

Dominik Krężolek

Uniwersytet Ekonomiczny w Katowicach

MIARY ZALEŻNOŚCI – ANALIZA STATYSTYCZNA NA PRZYKŁADZIE WYBRANYCH WALORÓW RYNKU METALI NIEŻELAZNYCH

Wprowadzenie

Szeregi czasowe obserwowane na rynkach kapitałowych oraz towarowych cechują się wysokim poziomem zmienności w czasie, a także wykazują charakter nieliniowy. W przypadku zwiększania się stopnia zmienności cen aktywów na rynku oraz gdy ich stopy zwrotu generują wartości ujemne można obserwować właśnie jednokierunkowe zmiany cen analizowanych walorów. Związki tego rodzaju występują również w sytuacji pojawienia się obserwacji w szeregu czasowym, które w istotnym stopniu odchylają się od wartości oczekiwanej. Dodatkowo stopy zwrotu cechują się asymetrią, wysokim wskaźnikiem leptokurtozy¹ oraz brakiem normalności rozkładu².

Wspomniane charakterystyki, stanowiące istotny punkt przy wyborze modelu opisującego charakter szeregu, powodują, iż współczynnik korelacji liniowej Pearsona, mierzący liniowy związek pomiędzy dwiema zmiennymi losowymi, nie jest do końca odpowiednią miarą zależności. Problem ten może zostać rozwiązany poprzez zastosowanie alternatywnych mierników, takich jak *tau* Kendalla czy *rho* Spearmana³. Ponadto ważnym zagadnieniem jest analiza zależności występujących w ogonach rozkładów prawdopodobieństwa analizowanych stóp zwrotu, co jest szczególnie istotne z punktu widzenia zarządzania ryzykiem inwestycyjnym.

¹ E.F. Fama: *The Behavior of Stock Market Prices*. „Journal of Business” 1965, Vol. 38, No. 1, s. 34-105.

² B. Mandelbrot: *The Variation of Certain Speculative Prices*. „Journal of Business” 1963, Vol. 36, No. 4, s. 394-419.

³ Określany także w literaturze jako współczynnik korelacji rang [przyp. aut.].

1. Korelacje liniowe

Współczynnik korelacji liniowej jest jedną z najpowszechniej wykorzystywanych miar zależności zarówno w analizie rynków kapitałowych, towarowych, walutowych, jak i w ubezpieczeniach⁴. Mając dwie zmienne losowe X oraz Y , współczynnik korelacji liniowej jest określony jako:

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{D(X)D(Y)}. \quad (1.1)$$

Powyższy wzór zachodzi pod warunkiem, że wariancje zmiennych losowych X oraz Y istnieją. Wyrażenie $\text{cov}(X, Y)$ określa kowariancję pomiędzy zmiennymi losowymi X oraz Y . Współczynnik $\rho(X, Y)$ jest określony jako współczynnik korelacji liniowej, ponieważ jego znajomość jest równoważna znajomości współczynnika β równania regresji liniowej $Y = \beta X + \varepsilon$, gdzie ε jest resztą, która jest liniowo nieskorelowana z X . Zachodzi zatem:

$$\rho = \beta \frac{D(X)}{D(Y)}. \quad (1.2)$$

Zmienne losowe cechujące się wartościami indeksu ogona⁵ poniżej dwóch nie posiadają skończonej wariancji, stąd współczynnik korelacji nie może być zastosowany. Natomiast gdy istnieje czwarty moment centralny, wtedy współczynnik korelacji istnieje, ale jego estymator, oparty na próbie o rozmiarze T $\{(X_i, Y_i)\}_{i=1}^T$, dany wzorem:

$$\hat{\rho}_T = \frac{\frac{1}{T} \sum_{i=1}^T (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\frac{1}{T} \sum_{i=1}^T (X_i - \bar{X})^2 \frac{1}{T} \sum_{i=1}^T (Y_i - \bar{Y})^2}}, \quad (1.3)$$

gdzie $\bar{X}$ oraz $\bar{Y}$ oznaczają odpowiednio wartości średnie zmiennych losowych X oraz Y , nie jest do końca odpowiedni, ponieważ jego asymptotyczny rozkład jest rozkładem Lévy'ego, a nie Gaussa. Zatem współczynnik korelacji z próby może wykazywać istotne odchylenia od swojej rzeczywistej wartości,

⁴ Y. Malevergne, D. Sornette: *Extreme Financial Risk. From Dependence to Risk Management*. Springer-Verlag, Berlin, Heidelberg 2006.

⁵ Określany w literaturze jako parametr kształtu, indeks stabilności; miara grubości ogona rozkładu zmiennej losowej [przyp. aut.].

dostarczając niewłaściwych oszacowań. Jest to duży problem w modelowaniu danych finansowych⁶.

Rozważając dwie niezależne zmienne losowe łatwo wskazać, że współczynnik korelacji pomiędzy nimi wynosi zero. Niemniej jednak zależność odwrotna nie istnieje. Zatem mając zmienną losową ω posiadającą rozkład jednostajny na przedziale $\langle 0, 2\pi \rangle$, można zdefiniować parę zmiennych losowych U oraz V następująco⁷:

$$(U, V) = (\cos \omega, \sin \omega). \quad (1.4)$$

Przyjmując określenie dane wzorem 1.4, zachodzi, że $\rho(U, V) = 0$, nawet jeśli dwie zmienne losowe U oraz V nie są niezależne.

Bardziej wyrazista jest sytuacja, gdy znajomość jednej zmiennej losowej całkowicie określa drugą zmienną losową. Dla przykładu zmienna losowa U ma rozkład jednostajny na przedziale $\langle 0, 1 \rangle$, natomiast zmienna losowa V jest zdefiniowana jako:

$$\begin{cases} V = \frac{U}{\theta}, & U \in \langle 0, \theta \rangle \\ V = \frac{1-U}{1-\theta}, & U \in \langle \theta, 1 \rangle \end{cases}, \quad (1.5)$$

dla pewnego $\theta \in \langle 0, 1 \rangle$. Można łatwo wykazać, że zmienna losowa V również ma rozkład jednostajny na przedziale $\langle 0, 1 \rangle$ oraz że zachodzi:

$$\rho(U, V) = 2\theta - 1. \quad (1.6)$$

Zatem U oraz V są nieskorelowane dla $\theta = 0,5$, podczas gdy zmienna losowa V może być jednoznacznie wyznaczona za pomocą zmiennej losowej U ⁸.

Jeśli dwie zmienne losowe X oraz Y są zależne w sposób liniowy za pomocą relacji:

$$Y = \alpha + \beta X, \quad (1.7)$$

⁶ M.M. Meerschaert, H. Scheffler: *Sample Cross-correlations for Moving Averages with Regularly Varying Tails*. „Journal of Time Series Analysis” 2001, Vol. 22, No. 4, s. 482.

⁷ Y. Malevergne, D. Sornette: Op. cit., s. 148.

⁸ Ibid., s. 149.

to współczynnik korelacji $\rho(X, Y) = \pm 1$ zależnie od tego, czy parametr β przyjmuje dodatnią lub ujemną wartość. W tym przypadku zachodzi także relacja odwrotna. Wynika to z określenia:

$$\rho(X, Y)^2 = 1 - \min_{\alpha, \beta} \frac{E[(Y - (\alpha + \beta \cdot X))^2]}{D^2(Y)}, \quad (1.8)$$

gdzie $E[\cdot]$ oznacza wartość oczekiwaną łącznego rozkładu zmiennych losowych X oraz Y , natomiast $\rho(X, Y)^2$ jest określone jako współczynnik determinacji i wskazuje procent wariancji zmiennej losowej Y , która jest tłumaczona poprzez zmienną losową X .

Za pomocą nierówności Cauchy'ego-Schwartz'a wynika, że we wzorze 2.1 współczynnik korelacji $\rho \in \langle -1, 1 \rangle$. Jednakże mając dane dwie zmienne losowe X oraz Y o ustalonych rozkładach brzegowych, posiadających funkcje dystrybuanty F_X oraz F_Y , nie zawsze współczynnik korelacji może osiągnąć swoje wartości graniczne ± 1 . Można wykazać, że dla dowolnego dwuwymiarowego rozkładu o dystrybuancie F zachodzi nierówność Frécheta-Hoeffdinga⁹:

$$\begin{cases} \max\{F_X(x) + F_Y(y) - 1, 0\} \leq F(x, y) \\ F(x, y) \leq \min\{F_X(x), F_Y(y)\} \end{cases}. \quad (1.9)$$

Tym samym, stosując tożsamość Hoeffdinga:

$$\rho(X, Y) = \iint [F(x, y) - F_X(x) \cdot F_Y(y)] dx dy, \quad (1.10)$$

można dowieść, że mając dane funkcje dystrybuanty F_X oraz F_Y , współczynnik korelacji ρ zawiera się pomiędzy $\rho_{\min}$ oraz $\rho_{\max}$, gdzie $\rho_{\min}$ jest osiągane dla przeciwnie monotonicznych zmiennych losowych X oraz Y , natomiast $\rho_{\max}$ jest osiągane dla współmonotonicznych zmiennych losowych X oraz Y .

