

Katarzyna Budny

Jan Tatar

Wielowymiarowa analiza statystyczna w badaniach rynku kapitałowego*

Streszczenie

We wcześniejszych pracach autorzy artykułu przedstawili odmienne od klasycznego podejście do opisu i badania wielowymiarowych rozkładów prawdopodobieństwa. Było to możliwe dzięki uprzedniemu zdefiniowaniu potęgi wektora w przestrzeni z iloczynem skalarnym. W niniejszym artykule zaproponowane wcześniej nowe narzędzia wykorzystano do badania i analizy wybranych dwu-, trzy- oraz czterowymiarowych wielkości (o charakterze wektorów losowych) występujących na polskim rynku kapitałowym. Współrzednymi wektorów są indeksy giełdowe WIG, WIG-20, WIG-Banki, WIG-Paliwa oraz rentowności tych indeksów. Wykorzystując dane rynkowe z okresu od 4 stycznia 2016 r. do 7 lipca 2017 r., wyznaczono oraz zinterpretowano estymatory następujących parametrów badanych rozkładów: wartość oczekiwana, wariancja łączna, odchylenie standardowe łączne, współczynnik asymetrii, norma (długość) współczynnika asymetrii, kwadrat współczynnika asymetrii, kurtoza oraz współczynnik ekscesu. W celach poglądowych oraz porównawczych dla każdego wektora wyznaczono również macierz kowariancji, macierz współczynników korelacji cząstkowych oraz klasycznie rozumiane następujące charakte-

Katarzyna Budny, Uniwersytet Ekonomiczny w Krakowie, Wydział Finansów i Prawa, Katedra Matematyki, ul. Rakowicka 27, 31-510 Kraków, e-mail: budnyk@uek.krakow.pl

Jan Tatar, Uniwersytet Ekonomiczny w Krakowie, Wydział Finansów i Prawa, Katedra Matematyki, ul. Rakowicka 27, 31-510 Kraków, e-mail: tatarj@uek.krakow.pl

* Artykuł powstał w wyniku realizacji projektu badawczego sfinansowanego ze środków przyznanych Wydziałowi Finansów i Prawa Uniwersytetu Ekonomicznego w Krakowie w ramach dotacji na utrzymanie potencjału badawczego.

rystyki rozkładów brzegowych: wartość oczekiwana, wariancja, odchylenie standardowe, współczynnik asymetrii, kurtoza oraz współczynnik ekscesu. Ponadto dla wybranych par badanych finansowych wektorów losowych wyznaczono estymator kwadratu współczynnika korelacji wielowymiarowej jako jedną z możliwych miar ich zależności.

Słowa kluczowe: parametry rozkładu prawdopodobieństwa, wielowymiarowy wektor losowy, estymator, rynek kapitałowy, indeks giełdowy, rentowność.

Klasyfikacja JEL: C15, C18, G10, G11.

1. Wprowadzenie

Zdecydowana większość wielkości rozważanych w procesie badania i analizy sytuacji społeczno-demograficzno-gospodarczej konkretnego kraju, grupy krajów bądź konkretnego regionu ma charakter zmiennych losowych. Oznacza to, że są one funkcjami określonymi na przestrzeni probabilistycznej o wartościach należących do znanego zbioru. Jeżeli wartości te są ponadto kwantyfikowalne, to mamy wówczas do czynienia ze zmiennymi losowymi o wartościach liczbowych. Często badaniu poddawane są nie pojedyncze zmienne losowe, ale ich odpowiednio dobrane zestawy, czyli zmienne losowe wielowymiarowe, nazywane także wektorami losowymi.

W klasycznie rozumianej statystyce matematycznej, w konsekwencji także w klasycznej statystyce opisowej, wypracowana została stosowna metodologia opisu, badania i analizy wektorów losowych. Charakterystyczną cechą tej metodologii jest, że badając dany wektor losowy, w istocie badaniu poddaje się jednowymiarową zmienną losową będącą odpowiednio zdefiniowaną funkcją zmiennych losowych stanowiących „współrzędne” tego wektora.

W wielu wcześniejszych pracach autorzy niniejszego opracowania przedstawili propozycję odmiennego od klasycznego sposobu opisu i analizy wielowymiarowych rozkładów prawdopodobieństwa oraz wykorzystali zaproponowane nowe narzędzia do modelowania i badania tych wielkości, które mają charakter wielowymiarowych wektorów losowych.

W prezentowanej pracy, należącej także do nurtu badań nad wielowymiarowymi charakterystykami społeczno-gospodarczymi, zaproponowane wcześniej nowe narzędzia probabilistyczne wykorzystane zostały do opisu i badania wybranych wektorowych charakterystyk polskiego rynku kapitałowego. Współrzędnymi badanych wektorów były poziomy indeksów giełdowych: WIG, WIG-20, WIG-Banki oraz WIG-Paliwa, a także rentowności tych indeksów. Dane, które wykorzystane zostały w części merytorycznej artykułu, pochodzą z okresu od 4 stycznia 2016 r. do 7 lipca 2017 r.

W pierwszej, metodologicznej części artykułu przypomniano, wskazując jednocześnie konkretne pozycje literaturowe, probabilistyczne i statystyczne

narzędzia, czyli definicje oraz postaci zarówno samych parametrów wielowymiarowych rozkładów prawdopodobieństwa, jak i ich estymatorów, których wartości, czyli oszacowania (oceny) parametrów, zostały wyznaczone i zinterpretowane w części merytorycznej.

2. Wybrane parametry wielowymiarowych rozkładów prawdopodobieństwa oraz ich estymatory

Podstawowymi charakterystykami rozkładu jednowymiarowej zmiennej losowej są jego momenty zwykłe oraz centralne (por. np. Feller 1969, Shao 2003). W pracach (Tatar 1996, 1999), wykorzystując definicję potęgi wektora w przestrzeni z iloczynem skalarnym, zaproponowano wielowymiarowe uogólnienie tych pojęć.

Niech $n, r \in N_0 = N \cup \{0\}$ oraz niech

$$L_n^r(\Omega) = \left\{ X: \Omega \rightarrow R^n: X \text{ wektor losowy i } \int_{\Omega} \|X\|^r dP < +\infty \right\}$$

będzie przestrzenią wektorów losowych sumowalnych (całkowalnych) w r -tej potędze.

Należy nadmienić w tym miejscu, że w literaturze probabilistycznej wielkość $E[\|X\|^r] = \int_{\Omega} \|X\|^r dP$ jest niekiedy nazywana (por. np. Bilodeau i Brenner 1999) momentem rzędu r wektora losowego X i oznaczana symbolem $E[X^r]$. Natomiast w pracy (Tatar 2002), przez analogię do przypadku jednowymiarowego, wyrażenie to określono jako moment absolutny rzędu r wektora losowego X .

Rozważmy zatem wektor $X: \Omega \rightarrow R^n$ należący do przestrzeni $L_n^r(\Omega)$, taki że jego moment absolutny rzędu r jest skończony. Przyjmijmy ponadto, że

$$\Sigma_X = \begin{bmatrix} \sigma_1^2 & \dots & \rho_{1n} \sigma_1 \sigma_n \\ \vdots & \ddots & \vdots \\ \rho_{1n} \sigma_1 \sigma_n & \dots & \sigma_n^2 \end{bmatrix}$$

jest klasycznie rozumianą macierzą kowariancji wektora X , zaś

$$R_X = \begin{bmatrix} 1 & \dots & \rho_{1n} \\ \vdots & \ddots & \vdots \\ \rho_{1n} & \dots & 1 \end{bmatrix}$$

macierzą jego korelacji cząstkowych.

