sum_{i=1}^d L_i(\alpha_i)) = \\ &= \arg \max_{\alpha_1, \dots, \alpha_d, \theta} \left( \sum_{j=1}^n \log c(F_1(y_1^{(j)}; \alpha_1), \dots, F_d(y_d^{(j)}; \alpha_d); \theta) + \sum_{i=1}^d \sum_{j=1}^n \log f_i(x_i^{(j)}; \alpha_i) \right) \end{aligned} \quad (4)$$

Występujące w (4) parametry można estymować w dwóch etapach<sup>20</sup>. W pierwszym kroku są estymowane parametry  $\alpha_1, \dots, \alpha_d$  za pomocą funkcji maksymalizacji:

$$\hat{\alpha}_i^{IFM} = \arg \max_{\alpha_i} L_i(\alpha_i) \quad (5)$$

W drugim kroku otrzymuje się oszacowanie parametru funkcji łączącej  $\theta$  poprzez:

$$\hat{\theta}^{IFM} = \arg \max_{\theta} L_C(\hat{\alpha}_1^{IFM}, \dots, \hat{\alpha}_d^{IFM}, \theta) \quad (6)$$

Dwuetaповe podejście do estymacji parametrów funkcji łączących jest nazywane metodą funkcji wnioskowania dla rozkładów brzegowych (ang. *the method of Inference Functions for Margins*, IFM). Procedura IFM jest obliczeniowo prostsza niż jednoetapowa estymacja wektora parametrów  $(\alpha_1, \dots, \alpha_d, \theta)$ , gdyż numeryczna optymalizacja z wieloma parametrami jest o wiele bardziej czasochłonna od kilku optymalizacji z mniejszą liczbą parametrów. Jeśli każdy rozkład brzegowy  $F_i = F_i(y_i; \alpha_i)$ ,  $i = 1, 2, \dots, d$ , jest opisany przez oddzielny wektor parametrów charakteryzujących strukturę zależności, to można uprościć estymację w drugim kroku tak, by nie było potrzeby numerycznej optymalizacji z dużą liczbą parametrów również w przypadku struktury zależności<sup>21</sup>. Przy spełnionych warunkach regularności<sup>22</sup> estymator IFM, podobnie jak estymator największej wiarygodności, jest zgodny i ma własność asymptotycznej normalności. Z badań, jakie przeprowadzili Xu i Joe<sup>23</sup>, metoda IFM cechuje się wysoką efektywnością w odniesieniu do pełnej metody największej wiarygodności.

Kolejnym podejściem do estymacji parametrów funkcji łączących jest metoda semiparametryczna<sup>24</sup>. Nie zakłada się znajomości rozkładu dla dystrybuant brzegowych. Procedura, podobnie jak w IFM, przebiega w dwóch etapach, z różnicą w sposobie estymacji dystrybuant brzegowych. W pierwszym kroku estymuje się metodą rozkładu brzegowe  $F_i$  (wykorzystując metody nieparametryczne, na ogół empiryczne dystrybuanty  $\hat{F}_i$ ), by w drugim kroku wyestymować parametr funkcji łączących:

$$\hat{\theta} = \arg \max_{\theta} L_C(\theta) = \arg \max_{\theta} \sum_{j=1}^n \log c(\hat{F}_1(y_1^{(j)}), \dots, \hat{F}_d(y_d^{(j)}); \theta) \quad (7)$$

<sup>20</sup> H. Joe, J.J. Xu: *The Estimation Method of Inference Functions for Margins for Multivariate Models*. Technical Report No. 166, Department of Statistics, University of British Columbia, 1996.

<sup>21</sup> R. Doman: Op. cit.

<sup>22</sup> R.J. Serfling: *Approximation Theorems of Mathematical Statistics*. Wiley-Blackwell, New York 1980.

<sup>23</sup> H. Joe, J.J. Xu: Op. cit.

<sup>24</sup> X. Chen, Y. Fan: *Estimation and Model Selection of Semiparametric Copula-based Multivariate Dynamic Models under Copula Misspecification*. „Journal of Econometrics” 2006, 135, 125-154.



