PRACE NAUKOWE UNIWERSYTETU EKONOMICZNEGO WE WROCŁAWIU nr 107
Taksonomia 17 2010

Anna Czapkiewicz
Akademia Górniczo-Hutnicza w Krakowie

Beata Basiura
Wyższa Szkoła Zarządzania i Bankowości w Krakowie

GRUPOWANIE INDEKSÓW ŚWIATOWYCH NA PODSTAWIE MODELI COPULA-GARCH

Streszczenie: W pracy zaprezentowana została próba pogrupowania danych, którymi są dzienne stopy zwrotu 42 indeksów światowych. Celem badania jest wyodrębnienie podgrup, w obrębie których istnieje silne powiązanie między indeksami. Problem grubych ogonów rozkładów dziennych stóp zwrotu częściowo udało się ominąć, stosując model GARCH(1,1) z warunkowym rozkładem *t*-Studenta oraz GED dla modelowania zmian tych indeksów. Jako miarę powiązań między poszczególnymi indeksami przyjęto współczynnik korelacji, który jest parametrem funkcji połączeń i *t*-Studenta. Na podstawie tego współczynnika zdefiniowano miarę odległości, pozwalającą utworzyć podział na grupy taksonomiczne.

1. Wstęp

Celem badań jest pogrupowanie głównych indeksów światowych w grupy taksonomiczne. Trudnością jest wybór miary, która determinuje siłę związku między indeksami. Trudność ta wynika z charakterystyki rozkładów dziennych stóp zwrotu, dla których istnieją tzw. grube ogony. Ponadto rozkłady te cechuje duża kurtoza i silna asymetria. Dla modelowania wielowymiarowych rozkładów stóp zwrotu Embreecht [Embreecht, McNeil, Straumann 2001] zaproponował zastosowanie funkcji połączeń (*copula function*). Podejście to umożliwia rozważanie osobno rozkładów brzegowych i łącznego wielowymiarowego ciągłego rozkładu. Miary zależności są reprezentowane przez funkcje połączeń. Przy zastosowaniu tego podejścia badanie przebiega w dwóch etapach. Pierwszy etap to modelowanie rozkładów brzegowych, drugi zaś to modelowanie związków pomiędzy rozkładami brzegowymi.

W prezentowanej pracy do modelowania rozkładów brzegowych zastosowano model GARCH(1,1) [Bollerslev 1986] z warunkowym rozkładem normalnym, *t*-Studenta oraz GED. Jako funkcję połączeń przyjęto funkcję połączeń Gaussa oraz *t*-Studenta. Zastosowanie takich modeli do badania siły powiązań między stopami zwrotu było dyskutowane w literaturze. Modele *copuli* z rozkładami brzegowymi GARCH(1,1) były dyskutowane m.in. w pracy [Yan, Luger 2009], w której uwagę koncentruje się na sposobie estymacji nieznanych parametrów. Praca [Roch, Alegre 2006] ma charakter przeglądowy i dotyczy testowania funkcji połączeń dwuwymiarowych rozkładów przy rozkładach brzegowych typu ARMA(1,1)-GARCH(1,1). Przegląd modeli GARCH(1,1) z różnymi warunkowymi rozkładami modelującymi kurtozę i asymetrię, a następnie powiązanie ich funkcją połączeń można znaleźć w pracy [Jondeau, Rockinger 2006].

Postawiono hipotezę, że w wyniku zastosowania miary budowanej na podstawie współczynnika korelacji uzyskanego z modelu Copula-Garch i zastosowania metody aglomeracyjnej Warda otrzymamy grupy rynków mocno ze sobą powiązanych. Celem prezentowanego artykułu jest pogrupowanie głównych indeksów światowych reprezentujących zarówno rynki rozwinięte, jak i rynki rozwijające się. Próbowano również znaleźć odpowiedź na pytanie, z którymi rynkami powiązany jest rynek polski.

