

**Jerzy Korzeniewski**

Uniwersytet Łódzki  
e-mail: jurkor@wp.pl

---

## ZASTOSOWANIE ANALIZY SKUPIEŃ DO KONSTRUOWANIA PORTFELI AKCJI NA WGPW

---

## APPLICATION OF CLUSTER ANALYSIS TO THE CONSTRUCTION OF INVESTMENT PORTFOLIO ON THE WARSAW STOCK EXCHANGE

---

DOI: 10.15611/pn.2017.468.11

JEL Classification: G20, G29, G35

**Streszczenie:** Metody analizy skupień zastosowane do konstruowania portfeli papierów wartościowych mogą być konkurencyjne dla innych, bardziej tradycyjnych, metod badania ryzyka inwestycyjnego. Takie wnioski można wyciągnąć z badań amerykańskiego rynku kapitałowego z początku XXI wieku (por. [Marvin 2015; Craighead, Klemesrud 2002]). W artykule są przedstawione badania możliwości zastosowania metod analizy skupień na warszawskim rynku GPW. Badanie ma na celu zbadanie racjonalności stosowania tej grupy metod pod kątem możliwości wyboru optymalnych metod grupowania spółek niezależnie od koniunktury giełdowej, optymalnego typu danych opisujących notowania spółek, sensu przenoszenia wzorców ustalonych na rynku amerykańskim na rynek warszawski. Badanie zostało przeprowadzone na notowaniach z pięciu ostatnich lat, w okresach reprezentujących różne poziomy koniunktury giełdowe. Zbadano kilka metod grupowania danych od metod partycjonujących ( $k$ -średnich oraz PAM) do metod aglomeracyjnych.

**Słowa kluczowe:** portfel inwestycyjny, metoda  $k$ -średnich, metoda PAM, metoda aglomeracyjna, indeks Calińskiego-Harabasz, wskaźnik Sharpa.

**Summary:** Cluster analysis used for portfolio construction can be competitive for other methods of investment risk analysis. Such conclusions follow from the American stock market research from the beginning of this century. In this article an investigation of the possibility of applying cluster analysis on the Warsaw Stock Exchange is presented. The investigation is aimed at finding answers to the following questions. Are the patterns from the American market also present on the Warsaw market? In particular, is it better to analyze stock price series or return rates series? Is it reasonable to exclude one-element clusters from further analysis i.e. portfolio? Is it possible to select clustering methods that would be optimal with respect to the quality of portfolios constructed independently of the stock exchange situation? In particular, is the PAM method also recommendable? The investigation was carried out on the data from the recent five years in times of different economic situations.

**Keywords:** investment portfolio,  $k$ -means method, PAM method, agglomerative method, Caliński-Harabasz index, Sharp index.

