A C T A   U N I V E R S I T A T I S   L O D Z I E N S I S
FOLIA OECONOMICA 269, 2012

Joanna Tomanek[*]

TESTS OF MULTIVARIATE INDEPENDENCE BASED ON COPULA

Abstract. Very often the aim of statistical analysis is to identify dependencies among variables. More and more multidimensional variables and processes are in focus. This paper presents the tests of multivariate independence based on the empirical copula and the Möbius transform. The important contribution to the development of this test had works of Blum, Kiefer, Rosenblatt (1961), Dugue (1975), Deheuvels (1981), Ghoudi, Kulperger, Remillard (2001), Genest, Rémillard (2004) and Kojadinovic, Holmes (2009). The first section of the article presents the copula function and the empirical copula. The next section introduces the multivariate independence tests and the last section gives the empirical example.
Key words: Multivariate independence test, empirical copula process, Möbius transform.

I. COPULA FUNCTION

According to Sklar (1959), the joint density of a continuous multidimensional variable can be expressed uniquely as a product of the marginal densities and a copula function, which is a function of corresponding probability distribution functions of margins. The d-dimensional cumulative distribution function F with continuous margins $F_1, \dots, F_d$ can be presented as

$$F(x_1, x_2,\dots, x_d) = C(F_1(x_1), F_2(x_2),\dots, F_d(x_d)) \qquad (1)$$

where $F_1, \dots, F_d$ denote the cumulative distribution functions of the d random variables $x_1, \dots, x_d$ and C is a copula function.

A simple non-parametric estimator of the copula function is empirical copula proposed by Deheuvels (1979). For a random sample $(X_{11},\dots,X_{1d}),\dots, (X_{n1},\dots,X_{nd})$ empirical copula can be expressed as in Bouyé et al. (2000), Fermanian et al (2004):

[*] M.Sc., PhD student at the Independent Department of Demography and Economics Statistics, University of Economics in Katowice

$$C_n(u_1,\ldots,u_d) = \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{d} \mathrm{I}\{F_{j,n}(X_{ij}) \le u_{ij}\} \tag{2}$$

where I denotes the indicator function and $F_{j,n}$ is n/(n + 1) times the empirical distribution function of X_j based on the random sample

$$(X_{11},\ldots,X_{1d}),\ldots,(X_{n1},\ldots,X_{nd}):\ F_{j,n}(x_j) = \frac{1}{n+1}\sum_{i=1}^{n} I(X_{ij} \le x_{ij}),\ 1 \le j \le d \tag{3}$$

One can observe that empirical copula C_n is actually a function of the ranks of these observations.

II. TEST OF MULTIVARIATE INDEPENDENCE

Since the dependence structure among the variables is completely summarized by the copula and mutual independence occurs if and only if $C(u_1, u_2,\ldots, u_d) = \prod_{j=1}^{d} u_j$, a test of the mutual independence of the components of X can be the statistic I_n as proposed by Genest and Rémillard (2004):

$$I_n = \int_{[0,1]^d} n\left\{C_n(u) - \prod_{j=1}^{d} u_j\right\}^2 du \tag{4}$$

What is interesting here is the fact that, under the mutual independence of the components $X_1, \ldots, X_d$ of X, the empirical process $\sqrt{n}\{C_n - \prod\}$ can be decomposed, using the Möbius transform into 2^d – d- 1 sub-processes $\sqrt{n}M_A(C_n), A \subseteq \{1,\ldots,d\}, |A| > 1$, that converge jointly to tight centered mutually independent Gaussian processes. Möbius transform has a form of :

$$M_A(C_n) = G_{A,n}, A \subseteq \{1,\ldots,d\}, |A| > 1$$

$$G_{A,n}(u_1,\ldots,u_d) = \sum_{B \subset A} (-1)^{|A \setminus B|} C_n(u^B) \prod_{j \in A \setminus B} u_j \tag{5}$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \prod_{j \in A} [I\{R_{ij} \le (n+1)u_j\} - u_j]$$

Where

$$u^B \in [0,1]^d \text{ and } u_j^B = \begin{cases} u_j & \text{if} \quad j \in B \\ 1 & \text{if} \quad j \notin B \end{cases}$$

