PRZEGLĄD STATYSTYCZNY
R. LX – ZESZYT 4 – 2013

KATARZYNA OSTASIEWICZ

ADEKWATNOŚĆ WYBRANYCH ROZKŁADÓW TEORETYCZNYCH DOCHODÓW W ZALEŻNOŚCI OD METODY APROKSYMACJI[1]

1. WPROWADZENIE

Ostatecznym wynikiem badań statystycznych jest często pewien parametr liczbowy, syntetycznie charakteryzujący badaną zbiorowość ze względu na określoną cechę statystyczną. W przypadku dochodów ludności takim parametrem jest na przykład współczynnik Giniego. Współczynnik ten zwykle obliczany jest bezpośrednio na podstawie szeregów rozdzielczych. Możliwe jest jednak alternatywne podejście: można szereg rozdzielczy aproksymować jakimś rozkładem teoretycznym, a następnie rozkład ten wykorzystać do obliczenia (aproksymacji) współczynnika Giniego. Wyniki obu postępowań mogą być – i zwykle są – różne, podobnie jak różnić się mogą wyniki otrzymywane przy użyciu różnych rozkładów teoretycznych.

Inspiracją do przeprowadzenia badań przedstawionych w niniejszej pracy były wyniki uzyskane przez McDonalda i Ransoma (1979) na podstawie danych angielskich. W cytowanej pracy autorzy ci porównywali wartości dwóch charakterystyk liczbowych rozkładów dochodów (wartości średniej i współczynnika Giniego), uzyskiwanych przy aproksymacji szeregu rozdzielczego dochodów kilkoma różnymi rozkładami teoretycznymi i kilkoma różnymi technikami szacowania badanych wielkości. Jako miara „dobroci dopasowania" danego rozkładu teoretycznego przyjęta była wielkość sumy odchyleń kwadratowych oraz wartości statystyki χ^2. Następnie autorzy porównywali otrzymane rezultaty (wartości średnie oraz wartości współczynnika Giniego) pomiędzy sobą (czyli dla różnych rozkładów teoretycznych i różnych technik obliczeniowych), w przypadku współczynnika Giniego odnosząc je dodatkowo do ograniczeń dla tej wielkości wyznaczonych przez Gastwirtha (1972).

Niniejsza praca pod względem metodycznym jest niemal dokładnym powtórzeniem pracy McDonalda i Ransoma z tą różnicą, że dane uwzględnione w badaniu dotyczyły całej populacji (wspólnie rozliczających podatki par małżeńskich w określo-

[1] Autorka wyraża wdzięczność anonimowym Recenzentom za krytyczne uwagi do pierwszej wersji pracy, które pomogły spojrzeć na przeprowadzone badanie z zupełnie innej perspektywy oraz pozwoliły dostrzec ułomność prezentacji uzyskanych wyników.

nym regionie Dolnego Śląska), zatem wyniki uzyskane za pomocą różnych rozkładów teoretycznych i różnych technik obliczeniowych mogły być porównywane nie tylko pomiędzy sobą, ale również z wynikami ścisłymi, obliczonymi na podstawie szeregu szczegółowego dla całej populacji. Ponadto, poza współczynnikiem Giniego, analizie poddano takie charakterystyki, jak odchylenie standardowe, współczynnik Theila oraz trzy wskaźniki z rodziny współczynników Atkinsona.

Rozkłady dochodów modelowane są za pomocą różnych rozkładów teoretycznych (por. np. Kot, 1999; Kośny, 2001; Ulman, 2011). Wśród nich szczególną rolę odgrywa rozkład logarytmiczno-normalny, prawdopodobnie najczęściej przyjmowany jako teoretyczny model dochodów. Spośród innych wymienić można rozkład Pareto, Weibulla, czy też zdobywającą ostatnio coraz większą popularność funkcję logarytmiczno-logistyczną (np. Kot, Adamkiewicz-Drwiłło, 2013). Singh i Maddala (1976) zaproponowali rozkład będący uogólnieniem rozkładów Pareto z jednej i Weibulla z drugiej strony, który, jak się okazało, bardziej pasował do badanych przez nich dochodów niż rozkłady gamma i log-normalny. Rozkład ten był także badany przez polskich autorów (por. Kot, 1995). Do modelowania rozkładów dochodów stosowano również i znacznie mniej „popularne" rozkłady, takie jak rozkład Champernowne'a (por. Champernowne, 1953) czy uogólniony rozkład beta (por. Thurow, 1970). Rozkład beta uważany jest za najlepszy model teoretyczny rozkładu dochodów, mimo to – prawdopodobnie z powodu dużych wymagań obliczeniowych wynikających z konieczności określenia wielu parametrów – wykorzystywany jest stosunkowo rzadko w pracach dotyczących zastosowań. Warto zwrócić uwagę na propozycję przedstawioną przez Daguma (1977). Proponuje on wykorzystać uogólniony rozkład log-logistyczny. Dobre dopasowanie tej funkcji do danych empirycznych zostało zbadane, potwierdzone i przedstawione w literaturze (por. Dagum, Lemmi, 1987). Rozkład ten stosowany był również do modelowania rozkładów dochodów w Polsce (np. Panek, Szulc, 1991).

Jako że celem niniejszej pracy jest badanie wpływu przyjęcia takiego lub innego modelu na ocenę dokładności miar nierówności, zbadano typowe rozkłady stosowane w praktyce: rozkład log-normalny, gamma, Singh-Maddala, log-logistyczny oraz rozkład Daguma. Parametry badanych rozkładów wyznaczano za pomocą pięciu różnych metod. Trzy z nich stosowane były w pracy McDonalda i Ransoma (metoda najmniejszych kwadratów, metoda największej wiarygodności, kryterium minimum statystyki χ^2), zaś dwie dodatkowe metody to: zmodyfikowana metoda minimum χ^2 oraz „mechaniczna" metoda dopasowania rozkładu teoretycznego do danych empirycznych.

Dalszy układ pracy jest następujący. W części następnej zdefiniowane są teoretyczne funkcje badanych rozkładów oraz definicje badanych charakterystyk liczbowych. W części trzeciej przedstawiono metody obliczania tych parametrów. Po tej części wprowadzającej przedstawione są wyniki przeprowadzonych obliczeń. Artykuł kończy krótkie podsumowanie całości badań.

2. WIELKOŚCI CHARAKTERYZUJĄCE ROZKŁAD I TEORETYCZNE ROZKŁADY DOCHODÓW

W części tej przedstawiona jest postać rozkładów teoretycznych, stanowiących przedmiot badania, i ich wybranych charakterystyk. Na początek wybrane charakterystyki zdefiniowane zostaną w sposób ogólny: w przypadku niektórych funkcji rozkładu mogą one przybrać bardziej zwięzłą postać, wyrażoną poprzez parametry rozkładu.

Analizie poddano pięć charakterystyk liczbowych rozkładu dochodów:

– wartość średnia, na podstawie szeregu szczegółowego liczona następująco:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i, \tag{1}$$

gdzie x_i oznaczają poszczególne indywidualne obserwacje, a N – liczebność badanej populacji;

oraz na podstawie dopasowanej funkcji gęstości:

$$\mu = \int_{-\infty}^{\infty} x f(x) dx, \tag{2}$$

gdzie $f(x)$ oznacza funkcję gęstości prawdopodobieństwa rozkładu teoretycznego.

– odchylenia standardowe, liczone odpowiednio według wzorów:

$$s = \left[\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2\right]^{1/2} \tag{3}$$

oraz:

$$\sigma = \left[\int_{-\infty}^{\infty}(x-\mu)^2 f(x) dx\right]^{1/2}. \tag{4}$$

– współczynnik Giniego, liczony ze wzorów:

$$G = \frac{1}{2N^2\bar{x}}\sum_{i,j=1}^{N}\left|x_i - x_j\right| \tag{5}$$

oraz:

$$G_f = \frac{1}{\mu}\int_{-\infty}^{\infty} F(x)[1 - F(x)] dx, \tag{6}$$

gdzie $F(x)$ oznacza dystrybuantę rozkładu teoretycznego.

