

Zastosowanie rozkładu Pareta do aproksymacji rozkładów dochodów zamożnych gospodarstw domowych

Dorota Pekasiewicz^a

Streszczenie. Celem badania omawianego w artykule jest aproksymacja rozkładów ekwiwalentnych dochodów zamożnych gospodarstw domowych w grupach społeczno-ekonomicznych za pomocą rozkładu Pareta z parametrami oszacowanymi metodą największej wiarygodności. Za zamożne gospodarstwa domowe uznano te, których dochody przekroczyły ustalony próg bogactwa (zamożności). Rozkłady dochodów zamożnych gospodarstw domowych są zazwyczaj zeromodalne i cechują się grubymi ogonami, dlatego do ich aproksymacji wykorzystano rozkład Pareta. Analizowano ekwiwalentne dochody zamożnych gospodarstw w Polsce ogółem i w poszczególnych grupach społeczno-ekonomicznych w latach 2014–2017. Dane uzyskano z badania budżetów gospodarstw domowych. Stopień dopasowania rozkładów teoretycznych do empirycznych zbadano za pomocą wybranych miar podobieństwa.

Wyniki potwierdziły dużą zgodność empirycznych rozkładów dochodów z modelem Pareta. Bardzo dobre aproksymacje uzyskano w szczególności dla zamożnych gospodarstw domowych pracowników, pracujących na własny rachunek oraz emerytów i rencistów. Nieznacznie gorsze wyniki otrzymano dla gospodarstw rolników. Rozkłady teoretyczne dobrze dopasowane do danych empirycznych zostały wykorzystane do szacowania wybranych charakterystyk rozkładu, m.in. miar położenia, zróżnicowania i nierównomierności, oraz do porównywania grup pod względem ich zamożności.

Słowa kluczowe: dochód, gospodarstwo domowe, rozkład Pareta, zamożność

JEL: C10, J30

The application of the Pareto distribution to approximate income distributions of wealthy households in Poland

Abstract. The aim of the paper is to approximate the equivalent income distributions of wealthy households in particular socio-economic groups using the Pareto distribution, with parameters estimated by means of the maximum likelihood estimation method. Households whose income exceeded the established wealth threshold were classified as wealthy households. Income distributions of wealthy households are usually non-modal and heavy-tailed, thus, the Pareto distribution was applied as their theoretical model. The equivalent income of wealthy households in Poland was analysed in total and in particular socio-economic groups. The research was based on data from the 2014–2017 Household Budget Survey. Selected similarity measures were used to examine the degree to which the theoretical distributions proved consistent with the empirical ones.

The obtained results confirmed the high level of consistency of empirical income distributions with the Pareto model. Moreover, very good approximations were obtained especially for

^a Uniwersytet Łódzki, Wydział Ekonomiczno-Socjologiczny, Polska / University of Lodz, Faculty of Economics and Sociology, Poland. ORCID: <https://orcid.org/0000-0001-8275-3345>.
E-mail: dorota.pekasiewicz@uni.lodz.pl.

wealthy households of employees and self-employed, as well as pensioners. Slightly worse results were obtained for the farmers group. Theoretical distributions well fitted to empirical data were used to estimate selected distribution characteristics, including measures of location, dispersion and inequality, and to compare the different groups in terms of their wealth.

Keywords: income, household, Pareto distribution, wealth

1. Wprowadzenie

Rozkłady dochodów, w tym ich nierównomierność, są przedmiotem badań zarówno statystyków, jak i ekonomistów. Badania te obejmują takie nurty, jak poszukiwanie modelu teoretycznego opisującego rozkłady empiryczne, tworzenie teorii wyjaśniających proces kształtowania się rozkładów dochodów i ich nierównomierności oraz doskonalenie metod badania nierównomierności rozkładów.

Dotychczasowe badania zaowocowały dużą liczbą rozkładów teoretycznych, które znajdują zastosowanie w analizie dochodów. W pracach Jędrzejczak (2006), Jędrzejczak i Pekasiewicz (2020), Kleibera (1996), Kleibera i Kotza (2003), Ostasiewicz (2013) i Salamagi (2016) można znaleźć wyniki analiz wykorzystujących wiele rozkładów, w tym Daguma, Singha-Maddali czy Fiska, do aproksymacji rozkładów dochodów, ze wskazaniem tych, które najlepiej odwzorowują rozkłady empiryczne dla różnych grup gospodarstw domowych. Rozkłady dochodów gospodarstw najczęściej są jednomodalne i charakteryzują się prawostronną asymetrią oraz dodatnią kurtozą (Kot, 1999).

Inny typ rozkładu cechuje gospodarstwa domowe o wysokich dochodach. Są one zazwyczaj zeromodalne i mają nieskończone wariancje. Już wyniki badania przedstawione w pracy Mandelbrota (1963) pokazały, że w przypadku grup o wysokich dochodach dobre dopasowanie do danych empirycznych wykazuje rozkład Pareta, dlatego jest on wykorzystywany do aproksymacji miesięcznych dochodów ekwiwalentnych gospodarstw domowych w grupach społeczno-ekonomicznych. Problematykę zamożności i bogactwa oraz modelowania rozkładów indywidualnych i ekwiwalentnych dochodów zamożnych gospodarstw domowych za pomocą rozkładu Pareta rozważają m.in. Brzeziński (2014), Brzeziński i in. (2020), Charpentier i Flachaire (2019), Jagielski i Kutner (2011), Peichl i in. (2008), Sączewska-Piotrowska (2018) oraz Włodarczyk (2013).

W analizach dotyczących dochodów przypadających na osobę i porównywania dochodów w gospodarstwach domowych istotnym zagadnieniem jest uwzględnienie struktury demograficznej gospodarstwa. Koszty utrzymania nie są proporcjonalne do liczby członków gospodarstwa, dlatego wielkość dochodów ekwiwalentnych wyznacza się z zastosowaniem wybranej skali ekwiwalentności. W analizach przeprowadzonych w ramach badania omawianego w niniejszym artykule wykorzystano skalę potęgową z elastycznością ekwiwalentności równą 0,5.

