

**Alina Jędrzejczak
Dorota Pekasiewicz**

Teoretyczne rozkłady dochodów gospodarstw domowych i ich estymacja



**Teoretyczne
rozkłady dochodów
gospodarstw
domowych
i ich estymacja**



WYDAWNICTWO
UNIWERSYTETU
ŁÓDZKIEGO

Alina Jędrzejczak
Dorota Pekasiewicz

**Teoretyczne
rozkłady dochodów
gospodarstw
domowych
i ich estymacja**

Alina Jędrzejczak, Dorota Pekasiewicz – Uniwersytet Łódzki
Wydział Ekonomiczno-Socjologiczny, Katedra Metod Statystycznych
90-214 Łódź, ul. Rewolucji 1905 r. nr 41

RECENZENT

Grzegorz Kończak

REDAKTOR INICJUJĄCY

Beata Koźniewska

REDAKTOR WYDAWNICTWA UŁ

Dorota Stępień

SKŁAD I ŁAMANIE

Munda – Maciej Torz

PROJEKT OKŁADKI

Polkadot Studio Graficzne

Aleksandra Woźniak, Hanna Niemierowicz

Zdjęcie wykorzystane na okładce: © Depositphotos.com/Supertrooper

© Copyright by Alina Jędrzejczak, Dorota Pekasiewicz, Łódź 2020

© Copyright for this edition by Uniwersytet Łódzki, Łódź 2020

Wydane przez Wydawnictwo Uniwersytetu Łódzkiego

Wydanie I. W.08891.18.0.K

Ark. wyd. 10,0; ark. druk. 11,25

ISBN 978-83-8142-899-6

e-ISBN 978-83-8142-900-9

<https://doi.org/10.18778/8142-899-6>

Wydawnictwo Uniwersytetu Łódzkiego

90-131 Łódź, ul. Lindleya 8

www.wydawnictwo.uni.lodz.pl

e-mail: ksiegarnia@uni.lodz.pl

tel. (42) 665 58 63

Spis treści

Wprowadzenie	7
1. Teoretyczne rozkłady płac i dochodów	17
1.1. Uwagi wstępne	17
1.2. Rozkład Pareto i jego modyfikacje	22
1.3. Rozkład logarytmiczno-normalny	36
1.4. Rozkład gamma	43
1.5. Rozkład Daguma	47
1.6. Rozkład Singha-Maddali	54
1.7. Rozkład Fiska	59
1.8. Rozkład Weibulla	62
1.9. Rozkład Zengi	65
2. Podejście systemowe do modeli rozkładów dochodów	71
2.1. Uwagi wstępne	71
2.2. System Daguma	72
2.3. System Pearsona	74
2.4. System D'Addario	78
2.5. System Burra	81
2.6. System Johnsa	82
2.7. Rodzina uogólnionego rozkładu beta drugiego rodzaju	85
3. Metody estymacji parametrów rozkładu	91
3.1. Uwagi wstępne	91
3.2. Metoda największej wiarygodności	92
3.3. Metoda momentów	95
3.4. Metoda momentów ważonych prawdopodobieństwami	98
3.5. Metoda percentyli	101
3.6. Kwantylowa metoda najmniejszych kwadratów i jej modyfikacje	103
3.7. Metoda najmniejszych kwadratów	105
3.8. Wazona metoda najmniejszych kwadratów	106
3.9. Metody D'Addario	107
4. Estymacja rozkładów dochodów gospodarstw domowych w Polsce	109
4.1. Uwagi wstępne	109
4.2. Charakterystyka zbioru danych	111

4.3. Aproksymacja rozkładów dochodów gospodarstw domowych w Polsce w podziale na regiony	114
4.4. Aproksymacja rozkładów dochodów zamożnych gospodarstw domowych w Polsce	136
4.5. Analiza symulacyjna własności estymatorów parametrów rozkładu Daguma	147
4.6. Analiza symulacyjna własności estymatorów parametrów rozkładu Pareto	152
Zakończenie	157
Literatura	161
Spis rysunków	169
Spis tabel	171
Aneks. Wybrane charakterystyki funkcyjne i liczbowe podstawowych rozkładów dochodów	173

Wprowadzenie

Zagadnienia rozkładu dochodów i bogactwa oraz związane z nimi pojęcie nierówności ekonomicznej mają stosunkowo długą historię. Obie te kwestie przeplatają się ze sobą w rozważaniach filozofów i ekonomistów, choć przez długi czas zainteresowanie nimi było raczej niewielkie. Pierwsze wzmianki na ten temat znajdujemy już w starożytności w pismach Platona i Arystotelesa. Są one nieliczne, gdyż poniżający charakter pracy w systemie niewolniczym sprzyjał raczej pomijaniu jej w badaniach naukowych. Częściej zajmowano się zagadnieniami sprawiedliwości społecznej z punktu widzenia etyki, choć należy zaznaczyć, że pojęcie równości, sprawiedliwości i bogactwa było rozumiane inaczej niż dzisiaj. Platon w swojej pracy *Państwo* odwoływał się do idei państwa hierarchicznego, w którym sprawiedliwość nie miała nic wspólnego z równością, a ideałem dla stanów wyższych było wyrzeczenie się bogactwa.

W średniowieczu zainteresowanie problemami ekonomicznymi, w tym problemem podziału dochodu, było nadal niewielkie, gdyż praca, choć przeważnie nie była już niewolnicza, nadal nie stanowiła przedmiotu gry rynkowej. Chłopi przywiązani do ziemi nie otrzymywali wynagrodzenia za pracę i inne świadczenia na rzecz feudałów. Pewien wzrost zainteresowania problemami ekonomicznymi, a co za tym idzie także problemami rozkładu dochodów i bogactwa, nastąpił dopiero w XIII wieku, o czym świadczą pisma Tomasza z Akwinu. Można w nich znaleźć poglądy popierające ideę równości, a także niechętny stosunek do bogactwa jako źródła różnic społecznych.

W drugiej połowie XV wieku rozwinął się merkantylizm, który doprowadził do zerwania z podstawami moralnymi i naukowymi Arystotelesa i Tomasza z Akwinu – rosło bowiem znaczenie społeczne dążących do bogactwa kupców. Wynikiem krytyki starego systemu były m.in. prace Jeana-Jacques’a Rousseau poświęcone w dużej mierze ideom równości ekonomicznej i sprawiedliwości. Poglądy te zostały zawarte w dziele *O umowie społecznej*.

W ostatnim trzydziestoleciu XVIII wieku, wraz z nastaniem rewolucji przemysłowej, zaczęła kształtować się klasa robotnicza, a wraz z nią bardzo istotne stały się czynniki określające płace. Na ten okres przypada działalność Adama Smitha, której efektem była praca *Badania nad naturą i bogactwem narodów*. Dla Smitha płaca była kosztem wychowania robotnika do chwili podjęcia pracy oraz utrzymania go przez okres jej trwania. Przedstawiona przez Smitha teoria płacy, oparta na kosztach utrzymania robotnika przy życiu, została później przekształcona przez Davida Ricardo w „żelazne prawo płac”, które mówiło, że płace są ceną płaconą przez pracodawców, umożliwiającą ogółowi robotników utrzymanie, a jednocześnie niepowodującą ani zwiększenia, ani zmniejszenia ich liczby. Ricardo twierdził ponadto, że płace powinny być całkowicie pozostawione wolnej konkurencji, a państwo nie powinno ich kontrolować. W XIX wieku wraz z rozwojem ekonomii klasycznej nastąpił więc okres, w którym płace kształtowane były na zasadzie minimum egzystencji, a często, w wyniku napływu robotników do miast, nawet poniżej tego minimum.

Przełom XIX i XX wieku, kiedy coraz częściej zaczęły się pojawiać recesje, spowodował presję na interwencję państwa w gospodarkę, w tym także w zasady podziału wytworzonego bogactwa. Chodziło między innymi o zmniejszenie nierówności dochodowych, czego domagali się ekonomiści i filozofowie o orientacji socjalistycznej. Właśnie w takich warunkach powstała przełomowa dla teorii rozkładów dochodów praca Vilfredo Pareto *Kurs ekonomii politycznej* (1897), a jej głównym celem była obrona klasycznego systemu ekonomicznego. Pozycja ta zawierała pierwszą sformalizowaną teorię rozkładu dochodów i nierówności dochodowych. Wcześniej zainteresowanie badaczy rozkładami dochodów i bogactwa oraz wynikającymi z nich nierównościami nie było poparte merytoryczną specyfikacją modeli, szacunkami parametrów oraz analizą społeczno-ekonomiczną. Pareto, obserwując rozkłady dochodów w różnych krajach (opierając się na statystykach podatkowych) zauważył prawidłowość, że 20% najbogatszych osób skupiało aż 80% bogactwa, jednak uznał, że nierówności te są zjawiskiem całkowicie naturalnym i zapewniają stabilność gospodarki oraz makroekonomiczne korzyści. Podział ten, choć bardzo nierównomierny, odzwierciedlał jego zdaniem naturalny rozkład zdolności i talentów w systemie społecznym. Na skutek

działania praw rynkowych system taki osiągał stan zwany optimum Pareto (lub optimum w sensie Pareto), w którym niemożliwa była poprawa dobrobytu żadnej jednostki bez pogorszenia sytuacji innych. Na podstawie obserwacji empirycznych Pareto sformułował pierwszy rozkład teoretyczny, który do dziś uważany jest za idealny model rozkładu dochodów dla najbogatszej części populacji. Mimo powstawania nowych nurtów odpowiadających na nowe wyzwania związane z funkcjonowaniem systemu kapitalistycznego jeszcze w latach trzydziestych XX wieku ekonomiści sądzili, że można posługiwać się tzw. optymalnością Pareto do oceny gospodarki bez wypowiadania sądów wartościujących. Próbowano formułować warunki, w których równowaga ogólna osiągnie optimum Pareto, co miało mieć znaczenie dla dobrobytu społecznego i efektywnej alokacji zasobów.

Można stwierdzić, że praca Pareto (1897) poruszająca w sposób sformalizowany problemy podziału dochodu, choć niepozabawiona poglądów kontrowersyjnych, zapoczątkowała szerokie badania nad kształtowaniem się rozkładów dochodów i ich nierównomiernością, które prowadzone były przez wielu statystyków i ekonomistów. W badaniach tych można wyodrębnić trzy podstawowe nurty:

- 1) poszukiwania modelu teoretycznego opisującego rozkłady empiryczne;
- 2) tworzenie teorii wyjaśniających proces kształtowania się rozkładów dochodów i ich nierównomierności;
- 3) doskonalenie metod badania nierównomierności rozkładów.

Do pierwszych prac dotyczących teoretycznych rozkładów płac i dochodów, obok wspomnianej już pracy Pareto, można zaliczyć dokonania Marcha (1898) i Edgewortha (1898). March zaproponował wykorzystanie rozkładu gamma, zaś Edgeworth zastosował rozkład normalny jako funkcję generującą rozkłady dochodów, co później zostało wykorzystane między innymi przez D'Addario (1949) do stworzenia pierwszego systemu funkcji gęstości opisujących rozkłady płac i dochodów. Na początku XX wieku Benini (1906) przedstawił interesującą propozycję modyfikacji krzywej Pareto, natomiast Vinci (1921), opierając się na pewnych przesłankach probabilistycznych, dowodził, że krzywa Pearsona V typu powinna dobrze aproksymować empiryczne rozkłady dochodów. W tym samym czasie powstała praca Amoroso (1925), w której przedstawiona została propozycja uogólnienia rozkładu gamma. Jednocześnie przeprowadzano

coraz szersze badania empiryczne płac i dochodów – należy tu wymienić m.in. pracę Macaulaya (1922) dotyczącą badania rozkładów w USA oraz analizy Gibrata (1931), który badając rozkłady dochodów we Francji, zastosował po raz pierwszy rozkład logarytmiczno-normalny.

W Polsce pierwsze prace na temat rozkładów płac i dochodów pochodzą z okresu międzywojennego i bazują na informacjach czerpanych ze statystyk ubezpieczeń społecznych (por. np. Wiśniewski, 1934; Piltz, 1929). Na szczególną uwagę zasługują badania Wiśniewskiego (1934), który jako pierwszy wykazał przydatność rozkładu logarytmiczno-normalnego do badań płac pracowników biurowych. Po II wojnie światowej do rozwoju badań nad rozkładami płac i dochodów przyczynił się Vielrose (1960), który w swojej pracy przedstawił metody analizy rozkładów dochodów w oparciu o rozkłady teoretyczne. Autor ten wykazał między innymi przydatność rozkładu logarytmiczno-normalnego do aproksymacji płac pracowników w gospodarce uspołecznionej, a także udowodnił, że rozkład płac w Polsce nie jest zgodny z rozkładem Pareto. Rozważania nad przydatnością rozkładu Pareto i logarytmiczno-normalnego prowadzone były w tym czasie także przez Langego (1967), Kordosa (1968) oraz Kordosa i Stroińską (1971). W kolejnych latach badano również możliwość zastosowania innych rozkładów teoretycznych. W szczególności Kordos (1973) oraz Kordos i Stroińska (1971) rozważali zgodność rozkładów płac w Polsce w podziale na pracowników fizycznych i umysłowych oraz w podziale według płci z krzywymi Pearsona typu I, III, IV i VI oraz z rozkładem normalnym. Dalsze badania rozkładów płac w przekroju dynamicznym prowadzone były przez Wąsika (1967), który porównywał zgodność rozkładów empirycznych z dwu- i trójparametrowym rozkładem logarytmiczno-normalnym. Domański (1973) analizował rozkłady płac w przekroju terytorialnym oraz rozkłady dochodów gospodarstw domowych emerytów (por. także: Domański, Michałkiewicz, 1974).

W ramach drugiego nurtu powstało wiele teorii wyjaśniających proces kształtowania się rozkładu dochodów, który według badań empirycznych był na ogół jednomodalny, dodatnio asymetryczny i charakteryzował się tzw. grubym ogonem.

W latach trzydziestych XX wieku Champernowne (1937) zaproponował pierwszy model stochastyczny dotyczący rozkładu dochodów, zdążający do rozkładu Pareto. Opierał się on na szeregu upraszczających

założeń dotyczących na przykład stałości liczby osób otrzymujących dochód oraz stałości w czasie macierzy prawdopodobieństw przejścia do następnej grupy dochodowej. Kolejnym ważnym krokiem do wyjaśnienia mechanizmu kształtowania się rozkładu dochodów była cytowana już praca Gibrata (1931), który wykorzystał prawo proporcjonalnego efektu do zbudowania procesu stochastycznego, którego wynikiem był rozkład logarytmiczno-normalny. Prawo proporcjonalnego efektu w odniesieniu do dochodów zakładało, że zmiany dochodu w jakimkolwiek czasie są proporcjonalne do jego wartości w poprzednim okresie. Model Gibrata był następnie modyfikowany między innymi przez Kaleckiego (1945) oraz Aitchisona i Browna (1957).

Inną znaną teorią stochastyczną prowadzącą do rozkładu Pareto była teoria Lydalla (1968). Zasadniczym jej elementem jest założenie, że pracownicy w przedsiębiorstwie tworzą pewnego rodzaju piramidę, na szczycie której stoi dyrektor, zaś na samym dole znajdują się robotnicy. Lydall zakładał, że pracownicy na jakimkolwiek stopniu piramidy z wyjątkiem ostatniego kontrolują stałą liczbę osób, zaś suma całkowitego zarobku na dowolnym stopniu jest proporcjonalna do sumy płac na niższym szczeblu, przy czym współczynnik proporcjonalności jest stały.

Obok teorii stochastycznych powstawały też teorie socjologiczne i ekonomiczne próbujące wyjaśnić problem nierównomierności rozkładu dochodów oraz ich skośności. Najwcześniejszą z nich była teoria Taussiga (1915). Według niego o nierównomierności rozkładu dochodów decyduje zróżnicowanie wrodzonych zdolności oraz wpływ otoczenia i dziedziczenia. Podobne poglądy reprezentował Pigou (1920). Sądził, że skośność rozkładów dochodów jest wynikiem wpływu cech osobistych jednostek (zdolności) oraz ich zamożności. Rozkład zależny jedynie od cech osobistych, a więc działających addytywnie, byłby rozkładem normalnym. Garvey (1952) wymienił szereg czynników, które mogą określać rozkład dochodów. Należą do nich, obok czynników społeczno-ekonomicznych, także czynniki demograficzne, takie jak: płeć, wiek, fluktuacje zatrudnienia, czynniki geograficzne, a także jednostka czasu, do której odnosi się badany rozkład. Interesującym wyjaśnieniem nierównomierności rozkładu dochodów była teoria Tinbergena (1956). Zgodnie z nią rozkład dochodów jest wynikiem podaży i popytu na określone cechy jakościowe pracowników, zaś jego nierównomierność powodują „napięcia” między

tym, co dostępne na rynku pracy, a tym co pożądane (pracownicy o określonych kwalifikacjach). Inną koncepcją Tinbergena (1972) wyjaśniającą zróżnicowanie i asymetrię rozkładu dochodów była teoria o nakładaniu się dwóch rozkładów normalnych o różnych wartościach oczekiwanych i różnym stopniu zróżnicowania. Jednym z tych rozkładów miał być rozkład zdolności, drugim zaś rozkład skłonności do pracy. Podobne podejście reprezentował Friedman (1953), jednak mówił on o dwóch rozkładach, reprezentujących zbiorowości o różnym stopniu skłonności do ryzyka. Uważał, że w każdym społeczeństwie istnieje nieliczna grupa o dużej skłonności do ryzyka oraz duża grupa wykazująca niską skłonność do ryzyka. W wyniku nakładania się rozkładów dochodów obu tych grup powstaje rozkład charakteryzujący się dodatnią asymetrią.

W ramach trzeciego nurtu, zajmującego się badaniem nierównomierności rozkładów, do najwcześniejszych dokonań należą: praca Lorenza (1905), w której zaproponowana została krzywa koncentracji zwana później krzywą Lorenza, oraz dzieła włoskiego ekonomisty i matematyka Giniego (1909; 1912), który koncentrował się głównie na problemach nierównomierności i interakcji pomiędzy rozkładami dochodów. Należy tu także wymienić pionierską pracę Daltona (1920), zapoczątkowującą rozważania nad nierównomiernością dochodów w powiązaniu z zagadnieniem społecznego dobrobytu.

Podsumowując, można stwierdzić, że począwszy od czasów Pareto, problem znalezienia modelu rozkładu dochodów, a także wyjaśnienia i badania ich nierównomierności stał się obecny w rozważaniach statystyków, ekonomistów i socjologów. W poglądach na nierównomierność zaobserwować możemy dwa przeciwstawne stanowiska: pierwsze, związane z ekonomią klasyczną i ortodoksyjnym liberalizmem oraz drugie, występujące w różnych nurtach krytycznych do ekonomii klasycznej – od socjalizmu utopijnego do keynesizmu. Zgodnie z ekonomią liberalną dążenie do sprawiedliwości społecznej jest sprzeczne z wymogami wydajności ekonomicznej, bowiem podział będący wynikiem niezakłóconej gry rynkowej stwarza większe możliwości oszczędzania. W wyniku modyfikacji systemu rynkowego może dojść do zakłócenia rynkowego procesu alokacji zasobów zapewniającego maksymalizację wydajności (por. np. Hayek, 1981). W poglądach liberalnych dominuje więc teza, że wydajność i sprawiedliwość są nie do pogodzenia lub też przyjmowana jest specyficz-

na koncepcja sprawiedliwości, polegająca na respektowaniu żywiołowego podziału rynkowego. (Znalazło to odbicie w wielu poglądach na temat źródła nierównomierności dochodów). Keynes i jego kontynuatorzy będący przedstawicielami drugiego nurtu dowodzili, że zmniejszenie niepewności ekonomicznej i wzrost gospodarczy nie są sprzeczne, lecz ściśle ze sobą powiązane. Tezę tę potwierdził pomyślny rozwój gospodarek zachodnich w latach pięćdziesiątych i sześćdziesiątych XX wieku. Załamanie się korzystnego trendu „epoki Keynesa” związane z wystąpieniem wysokiej stopy inflacji połączonej ze spadkiem produkcji i wysokim bezrobociem, a także skutki ostatniego kryzysu finansowego z 2008 roku, podważają jednak bezwarunkową słuszność takich poglądów. W różnych społeczeństwach, w różnych sytuacjach historycznych, przy różnym poziomie sił wytwórczych wzrost gospodarczy może wymagać różnych rozwiązań w tej dziedzinie. Wciąż aktualny wydaje się wkład noblisty Amartya Sena, przedstawiciela nurtu ekonomii dobrobytu, twórcy tzw. teorii wyboru społecznego (Sen, 1973), który rozważa, w jaki sposób jednostkowe preferencje mogą prowadzić do decyzji optymalnych i sprawiedliwych dla ogółu. Problem wpływu decyzji indywidualnych i społecznych na rozkład dochodu i jego nierównomierność podejmuje także dedykowana Camilo Dagumowi praca Chotikapanicha (2008), gdzie znajdujemy uzasadnienie podejścia parametrycznego w badaniu rozkładów dochodów oraz dokładny przegląd metodologii modelowania podziału dochodu.

Mimo różnych poglądów na związki między nierównością a wzrostem gospodarczym, które podlegały także weryfikacji empirycznej (por. Kumor, Sztandynger, 2007; Jędrzejczak, 2015), globalizacja stosunków ekonomicznych oraz presja krajów rozwijających się prowadzi do wzrostu znaczenia problemów podziału dochodu i równości społecznej w badaniach ekonomicznych. Wiąże się to z kolei z potrzebą badania rozkładów dochodów w różnych przekrojach, strukturalnych i koniunkturalnych sfer ubóstwa, a także dobrobytu społecznego. Problemy podziału dochodu i narastających nierówności we współczesnym świecie, uzupełnione pogłębioną analizą tego zjawiska w ujęciu dynamicznym, rozważane były ostatnio w bestsellerowej pracy Piketty (2014). Należy podkreślić znaczenie takich badań dla ekonomii rozwoju oraz dla decydentów, którzy projektują i oceniają politykę ograniczania ubóstwa i redystrybucji dochodów.

Jednym z głównych kierunków badań w dziedzinie badania rozkładów dochodów powinno być poszukiwanie modelu teoretycznego opisującego empiryczne rozkłady płac i dochodów.

Analiza rozkładów dochodów za pomocą rozkładów teoretycznych jest uzasadniona z kilku powodów:

- 1) zastosowanie modelu teoretycznego upraszcza analizę, ponieważ różne charakterystyki rozkładu, w tym skomplikowane punktowe i syntetyczne miary nierównomierności, można oszacować przy użyciu niewielkiej liczby tych samych parametrów. Podejście takie ułatwia także badanie własności tych charakterystyk wyrażonych jako funkcje określonych parametrów rozkładu teoretycznego;
- 2) model teoretyczny dobrze dopasowany do danych można wykorzystać do prognozowania rozkładu płac i dochodów w różnych przekrojach, zarówno w czasie, jak i w przestrzeni, co może znaleźć zastosowanie na przykład w statystyce małych obszarów w ramach tzw. podejścia modelowego;
- 3) aproksymacja empirycznych rozkładów płac i dochodów za pomocą odpowiednich rozkładów teoretycznych może służyć jako metoda wyrównywania nieprawidłowości wynikających z metody gromadzenia danych.

Dotychczasowe badania zaowocowały dużą liczbą rozkładów teoretycznych, które z mniejszym lub większym powodzeniem mogą być wykorzystywane do analizy płac, dochodów czy zamożności; większość z nich nie spełnia jednak wymagań stawianych rozkładowi teoretycznym, które umożliwiają ich wykorzystanie do pogłębionych analiz empirycznych w różnych przekrojach.

Celem niniejszej pracy jest prezentacja własności najważniejszych rozkładów teoretycznych oraz ich systemów, które mogą być zastosowane do analizy rozkładów płac, dochodów oraz ich nierównomierności. Rozkłady wybrane na podstawie rozważań teoretycznych i badań symulacyjnych zostały zastosowane do analizy dochodów w Polsce. Estymacja parametrów rozkładów była przeprowadzona w oparciu o dane indywidualne pochodzące z badania budżetów gospodarstw domowych GUS.

Publikacja składa się z czterech rozdziałów. W rozdziale 1 zostały omówione teoretyczne rozkłady płac i dochodów, m.in. rozkład Pareto, Daguma, logarytmiczno-normalny, gamma oraz Singha-Maddali. Przed-

stawiono ich charakterystyki funkcyjne i liczbowe, z uwzględnieniem miar nierównomierności rozkładów. Badanie struktury i własności rozkładów ułatwia ich pogrupowanie w systemy lub rodziny, dlatego w rozdziale 2 zaprezentowano podstawowe systemy rozkładów: Daguma, Pearsona, D'Addario, Burra, Johnsona oraz uogólniono rozkład beta drugiego rodzaju. W rozdziale 3 przedstawiono wybrane metody szacowania parametrów rozkładów, które mogą znaleźć zastosowanie przy aproksymacji rozkładów empirycznych różnymi typami rozkładów teoretycznych do estymacji ich parametrów. Ostatni rozdział ma charakter aplikacyjny, ukazano w nim rezultaty badań rozkładów dochodów w Polsce w ujęciu regionalnym. Badano zarówno rozkłady dochodów wszystkich gospodarstw domowych, jak i rozkłady dochodów gospodarstw domowych o wysokich dochodach.

ROZDZIAŁ 1

Teoretyczne rozkłady płac i dochodów

1.1. Uwagi wstępne

Aitchison i Brown (1957) podali cztery podstawowe cechy jako wskazówkę do identyfikacji najbardziej reprezentatywnego modelu nieznanego procesu stochastycznego, który generuje rozkład dochodu. Następnie ten sam problem rozważali Metcalf (1969) i Dagum (1977). Dagum i Lemmi (1989), opierając się na wcześniejszych pracach, przedstawili szereg własności rozkładu teoretycznego, który można uznać za dobry model rozkładu dochodów. Są one następujące:

- a) podstawy teoretyczno-empiryczne;
- b) zbieżność z prawem Pareto;
- c) istnienie tylko niewielkiej liczby skończonych momentów rozkładu,
- d) możliwość interpretacji ekonomicznej parametrów rozkładu;
- e) dobre dopasowanie do rozkładów empirycznych;
- f) przydatność do analizy zarówno rozkładów jednomodalnych, jak i zeromodalnych;
- g) wysoka zgodność z rozkładami o dochodach ujemnych i zerowych;
- h) proste i efektywne metody szacowania parametrów;
- i) możliwość wyznaczenia krzywej Lorenza;
- j) możliwość wyprowadzenia współczynnika Giniego;
- k) niezmienność modelu w przypadku zmian skali pomiaru dochodów;
- l) zasada oszczędności.

Punkty e, f, g dobrego modelu rozkładu dochodów dotyczą możliwości dostosowania modelu do sytuacji, jakie mogą zaistnieć w rzeczywistości.

Dobry rozkład teoretyczny powinien dostosowywać się do zmian w kształcie rozkładów empirycznych poprzez zmiany wielkości parametrów, wtedy będzie dobrze opisywał rozkłady w różnych przekrojach i okresach. Innym problemem jest możliwość odzwierciedlenia dochodów zerowych i ujemnych. Większość znanych rozkładów teoretycznych może być stosowana jedynie dla dochodów dodatnich – wprowadzenie dochodów ujemnych i zerowych wymaga dodatkowych zabiegów (por. np. Aitchison i Brown, 1957). Do wyjątków należy rozkład Daguma typu II. Nie mniej istotnym zagadnieniem jest dopasowywanie rozkładu teoretycznego do sytuacji, w których minimalny dochód jest większy od zera. Powinno być to możliwe przez zmiany w parametrach rozkładu bez konieczności „ucinania go” (przykładem może być rozkład Daguma typu III lub trójparymetryczny rozkład logarytmiczno-normalny).

Za jedną z cech teoretycznych rozkładu prawdopodobieństwa, która może być przydatna w analizie płac i dochodów, Dagum i Lemmi (1989) uznali możliwość przejścia w rozkład zeromodalny. W praktyce zdarzają się sytuacje, kiedy rozkłady dochodów nie osiągają w żadnym punkcie maksimum. Może to mieć miejsce, gdy rozważamy rozkłady dochodów krajów rozwijających się o dużej liczbie ludności, mających znaczny udział zatrudnionych w rolnictwie lub w przypadku krajów z niewielkim odsetkiem osób bardzo bogatych i dużym odsetkiem bardzo biednych. Warto również dodać, że rozkłady zamożności są zawsze zeromodalne.