Estymator semiparametryczny jest zgodny i asymptotycznie normalny przy spełnieniu warunków regularności<sup>25</sup>. Genest i in.<sup>26</sup> pokazali, że metoda ta przy dodatkowych warunkach, które spełniają dwuwymiarowe funkcje łączące, np. gaussowska, Eyraud-Farlie-Gumbel, rodzina funkcji Claytona i Franka, estymator  $\hat{\theta}$  jest w pełni efektywny, gdy istnieje niezależność między zmiennymi.

W literaturze istnieje wiele metod alternatywnych do metody największej wiarygodności i na ogół są to metody nieparametryczne, w których najczęściej wykorzystuje się fakt wydzielenia funkcji łączącej z dystrybucyjną łączną w przypadku ciągłych dystrybucyj brzegowych, tzn.:

$$\hat{C}(u_1, \dots, u_d) = \hat{F}(\hat{F}_1^{-1}(u_1), \dots, \hat{F}_d^{-1}(u_d)) \quad (8)$$

gdzie  $\hat{F}$  jest nieparametrycznym estymatorem  $d$ -wymiarowej dystrybucyj  $F$ , a  $\hat{F}_1^{-1}, \dots, \hat{F}_d^{-1}$  są nieparametrycznymi estymatorami  $F_i^{-1}(s) = \{t \mid F_i(t) \geq s\}$  dystrybucyj brzegowych  $F_1, \dots, F_d$ . Jak już wcześniej wspomniano, naturalnym sposobem estymacji parametrów funkcji łączących jest zastosowanie empirycznej funkcji łączącej przez estymację empirycznej dystrybucyj  $F$ :

$$\hat{F}(y_1, \dots, y_d) = \frac{1}{T} \sum_{i=1}^d \mathbf{1}(Y_i \leq y_1, \dots, Y_i \leq y_d) \quad (9)$$

oraz estymację empirycznych brzegowych dystrybucyj:

$$\hat{F}_i(y_i) = \sum_{i=1}^d \mathbf{1}(Y_i \leq y_i) \quad (10)$$

Kolejne podejście do estymacji parametrów funkcji łączącej poprawia odporność estymatorów na obserwacje odstające. Najprostszą metodą poprawienia odporności metody estymacji funkcji łączącej jest zastosowanie procedury wcześniejszej identyfikacji obserwacji odstających, a następnie przeprowadzenie estymacji metodą największej wiarygodności. Do identyfikacji obserwacji odstającej można wykorzystać estymator macierzy kowariancji (o wysokim punkcie załamania<sup>27</sup>), np. MCD, Donoho-Stahel.

<sup>25</sup> C. Genest, K. Ghoudi, L. Rivest: *A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions*. „Biometrika” 1995, 82, 543-552.

<sup>26</sup> Ibid.

<sup>27</sup> Punkt załamania jest miarą globalnej odporności estymatora. Punkt załamania próby skończonej wskazuje na maksymalny odsetek obserwacji nietypowych w próbie, która nie sprawia, że estymator „nie załamuje się”.

Mendes i in.<sup>28</sup> zaprezentowali odporne estymatory parametrów funkcji łączącej polegające na minimalizacji empirycznej funkcji łączącej na podstawie dopasowania statystyk, np. statystykę odległości Kołmogorova  $K$ , Andersona-Darlinga  $AD$  czy Cramera-von Misesa  $W^2$ .