– współczynnik Theila, liczony ze wzorów:

$$T = \frac{1}{N}\sum_{i=1}^{N}\frac{x_i}{\bar{x}}\ln\frac{x_i}{\bar{x}} \tag{7}$$

oraz:

$$T_f = \int_{-\infty}^{\infty} f(x) \frac{x}{\mu} \ln \frac{x}{\mu} dx. \tag{8}$$

– współczynnik Atkinsona, liczony jako:

$$A_\varepsilon = 1 - \frac{1}{\bar{x}} \left[\frac{1}{N} \sum_{i=1}^{N} x_i^{1-\varepsilon} \right]^{1/(1-\varepsilon)} \tag{9}$$

oraz:

$$A_\varepsilon = 1 - \frac{1}{\mu} \left[\int_{-\infty}^{\infty} x^{1-\varepsilon} f(x) \right]^{1/(1-\varepsilon)}. \tag{10}$$

Wartości współczynnika Atkinsona obliczane są dla trzech różnych wartości parametru ε: 0,1, 0,5 i 0,8.

Do analizy wybrano pięć podanych niżej rozkładów:

– **rozkład logarytmiczno-normalny** (por. Aitchison, Brown, 1969):

$$f_{LN}(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2} \right], \; x > 0, \tag{11}$$

gdzie μ jest parametrem położenia, a σ – kształtu.
Wartość średnia i odchylenie standardowe dla rozkładu log-normalnego wyrażają się poprzez jego parametry w następujący sposób:

$$\mu_{LN} = e^{\mu + \frac{\sigma^2}{2}}, \sigma_{LN} = e^{\mu + \frac{\sigma^2}{2}} \sqrt{e^{\sigma^2} - 1}. \tag{12}$$

– **rozkład gamma** (por. Salem, Mount, 1974):

$$f_G(x) = \frac{1}{\beta^\alpha} x^{\alpha - 1} \frac{\exp[-x/\beta]}{\Gamma(\alpha)}, \; x > 0, \tag{13}$$

gdzie Γ oznacza funkcję gamma Eulera, β jest parametrem położenia, a α kształtu.
Wartość średnia i odchylenie standardowe wyrażają się poprzez parametry rozkładu w sposób następujący:

$$\mu_G = \alpha\beta, \sigma_G = \beta\sqrt{\alpha}. \tag{14}$$

Znana jest również analityczna postać współczynnika Giniego dla tego rozkładu:

$$G_G = \frac{\Gamma\left(\alpha + \frac{1}{2}\right)}{\Gamma(\alpha + 1)\sqrt{\pi}}. \tag{15}$$

– **rozkład logarytmiczno-logistyczny** (Fisk, 1961):

$$f_{LL}(x) = \frac{(a/b)(x/b)^{a-1}}{[1+(x/b)^a]^2},\ x > 0, \tag{16}$$

gdzie b jest parametrem położenia, i a – parametrem kształtu.
Dla rozkładu log-logistycznego wartość średnia istnieje, gdy $a > 1$, a odchylenie standardowe gdy $a > 2$, i wyrażają się wzorami:

$$\mu_{LL} = \frac{b\pi}{a\sin(x/a)}, \sigma_{LL} = b\sqrt{\frac{2\pi/a}{\sin(2\pi/a)} - \frac{(\pi/a)^2}{\sin^2(\pi/a)}}. \tag{17}$$

W rozkładzie tym współczynnik Giniego ma wyjątkowo prostą postać:

$$G_{LL} = \frac{1}{a}. \tag{18}$$

– **rozkład Daguma**, będący uogólnieniem rozkładu log-logistycznego (por. Dagum, 1977):

$$f_D(x) = \frac{(ac/x)(x/b)^{ac}}{[1+(x/b)^a]^{c+1}},\ x > 0. \tag{19}$$

Jak widać, w porównaniu do funkcji log-logistycznej ma ona dodatkowy parametr kształtu, c. Z porównania funkcji (16) i (19) wynika, że rozkład log-logistyczny jest szczególnym przypadkiem rozkładu Daguma, dla $c = 1$. Wartość średnia i odchylenie standardowe dla rozkładu Daguma mają następującą postać:

$$\mu_D = -\frac{b}{a}\frac{\Gamma\left(-\frac{1}{a}\right)\Gamma\left(\frac{1}{a}+c\right)}{\Gamma(c)}, \sigma_D = \frac{b}{a}\sqrt{-2a\frac{\Gamma\left(-\frac{2}{a}\right)\Gamma\left(\frac{2}{a}+c\right)}{\Gamma(c)} - \left(\frac{\Gamma\left(-\frac{1}{a}\right)\Gamma\left(\frac{1}{a}+c\right)}{\Gamma(c)}\right)^2}, \tag{20}$$

a warunkiem ich istnienia jest $a > 1$ (dla wartości średniej) i $a > 2$ (dla odchylenia standardowego). Współczynnik Giniego wyraża się wzorem:

$$G_D = \frac{\Gamma\left(\frac{1}{a}+2c\right)\Gamma(c)}{\Gamma\left(\frac{1}{a}+c\right)\Gamma(2c)} - 1. \tag{21}$$

– **rozkład Singh-Maddala** (por. Singh, Maddala, 1976):

$$f_{SM}(x) = \frac{(a/b)c(x/b)^{a-1}}{\left(1+\left(\frac{x}{b}\right)^{a}\right)^{1+c}},\ x > 0. \tag{22}$$

W tym przypadku również można zauważyć, że rozkład log-logistyczny jest szczególnym przypadkiem rozkładu Singh-Maddala, dla $c = 1$. Wartość średnia istnieje, gdy $ac > 1$, a odchylenie standardowe gdy $ac > 2$. Określone są one wyrażeniami:

$$\mu_{SM} = \frac{b\Gamma\left(\frac{1}{a}+1\right)\Gamma\left(-\frac{1}{a}+c\right)}{\Gamma(c)},$$

$$\sigma_{SM} = \frac{b}{\Gamma(c)}\sqrt{\Gamma\left(1+\frac{2}{a}\right)\Gamma(c)\Gamma\left(-\frac{2}{a}+c\right) - \Gamma^2\left(1+\frac{1}{a}\right)\Gamma^2\left(-\frac{1}{a}+c\right)}. \tag{23}$$

Znana jest również analityczna postać współczynnika Giniego dla tego rozkładu:

$$G_{SM} = 1 - \frac{\Gamma\left(2c-\frac{1}{a}\right)\Gamma(c)}{\Gamma(2c)\Gamma\left(c-\frac{1}{a}\right)}. \tag{24}$$

Kleiber (1996) wykazał, iż pomiędzy rozkładem Daguma a rozkładem Singh-Maddala istnieje ścisłe powiązanie. Gdy zmienna losowa X ma rozkład Daguma z parametrami a, b i c wówczas jej odwrotność, czyli zmienna losowa $1/X$, ma rozkład Singh-Maddala z parametrami a, $1/b$ i c.

Pozostałe charakterystyki liczbowe, których postaci analityczne nie zostały tu podane, mogą być wyznaczone numerycznie na podstawie ogólnych wzorów, (6), (8) i (10).

Analizowane dane o dochodach pochodzą z Izby Skarbowej Wrocław-Fabryczna i dotyczą 19487 małżeństw wspólnie rozliczających podatki za rok 2007. Na podstawie danych dotyczących dochodów w tej zbiorowości skonstruowano szereg rozdzielczy, oznaczony Sc. Następnie z populacji tej wyodrębniono dwie podpopulacje: małżeństw bezdzietnych, liczącą 10625 małżeństw (skonstruowany szereg rozdzielczy oznaczono przez S0) oraz podpopulację o liczebności 5458 małżeństw z jednym dzieckiem (skonstruowany szereg rozdzielczy oznaczono przez S1). Każdy z szeregów rozdzielczych składał się z 13 przedziałów klasowych o różnych szerokościach (zwiększających się dla ostatnich 6 klas, by zapewnić minimalną liczebność obserwacji w każdej klasie).

