

Wiesław Szczesny¹, Teresa Kowalczyk², Marek Wiech²,
¹Katedra Ekonometrii i Informatyki SGGW, ²Instytut Podstaw Informatyki PAN
e-mail: wszczesny@mors.sggw.waw.pl,

KONCEPCJE POMIARU NIERÓWNOŚCI DLA WIELU ZMIENNYCH¹.

Streszczenie: Praca składa się z trzech głównych części. W pierwszej z nich wprowadzone zostają pojęcia kierunkowej krzywej Lorenza i kierunkowego wskaźnika Giniego, uogólnione następnie na odpowiednie pojęcia przy pomiarze koncentracji. W części drugiej wprowadzone pojęcia wykorzystane są do przedstawienia pewnych metod pomiaru nierówności, określanych jako podejście gradacyjne dla tablic z wieloma zmiennymi. Jako przykłady służą trzy tablice z danymi dla pięciu banków i trzech ich cech: stanu depozytów oraz stanów kredytów złotówkowych i kredytów dewizowych. Końcowa część pracy to porównanie wskaźników gradacyjnych z proponowanymi przez Moslera i Koshevoya oraz krótkie odniesienie do innych koncepcji pomiaru nierówności.

Słowa kluczowe: kierunkowa krzywa Lorenza, kierunkowy wskaźnik Giniego, koncentracja, nierówność, zróżnicowanie, tau Kendalla, gradacyjna analiza danych

WSTĘP

Tabela 1 przedstawia przykładowe tabele danych dla pięciu banków i trzech cech: D – stan depozytów, KZ – stan kredytów złotych, KW – stan kredytów walutowych. Na tych przykładach zilustrujemy różne koncepcje pomiaru nierówności, wprowadzane jako szczególny przypadek pomiaru koncentracji. W ujęciu intuicyjnym koncentracja środków finansowych jest w pierwszej i drugiej tabeli większa niż w trzeciej, ale trudno wydać intuicyjny osąd porównujący koncentracje w pierwszej i drugiej tabeli, chociaż można się raczej skłaniać ku temu, że większa jest nierówność w pierwszej tabeli. Ten intuicyjny osąd będzie w pracy zestawiony z ocenami koncentracji dokonanyymi za pomocą paru wybranych wskaźników.

W rozdziale 2 przypomnimy znane od ponad stu lat pojęcia krzywej Lorenza i wskaźnika Giniego, które posłużą następnie do wprowadzenia *kierunkowej krzywej Lorenza i kierunkowego wskaźnika Giniego*, po czym uogólnimy te pojęcia na kierunkową krzywą i wskaźnik *koncentracji* (por. rozdział 4 w [Kowalczyk i in.

¹ Artykuł częściowo finansowany z projektu badawczego nr 3 T11C 053 28, przyznanego przez Ministerstwo Nauki i Informatyzacji

2004]). W rozdziale 3 wykorzystamy pojęcia wprowadzone w rozdziale 2 do przedstawienia pewnych metod pomiaru nierówności dla tablic z wieloma zmiennymi opisującymi zbiór obiektów (banków). To podejście do pomiaru nierówności dla wielu zmiennych (wprowadzone w [Kowalczyk i in. 2006]) będziemy nazywać podejściem *gradacyjnym*. Następnie w rozdziale 4 porównamy podejście gradacyjne z wysoko cenionym w środowisku statystycznym podejściem proponowanym przez Koshevoya i Moslera [1997] i omówimy tabelę zawierającą wartości wszystkich rozpatrywanych wskaźników dla tabel 1a, 1b i 1c. Inne koncepcje pomiaru nierówności zostaną krótko wspomniane w rozdziale 5.

W całym opracowaniu zajmujemy się tylko k -tkami (wektorami) addytywnych zmiennych o wartościach nieujemnych, które są mierzone na skali ilorazowej. Trójka zmiennych D , KZ i KW spełnia te założenia. Chcemy również podkreślić, że używając w rozdziałach 2-4 słowa *zmienna* mamy zawsze na myśli zmienne o podanych wyżej własnościach.

