

Anna Czapkiewicz, Beata Basiura

Akademia Górniczo-Hutnicza w Krakowie

GRUPOWANIE INDEKSÓW ŚWIATOWYCH Z UWZGLĘDNIENIEM PRZESUNIĘĆ CZASOWYCH NA PODSTAWIE MODELI *COPULA*-GARCH

Streszczenie: W pracy zaprezentowana została próba pogrupowania danych, którymi są dzienne stopy zwrotu 42 indeksów światowych. Jako miarę powiązań między poszczególnymi indeksami przyjęto współczynnik korelacji, który jest parametrem funkcji połączeń *t*-Studenta. W oparciu o ten współczynnik zdefiniowano miarę odległości, pozwalającą utworzyć podział na grupy taksonomiczne. Celem badania jest określenie, czy istnieje wpływ przesunięcia czasowego na wyniki grupowania.

Słowa kluczowe: klasyfikacja indeksów giełdowych, miara zależności, model *Copula*-GARCH, funkcje połączeń, skośne rozkłady brzegowe.

1. Wstęp

Prezentowana praca jest kontynuacją badania, którego wyniki zamieszczono w publikacji Czapkiewicz, Basiura [Czapkiewicz, Basiura 2010, 2010a]. Celem niniejszego opracowania jest sprawdzenie, czy uwzględnienie różnic w strefach czasowych ma wpływ na grupowanie stóp zwrotu światowych indeksów giełdowych.

Wiadomo, że między rynkami azjatyckimi, europejskimi oraz amerykańskimi występują różnice czasowe. Fakt ich występowania jest często uwzględniany w badaniach empirycznych (np. [Jondeau, Rockinger 2006]). Nie ma natomiast badań, czy uwzględnienie tych zmian czasowych jest konieczne w celu poprawnego diagnozowania powiązań pomiędzy rynkami. W prezentowanej pracy autorki analizowały związki pomiędzy głównymi indeksami giełd światowych w dwóch przypadkach: przy uwzględnieniu oraz przy braku uwzględnienia różnic czasowych w rynkach międzynarodowych.

Badanie zależności między stopami zwrotu różnych indeksów jest jednym z ważniejszych zagadnień w modelowaniu rynków kapitałowych. Popularna, ze względu na swoje właściwości, miara zależności, jaką jest współczynnik korelacji Pearsona, może być stosowana tylko wtedy, gdy stopy zwrotu mają rozkład elip-

tyczny. W przypadku braku takich założeń wnioskowanie w oparciu o współczynnik korelacji liniowej Pearsona może nie być właściwe. Zdarza się, że analiza powiązań rozważana na tej podstawie wskazuje na brak zależności w przypadku, gdy w rzeczywistości związek między badanymi szeregami czasowymi istnieje.

Szeregi utworzone z dziennych stóp zwrotu charakteryzują się tzw. grubymi ogonami. Dodatkowo rozkłady te cechuje duża kurtoza i silna asymetria. Do modelowania rozkładów stóp zwrotu często stosuje się model GARCH(1,1) zaproponowany przez Bollersleva [1986]. Jako rozkład warunkowy w modelu GARCH przyjmowane są rozkłady skośne, najczęściej rozkład skośny t -Studenta lub rozkład skośny GED.

Do modelowania wielowymiarowych rozkładów stóp zwrotu Embreecht i in. [2001] zaproponowali zastosowanie funkcji połączeń (*copula function*). Podejście to umożliwia rozważanie osobno rozkładów brzegowych i łącznego wielowymiarowego ciągłego rozkładu. Badanie wówczas przebiega w dwóch etapach. Pierwszy etap to modelowanie rozkładów brzegowych, natomiast drugi to modelowanie związków pomiędzy rozkładami brzegowymi.