Genest and Remillard (2004) showed that mutual independence among X_1, ..., X_d is equivalent to having $M_A(C) = 0$, for all $u \in [0,1]^d$ and all $A \subseteq \{1,...,d\}$ such that $|A| > 1$. Instead of the single test statistic I_n one can consider $2^d - d - 1$ test statistics of the form

$$\mathrm{M}_{A,n} = \int_{[0,1]^d} n\{M_A(C_n)(u)\}^2 du \text{ , } A \subseteq \{1,...,d\}, |A| > 1 \tag{6}$$

$\mathrm{M}_{A,n}$ are asymptotically mutually independent under the null hypothesis of independence. Each statistic $\mathrm{M}_{A,n}$ can be seen as focusing on the dependence among the components of X whose indices are in A. Kojadinovic and Holmes (2009) has been recently extended the above decomposition to the situation where one wants to test the mutual independence of several continuous random vectors.

Under the null hypothesis of independence or randomness statistics the limiting distribution of $\mathrm{M}_{A,n}$ depends only on the number of elements in A:

$$\xi_k = \sum_{(i_1,...,i_k) \in N^k} \frac{1}{\pi^{2k}(i_1 \cdots i_k)^2} Z_{i_1,...,i_k}^2 \tag{7}$$

Where $Z_{i_1,...,i_k}$ are independent N(0, 1) random variables.

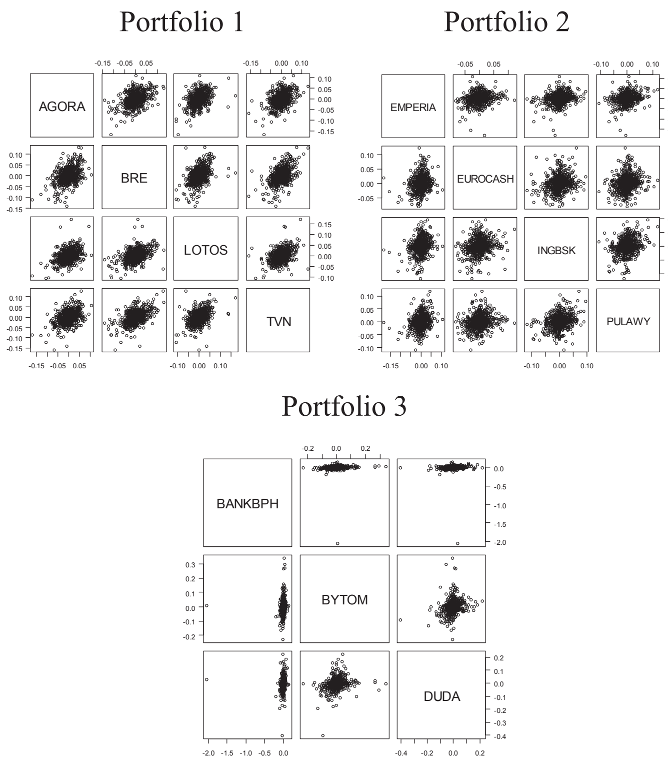
One can combine the 2d − d − 1 statistics $\mathrm{M}_{A,n}$ into one global statistic for testing independence in a way to achieve a desired significance level or one can find first individual p-values and afterwards aggregate them using for example the combination rules of Fisher (1932) or Tippett (1931).

The results of the independence test can be visualized in a graphical representation, called a dependogram. For each subset $A \subseteq \{1,...,d\}$, $|A| > 1$, a vertical bar is drawn whose height is proportional to the value of $\mathrm{M}_{A,n}$. The approximate critical values of $\mathrm{M}_{A,n}$ are represented on the bars by the bullets. Subsets for which the bar exceeds the critical value can be considered as being composed of dependent variables.