Zasada „oszczędności” polega na takim dobraniu modelu teoretycznego, aby opisywał on empiryczne rozkłady dochodów w sposób możliwie najprostszy, w szczególności z niewielką liczbą nieznanymi parametrów rozkładu, które będą musiały być oszacowane. Ogólnie wiadomo, że modele najprostsze – dwuparametrowe – wykazują często dość niski stopień zgodności z rzeczywistymi rozkładami płac i dochodów. Wyjątkiem jest krzywa Pareto, której stopień dopasowania do danych empirycznych jest bardzo wysoki, oczywiście jeśli weźmiemy pod uwagę grupy o wysokich dochodach. Ogólnie można jednak stwierdzić, że do odzwierciedlenia zjawisk ekonomicznych o takiej złożoności, jaką mają rozkłady dochodów, należy wykorzystać krzywe o liczbie parametrów większej od dwóch. Rozkłady trójparymetryczne (Daguma, Singha-Maddali), wykazujące bardzo wysoki stopień zgodności z rozkładami empi-

rycznymi, potwierdzają tezę, że liczba parametrów równa trzy na ogół wystarcza do opisania rzeczywistych rozkładów płac i dochodów.

Możliwość interpretacji ekonomicznej parametrów proponowanych rozkładów teoretycznych jest bardzo istotna ze względu na to, że wszystkie charakterystyki rozkładów dochodów, takie jak średnia, odchylenie standardowe czy współczynniki koncentracji, są funkcjami tych parametrów. Wśród parametrów danego rozkładu teoretycznego powinno się dać wyróżnić parametry kształtu (nierównomierności) i parametry skali lub też parametry położenia. Parametry nierównomierności określają kształt rozkładu, a więc mają wpływ na wielkość współczynników zróżnicowania, koncentracji i asymetrii, natomiast parametry skali zależą od jednostek pieniężnych, w których wyrażone są dochody i mają wpływ głównie na miary średnie badanych rozkładów.

Przyjmując za kryterium idee, jakie leżały u podstaw konstrukcji lub wyboru konkretnej funkcji opisującej rozkład dochodów, można podzielić je na cztery grupy:

- rozkłady „ad-hoc”;
- rozkłady utworzone przez analogię;
- rozkłady oparte na przesłankach ekonomicznych;
- rozkłady oparte na procesach stochastycznych.

Podstawą wyboru rozkładów „ad-hoc” jest to, że charakteryzują się dodatnią asymetrią. Do tej grupy zalicza się: rozkład gamma proponowany przez wielu autorów, m.in. Marcha (1898), Amoroso (1925), Salema i Mounta (1974), rozkład beta zaproponowany przez Thurowa (1970) i Mc Donalda (1984), logarytmiczną transformację rozkładu *t*-Studenta, podaną przez Kloeka i van Dijka (1977), rozkład Zengi (Zenga, 2010; Zenga i in., 2010a; 2010b), rozkłady Champernowne’a (1952) i Fiska (1961), krzywe Pareto I i III typu (Pareto, 1897), krzywą Pearsona typu V (Vinci, 1921) i typu VI (Mc Donald, 1984). Nie wszystkie z wymienionych rozkładów znalazły powszechne zastosowanie. W szczególności rozkład beta był rzadko wykorzystywany głównie z powodu trudnych do interpretacji parametrów, natomiast krzywe Pearsona ze względu na zbyt dużą liczbę parametrów i trudności przy ich szacowaniu (krzywa zaproponowana przez Mc Donalda ma aż 4 parametry niezależne).

Teoretyczny rozkład dochodów może być także wynikiem formalnej adaptacji lub dostosowania zmiennych modelu, który był pierwotnie wykorzystywany w innej dziedzinie. Do takich rozkładów możemy zaliczyć model Davisa (1941) – oparty na analogii do prawa Plancka, dotyczącego rozkładu energii widmowej ciała będącego w stałej temperaturze (prawo to stało się początkiem teorii kwantów). Davis porównał ciało o stałej temperaturze do gospodarki w danym momencie czasu, natomiast energię emitowaną przez to ciało interpretował jako strumień dochodów. Funkcja Davisa nie znalazła szerszego zastosowania głównie ze względu na skomplikowaną formułę i trudne do interpretacji parametry. Singh i Maddala (1976) dokonali adaptacji teorii stopy ryzyka opartej na długości życia. Zamiast prawdopodobieństwa przeżycia co najmniej do danego wieku wprowadzili prawdopodobieństwo osiągnięcia dochodu co najmniej „ y ”. W efekcie otrzymali dystrybuantę rozkładu dochodów zależną od trzech parametrów i dającą bardzo wysoki stopień zgodności z rozkładami empirycznymi.

Bardzo ważną grupę rozkładów teoretycznych opisujących płace i dochody skonstruowano, opierając się na przesłankach ekonomicznych wynikających z badań empirycznych. Pierwszym takim rozkładem była krzywa Pareto typu I zaproponowana w 1895 roku. Obserwując rozkłady dochodów różnych krajów europejskich, Pareto (1895; 1897) zauważył pewne prawidłowości w zachowaniu się elastyczności dochodowej dopełnienia do dystrybuanty dla grup o wysokich dochodach. Dla Daguma (1977) analiza empirycznych rozkładów dochodów również stała się podstawą do wyprowadzenia modelu teoretycznego. Zaobserwował, że elastyczność dochodowa dystrybuanty jest malejąca, przeważnie wypukła i, w odróżnieniu od modelu Pareto, ograniczona funkcją dystrybuanty dla wszystkich poziomów dochodu. Dystrybuenta rozkładu Daguma została znaleziona jako wynik rozwiązania równania różniczkowego spełniającego wyżej wymienione kryteria (elastyczność malejąca, wypukła i ograniczona). Teoretyczny model rozkładów płac dochodów otrzymany przez Daguma wykazuje bardzo wysoki stopień zgodności z rozkładami empirycznymi, natomiast krzywa, którą zaproponował Pareto, jest do dnia dzisiejszego uważana za idealną dla wysokich grup dochodowych.

Kolejną grupę rozkładów teoretycznych stosowanych w analizie płac i dochodów stanowią rozkłady będące wynikiem pewnych procesów stochastycznych. Do najważniejszych należy rozkład logarytmiczno-normalny zaproponowany przez Gibrata (1931), który został oparty na prawie proporcjonalnego efektu. Gibrat zakładał, że relacja pomiędzy logarytmem dochodu a zmienną dochodową o rozkładzie normalnym jest liniowa. Innego zdania był Rutherford (1955), który zauważył, że rozkłady dochodów z różnych krajów nie są zgodne ani z rozkładem logarytmiczno-normalnym, ani z krzywą Pareto. Podjął więc próbę zbudowania innego modelu oraz procesu stochastycznego, który by do niego zdążał. Wysunął przypuszczenie, że zależność określona przez Pareto jako liniowa jest w rzeczywistości wielomianem trzeciego stopnia. Efektem jego prac był czteroparametrowy rozkład teoretyczny, który, jak wynika z dowodów samego autora (Rutherford, 1955), dąży do rozkładu Gram-Chaliera typu A (por. Cramer, 1958). Z kolei Champernowne (1937; 1953) zbudował model stochastyczny oparty na łańcuchach Markowa, którego efektem jest rozkład dochodów zgodny z krzywą Pareto. Model ten był później modyfikowany przez Aitchisona i Browna (1957), którzy dowiedli jego zbieżność z rozkładem logarytmiczno-normalnym. Należy zaznaczyć, że również rozkład zaproponowany przez Daguma (1977), mimo że został utworzony na podstawie analizy rozkładów empirycznych, może być przedstawiony jako wynik procesu stochastycznego. Fattorini i Lemmi (1979) otrzymali dystrybucję rozkładu Daguma, korzystając z równania Kołmogorowa (znanego w fizyce jako równanie Fokkera-Plancka) oraz czyniąc szereg założeń dotyczących stałości średniej i wariancji logarytmów zmian dochodowych (por. Fisz, 1967).

W rozdziale tym przedstawione zostały własności rozkładów wykorzystywanych w analizach płac i dochodów, w szczególności charakterystyki funkcyjne i wybrane charakterystyki liczbowe, z uwzględnieniem współczynników określających stopień nierównomierności rozkładów. Omówione rozkłady reprezentują wszystkie cztery wyszczególnione, ze względu na idee konstrukcji, grupy rozkładów.

1.2. Rozkład Pareto i jego modyfikacje

Rozkład Pareto należy do rozkładów, które zaproponowano w wyniku obserwacji empirycznych. Pareto obserwował rozkłady dochodów ludności uzyskane ze statystyk podatkowych różnych krajów europejskich. Stwierdził, że między wysokością dochodów (y) a frakcją osób posiadających dochody nie mniejsze od określonej wysokości θ istnieje pewna prawidłowość, która została ujęta jako krzywa Pareto pierwszego typu (Pareto, 1895; 1897). Jej klasyczna postać jest następująca:

$$P(y) = A(y - \theta)^{-a}, \quad (1.1)$$

gdzie:

$P(y)$ – dopełnienie do dystrybuanty,

θ – minimalny dochód,

A, a – parametry rozkładu (a – stała Pareto).

Funkcja wyrażona wzorem (1.1) przedstawiona w skali podwójnie logarytmicznej przyjmuje postać linii prostej, której współczynnik kierunkowy wynosi a :

$$\ln P(y) = \ln A - a \ln (y - \theta). \quad (1.2)$$

Jeżeli przyjmiemy, że $a > 0$ i dochody powyżej wielkości θ przedstawimy jako jej wielokrotności, to dystrybuantę rozkładu Pareto $Pa(\theta, a)$ możemy wyrazić w następującej postaci:

$$F(y) = \begin{cases} 1 - \left(\frac{y}{\theta}\right)^{-a} & \text{dla } y \geq \theta, \\ 0 & \text{dla } y < \theta. \end{cases} \quad (1.3)$$

Funkcja gęstości rozkładu Pareto określona jest wzorem:

$$f(y) = \begin{cases} \frac{a\theta^a}{y^{a+1}} & \text{dla } y \geq \theta, \\ 0 & \text{dla } y < \theta. \end{cases} \quad (1.4)$$

Pareto (1897) na podstawie szeroko prowadzonych badań empirycznych zauważył, że parametr a rozkładu Pareto przyjmuje wartości w granicach od 1,24 do 1,89. Wiadomo, że dla $a < 2$ rozkład Pareto nie ma skończonej wariancji. Opierając się na podobnych obserwacjach empirycznych oraz na rozważaniach teoretycznych, Mandelbrot (1963) stwierdził, że rozkłady zmiennych wyrażonych w jednostkach pieniężnych powinny mieć nieskończone wariancje. Jednak badania empiryczne prowadzone przez innych autorów (Dagum, 1977; Fattorini, Lemmi, 1979) nie potwierdziły tej prawidłowości. Rozkłady płac i dochodów mają najczęściej skończone wariancje i momenty trzeciego rzędu, natomiast w przypadku rozkładów zamożności wariancje rzeczywiście dążą do nieskończoności. Ogólnie można stwierdzić, że rozkłady płac, dochodów i zamożności charakteryzują się niewielką liczbą skończonych momentów, ale nie muszą mieć nieskończonych wariancji. Oznacza to, że mają one „ciężkie ogony” (*heavy tails*) i im mniejsza liczba skończonych momentów rozkładu, tym grubszy jego prawy kraniec. Dobry model rozkładu dochodów powinien więc posiadać podobne cechy.

Dla rozkładu Pareto można wyznaczyć elastyczność dochodową dystrybuantry:

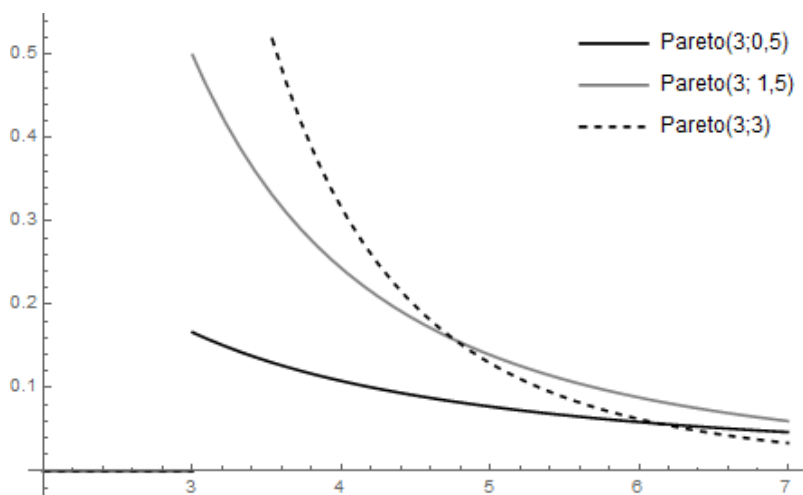
$$\varepsilon(y, F(y)) = \frac{F'(y)}{F(y)} y = \frac{f(y)}{F(y)} y = \frac{\frac{a\theta^a}{y^{a+1}}}{\frac{1}{F(y)}} y = \frac{a\left(\frac{y}{\theta}\right)^{-a}}{F(y)} = a \frac{1-F(y)}{F(y)}. \quad (1.5)$$

Jest ona malejącą i nieograniczoną z góry funkcją dystrybuantry $F(y)$. Natomiast elastyczność dochodowa dopełnienia do dystrybuantry ma postać:

$$\varepsilon(y, 1-F(y)) = \frac{(1-F(y))'}{1-F(y)} y = \frac{-f(y)}{1-F(y)} y = \frac{-\frac{a\theta^a}{y^{a+1}}}{\left(\frac{y}{\theta}\right)^{-a}} y = -a, \quad (1.6)$$

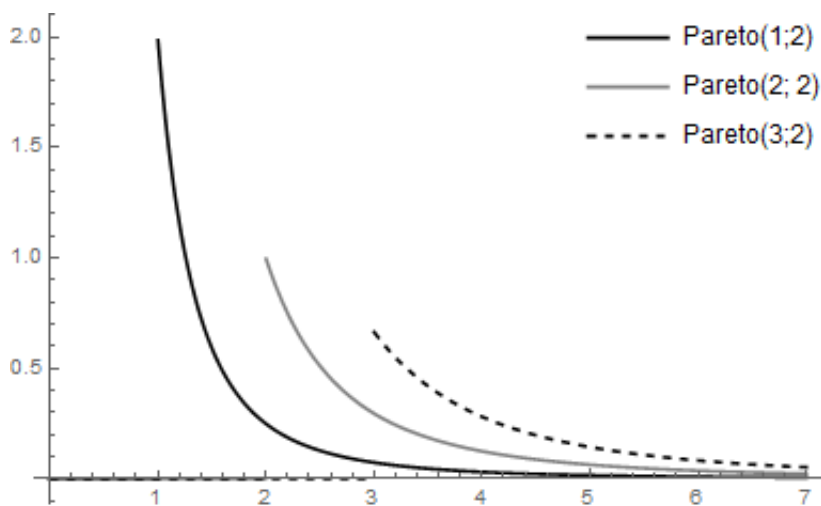
czyli jest funkcją stałą.

Przykładowe wykresy gęstości rozkładu Pareto, zwanego również rozkładem Pareto typu I, przedstawione są na rys. 1.1 (dla ustalonej wartości θ i różnych wartości a) oraz na rys. 1.2 (dla ustalonej wartości a i różnych wartości θ).



Rysunek 1.1. Gęstości rozkładów Pareto dla $\theta = 3$ i wybranych wartości a

Źródło: opracowanie własne



Rysunek 1.2. Gęstości rozkładów Pareto dla $a = 2$ i wybranych wartości θ

Źródło: opracowanie własne

Wartość oczekiwana zmiennej dochodowej o rozkładzie Pareto istnieje, gdy $a > 1$ i jest wyrażona wzorem:

$$E(Y) = \frac{a\theta}{a-1} \quad \text{dla } a > 1. \quad (1.7)$$

Z wzoru (1.7) wynika, że wartość oczekiwana wszystkich dochodów powyżej określonej wielkości θ , dla której spełnione jest prawo Pareto, jest wprost proporcjonalna do wartości minimalnej θ .

Uogólniając, moment zwykły rzędu r istnieje, gdy parametr skali spełnia nierówność $a > r$. Przyjmuje on postać:

$$E(Y^r) = \theta^r \frac{a}{a-r} \quad \text{dla } a > r. \quad (1.8)$$

Korzystając z własności $D^2(Y) = E(Y^2) - (EY)^2$, otrzymujemy odchylenie standardowe wyrażone wzorem:

$$D(Y) = \frac{\theta}{a-1} \sqrt{\frac{a}{a-2}} \quad \text{dla } a > 2. \quad (1.9)$$

Zatem rozkład dochodów określony krzywą Pareto ma skończoną wariancję tylko dla $a > 2$, zaś w innych przypadkach wariancja oraz wszystkie momenty wyższych rzędów dążą do nieskończoności i wtedy rozkład ten należy do klasy rozkładów stabilnych Pareto-Levy'ego (por. Mandelbrot, 1963).

Mediana rozkładu Pareto ma postać:

$$Me = \theta \sqrt[a]{2}, \quad (1.10)$$

zaś funkcja kwantylowa:

$$Q(p) = \frac{\theta}{\sqrt[a]{1-p}} \quad \text{dla } 0 \leq p < 1. \quad (1.11)$$

Dla rozkładów, które posiadają momenty centralne rzędu trzeciego i czwartego, można wyznaczyć współczynniki skośności (g_1) i spłaszczenia (g_2):

$$g_1 = \frac{E(Y - E(Y))^3}{D^3(Y)} = \frac{2(1+a)}{a-3} \sqrt{\frac{a-2}{a}} \quad \text{dla } a > 3, \quad (1.12)$$

$$g_2 = \frac{E(Y - E(Y))^4}{D^4(Y)} = \frac{6(a^3 + a^2 - 6a - 2)}{a(a-3)(a-4)} \quad \text{dla } a > 4. \quad (1.13)$$

Badania wielu autorów, m.in. Mandelbrota (1963), wykazały, że rozkład Pareto bardzo dobrze pasuje do danych empirycznych dla wysokich grup dochodowych. Dlatego też idealny teoretyczny rozkład dochodów dla wysokich grup dochodowych powinien dążyć do rozkładu Pareto pierwszego typu.

Wartość parametru a posiada swoją interpretację, jest związana z odsetkiem $p \cdot 100\%$ dochodów skupionych w grupie $q \cdot 100\%$ najbogatszych gospodarstw domowych.

Z warunku:

$$\frac{\sum_{y_i > y_{1-q}} y_i}{\sum_i y_i} = p, \quad (1.14)$$

gdzie y_{1-q} jest kwantylem rzędu $1 - q$ oraz funkcji kwantylowej rozkładu Pareto, otrzymujemy:

$$\frac{\int_{1-q}^1 \frac{\theta}{\sqrt[q]{1-x}} dx}{\int_0^1 \frac{\theta}{\sqrt[q]{1-x}} dx} = p. \quad (1.15)$$

Rozwiązując (1.15), mamy:

$$\begin{aligned} \int_{1-q}^1 (1-x)^{-\frac{1}{a}} dx &= p \int_0^1 (1-x)^{-\frac{1}{a}} dx \\ \left[-\frac{(1-x)^{1-\frac{1}{a}}}{1-\frac{1}{a}} \right]_{1-q}^1 &= \left[-p \frac{(1-x)^{1-\frac{1}{a}}}{1-\frac{1}{a}} \right]_0^1 \\ p &= q^{1-\frac{1}{a}}. \end{aligned} \quad (1.16)$$

Wyznaczając wartość parametru a , otrzymujemy:

$$\left(1 - \frac{1}{a}\right) \ln q = \ln p$$

$$\frac{a-1}{a} = \frac{\ln p}{\ln q},$$

czyli

$$a = \frac{\ln q}{\ln q - \ln p}. \quad (1.17)$$

W tab. 1.1 przedstawione są wartości parametru a dla wybranych wartości p i q . Jeśli wartość parametru a jest równa np. 1,07, oznacza to na przykład, że 20% najbogatszych gospodarstw domowych skupia 90% sumy dochodów wszystkich gospodarstw domowych. Gdy zaś $a = 2,322$, to 20% najbogatszych gospodarstw domowych skupia tylko 40% sumy dochodów.

Jeśli chcemy wyznaczyć odsetek sumy dochodów, jaki skupia grupa $100 \cdot q\%$ najbogatszych gospodarstw domowych, dla danej wartości parametru a , korzystamy z wzoru (1.16).

Tabela 1.1. Wartości parametru a rozkładu Pareto

q	p					
	0,40	0,50	0,60	0,70	0,80	0,90
0,10	1,661	1,431	1,285	1,183	1,107	1,048
0,15	1,934	1,576	1,368	1,232	1,133	1,059
0,20	2,322	1,756	1,465	1,285	1,161	1,070
0,25	2,950	2,000	1,583	1,346	1,192	1,082

Źródło: obliczenia własne

Zastosowanie rozkładu Pareto do analiz rozkładów dochodów wiąże się z określeniem krzywej koncentracji Lorenza oraz współczynników określających stopień nierównomierności rozkładu, takich jak współczynnik koncentracji Lorenza, indeks Giniego czy indeks Pietry, zwany również indeksem Schutza. Funkcję Lorenza wyznaczamy z wzoru:

$$L(p) = \mu^{-1} \int_0^p F^{-1}(t) dt \quad \text{dla } 0 \leq p \leq 1, \quad (1.18)$$

gdzie μ jest wartością oczekiwaną rozkładu.

Ma ona postać:

$$L(p) = 1 - (1 - p)^{1 - \frac{1}{a}} \quad \text{dla } a > 1. \quad (1.19)$$

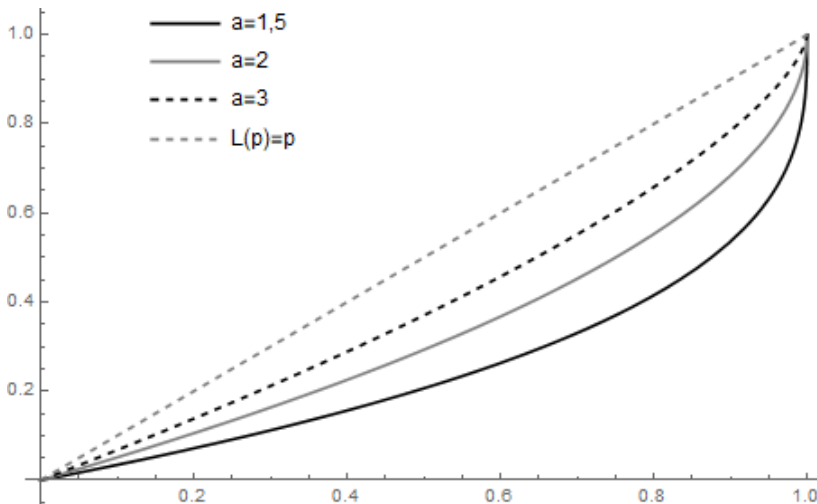
Wykresy krzywych Lorenza dla wybranych rozkładów przedstawione są na rys. 1.3.

Współczynnik Giniego wyznaczamy z wzoru:

$$G = 2 \int_0^1 (p - L(p)) dp \quad \text{dla } a > 1, \quad (1.20)$$

otrzymując

$$G = 2 \int_0^1 \left(p - 1 + (1 - p)^{1 - \frac{1}{a}} \right) dp = \frac{1}{2a - 1} \quad \text{dla } a > 1. \quad (1.21)$$



Rysunek 1.3. Funkcje Lorenza dla rozkładów Pareto o wybranych wartościach parametru a

Źródło: opracowanie własne

Indeks Pietry jest miarą nierówności równą maksymalnej odległości pionowej funkcji Lorenza od linii egalitarnej:

$$SP = \max_{0 \leq p \leq 1} (p - L(p)). \quad (1.22)$$

Dla rozkładu Pareto funkcja $f(p) = p - L(p)$ osiąga maksimum dla $p = 1 - \left(\frac{a-1}{a}\right)^a$, zatem indeks Pietry wyraża się wzorem:

$$SP = 1 - \left(\frac{a-1}{a}\right)^a, \quad \text{dla } a > 1. \quad (1.23)$$

Im większa jest wartość parametru a , tym wartości współczynników Giniego i Pietry są mniejsze, a tym samym mniejsze są nierównomierności rozkładu, co potwierdzają wykresy krzywych Lorenza przedstawione na rys. 1.3 i wartości współczynników w tab. 1.2 wyznaczone dla wybranych rozkładów Pareto.

Należy podkreślić, że rozkład Pareto, którego funkcja gęstości jest hiperbolą, nie nadaje się do całosciowej analizy rozkładów płac i dochodów, które są zwykle jednomodalne i dodatnio asymetryczne. Jednak ze względu na bardzo wysoką zgodność z rozkładami dochodów średnich i wysokich jest on traktowany jako wzorcowy dla prawej części rozkładu. Na uwagę zasługuje także jego prostota oraz możliwość wykorzystania do aproksymacji rozkładów zeromodalnych. Z tych względów rozkład Pareto był często wykorzystywany w literaturze do aproksymacji rozkładów zamożności. Próbowano także modyfikować go lub uogólnić tak, aby mógł być przydatny do badania rozkładów dochodów w całości.

Tabela 1.2. Wartości współczynników Giniego i Pietry dla wybranych rozkładów Pareto

Parametr a rozkładu Pareto	Współczynnik Giniego	Indeks Pietry
1,5	0,500	0,808
2,0	0,333	0,750
2,5	0,250	0,721
3,0	0,200	0,704
3,5	0,167	0,692
4,0	0,143	0,684

Źródło: obliczenia własne

Pareto (1895) zaproponował również dwa inne typy rozkładów.
Rozkład Pareto typu II $Pa(\theta, a, c)$ określony jest przez dystrybuantę postaci:

$$F(y) = \begin{cases} 1 - \left(1 + \frac{y-c}{\theta}\right)^{-a} & \text{dla } y \geq c, \\ 0 & \text{dla } y < c, \end{cases} \quad (1.24)$$

gdzie $c, \theta, a > 0$.

Dystrybuanta rozkładu Pareto typu III $Pa(\theta, c, \gamma)$ wyraża się wzorem:

$$F(y) = \begin{cases} 1 - \left(1 + \left(\frac{y-c}{\theta}\right)^{\frac{1}{\gamma}}\right)^{-1} & \text{dla } y \geq c, \\ 0 & \text{dla } y < c, \end{cases} \quad (1.25)$$

gdzie $c, \theta, \gamma > 0$.

W literaturze znaleźć można również czteroparametrowy rozkład Pareto $Pa(\theta, a, c, \gamma)$ (Pareto typu IV) stanowiący uogólnienie rozkładów Pareto I–III (dla $\gamma = 1, c = 0$ mamy rozkład Pareto I, dla $\gamma = 1$ – Pareto II, natomiast dla $a = 1$ – Pareto III). Dystrybuanta rozkładu $Pa(\theta, a, c, \gamma)$ ma postać:

$$F(y) = \begin{cases} 1 - \left(1 + \left(\frac{y-c}{\theta}\right)^{\frac{1}{\gamma}}\right)^{-a} & \text{dla } y \geq c, \\ 0 & \text{dla } y < c. \end{cases} \quad (1.26)$$

Zestawienie funkcji gęstości rozkładów Pareto przedstawione zostało w tab. 1.3, zaś ich podstawowe charakterystyki liczbowe w tab. 1.4.