Definicja 1 (Tatar 1996, 1999). Momentem zwykłym rzędu r wektora losowego $X: \Omega \rightarrow R^n$ nazywamy wyrażenie $\alpha_{r,n}(X) = E[X^r]$.

Zauważmy, że moment zwykły rzędu pierwszego wektora losowego jest wektorem wartości oczekiwanych jego składowych, czyli $\alpha_{1,n}(X) = m(X) = EX$; moment ten będziemy nazywać wartością oczekiwaną wektora losowego X .

Definicja 2 (Tatar 1996, 1999). Momentem centralnym rzędu r wektora losowego $X: \Omega \rightarrow R^n$ nazywamy wielkość $\mu_{r,n}(X) = E[(X - EX)^r]$.

Szczególne znaczenie ma moment centralny rzędu drugiego wektora X , czyli $\mu_{2,n}(X) = E[(X - EX)^2]$. Moment ten będziemy nazywać wariancją wektora losowego X oraz oznaczać także symbolem D^2X .

Za pomocą momentów centralnych wektora losowego definiujemy z kolei takie charakterystyki wielowymiarowego rozkładu prawdopodobieństwa, jak współczynnik asymetrii czy kurtoza.

Definicja 3 (Tatar 2000). Współczynnik asymetrii wektora losowego X definiujemy jako:

$$\gamma_{1,n}(X) = SkewX = \frac{\mu_{3,n}(X)}{(\mu_{2,n}(X))^{\frac{3}{2}}} = \frac{E[(X - EX)^3]}{(D^2X)^{\frac{3}{2}}}.$$

Zwróćmy uwagę, że zdefiniowana powyżej miara asymetrii rozkładu wektora losowego dostarcza także wartości wektorowej. Wskazuje ona zatem kierunek, na którym występuje ewentualna „skośność” (asymetria). Do pomiaru i wyrażenia jej wielkości można z kolei wykorzystać długość lub kwadrat wektora $\gamma_{1,n}(X)$, czyli $\|\gamma_{1,n}(X)\|$ lub $\beta_{1,n}(X) = \gamma_{1,n}^2(X)$.

Definicja 4 (Budny 2009, Budny i Tatar 2009). Kurtozą wektora losowego X nazywamy wielkość $\beta_{2,n}(X)$ wyrażoną jako:

$$\beta_{2,n}(X) = KurtX = \frac{\mu_{4,n}(X)}{(\mu_{2,n}(X))^2} = \frac{E[(X - EX)^4]}{(D^2X)^2}.$$

W pracy (Budny 2012) wykazano, że kurtoza wektora losowego $N: \Omega \rightarrow R^n$ o wielowymiarowym rozkładzie normalnym z tą samą wektorową wartością oczekiwaną i tą samą macierzą kowariancji, jakimi charakteryzuje się wektor losowy X , przyjmuje postać:

$$\beta_{2,n}(N) = KurtN = 1 + \frac{2 \sum_{i=1}^n (D^2 N_i)^2 + 2 \sum_{\substack{i,j=1 \\ i \neq j}}^n \rho_{ij}^2 D^2 N_i \cdot D^2 N_j}{\sum_{i,j=1}^n D^2 N_i \cdot D^2 N_j} =$$

$$= 1 + \frac{2 \sum_{i=1}^n \sigma_i^4 + 2 \sum_{\substack{i,j=1 \\ i \neq j}}^n \rho_{ij}^2 \sigma_i^2 \sigma_j^2}{\sum_{i,j=1}^n \sigma_i^2 \sigma_j^2},$$

gdzie ρ_{ij} oznacza współczynnik korelacji składowych N_i oraz N_j wektora N .

Wykorzystując pojęcie kurtozy, w pracy (Budny 2014b) zdefiniowano kolejną charakterystykę wielowymiarowego rozkładu prawdopodobieństwa.

Definicja 5 (Budny 2014b). Współczynnikiem ekscesu (ekscesem) wektora losowego $X: \Omega \rightarrow R^n$ nazywamy wielkość $\gamma_{2,n}(X)$ określoną następująco:

$$\begin{aligned} \gamma_{2,n}(X) &= \text{Ekscess } X = \\ &= \text{Kurt } X - \left(1 + \frac{2 \sum_{i=1}^n (D^2 X_i)^2 + 2 \sum_{\substack{i,j=1 \\ i \neq j}}^n \rho_{ij}^2 D^2 X_i \cdot D^2 X_j}{\sum_{i,j=1}^n D^2 X_i \cdot D^2 X_j} \right) = \beta_{2,n}(X) - \beta_{2,n}(N). \end{aligned}$$

W badaniu wielowymiarowych wielkości losowych istotne znaczenie mają pomiar oraz analiza ich wzajemnych zależności. Jedną z ważniejszych miar w tym zakresie jest współczynnik korelacji wielowymiarowej.

Niech $X: \Omega \rightarrow R^m$ oraz $Y: \Omega \rightarrow R^n$ będą dowolnymi wektorami losowymi, przy czym o wektorze Y zakładamy, że $\mu_{2,n}(Y) = D^2 Y \neq 0$.

Definicja 6 (Budny 2018). Współczynnik korelacji wielowymiarowej (współczynnik korelacji wektorów losowych X oraz Y) definiujemy jako:

$$\rho_{multi}(X, Y) = \sqrt{\frac{D^2(A_{X,Y} X)}{D^2 Y}},$$

gdzie $A_{X,Y}$ jest macierzą realizującą warunek $\min_{\substack{A \in M_{(n \times m)}(R) \\ B \in M_{(n \times 1)}(R)}} E[(Y - (AX + B))^2]$.

Ważnymi własnościami współczynnika korelacji wielowymiarowej są jego nieujemność oraz niezmienniczość względem skali i translacji wektorów. Inną istotną własnością miary ρ_{multi} jest brak jej symetrii; tzn. mogą istnieć pary wektorów X oraz Y , dla których zachodzi nierówność $\rho_{multi}(X, Y) \neq \rho_{multi}(Y, X)$.

W pracy (Budny 2018) wykazano, że kwadrat zdefiniowanego powyżej współczynnika korelacji wielowymiarowej można, w sposób równoważny, zapisać jako:

$$\rho_{multi}^2(X, Y) = \sum_{i=1}^n \frac{D^2 Y_i}{D^2 Y} \rho^2(X, Y_i),$$

gdzie $\rho(X, Y_i)$ jest współczynnikiem korelacji wielokrotnej (wielorakiej) między zmienną losową Y_i a wektorem losowym $X = (X_1, \dots, X_m)$, dla wszystkich $i \in \{1, \dots, n\}$.

W praktyce, w analizie wielowymiarowych danych empirycznych – badając ich charakterystyki oparte na definicji potęgi wektora – wykorzystuje się (podobnie jak ma to miejsce w przypadku jednowymiarowym) ich odpowiednie estymatory, czyli oszacowania uzyskane na podstawie dostępnej próby statystycznej. W pracach (Budny 2014a, 2017, 2018) zaproponowano postać estymatorów momentów zwykłych, momentów centralnych, współczynnika asymetrii, kwadratu współczynnika asymetrii, kurtozy i ekscesu wektora losowego oraz wykazano, że są to statystyki zgodne i (co najmniej) asymptotycznie nieobciążone. Dla potrzeb rozważań prowadzonych w dalszej części pracy przypomnijmy ich definicje.

