

Zastosowanie analizy skupień do określania ram czasowych ostatniego kryzysu finansowego

Eliza Buszkowska*

Streszczenie: W artykule zastosowano metody analizy skupień z zakresu statystyki wielowymiarowej do badania okresu kryzysu finansowego lat 2007–2013. Zgodnie z tymi algorytmami grupowano spready pomiędzy stopami rynku międzybankowego a OIS. Przeprowadzono dyskusję nad możliwością zastosowania tej metody do generowania ram czasowych różnych okresów kryzysów finansowych. Analiza skupień umożliwiła poczynienie nowych obserwacji na temat kryzy subprime i występującego po nim kryzysu zadłużeniowego. Praca nawiązuje do wcześniejszej publikacji autorki. Uwzględnia inne dane, inny okres kryzysu i zmodyfikowaną metodologię – zamiast podziału próby na lata, podział na okresy równej długości. Dzięki temu nie skraca się znacząco szeregów i nie traci się dużej części informacji.

Słowa kluczowe: analiza skupień, kryzys finansowy, korelacje, zmienność, kryzys zadłużeniowy

Wprowadzenie

W niniejszej pracy zastosowano jedną z metod grupowania – metodę analizy skupień, z zakresu statystyki wielowymiarowej – do analizy okresu ostatniego kryzysu finansowego. Dokonano tego poprzez grupowanie spreadów pomiędzy stopami rynku międzybankowego a OIS. Użyta metoda statystyczna znalazła zastosowanie na przykład w medycynie do grupowania chorób, szukania metod leczenia lub symptomów paranoi i schizofrenii. W marketingu ma zastosowanie do segmentacji rynku. W archeologii techniki analizy skupień wykorzystuje się do grupowania narzędzi kamiennych i innych znalezisk. W finansach stosuje się je między innymi w analizie portfelowej do badania spółek giełdowych (Pietrzykowski, Kobus 2006: 301). Techniki analizy wielowymiarowej są narzędziem do klasyfikacji zbiorów wielu zmiennych, którymi będą w tym badaniu spready pomiędzy stopami międzybankowymi a OIS. Autorka sprawdziła, czy przy użyciu metody analizy skupień można wyróżnić okresy kryzysu, tzn. czy spready w okresie kryzysu oraz sąsiednich tworzyły osobne charakterystyczne skupienia. Zweryfikowała, czy analiza skupień wnosi nowe informacje do naszej wiedzy na temat kryzysu finansowego subprime i kryzysu zadłużeniowego.

* dr Eliza Buszkowska, Wydział Prawa i Administracji UAM, e-mail: eliza_b2@o2.pl.

Autorka stawia hipotezy, że analiza skupień jest techniką dobrą do określania lat kryzysów finansowych. Pozwala ująć na jednym wykresie istotne informacje o okresach podwyższonej zmienności i korelacji, co byłoby trudne do zaprezentowania na klasycznych wykresach. Dzięki niej można zweryfikować, czy okres kryzysu różnił się od pozostałych ze względu na rozpatrywane zmienne. Analiza skupień jest konkurencyjnym sposobem analizowania cech kryzysów finansowych w oparciu o dane finansowe. Inaczej można tego dokonać poprzez klasyczne porównania indywidualne danych. Co więcej, autorka potwierdzi nowym sposobem, że zgodnie z literaturą przedmiotu kryzys finansowy charakteryzował się wzrostem zmienności oraz wzrostem korelacji pomiędzy spreadami strefy euro i USD.

Zdaniem Płuciennika „Światowy kryzys gospodarczy natychmiast przeniósł się ze Stanów Zjednoczonych na inne rynki międzybankowe. Najbardziej znaczące skutki kryzysu wystąpiły w Kanadzie, Japonii i Australii, ponieważ banki pochodzące z tych krajów były w największym stopniu zaangażowane na rynku amerykańskich papierów *subprime*. Pogorszenie kondycji sektora międzybankowego było także obserwowane w strefie euro. W Polsce w początkowej fazie skutki kryzysu odczuły jedynie rynki kapitałowe. W drugiej połowie 2009 roku dynamika produktu krajowego zaczęła się stabilizować, a Polska była jedynym krajem regionu, który zakończył najgorszy, 2009 rok z dodatnim PKB” (Płuciennik 2015). W tym badaniu autorka sprawdzi między innymi, czy korelacje wyznaczone metodą analizy skupień potwierdzają wyniki Płuciennika i świadczą o odmiennej specyfice polskiego rynku międzybankowego w okresie kryzysu, w stosunku do dynamiki spreadów strefy euro i dolara amerykańskiego.

