PRACE NAUKOWE
Uniwersytetu Ekonomicznego we Wrocławiu

RESEARCH PAPERS
of Wrocław University of Economics

Nr 328

Taksonomia 23
Klasyfikacja i analiza danych – teoria i zastosowania

Redaktorzy naukowi
Krzysztof Jajuga, Marek Walesiak

Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2014

Redaktor Wydawnictwa: Barbara Majewska
Redaktor techniczny: Barbara Łopusiewicz
Korektor: Barbara Cibis
Łamanie: Beata Mazur
Projekt okładki: Beata Dębska

Publikacja jest dostępna w Internecie na stronach:
www.ibuk.pl, www.ebscohost.com,
w Dolnośląskiej Bibliotece Cyfrowej www.dbc.wroc.pl,
The Central and Eastern European Online Library www.ceeol.com,
a także w adnotowanej bibliografii zagadnień ekonomicznych BazEkon
http://kangur.uek.krakow.pl/bazy_ae/bazekon/nowy/index.php

Informacje o naborze artykułów i zasadach recenzowania znajdują się
na stronie internetowej Wydawnictwa
www.wydawnictwo.ue.wroc.pl

Tytuł dofinansowany ze środków Narodowego Banku Polskiego
oraz ze środków Sekcji Klasyfikacji i Analizy Danych PTS

ISSN 1899-3192 (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)
ISSN 1505-9332 (Taksonomia)

Wersja pierwotna: publikacja drukowana

Druk: Drukarnia TOTEM

Spis treści

Summaries

PRACE NAUKOWE UNIWERSYTETU EKONOMICZNEGO WE WROCŁAWIU
RESEARCH PAPERS OF WROCŁAW UNIVERSITY OF ECONOMICS nr 328 • 2014
Taksonomia 23 ISSN 1899-3192
Klasyfikacja i analiza danych – teoria i zastosowania

Iwona Konarzewska
Uniwersytet Łódzki

MODEL PCA DLA RYNKU AKCJI – STUDIUM PRZYPADKU

Streszczenie: W pracy podjęty został temat konstrukcji modeli statystycznych czynnikowych, wykorzystujących analizę głównych składowych (PCA) macierzy kowariancji lub korelacji stóp zwrotu z akcji za pomocą ograniczonej liczby ortogonalnych czynników ryzyka. Przedstawiono założenia teoretyczne modelowania oraz wybrane wyniki badania empirycznego przeprowadzonego dla rynku średnich i dużych spółek na GPW w Warszawie, obserwowanego w latach 2009-2013. Wskazano na możliwości wykorzystania modeli PCA do klasyfikacji spółek ze względu na wrażliwość stóp zwrotu na główne czynniki ryzyka oraz do dekompozycji ryzyka rynkowego portfeli inwestycyjnych.

Słowa kluczowe: analiza głównych składowych, statystyczne modele czynnikowe, ryzyko rynkowe portfeli akcji.

1. Wprowadzenie – statystyczny model PCA dla stóp zwrotu z akcji

Podstawy analizy głównych składowych (PCA – *Principal Component Analysis*) zawarte są w pracach [Pearson 1901] oraz [Hotteling 1933]. W pracy [Jolliffe 2002] podjęto próbę całościowego opisu problemu poszukiwania składowych głównych, ich własności statystycznych, powiązań z analizą czynnikową i innymi metodami statystyki wielowymiarowej i analizy regresji. Alexander [2008] wiele miejsca poświęca metodzie PCA stosowanej w powiązaniu z modelami czynnikowymi (*factor models*) do badania ryzyka inwestycji finansowych.

Celem niniejszej pracy jest pokazanie zastosowania analizy głównych składowych do analizy ryzyka inwestycji na polskim rynku kapitałowym. Zadaniem analizowanej tu wersji modelu PCA – statystycznego modelu PCA, jest przybliżenie stóp zwrotu z akcji dla wybranego segmentu rynku kapitałowego za pomocą ograniczonej liczby czynników, utożsamianych z głównymi składowymi macierzy kowariancji bądź korelacji stóp zwrotu. Wykorzystanie metody prowadzi do konstrukcji specyficznych modeli czynnikowych, opisujących stopy zwrotu z akcji spółek analizowanego segmentu rynku, w których czynnikami ryzyka (utożsamianego tu ze zmiennością i mierzonego za pomocą wariancji) stają się główne składowe macierzy kowariancji lub korelacji stóp zwrotu.