Zagadnienia poruszane w artykule dotyczą aproksymacji empirycznych rozkładów dochodów ekwiwalentnych zamożnych gospodarstw domowych za pomocą rozkładu Pareta. Wiadomo, że o zamożności gospodarstw decydują nie tylko dochody. Duże znaczenie ma również wielkość dochodów niewydatkowanych na konsumpcję i kumulowanych w postaci oszczędności rzeczowych oraz finansowych, generujących posiadanie aktywów. Niemniej wielkość dochodu jest jednym z najważniejszych czynników wpływających na sytuację finansową gospodarstw. Za gospodarstwo domowe o wysokich dochodach uznaje się to, którego dochody przekraczają ustalony próg (granice) bogactwa, zwany też granicą zamożności.

Wartość granicy zamożności jest ustalana różnie, m.in. na poziomie dwu- lub trzykrotności mediany rozkładu dochodów ekwiwalentnych albo wybranego kwantyla rozkładu dochodów (zob. Brzeziński, 2014; Peichl i in., 2008). W literaturze czasem odróżnia się zamożnych od bogatych i bardzo bogatych. Według Franziniego i in. (2016) gospodarstwa domowe osiągające dochody powyżej trzykrotności mediany traktuje się jako zamożne, osiągające dochody pięciokrotnie przewyższające medianę – jako bogate, a osiągające dochody przynajmniej 10 razy wyższe od mediany – jako bardzo bogate. W omawianym badaniu za zamożne uznano gospodarstwa o miesięcznych dochodach ekwiwalentnych przekraczających wartość dwukrotności mediany i aproksymowano rozkłady ich dochodów za pomocą rozkładu Pareta o parametrach szacowanych metodą największej wiarygodności.

Celem badania było dopasowanie rozkładu teoretycznego do rozkładów ekwiwalentnych dochodów zamożnych gospodarstw domowych w grupach społeczno-ekonomicznych. Rozkłady Pareta aproksymujące empiryczne rozkłady dochodów zamożnych gospodarstw wyznaczono dla grup społeczno-ekonomicznych w latach 2014–2017. Rozkłady teoretyczne umożliwiają dokładniejszą analizę oraz opracowanie prognoz dotyczących rozkładów dochodów. W omawianym badaniu zostały wykorzystane do oszacowania miar położenia, zróżnicowania oraz nierównomierności ekwiwalentnych dochodów zamożnych gospodarstw domowych w grupach społeczno-ekonomicznych w 2017 r.

2. Rozkład Pareta i jego wybrane charakterystyki

Wśród rozkładów, które opracowano w wyniku obserwacji rozkładów dochodów ludności uzyskanych ze statystyk podatkowych różnych krajów europejskich, znajduje się rozkład Pareta (zob. Pareto, 1897).

Funkcja gęstości rozkładu Pareta jest określona wzorem

$$f(y) = \begin{cases} \frac{a\theta^a}{y^{a+1}} & \text{dla } y \geq \theta, \\ 0 & \text{dla } y < \theta, \end{cases} \quad (1)$$

a dystrybuanta – wzorem

$$F(y) = \begin{cases} 1 - \left(\frac{y}{\theta}\right)^{-a} & \text{dla } y \geq \theta, \\ 0 & \text{dla } y < \theta, \end{cases} \quad (2)$$

gdzie:

θ – minimalny dochód,

a – parametr zwany stałą Pareta.

Momenty zwykłe rzędu r istnieją, gdy parametr kształtu a spełnia nierówność $a > r$. Wówczas przyjmują one postać (zob. np. Krishnamoorthy, 2006)

$$E(Y^r) = \theta^r \frac{a}{a - r} \quad \text{dla } a > r. \quad (3)$$

Wartość oczekiwana wyraża się wzorem

$$E(Y) = \theta \frac{a}{a - 1} \quad \text{dla } a > 1. \quad (4)$$

Rozkład dochodów określony krzywą Pareta ma skończoną wariancję tylko dla $a > 2$ i przyjmuje postać

$$D(Y) = \frac{\theta^2}{(a - 1)^2} \cdot \frac{a}{a - 2} \quad \text{dla } a > 2. \quad (5)$$

Rozkład Pareta może charakteryzować się niewielką liczbą skończonych momentów, co oznacza, że może być rozkładem z ciężkim ogonem (*heavy tail*). Im mniejsza liczba skończonych momentów rozkładu, tym grubszy jest jego prawy koniec.

Dobre dopasowanie rozkładu Pareta do rozkładu dochodów może być wykorzystane do oszacowania funkcji Lorenza oraz miar określających stopień nierównomierności rozkładu, takich jak współczynnik Giniego czy współczynnik Schutza-Pietry.

Funkcja Lorenza dla rozkładu Pareta wyraża się wzorem

$$L(p) = 1 - (1 - p)^{1 - \frac{1}{a}} \quad \text{dla } a > 1, \quad (6)$$

gdzie $p \in [0, 1]$.

Współczynnik Giniego i współczynnik Schutza-Pietry przyjmują postać odpowiednio:

$$G = \frac{1}{2a - 1} \quad \text{dla } a > 1, \quad (7)$$

$$SP = \frac{1}{a - 1} \left(\frac{a - 1}{a} \right)^a \quad \text{dla } a > 1. \quad (8)$$

Oszacowanie \hat{a} parametru a pozwala zatem wyznaczyć zarówno funkcję Lorenza $\hat{L}(p) = 1 - (1 - p)^{1 - \frac{1}{\hat{a}}}$ dla $p \in [0, 1]$, jak i współczynniki nierówności: $\hat{G} = \frac{1}{2\hat{a} - 1}$ oraz $\hat{SP} = \frac{1}{\hat{a} - 1} \left(\frac{\hat{a} - 1}{\hat{a}} \right)^{\hat{a}}$ dla $a > 1$.

Im większa jest wartość parametru a , tym wartości współczynników Giniego i Pietry są mniejsze, a tym samym mniejsze są nierównomierności rozkładu.

3. Metoda badania

Podstawowym źródłem informacji o dochodach, wydatkach, spożyciu oraz innych aspektach warunków życia określonych grup ludności są badania budżetów gospodarstw domowych (BBGD), prowadzone przez Główny Urząd Statystyczny (Główny Urząd Statystyczny [GUS], 2018b). Ich wyniki pozwalają na analizę wpływu różnych czynników na kształtowanie się poziomu i zróżnicowania dochodów gospodarstw pogrupowanych według określonych kryteriów, m.in. źródła utrzymania, klasy miejscowości zamieszkania albo typu biologicznego gospodarstwa, województwa czy regionu. Ze względu na źródła utrzymania wyróżnia się następujące grupy społeczno-ekonomiczne gospodarstw:

- pracowników;
- rolników;
- pracujących na własny rachunek;
- emerytów i rencistów;
- utrzymujących się ze źródeł niezarobkowych.