3. OCENA BADANYCH PARAMETRÓW

Zastosowano pięć różnych metod oceny parametrów rozkładów teoretycznych. Pierwszą z nich jest metoda największej wiarygodności (MNW), zaadaptowana do danych zagregowanych, czyli przedstawionych w postaci szeregów rozdzielczych. Niech n_i, i = ,..., g, oznaczają liczebności poszczególnych przedziałów klasowych szeregu rozdzielczego, $N = \sum_{i=1}^{g} n_i$, i (x_{i-1}, x_i), oznaczają granice przedziałów klasowych. Teoretyczne prawdopodobieństwa, że obserwacja znajdzie się w danym przedziale, wyrażają się następująco:

$$p_i(\{a\}) = \int_{x_{i-1}}^{x_i} f(x, \{a\})dx, \tag{25}$$

gdzie argument $\{a\}$ podkreśla zależność zarówno funkcji gęstości ($f(x, \{a\})$, jak i otrzymanych prawdopodobieństw ($p_i(\{a\})$), od zestawu parametrów oznaczonych symbolicznie poprzez $\{a\}$. Funkcja wiarygodności, czyli prawdopodobieństwo otrzymania w poszczególnych przedziałach n_i obserwacji przy założeniu danej funkcji gęstości $f(x, \{a\})$ ma następującą postać:

$$L\left(n_1, \dots, n_g \middle| \{a\}\right) = N! \prod_{i=1}^{g} \left(\frac{{p_i}^{n_i}}{n_i!}\right). \tag{26}$$

Maksymalizując funkcję $L(n_1,...,n_g| \{a\})$ ze względu na wektor parametrów $\{a\}$ uzyskuje się asymptotycznie efektywne (por. np. Cox, Hinckley, 1974) estymatory tych parametrów. Zadanie maksymalizowania funkcji (26) może zostać sprowadzone do prostszego obliczeniowo szukania maksimum następującej funkcji (po zlogarytmowaniu funkcji wiarygodności):

$$\tilde{L}\left(n_1, \dots, n_g \middle| \{a\}\right) = \sum_{i=1}^{g} n_i \ln p_i. \tag{27}$$

Drugą z metod stosowanych w tej pracy jest metoda najmniejszych kwadratów (MNK), polegająca na takim doborze parametrów $\{a\}$, dla których suma kwadratów odchyleń teoretycznych prawdopodobieństw od zaobserwowanych częstości jest minimalna:

$$S_{MNK}(\{a\}) = \sum_{i=1}^{g} \left(\frac{n_i}{N} - p_i\right)^2. \tag{28}$$

Trzecia metoda aproksymacji polega na minimalizacji wartości χ^2 (min χ^2), to znaczy, minimalizacji następującej sumy:

$$S_{\chi^2}(\{a\}) = N \sum_{i=1}^{g} \frac{\left(\frac{n_i}{N} - p_i\right)^2}{p_i}, \tag{29}$$

gdzie p_i obliczane są według wzoru (25).

Otrzymane w ten sposób wielkości parametrów są asymptotycznie równoważne estymatorom otrzymanym metodą największej wiarygodności (por. np. Cox, Hinckley, 1974). Z drugiej strony, metoda ta może być postrzegana jako uogólnienie metody najmniejszych kwadratów, polegające na nadaniu wag poszczególnym kwadratom odchyleń.

Stosowana jest również modyfikacja metody χ^2, polegająca na tym, że wagi zależą nie od teoretycznych prawdopodobieństw ale od częstości empirycznych (np. Stuart, Ord, 1991) (metoda oznaczana będzie jako min$\chi^{2'}$). W tej metodzie zadanie sprowadza się do minimalizacji wartości następującej statystyki (por. Neyman, 1949):

$$S_{\chi^2\prime}(\{a\}) = N \sum_{i=1}^{g} \frac{\left(\frac{n_i}{N} - p_i\right)^2}{\frac{n_i}{N}}. \tag{30}$$

Zaletą tej modyfikacji jest mniejsza złożoność obliczeniowa, jako, że parametry $\{a\}$, ze względu na które minimalizujemy wyrażenie (30), występują tylko w liczniku.

Ostatnia, piąta metoda ma charakter raczej „mechaniczny" niż statystyczny, jako, że polega ona na prostym dopasowywaniu krzywej dystrybuanty rozkładu teoretycznego do punktów empirycznych. Dopasowywanie to uzyskuje się metodą najmniejszych kwadratów, tzn. minimalizując sumę kwadratów odchyleń wartości dystrybuanty teoretycznej od wartości skumulowanych częstości (w odróżnieniu od metody MNK, por. wzór (28), w której minimalizowana jest suma kwadratów odchyleń prawdopodobieństw teoretycznych od częstości). Metoda oznaczana jest jako MNK'. Minimalizowana funkcja ma następującą postać:

$$S_{MNK\prime}(\{a\}) = \sum_{i=1}^{g} \left(\frac{\sum_{j=1}^{i} n_j}{N} - F(x_i, \{a\}) \right)^2. \tag{31}$$

Funkcja (31) często stosowana jest jako kryterium dobroci dopasowania rozkładu teoretycznego do zaobserwowanych danych empirycznych. Z definicji, to właśnie przy użyciu metody MNK' suma ta osiąga wartość minimalną, a zatem dobroć dopasowania mierzona za pomocą tego kryterium jest największa. Jednym z celów pracy jest zatem porównanie metod statystycznych z metodą „mechaniczną" pod względem dokładności aproksymacji oraz oszacowań wartości charakterystyk rozkładów.

4. WYNIKI

We wstępnym etapie przeprowadzono, dla każdego z rozkładów opisanych w części 2, badanie zgodności rozkładu empirycznego z rozkładem teoretycznym, przy użyciu testu zgodności χ^2. W zależności od liczby parametrów (2 dla rozkładów logarytmiczno-normalnego, gamma i logarytmiczno-logistycznego oraz 3 dla rozkładów Daguma i Singh-Maddala) statystyka testowa ma asymptotyczny rozkład χ^2 o 9 lub 10 stopniach swobody. Dla poziomu istotności 0,01 wartość krytyczna wynosi 21,7 lub 23,2 (dla 2 lub 3 parametrów), i jest ona na tyle duża, że w przypadku każdego z rozkładów wartość statystyki testowej była nawet o kilka rzędów wielkości od niej mniejsza. Zatem hipoteza o zgodności rozkładu empirycznego z żadnym z rozkładów teoretycznych nie została odrzucona.

W tabelach poniżej przedstawione zostały wyniki dalszych obliczeń. W tabeli 1 zamieszczone zostały oszacowane wartości parametrów rozkładów teoretycznych, wykorzystywanych do aproksymacji szeregów S0, S1 i Sc. W ostatnich wierszach zamieszczone zostały wartości średniego odchylenia kwadratowego, czyli wielkości określone wzorem: $d^2 = S_{MNK'}/g$ (dla przypomnienia, g oznacza liczbę przedziałów klasowych, czyli jednocześnie liczbę empirycznych częstości skumulowanych, por. też wzór (31)). Kursywą wyróżnione zostały te, które przyjmują wartość najmniejszą (nie licząc metody MNK', dla której z definicji średnie odchylenie kwadratowe jest najmniejsze). W każdej kolumnie (odpowiadającej danej metodzie) czcionką pogrubioną wyróżnione zostały trzy wartości (wartości najmniejsze dla każdego z trzech szeregów z osobna), identyfikujące rozkłady, dające dla danej metody obliczania najlepszą dokładność dla poszczególnych szeregów.

Jak widać, prawie we wszystkich przypadkach, (poza jednym), rozkład Singh-Maddala daje najlepsze dopasowanie do danych empirycznych, jeśli dokładność mierzona jest jako średnie odchylenie kwadratowe. Dla rozkładu gamma najlepsze dopasowanie (po metodzie MNK') daje metoda najmniejszych kwadratów, dla rozkładu log-normalnego (również po metodzie MNK') daje stosowanie metody min χ^2, w przypadku pozostałych rozkładów nie jest to jednoznaczne. Oczywiście, w każdym przypadku średnie odchylenia kwadratowe jest najmniejsze dla metody MNK', gdyż wynika to z istoty tej metody.

Tabela 1.