Tabele 1a, 1b, 1c są uzupełnione wierszami zawierającymi sumy i średnie kolumn oraz kolumną wartości zmiennej Z , która jest sumą zmiennych D , KZ i KW :

$$Z = D + KZ + KW.$$

Tabela 1. Zbiory danych dla pięciu banków i trzech zmiennych (D – depozyty, KZ – kredyty złotówkowe, KW – kredyty walutowe).

a)		D	KZ	KW	Z
	B ₁	400	300	80	780
	B ₂	50	30	6	86
	B ₃	15	25	8	48
	B ₄	10	30	4	44
	B ₅	25	15	2	42
	suma	500	400	100	1000
	średnia	100	80	20	200

b)		D	KZ	KW	Z
	B ₁	400	15	0	415
	B ₂	50	30	3	83
	B ₃	15	25	5	45
	B ₄	10	30	90	130
	B ₅	25	300	2	327
	suma	500	400	100	1000
	średnia	100	80	20	200

c)		D	KZ	KW	Z
	B ₁	200	150	40	390
	B ₂	100	100	25	225
	B ₃	80	80	20	180
	B ₄	70	50	10	130
	B ₅	50	20	5	75
	suma	500	400	100	1000
	średnia	100	80	20	200

POMIAR NIERÓWNOŚCI DLA POJEDYNCZEJ ZMIENNEJ

Tabela 2. Wartości z tabeli 1 zamienione na udziały zmiennych; składowe wektora π u dołu tabel są udziałami sum w kolumnach w ogólnej sumie wartości zmiennych D , KZ i KW .

	uD	uKZ	uKW	uZ
B_1	0.80	0.75	0.80	0.78
B_2	0.10	0.075	0.06	0.086
B_3	0.03	0.063	0.08	0.048
B_4	0.02	0.075	0.04	0.044
B_5	0.05	0.038	0.02	0.042
suma	1	1	1	1
π	0.50	0.40	0.10	1

	uD	uKZ	uKW	uZ
B_1	0.80	0.04	0.00	0.415
B_2	0.10	0.08	0.03	0.083
B_3	0.03	0.06	0.05	0.045
B_4	0.02	0.08	0.9	0.13
B_5	0.05	0.75	0.02	0.327
suma	1	1	1	1
π	0.50	0.40	0.10	1

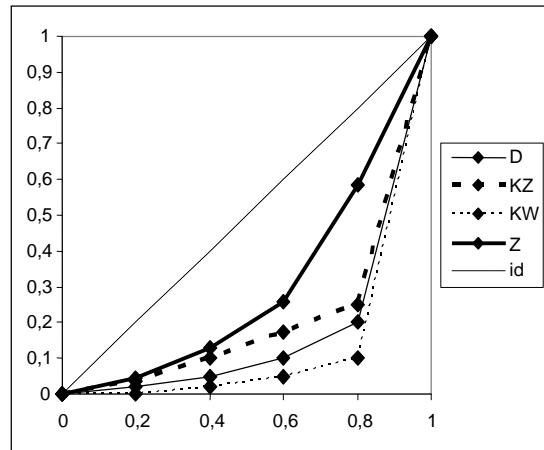
	uD	uKZ	uKW	uZ
B_1	0.40	0.38	0.40	0.39
B_2	0.20	0.25	0.25	0.225
B_3	0.16	0.2	0.2	0.18
B_4	0.14	0.125	0.1	0.13
B_5	0.1	0.05	0.05	0.075
suma	1	1	1	1
π	0.50	0.40	0.10	1

W tabelach 2a, 2b, 2c zawarte są udziały wartości każdej ze zmiennych D , KZ i KW w bankach $B_1 - B_5$ w sumie jej wartości, oznaczone odpowiednio uD , uKZ i uKW . W uzupełniających tabelach dolnych wierszach zapisany jest wektor π , którego składowymi są udziały sum każdej zmiennej w ogólnej sumie wartości tych trzech zmiennych. Zauważmy, że w ten sposób każda z tabel 2a, 2b, 2c wyznacza pewien dwuwymiarowy rozkład, z rozkładem brzegowym $\pi = (\pi_D, \pi_{KZ}, \pi_{KW})$, jednakowym w tabelach 2a, 2b, 2c i równym $\pi = (0.5, 0.4, 0.1)$ oraz z rozkładami warunkowymi zapisanymi w kolumnach (nazywanymi profilami danej zmiennej; na przykład rozkład (0.8, 0.1, 0.03, 0.02, 0.05) jest profilem dla depozytów D w tabelach 1a i 1b).