W badaniu przedyskutowano 42 indeksy światowe z krajów Europy, Azji, Ameryki i Australii. Za zmienną losową charakteryzującą zachowanie rynku przyjęto logarytmiczną dzienną stopę zwrotu. Za miarę podobieństwa pomiędzy dwiema zmiennymi losowymi przyjęto współczynnik korelacji funkcji połączeń Gaussa oraz *t*-Studenta przy jednoczesnym modelu GARCH(1,1) dla rozkładów brzegowych. Te dwie funkcje połączeń różnią się w modelowaniu asymptotycznej zależności. Kopula Gaussa nie wykazuje takiej zależności, natomiast kopula *t*-Studenta modeluje tę zależność, jednakże tylko symetryczną. Kopula *t*-Studenta rekomendowana jest przez autorów, takich jak Mashal, Zeevi [2002] oraz Breymann, Dias, Embbrechts [2003].

Do wyodrębnienia podgrup zastosowano metodę Warda z macierzą odległości zdefiniowaną na podstawie współczynnika korelacji.

2. Modele funkcji połączeń

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

(1) dla wszystkich $u_i \in [0,1]$, $C(1,...,1, u_i,1,...,1) = u_i$,

(2) dla każdego $u \in [0,1]^d, C(u_1,...,u_d) = 0$, jeśli co najmniej jedna współrzędna $u_i = 0$,

(3) C jest funkcją d-rosnącą.

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

$$F(x_1,...,x_d) = P(X_1 \le x_1,...,X_d \le x_d),$$

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

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

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

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

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. Ważną własnością tych funkcji jest to, że są niezmiennicze przez obłożenie funkcją rosnącą. Niech $X = (X_1,...,X_d)$ będzie zmienną losową typu ciągłego z funkcją połączeń *C*. Jeśli $g_1,...,g_d : R \to R$ są ściśle monotoniczne odpowiednio na $X_1,...,X_d$, to zmienna losowa $\left(g_1(X_1),...,g_d(X_d)\right)$ ma też tę samą funkcję połączeń *C*. Z własności tej wynika fakt, że zależności strukturalne pomiędzy zmiennymi mogą być wyjaśniane przez funkcję połączeń niezależnie od rozkładów brzegowych.

Podstawową klasę funkcji połączeń stanowią tzw. *copule* eliptyczne, do których należy funkcja połączeń Gaussa i *t-copula*. Postać analityczna tych funkcji połączeń wynika bezpośrednio z twierdzenia Sklara. Funkcja połączeń Gaussa jest postaci:

$$C(u,v) = \Phi_\rho(\Phi^{-1}(u)\Phi^{-1}(v)),$$

gdzie Φ jest dystrybuantą standaryzowanego rozkładu normalnego, natomiast Φ_ρ jest dystrybuantą dwuwymiarowego rozkładu normalnego z współczynnikiem korelacji ρ.

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.

3. Rozkłady brzegowe

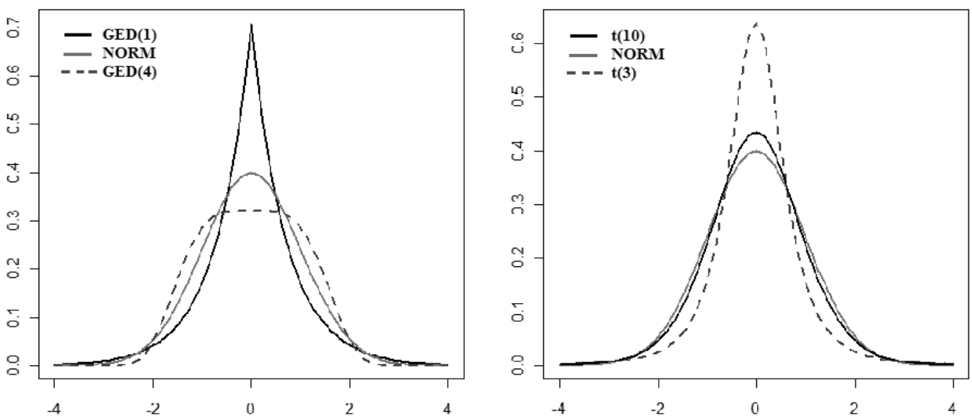
W punkcie tym przedstawiono model, który został przyjęty do opisu rozkładów brzegowych, i sposoby jego estymacji. Niech y_t, $t = 1,...,T$ będzie stopą zwrotu danego indeksu. Podobnie jak w wielu pracach poświęconych tej tematyce założono, że spełnia ona model GARCH(1,1). W klasycznym modelu GARCH(1,1) zakłada się, że warunkowy rozkład ε_i jest normalny. Okazuje się jednak, że rzeczywiste reszty modelu mają warunkowy rozkład o grubszych ogonach niż normalny. W literaturze przedmiotu na ten temat zasugerowano wiele innowacji dotyczących założeń o kształcie rozkładów warunkowych. W prezentowanej pracy dodatkowo poddano próbie dwa rozkłady warunkowe: *t*-Studenta oraz rozkład GED. Gęstości tych funkcji wyrażają się następującymi wzorami. Dla rozkładu *t*-Studenta:

