PRACE NAUKOWE UNIWERSYTETU EKONOMICZNEGO WE WROCŁAWIU
Nr 7 (1207) 2008
TAKSONOMIA 15
Klasyfikacja i analiza danych – teoria i zastosowania

Daniel Papla

Uniwersytet Ekonomiczny we Wrocławiu

PRZYKŁAD ZASTOSOWANIA FUNKCJI POWIĄZAŃ FARLIEGO-GUMBELA-MORGENSTERNA W ANALIZIE ZALEŻNOŚCI MIĘDZY WYBRANYMI AKCJAMI NOTOWANYMI NA GPW W WARSZAWIE

1. Wstęp

Celem artykułu jest zbadanie charakteru zależności występujących między akcjami na polskiej giełdzie. Ze względu na czasochłonność i złożoność obliczeń ograniczono się do wybranych spółek z polskiej giełdy. Z tego samego powodu w obliczeniach wykorzystano jedynie portfele składające się z dwóch i trzech elementów. Dodatkowym ograniczeniem liczby składników portfela jest specyfika wykorzystanej przez autora funkcji powiązań.

Hipotezę badawczą tego artykułu sformułowano w następujący sposób: na polskiej giełdzie występują nie tylko zależności między parami akcji (które można analizować za pomocą np. klasycznego współczynnika korelacji), ale również zależności między trójkami akcji, które można badać za pomocą wielowymiarowych (w tym przypadku trójwymiarowych) funkcji powiązań. Potwierdzenie tej hipotezy oznaczałoby, że zależność między akcjami polskich spółek jest nieco bardziej skomplikowana, niż to wynika z klasycznych modeli rynku kapitałowego, i ma charakter nieliniowy.

W celu weryfikacji tej hipotezy autor wykorzystał funkcję powiązań Farliego-Gumbela-Morgensterna (FGM). Analiza składała się z dwóch etapów. W pierwszym estymowano współczynniki funkcji powiązań FGM dla par akcji, w drugim zaś dla trójek (czyli portfeli trójskładnikowych). Porównanie wyników tych estymacji stanowi podstawę weryfikacji hipotezy badawczej.

Praca składa się z trzech części. W pierwszej omówiono funkcje powiązań. W drugiej przedstawiono zastosowaną funkcję Farliego-Gumbela-Morgensterna. W ostatniej części przedstawiono wyniki badań i ich interpretację.

2. Funkcje powiązań

Definicja funkcji powiązań wygląda następująco: k-wymiarową funkcję C: $[0, 1]^k \to [0, 1]$ nazywamy *copula*, jeśli spełnia następujące warunki:

a) $C(u_1, u_2, ..., u_k)$ jest funkcją k-rosnącą,

b) $C(u_1, u_2, ..., u_{i-1}, 0, u_{i+1}, ..., u_k) = 0$,

c) $C(1, ..., 1, u_i, 1, ..., 1) = u_i$.

Znaczenie funkcji powiązań w analizie zależności wielowymiarowych wynika z twierdzenia Sklara:

Niech H będzie dystrybuantą łączną rozkładu wielowymiarowego, którego rozkłady brzegowe oznaczmy odpowiednio przez F_i. Wtedy istnieje funkcja powiązań C taka, że

$$H(x_1, x_2, ..., x_k) = C(F_1(x_1), F_2(x_2), ..., F_k(x_k)).$$

Jeśli F_i są ciągłe, to C jest jednoznacznie określona. Ponadto jeśli F_i są dystrybuantami, to funkcja H określona powyższym równaniem jest dystrybuantą łączną rozkładu wielowymiarowego.

Dowód dwuwymiarowego przypadku tego twierdzenia można znaleźć w pracy Sklara z 1959 r. Schweizer i Sklar [1974] przedstawili nieco prostszą postać. Dowody na przypadek dwu- i wielowymiarowy można znaleźć również w pracy Nelsena [1999]. O tym, jak sobie radzić w przypadku, kiedy dystrybuanty F_i nie są ciągłe, traktuje m.in. praca Joego [1997].

Jeśli oznaczymy $u_i = F_i(x_i)$, to z twierdzenia Sklara wypływa następujący wniosek:

$$C(u_1, u_2, ..., u_k) = H\left(F_1^{-1}(u_1), F_2^{-1}(u_2), ..., F_k^{-1}(u_k)\right).$$

Z twierdzenia tego wynika, że funkcją powiązań możemy przybliżać nieznany wielowymiarowy rozkład łączny. Oczywiście problemem jest znalezienie odpowiednio dobrze dopasowanej funkcji.