III. EXAMPLE ON THE POLISH STOCKS MARKET

The described test was employed to analyze of several portfolios of stocks traded on Warsaw Stock Exchange. On the stage of risk optimization the investor was interested in 3 portfolios. It was important to determine if there is dependence between portfolios and between the components of each portfolio. The first portfolio consisted of 4 stocks: AGORA, BRE, LOTOS, TVN, the second of: EMPERIA, EUROCASH, INGBSK, PULAWY and the last one of: BANKBPH, BYTOM, DUDA. The research was carried out on the basis of historical daily quotes between 02.01.2007 and 31.12.2009.

Below are presented the scatter plots of the daily returns of each stock in the portfolio in the period and the values of correlation coefficients calculated accordingly to Pearson, Kendall and Spearman method.



Graph 1. Scatterplots of the daily returns of each stock in the portfolio in the period from 02.01.2007 to 31.12.2009

Source: Own calculation.

Table 1. Values of correlation coefficients calculated accordingly to Pearson, Kendall and Spearman method

Portfolio 1	*Pearson*				*Kendall*				*Spearman*			
	AGO	BRE	LOT	TVN	AGO	BRE	LOT	TVN	AGO	BRE	LOT	TVN
1. AGORA	1.00	0.43	0.41	0.42	1.00	0.26	0.26	0.25	1.00	0.37	0.37	0.37
2. BRE	0.43	1.00	0.45	0.49	0.26	1.00	0.30	0.30	0.37	1.00	0.42	0.43
3. LOTOS	0.41	0.45	1.00	0.41	0.26	0.30	1.00	0.27	0.37	0.42	1.00	0.39
4. TVN	0.42	0.49	0.41	1.00	0.25	0.30	0.27	1.00	0.37	0.43	0.39	1.00
Portfolio 2	*Pearson*				*Kendall*				*Spearman*			
	EMP	EUR	ING	PUL	EMP	EUR	ING	PUL	EMP	EUR	ING	PUL
1. EMPERIA	1.00	0.21	0.25	0.26	1.00	0.13	0.16	0.16	1.00	0.20	0.23	0.23
2. EUROCASH	0.21	1.00	0.19	0.18	0.13	1.00	0.14	0.14	0.20	1.00	0.20	0.21
3. INGBSK	0.25	0.19	1.00	0.28	0.16	0.14	1.00	0.18	0.23	0.20	1.00	0.26
4. PULAWY	0.26	0.18	0.28	1.00	0.16	0.14	0.18	1.00	0.23	0.21	0.26	1.00

Portfolio 3	*Pearson*			*Kendall*			*Spearman*		
	BAN	BYT	DUD	BAN	BYT	DUD	BAN	BYT	DUD
1. BANKBPH	1.00	0.07	0.05	1.00	0.15	0.15	1.00	0.22	0.21
2. BYTOM	0.07	1.00	0.24	0.15	1.00	0.19	0.22	1.00	0.28
3. DUDA	0.05	0.24	1.00	0.15	0.19	1.00	0.21	0.28	1.00

Source: Own calculation.

One can notice that the analyzed stocks are characterized by a relatively low linear correlation coefficient, however graphs indicate some dependence. The returns values are close to zero, which is typical for the stocks returns.

The popular Pearson coefficient measures only linear correlation and its value close to zero does not mean independence between the variables. This measure also assumes the normal distribution, that often is not true. This assumption is not required in case Kendall (τ) and Spearman (ρ rank correlation coefficients. The both measures unlike Pearson coefficient indicate not only linear but also monotonical dependence. Copula function represents the structure of the dependence in a more general way, as a function.

In the example one can notice that there are differences in the assessment of dependence given by different measures. Correlation between returns of BYTOM and BANKBPH was measured by the Pearson coefficient on the level of 0.07, with the Kendall coefficient on the level of 0.15, and with the Spearman coefficient as 0.22. As the number of observations is high, excluding the coefficient for the pair of stocks DUDA and BANKBPH, all the other coefficients are significant.

Regardless of the method, the highest values of correlation coefficient were received for the first portfolio. These values are in range 0.4–0.5. The biggest discrepancies in the measures of the correlation appeared in the third port folio. According to the Pearson coefficient the correlation was low, and in case of pairs (BANKBPH, BYTOM) and (BANKBPH, DUDA) – close to zero.