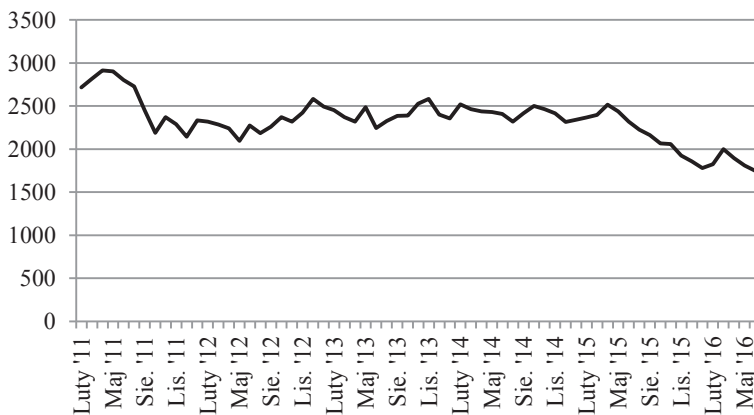
## 1. Wstęp

Pomysłów na wykorzystanie analizy skupień do konstruowania portfela inwestycyjnego można, w literaturze przedmiotu znaleźć kilka. Ich cechą wspólną jest jednak to, że analizie skupień z reguły towarzyszą metody statystyczne lub finansowe, które pozwalają na wybranie reprezentantów otrzymanych skupień podobnych do siebie spółek giełdowych, tak, by było możliwe skonstruowanie niezbyt obszernego portfela inwestycyjnego. Craighead i Klemesrud [2002] badali szeregi danych tygodniowych dotyczących 138 spółek z giełdy nowojorskiej w okresie od początku 1998 roku do końca 2002 roku. Oprócz analizy skupień stosowali jeszcze metodę filtrowania Kalmana w celu usunięcia niektórych spółek. Usuwali również spółki przypisane do skupień jednoelementowych. Ta metoda dała dobre efekty, ponieważ spółka, która w ciągu następnych kilku miesięcy zbankrutowała, została tą metodą właśnie usunięta. W celu wyróżnienia portfela składającego się z co najwyżej kilkunastu spółek posłużyli się „doświadczeniem inwestorskim”. To podejście wydaje się mało statystyczne. Ogólne wnioski były jednak pozytywne dla przydatności analizy skupień. Marvin [2015] badała również spółki z giełdy nowojorskiej w zupełnie inny sposób. Analizę skupień zastosowała do zmiennych o charakterze finansowym, tzn. spółki były scharakteryzowane przez wskaźniki finansowe, takie jak przychód podzielony przez wielkość aktywów oraz dochód netto podzielony przez wielkość aktywów. Następnie, w celu skonstruowania konkretnego niezbyt obszernego portfela, posłużyła się wskaźnikiem również czysto finansowym, tj. wskaźnikiem Sharpa. Wyniki tych badań również przemawiają na korzyść analizy skupień. Należy zauważyć, że wykorzystanie wskaźnika Sharpa wydaje się godne uwagi, gdyż wskaźnik ten w najprostszy sposób odnosi stopę zwrotu inwestycji do jej ryzyka. W kilku pracach wykorzystywano miary skorelowania szeregów czasowych do wyznaczania skupień spółek podobnych do siebie. Ren [2005] używał granicy 0,2 dla współczynnika korelacji od której grupował spółki jako „podobne”. Ta analiza nie jest dość jasna, samo grupowanie na bazie wspomnianej zasady może doprowadzić do włączenia wszystkich spółek do tego samego skupienia. Potrzebne jest narzucenie jakichś dodatkowych warunków. Rosen [2006] grupował spółki na giełdzie sztokholmskiej w oparciu wyłącznie o ich wzajemne współczynniki korelacji. Podejście oparte na analizie współczynników korelacji oprócz innych wad ma, jak zauważyła Marvin [2015], następującą wadę zasadniczą. Skorelowanie notowań spółek zmienia się zasadniczo wraz z koniunkturą giełdową. Na przykład bardzo silnie skorelowane stopy zwrotu dwóch spółek z tej samej branży, ExxonMobil i Chevron, (współczynnik korelacji liniowej dziennych stóp zwrotu równy 0,85) w okresie od 2000 do 2011 roku miały takie same znaki stóp zwrotu przez 2541 dni, zaś przeciwne znaki przez 589 dni. Sądząc po tym przykładzie, można oczekiwać, że dwie spółki skorelowane na poziomie 0,2 pozostawiają niewielką nadzieję na to, że będą charakteryzowały się takim samym znakiem stóp zwrotu w znacznie wyższej od połowy liczbie dni.

W pracy niniejszej są przedstawione badania możliwości zastosowania metod analizy skupień na warszawskim rynku GPW. W szczególności, celem było znalezienie odpowiedzi na następujące pytania. Czy wzorce ustalone na rynku amerykańskim są zachowane na rynku warszawskim? W szczególności czy lepsze do analizy są szeregi danych cen akcji czy też stóp zwrotu? Ponadto, czy ma sens pomijanie spółek, które zostały przydzielone przez metodę grupowania do skupienia jednoelementowego? Czy można wybrać metody analizy skupień, które będą zachowywały swoje optymalne własności odnoszące się do jakości konstruowanych portfeli niezależnie od koniunktury giełdowej? W szczególności, czy metoda PAM na rynku WGPW również jest godna polecenia?

## 2. Materiały źródłowe i założenia metodyczne

Na rys. 1 widoczny jest przebieg wartości indeksu WIG20 charakteryzującego parkiet warszawski. Badaniu poddane zostały dane dzienne wszystkich spółek o kapitalizacji przekraczającej 500 mln złotych (w dniu 31 lipca 2016 roku). Tych spółek notowanych w całym wspomnianym okresie było 85. Naszym zbiorem danych jest zatem zbiór 85 szeregów czasowych o długości 1360 obserwacji. Szeregi będą rozważane w dwóch wariantach: dziennych stóp zwrotu oraz znormalizowanych cen (cena minus cena minimalna danej spółki podzielć przez rozstęp cen spółki). W celu zbadania w miarę szerokiego wachlarza metod grupowania spółek wybrane zostały trzy metody. Pierwsza to klasyczna metoda  $k$ -średnich z losowym wyborem obiektów startowych. Grupowanie powtarzane było 100 razy i jako ostateczne wybierane było to, które miało najmniejszą wariancję wewnątrzklasową (por. [Gatnar, Walesiak 2004]). Ze wszystkich wariantów metody  $k$ -średnich ta polecana jest często jako najbardziej efektywna.