Dla przykładu¹⁰, mając dane dwie zmienne losowe o logarytmiczno-normalnych rozkładach brzegowych: $X \sim \log N(0, 1)$ oraz $Y \sim \log N(0, \sigma)$,

⁹ J. Dhaene, M.J. Goovaerts: *Dependency of Risks and Stop-loss Order*. „Austin Bulletin” 1996, Vol. 26, No. 2, s. 204.

¹⁰ P. Embrechts, A.J. McNeil, D. Straumann: *Correlation and Dependence in Risk Management: Properties and Pitfalls*. W: *Risk Management: Value at Risk and Beyond*. Ed. M.A.H. Dempster. Cambridge 2002, s. 176-223.

dolna oraz górna granica dla współczynnika korelacji liniowej $\rho(X, Y)$ są dane odpowiednio wzorami:

$$\rho_{\min} = \rho(e^Z, e^{-\sigma Z}) \text{ oraz } \rho_{\max} = \rho(e^Z, e^{\sigma Z}), \quad (1.11)$$

gdzie Z jest zmienną losową o rozkładzie normalnym standardowym. Bezpośrednie obliczenia z wykorzystaniem związku¹¹:

$$E[e^{\alpha Z}] = e^{\frac{\alpha^2}{2}}, \quad (1.12)$$

umożliwiają wyznaczenie wartości granicznych jak poniżej:

$$\begin{cases} \rho_{\min} = \frac{e^{-\sigma} - 1}{\sqrt{(e-1)(e^{\sigma^2} - 1)}} \\ \rho_{\max} = \frac{e^{\sigma} - 1}{\sqrt{(e-1)(e^{\sigma^2} - 1)}} \end{cases}. \quad (1.13)$$

Współczynnik korelacji jest niezmienniczy ze względu na rosnące przekształcenie afiniczne zmiennych w postaci:

$$X' = a \cdot X + b, \quad a > 0, \quad (1.14)$$

$$Y' = c \cdot Y + d, \quad c > 0, \quad (1.15)$$

ponieważ $\rho(X', Y') = \rho(X, Y)$.

Niemniej jednak ta własność nie może zostać uogólniona na jakiekolwiek (nieliniowe) rosnące przekształcenie. Ten brak niezmienniczości w związku z nieliniowymi zmianami zmiennych losowych wynika z faktu, że współczynnik korelacji agreguje informacje z rozkładów brzegowych obu zmiennych losowych oraz rzeczywistą strukturę zależności, którą można analizować z wykorzystaniem funkcji kopuli.

2. Korelacje lokalne

Zamiast koncentrować się na definiowaniu korelacji w kontekście ogólnym, można także się skupić na lokalnej liniowej zależności pomiędzy dwiema zmiennymi losowymi. Podejście to, zaproponowane przez Doksuma et al., po-

¹¹ Y. Malevergne, D. Sornette: Op. cit., s. 150.

zwala badać siłę korelacji jako funkcję wartości realizacji zmiennych losowych. Umożliwia to znalezienie odpowiedzi na pytanie, czy korelacja jest stała czy też ulega zmianie, gdy realizacje zmiennych losowych są typowe lub nie. Problem ten ma szczególne znaczenie w przypadku zjawiska zarażania się rynków.

Definiowanie lokalnego współczynnika korelacji jest sprawą naturalną. Należy wyjść z założenia, że w przypadku podejścia liniowego, jeśli dwie zmienne losowe X oraz Y są zależne poprzez relację¹²:

$$Y = \alpha + \beta X + \varepsilon, \quad (2.1)$$

gdzie składnik losowy ε jest niezależny (lub przynajmniej nieskorelowany) od zmiennej losowej X , wówczas współczynnik korelacji ma postać:

$$\rho = \frac{\beta \cdot \sigma_X}{\sqrt{\beta^2 \cdot \sigma_X^2 + \sigma_\varepsilon^2}}, \quad (2.2)$$

gdzie σ_X^2 oraz σ_ε^2 oznaczają odpowiednio wariancje zmiennej losowej X oraz składnika losowego ε .

Niemniej jednak można przyjąć bardziej ogólną relację pomiędzy zmiennymi losowymi X oraz Y ¹³:

$$Y = f(X) + \sigma(X) \cdot \varepsilon \quad (2.3)$$

oraz $\sigma_\varepsilon = 1$ i $f(\cdot)$ jest funkcją różniczkowalną. W sąsiedztwie $X = x_0$ można dokonać linearyzacji relacji określonej wzorem 2.3 jako:

$$Y = [f(x_0) - x_0 \cdot f'(x_0)] + f'(x_0)X + \sigma(x_0)\varepsilon \quad (2.4)$$

oraz, poprzez analogię do wzoru 2.2, zdefiniować lokalny współczynnik korelacji liniowej jako:

$$\rho(x_0) = \frac{f'(x_0) \cdot \sigma(x_0)}{\sqrt{f'(x_0)^2 \cdot \sigma(x_0)^2 + \sigma(x_0)^2}}. \quad (2.5)$$

Można bezpośrednio sprawdzić, czy lokalny współczynnik korelacji redukuje się do współczynnika korelacji liniowej, gdy $f(\cdot)$ jest odwzorowaniem afi-

¹² S. Bjerre, K. Doksum: *Correlation Curves: Measures of Association as Functions of Covariate Values*. „The Annals of Statistics” 1993, Vol. 21, No. 2, s. 892.

¹³ Ibid., s. 893.

nicznym oraz $\sigma(x)$ jest stałe. Dodatkowo lokalny współczynnik korelacji $\rho(x)$ posiada te same własności, co współczynnik korelacji liniowej ρ :

- $\rho(x) \in \langle -1, 1 \rangle$,
- $\rho(x)$ jest niezmienny ze względu na (rosnące) odwzorowanie liniowe zarówno zmiennej losowej X , jak i zmiennej losowej Y , $\rho(x) = 0$ dla każdego x , jeśli zmienne losowe X oraz Y są niezależne.

Ponadto, i jednocześnie w opozycji do współczynnika korelacji liniowej, lokalny współczynnik korelacji przyjmuje wartości graniczne tylko wtedy, gdy $\sigma(x)$ jest równe zero (znak zależy od pochodnej funkcji $f(\cdot)$), zatem gdy $Y = f(X)$. Tym samym lokalny współczynnik korelacji nie uwzględnia negatywnej właściwości współczynnika korelacji liniowej, iż zanikające wartości bliskie zero mogą zostać osiągnięte nawet wtedy, gdy zmienne losowe X oraz Y są deterministycznie zależne jedna od drugiej¹⁴.

3. Uogólniona korelacja pomiędzy $N > 2$ zmiennymi losowymi

Współczynnik korelacji ρ jest miarą liniowej zależności pomiędzy dwiema zmiennymi losowymi. Można przedstawić jego uogólnienie dla N zmiennych losowych.

Niech $\mathbf{X}(t)$ oznacza losowy wektor składający się z N składników, przykładowo wektor stóp zwrotu N aktywów finansowych, tworzących portfel inwestycyjny. W pierwszej kolejności należy oszacować wartości średnie składników wektora losowego $\mathbf{X}(t)$, a następnie odjąć te wartości od wektora $\mathbf{X}(t)$ dla $t = 1, \dots, L$, gdzie L oznacza rozmiar próby (równy przykładowo wybranej długości przedziału czasowego wykorzystanego do estymacji). Celem uproszczenia zapisu niech $\mathbf{X}(t)$ oznacza teraz wektor scentrowany. Estymator z próby macierzy kowariancji N zmiennych losowych na próbie o długości L można zapisać jako¹⁵:

$$\mathbf{S}_{\mathbf{X}(t)} = \frac{1}{L} \sum_{t=1}^L \mathbf{X}(t) \cdot \mathbf{X}(t)^T. \quad (3.1)$$

Dokonując podziału zbioru N składowych wektora $\mathbf{X}(t)$ na dwie części, otrzymano: skalar $X_i(t)$ utworzony z jednej ze składowych oraz $(N-1)$ -

¹⁴ Ibid., s. 896.

¹⁵ A.A. Lyubushin Jr.: *Robust Wavelet-aggregated Signal for Geophysical Monitoring Problems*. „Izvestiya, Physics of the Solid Earth” 2002, Vol. 38, No. 9, s. 5.