Tabela 1.3. Funkcje gęstości rozkładów Pareto I–IV

Rozkład	Funkcja gęstości
Pareto I ($Pa(\theta, a)$)	$f(y) = \begin{cases} \frac{a\theta^a}{y^{a+1}} & \text{dla } y \geq \theta \\ 0 & \text{dla } y < \theta \end{cases}$

Rozkład	Funkcja gęstości
Pareto II ($Pa(\theta, a, c)$)	$f(y) = \begin{cases} \frac{a\theta^{a+2}}{(y+\theta-c)^{a+1}} & \text{dla } y \geq c \\ 0 & \text{dla } y < c \end{cases}$
Pareto III ($Pa(\theta, c, \gamma)$)	$f(y) = \begin{cases} \frac{(y-c)^{\frac{1}{\gamma}-1} \left[1 + \left(\frac{y-c}{\theta} \right)^{\frac{1}{\gamma}} \right]^{-2}}{\gamma^{\frac{1}{\gamma}} \sqrt{\theta}} & \text{dla } y \geq c \\ 0 & \text{dla } y < c \end{cases}$
Pareto IV ($Pa(\theta, a, c, \gamma)$)	$f(y) = \begin{cases} \frac{a(y-c)^{\frac{1}{\gamma}-1} \left[1 + \left(\frac{y-c}{\theta} \right)^{\frac{1}{\gamma}} \right]^{-a-1}}{\gamma^{\frac{1}{\gamma}} \sqrt{\theta}} & \text{dla } y \geq c \\ 0 & \text{dla } y < c \end{cases}$

Źródło: opracowanie własne

Tabela 1.4. Podstawowe charakterystyki rozkładów Pareto I–IV

Rozkład	$E(Y)$	$D(Y)$	Me
Pareto I	$\frac{a\theta}{a-1}$ dla $a > 1$	$\frac{\theta}{a-1} \sqrt{\frac{a}{a-2}}$ dla $a > 2$	$\theta^{\frac{1}{a}} \sqrt{2}$
Pareto II	$c + \frac{\theta}{a-1}$ dla $a > 1$	$\frac{\theta}{a-1} \sqrt{\frac{a}{a-2}}$ dla $a > 2$	$c + \theta(\sqrt[2]{2} - 1)$
Pareto III	$c + \theta\Gamma(1-\gamma)\Gamma(1+\gamma)$ dla $-1 < \gamma < 1$	$-\theta^2\Gamma^2(1-\gamma)\Gamma^2(1+\gamma) +$ $+\Gamma(1-2\gamma)\Gamma(1+2\gamma)$ dla $2\gamma < 1$	$c + \theta$

Tabela 1.4 (cd.)

Rozkład	$E(Y)$	$D(Y)$	Me
Pareto IV	$c + \frac{\theta \Gamma(a-\gamma) \Gamma(1+\gamma)}{\Gamma(a)}$ <p>dla $-1 < \gamma < a$</p>	$-\frac{\theta^2 \Gamma^2(a-\gamma) \Gamma^2(1+\gamma)}{\Gamma^2(a)} +$ $+ \frac{\Gamma(a) \Gamma(a-2\gamma) \Gamma(1+2\gamma)}{\Gamma^2(a)}$ <p>dla $a > 2\gamma$</p>	$c + \theta(\sqrt[4]{2}-1)^\gamma$

Źródło: opracowanie własne

Benini (1906), opierając się głównie na obserwacjach wysokich grup dochodowych, zauważył, że elastyczność dochodowa dopełnienia do dystrybucyjności rozkładu dochodów jest funkcją malejącą i wklęsłą, nie zaś stałą – jak twierdził Pareto. Zaproponował więc modyfikację modelu Pareto I typu. Dystrybucyjność rozkładu Beniniego przyjmuje postać:

$$F(y) = 1 - \left(\frac{y}{\theta} \right)^{-a \ln \left(\frac{y}{\theta} \right)} \quad \text{dla } y \geq \theta \text{ i } a, \theta > 0, \quad (1.27)$$

a jego funkcja gęstości:

$$f(y) = \frac{2a}{\theta} \left(\frac{y}{\theta} \right)^{-a \ln \left(\frac{y}{\theta} \right)} \ln \left(\frac{y}{\theta} \right) \quad \text{dla } y \geq \theta \text{ i } a, \theta > 0. \quad (1.28)$$

Champernowne (1952; 1953), również wykorzystując rozkład Pareto, zaproponował krzywą opisującą rozkłady dochodów. Zdefiniował zmienną $Z = \ln Y$ jako „potęgę dochodu” i otrzymał następującą funkcję gęstości rozkładu zmiennej Z :

$$f(z) = \frac{k}{\cosh[a(z - z_0)] + \lambda}, \quad (1.29)$$

gdzie: k, a, z_0, λ są parametrami rozkładu (przy czym k jest funkcją pozostałych parametrów).

Korzystając z zależności: $\cosh x = \frac{1}{2}(e^x + e^{-x})$, otrzymujemy:

$$f(z) = \frac{k}{0,5e^{a(z-z_0)} + 0,5e^{-a(z-z_0)} + \lambda}. \quad (1.30)$$

Dla $\lambda = 1$ z funkcji (1.30) można wyprowadzić dystrybuantę rozkładu Burra typu XII, a dodatkowo, gdy $a = 1$ oraz $z_0 = 0$ – rozkład logistyczny.

Jeśli zmienna losowa Z ma rozkład Champernowne'a, to zmienna $Y = e^Z$ określająca wielkość dochodów ma funkcję gęstości postaci:

$$f(y) = \frac{k}{y \left(0,5 \left(\frac{y}{\theta} \right)^{-a} + \lambda + 0,5a \left(\frac{y}{\theta} \right)^a \right)} \quad \text{dla } y > 0, \quad (1.31)$$

gdzie $\theta = e^{z_0}$.

Parametr a jest odpowiednikiem stałej Pareto dla wysokich grup dochodowych, zaś θ jest medianą dochodów.

Jeśli $\lambda = 1$, wzór (1.31) określa funkcję gęstości rozkładu Fiska, który został omówiony w paragrafie 1.7.

Rozkład Champernowne'a wykazuje wyższy stopień zgodności z empirycznymi rozkładami dochodów niż rozkłady Pareto czy logarytmiczno-normalny, jednakże rozbieżności pomiędzy rozkładami są nadal znaczne. Lepsze rezultaty uzyskał Fisk, rozpatrując szczególny przypadek rozkładu Champernowne'a i otrzymując rozkład sech^2 (secant hiperboliczny) postaci:

$$f(\phi) = \frac{e^\phi}{(1 + e^\phi)^2}, \quad (1.32)$$

gdzie $\phi = a \ln \frac{y}{\theta}$.

Dystrybuanta rozkładu określonego wzorem (1.32) daje krzywą logistyczną:

$$F(\phi) = \frac{e^\phi}{1 + e^\phi}. \quad (1.33)$$

Dla $y \rightarrow \infty$ rozkład sech^2 dąży do krzywej Pareto.

Funkcja (1.32) jest traktowana jako funkcja generująca rozkład dochodów, natomiast ϕ jako funkcja transformująca. Funkcja generująca jest zbliżona kształtem do gęstości rozkładu normalnego, ale krańcom rozkładu $sech^2$ odpowiadają większe częstości, czyli rozkład ma cięższe „ogony”, co może prowadzić (przy odpowiedniej transformacji) do lepszego dopasowania tego rozkładu do empirycznych rozkładów płac i dochodów.

Fisk, opierając się na rozkładach dochodów ogółem w USA i Rumunii, doszedł do wniosku, że funkcja transformująca ϕ powinna być funkcją podwójnie logarytmiczną.

Rozkład $sech^2$ bardziej nadaje się do badania rozkładów grup jednorodnych pod pewnym względem, np. stanowiska pracy. W przypadku takich zbiorowości funkcja transformująca może mieć postać funkcji liniowej lub logarytmicznej, co można stwierdzić, posługując się wykresem zależności $\phi(y) = \ln\left(\frac{F}{1-F}\right)$.

Stoppa (1990) przedstawił propozycje uogólnienia krzywej Pareto pierwszego i drugiego rodzaju, które w odróżnieniu od podstawowych form tych krzywych wykazują wysoką zgodność z danymi empirycznymi, także w wysokich grupach dochodowych. Określił zbiór cech, którymi charakteryzują się obserwowane rozkłady płac i dochodów. Zauważył, że rozkłady płac i dochodów mogą być zeromodalne albo jednomodalne i dodatnio asymetryczne. Ponadto elastyczność dochodowa dystrybucyjny rozkładu dochodów $\varepsilon(y, F(y))$ jest malejącą funkcją dystrybucyjną, która zależy od względnego przyrostu funkcji $1 - F(y)$ (por. też Dagum, 1990). Jeśli za $H(y)$ przyjmiemy funkcję $F^{1/\alpha}$, gdzie $\alpha \in (0, \infty)$, która może być interpretowana jako stopień rozwoju społeczno-ekonomicznego lub poziomu ogólnego dobrobytu, to wyżej wymienione cechy rozkładów płac i dochodów mogą być ujęte w formie następującego równania:

$$\varepsilon(y, F(y)) = \alpha \delta \frac{1 - H(y)}{H(y)}, \quad \alpha, \delta > 0. \quad (1.34)$$

Dla $F(y) \neq 0$ oraz $F(y) \neq 1$ otrzymujemy następujące równanie różniczkowe:

$$\frac{dF^{1/\alpha}}{1-F^{1/\alpha}} = \frac{\delta}{y} dy, \quad \alpha, \delta > 0. \quad (1.35)$$

Rozwiązaniem tego równania dla $F(y) \neq 1$ jest funkcja $F(y)$ określająca dystrybuantę rozkładu dochodów postaci:

$$F(y) = (1 - \theta^\delta y^{-\delta})^\alpha \quad \text{dla } y > \theta > 0, \quad (1.36)$$

gdzie:

θ^δ – antylogarytm stałej całkowania,

α, δ, θ – parametry rozkładu.

Dystrybuanta (1.36) określa rozkład dochodów spełniający warunki wymienione wstępnie przez Stoppe i jest uogólnieniem krzywej Pareto I-go typu. Parametr θ^δ można interpretować jako parametr skali, natomiast parametry δ i α są parametrami nierównomierności rozkładu. Jeśli $\alpha \in (0, 1]$, to funkcja gęstości rozkładu jest L -kształtna, natomiast gdy $\alpha > 1$ funkcja gęstości jest jednomodalna. Wraz ze wzrostem parametru α (dla $\alpha > 1$) funkcja gęstości ma cieńszy lewy, a grubszy prawy „ogon”. Pozwala to na lepsze dopasowanie do rozkładów empirycznych w porównaniu z krzywą Pareto, gdzie parametr ten nie występował ($\alpha = 1$, $\delta = a$). Jednocześnie wraz ze wzrostem α zmniejsza się stopień asymetrii rozkładu. Ponadto rozkład ten posiada nie tylko bezpośrednią analityczną postać dystrybuanty, lecz także funkcję odwrotną do dystrybuanty, co daje wzór na kwantyle Q_p rzędu p postaci:

$$Q_p = \theta(1 - p^{1/\alpha})^{-1/\delta}. \quad (1.37)$$

Funkcję gęstości odpowiadającą dystrybuancie (1.36) określa wzór:

$$f(y) = \alpha \delta \theta^\delta y^{-\delta-1} (1 - \theta^\delta y^{-\delta})^{\alpha-1} \quad \text{dla } y > \theta > 0 \quad (1.38)$$

Momenty zwykłe rzędu r przyjmują postać:

$$EY^r = \alpha \theta^r B\left(1 - \frac{r}{\delta}, \alpha\right) \text{ dla } r < \delta, \quad (1.39)$$

a funkcja Lorenza dla rozkładu Stoppy (1990) jest określona następująco:

$$L(p) = \frac{B\left(p^{1/\alpha}; \alpha, 1 - \frac{1}{\delta}\right)}{B\left(\alpha, 1 - \frac{1}{\delta}\right)} \text{ dla } \delta > 1, \quad (1.40)$$

gdzie $B(\cdot; \cdot, \cdot)$ jest niekompletną funkcją beta, zaś $B(\cdot, \cdot)$ jest (kompletną) funkcją beta.

Współczynnik koncentracji Giniego G odpowiadający funkcji Lorenza przyjmuje postać:

$$G = -1 + \frac{B\left(2\alpha, 1 - \frac{1}{\delta}\right)}{B\left(\alpha, 1 - \frac{1}{\delta}\right)} \text{ dla } \delta > 1. \quad (1.41)$$

Współczynnik G jest funkcją parametrów nierównomierności rozkładu: α oraz δ i jego wartość rośnie wraz ze wzrostem parametru α , natomiast maleje wraz ze wzrostem parametru δ .

1.3. Rozkład logarytmiczno-normalny

Ogólną teorię rozkładu logarytmiczno-normalnego sformułował Mc Alister w końcu XIX wieku. Rozkład ten został po raz pierwszy zastosowany do badania płac i dochodów przez francuskiego ekonomistę i statystyka Gibrata (1931), który jednocześnie sformułował prawo proporcjonalnego efektu, dające podwaliny stochastyczne omawianego rozkładu. W Polsce rozkład logarytmiczno-normalny po raz pierwszy zastosował Wiśniewski (1934). Z jego badań wynikało, że rozkład ten może być wykorzystywany do badania płac pracowników biurowych. Pawłowski (1960)

i Vielrose (1960) wykazali przydatność tego rozkładu do aproksymacji płac wszystkich pracowników w gospodarce społecznej. Z kolei Kordos i Stroińska (1971), badając płace w Polsce w latach 1955–1970, wykazali, że rozkład logarytmiczno-normalny dobrze opisuje rozkłady płac w różnych przekrojach (gałęziowym, działowym), a także w podziale na pracowników fizycznych i umysłowych. Przydatność rozkładu logarytmiczno-normalnego do badania rozkładów płac w Polsce była także tematem prac m.in. Wąsika (1967), Domańskiego (1973) i Kordosa (1968; 1973). Własności rozkładu logarytmiczno-normalnego zostały szeroko omówione w monografii Aitchisona i Browna (1957).

Zmienna losowa Y ma rozkład logarytmiczno-normalny $LN(\mu, \sigma)$, jeśli logarytmy tej zmiennej układają się według krzywej normalnej. Funkcję gęstości takiej zmiennej można wyrazić za pomocą następującego wzoru:

$$f(y) = \frac{1}{y\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln y - \mu)^2}{2\sigma^2}\right) \text{ dla } y \in (0, +\infty), \quad (1.42)$$

gdzie: $\mu \in R$ – wartość oczekiwana logarytmów zmiennej losowej Y ,
 $\sigma > 0$ – odchylenie standardowe logarytmów zmiennej losowej Y .

Dystrybuanty zmiennej losowej o rozkładzie logarytmiczno-normalnym nie da się wyrazić wzorem, można ją tylko zapisać jako:

$$F(y) = \frac{1}{2} + \frac{1}{2} \operatorname{erf}\left(\frac{\ln y - \mu}{\sigma\sqrt{2}}\right), \text{ gdzie } \operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt.$$

Przykładowe wykresy funkcji gęstości dwuparametrowego rozkładu logarytmiczno-normalnego przedstawione są na rys. 1.4–1.5.

Przy wykorzystaniu parametrów μ oraz σ można wyrazić wszystkie pozostałe charakterystyki rozkładu.

Rozkład logarytmiczno-normalny posiada momenty zwykłe i centralne wszystkich rzędów. Momenty zwykłe rzędu r określa wzór:

$$E(Y^r) = e^{r\mu + 0,5r^2\sigma^2}, \quad (1.43)$$

czyli w szczególności wartość oczekiwana zmiennej losowej Y ma postać:

$$E(Y) = e^{\mu + 0,5\sigma^2}, \quad (1.44)$$

zaś odchylenie standardowe:

$$D(Y) = \sqrt{(e^{\sigma^2} - 1)e^{2\mu + \sigma^2}} = E(Y)\sqrt{e^{\sigma^2} - 1}. \quad (1.45)$$

Trzeci i czwarty moment centralny są równe, odpowiednio:

$$E(Y - E(Y))^3 = (E(Y))^3(V^6 + 3V^4), \quad (1.46)$$

$$E(Y - E(Y))^4 = (E(Y))^4(V^{12} + 6V^{10} + 15V^8 + 16V^6 + 3V^4), \quad (1.47)$$

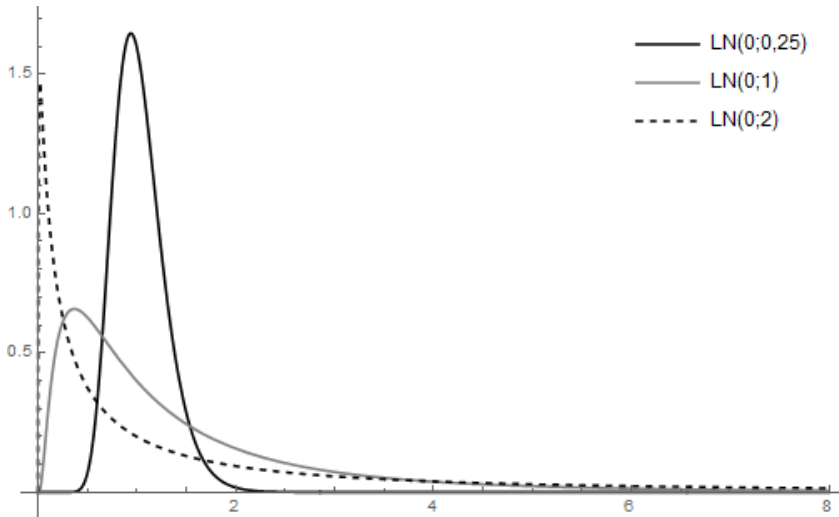
gdzie $V = \sqrt{e^{\sigma^2} - 1}$.

Stąd możemy otrzymać wzory na współczynniki: skośności (g_1) i spłaszczenia (g_2):

$$g_1 = V^3 + 3V, \quad (1.48)$$

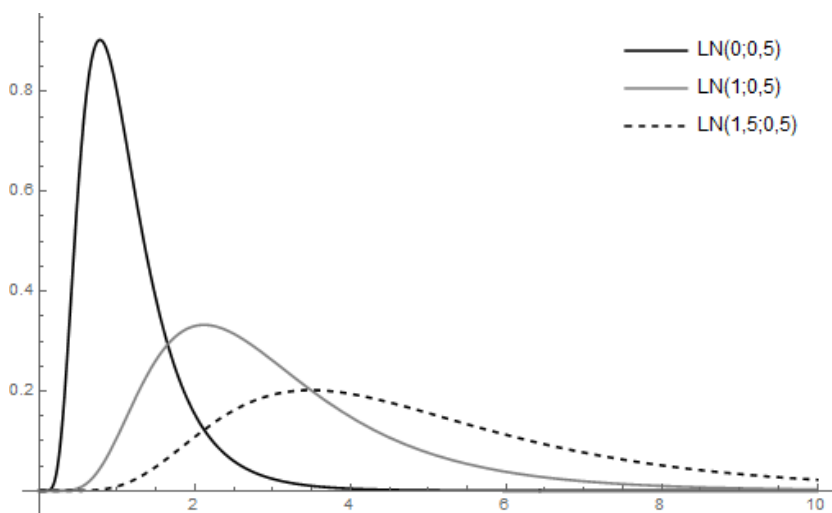
$$g_2 = V^8 + 6V^6 + 15V^4 + 16V^2 + 3. \quad (1.49)$$

Łatwo zauważyć, że obydwa współczynniki przyjmują tylko wartości dodatnie i wzrastają wraz ze wzrostem wariancji.



Rysunek 1.4. Gęstości rozkładów logarytmiczno-normalnych dla $\mu = 0$ i wybranych wartości σ

Źródło: opracowanie własne



Rysunek 1.5. Gęstości rozkładów logarytmiczno-normalnych dla $\sigma = 0,5$ i wybranych wartości μ

Źródło: opracowanie własne

Mediana i dominanta rozkładu logarytmiczno-normalnego przyjmują, odpowiednio, postaci:

$$Me = e^{\mu}, \quad (1.50)$$

$$Do = e^{\mu - \sigma^2}. \quad (1.51)$$

Kwantyle rozkładu logarytmiczno-normalnego można wyrazić za pomocą kwantyli rozkładu normalnego standaryzowanego. Kwantyl rzędu p rozkładu logarytmiczno-normalnego określa wzór:

$$Q_p = e^{\mu + u_p \sigma^2}, \quad (1.52)$$

gdzie u_p jest kwantylem rzędu p rozkładu $N(0,1)$.

Istnieje wiele charakterystyk rozkładu logarytmiczno-normalnego zależnych jedynie od parametru σ . Stwarza to możliwość różnorodnych analiz, w tym badania nierówności rozkładów w badaniach dochodów gospodarstw domowych. Obok już wymienionych współczynników zmienności, skośności i spłaszczenia należą do nich:

- stosunek średniej arytmetycznej do mediany:

$$\frac{E(Y)}{Me} = e^{0,5\sigma^2}, \quad (1.53)$$

- stosunek średniej arytmetycznej do dominanty:

$$\frac{E(Y)}{Do} = e^{1,5\sigma^2}, \quad (1.54)$$

- frakcja obserwacji o wartościach nie większych od średniej arytmetycznej:

$$W = \Phi\left(\frac{\sigma}{2}\right), \quad (1.55)$$

- współczynnik Giniego:

$$G = 2\Phi\left(\frac{\sigma}{\sqrt{2}}\right) - 1, \quad (1.56)$$

- indeks Pietry:

$$SP = 2\Phi\left(\frac{\sigma^2}{2}\right) - 1, \quad (1.57)$$

gdzie: $\Phi(\cdot)$ jest dystrybuantą rozkładu normalnego standaryzowanego.

Funkcja Lorenza ma postać:

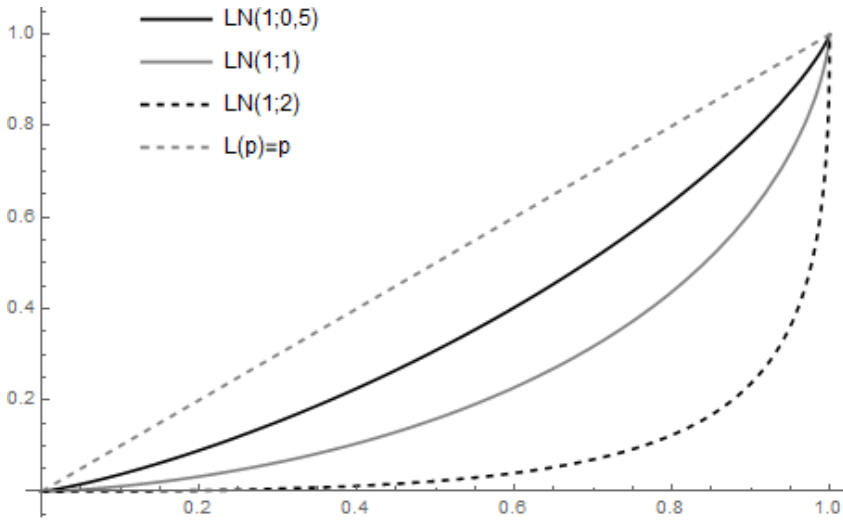
$$L(p) = \Phi(\Phi^{-1}(p) - \sigma). \quad (1.58)$$

Wartości współczynników Giniego i Pietry, a tym samym nierównomierności rozkładu, rosną wraz ze wzrostem parametru σ . Przykładowe wartości tych współczynników zawarte są w tab. 1.5, a wykresy funkcji Lorenza na rys. 1.6.

Tabela 1.5. Wartości współczynników Giniego i Pietry dla wybranych rozkładów logarytmiczno-normalnych

Parametr σ rozkładu logarytmiczno-normalnego	Współczynnik Giniego	Indeks Pietry
0,5	0,276	0,099
1,0	0,520	0,386
1,5	0,711	0,739
2,0	0,843	0,955
2,5	0,923	0,998

Źródło: obliczenia własne



Rysunek 1.6. Funkcje Lorena dla rozkładów logarytmiczno-normalnych o wybranych wartościach parametru σ

Źródło: opracowanie własne

W badaniach rozważa się także inne warianty rozkładu logarytmiczno-normalnego: trój- i czteroparametrowe. Rozkład trójparametrowy $LN(\mu, \sigma, \gamma)$ określony jest za pomocą funkcji gęstości:

$$f(y) = \frac{1}{(y - \gamma)\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln(y - \gamma) - \mu)^2}{2\sigma^2}\right), \quad (1.59)$$

gdzie: $y > \gamma \geq 0, \sigma > 0, \mu \in R$.

Oczywiście rozkład dwuparametrowy jest szczególnym przypadkiem trójparametrowego rozkładu logarytmiczno-normalnego ($\gamma = 0$).

Funkcja gęstości rozkładu czteroparametrowego $LN(\mu, \sigma, \gamma, \lambda)$ ma postać:

$$f(y) = \frac{1}{(y - \gamma)\sigma\sqrt{2\pi}} \exp\left(-\frac{\left(\ln\left(\frac{y - \gamma}{\lambda}\right) - \mu\right)^2}{2\sigma^2}\right), \quad (1.60)$$

gdzie: $0 \leq \gamma < y < \gamma + \lambda, \sigma > 0, \mu \in R$.

Rozkład logarytmiczno-normalny ma szereg zalet, jeśli chodzi o możliwość wykorzystania go w analizach rozkładów płac i dochodów. Jedną z nich jest niewątpliwie jego prostota oraz możliwość jasnej interpretacji ekonomicznej parametrów (parametr μ można także interpretować jako logarytm mediany rozkładu dochodów). Dużym ułatwieniem mogą być również bezpośrednio i proste formuły matematyczne, za pomocą których można określić większość charakterystyk tego rozkładu, takich jak np. kwantyle, momenty, współczynnik Giniego. Trzeba jednak zaznaczyć, że dla rozkładu logarytmiczno-normalnego niemożliwe jest otrzymanie prostej formuły określającej krzywą koncentracji Lorenza; nieznana jest też postać dystrybucyjny rozkładu, co utrudnia niektóre procedury estymacji parametrów oraz komplikuje proces generowania liczb losowych. Rozkład logarytmiczno-normalny może być stosowany dla zmiennej dochodowej, której wartość najmniejsza jest większa od zera (rozkład trój- i czteroparametrowy), jednak nie można go wykorzystać do opisywania rozkładów zeromodalnych ani też rozkładów o dochodach zerowych i ujemnych. Najistotniejszą wadą tego rozkładu jest jednak fakt, że nie spełnia on warunków zbieżności do prawa Pareto dla wysokich grup dochodowych, a co za tym idzie posiada skończone momenty wszystkich rzędów. Konsekwencją tego jest gorsza zgodność rozkładu logarytmiczno-normalnego z rozkładami empirycznymi na prawym krańcu rozkładu. Analizy rozkładów płac w Polsce prowadzone do 1989 roku wykazywały, że rozkład ten charakteryzuje się wysokim stopniem zgodności z danymi empirycznymi. Jednak późniejsze badania wykazały, że wraz z odejściem od gospodarki centralnie planowanej rozkład logarytmiczno-normalny gorzej aproksymuje rozkłady płac w Polsce (por. Jędrzejczak, 1990; 1993; 2006), co potwierdzają wyniki badań prowadzonych w krajach o wysoko rozwiniętych gospodarkach rynkowych (por. np. Singh, Maddala, 1976; Kloek, van Dijk, 1977).

Podejmowane były także próby modyfikacji rozkładu logarytmiczno-normalnego w celu poprawienia jego zgodności z danymi empirycznymi. Metcalf (1969) zaproponował wykorzystanie przesuniętego rozkładu logarytmiczno-normalnego, przyjmując, że zmienna postaci $\ln(y + c)$ ma rozkład normalny. Tak zmodyfikowany rozkład logarytmiczno-normalny lepiej odzwierciedla dodatnią asymetrię rozkładów płac i dochodów. Nie poprawia to jednak dopasowania rozkładu w skrajnych grupach dochodowych. Steyn (1966) zaproponował wykorzystanie kombinacji dwóch rozkładów logarytmiczno-normalnych, co dało lepsze

dopasowanie niż rozkład pojedynczy, natomiast Kloek i van Dijk (1977) zaproponowali wykorzystanie logarytmicznej transformacji rozkładu t -Studenta ($\log t$) zamiast transformacji rozkładu normalnego. Poprawia to dopasowanie krańców rozkładu do danych empirycznych, pozostawiając prostą interpretację ekonomiczną parametrów.

Funkcja gęstości rozkładu $\log t$ ma następującą postać:

$$f(y) = \frac{\frac{1}{v^2}}{B\left(\frac{1}{2}, \frac{1}{2}v\right)\tau} y \left[v + \frac{(\ln y - \ln \mu)^2}{\tau^2} \right]^{-\frac{1}{2}(v+1)} \quad \text{dla } y > 0, \quad (1.61)$$

gdzie: $\mu, \tau, v > 0$ – parametry rozkładu.

Parametr μ można zinterpretować tak samo jak e^μ rozkładu logarytmiczno-normalnego. Jest to mediana dochodów. Parametry τ i v są parametrami nierównomierności, przy czym v jest liczbą stopni swobody. Wariancję logarytmów dochodu dla rozkładu $\log t$ wyraża wzór:

$$\sigma^2 = \frac{v\tau^2}{v-2} \quad \text{dla } v > 2, \quad (1.62)$$

natomiast współczynnik spłaszczenia przyjmuje postać:

$$g_2 = \frac{6}{v-4} \quad \text{dla } v > 4. \quad (1.63)$$

Parametr v oznacza liczbę stopni swobody rozkładu i charakteryzuje jego „ogony”. Im większa wartość v , tym cieńszy „ogon” rozkładu. Dla $v \rightarrow \infty$ rozkład $\log t$ dąży do rozkładu logarytmiczno-normalnego.

