

**Kamila TRZCIŃSKA<sup>1</sup>**

## Aproksymacja rozkładu dochodów ludności Polski za pomocą modeli Daguma i Zengi

**Streszczenie.** *Celem pracy jest analiza dochodów ludności Polski za pomocą rozkładów Daguma i Zengi. Rozkład Zengi zaproponowany w 2010 r. jest nowym, konkurencyjnym rozkładem, który nie był dotychczas stosowany do badania płac i dochodów w Polsce. W artykule przedstawiono rozkłady Daguma i Zengi oraz wyniki aproksymacji rozkładów płac pochodzących z badania budżetów gospodarstw domowych ludności Polski w 2014 r. Obliczone miary zgodności dopasowania rozkładów teoretycznych do empirycznych wskazują, że rozkład Zengi lepiej opisuje rozkład płac ludności Polski niż – uważany dotychczas za jeden z najlepszych – rozkład Daguma.*

**Słowa kluczowe:** rozkład dochodów, model Daguma, model Zengi, metody estymacji

## Approximation of distribution of Poland's population's income according to Dagum and Zenga models

**Abstract.** *The aim of this paper is to analyse the income of the population of Poland using the Dagum and the Zenga distributions. The Zenga distribution, introduced in 2010, is a new distribution which has not been yet applied to analysing wages and income in Poland. The paper presents both the Dagum and Zenga distributions, as well as the results of the approximations of wage distributions drawn from the 2014 household budget survey. The calculated degree of compliance of the theoretical distribution with the empirical one demonstrates that the Zenga distribution describes the distribution of the income of the Polish population more reliably than the Dagum distribution, which so far has been regarded as one of the best.*

**Keywords:** income distribution, Dagum distribution, Zenga distribution, estimation methods

**JEL Classification:** C1, C10, C13, C15

---

<sup>1</sup> Uniwersytet Łódzki, Wydział Ekonomiczno-Socjologiczny, Katedra Metod Statystycznych, ul. Rewolucji 1905 r. 41/43, 90-214 Łódź, Polska, e-mail: kamila.trzcinska@uni.lodz.pl, ORCID: <https://orcid.org/0000-0002-4714-4074>.

## 1. WPROWADZENIE

Pierwszą usystematyzowaną teorię dotyczącą rozkładu i nierównomierności dochodów przedstawił Pareto w przełomowej pracy z 1897 r. Następnie pojawiły się prace dotyczące teoretycznych rozkładów płac i dochodów. March (1898) zaproponował wykorzystanie rozkładu gamma, a Edgeworth (1898) – zastosowanie rozkładu normalnego jako funkcji generującej rozkłady dochodów. Podejście to przyczyniło się do stworzenia m.in. przez D'Addaria (1939) pierwszego systemu funkcji gęstości opisujących rozkłady płac i dochodów. Zagadnienie rozwoju badań dotyczących dochodów i bogactwa na przestrzeni lat poruszają Domański i Jędrzejczak (2005). Obecnie badania rozkładów dochodów skupiają się głównie na próbach dopasowania rozkładów teoretycznych do empirycznego rozkładu płac i dochodów w różnych przekrojach oraz na analizie nierównomierności tych rozkładów.

Istnieje wiele teoretycznych rozkładów dochodów, które zostały sformułowane na podstawie obserwacji empirycznych (rozkład Pareta, rozkład Daguma) lub też jako wynik pewnego procesu stochastycznego (rozkład Champernowne'a). Przegląd większości z nich przedstawiono w pracy Kleibera i Kotza (2003). Dagum (1990) określił kryteria, które powinny spełniać teoretyczne modele rozkładów dochodów. Do najważniejszych z nich należą:

- zbieżność modelu do rozkładu Pareta typu I dla wysokich dochodów;
- istnienie małej liczby skończonych momentów;
- ekonomiczna interpretacja parametrów;
- możliwość dopasowania modelu do dwóch rodzajów rozkładów dochodów (w zależności od wartości parametrów powinien generować rozkłady jednomodalne lub zeromodalne).

Wiele rozkładów stosowanych w analizie płac i dochodów wybrano ad hoc, ze względu na ich kształt (rozkład gamma, rozkład beta, krzywą Pearsona V i VI typu). Empiryczne rozkłady bardzo często cechują się bowiem jednomodalnością, prawostronną asymetrią oraz dodatnią kurtozą. Są to prawidłowości, które zaobserwowano niezależnie od kraju oraz czasu. Do opisywania dochodów empirycznych poszukuje się takich rozkładów teoretycznych, które przy odpowiedniej parametryzacji odzwierciedlają te własności.

Zastosowanie rozkładów teoretycznych do analizy płac i dochodów jest uzasadnione z wielu powodów. Podejście takie umożliwia wyrównywanie nieregularności w obserwowanych rozkładach, które powstały w wyniku gromadzenia danych. Ponadto ułatwia analizę rozkładów, ponieważ wszystkie charakterystyki rozkładów teoretycznych, w tym także współczynniki nierównomierności, mogą być wyrażone za pomocą tych samych parametrów. Znajomość parametrów rozkładu może być też wykorzystana do porównania tych samych rozkładów w różnych okresach oraz do konstrukcji modeli rozkładów dochodów. Dane, którymi dysponujemy, pochodzą na ogół z badań częściowych, dlatego znajomość funkcji gęstości lub dystrybuanty teoretycznej ułatwia proces estymacji

miar nierównomierności dla zbiorowości generalnej. Wielu badaczy koncentruje się na problemach porównań rozkładów dochodów w czasie i przestrzeni. Stosowanie rozkładów teoretycznych do opisu empirycznych rozkładów dochodów stwarza możliwość prognozowania.

Do aproksymacji rozkładów dochodów ludności Polski stosowano różne rozkłady teoretyczne, np. rozkłady Pareta, logarytmiczno-normalny, gamma, Daguma i wiele innych (Wiśniewski, 1934; Vielrose, 1960; Lange, 1967; Wąsik, 1967; Kordos, 1968, 1973; Domański, 1973; Kot, 1999, 2000). Zastosowaniem dwuparametrowego rozkładu log-normalnego zajmowali się Kordos (1990) oraz Jagielski i Kutner (2010).

Badania nad rozkładami dochodów i płac w Polsce wskazują, że do aproksymacji empirycznych rozkładów płac i dochodów mogą być z powodzeniem stosowane rozkłady teoretyczne trójparametrowe – w szczególności rozkłady Singha-Maddali i Daguma dobrze dopasowują się do danych empirycznych w różnych przekrojach (Łukasiewicz, Orłowski, 2004; Ostasiewicz, 2013; Salamağa, 2016). Rozkład Daguma, należący do krzywych Burra III typu (Kleiber, Kotz, 2003), jest najczęściej stosowany nie tylko w Polsce. Jest on nieco lepiej dopasowany od rozkładu Singha-Maddali, należącego do krzywych Burra XII typu (Jędrzejczak, 1999), a oprócz tego jest zbieżny z krzywą Pareta dla wysokich grup dochodowych.

W 2010 r. Zenga zaproponował trójparametrowy rozkład oparty na uciętym rozkładzie Pareta przedstawionym w pracy Poliscichia (2008), który bardzo dobrze aproksymuje rozkłady dochodów wielu krajów. Rozkład ten nie był dotąd stosowany do analizy rozkładów ludności Polski. Funkcja gęstości Zengi, dzięki swoim własnościom, łatwo dopasowuje się do różnego rodzaju dochodów.