-wymiarowy wektor kolumnowy przedstawiony w postaci $\xi_i(t) = [X_1(t), \dots, X_{i-1}(t), X_{i+1}(t), \dots, X_N(t)]^T$ i stworzony z pozostałych składników. Mnożąc (iloczyn skalarny) każdy wektor ξ_i przez nieznany wektor $\boldsymbol{\varphi}$, otrzymano zbiór wartości skalarów $\zeta_i = \boldsymbol{\varphi}^T \cdot \xi_i$.

Zagadnienie polega na wyszukaniu wektora $\boldsymbol{\varphi}$, który tworzy kwadrat współczynnika korelacji pomiędzy dwoma zmiennymi losowymi X_i oraz ζ_i jako problem maksimum. Podejście to stanowi przykład zastosowania klasycznego rozwiązania zaproponowanego przez Hotellinga dotyczącego korelacji kanonicznej: wektor $\boldsymbol{\varphi}$ jest zdefiniowany jako wektor własny odpowiadający maksymalnej wartości własnej (która jest równa maksymalnemu współczynnikowi korelacji pomiędzy dwiema zmiennymi losowymi X_i oraz ζ_i) w postaci macierzy o wymiarach $(N-1) \times (N-1)$ ¹⁶:

$$\mathbf{S}_{\xi_i \xi_i}^{-1} \mathbf{S}_{\xi_i X_i} \mathbf{S}_{X_i X_i}^{-1} \mathbf{S}_{X_i \xi_i}, \quad (3.2)$$

gdzie:

$$\begin{aligned} \mathbf{S}_{X_i X_i} &= \text{cov}(X_i, X_i), \\ \mathbf{S}_{X_i \xi_i} &= \mathbf{S}_{\xi_i X_i}^T = \text{cov}(X_i, \xi_i^T), \\ \mathbf{S}_{\xi_i \xi_i} &= \text{cov}(\xi_i, \xi_i). \end{aligned} \quad (3.3)$$

Macierze we wzorach 3.2 oraz 3.3 są podmacierzami ogólnej $N \times N$ -wymiarowej macierzy kowariancji $\mathbf{S}_{XX} = \text{cov}(\mathbf{X}, \mathbf{X}^T)$, której oszacowanie jest dane wzorem 3.1. Zatem zastępując macierz $\mathbf{S}_{XX}$ oraz jej podmacierze we wzorach 3.2 oraz 3.3 przez jej oszacowanie dane wzorem 3.1, można wyznaczyć wektor $\boldsymbol{\varphi}$ oraz zbiór wartości skalarów ζ_i dla $i = 1, \dots, N$. Maksymalna wartość własna macierzy danej wzorem 3.2 jest określana jako współczynnik kanoniczny N -korelacji pomiędzy zmienną losową X_i i pozostałymi $(N-1)$ zmiennymi losowymi. Wykonując tę samą operację na wszystkich pozostałych składowych wektora $\mathbf{X}$, można otrzymać N -wymiarowy wektor współczynników kanonicznych N -korelacji równy największej wartości własnej macierzy określonej wzorem 3.2 dla $i = 1, \dots, N$. Dla $N = 2$ $(N-1)$ -wymiarowa macierz dana wzorem 3.2 redukuje się do kwadratu standardowego współczynnika korelacji pomiędzy dwiema zmiennymi losowymi.

¹⁶ Ibid., s. 5.

Odminną, ale równoważną formę zapisu uprzednio zaprezentowanych zależności przedstawiono poniżej. Niech dane będzie zagadnienie regresji zmiennej losowej X_i na $(N-1)$ -wymiarowym wektorze losowym określonym jako $\xi_i(t) = [X_1(t), \dots, X_{i-1}(t), X_{i+1}(t), \dots, X_N(t)]^T$, tj. rozwinięcie wektora ϕ współczynników regresji w postaci liniowej¹⁷:

$$X_i = \sum_{j \neq i} \phi_j X_j + \varepsilon_i = \phi^T \cdot \xi_i + \varepsilon_i, \quad (3.4)$$

gdzie ε_i reprezentuje składnik resztowy równania regresji. Jeśli wektor ϕ jest zdefiniowany poprzez metodę MNK, rozwiązując problem optymalizacyjny:

$$\sum_{i=1}^L (\phi^T \cdot \xi_i - X_i)^2 \rightarrow \min, \quad (3.5)$$

w odniesieniu do wektora ϕ , wówczas jego oszacowanie można w łatwy i jednoznaczny sposób uzyskać jako:

$$\hat{\phi} = S_{\xi_i \xi_i}^{-1} \cdot S_{\xi_i X_i}. \quad (3.6)$$

Niech $\hat{\xi}_i = \hat{\phi}^T \cdot \xi_i$ oznacza udział w regresji określonej wzorem 3.4 oszacowania zadanego wzorem 3.6. Ponieważ:

$$\text{cov}(X_i, \hat{\xi}_i) = \text{cov}(X_i, S_{\xi_i \xi_i}^{-1} \cdot S_{\xi_i X_i} \cdot \xi_i) = S_{X_i \xi_i} \cdot S_{\xi_i \xi_i}^{-1} \cdot S_{\xi_i X_i}, \quad (3.7)$$

wynika stąd, że współczynnik korelacji pomiędzy $\hat{\phi}$ a X_i jest równy skalarowi $S_{X_i \xi_i} \cdot S_{\xi_i \xi_i}^{-1} \cdot S_{\xi_i X_i} \cdot S_{X_i X_i}^{-1}$, czyli jest maksymalną wartością macierzy własnej określonej wzorem 3.2. Stąd kanoniczny współczynnik N -korelacji¹⁸:

$$\rho_N^i = S_{X_i \xi_i} \cdot S_{\xi_i \xi_i}^{-1} \cdot S_{\xi_i X_i} \cdot S_{X_i X_i}^{-1} \quad (3.8)$$

może być określony z rozwiązania problemu regresji danego wzorami 3.4 oraz 3.5. Zależność pomiędzy dwoma wskazanymi powyżej rozwiązaniami ma swoje korzenie w podobnej zależności pomiędzy korelacją liniową a współczynnikiem regresji liniowej.

Ponownie, kanoniczny współczynnik N -korelacji jest niezmienny ze względu na transformację liniową każdej ze zmiennych losowych X_i , jednakże nie pozostaje on niezmienny ze względu na nieliniowe, monotoniczne transformacje.

¹⁷ Ibid., s. 6.

¹⁸ Y. Malevergne, D. Sornette: Op. cit., s. 154.

4. Miary zgodności i podobieństwa: *tau* Kendalla oraz *rho* Spearmana

Fundamentalny problem związany z zarządzaniem ryzykiem na rynkach finansowych dotyczy zagadnienia jednoczesnych jednokierunkowych zmian cen analizowanych walorów. Jeżeli rozwiązanie tej kwestii jest twierdzące, wówczas dywersyfikacja ryzyka byłaby prawdopodobnie trudna, gdyż opiera się ona na założeniu, że spadek ceny jednego aktywu jest statystycznie zbilansowany wzrostem ceny innego. Naturalną drogą do określenia tendencji aktywów finansowych do wspólnych, jednokierunkowych zmian jest porównanie prawdopodobieństwa wspólnego wzrostu (lub spadku) ich ceny z prawdopodobieństwem, że cena jednego z tych dwóch aktywów wzrośnie (lub spadnie), podczas gdy cena drugiego odpowiednio spadnie (lub wzrośnie). Można to zapisać matematycznie w następujący sposób:

Zaczynając od dwóch niezależnych realizacji (X_1, Y_1) oraz (X_2, Y_2) pary zmiennych losowych (X, Y) , można wyznaczyć wielkość¹⁹:

$$\tau(X, Y) = \Pr[(X_1 - X_2) \cdot (Y_1 - Y_2) > 0] - \Pr[(X_1 - X_2) \cdot (Y_1 - Y_2) < 0]. \quad (4.1)$$

Rozważając prawą stronę wzoru 4.1, pierwsza jego część określa prawdopodobieństwo zgodności (konkordancji), tj. prawdopodobieństwo tego, że zmienne losowe X oraz Y zmieniają się w jednym kierunku (wzrost lub spadek). Z kolei druga część wzoru 4.1 określa prawdopodobieństwo niezgodności, tj. prawdopodobieństwo tego, że zmienne losowe X oraz Y zmieniają się w przeciwnym kierunku.