W omawianym badaniu wymienione grupy gospodarstw analizowano pod kątem ich zamożności.

Badanie budżetów gospodarstw domowych prowadzone jest metodą reprezentacyjną, która daje możliwość uogólnienia (z określonym błędem) uzyskanych wyników na wszystkie gospodarstwa domowe w Polsce lub na wybrane grupy gospodarstw. Jednostką obserwacji jest gospodarstwo domowe jedno- lub wieloosobowe. Za gospodarstwo zamożne uznaje się gospodarstwo o dochodach przekraczających próg bogactwa (zamożności). Na potrzeby pracy badawczej opisanej w artykule próg ten ustalono na poziomie dwukrotności mediany.

Do analiz wykorzystano dane za lata 2014–2017 pochodzące z BBDG (GUS, 2015, 2016, 2017, 2018a)¹. Dane te charakteryzują się jednak nadreprezentatywnością grup o niskich dochodach. Występuje tu także problem niedostatecznej reprezentacji najbogatszych gospodarstw domowych. Do zaniżania dochodów dochodzi nawet w przypadku danych uzyskiwanych z legalnych źródeł, co jest argumentem przemawiającym za analizowaniem wielkości wydatków, a nie dochodów w badaniach dotyczących dobrobytu (zob. Kot, 2003).

Przystępując do badania rozkładów dochodów ekwiwalentnych, należy podjąć decyzję o wyborze skali ekwiwalentności, która odzwierciedla wpływ struktury demograficznej gospodarstwa domowego na koszty utrzymania gospodarstw. W niniejszej pracy wykorzystano skalę potęgową – często stosowaną oprócz zmodyfikowanej i oryginalnej skali OECD – z elastycznością ekwiwalentności równą 0,5, czyli pierwiastek z liczby osób w gospodarstwie domowym.

Parametry rozkładu Pareta wykorzystywanego do aproksymacji rozkładów dochodów zamożnych gospodarstw domowych szacowano metodą największej wiarygodności (MNW). Metoda ta została wybrana na podstawie badań symulacyjnych, których rezultaty przedstawiono w pracy Jędrzejczak i Pekasiewicz (2020). Świadczą one o dobrych własnościach estymatorów otrzymanych za pomocą MNW w porównaniu z własnościami (obciążonością, średnim kwadratem błędu) estymatorów uzyskanych innymi metodami, np. metodą momentów ważonych prawdopodobieństwami czy metodą najmniejszych kwadratów lub ważoną metodą najmniejszych kwadratów.

Dobroć, czyli trafność dopasowania rozkładu teoretycznego do empirycznego, oceniona została za pomocą współczynników zgodności: wskaźnika podobieństwa struktur (Wps), współczynnika Mortary² (IM) oraz współczynnika uwzględniającego różnicę między wartością średnią rozkładu empirycznego i wartością oczekiwaną rozkładu teoretycznego (I_r).

Do obliczenia wskaźnika podobieństwa struktur posłużono się wzorem (zob. np. Chomątowski i Sokołowski, 1978; Kruszka, 1989)

$$Wps = \sum_{j=1}^k \min(w_j, \hat{w}_j), \quad (9)$$

gdzie:

k – liczba przedziałów, na które zostały pogrupowane uporządkowane dane,

w_j – empiryczne wskaźniki struktury, $j = 1, \dots, k$,

\hat{w}_j – teoretyczne wskaźniki struktury, $j = 1, \dots, k$.

¹ W trakcie pisania artykułu dane za 2017 r. były najnowszymi dostępnymi danymi.

² Giorgio Mortara (1885–1967) – włoski ekonomista, demograf i statystyk.

Współczynnik Mortary ma postać (zob. np. Zenga i in., 2010)

$$IM = \frac{1}{n} \sum_{j=1}^k |n_j - \hat{n}_j|, \quad (10)$$

gdzie:

k – liczba przedziałów,

n_j – częstości empiryczne,

\hat{n}_j – częstości teoretyczne.

Im wskaźnik podobieństwa struktur jest bliższy 1, a współczynnik Mortary – bliższy 0, tym dopasowanie jest lepsze.

W omawianym badaniu liczba przedziałów, na które zostały podzielone dane z poszczególnych grup, służąca do obliczenia wskaźnika podobieństwa struktur i współczynnika Mortary, wynosiła od 15 do 20, w zależności od liczebności grup.

Zgodność rozkładu empirycznego z wybranym modelem teoretycznym bada się również poprzez porównanie najważniejszych charakterystyk liczbowych rozkładu teoretycznego z odpowiednimi charakterystykami empirycznymi – mogą to być średnie, odchylenia standardowe lub współczynniki nierównomierności. W przeprowadzonych analizach uwzględniono współczynnik będący względną różnicą między średnią wartością rozkładu empirycznego i oczekiwaną wartością rozkładu teoretycznego:

$$I_r = \frac{|\bar{Y} - EY|}{EY} \cdot 100\%, \quad (11)$$

gdzie:

\bar{Y} – średnia arytmetyczna wartości dochodów Y_1, Y_2, \dots, Y_n ,

EY – wartość oczekiwana teoretycznego rozkładu Pareta.

Przyjmuje się, że w przypadku bardzo dobrego dopasowania różnica między empirycznymi i teoretycznymi charakterystykami rozkładu powinna być mniejsza niż 5%.

4. Wyniki analiz rozkładów dochodów zamożnych gospodarstw domowych w grupach społeczno-ekonomicznych

W badaniu analizowano rozkłady miesięcznych dochodów ekwiwalentnych zamożnych gospodarstw domowych w latach 2014–2017.

Na podstawie prób pochodzących z BBGD (GUS, 2015, 2016, 2017, 2018a) dla każdego analizowanego roku wyznaczono granicę zamożności (y_b^*) jako wartość dwukrotności mediany empirycznego rozkładu miesięcznych dochodów ekwiwalentnych. Otrzymano następujące wartości:

- 2014 r. – 4041,46 zł;
- 2015 r. – 4195,36 zł;
- 2016 r. – 4477,12 zł;
- 2017 r. – 4695,74 zł.