Do opisu rozkładów dochodów ludności Polski stosowano również czteroparametrowy rozkład GB2. Badania prowadzone przez Banduriana i in. (2003), Dastrupa i in. (2007) oraz Brzezińskiego (2013) potwierdzają bardzo dobre dopasowanie tego rozkładu do danych empirycznych.

W niniejszym artykule przedstawiono porównanie rozkładu Daguma, uważanego do tej pory za rozkład trójparametrowy najlepiej opisujący rozkłady empiryczne w Polsce, z rozkładem Zengi, który pozwala na bezpośrednią interpretację parametrów.

## 2. TEORETYCZNE ROZKŁADY DOCHODÓW

### 2.1. ROZKŁAD DAGUMA

Rozkład Daguma (Dagum, 1977) powstał na podstawie obserwacji empirycznych i jest jednym z najczęściej stosowanych do opisu rozkładów płac i dochodów. Rozkład ten może być zarówno jednomodalny, jak i zeromodalny. W przypadku wysokich wartości dochodów rozkład Daguma jest zbieżny z rozkładem Pareta, dzięki czemu posiada niewielką liczbę skończonych momentów, co sprawia, że charakteryzuje się wysoką zgodnością z danymi empirycznymi. Do

opisu rozkładu dochodów w Polsce został po raz pierwszy zastosowany przez Jędrzejczak (1990). W artykule zostanie rozważony rozkład Daguma typu I (Kleiber, Kotz, 2003). Funkcja gęstości rozkładu Daguma ma postać:

$$f(x; a; b; p) = \frac{apx^{ap-1}}{b^{ap} \left[ 1 + \left(\frac{x}{b}\right)^a \right]^{p+1}}, x > 0, \tag{1}$$

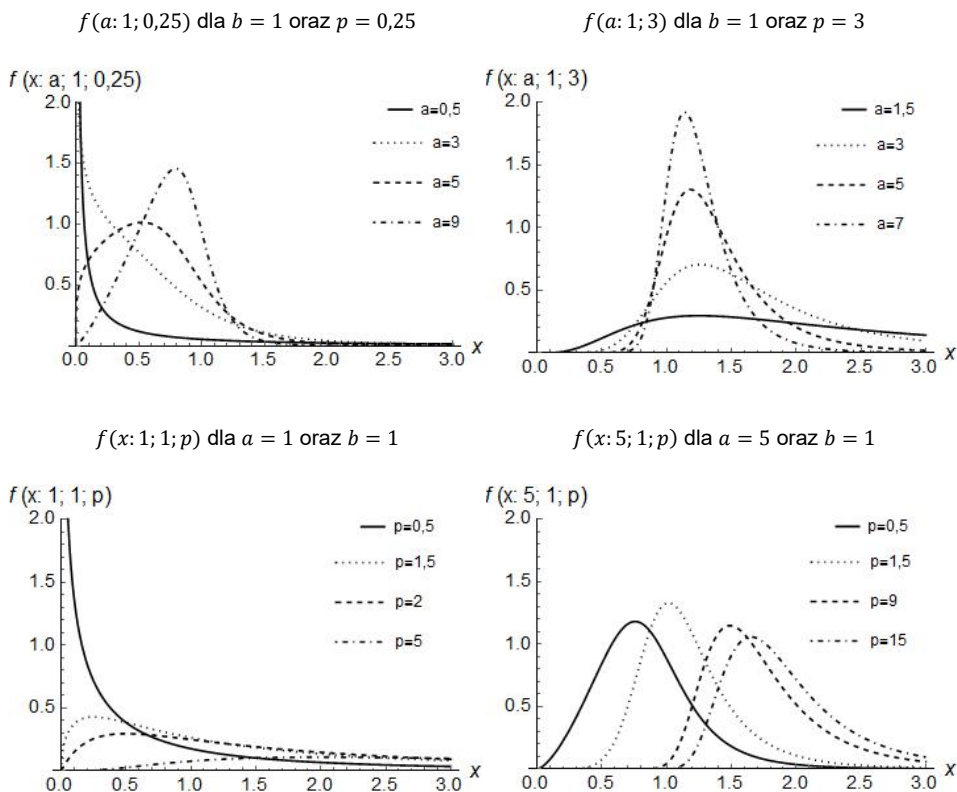
gdzie  $a, b, p > 0$ .

Natomiast dystrybuanta rozkładu Daguma jest następująca:

$$F(x; a; b; p) = \left[ 1 + \left(\frac{x}{b}\right)^a \right]^{-p}, x > 0, \tag{2}$$

gdzie  $a, b, p > 0$ . Parametr  $b$  jest parametrem skali, natomiast  $a$  oraz  $p$  są parametrami kształtu.

Rysunek 1. Funkcja gęstości rozkładu Daguma



## 2.2. ROZKŁAD ZENGI

W 2010 r. Zenga skonstruował trzyparametrową funkcję gęstości  $f(x; \mu; \alpha; \theta)$ , ( $\mu > 0; \alpha > 0; \theta > 0$ ) dla nieujemnych wartości dochodów (Zenga, 2010; Zenga i in., 2010a, 2010b), która powstała w oparciu o rozkład Pareta przedstawiony przez Polisicchia (2008):

$$v(x; \mu; k) = \begin{cases} \frac{\sqrt{\mu}}{2} k^{0,5} (1-k)^{-1} x^{-1,5} & \text{dla } \mu k \leq x \leq \frac{\mu}{k}; \quad \mu > 0, \quad 0 < k < 1 \\ 0 & \text{dla pozostałych przypadków,} \end{cases} \quad (3)$$

oraz rozkład beta

$$g(k; \alpha; \theta) = \begin{cases} \frac{k^{\alpha-1} (1-k)^{\theta-1}}{\beta(\alpha; \theta)} & \text{dla } 0 < k < 1; \quad \theta > 0, \alpha > 0 \\ 0 & \text{dla pozostałych przypadków,} \end{cases} \quad (4)$$

gdzie  $B(\alpha; \theta)$  jest funkcją beta.