W prezentowanej pracy do rozkładów brzegowych zastosowano model GARCH(1,1) z warunkowym rozkładem skośnym t -Studenta oraz skośnym GED. Jako funkcję połączeń przyjęto funkcję połączeń t -Studenta. Należy nadmienić, że *copula* t -Studenta rekomendowana jest przez takich autorów, jak Mashal, Zeevi [2002] oraz Breymann [Breymann i in. 2003].

W badaniu uwzględniono losowo wybrane czterdzieści dwa indeksy giełdowe, których notowania datowane są od stycznia 2000 do grudnia 2009. Jako zmienną losową charakteryzującą zachowanie rynku przyjęto logarytmiczną dzienną stopę zwrotu. Do wyodrębnienia podgrup zastosowano metodę Warda z macierzą odległości zdefiniowaną w oparciu o wyznaczony współczynnik korelacji uzyskany z kopuli t -Studenta. Wybór tej kopuli poprzedzony był odpowiednim testowaniem.

2. Rozkłady brzegowe

W paragrafie tym przedstawiono model, który został przyjęty do opisu rozkładów brzegowych. Niech y_t dla $t = 1, 2, 3, \dots, T$ będzie stopą zwrotu danego indeksu. Podobnie jak w wielu pracach poświęconych tej tematyce założono, że spełnia ona model GARCH(1,1):

$$y_t = \mu + \varepsilon_t, \quad \varepsilon_t = \sqrt{h_t} \eta_t, \quad \eta_t \sim iid(0,1), \quad h_t = a_0 + a_1 \varepsilon_{t-1}^2 + a_2 h_{t-1}.$$

Zakłada się, że $a_0 > 0$, $a_1, a_2 > 0$, oraz $a_1 + a_2 < 1$. W klasycznym modelu GARCH(1,1), zakłada się także, że warunkowy rozkład ε_t jest normalny. Okazuje się jednak, że rzeczywiste reszty modelu posiadają warunkowy rozkład o grubszych ogonach niż rozkład normalny oraz cechuje je asymetria. Stąd zastosowanie

skośnego rozkładu warunkowego, którego gęstość ogólnie można przedstawić następująco:

$$f(x; \xi, \nu) = \frac{2}{a(\xi) + b(\xi)} \left[g\left(\frac{x}{a(\xi)}\right) I_{(x < 0)} + g\left(\frac{x}{b(\xi)}\right) I_{(x > 0)} \right],$$

gdzie ξ oznacza parametr asymetrii. Jako normalizujące funkcje $a(\xi)$, $b(\xi)$ przyjęto $a(\xi) = \xi$ oraz $b(\xi) = \xi^{-1}$ [Fernández, Steel 1998]. Funkcja $g(\cdot)$ oznacza gęstość rozkładu t -Studenta lub GED. Funkcja gęstości rozkładu t -Studenta jest postaci:

$$f_S(x, \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) \sqrt{(\nu-2)\pi}} \left(1 + \frac{x^2}{(\nu-2)}\right)^{-\frac{\nu+1}{2}},$$

natomiast gęstość rozkładu GED ma postać:

$$f_G(x, \nu) = \frac{\nu \exp\left\{-\frac{1}{2}\left|\frac{x}{\lambda}\right|^\nu\right\}}{\lambda \Gamma\left(\frac{1}{\nu}\right)} 2^{\frac{\nu+1}{\nu}}, \text{ przy czym } \lambda = \left[\frac{\Gamma\left(\frac{1}{\nu}\right)}{\Gamma\left(\frac{3}{\nu}\right)} 2^{\frac{2}{\nu}}\right]^{1/2},$$

gdzie ν oznacza liczbę stopni swobody.