W pracy przyjęto następujące oznaczenia:
$\mathbf{R} = [R_{tj}]$, t = 1,…, T, j = 1,…, N, ,T ≥ N – macierz T-elementowych prób szeregów czasowych stóp zwrotu z akcji N spółek,
$\overline{R}_j$ – średnia z próby dla j-tego szeregu, j = 1,…, N,
$\mathbf{R}_c = [R_{tj}^c]$ – macierz T × N scentrowanych stóp zwrotu, $R_{tj}^c = R_{tj} - \overline{R}_j$,
$\hat{\mathbf{\Sigma}} = \frac{1}{\mathrm{T}}\mathbf{R}_c^{\mathrm{T}}\mathbf{R}_c = \left[\hat{\sigma}_{ij}\right]$, $\hat{\sigma}_{ij} = \frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}}(R_{ti} - \overline{R}_i)(R_{tj} - \overline{R}_j)$, i, j =1,..., N – macierz kowariancji z próby[1].

Dekompozycję macierzy kowariancji stóp zwrotu według wartości własnych przedstawia wzór (1) :

$$\hat{\mathbf{\Sigma}} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{\mathrm{T}}, \tag{1}$$

gdzie: $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, ..., \lambda_{\mathrm{N}})$ – macierz diagonalna z wartościami własnymi macierzy $\hat{\mathbf{\Sigma}}$ na głównej przekątnej; $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_{\mathrm{N}}$,
$\mathbf{V} = \left[\mathbf{v}_1, ..., \mathbf{v}_{\mathrm{N}}\right]$ – macierz znormalizowanych wektorów własnych związanych z $\lambda_1, \lambda_2, ..., \lambda_{\mathrm{N}}$; $\mathbf{V}^{\mathrm{T}}\mathbf{V} = \mathbf{V}\mathbf{V}^{\mathrm{T}} = \mathbf{I}_{\mathrm{N}}$.

Macierz $\mathbf{Y}$ o wymiarach T × N składowych głównych macierzy kowariancji zdefiniowana jest wzorem (2):

$$\mathbf{\Lambda} = \frac{1}{\mathrm{T}}\mathbf{V}^{\mathrm{T}}(\mathbf{R}_c)^{\mathrm{T}}\mathbf{R}_c\mathbf{V} = \mathbf{Y}^{\mathrm{T}}\mathbf{Y} \Rightarrow \mathbf{Y} = \frac{1}{\sqrt{\mathrm{T}}}\mathbf{R}_c\mathbf{V}. \tag{2}$$

Konstruując model, należy dokonać wyboru liczby K ≤ N głównych składowych – czynników ryzyka. W tym celu można zastosować następujące, przykładowe, kryteria: procent objaśnienia wariancji, pominięcie składowych związanych z wartościami własnymi mniejszymi od średniej wartości własnej, wykorzystać tzw. wykres osypiska (zwany też wykresem piargowym), zastosować statystyczne testy istotności i inne [por. Gatnar, Walesiak 2004, s. 208-211; Krzyśko 2009, s. 241-243]. Gdy podstawą konstrukcji modelu jest macierz korelacji stóp zwrotu, można wykorzystać także propozycję sformułowaną w oparciu o teorię macierzy losowych w pracy [Plerou i in. 2002] dla nieskorelowanych szeregów standaryzowanych stóp zwrotu o rozkładzie normalnym przy N→∞ oraz T→∞ wartości własne macierzy korelacji zawierają się w przedziale $\left[\lambda_{\min}, \lambda_{\max}\right]$, gdzie

$$\lambda_{\min} = 1 + \frac{\mathrm{N}}{\mathrm{T}} - 2\sqrt{\frac{\mathrm{N}}{\mathrm{T}}} \text{ oraz } \lambda_{\max} = 1 + \frac{\mathrm{N}}{\mathrm{T}} + 2\sqrt{\frac{\mathrm{N}}{\mathrm{T}}}.$$

[1] W przypadku wykorzystywania do analiz macierzy korelacji zamiast centrowania stosujemy operację standaryzacji.

Model PCA dla K składowych głównych można zapisać następująco:

$$R_{tj} = \alpha_j + \sum_{i=1}^{K} \beta_{ij} Y_{ti} + \varepsilon_{tj}, \quad j = 1,..., N, t = 1,...,T, \tag{3}$$

gdzie ε_{tj} oznacza zaburzenie losowe, o którym przyjmujemy standardowe założenia jak w liniowym modelu ekonometrycznym. Parametry modelu można oszacować za pomocą metody najmniejszych kwadratów, uzyskując przybliżenie

$$\hat{R}_{tj} = \hat{\alpha}_j + \sum_{i=1}^{K} \hat{\beta}_{ij} Y_{ti}, \quad j = 1,..., N, t = 1,..., T. \tag{4}$$

Wartość oczekiwana *j*-tej stopy zwrotu wynosi $E(\hat{R}_j) = \alpha_j$ [2].