Funkcja gęstości rozkładu Zengi ma postać:

$$\begin{aligned} f(x; \mu; \alpha; \theta) &= \int_0^1 v(x; \mu, k) g(k; \alpha, \theta) dk \\ &= \begin{cases} \frac{1}{2\mu B(\alpha; \theta)} \left(\frac{x}{\mu}\right)^{-1,5} \int_0^{\frac{x}{\mu}} k^{\alpha-0,5} (1-k)^{\theta-2} dk & \text{dla } 0 < x < \mu \\ \frac{1}{2\mu B(\alpha; \theta)} \left(\frac{\mu}{x}\right)^{1,5} \int_0^{\frac{\mu}{x}} k^{\alpha-0,5} (1-k)^{\theta-2} dk & \text{dla } \mu < x. \end{cases} \end{aligned} \quad (5)$$

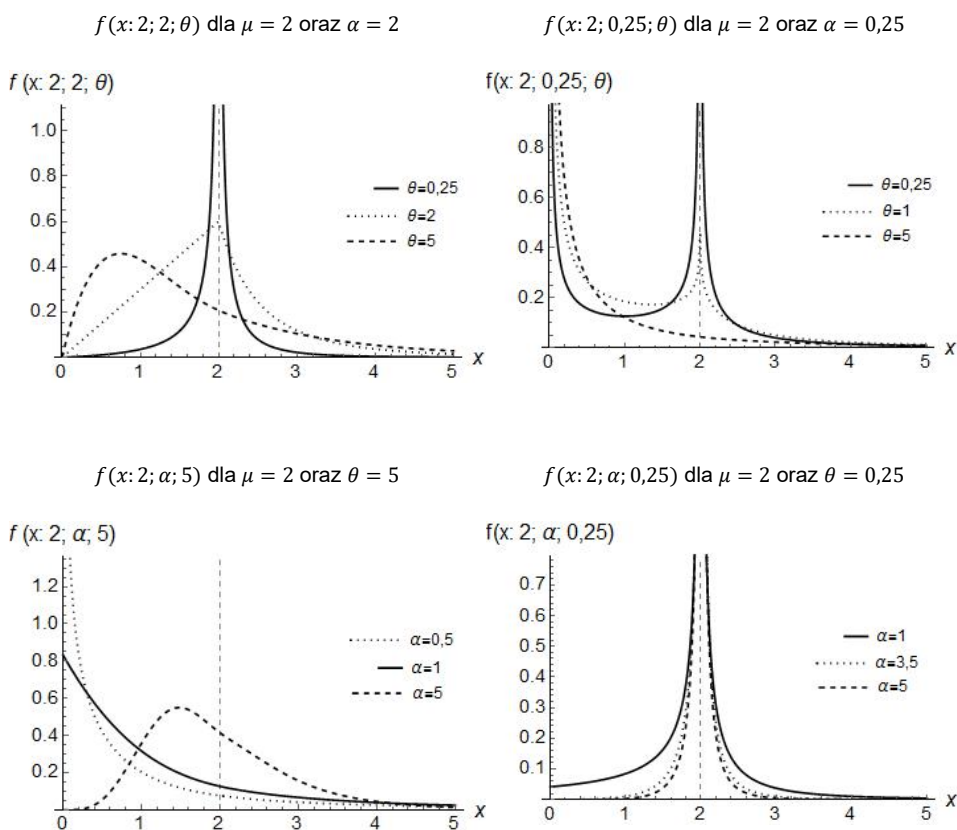
W przypadku gdy parametr  $\theta > 0$ , wówczas dystrybuanta rozkładu Zengi jest następująca:

$$\begin{aligned} F(x; \mu; \alpha; \theta) &= \\ &= \begin{cases} \frac{1}{B(\alpha, \theta)} \sum_{i=1}^{\infty} \left\{ IB\left(\frac{x}{\mu}; \alpha + i - 1; \theta\right) - \left(\frac{\mu}{x}\right)^{0,5} IB\left(\frac{x}{\mu}; \alpha + i - 0,5; \theta\right) \right\} & \text{dla } 0 < x \leq \mu \\ 1 - \frac{1}{B(\alpha, \theta)} \sum_{i=1}^{\infty} \left\{ \left(\frac{\mu}{x}\right)^{0,5} IB\left(\frac{\mu}{x}; \alpha + i - 0,5; \theta\right) - IB\left(\frac{\mu}{x}; \alpha + i; \theta\right) \right\} & \text{dla } \mu < x, \end{cases} \end{aligned} \quad (6)$$

gdzie  $IB(x; \alpha; \theta)$  jest niekompletną funkcją beta. Ponadto  $F(\mu; \mu; \alpha; \theta) \geq \frac{1}{2}$ , dla  $\alpha > 0; \theta > 0$ , dlatego też rozkład ma dodatnią asymetrię.

Łatwo zauważyć, że dla wszystkich wartości parametrów  $\mu > 0$ ;  $\alpha > 0$ ;  $\theta > 0$  wartość oczekiwana rozkładu Zengi wynosi  $\mu$ . Liczba skończonych momentów rozkładu jest określona jako mniejsza niż  $\alpha + 1$ , tak więc dla typowych wartości tego parametru 2 lub 3 rozkład Zengi posiada 3 lub 4 momenty. Funkcja gęstości rozkładu Zengi przyjmuje różne kształty, bardziej różnorodnie niż tradycyjne trójparametrowe modele rozkładów dochodów, jak np. rozkład Daguma (por. rys. 1 i rys. 2), a cecha ta pozwala na dobre dopasowanie także w przypadku niskich grup dochodowych.

Rysunek 2. Funkcja gęstości rozkładu Zengi



Źródło: opracowanie własne.

Parametr  $\mu$  jest parametrem skali rozkładu Zengi, natomiast  $\alpha$  i  $\theta$  są parametrami kształtu. W pracy zastosujemy trójparametrową funkcję gęstości Zengi do opisu empirycznego rozkładu dochodów ludności Polski na podstawie danych indywidualnych dochodu rozporządzalnego netto w gospodarstwach domowych w 2014 r. Źródłem danych są badania budżetów gospodarstw domowych prowadzone przez Główny Urząd Statystyczny.

### 3. METODY ESTYMACJI PARAMETRÓW ROZKŁADU ZENGI

Do estymacji parametrów funkcji Zengi  $f(x; \mu; \alpha; \theta)$  użyjemy m.in. metod zaproponowanych przez D'Addaria (1934, 1939) oraz metody momentów. Przyjmujemy, że parametr  $\mu = E(X)$  jest estymowany przez średnią z próby  $\bar{x}$ . Wówczas znalezienie parametrów funkcji Zengi  $\alpha$  oraz  $\theta$  polega na rozwiązaniu odpowiednich układów równań (Zenga i in., 2010a). Przyjmijmy następujące oznaczenia dla każdego z pięciu wariantów metody D'Addaria przedstawionych przez Zengę (2010):

- metoda 1:  $F(\hat{m}: \bar{x}, \alpha, \theta) = 0,5$  oraz  $F(\bar{x}: \bar{x}, \alpha, \theta) = \hat{F}(\bar{x})$ , gdzie  $\hat{m}$  jest empiryczną medianą, a  $\hat{F}(\bar{x})$  – empiryczną dystrybuantą obliczoną dla średniej;
- metoda 2:  $F(\hat{m}: \bar{x}, \alpha, \theta) = 0,5$  oraz  $P(\bar{x}; \alpha; \theta) = \hat{P}$ , gdzie  $P(\bar{x}; \alpha; \theta) = 2F(1; 1; \alpha; \theta) - 1$  jest indeksem Pietry, a  $\hat{P} = \frac{1}{2\bar{x}n} \sum_{i=1}^n |x_i - \bar{x}|$  – empirycznym indeksem Pietry;
- metoda 3:  $F(\hat{m}: \bar{x}, \alpha, \theta) = 0,5$  oraz  $Var(\bar{x}, \alpha, \theta) = m_2$ , gdzie  $m_2$  jest wariancją oraz  $Var(\bar{x}, \alpha, \theta) = \frac{\mu^2 \theta (\theta + 1)}{3(\alpha - 1)(\alpha + \theta)}$ ;
- metoda 4:  $F(\bar{x}: \bar{x}, \alpha, \theta) = \hat{F}(\bar{x})$  oraz  $Var(\bar{x}, \alpha, \theta) = m_2$ ;
- metoda 5:  $F(\hat{m}: \bar{x}, \alpha, \theta) = 0,5$  oraz  $A(\bar{x}; \alpha; \theta) = \hat{A}(\bar{x})$ , gdzie  $A(\bar{x}; \alpha; \theta) = 1 - \left\{ \frac{1 - F(1; 1; \alpha; \theta)}{F(1; 1; \alpha; \theta)} \right\}^2$  jest indeksem punktowym Zengi (2007) dla  $\bar{x}$  oraz  $\hat{A}(\bar{x}) = \frac{\sum_{x_i \leq \bar{x}} x_i}{\hat{F}(\bar{x})} \times \frac{1 - \hat{F}(\bar{x})}{\sum_{x_i > \bar{x}} x_i}$ .