Wyrażenie określone wzorem 4.1 definiuje tzw. współczynnik *tau* Kendalla (dotyczący populacji). Wielkość ta jest niezmienna ze względu na przekształcenia rozkładów brzegowych. W istocie zakładając dowolne rosnące odwzorowanie G_X oraz G_Y , zachodzi²⁰:

$$X_1 \geq X_2 \Leftrightarrow G_X(X_1) \geq G_X(X_2), \quad (4.2)$$

$$Y_1 \geq Y_2 \Leftrightarrow G_Y(Y_1) \geq G_Y(Y_2). \quad (4.3)$$

Zakładając ciągłą postać zmiennych losowych X oraz Y , określenie *tau* Kendalla dane wzorem 4.1 można przedstawić jako:

$$\tau(X, Y) = 2P[(X_1 - X_2) \cdot (Y_1 - Y_2) > 0] - 1. \quad (4.4)$$

¹⁹ R. Doman: *Zastosowanie kopuli w modelowaniu dynamiki zależności na rynkach finansowych*. Wydawnictwo Uniwersytetu Ekonomicznego, Poznań 2011, s. 16.

²⁰ Y. Malevergne, D. Sornette: Op. cit., s. 155.

Współczynnik *tau* Kendalla przyjmuje wartości określone na przedziale $\langle -1, 1 \rangle$. Dolna granica określoności współczynnika jest osiągana wtedy i tylko wtedy, gdy zmienne losowe (X, Y) są przeciwnie monotoniczne, natomiast górna granica wtedy i tylko wtedy, gdy zmienne losowe (X, Y) są współmonotoniczne. Ponadto współczynnik *tau* Kendalla przyjmuje wartość zero dla niezależnych zmiennych losowych. Jednakże, jak w przypadku współczynnika korelacji (liniowej), współczynnik *tau* może przyjmować wartości oscylujące wokół zera dla zmiennych losowych niebędących zmiennymi niezależnymi.

Inna miara zgodności oraz podobieństwa zmiennych losowych została zaproponowana przez Spearmana. Niech dana będzie para niezależnych realizacji zmiennych losowych (X, Y) , tj. (X_1, Y_1) oraz (X_2, Y_2) . Tym samym współczynnik *rho* Spearmana jest określony następującym wzorem²¹:

$$\rho_s(X, Y) = 3[\Pr[(X - X_1) \cdot (Y - Y_2) > 0] - \Pr[(X - X_1) \cdot (Y - Y_2) < 0]]. \quad (4.5)$$

Dodatkowo, rozważając dwie zmienne losowe (X, Y) o rozkładach brzegowych określonych dystrybuantami F_X oraz F_Y , współczynnik *rho* Spearmana można zapisać za pomocą formuły:

$$\rho_s(X, Y) = \frac{\text{cov}(F_X, F_Y)}{\sqrt{D^2(F_X) \cdot D^2(F_Y)}}. \quad (4.6)$$

Pomiędzy współczynnikiem *tau* Kendalla a *rho* Spearmana zachodzi związek²²:

$$\frac{3\tau(X, Y) - 1}{2} \leq \rho_s(X, Y) \leq -\frac{\tau^2(X, Y) - 2\tau(X, Y) - 1}{2}, \quad \tau(X, Y) \geq 0, \quad (4.7)$$

$$\frac{\tau^2(X, Y) + 2\tau(X, Y) - 1}{2} \leq \rho_s(X, Y) \leq \frac{3\tau(X, Y) + 1}{2}, \quad \tau(X, Y) \leq 0. \quad (4.8)$$

Współczynnik *rho* Spearmana, podobnie jak *tau* Kendalla, przyjmuje wartości określone na przedziale $\langle -1, 1 \rangle$. Dolna granica określoności współczynnika jest osiągana wtedy i tylko wtedy, gdy zmienne losowe (X, Y) są przeciwnie monotoniczne, natomiast górna granica wtedy i tylko wtedy, gdy zmienne losowe (X, Y) są współmonotoniczne. Współczynnik *rho* Spearmana przyjmuje wartość zero dla niezależnych zmiennych losowych.

²¹ R. Doman: Op. cit., s. 16.

²² W.H. Kruskal: *Ordinal Measures of Association*. „Journal of the American Statistical Association” 1958, No. 53, s. 814-861.

5. Zależności w ogonach rozkładów

Jak wspomniano we wprowadzeniu, ważnym zagadnieniem w analizie ryzyka inwestycyjnego jest wykrywanie i badanie zależności występujących w ogonach rozkładów prawdopodobieństwa analizowanych stóp zwrotu. Problem polega na szacowaniu prawdopodobieństwa tego, że zmienne losowe X oraz Y będą jednocześnie przyjmować wartości istotnie oddalone od centralnej części ich rozkładu. Zagadnienie sprowadza się zatem do oszacowania prawdopodobieństwa tego, że zmienna losowa Y przekroczy kwantyl rzędu α pod warunkiem, że przekroczy go także zmienna losowa X . Tym samym można zdefiniować współczynniki określające zależności w górnym oraz dolnym ogonie rozkładu.

Zakładając, że zmienne losowe X oraz Y posiadają dystrybuanty brzegowe określone odpowiednio jako F_X oraz F_Y , można zdefiniować²³:

- współczynnik zależności w dolnym ogonie:

$$\lambda^L = \lim_{\alpha \rightarrow 0^+} P(Y \leq F_Y^{-1}(\alpha) | X \leq F_X^{-1}(\alpha)), \quad (5.1)$$

- współczynnik zależności w górnym ogonie:

$$\lambda^U = \lim_{\alpha \rightarrow 1^-} P(Y > F_Y^{-1}(\alpha) | X > F_X^{-1}(\alpha)), \quad (5.2)$$

gdzie F_X^{-1} oraz F_Y^{-1} oznaczają funkcje kwantylowe reprezentujące zmienne losowe X oraz Y . Powyższe wzory zachodzą przy założeniu istnienia rozważanych granic.

Przedstawione współczynniki zależności w dolnym i górnym ogonie rozkładu pozwalają określić prawdopodobieństwo jednoczesnego wystąpienia ekstremalnych realizacji zmiennych losowych X oraz Y . Jeżeli wartości współczynnika λ (niezależnie od ogona rozkładu) przyjmują wartości dodatnie, ekstremalne realizacje obu zmiennych losowych wystąpią równocześnie z prawdopodobieństwem warunkowym równym λ . Natomiast jeśli $\lambda = 0$, wówczas zmienne losowe X oraz Y nie wykazują zależności w ogonach rozkładów (są niezależne). Dodatkowo warto nadmienić, że kwantyle wyznaczone jako $F_X^{-1}(\alpha)$ oraz $F_Y^{-1}(\alpha)$ reprezentują wartość zagrożoną Value-at-Risk na poziomie istotności α (alternatywnie na poziomie tolerancji $(1 - \alpha)$)²⁴.

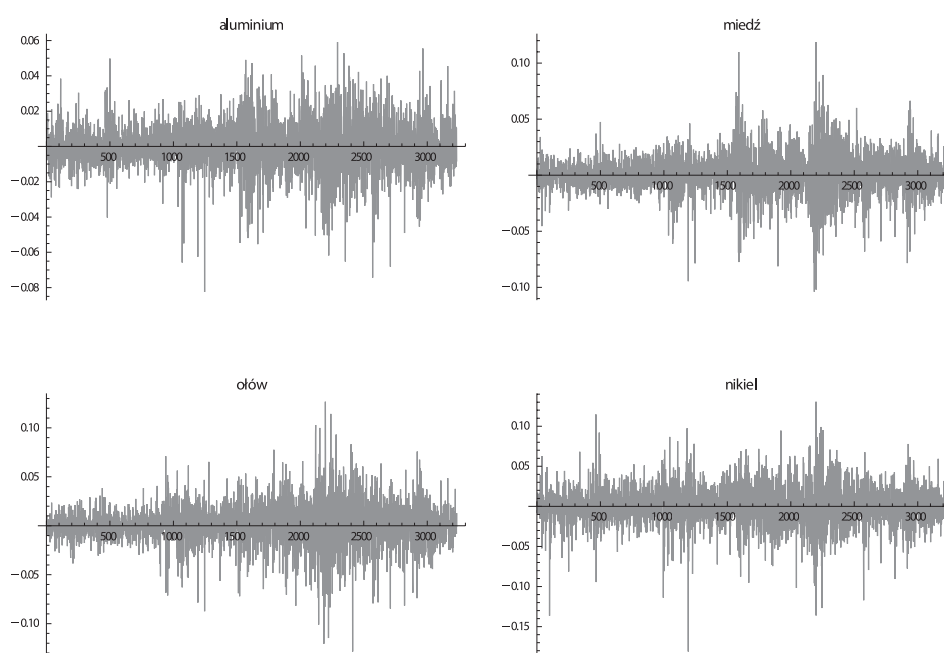
²³ R. Doman: Op. cit., s. 18.