Ocena macierzy systematycznych kowariancji stóp zwrotu wyraża się wzorem:

$$\tilde{\boldsymbol{\Sigma}} = \mathbf{B}^{\mathrm{T}} \boldsymbol{\Omega} \mathbf{B}, \tag{5}$$

gdzie: $\mathbf{B} = [\hat{\beta}_{ij}]$ macierz K × N ocen parametrów β dla N spółek i K czynników,

$\boldsymbol{\Omega} = \mathrm{diag}(\lambda_1,..., \lambda_K)$ K × K macierz kowariancji czynników ryzyka.

Cechą charakterystyczną modeli czynnikowych tworzonych przy udziale analizy głównych składowych jest ortogonalność czynników ryzyka. Modele te mogą służyć zarówno do celów opisowych, jak i do szacowania zmienności wspólnej i specyficznej stóp zwrotu dla spółek wybranego segmentu rynku. Mogą być stosowane do wyceny spółek indywidualnych i portfeli inwestycyjnych. Niedogodnością jest trudność interpretacji wyodrębnionych czynników ryzyka w kategoriach ekonomicznych. W literaturze spotyka się próby wyodrębniania segmentów spółek związanych z tak konstruowanymi czynnikami ryzyka[3].

Przed przystąpieniem do etapu modelowania należy zdecydować, czy analizie poddajemy próbkową macierz kowariancji, czy też korelacji stóp zwrotu. Wyniki analiz przeprowadzanych przy zastosowaniu macierzy kowariancji i korelacji różnią się istotnie[4]. Argumentem za stosowaniem macierzy korelacji jest wrażliwość głównych składowych macierzy kowariancji na sposób pomiaru analizowanych zmiennych. Jolliffe [2002, s. 22, 42] podkreśla, że gdy występują istotne różnice w wariancjach badanych zmiennych, to te o największej wariancji wywierają dominujący wpływ na wartości pierwszych głównych składowych. Stosowanie PCA w przypadku macierzy kowariancji jest zasadne, gdy dokonuje się pomiaru zmien-

[2] Wartość oczekiwana każdej z głównych składowych, zarówno w przypadku macierzy kowariancji, jak i korelacji, na mocy definicji wynosi zero.

[3] Na przykład w pracy [Plerou i in. 2002] znajdujemy raport z badania przeprowadzonego dla spółek rynku NASDAQ, a w pracy [Alexander 2008, s. 82-84] – wyniki przykładowego badania wybranych spółek indeksu DJIA.

[4] W pracy [Konarzewska 2012, s. 71] pokazano, jak wygląda zależność pomiędzy wartościami własnymi macierzy korelacji i kowariancji.

nych w tych samych jednostkach, ale nawet wówczas często stosuje się analizę opartą na macierzy korelacji[5]. Badanie przeprowadzono w oparciu o PCA dla macierzy kowariancji i korelacji.

Następny pojawiający się problem to wybór sposobu szacowania elementów macierzy kowariancji czy też korelacji. Dane wykorzystywane w badaniu to szeregi czasowe stóp zwrotu. Wnikliwe studium modeli w zastosowaniu do analizy polskiego rynku finansowego, ze szczególnym naciskiem na modelowanie zmienności, zawiera praca [Doman, Doman 2004]. Właściwymi narzędziami dla modelowania zmienności, tj. niepewności co do przyszłych zmian ceny instrumentu finansowego, są rozkłady i momenty warunkowe[6]. W niniejszej pracy w celu przedstawienia zastosowania modelu PCA jedynie oszacowano elementy macierzy kowariancji i korelacji na podstawie danych dziennych i tygodniowych z okresu I 2009 – VII 2013 (odpowiednio 1149 oraz 238 obserwacji). Wykonano także, aby sprawdzić stabilność uzyskanych oszacowań, obliczenia dla okresu I 2009 – XII 2012 oraz VIII 2009 – VII 2013. Wyniki szacunków dla macierzy kowariancji były zbliżone, również elementy odpowiednich wektorów własnych nie różniły się istotnie.

2. Wybrane wyniki analizy składowych głównych dla rynku akcji dużych i średnich spółek na GPW w Warszawie w okresie I 2009 – VII 2013