Wartości oszacowanych parametrów rozkładów teoretycznych i średniego odchylenia kwadratowego

			MNK	minχ^2	min$\chi^{2'}$	MNW	MNK’
		α	2,57555	2,02723	2,04128	2,18857	2,35942
	S0	β	21758,4	30526,4	30013,8	27238,2	24417,8
		d^2	*0,000219*	0,000698	0,00056	0,000249	0,000112
		α	2,94479	2,45858	2,55739	2,64932	2,82227
Gamma	S1	β	23429,1	30445,3	29161,4	27463,2	25004,5
		d^2	*0,00018*	0,000618	0,00054	0,000242	0,000105
		α	2,64322	2,16463	2,19187	2,30966	2,46676
	Sc	β	23838,1	31803,3	31150,1	28891,2	26206,8
		d^2	*0,000187*	0,000594	0,000476	0,000223	9,36E-05
		μ	10,8311	10,7616	10,779	10,7659	10,7852
	S0	σ	0,68007	0,704215	0,686921	0,700012	0,663488
		d^2	0,00039	0,000169	*0,000106*	0,000149	7,97E-05
		μ	11,045	11,0013	11,0042	11,0019	11,0172
log-Normal	S1	σ	0,634982	0,643675	0,636354	0,641142	0,609046
		d^2	0,000322	0,000186	*0,000157*	0,000176	0,000107
		μ	10,9462	10,885	10,8993	10,8886	10,9082
	Sc	σ	0,66926	0,692638	0,673396	0,687488	0,650544
		d^2	0,000341	0,000189	*0,000111*	0,000163	8,27E-05
		a	2,15292	2,14727	2,15003	2,14748	2,15897
	S0	b	68211,5	66479,9	66424,2	66543,2	66765,6
		c	1,71608	1,65317	1,6509	1,6557	1,6665
		d^2	**4,45E-06**	**4,32E-06**	***4,08E-06***	**4,16E-06**	**3,12E-06**
		a	2,26934	2,44438	2,46964	2,45298	2,35454
S-M	S1	b	89828,6	75267,3	74709,3	75101,7	82720,6
		c	1,95065	1,45138	1,426	1,44383	1,69296
		d^2	0,0000313765	**2,85E-05**	**2,98E-05**	***2,84E-05***	3,12E-06
		a	2,16818	2,21258	2,21358	2,21217	2,18199
	Sc	b	77989,1	73462,4	73333,2	73518,2	76986,8
		c	1,76736	1,61222	1,60988	1,6151	1,73347
		d^2	***3,61E-06***	**5,30E-06**	**5,37E-06**	**5,17E-06**	**3,47E-06**
		a	2,87815	2,97399	2,97249	2,97441	2,9599
	S0	b	60318,8	62027,	61970,6	62021,1	61950,1
		c	0,674986	0,636064	0,636587	0,635919	0,63786
		d^2	7,59E-06	*4,60E-06*	4,61E-06	4,62E-06	4,33E-06
Daguma		a	3,19279	3,10215	3,09333	3,09666	3,2513
	S1	b	76434,9	72048,3	71431,3	71738,9	76797,1
		c	0,637454	0,720323	0,74194	0,729592	0,633004
		d^2	***2,66E-05***	3,06E-05	3,82E-05	3,21E-05	2,07E-05

	Sc	*a*	2,90087	3,00015	2,99846	3,00006	3,04148
		b	67647,5	68582,8	68507,3	68563,6	70329,1
		c	0,68119	0,658424	0,659958	0,658791	0,626864
		d^2	1,73E-05	*9,52E-06*	9,71E-06	9,59E-06	7,64E-06
log-logist	S0	*a*	2,44163	2,46869	2,51718	2,49585,	2,56436
		b	49939,4	48207,8	48421,4	48172,5	48426,1
		d^2	0,000316	0,000117	*9,52E-05*	0,000102	8,69E-05
	S1	*a*	2,6245	2,73064	2,73649	2,73943	2,78189
		b	62450,3	60821,8	61026,6	60828,9	61065
		d^2	0,00033	0,000124	0,000122	*0,000121*	0,000116
	Sc	*a*	2,47397	2,52628	2,55576	2,54687	2,60574
		b	56145	54575,6	54599	54482,5	54752,3
		d^2	0,000307	0,000123	*0,00011*	0,000113	0,000102

Źródło: opracowanie własne.

Przejdźmy teraz do analizy dokładności oszacowania za pomocą teoretycznych funkcji rozkładu zdefiniowanych w części 2. charakterystyk rozkładu i porównania dokładności tych oszacowań z dokładnością dopasowania rozkładu teoretycznego do danych empirycznych, mierzoną jako d^2.

Konstrukcja kolejnych tabel jest następująca. W kolumnach umieszczone zostały wyniki uzyskane dla różnych metod, natomiast w wierszach – dla różnych funkcji w podziale na trzy rozpatrywane szeregi. Dodatkowo, w ostatnim wierszu odpowiadającym poszczególnym rozkładom teoretycznym umieszczone zostały uśrednione (po trzech różnych szeregach) wartości względnych odchyleń wyznaczonych parametrów od ich wartości dokładnych, tj.:

$$\bar{d}_{w,f} = \frac{1}{3}\left(\left|\frac{w_{f,S0}-w_{S0}}{w_{S0}}\right| + \left|\frac{w_{f,S1}-w_{S1}}{w_{S1}}\right| + \left|\frac{w_{f,Sc}-w_{Sc}}{w_{Sc}}\right|\right),$$

gdzie *w* oznacza badaną wielkość, indeks dolny *f* oznacza wartość obliczoną za pomocą funkcji *f*, a indeksy S0, S1 i Sc odnoszą się do trzech różnych szeregów. W ostatniej kolumnie każdej tabeli umieszczone zostały wielkości $\bar{\bar{d}}_w$, czyli wartości średnie $\bar{d}_w$ uśrednione po wszystkich stosowanych metodach. Kursywą wyróżnione zostały wielkości najlepsze (najmniejsze odchylenia) w każdym z wierszy odpowiadających $\bar{d}_w$ (identyfikując w ten sposób metodę najlepszą dla danego rozkładu przy obliczaniu konkretnej wielkości), a czcionką pogrubioną – wartość najmniejszą (czyli wynik najdokładniejszy) w odpowiedniej kolumnie, identyfikując zatem rozkład, który przy danej metodzie daje najdokładniejsze wyniki. Wielkość pogrubiona w ostatniej kolumnie określa funkcję, która daje najlepsze (w sensie średnim) wyniki dla wszystkich zastosowanych metod. Na dole tabel podane są, dla porównania, wartości dokładne (obliczone na podstawie obserwacji niezgrupowanych) badanych wielkości, w_{S0}, w_{S1} i w_{Sc}. W ramce znajdują się wyniki najlepsze (najdokładniejsze) ze wszystkich zawartych w danej tabeli.

Tabela 2.

Wartość średnia dla różnych metod i rozkładów

metoda / rozkład		MNK	minχ^2	min$\chi^{2'}$	MNW	MNK’	$\bar{\bar{d}}_{\bar{x}}$
gamma	S0	56039,8	61884,3	61266,6	59612,8	57611,8	
	S1	68993,8	74852,2	74577,1	72759,	70569,5	
	Sc	63009,4	68842,3	68277	66728,9	64645,9	
	$\bar{d}_{\bar{x},G}$	0,069232	0,018179	*0,010743*	0,01421	0,045389	0,031551
log-norm	S0	63726,1	60446,3	60772,3	60530,2	60190,5	
	S1	76621,6	73750,7	73621,6	73679,6	73324,9	
	Sc	70979,7	67838,1	67916,7	67842,2	67493,4	
	$\bar{d}_{\bar{x},LN}$	0,046851	***0,001974***	0,004744	0,002778	**0,004705**	0,01221
S-M	S0	60030,3	60247,8	60251,2	60231,9	60082,7	
	S1	71975,8	73654,6	73903,6	73722,3	72943,3	
	Sc	67079,5	67589,4	67541,3	67549,8	67149,3	
	$\bar{d}_{\bar{x},SM}$	0,013709	0,00243	***0,001525***	**0,002407**	0,008714	**0,005757**
Daguma	S0	61367,4	60717,7	60694,7	60702,8	60797,5	
	S1	73788,4	74171,8	74620,4	74332,8	73650,8	
	Sc	68988,5	68161	68173,9	68161,3	67999,5	
	$\bar{d}_{\bar{x},D}$	**0,012554**	*0,005311*	0,00727	0,005956	*0,005159*	0,00725
log-logist	S0	66939,5	64180,9	63722,8	63713,3	63057,3	
	S1	80298,6	76646	76824,3	76535,9	76274,4	
	Sc	74651	71671,4	71229,2	71217,4	70671,1	
	$\bar{d}_{\bar{x},LL}$	0,09925	0,052877	0,048973	0,047563	*0,04007*	0,057747
dokładne	$\bar{x}_{S0}$	60302,78					
	$\bar{x}_{S1}$	73933,62					
	$\bar{x}_{Sc}$	67765,83					

Źródło: opracowanie własne.