Zauważmy, że w celu wyznaczenia parametrów funkcji Zengi wszystkie z pięciu metod nakładają po trzy ograniczenia. Narzucając jedynie dwa ograniczenia –  $\mu = \bar{x}$  oraz:  $F(\hat{m}: \bar{x}, \alpha, \theta) = 0,5$ ,  $F(\bar{x}: \bar{x}, \alpha, \theta) = \hat{F}(\bar{x})$ ,  $P(\bar{x}; \alpha; \theta) = \hat{P}$ ,  $A(\bar{x}; \alpha; \theta) = \hat{A}(\bar{x})$ , otrzymujemy inwariantną metodę D'Addaria.

Zenga i in. (2010c) otrzymali analityczną metodę estymacji parametrów funkcji gęstości modelu Zengi poprzez rozwiązanie następującego układu równań:

$$\begin{cases} \mu = \bar{x} \\ \frac{\bar{x}^2 \theta (\theta + 1)}{3(\alpha - 1)(\alpha + \theta)} = m_2 \\ \frac{3\bar{x} m_2 (\theta + 3)(\theta + 2)}{5(\alpha + \theta + 1)(\alpha - 2)} = m_3. \end{cases}$$

Lewe strony równań oznaczają momenty teoretyczne, a  $\bar{x}$ ,  $m_2$  oraz  $m_3$  są odpowiednio średnią arytmetyczną, wariancją oraz trzecim momentem centralnym

z próby. Aby układ równań miał rozwiązanie, trzeci moment centralny musi być skończony, czyli  $\alpha > 2$ . Wówczas estymowane parametry mają postać:

$$\hat{\mu} = \bar{x},$$

$$\hat{\theta}$$

$$= \frac{-\left[\frac{1}{3} \frac{\bar{x}^2}{m_2} - 3 \frac{\bar{x}m_2}{m_3} - 1\right] + \sqrt{\left[\frac{1}{3} \frac{\bar{x}^2}{m_2} - 3 \frac{\bar{x}m_2}{m_3} - 1\right]^2 + 4 \left[\frac{1}{3} \frac{\bar{x}^2}{m_2} - \frac{3}{5} \bar{x} \frac{m_2}{m_3}\right] \left[\frac{18}{5} \frac{\bar{x}m_2}{m_3} + 2\right]}}{2 \left[\frac{1}{3} \frac{\bar{x}^2}{m_2} - \frac{3}{5} \bar{x} \frac{m_2}{m_3}\right]},$$

$$\hat{\alpha} = \frac{-\left(\hat{\theta} - 1\right) + \sqrt{\left(\hat{\theta} - 1\right)^2 + 4 \left[\frac{1}{3} \frac{\bar{x}^2}{m_2} \hat{\theta} \left(\hat{\theta} + 1\right) + \hat{\theta}\right]}}{2}.$$

Do estymacji parametrów rozkładu Daguma stosowane są różne metody, które zostały omówione w pracach Domańskiego i Jędrzejczak (1998) oraz Jędrzejczak i in. (2018, 2019). W celu określenia stopnia zgodności rozkładów empirycznych z teoretycznymi wykorzystuje się rozbieżności pomiędzy empirycznymi wartościami różnych charakterystyk statystycznych rozkładu średniej arytmetycznej oraz odpowiednich indeksów (np. Giniego, Pietry, Theila).

W niniejszym artykule zgodność dopasowania rozważanych rozkładów została oceniona również za pomocą wskaźnika podobieństwa struktur, indeksu Mortara  $A_1$ , indeksu Pearsona kwadratów różnic częstości względnych  $A_2$  oraz zmodyfikowanego indeksu kwadratów różnic częstości względnych  $A'_2$ . Indeksy te mają postać:

$$A_1 = \frac{1}{n} \sum_{j=1}^s |n_j - \hat{n}_j|, \quad (7)$$

$$A_2 = \sqrt{\frac{1}{n} \sum_{j=1}^s \frac{(n_j - \hat{n}_j)^2}{\hat{n}_j}}, \quad (8)$$

$$A'_2 = \sqrt{\frac{1}{n} \sum_{j=1}^s \frac{(n_j - \hat{n}_j)^2}{n_j}}, \quad (9)$$

gdzie  $s$  to liczba przedziałów, na które zostały pogrupowane uporządkowane dane, a  $n_j$  oraz  $\hat{n}_j$  to rzeczywiste oraz szacowane częstości poszczególnych przedziałów.



Dla rozkładu Zengi szacowane częstości wyznaczamy w następujący sposób:

$$\hat{n}_j = \begin{cases} nF(x_1; \hat{\mu}; \hat{\alpha}; \hat{\theta}), & j = 1 \\ n\{F(x_j; \hat{\mu}; \hat{\alpha}; \hat{\theta}) - F(x_{j-1}; \hat{\mu}; \hat{\alpha}; \hat{\theta})\}, & j = 2, 3, \dots, s, \end{cases}$$

gdzie  $\hat{\mu}, \hat{\alpha}, \hat{\theta}$  są estymowanymi wartościami parametrów funkcji Zengi  $f(x; \mu; \alpha; \theta)$ .

Analizowane dane zostały pogrupowane na 20 przedziałów.

#### 4. ZASTOSOWANIE ROZKŁADÓW DAGUMA I ZENGI DO ANALIZY ROZKŁADU DOCHODÓW GOSPODARSTW DOMOWYCH

Rozkłady Daguma i Zengi zostały wykorzystane do analizy dochodów gospodarstw domowych ludności Polski w 2014 r. Rysunek 3 obrazuje różne metody estymacji rozkładu Zengi do danych empirycznych. W tabl. 1 zestawiono miary zgodności dopasowania do danych empirycznych, takie jak wskaźnik podobieństwa struktur oraz indeksy określone wzorami (7), (8), (9) dla rozkładu Daguma i rozkładu Zengi. Tablica 2 zawiera empiryczne oraz teoretyczne charakterystyki rozkładu dochodów mieszkańców Polski.

Do estymacji parametrów rozkładu Zengi użyto metod D'Addaria opartych na dwóch oraz trzech ograniczeniach oraz metody momentów. Wykorzystanie metody największej wiarygodności dla rozkładu Zengi rozważano w pracach Arcagniego (2014), Arcagniego i Porra (2013) oraz Arcagniego i Zengi (2013), jednak otrzymane wyniki były nieznacznie gorsze od wyników otrzymanych prostszymi metodami minimalizującymi różnice między empirycznymi i teoretycznymi charakterystykami rozkładu.