$$f_S(\varepsilon_{ti}, h_{ti}, \eta) = \frac{\Gamma\left(\frac{\eta+1}{2}\right) h_{ti}^{-1/2}}{\Gamma\left(\frac{\eta}{2}\right)\sqrt{(\eta-2)\pi}} \left(1 + \frac{\varepsilon_{ti}^2}{(\eta-2)h_{ti}}\right)^{\frac{\eta+1}{2}},$$

dla rozkładu GED:

$$f_G(\varepsilon_{ti}, h_{ti}, \eta) = \frac{\eta \exp\left\{-\frac{1}{2}\left|\frac{\varepsilon_{ti}}{\lambda h_{ti}^{1/2}}\right|^{\eta}\right\}}{\lambda \Gamma\left(\frac{1}{\eta}\right) h_{ti}^{1/2}} 2^{\frac{\eta+1}{\eta}}, \quad \text{gdzie} \quad \lambda = \left[\frac{\Gamma\left(\frac{1}{\eta}\right)}{\Gamma\left(\frac{3}{\eta}\right)} 2^{-\frac{2}{\eta}}\right]^{1/2}.$$

Rozkłady *t*-Studenta oraz GED, w zależności od przyjętej liczby stopni swobody η, mogą mieć grubsze ogony niż rozkład normalny. Przykładowe funkcje gęstości wyżej wymienionych rozkładów prezentuje rys. 1.



Rys. 1. Wykresy gęstości rozkładów normalnych, *t*-Studenta oraz GED

Źródło: opracowanie własne.

4. Estymacja parametrów funkcji połączeń

W poprzednim punkcie poddano rozważaniom rozkłady brzegowe. W tym uwaga skupiona będzie na estymacji funkcji połączeń. W czasie szukania macierzy korelacji między k szeregami danych idealnym rozwiązaniem byłaby estymacja parametru z kopuli k- wymiarowej. Jednakże procedura taka, o ile może być technicznie wykonana w przypadku, gdy wymiar kopuli jest stosunkowo mały, jest niewykonalna w przypadku bardzo dużej liczby szeregów danych. W takim przypadku najczęściej stosuje się dwuwymiarową funkcję połączeń dla kolejno branych par indeksów. Pomimo tego ułatwienia nadal problemem jest sposób estymacji nieznanych parametrów, których np. w przypadku funkcji połączeń Gaussa z rozkładami brzegowymi GARCH(1,1) z rozkładem warunkowym normalnym jest dziewięć.

Jeśli z_1, z_2 będą zmiennymi losowymi odpowiednio o dystrybuantach $F_i(z_i,\alpha_i)$ wraz z odpowiadającymi im funkcjami gęstości $f_i(z_i,\alpha_i)$, nieznane parametry rozkładów brzegowych oznaczymy jako $\theta_1 = (\alpha_1', \alpha_2')'$, to dystrybuanta dwuwymiarowego rozkładu ma postać: $F(z_1, z_2; \theta) = C(F_1(z_1,\alpha_1), F_2(z_2,\alpha_2); \theta_2)$, gdzie C jest funkcją połączeń z nieznanym parametrem θ_2. Celem staje się estymacja parametrów $\theta = (\theta_1, \theta_2)$.