Zaobserwowano wzrost granicy zamożności z roku na rok, a średnie tempo zmian w analizowanym okresie wynosiło 5,13%. Udziały gospodarstw zamożnych w Polsce i w poszczególnych grupach społeczno-ekonomicznych, obliczone na podstawie wyznaczonych wartości y_b^* , przedstawiono w tabl. 1.

Tabl. 1. Odsetek gospodarstw o wysokich dochodach w grupach społeczno-ekonomicznych

L a t a	Gospodarstwa					
	ogółem	pracowników	rolników	pracujących na własny rachunek	emerytów i rencistów	utrzymujących się ze źródeł niezarobkowych
2014	8,67	11,05	16,32	19,65	2,71	1,82
2015	8,12	10,16	14,51	20,98	2,82	1,16
2016	7,34	9,33	15,57	19,49	2,07	1,76
2017	7,42	9,32	17,85	19,74	1,90	1,87

Źródło: obliczenia własne na podstawie: GUS (2015, 2016, 2017, 2018a).

W latach 2014–2017 gospodarstwa domowe osiągające miesięczne dochody ekwiwalentne powyżej dwukrotności mediany stanowiły niewiele ponad 7% wszystkich gospodarstw w Polsce, przy czym w analizowanym okresie zaobserwowano zmniejszenie się udziału gospodarstw o wysokich miesięcznych dochodach ekwiwalentnych. W grupie pracujących na własny rachunek i w grupie rolników odsetek gospodarstw domowych o dochodach ekwiwalentnych przekraczających granicę zamożności był najwyższy, a znacznie mniejszy – w grupach emerytów i rencistów oraz utrzymujących się ze źródeł niezarobkowych.

W tabl. 2 i na wyk. 1 przedstawiono charakterystyki empirycznych rozkładów miesięcznych dochodów ekwiwalentnych zamożnych gospodarstw domowych w 2017 r.

Miesięczne dochody ekwiwalentne zamożnych gospodarstw domowych w grupie rolników okazały się bardzo zróżnicowane. W wylosowanej próbie pojawiły się wartości odstające. Można je zidentyfikować na podstawie wykresu ramkowego, który

według Safariego i in. (2019) jest dobrą metodą wyznaczania nietypowych (odstających) wielkości dochodów, lub za pomocą metody kwartyłowej (zob. Chromiński i Tkacz, 2010). Za ekstremalnie nietypowe uznaje się wartości większe od $Q_3 + 3RQ$ lub mniejsze od $Q_1 - 3RQ$, gdzie Q_1 , Q_3 są odpowiednio kwartylem pierwszym i trzecim, a RQ oznacza rozstęp ćwiartkowy.

Tabl. 2. Charakterystyki empirycznych rozkładów dochodów zamożnych gospodarstw domowych w grupach społeczno-ekonomicznych w 2017 r.

Gospodarstwa	Liczba gospodarstw	Minimum	Maksimum	Średnia	Mediana	Odchylenie standardowe	Współczynnik zmienności w %
Ogółem	2721	4696,10	264261,36	6965,53 (6870,94)	5773,50	6404,61 (4083,82)	91,95 (59,43)
Pracowników	1650	4696,10	66178,50	6444,46	5673,62	3031,41	47,12
Rolników	296	4715,67	264261,36	10692,10 (9832,57)	7479,66	16812,30 (7997,43)	157,24 (81,34)
Pracujących na własny rachunek	495	4698,94	57000,00	7048,34	6010,41	3854,53	54,69
Emerytów i rencistów	252	4702,26	22208,00	5857,17	5391,90	1584,69	27,06
Utrzymujących się ze źródeł niezarobkowych	28	4843,97	12625,50	6787,66	5859,72	2052,50	30,24

Uwaga. W nawiasach podano charakterystyki liczbowe po odrzuceniu wartości ekstremalnie odstającej, tzn. 264261,36 zł.

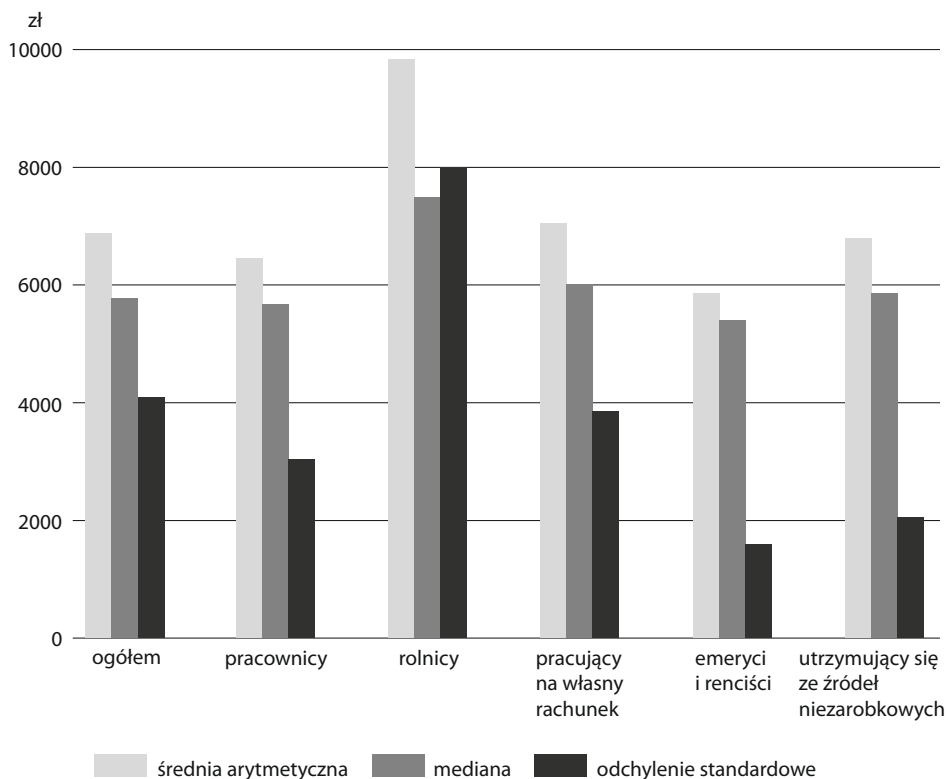
Źródło: obliczenia własne na podstawie: GUS (2018a).