Niech $X: \Omega \rightarrow R^n$ w dalszym ciągu będzie badanym n -wymiarowym wektorem losowym oraz niech $k \in N$ oznacza liczebność wykorzystywanej próby losowej. Sama zaś próba losowa prosta z rozkładu n -wymiarowego niech będzie ciągiem wektorów:

$$X^1: \Omega \rightarrow R^n, \dots, X^k: \Omega \rightarrow R^n.$$

Definicja 7 (Budny 2014a, 2018). Estymatorem momentu zwykłego rzędu r wektora losowego (inaczej: momentem zwykłym rzędu r w próbie wielowymiarowej) nazywamy statystykę:

$$a_{r,n} = \frac{\sum_{i=1}^k (X^i)^r}{k}. \quad (1)$$

Szczególne znaczenie ma estymator momentu zwykłego rzędu pierwszego (estymator wartości oczekiwanej), czyli $a_{1,n}$ oznaczony także symbolem \bar{X} .

Definicja 8 (Budny 2017, 2018). Estymatorem momentu centralnego rzędu r wektora losowego (momentem centralnym rzędu r w próbie wielowymiarowej) nazywamy statystykę:

$$m_{r,n} = \frac{\sum_{i=1}^k (X^i - \bar{X})^r}{k}. \quad (2)$$

W tym przypadku szczególnie istotny jest estymator momentu centralnego rzędu drugiego (tzn. estymator wariancji) $m_{2,n}$.

W konsekwencji pierwiastek kwadratowy z wariancji, czyli

$$s = \sqrt{m_{2,n}} = \sqrt{\frac{\sum_{i=1}^k (X^i - \bar{X})^2}{k}} \quad (3)$$

jest estymatorem odchylenia standardowego.

Estymatory momentów zwykłych i centralnych posłużyły do zdefiniowania kolejnych ważnych statystyk.

Definicja 9 (Budny 2018). Estymatorem współczynnika asymetrii wektora losowego (inaczej: współczynnikiem asymetrii w próbie wielowymiarowej) nazywamy wielkość:

$$\hat{\gamma}_{1,n} = \frac{m_{3,n}}{(m_{2,n})^{\frac{3}{2}}}. \quad (4)$$

Naturalną konsekwencją powyższej definicji są postacie:

– estymatora długości wektora (współczynnika) asymetrii:

$$\|\hat{\gamma}_{1,n}\| = \left\| \frac{m_{3,n}}{(m_{2,n})^{\frac{3}{2}}} \right\|, \quad (5)$$

oraz

– estymatora kwadratu współczynnika asymetrii wektora losowego (Budny 2018):

$$\hat{\beta}_{1,n} = (\hat{\gamma}_{1,n})^2. \quad (6)$$

W kolejnych dwóch definicjach podano postaci estymatorów współczynników koncentracji i spłaszczenia wielowymiarowego rozkładu prawdopodobieństwa.

Definicja 10 (Budny 2018). Estymatorem kurtozy wektora losowego (kurtozą w próbie wielowymiarowej) nazywamy statystykę:

$$\hat{\beta}_{2,n} = \frac{m_{4,n}}{(m_{2,n})^2}. \quad (7)$$

Definicja 11 (Budny 2018). Estymatorem współczynnika ekscesu wektora losowego (współczynnikiem ekscesu w próbie wielowymiarowej) nazywamy wielkość:

$$\hat{\gamma}_{2,n} = \hat{\beta}_{2,n} - \left(1 + \frac{2 \sum_{i=1}^n (m_{2,n}^i)^2 + 2 \sum_{i,j=1}^n r_{ij}^2 m_{2,n}^i m_{2,n}^j}{(m_{2,n})^2} \right), \quad (8)$$

przy czym $m_{2,n}^i$ oznacza „jednowymiarową” wariancję w próbie i -tej współrzędnej wektora losowego, a r_{ij}^2 współczynnik korelacji cząstkowej w próbie między i -tą oraz j -tą współrzędną, dla wszystkich $i, j \in \{1, \dots, n\}$.

Na koniec tej części pracy przypomnijmy postać estymatora zaproponowanej w definicji 6 miary zależności wektorów losowych $X: \Omega \rightarrow R^m$ oraz $Y: \Omega \rightarrow R^n$.

Definicja 12 (Budny 2018). Estymatorem kwadratu współczynnika korelacji wielowymiarowej wektorów X oraz Y nazywamy wielkość:

$$R_{multi}^2(X, Y) = \sum_{i=1}^n \frac{m_{2,n}^{Y_i}}{m_{2,n}^Y} R_{X, Y_i}^2, \quad (9)$$

gdzie $m_{2,n}^{Y_i}$ jest klasycznym estymatorem „jednowymiarowej” wariancji zmiennej losowej Y_i , $m_{2,n}^Y = \sum_{i=1}^n m_{2,n}^{Y_i}$, natomiast R_{X, Y_i}^2 jest estymatorem kwadratu współczynnika korelacji wielokrotnej między zmienną losową Y_i a zestawem zmiennych losowych X_1, \dots, X_m , czyli wektorem losowym $X = (X_1, \dots, X_m)$, dla wszystkich $i \in \{1, \dots, n\}$.

3. Wybrane charakterystyki polskiego rynku kapitałowego – analiza

3.1. Uwagi ogólne

W tej części pracy wykorzystano przedstawione w punkcie 2 charakterystyki wielowymiarowych rozkładów prawdopodobieństwa (oraz ich estymatory) do badania kilku wektorów losowych wybranych z polskiego rynku finansowego. Współrzędnymi (składowymi) tych wektorów są wielkości wymienione we wprowadzeniu do niniejszego opracowania.

Próbkę, którą wykorzystano do uzyskania ocen wybranych parametrów rozkładu badanych wektorów, stanowiły dane pobrane ze strony www.money.pl. Są to poziomy „zamknięcia” w każdym dniu notowań w okresie od 4 stycznia 2016 r. do 10 lipca 2017 r. W przypadku danych nominalnych wykorzystano zatem próbę 380-elementową, zaś w badaniu rentowności wybranych instrumentów (indeksów giełdowych), z oczywistych względów, próbę 379-elementową. Dla każdego z wyspecyfikowanych wektorów finansowych wyznaczono estymatory następujących parametrów ich rozkładów:

- wartość oczekiwana (wektor), według wzoru (1),
- wariancja łączna (skalar), według wzoru (2),
- odchylenie standardowe łączne (skalar), według wzoru (3),
- współczynnik asymetrii (wektor), według wzoru (4),
- norma/długość współczynnika asymetrii (skalar), według wzoru (5),
- kwadrat współczynnika asymetrii (skalar), według wzoru (6),
- kurtoza, według wzoru (7),
- współczynnik ekscesu, według wzoru (8).

W celach poglądowych, poznawczych oraz porównawczych w każdym przypadku wyznaczono również macierz kowariancji, macierz współczynników korelacji cząstkowych oraz klasycznie rozumiane następujące charakterystyki

rozkładów brzegowych: wartość oczekiwana, wariancja, odchylenie standardowe, współczynnik asymetrii, kurtoza oraz współczynnik ekscesu.

Dla wybranych par badanych finansowych wektorów losowych wyznaczono także (według wzoru (9)) estymator kwadratu współczynnika korelacji wielowymiarowej jako jedną z miar ich zależności.

3.2. Estymatory wybranych charakterystyk rozkładu czterowymiarowego wektora poziomu indeksów

Niech X_1 oznacza wartość indeksu WIG, X_2 – wartość indeksu WIG-20, X_3 – wartość indeksu WIG-Banki oraz X_4 – wartość indeksu WIG-Paliwa na Giełdzie Papierów Wartościowych w Warszawie (GPW). Jako pierwszy poddamy badaniu wektor postaci $X = (X_1, X_2, X_3, X_4)$. Korzystając z danych, o których mowa w punkcie 3.1, tj. z próby 380-elementowej, otrzymujemy podane poniżej wyniki.