1.4. Rozkład gamma

Po raz pierwszy rozkład gamma został zastosowany do analizy dochodów przez Marcha (1898), a później przez Amoroso (1925). Jego zgodność z danymi empirycznymi jest w niektórych przypadkach lepsza niż logarytmiczno-normalnego, a poza tym parametry tego rozkładu posiadają interpretację ekonomiczną. Kolejną zaletą jest fakt, że oba parametry

tego rozkładu można otrzymać w sposób analityczny przy wykorzystaniu metody największej wiarygodności.

Rozkład gamma $G(\alpha, \lambda)$ należy do rodziny krzywych Pearsona III typu i może być określony wzorem (Salem, Mount, 1974):

$$f(y) = \frac{\lambda^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\lambda y} \quad \text{dla } y > 0 \quad (1.64)$$

gdzie:

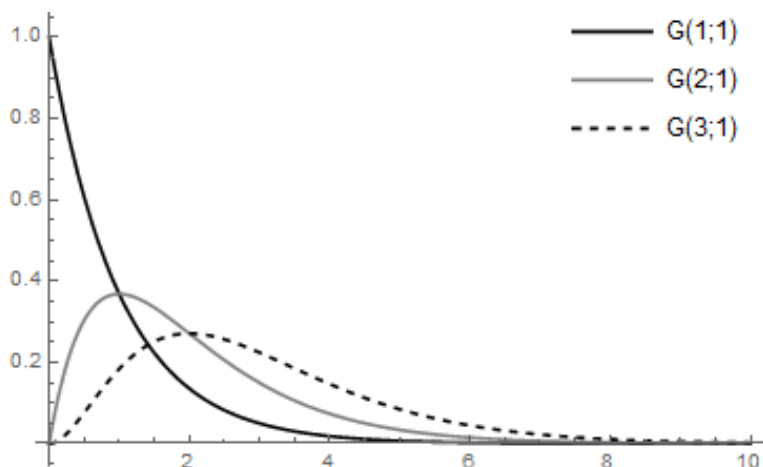
$\lambda, \alpha > 0$ – parametry rozkładu,

$\Gamma(\cdot)$ – funkcja gamma.

Parametr λ może być interpretowany jako parametr skali. Jeśli wszystkie wartości dochodu Y pomnożymy przez stałą c i otrzymamy $X = c \cdot Y$, to funkcja gęstości zmiennej X będzie równoważna funkcji gęstości zmiennej Y określonej wzorem (1.64), gdy za λ postawimy $\frac{\lambda}{c}$.

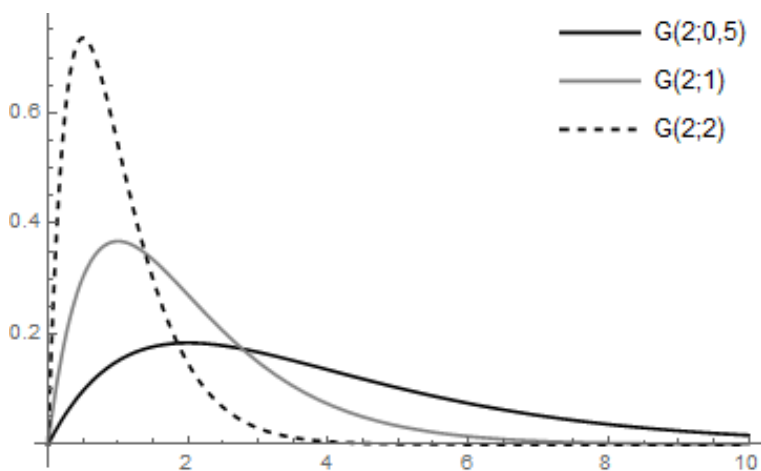
Parametr α jest natomiast parametrem nierównomierności – współczynnik skośności i współczynnik koncentracji Giniego są funkcjami tylko tego parametru.

Przykładowe wykresy funkcji gęstości rozkładów gamma dla ustalonej wartości jednego z parametrów i różnych wartości parametru drugiego przedstawione są na rys. 1.7–1.8.



Rysunek 1.7. Gęstości rozkładów gamma dla $\lambda = 1$ i wybranych wartości α

Źródło: opracowanie własne



Rysunek 1.8. Gęstości rozkładów gamma dla $\alpha = 2$ i wybranych wartości λ

Źródło: opracowanie własne

Główne charakterystyki rozkładu gamma określają wzory:

- wartość oczekiwana:

$$E(Y) = \frac{\alpha}{\lambda}, \quad (1.65)$$

- odchylenie standardowe:

$$D(Y) = \frac{\sqrt{\alpha}}{\lambda}, \quad (1.66)$$

- dominanta:

$$D = \frac{\alpha - 1}{\lambda}, \quad (1.67)$$

- współczynnik skośności:

$$g_1 = \frac{E(Y - E(Y))^3}{D^3(Y)} = \frac{2}{\sqrt{\alpha}}, \quad (1.68)$$

- kurtoza:

$$g_2 = \frac{E(Y - E(Y))^4}{D^4(Y)} = \frac{6}{\alpha}. \quad (1.69)$$

Współczynnik Giniego określony jest wzorem:

$$G = \frac{\Gamma(\alpha + 0,5)}{\sqrt{\pi}\Gamma(\alpha + 1)}, \quad (1.70)$$

gdzie $\Gamma(\cdot)$ jest funkcją gamma.

Koncentracja rozkładu (współczynnik Giniego) zależy od parametru α rozkładu i im większa jest jego wartość, tym mniejsze są nierównomierności rozkładu (por. tab. 1.6).

Tabela 1.6. Wartości współczynnika Giniego dla wybranych rozkładów gamma

Parametr α rozkładu gamma	Współczynnik Giniego
0,5	0,564
1,0	0,500
1,5	0,424
2,0	0,375
2,5	0,340
3,0	0,313

Źródło: obliczenia własne

Amoroso (1925) we wspomnianej już pracy przedstawił uogólniony rozkład gamma o funkcji gęstości postaci:

$$f(y) = \frac{\lambda^p}{\Gamma(p)} (y - \theta)^{p-1} \exp(-\lambda(y - \theta)) \quad \text{dla } 0 < \theta \leq y < \infty, \quad (1.71)$$

gdzie: $\lambda > 0$, $s \neq 0$, $\theta, p \in R$ – parametry rozkładu.

Jeżeli podstawimy $\theta = 0$, $p = \alpha$ oraz $s = 1$, to otrzymamy dwuparametrowy rozkład gamma $G(\alpha, \lambda)$ (por. wzór (1.64)). Podstawiając $\theta = 0$ oraz $s = -1$, otrzymujemy krzywą Pearsona V typu (Vinci, 1921), natomiast dla $p = 1$ oraz $\frac{1}{s} = \alpha$ i $\lambda = k^\alpha$ krzywa określona wzorem (1.71) jest funkcją gęstości rozkładu Weibulla $W(\alpha, k)$.

Rozkład gamma był także rozważany przez Stacy'ego (1962). Uogólnienie polegało na wprowadzeniu parametru p jako potęgi w czynniku wykładniczym rozkładu gamma. Funkcja gęstości przyjmuje wtedy postać:

$$f(y) = \frac{p}{\alpha^d} y^{d-1} \frac{\exp\left(-\frac{y}{\alpha}\right)^p}{\Gamma\left(\frac{d}{p}\right)}, \quad (1.72)$$

gdzie: $\alpha, d, p > 0$ są parametrami rozkładu.

Funkcja zaproponowana przez Stacy'ego może być również traktowana jako szczególny przypadek funkcji określonej przez Amoroso. Uogólnione formy rozkładu gamma poprawiają co prawda dopasowanie tego rozkładu do danych empirycznych, jednak ze względu na bardziej skomplikowaną postać parametry ich są trudniejsze do oszacowania i interpretacji.

1.5. Rozkład Daguma

Rozkład Daguma powstał na podstawie pogłębionych obserwacji i analiz empirycznych rozkładów płac i dochodów. Analizy te wykazały, że elastyczność dochodową dystrybuan-ty można przedstawić w postaci następującego równania różniczkowego (Dagum, 1977; Dagum, Lemmi, 1989):

$$\varepsilon(y, F(y)) = \frac{d \ln F(y)}{d \ln y} = \beta_1 (1 - (F(y))^{\beta_2}) \quad \text{dla } y \geq 0, \beta_1, \beta_2 > 0. \quad (1.73)$$

Z równania (1.73) wynika, że elastyczność dochodowa dystrybuan-ty rozkładu dochodów jest funkcją malejącą, na ogół wklęsłą oraz ograniczoną ($\lim_{F(y) \rightarrow 0^+} \varepsilon(y, F(y)) = \beta_1$), nie zaś nieograniczoną – jak twierdził Pareto (1897).

Rozwiązując równanie różniczkowe (1.73), Dagum otrzymał dystrybuan-ty rozkładu dochodów:

$$F(y) = \begin{cases} (1 + \lambda y^{-\nu})^{-a} & \text{dla } y > 0, \\ 0 & \text{dla } y \leq 0, \end{cases} \quad (1.74)$$

gdzie: $\lambda = e^c > 0$ jest parametrem skali (c jest stałą całkowania wynikającą z rozwiązania równania (1.73)), $\nu = \beta_1 \beta_2 > 0$ i $a = \frac{1}{\beta_2} > 0$ są parametrami kształtu.

Funkcja gęstości rozkładu Daguma odpowiadająca dystrybuancie (1.74) określona jest wzorem:

$$f(y) = \begin{cases} avy\lambda^{-v-1}(1+\lambda y^{-v})^{-a-1} & \text{dla } y > 0 \\ 0 & \text{dla } y \leq 0. \end{cases} \quad (1.75)$$

Często stosowana jest parametryzacja oparta na pracy Kleibera i Kotza (2003), którą otrzymujemy po podstawieniu $\lambda = \left(\frac{1}{b}\right)^{-v}$.

Dystrybuanta rozkładu Daguma jest wtedy postaci:

$$F(y) = \begin{cases} \left(1 + \left(\frac{y}{b}\right)^{-v}\right)^{-a} & \text{dla } y > 0, \\ 0 & \text{dla } y \leq 0, \end{cases} \quad (1.76)$$

natomiast funkcja gęstości:

$$f(y) = \begin{cases} \frac{av}{b} \left(\frac{y}{b}\right)^{av-1} \left(1 + \left(\frac{y}{b}\right)^v\right)^{-a-1} & \text{dla } y > 0, \\ 0 & \text{dla } y \leq 0, \end{cases} \quad (1.77)$$

gdzie: $v > 0$, $a > 0$ i $b > 0$.

Funkcje gęstości rozkładu Daguma $D(a, v, b)$ dla wybranych parametrów przedstawione są na rys. 1.9–1.11.

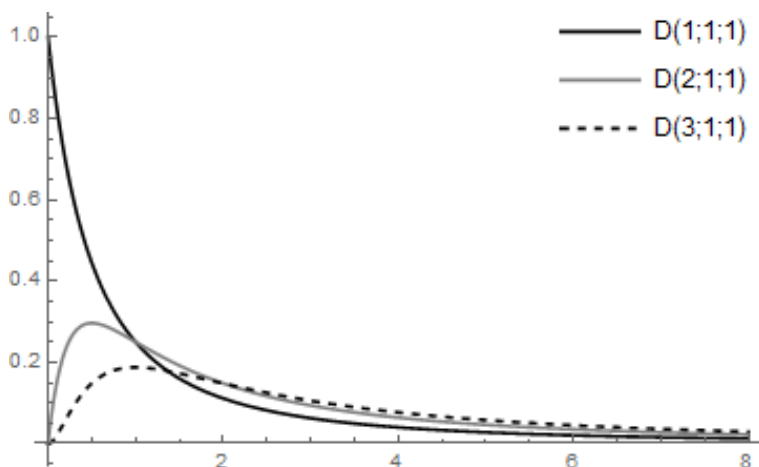
Momenty zwykłe rzędu r zmiennej losowej Y o rozkładzie Daguma $D(a, v, b)$ określone są wzorem:

$$E(Y^r) = b^r \frac{\Gamma\left(1 - \frac{r}{v}\right) \Gamma\left(a + \frac{r}{v}\right)}{\Gamma(a)} \quad \text{dla } r < v, \quad (1.78)$$

gdzie Γ jest funkcją gamma.

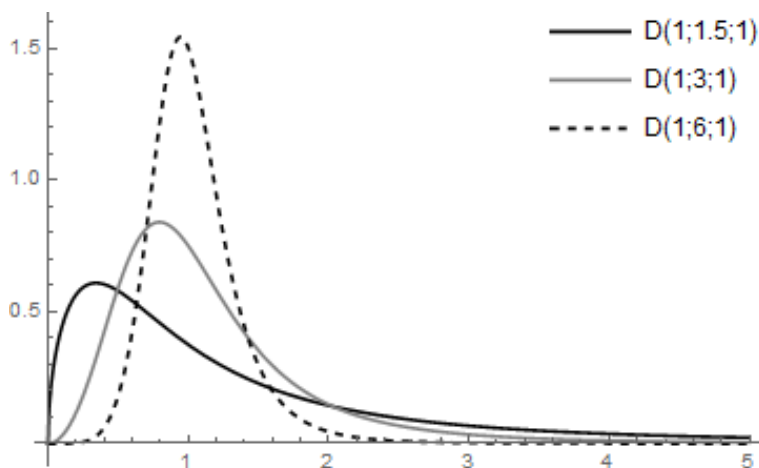
Jak wynika z wzoru (1.78), momenty rzędu r są skończone tylko dla $r < v$. Ponieważ oszacowania parametru v wahają się w granicach 2–4, to rozkład Daguma charakteryzuje się niewielką liczbą skończonych momentów, ma jednak zwykle skończoną wariancję. Jest to konsekwencją

faktu, że funkcja Daguma dąży do funkcji Pareto dla wysokich grup dochodowych, czyli warunek $\lim_{y \rightarrow \infty} (1 - F(y)) / \left(\frac{y}{\theta}\right)^{-a} = 1$ jest spełniony (por. Mandelbrot, 1963).



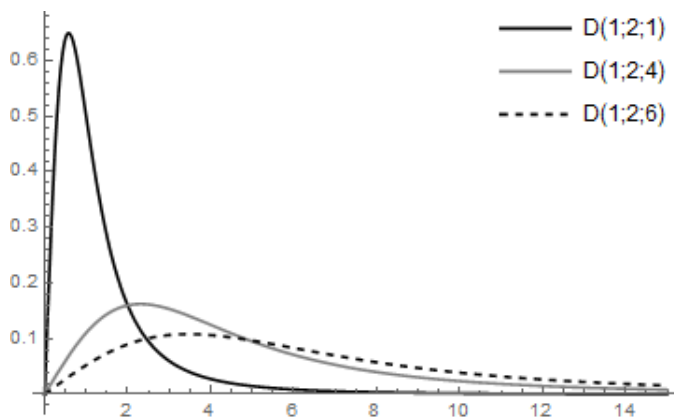
Rysunek 1.9. Gęstości rozkładów Daguma dla wybranych wartości a i ustalonych pozostałych parametrów

Źródło: opracowanie własne



Rysunek 1.10. Gęstości rozkładów Daguma dla wybranych wartości v i ustalonych pozostałych parametrów

Źródło: opracowanie własne



Rysunek 1.11. Gęstości rozkładów Daguma dla wybranych wartości b i ustalonych pozostałych parametrów

Źródło: opracowanie własne

Z wzoru (1.78) otrzymujemy postać wartości oczekiwanej:

$$E(Y) = b \frac{\Gamma\left(1 - \frac{1}{\nu}\right) \Gamma\left(a + \frac{1}{\nu}\right)}{\Gamma(a)} \quad \text{dla } \nu > 1. \quad (1.79)$$

Wariancja rozkładu Daguma wyraża się wzorem:

$$D^2(Y) = b^2 \frac{\Gamma\left(1 - \frac{2}{\nu}\right) \Gamma(a) \Gamma\left(a + \frac{2}{\nu}\right) - \Gamma^2\left(1 - \frac{1}{\nu}\right) \Gamma^2\left(a + \frac{1}{\nu}\right)}{\Gamma^2(a)} \quad \text{dla } \nu > 2. \quad (1.80)$$

Funkcja kwantylowa rozkładu Daguma jest postaci:

$$Q_D(p) = b \left(p^{-\frac{1}{a}} - 1 \right)^{-\frac{1}{\nu}} \quad \text{dla } 0 < p < 1, \quad (1.81)$$

w szczególności mediana rozkładu Daguma ma postać:

$$Me = b \left(2^{\frac{1}{a}} - 1 \right)^{-\frac{1}{\nu}}. \quad (1.82)$$

Funkcja Lorenza odpowiadająca dystrybucji określonej wzorem (1.76) może być przedstawiona w następującej postaci:

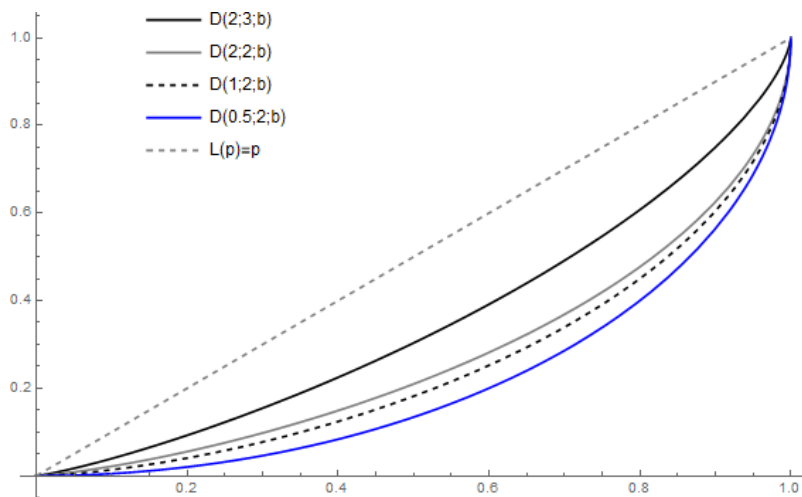
$$L(p) = B\left(p^{\frac{1}{a}}; a + \frac{1}{v}, 1 - \frac{1}{v}\right) \quad \text{dla } v > 1, \quad (1.83)$$

gdzie:

$B(\cdot; \cdot, \cdot)$ – niekompletna funkcja beta,

a, v – parametry rozkładu.

Forma parametryczna funkcji Lorenza dla rozkładu Daguma jest więc dystrybuantą rozkładu beta względem transformacji zmiennej losowej Y równej $(F(Y))^{\frac{1}{a}}$. Parametrami są tu $a + \frac{1}{v}$ oraz $1 - \frac{1}{v}$.



Rysunek 1.12. Funkcje Lorenza dla wybranych rozkładów Daguma

Źródło: obliczenia własne

Współczynnik koncentracji Giniego odpowiadający funkcji Lorenza określonej wzorem (1.83) przyjmuje postać:

$$G = 1 - 2 \int_0^1 L(p) dp = -1 + \frac{B(a, a)}{B\left(a, a + \frac{1}{v}\right)} \quad \text{dla } v > 1. \quad (1.84)$$

Jak wynika z wzorów (1.83) i (1.84), parametry a i v są parametrami nierównomierności rozkładu Daguma, określają one bowiem postać krzywej koncentracji Lorenza oraz współczynnika Giniego. Można wykazać, że im większa wartość parametru v , a mniejsza parametru a , tym mniejszy jest stopień koncentracji dochodów w badanej zbiorowości. Wartości współczynnika Giniego dla wybranych parametrów a i v rozkładu Daguma zawarte są w tab. 1.7.

Parametr b jest parametrem skali, będącym funkcją jednostki pieniężnej, w której mierzone są dochody. Dlatego też wartość oczekiwana, wariancja oraz wszystkie inne momenty rozkładu zależą przede wszystkim od wartości tego parametru. Wykorzystując model (1.76) do analizy porównawczej rozkładów w różnych okresach czasu lub też w różnych krajach, wystarczy zmienić jedynie wartość parametru b , aby odpowiadała ona jednostce pieniężnej, którą chcemy użyć do porównań. Oszacowania parametrów a i v są niezmiennie, jeśli zmienia się tylko skala, w której wyrażona jest dana dochodowa.

Parametry a i v , będące parametrami określającymi kształt rozkładu, mają także wpływ na to, czy funkcja określona równaniem (1.77) osiąga w pewnym punkcie maksimum. Aby tak było, iloczyn tych parametrów musi być większy od jedności, tzn. $av > 1$. Jeśli zaś $0 < av < 1$, to rozkład pozostaje zeromodalny. Z rozważań tych wynika, że rozkład Daguma może być wykorzystywany zarówno do analizy rozkładów jednomodalnych, jak i zeromodalnych, co bardzo rozszerza zakres jego zastosowań.

Tabela 1.7. Wartości współczynnika Giniego dla wybranych rozkładów Daguma

Parametr a	Parametr v	Współczynnik Giniego
2,0	3,0	0,296
1,0	3,0	0,333
0,5	3,0	0,402
2,0	2,0	0,458
1,0	2,0	0,500
0,5	2,0	0,571

Źródło: obliczenia własne

Obok rozkładu trójparametrowego wyrażonego wzorem (1.77), który określa się jako rozkład Daguma typu I, Dagum (1977) zaproponował także rozkłady czteroparametrowe, uwzględniające przypadki, kiedy w zbiorowości występują dochody zerowe i ujemne (rozkład typu II) lub tylko dochody większe od zera (rozkład typu III). Warto zauważyć, że rozkład typu II może być stosowany nie tylko do rozkładów, w których występują dochody zerowe i ujemne, lecz także do aproksymacji rozkładów zamożności, charakteryzujących się znacznym odsetkiem ujemnych aktywów brutto, a jednocześnie zerowych i ujemnych aktywów netto.

Dystrybuenta rozkładu czteroparametrowego może być wyrażona w następującej postaci:

$$F(y) = \beta + (1 - \beta) \left(1 + \left(\frac{y}{b} \right)^{-\nu} \right)^{-a} \quad (1.85)$$

dla $a, \nu, b > 0$.

Typ rozkładu określonego wzorem (1.85) zależy od wielkości parametru β :

- $\beta = 0$ – rozkład typu I;
- $0 < \beta < 1$ – rozkład typu II;
- $\beta < 0$ – rozkład typu III.

Parametr $\beta \in (0, 1)$ charakteryzujący rozkład typu II posiada interesującą interpretację ekonomiczną. W przypadku rozkładu dochodów osobistych można go traktować jako oszacowanie czystej stopy bezrobocia występującej w badanej zbiorowości, gdyż jego wysokość zależy od odsetka osób nieosiągających żadnego dochodu, nawet zasiłku.

Podsumowując, można stwierdzić, że rozkład Daguma charakteryzuje się szeregiem korzystnych z punktu widzenia analizy rozkładów płac i dochodów. Powstał na podstawie obserwacji rozkładów empirycznych, ale ma podbudowę teoretyczną opartą na modelu stochastycznym. Jest zbieżny z rozkładem Pareto dla dochodów wysokich, a momenty wyższych rzędów tego rozkładu dążą do nieskończoności, co zapewnia wysoką zgodność z danymi empirycznymi. Pod tym względem rozkład Daguma znacznie przewyższa często stosowane rozkłady teoretyczne, takie jak logarytmiczno-normalny czy gamma

(por. Jędrzejczak, 1990). Rozkład ten może być zarówno jedno-, jak i zeromodalny; nadaje się także do analizy populacji, w których występują osoby o dochodach zerowych i ujemnych. W odróżnieniu od rozkładu logarytmiczno-normalnego znana jest postać dystrybuanty tego rozkładu oraz krzywej koncentracji Lorenza.

1.6. Rozkład Singha-Maddali

Singh i Maddala (1976) zaproponowali nowy teoretyczny rozkład dochodów, który pod względem zgodności z danymi empirycznymi przewyższył stosowane powszechnie modele dwu- i trójparametrowe, takie jak rozkład logarytmiczno-normalny, gamma czy $\log t$. Punktem wyjścia do konstrukcji tego modelu była formalna analogia do krzywej długości życia. W Polsce rozkład Singha-Maddali był wykorzystywany do badania dochodów ludności, m.in. przez Kota (1995).

Niech $f(x)$ będzie funkcją gęstości zmiennej losowej X określonej jako długość życia natomiast F dystrybuantą tej zmiennej. Gęstość $f(x)$ będzie więc oznaczała prawdopodobieństwo śmierci w małym okresie czasu dx , natomiast $1 - F(x)$ prawdopodobieństwo przeżycia co najmniej do wieku x . Stopa ryzyka śmierci w krótkim okresie czasu określona jest następująco:

$$r(x) = \frac{f(x)}{1 - F(x)}. \quad (1.86)$$

Dla rozkładów długości życia $r(x)$ jest rosnącą funkcją zmiennej X . Rosnącą funkcję $r(x)$ zwaną też funkcją hazardu mają rozkłady: gamma, wykładniczy, normalny.

Singh i Maddala poszukiwali postaci $r(x)$, która byłaby najbardziej odpowiednia dla rozkładów dochodów. Wiadomo, że rozkład Pareto, dobrze dopasowany do wysokich grup dochodowych, charakteryzuje się malejącą funkcją hazardu, podczas gdy dla rozkładu logarytmiczno-normalnego funkcja ta jest rosnąca, a następnie malejąca. Ogólnie można więc przyjąć, że powinna to być funkcja najpierw rosnąca, potem ma-

lejąca. W przypadku rozkładów dochodów wygodniej jest jednak rozważać proporcjonalną stopę ryzyka określoną wzorem:

$$r^*(z) = \frac{\frac{dF}{dz}}{1 - F(z)} \quad (1.87)$$

gdzie: $z = \ln y$, y – dochód,
 $F(z)$ – dystrybuanta zmiennej losowej Z .

Przyjęto, że funkcja $r^*(z)$ powinna najpierw rosnąć z rosnącą stopą wzrostu, następnie z malejącą stopą wzrostu dążyć do funkcji stałej. Ponadto stopa wzrostu powinna wynosić zero, gdy $r^*(z) = 0$. Warunki te spełnia rozwiązanie równania różniczkowego postaci:

$$g''(z) = ag'(z)[\alpha - g'(z)] \quad \text{dla} \quad g''(z) > 0, \quad (1.88)$$

gdzie $g(z) = -\ln(1 - F(z))$, α jest stałą Pareto, zaś a jest pewną stałą.

Rozwiązanie tego równania daje dystrybuantę rozkładu dochodów postaci:

$$F(y) = \begin{cases} 1 - \frac{1}{(1 + a_1 y^{a_2})^{a_3}} & \text{dla } y \geq 0, \\ 0 & \text{dla } y < 0, \end{cases} \quad (1.89)$$

gdzie a_1, a_2, a_3 są dodatnimi parametrami rozkładu.

Funkcja gęstości odpowiadająca dystrybuancie określonej wzorem (1.89) jest postaci:

$$f(y) = \begin{cases} a_1 a_2 a_3 \frac{y^{a_2-1}}{(1 + a_1 y^{a_2})^{a_3+1}} & \text{dla } y \geq 0, \\ 0 & \text{dla } y < 0. \end{cases} \quad (1.90)$$

Rozkład o funkcji gęstości (1.90) i dystrybuancie (1.89) jest nazywany rozkładem Singha-Maddali, $S - M(a_1, a_2, a_3)$.

Podstawiając: $a_3 = q$, $a_2 = a$ oraz $a_1 = b^{-a}$ ($q, a, b > 0$) otrzymujemy alternatywną parametryzację rozkładu $S - M(q, a, b)$ i wzór funkcji gęstości postaci:

$$f(y) = \begin{cases} ab^{-a} q y^{a-1} \left(1 + \left(\frac{y}{b} \right)^a \right)^{-q-1} & \text{dla } y \geq 0, \\ 0 & \text{dla } y < 0. \end{cases} \quad (1.91)$$

oraz dystrybuanty:

$$F(y) = \begin{cases} 1 - \frac{1}{\left(1 + \left(\frac{y}{b} \right)^a \right)^q} & \text{dla } y \geq 0, \\ 0 & \text{dla } y < 0. \end{cases} \quad (1.92)$$

Przykładowe wykresy gęstości rozkładu $S - M(q, a, b)$ przedstawione są na rys. 1.13–1.15.