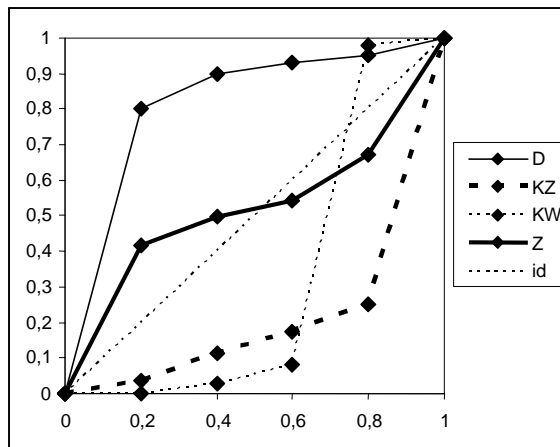
Do obliczenia krzywej Lorenza dla wybranej zmiennej potrzebny jest jedynie *ustawiony rosnąco* (niemalejąco) ciąg udziałów tej zmiennej, który porównujemy z rozkładem jednostajnym. Przykładowo dla zmiennej D w tabelach 1a i 1b rosnący ciąg udziałów jest równy: 0.02, 0.03, 0.05, 0.1, 0.8, a wtedy krzywa Lorenza L_D dla zmiennej D jest wypukłą łamaną o wierzchołkach $((0, 0), (0.2, 0.02), (0.2+0.2, 0.02+0.03), \dots, (1, 1))$.

Jeśli dokonamy permutacji zbioru banków, a więc spermutujemy również wartości udziałów, to postępując analogicznie jak przy wyznaczaniu L_D

otrzymamy łamaną, która nie będzie wypukła. Dokonując wszystkich możliwych permutacji utworzymy zbiór 120 łamanych, które mieszczą się w kwadracie jednostkowym w obszarze ograniczonym od dołu krzywą Lorenza L_D , a od góry tzw. górną krzywą Lorenza \bar{L}_D , będącą odbiciem krzywej L_D względem punktu $(0.5, 0.5)$. Poszczególne krzywe z tego zbioru nazywamy *kierunkowymi krzywymi Lorenza* (wg funkcji φ porządkującej banki) i oznaczamy $L_{D;\varphi}$.



Rysunek 1. Krzywe Lorenza L_D , L_{KZ} , L_{KW} i L_Z dla zmiennych w tabeli 1b.



Rysunek 2. Kierunkowe krzywe Lorenza $L_{D;\varphi}$, $L_{KZ;\varphi}$, $L_{KW;\varphi}$ i $L_{Z;\varphi}$ dla zmiennych w tabeli 1b przy uporządkowaniu od B_1 do B_5 .

Na rysunku 1 przedstawiono krzywe Lorenza L_D , L_{KZ} , L_{KW} i L_Z dla zmiennych w tabeli 1b. Uporządkowania φ są tutaj dostosowane osobno do każdej ze zmiennych: $(B_4, B_3, B_5, B_2, B_1)$ dla L_D , $(B_1, B_3, B_2, B_4, B_5)$ dla L_{KZ} , etc. Natomiast na rys. 2 przedstawiono krzywe $L_{D;\varphi}$, $L_{KZ;\varphi}$, $L_{KW;\varphi}$ i $L_{Z;\varphi}$ dla uporządkowania $(B_1, B_2,$

B_3, B_4, B_5) przyjętego w tabeli 1 (identycznego dla każdej z krzywych). Łatwo zauważyć, że jak wcześniej wspomniano każda z krzywych na rysunku 2 zawiera się w obszarze zawartym między krzywą Lorenza danej zmiennej a odbiciem tej krzywej względem punktu $(0.5, 0.5)$.

Tabela 1c wyróżnia się tym, że każda ze zmiennych ma wartości uporządkowane malejąco, poczynając od banku B_1 do B_5 (jest to szczególnie przykład zestawu zmiennych *komonotonicznych*). Wskutek tego krzywe Lorenza dla zmiennych D, KZ, KW pomnożone przez przypisaną im składową wektora π są w sumie równe krzywej Lorenza dla ich sumy (krzywa L_Z). Tak nie jest w tabelach 1a i 1b; natomiast we wszystkich tabelach dla dowolnego uporządkowania banków zachodzi równość (zob. także [Rietveld 1990]):

$$L_{Z;\varphi}(t) = \pi_D L_{D;\varphi}(t) + \pi_{KZ} L_{KZ;\varphi}(t) + \pi_{KW} L_{KW;\varphi}(t), \quad t \in [0,1] \quad (1)$$

co dla danych z tabeli 1b można sprawdzić na rysunku 2.