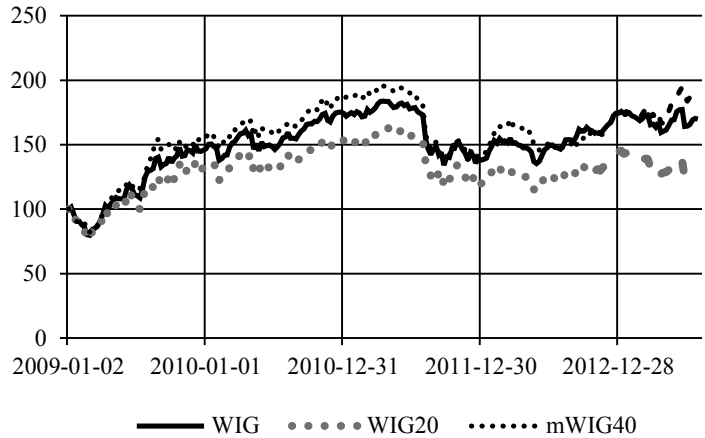
Zachowanie się rynku akcji średnich i dużych spółek w Polsce w latach 2009-2013 ilustrują indeksy dynamiki dla mWIG40, WIG20 oraz indeksu całego rynku akcji – WIG, prezentowane na rys. 1[7]. W ostatnim okresie próby widać wyraźnie silny pozytywny wpływ średnich spółek na kondycję całego rynku. Badaniu poddano stopy zwrotu 50 spółek – w dniu 6.09.2013 r. ich udział w kapitalizacji całego rynku wynosił ponad 75%.

Badanie przeprowadzono na podstawie danych dziennych i tygodniowych. W niniejszej pracy przedstawiono tylko niektóre, wybrane wyniki. Na rys. 2 zaprezentowane są wykresy „osypiska" wartości własnych uzyskane na podstawie danych tygodniowych: a) dla macierzy kowariancji i b) dla macierzy korelacji. Kształty uzyskane dla danych dziennych były bardzo podobne. Osypisko w przy-

[5] M. Krzyśko [2009, s. 238-239] opowiada się za stosowaniem PCA do macierzy kowariancji, wskazując na niejednoznaczność wyników analizy składowych głównych dla macierzy korelacji – „... zależą jedynie od ilorazów współczynników korelacji, a nie od ich rzeczywistych wartości...".

[6] M. Doman i R. Doman [2004, s. 173] zwrócili uwagę na sprzeczność występującą przy próbie prognozowania zmienności na podstawie informacji historycznych i ruchomego „okna" obserwacji, przy założeniu, że stopy zwrotu są niezależne w czasie i o jednakowym rozkładzie i jednocześnie, że poziom zmienności nie jest stały w czasie.

[7] W celu ilustracji dynamiki przyjęto, że wartości indeksów mWIG40, WIG20 i WIG 2.01.2009 są równe 100.



Rys. 1. Indeksy dynamiki dla mWIG40, WIG20 i WIG w latach 2009-2013

Źródło: obliczenia własne.

padku macierzy korelacji jest bardziej strome. Szeregi danych tygodniowych wykazują silniejszą współzależność mierzoną za pomocą stopnia uwarunkowania[8] niż szeregi danych dziennych: dla macierzy kowariancji wyniósł on 164,72, natomiast dla macierzy korelacji 99,72. Odpowiednio dla danych dziennych otrzymano wartości 71,45 oraz 50,13. Liczba składowych głównych objaśniających co najmniej 60% wariancji całkowitej wynosiła dla danych tygodniowych w przypadku macierzy kowariancji 10, w przypadku macierzy korelacji 13. W przypadku danych dziennych uzyskano wynik 11 dla macierzy kowariancji oraz 17 dla macierzy korelacji.

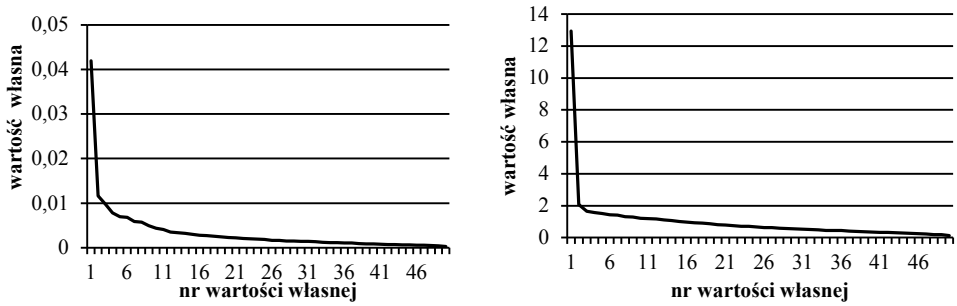
Wartości własne macierzy korelacji porównano z wartościami progowymi ustalonymi dla losowych macierzy korelacji. Progi te wynosiły:

- dane tygodniowe (0,2934; 2,1268) – zaobserwowano tylko 1 wartość własną powyżej górnego progu oraz 7 poniżej dolnego progu,
- dane dzienne (0,6263;1,4607) – zaobserwowano 2 wartości powyżej górnego progu oraz 15 poniżej.