3. Poprawność doboru rozkładów brzegowych

Badając poprawność doboru modelu, zastosowano rozumowanie Diebolda i in. [1998]. Niech ε_t będą standaryzowanymi resztami w modelu GARCH(1,1) oraz $F(x)$ będzie dystrybucją rozkładu skośnego t -Studenta lub rozkładu skośnego GED. Właściwy wybór rozkładu warunkowego implikuje jednostajność rozkładu $u_{it} = F_k(\varepsilon_{it})$, $k = 1, 2$. Natomiast właściwy model implikuje niezależność u_{it} . Skorzystano z twierdzenia, że jeśli szereg $\{y\}_{t=1}^T$ ma rozkład warunkowy o dystrybuc-

cie F_t , to przekształcona zmienna losowa $u_t = F_t\left(\frac{\varepsilon_t}{h_t}\right)$ jest niezależna o tym samym rozkładzie jednostajnym w przedziale $[0, 1]$.

W celu testowania niezależności u_t została zastosowana cytowana wcześniej procedura [Diebold i in. 1998]. Dla modelu regresji wielorakiej, gdzie zmienną zależną była wielkość $(u_t - \bar{u})^k$, $k = 1, \dots, 4$, a zmiennymi niezależnymi $(u_{t-s} - \bar{u})^k$, przy czym s oznacza rząd opóźnienia od 1 do 20, który został ustalony arbitralnie.

Z modelu tego wyznaczono wartość współczynnika determinacji R^2 . Statystyka testowa $R(k) = (T - 20)R^2$, przy założeniu poprawności hipotezy zerowej, ma rozkład chi-kwadrat (χ_{20}^2) z dwudziestoma stopniami swobody. Test zgodności z rozkładem jednostajnym przeprowadzono w oparciu o tradycyjny test chi-kwadrat.

4. Modele funkcji połączeń

Funkcja połączeń (*copula function*) jest wielowymiarową dystrybuantą z jednostajnymi w przedziale $[0,1]$ rozkładami brzegowymi. Funkcja $C : [0,1]^d \rightarrow [0,1]$ jest d -wymiarową funkcją połączeń, jeśli spełnia następujące warunki:

- (1) Dla wszystkich $u_i \in [0, 1]$, $C(1, \dots, 1, u_i, 1, \dots, 1) = u_i$.
- (2) Dla każdego $u \in [0, 1]^d$, $C(u_1, \dots, u_d) = 0$, jeśli co najmniej jedna współrzędna $u_i = 0$.
- (3) C jest funkcją d -rosnącą.

Jeśli $X = (X_1, \dots, X_n) \in R^d$ będzie d -wymiarową zmienną losową o ciągłej dystrybuancie F :

$$F(x_1, \dots, x_d) = P(X_1 \leq x_1, \dots, X_d \leq x_d),$$

to według twierdzenia Sklara [1959] istnieje jednoznaczna funkcja połączeń $C : [0,1]^d \rightarrow [0,1]$, taka że:

$$F(x_1, \dots, x_d) = P(X_1 \leq x_1, \dots, X_d \leq x_d) = C(F_1(x_1), \dots, F_d(x_d)),$$

gdzie $F_n(x)$ jest dystrybuantą rozkładu brzegowego, czyli

$$F_n(x_n) = P(X_n \leq x_n), \quad x_n \in R, \quad n = 1, \dots, d.$$

Fundamentalnym wnioskiem z twierdzenia Sklara jest fakt, że wielowymiarowy ciągły rozkład i rozkłady brzegowe mogą być rozważane osobno, a miara zależności między nimi może być reprezentowana funkcją połączeń.

Podstawową klasę funkcji połączeń stanowią tzw. *copule* eliptyczne, do których należy między innymi funkcja połączeń t -Studenta. Dwuwymiarowa funkcja połączeń t -Studenta wyraża się jako:

$$C(u, v) = t_{\rho, \eta}(t_{\eta}^{-1}(u)t_{\eta}^{-1}(v)),$$

gdzie t_{η} jest dystrybuantą rozkładu t -Studenta z η stopniami swobody, natomiast $t_{\rho, \eta}$ jest dystrybuantą dwuwymiarowego rozkładu t -Studenta z η stopniami swobody i współczynnikiem korelacji ρ .