Krzywa Lorenza jest funkcyjnym wskaźnikiem nierówności dowolnej zmiennej X . Częściowy porządek nierówności dla par zmiennych X i Y określa warunek: „ L_X leży pod L_Y ”; np. na rysunku 1 krzywa L_{KW} leży pod krzywą L_D , czyli nierówność zmiennej KW jest większa niż nierówność zmiennej D . Natomiast *liczbowym* wskaźnikiem nierówności dowolnej zmiennej X jest wskaźnik *Gini*(X), który jest równy podwojonemu polu między przekątną łączącą w kwadracie jednostkowym punkty $(0, 0)$ i $(1, 1)$ a krzywą Lorenza L_X . Dla zmiennych z tabeli 1b mamy: $Gini(KW)=0.732$, $Gini(D)=0.652$, $Gini(KZ)=0.575$, $Gini(Z)=0.394$, co zgadza się z oglądem pól na rysunku 1. Ponadto dla dowolnej zmiennej X i uporządkowania φ wprowadza się *kierunkowy wskaźnik Gini*($X; \varphi$), który jest równy podwojonej różnicy dwóch pól: pola między wcześniej wspomnianą przekątną a tą częścią kierunkowej krzywej $L_{X;\varphi}$, która jest położona *pod* przekątną, oraz pola między przekątną a tą częścią krzywej $L_{X;\varphi}$, która przebiega *nad* przekątną. Kierunkowy wskaźnik Giniego jest największy przy uporządkowaniu $\varphi = X$ (przy którym $L_{X;\varphi} = L_X$). Dla zmiennych z tabeli 1b i uporządkowania φ od B_1 do B_5 otrzymujemy następujące wartości kierunkowych wskaźników Giniego: $Gini(KZ; \varphi)=0.570$, $Gini(KW; \varphi)=0.364$, $Gini(Z; \varphi)=-0.052$, $Gini(D; \varphi)=-0.632$, co zgadza się z oglądem pól na rysunku 2.

Kierunkowa krzywa Lorenza i jej wskaźnik Giniego są szczególnymi przypadkami *krzywej koncentracji* i *wskaźnika koncentracji*, o których mowa dalej. Niech $\mathbf{p} = (p_1, \dots, p_m)$ oraz $\mathbf{q} = (q_1, \dots, q_m)$, $p_1 + \dots + p_m = q_1 + \dots + q_m = 1$ będą dwoma rozkładami na zbiorze m elementowym, odpowiadającymi dowolnej parze zmiennych losowych X i Y określonych na tym zbiorze. *Krzywa koncentracji* $C(Y; X; \varphi)$ *zmiennej* Y *względem zmiennej* X według uporządkowania φ zadającego permutację (i_1, \dots, i_m) ciągu $(1, \dots, m)$ jest łamaną o wierzchołkach $\{(0,0), (p_{i_1}, q_{i_1}), (p_{i_1}+p_{i_2}, q_{i_1}+q_{i_2}), \dots, (1,1)\}$. *Kierunkowy wskaźnik koncentracji* $ar(Y; X; \varphi)$ i *maksymalny wskaźnik koncentracji* $ar_{\max}(Y; X)$ dla krzywych $C(Y; X; \varphi)$ i $C_{\max}(Y; X)$ definiuje się analogicznie jak wskaźniki nierówności.

Tak więc na przykład dla pary zmiennych (D, KZ) z tabeli 1b i przyjętego w niej uporządkowania banków od B_1 do B_5 , krzywa koncentracji $C(KZ:D;\varphi)$ jest łamaną o wierzchołkach $\{(0, 0), (0.80, 0.04), (0.80+0.10, 0.04+0.08), \dots, (1, 1)\}$. Gdy uporządkowanie banków jest takie, że krzywa koncentracji staje się wypukła, jest ona *maksymalną krzywą koncentracji* $C_{max}(KZ:D)$, położoną pod wszystkimi innymi krzywymi $C(KZ:D;\varphi)$.

Dla danych z tabeli 1b zachodzi: $ar(KZ:D;\varphi) = ar_{max}(KZ:D) = 0.874$, czyli uporządkowanie banków od B_1 do B_5 ustawią pary wartości obu zmiennych tak, że zróżnicowanie odpowiadających tym zmiennym kolumn jest w tabeli 1b największe. Można sprawdzić, że równość ar i ar_{max} nie zachodzi dla tych samych zmiennych w tabeli 1a.

Wskaźnik Giniego ma wiele przedstawień, prowadzących do różnych uogólnień wielowymiarowych. Jednym z nich jest unormowana średnia różnica Giniego (ang. *distance Gini index*):

$$Gini(X) = \frac{E(|X - X'|)}{2E(X)} \quad (2)$$

gdzie X' jest niezależną kopią zmiennej X , a E oznacza wartość oczekiwaną. Zatem $Gini(X)$ mierzy średnie zróżnicowanie wartości zmiennej X unormowanej za pomocą $E(X)$.