**Rys. 1.** Wykres wartości indeksu WIG20 w okresie od początku 2011 roku do połowy 2016 roku

Źródło: opracowanie własne, na podst. danych z [www.gpw.pl](http://www.gpw.pl) (1.09.2016).

Druga metoda grupowania to algorytm PAM (*Partitioning Around Medoids* (por. [Gatnar, Walesiak 2004]), który jest bardziej od metody  $k$ -średnich ukierunkowany na wyszukiwanie skupień o kształtach bardziej złożonych od kulistych. Trzecia metoda to grupowanie aglomeracyjne metodą całkowitego łączenia (por. [Gatnar, Walesiak 2004]). We wszystkich metodach miarą odległości obiektów jest odległość euklidesowa odniesiona do postaci danych (dzienne stopy zwrotu lub ceny) i liczby współrzędnych wyznaczonych przez okres półroczny zbierania danych.

Wykorzystywanie daleko idących i dość złożonych aspektów związanych z poszukiwaniem jak najlepszej strategii inwestowania (np. łączenia inwestycji w akcje z innymi inwestycjami, żonglowania częściowymi sprzedażami akcji) ma zbyt duży wpływ na wyniki eksperymentów, których zasadniczym celem jest ocenianie możliwości stosowania analizy skupień. Nie mniej jednak, konieczne jest przyjęcie jakiejś konkretnej formy inwestowania w celu uzyskania miary liczbowej możliwości stosowania analizy skupień. W tym kontekście ważnym problemem jest ustalenie sposobu inwestowania, tzn. określenie momentów zakupu i sprzedaży. Sposobów inwestowania może być bardzo wiele i jeśli chcemy skupić się na badaniu przydatności analizy skupień do konstruowania portfela, to nie możemy w zbyt dużym stopniu uzależniać się od przeróżnych możliwości określania tych dat. Takie analizy raczej utrudniają racjonalną ocenę możliwości wykorzystania analizy skupień do konstrukcji portfela inwestycyjnego, gdyż zbyt silnie uzależniamy tę ocenę od przyjętej strategii. Prostą metodą pozwalającą uniknąć wspomnianych trudności jest jednorazowy zakup i jednorazowa sprzedaż wszystkich kupionych walorów. Wydaje się, że logiczne byłoby ustalenie okresu inwestowania czyli przetrzymywania zakupionych walorów na czas w przybliżeniu równy okresowi, w którym dane na temat walorów były zbierane. Z kolei ten okres zbierania danych nie może być zbyt długi, gdyż w długiej perspektywie mogą zmienić się znacznie charakterystyki ekonomiczne spółek. Wobec tego przyjęto założenie, że dane potrzebne do skonstruowania portfela były zbierane z okresu około pół roku (128 dni giełdowych, okres przeszły) oraz, że zakupione walory były sprzedawane po upływie kolejnego okresu półrocznego (okres nowy).

Problemem jest również ustalenie liczby skupień, na które podzielimy wszystkie 85 spółek. Przeprowadzono w tym celu badanie wstępne przy pomocy indeksu Calińskiego-Harabasz, który cieszy się dość dobrą opinią (por. [Korzeniewski 2014]). Indeks ten wskazywał różne wartości z przedziału od 2 do nawet 19 skupień w zależności od przyjętej metody grupowania spółek oraz okresu półrocznego. Wobec tego zdecydowano się na ustalenie trzech różnych liczb skupień i co za tym idzie, spółek w portfelu, gdyż zasadniczym celem badania było ocenienie przydatności stosowania analizy skupień do konstrukcji portfela. Założono, że będziemy próbowali konstruować portfele składające się z 4, 7 oraz 12 spółek. W celu wybrania reprezentanta każdego skupienia zastosowano wskaźnik Sharpa tzn. z każdego skupienia do portfela włączamy tylko jedną spółkę, która w okresie przeszłym miała najwyższą wartość wskaźnika Sharpa.

W celu zbadania okresów różnych koniunktur giełdowych wybrano do prognozowania następujące okresy półroczne (okresy przeszłe).