1. Macierz kowariancji:

$$\Sigma_X = \begin{bmatrix} 36590093,7 & 1265614,20 & 3773372,28 & 6361990,66 \\ 1265614,20 & 46583,96 & 134004,30 & 220345,67 \\ 3773372,28 & 134004,30 & 413766,16 & 630121,41 \\ 6361990,66 & 220345,67 & 630121,41 & 1169543,61 \end{bmatrix}.$$

2. Macierz współczynników korelacji cząstkowej:

$$R_X = \begin{bmatrix} 1 & 0,9694 & 0,9698 & 0,9725 \\ 0,9694 & 1 & 0,9652 & 0,9440 \\ 0,9698 & 0,9652 & 1 & 0,9058 \\ 0,9725 & 0,9440 & 0,9058 & 1 \end{bmatrix}.$$

3. Charakterystyki rozkładów brzegowych wektora X (tabela 1).

Tabela 1. Estymatory charakterystyk rozkładów brzegowych

Charakterystyka	WIG	WIG-20	WIG-Banki	WIG-Paliwa
Wartość oczekiwana	50976,24	1955,92	6376,48	5323,85
Wariancja	36 590 093,47	46 583,96	413 766,16	1 169 543,61
Odchylenie standardowe	6048,98	215,83	643,25	1 081,45
Współczynnik asymetrii	0,6267	0,6909	0,5168	0,8203
Kurtoza	1,8384	1,9625	1,9341	2,1897
Współczynnik ekscesu	-1,1616	-1,0375	-1,0659	-0,8103

Źródło: opracowanie własne.

4. Charakterystyki łączne wektora X (tabela 2).Tabela 2. Estymatory charakterystyk łącznych rozkładu wektora losowego X (opartych na definicji potęgi wektora)

Charakterystyka	Wartość estymatora
EX	(50 976,24; 1955,92; 6376,48; 5323,85)
D^2X	38 219 987,20
Odchylenie standardowe	6182,23
$\gamma_{1,n}(X)$	(0,6162; 0,0238; 0,0625; 0,1197)
Norma (długość) $\gamma_{1,n}$	0,6313
$\beta_{1,n}(X)$	0,3985
$\beta_{2,n}(X)$	1,8419
$\gamma_{2,n}(X)$	-1,1488

Źródło: opracowanie własne.

Interpretacja wyników

1. Wariancja łączna wektora losowego X jest równa sumie wariancji rozkładów brzegowych. Nie jest to zaskoczeniem, ponieważ w pracach (Tatar 1996, Tatar 1999) udowodniono prawdziwość tego związku w przypadku każdego wektora losowego.

2. Największy wpływ na wartość łącznego odchylenia standardowego wektora losowego X ma odchylenie standardowe zmiennej losowej X_1 , czyli poziomu indeksu WIG.

3. Każdy z rozkładów brzegowych wektora X charakteryzuje się wyraźną asymetrią prawostronną („dodatnią”). Najbardziej „wydłużone prawe ramię” ma rozkład zmiennej WIG-Paliwa (0,8203), najmniej zaś rozkład indeksu WIG-Banki.

4. Łączna asymetria wektora X wyrażona czterowymiarowym wektorem $\gamma_{1,n} = (0,6162; 0,0238; 0,0625; 0,1197)$ również wskazuje na „dodatnią skośność” jego rozkładu na każdym z czterech kierunków. Zdecydowanie największą wartość ma jednak ta współrzędna wektora asymetrii, która odpowiada „kierunkowi” (zmiennej losowej) WIG. Co więcej, długość wektora asymetrii łącznej (0,6313) jest zbliżona do wartości współczynnika asymetrii brzegowej zmiennej losowej WIG. Spostrzeżenie to skłania do konkluzji, że próba analizy asymetrii wektora losowego poprzez analizę asymetrii jego rozkładów brzegowych może prowadzić do nieuprawnionych wniosków. W rozważanym w artykule przypadku takie podejście mogłoby błędnie sugerować, że wektor X będzie charakteryzował się największą asymetrią (skośnością) na kierunku WIG-Paliwa. Można zatem sformułować wniosek, że analizując asymetrię rozkładów wielowymiarowych,

nie wystarczy zbadać, jak bardzo (i w którą stronę) „wydłużone” są ramiona rozkładów brzegowych, ale trzeba ponadto zmierzyć, jak duża „masa prawdopodobieństwa” mieści się pod tymi ramionami, a na to pozwala współczynnik asymetrii łącznej $\gamma_{1,n}$.

5. Ostatnie dwa wskaźniki, tzn. kurtoza oraz (będący jej konsekwencją) współczynnik ekscesu, mogą być interpretowane jako miary, odpowiednio, „spłaszczenia” badanego rozkładu oraz jego różnienia się (odbiegania) od rozkładu normalnego o tej samej wartości oczekiwanej oraz wariancji. Interpretacja ta odnosi się do rozkładów zarówno jednowymiarowych, jak i wielowymiarowych. W przypadku badanego wektora indeksów giełdowych największym spłaszczeniem charakteryzuje się rozkład brzegowy WIG-Paliwa (2,1897), chociaż odbiega on w najmniejszym stopniu (spośród wszystkich czterech rozkładów brzegowych) od rozkładu normalnego (eksces równy $-0,8103$).

Kurtoza łącznego rozkładu wektora X przyjmuje wartość (1,84190) zbliżoną do wartości kurtozy głównego indeksu WGPW, czyli WIG-u. Podobnie jest w przypadku wskaźnika ekscesu: czterowymiarowy rozkład wektora indeksów giełdowych różni się od odpowiadającego mu rozkładu normalnego w przybliżeniu w tym samym stopniu ($-1,1488$), w jakim rozkład WIG-u różni się od rozkładu normalnego o wartości oczekiwanej 50 976,24 oraz odchyleniu standardowym 6048,98.

Interpretując wartość estymatora współczynnika ekscesu, należy również zwrócić uwagę na jego znak. Ujemna wartość wskazuje na to, że odbieganie od rozkładu normalnego ma miejsce przede wszystkim w przypadku realizacji bardziej odległych od wartości oczekiwanej (czyli w przypadku jednowymiarowym na tzw. ogonach), zaś dodatni współczynnik ekscesu świadczy o różnicach między badanym rozkładem a rozkładem normalnym dla realizacji bliskich wartości średniej.

3.3. Estymatory wybranych charakterystyk rozkładu dwuwymiarowego wektora poziomu indeksów

Biorąc pod uwagę oznaczenia podane w punkcie 3.2, będziemy teraz rozważać wektor losowy $Y = (X_1, X_2)$, czyli wektor, którego brzegowymi zmiennymi losowymi są WIG oraz WIG-20. Korzystając z tej samej próby historycznej (ciąg 380-elementowy), wyznaczmy estymatory charakterystyk wymienionych w punkcie 3.1. Stosując odpowiednie formuły, otrzymujemy podane poniżej wyniki.

1. Macierz kowariancji:

$$\Sigma_Y = \begin{bmatrix} 36590093,47 & 1265614,20 \\ 1265614,20 & 46583,96 \end{bmatrix}.$$

2. Macierz współczynników korelacji cząstkowej:

$$R_Y = \begin{bmatrix} 1 & 0,694 \\ 0,9694 & 1 \end{bmatrix}.$$

3. Charakterystyki rozkładów brzegowych wektora Y ; wartości estymatorów zostały obliczone w punkcie 3.2 i podano je w tabeli 1.