W przypadku zamożnych gospodarstw rolników i wszystkich gospodarstw o wysokich dochodach ekstremalnie odstającą wartością jest miesięczny dochód ekwiwalentny w wysokości 264261,36 zł (dla grupy rolników otrzymano $Q_3 + 3RQ = 26056,977$ zł, a dla ogółu – $Q_3 + 3RQ = 22435,088$ zł), który został pominięty przy aproksymacji rozkładów miesięcznych dochodów zamożnych gospodarstw rolników i ogółu zamożnych gospodarstw domowych w Polsce. W przypadku rolników tak duża wartość miesięcznego dochodu może wynikać z sezonowości ich dochodów. W dalszych analizach ograniczono się do usunięcia jednej wartości odstającej (ekstremalnie odstającej), ponieważ pominięcie zbyt wielu informacji o wysokich dochodach może zaburzyć obraz zamożności.

Porównując wielkości dochodów zamożnych gospodarstw domowych w grupach społeczno-ekonomicznych w 2017 r., można zauważyć, że średnie dochody (wyrażone zarówno średnią arytmetyczną, jak i medianą dochodów) są najwyższe w grupie rolników, a najniższe w grupie emerytów i rencistów. Stopień zróżnicowania

dochodów gospodarstw domowych osiągających wysokie dochody ekwiwalentne w grupach społeczno-ekonomicznych jest niejednakowy. Najbardziej niejednorodna jest grupa gospodarstw rolników, czyli takich, których wyłącznym lub głównym źródłem utrzymania jest dochód z użytkowanego gospodarstwa indywidualnego w rolnictwie, oraz grupa gospodarstw, w której źródłem utrzymania jest praca na własny rachunek poza gospodarstwem indywidualnym w rolnictwie lub wykonywanie wolnego zawodu (współczynniki zmienności odchylenia standardowego odpowiednio ponad 81% i 54%). Zamożne gospodarstwa emerytów i rencistów stanowią niewielki odsetek i są dość jednorodne (współczynnik zmienności – 27%).

Wykr. 1. Podstawowe miary położenia i zróżnicowania dochodów gospodarstw domowych w grupach społeczno-ekonomicznych w 2017 r.



Źródło: opracowanie własne na podstawie: GUS (2018a).

Podobnie kształtowała się sytuacja w latach 2014–2016.

Do oszacowania parametrów rozkładu Pareta aproksymującego rozkłady miesięcznych dochodów ekwiwalentnych wykorzystano MNW. Otrzymane wartości

parametrów rozkładu oraz miary określające dobroć dopasowania rozkładu teoretycznego do danych empirycznych przedstawiono w tabl. 3. W przypadku gospodarstw domowych utrzymujących się ze źródeł niezarobkowych nie można było dopasować rozkładu teoretycznego ze względu na zbyt małą liczebność próby.

Tabl. 3. Oszacowania parametrów rozkładu Pareta aproksymującego rozkład dochodów zamożnych gospodarstw domowych w grupach społeczno-ekonomicznych oraz miary dopasowania

Gospodarstwa	Lata	Parametry rozkładu		Wps	IM	I_r w %
		θ	a			
Ogółem	2014	4016,200	3,151	0,991	0,018	0,757
	2015	4195,410	3,225	0,994	0,012	1,292
	2016	4477,200	3,318	0,998	0,003	1,712
	2017	4696,101	3,235	0,995	0,009	1,084
Pracowników	2014	4016,197	3,361	0,996	0,008	0,237
	2015	4195,406	3,539	0,993	0,014	0,486
	2016	4477,195	3,666	0,991	0,017	0,158
	2017	4696,101	3,728	0,995	0,010	0,419
Rolników	2014	4029,094	1,856	0,942	0,112	7,111
	2015	4207,700	1,832	0,953	0,091	5,378
	2016	4482,770	1,928	0,988	0,023	4,867
	2017	4715,672	1,736	0,954	0,084	11,600
Pracujących na własny rachunek	2014	4017,000	2,619	0,976	0,042	1,586
	2015	4200,000	2,714	0,983	0,032	0,707
	2016	4483,057	2,940	0,958	0,074	2,070
	2017	4698,938	2,949	0,983	0,034	0,866
Emerytów i rencistów	2014	4017,020	5,030	0,925	0,143	0,457
	2015	4197,188	4,782	0,960	0,079	0,081
	2016	4486,593	4,641	0,971	0,057	0,025
	2017	4702,260	5,097	0,968	0,063	0,123

Uwaga. Wps – wskaźnik podobieństwa struktur, IM – współczynnik Mortary, I_r – współczynnik będący względną różnicą średniej wartości rozkładu empirycznego i oczekiwanej wartości rozkładu teoretycznego.

Źródło: obliczenia własne na podstawie: GUS (2015, 2016, 2017, 2018a).

Rozkład Pareta aproksymujący rozkład dochodów zamożnych gospodarstw rolników ma parametr $a < 2$, co oznacza, że momenty rzędów wyższych niż 1 nie istnieją, a tym samym nie istnieje odchylenie standardowe. Rozkład dochodów w tej grupie charakteryzuje się grubszym prawym ogonem niż rozkłady dochodów w pozostałych grupach społeczno-ekonomicznych.

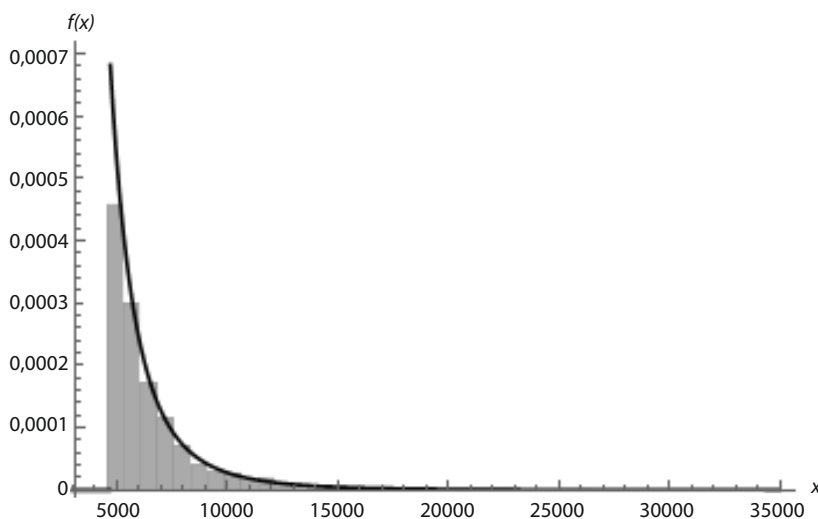
Obliczone miary (Wps, IM, I_r) świadczą o tym, że rozkład Pareta w większości przypadków bardzo dobrze aproksymuje rozkłady empiryczne. Problematyczne może być jedynie dopasowanie rozkładu teoretycznego w grupie gospodarstw rolników w całym analizowanym okresie oraz w grupie gospodarstw emerytów i rencistów w 2014 r. W przypadku grupy rolników o wysokich dochodach ekwiwalentnych współczynnik I_r , określony wzorem (11), jest bliski 5% w 2016 r. (4,87%) lub prze-