Dla $a > 1$ rozkład jest jednomodalny, natomiast dla $a \leq 1$ funkcja gęstości jest L -kształtna. Funkcja ta spełnia także warunek zbieżności rozkładu do krzywej Pareto dla $y \rightarrow \infty$.

Momenty rzędu r istnieją, gdy spełniony jest warunek $r < aq$. Przyjmują one postać:

$$E(Y^r) = \frac{b^r \Gamma\left(1 + \frac{r}{a}\right) \Gamma\left(q - \frac{r}{a}\right)}{\Gamma(q)}, \quad (1.93)$$

gdzie $B(\cdot, \cdot)$ jest funkcją beta.

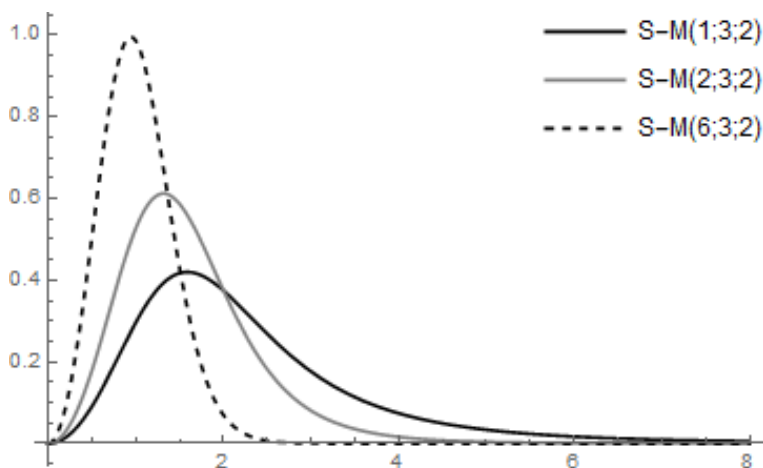
W szczególności wartość oczekiwana określona jest wzorem:

$$E(Y) = \frac{b \Gamma\left(1 + \frac{1}{a}\right) \Gamma\left(q - \frac{1}{a}\right)}{\Gamma(q)}, \quad \text{gdzie } aq > 1. \quad (1.94)$$

Odchylenie standardowe rozkładu Singha-Maddali jest postaci:

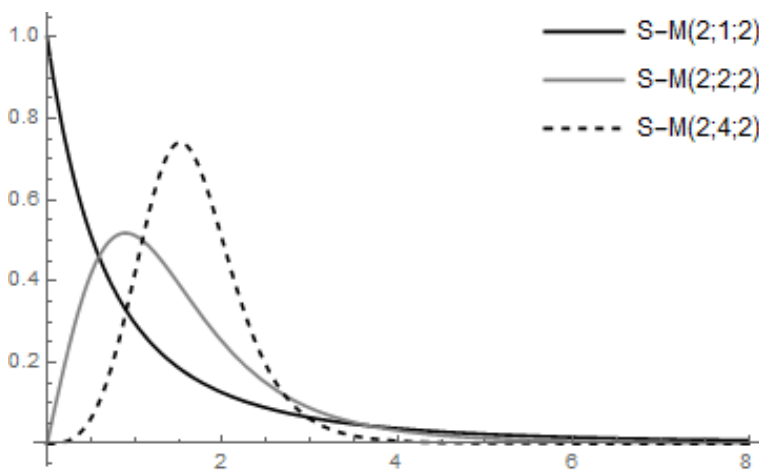
$$D(Y) = \frac{b}{\Gamma(q)} \sqrt{\Gamma\left(1 + \frac{2}{a}\right) \Gamma\left(q - \frac{2}{a}\right) \Gamma(q) - \Gamma^2\left(1 + \frac{1}{a}\right) \Gamma^2\left(q - \frac{1}{a}\right)}, \quad (1.95)$$

gdy $aq > 2$.



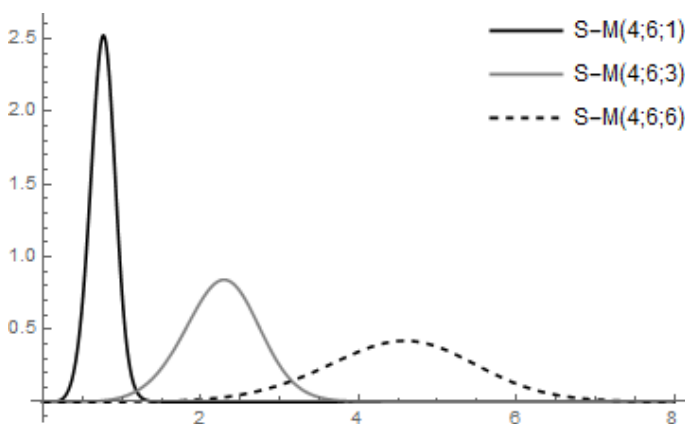
Rysunek 1.13. Gęstości rozkładów Singha-Maddali dla wybranych wartości q i ustalonych pozostałych parametrów

Źródło: opracowanie własne



Rysunek 1.14. Gęstości rozkładów Singha-Maddali dla wybranych wartości a i ustalonych pozostałych parametrów

Źródło: opracowanie własne



Rysunek 1.15. Gęstości rozkładów Singha-Maddali dla wybranych wartości b i ustalonych pozostałych parametrów

Źródło: opracowanie własne

Mediana rozkładu Singha-Maddali wyraża się wzorem:

$$Me = b \left(\frac{1}{2^q - 1} \right)^{\frac{1}{a}}. \quad (1.96)$$

Funkcja Lorenza istnieje, gdy $aq > 1$. Ma ona postać:

$$L(p) = B \left(1 - (1-q)^{\frac{1}{q}}; 1 + \frac{1}{a}, q - \frac{1}{a} \right), \quad (1.97)$$

gdzie $B(\cdot; \cdot, \cdot)$ jest niekompletną funkcją beta.

Współczynnik Giniego wyraża się wzorem:

$$G = 1 - \frac{\Gamma(q)\Gamma\left(2q - \frac{1}{a}\right)}{\Gamma\left(q - \frac{1}{a}\right)\Gamma(2q)}, \quad (1.98)$$

gdzie Γ jest funkcją gamma.

Wartości współczynnika Giniego dla wybranych parametrów a i q rozkładu Singha-Maddali przedstawione są w tab. 1.8.

Tabela 1.8. Wartości współczynnika Giniego dla wybranych rozkładów Singha-Maddali

Parametr a	Parametr q	Współczynnik Giniego
1,0	1,0	0,75
1,0	2,0	0,67
1,0	20,0	0,51
2,0	1,0	0,50
2,0	2,0	0,38
2,0	20,0	0,30
6,0	1,0	0,17
6,0	2,0	0,13
6,0	20,0	0,11

Źródło: obliczenia własne

Na wartość współczynnika Giniego mają wpływ obydwa parametry (wzór (1.98)), ale jest on bardziej wrażliwy na wartość parametru a . Nie-równomierności rozkładu są mniejsze, im większe są wartości a oraz q .

Kleiber (1996) wykazał, iż pomiędzy rozkładem Daguma a rozkładem Singha-Maddali istnieje ściśle powiązanie. Gdy zmienna losowa Y ma rozkład Daguma z parametrami a , v i b , wówczas jej odwrotność, czyli zmienna losowa $\frac{1}{Y}$, ma rozkład Singha-Maddali z parametrami a , $\frac{1}{v}$ i b .

1.7. Rozkład Fiska

Rozkład Fiska jest szczególnym przypadkiem rozkładu Singha-Maddali. Podstawiając $q = 1$ do wzoru na funkcję gęstości rozkładu Singha-Maddali, otrzymujemy gęstości rozkładu Fiska $F(a, b)$ określoną wzorem:

$$f(y) = \begin{cases} \frac{a \left(\frac{y}{b}\right)^{a-1}}{b \left(1 + \left(\frac{y}{b}\right)^a\right)^2} & \text{dla } y \geq 0, \\ 0 & \text{dla } y < 0 \end{cases} \quad (1.99)$$

oraz dystrybuantę:

$$F(y) = \begin{cases} 1 - \frac{1}{1 + \left(\frac{y}{b}\right)^a} & \text{dla } y \geq 0, \\ 0 & \text{dla } y < 0. \end{cases} \quad (1.100)$$

Ze względu na fakt, że rozkład Fiska jest szczególnym przypadkiem rozkładu Singha-Maddali wartość oczekiwaną oraz odchylenie standardowe otrzymujemy z wzorów (1.94) i (1.95), podstawiając $q = 1$.

Inne postaci wzorów na wartość oczekiwaną i odchylenie standardowe rozkładu Fiska otrzymujemy, wykorzystując własności funkcji gamma: $\Gamma(x + 1) = x\Gamma(x)$ oraz $\Gamma(x)\Gamma(1 - x) = \frac{\pi}{\sin \pi x}$:

$$EY = \frac{b\pi}{a \sin\left(\frac{\pi}{a}\right)} \quad \text{dla } a > 1, \quad (1.101)$$

$$DY = b \sqrt{\frac{2\pi}{a \sin\left(\frac{2\pi}{a}\right)} - \frac{\pi^2}{a^2 \sin^2\left(\frac{\pi}{a}\right)}} \quad \text{dla } a > 2. \quad (1.102)$$

Mediana rozkładu Fiska ma postać:

$$Me = b, \quad (1.103)$$

zaś dominanta wyraża się wzorem:

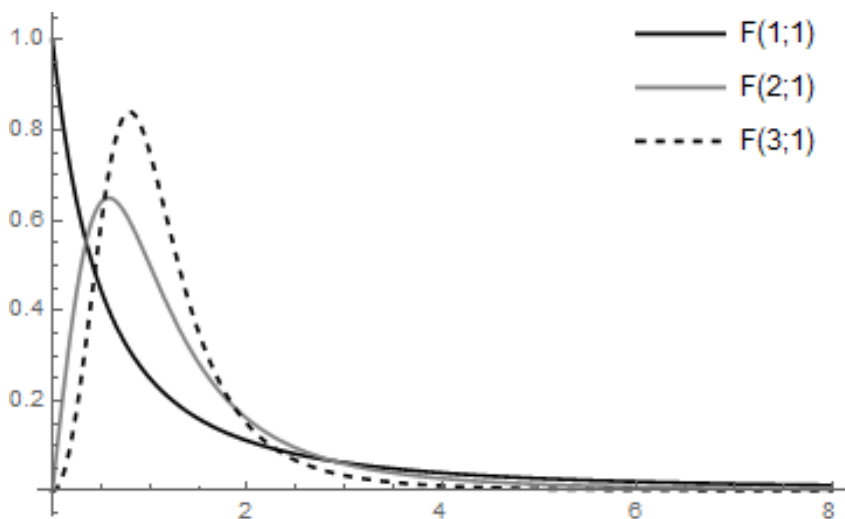
$$D_0 = b \left(\frac{a-1}{a+1} \right)^{\frac{1}{a}}. \quad (1.104)$$

Współczynnik Giniego wyznaczany z wzoru (1.98) i własności $\Gamma(1 + x) = x\Gamma(x)$ ma następującą postać:

$$G = 1 - \frac{\Gamma\left(2 - \frac{1}{a}\right)}{\Gamma\left(1 - \frac{1}{a}\right)} = 1 - \frac{\left(1 - \frac{1}{a}\right)\Gamma\left(1 - \frac{1}{a}\right)}{\Gamma\left(1 - \frac{1}{a}\right)} = \frac{1}{a}. \quad (1.105)$$

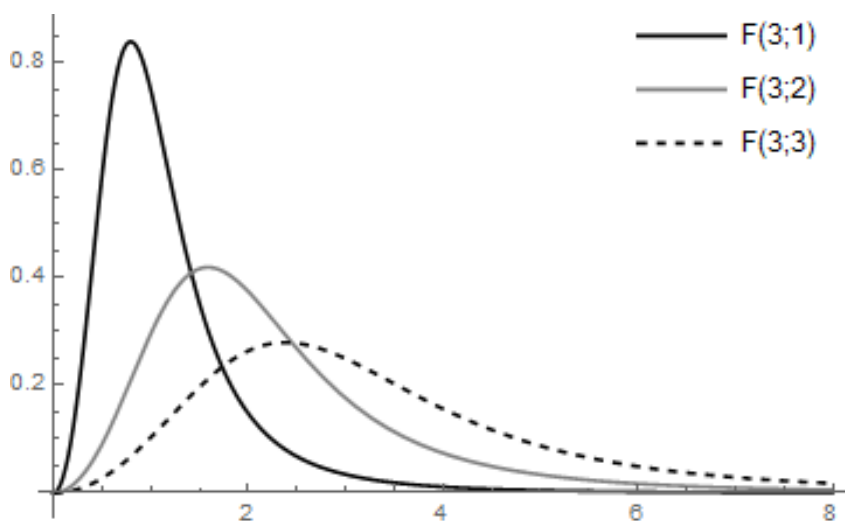
Rozkład Fiska zwany jest również rozkładem logarytmiczno-logistycznym.

Przykładowe wykresy gęstości rozkładów Fiska przedstawione są na rys. 1.16 i 1.17.



Rysunek 1.16. Gęstości rozkładów Fiska dla wartości $b = 1$ i wybranych wartości a

Źródło: opracowanie własne



Rysunek 1.17. Gęstości rozkładów Fiska dla wartości $a = 3$ i wybranych wartości b

Źródło: opracowanie własne

1.8. Rozkład Weibulla

Dwuparametrowy rozkład Weibulla jest szczególnym przypadkiem rozkładu Singha-Maddali, gdy $a_3 \rightarrow \infty$. Dystrybuenta rozkładu Weibulla $W(\alpha, k)$ określona jest wzorem:

$$F(y) = \begin{cases} 1 - e^{-\left(\frac{y}{k}\right)^\alpha} & \text{dla } y \geq 0, \\ 0 & \text{dla } y < 0. \end{cases} \quad (1.106)$$

gdzie $k, \alpha > 0$, zaś funkcja gęstości rozkładu ma postać:

$$f(y) = \begin{cases} \frac{\alpha}{k} \left(\frac{y}{k}\right)^{\alpha-1} e^{-\left(\frac{y}{k}\right)^\alpha} & \text{dla } y \geq 0, \\ 0 & \text{dla } y < 0. \end{cases} \quad (1.107)$$

Przykładowe wykresy gęstości rozkładu Weibulla przedstawione są na rys. 1.18 i 1.19. Jeśli $\alpha < 1$, to rozkład jest zeromodalny (kształt rozkładu Pareto), jeśli $\alpha = 1$, to funkcja gęstości przecina oś OY, zaś dla $\alpha > 1$ wartości koncentrują się coraz bardziej wartości średniej. Dla rozkładów dochodów parametr $\alpha \in [1,5; 2,5]$.

Momenty zwykłe r -tego rzędu zawsze istnieją i są postaci:

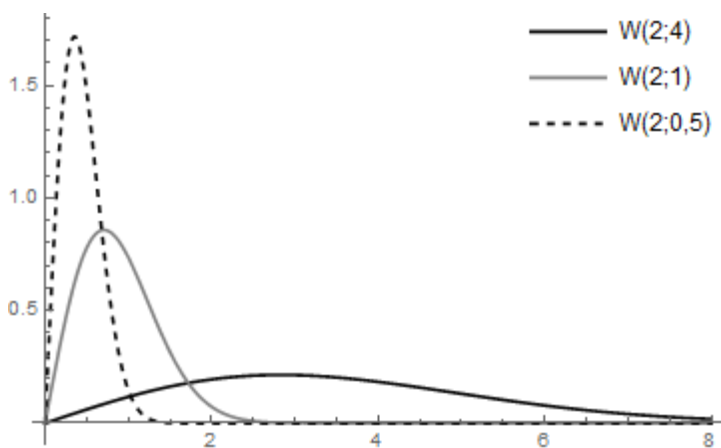
$$EY^r = k^r \Gamma\left(1 + \frac{r}{\alpha}\right), \quad (1.108)$$

w szczególności istnieją wartość oczekiwana rozkładu oraz wariancja:

$$EY = k \Gamma\left(1 + \frac{1}{\alpha}\right), \quad (1.109)$$

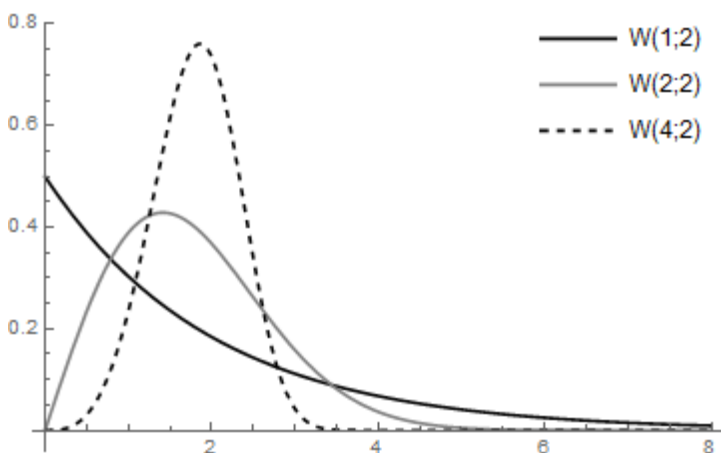
$$DY^2 = k^2 \left(\Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma^2\left(1 + \frac{1}{\alpha}\right) \right), \quad (1.110)$$

gdzie $\Gamma(\cdot)$ jest funkcją gamma.



Rysunek 1.18. Gęstości rozkładów Weibulla dla wartości $\alpha = 2$ i wybranych wartości k

Źródło: opracowanie własne



Rysunek 1.19. Gęstości rozkładów Weibulla dla wartości $k = 2$ i wybranych wartości α

Źródło: opracowanie własne

Mediana rozkładu Weibulla wyraża się wzorem:

$$Me = k(\ln 2)^{\frac{1}{\alpha}}, \quad (1.111)$$

a dominanta:

$$D_0 = k \left(\frac{\alpha - 1}{\alpha} \right)^{\frac{1}{\alpha}} \quad \text{dla } \alpha > 1. \quad (1.112)$$

W analizach dochodów wykorzystuje się funkcję Lorenza, która ma postać:

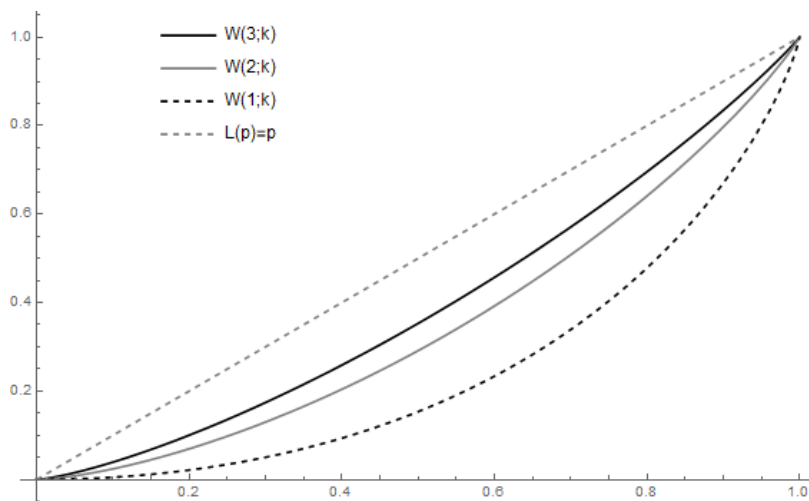
$$L(p) = 1 - \frac{\Gamma\left(-\ln(1-p); 1 + \frac{1}{\alpha}\right)}{\Gamma\left(1 + \frac{1}{\alpha}\right)} \quad \text{dla } 0 \leq p \leq 1, \quad (1.113)$$

gdzie $\Gamma(\cdot; \cdot)$ jest niekompletną funkcją gamma.

Współczynnik Giniego określający stopień nierównomierności rozkładu wyraża się wzorem:

$$G = 1 - 2^{-1/\alpha}. \quad (1.114)$$

Funkcje Lorenza dla wybranych rozkładów przedstawione są na rys. 1.20, a wartości współczynników Giniego w tab. 1.9.



Rysunek 1.20. Funkcje Lorenza dla wybranych rozkładów Weibulla

Źródło: obliczenia własne

Tabela 1.9. Wartości współczynnika Giniego dla wybranych rozkładów Weibulla

Parametr α	Współczynnik Giniego
0,5	0,75
1,0	0,50
1,5	0,37

Parametr α	Współczynnik Giniego
2,0	0,29
2,5	0,24
3,0	0,21

Źródło: obliczenia własne

1.9. Rozkład Zengi

W 2010 roku Zenga dla nieujemnych wartości dochodów skonstruował trzyparametrową funkcję gęstości f , która dobrze aproksymuje rozkłady dochodów w wielu krajach (por. Zenga, 2010; Zenga i in., 2010a). Funkcja ta powstała w oparciu o ucięty rozkład Pareto przedstawiony przez Poliscchio (2008). Rozkład Zengi jest w istocie mieszaniną rozkładów Poliscchio z wagami otrzymanymi na podstawie funkcji beta.

Funkcja gęstości rozkładu Zengi przyjmuje postać:

$$f(y) = \int_0^1 v(y; \mu, k) g(k; \alpha, \theta) dk =$$

$$= \begin{cases} \frac{1}{2\mu B(\alpha, \theta)} \left(\frac{y}{\mu}\right)^{-1,5} \int_0^{\frac{y}{\mu}} k^{\alpha-0,5} (1-k)^{\theta-2} dk, & \text{gdymy } 0 < y \leq \mu \\ \frac{1}{2\mu B(\alpha, \theta)} \left(\frac{\mu}{y}\right)^{1,5} \int_0^{\frac{\mu}{y}} k^{\alpha-0,5} (1-k)^{\theta-2} dk, & \text{gdymy } y > \mu, \end{cases} \quad (1.115)$$

gdzie

$$v(y; \mu, k) = \begin{cases} \frac{\sqrt{\mu}}{2} k^{0,5} (1-k)^{-1} y^{-1,5} & \text{dla } \mu k \leq y \leq \frac{\mu}{k}; \mu > 0, 0 < k < 1 \\ 0 & \text{dla pozostałych przypadków,} \end{cases} \quad (1.116)$$

$$g(k; \alpha, \theta) = \begin{cases} \frac{k^{\alpha-1} (1-k)^{\theta-1}}{B(\alpha, \theta)} & \text{gdymy } 0 < k < 1; \theta > 0, \alpha > 1 \\ 0 & \text{dla pozostałych przypadków,} \end{cases} \quad (1.117)$$

oraz $B(\alpha, \theta)$ jest funkcją beta.

Parametr $\mu > 0$ jest parametrem skali rozkładu Zengi, natomiast $\alpha > 0$ i $\theta > 0$ są parametrami kształtu.

Gdy parametr $\theta > 1$ dystrybuenta rozkładu Zengi ma postać (por. Zenga i in., 2010a):

$$F(y) = F(y; \mu, \alpha, \theta) = \begin{cases} \frac{1}{B(\alpha; \theta)} \left[B\left(\frac{y}{\mu}; \alpha, \theta-1\right) - \left(\frac{\mu}{y}\right)^{1/2} B\left(\frac{y}{\mu}; \alpha + \frac{1}{2}, \theta-1\right) \right] \\ \text{gdy } 0 < y \leq \mu \\ 1 - \frac{1}{B(\alpha; \theta)} \left[\left(\frac{\mu}{y}\right)^{1/2} B\left(\frac{y}{\mu}; \alpha + \frac{1}{2}, \theta-1\right) - B\left(\frac{y}{\mu}; \alpha+1, \theta-1\right) \right] \\ \text{gdy } y > \mu, \end{cases} \quad (1.118)$$

gdzie $B(\cdot, \cdot)$ jest funkcją beta, natomiast $B(\cdot; \cdot, \cdot)$ jest niekompletną funkcją beta.

Uogólnienie wzoru (1.118) dla wartości $\theta > 0$ przyjmuje postać (por. Zenga i in., 2010b):

$$F(y) = F(y; \mu, \alpha, \theta) = \begin{cases} \frac{1}{B(\alpha; \theta)} \sum_{i=1}^{\infty} \left[B\left(\frac{y}{\mu}; \alpha+i-1, \theta\right) - \left(\frac{\mu}{y}\right)^{1/2} B\left(\frac{y}{\mu}; \alpha+i-\frac{1}{2}, \theta\right) \right] \\ \text{gdy } 0 < y \leq \mu \\ 1 - \frac{1}{B(\alpha; \theta)} \sum_{i=1}^{\infty} \left[\left(\frac{\mu}{y}\right)^{1/2} B\left(\frac{y}{\mu}; \alpha+i-\frac{1}{2}, \theta\right) - B\left(\frac{y}{\mu}; \alpha+i, \theta\right) \right] \\ \text{gdy } y > \mu. \end{cases} \quad (1.119)$$

Momenty zwykłe rzędu r rozkładu Zengi wyrażają się wzorem:

$$E(Y^r) = \frac{\mu^r}{2r-1} \cdot \frac{1}{B(\alpha, \theta)} [(B(\alpha-r+1, \theta-1) - B(\alpha+r, \theta-1))], \quad (1.120)$$

gdy $r \in N, r < \alpha + 1$ oraz $\theta > 1$.

Uogólniony wzór na momenty zwykłe rzędu r , dla $\theta > 0$, jest postaci:

$$E(Y^r) = \frac{\mu^r}{2r-1} \cdot \frac{1}{B(\alpha, \theta)} \sum_{i=1}^{2r-1} B(\alpha-r+i, \theta), \quad (1.121)$$

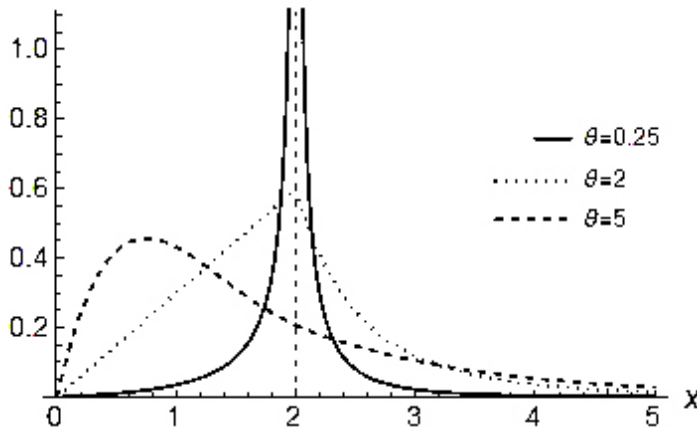
gdy $r \in \mathbb{N}$, $r < \alpha + 1$.

Z wzorów (1.120) i (1.121) wynika, że wartość oczekiwana zmiennej losowej o rozkładzie Zengi jest równa parametrowi skali μ , tzn.:

$$E(Y) = \mu. \quad (1.122)$$

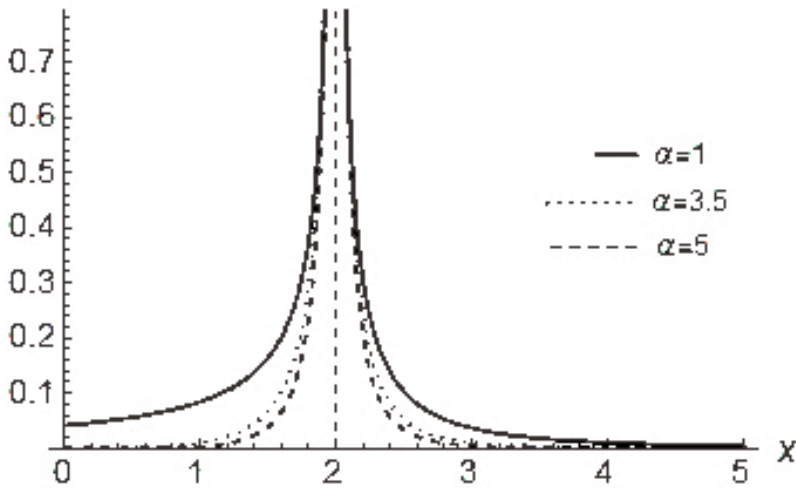
Wzory na momenty zwykłe rzędu r rozkładu Zengi ułatwiają estymację jego parametrów i umożliwiają bezpośrednią interpretację parametru skali. Należy podkreślić, że parametr skali rozkładu Zengi μ jest także średnią rozkładu Polisicchio (2008). Ponadto parametr ten w przypadku rozkładu Zengi jest zawsze skończony, w odróżnieniu od wielu innych rozkładów, w tym rozkładu Daguma, gdzie w zależności od parametrów kształtu może być skończony lub nieskończony.