4. Charakterystyki łączne wektora Y (tabela 3).

Tabela 3. Estymatory charakterystyk łącznych rozkładu wektora losowego Y (opartych na definicji potęgi wektora)

Charakterystyka	Wartość
EY	(50 976,24; 1955,92)
D^2Y	36 636 677,43
Odchylenie standardowe	6052,82
$\gamma_{1,n}(Y)$	(0,6264; 0,0242)
Norma (długość) $\gamma_{1,n}$	0,6269
$\beta_{1,n}(Y)$	0,3930
$\beta_{2,n}(Y)$	1,8384
$\gamma_{2,n}(Y)$	-1,1613

Źródło: opracowanie własne.

Interpretacja wyników

1. Podobnie jak w przypadku wektora badanego w punkcie 3.2 wariancja łączna jest liczbą, czyli skalarem, nie zaś wektorem. Jest to immanentna własność wariancji liczonej według nowej koncepcji charakteryzacji wielowymiarowych rozkładów prawdopodobieństwa. Co więcej, według tej koncepcji wszystkie momenty rzędu parzystego są skalarami, wszystkie zaś momenty rzędu nieparzystego są wektorami. Nie powinno budzić to zaniepokojenia; przeciwnie, skoro momenty rzędu nieparzystego mają służyć (i służą) do definiowania parametrów położenia rozkładu, to w przestrzeni wielowymiarowej powinny być wektorami; skoro zaś momenty rzędu parzystego wykorzystuje się do określania parametrów rozproszenia danego rozkładu, to – także w przestrzeni wielowymiarowej – powinny być skalarami (liczbami).

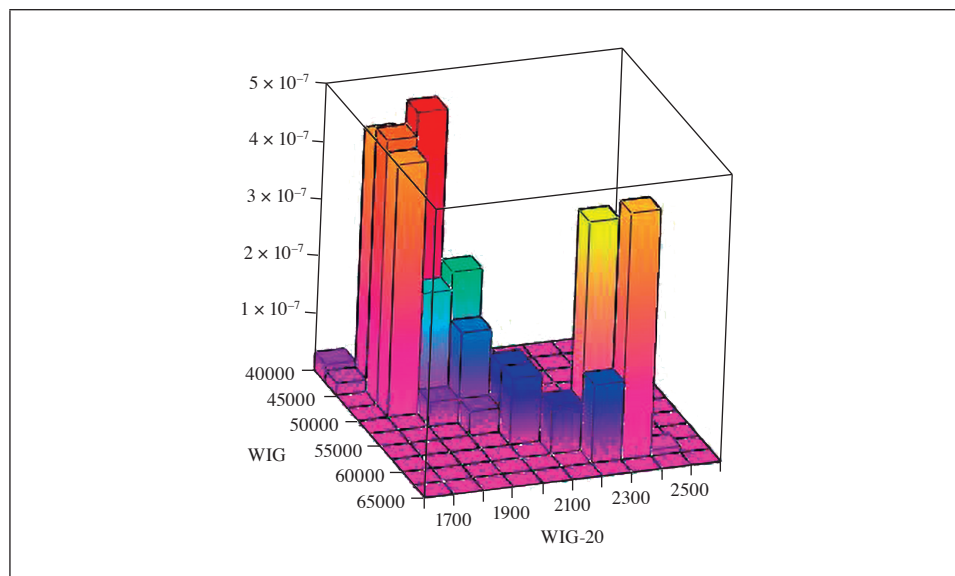
2. Również w tym przypadku współczynniki asymetrii obliczone oddzielnie dla obydwu rozkładów brzegowych mogłyby skłaniać do błędnego wniosku, że badany wektor losowy jest „bardziej asymetryczny” na kierunku X_2 , tzn. WIG-20 (0,6909 vs 0,6267). Tymczasem wyznaczony łączny (wektorowy) współczynnik

(0,6264; 0,0242) wskazuje na zdecydowaną skośność na kierunku WIG. Przyczyny takiego stanu rzeczy wskazano, interpretując wyniki uzyskane w punkcie 3.2.

3. Rozkład prawdopodobieństwa wektora dwóch indeksów giełdowych rozważanego w tym punkcie pracy także charakteryzuje się znacznym spłaszczeniem (kurtoza równa 1,8384) oraz odbieganiem od normalności (-1,1613) przede wszystkim dla tych realizacji, które są dość odległe od wartości oczekiwanej (ujemny wskaźnik ekscesu).

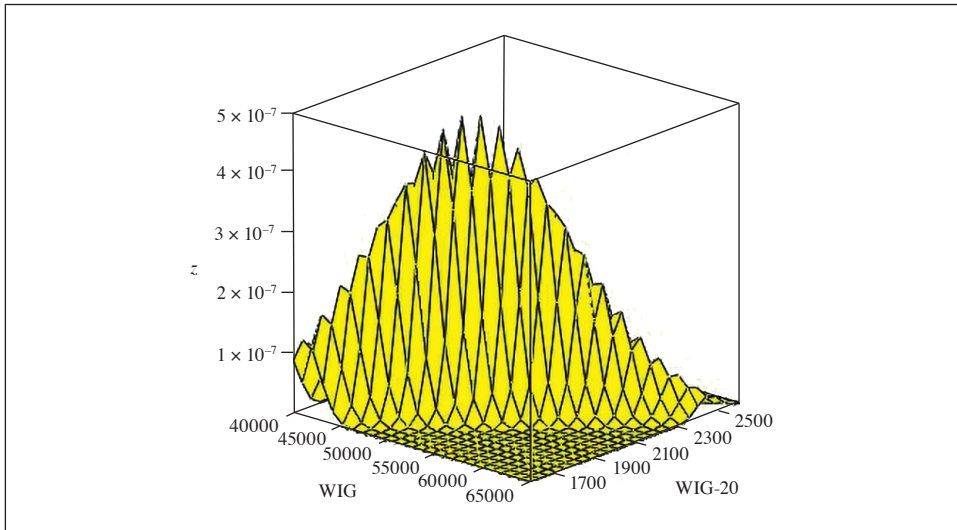
4. Dwuwymiarowość wektora Y stwarza możliwość graficznego przedstawienia jego rozkładu. Poniżej zaprezentowano trzy wykresy: histogram rozkładu wektora Y skonstruowany na podstawie 380-elementowej próby (rys. 1), dwuwymiarowy rozkład normalny o tej samej wartości oczekiwanej oraz wariancji (rys. 2), a także zestawienie obydwu rysunków (rys. 3). Wykresy te potwierdzają słuszność sformułowanych wcześniej wniosków.

Niestety, z oczywistych przyczyn, nie jest możliwe graficzne przedstawienie rozkładu (oraz jego charakterystyk) czterowymiarowego wektora badanego w poprzednim punkcie pracy – wymagałoby to sporządzenia wykresu w przestrzeni pięciowymiarowej.