W celu uzasadnienia, czy dana funkcja połączeń jest dobrze dopasowana do danych, w tej pracy zastosowano test zgodności. Zaproponowano procedurę testowania opartą na dwuwymiarowym teście zgodności chi-kwadrat. Algorytm stosowania tego testu wyglądał następująco: Niech A_{ij} będzie częścią obszaru $[0, 1]^2$ (oś y została podzielona na r części oraz oś x na s części), O_{ij} niech będzie obserwowaną liczebnością dla obszaru A_{ij} oraz niech $E_{ij}(\hat{\theta}) = n \cdot p_{ij}(\hat{\theta})$ będzie teoretyczną liczebnością dla tego obszaru A_{ij} , przy czym $p_{ij}(\hat{\theta}) = \iint_{A_{ij}} dC(u, v; \hat{\theta})$. Wtedy statystyka testowa ma postać:

$$\chi^2(\hat{\theta}) = \sum_{i=1}^r \sum_{j=1}^s \frac{(O_{ij} - E_{ij}(\hat{\theta}))^2}{E_{ij}(\hat{\theta})},$$

gdzie $\hat{\theta}$ jest d -wymiarowym wektorem oszacowanych parametrów.

Statystyka ta ma asymptotycznie rozkład chi-kwadrat z $(rs - d - 1)$ stopniami swobody. Wadą tego testu jest bardzo duża czułość na sposób wyboru podziałów [Roch, Alegre 2006]. W części empirycznej zastosowano podział na sto równych obszarów, przyjmując $r = s = 10$.

5. Badanie empiryczne

W pierwszym etapie badania dla utworzonych szeregów stóp zwrotu dokonano estymacji ich rozkładów, przyjmując model GARCH(1,1) z dwoma wymienionymi wcześniej rozkładami warunkowymi: rozkładem skośnym t -Studenta i rozkładem skośnym GED. W dalszej analizie uwzględniane były te rozkłady, dla których funkcja największej wiarygodności była większa. Tabela 1 zawiera wyniki testu sprawdzającego poprawność doboru modelu odpowiednio dla skośnego rozkładu t -Studenta lub dla skośnego GED. O wyborze rozkładu zadecydowano na podstawie kryterium AIC. Wielkości $p(1)$, $p(2)$, $p(3)$ i $p(4)$ oznaczają odpowiednio oszacowane poziomy istotności dla statystyki testowej $R(1)$, $R(2)$, $R(3)$ i $R(4)$. Natomiast wyestymowany poziom istotności dla testu zgodności chi-kwadrat w tabeli oznaczamy jako p . Gwiazdką zaznaczono kraje, z których stopy zwrotu indeksów były modelowane rozkładem warunkowym t -Studenta. Wnioskowanie prowadzono dla przyjętego poziomu istotności 1%. Tabela 1 podaje tylko te kraje, dla których testowanie rozkładów dziennych stóp zwrotu indeksów giełdowych nie odrzuciło hipotezy na poziomie istotności 1%.