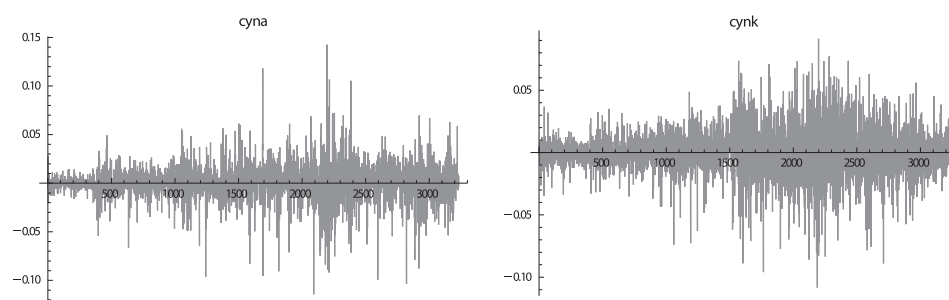
²⁴ Y. Malevergne, D. Sornette: Op. cit., s. 168.

6. Przykład empiryczny

Wybrane z omówionych w pracy mierników zależności zastosowano do opisu związku pomiędzy wybranymi walorami rynku metali nieżelaznych. Rynek ten jest szczególnie ważny z punktu widzenia gospodarki krajów rozwiniętych oraz wschodzących. Ogromne zróżnicowanie w gatunkach stali wynika z jej różnorodnego zastosowania: w przemyśle konstrukcyjnym, motoryzacyjnym, lotniczym (w tym projekty związane z eksploracją przestrzeni kosmicznej). Im bardziej specyficzne jest wykorzystanie stali, tym większe znaczenie ma jej jakość, na którą w ogromnym stopniu wpływają dodatki stopowe, które mają na celu uszlachetnienie jakościowe produktu finalnego.

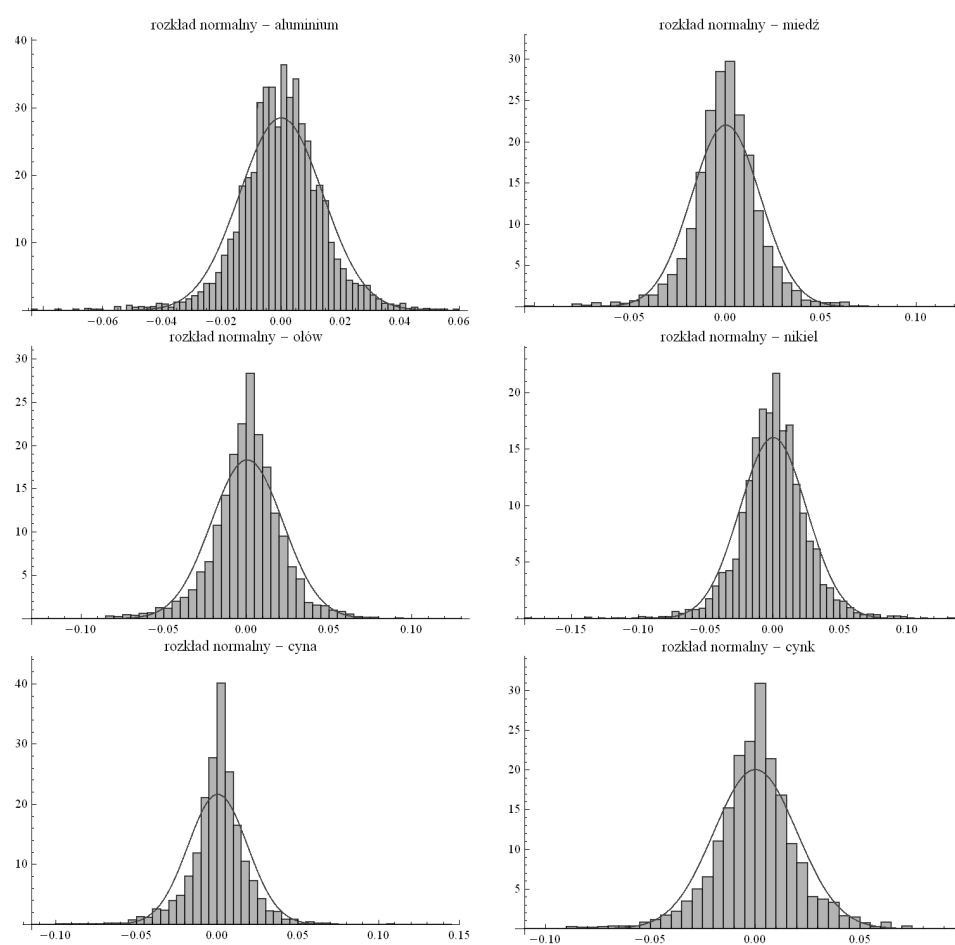
W opracowaniu rozważono dodatki stopowe notowane na giełdzie towarowej w Londynie (London Metal Exchange) w okresie styczeń 2000 – grudzień 2012. Przedmiotem badania objęto dzienne logarytmiczne stopy zwrotu następujących metali: aluminium, miedź, nikiel, ołów, cyna oraz cynk. Na poniższych wykresach przedstawiono empiryczne szeregi czasowe oraz histogramy wraz z dopasowanymi funkcjami gęstości rozkładu normalnego.





Rys. 1. Empiryczne szeregi czasowe stóp zwrotu analizowanych metali

Źródło: Obliczenia własne.



Rys. 2. Histogramy wraz z dopasowaną funkcją gęstości

Źródło: Obliczenia własne.

Szeregi czasowe stóp zwrotu analizowanych metali odznaczają się wysokim poziomem zmienności, a dodatkowo uwidaczniają zjawisko skupiania się danych. Dodatkowo empiryczne rozkłady okazują się być leptokurtyczne, co w połączeniu z wcześniej wspomnianymi cechami prowadzi do odrzucenia hipotezy o normalności rozkładu. Potwierdzają to testy Andersona-Darlinga (A-D) oraz Jarque’a-Bera (J-B).

Tabela 1

Parametry rozkładu normalnego

Zmienna	Średnia	Odchylenie standardowe
Aluminium	0,000072	0,013989
Miedź	0,000446	0,018106
Ołów	0,000481	0,021748
Nikiel	0,000224	0,024885
Cyna	0,000418	0,018434
Cynk	0,000160	0,019889

Źródło: Obliczenia własne.

Tabela 2

Testy zgodności dla rozkładu normalnego

Zmienna	Statystyka AD	p-value	Statystyka J-B	p-value
Aluminium	14,021	$3,95824 \times 10^{-7}$	835,658	0
Miedź	29,891	0	2184,381	0
Ołów	28,001	0	1529,642	0
Nikiel	17,399	$1,92084 \times 10^{-7}$	1488,511	0
Cyna	57,159	0	4825,693	0
Cynk	26,284	0	834,623	0

Źródło: Obliczenia własne.

Na podstawie realizacji stóp zwrotu wyznaczono współczynniki korelacji pomiędzy analizowanymi zmiennymi. Wyniki przedstawia tabela 3.