WSKAŹNIK NIERÓWNOŚCI DLA TABLIC Z WIELOMA ZMIENNYMI OPISUJĄCYMI ZBIÓR OBIEKTÓW

Kierunkowe wskaźniki Giniego dla każdej z cech D , KZ i KW z osobna oraz dla ich sumy Z zastąpimy teraz kierunkowym wskaźnikiem nierówności dla trójki cech łącznie. Ma on tę własność, że jeśli zamiast trójki weźmie się pod uwagę tylko jedną zmienną, np. D , to wskaźnik zdefiniowany dla wielu zmiennych zredukuje się do $Gini(D;\varphi)$. Konstrukcję wskaźnika przedstawimy na przykładzie trójki $X = (D, KZ, KW)$ w tabeli 1b przy pierwotnym ustawieniu banków od B_1 do B_5 . Niech T_φ będzie macierzą powstałą z tabeli 2b, a \tilde{T}_φ macierzą T_φ poprzedzoną kolumną \mathbf{p} o wartościach $(0.2, 0.2, 0.2, 0.2, 0.2)$:

$$T_\varphi = \begin{bmatrix} 0.80 & 0.04 & 0.00 \\ 0.10 & 0.08 & 0.03 \\ 0.03 & 0.06 & 0.05 \\ 0.02 & 0.08 & 0.90 \\ 0.05 & 0.75 & 0.02 \end{bmatrix}, \quad \tilde{T}_\varphi = \begin{bmatrix} 0.2 & 0.80 & 0.04 & 0.00 \\ 0.2 & 0.10 & 0.08 & 0.03 \\ 0.2 & 0.03 & 0.06 & 0.05 \\ 0.2 & 0.02 & 0.08 & 0.90 \\ 0.2 & 0.05 & 0.75 & 0.02 \end{bmatrix}.$$

Kolumny w macierzach T_φ i \tilde{T}_φ traktujemy teraz jako profile zmiennych dla rozkładów dwuwymiarowych, które oznaczymy odpowiednio przez P_φ i \tilde{P}_φ , przyjmując jako rozkłady brzegowe odpowiednio wektory $\boldsymbol{\pi}$ oraz $(0.5, 0.5\pi_D, 0.5\pi_{KZ}, 0.5\pi_{KW})$. Kolumny udziałów zmiennych D , KZ i KW w rozkładzie \tilde{P}_φ mają łącznie wagę 0.5, taką samą jak kolumna określonego na bankach jednostajnego rozkładu, z którym chcemy porównać macierz T_φ . Rozkład P_φ jest więc zadany przez parę $(T_\varphi, \boldsymbol{\pi})$, a rozkład \tilde{P}_φ przez parę $(\tilde{T}_\varphi, [0.5, 0.5\boldsymbol{\pi}])$; w tablicach 1a, 1b, 1c wektor $[0.5, 0.5\boldsymbol{\pi}]$ jest jednakowy i równy $[0.5, 0.25, 0.20, 0.05]$.

Znajdziemy najpierw związek między wskaźnikami *tau Kendalla* dla P_φ i \tilde{P}_φ . Na podstawie wzoru (8.23) w [Kowalczyk i in. 2004]:

$$\tau(P_\varphi) = 2 \times (\pi_D \pi_{KZ} ar(KZ : D; \varphi) + \pi_D \pi_{KW} ar(KW : D; \varphi) + \pi_{KZ} \pi_{KW} ar(KW : KZ; \varphi)). \quad (3)$$

Rozpisując analogicznie $\tau(\tilde{P}_\varphi)$ i korzystając ze wzoru (1) otrzymujemy równość:

$$\tau(\tilde{P}_\varphi) = 0.5 \times Gini(Z; \varphi) + 0.25 \times \tau(P_\varphi). \quad (4)$$

Moduły obu wskaźników *tau Kendalla* mają następujące ograniczenia:

$$|\tau(P_\varphi)| \leq M = 1 - \pi_D^2 - \pi_{KZ}^2 - \pi_{KW}^2, \quad |\tau(\tilde{P}_\varphi)| \leq \tilde{M} = 0.25 \times (3 - \pi_D^2 - \pi_{KZ}^2 - \pi_{KW}^2). \quad (5)$$