1. Metodologia

Przeprowadzono badanie, które umożliwia grupowanie obiektów przy użyciu analizy skupień (Stanisz 2007). Analiza skupień jest to zbiór metod służących do wyodrębniania jednorodnych podzbiorów obiektów populacji. Jest działem statystyki wielowymiarowej. W oparciu o pewne zmienne wyznacza grupy obiektów bardziej podobnych do obiektów należących do tego samego skupienia aniżeli do obiektów z innych skupień.

Istnieje wiele procedur grupowania obiektów za pomocą analizy skupień. Najczęściej wykorzystuje się metody hierarchiczne, których wyniki prezentowane są na tzw. dendrogramach. Innym często stosowanym algorytmem jest metoda *k*-średnich, w której obiekty przydzielane są do *k* skupień, każdy tylko do jednego skupienia. Liczba skupień *k* jest ustalana przez badacza.

Metody hierarchiczne dzielą się na dwie grupy technik – techniki aglomeracyjne i techniki podziałowe. W technikach aglomeracyjnych początkowo każdy obiekt tworzy odrębne skupienie. Następnie sukcesywnie łączy się najbliższe sobie obiekty w nowe skupienia, aż do uzyskania jednego skupienia. Techniki podziałowe to algorytmy odwrotne. W praktyce częściej stosowane są techniki aglomeracyjne, również w tym badaniu. Istnieje problem, jak określić odległość między nowymi skupieniami, innymi słowy, kiedy dwa skupienia są

dostatecznie podobne, żeby je połączyć. Można wymienić wiele różnych zasad wiązania. W tym badaniu zastosowano metodę pojedynczego wiązania, zwaną również metodą najbliższego sąsiedztwa. Odległość między dwoma skupieniami jest w tej metodzie określona przez odległość między dwoma najbliższymi sąsiadami należącymi do różnych skupień. Jest to metoda najbardziej zgodna z intuicją. Inne istniejące metody to metoda pełnego wiązania, metoda średnich połączeń, metoda średnich połączeń ważonych, metoda środków ciężkości, metoda ważonych środków ciężkości i metoda Warda opisane w podręczniku do nauki programu STATISTICA (Stanisz 2007).

Metoda pełnego wiązania (najdalszego sąsiedztwa) to metoda, w której odległość między klastrami to odległość największa. Metoda nie została zastosowana w tej pracy, ponieważ nie jest to konieczne, a wydaje się być mniej zgodne z intuicją. Zamiast poprzedniej metody można było wykorzystać trudniejszą, lecz bardziej dokładną, a zarazem uniwersalną metodę średnich połączeń. Kolejny sposób – metodę średnich połączeń ważonych powinno się stosować, gdy skupienia są zdecydowanie różnoliczne. W tym badaniu w większości było inaczej. Metoda środków ciężkości polega na wyznaczeniu średniego punktu w przestrzeni wielowymiarowej, zdefiniowanej przez wymiary obserwacji. Byłaby również odpowiednia do tego badania, ponieważ nie było obserwacji odstających. Metoda ważonych środków ciężkości (mediana) wprowadza ważenie, żeby uwzględnić różne wielkości skupień. Zatem w tym badaniu nie ma zastosowania, gdyż można otrzymać wnioski podobnej jakości w prostszy sposób. Można też było oszacować odległość między skupieniami wykorzystując analizę wariancji. Metoda ta w literaturze jest opisywana, jako „bardzo efektywna, chociaż zmierza do tworzenia skupień o małej wielkości”.