These results one can compare with the information received from the tests of multivariate independence based on the empirical process and $\sqrt{n}\{C_n - \Pi\}$ the Möbius transform. The table below presents the results of the test.

Table 2. Results of the multivariate independence tests based on the empirical process $\sqrt{n}\{C_n - \Pi\}$ and the Möbius transform

Portfolio 1			
Subset A	Statistic $M_{A,n}$	Critical value	P-value
{1,2}	0.945	0.101	0
{1,3}	0.973	0.101	0
{1,4}	0.954	0.101	0
{2,3}	1.254	0.101	0
{2,4}	1.288	0.101	0
{3,4}	1.113	0.101	0
{1,2,3}	0.037	0.011	0
{1,2,4}	0.034	0.011	0
{1,3,4}	0.024	0.011	0
{2,3,4}	0.034	0.011	0
{1,2,3,4}	0.013	0.002	0

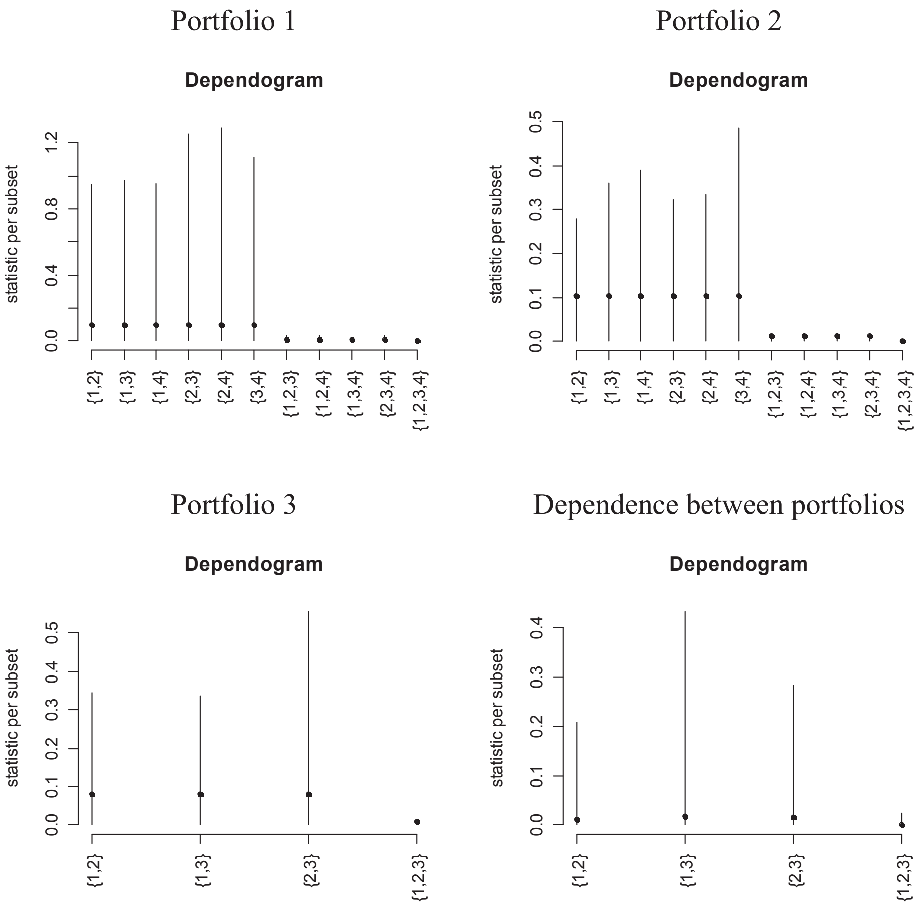
Portfolio 2			
Subset A	Statistic $M_{A,n}$	Critical value	P-value
{1,2}	0.279	0.104	0
{1,3}	0.361	0.104	0
{1,4}	0.389	0.104	0
{2,3}	0.323	0.104	0
{2,4}	0.335	0.104	0
{3,4}	0.485	0.104	0
{1,2,3}	0.01	0.013	0.018
{1,2,4}	0.009	0.013	0.035
{1,3,4}	0.005	0.013	0.228
{2,3,4}	0.012	0.013	0.007
{1,2,3,4}	0.002	0.002	0