Możemy więc wprowadzić ilorazy

$$DIV(\mathbf{X}; \varphi) = \frac{|\tau(P_\varphi)|}{M}, \quad INEQ(\mathbf{X}; \varphi) = \frac{|\tau(\tilde{P}_\varphi)|}{\tilde{M}} \quad (6)$$

nazwane w [Kowalczyk i in. 2006] kierunkowym wskaźnikiem *zróznicowania* wektora \mathbf{X} i kierunkowym wskaźnikiem *nierówności* wektora \mathbf{X} (*DIV* od ang. *divergence*, *INEQ* od ang. *inequality*). Zatem:

$$INEQ(\mathbf{X}; \varphi) = \alpha \times Gini(Z; \varphi) + (1 - \alpha) \times DIV(\mathbf{X}; \varphi), \quad (7)$$

$$\text{gdzie } \alpha = \frac{2}{3 - (\pi_D^2 + \pi_{KZ}^2 + \pi_{KW}^2)}, \quad \alpha \in [0, 1]. \quad (8)$$

Gdy zamiast trójki zmiennych (D, KZ, KW) weźmiemy pod uwagę tylko jedną zmienną, np. D , to $\pi_D = 1$, $\alpha = 1$, $1 - \alpha = 0$, $Z = D$, a więc $INEQ(D; \varphi) = Gini(D; \varphi)$.

Wskaźnik $INEQ(X; \varphi)$ zależy od ustawienia φ dla wierszy i ustawienia ψ dla zmiennych w wektorze X . Maksymalny wskaźnik nierówności $INEQ_{max}$ definiujemy za [Kowalczyk i in. 2006] jako:

$$INEQ_{max}(X) = \max_{\varphi, \psi} INEQ(X; \varphi, \psi) = INEQ(X; \varphi_{opt}, \psi_{opt}). \quad (9)$$

Wartości $INEQ_{max}(D, KZ, KW)$ dla danych z tabel 1a, 1b i 1c podane są w tabeli 3 łącznie z optymalnymi ustawieniami ($\varphi_{opt}, \psi_{opt}$), współczynnikiem α i wartościami wskaźników $Gini(Z; \varphi_{opt})$ i $DIV(X; \varphi_{opt}, \psi_{opt})$, które występują we wzorze (7) po wstawieniu do tego wzoru pary ustawień φ i ψ maksymalizujących $INEQ(X; \varphi)$.

PORÓWNANIE UJĘCIA MOSLERA-KOSHEVOYA Z UJĘCIEM GRADACYJNYM

Dwie propozycje uogólnienia wskaźnika nierówności dla pojedynczej zmiennej, podane w [Koshevoy i Mosler 1997], odnoszą się do dwóch różnych przedstawień wskaźnika Giniego dla pojedynczej zmiennej, o których była mowa w rozdziale 2. W pierwszym przedstawieniu $Gini(X)$ jest powierzchnią pola między L_X i \bar{L}_X , w drugim zaś uśrednionym i unormowanym modułem różnicy wartości zmiennej X i jej kopii X' . Mosler i Koshevoy zastępują powierzchnię obszaru między L_X i \bar{L}_X objętością pewnej bryły w przestrzeni $(k+1)$ -wymiarowej nazwanej Zonoidem Lorenza, jak również zastępują uśrednione odległości między wartościami zmiennych X i X' uśrednionymi odległościami dla wektorów. Pierwszy wskaźnik nosi nazwę GZI (od *Gini Zonoid Index*). Drugi wskaźnik, nazwany DGI (od *distance Gini index*), jest uogólnieniem unormowanej średniej różnicy Giniego (wzór (2)). Wartości $GZI(X)$ i $DGI(X)$ dla danych z tabel 1a, 1b i 1c są zamieszczone w tabeli 3.

W ujęciu gradacyjnym opisanym w rozdziale 3 mamy do czynienia z wypukłą kombinacją nieco innych uogólnień obu przedstawień wskaźnika Giniego dla pojedynczej zmiennej. Pierwszy składnik, $\alpha \times Gini(Z; \varphi_{opt})$, odnosi się do różnicy pól między kierunkową krzywą Lorenza dla sumy zmiennych a diagonalą. Drugi składnik, $(1 - \alpha) \times DIV(X; \varphi_{opt}, \psi_{opt})$ uśrednia kierunkowe wskaźniki koncentracji dla par zmiennych; są one gradacyjnymi wskaźnikami różnicowania zastępującymi odległości euklidesowe. Jest godne uwagi, że posłużenie się dekompozycją tau Kendalla, która uśrednia kierunkowe wskaźniki koncentracji, prowadzi do jednoczesnego uwzględnienia obu nurtów uogólnień: „geometrycznego” i „odległościowego”.