O ile w przypadku dwuwymiarowym w literaturze występuje bardzo wiele funkcji powiązań (por. np. spis zamieszczony w pracy [Nelsen 1999]), o tyle w przypadku, gdy mamy do czynienia z więcej niż dwoma wymiarami, takich funkcji jest znacznie mniej. Bardzo popularna w analizie dwuwymiarowej rodzina funkcji archimedesowskich [Nelsen 1999; Armstrong 2003; Bouyé i. in. 2000; Embrechts, Lindskog, McNeil 2001; Genest, MacKay 1986; Genest, Rivest 1993; Guegan, Ladoucette 2004], choć możliwa do uogólnienia na przypadek wielowymiarowy, jest rzadko stosowana ze względu na swoje właściwości (por. np. [Rokita 2006]). Nie wchodząc w szczegóły, ogólnie można stwierdzić, że funkcje archimedesowskie zakładają, że powiązanie każdej z par zmiennych jest równe tej samej liczbie, co oznaczałoby, że wszystkie pary są powiązane w ten sam sposób, uniemożliwiający ich rozróżnienie [Nelsen 1999; Rokita 2006].

Przedstawione w tym artykule wyniki badań polskiej giełdy z wykorzystaniem funkcji powiązań nie są jedynymi występującymi w literaturze przedmiotu. Według wiedzy autora większość tych badań została przeprowadzona przez pracowników ośrodka wrocławskiego. Znaczna ich część jednak dotyczy przypadków dwuwymiarowych. Należy tu wymienić takie prace, jak: [Jajuga, Kuziak 2003; Jajuga, Papla 2005; 2006; Rokita 2005; Papla, Piontek 2005; Papla 2005a; 2005b; 2006]. W większości z tych prac analiza rynków kapitałowych (zwłaszcza polskiej giełdy) za pomocą funkcji powiązań pozwoliła wykazać takie ich zalety, jak:

- uwzględnianie innej struktury zależności niż liniowa, w tym uwzględnianie zależności w ogonach rozkładu,
- uwzględnianie grubych ogonów rozkładów brzegowych,
- uchwycenie pełnej struktury zależności między badanymi zmiennymi – postać funkcji powiązań w pełni oddaje zależność między zmiennymi,
- zastosowanie funkcji powiązań pozwala „ominąć" problem nieznajomości postaci analitycznej rozkładu łącznego.

W omawianej już pracy Rokity [2006] przedstawiono próbę zastosowania funkcji powiązań dla liczby wymiarów większej niż dwa, która jednak nie zakończyła się sukcesem ze względu na przedstawione wady funkcji archimedesowskich w przypadku wielowymiarowym.

3. Funkcja powiązań FGM

Funkcję Farliego-Gumbela-Morgensterna w przypadku wielowymiarowym można przedstawić w następujący sposób [Armstrong, Galli 2002]:

$$C(u_1,u_2,\ ...,\ u_k)=u_1,u_2,\ ...,\ u_kP(u_1,u_2,\ ...,\ u_k),$$

gdzie: $P(u_1,u_2,\ ...,\ u_k)=1+\sum_{l=2}^{k}\ \sum_{1\le j_1<...<j_l\le k}\theta_{j_1j_2...j_l}(1-u_{j_1})(1-u_{j_2})...(1-u_{j_l}),$

$\theta_{j_1j_2...j_l}$ – współczynniki funkcji FGM.

Funkcja gęstości wygląda zaś tak:

$$c(u_1,u_2,\ ...,\ u_k)=P(2u_1,2u_2,\ ...,\ 2u_k).$$

Dla $k=3$ wielomian P ma następującą postać:

$$P(u_1,u_2,u_2)=1+\theta_{12}(1-u_1)(1-u_2)+\theta_{13}(1-u_1)(1-u_3)+$$
$$+\theta_{23}(1-u_2)(1-u_3)+\theta_{123}(1-u_1)(1-u_2)...(1-u_3).$$

Funkcja gęstości dla $k=3$:

$$c(u_1,u_2,u_2)=1+\theta_{12}(1-2u_1)(1-2u_2)+\theta_{13}(1-2u_1)(1-2u_3)+$$
$$+\theta_{23}(1-2u_2)(1-2u_3)+\theta_{123}(1-2u_1)(1-2u_2)...(1-2u_3).$$

Estymacja parametrów takiej funkcji przebiega w dwóch etapach [Nelsen 1999; Armstrong, Galli 2002]:

1. W pierwszym etapie estymuje się parametry poszczególnych rozkładów brzegowych. Te parametry służą następnie do wyznaczenia wartości dystrybuant dla obserwacji z każdego rozkładu brzegowego.