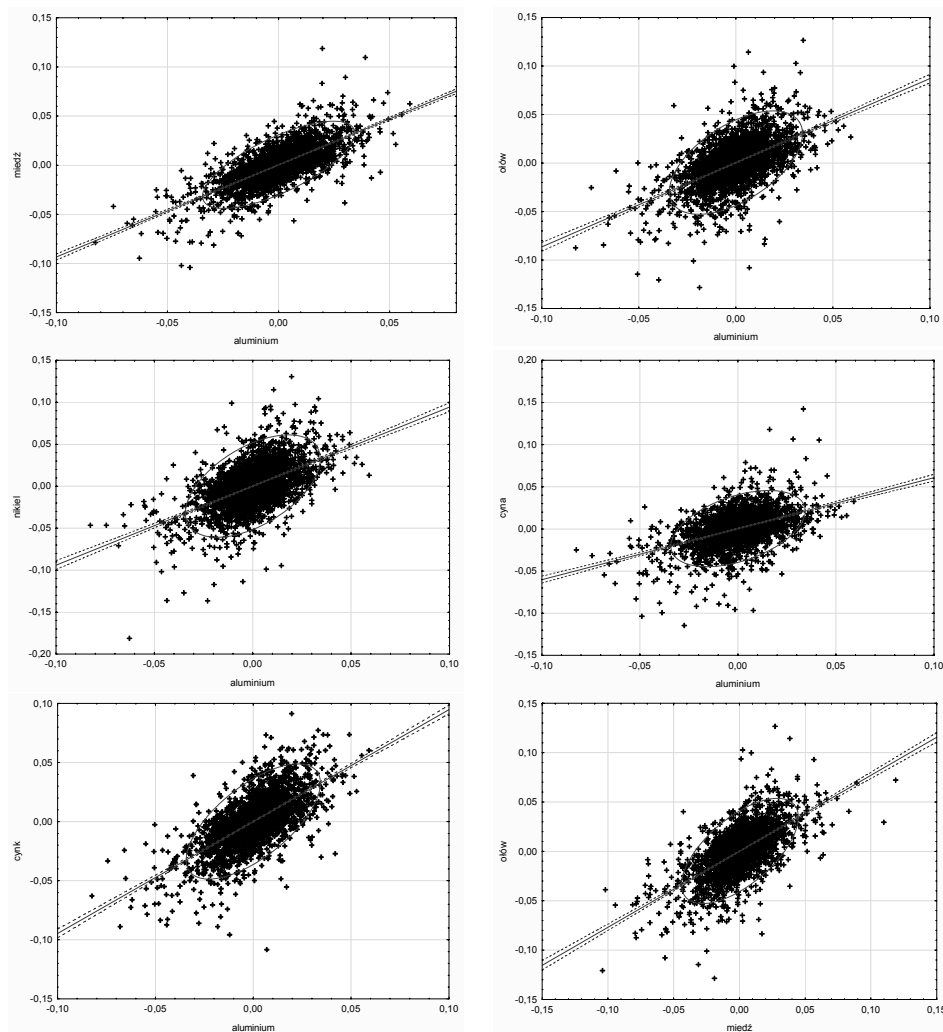
Tabela 3

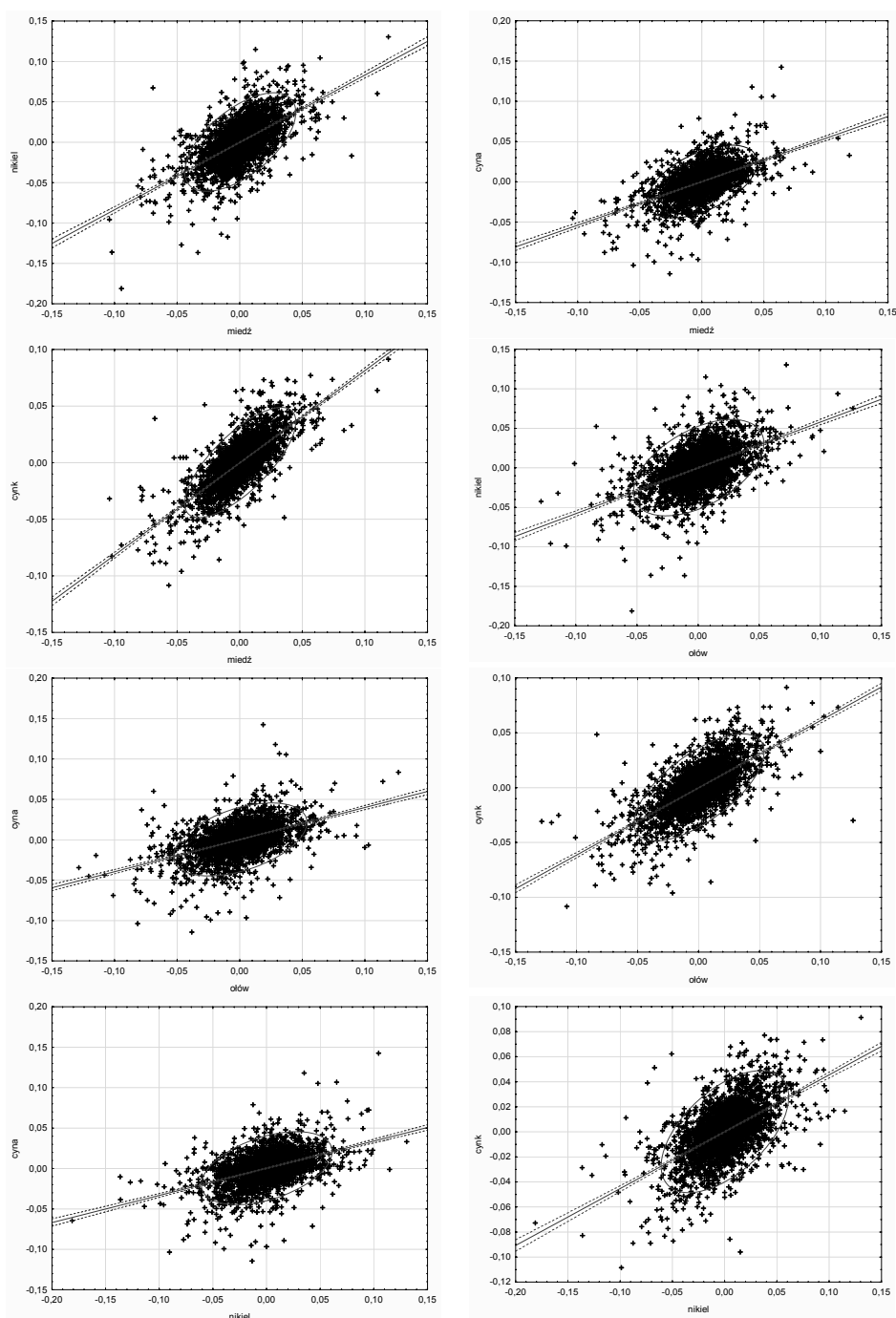
Współczynniki korelacji liniowej pomiędzy analizowanymi zmiennymi

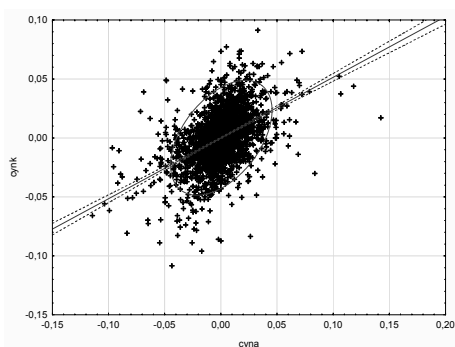
Współczynnik korelacji Pearsona	Aluminium	Miedź	Ołów	Nikiel	Cyna	Cynk
Aluminium	1,000000	0,723646	0,558070	0,529096	0,459054	0,666113
Miedź	0,723646	1,000000	0,640607	0,606764	0,530213	0,742348
Ołów	0,558070	0,640607	1,000000	0,506440	0,467962	0,670301
Nikiel	0,529096	0,606764	0,506440	1,000000	0,454081	0,567735
Cyna	0,459054	0,530213	0,467962	0,454081	1,000000	0,477203
Cynk	0,666113	0,742348	0,670301	0,567735	0,477203	1,000000

Źródło: Obliczenia własne.

Biorąc pod uwagę miarę zależności, jaką jest współczynnik korelacji liniowej Pearsona, najsilniejszy związek wykazano dla pary zmiennych miedź oraz cynk, natomiast najsłabszy dla pary zmiennych nikiel oraz cyna. Wszystkie współczynniki korelacji wykazywały statystyczną istotność na poziomie $\alpha = 0,05$. Wykresy rozrzutu pomiędzy poszczególnymi parami zmiennych zaprezentowano na rysunku 3.