Odległość między obiektami wyznaczono za pomocą dwóch metryk:

- klasycznej metryki euklidesowej zdefiniowanej następująco:

$$d(x, y) = \sqrt{\sum_{i=1}^p (x_i - y_i)^2} \quad (1)$$

gdzie:

$$x = (x_1, \dots, x_p), y = (y_1, \dots, y_p),$$

- oraz metryki uwzględniającej współczynnik korelacji Pearsona w postaci następującego wyrażenia:

$$d(x, y) = 1 - r \quad (2)$$

gdzie r jest współczynnikiem korelacji Pearsona (Stanisz 2007).

2. Dane

W badaniu wykorzystano dzienne obserwacje notowań wybranych stóp rynku międzybankowego EURIBOR 3M, WIBOR 3M oraz LIBORUSD 3M, oraz trzymiesięczne spready

OIS dla strefy euro, Polski i USA. Okres badawczy to przedział: 7.07.2004–25.01.2013. Wiadomo z literatury, że okres ten obejmuje ostatni kryzys finansowy 8.08.2007–10.03.2009 (Thornton 2009) oraz okres kryzysu zadłużeniowego do daty 21.01.2009. Szeregi czasowe zostały podzielone na krótkie, 24-elementowe odcinki. Ponieważ analiza skupień ma zastosowanie tylko do równolicznych zbiorów danych, szeregi czasowe musiały być podzielone na interwały jednakowej długości. Krótkie okresy podziału były szansą na otrzymanie bardziej dokładnych ram czasowych kryzysu. Podczas uzgadniania dat dla różnych walut autorka skróciła fragmenty 30-elementowe o 7 ostatnich obserwacji. W innej publikacji (Buszkowska 2015) autorka podzieliła szeregi danych na lata. Żeby ciągi danych miały równą długość, musiała w poprzedniej pracy usunąć znaczną część obserwacji. Straciła w ten sposób dużą część informacji, gdyż w poszczególnych latach liczby notowań znacznie się różniły. Ponieważ musiała przyjąć konsekwentny algorytm redukcji danych, zdecydowała się wtedy na pobieranie początkowych dłuższych fragmentów szeregów czasowych (Buszkowska 2015), przy założeniu, że w kryzysie korelacje istniały w całym okresie. Niestety kryzys subprime rozpoczął się w drugiej połowie 2007 roku i taka technika zmniejszała otrzymane ramy czasowe badanego okresu.

3. Wyniki empiryczne

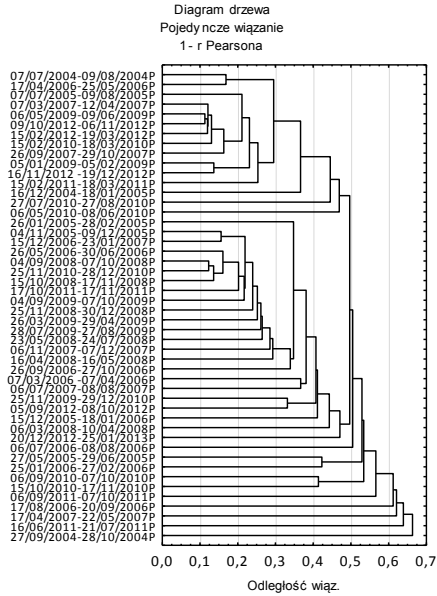
Na pierwszym etapie badania sprawdzono, czy kryzys finansowy wiązał się ze wzrostem korelacji pomiędzy spreadami. Wyniki w postaci dendrogramów metody hierarchicznej są zaprezentowane na rysunkach 1–4.

Zauważamy, że niektóre polskie spready z okresu kryzysu finansowego, lat 2007–2013 utworzyły wyróżniające się skupienie, charakteryzujące się podwyższoną korelacją (rys. 1).

Dla spreadów strefy euro (rys. 2) było podobnie. Niektóre różnice EURIBOR 3M – OIS 3M z okresu ostatniego kryzysu finansowego, lat 2007–2013 były zdecydowanie bardziej skorelowane ze sobą niż pozostałe obserwacje.