2. Wyznaczone w ten sposób wartości dystrybuant wykorzystuje się następnie w estymacji metodą największej wiarygodności parametrów funkcji powiązań.

Jednym z podstawowych problemów występujących przy zastosowaniu funkcji FGM w analizie szeregów wielowymiarowych jest następujące ograniczenie nałożone na wartości

$$\sum_{l=2}^{k} \sum_{1 \le j_1 < \ldots < j_l \le k} \varepsilon_{j_1} \varepsilon_{j_2} \ldots \varepsilon_{j_l} \theta_{j_1 j_2 \ldots j_l} \le 1,$$

$$\varepsilon_{j_1}, \varepsilon_{j_2}, \ldots, \varepsilon_{j_l} \in \{-1, 1\},$$

$$\left|\theta_{j_1 j_2 \ldots j_l}\right| \le 1.$$

Oznacza to, że wraz ze wzrostem liczby wymiarów rząd wielkości parametrów funkcji FGM maleje bardzo szybko (w tempie 2^k), co zmniejsza jej użyteczność w badaniach wielowymiarowych.

4. Dane, wyniki estymacji, interpretacja

Jak już wspomniano, ze względu na czasochłonność obliczeń i trudną do ogarnięcia obszerność wyników w tej pracy zostaną zaprezentowane wyniki dotyczące tylko wybranej grupy akcji. Są to akcje 10 najdłużej notowanych spółek: BRE, Elektrim, Irena, Kable, Krosno, Millennium, Mostostal Export, Próchnik, Swarzędz i Żywiec. W obliczeniach wykorzystano dzienne logarytmiczne stopy zwrotu z okresu od 2 stycznia 1995 r. do 14 czerwca 2007 r. Funkcję FGM estymowano dla wszystkich dwu- i trójelementowych portfeli, które można otrzymać z wybranych 10 spółek. Obliczenia wykonano w pakiecie Matlab.

Zwraca uwagę duża różnica między wartościami współczynników θ estymowanymi dla par w przypadku dwuwymiarowym (θ_2) a średnią współczynników estymowanych dla tych samych par w przypadku trójwymiarowym (θ_3) (tab. 1 i 2). Różnica jest tak duża, że współczynnik korelacji rang Spearmana między tymi dwiema wielkościami jest ujemny i wynosi –0,5675. Z kolei współczynnik korelacji rang Spearmana między estymatorami θ_2 a estymatorem współczynnika korelacji liniowej (r) wynosi 0,8782 (tab. 3), co oznacza dużą zgodność obydwu miar.