### 3. Opis i wynik badania symulacyjnego

Badanie symulacyjne pozwoliło na porównanie wybranych estymatorów parametrów dwuwymiarowych funkcji łączących pod względem odporności i efektywności<sup>29</sup>. W badaniu wykorzystano koncepcję modelu  $\varepsilon$ -zaburzonego. Skonstruowano rozkłady dwuwymiarowe, z których obie zmienne mają rozkład normalny z wprowadzanymi poziomami zaburzenia<sup>30</sup> odpowiednio:  $\varepsilon = 0,2\%$  oraz  $\varepsilon = 2\%$ . Struktura zależności jest opisywana funkcją łączącą Clayтона z parametrami  $\theta = 0,5$  i  $\theta = 8$ , Gumbela z parametrami  $\theta = 1,25$  i  $\theta = 5$  oraz Franka z parametrami  $\theta = 1,86$  i  $\theta = 18,2$ . Założona postać funkcji łączącej oznacza, że współczynnik korelacji  $\tau$  Kendalla między zmiennymi<sup>31</sup> wynosi odpowiednio:  $0,2$  i  $0,8$ . Związki parametru  $\theta$  z współczynnikiem  $\tau$  Kendalla są następujące:

- dla funkcji Clayтона  $\tau = \frac{\theta}{\theta + 2}$ ,  $-1 \leq \tau \leq 1$ ,
- dla funkcji Gumbela  $\tau = \frac{\theta - 1}{\theta}$ ,  $0 \leq \tau \leq 1$ ,
- dla funkcji Franka  $\tau = 1 - \frac{4}{\theta} + \frac{4D_1(\theta)}{\theta}$ ,  $-1 \leq \tau \leq 1$ ,  $\theta \neq 0$ ,  $\tau \neq 0$ ,

$$D_1(\theta) = \int_0^{\theta} \frac{x}{\theta(e^x - 1)} dx.$$

Eksperyment polegający na konstrukcji rozkładów dwuwymiarowych powtórzono 250 razy. Długość szeregu wygenerowanych danych wyniosła 1000 dla każdego rozkładu brzegowego.

<sup>28</sup> B. Mendes, E. de Melo, R. Nelsen: *Robust Fits for Copula Models*. „Communications in Statistics – Simulation and Computation” 2007, 36(5), 997-1017.

<sup>29</sup> Wszystkie obliczenia zostały wykonane w programie R project 3.1.0.

<sup>30</sup> Zaburzenie wprowadzono za zasadzie losowego zastąpienia  $0,2\%$  oraz  $2\%$  losowo wybranych danych przez dobraną wielkość. Wielkość ta jest sumą średniej wszystkich obserwacji oraz trzech odchyłeń standardowych odpowiadających. Ze względu na przewidywane rezultaty w zakresie efektywności estymatorów nie rozważamy przypadku, gdy  $\varepsilon = 0\%$ .

<sup>31</sup> Wykorzystując różne funkcje łączące, można otrzymać różne rozkłady dwuwymiarowe z zadanymi rozkładami brzegowymi i ustalonym współczynnikiem  $\tau$ -Kendalla.

Zastosowano dwuetapowe metody estymacji funkcji łączących, co pozwoliło na kombinację metod szacowania dystrybuant brzegowych i parametrów funkcji łączących. Analizowano estymację metodą największej wiarygodności (metoda oznaczona przez MLE), nieparametryczną estymację z wykorzystaniem empirycznej dystrybuanty (metoda oznaczona przez NP), estymację odporną (metoda oznaczona przez ODP). Oceny metod estymacji dokonano porównując średnie błędy kwadratowe  $MSE = \frac{1}{N} \sum_{i=1}^N (\hat{\theta}^{(i)} - \theta)^2$  oraz relację MSE-efektywność względem klasycznej metody IFM.  $\hat{\theta}^{(i)}$  jest estymatorem parametru  $\theta$  dla  $i$ -tej próby. Wyniki eksperymentów przedstawiają tabele 1-3.

Tabela 1

Błędy MSE oraz efektywność estymatorów w odniesieniu do IFM estymatora dla funkcji Clayтона

Metoda estymacji		$\theta = 0,5$		$\theta = 8$	
		$\varepsilon = 0,2\%$			
Etap1	Etap2	MSE	Eff	MSE	Eff
MLE	MLE	0,0164	100,00	1,9169	100,00
NP	MLE	0,0158	100,64	1,6320	124,90
NP	ODP	0,0148	101,59	0,6834	233,95
ODP	ODP	0,0124	109,51	0,5081	295,83
		$\varepsilon = 2\%$			
MLE	MLE	0,0243	100,00	23,761	100,00
NP	MLE	0,0287	131,87	20,740	423,87
NP	ODP	0,0215	138,45	1,7328	643,89
ODP	ODP	0,0194	142,99	5,8311	499,65