kracza 5% w latach 2014, 2015 i 2017, co świadczy o gorszym dopasowaniu rozkładu teoretycznego do danych empirycznych. Pozostałe miary, czyli wskaźnik podobieństwa struktur (większy niż 0,95) i współczynnik Mortary (bliski 0), również pozwalają uznać, że rozkład Pareta z oszacowanymi parametrami dobrze aproksymuje empiryczny rozkład dochodów zamożnych rolników w latach 2015–2017. Tylko w przypadku roku 2014 wszystkie trzy miary przyjmują niesatysfakcjonujące wartości: Wps wynosi nieco poniżej 0,95, I_r jest większy niż 5%, a IM wynosi 0,112. Rozkład Pareta aproksymujący empiryczny rozkład ekwiwalentnych dochodów zamożnych gospodarstw emerytów i rencistów można uznać za gorzej dopasowany, aczkolwiek miara $I_r < 5\%$.

Na wykresach 2 i 3 przedstawiono rozkłady empiryczne wraz z aproksymującymi je teoretycznymi funkcjami gęstości (f) rozkładu Pareta dla dochodów ogółu zamożnych gospodarstw domowych i dla poszczególnych grup społeczno-ekonomicznych w 2017 r.

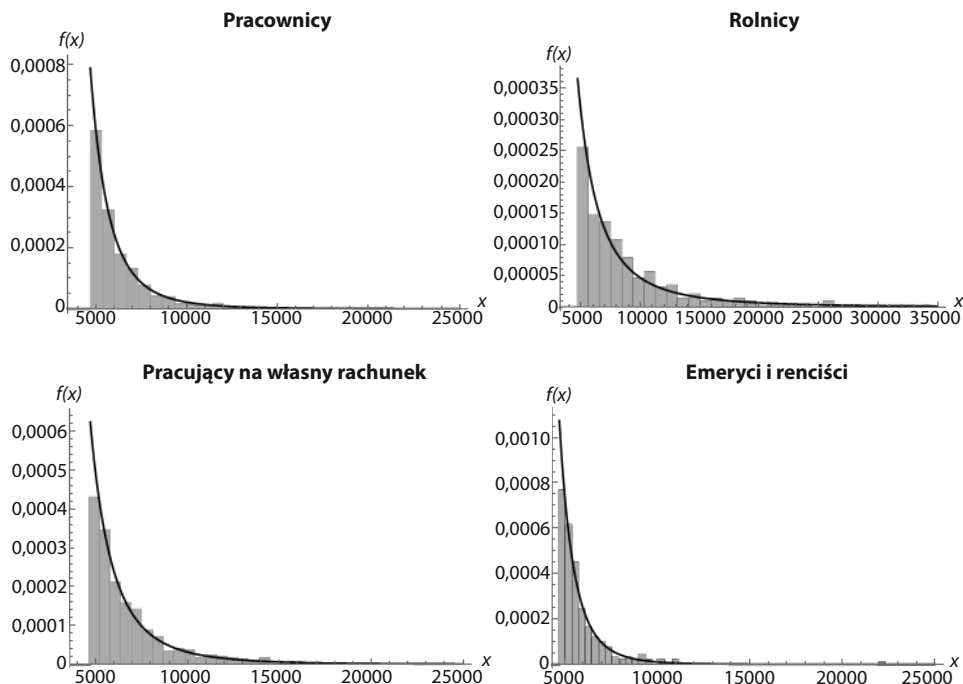
Dobre dopasowanie rozkładu teoretycznego do danych empirycznych prowadzi do alternatywnego sposobu szacowania parametrów charakteryzujących rozważaną populację, polegającego najpierw na aproksymacji rozkładu empirycznego rozkładem teoretycznym, a następnie na wyznaczeniu odpowiednich wielkości na podstawie oszacowanych parametrów funkcji gęstości rozkładu teoretycznego.

Wykr. 2. Aproksymacja rozkładu dochodów zamożnych gospodarstw domowych ogółem w 2017 r.



Uwaga. x – wielkość dochodu ekwiwalentnego gospodarstwa domowego w zł, $f(x)$ – funkcja gęstości rozkładu Pareta dochodu ekwiwalentnego.

Źródło: opracowanie własne na podstawie: GUS (2018a).

Wykr. 3. Aproksymacja rozkładów dochodów zamożnych gospodarstw domowych w grupach społeczno-ekonomicznych w 2017 r.

Uwaga. Jak przy wykr. 2.

Źródło: opracowanie własne na podstawie: GUS (2018a).

Korzystając z otrzymanych rozkładów Pareta, wyznaczono następujące charakterystyki liczbowe rozkładów miesięcznych dochodów ekwiwalentnych w 2017 r.: wartość oczekiwana, mediana, odchylenie standardowe (z wyjątkiem gospodarstw rolników) i miary nierówności (tabl. 4). Różnice między wartościami z próby a teoretycznymi charakterystykami liczbowymi rozkładów są tym mniejsze, im dopasowanie rozkładu teoretycznego do danych empirycznych jest lepsze.

Tabl. 4. Charakterystyki rozkładów Pareta aproksymujących rozkłady dochodów zamożnych gospodarstw domowych w grupach społeczno-ekonomicznych w 2017 r.

Gospodarstwa	Wartość oczekiwana	Mediana	Odchylenie standardowe	Współczynnik Giniego	Współczynnik Schutza-Pietry
	w zł				
O g ó ł e m	6797,27	5818,24	3400,66	0,18	0,14
Pracowników	6417,55	5655,69	2528,48	0,15	0,11
Rolników	11122,84	7029,88	.	0,41	0,31
Pracujących na własny rachunek	7109,89	5943,99	4250,03	0,20	0,15
Emerytów i rencistów	5849,99	5387,25	1472,41	0,11	0,08

Źródło: obliczenia własne na podstawie: GUS (2018a).