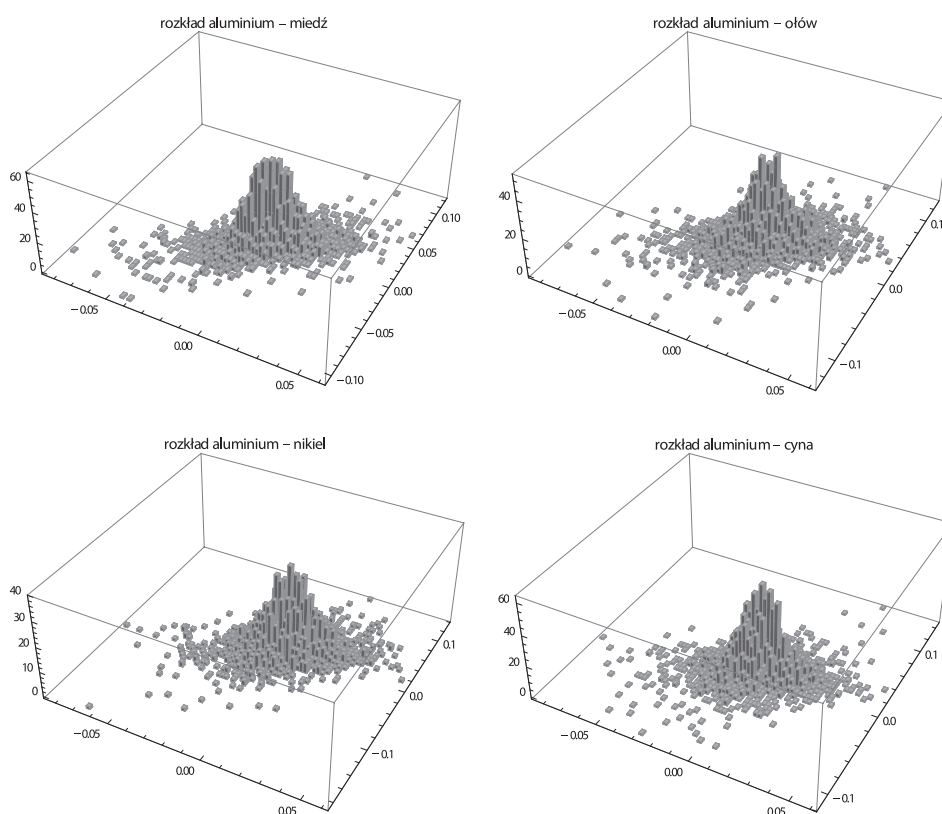




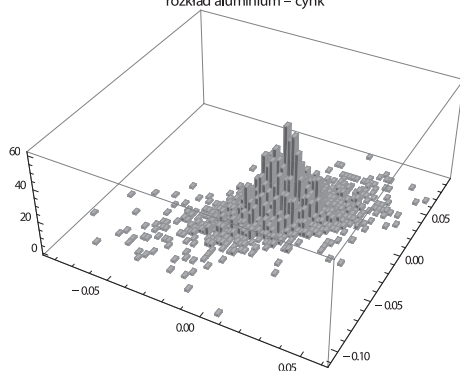
Rys. 3. Wykresy rozrzutu

Źródło: Obliczenia własne.

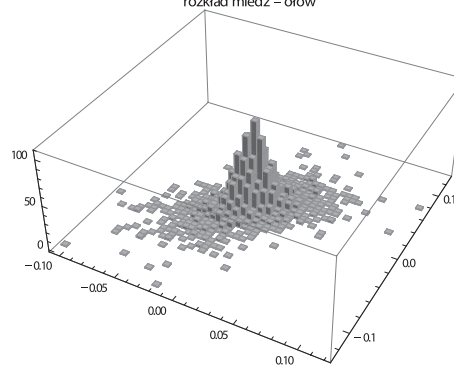
Na wykresach rozrzutu zaznaczono dodatkowo linię regresji wraz z 95% przedziałem ufności oraz elipsę rozkładu normalnego. Wydać tym samym występowanie obserwacji oddalonych. Ponadto na rysunku 4 przedstawiono dwuwymiarowe rozkłady empiryczne dla analizowanych par aktywów rynku metali.



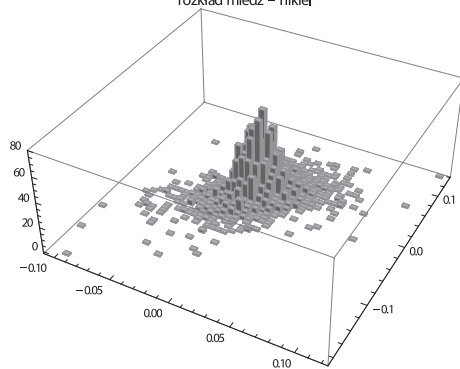
rozkład aluminium – cynk



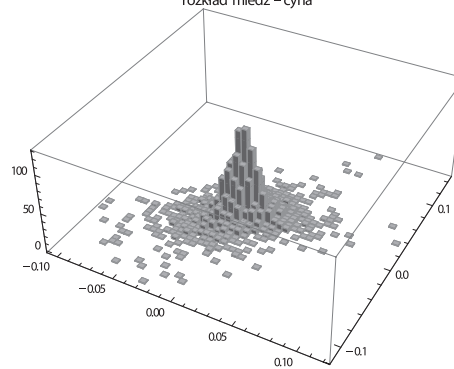
rozkład miedź – ołów



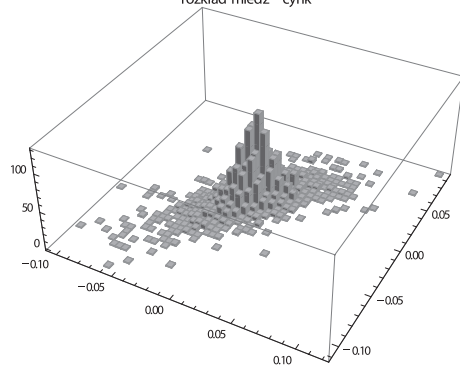
rozkład miedź – nikiel



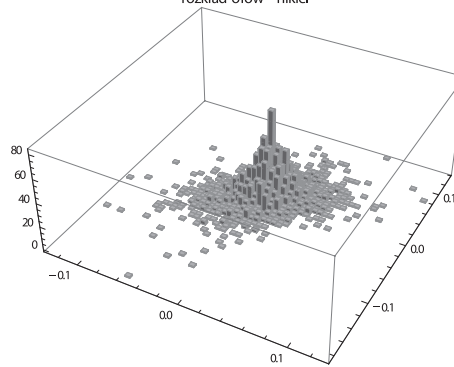
rozkład miedź – cyna

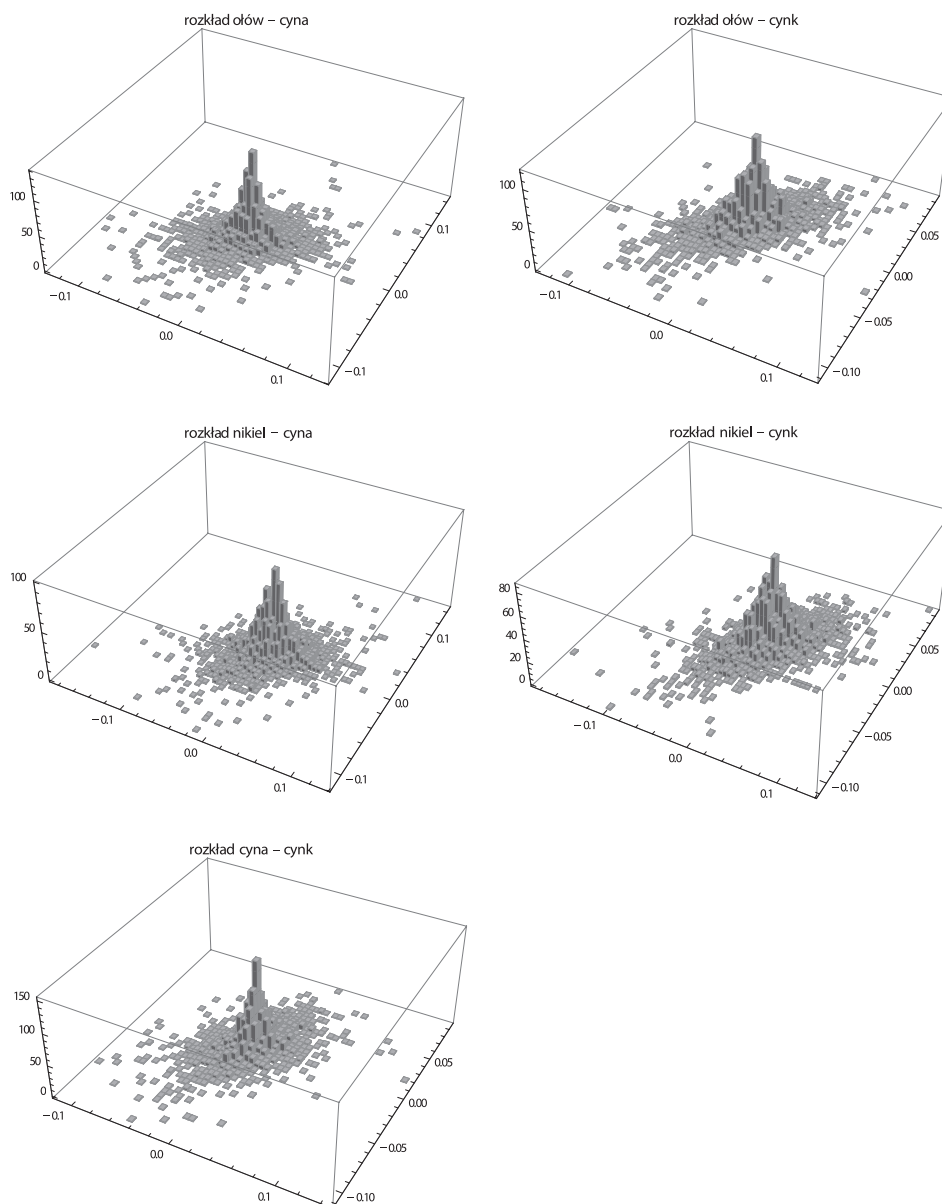


rozkład miedź – cynk



rozkład ołów – nikiel





Rys. 4. Dwuwymiarowe rozkłady normalne

Źródło: Obliczenia własne.