1. Rynek **bessy**: maj 2011 (84 sesja), WIG20 2903 pkt, do listopada 2011 (212 sesja), WIG20 2288 pkt.

2. Rynek **hossy**: lipiec 2012 (377 sesja), WIWG20 2185 pkt, do stycznia 2013 (505 sesja), WIG20 2492 pkt.

3. Rynek **stabilny**: kwiecień 2014 (810 sesja), WIG20 2439 pkt, do października 2014 (938 sesja), WIG20 2463 pkt.

Idea odrzucania spółek ze skupień jednoelementowych wydaje się intuicyjnie uzasadniona – ryzyko inwestowania w spółkę, której notowania nie mają podobnych sobie, jest duże, ale przecież takie spółki mogą być przy sprzyjającej koniunkturze źródłem wysokich dochodów. W celu sprawdzenia jakości tego pomysłu przeprowadzono dwa warianty grupowania: z odrzucaniem skupień jednoelementowych (kolejno aż do momentu, gdy każde ze skupień będzie miało co najmniej dwa elementy) oraz z uwzględnianiem dowolnych skupień.

Po wybraniu reprezentantów każdego skupienia, inwestujemy (na koniec okresu przeszłego) taką samą kwotę pieniędzy w każdego reprezentanta i sprzedajemy wszystkie akcje po upływie pół roku (koniec okresu nowego). Miarą efektywności będzie stopa zwrotu z tej inwestycji liczona w postaci średniej arytmetycznej z pięciu stóp zwrotu otrzymanych przy początkowym dniu inwestycji w pięciu kolejnych początkowych dniach każdego z trzech okresów (bessy, hossy i rynku stabilnego).

### 3. Wyniki i wnioski

Przyjęta prosta metoda wykorzystania analizy skupień do konstruowania portfela inwestycyjnego daje zdecydowanie pozytywną ocenę możliwości takiego wykorzystania. Stosując grupowanie spółek właściwą metodą, w połączeniu z wyborem reprezentanta w postaci spółki z najwyższym wskaźnikiem Sharpa, można zawsze uzyskać o wiele wyższą stopę zwrotu od stopy zwrotu całego rynku.

Starając się odpowiedzieć na postawione pytania można sformułować następujące wnioski. Pomijanie skupień jednoelementowych zupełnie nie zdało egzaminu. Przy pomijaniu takich skupień uzyskano nieco gorsze rezultaty niż w przypadku ogólnym. Wobec tego w tabelach 1-3 przedstawione są tylko wyniki wariantu ogólnego. Należy jednak dodać, że metody partycjonujące tzn. *k*-średnich oraz PAM miały o wiele mniejszą skłonność do konstruowania skupień jednoelementowych.

Zdecydowanie nie potwierdziły się sugestie dotyczące tego, że szeregi czasowe cen lepiej opisują zachowania rynku od szeregów stóp zwrotu. W całym badaniu można znaleźć jedynie kilka spośród 27 przypadków, kiedy to średnia stopa zwrotu portfela w nowym okresie była wyższa w przypadku cen.

Zasadniczym wnioskiem dotyczącym uzależnienia efektywności od metod grupowania może być tylko jeden wniosek: zdecydowanie najbardziej godna polecenia jest metoda PAM. Ten wniosek, w pewnym sensie pokrywa się z zaleceniami

**Tabela 1.** Rynek bessy. Średnie arytmetyczne stóp zwrotu (w %) wybranych do portfela spółek na tle stóp zwrotu z całego rynku. Stopa zwrotu rynku z okresu nowego jest równa 2%

Liczba spółek	Metoda grupowania spółek	Postać danych	Średnia stopa zwrotu z okresu przeszłego	Średnia stopa zwrotu z okresu nowego
4 spółki	<i>k</i> -średnich	ceny	9	0
		stopy zwrotu	-5	21
	PAM	ceny	9	0
		stopy zwrotu	2	<b>14</b>
	aglomeracja	ceny	8	-1
		stopy zwrotu	-19	6
7 spółek	<i>k</i> -średnich	ceny	3	0
		stopy zwrotu	-13	9
	PAM	ceny	2	5
		stopy zwrotu	-5	14
	aglomeracja	ceny	-7	1
		stopy zwrotu	-27	4
12 spółek	<i>k</i> -średnich	ceny	-3	1
		stopy zwrotu	-11	10
	PAM	ceny	-3	-1
		stopy zwrotu	-14	<b>14</b>
	aglomeracja	ceny	-10	-2
		stopy zwrotu	-28	5

Źródło: obliczenia własne.