Porównując rozkłady dochodów zamożnych gospodarstw domowych w grupach społeczno-ekonomicznych, należy stwierdzić, że rozkład dochodów zamożnych gospodarstw rolników charakteryzuje największa wartość oczekiwana i mediana w porównaniu z rozkładami dochodów zamożnych gospodarstw w pozostałych grupach społeczno-ekonomicznych. W grupie rolników obserwuje się również największe nierówności dochodowe, o czym świadczy najwyższa wartość oszacowanego współczynnika Giniego i potwierdza to najwyższa wartość współczynnika Schutza-Pietry. Zamożne gospodarstwa emerytów i rencistów są grupą najbardziej jednorodną (małe odchylenie standardowe, współczynnik zmienności odchylenia standardowego równy 25,17%, nieduża wartość współczynników Giniego i Schutza-Pietry).

Współczynnik Schutza-Pietry określa, jaki odsetek sumy dochodów ekwiwalentnych wszystkich badanych gospodarstw powinien być transferowany z grupy gospodarstw o dochodach powyżej wartości oczekiwanej do grupy gospodarstw o dochodach równych wartości oczekiwanej lub niższych od niej, aby zniknęły nierówności. W przypadku zamożnych gospodarstw rolników transfer powodujący zniwelowanie nierówności dochodowych musiałby wynieść 31% dochodów, a w grupie emerytów – tylko 8%.

W pracy Włodarczyk (2013) przedstawiono analizy dotyczące rozkładów indywidualnych miesięcznych dochodów netto w Polsce w latach 2003–2011, w których najwyższe dochody modelowano rozkładem Pareta. Nierówności dochodowe mierzone współczynnikiem Giniego wzrastały od 0,156 do 0,216 w poszczególnych latach. W przypadku miesięcznych dochodów ekwiwalentnych gospodarstw domowych w latach 2014–2017 nie obserwuje się takiej tendencji. Różnice pomiędzy nimi w poszczególnych latach są znacznie mniejsze: $\hat{G}_{2014} = 0,189$, $\hat{G}_{2015} = 0,183$, $\hat{G}_{2016} = 0,177$ oraz $\hat{G}_{2017} = 0,183$.

5. Podsumowanie

W badaniu omówionym w artykule analizowano rozkłady miesięcznych dochodów ekwiwalentnych zamożnych gospodarstw domowych w poszczególnych grupach społeczno-ekonomicznych w latach 2014–2017. Badano zgodność rozkładów empirycznych z rozkładem Pareta o parametrach oszacowanych na podstawie próby MNW. Dobre dopasowanie rozkładów teoretycznych (w tym przypadku rozkładów Pareta) do rozkładów empirycznych i wykorzystanie parametrów tych rozkładów pozwoliło na oszacowanie istotnych miar charakteryzujących rozkłady ekwiwalentnych dochodów gospodarstw domowych, w tym miar nierównomierności.

W niniejszej pracy za zamożne gospodarstwa domowe uznano te, których miesięczne dochody ekwiwalentne przekroczyły wartość dwukrotności mediany. Innych

granic zamożności nie rozważano (np. trzykrotności mediany), gdyż dane uzyskane z BBGD nie pozwoliłyby na aproksymowanie dochodów w grupie gospodarstw utrzymujących się ze źródeł niezarobkowych, a także w innych grupach, m.in. gospodarstwach emerytów i rencistów, ze względu na zbyt małe próby, co prowadziło do błędów związanych z estymacją parametrów i dopasowaniem rozkładu Pareta do rozkładów empirycznych.

W przypadku zamożnych gospodarstw domowych ogółem, a także gospodarstw: pracowników, pracujących na własny rachunek oraz emerytów i rencistów rozkłady ekwiwalentnych dochodów zostały bardzo dobrze aproksymowane rozkładem Pareta z parametrami oszacowanymi za pomocą MNW, o czym świadczą wartości wszystkich zastosowanych miar dopasowania. Nieznacznie gorsze wyniki otrzymano dla gospodarstw rolników, szczególnie w 2014 r. i 2017 r., oraz gospodarstw emerytów i rencistów w 2014 r., ale i w tych grupach co najmniej jedna spośród trzech miar określających dobroć dopasowania pozwala uznać rozkład Pareta za dobrze aproksymujący dochody zamożnych gospodarstw domowych.

Rozkład dochodów gospodarstw domowych rolników osiągających wysokie dochody różni się od innych grup, gdyż charakteryzuje się grubszym prawym ogonem, a poza tym w tej grupie występuje większe zróżnicowanie wielkości dochodów ekwiwalentnych, o czym świadczą zarówno odchylenie standardowe i współczynnik zmienności odchylenia standardowego, jak i współczynniki Giniego oraz Schutza-Pietry oszacowane na podstawie dopasowanego rozkładu Pareta.

Parametry kształtu rozkładów Pareta były większe od 3 lub bardzo bliskie tej wartości, z wyjątkiem rozkładu dla zamożnych gospodarstw rolników, co świadczy o stosunkowo niewielkim odsetku gospodarstw domowych o skrajnie wysokich dochodach w badanych grupach. Współczynniki Giniego i Schutza-Pietry wskazują, że nierównomierności dochodów zamożnych gospodarstw domowych w badanych grupach społeczno-ekonomicznych są niezbyt duże.

Bibliografia

- Brzeziński, M. (2014). Statistical Inference for Richness Measures. *Applied Economics*, 46(14), 1599–1608. <https://doi.org/10.1080/00036846.2014.880106>.
- Brzeziński, M., Salach, K., Wroński, M. (2020). Wealth inequality in Central and Eastern Europe: Evidence from household survey and rich lists' data combined. *Economics of Transition and Institutional Change*, 28(4), 637–660. <https://doi.org/10.1111/ecot.12257>.
- Charpentier, A., Flachaire, E. (2019). *Pareto Models for Top Incomes* (Working Papers). <https://hal.archives-ouvertes.fr/hal-02145024/document>.
- Chomątowski, S., Sokołowski, A. (1978). Taksonomia struktur. *Przegląd Statystyczny*, 25(2), 217–226.