Ze względu na odrzucenie hipotezy o normalności rozkładu badanych zmiennych wyznaczono dodatkowe mierniki zależności: τ Kendalla oraz ρ Spearmana. Wyniki przedstawiono w tabelach 4 i 5.

Tabela 4

Współczynnik τ Kendalla

Współczynnik τ Kendalla	Aluminium	Miedź	Ołów	Nikiel	Cyna	Cynk
Aluminium	1,000000	0,532613	0,388139	0,359145	0,314182	0,473709
Miedź	0,532613	1,000000	0,455021	0,416378	0,364273	0,536839
Ołów	0,388139	0,455021	1,000000	0,339585	0,324095	0,486198
Nikiel	0,359145	0,416378	0,339585	1,000000	0,308108	0,389886
Cyna	0,314182	0,364273	0,324095	0,308108	1,000000	0,326339
Cynk	0,473709	0,536839	0,486198	0,389886	0,326339	1,000000

Źródło: Obliczenia własne.

Tabela 5

Współczynnik ρ Spearmana

Współczynnik ρ Spearmana	Aluminium	Miedź	Ołów	Nikiel	Cyna	Cynk
Aluminium	1,000000	0,716383	0,547485	0,512659	0,448554	0,650970
Miedź	0,716383	1,000000	0,625180	0,580379	0,512819	0,717903
Ołów	0,547485	0,625180	1,000000	0,482420	0,457977	0,656908
Nikiel	0,512659	0,580379	0,482420	1,000000	0,440251	0,546062
Cyna	0,448554	0,512819	0,457977	0,440251	1,000000	0,461653
Cynk	0,650970	0,717903	0,656908	0,546062	0,461653	1,000000

Źródło: Obliczenia własne.

W przypadku wyników pomiaru zależności za pomocą współczynników τ Kendalla oraz ρ Spearmana wyciągnięto podobne wnioski, jak przy zastosowaniu współczynnika korelacji liniowej Pearsona: najsilniejszą zależność wskazano dla pary miedź oraz cynk, natomiast najsłabszą dla pary zmiennych nikiel oraz cyna. Wszystkie współczynniki korelacji wykazywały statystyczną istotność na poziomie $\alpha = 0,05$. Dodatkowo wskazano, że zachodzi związek: $\tau(X, Y) < \rho_s(X, Y) < \rho(X, Y)$.

W ostatnim etapie analizy oszacowano współczynniki zależności w dolnym (λ_L) i górnym (λ_U) ogonie dwuwymiarowych rozkładów badanych watorów. Wyniki przedstawiono w tabelach 6 oraz 7.

Tabela 6

Współczynniki zależności w dolnym ogonie rozkładu

λ_L	Aluminium	Miedź	Ołów	Nikiel	Cyna	Cynk
Aluminium	1,000000	0,192215	0,154478	0,124351	0,142257	0,182245
Miedź	0,192215	1,000000	0,172248	0,151245	0,145598	0,312254
Ołów	0,154478	0,172248	1,000000	0,161124	0,154642	0,212429
Nikiel	0,124351	0,151245	0,161124	1,000000	0,092257	0,192254
Cyna	0,142257	0,145598	0,154642	0,092257	1,000000	0,132235
Cynk	0,182245	0,312254	0,212429	0,192254	0,132235	1,000000

Źródło: Obliczenia własne.

Tabela 7

Współczynniki zależności w górnym ogonie rozkładu

λ_U	Aluminium	Miedź	Ołów	Nikiel	Cyna	Cynk
Aluminium	1,000000	0,172534	0,149987	0,132455	0,151234	0,191134
Miedź	0,172534	1,000000	0,162249	0,142238	0,150122	0,292214
Ołów	0,149987	0,162249	1,000000	0,159984	0,139958	0,201139
Nikiel	0,132455	0,142238	0,159984	1,000000	0,085564	0,162245
Cyna	0,151234	0,150122	0,139958	0,085564	1,000000	0,123348
Cynk	0,191134	0,292214	0,201139	0,162245	0,123348	1,000000

Źródło: Obliczenia własne.

Wyniki przedstawione w tabelach 6 oraz 7 wskazują, że występują zależności zarówno w dolnym, jak i górnym ogonie dwuwymiarowych rozkładów analizowanych stóp zwrotu metali. Podobnie jak w przypadku współczynnika Pearsona, *tau* Kendalla oraz *rho* Spearmana, wykazano najsilniejsze zależności w obu ogonach dla pary miedź oraz cynk, natomiast najsłabsze dla pary zmiennej nikiel oraz cyna.

Podsumowanie

Badania empiryczne danych finansowych prowadzone przez ośrodki naukowe na świecie jednoznacznie wskazują, iż występują pomiędzy nimi zależności nie tylko w obrębie samej klasy aktywów, ale także pomiędzy różnymi rynkami (w sensie przedmiotowym i geograficznym). Powszechnie wykorzystywane założenie normalności rozkładów empirycznych nie jest spełnione

w praktyce, implikując tym samym konieczność wykorzystania innych mierników zależności, odpornych na takie cechy finansowych szeregów czasowych, jak asymetria, tworzenie się skupisk danych czy też występowanie obserwacji ekstremalnych. W opracowaniu tym, oprócz klasycznego i nadal powszechnie wykorzystywanego miernika zależności, jakim jest współczynnik korelacji Pearsona, zaprezentowano teoretycznie i praktycznie takie miary, jak τ Kendalla, ρ Spearmana oraz współczynniki mierzące związki występujące w ogonach rozkładów. Wyniki wskazują na występowanie różnic w wartościach współczynników, przy czym pokazano, iż zachodzi związek $\tau(X, Y) < \rho_s(X, Y) < \rho(X, Y)$. Dodatkowo występują także zależności zarówno w dolnym, jak i górnym ogonie rozkładu dwuwymiarowego dla rozważanych par metali. Analiza tego rodzaju związków jest szczególnie istotna z punktu widzenia zarządzania ryzykiem, gdyż pozwala oszacować prawdopodobieństwo równoczesnego wystąpienia realizacji stopy zwrotu na poziomie oddalonym od centralnej części rozkładu dla obu rozważanych aktywów.

Literatura

- Bjerve S., Doksum K.: *Correlation Curves: Measures of Association as Functions of Covariate Values*. „The Annals of Statistics” 1993, Vol. 21, No. 2, s. 890-902.
- Dhaene J., Goovaerts M.J.: *Dependency of Risks and Stop-loss Order*. „Austin Bulletin” 1996, Vol. 26, No. 2, s. 201-212.
- Doman R.: *Zastosowanie kopuli w modelowaniu dynamiki zależności na rynkach finansowych*. Wydawnictwo Uniwersytetu Ekonomicznego, Poznań 2011.
- Embrechts P., McNeil A.J., Straumann D.: *Correlation and Dependence in Risk Management: Properties and Pitfalls*. W: *Risk Management: Value at Risk and Beyond*. Ed. M.A.H. Dempster. Cambridge 2002, s. 176-223.
- Fama E.F.: *The Behavior of Stock Market Prices*. „Journal of Business” 1965, Vol. 38, No. 1, s. 34-105.
- Gurgul P., Syrek R.: *Zastosowanie mieszanki kopuł do modelowania współzależności pomiędzy wybranymi sektorami gospodarki*. „Ekonomia Menedżerska” 2009, nr 6, s. 129-139.
- Kruskal W.H.: *Ordinal Measures of Association*. „Journal of the American Statistical Association” 1958, No. 53, s. 814-861.
- Lyubushin Jr. A.A.: *Robust Wavelet-aggregated Signal for Geophysical Monitoring Problems*. „Izvestiya, Physics of the Solid Earth” 2002, Vol. 38, No. 9, s. 1-17.
- Malevergne Y., Sornette D.: *Extreme Financial Risk. From Dependence to Risk Management*. Springer-Verlag, Berlin, Heidelberg 2006.

Mandelbrot B.: *The Variation of Certain Speculative Prices*. „Journal of Business” 1963, Vol. 36, No. 4, s. 394-419.

Meerschaert M.M., Scheffler H.: *Sample Cross-correlations for Moving Averages with Regularly Varying Tails*. „Journal of Time Series Analysis” 2001, Vol. 22, No. 4, s. 481-492.

DEPENDENCY MEASURES – STATISTICAL ANALYSIS USING SELECTED ASSETS FROM NON-FERROUS METALS MARKET

Summary

The analysis of investment risk is strongly associated with the knowledge of dependency structure between assets. The aim of this paper is to present some dependency and concordance measures between random variables. Such measures as Pearson coefficient of correlation, Kendall's *tau* and Spearman's *rho* has been discussed. Moreover, the measures of tail dependencies has been presented. An empirical analysis has been carried out using selected assets from non-ferrous metals market.