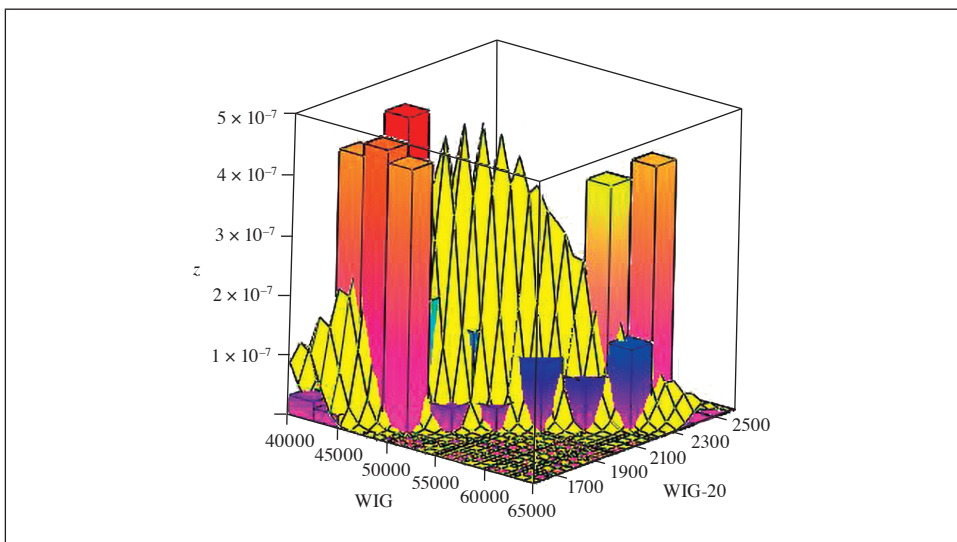
Rys. 1. Histogram rozkładu wektora Y

Źródło: opracowanie własne.



Rys. 2. Dwuwymiarowy rozkład normalny o wektorze wartości oczekiwanych EY i macierzy kowariancji Σ_Y

Źródło: opracowanie własne.



Rys. 3. Zestawienie histogramu oraz dwuwymiarowego rozkładu normalnego

Źródło: opracowanie własne.

3.4. Estymacja wielowymiarowego współczynnika korelacji

Niech w dalszym ciągu obowiązują ustalenia podane w punkcie 3.2 oraz niech $Y = (X_1, X_2)$ i $Z = (X_3, X_4)$. Będziemy zatem rozważać dwa dwuwymiarowe wektory: $Y = (WIG, WIG-20)$ oraz $Z = (WIG-Banki, WIG-Paliwa)$. Zbadamy ich wzajemną zależność, obliczając, według wzoru (9), estymator kwadratu wielowymiarowego współczynnika korelacji. Otrzymujemy:

$$\rho_{multi}^2(Y, Z) = 0,9473 \quad \text{oraz} \quad \rho_{multi}^2(Z, Y) = 0,9898.$$

Powyższy rezultat potwierdza, że wykorzystywany w niniejszej pracy współczynnik korelacji (jako miara zależności wektorów losowych) nie ma własności symetrii. Nie należy w tym upatrywać niczego zaskakującego: pozostając we wzajemnym związku, wektory Y i Z odgrywają w nim różne role, a zatem siła zależności Y od Z nie musi być (i na ogół nie jest) taka sama jak siła zależności Z od Y . Uwaga ta dotyczy także zmiennych losowych jednowymiarowych: jeżeli np. zysk danego przedsiębiorstwa w jakimś stopniu zależy od ogólnego poziomu rynkowych stóp procentowych, to przecież wysokość stóp procentowych zapewne w innym stopniu (jeżeli w ogóle) zależy od zysku tego konkretnego przedsiębiorstwa. W rozważanym w artykule przypadku okazuje się, że obie zależności (Y od Z oraz Z od Y) są bardzo silne, co – biorąc pod uwagę składowe oraz strukturę tych wektorów Y i Z – nie jest zaskoczeniem. Nie zaskakuje również, że w tym przypadku zależność między Z a Y jest wyraźnie silniejsza niż zależność między Y a Z .

3.5. Estymatory wybranych charakterystyk rozkładu czterowymiarowego wektora rentowności

Niech w dalszym ciągu $X = (X_1, X_2, X_3, X_4)$ będzie wektorem ustalonym na początku punktu 3.2, czyli wektorem czterech wybranych indeksów notowanych na WGPW. Określmy ponadto wektor $U = (U_1, U_2, U_3, U_4)$, gdzie U_1 jest dobową (lepiej: jednosesyjną) rentownością indeksu WIG, U_2 – dobową rentownością indeksu WIG-20, U_3 – dobową rentownością indeksu WIG-Banki, a U_4 – dobową rentownością wskaźnika WIG-Paliwa. Rozważamy zatem czterowymiarowy wektor rentowności (stóp zwrotu) wybranych indeksów. Ponieważ mówimy o klasycznie rozumianych rentownościach jednosesyjnych, więc historyczna próba, którą wykorzystamy do dalszych obliczeń, jest próbą 379-elementową.

Postępując podobnie jak dotychczas oraz korzystając z odpowiednich formuł, otrzymujemy podane poniżej wyniki.

1. Macierz kowariancji:

$$\Sigma_U = \begin{bmatrix} 0,000086 & 0,000103 & 0,000118 & 0,000091 \\ 0,000103 & 0,000129 & 0,000145 & 0,000114 \\ 0,000118 & 0,000145 & 0,000205 & 0,000093 \\ 0,000091 & 0,000114 & 0,000093 & 0,000222 \end{bmatrix}.$$

2. Macierz współczynników korelacji cząstkowej:

$$R_U = \begin{bmatrix} 1 & 0,9825 & 0,8892 & 0,587 \\ 0,9825 & 1 & 0,8937 & 0,6765 \\ 0,8892 & 0,8937 & 1 & 0,4381 \\ 0,6587 & 0,6765 & 0,4381 & 1 \end{bmatrix}.$$

3. Charakterystyki rozkładów brzegowych wektora U (tabela 4).

Tabela 4. Estymatory charakterystyk rozkładów brzegowych

Charakterystyka	rent.WIG	rent.WIG-20	rent.WIG- -Banki	rent.WIG- -Paliwa
Wartość oczekiwana	0,000818	0,000706	0,000659	0,001463
Wariancja	0,000086	0,000129	0,000205	0,000222
Odchylenie standardowe	0,009274	0,026571	0,014318	0,014900
Współczynnik asymetrii	-0,2429	-0,0012	0,3592	0,0646
Kurtoza	4,6668	3,8779	4,9574	3,5143
Współczynnik ekscesu	1,6668	0,8779	1,9574	0,5143

Źródło: opracowanie własne.

4. Charakterystyki łączne wektora U (tabela 5).Tabela 5. Estymatory charakterystyk łącznych rozkładu wektora losowego U (opartych na definicji potęgi wektora)

Charakterystyka	Wartość
EU	(0,000818; 0,000706; 0,000659; 0,001463)
D^2U	0,000641
Odchylenie standardowe	0,025318
$\gamma_{1,n}(U)$	(-0,002488; 0,023501; 0,072290; 0,004629)
Norma (długość) $\gamma_{1,n}$	0,076197
$\beta_{1,n}(U)$	0,005806
$\beta_{2,n}(U)$	3,0710
$\gamma_{2,n}(U)$	0,7748

Źródło: opracowanie własne.

Interpretacja uzyskanych wyników

1. Rozważając odrębnie rozkłady brzegowe wektora U zauważamy, że największą asymetrią (prawostronną) charakteryzuje się zmienna losowa „rent.WIG-Banki” (0,3592), następnie zmienna „rent.WIG” (−0,2429), a najmniejszą asymetrią brzegową wykazuje zmienna „rent.WIG-20” (−0,0016). Z kolei asymetria łączna (wektorowa) wskazuje, że wprawdzie wektor U charakteryzuje się największą skośnością także na kierunku „rent.WIG-Banki” (0,0723), ale kolejnym jest kierunek „rent.WIG-20” ze skośnością równą 0,0235.

2. Kurtoza łączna wektora U wynosi 3,071 i jest zdecydowanie mniejsza niż kurtoza każdego z rozkładów brzegowych. Oznacza to, że łącznie rozpatrywany wektor rentowności indeksów giełdowych nie wykazuje aż tak dużego „spłaszczenia” jak każdy z jego rozkładów brzegowych.