Liczba skończonych momentów rozkładu zależy od parametru α i dla typowych wartości tego parametru większych od 2 rozkład Zengi posiada 3 lub więcej momentów. Funkcja gęstości rozkładu Zengi przyjmuje różne kształty, bardziej różnorodne niż tradycyjne trójparametrowe modele rozkładów dochodów, jak np. rozkład Daguma i cecha ta pozwala na dobre dopasowanie także dla niskich grup dochodowych. Przykładowe wykresy funkcji gęstości przedstawione są na rys. 1.21–1.22.



Rysunek 1.21. Gęstości rozkładu Zengi dla $\mu = 2$ oraz $\alpha = 2$ i wybranych wartości parametru θ

Źródło: obliczenia własne



Rysunek 1.22. Gęstości rozkładu Zengi dla $\mu = 2$ oraz $\theta = 0,25$ i wybranych wartości parametru α

Źródło: opracowanie własne

Parametry rozkładu Zengi (2010), który po zastosowaniu odpowiednich miar zgodności okazał się dostatecznie dobrze dopasowany do rozważanego empirycznego rozkładu dochodu, mogą być następnie wykorzystane do oceny różnych charakterystyk statystycznych rozkładu, w szczególności punktowych i syntetycznych miar nierówności dochodów. W przypadku tego rozkładu popularny współczynnik nierównomierności Giniego może być otrzymany w oparciu o numeryczne oszacowanie pola między krzywą Lorenza $L(p)$ a linią równomiernego podziału $p = L(p)$.

W celu otrzymania dla funkcji gęstości $f(y; \mu, \alpha, \theta)$ wykresu krzywej Lorenza należy obliczyć dla wygenerowanych wartości zmiennej Y wartości pierwszego niekompletnego momentu $H(y; 1, \alpha, \theta)$ i dystrybuanty $F(y; 1, \alpha, \theta)$. Wiadomo, że zarówno funkcja Lorenza, jak i współczynnik nierównomierności Giniego są niezależne od przyjętej skali dochodu (*invariant to scale transformation*), wystarczy więc rozważyć najprostszy przypadek $f(y; 1, \alpha, \theta)$. Wiadomo, że:

$$\begin{aligned}
F(y; 1, \alpha, \theta) &= \\
&= \begin{cases} F_1(y; 1, \alpha, \theta) = \frac{1}{B(\alpha, \theta)} \sum_{i=1}^{\infty} \left[B(y; \alpha+i-1, \theta) - y^{-1/2} B\left(y; \alpha+i-\frac{1}{2}, \theta\right) \right] \\ \text{gdy } 0 < y \leq 1 \\ F_2(y; 1, \alpha, \theta) = 1 - \frac{1}{B(\alpha, \theta)} \sum_{i=1}^{\infty} \left[y^{-1/2} B\left(\frac{1}{y}; \alpha+i-\frac{1}{2}, \theta\right) + B\left(\frac{1}{y}; \alpha+i, \theta\right) \right] \\ \text{gdy } y > 1. \end{cases} \quad (1.123)
\end{aligned}$$

Pierwszy niekompletny moment zmiennej losowej Y o gęstości $f(y; \mu, \alpha, \theta)$ określony jest następująco:

$$H(y; 1, \alpha, \theta) = \int_0^y t f(t; 1, \alpha, \theta) dt. \quad (1.124)$$

W wyniku elementarnych przekształceń dla rozkładu Zengi otrzymujemy następującą postać funkcji H :

$$\begin{aligned}
H(y; 1, \alpha, \theta) &= \\
&= \begin{cases} \frac{1}{B(\alpha, \theta)} \sum_{i=1}^{\infty} \left[y^{1/2} B\left(y; \alpha+i-\frac{1}{2}, \theta\right) - B(y; \alpha+i, \theta) \right] \\ \text{gdy } 0 < y \leq 1 \\ 1 - \frac{1}{B(\alpha, \theta)} \sum_{i=1}^{\infty} \left[B\left(\frac{1}{y}; \alpha+i-1, \theta\right) + y^{1/2} B\left(\frac{1}{y}; \alpha+i-\frac{1}{2}, \theta\right) \right] \\ \text{gdy } y > 1. \end{cases} \quad (1.125)
\end{aligned}$$

Wykres punktowej miary koncentracji opartej na krzywej Lorenza dla rozkładu Zengi otrzymujemy, odkładając na osi odciętych $F(y; 1, \alpha, \theta)$, zaś na osi rzędnych wartości $H(y; 1, \alpha, \theta)$, dla dostatecznej liczby wartości y . Można również dokonać transformacji $p = F(y)$ i dla dostatecznej dużej liczby wartości $p \in [0, 1]$ wykreślić funkcję Lorenza $L(p)$. W takim przypadku wartości otrzymane na osiach OX i OY będą miały jednakową skalę.

Współczynnik Giniego, określony jako podwojone pole między krzywą Lorenza a linią $p = L(p)$, możemy otrzymać za pomocą całkowania numerycznego. Przykładowe wartości współczynnika Giniego podane zostały w tab. 1.10.

Tabela 1.10. Wartości współczynnika Giniego dla wybranych rozkładów Zengi

Parametr a	Parametr θ	Współczynnik Giniego
4,5	5,2	0,29
3,2	3,8	0,30
5,0	6,5	0,31
2,8	4,0	0,34
2,0	3,0	0,37
1,5	3,5	0,49
1,2	4,0	0,56

Źródło: obliczenia własne

ROZDZIAŁ 2

Podejście systemowe do modeli rozkładów dochodów

2.1. Uwagi wstępne

Przeprowadzając przegląd różnych rozkładów płac i dochodów, powstaje problem badania ich własności. W tym celu wygodnie jest pogrupować rozkłady w systemy lub rodziny, które ułatwiają badanie ich struktury i rozważanie ich wad oraz zalet. Niektóre z systemów, jak system D'Addario czy system Daguma, mają charakter systemów generujących rozkłady i zostały stworzone specjalnie z myślą o rozkładach dochodów; inne zaś, jak systemy krzywych Pearsona, Burra czy Johnsona systematyzują rozkłady prawdopodobieństwa w pewne jednorodne grupy i rozważają różne funkcje rozkładów prawdopodobieństwa, a wśród nich także funkcje rozkładów dochodów. Dagum (1980; 1985; 1990) wyprowadził prawie wszystkie modele rozkładów dochodów z jednego z trzech systemów: D'Addario, Daguma, Pearsona. Oczywiście systemy te nie są rozłączne, niektóre funkcje rozkładów należą do kilku systemów jednocześnie: rozkład logarytmiczno-normalny należy do systemów D'Addario i Johnsona, rozkład gamma – do systemów Pearsona, D'Addario i Johnsona, rozkład Daguma – do systemów Daguma, Pearsona oraz Burra, rozkład Pareto – do systemów Daguma i D'Addario, a Singha-Maddali – do systemów Daguma i Burra.

W poszukiwaniu coraz bardziej uniwersalnych rozkładów, które mogą przyjmować różne kształty i łatwo dopasowywać się do danych empirycznych w różnych przekrojach, częstą praktyką jest uogólnianie pewnych rozkładów prawdopodobieństwa poprzez dodawanie kolejnych

parametrów. W ten sposób powstają rodziny rozkładów, które jako szczególne lub graniczne przypadki mogą zawierać bardzo wiele znanych rozkładów o mniejszej liczbie parametrów. W ten sposób, poprzez dodanie czwartego parametru do standardowego rozkładu beta drugiego rodzaju, powstała rodzina zwana uogólnionym rozkładem beta drugiego rodzaju (GB2), która ma szczególne znaczenie z punktu widzenia badania rozkładów różnych rozkładów według wielkości (*size distributions*), do których należą rozkłady płac i dochodów.

Johnson i inni (1994), dokonując przeglądu najczęściej używanych rozkładów statystycznych, stwierdzili, że do dobrej aproksymacji rozkładów empirycznych potrzebne są co najmniej trzy parametry, zaś zastosowanie rozkładów czteroparametrowych, do których należy rozkład GB2, powinno w zupełności wystarczyć do większości praktycznych celów. Przyjmuje się, że nie można oczekiwać zauważalnej poprawy wynikającej z włączenia piątego lub szóstego parametru.

2.2. System Daguma

System Daguma opiera się na równaniu różniczkowym, które jest efektem obserwacji pewnych prawidłowości związanych z kształtowaniem się elastyczności dochodowej dystrybuanty rozkładów dochodów, płac i zamożności. Równanie to ma następującą postać (por. Dagum, 1990):

$$\varepsilon(F(y) - \alpha, y) = \frac{d \ln(F(y) - \alpha)}{d \ln y} = \psi(y)\phi(F) \leq k \quad (2.1)$$

przy czym: $k > 0$, $0 \leq y_0 < y < \infty$ oraz $\alpha < 1$, $\psi(y) > 0$, $\phi(F) > 0$, $\frac{d(\psi(y)\phi(F))}{dy} < 0$, gdzie y_0 jest dochodem minimalnym.

Zgodnie z równaniem (2.1) elastyczność dochodowa dystrybuanty rozkładu dochodów $\varepsilon(F(y) - \alpha, y)$ jest dodatnią, malejącą i ograniczoną funkcją dystrybuanty, a więc i dochodów. Podstawiając za funkcje $\psi(y)$ i $\phi(F)$ odpowiednie formuły i rozwiązując równanie (2.1), otrzymujemy wiele modeli rozkładów dochodów (por. tab. 2.1).

Tabela 2.1. Rozkłady dochodów należące do systemu Daguma

Rozkład dochodu	$\psi(y)$	$\phi(F)$	δ	(α, β)	Zakres
Pareto typu I	δ	$\frac{1-F}{F}$	> 1	$(0,0)$	$0 \leq y_0 < y < \infty$
Pareto typu II	$\frac{\delta y}{y-c}$	$\frac{1-F}{F}$	> 1	$(0,0)$	$0 \leq y_0 < y < \infty$
Pareto typu III	$\beta y + \frac{\delta y}{y-c}$	$\frac{1-F}{F}$	> 0	$(0, +)$	$0 \leq y_0 < y < \infty$
Beniniego	$2\delta \ln y$	$\frac{1-F}{F}$	> 0	$(0,0)$	$0 \leq y_0 < y < \infty$
Weibulla	$\beta y(y-c)^{\delta-1}$	$\frac{1-F}{F}$	> 0	$(0, +)$	$c \leq y < \infty$
Fisk	δ	$1-F$	> 1	$(0,0)$	$0 \leq y < \infty$
Singha-Maddali	δ	$\frac{1-(1-F)^\beta}{F(1-F)^{-1}}$	> 0	$(0, +)$	$0 \leq y < \infty$
Log-Gompertza	$-\ln \delta$	$-\ln F$	$(0, 1)$	$(0,0)$	$0 \leq y < \infty$
Daguma typu I	δ	$1-F^{1/\beta}$	> 1	$(0, +)$	$0 \leq y < \infty$
Daguma typu II	δ	$1-\left(\frac{F-\alpha}{1-\alpha}\right)^{1/\beta}$	> 1	$(+, +)$	$0 \leq y < \infty$
Daguma typu III	δ	$1-\left(\frac{F-\alpha}{1-\alpha}\right)^{1/\beta}$	> 1	$(-, +)$	$0 \leq y_0 < y < \infty$

Źródło: Kleiber, Kotz (2003)

Parametr oznaczony jako δ jest parametrem a rozkładów Pareto i Fiska, parametrem v Daguma oraz parametrem równym aq rozkładu Singha-Maddali. Parametr β jest równy parametrowi a rozkładu Daguma i parametrowi q rozkładu Singha-Maddali. Ponadto znaki i wartości parametrów α, β związane są z ograniczeniami równania (2.1).

Rozkłady zamieszczone w tab. 2.1 powstały przeważnie jako efekt obserwacji empirycznych dotyczących elastyczności dochodowej dystrybucyj. W ten sposób powstały krzywe: Pareto, Beniniego, Daguma I i II typu. Funkcje zaproponowane przez Fiska (*sech*²) oraz Singha i Maddalę nie były co prawda efektem obserwacji takich prawidłowości, jednak

spełniają one także warunek zbieżności do rozkładu Pareto dla wysokich grup dochodowych.

Rozkłady należące do systemu Daguma wykazują przeważnie bardzo wysoki stopień zgodności z danymi empirycznymi dla całego przedziału dochodów (rozkłady Daguma, Singha-Maddali oraz Fiska) lub też dla wysokich grup dochodowych (rozkład Pareto). Łatwo zauważyć, że w ramach systemu Daguma możemy otrzymać wiele innych rozkładów, wprowadzając parametry skali lub dodatkowe parametry kształtu, lub też stosując inne postaci funkcji $\phi(F)$.

2.3. System Pearsona

Wiele krzywych wykorzystywanych w analizie płac i dochodów należy do systemu krzywych Pearsona (1895; 1948). System ten powstał, aby opisywać różnorodne zjawiska o charakterze biologicznym lub społeczno-ekonomicznym. Ideą Pearsona było stworzenie systemu funkcji gęstości, które byłyby na tyle elastyczne, aby przechodziły jedna w drugą oraz aby obejmowały rozkłady zarówno symetryczne, jak i niesymetryczne o różnej dziedzinie.

Punktem wyjścia rozważań Pearsona były obserwacje dotyczące funkcji gęstości stosowanych w różnych dziedzinach badań, nie zaś obserwacje prawidłowości w dziedzinie rozkładów dochodów, co było tematem rozważań Daguma. Pearson zauważył, że główną cechą często stosowanych funkcji gęstości $f(y)$ jest to, że ich pierwsza pochodna dąży do zera dla pewnych wartości y . Ponadto zauważył, że funkcje jednomodalne najczęściej dążą do zera w jednym lub obu skrajnych punktach swej dziedziny, natomiast funkcje nieposiadające ekstremum – w jednym skrajnym punkcie dziedziny. Obie te własności spełniają funkcje będące rozwiązaniem następującego równania różniczkowego:

$$\frac{df(y)}{dy} = \frac{(y - c)f(y)}{a_0 + a_1 y + a_2 y^2}. \quad (2.2)$$

Funkcja $f(y)$ osiąga ekstremum w punkcie $y = c$. Jeśli jest to maksimum, to c odpowiada wartości dominanty. Jeśli jest to minimum, to rozkład jest

U-kształtny. Forma matematyczna funkcji $f(y)$ zależy od pierwiastków równania kwadratowego zawartego w mianowniku równania (2.2), czyli od wartości $\Delta = a_1^2 - 4a_0a_2$.

Parametry równania (2.2) można przedstawić, jak wykazał Pearson, za pomocą współczynnika skośności (g_1) i współczynnika spłaszczenia (g_2), a więc zależą one od czterech pierwszych momentów rozkładu:

$$\begin{aligned}a_0 &= 4g_2 - 3g_1^2, \\a_1 &= g_1(g_2 + 3), \\a_2 &= 2g_2 - 3g_1^2 - 6,\end{aligned}$$

gdzie: $g_1 = \frac{\mu_3}{\sigma^3}$, $g_2 = \frac{\mu_4}{\sigma^4}$ oraz μ_3, μ_4 są momentami centralnymi odpowiednio trzeciego i czwartego rzędu, zaś σ – odchyleniem standardowym.

Krzywe Pearsona klasyfikuje się w oparciu o wartość k :

$$k = \frac{a_1^2}{4a_0a_2} = \frac{g_1^2(g_2 + 3)^2}{4(4g_2 - 3g_1^2)(2g_2 - 3g_1^2 - 6)}. \quad (2.3)$$

Wyróżnia się trzy następujące typy główne krzywych Pearsona:

- Typ I – dla $k < 0$.
Dla równania $a_0 + a_1y + a_2y^2 = 0$ mamy $\Delta > 0$ oraz $\frac{a_2}{a_0} < 0$, czyli równanie ma dwa pierwiastki rzeczywiste przeciwnych znaków.
Rozkład beta jest szczególnym przypadkiem krzywej I typu.
- Typ IV – dla $0 < k < 1$.
W tym przypadku $a_0a_2 < 0$ i $\Delta < 0$, czyli istnieją pierwiastki zespolone równania $a_0 + a_1y + a_2y^2 = 0$.
Do tej grupy należy m.in. rozkład Cauchy'ego.
- Typ VI – dla $k > 1$.
Równanie $a_0 + a_1y + a_2y^2 = 0$ ma pierwiastki rzeczywiste tych samych znaków, gdyż $\Delta > 0$ oraz $\frac{a_2}{a_0} > 0$.
Do tej grupy zalicza się rozkład Fishera.
Oprócz typów głównych wyróżnia się także typy przechodnie, m.in.:
 - Typ II – dla $k = 0$ oraz $g_2 < 3$ ($a_2 < 0$).
Rozkład jednostajny jest przykładem krzywej II typu.

- Typ III – dla $k = \pm\infty$, $a_2 = 0$.
Rozkłady gamma oraz chi-kwadrat są szczególnymi przypadkami krzywej III typu.
- Typ V – dla $k = 1$.
Odwrócone rozkłady gamma i chi-kwadrat należą do grupy krzywych Pearsona V typu.
- Typ VII – dla $k = 0$, $g_2 > 3$ ($a_2 > 0$).
Do grupy tej zalicza się rozkład t -Studenta.

Jeśli $a_1 = a_2 = 0$ oraz $a_0 > 0$, to rozwiązaniem równania różniczkowego (2.2) są funkcje gęstości nazywane czasem „typu 0”. Są to m.in. funkcje gęstości rozkładu normalnego.

Z równania (2.2) wynika, że jeśli funkcja gęstości posiada dominantę, to jest nią $D_0 = c$.

Do systemu Pearsona zostały włączone inne krzywe, m.in. w 1916 roku krzywa Pareto, jako szczególny przypadek krzywej XI typu. Zestawienie krzywych Pearsona zawiera tab. 2.2.

Tabela 2.2. Funkcje gęstości krzywych Pearsona

Typ	Postać funkcji gęstości	Zakres	Parametry
I	$f(y) = h \left(1 + \frac{y}{\alpha_1} \right)^{m_1} \left(1 - \frac{y}{\alpha_2} \right)^{m_2}$	$-\alpha_1 \leq y \leq \alpha_2$	$m_1, m_2 > 0$
II	$f(y) = h \left(1 - \frac{y^2}{\alpha^2} \right)^m$	$-\alpha \leq y \leq \alpha$	$m > -1$
III	$f(y) = h \left(1 + \frac{y}{\alpha} \right)^{\mu\alpha} \exp(-\mu y)$	$-\alpha \leq y < \infty$	$\mu, \alpha > -1$
IV	$f(y) = h \left(1 + \frac{y^2}{\alpha^2} \right)^{-m} \exp\left(-\mu \arctg \frac{y}{\alpha}\right)$	$-\infty < y < \infty$	$\mu, \alpha, m > 0$
V	$f(y) = h y^{-m} \exp\left(-\frac{\alpha}{y}\right)$	$0 < y < \infty$	$\alpha > 0, m > 1$
VI	$f(y) = h y^{-m_1} (y - \alpha)^{m_2}$	$\alpha \leq y < \infty$	$m_1 < 1, m_2 > -1$ $m_1 > m_2 - 1$

Typ	Postać funkcji gęstości	Zakres	Parametry
VII	$f(y) = h \left(1 + \frac{y^2}{\alpha^2} \right)^{-m}$	$-\infty < y < \infty$	$m > \frac{1}{2}$
VIII	$f(y) = h \left(1 + \frac{y}{\alpha} \right)^{-m}$	$-\alpha < y \leq 0$	$m > 1$
IX	$f(y) = h \left(1 + \frac{y}{\alpha} \right)^m$	$-\alpha < y \leq 0$	$m > -1$
X	$f(y) = h \exp \left(-\frac{y-m}{\sigma} \right)$	$m \leq y < \infty$	$\sigma > 0$
XI	$f(y) = hy^{-m}$	$b \leq y < \infty$	$m > 1$
XII	$f(x) = \left(1 + \frac{y}{\alpha_1} \right)^m \left(1 - \frac{y}{\alpha_2} \right)^{-m}$	$-\alpha_1 \leq y \leq \alpha_2$	$ m < 1$

Źródło: opracowanie własne na podstawie Lahcene (2013)

Parametry krzywych Pearsona szacowane są przeważnie metodą największej wiarygodności lub momentów, jednak istotnym utrudnieniem w ich stosowaniu jest to, że w typach głównych występują po cztery parametry niezależne, trudne do interpretacji ekonomicznej.

Rozważa się również uogólniony system Pearsona, do którego należą funkcje spełniające równanie różniczkowe postaci:

$$\frac{df(y)}{dy} = g(y)f(y), \quad (2.4)$$

gdzie

$$g(y) = \frac{\sum_{j=0}^m a_j y^j}{\sum_{j=0}^n b_j y^j} \quad (2.5)$$

oraz $m, n \in N \setminus \{0\}$ i a_j, b_j są liczbami rzeczywistymi.

W zależności od przyjętych wartości parametrów otrzymano krzywe charakteryzujące się równymi własnościami (por. Lahcene, 2013; Shakil i in., 2016).

Cobb i in. (1983), przyjmując w liczniku wielomian stopnia wyższego niż jeden, a w mianowniku funkcje postaci:

- $v(y) = 1, \quad -\infty < y < \infty,$
- $v(y) = y, \quad 0 < y < \infty,$
- $v(y) = y^2, \quad 0 < y < \infty,$
- $v(y) = y(1 - y), \quad 0 < y < 1.$

rozszerzyli klasę krzywych Pearsona o rozkłady wielomodalne.

Roy (1971) analizował rozwiązania równania (2.4), gdy $m = 3, n = 2$ i $b_0 = 0$, uzyskując pięć krzywych, których parametry zależą od siedmiu pierwszych momentów. Natomiast Chaudhry i Ahmad (1993) rozważali klasę rozkładów spełniających równanie (2.4) dla $m = 4, n = 3$ i $b_0 = b_1 = b_2 = 0, b_3 \neq 0$ oraz $\frac{a_4}{2b_3} = -2\alpha, \frac{a_0}{2b_3} = 2\beta$.

Część z wymienionych rozkładów stosowana jest w analizach płac i dochodów, np. rozkład gamma (por. rozdział 1) czy rozkład t -Studenta wykorzystywany jako funkcja generująca rozkład logarytmów dochodów.

2.4. System D'Addario

Funkcja rozkładu dochodów może być także tworzona za pomocą procesu probabilistycznego, jaki jest wykorzystywany do opisu zjawisk występujących w fizyce molekularnej. System funkcji opisujących rozkłady dochodów zaproponowany przez D'Addario (1949) opiera się właśnie na formalnej analogii do rozkładu najbardziej prawdopodobnej statystyki kwantów Brillouina. Statystyka ta jest uogólnieniem statystyk kwantowych: Boltzmannna, Bosego-Einsteina i Fermi-Diraca. Ze statystyki Bosego-Einsteina można otrzymać rozkład długości fal emitowanych przez ciało doskonale czarne, będące w stałej temperaturze, otrzymany przez Plancka (prawo Plancka). Davis (1941) zaproponował model rozkładu dochodów utworzony przez analogię do tego prawa. W obszarze większych długości fal (odpowiednik wysokich dochodów) prawo Plancka przechodzi w prawo promieniowania Rayleigha-Jeansa, z którego można wyprowadzić model Pareto, natomiast dla małych długości fal (odpowiednik niskich grup dochodowych) prawo Plancka jest zbieżne do prawa Wiena, które było podstawą do utworzenia modelu Vinci'ego (1921).

Równanie Brillouina przyjmuje postać:

$$g(z) = K \left(b + e^{z^{\frac{1}{p}}} \right)^{-1}, \quad z \in (z_0, z_1), \quad (2.6)$$

gdzie $p > 0$, $b \in \mathbb{R}$, zaś K jest znormalizowaną stałą.

Jeżeli $b = 0$, to otrzymujemy równanie najbardziej prawdopodobnej statystyki Boltzmann, dla $b = -1$ równanie Bosego-Einsteina, natomiast dla $b = 1$ równanie Fermi-Diraca. D'Addario wykorzystał metodę translacji zaproponowaną przez Edgewortha (1898)¹.

Funkcja $g(z)$ jest traktowana jako funkcja generująca rozkłady dochodów.

Funkcja transformująca $\phi(y)$ spełnia równanie różniczkowe postaci:

$$\phi'(y)[\phi(y)]^q = \frac{\alpha}{y-c} \quad \text{dla } c \leq y_0 < y < \infty \quad (2.7)$$

gdzie $\alpha \neq 0$, $q \in \mathbb{R}$.

Przyjmuje ona jedną postać (por. D'Addario, 1949):

$$\phi(y) = h(y-c)^\alpha \quad \text{dla } q = -1, d > 0, \alpha \neq 0 \quad (2.8)$$

lub

$$\phi(y) = [(1+q)(\alpha \ln(y-c) + d)]^{1/(1+q)} \quad \text{dla } q \neq -1, \alpha \neq 0 \quad (2.9)$$

gdzie: d, h – stałe całkowania.

Funkcje rozkładu dochodów systemu D'Addario otrzymujemy, rozwiązując równanie różniczkowe postaci:

$$f(y) = \phi'(y) \cdot g[\phi(y)]. \quad (2.10)$$

¹ U Edgewortha rolę funkcji generującej odgrywał rozkład normalny, natomiast logarytm dochodu był funkcją transformującą. W ten sposób otrzymał on szczególny przypadek rozkładu logarytmiczno-normalnego. Jako funkcja generująca rozkłady dochodów wykorzystywana była także funkcja Laplace'a I typu, rozkład t -Studenta oraz krzywa logistyczna.

Do systemu tego należy szereg funkcji, z których wymienić należy rozkład Pareto I-go i II-go typu, krzywą Pearsona V typu zaproponowaną przez Vinci'ego, rozkład logarytmiczno-normalny, rozkład gamma uogólniony zaproponowany przez Amoroso oraz rozkład Davisa (por. tab. 2.3).

Tabela 2.3. Wybrane rozkłady dochodów należące do systemu D'Addario

Rozkład	Funkcja generująca		Funkcja transformująca			Zakres
	b	p	q	α	c	
Pareto I	0	1	0	> 0	0	$y_0 \leq y < \infty$
Pareto II	0	1	0	> 0	$\neq 0$	$c < y_0 \leq y < \infty$
Log – normalny (dwuparametrowy)	0	0,5	0	> 0	0	$0 \leq y < \infty$
Log – normalny (trójparametrowy)	0	0,5	0	> 0	$\neq 0$	$c \leq y < \infty$
Amoroso (uogólniony gamma)	0	> 0	-1	$\neq 0$	$\neq 0$	$c < y_0 \leq y < \infty$
Davisa	-1	> 0	-1	$-p$	$\neq 0$	$c < y_0 \leq y < \infty$

Źródło: Kleiber, Kotz (2003)

Krzywe należące do systemu D'Addario charakteryzują się dużą różnorodnością w porównaniu z systemami Pearsona czy Daguma, jeśli weźmiemy pod uwagę przesłanki, które odegrały rolę przy wyborze tych właśnie funkcji do analizy płac i dochodów. Do tego systemu wchodzi zarówno funkcje wybrane „ad hoc”, rozkład gamma oraz krzywa Pearsona V typu, jak i otrzymane na bazie obserwacji empirycznych (rozkład Pareto) czy będące wynikiem procesu stochastycznego (rozkład logarytmiczno-normalny). Jedynie rozkład Davisa został utworzony całkowicie przez analogię. Tym bardziej interesująca wydaje się praca D'Addario, który wykazał, że procesy powodujące powstawanie rozkładów dochodów są podobne do tych zaobserwowanych w fizyce kwantowej i dotyczących długości fal.

2.5. System Burra

System Burra, podobnie jak system krzywych Pearsona, obejmuje szeroki zakres funkcji odpowiednich do opisu różnych danych jednomodalnych.

Dystrybuanty rozkładów należące do systemu Burra powstały jako rozwiązanie równania różniczkowego postaci:

$$\frac{dF(y)}{dy} = F(y)(1 - F(y))g(y, F(y)). \quad (2.11)$$

Gdy $g(x, y) = g(x)$, to dystrybuanty można zapisać jako:

$$F(y) = \left[e^{-\int g(y)dy} + 1 \right]^{-1}. \quad (2.12)$$

Dystrybuanty i funkcje gęstości rozkładów Burra przedstawione są w tab. 2.4.