Portfolio 3			
Subset A	Statistic $M_{A,n}$	Critical value	P-value
{1,2}	0.344	0.082	0
{1,3}	0.337	0.082	0
{2,3}	0.555	0.082	0
{1,2,3}	0.016	0.01	0.001

Dependence between portfolios			
Subset A	Statistic $M_{A,n}$	Critical value	P-value
{1,2}	0.208	0.013	0
{1,3}	0.431	0.019	0
{2,3}	0.282	0.016	0
{1,2,3}	0.025	0.001	0

Source: Own calculation.

Very low p-value levels indicate the rejection of the null hypothesis of independence between the components of each portfolio and in the case of the last test, between the portfolios. The only case where rejection of the null hypothesis is unjustified is a triplet: EMPERIA (1), INGBSK (3), PULAWY (4), where the p-value is on the level of 0.228. The results of the test are visualized on the dependograms below.



Graph 2 Dependograms – visualization of the multivariate independence tests
Source: Own calculation.

With the dependorams, where for each subset $A \subseteq \{1,...,d\}$, $|A| > 1$, a vertical bar is proportional to the value of $\mathrm{M}_{A,n}$ and the approximate critical values of $\mathrm{M}_{A,n}$

are represented on the bars by the black bullets, it is possible to quickly realize where the critical value was exceeded. The excess of the critical value in pairs is evident, however in case of triples the distance to the critical value is smaller.

IV. CONCLUSIONS

The tests of multivariate independence based on the empirical copula and the Mobius transform that was presented in this paper is an interesting proposal for testing independence. The copula summarizes completely the dependence structure among the variables, therefore it can be used for testing independence. On the other hand, Möbius transform enables obtaining useful test statistics. The results of the independence test can be visualized on a dependogram.

REFERENCES

Bouyé E., Durrleman V., Nikeghbali A., Riboulet G., Roncalli T. (2000), *Copulas for Finance – A Reading Guide and Some Applications*, http://ssrn.com/abstract=1032533

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

Fermanian J.-D., Radulovic D., Wegkamp M.(2004), Weak convergence of empirical copula processes, *Bernoulli,* 10(5), 847–860

Franke J., Haerdle W., Hafner Ch.(2004), *Statistics of Financial Markets*, Springer, Berlin

Genest, C., Rémillard, B. (2004) Tests of Independence or Randomness Based on the Empirical Copula Process, *Test*, 13, 335–369

Heilpern S.(2007), Eliptyczne funkcje łączące, *Prace naukowe Akademii Ekonomicznej we Wrocławiu*, 1189, Wrocław

Jajuga K.(2207), *Zarządzanie ryzykiem*, PWN, Warszawa

Kojadinovic I., Holmes M. (2009), Tests of Independence among Continuous Random Vectors Based on Cramer-von Mises Functionals of the Empirical Copula Process, *Journal of Multivariate Analysis*, 100(6), 1137–1154

Kojadinovic I, Yan J.(2010), Modeling Multivariate Distributions with Continuous Margins using the copula R Package, *Journal of Statistical Software*, 34(9)

Joanna Tomanek

TEST NIEZALEŻNOŚCI WEKTORÓW LOSOWYCH W OPARCIU O FUNKCJĘ POŁĄCZEŃ

W ostatnim czasie w centrum zainteresowania stoją procesy i zmienne wielowymiarowe. Prezentowany w artykule wielowymiarowy test niezależności pozwala na weryfikację istnienia zależności pomiędzy składowymi danego wektora, zależności pomiędzy wieloma wektorami czy badanie losowości wielowymiarowego szeregu w czasie. Jego istota polega na wykorzystaniu własności funkcji połączeń oraz dekompozycji Mobiusa. W pierwszej części artykułu wprowadzone zostały pojęcia funkcji połączenia oraz empirycznej funkcji połączenia. W dalszej kolejności przedstawione zostały główne założenia wielowymiarowego testu niezależności, a w ostatniej przykład empiryczny dotyczący zależności na polskim rynku akcji.