Rozkład Singh-Maddala zawsze niedoszacowuje, w ramach przeprowadzonych badań, wartość średnią, w przeciwieństwie do rozkładu log-logistycznego, który zawsze przeszacowuje tę wielkość. Rozkład Daguma przeszacowuje wartość średnią w każdym poza dwoma przypadkami. Przy użyciu funkcji log-logistycznej zawsze popełniany jest większy błąd niż przy wykorzystaniu funkcji Daguma; różnica ta jest znaczna, przeważnie jest to rząd wielkości jeśli uwzględni się względne odchylenie modułów. Metoda, przy której uzyskuje się najlepsze przybliżenia, zależy od postaci rozkładu. W żadnym jednak przypadku nie jest nią ani metoda najmniejszych kwadratów ani metoda największej wiarygodności. Widać, że nie zawsze pokrywa się to z metodą dającą najlepsze dopasowanie, wg tabeli 1. Przykładowo, najlepsze dopasowanie rozkładu gamma uzyskuje się metodą najmniejszych kwadratów, MNK, natomiast najlepsze oszacowanie wartości średniej – metodą min$\chi^{2'}$. Ogólnie, najlepszą dokładność oszacowania wartości średniej uzyskuje się przy pomocy rozkładu Singh-Maddala oraz metody min$\chi^{2'}$. Jeśli nie patrzeć na stosowaną metodę wyznaczania parametrów rozkładu, największą dokładność oszacowania średniej uzyskuje się przy wykorzystaniu rozkładu Singh-Maddala. Drugim po nim jest rozkład Daguma.

Tabela 3.

Odchylenie standardowe dla różnych metod i rozkładów

metoda / rozkład		MNK	minχ^2	min$\chi^{2'}$	MNW	MNK'	$\bar{\bar{d}}_s$
gamma	S0	349195	43463,9	42881,7	40295,8	37506,7	
	S1	40205,3	47737,8	46634,5	44701,2	42006,6	
	Sc	38755,9	46791,1	46117,7	43907,6	41160,2	
	$\bar{d}_{s,G}$	0,288604	*0,136991*	0,151605	0,194089	0,229425	0,200143
log-norm	S0	48867,2	48432,7	47190,5	48133,6	44761,8	
	S1	53995,6	52840,6	52017,9	52536,2	49138,2	
	Sc	53354,6	53228,8	51444,4	52735	48989,8	
	$\bar{d}_{s,LN}$	***0,024876***	0,033932	0,058036	0,040756	0,10657	0,052834
S-M	S0	47376,8	49085,6	49050,9	49006,9	48272,2	
	S1	49053,4	55375,4	55567,8	55389	51534,1	
	Sc	51370	53807,9	53796,8	53715,2	51662	
	$\bar{d}_{s,SM}$	0,074938	**0,015401**	***0,014079***	**0,015337**	**0,05223**	**0,034397**
Daguma	S0	57735,9	54564,	54580,9	54539,7	55058,	
	S1	58912,6	60303,8	60565,7	60443,5	57263,3	
	Sc	63838	59844,9	59882,6	59840,7	58997,9	
	$\bar{d}_{s,D}$	0,12955	0,092262	0,09417	0,092907	*0,07238*	0,096254
log-logist	S0	85270,8	79341,4	74840,6	76502,4	70684,8	
	S1	85131	74423,2	74258,4	73812,3	71246,1	
	Sc	91694,9	83372,4	80485,1	81170,8	76207,9	
	$\bar{d}_{s,LL}$	0,64151	0,487067	0,438011	0,450813	*0,365957*	0,476672
dokładne	S_{S0}	48751,93					
	S_{S1}	55731,32					
	S_{Sc}	55642,52					

Źródło: opracowanie własne.

Jak widać, rozkłady log-logistyczny i Daguma zawsze przeszacowują, w ramach przeprowadzonych badań, odchylenie standardowe, natomiast pozostałe rozkłady mają wyraźną tendencję (z odstępstwami) do niedoszacowywania tej wielkości. Niedokładność wyników przy użyciu funkcji log-logistycznej znacznie przewyższa tę uzyskiwaną przy użyciu funkcji Daguma, i jest to niedokładność sięgająca, a nawet przewyższająca 50%. Najdokładniejsza dla badanych przykładów prawie zawsze jest funkcja Singh-Maddala i metoda minχ^2. Funkcja Daguma oraz Singh-Maddala dają dokładność największą dla każdej metody, za wyjątkiem metody najmniejszych kwadratów. Podobnie jak poprzednio, metoda, dla której wyniki były najdokładniejsze, zależała od wykorzystywanego rozkładu, (chociaż nigdy nie była nią metoda największej wiarygodności). Największa dokładność oszacowania odchylenia standardowego nie zawsze pokrywała się z najlepszym dopasowaniem rozkładu teoretycznego do danych empirycznych, np. funkcja gamma najlepiej dopasowana jest przy wykorzystaniu metody MNK, natomiast największą dokładność oszacowania odchylenia standardowego otrzymuje się przy użyciu metody minχ^2. Generalnie, najdokładniejsze oszacowania odchylenia standardowego uzyskuje się przy użyciu funkcji Singh-Maddala, a następnie log-normalnej i Daguma. Funkcja gamma i log-logistyczna dają oszacowania o rząd wielkości gorsze.

Skoro rozkłady Daguma i log-logistyczny zawsze przeszacowują zarówno średnią jak i odchylenia standardowe, natomiast rozkład Singh-Maddala ma silną tendencje do niedoszacowywania obu tych wielkości, można zastanawiać się, czy te wielkości przeszacowania (w przypadku rozkładu Daguma i log-logistycznego) oraz wielkości niedoszacowania (w przypadku rozkładu Singh-Maddala) nie kompensują się, dając w efekcie lepsze oszacowanie współczynnika zmienności. Tak jednak nie jest, gdyż względne odchylenia wartości średniej są o rząd wielkości mniejsze niż wielkości odchylenia standardowego, i współczynnik zmienności wykazuje nieco tylko słabszą tendencję do przeszacowania (dla rozkładu Daguma i log-logistycznego) czy niedoszacowania (dla rozkładu Singh-Maddala).

W kolejnych tabelach zawarte są rezultaty szacowania miar nierówności dla rozkładów teoretycznych.

Oszacowanie współczynnika Giniego w przypadku rozkładu log-logistycznego oraz Daguma w każdym przypadku przewyższa rzeczywistą wartość tego wskaźnika, przy czym błąd popełniany przy wykorzystaniu funkcji log-logistycznej jest w każdym przypadku większy niż w przypadku funkcji Daguma (średnio rzecz biorąc błąd ten jest kilkakrotnie większy). W przypadku rozkładów gamma i Singh-Maddala występuje tendencja do niedoszacowania współczynnika Giniego, choć nie dzieje się tak w każdym przypadku. Nie zaobserwowano żadnej tendencji w przypadku stosowania rozkładu log-normalnego. Metoda szacowania badanych wielkości, przy której uzyskuje się największą dokładność, zależy od postaci rozkładu teoretycznego. Nigdy nie jest to jednak metoda najmniejszych kwadratów. Najmniejszy średni błąd popełniany jest przy użyciu funkcji log-normalnej i metody minχ^2 lub rozkładu Singh-Maddala i metody minχ^2. Jeżeli uśredni się wyniki otrzymywane za pomocą wszystkich stosowanych metod, największa zgodność występuje dla rozkładu Singh-Maddala. Potwierdza się też fakt, że metoda najlepsza do szacowania konkretnej charakterystyki niekoniecznie pokrywa się z metodą dającą najlepszą zgodność dopasowania.