Tabela 2.4. Dystrybuanty i funkcje gęstości rozkładów Burra

Typ	Dystrybuanta	Funkcja gęstości	Zakres
I	$F(y) = y$	$f(y) = 1$	$0 < y < 1$
II	$F(y) = (e^{-y} + 1)^{-r}$	$f(y) = \frac{r(e^{-y} + 1)^{-r-1}}{e^y}$	$-\infty < y < \infty$
III	$F(y) = (y^{-c} + 1)^{-r}$	$f(y) = \frac{rc(y^{-c} + 1)^{-r-1}}{y^{c+1}}$	$0 < y < \infty$
IV	$F(y) = \left[\left(\frac{c-y}{y} \right)^{\frac{1}{c}} + 1 \right]^{-r}$	$f(y) = \frac{r}{x^2} \left[\left(\frac{c-y}{y} \right)^{\frac{1}{c}} + 1 \right]^{-r-1} \left(\frac{c-y}{y} \right)^{\frac{1}{c}-1}$	$0 < y < c$
V	$F(y) = (ce^{-\operatorname{tgy}} + 1)^{-r}$	$f(x) = \frac{rc(ke^{-\operatorname{tgy}} + 1)^{-r-1}}{\cos^2 y e^{\operatorname{tgy}}}$	$-\frac{\pi}{2} < y < \frac{\pi}{2}$
VI	$F(y) = (e^{-c \sinh y} + 1)^{-r}$	$f(y) = \frac{rc(e^{-c \sinh y} + 1)^{-r-1}}{e^{c \sinh y}} \cosh y$	$-\infty < y < \infty$
VII	$F(y) = 2^{-r} (1 + \operatorname{tgh} y)^r$	$f(y) = \frac{2^{-r} r (1 + \operatorname{tgh} y)^{r-1}}{\cosh^2 y}$	$-\infty < y < \infty$

Tabela 2.4 (cd.)

Typ	Dystrybuanta	Funkcja gęstości	Zakres
VIII	$F(y) = \left(\frac{2}{\pi} \arctg e^y \right)^r$	$f(y) = \frac{2re^y \left(\frac{2}{\pi} \arctg e^y \right)^{r-1}}{\pi(1+e^{2y})}$	$-\infty < y < \infty$
IX	$F(y) = 1 - \frac{2}{c[(1+e^y)^r - 1] + 2}$	$f(y) = \frac{2cr(1+e^y)^{r-1}e^y}{(c[(1+e^y)^r - 1] + 2)^2}$	$-\infty < y < \infty$
X	$F(y) = (1 - e^{-y^2})^r$	$f(x) = 2xr(1 - e^{-x^2})^{r-1}e^{-x^2}$	$0 \leq y < \infty$
XI	$F(y) = \left(y - \frac{1}{2\pi} \sin 2\pi y \right)^r$	$f(x) = r \left(x - \frac{1}{2\pi} \sin 2\pi x \right)^{r-1} (1 - \cos 2\pi x)$	$0 < y < 1$
XII	$F(y) = 1 - (1 + y^c)^{-r}$	$f(y) = r(1 + y^c)^{-r-1} y^{c-1}$	$0 \leq y < \infty$

Źródło: opracowanie własne na podstawie Kleiber, Kotz (2003)

Gdy na przykład $g(y) = c$, otrzymujemy rozkład typu II, gdy $g(y) = \frac{c}{y}$ – rozkład III, gdy zaś $g(y) = [(c - y)y]^{-1}$ – rozkład typu IV.

Do systemu Burra zalicza się m.in. rozkłady Daguma, Fiska i Singha-Maddali (por. Tadikamalla, 1980).

Funkcja Singha-Maddali to krzywa Burra XII typu z parametrem skali. To właśnie krzywym typu XII Burr poświęcił szczególnie dużo uwagi, uważając je za przydatne do analizy wielu zjawisk ekonomicznych. Funkcja $sech^2$ zaproponowana przez Fiska do badania dochodów to graniczny rozkład Burra XII typu.

Rozkład Burra typu III odpowiada natomiast dystrybuancie rozkładu Daguma określonej wzorem (1.74).

2.6. System Johnsona

Johnson opisał system krzywych, które reprezentują transformacje standaryzowanych krzywych normalnych (por. George, Ramachandran, 2011). Dla ciągłej zmiennej losowej Y o nieznanym rozkładzie ogólna postać transformacji jest następująca:

$$Z = \gamma + \delta g\left(\frac{Y - \xi}{\lambda}\right), \quad (2.13)$$

gdzie $g(\cdot)$ jest funkcją transformującą, Z jest zmienną losową o rozkładzie normalnym, γ i $\delta > 0$ są parametrami kształtu, $\lambda > 0$ jest parametrem skali, zaś ξ parametrem położenia.

Korzystając z funkcji odwrotnej g^{-1} , otrzymujemy:

$$Y = \xi + \lambda g^{-1}\left(\frac{Z - \gamma}{\delta}\right). \quad (2.14)$$

W zależności od postaci funkcji g Johnson zaproponował cztery typy rozkładów, określone w tab. 2.5.

Tabela 2.5. Funkcje transformujące i ich funkcje odwrotne dla rozkładów Johnsona

Typ	Funkcja $g(y)$	Funkcja $g^{-1}(z)$
S_L	$\ln y$	e^z
S_B	$\ln\left(\frac{y}{1-y}\right)$	$(1 + e^{-z})^{-1}$
S_U	$\ln(y + (1 + y^2)^{1/2})$	$\frac{e^z - e^{-z}}{2}$
S_N	y	z

Źródło: opracowanie własne

Pierwsza transformacja (S_L) zaproponowana przez Johnsona zawiera rozkłady logarytmiczno-normalne i jest postaci:

$$Z = \gamma + \delta \ln\left(\frac{Y - \xi}{\lambda}\right) = \gamma^* + \delta \ln(Y - \xi), \quad (2.15)$$

gdzie $\gamma^* = \gamma - \delta \ln \lambda$.

Zmienna losowa Y ma rozkład o funkcji gęstości:

$$f(y) = \frac{\delta}{(y - \xi)\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\gamma + \delta \ln\left(\frac{y - \xi}{\lambda}\right)\right)^2\right] \quad \text{dla } y > \xi. \quad (2.16)$$

Druga z transformacji (S_B) zdefiniowana jest następująco:

$$Z = \gamma + \delta \ln \left(\frac{Y - \xi}{\xi + \lambda - Y} \right). \quad (2.17)$$

Zatem zmienne losowe Y opisane są przez funkcje gęstości:

$$f(y) = \frac{\delta \lambda}{(y - \xi)(\xi + \lambda - y)\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left[\gamma + \delta \ln \left(\frac{y - \xi}{\xi + \lambda - y} \right) \right]^2 \right\} \quad (2.18)$$

dla $\xi < y < \xi + \lambda$.

Rozkłady uzyskane przy jej zastosowaniu są określone na zbiorze ograniczonym z dołu lub góry albo i z dołu, i z góry. Do tej grupy rozkładów należą m.in. rozkłady gamma i beta.

Trzecia transformacja Johnsona (S_U) jest postaci:

$$Z = \gamma + \delta \ln \left\{ \left(\frac{Y - \xi}{\lambda} \right) + \left[1 + \left(\frac{Y - \xi}{\lambda} \right)^2 \right]^{1/2} \right\} = \gamma + \delta \sinh^{-1} \left(\frac{Y - \xi}{\lambda} \right), \quad (2.19)$$

a funkcja gęstości zmiennej Y wyraża się wzorem:

$$f(y) = \frac{\delta \lambda}{(y - \xi)(\xi + \lambda - y)\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left[\gamma + \delta \ln \left[\left(\frac{y - \xi}{\lambda} \right) + \left(1 + \left(\frac{y - \xi}{\lambda} \right)^2 \right)^{1/2} \right] \right]^2 \right\} \quad (2.20)$$

dla $-\infty < y < \infty$

Krzywe z tej grupy są określone na zbiorze nieograniczonym, np. rozkład t -Studenta.

W przypadku ostatniej transformacji:

$$Z = \gamma + \delta \left(\frac{Y - \xi}{\lambda} \right) \quad (2.21)$$

otrzymujemy rozkład normalny:

$$f(y) = \frac{\delta}{\lambda\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left[y + \delta \left(\frac{y-\xi}{\lambda} \right) \right]^2 \right\} \quad \text{dla } y > \xi, \quad (2.22)$$

o wartości oczekiwanej $E(Y) = \xi - \gamma \frac{\lambda}{\delta}$ i odchyleniu standardowym $D(Y) = \frac{\lambda}{\delta}$.

2.7. Rodzina uogólnionego rozkładu beta drugiego rodzaju

Uogólniony rozkład beta drugiego rodzaju (GB2) jest rozkładem czteroparametrowym, który powstał przez dodanie parametru skali i trzeciego parametru kształtu do funkcji gęstości rozkładu beta (Gini, 1911), w celu jej uelastycznienia. Po raz pierwszy został zaproponowany w kontekście badania rozkładów dochodów przez Mc Donalda (1984) w pracy poświęconej uogólnionym rozkładom stosowanym do badania dochodów, gdzie oprócz rozkładu GB2 przedstawiony został także uogólniony rozkład gamma. Obie te rodziny zawierają wiele użytecznych rozkładów, jako przypadki szczególne lub graniczne, z których większość znalazła zastosowanie w analizach rozkładów płac, dochodów czy zamożności.

Dystrybuantę uogólnionego rozkładu beta drugiego rodzaju otrzymujemy z dystrybuanty rozkładu beta (znanego także w literaturze jako *rozkład beta drugiego rodzaju*, *beta prime* lub *odwrotny rozkład beta*) postaci:

$$F(z) = \begin{cases} \frac{z}{B(p,q)} \int_0^z \frac{t^{p-1}}{(1+t)^{p+q}} dt & \text{dla } z > 0, \\ 0 & \text{dla } z \leq 0. \end{cases} \quad (2.23)$$

poprzez podstawienie $z = \left(\frac{y}{b} \right)^a$.

Funkcja gęstości uogólnionego rozkładu beta drugiego rodzaju z dodatnimi parametrami a, b, p i q , czyli GB2 (a, b, p, q) przyjmuje postać:

$$f(y) = \begin{cases} \frac{ay^{ap-1}}{b^{ap}B(p,q)\left[1+\left(\frac{y}{b}\right)^a\right]^{p+q}} & \text{dla } y > 0, \\ 0 & \text{dla } y \leq 0. \end{cases} \quad (2.24)$$

Parametr b jest parametrem skali rozkładu GB2, natomiast pozostałe parametry: a, p i q są parametrami kształtu. Parametr a pełni podobną funkcję w rozkładzie GB2 jak stała Pareto, czyli odpowiada za grubość ogona rozkładu (por. rys. 2.1). Dokładniej: czym większa wartość tego parametru, tym cieńszy ogon rozkładu, a więc większa jest liczba skończonych momentów rozkładu.

Parametr p kontroluje lewy ogon rozkładu (por. rys. 2.2), parametr q dodatkowo kontroluje prawy ogon, co można zaobserwować na rys. 2.3–2.4. Parametry p i q razem odpowiadają za asymetrię rozkładu.

Momenty zwykłe rzędu r dla rozkładu GB2 istnieją tylko dla $-ap < r < aq$ i są określone następującym wzorem:

$$E(Y^r) = \frac{b^r B\left(p + \frac{r}{a'}, q - \frac{r}{a}\right)}{B(p, q)}, \quad (2.25)$$

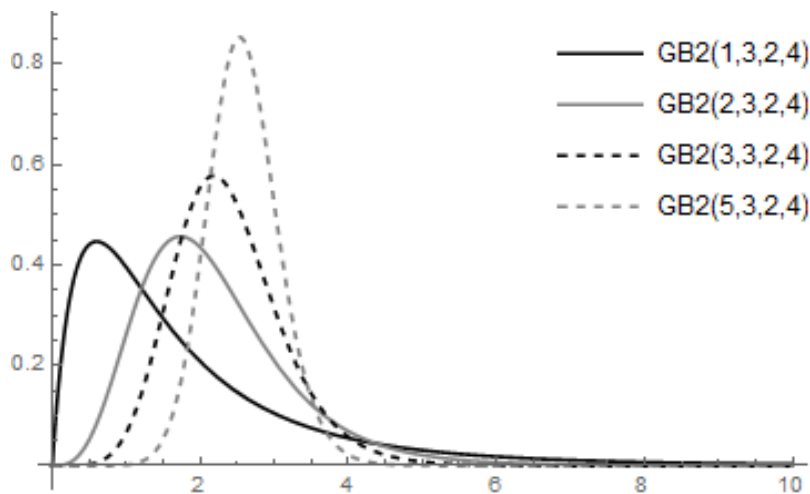
w szczególności wartość oczekiwana wynosi:

$$E(Y) = \frac{bB\left(p + \frac{1}{a'}, q - \frac{1}{a}\right)}{B(p, q)}, \quad (2.26)$$

a wariancja:

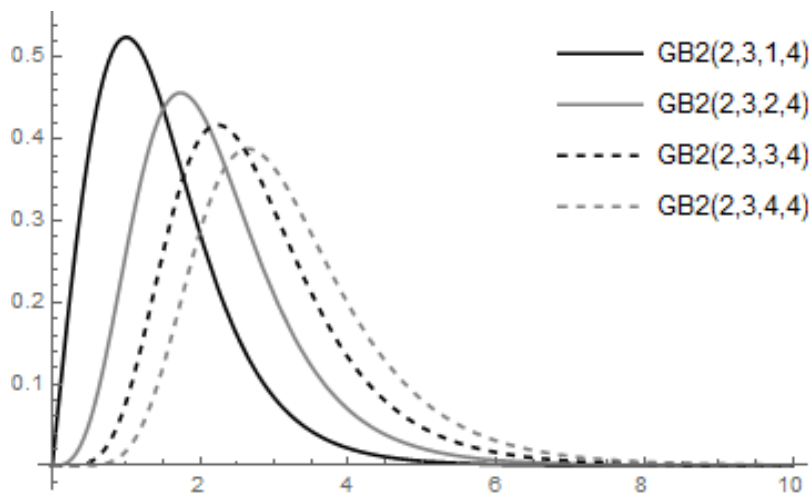
$$D^2(Y) = b^2 \frac{B\left(p + \frac{2}{a'}, q - \frac{2}{a}\right)B(p, q) - B^2\left(p + \frac{1}{a'}, q - \frac{1}{a}\right)}{B^2(p, q)}. \quad (2.27)$$

Rozkład GB2 oprócz swojej niezwyklej elastyczności ma szereg interesujących własności, które umożliwiają wykorzystywanie go do różnych transformacji zmiennych losowych i są przydatne w procesie wnioskowania statystycznego.



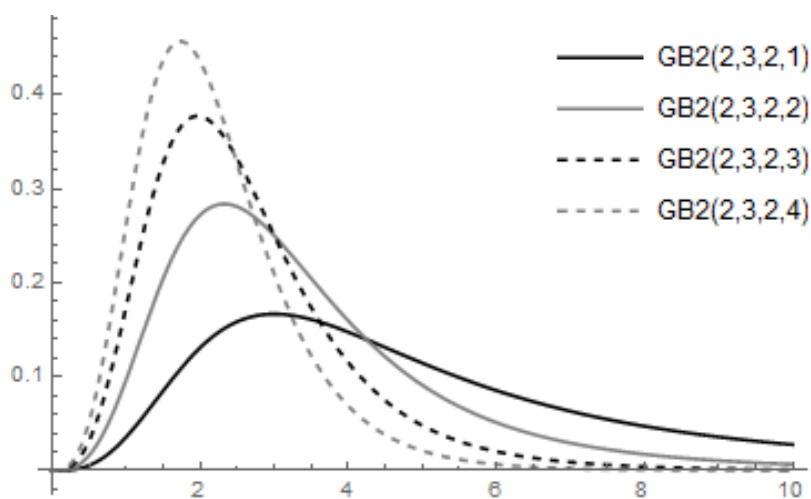
Rysunek 2.1. Gęstości rozkładów GB2 dla $b = 3, p = 2, q = 4$ i wybranych wartości a

Źródło: opracowanie własne



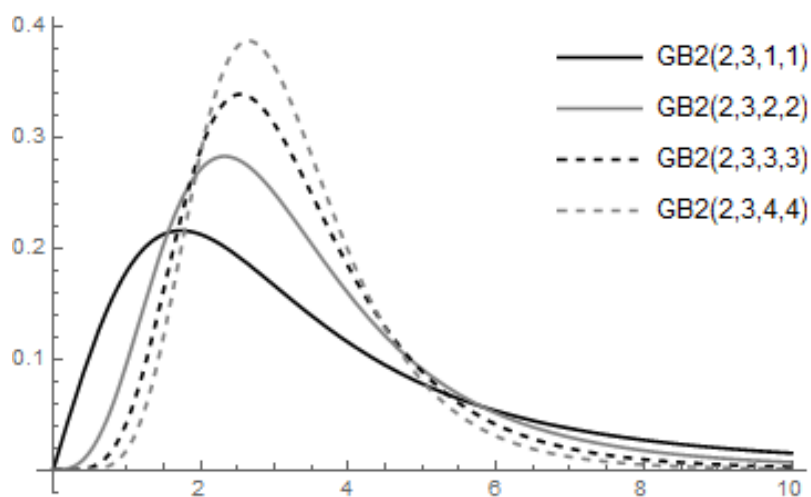
Rysunek 2.2. Gęstości rozkładów GB2 dla $a = 2, b = 3, q = 4$ i wybranych wartości p

Źródło: opracowanie własne



Rysunek 2.3. Gęstości rozkładów GB2 dla $a = 2$, $b = 3$, $p = 2$ i wybranych wartości q

Źródło: opracowanie własne



Rysunek 2.4. Gęstości rozkładów GB2 dla $a = 2$, $b = 3$ i wybranych wartości p, q

Źródło: opracowanie własne

W tab. 2.6 przedstawione zostały najważniejsze i najczęściej stosowane rozkłady, które są szczególnymi przypadkami uogólnionego rozkładu beta drugiego rodzaju.

Tabela 2.6. Rozkłady należące do rodziny GB2

Symbol rozkładu	Nazwa rozkładu	Związek z GB2 (a, b, p, q)
B2 (b, p, q)	Beta drugiego rodzaju	GB2 ($1, b, p, q$)
Fisher (u, v)	Fishera	GB2 ($1, v/u, u/2, v/2$)
S-M (q, a, b)	Singh-Maddala (Burr XII typu)	GB2 ($a, b, 1, q$)
D (a, v, b)	Dagum (Burr III typu)	GB2 ($v, b, a, 1$)
L (b, q)	Lomaxa	GB2 ($1, b, 1, q$)
IL (b, q)	Odwrotny Lomaxa	GB2 ($1, b, p, 1$) S – M ($a = -1$) = Dagum ($v = 1$)
F (a, b)	Fiska	GB2 ($a, b, 1, 1$) S – M ($q = 1$) = Dagum ($a = 1$)
LL (b)	Logarytmiczno-logistyczny	GB2 ($1, b, 1, 1$) = Fisk ($a = 1$)

Źródło: opracowanie własne

Łatwo zauważyć, że rodzina GB2 zawiera bardzo różne funkcje gęstości, które w zależności od parametrów mogą przyjmować różne kształty, jednak większość z nich może być stosowana w badaniach rozkładów dochodów według wysokości oraz innych rozkładów tego typu, jak np. rozkłady wykorzystywane w badaniach finansowych i aktuarialnych (por. Kleiber, Kotz, 2003). Przedstawione na rysunkach 2.1–2.4 różne kształty funkcji gęstości rozkładu GB2 pozwalają dodatkowo wyjaśnić złożoność rozkładu, jego elastyczność i wpływ poszczególnych parametrów na jego asymetrię, spłaszczenie i grubość ogonów.

ROZDZIAŁ 3

Metody estymacji parametrów rozkładu

3.1. Uwagi wstępne

Do wyznaczania estymatorów parametrów rozkładu zmiennej losowej wykorzystuje się zarówno parametryczne, jak i nieparametryczne metody estymacji. Do najpopularniejszych należą: metoda największej wiarygodności, momentów i metoda najmniejszych kwadratów. Istnieją jednak ograniczenia w wykorzystywaniu tych metod ze względu na brak możliwości wyznaczenia maksimum funkcji wiarygodności czy brak momentów centralnych odpowiednich rzędów, a to sprawia, że poszukuje się innych procedur szacowania parametrów rozkładu analizowanych zmiennych. Bardziej uniwersalne w zastosowaniu są metody oparte na percentylach i metoda momentów ważonych prawdopodobieństwami.

Wybór metody estymacji wpływa na własności estymatorów, ich obciążoność, zgodność i efektywność, dlatego w praktycznym zastosowaniu istotne są informacje dotyczące własności otrzymywanych estymatorów.

W rozdziale tym przedstawione są metody estymacji, które mogą być stosowane przy szacowaniu parametrów rozkładów wykorzystywanych w analizach dochodów, płac i wydatków gospodarstw domowych. W przypadku każdej z procedur szacowania parametrów podane są przykłady zastosowania dla wybranych rozkładów teoretycznych.

3.2. Metoda największej wiarygodności

Do szacowania parametrów rozkładu często wykorzystuje się metodę największej wiarygodności (MNW), której zastosowanie związane jest z wyznaczeniem funkcji wiarygodności lub logarytmu funkcji wiarygodności, a następnie maksimum globalnego.

Oznaczmy przez Y ciągłą zmienną losową o rozkładzie określonym przez funkcję gęstości $f(y; \theta_1, \theta_2, \dots, \theta_k)$, gdzie $\theta_1, \theta_2, \dots, \theta_k$ są parametrami rozkładu tej zmiennej. Funkcja wiarygodności wyznaczona na podstawie n -elementowej próby prostej Y_1, Y_2, \dots, Y_n wyraża się wzorem:

$$L(Y_1, Y_2, \dots, Y_n) = \prod_{i=1}^n f(Y_i; \theta_1, \theta_2, \dots, \theta_k), \quad (3.1)$$

zaś funkcja logarytmu ilorazu wiarygodności ma postać:

$$\ln L(Y_1, Y_2, \dots, Y_n) = \sum_{i=1}^n \ln f(Y_i; \theta_1, \theta_2, \dots, \theta_k). \quad (3.2)$$

Estymatory parametrów rozkładu $\theta_1, \theta_2, \dots, \theta_k$ wyznacza się, rozwiązując układ k równań:

$$\begin{cases} \frac{\partial L}{\partial \theta_1} = 0, \\ \dots \\ \frac{\partial L}{\partial \theta_k} = 0 \end{cases} \quad (3.3)$$

lub

$$\begin{cases} \frac{\partial \ln L}{\partial \theta_1} = 0, \\ \dots \\ \frac{\partial \ln L}{\partial \theta_k} = 0, \end{cases} \quad (3.4)$$

a następnie sprawdzając warunek dostateczny istnienia ekstremum.

Metodę największej wiarygodności można wykorzystać m.in. do estymacji parametrów rozkładu logarytmiczno-normalnego $LN(\mu, \sigma)$. Estymatory parametrów tego rozkładu wyznaczone w oparciu o próbę losową Y_1, Y_2, \dots, Y_n wyrażają się wzorami:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n \ln Y_i, \quad (3.5)$$

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (\ln Y_i - \hat{\mu})^2. \quad (3.6)$$

Obydwa estymatory są nieobciążone i najbardziej efektywne, a ich wariancje dla dużych n są równe, odpowiednio:

$$D^2(\hat{\mu}) = \frac{\sigma^2}{n} \text{ oraz } D^2(\hat{\sigma}^2) = \frac{2\sigma^4}{n}.$$

W przypadku danych pogrupowanych konieczne są jednak modyfikacje wzorów (3.5) i (3.6) (por. Kordos, 1973):

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^m g_i \ln \bar{Y}_i, \quad (3.7)$$

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^m (g_i \ln \bar{Y}_i - \hat{\mu})^2, \quad (3.8)$$

gdzie:

$$\ln \bar{Y}_i = \frac{\ln Y_{i-1} - \ln Y_i}{2},$$

Y_{i-1} – dolna granica i -tej grupy,

Y_i – górna granica i -tej grupy,

g_i – frakcja jednostek w i -tej grupie,

m – liczba grup.

Z wykorzystaniem tej metody do estymacji parametrów rozkładu logarytmiczno-normalnego należy uważać w przypadku danych pogrupowanych, gdy skrajne grupy są nieograniczone i liczne, gdyż wtedy możemy popełnić duży błąd przy obliczaniu średnich grupowych, a tym samym otrzymane estymatory będą charakteryzowały się dużymi błędami średniokwadratowymi.

Zastosowanie metody największej wiarygodności do estymacji parametrów rozkładu gamma $G(\alpha, \lambda)$ na podstawie n -elementowej próby prostej Y_1, Y_2, \dots, Y_n związane jest z rozwiązaniem układu równań:

$$\begin{cases} \lambda = \frac{\alpha}{\bar{Y}}, \\ \ln \alpha - \Psi(\alpha) = \ln \bar{Y} - \ln \tilde{Y}, \end{cases} \quad (3.9)$$

gdzie \bar{Y} jest średnią arytmetyczną z próby, \tilde{Y} jest średnią geometryczną z próby, zaś $\Psi(\alpha)$ jest funkcją digamma.

Drugie równanie z układu (3.9) może być rozwiązane graficznie lub interpolacyjnie po obliczeniu z próby ilorazu średniej arytmetycznej i geometrycznej. Oszacowany w ten sposób parametr α wstawiamy następnie do równania pierwszego i otrzymujemy oszacowanie parametru λ .

Zastosowanie metody największej wiarygodności do szacowania parametrów rozkładu Pareto $Pa(\theta, a)$ o dystrybucie określonej wzorem (1.5), na podstawie próby losowej Y_1, Y_2, \dots, Y_n , związane jest z przyjęciem minimum z próby jako estymatora parametru θ , a następnie wyznaczeniem estymatora parametru a .

Wzory na estymatory parametrów rozkładu Pareto są następujące:

$$\hat{\theta} = \min_{1 \leq i \leq n} Y_i, \quad (3.10)$$

$$\hat{a} = \frac{n}{\sum_{i=1}^n \ln \left(\frac{Y_i}{\hat{\theta}} \right)}. \quad (3.11)$$

Metodę największej wiarygodności można też wykorzystać do estymacji parametrów rozkładu Daguma $D(a, \nu, b)$ (por. Domma i in., 2011; Domański, Jędrzejczak, 1998), przy czym liczebność próby, na podstawie której szacujemy parametry, musi być co najmniej 1000. W tym przypadku funkcja wiarygodności ma postać:

$$L(Y_1, Y_2, \dots, Y_n) = \left(\frac{a\nu}{b} \right)^n \prod_{i=1}^n \left(\frac{Y_i}{b} \right)^{a\nu-1} \left[1 + \left(\frac{Y_i}{b} \right)^\nu \right]^{-a-1}. \quad (3.12)$$

Niestety, w tym przypadku, aby znaleźć wartości estymatorów para-

$$\text{metrów } a, v, b, \text{ należy numerycznie rozwiązać układ równań: } \begin{cases} \frac{\partial L}{\partial a} = 0 \\ \frac{\partial L}{\partial v} = 0 \\ \frac{\partial L}{\partial b} = 0 \end{cases}.$$

3.3. Metoda momentów

Do szacowania parametrów zmiennej losowej Y o rozkładzie określonym przez funkcję gęstości $f(y; \theta_1, \theta_2, \dots, \theta_k)$ metodą momentów (MM) wykorzystuje się k momentów zwykłych rozkładu i ich estymatorów.

Momenty zwykłe $\mu^{(r)}$ rozkładu są funkcjami parametrów $\theta_1, \theta_2, \dots, \theta_k$, natomiast ich estymatory wyrażają się wzorami:

$$m^{(r)} = \frac{1}{n} \sum_{i=1}^n Y_i^r, \quad r = 1, 2, \dots, \quad (3.13)$$

gdzie Y_1, Y_2, \dots, Y_n jest n -elementową próbą prostą.

Estymatory parametrów rozkładu $f(y; \theta_1, \theta_2, \dots, \theta_k)$ wyznacza się, rozwiązując układ k równań postaci:

$$\begin{cases} \mu^{(1)} = m^{(1)}, \\ \dots \\ \mu^{(k)} = m^{(k)}, \end{cases} \quad (3.14)$$

Zastosowanie tej metody związane jest z istnieniem momentów do rzędu k -tego włącznie.

Metodę momentów można wykorzystać między innymi do szacowania parametrów niektórych grup rozkładów Pareto i Daguma oraz rozkładu logarytmiczno-normalnego.