Zebrane wyniki pokazują, że najlepsze dopasowanie do danych empirycznych dla metody D'Addaria z trzema ograniczeniami uzyskuje się, stosując metodę 2:

$$\begin{cases} \hat{\mu} = \bar{x} \\ F(\hat{m}; \bar{x}, \alpha, \theta) = 0,5 \\ P(\bar{x}, \alpha, \theta) = \hat{P} \end{cases}$$

Natomiast stosując metodę D'Addaria z dwoma ograniczeniami, najlepsze dopasowanie otrzymujemy przy użyciu następującej metody:

$$\begin{cases} \hat{\mu} = \bar{x} \\ F(\hat{m}; \bar{x}, \alpha, \theta) = 0,5. \end{cases}$$

Metoda momentów daje również zadowalające rezultaty, a wyznaczona analitycznie postać parametrów funkcji gęstości Zengi jest niewątpliwie zaletą stosowania tej metody. Warto zauważyć, że używając metody D'Addaria należy zastosować metody numeryczne.

W wyniku estymacji parametrów metodą największej wiarygodności w pracy Jędrzejczak i Pekasiewicz (2018) wyznaczono parametry rozkładu Daguma:  $p = 0,747$ ,  $a = 3,12$ ,  $b = 3611,02$  dla budżetów gospodarstw domowych ludności Polski w 2014 r.

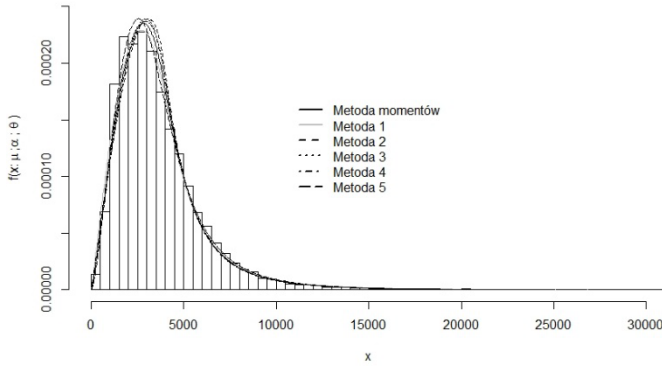
**TABLICA 1. WYNIKI ESTYMACJI ROZKŁADÓW DAGUMA I ZENGI**

Metoda	Parametr funkcji gęstości			Indeks dopasowania			Współczynnik podobieństwa struktur
	$\hat{\mu}$	$\hat{\alpha}$	$\hat{\theta}$	$A_1$	$A_2$	$A_2'$	$w_p$
Model Zengi							
Metoda momentów	3755,33	2,2427	3,0303	0,0952	0,1271	0,1375	0,9524
Metody z trzema ograniczeniami: $F(\hat{m}: \bar{x}; \alpha; \theta) = 0,5$ $F(\bar{x}: \bar{x}; \alpha; \theta) = \hat{F}(\bar{x})$	3755,33	2,2502	2,9101	0,1112	0,1415	0,1435	0,9444
$F(\hat{m}: \bar{x}; \alpha; \theta) = P$ $P(\bar{x}; \alpha; \theta) = \hat{P}(\bar{x})$	3755,33	2,5693	3,6254	0,0666	0,0905	0,0929	0,9667
$F(\hat{m}: \bar{x}; \alpha; \theta) = 0,5$ $Var(\bar{x}; \alpha; \theta) = m_2$	3755,33	2,1202	3,0247	0,0857	0,1338	0,1684	0,9572
$F(\bar{x}: \bar{x}; \alpha; \theta) = \hat{F}(\bar{x})$ $Var(\bar{x}; \alpha; \theta) = m_2$	3755,33	2,1132	2,6991	0,1235	0,1596	0,1684	0,9383
$F(\hat{m}: \bar{x}; \alpha; \theta) = 0,5$ $A(\bar{x}; \alpha; \theta) = \hat{A}(\bar{x})$	3755,33	2,1723	2,8501	0,1627	0,2172	0,2779	0,9338
Metody z dwoma ograniczeniami: $F(\hat{m}: \bar{x}; \alpha; \theta) = 0,5$	3755,33	2,7943	3,9244	0,0599	0,0815	0,0052	0,9700
$F(\bar{x}: \bar{x}; \alpha; \theta) = \hat{F}(\bar{x})$	3755,33	3,6195	5,0377	0,0733	0,1205	0,1062	0,9633
$P(\bar{x}; \alpha; \theta) = \hat{P}(\bar{x})$	3755,33	2,1469	2,9309	0,0973	0,6023	0,1553	0,9513
$P(\bar{x}; \alpha; \theta) = \hat{P}(\bar{x})$	3755,33	3,1602	1,6269	0,6700	1,0572	0,7787	0,6649
Model Daguma							
Metoda największej wiarygodności	$p$	$a$	$b$	$A_1$	$A_2$	$A_2'$	$w_p$
	0,747	3,12	3611,02	0,0745	0,0965	0,0974	0,9627

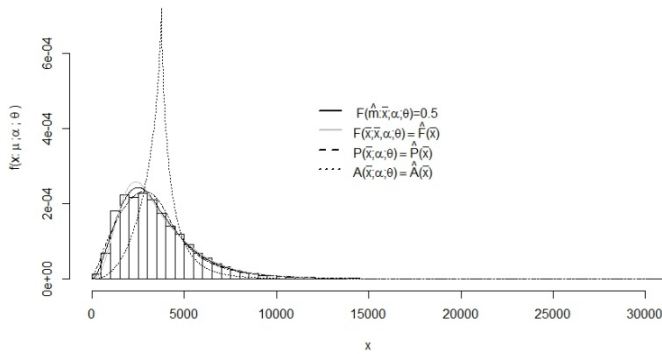
Źródło: opracowanie własne.

Rysunek 3. Aproksymacja modelu Zengi do empirycznego rozkładu dochodów ludności Polski (36929 obserwacji)

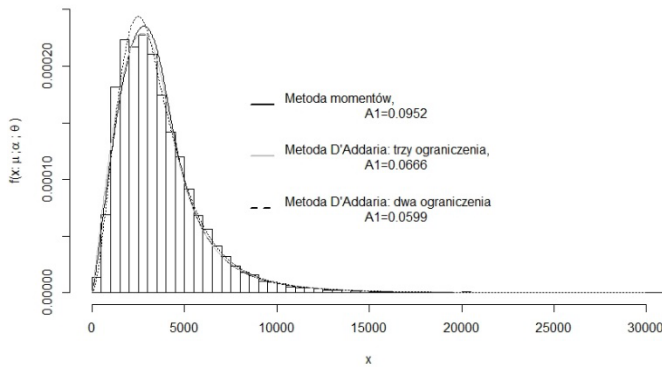
Metoda momentów oraz metoda D'Addaria z trzema ograniczeniami