Tabela 4.

Współczynnik Giniego

metoda / rozkład		MNK	minχ^2	min$\chi^{2'}$	MNW	MNK’	$\bar{\bar{d}}_G$
gamma	S0	0,334996	0,372777	0,371645	0,360366	0,34848	
	S1	0,315173	0,342096	0,336069	0,330738	0,321357	
	Sc	0,331084	0,36213	0,360125	0,351821	0,341585	
	$\bar{d}_{G,G}$	0,095307	*0,00908*	0,015932	0,038726	0,067718	0,045353
log-norm	S0	0,3694	0,381484	0,372839	0,379388	0,361042	
	S1	0,346568	0,350997	0,347268	0,349707	0,333285	
	Sc	0,363957	0,375703	0,366042	0,373123	0,354487	
	$\bar{d}_{G,LN}$	**0,00454**	0,02142	***0,00213***	0,015953	0,033245	0,015457
S-M	S0	0,366879	0,37218	0,371911	0,371961	0,369372	
	S1	0,337054	0,345279	0,344186	0,344799	0,339451	
	Sc	0,361248	0,365113	0,36514	0,364959	0,361295	
	$\bar{d}_{G,SM}$	0,018236	***0,002769***	0,003555	**0,003175**	0,013652	**0,008278**
Daguma	S0	0,382238	0,376908	0,376995	0,376882	0,37827	
	S1	0,352563	0,349808	0,347947	0,349176	0,347367	
	Sc	0,37848	0,370188	0,37014	0,37014	0,370689	
	$\bar{d}_{G,SM}$	0,026128	0,011157	*0,009401*	0,010482	**0,010485**	0,013531
log-logist	S0	0,409562	0,405073	0,397271	0,400665	0,389962	
	S1	0,381025	0,366215	0,365432	0,36504	0,359468	
	Sc	0,404208	0,395839	0,391273	0,392639	0,383768	
	$\bar{d}_{G,LL}$	0,101408	0,075524	0,063622	0,067531	*0,044506*	0,070518
dokładne	G_{S0}	0,37178					
	G_{S1}	0,34647					
	G_{Sc}	0,36650					

Źródło: opracowanie własne.

Tabela 5.

Współczynnik Theila

metoda / rozkład		MNK	minχ^2	min$\chi^{2'}$	MNW	MNK'	$\bar{\bar{d}}_T$
gamma	S0	0,181748	0,226811	0,22538	0,211395	0,197196	
	S1	0,160287	0,189796	0,182953	0,177014	0,166824	
	Sc	0,177396	0,213549	0,211101	0,201131	0,18921	
	$\bar{d}_{T,G}$	0,250165	*0,09131*	0,107473	0,149956	0,202155	0,160212
log-norm	S0	0,231247	0,247959	0,23593	0,245008	0,220108	
	S1	0,201601	0,207159	0,202473	0,205531	0,185468	
	Sc	0,223954	0,239874	0,226731	0,23632	0,211604	
	$\bar{d}_{T,LN}$	**0,052087**	*0,015968*	0,040405	*0,015691*	0,110045	0,046839
S-M	S0	0,233806	0,242331	0,241984	0,241972	0,238097	
	S1	0,192363	0,208882	0,207921	0,20839	0,197767	
	Sc	0,225333	0,233017	0,233095	0,23275	0,225852	
	$\bar{d}_{T,SM}$	0,061197	***0,01274***	**0,014621**	**0,01438**	0,04607	**0,029802**
Daguma	S0	0,268126	0,257885	0,258045	0,257838	0,260087	
	S1	0,22209	0,221078	0,219151	0,220488	0,21475	
	Sc	0,26244	0,24859	0,248582	0,248533	0,247975	
	$\bar{d}_{T,D}$	0,084588	0,049582	*0,046756*	0,048509	**0,041779**	0,054243
log-logist	S0	0,331278	0,322633	0,308011	0,31431	0,294762	
	S1	0,279133	0,254556	0,253301	0,252675	0,243884	
	Sc	0,320986	0,305382	0,297108	0,299566	0,283866	
	$\bar{d}_{T,LL}$	0,342933	0,270552	0,236941	0,248037	*0,185418*	0,256776
dokładne	T_{S0}	0,242836					
	T_{S1}	0,211639					
	T_{Sc}	0,23853					

Źródło: opracowanie własne.

W przypadku współczynnika Theila, podobnie jak poprzednio, rozkłady log-logistyczny i Daguma w każdym przypadku przeszacowują rzeczywistą wartość. Błąd występujący przy wykorzystaniu funkcji log-logistycznej jest większy niż przy użyciu funkcji Daguma, przy czym różnica ta jest znaczna: średnio 5,4% w przypadku rozkładu Daguma i aż ponad 25% przy użyciu funkcji log-logistycznej. Rozkłady gamma i Singh-Maddala niedoszacowują rzeczywistą wartość współczynnika Theila, przy czym niedoszacowanie to jest rzędu kilku procent wartości rzeczywistej. Rozkład log-normalny również ma tendencję do niedoszacowania tej wielkości, choć nie w każdym przypadku. Podobnie jak poprzednio, metoda dająca największą zgodność z rzeczywistą wartością zależy od użytej funkcji. Warto zauważyć, że oprócz przypadku funkcji log-normalnej (zmiana z metody min$\chi^{2'}$ na minχ^2 jako najlepszej) są to te same metody, które okazały się najdokładniejsze w przypadku współczynnika Giniego. Najmniejszy błąd występuje przy wykorzystaniu funkcji Singh-Maddala, zarówno pod względem minimalnej wartości (z rozróżnieniem na metody), jak i po uśrednieniu po wszystkich zastosowanych sposobach aproksymacji. Warto podkreślić, że dla funkcji gamma i log-logistycznej występują duże błędy, nawet do 20% i więcej. Przy wykorzystaniu metody najmniejszych kwadratów powstają największe błędy. Odnosi się to do wszystkich funkcji oprócz log-normalnej, gdzie jeszcze większy błąd uzyskuje się przy użyciu „mechanicznej" metody dopasowywania funkcji.

Tabela 6.

Współczynnik Atkinsona $A_{0,1}$

rozkład \ metoda		MNK	minχ^2	min$\chi^{2'}$	MNW	MNK'	$\bar{\bar{d}}_{A_{0,1}}$
gamma	S0	0,0182159	0,0227392	0,0225956	0,0211917	0,019767	
	S1	0,016062	0,0190236	0,0183368	0,0177407	0,0167181	
	Sc	0,017779	0,021408	0,021162	0,020162	0,018965	
	$\bar{d}_{A_{0.1},G}$	0,23817	*0,076543*	0,092993	0,136216	0,189327	0,14665
log-norm	S0	0,0228594	0,024491	0,0233169	0,0242031	0,02177	
	S1	0,019958	0,020503	0,020044	0,020343	0,018376	
	Sc	0,022147	0,023702	0,022418	0,023355	0,020938	
	$\bar{d}_{A_{0,1},LN}$	**0,049471**	0,015687	0,037897	*0,013586*	0,107016	0,044732
S-M	S0	0,0231248	0,0239342	0,0238998	0,0239002	0,023527	
	S1	0,019122	0,020622	0,020521	0,020572	0,0196039	
	Sc	0,022313	0,023017	0,023024	0,022992	0,022355	
	$\bar{d}_{A_{0,1},SM}$	0,05677	***0,011581***	**0,013585**	**0,01321**	0,04289	**0,027607**
Daguma	S0	0,0262859	0,0253446	0,0253594	0,0253402	0,0255539	
	S1	0,021897	0,0217365	0,0215367	0,0216734	0,0211909	
	Sc	0,025736	0,024431	0,024429	0,024425	0,024402	
	$\bar{d}_{A_{0,1},D}$	0,080025	0,045863	0,042846	0,044712	***0,039648***	0,050619
log-logist	S0	0,0319743	0,031169	0,0298019	0,030391	0,0285601	
	S1	0,027091	0,02477	0,024651	0,024592	0,023758	
	Sc	0,031015	0,029556	0,02878	0,029011	0,027536	
	$\bar{d}_{A_{0,1},LL}$	0,316885	0,247931	0,216042	0,226551	*0,166883*	0,234859
dokładne	$A_{0,1,S0}$	0,023984					
	$A_{0,1,S1}$	0,02087					
	$A_{0,1,Sc}$	0,023506					

Źródło: opracowanie własne.