Ujęcie gradacyjne przedstawione w tabelicy 3 wydaje się zgodne z intuicjami dotyczącymi nierówności w tabelach 1a, 1b i 1c, o których była mowa we Wstępie.

Dla danych 1a widać „gołym okiem” bardzo duże zróżnicowanie pomiędzy rekordami dla banków (bank B_1 dominuje pozostałe banki, czyli koncentracja jest bardzo silna), a zatem należy się spodziewać dużej wartości $Gini(Z; \varphi_{opt})$. Natomiast profile kolumn są zbliżone i stąd należy się spodziewać małej wartości $DIV(X; \varphi_{opt}, \psi_{opt})$. Ogólnie koncentracja w tabeli 1a jest duża, oparta głównie na zróżnicowaniu rekordów dla pierwszego banku i banków pozostałych. Te intuicje są dobrze oddane przez wartości podane w tabeli 3.

Tabela 3: Wartości wyliczanych w artykule wskaźników nierówności.

Wskaźniki oceniające nierówność w tabelach danych	Tabela 1a	Tabela 1b	Tabela 1c	Ustawienie tabel wg wartości wskaźników
w ujęciu gradacyjnym:				
Optymalne ustawienie φ_{opt}	$B_4, B_5, B_3,$ B_2, B_1	$B_3, B_2, B_4,$ B_5, B_1	$B_5, B_4, B_3,$ B_2, B_1	-
Optymalne ustawienie ψ_{opt}	KZ, KW, D	KW, KZ, D	D, KZ, KW	-
α	0.775	0.775	0.775	-
$INEQ_{max}(X)$	0.482	0.456	0.232	$1a \succ 1b \succ 1c$
$Gini(Z; \varphi_{opt})$	0.606	0.393	0.290	$1a \succ 1b \succ 1c$
$DIV(X; \varphi_{opt}, \psi_{opt})$	0.053	0.671	0.031	$1b \succ \succ 1a \succ 1c$
w ujęciu Moslera i Koshevoya:				
$GZI(X)$	0.0011	0.1833	0.0003	$1b \succ \succ 1a \succ 1c$
$DGI(X)$	0.363	0.531	0.180	$1b \succ 1a \succ 1c$

Dla danych w tabeli 1b, zamiast koncentracji kapitału w jednym banku jak w 1a, mamy do czynienia z wyraźną koncentracją w dwóch bankach (B_1 i B_5), co każe się spodziewać dużej wartości wskaźnika $Gini(Z; \varphi_{opt})$, mniejszej jednak niż w 1a. Natomiast profile zmiennych w 1b są bardzo zróżnicowane w każdej z par zmiennych (D, KZ), (D, KW) i (KZ, KW), co każe się spodziewać dużej wartości $DIV(X; \varphi_{opt}, \psi_{opt})$, znacznie większej niż w przykładzie 1a. Te intuicje są ponownie zgodne z wartościami wskaźników podanymi w tabeli 3. Uśrednienie $Gini(Z; \varphi_{opt})$ i $DIV(X; \varphi_{opt}, \psi_{opt})$ za pomocą α i $(1 - \alpha)$ prowadzi do wartości maksymalnej nierówności $INEQ_{max}(X)$ nieco mniejszej niż w przykładzie 1a.

Dane w tabeli 1c wykazują umiarkowane zróżnicowanie pomiędzy bankami i bardzo małe zróżnicowanie profili poszczególnych zmiennych (np. udziały depozytów dla poszczególnych banków są bardzo podobne do udziałów kredytów wśród tych banków). Można się więc spodziewać umiarkowanej (ale mniejszej niż w tabelach 1a i 1b) wartości $Gini(Z; \varphi_{opt})$ oraz bardzo małej wartości $DIV(X; \varphi_{opt}, \psi_{opt})$. Tak właśnie jest w tabeli 3.

Jak się wydaje, Mosler i Koshevoy skłaniają się ku opisywaniu nierówności wielocechowej dwoma wskaźnikami, $GZI(X)$ i $DGI(X)$, a nie jednym globalnym. Próba łącznej interpretacji wartości obu tych wskaźników w zestawieniu ze wskaźnikami gradacyjnymi będzie podjęta w [Kowalczyk i in. 2006]. Tutaj zwrócimy tylko uwagę, że w przykładach 1a i 1c $GZI(X)$ przyjmuje małe wartości, co wynika z faktu, że zmienne są w tych tabelach silnie skorelowane. Gdyby którekolwiek dwie zmienne były liniowo zależne, to wartość $GZI(X)$ byłaby równa 0.