3. Miara odbiegania rozkładu łącznego wektora rentowności od rozkładu normalnego (współczynnik ekscesu) wynosi 0,7748 i jest znacznie niższa od ekscesu wyznaczonego w punkcie 3.2 dla wektora wartości nominalnych wybranych indeksów. Potwierdza to słuszność podejścia stosowanego w badaniu jednowymiarowych cen aktywów oraz ich rentowności, zgodnie z którym o ile często uzasadnione jest założenie o normalności rozkładu stóp zwrotu, o tyle na ogół błędne byłoby analogiczne przypuszczenie w odniesieniu do cen nominalnych rozważanych aktywów.

3.6. Estymatory wybranych charakterystyk rozkładu dwuwymiarowego wektora rentowności

Niech $U = (U_1, U_2, U_3, U_4)$ będzie nadal wektorem rentowności wybranych indeksów GPW określonym w punkcie 3.5. Rozważmy teraz wektor $V = (U_1, U_2)$, czyli $V = (\text{rent.WIG}, \text{rent.WIG-20})$. Wykorzystując w dalszym ciągu 279-elementowy ciąg danych historycznych, otrzymujemy podane poniżej estymatory wybranych charakterystyk.

1. Macierz kowariancji:

$$\Sigma_V = \begin{bmatrix} 0,00086 & 0,000103 \\ 0,000103 & 0,000129 \end{bmatrix}.$$

2. Macierz współczynników korelacji cząstkowej:

$$R_V = \begin{bmatrix} 1 & 0,9825 \\ 0,9825 & 1 \end{bmatrix}.$$

3. Charakterystyki rozkładów brzegowych wektora V ; wartości ich estymatorów zostały obliczone w punkcie 3.5 i podane są w tabeli 4.

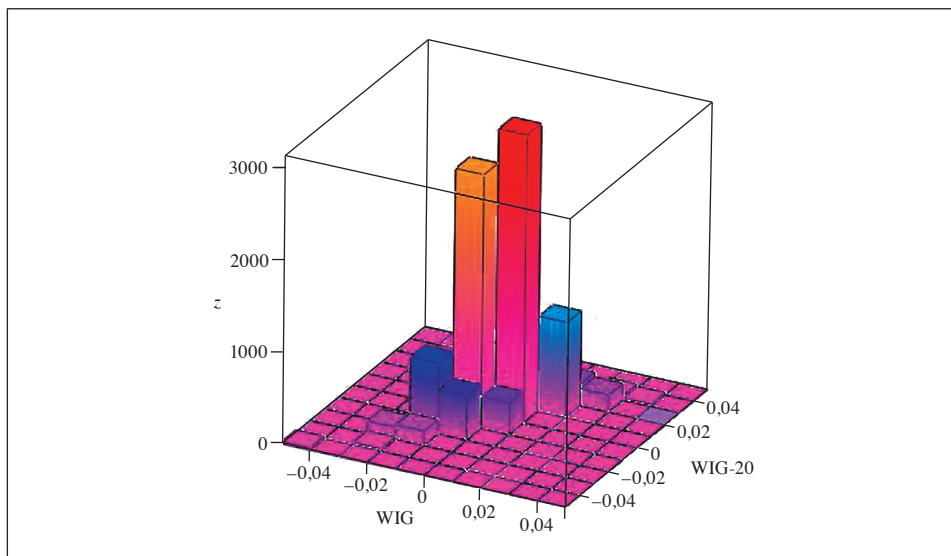
4. Charakterystyki łączne wektora V (tabela 6).Tabela 6. Estymatory charakterystyk łącznych rozkładu wektora losowego V (opartych na definicji potęgi wektora)

Charakterystyka	Wartość
EV	(0,000818; 0,000706)
D^2V	0,000214
Odchylenie standardowe	0,014629
$\gamma_{1,n}(U)$	(-0,088765; -0,047867)
Norma (długość) $\gamma_{1,n}$	0,100849
$\beta_{1,n}(V)$	0,010171
$\beta_{2,n}(V)$	4,1276
$\gamma_{2,n}(V)$	1,1609

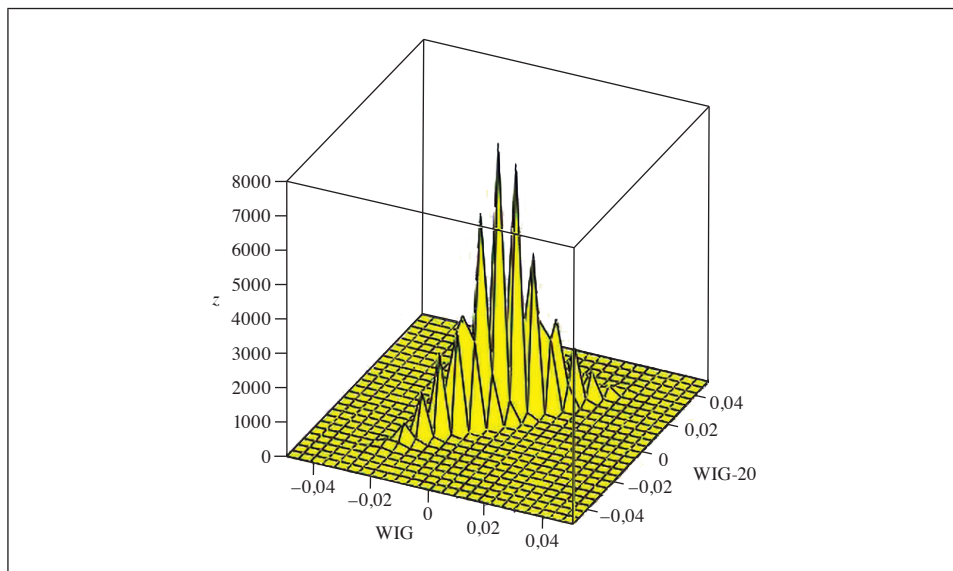
Źródło: opracowanie własne.

Interpretacja uzyskanych wyników

1. Kolejny raz zauważamy, że łączny wektor asymetrii $\gamma_{1,n}(V)$ wskazuje na zupełnie inny jej kierunek niż ten, który mógłby być odczytany ze współczynników asymetrii rozkładów brzegowych. Także poziom (wielkość) asymetrii łącznej jest różny od poziomu asymetrii rozkładów brzegowych.

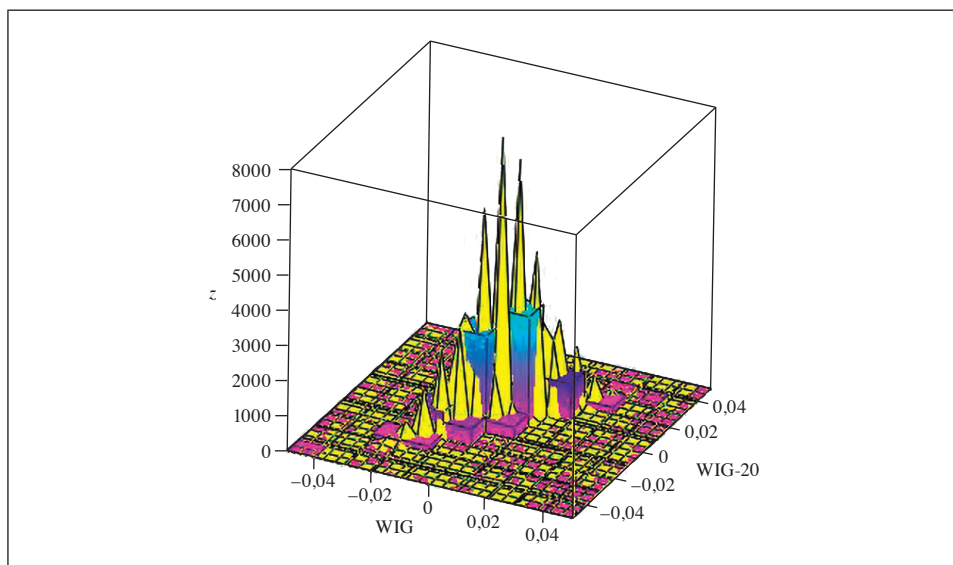
Rys. 4. Histogram rozkładu wektora V

Źródło: opracowanie własne.