Tabela 7.

Współczynnik Atkinsona $A_{0,5}$

rozkład \ metoda		MNK	minχ^2	min$\chi^{2'}$	MNW	MNK'	$\bar{\bar{d}}_{A_{0,5}}$
gamma	S0	0,0919735	0,114983	0,114252	0,107109	0,099859	
	S1	0,0810291	0,0960805	0,0925882	0,0895583	0,0843616	
	Sc	0,089753	0,108209	0,106959	0,101868	0,095782	
	$\bar{d}_{A_{0,5},G}$	0,193765	*0,0216*	0,039081	0,085137	0,141748	0,096266
log-norm	S0	0,10919	0,116602	0,111273	0,115298	0,104214	
	S1	0,095887	0,098396	0,096281	0,097662	0,088564	
	Sc	0,105935	0,113024	0,107176	0,111446	0,100397	
	$\bar{d}_{A_{0,5},LN}$	0,045733	0,011161	0,034672	***0,007557***	0,101253	0,040075
S-M	S0	0,111407	0,114734	0,114567	0,114595	0,112974	
	S1	0,093814	0,098698	0,098097	0,098429	0,0951848	
	Sc	0,107949	0,110381	0,110399	0,110284	0,107988	
	$\bar{d}_{A_{0,5},SM}$	**0,040264**	***0,00697***	**0,009419**	0,008564	**0,031001**	**0,019244**
Daguma	S0	0,123172	0,119817	0,119872	0,119801	0,120693	
	S1	0,104623	0,102752	0,101611	0,10236	0,101539	
	Sc	0,120705	0,115441	0,115406	0,11541	0,11587	
	$\bar{d}_{A_{0,5},D}$	0,068159	0,036461	*0,032685*	0,035006	0,036206	0,041703
log-logist	S0	0,141924	0,138742	0,133305	0,135655	0,128317	
	S1	0,122358	0,112816	0,112323	0,112077	0,108607	
	Sc	0,138133	0,13232	0,129205	0,130132	0,124171	
	$\bar{d}_{A_{0,5},LL}$	0,234019	0,175428	0,148726	0,157476	*0,106783*	0,164486
dokładne	$A_{0,5,S0}$	0,114967					
	$A_{0,5,S1}$	0,099267					
	$A_{0,5,Sc}$	0,111851					

Źródło: opracowanie własne.

Tabela 8.

Współczynnik Atkinsona $A_{0,8}$

metoda / rozkład		MNK	minχ^2	min$\chi^{2'}$	MNW	MNK’	$\bar{\bar{d}}_{A_{0,8}}$
gamma	S0	0,148375	0,185777	0,184588	0,172971	0,161185	
	S1	0,130612	0,155046	0,149373	0,144453	0,136018	
	Sc	0,14477	0,174759	0,172727	0,16445	0,15456	
	$\bar{d}_{A_{0,8},G}$	0,164318	*0,015482*	0,021289	0,050983	0,110077	0,07243
log-norm	S0	0,168894	0,179931	0,172002	0,177993	0,161455	
	S1	0,148947	0,152723	0,149541	0,151619	0,137892	
	Sc	0,164031	0,17461	0,165886	0,17226	0,15573	
	$\bar{d}_{A_{0,8},LN}$	0,049615	*0,004253*	0,039006	0,011174	0,103274	0,04146
S-M	S0	0,174535	0,179172	0,178909	0,17898	0,176614	
	S1	0,148722	0,153919	0,152863	0,15347	0,149848	
	Sc	0,169577	0,172407	0,17242	0,172278	0,169468	
	$\bar{d}_{A_{0,8},SM}$	**0,029008**	***0,003967***	**0,006473**	**0,005298**	**0,02292**	**0,013533**
Daguma	S0	0,1902	0,18613	0,186199	0,18611	0,187381	
	S1	0,163526	0,159334	0,157326	0,158614	0,158985	
	Sc	0,186463	0,179192	0,179108	0,179138	0,180498	
	$\bar{d}_{A_{0,8},D}$	0,064487	0,033921	*0,029538*	0,032221	0,037989	0,039631
log-logist	S0	0,211539	0,207122	0,199541	0,202824	0,192551	
	S1	0,184152	0,170598	0,169894	0,169542	0,164576	
	Sc	0,206275	0,198163	0,193797	0,195099	0,186713	
	$\bar{d}_{A_{0,8},LL}$	0,186731	0,133627	0,109669	0,117493	*0,071603*	0,123825
dokładne	$A_{0,8,S0}$	0,179635					
	$A_{0,8,S1}$	0,153862					
	$A_{0,8,Sc}$	0,173965					

Źródło: opracowanie własne.

W przypadku wszystkich trzech rozważanych tutaj współczynników Atkinsona można zauważyć, że rozkład Daguma oraz log-logistyczny konsekwentnie przeszacowują te wielkości, rozkład Singh-Maddala konsekwentnie niedoszacowuje, natomiast log-normalny oraz gamma mają tendencje do niedoszacowywania, aczkolwiek z wyjątkami. Rozkład Daguma daje zawsze dużo lepszą dokładność niż rozkład log-logistyczny. Metoda najmniejszych kwadratów, jak poprzednio, nie jest optymalna dla żadnej z funkcji, a pozostałe metody zachowują się różnie w zależności od funkcji.

5. WNIOSKI I PODSUMOWANIE

W niniejszej pracy zbadane zostało dopasowanie do danych empirycznych dotyczących dochodów pięciu różnych typów rozkładów za pomocą pięciu różnych metod. Miarą dobroci dopasowania było z jednej strony tradycyjnie stosowane średnie odchylenie kwadratowe, z drugiej – dokładność oszacowania kilku wielkości, których ścisłe wartości są znane. Okazuje się, że najlepsze dopasowanie do danych empirycznych, mierzone wielkością średniego odchylenia kwadratowego, nie zawsze przekłada się na największą dokładność oszacowania takich wielkości, jak średnia czy miary nierówności. Niektóre z otrzymanych tutaj rezultatów potwierdzają spostrzeżenia McDonalda i Ransoma (1979) poczynione na podstawie danych angielskich: że dla rozkładów log-normalnego i Singh-Maddala metoda najmniejszych kwadratów daje zazwyczaj najmniejszą wartość współczynnika Giniego (w przypadku rozkładu Daguma i log-logistycznego taka zależność nie zachodzi, ale te rozkłady nie były przez McDonalda i Ransoma badane), lub że przy aproksymacji rozkładem Singh-Maddala otrzymuje się przeważnie zaniżoną wartość współczynnika Giniego. Inne spostrzeżenia przeczą tym, które poczynił McDonald, przykładowo, dla przypadków badanych w niniejszej pracy nie jest prawdą, jak u McDonalda i Ransoma, iż największą wartość średnią dla rozkładu gamma uzyskuje się przy użyciu metody najmniejszych kwadratów.

Ponieważ analizowane w tej pracy dane dotyczą tylko jednego okresu w jednej jednostce administracyjnej kraju, trudno byłoby otrzymane wyniki generalizować. Niemniej, mogą one posłużyć co najmniej jako sugestie pewnych efektów.

Po pierwsze, stosowane powszechnie miary dobroci dopasowania rozkładów mogą nie mieć jednoznacznego przełożenia na dokładność otrzymywanych wyników. Metoda, przy której otrzymuje się najmniejsze (po MNK') średnie odchylenie kwadratowe czasami tylko pokrywa się z metodą dającą najlepsze oszacowanie którejś z wielkości.