Tabela 1. Wartości p dla odpowiednich statystyk testowych

Kraj	$p(1)$	$p(2)$	$p(3)$	$p(4)$	p
Argentyna	0,046	0,157	0,651	0,322	0,898
Australia*	0,155	0,889	0,305	0,909	0,364
Austria	0,545	0,911	0,393	0,849	0,434
Belgia	0,734	0,550	0,021	0,722	0,135
Kanada	0,718	0,270	0,595	0,167	0,205
Czechy	0,010	0,930	0,012	0,669	0,033
Finlandia*	0,478	0,180	0,261	0,029	0,641
Niemcy	0,241	0,011	0,558	0,012	0,039
Holandia*	0,069	0,012	0,208	0,010	0,137
Hongkong	0,065	0,308	0,100	0,535	0,247
Węgry	0,068	0,885	0,039	0,868	0,556
Włochy	0,377	0,227	0,729	0,414	0,697
Japonia	0,693	0,095	0,633	0,082	0,194
Singapur*	0,216	0,275	0,539	0,194	0,756
Norwegia	0,566	0,684	0,335	0,749	0,033
Polska	0,241	0,044	0,035	0,112	0,627
Turcja*	0,078	0,386	0,706	0,136	0,365
Zelandia*	0,325	0,031	0,950	0,104	0,906
Korea Pd.	0,523	0,539	0,290	0,669	0,675
Hiszpania	0,597	0,323	0,388	0,376	0,357
Szwajcaria*	0,025	0,011	0,178	0,015	0,208
USA DJIA	0,176	0,011	0,463	0,011	0,708

Źródło: obliczenia własne.

Analizując wyniki tab. 1, zauważamy, że model GARCH(1,1) z warunkowym rozkładem skośnym jest dobrze dopasowany do ponad połowy analizowanych stóp zwrotu z indeksów giełdowych.

Do sprawdzenia wpływu różnic czasowych zastosowano następującą procedurę. Notowania grupy krajów europejskich zostały uznane za bazowe. Notowania grupy krajów amerykańskich zostały przesunięte o jedno notowanie w dół, natomiast notowania grupy krajów azjatyckich i Australii zostały przesunięte o jeden do góry.

Następnym krokiem jest estymacja współczynnika korelacji funkcji połączeń ρ . Idealnym rozwiązaniem byłaby estymacja w jednym kroku całej macierzy korelacji dla wszystkich zmiennych. Jednakże ze względu na duży wymiar tej macierzy estymację parametrów przeprowadzono dla poszczególnych par oddzielnie. Rozpatrywano zatem tylko dwuwymiarowe funkcje połączeń. Stosując algorytm IFM [Joe, Xu 1996], otrzymano estymatory tego parametru pomiędzy kolejno branymi parami indeksów. Poprawność wyboru kopuli t -Studenta potwierdzona została testem zgodności chi-kwadrat. Wyniki testu przedstawia tab. 2, gdzie zamieszczono procent z wszystkich analizowanych

par indeksów, dla których przeprowadzony test zgodności z wybraną funkcją połączeń nie dał podstaw do odrzucenia hipotezy zerowej. W panelu pierwszym analizowano pary utworzone z wszystkich rozważanych indeksów, natomiast w panelu drugim – tylko te pary indeksów, dla których rozkłady brzegowe zostały potwierdzone w tab. 1.

Tabela 2. Procent z wszystkich możliwych par, dla których przeprowadzony test zgodności z wybraną funkcją połączeń nie dał podstaw do odrzucenia hipotezy zerowej

Funkcja połączeń <i>t</i> -Studenta	Poziom istotności $\alpha = 0,05$	Poziom istotności $\alpha = 0,01$
	PANEL I	
	67	83
PANEL II		
	74	92

Źródło: obliczenia własne.