Analizie poddano także elementy wektorów własnych. Do określenia ilości istotnych składowych wektora własnego zaproponowano [za: Plerou i in. 2002] wykorzystanie miernika IPR (*Inverse Participation Ratio*), stosowanego w teorii lokalizacji:

$$\mathrm{IPR}(k) = \sum_{j=1}^{\mathrm{N}} [v_{jk}]^4 \quad k - \text{numer wartości własnej.} \tag{6}$$

[8] Stopień uwarunkowania macierzy kwadratowej symetrycznej obliczany jest jako iloraz największej do najmniejszej wartości własnej.



a) b)

Rys. 2. Wykres „osypiska" dla a) danych tygodniowych i macierzy kowariancji stóp zwrotu; b) danych tygodniowych i macierzy korelacji stóp zwrotu

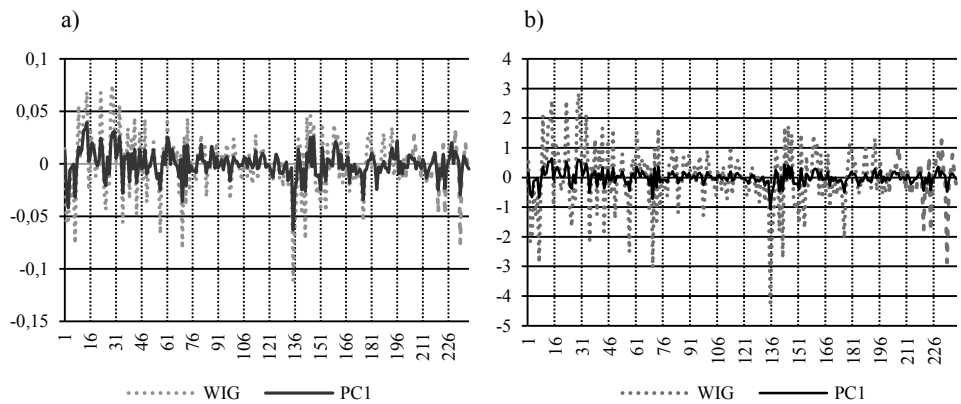
Źródło: obliczenia własne.

Miernik IPR można interpretować jako odwrotność liczby istotnych składowych wektora własnego. Największej wartości własnej odpowiada największa liczba istotnych elementów wektora własnego – w przypadku danych tygodniowych, o silniejszej współzależności, wartość 1/IPR wynosiła 39, dla danych dziennych – ok. 37. Zaobserwowano, że wraz ze spadkiem wartości własnych przekraczających $\lambda_{\max}$ zmniejsza się liczba istotnych elementów wektorów własnych. Dotyczy to także wartości własnych mniejszych od $\lambda_{\min}$. W tym przypadku „istotne" elementy wektorów własnych identyfikują przybliżone liniowe związki między szeregami stóp zwrotu.

Wykresy na rys. 3 przedstawiają pierwsze składowe główne dla danych tygodniowych na tle stóp zwrotu z indeksu rynku WIG – scentrowanych w przypadku dekompozycji macierzy kowariancji i zestandaryzowanych w przypadku macierzy korelacji stóp zwrotu. Pierwsza główna składowa macierzy kowariancji wyraźnie lepiej charakteryzuje ryzyko inwestowania niż w pierwsza główna składowa macierzy korelacji.

Analiza elementów wektorów własnych macierzy kowariancji, oszacowanej na podstawie danych tygodniowych związanych z trzema największymi wartościami własnymi, wskazała na kilka ciekawych obserwacji. Wartości elementów 1. wektora własnego wskazują , że na zmienność rynkowych stóp zwrotu silny wpływ wywiera cały sektor finansowy i przemysłowy (z wyjątkiem części spółek przemysłu elektromaszynowego i metalowego). Wpływ sektora handlowego na pierwszą składową główna jest niewielki. 1. wartości własnej odpowiada 25% zmienności całkowitej. Największe dodatnie elementy 2. wektora własnego odpowiadają spółce OIL sektora przemysłu paliwowego oraz MDS – spółce sektora finansowego. Stosunkowo duże co do modułu elementy obserwujemy dla spółek sektora bankowego. Zmienność związaną z 3. wartością własną można przypisać przede wszystkim spółkom: MDS (finanse), CDR (informatyka), BRS (przemysł metalowy).

Elementy wektora własnego związanego z najmniejszą wartością własną macierzy kowariancji pokazują silny związek o charakterze liniowym stóp zwrotu z akcji PKO oraz PEO.



Rys. 3. Pierwsza składowa główna, dane tygodniowe:
a) dla macierzy kowariancji wraz ze scentrowaną stopą zwrotu z WIG,
b) dla macierzy korelacji wraz ze standaryzowaną stopą zwrotu z WIG

Źródło: obliczenia własne.