Tabela 2

Błędy MSE oraz efektywność estymatorów w odniesieniu do IFM estymatora dla funkcji Franka

Metoda estymacji		$\theta = 2$		$\theta = 1,25$	
		$\varepsilon = 0,2\%$			
Etap1	Etap2	MSE	Eff	MSE	Eff
MLE	MLE	0,0112	100,00	0,1649	100,00
NP	MLE	0,0175	96,23	0,1754	95,87
NP	ODP	0,1296	100,76	0,1952	87,32
ODP	ODP	0,0120	126,76	0,1388	119,69
		$\varepsilon = 2\%$			
MLE	MLE	0,1287	100,00	8,6965	100,00
NP	MLE	0,1264	97,85	7,9956	108,51
NP	ODP	0,0748	108,92	0,8187	2101,8
ODP	ODP	0,0065	127,48	0,6376	1403,7

Tabela 3

Błędy MSE oraz efektywność estymatorów w odniesieniu do IFM estymatora dla funkcji Gumbela

Metoda estymacji		$\theta = 1,25$		$\theta = 5$	
		$\varepsilon = 0,2\%$			
Etap1	Etap2	MSE	Eff	MSE	Eff
MLE	MLE	0,0051	100,00	2,2857	100,00
NP	MLE	0,0041	121,83	1,0286	216,21
NP	ODP	0,0039	111,37	0,1733	1563,90
ODP	ODP	0,0030	115,84	0,1395	1672,69
		$\varepsilon = 2\%$			
MLE	MLE	0,0778	100,00	16,919	100,00
NP	MLE	0,0155	439,05	7,9390	181,92
NP	ODP	0,0078	657,78	0,7197	2745,9
ODP	ODP	0,0107	497,11	0,3075	6922,0

Na podstawie wyników z tabel 1-3 wnioskuje się, że występowanie zaledwie 0,2% obserwacji (czyli 2 obserwacji w 1000-elementowym szeregu) powoduje zwiększenie obciążenia estymatorów metody największej wiarygodności. W szczególności zjawisko nabiera na sile w przypadku istnienia silniejszej zależności i większej liczby obserwacji ekstremalnych. Nieznacznie większe błędy otrzymuje się w przypadku zastąpienia w pierwszym etapie metody MLE metodą nieparametryczną. Zadawalające efekty otrzymuje się w przypadku zastosowania kombinacji estymacji odpornych. Zbliżone wartości błędów odnotowano dla kombinacji estymacji nieparametrycznej oraz odpornej.

## Zakończenie

Funkcje łączące są bez wątpienia potężnym narzędziem służącym do modelowania struktur zależności. Wykorzystywanie tych funkcji bez staranności i właściwego zrozumienia może doprowadzić do nieprawidłowości we wnioskowaniu statystycznym dotyczącym struktury zależności, co może się przełożyć na błędne wnioskowanie w zakresie zarządzania ryzykiem finansowym. Istotnym wnioskiem z badania jest wskazanie, iż w szczególności w przypadku występowania pojedynczych obserwacji ekstremalnych odporne metody estymacji funkcji łączących nie tylko pozwalają uzyskać mniejsze błędy estymatorów, ale również uwzględniają właściwą siłę zależności między zmiennymi.