Tabela 1. Porównanie estymatorów

Pary akcji	θ_2	θ_3	r	Pary akcji	θ_2	θ_3	r
Elektrim BRE	0,9990	0,5375	0,2580	Próchnik Irena	0,5501	0,2703	0,1105
Irena BRE	0,5506	0,3662	0,1607	Próchnik Kable	0,5496	0,1479	0,1063
Irena Elektrim	0,6253	0,5132	0,1354	Próchnik Krosno	0,3616	0,2591	0,0784
Kable BRE	0,4195	0,4450	0,1192	Próchnik Millennium	0,4907	0,3195	0,1004
Kable Elektrim	0,4425	0,3826	0,0794	Próchnik Mostostal Exp.	0,5451	0,2489	0,1296
Kable Irena	0,7509	0,3531	0,1935	Swarzędz BRE	0,4776	0,1604	0,1373
Krosno BRE	0,6459	0,2663	0,2356	Swarzędz Elektrim	0,6526	0,1840	0,1360
Krosno Elektrim	0,7155	0,2452	0,1480	Swarzędz Irena	0,8070	0,1917	0,2071
Krosno Irena	0,6571	0,1638	0,2176	Swarzędz Kable	0,6911	0,2517	0,1637
Krosno Kable	0,6777	0,1501	0,1819	Swarzędz Krosno	0,6113	0,2538	0,1497
Millennium BRE	0,9990	0,4388	0,3573	Swarzędz Millennium	0,6917	0,2592	0,1858
Millennium Elektrim	0,9849	0,4306	0,2225	Swarzędz Mostostal Exp.	0,7372	0,1418	0,2243
Millennium Irena	0,6621	0,3640	0,1823	Swarzędz Próchnik	0,7369	0,2517	0,1727
Millennium Kable	0,6093	0,2949	0,1331	Żywiec BRE	0,5610	0,3215	0,1746
Millennium Krosno	0,7753	0,2667	0,2591	Żywiec Elektrim	0,6197	0,4548	0,1091
Mostostal Exp. BRE	0,9630	0,4314	0,3307	Żywiec Irena	0,4491	0,3909	0,1114
Mostostal Exp. Elektrim	0,9411	0,3981	0,2753	Żywiec Kable	0,3705	0,3207	0,0977
Mostostal Exp. Irena	0,6989	0,5807	0,1944	Żywiec Krosno	0,4364	0,5709	0,1352
Mostostal Exp. Kable	0,5417	0,5131	0,1281	Żywiec Millennium	0,5321	0,4530	0,1803
Mostostal Exp. Krosno	0,6954	0,4300	0,2093	Żywiec Mostostal Exp.	0,5756	0,4347	0,1514
Mostostal Exp. Millennium	0,9482	0,3095	0,3136	Żywiec Próchnik	0,3593	0,3417	0,0730
Próchnik BRE	0,3995	0,3161	0,0895	Żywiec Swarzędz	0,3932	0,1904	0,1121
Próchnik Elektrim	0,4853	0,3617	0,1058	Próchnik Irena	0,5501	0,2703	0,1105

Źródło: obliczenia własne.

Tabela 2. Analiza wyników estymacji

	θ_2	θ_3	r
Średnia	0,6308	0,3328	0,1684
Odchylenie standardowe	0,1792	0,1183	0,0680
Maksimum	0,9990	0,5807	0,3573
Minimum	0,3593	0,1418	0,0730

Źródło: obliczenia własne.

Tabela 3. Współczynniki korelacji rang Spearmana

Współczynnik korelacji rang Spearmana	θ_2	θ_3	r
θ_2	1,0000	–0,5680	0,8782
θ_3	–0,5680	1,0000	–0,4968
r	0,8782	–0,4968	1,0000

Źródło: obliczenia własne.

Zdaniem autora otrzymane wyniki nie potwierdzają hipotezy badawczej. Świadczy o tym:

- podana wyżej zgodność wyników estymacji współczynników funkcji FGM i korelacji liniowej,
- fakt, że w żadnym przypadku nie stwierdzono występowania istotnie różnego od zera wartości współczynnika θ_{123} (nie przedstawionego w tabelach).

Analizując wyniki, należy jednak mieć na uwadze ograniczenia funkcji FGM. Badania innych autorów, jak również inne badania autora niniejszego artykułu potwierdzają występowanie nieliniowości na GPW w Warszawie. Te badania świadczą raczej przeciw wykorzystywaniu FGM w analizie finansowych szeregów czasowych.

Literatura

Armstrong M. (2003), *Copula Catalogue. Part 1: Bivariate Archimedean Copulas*, maszynopis, CERNA, Paryż, http://www.cerna.ensmp.fr.

Armstrong M., Galli A. (2002), *Sequential Nongaussian Simulations using the FGM Copula*, Working Paper.

Bouyé E. i. in. (2000), *Copulas for Finance. A Reading Guide and Some Applications*, maszynopis, City University Business School, London, Crédit Lyonnais, Paris.

Embrechts P., Lindskog F., McNeil A. (2001), *Modelling Dependence with Copulas and Applications to Risk Management*, maszynopis, ETH, Zürich.

Genest C., MacKay J. (1986), *The Joy of Copulas: Bivariate Distributions with Uniform Marginals*, „The American Statistician"nr 40, s. 280-283.

Genest C., Rivest L.P. (1993), *Statistical Inference Procedures for Bivariate Archimedean Copulas*, „Journal of the American Statistical Association" nr 88, s. 1034-1043.

Guegan D., Ladoucette S. (2004), *Dependence Modelling of the Joint Extremes in a Portfolio Using Archimedean Copulas: Application to MSCI Indices*, raport badawczy, IDHE-MORA, nr 05--2004.