W modelu PCA wykorzystano, po analizie, 10 głównych składowych – dotyczy to zarówno danych tygodniowych, jak i dziennych oraz macierzy kowariancji i korelacji. W przypadku dekompozycji macierzy kowariancji oceny parametrów modeli regresji są równe odpowiednim elementom wektorów własnych przemnożonym przez pierwiastek z liczby obserwacji. Wyraz wolny każdego z modeli to odpowiednia średnia stopa zwrotu. W przypadku korzystania z dekompozycji macierzy korelacji nie ma tak bezpośredniej zależności. Wyniki analizy regresji ilustrują zawarte w tabeli 1 oraz 2 uzyskane wartości współczynników determinacji (reprezentujących udział zmienności objaśnionej przez 10 wyróżnionych czynników ryzyka) oraz średnie błędy szacunku (charakteryzujące zmienność nieobjaśnioną). Zaobserwowano, że w przypadku korzystania z dekompozycji macierzy kowariancji modele dla stóp zwrotu z akcji średnich spółek (składowych indeksu mWIG40) o relatywnie wysokiej zmienności charakteryzuje wysoki poziom objaśnienia. Znacznie niższe współczynniki determinacji uzyskano, gdy dekompozycji poddano macierz korelacji. We wszystkich modelach konstruowanych dla dużych spółek, składowych indeksu WIG20, uzyskano wyższy stopień objaśnienia zmienności stóp zwrotu, wykorzystując dekompozycję macierzy korelacji.

W przypadku konstrukcji modelu PCA do opisu stopy zwrotu z akcji z wykorzystaniem dekompozycji według wartości własnych macierzy korelacji można sformułować wniosek o dodatniej liniowej zależności między średnim błędem szacunku (miernikiem zmienności specyficznej) a odchyleniem standardowym stopy

zwrotu. W przypadku korzystania z dekompozycji macierzy kowariancji takiego związku nie obserwuje się.

Tabela 1. Wybrane charakterystyki stóp zwrotu z akcji średnich spółek oraz wyniki modelowania PCA

Spółka	Średnia stopa zwrotu	Odchylenie standardowe	Model PCA			
			dla macierzy kowariancji		dla macierzy korelacji	
			R^2	S_e	R^2	S_e
AGO	–0,0009	0,0565	0,4227	0,044	0,4792	0,0418
ALC	–0,0011	0,0403	0,2061	0,0368	0,5417	0,0279
APT	0,0062	0,0380	0,2862	0,0329	0,4831	0,0208
AST	0,0102	0,0685	0,7772	0,0331	0,5090	0,0491
ATT	0,0115	0,0549	0,4705	0,0409	0,6708	0,0323
BDX	0,0055	0,0442	0,2833	0,0383	0,3580	0,0363
BRS	0,0117	**0,0887**	**0,9581**	**0,0186**	0,4880	0,0650
CAR	0,0083	0,0516	0,3414	0,0429	0,4826	0,0380
CCC	0,0051	0,0402	0,2004	0,0368	0,5637	0,0272
CCI	0,0037	0,0434	0,2133	0,0394	0,4829	0,0319
CDR	0,0122	0,0767	0,8162	0,0337	0,4506	0,0582
CIE	0,0029	0,0625	0,4424	0,0477	0,4891	0,0457
CPS	0,0030	0,0385	0,2243	0,0347	0,4934	0,0281
CRM	0,0127	**0,0822**	**0,9565**	0,0176	0,5085	0,0590
EAT	0,0036	0,045	0,3223	0,0379	0,4509	0,0341
ECH	0,0061	0,0512	0,4101	0,0403	0,4840	0,0377
EMP	0,0053	0,0432	0,124	0,0414	0,3429	0,0358
ENA	0,0012	0,0397	0,1649	0,0371	0,5238	0,0280
GNB	0,0036	0,0585	0,5347	0,0409	0,5693	0,0393
GTN	0,0034	0,0614	0,6344	0,038	0,6372	0,0379
HWE	0,0056	0,0731	0,6401	0,0449	0,4960	0,0531
IDM	–0,0055	**0,0816**	**0,8212**	0,0353	0,5558	0,0557
ING	0,0046	0,0442	0,5009	0,0319	0,6087	0,0283
ITG	0,0138	0,0547	0,2535	0,0484	0,4079	0,0431
KPX	0,0017	0,0569	0,3061	0,0485	0,3678	0,0463
KTY	0,0052	0,0387	0,3081	0,0329	0,5339	0,0270
LPP	0,0086	0,0372	0,1784	0,0346	0,5597	0,0253
MDS	0,0007	**0,1005**	**0,9682**	0,0183	0,5067	0,0722
MIL	0,0055	0,0614	0,6398	0,0377	0,6184	0,0388
NET	0,0034	0,0399	0,1926	0,0367	0,6354	0,0247
OIL	–0,0130	**0,1114**	**0,9939**	0,0089	0,2769	0,0970
ORB	0,0016	0,047	0,2974	0,0403	0,5348	0,0328
PXM	–0,0081	**0,0800**	**0,9510**	0,0181	0,6236	0,0502
RSE	–0,0013	**0,0834**	**0,9367**	0,0215	0,5852	0,0550
TVN	0,0014	0,0593	0,4474	0,0451	0,4743	0,0440

Źródło: obliczenia własne.