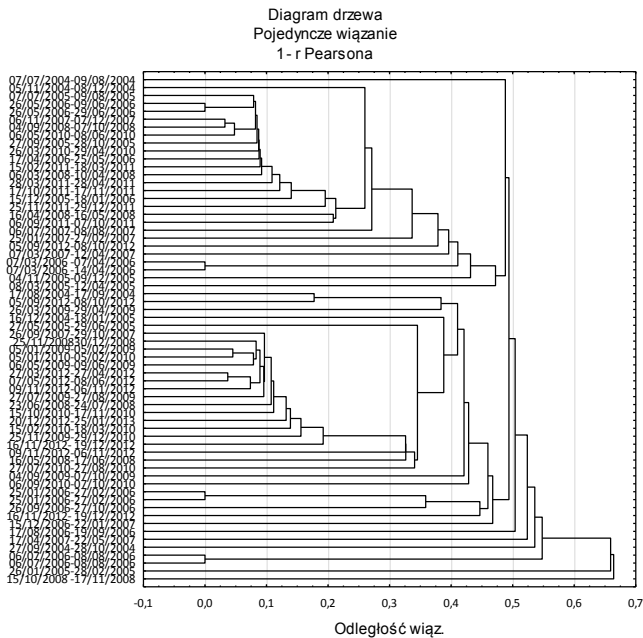
Można potwierdzić (rys. 3), że w okresie kryzysu lat 2007–2013 spready miały większą zmienność.

Potwierdzamy, że istniały znaczne powiązania pomiędzy kryzysem strefy euro i Stanów Zjednoczonych w latach 2007–2013 (rys. 4). Polski kryzys miał natomiast inną dynamikę (Płuciennik 2015).



Rysunek 1. Odległości korelacyjne dla polskich spreadów

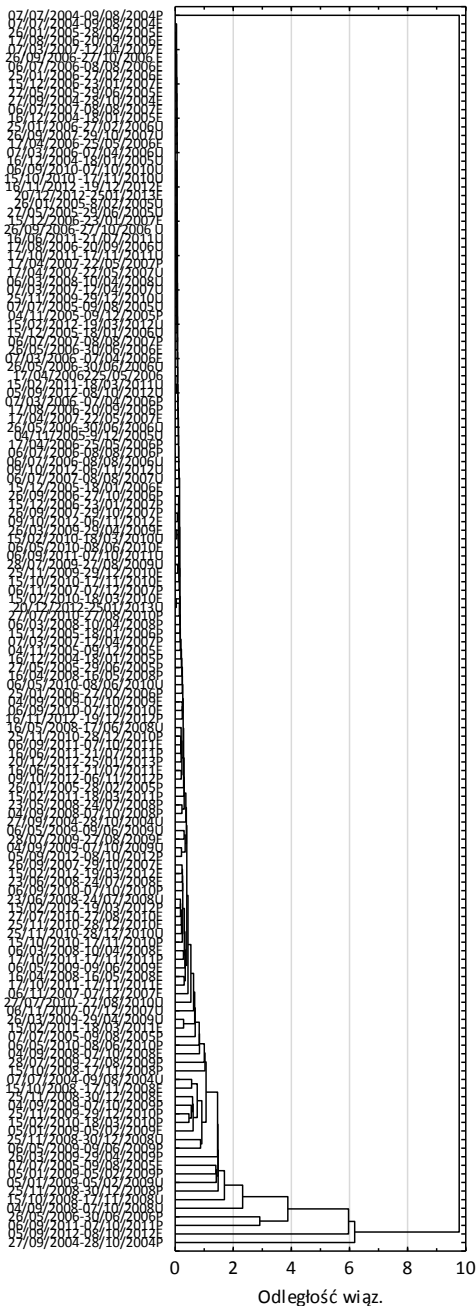
Źródło: opracowanie własne.