Rys. 5. Dwuwymiarowy rozkład normalny o wektorze wartości oczekiwanych EV i macierzy kowariancji Σ_V

Źródło: opracowanie własne.



Rys. 6. Zestawienie histogramu oraz dwuwymiarowego rozkładu normalnego

Źródło: opracowanie własne.

2. Kurtoza oraz współczynnik ekscesu rozkładu wektora V sugerują, że znacznie bardziej odbiega on od rozkładu normalnego niż było to w przypadku rozważanego w punkcie 3.5 wektora U . Można to także dostrzec na rys. 4–6 – zauważamy, że histogram rozkładu wektora V jest dużo bardziej „spłaszczony” niż rozkład normalny o tej samej wartości oczekiwanej oraz wariancji.

3.7. Estymatory wielowymiarowych współczynników korelacji

W ostatnim punkcie prowadzonej analizy przyjmijmy, że $V = (U_1, U_2)$ oraz $W = (U_3, U_4)$, gdzie U_1, U_2, U_3, U_4 mają znaczenie określone w punkcie 3.5. Rozważamy zatem dwa dwuwymiarowe wektory:

$$V = (\text{rent.WIG}, \text{rent.WIG}-20) \text{ oraz } W = (\text{rent.WIG-Banki}, \text{rent.WIG-Paliwa}).$$

Obliczając estymatory kwadratu współczynników korelacji wielowymiarowej (korzystając ze wzoru (9)), otrzymujemy:

$$\rho_{multi}^2(V, W) = 0,6249 \text{ oraz } \rho_{multi}^2(W, V) = 0,8919.$$

Komentarz do uzyskanych wyników jest analogiczny do sformułowanego na zakończenie punktu 3.4.

4. Uwagi końcowe i perspektywy badawcze

Zaprezentowane w artykule przykłady wykorzystania łącznych charakterystyk wielowymiarowych rozkładów prawdopodobieństwa pozwalają na sformułowanie wniosku, że odmienne od klasycznego podejście do analizy rozkładów wektorów losowych pozwala dokładniej opisać badaną rzeczywistość niż badanie ich jednowymiarowych wektorów losowych. Odnosi się to m.in. do takich ich charakterystyk, jak asymetria, kurtoza czy współczynnik ekscesu. W szczególności eksces rozkładu wielowymiarowego może być ważną przesłanką formułowania hipotezy o jego normalności. Oczywiście, taka hipoteza musiałaby zostać poddana weryfikacji, do tego zaś potrzebne będą odpowiednie testy statystyczne. Konstrukcja takich testów jawi się jako kolejne zadanie w procesie rozwijania koncepcji nowej charakteryzacji wielowymiarowych rozkładów prawdopodobieństwa. Zadanie to nie musi być łatwe, ale – zdaniem autorów – dogodnym punktem wyjścia do jego realizacji mogą być uzyskane już wcześniej postaci estymatorów szeregu charakterystyk. Estymatory te z powodzeniem zostały wykorzystane także w niniejszym opracowaniu.

Literatura

- Bilodeau M., Brenner D. (1999), *Theory of Multivariate Statistics*, Springer-Verlag, New York.
- Budny K. (2009), *Kurtoza wektora losowego*, „Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu”, nr 78, seria: Ekonometria, nr 26.
- Budny K. (2012), *Kurtoza wektora losowego o wielowymiarowym rozkładzie normalnym (w:) Zastosowanie metod ilościowych w finansach i ubezpieczeniach*, red. S. Folrlicz, CeDeWu, Warszawa.
- Budny K. (2014a), *Estymacja momentów zwykłych wektora losowego opartych na definicji potęgi wektora*, „Folia Oeconomica Cracoviensia”, vol. 55.
- Budny K. (2014b), *Współczynnik ekscesu wektora losowego*, „Studia Ekonomiczne. Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach”, nr 203.
- Budny K. (2017), *Estimation of the Central Moments of a Random Vector Based on the Definition of the Power of a Vector*, „Statistics in Transition – New Series”, vol. 18, nr 1, <https://doi.org/10.21307/stattrans-2016-061>.
- Budny K. (2018), *Nowe charakterystyki rozkładu i zależności wektorów losowych – konstrukcja, estymacja, zastosowania*, Monografie: Prace Doktorskie, Wydawnictwo Uniwersytetu Ekonomicznego w Krakowie, Kraków.
- Budny K., Tatar J. (2009), *Kurtosis of a Random Vector – Special Types of Distributions*, „Statistics in Transition – New Series”, vol. 10, nr 3.
- Feller W. (1969), *Wstęp do rachunku prawdopodobieństwa*, t. 2, PWN, Warszawa.
- Shao J. (2003), *Mathematical Statistics*, 2nd ed., Springer, New York.
- Tatar J. (1996), *O niektórych miarach rozproszenia rozkładów prawdopodobieństwa*, „Przegląd Statystyczny”, vol. 43, nr 3–4.
- Tatar J. (1999), *Moments of a Random Variable in a Hilbert Space*, „Przegląd Statystyczny”, vol. 46, nr 2.
- Tatar J. (2000), *Asymetria wielowymiarowych rozkładów prawdopodobieństwa*, Materiały z XXXV Konferencji Statystyków, Ekonometryków i Matematyków Akademii Ekonomicznych Polski Południowej zorganizowanej przez Katedrę Statystyki Akademii Ekonomicznej w Krakowie (Osieczany, 23–25 marca 1999 r.), Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków.
- Tatar J. (2002), *Nierówność Lapunowa dla wielowymiarowych rozkładów prawdopodobieństwa*, „Zeszyty Naukowe Akademii Ekonomicznej w Krakowie”, nr 549.

Multivariate Statistical Analysis in the Study of Capital Markets

(Abstract)

In their previous work, the authors have presented an approach to describing and researching multivariate probability distributions that departs from the standard. In this paper, the new tools that resulted have been used to research and analyse selected two-, three- and fourth-dimension random vectors that have appeared on the Polish capital market. Coordinates of these vectors are stock market indices: WIG, WIG-20, WIG-Banks, WIG-Fuels and profitability of these indices. Using market data for the period 4/1/16–7/7/17, the following estimators of parameters of analysed distributions

were calculated and interpreted: expected value, total variance, total standard deviation, skewness coefficient, norm of the skewness coefficient, square of the skewness coefficient, kurtosis and excess coefficient. In order to overview and compare, for each vector a covariance matrix and a matrix of correlation coefficients are indicated. The following characteristics of marginal distributions are also used: expected value, variance, standard deviation, skewness coefficient, kurtosis and excess coefficient. For each pair of financial random vectors researched, the estimator of the square of correlation coefficient was also calculated, as that is one of the possible measures of their dependence.

Keywords: parameters of the probability distribution, multivariate random vector, estimator, capital market, stock index, profitability.