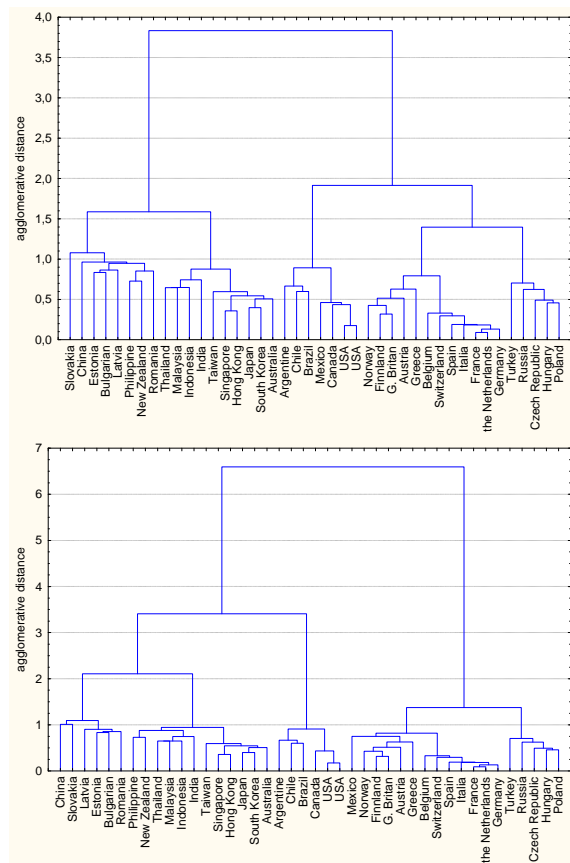
Uwaga: PANEL I – wszystkie możliwe przypadki, PANEL II – pary utworzone tylko z tych indeksów, dla których rozkłady brzegowe zostały dobrze wyspecyfikowane.

Mając oszacowaną macierz korelacji, przystąpiono do grupowania szeregów. Do grupowania wybrano algorytm aglomeracji metodą Warda [Grabiński i in. 1989]. Jako miarę odległości przyjęta została odległość oparta na współczynniku korelacji, który jest parametrem kopuli *t*-Studenta: $d_{ij} = 1 - \rho_{ij}$.

Grupowanie przeprowadzono dla dwóch zestawów szeregów. W pierwszym zestawie nie uwzględniano przesunięcia czasowego, natomiast w drugim zestawie uwzględniono różnicę czasową. Wyniki obu grupowań przedstawiono na rys. 1.

Analizując dendrogramy (rys. 1), widzimy, że uwzględnienie przesunięcia czasowego generalnie nie miało wpływu na grupowanie. Można też zauważyć, że indeks meksykański oraz indeksy z Filipin i Nowej Zelandii przesunęły się do innej grupy. Wydaje się jednakże, że występujące delikatne różnice w klasyfikacji mają raczej charakter przypadkowy i występują pomiędzy indeksami krajów słabo ze sobą powiązanych.

W obu przypadkach, gdy rozważano przesunięcie czasowe oraz gdy tego przesunięcia nie brano pod uwagę, uzyskano podobne grupy. Do grupy pierwszej należą: Polska, Węgry, Czechy oraz Rosja i Turcja. Druga grupa, zawierająca podgrupę krajów najsilniej powiązanych ze sobą, obejmuje Norwegię, Finlandię, Wielką Brytanię, Austrię, Grecję, Belgię, Szwajcarię, Hiszpanię, Włochy, Francję, Holandię, Niemcy. Trzecia grupa reprezentuje kraje Ameryki Północnej i Południowej. Kolejna grupa to kraje azjatyckie i Australia. W ostatnim zbiorze znajdują się rynki bardzo słabo ze sobą powiązane.



Rys. 1. Wyniki grupowania otrzymane metodą Warda. Lewy dendrogram nie uwzględnia przesunięcia czasowego. Prawy dendrogram uwzględnia przesunięcie czasowe

Źródło: opracowanie własne.

6. Podsumowanie

W pracy przedstawiono grupowanie indeksów światowych metodą Warda. Jako miarę podobieństwa między danymi szeregami zastosowano parametr kopuli t -Studenta z modelu Copula-GARCH. Miarę odległości między dwoma szeregami czasowymi zdefiniowano w oparciu o współczynnik korelacji.

Grupowanie przeprowadzono dla szeregów, w których nie uwzględniono przesunięcia czasowego, oraz dla szeregów, w których uwzględniono przesunięcie czasowe. Wyniki przeprowadzonych badań wskazały na brak istotnego wpływu stref czasowych. Należy zauważyć, że w otrzymanych grupach znalazły się indeksy krajów z tych samych stref czasowych. Nie wystąpiła sytuacja, w której rynek z jednej

strefy czasowej jest silnie powiązany z rynkiem z innej strefy czasowej. Wydaje się, że delikatne różnice w klasyfikacji mają raczej charakter przypadkowy i występują pomiędzy indeksami krajów słabo ze sobą powiązanych.