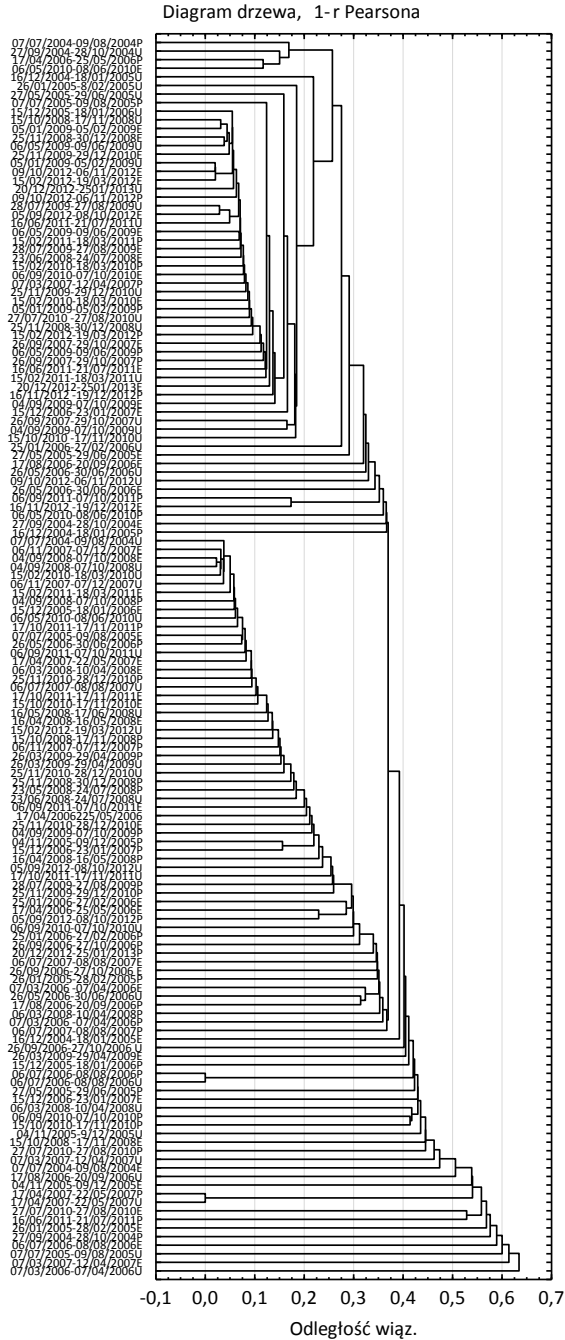
Rysunek 2. Odległości korelacyjne spreadów dla strefy euro

Źródło: opracowanie własne.

Diagram drzewa. Pojedyncze wiązanie Odł. euklidesowa



Rysunek 3. Zmienność spreadów trzech obszarów: Polski, Europy i Stanów Zjednoczonych
 Źródło: opracowanie własne.



Rysunek 4. Korelacje Pearsona dla spreadów Polski, Europy i Stanów Zjednoczonych

Źródło: opracowanie własne

Uwagi końcowe

Analiza skupień może być wykorzystywana do weryfikowania własności charakterystycznych okresów na rynku finansowym, nie zaś do określania ich ram czasowych, gdy za jednostkę podziału próby przyjmie się przyległe odcinki równej długości. Potwierdzono nową metodą, że w okresie kryzysu istniały większe powiązania pomiędzy strefą euro a strefą dolara amerykańskiego. Należy dodać, że otrzymane korelacje świadczyły o współwystępowaniu obserwacji (podobnych tendencjach na wykresie), a nie o zależnościach przyczynowo-skutkowych, gdyż nie operowano szeregami stacjonarnymi. Jednak dla badacza, który określa ramy czasowe na rynku finansowym może być bardziej interesująca sama relacja wykresów spreadów, a nie wzajemne wpływy jednych spreadów na drugie, rozumiane jakościowo, dlatego w tym artykule autorka rezygnuje z analizy stacjonarności. Najważniejsza hipoteza o możliwości zastosowania analizy skupień do generowania ram czasowych charakterystycznych okresów na rynku finansowym, przy podziale próby na równe, prawie przyległe interwały, bez znacznej utraty informacji, została odrzucona. Tak samo byłoby, gdyby dokonano w inny sposób podziału na okresy – odcinki roczne (Buszkowska 2015), co powoduje większą utratę informacji. Trudno jest, stosując analizę skupień, precyzyjnie określić początek kryzysu, gdyż nie wiadomo, jak duże muszą być odległości między klastrami na dendrogramie, żeby można już było wnioskować o kryzysie. Innymi słowy, otrzymane wykresy pokazują, że korelacje na dendrogramie nie zwiększyły się w okresie kryzysu gwałtownie.