**Tabela 2.** Rynek stabilny. Średnie arytmetyczne stóp zwrotu (w %) wybranych do portfela spółek na tle stóp zwrotu z całego rynku. Stopa zwrotu rynku z okresu nowego jest równa 10%

Liczba spółek	Metoda grupowania spółek	Postać danych	Średnia stopa zwrotu z okresu przeszłego	Średnia stopa zwrotu z okresu nowego
1	2	3	4	5
4 spółki	<i>k</i> -średnich	ceny	24	6
		stopy zwrotu	18	5
	PAM	ceny	22	-2
		stopy zwrotu	1	<b>24</b>
	aglomeracja	ceny	21	6
		stopy zwrotu	-7	5

Tabela 2, cd.

1	2	3	4	5
7 spółek	<i>k</i> -średnich	ceny	20	13
		stopy zwrotu	12	10
	PAM	ceny	19	12
		stopy zwrotu	1	17
	aglomeracja	ceny	15	3
		stopy zwrotu	-7	5
12 spółek	<i>k</i> -średnich	ceny	15	13
		stopy zwrotu	8	15
	PAM	ceny	11	13
		stopy zwrotu	-3	14
	aglomeracja	ceny	9	7
		stopy zwrotu	-6	13

Źródło: obliczenia własne.

**Tabela 3.** Rynek hossy. Średnie arytmetyczne stóp zwrotu (w %) wybranych do portfela spółek na tle stóp zwrotu z całego rynku. Stopa zwrotu rynku z okresu nowego jest równa 7%

Liczba spółek	Metoda grupowania spółek	Postać danych	Średnia stopa zwrotu z okresu przeszłego	Średnia stopa zwrotu z okresu nowego
4 spółki	<i>k</i> -średnich	ceny	41	12
		stopy zwrotu	43	17
	PAM	ceny	41	12
		stopy zwrotu	18	13
	aglomeracja	ceny	38	1
		stopy zwrotu	15	6
7 spółek	<i>k</i> -średnich	ceny	33	13
		stopy zwrotu	35	13
	PAM	ceny	36	3
		stopy zwrotu	19	8
	aglomeracja	ceny	24	2
		stopy zwrotu	8	0
12 spółek	<i>k</i> -średnich	ceny	32	19
		stopy zwrotu	22	-3
	PAM	ceny	32	1
		stopy zwrotu	18	4
	aglomeracja	ceny	26	6
		stopy zwrotu	15	2

Źródło: obliczenia własne.

Craigheada i Klemesruda [2002]. Metoda PAM okazała się najstabilniejsza i efektywna. W niektórych przypadkach lepsza była metoda  $k$ -średnich, ale ta metoda jest bardzo niestabilna w sensie dużego uzależnienia od liczby skupień oraz dużego wpływu losowości na wyniki. Metoda PAM, polecana jako zdecydowanie najlepsza, niezależnie od koniunktury giełdowej, prawie zawsze uzyskiwała portfel ze stopą zwrotu o kilka punktów procentowych wyższą od portfela rynkowego.

Różnice wyników w zależności od stosowanej metody grupowania maleją wraz ze wzrostem liczby skupień czyli spółek w portfelu. Taki związek jest jak najbardziej prawidłowy, na przykład różnice pomiędzy metodą  $k$ -średnich a metodą PAM ukierunkowaną na wyszukiwanie skupień o bardziej od kulistego złożonym kształcie, powinny zanikać przy wzroście liczby skupień.

## Literatura

- Craighead S., Klemesrud B., 2002, *Stock Selection Based on Cluster and Outlier Analysis*, Fifteenth International Symposium on Mathematical Theory of Networks and Systems, University of Notre Dame, August 12-16.
- Gatnar E., Walesiak M., 2004, *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*, Wydawnictwo AE, Wrocław.
- Korzeniewski J., 2014, *Indeks wyboru liczby skupień w zbiorze danych*, Przegląd Statystyczny, nr 61, 2, s. 169-180.
- Marvin K., 2015, *Creating Diversified Portfolios Using Cluster Analysis*, unpublished research, s. 1-15.
- Ren Z., 2005, *Portfolio Construction Using Clustering Methods*, Thesis at The Worcester Polytechnic Institute.
- Rosen F., 2006, *Correlation based clustering of the Stockholm Stock Exchange*, PhD thesis.