Tabela 1c – zgodnie z intuicjami we Wstępie – uzyskała najmniejsze wartości wszystkich wskaźników występujących w tabeli 3. Z kolei tabela 1a ma wartości wskaźników $INEQ_{max}(X)$ i $Gini(Z; \varphi_{opt})$ większe niż tabela 1b, a dla wskaźników $DGI(X)$, $GZI(X)$ i $DIV(X; \varphi_{opt}, \psi_{opt})$ jest przeciwnie. Przypomnijmy, że intuicyjny osąd we Wstępie skłaniał się raczej ku wyższej ocenie nierówności w tabeli 1a.

INNE KONCEPCJE POMIARÓW NIERÓWNOŚCI

Poza omówionym już podejściem Moslera-Koshevoya używającym narzędzi analizy wypukłej, w literaturze wyróżnia się jeszcze dwa podejścia: jedno oparte na tzw. funkcjach dobrobytu (ang. *social welfare functions*), a drugie oparte na funkcjach użyteczności [Savaglio 2004]. Funkcje te służą do sprawdzania macierzy danych w tablicy $m \times k$ do pojedynczej zmiennej o m -wartościach, której nierówność ocenia się następnie za pomocą wskaźników jednowymiarowych. Przegląd takich wskaźników zawiera [Weymark 2004]. Prace Savaglio oraz [Gajdos i Weymark 2005] zawierają aksjomatyczne ujęcia pomiaru wielowymiarowej nierówności. W drugiej z tych prac podane są przykłady wskaźników wielowymiarowej nierówności, które powstały z rozmaitych uogólnień wskaźników jednowymiarowych.

Użyteczność badań nad pomiarem koncentracji w działalności sektora bankowego przedstawia raport Jackowicza i Kowalewskiego [2002], w którym rozpatrywana jest koncentracja w ujęciu jednowymiarowym. Autorzy piszą we wstępie, że miary koncentracji są w praktyce wykorzystywane do wyznaczenia reguł prowadzonej polityki ochrony konkurencji, jak również do objaśnienia przyczyn zachowań niekonkurencyjnych lub do oceny skutków niejednakowej efektywności uczestników rynku.

LITERATURA

- Gajdos T., Weymark J. A. (2005) Multidimensional generalized Gini indices, *Economic Theory*, 26, str. 471 – 496
- Jackowicz K., Kowalewski O. (2002) Koncentracja działalności sektora bankowego w Polsce w latach 1994-2000, Departament Analiz i Badań, *Narodowy Bank Polski*, 143, Warszawa
- Koshevoy G., Mosler K. (1997) Multivariate Gini indices, *Journal of Multivariate Analysis*, 60, str. 252 – 276
- Kowalczyk T., Pleszczyńska E., Ruland, F. (red.). (2004) *Grade Models and Methods for Data Analysis*, Studies in Fuzziness and Soft Computing No 151, str. 1 – 477, Springer: Berlin-Heidelberg-New York
- Kowalczyk T., Pleszczyńska E., Szczesny W., Wiech M (2006) Measurement of multivariate inequality based on Kendall tau or Spearman rho (*w przygotowaniu*)
- Rietveld P. (1990) Multidimensional Inequality Comparisons, On Aggravation and Mitigation of Inequalities, *Economic Letters*, 32, str. 187 – 192
- Savaglio E. (2004) *Multidimensional inequality: a survey*, w: Farina F., Savaglio E. (red.) Inequality and economic integration, London: Routledge
- Weymark J. A. (2004) *The normative approach to the measurement of multidimensional inequality*, w: Farina F., Savaglio E. (red.) Inequality and economic integration, London: Routledge

Concepts for measures of inequality in multivariate data

Summary: The article consists of three main parts. The first part introduces concepts of directional Lorenz curve and directional Gini index, generalized then to obtain suitable terms concerning concentration measures. In the second part some inequality indices are introduced, which present a grade approach to multivariate datasets. As an example three datasets concerning five banks and three variables (deposits, Polish currency loans and foreign currency loans) are considered. The final part consists of comparison of grade approach to approach proposed by Mosler and Koshevoy, followed by a short account of some other concepts of inequality measurement.

Key words: directional Lorenz curve, directional Gini index, concentration, inequality, diversity, Kendall's tau, grade data analysis