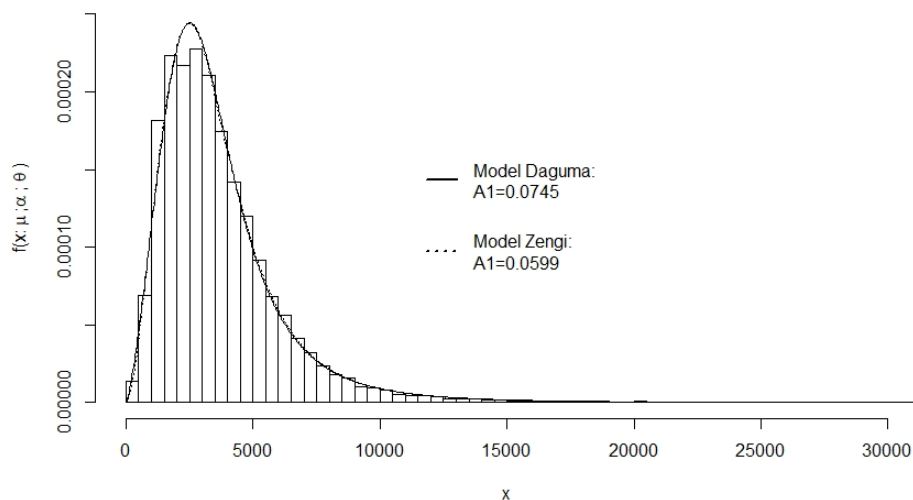
Metoda D'Addaria z dwoma ograniczeniami



Zestawienie najlepszych metod estymacji



Rysunek 4. Aproksymacja rozkładu Daguma oraz rozkładu Zengi do empirycznego rozkładu dochodów ludności Polski (36929 obserwacji)



Źródło: opracowanie własne.

Z przedstawionych rezultatów aproksymacji wynika, że możliwe jest określenie parametrów rozkładu Zengi dla dochodów indywidualnych. O zgodności tego rozkładu z obserwacjami empirycznymi świadczą wysokie wartości współczynników podobieństwa struktur oraz niskie wartości indeksów dopasowania. Podobne rezultaty otrzymał Zenga (2010). Porównując indeksy dopasowania rozkładu dochodów ludności Polski (Jędrzejczak, Trzcińska, 2018) z indeksami dla Włoch, Stanów Zjednoczonych oraz Szwajcarii wyznaczonymi przez Zengę, można zaobserwować, że dla Polski są nieco wyższe. Przyczyn tych rozbieżności należy upatrywać przede wszystkim w różnicach definicji zmiennej dochodowej, która w przytoczonych badaniach jest dochodem ekwiwalentnym, natomiast w badaniu polskim zastosowano dochód rozporządzalny gospodarstwa domowego. Dodatkowy wpływ na wynik mają odmienne schematy losowania próby oraz inne liczebności. Z przeprowadzonej analizy można wywnioskować, że zmiany parametrów  $\alpha$  oraz  $\theta$  rozkładu Zengi objaśniają zmiany wartości rozważonych miar statystycznych, co ilustruje tabl. 2. Parametry  $\alpha$  oraz  $\theta$  są parametrami nierównomierności rozkładu Zengi, określają one postać krzywej koncentracji Lorenza, a tym samym współczynnik Giniego. Trzeci parametr rozkładu Zengi  $\mu$  jest parametrem skali, więc nie wpływa na wartość współczynnika Giniego (Zenga i in., 2010b).

**TABLICA 2. EMPIRYCZNE I TEORETYCZNE CHARAKTERYSTYKI  
ROZKŁADU DOCHODÓW W POLSCE**

Metoda	Współczynnik Giniego	Średnia arytmetyczna	Odchylenie standardowe	Mediana	Kwantyl rzędu 0,1	Kwantyl rzędu 0,95
Model Zengi						
Metoda momentów	0,2704	3755,33	2959,97	3202,09	1319,16	8299,97
$F(\hat{m}: \bar{x}; \alpha; \theta) = 0,5$ $F(\bar{x}: \bar{x}; \alpha; \theta) = \hat{F}(\bar{x})$	0,3399	3755,33	2879,44	3237,46	1359,05	8168,20
$F(\hat{m}: \bar{x}; \alpha; \theta) = P$ $P(\bar{x}; \alpha; \theta) = \hat{P}(\bar{x})$	0,3484	3755,33	2847,59	3158,97	1341,34	8379,67
$F(\hat{m}: \bar{x}; \alpha; \theta) = 0,5$ $Var(\bar{x}; \alpha; \theta) = m_2$	0,3611	3755,33	3151,08	3158,99	1245,80	8487,26
$F(\bar{x}: \bar{x}; \alpha; \theta) = \hat{F}(\bar{x})$ $Var(\bar{x}; \alpha; \theta) = m_2$	0,3414	3755,33	2959,93	3251,09	1339,54	8159,10
$F(\hat{m}: \bar{x}; \alpha; \theta) = 0,5$ $A(\bar{x}; \alpha; \theta) = \hat{A}(\bar{x})$	0,3414	3755,33	2959,91	3227,95	1329,58	8224,76
$F(\hat{m}: \bar{x}; \alpha; \theta) = 0,5$	0,3458	3755,33	2745,12	3158,99	1378,89	8332,40
$F(\bar{x}: \bar{x}; \alpha; \theta) = \hat{F}(\bar{x})$	0,3308	3755,33	2510,97	3155,49	1475,78	8207,88
$P(\bar{x}; \alpha; \theta) = \hat{P}(\bar{x})$	0,3558	3755,33	3049,53	3195,90	1289,44	8348,12
$A(\bar{x}; \alpha; \theta) = \hat{A}(\bar{x})$	0,1300	3755,33	1393,82	3663,77	2356,82	5938,10
Model Daguma						
Metoda największej wiarygodności	0,3445	3775,76	3016,78	3151,42	1364,88	8426,81
Wartości empiryczne	0,3410	3755,33	2959,95	3158,99	1352,47	8150,00

Źródło: opracowanie własne.

Tablica 2 przedstawia oszacowane wartości charakterystyk liczbowych rozkładów teoretycznych oraz ich empiryczne wartości. Jak widać, za pomocą rozkładu Zengi otrzymujemy najlepsze dopasowanie do danych empirycznych (rys. 4). W ramach przeprowadzonych badań rozkład Daguma przeszacowuje wybrane miary, wyjątek stanowi mediana, gdyż w tym przypadku rozkład Daguma nie doszacowuje tej wartości. Rozkład Zengi idealnie szacuje średnią arytmetyczną, wynika to z konstrukcji metod estymacji. Idealną zgodność z danymi empirycznymi dla mediany otrzymujemy dzięki metodzie D'Addaria z dwoma ograniczeniami  $\left\{ \begin{array}{l} \mu = \bar{x} \\ F(\hat{m}: \bar{x}; \alpha; \theta) = 0,5 \end{array} \right.$  oraz trzema ograniczeniami  $\left\{ \begin{array}{l} \mu = \bar{x} \\ F(\hat{m}: \bar{x}; \alpha; \theta) = 0,5 \\ Var(\bar{x}; \alpha; \theta) = m_2 \end{array} \right.$ . Najlepsze oszacowanie odchylenia standardowego otrzymujemy przy zastosowaniu

metody momentów oraz metody D'Addaria z trzema ograniczeniami:

$$\begin{cases} \mu = \bar{x} \\ F(\bar{x}; \bar{x}; \alpha; \theta) = \hat{F}(\bar{x}); \\ Var(\bar{x}; \alpha; \theta) = m_2 \end{cases}; \begin{cases} \mu = \bar{x} \\ F(\hat{m}; \bar{x}; \alpha; \theta) = 0,5. \\ A(\bar{x}; \alpha; \theta) = \hat{A}(\bar{x}) \end{cases}$$

empirycznych kwantyla rzędu 0,1 otrzymujemy za pomocą metody D'Addaria

$$\text{z trzema ograniczeniami } \begin{cases} \mu = \bar{x} \\ F(\hat{m}; \bar{x}; \alpha; \theta) = 0,5, \\ F(\bar{x}; \bar{x}; \alpha; \theta) = \hat{F}(\bar{x}) \end{cases}$$

nie kwantyla rzędu 0,95 otrzymujemy, stosując metodę D'Addaria z trzema

$$\text{ograniczeniami } \begin{cases} \mu = \bar{x} \\ F(\bar{x}; \bar{x}; \alpha; \theta) = \hat{F}(\bar{x}). \\ Var(\bar{x}; \alpha; \theta) = m_2 \end{cases}$$

ka Giniego, to na najlepsze dopasowanie do danych empirycznych pozwala

$$\text{metoda D'Addaria z trzema ograniczeniami: } \begin{cases} \mu = \bar{x} \\ F(\bar{x}; \bar{x}; \alpha; \theta) = \hat{F}(\bar{x}) \\ Var(\bar{x}; \alpha; \theta) = m_2 \end{cases} \text{ oraz}$$

$$\begin{cases} \mu = \bar{x} \\ F(\hat{m}; \bar{x}; \alpha; \theta) = 0,5. \\ A(\bar{x}; \alpha; \theta) = \hat{A}(\bar{x}) \end{cases}$$

metody D'Addaria z dwoma ograniczeniami  $\begin{cases} \mu = \bar{x} \\ A(\bar{x}; \alpha; \theta) = \hat{A}(\bar{x}) \end{cases}$ , co nie jest zaskoczeniem, gdyż ta metoda estymacji ma najniższą wartość współczynnika podobieństwa struktur oraz najwyższe wartości indeksów dopasowania.

Na podstawie przeprowadzonej analizy można stwierdzić, że metoda zapewniająca najlepszą zgodność dopasowania do danych empirycznych niekoniecznie pokrywa się z metodą najlepszą do szacowania konkretnej charakterystyki. Spostrzeżenie to jest zgodne z wynikami pracy Ostasiewicz (2013). Zauważmy, że stosując rozkład Zengi, w każdym przypadku otrzymujemy lepsze oszacowanie do danych empirycznych wybranych miar niż za pomocą rozkładu Daguma. Potwierdza to zasadność stosowania rozkładu Zengi do opisu dochodów Polaków. Słabą stroną rozkładu Zengi może być jedynie to, że klasyczne metody estymacji parametrów są dość skomplikowane. Chodzi tu m.in. o zastosowanie metody największej wiarygodności, która wiąże się z koniecznością bardzo czasochłonnym obliczeń z użyciem metod numerycznych.

## 5. WNIOSKI

W artykule przedstawiono aproksymację rozkładu płac i dochodów ludności Polski za pomocą rozkładów Daguma i Zengi. Porównując wykresy funkcji gęstości obu rozkładów, można zauważyć, że funkcja gęstości rozkładu Zengi przyjmuje bardziej różnorodne kształty niż funkcja gęstości rozkładu Daguma. Dzięki temu rozkład Zengi łatwiej dopasowuje się do różnego rodzaju danych

empirycznych. Na podstawie rozważań teoretycznych i przedstawionych wyników estymacji zauważamy, że w odróżnieniu od innych rozkładów trójparametrycznych, np. rozkładu Daguma, parametry rozkładu Zengi można interpretować w kategoriach ekonomicznych. Parametr  $\mu$  jest bezpośrednio interpretowalny jako średni dochód indywidualny lub średni dochód gospodarstwa domowego, natomiast parametry  $\alpha$  i  $\theta$  mogą być interpretowane w kategoriach nierówności dochodowych. W szczególności wartość parametru  $\alpha$  jest odwrotnym wskaźnikiem nierówności, który kontroluje kształt rozkładu, podczas gdy  $\theta$  jest bezpośrednim wskaźnikiem nierówności, który kontroluje rozkład wokół wartości oczekiwanej  $\mu$  (Arcagni, Porro, 2013).

Do estymacji parametrów rozkładu Daguma zastosowano metodę największej wiarygodności, a dla rozkładu Zengi – metodę momentów i metodę D’Addaria z dwoma i trzema ograniczeniami. Z analizy metod estymacji przeprowadzonych dla parametrów funkcji gęstości Zengi wynika, że najlepsze dopasowanie uzyskuje się za pomocą metody D’Addaria z dwoma ograniczeniami:  $\hat{\mu} = \bar{x}$  oraz  $F(\hat{m}; \bar{x}, \alpha, \theta) = 0,5$ . Metoda momentów przynosi również zadowalające efekty. Na podstawie przeprowadzonej analizy można stwierdzić, że metoda zapewniająca najlepszą zgodność dopasowania do danych empirycznych niekoniecznie pokrywa się z metodą najlepszą do szacowania konkretnej charakterystyki. Spostrzeżenie to jest zgodne z wynikami pracy Ostasiewicz (2013).

Z przedstawionych rezultatów wynika, że możliwe jest efektywne dopasowanie rozkładu dochodów z użyciem modelu Zengi. Ten wniosek jest zgodny z wynikami podobnych badań prowadzonych nad rozkładami dochodów przez Zengę (2010) oraz Zengę i in. (2010a, 2010b, 2010c).

Porównanie rozkładów Daguma i Zengi pokazuje, że rozkład Zengi lepiej opisuje rozkład dochodów ludności Polski. Rozkład Zengi spełnia kryteria stawiane teoretycznym rozkładom dochodów. Przejrzysta interpretacja ekonomiczna parametrów rozkładu Zengi jest dodatkowym argumentem skłaniającym do dalszych badań dotyczących przydatności tego rozkładu do opisu rozkładu płac i dochodów ludności Polski.

## LITERATURA

- Arcagni A., (2014), Zenga Distribution: Parameters Estimation with Constraints on Synthetic Inequality Measures, *Statistica & Applicazioni*, 12(1), 17–39.
- Arcagni A., Porro F., (2013), On the Parameters of Zenga Distribution, *Statistical Methods and Applications*, 22(3), 285–303. DOI: 10.1007/s10260-012-0219-y.
- Arcagni A., Zenga M., (2013), Application of Zenga’s Distribution to a Panel Survey on Household Income of European Member States, *Statistica & Applicazioni*, 2(1), 79–102. DOI: 10.1400/209741.
- Bandurian R., McDonald J. B., Turley R. S., (2003), A comparison of parametric models of income distribution across countries and over time, *Estadística*, 55, 135–152.