Logarytm funkcji największej wiarogodności dla próby $\{z_{1t}, z_{2t}\}_{t=1}^{T}$ jest postaci:

$$l(\theta) = \sum_{t=1}^{T} \ln c\left(F_1(z_{t1};\alpha_1), F_2(z_{t2};\alpha_2);\theta_2\right) + \sum_{t=1}^{T}\sum_{i=1}^{2} \ln f_i(z_{ti};\alpha_i),$$

gdzie $c(u,v)$ jest funkcją gęstości dla funkcji połączeń $C(u,v;\theta_2)$. Przedstawiając funkcję jako dekompozycję dwóch składników, otrzymuje się:

$$l(\theta) = l_c(\theta_1, \theta_2) + l_m(\theta_1).$$

W modelu, który został zastosowany do obliczeń empirycznych, taka dekompozycja jest możliwa. W ogólnym przypadku nie zawsze istnieje możliwość przedstawienia logarytmu funkcji wiarogodności jako sumy komponentów z parametrami związanymi tylko z rozkładami brzegowymi i parametrami związanymi tylko z funkcją połączeń [Patton 2006].

Wybrana metoda IFM [Joe, Xu 1996] polega na estymacji w pierwszym kroku nieznanych parametrów $\theta_1 = (\alpha_1, \alpha_2)$ dla rozkładów brzegowych, a następnie, w etapie drugim, po uzyskaniu estymatora $\hat{\theta}_1$ z kroku pierwszego, estymowaniu parametru θ_2 jako $\hat{\theta}_2 = \arg\max l_c\left(\hat{\theta}_1, \theta_2\right)$.

Przy nałożeniu warunków regularności można wykazać [Patton 2006; Joe, Xu 1996], zbieżność i asymptotyczną normalność estymatorów uzyskanych metodą

IFM. Wadą tego podejścia jest strata efektywności estymatora, ponieważ krok pierwszy ignoruje zależność pomiędzy z_1 i z_2.

Modyfikację problemu estymacji kopul, która poprawia efektywność estymatorów, można znaleźć w pracy [Yan, Luger 2009].

5. Badanie empiryczne

Badaniem objęto notowania 42 indeksów giełd światowych od dnia 2 kwietnia 2002 r. do dnia 30 października 2008 r. Wybrano do analizy dane dotyczące wybranych państw Europy, Ameryki i Azji. W rozważaniach przyjęto logarytmiczną dzienną stopę zwrotu, definiowaną jako:

$$z_{it} = \ln \frac{P_{i,t}}{P_{i,t-1}},$$

gdzie $P_{i,t}$ jest ceną zamknięcia *i*-tego indeksu w chwili *t*, natomiast $P_{i,t-1}$ jest ceną zamknięcia dnia poprzedniego.

W pierwszym etapie badania dla utworzonych szeregów stóp zwrotu dokonano estymacji ich rozkładów, przyjmując model GARCH(1,1) z trzema wymienionymi wcześniej rozkładami warunkowymi: normalnym, *t*-Studenta i GED.