Literatura

- Blackburn R. (2008), *The Subprime Crisis*, „New Left Review” vol. 50, March–April.
- Buszkowska E. (2010), *Badanie zależności między indeksami giełdowymi a kursami walutowymi*, Zeszyty Naukowe Uniwersytetu Ekonomicznego w Krakowie, Kraków.
- Buszkowska E. (2015), *Zastosowanie statystyki wielowymiarowej do badania kryzysu subprime*, „Optimum. Studia Ekonomiczne” nr 4 (76), s. 85–102.
- Buszkowska E., Pluciennik P. (2013), *Wpływ kryzysu subprime na polski rynek kapitałowy*, „Ekonomiczne etyczne aspekty kryzysu gospodarczego” nr 1.
- Doman M., Doman R. (2009), *Modelowanie zmienności i ryzyka. Metody ekonometrii finansowej*, Oficyna a Wolter Kluwer Business, Kraków.
- Doman M., Doman R. (2014), *Dynamika zależności na globalnym rynku finansowym*, Difin, Warszawa.
- Foster J.B., Magdoff F. (2009), *The great financial crisis, causes and consequences*, Monthly Review Press.
- Konopczak M., Sieradzki R., Wiernicki M. (2010), *Kryzys na światowych rynkach finansowych – wpływ na rynek finansowy w Polsce oraz implikacje dla sektora realnego*, „Bank i Kredyt”, nr 6.
- Migut G. (2009), *Zastosowanie technik analizy skupień i drzew decyzyjnych do segmentacji rynku*, http://www.statsoft.pl/portals/0/Downloads/Zastosowanie_teknik.pdf.
- Pietrzykowski R., Kobus P. *Zastosowanie modyfikacji metody k-średnich w analizie portfelowej*, http://www.wne.sggw.pl/czasopisma/pdf/EIOGZ_2006_nr60_s301.pdf.
- Pluciennik P. (2012), *Influence of the American Financial Market on Other Markets During the Subprime Crisis*, „Folia Oeconomica Stetinensia”.
- Pluciennik P. (2015), *Wpływ światowego kryzysu gospodarczego na sytuację w polskim sektorze bankowym*, Zeszyty Naukowe Uniwersytetu Szczecińskiego nr 862, „Finanse, Rynki Finansowe, Ubezpieczenia” nr 75.

- Stanisz A. (2007), *Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny. Tom 3. Analizy wielowymiarowe*, Statsoft, Kraków.
- Thornton D.L. (2009), *What the Libor-OIS Spread Says*, „Economic Synopses” vol. 24, Federal Reserve Bank of St. Louis, <https://research.stlouisfed.org/publications/es/09/ES0924.pdf> (11.05.2009).

THE USE OF CLUSTER ANALYSIS TO IDENTIFY THE TIME FRAMEWORK OF THE LAST FINANCIAL CRISIS

Abstract: This paper uses cluster analysis in the field of multivariate statistics to test the period of the financial crisis of years 2007–2013. With these algorithms were grouped spreads between interbank market rate and the OIS. The author debates on the possibility of using this method to generate the timeframe of different periods on the interbank market. A real cluster analysis enabled the new observations about the subprime crisis and the occurring after debt crisis. The work refers to the previous publication of the author. It takes into account another data, another period of crisis and modified methodology – instead of attempting division for years, the time periods are of equal length. This algorithm does not shorten the series with losing part of information.

Keywords: cluster analysis, financial crisis, correlations, volatility, debt crisis

Cytowanie

- Buszkowska E. (2016). Zastosowanie analizy skupień do określania ram czasowych ostatniego kryzysu finansowego. *Finanse, Rynki Finansowe, Ubezpieczenia*, 1 (79), 385–393; www.wneiz.pl/frfu.