Po drugie, metoda najmniejszych kwadratów nigdy (poza jednym przypadkiem) nie jest najlepszą metodą szacowania wartości średniej, odchylenia standardowego i miar nierówności; również metoda największej wiarygodności rzadko tylko okazuje się najlepsza. Wydaje się, że na większą uwagę zasługuje stosunkowo rzadko wykorzystywana metoda minimalizacji wartości χ^2. Ponadto, mechaniczna metoda dopasowywania krzywej do danych empirycznych, metoda MNK', okazuje się być w każdym przypadku najlepszą dla rozkładu log-logistycznego.

Po trzecie, dwa rozkłady zwracają uwagę ze względu na dokładność oszacowań, w większości przypadków znacznie przewyższając pozostałe. Są to rozkład Singh-Maddala i rozkład Daguma. Jak można było oczekiwać, ponieważ rozkład Daguma jest uogólnieniem rozkładu log-logistycznego, w każdym rozpatrywanym przypadku powinien on dawać wyniki lepsze niż ten ostatni. Okazuje się, że poprawa dokładności w wyniku zastąpienia rozkładu log-logistycznego rozkładem Daguma jest bardzo duża. Wydaje się zatem, że warto korzystać z rozkładu Daguma zamiast log-logistycznego, gdyż kosztem wprowadzenia jednego dodatkowego parametru zyskujemy znacznie lepsze dopasowanie do rzeczywistych danych, czy to mierzone średnim odchyleniem kwadratowym, czy też dokładnością oszacowań różnych miar charakteryzujących rozkład.

Na koniec, wybierając jeden z tych dwóch najdokładniejszych rozkładów: Daguma lub Singh-Maddala, warto pamiętać, iż obok porównywalnej dokładności różnią się wyraźnie charakterem swoich oszacowań. Rozkład Daguma konsekwentnie przeszacowuje wszystkie badane wielkości, rozkład Singh-Maddala natomiast – niedoszacowuje. Ten ostatni wniosek, dotyczący rozkładu Singh-Maddala, znajduje potwierdzenie również w wynikach McDonalda i Ransoma (rozkład Daguma nie był przez nich badany). Ponieważ efekt ten uzyskano dla wszystkich pięciu rozpatrywanych miar nierówności wydaje się on prawdopodobny w odniesieniu do szeroko ujmowanej nierówności i możliwych jej miar.

Uniwersytet Ekonomiczny we Wrocławiu

LITERATURA

[1] Aitchison J., Brown J. A. C., (1969), *The Lognormal Distribution*, Cambridge University Press, Cambridge.
[2] Champernowne D. G., (1953), A Model of Income Distribution, *The Economic Journal*, 63, 318–351.
[3] Cox D. R., Hinckley D. V., (1974), *Theoretical Statistics*, Chapman and Hall, London.
[4] Dagum C., (1977), A New Model of Personal Income Distribution: Specification and Estimation, *Economie Appliquée,* 30, 413–437.
[5] Dagum C., Lemmi A., (1987), *A Contribution to the Analysis of Income Distribution and Income Inequality, and a case study: Italy*, University of Ottawa, Department of Economics.
[6] Fisk P. R., (1961), The Graduation of Income Distributions, *Econometrica,* 29, 171–185.
[7] Gastwirth J. L., (1972), The Estimation of the Lorenz Curve and Gini Index, *The Review of Economics and Statistics,* 54, 306–316.
[8] Kleiber Ch., (1996), Dagum vs. Singh-Maddala Income Distributions, *Economics Letters*, 53, 265–268.
[9] Kośny M., (2001), Probabilistyczne modele rozkładu dochodów, *Wiadomości Statystyczne*, 7, 20–35.
[10] Kot S. M. (red.), (1999), *Analiza ekonometryczna kształtowania się płac w Polsce w okresie transformacji*, Wydawnictwo Naukowe PWN.
[11] Kot S. M., (1995), Probabilistyczny model dystrybucji dochodu wraz ze współwariantami, *Przegląd Statystyczny,* 42, 155–180.
[12] Kot S. M., Adamkiewicz-Drwiłło H., (2013), Rekonstrukcja światowego rozkładu dochodów na podstawie minimalnej informacji statystycznej, *Śląski Przegląd Statystyczny,* 11 (17), 179–200.
[13] McDonald J.B., Ransom M.R., (1979), Functional Forms, Estimation Techniques and the Distribution of Income, *Econometrica*, 47, 1513–1525.
[14] Neyman J., (1949), Contribution to the Theory of the χ^2 Test, *Proceedings of the First Berkeley Symposium on Mathematical Statistics and Probability,* 1, 239–273.

[15] Panek T., Szulc A., (1991), *Income Distribution and Poverty. Theory and a Case Study of Poland in the Eighties*, Research Centre for Statistical and Economic Analysis of the Central Statistical Office and Academy of Science.

[16] Salem A. B. Z., Mount T. D., (1974), A Convenient Descriptive Model of Income Distribution: The Gamma Density, *Econometrica,* 42, 1115–1127.

[17] Singh S. K., Maddala G. S., (1976), A Function for Size Distribution of Incomes, *Econometrica,* 44, 963–970.

[18] Stuart A., Ord J. K., (1991), *Classical Inference and Relationship, Vol. 2., Kendall's Advanced Theory of Statistics,* Edward Arnold, New York.

[19] Thurow L. C., (1970), Analyzing the American Income Distribution, *American Economic Review,* 60, 261–269.

[20] Ulman P., (2011), *Rozkłady dochodów gospodarstw domowych,* w: *Sytuacja osób niepełnosprawnych i ich gospodarstw domowych w Polsce,* Wydawnictwo Uniwersytetu Ekonomicznego w Krakowie.

ADEKWATNOŚĆ WYBRANYCH ROZKŁADÓW TEORETYCZNYCH DOCHODÓW W ZALEŻNOŚCI OD METODY APROKSYMACJI

Streszczenie

Rozkłady dochodów modelowane są za pomocą wielu rozkładów teoretycznych, których parametry wyznaczane są za pomocą różnych metod. Wśród rozkładów tych wymienić można rozkład logarytmiczno-normalny, gamma, czy logarytmiczno-logistyczny. Najczęściej stosowanymi metodami są metoda najmniejszych kwadratów oraz metoda największej wiarygodności. Miarą dobroci dopasowania rozkładu teoretycznego jest zazwyczaj średnie odchylenie kwadratowe lub wartość statystyki χ^2. Tak zdefiniowana jakość dopasowania nie musi jednakże dokładnie przekładać się na jakość oszacowania takich charakterystyk rozkładu, jak wartość średnia czy miary nierówności. Celem pracy jest aproksymacja dochodów różnymi rozkładami teoretycznymi (log-normalny, gamma, log-logistyczny, Daguma i Singh-Maddala) oraz różnymi technikami i porównanie dobroci dopasowania mierzonej jako średnie odchylenie kwadratowe z dokładnością oszacowania kilku charakterystyk liczbowych rozkładu (wartości średniej, odchylenia standardowego, współczynnika Giniego, współczynnika Theila i trzech współczynników Atkinsona), których wartości porównywane są z wartościami dokładnymi, policzonymi na podstawie danych niezgrupowanych.

Słowa kluczowe: teoretyczny rozkład dochodów; aproksymacja rozkładu empirycznego; dokładność dopasowania

ADEQUACY OF SOME CHOSEN THEORETICAL DISTRIBUTIONS CONDITIONED ON THE METHOD OF APPROXIMATION

Abstract

There are various theoretical distributions which are used as models for the distribution of income. Among them the most commonly used is probably log-normal distribution, but also gamma distribution or log-logistic one. There are also a number of approximation methods of these distributions. The goodness of fit is commonly measured by mean squared deviation or value of χ^2 statistics. However, such measures of quality of approximation do not necessarily coincide with accuracy of some distribution characteristics, like inequality measures. The aim of this paper is to investigate the goodness of approximation of income distribution, given in the form of frequency distribution, by means of chosen theoretical distributions, namely, log-normal, gamma, log-logistic, Dagum and Singh-Maddala distribution. The goodness is measured both by mean squared error and deviation of some distribution characteristics. Values calculated on ungrouped data are used as a reference for comparisons.

Keywords: theoretical income distribution; approximation of distribution; goodness of fit