Z przeprowadzonych klasyfikacji wynika wniosek, że najsilniej powiązane są rynki krajów Europy Zachodniej, takich jak Belgia, Szwajcaria, Hiszpania, Włochy, Francja, Holandia i Niemcy. Natomiast najsłabsze związki występują wśród rynków takich, jak Bułgaria, Estonia, Łotwa, Rumunia, Słowacja i Chiny.

Literatura

- Bollerslev T., *Generalized autoregressive conditional heteroskedasticity*, „Journal of Econometrics” 1986, no. 31, s. 307-327.
- Breymann W., Dias A., Embrechts P., *Dependence structures for multivariate high-frequency data in finance*, „Quantitative Finance” 2003, 3(1), s.1-16.
- Czapkiewicz A., Basiura B., *Clustering Financial Data Using Copula-GARCH Model in an Application for Main Market Stock Returns*, „Statistics in Transition (New Series)” 2010, vol. 11, no. 1, s. 25-45.
- Czapkiewicz A., Basiura B., *Grupowanie indeksów światowych na podstawie modeli copula-GARCH*, Prace Naukowe Uniwersytetu Ekonomicznego, Taksonomia 17, Wrocław 2010a, s. 81-89.
- Diebold F.X., Gunther T.A., Tay A.S., *Evaluating density forecasts with application to financial risk management*, „International Economic Review” 1989, 39(4), s. 863-883.
- Embreecht P., McNeil A.J., Straumann D., *Correlation and Dependency in Risk Management: Properties and Pitfalls*, [w:] M. Dempster, H. Moffant (red.), *Risk Management*, Cambridge University Press, New York 2001, s. 176-223.
- Fernández C., Steel M., *On Bayesian modelling of fat tails and skewness*, „Journal of the American Statistical Association” 1998, no. 93, s. 359-371.
- Grabiński T., Wydymus S., Zeliaś A., *Metody taksonomii numerycznej w modelowaniu zjawisk społeczno-gospodarczych*, Wydawnictwo Naukowe PWN, Warszawa 1989.
- Joe H., Xu J.J., *The Estimation Method of Inference Function for Margins for Multivariate Models*, Technical Report, Departments of Statistics, University of British Columbia 1996.
- Jondeau E., Rockinger M., *The Copula-Garch model of conditional dependencies: An international stock market application*, „Journal of International Money and Finance” 2006, no. 25, s. 827-853.
- Mashal R., Zeevi A., *Beyond Correlation: Extreme Co-movements Between Financial Assets* (October 14, 2002), Available at SSRN: <http://ssrn.com/abstract=317122> or DOI: 10.2139/ssrn.317122.
- Roch O., Alegre A., *Testing the bivariate distribution of daily equity returns using copulas. An application to the Spanish stock market*, „Computational Statistics & Data Analysis” 2006, no. 51, s. 1312-1329.
- Sklar A., *Fonction de répartition à n dimensions et leur marges*, Publications de L'Institut de Statistiques de L'Université de Paris, Paris 1959, s. 229-231.

WORLD INDEXES CLUSTERING USING THE *COPULA*-GARCH MODEL INCLUDING TIME DIFFERENCES

Summary: A fundamental problem in cluster analysis of financial time series is the choice of a relevant metric. In this paper we cluster the world indexes using the Ward method. We use the measure of distance based on the correlation of t -Student Copula with the marginal GARCH(1,1) model. Data used in this study are daily stock markets returns for 42 major international stock markets. We created five groups where the dependence of the market indexes were strongly dependent. We considered two cases: the case when the time differences were included and the case when the time difference was not included. The obtained results were similar.