3. Uwagi o możliwościach wykorzystania statystycznego modelu PCA do przybliżania ryzyka portfeli akcji

Statystyczny model PCA z K czynnikami ryzyka może być wykorzystany do wyceny ryzyka dowolnego portfela inwestycyjnego. Jeżeli $\mathbf{X} = [x_1,...,x_N]$ oznacza wektor udziałów spółek w portfelu inwestycyjnym, to posługując się ocenami parametrów modeli PCA można uzyskać oceny odpowiednich parametrów charakteryzujących średnią stopę zwrotu z portfela oraz współczynniki wrażliwości względem czynników ryzyka:

$$\hat{\alpha}^{port} = \sum_{j=1}^{N} x_j \hat{\alpha}_j , \quad \hat{\beta}_i^{port} = \sum_{j=1}^{N} x_j \hat{\beta}_{ij} , \quad i = 1,..., K. \tag{7}$$

Tabela 2. Wybrane charakterystyki stóp zwrotu z akcji dużych spółek oraz wyniki modelowania PCA

Spółka	Średnia stopa zwrotu	Odchylenie standardowe	Model PCA			
			dla macierzy kowariancji		dla macierzy korelacji	
			R^2	S_e	R^2	S_e
ACP	0,0013	0,0379	0,3047	0,0324	0,493	0,0277
BHW	0,0047	0,0439	0,4253	0,0341	0,5183	0,0312
BRE	0,0051	0,0525	0,6506	0,0318	0,697	0,0296
BZW	0,0057	0,0419	0,4446	0,032	0,5647	0,0283
EUR	0,0092	0,0518	0,2183	0,0469	0,3694	0,0421
GTC	–0,0006	0,0587	0,5301	0,0412	0,5442	0,0406
KER	0,0077	0,0575	0,5317	0,0403	0,6205	0,0362
KGH	0,0098	0,0568	0,5288	0,0399	0,6414	0,0349
LTS	0,0062	0,0595	0,5336	0,0416	0,6090	0,0381
PEO	0,0027	0,0449	0,6150	0,0285	0,7097	0,0248
PGN	0,0032	0,0337	0,2841	0,0292	0,3784	0,0272
PKN	0,0034	0,0480	0,6163	0,0304	0,6672	0,0283
PKO	0,0020	0,0418	0,6005	0,0271	0,6827	0,0241
SNS	0,0124	0,0585	0,3995	0,0464	0,5163	0,0416
TPS	–0,0009	0,0438	0,1052	0,0424	0,6109	0,0280

Źródło: obliczenia własne.

Wariancję systematyczną portfela można wyrazić za pomocą wzoru (8):

$$\hat{\sigma}^2_{port} = (\hat{\boldsymbol{\beta}}^{port})^{\mathrm{T}} \boldsymbol{\Omega} \hat{\boldsymbol{\beta}}^{port} , \tag{8}$$

gdzie: $\hat{\boldsymbol{\beta}}^{port} = [\beta_1^{port},...,\beta_K^{port}]^{\mathrm{T}}$, a macierz $\boldsymbol{\Omega}$ o wymiarach K × K jest diagonalną macierzą kowariancji K głównych składowych.

Ryzyko specyficzne portfela można przedstawić za pomocą wzoru (9):

$$SR_{port} = \left(\mathbf{X}^{\mathrm{T}} \hat{\mathbf{\Sigma}} \mathbf{X} - (\hat{\boldsymbol{\beta}}^{port})^{\mathrm{T}} \boldsymbol{\Omega} \hat{\boldsymbol{\beta}}^{port} \right)^{1/2}. \tag{9}$$

Wyniki optymalizacji składu portfela analizowanych akcji średnich i dużych spółek przy kryterium minimalizacji wariancji portfela i nieujemności udziałów zawiera tabela 3.

Tabela 3. Skład portfela o minimalnej wariancji

Spółka	Udział (w %)	Spółka	Udział (w %)	Spółka	Udział (w %)	Spółka	Udział (w %)	Spółka	Udział (w %)
AGO	0	CDR	0	HWE	0	OIL	0	GTC	0
ALC	11,72	CIE	0	IDM	0	ORB	0	KER	0
APT	4,38	CPS	6,16	ING	0	PXM	0	KGH	0
AST	0	CRM	1,46	ITG	0	RSE	0	LTS	0
ATT	0,75	EAT	0,09	KPX	0	TVN	0	PEO	0
BDX	3,17	ECH	0	KTY	0	ACP	1,67	PGN	13,52
BRS	0	EMP	8,08	LPP	8,21	BHW	0	PKN	0
CAR	0	ENA	11,91	MDS	0	BRE	0	PKO	0
CCC	3,65	GNB	0	MIL	0	BZW	0	SNS	0
CCI	5,12	GTN	0	NET	3,05	EUR	2,35	TPS	14,71

Źródło: obliczenia własne.