Jajuga K., Kuziak K. (2003), *Modeling Relationships in Multivariate Data*, [w:] Taksonomia 10, red. K. Jajuga, M. Walesiak, Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 988, AE, Wrocław, s. 461-471.

Jajuga K., Papla D. (2005), *Extreme Value Analysis and Copulas*, [w:] *Statistical Tools in Finance and Insurance*, red. P. Cizek, W. Härdle, R. Weron, Springer, Berlin, s. 45-64.

Jajuga K., Papla D. (2006), *Copula Functions in Model Based Clustering*, [w:] *From Data and Information Analysis to Knowledge Engineering*, red. M. Spiliopoulou i in., Springer Verlag, Berlin, s. 606-613.

Joe H. (1997), *Multivariate Models and Dependence Concepts*, Chapman & Hall, Boca Raton.

Nelsen R.B. (1999), *An Introduction to Copulas*, Springer Verlag, New York.

Papla D. (2005a), *Analiza notowań na giełdach w Polsce i na świecie z wykorzystaniem warunkowego rozkładu alfa-stabilnego i funkcji copula*, [w:] Taksonomia 12, red. K. Jajuga, M. Walesiak, Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 1076, AE, Wrocław, s. 423-433.

Papla D. (2005b), *Wykorzystanie funkcji copula w budowie optymalnego portfela na przykładzie wybranych spółek notowanych na GPW w Warszawie*, „Dynamiczne modele ekonometryczne" – materiały konferencyjne, UMK, Toruń, s. 101-110.

Papla D. (2006), *Klasyfikacja spółek notowanych na GPW w Warszawie z wykorzystaniem funkcji powiązań (copula functions)*, [w:] Taksonomia 13, red. K. Jajuga, M. Walesiak, Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 1126, AE, Wrocław, s. 271-279.
Papla D., Piontek K., *Zastosowanie rozkładów α-stabilnych i funkcji powiązań (copula) w obliczaniu wartości zagrożonej (VaR)*, praca złożona do Zeszytu Jubileuszowego Instytutu Zarządzania Finansami Akademii Ekonomicznej we Wrocławiu, Wrocław 2005 (w druku).
Rokita P. (2005), *Zastosowanie modeli wykorzystujących funkcje powiązań (copula functions) i teorię wartości ekstremalnych w analizie ryzyka portfeli akcji*, maszynopis AE Wrocław, referat wygłoszony na konferencji „Modelowanie preferencji a ryzyko".
Rokita P. (2006), *Zastosowanie archimedesowskich funkcji powiązań (Archimedean copulas) o liczbie wymiarów większej niż 2 w analizie ryzyka portfela na rynku polskim*, [w:] Taksonomia 13, red. K. Jajuga, M. Walesiak, Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 1126, AE, Wrocław, s. 280-288.
Schweizer B., Sklar A. (1974), *Operations on Distributions Functions not Derivable from Operations on Random Variables*, „Studia Mathematica", 52, s. 43-52.
Sklar A. (1959), *Fonctions de répartition á n dimensions et leurs marges*, „Publications de l'Institut Statistique de l'Université de Paris", 8, s. 229-231.

THE EXAMPLE OF USING FARLIE-GUMBEL-MORGENSTERN COPULA FUNCTION IN DEPENDENCE ANALYSIS AMONG CHOSEN STOCKS FROM WSE

Summary

The main goal of this paper is the analysis of dependence among stocks from the Polish stock market. The main hypothesis was formulated in the way: the dependence among stocks on the Polish stock market is not only two-dimensional, i.e. there is dependence between two stocks, but it is also three-dimensional, i.e. there is dependence between three stocks. This three-dimensional dependence one can analyze using multidimensional (in this case three-dimensional) copula functions, which are presented in the first part of this paper. If this hypothesis is positively verified, the dependence among stocks on the Polish stock market is more complicated than the classical capital market models imply. The second part of the paper presents Farlie-Gumbel-Morgenstern (FGM) function used to verify this hypothesis. The analysis consists of two stages. The estimation of the coefficients of two-dimensional FGM function for pairs of stocks was the first stage and the estimation of coefficients of three-dimensional FGM function for threes of stocks was the second stage. The comparison of results of those estimations was a basis for the verification of research hypothesis presented in the last part of the paper.