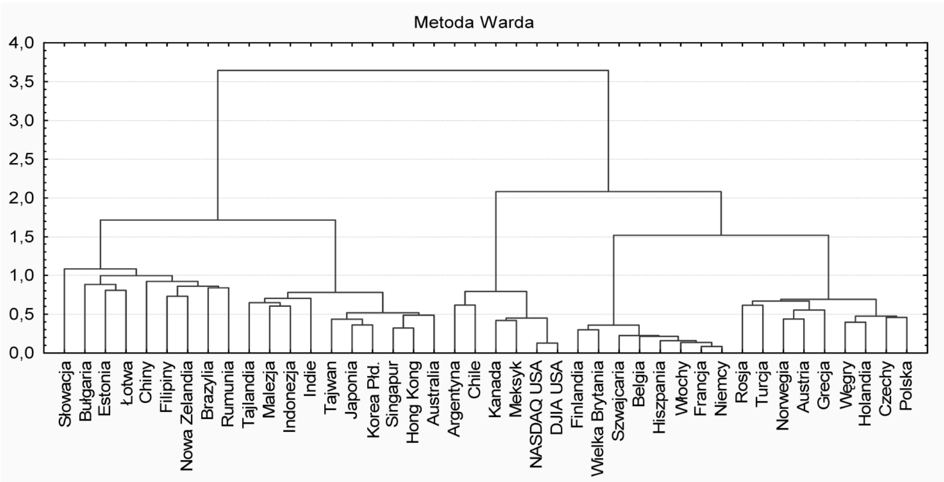
W dalszej analizie uwzględniane były te rozkłady, dla których funkcja największej wiarogodności była największa. Badając poprawność doboru modelu, zastosowano rozumowanie Diebold i in. [1998], które sugeruje badanie jednostajności u_{it} oraz ich niezależność. Jeśli ε_{it} będą standaryzowanymi resztami w modelu GARCH(1,1) oraz $F_1(x)$, $F_2(x)$, $F_3(x)$ będą odpowiednio dystrybuantami rozkładu normalnego, *t*-Studenta i GED, to właściwy wybór rozkładu warunkowego implikuje jednostajność rozkładu $u_{it} = F_k(\varepsilon_{it}),\ k = 1,2,3$. Wśród indeksów były takie, dla których najlepszy okazał się rozkład brzegowy GED (dla WIG-u otrzymano $p = 0{,}97$). Niestety znalazły się i takie indeksy, dla których żaden rozkład nie był dobry (wszystkie *p-value* mniejsze od 0,001).

Następnym krokiem jest estymacja parametru współczynnika korelacji funkcji połączeń ρ. Stosując algorytm IFM, otrzymano estymatory tego parametru pomiędzy kolejno branymi parami indeksów. Jako kryterium jakości dopasowania danych do dyskutowanych funkcji połączeń przyjęto kryterium AIC i BIC. Ponieważ wartości te sugerowały wybór kopuli *t*-Studenta, to na podstawie jej wyników przeprowadzono dalsze wnioskowanie.

Do grupowania wybrano algorytm aglomeracji metodą Warda [Grabiński, Wydymus, Zeliaś 1989]. Badania efektywności procedur taksonomicznych dowodzą wysokiej efektywności wykrywania struktury obiektów przez metodę Warda. Jako miarę odległości zdecydowano się przyjąć odległość opartą na współczynniku korelacji, daną wzorem:

$$d_{ij} = 1 - \rho_{ij} .$$

Przebieg aglomeracji otrzymany dla kopuli *t*-Studenta przedstawiono na rys. 2. Tabela 1 zawiera wyniki grupowania.



Rys. 2. Otrzymane grupy w wyniku zastosowania miary odległości na podstawie współczynnika korelacji z modelu Copula-GARCH

Źródło: opracowanie własne.

Tabela 1. Otrzymane grupy w wyniku zastosowania miary odległości na podstawie współczynnika korelacji z modelu Copula-GARCH

Grupy	Kraje
1	Polska, Czechy, Holandia, Węgry, Grecja, Austria, Norwegia, Turcja, Rosja
2	Niemcy, Francja, Włochy, Hiszpania, Belgia, Szwajcaria, Wielka Brytania, Finlandia
3	USA, Meksyk, Kanada, Chile, Argentyna
4	Australia, Hongkong, Singapur, Korea Płd., Japonia, Tajwan, Indie, Indonezja, Malezja, Tajlandia
5	Rumunia, Brazylia, Nowa Zelandia, Filipiny, Łotwa, Chiny, Estonia, Bułgaria, Słowacja

Źródło: opracowanie własne.