W tabeli 4 prezentowane są wyniki analizy ryzyka związanego z dziesięcioma wyróżnionymi czynnikami ryzyka oraz ryzyka specyficznego tego portfela.

Tabela 4. Analiza ryzyka portfela o minimalnej wariancji

Miernik ryzyka	Ocena dla PCA macierzy kowariancji	Ocena dla PCA macierzy korelacji
Wariancja całkowita	0,000315	0,000315
Wariancja „wspólna”	0,000194	0,000252
Ryzyko specyficzne	1,10%	0,8%

Źródło: obliczenia własne.

Zauważono, że skład optymalnego portfela odpowiada spółkom, dla których poziom objaśnienia (udział wariancji objaśnionej za pomocą 10 „głównych” czynników ryzyka) był niewielki, rzędu 0,2. W przypadku PCA dla macierzy korelacji i 10 czynników ryzyka uzyskano wyższą ocenę zmienności wspólnej, a tym samym niższą ocenę ryzyka specyficznego portfela niż w przypadku wykorzystania macierzy kowariancji stóp zwrotu.

4. Podsumowanie

Wyniki modelowania uzyskane z wykorzystaniem dekompozycji macierzy kowariancji i korelacji różnią się, zwłaszcza w przypadku spółek o wysokiej zmienności stóp zwrotu. W przypadku korzystania z macierzy kowariancji „bety" dla czynników ryzyka oszacowane za pomocą m.n.k. są proporcjonalne do elementów odpowiednich wektorów własnych macierzy kowariancji. Nie jest tak, gdy analizujemy macierz korelacji. Na podstawie identyfikacji „istotnych" elementów wektorów własnych związanych z największymi wartościami własnymi można dokonać próby powiązania czynników ryzyka uzyskanych w wyniku analizy głównych składowych z konkretnymi spółkami.

Model PCA w przypadku silnie współzależnego rynku akcji pozwala na szacowanie ryzyka specyficznego portfeli inwestycyjnych przy wykorzystaniu niewielkiej liczby „głównych" czynników ryzyka, co ważne, ortogonalnych. Analizując parametry równań modelu PCA, można próbować klasyfikować spółki ze względu na wrażliwość stóp zwrotu na „główne" czynniki ryzyka oraz diagnozować, które z nich odgrywają największą rolę w „tworzeniu" ryzyka rynkowego.

Literatura

Alexander C. (2008), *Market Risk Analysis*, vol. II: *Practical Financial Econometrics*, John Wiley & Sons, West Sussex.

Doman M., Doman R. (2004), *Ekonometryczne modelowanie dynamiki polskiego rynku finansowego*, Wydawnictwo Akademii Ekonomicznej w Poznaniu, Poznań.

Gatnar E., Walesiak M. (2004), *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*, Wydawnictwo Akademii Ekonomicznej im. O. Langego we Wrocławiu, Wrocław.

Hotteling H. (1933), *Analysis of a complex of statistical variables into principal components*, „J. Educ. Psychol.", 24, s. 417-441, 498-520.

Jolliffe I.T. (2002), *Principal Component Analysis*, 2nd ed. , Springer, New York.

Konarzewska I. (2012), *Niepewność i ryzyko rynkowe inwestycji w akcje*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź.

Krzyśko M. (2009), *Podstawy wielowymiarowego wnioskowania statystycznego*, Wydawnictwo Naukowe UAM, Poznań

Pearson K. (1901), *On lines and planes of closest fit to systems of points in space*, „Phil. Mag." (6) 2, s. 559-572.

Plerou V., Gopikrishnan P., Rosenow B., Amaral L.A.N., Guhr T., Stanley H.E. (2002), *Random matrix approach to cross correlations in financial data*, „Physical Review", vol. 65, s. 1-18.

MODELLING STOCK MARKET BY PCA FACTOR MODEL – CASE STUDY

Summary: The paper discusses the problem of developing statistical factor models describing returns on stocks. PCA model utilizes limited number of orthogonal risk factors being identified as principal components of covariance or correlation matrix. The article presents theoretical assumptions of modelling and chosen results of the empirical study conducted for 50 large and medium-size firms on the Stock Exchange in Warsaw for the period 2009--2013. The model can be applied to the classification of firms taking into account returns sensitivity with respect to risk factors as well as in performing the decomposition of investment portfolio market risk.

Keywords: principal component analysis PCA, statistical factor models, portfolio market risk.