## Literatura

- Chen X., Fan Y.: *Estimation and Model Selection of Semiparametric Copula-based Multivariate Dynamic Models under Copula Misspecification*. „Journal of Econometrics” 2006, 135, 125-154.
- Chen S., Poon S.: *Modelling International Stock Market Contagion Using Copula and Risk Appetite*. Working Paper 2007.
- Cherubini U., Luciano E., Vecchiato W.: *Copula Methods in Finance*. Wiley, 2004.
- Choros B., Ibragimov R., Permiakova E.: *Copula Estimation*. W: *Copula Theory and Its Applications*. Eds. P. Jaworski, F. Durante, W. Härdle, T. Rychlik. Springer, Dordrecht 2010 (Netherlands), 77-92.
- Clayton D.G.: *A Model for Association in Bivariate Life Tables and Its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence*. „Biometrika” 1978, 65, 141-151.
- Denuit M., Dhaene J., Goovaerts M., Kaas R.: *Actuarial Theory for Dependent Risks: Measures, Orders and Models*. Wiley, Chichester 2005.
- Doman R.: *Zastosowania funkcji łączących w modelowaniu dynamiki zależności na rynkach finansowych*. Wydawnictwo Uniwersytetu Ekonomicznego, Poznań 2011.
- Frank M.J.: *On the Simultaneous Associativity of  $F(x,y)$  and  $xyF(x,y)$* . „Aequationes Mathematicae” 1979, 19, 194-226.
- Genest C., Favre A.C.: *Everything You Always Wanted to Know about Copula but You Were Afraid to Ask*. „J Hydrol Eng” 2007, 12, 347-368.
- Genest C., Ghoudi K., Rivest L.: *A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions*. „Biometrika” 1995, 82, 543-552.
- Gumbel E.J.: *Bivariate Exponential Distributions*. „Journal of the American Statistical Association” 1960, 55, 698-707.
- Heilpern S.: *Modele odporne*. W: *Statystyczne metody analizy danych*. Red. W. Ostasiewicz. Wydawnictwo Akademii Ekonomicznej im. O. Langego, Wrocław 1998.
- Heilpern S.: *Funkcje łączące*. Wydawnictwo Akademii Ekonomicznej, Wrocław 2007.
- Joe H.: *Multivariate Models and Dependence Concepts*. Chapman & Hall, London 1997.
- Joe H., Xu J.J.: *The Estimation Method of Inference Functions for Margins for Multivariate Models*. Technical Report No. 166, Department of Statistics, University of British Columbia, 1996.
- Kim G., Silvapulle M., Silvapulle P.: *Comparison of Semiparametric and Parametric Methods for Estimating Copulas*. „Computational Statistics & Data Analysis” 2007, 51, 2836-2850.
- Li D.: *On Default Correlation: A Copula Approach*. „Journal of Fixed Income” 2000, 9, 43-54.
- Malevergne Y., Sornette D.: *Testing the Gaussian Copula Hypothesis for Financial Assets Dependence*. „Quantitative Finance” 2003, 3, 231-250.

- McNeil A.J., Frey R., Embrechts P.: *Quantitative Risk Management*. Princeton University Press, Princeton 2005.
- Mendes B., Melo E. de, Nelsen R.: *Robust Fits for Copula Models*. „Communications in Statistics – Simulation and Computation” 2007, 36(5), 997-1017.
- Nelsen R.: *An Introduction to Copulas*. 2nd ed. Springer, New York 2006.
- Patton A.: *A Review of Copula Models for Economic Time Series*. „J Multivariate Anal” 2012, 110, 4-18.
- Rodriguez J.: *Measuring Financial Contagion: A Copula Approach*. „Journal of Empirical Finance” 2007, 41, 401-423.
- Serfling R.J.: *Approximation Theorems of Mathematical Statistics*. Wiley-Blackwell, New York 1980.
- Sklar A.: *Fonctions de repartition a n dimensions et leurs margers*. Publications de l’Instut Statistique de l’Universite de Paris, 1959, 8, 229-231.
- Trzpiot G., Majewska J.: *Testing for Tail Independence in Extreme Value Models Application on Polish Stock Exchange*. „Acta Universitatis Lodzianis, Łódź, Folia Economica” 2011, 137-146.

## REVIEW OF CHOSEN METHODS OF COPULA ESTIMATION

### Summary

In this paper we provide a brief survey of some parametric estimation procedures for copula models. We review approaches to inference on copulas for random samples with dependent marginals and we also discuss the issue of robustness of estimation methods. The methods were considered in the context of the presence of outliers.