Analizując wyniki, można zauważyć, że druga grupa skupia indeksy najmocniej ze sobą skorelowane, natomiast piąta to kraje najsłabiej skorelowane z pozostałymi rynkami. Przeprowadzone badania nie uwzględniały asymptotycznej korelacji, która będzie tematem następnych rozważań.

6. Podsumowanie

W pracy zaprezentowana została metoda grupowania spółek na podstawie miary, jaką jest współczynnik korelacji funkcji połączeń Gaussa i *t*-Studenta. Struktury zmian stóp zwrotu są modelowane modelem GARCH(1,1). Jako warunkowy rozkład dla tego modelu przyjęto rozkład normalny, *t*-Studenta oraz GED. Estymacja współczynników funkcji połączeń jest bardzo wrażliwa na właściwy wybór rozkładów brzegowych. Ponieważ istnieją takie indeksy, które nie dały się modelować żadnym z zaproponowanych rozkładów warunkowych, należy poszukiwać innych rozkładów warunkowych, uwzględniających kurtozę i asymetrię.

Cel pracy został częściowo osiągnięty. Prezentowana praca jest wstępem do bardziej wnikliwych badań, które będą obejmować zarówno modelowanie rozkładów brzegowych jak i modelowanie asymptotycznej zależności. Dalsze badania będzie obejmować również uwzględnienie dynamicznego współczynnika korelacji w funkcji połączeń.

Literatura

Bollerslev T., *Generalized autoregressive conditional heteroskedasticity*, „Journal of Econometrics” 1986 nr 31, s. 307-327.

Breymann W., Dias A., Embbrechts P., *Dependence structures for multivariate high-frequency data in finance*, „Quantitative Finance” 3(1) 2003, s. 1-16.

Davidson R., MacKinnon H., *Estimation and Inference in Econometrics*, Oxford University Press, Oxford 1993.

Diebold F.X., Gunther T.A., Tay A.S., *Evaluating density forecasts with application to financial risk management*, „International Economic Review” 39(4) 1989, s. 863-883.

Embreecht P., McNeil A.J., Straumann D., *Correlation and Dependency in Risk Managment: Properties and Pitfalls*, [w:] M. Dempster, H. Moffant, *Risk Management*, Cambridge University Press, New York 2001, s. 176-223.

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, Departaments 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 nr 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.

Patton A.J., *Estimation of multivariate models for time series of possibly different lenghts,* „Journal of Applied Econometrics” 2006 nr 21, s. 147-173.

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 nr 51, s. 1312-1329.

Shih J.H., Louis T.A., *Inferences on the Associations Parameter in Copula Models for Bivariate Survival Data*, „Biometrics” 1995 nr 51, s. 1384-1399.

Sklar A., *Fonction de Repartition a n Dimention et Leur Marges*, Publication's de L'Institut de Statistiques de L'Unieversite de Paris, Paris 1959, s. 229-231.
Song K., *Multivariate dispersion models generated from Gaussian Copula*, Skandinavian „Journal of Statistics" 2000 nr 27, s. 305-320.
Yan L., Luger R., *Efficient estimation of copula-GARCH models 1*, „Computational Statistics & Data Analysis" 2009 nr 53, s. 2284-2297.
Zeliaś A. (red.), *Taksonomiczna analiza przestrzennego zróżnicowania poziomu życia w Polsce w ujęciu dynamicznym*, AE, Kraków 2000.

WORLD INDEXES CLUSTERING USING THE COPULA – GARCH MODELS

Summary: The identification of similarities or dissimilarities in financial time series has become an important research area in finance and empirical economics. In stock markets, the examination of mean and variance correlations between asset returns can be useful for portfolio diversification and risk management purposes. A fundamental problem in cluster analysis of financial time series is the choice of a relevant metric. In order to capture the correlation structure and the spectral behaviour of the time series, we may use measures of distance based on the correlation of t-Student Copula with the marginal GARCH(1,1) model.

In this paper, we clustered the word indexes using the Ward method. Data used in this study are daily stock markets returns based on daily data for 42 of the major international stock markets. We created five groups where the dependence of the market indexes was strong.