- Brzeziński M. (2013), Parametric modelling of income distribution in Central and Eastern Europe, *Central European Journal of Economic Modelling and Econometrics*, 35, 207–230.
- Dagum C., (1977), A New Model of Personal Income Distribution: Specification and Estimation, *Economie Appliquée*, 30, 413–437.
- Dagum C., (1990), Generation and Properties of Income Distribution Functions, *Proceedings of the Second International Conference on Income Distribution by Size: Generation, Distribution, Measurement and Applications*.
- Dastrup S. R., Hartshorn R., McDonald J. B. (2007), The Impact of Taxes and Transfer Payments on the Distribution of Income: A Parametric Comparison, *Journal of Economic Inequality*, 5, 353–369. DOI: 10.1007/s10888-006-9039-3.
- D'Addario R., (1934), *Sulla Misura Della Concentrazione dei Redditi*, Poligrafico dello stato, Roma.
- D'Addario R., (1939), Un Metodo Per la Rappresentazione Analitica Delle Distribuzioni Statistiche, *Annali dell' Istituto di Statistica dell' Università di Bari*, 16, 3–56.
- Domański C., (1973), Struktura płac pracowników w województwach grodzkich według działów gospodarki narodowej, *Studia Prawno-Ekonomiczne*, 11, 149–163.
- Domański C., Jędrzejczak A., (1998), Maximum Likelihood Estimation of the Dagum Model Parameters, *International Advances in Economic Research*, 4, 243–252.
- Domański C., Jędrzejczak A., (2005), *Analiza rozkładów płac i dochodów w Polsce w przekroju terytorialnym*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź.
- Edgeworth F. Y., (1898), On the Representation of Statistics by Mathematical Formulae, *Journal of the Royal Statistical Society*, 61, 670–700.
- Jagielski M., Kutner R., (2010), Study of Households' Income in Poland by Using the Statistical Physics Approach, *Acta Physica Polonica A*, 117(4), 615–618. DOI: 10.12693/APhysPolA.117.615.
- Jędrzejczak A., (1990), Uwagi o zastosowaniu rozkładu Daguma do badania rozkładów płac, *Wiadomości Statystyczne*, 7, 23–25.
- Jędrzejczak A. (1999, maszynopis), *Statystyczna analiza nierównomierności rozkładów płac i dochodów* [praca doktorska, Uniwersytet Łódzki].
- Jędrzejczak A., Pekasiewicz D., (2018), Analysis of the Properties of Selected Inequality Measures Based on Quantiles with the Application to the Polish Income Data, w: M. Papież, S. Śmiech (eds.), *The 11th Professor Aleksander Zeliaś International Conference on Modelling and Forecasting of Socio-Economic Phenomena, Conference Proceedings*, 113–122.
- Jędrzejczak A., Pekasiewicz D., Zieliński W., (2018), Comparison of Estimation Methods for the Dagum Distribution Parameters, w: M. Papież, S. Śmiech (eds.), *The 12th Professor Aleksander Zeliaś International Conference on Modelling and Forecasting of Socio-Economic Phenomena, Conference Proceedings*, 180–189.
- Jędrzejczak A., Pekasiewicz D., Zieliński W., (2019), Confidence Interval for Quantile Ratio of the Dagum Distribution, *Revstat – Statistical Journal*.
- Jędrzejczak A., Trzcińska K., (2018), Application of the Zenga distribution to the analysis of household income in Poland by socio-economic group, *Statistica & Applicazioni*, 16(2), 123–140.
- Kleiber C., Kotz S., (2003), *Statistical Size Distributions in Economics and Actuarial Sciences*, Wiley, Hoboken.
- Kordos J., (1968), *Metody matematyczne badania i analizy rozkładu dochodów ludności*, Główny Urząd Statystyczny, Warszawa.
- Kordos J., (1973), *Metody analizy i prognozowania rozkładów płac i dochodów ludności*, Główny Urząd Statystyczny, Warszawa.



- Kordos J., (1990), Research on Income Distribution by Size in Poland, w: C. Dagum, M. Zenga (eds.), *Income and Wealth Distribution, Inequality and Poverty*, Springer, New York, Berlin, London and Tokyo, 335–351.
- Kot S. M., (1999), *Analiza ekonometryczna kształtowania się płac w Polsce w okresie transformacji*, Wydawnictwo Naukowe PWN, Warszawa–Kraków.
- Kot S. M., (2000), *Ekonometryczne modele dobrobytu*, Wydawnictwo Naukowe PWN, Warszawa–Kraków.
- Lange O., (1967), *Wstęp do ekonometrii*, Państwowe Wydawnictwo Naukowe, Warszawa.
- Łukasiewicz P., Orłowski A., (2004), Probabilistic Models of Income Distributions, *Physica A*, 344, 146–151. DOI: 10.1016/j.physa.2004.06.106.
- March L., (1898), Quelques Exemples de Distribution de Salaries, *Journal de la Société Statistique de Paris*, 39, 193–206, 241–248.
- Ostasiewicz K. (2013), Adekwatność wybranych rozkładów teoretycznych dochodów w zależności od metody aproksymacji, *Przegląd Statystyczny*, 60(4), 499–521.
- Pareto V., (1897), *Cours d'economie Politique*, Rouge, Lausanne–Paris.
- Polisicchio M., (2008), The Continuous Random Variable with Uniform Point Inequality Measure  $I(p)$ , *Statistica & Applicazioni*, 6(2), 137–151.
- Salamaga M., (2016), Badanie wpływu metody estymacji teoretycznych modeli rozkładu dochodów na jakość aproksymacji rozkładu dochodów mieszkańców Krakowa, *Zeszyty Naukowe. Uniwersytet Ekonomiczny w Krakowie*, 3(951), 63–79. DOI: 10.15678/ZNUEK.2016.0951.0305.
- Vielrose E., (1960), *Rozkład dochodów według wielkości*, Polskie Wydawnictwo Gospodarcze, Warszawa.
- Wąsik B., (1967), Dwu- i trójparametryczny rozkład logarytmiczno-normalny jako aproksymacja rozkładów zarobków pracowników gospodarki społecznej 1955–1965, *Przegląd Statystyczny*, 14(4), 409–424.
- Wiśniewski J., (1934), *Rozkład dochodów według wysokości*, Instytut Badań Koniunktur Gospodarczych i Cen, Warszawa.
- Zenga M. M., (2007), Inequality Curve and Inequality Index Based on the Ratios Between Lower and Upper Arithmetic Means, *Statistica & Applicazioni*, 5(1), 3–28.
- Zenga M. M., (2010), Mixture of Polisicchios Truncated Pareto Distributions with Beta Weights, *Statistica & Applicazioni*, 8(1), 3–25.
- Zenga M. M., Pasquazzi L., Zenga Ma. (2010a), First Applications of a New Three Parameter Distribution for Non-Negative Variables, *Rapporto di Ricerca* N. 188, Dipartimento di Metodi Quantitativi per le Scienze Economiche ed Aziendali – Università degli Studi di Milano-Bicocca.
- Zenga M. M., Polisicchio M., Zenga Ma., Pasquazzi L., (2010b), More on a New Three-Parameter Distribution for Non-Negative Variables, *Rapporto di Ricerca* N. 187, Dipartimento di Metodi Quantitativi per le Scienze Economiche ed Aziendali – Università degli Studi di Milano-Bicocca.
- Zenga M. M., Porro F., Arcagni A., (2010c), Method of Moments for Zenga Distribution, *Rapporto di Ricerca* N. 193, Dipartimento di Metodi Quantitativi per le Scienze Economiche ed Aziendali – Università degli Studi di Milano-Bicocca.