- Chromiński, K., Tkacz, M. (2010). Comparison of outlier detection methods in biometrical data. *Journal of Medical Informatics and Technologies*, 16, 89–94. <http://jmit.us.edu.pl/cms/jmitjrn/16/11-18-Chrominski.pdf>.
- Franzini, M., Granaglia, E., Raitano, M. (2016). *Extreme Inequalities in Contemporary Capitalism: Should We Be Concerned About the Rich?*. Switzerland: Springer International Publishing. <https://doi.org/10.1007/978-3-319-28811-6>.
- Główny Urząd Statystyczny. (2015). *Budżety gospodarstw domowych w 2014 r.* Warszawa. <https://stat.gov.pl/obszary-tematyczne/warunki-zycia/dochody-wydatki-i-warunki-zycia-ludnosci/budzety-gospodarstw-domowych-w-2014-r-,9,9.html>.
- Główny Urząd Statystyczny. (2016). *Budżety gospodarstw domowych w 2015 r.* Warszawa. <https://stat.gov.pl/obszary-tematyczne/warunki-zycia/dochody-wydatki-i-warunki-zycia-ludnosci/budzety-gospodarstw-domowych-w-2015-r-,9,10.html>.
- Główny Urząd Statystyczny. (2017). *Budżety gospodarstw domowych w 2016 r.* Warszawa. <https://stat.gov.pl/obszary-tematyczne/warunki-zycia/dochody-wydatki-i-warunki-zycia-ludnosci/budzety-gospodarstw-domowych-w-2016-r-,9,11.html>.
- Główny Urząd Statystyczny. (2018a). *Budżety gospodarstw domowych w 2017 r.* Warszawa. <https://stat.gov.pl/obszary-tematyczne/warunki-zycia/dochody-wydatki-i-warunki-zycia-ludnosci/budzety-gospodarstw-domowych-w-2017-r-,9,12.html>.
- Główny Urząd Statystyczny. (2018b). *Zeszyt metodologiczny. Badanie budżetów gospodarstw domowych.* Warszawa. <https://stat.gov.pl/obszary-tematyczne/warunki-zycia/dochody-wydatki-i-warunki-zycia-ludnosci/zeszyt-metodologiczny-badanie-budzetow-gospodarstw-domowych,10,2.html>
- Jagielski, M., Kutner, R. (2011). Modelowanie zamożności polskich gospodarstw domowych metodami statystycznymi. *Ekonomia*, (25), 154–162. <http://ekonomia.wne.uw.edu.pl/ekonomia/getFile/716>.
- Jędrzejczak, A. (2006). The Characteristic of Theoretical Income Distributions and their Application to the Analysis of Wage Distributions in Poland by Regions. *Acta Universitatis Lodzianis. Folia Oeconomica*, 196, 183–198.
- Jędrzejczak, A., Pekasiewicz, D. (2020). *Teoretyczne rozkłady dochodów gospodarstw domowych i ich estymacja*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.
- Kleiber, C. (1996). Dagum vs. Singh-Maddala Income Distributions. *Economics Letter*, 53(3), 265–268. [https://doi.org/10.1016/S0165-1765\(96\)00937-8](https://doi.org/10.1016/S0165-1765(96)00937-8).
- Kleiber, C., Kotz, S. (2003). *Statistical Size Distributions in Economics and Actuarial Sciences*. Hoboken: John Wiley & Sons. <https://doi.org/10.1002/0471457175>.
- Kot, S. M. (1999). Empiryczna weryfikacja rozkładów płac w Polsce. W: S. M. Kot (red.), *Analiza ekonometryczna kształtowania się płac w Polsce w okresie transformacji* (s. 64–98). Warszawa, Kraków: Wydawnictwo Naukowe PWN.
- Kot, S. M. (2003). Metodologiczne dylematy pomiaru nierówności i dobrobytu. *Nierówności Społeczne a Wzrost Gospodarczy*, (1), 161–179.
- Krishnamoorthy, K. (2006). *Handbook of Statistical Distributions with Applications*. Boca Raton: Chapman & Hall/CRC, Taylor & Francis Group.

- Kruszka, K. (1989). Miary podobieństwa struktury obiektów społeczno-ekonomicznych (studium porównawcze). *Akademia Ekonomiczna w Poznaniu, Zeszyty Naukowe, Seria I*, (159), 48–65.
- Mandelbrot, B. (1963). New methods in statistical economics. *Journal of Political Economy*, 71(5), 421–440.
- Ostasiewicz, K. (2013). Adekwatność wybranych rozkładów teoretycznych dochodów w zależności od metody estymacji. *Przegląd Statystyczny*, 60(4), 499–521. <https://ps.stat.gov.pl/Article/2013/4/499-522>.
- Pareto, V. (1897). *Cours d'économie politique*. Lausanne: Rouge.
- Peichl, A., Schaefer, T., Scheicher, C. (2008). *Measuring Richness and Poverty: A Micro Data Application to Europe and Germany* (IZA Discussion Paper No. 3790). <http://ftp.iza.org/dp3790.pdf>. 26.11.2012.
- Safari, M. A. M., Masseran, N., Ibrahim, K. (2019). On the identification of extreme outliers and dragon-kings mechanisms in the upper tail of income distribution. *Journal of Applied Statistics*, 46(10), 1886–1902. <https://doi.org/10.1080/02664763.2019.1566447>.
- Salamaga, M. (2016). Modelowanie rozkładów dochodów kobiet i mężczyzn w województwie małopolskim. *Wiadomości Statystyczne*, 61(8), 32–44. <https://doi.org/10.5604/01.3001.0014.1065>.
- Sączewska-Piotrowska, A. (2018). *Zamożność dochodowa gospodarstw domowych: determinanty, dynamika, zróżnicowanie przestrzenne*. Katowice: Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach.
- Włodarczyk, J. (2013). Nierówności dochodowe w Polsce według rozkładów Pareta i Boltzmann-Gibbsa. *Studia Ekonomiczne*, 130, 76–87.
- Zenga, M. M., Pasquazzi, L., Zenga, M. (2010). *Rapporto n. 188: First applications of a new three-parameter distribution for non-negative variables*. Mediolan: Dipartimento di Metodi Quantitativi per le Scienze Economiche ed Aziendali, Università degli Studi di Milano Bicocca.