Estymatory parametrów rozkładu $Pa(\theta, a)$ gdy istnieje wartość oczekiwana i wariancja rozkładu, czyli gdy $a > 2$, mają postaci:

$$\hat{a} = 1 + \frac{1}{DY} \sqrt{\bar{Y}^2 + D^2 Y}, \quad (3.15)$$

$$\hat{\theta} = \frac{\bar{Y}\sqrt{\bar{Y}^2 + D^2Y}}{DY + \sqrt{\bar{Y}^2 + D^2Y}}, \quad (3.16)$$

gdzie $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$ oraz $D^2Y = \frac{1}{n} \sum_{i=1}^n Y_i^2 - \left(\frac{1}{n} \sum_{i=1}^n Y_i \right)^2$ są wyznaczone na podstawie n -elementowej próby prostej Y_1, Y_2, \dots, Y_n .

W przypadku szacowania parametrów rozkładu $D(a, v, b)$ zastosowanie metody momentów jest możliwe, gdy $v > 3$. Wtedy wartości estymatorów otrzymujemy z numerycznego rozwiązania układu równań:

$$\begin{cases} b \frac{\Gamma\left(1 - \frac{1}{v}\right) \Gamma\left(a + \frac{1}{v}\right)}{\Gamma(a)} = \bar{y}, \\ b^2 \frac{\Gamma\left(1 - \frac{2}{v}\right) \Gamma\left(a + \frac{2}{v}\right)}{\Gamma(a)} = \frac{1}{n} \sum_{i=1}^n y_i^2, \\ b^3 \frac{\Gamma\left(1 - \frac{3}{v}\right) \Gamma\left(a + \frac{3}{v}\right)}{\Gamma(a)} = \frac{1}{n} \sum_{i=1}^n y_i^3, \end{cases} \quad (3.17)$$

gdzie y_1, y_2, \dots, y_n są realizacjami próby prostej Y_1, Y_2, \dots, Y_n , zaś \bar{y} jest średnią arytmetyczną obliczoną na podstawie tej próby.

Dla rozkładu $LN(\mu, \sigma)$ estymatory parametrów otrzymane metodą momentów wyrażają się wzorami:

$$\hat{\mu} = 2 \ln \bar{Y} - \frac{1}{2} \ln \left(\frac{1}{n} \sum_{i=1}^n Y_i^2 \right), \quad (3.18)$$

$$\hat{\sigma}^2 = \ln \left(\frac{1}{n} \sum_{i=1}^n Y_i^2 \right) - 2 \ln \bar{Y}, \quad (3.19)$$

gdzie $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$.

Obydwa estymatory są zgodne i bardzo efektywne dla małych wartości σ^2 , jednak ich efektywność gwałtownie spada wraz ze wzrostem σ^2 . Dodatkowe problemy pojawiają się, gdy chcemy korzystać z danych pogrupowanych. Konieczne jest wtedy stosowanie tzw. poprawek Shepparda.

Mogą także występować trudności związane z obliczaniem średniej arytmetycznej dla grup skrajnych – podobnie jak przy zastosowaniu metody największej wiarygodności.

Metodę momentów można też wykorzystać do estymacji parametrów rozkładu Zengi. Zenga i in. (2010c) otrzymali formuły dla estymatorów parametrów rozkładu Zengi poprzez rozwiązanie następującego układu równań:

$$\begin{cases} \mu = \bar{Y}, \\ \frac{\bar{Y}^2 \theta (\theta + 1)}{3(\alpha - 1)(\alpha + \theta)} = \frac{1}{n} \sum_{i=1}^n Y_i^2 - \bar{Y}^2, \\ \frac{3m_2 \bar{Y} (\theta + 3)(\theta + 2)}{5(\alpha + \theta + 1)(\alpha - 2)} = \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^3, \end{cases} \quad (3.20)$$

gdzie Y_1, Y_2, \dots, Y_n jest n -elementową próbą prostą oraz $m_2 = \frac{1}{n} \sum_{i=1}^n Y_i^2$.

Oczywiście, aby zastosować tę metodę estymacji, parametr α musi być większy od 2. Wyznaczone z układu równań (3.20) estymatory parametrów mają postaci:

$$\hat{\mu} = \bar{Y}, \quad (3.21)$$

$$\begin{aligned} \hat{\theta} = & \frac{-\left[\frac{\bar{Y}^2}{3\alpha_2} - 3\frac{\bar{Y}\alpha_2}{m_3} - 1\right]}{2\left[\frac{1}{3}\frac{\bar{Y}^2}{\alpha_2} - \frac{3}{5}\bar{Y}\frac{\alpha_2}{m_3}\right]} + \\ & + \frac{\sqrt{\left[\frac{1}{3}\frac{\bar{Y}^2}{\alpha_2} - 3\frac{\bar{Y}\alpha_2}{m_3} - 1\right]^2 + 4\left[\frac{1}{3}\frac{\bar{Y}^2}{\alpha_2} - \frac{3}{5}\bar{Y}\frac{\alpha_2}{m_3}\right]\left[\frac{18}{5}\frac{\bar{Y}\alpha_2}{m_3} + 2\right]}}{2\left[\frac{1}{3}\frac{\bar{Y}^2}{\alpha_2} - \frac{3}{5}\bar{Y}\frac{\alpha_2}{m_3}\right]}, \end{aligned} \quad (3.22)$$

$$\hat{\alpha} = \frac{-\left(\hat{\theta} - 1\right) + \sqrt{\left(\hat{\theta} - 1\right)^2 + 4\left[\frac{1}{3}\frac{\bar{Y}^2}{\alpha_2}\hat{\theta}(\hat{\theta} + 1) + \hat{\theta}\right]}}{2}, \quad (3.23)$$

gdzie $\alpha_2 = \frac{1}{n} \sum_{i=1}^n Y_i^2 - \bar{Y}^2$, $m_3 = \frac{1}{n} \sum_{i=1}^n Y_i^3$.

Parametry innych rozkładów, np. Singha-Maddali czy Weibulla, można również wyznaczyć metodą momentów określając, odpowiednio, trzy lub dwa momenty rozkładu, odpowiednio, z wzorów (1.93) i (1.108) oraz porównując z momentami empirycznymi.

3.4. Metoda momentów ważonych prawdopodobieństwami

Metoda momentów ważonych prawdopodobieństwami (*MMWP*) jest procedurą estymacji parametrów wykorzystywaną w przypadku szacowania parametrów $\theta_1, \theta_2, \dots, \theta_k$ rozkładu określonego za pomocą dystrybucyj $F(y; \theta_1, \theta_2, \dots, \theta_k)$, gdy nie istnieje k momentów zwykłych. Wtedy można wykorzystać tzw. momenty rozkładu ważone prawdopodobieństwami (por. m.in. Furrer, Naveau, 2007; Rasool i in., 2002).

Momentami rozkładu zmiennej losowej Y ważonymi prawdopodobieństwami (o ile istnieją) nazywamy wyrażenia postaci:

$$M_{1,j,0} = E(YF^j(y; \theta_1, \dots, \theta_k)), \quad (3.24)$$

$$M_{1,0,s} = E(Y(1-F(y; \theta_1, \dots, \theta_k))^s), \quad (3.25)$$

dla $j, s = 0, 1, 2, \dots$

Momentami z próby Y_1, Y_2, \dots, Y_n ważonymi prawdopodobieństwami nazywamy następujące statystyki:

$$m_{1,j,0} = \frac{1}{n} \sum_{i=1}^n Y_{i:n} [F_n(y_{i:n}; \theta_1, \dots, \theta_k)]^j, \quad (3.26)$$

$$m_{1,0,s} = \frac{1}{n} \sum_{i=1}^n Y_{i:n} [1-F_n(y_{i:n}; \theta_1, \dots, \theta_k)]^s, \quad (3.27)$$

gdzie $Y_{i:n}$ są statystykami pozycyjnymi wyznaczonymi na podstawie n -elementowej próby losowej, natomiast $F_n(y_{i:n}; \theta_1, \dots, \theta_k)$ to wartość dystrybucyj empirycznej dla i -tej statystyki pozycyjnej, określająca prawdopodobieństwa nieprzekroczenia wartości $y_{i:n}$ statystyki $Y_{i:n}$ dla $i = 0, 1, 2, \dots, n$.

Estymatorami parametrów $\theta_1, \dots, \theta_k$ otrzymanymi metodą momentów ważonych prawdopodobieństwami nazywamy statystyki będące rozwiązaniem układu k równań postaci $m_{1,j,0} = M_{1,j,0}$ lub $m_{1,0,s} = M_{1,0,s}$ dla wybranych j, s .

Metoda MMWP, podobnie jak metoda momentów (MM), ma ograniczenia w zastosowaniu. Można ją wykorzystywać do estymacji parametrów rozkładów mających momenty wyrażone wzorami (3.24) i (3.25).

Nieobciążonymi estymatorami momentów rozkładu $M_{1,j,0}$ i $M_{1,0,s}$ są statystyki:

$$m_{1,0,0} = \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_{i:n}, \quad (3.28)$$

$$m_{1,j,0} = \frac{1}{n} \sum_{i=1}^n \frac{(i-1)(i-2)\dots(i-j)}{(n-1)(n-2)\dots(n-j)} Y_{i:n} \quad \text{dla } j = 1, 2, \dots \quad (3.29)$$

$$m_{1,0,s} = \frac{1}{n} \sum_{i=1}^n \frac{(n-i)(n-i-1)\dots(n-i-s+1)}{(n-1)(n-2)\dots(n-s)} Y_{i:n} \quad \text{dla } s = 1, 2, \dots \quad (3.30)$$

Metoda momentów ważonych prawdopodobieństwami może znaleźć zastosowanie na przykład w estymacji rozkładu Pareto, gdy $a \leq 2$. Wtedy istnieje tylko wartość oczekiwana ($1 < a \leq 2$) lub rozkład nie posiada obydwu momentów zwykłych ($a \leq 1$) i zastosowanie metody momentów jest niemożliwe.

Dla rozkładu Pareto $Pa(\theta, a)$, gdy $1 < a \leq 2$, estymatory parametrów θ i a otrzymujemy z układu równań:

$$\begin{cases} \frac{\theta a}{a-1} = \bar{Y}, \\ \frac{\theta a}{2a-1} = \frac{1}{n} \sum_{i=1}^n \frac{n-i}{n-1} Y_{i:n}, \end{cases} \quad (3.31)$$

gdzie $Y_{i:n}$ jest i -tą statystyką pozycyjną z próby losowej Y_1, Y_2, \dots, Y_n i są one postaci:

$$\hat{a} = \frac{n\bar{Y} - \sum_{i=1}^n \frac{n-i}{n-1} Y_{i:n}}{n\bar{Y} - 2 \sum_{i=1}^n \frac{n-i}{n-1} Y_{i:n}}, \quad (3.32)$$

$$\hat{\theta} = \frac{\bar{Y} \sum_{i=1}^n \frac{n-i}{n-1} Y_{i:n}}{n\bar{Y} - \sum_{i=1}^n \frac{n-i}{n-1} Y_{i:n}}. \quad (3.33)$$

Postaci estymatorów dla rozkładu Pareto, gdy $a \leq 1$, otrzymujemy z układu równań:

$$\begin{cases} \frac{\theta a}{(k+1)(a-1)} = m_{1,0,k}, \\ \frac{\theta a}{(j+1)(2a-1)} = m_{1,0,j} \end{cases} \quad (3.34)$$

i wyrażają się one wzorami:

$$\hat{a} = \frac{m_{1,0,k} - m_{1,0,j}}{(k+1)m_{1,0,k} - (j+1)m_{1,0,j}}, \quad (3.35)$$

$$\hat{\theta} = \frac{(l-k)m_{1,0,k}m_{1,0,j}}{m_{1,0,k} - m_{1,0,j}}. \quad (3.36)$$

Metodę momentów ważonych prawdopodobieństwami można też wykorzystać do estymacji parametrów rozkładu Daguma $D(a, \nu, b)$. W tym przypadku momenty ważne prawdopodobieństwami (o ile istnieją) są postaci:

$$M_{1,j,0} = E(YF^j(y, \theta_1, \dots, \theta_k)) = \lambda \frac{\Gamma\left[(j+1)a + \frac{1}{\nu}\right] \Gamma\left(1 - \frac{1}{\nu}\right)}{(j+1)\Gamma((j+1)a)} \quad (3.37)$$

Jeśli $\nu > 1$, estymatory parametrów rozkładu Daguma możemy wyznaczyć numerycznie z układu równań:

$$\begin{cases} b \frac{\Gamma\left(1 - \frac{1}{\nu}\right) \Gamma\left(a + \frac{1}{\nu}\right)}{\Gamma(a)} = \bar{Y} \\ b \frac{\Gamma\left(2a + \frac{1}{\nu}\right) \Gamma\left(1 - \frac{1}{\nu}\right)}{2\Gamma(2a)} = \frac{1}{n} \sum_{i=1}^n \frac{i-1}{n-1} Y_{i:n} \\ b \frac{\Gamma\left(3a + \frac{1}{\nu}\right) \Gamma\left(1 - \frac{1}{\nu}\right)}{3\Gamma(3a)} = \frac{1}{n} \sum_{i=1}^n \frac{(i-1)(i-2)}{(n-1)(n-2)} Y_{i:n}. \end{cases} \quad (3.38)$$

3.5. Metoda percentyli

Metoda percentyli (MP) posiada dość prostą konstrukcję i może być stosowana przy szacowaniu parametrów każdego rozkładu opisanego za pomocą dystrybuanty (por. Bartoszewicz, 1996; Wywił, 2004).

Przy wyznaczaniu metodą percentyli estymatorów nieznanych parametrów rozkładu zmiennej losowej Y wykorzystuje się definicję percentyla rozkładu i wartości percentyli z próby wybranych rzędów.

Niech Q_p będzie percentylem rzędu p rozkładu zmiennej Y o dystrybuancie $F(y; \theta_1, \theta_2, \dots, \theta_k)$, gdzie $p \in (0, 1)$. Z definicji percentyla rozkładu Q_p rzędu p wynika, że $F(Q_p; \theta_1, \theta_2, \dots, \theta_k) = p$, czyli w celu wyznaczenia estymatorów k parametrów tworzymy układ k równań postaci:

$$\begin{cases} F(Q_{p_1}; \theta_1, \theta_2, \dots, \theta_k) = p_1, \\ F(Q_{p_2}; \theta_1, \theta_2, \dots, \theta_k) = p_2, \\ \dots \\ F(Q_{p_k}; \theta_1, \theta_2, \dots, \theta_k) = p_k, \end{cases} \quad (3.39)$$

gdzie $p_i, p_j \in (0, 1), p_i \neq p_j$ dla $i \neq j$ oraz $i, j \in \{1, 2, \dots, k\}$.

Podstawienie do układu równań w miejsce kwantyli rozkładu rzędu p_i ($i = 1, 2, \dots, k$) wartości kwantyli z próby Y_{p_i} i rozwiązanie tego układu prowadzi do otrzymania estymatorów parametrów $\theta_1, \theta_2, \dots, \theta_k$.

Dla zmiennych losowych o rozkładach ciągłych z dwoma parametrami wykorzystuje się percentyle z próby rzędu p i $1 - p$.

Wadą metody kwantyli jest brak informacji o rzędach percentyli, które należy wybrać, aby obciążenia i błędy średniokwadratowe uzyskanych estymatorów były niewielkie. Okazuje się, że dla różnych klas rozkładów zmiennych losowych rzędy kwantyli pozwalające otrzymać estymatory o „dobrych” własnościach są różne. Dlatego istotna jest wstępna analiza, pozwalająca sformułować wnioski dotyczące wyboru rzędów kwantyli z próby.

Metoda kwantyli jest wykorzystywana do oceny parametrów μ i σ^2 rozkładu logarytmiczno-normalnego. Pozwala ona na oszacowanie tych parametrów z danych indywidualnych i pogrupowanych. Poza tym cechuje ją stosunkowo prosta procedura obliczeniowa.

Wybierając dwa kwantyle z n -elementowej próby: Y_p oraz Y_{1-p} , otrzymujemy układ równań:

$$\begin{cases} Y_p = e^{\mu + u_p \sigma}, \\ Y_{1-p} = e^{\mu + u_{(1-p)} \sigma}, \end{cases} \quad (3.40)$$

gdzie $\mu_p, \mu_{(1-p)}$ są kwantylami rozkładu normalnego standaryzowanego rzędu, odpowiednio, p oraz $1-p$. Jego rozwiązanie prowadzi do estymatorów postaci:

$$\hat{\mu} = \frac{u_p \ln Y_{1-p} - u_{(1-p)} \ln Y_p}{u_p - u_{(1-p)}}, \quad (3.41)$$

$$\hat{\sigma} = \frac{\ln Y_p - \ln Y_{1-p}}{u_p - u_{(1-p)}}. \quad (3.42)$$

Wielkości rzędów kwantyli, które należy stosować w estymacji parametrów rozkładu logarytmiczno-normalnego, określone są w pracy Aitchisona, Browna (1957). Okazuje się, że największą efektywność oceny parametru μ uzyskuje się, stosując kwantyle rzędów 0,27 i 0,73, natomiast największą efektywność estymatora parametru σ można otrzymać, wykorzystując kwantyle 0,07 i 0,93.

Estymatorami parametrów rozkładu Pareto $Pa(\theta, a)$ stosowanego w analizach dochodów grup zamożnych, uzyskanymi metodą percentyli, są statystyki:

$$\hat{\theta} = Y_{1-p} \cdot p^{\frac{\ln Y_{1-p} - \ln Y_p}{\ln(1-p) - \ln p}}, \quad (3.43)$$

$$\hat{a} = \frac{\ln(1-p) - \ln p}{\ln Y_{1-p} - \ln Y_p}. \quad (3.44)$$

W przypadku rozkładu Weibulla $W(\alpha, k)$, który również wykorzystywany jest w analizach dochodów i wydatków gospodarstw domowych, estymatory parametrów otrzymane metodą percentyli mają postaci:

$$\hat{\alpha} = \frac{\ln(-\ln(1-p)) - \ln(-\ln p)}{\ln Y_p - \ln Y_{1-p}}, \quad (3.45)$$

$$\hat{k} = Y_p(-\ln(1-p)) \frac{\ln Y_p - \ln Y_{1-p}}{\ln(-\ln(1-p)) - \ln(-\ln p)}. \quad (3.46)$$

W pracach Pekasiewicz (2012; 2015) można znaleźć wskazówki dotyczące wyboru wartości p dla różnych rozkładów. Przykładowo najmniejsze obciążenia parametru a rozkładu Pareto $Pa(\theta, a)$ uzyskuje się dla wartości p bliskich 0,2, zaś błędy średniokwadratowe są najmniejsze dla rzędów kwantyli $p \in \langle 0,1; 0,2 \rangle$. Estymatory parametru θ dla małych wartości p są praktycznie nieobciążone i charakteryzują się małymi błędami średniokwadratowymi (por. Pekasiewicz, 2015).

3.6. Kwantyłowa metoda najmniejszych kwadratów i jej modyfikacje

Estymacja parametrów rozkładu zmiennej losowej kwantyłową metodą najmniejszych kwadratów (KMNK) polega na minimalizacji sumy kwadratów odchyleń kwantyli z próby i kwantyli rozkładu badanej zmiennej losowej (por. Gilchrist, 2000).

Estymatorami parametrów $\theta_1, \theta_2, \dots, \theta_k$ rozkładu zmiennej losowej Y o dystrybucji $F(y; \theta_1, \theta_2, \dots, \theta_k)$ otrzymanymi kwantyłową metodą najmniejszych kwadratów nazywamy statystyki $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k$ będące funkcjami elementów próby losowej Y_1, Y_2, \dots, Y_n , dla których osiąga minimum globalne wyrażenie postaci:

$$G(\theta_1, \theta_2, \dots, \theta_k) = \sum_{i=1}^n (Y_{p_i} - Q_{p_i})^2, \quad (3.47)$$

gdzie statystyki Y_{p_i} są kwantylami z próby rzędu $p_i = \frac{1}{n}$, natomiast Q_{p_i} to kwantyle rozkładu rzędu p_i dla $i = 1, 2, \dots, n$.

Ta metoda estymacji nie jest efektywna dla rozkładów o tzw. grubych ogonach, m.in. dla niektórych rozkładów Daguma i Pareto, dlatego

można wykorzystać jej modyfikacje zaprezentowane w pracy Pekasiewicz (2015):

- kwantylową metodę najmniejszych kwadratów z uciętą liczbą kwantyli,
- medianowo-kwantylową metodę najmniejszych kwadratów.

Istotą pierwszej modyfikacji kwantylowej metody najmniejszych kwadratów jest pominięcie, w konstrukcji estymatorów parametrów rozkładu, ustalonej liczby s skrajnych statystyk pozycyjnych, czyli wykorzystanie $n - s$ kwantyli. Metoda ta nazywana kwantylową metodą najmniejszych kwadratów z uciętą liczbą kwantyli (*UKMNK*). Zatem jej zastosowanie związane jest z poszukiwaniem minimum globalnego wyrażenia postaci:

$$G(\theta_1, \theta_2, \dots, \theta_k) = \sum_{i \in I_s} (Y_{p_i} - Q_{p_i})^2, \quad (3.48)$$

gdzie s jest ustaloną liczbą naturalną, zaś I_s jest pewnym $(n - s)$ elementowym podzbiorem zbioru $\{1, 2, \dots, n\}$.

Podobnie jak w przypadku metody percentyli, dla różnych rozkładów zbiór I_s może być różny. W przypadku rozkładów symetrycznych lub zbliżonych do rozkładów symetrycznych sensowne jest pominięcie parzystej liczby s kwantyli: $\frac{s}{2}$ kwantyli najmniejszych i $\frac{s}{2}$ największych. Dla rozkładów asymetrycznych, charakteryzujących się tylko grubym prawym lub tylko grubym lewym ogonem, pomijamy s kwantyli z ogona rozkładu, odpowiednio, prawego bądź lewego i wyznaczamy minimum globalne tak otrzymanej funkcji.

Kolejną modyfikacją kwantylowej metody najmniejszych kwadratów jest medianowo-kwantylowa metoda najmniejszych kwadratów (*MKMNK*). Polega ona na wyznaczeniu najpierw oszacowań parametru θ_i , dla $i = 1, 2, \dots, k$ przy zastosowaniu kwantylowej metodę najmniejszych kwadratów z uciętą liczbą kwantyli dla różnych wielkości s , a następnie ich mediany. Zastosowanie tej metody nie wymaga żadnych dodatkowych informacji, w przeciwieństwie do metody poprzedniej, w której podejmujemy decyzję o liczbie pomijanych skrajnych kwantyli.

Przykłady zastosowań tych metod do estymacji parametrów wybranych rozkładów ciągłych wraz z wynikami analiz własności otrzymanych estymatorów zaprezentowane zostały w pracy Pekasiewicz (2015).

3.7. Metoda najmniejszych kwadratów

Estymatory parametrów rozkładu zmiennej losowej Y o dystrybuancie $F(y; \theta_1, \theta_2, \dots, \theta_k)$ wyznaczyć można, korzystając z charakterystyk liczbowych zmiennych losowych $F(Y_{i:n})$ dla $i = 1, 2, \dots, n$:

$$E[F(Y_{i:n})] = \frac{i}{n+1} \quad \text{oraz} \quad D^2[F(Y_{i:n})] = \frac{i(n-i+1)}{(n+1)^2(n+2)}, \quad (3.49)$$

gdzie $Y_{i:n}$ jest i -tą statystyką pozycyjną wyznaczoną na podstawie n -elementowej próby losowej Y_1, Y_2, \dots, Y_n .

Znalezienie minimum globalnego wyrażenia postaci (Swain i in., 1988):

$$G(\theta_1, \theta_2, \dots, \theta_k) = \sum_{i=1}^n \left(F(Y_{i:n}) - \frac{i}{n+1} \right)^2 \quad (3.50)$$

proceedzi do wyznaczenia estymatorów $\theta_1, \theta_2, \dots, \theta_k$ metodą najmniejszych kwadratów (MNK).

Niestety, wartości estymatorów wyznacza się zwykle numerycznie.

Dla rozkładu Daguma $D(a, v, b)$ o dystrybuancie określonej wzorem (1.76) wyznaczenie minimum globalnego wyrażenia (3.50) związane jest z rozwiązaniem układu równań:

$$\begin{cases} \sum_{i=1}^n \left[\left(1 + \left(\frac{y_{i:n}}{b} \right)^{-v} \right)^{-a} - \frac{i}{n+1} \right] \left[1 + \left(\frac{y_{i:n}}{b} \right)^{-v} \right]^{-a} \ln \left[1 + \left(\frac{y_{i:n}}{b} \right)^{-v} \right] = 0, \\ \sum_{i=1}^n \left[\left(1 + \left(\frac{y_{i:n}}{b} \right)^{-v} \right)^{-a} - \frac{i}{n+1} \right] a \left[1 + \left(\frac{y_{i:n}}{b} \right)^{-v} \right]^{-a-1} (y_{i:n})^{-v} \ln \left(\frac{y_{i:n}}{b} \right) = 0, \\ \sum_{i=1}^n \left[\left(1 + \left(\frac{y_{i:n}}{b} \right)^{-v} \right)^{-a} - \frac{i}{n+1} \right] \left[1 + \left(\frac{y_{i:n}}{b} \right)^{-v} \right]^{-a-1} (y_{i:n})^{-v} = 0 \end{cases} \quad (3.51)$$

natomiast dla rozkładu Pareto $Pa(\theta, a)$ o dystrybuancie (1.5) z rozwiązaniem układu równań:

$$\begin{cases} \sum_{i=1}^n \left[1 - \left(\frac{\theta}{y_{i:n}} \right)^a - \frac{i}{n+1} \right] \left(\frac{\theta}{y_{i:n}} \right)^a \ln \frac{\theta}{y_{i:n}} = 0, \\ \sum_{i=1}^n \left[1 - \left(\frac{\theta}{y_{i:n}} \right)^a - \frac{i}{n+1} \right] \left(\frac{1}{y_{i:n}} \right)^{a-2} = 0. \end{cases} \quad (3.52)$$

3.8. Ważona metoda najmniejszych kwadratów

Do wyznaczenia estymatorów parametrów $\theta_1, \theta_2, \dots, \theta_k$ rozkładu zmiennej losowej Y o dystrybucji $F(y; \theta_1, \theta_2, \dots, \theta_k)$ wykorzystuje się również ważoną metodę najmniejszych kwadratów (WMNK), która polega na poszukiwaniu minimum globalnego wyrażenia postaci:

$$G(\theta_1, \theta_1, \dots, \theta_k) = \sum_{i=1}^n \frac{(n+1)^2 (n+2)}{i(n-i+1)} \left(F(Y_{i:n}) - \frac{i}{n+1} \right)^2, \quad (3.53)$$

gdzie $Y_{i:n}$ jest i -tą statystyką pozycyjną wyznaczoną na podstawie n -elementowej próby losowej Y_1, Y_2, \dots, Y_n .

Przy zastosowaniu tej metody, podobnie jak metody najmniejszych kwadratów, wartości estymatorów wyznacza się zwykle numerycznie.

Dla rozkładu Daguma estymatory parametrów otrzymujemy z układu równań:

$$\begin{cases} \sum_{i=1}^n \frac{1}{i(n-i+1)} \left[\left(1 + \left(\frac{y_{i:n}}{b} \right)^{-v} \right)^{-a} - \frac{i}{n+1} \right] \left(1 + \left(\frac{y_{i:n}}{b} \right)^{-v} \right)^{-a} \ln \left(1 + \left(\frac{y_{i:n}}{b} \right)^{-v} \right) = 0, \\ \sum_{i=1}^n \frac{1}{i(n-i+1)} \left[\left(1 + \left(\frac{y_{i:n}}{b} \right)^{-v} \right)^{-a} - \frac{i}{n+1} \right] \left(1 + \left(\frac{y_{i:n}}{b} \right)^{-v} \right)^{-a-1} (y_{i:n})^{-v} \ln \left(\frac{y_{i:n}}{b} \right) = 0, \\ \sum_{i=1}^n \frac{1}{i(n-i+1)} \left[\left(1 + \left(\frac{y_{i:n}}{b} \right)^{-v} \right)^{-a} - \frac{i}{n+1} \right] \left(1 + \left(\frac{y_{i:n}}{b} \right)^{-v} \right)^{-a-1} (y_{i:n})^{-v} = 0. \end{cases} \quad (3.54)$$

natomiast dla rozkładu Pareto:

$$\begin{cases} \sum_{i=1}^n \frac{1}{i(n-i+1)} \left[1 - \left(\frac{\theta}{y_{i:n}} \right)^a - \frac{i}{n+1} \right] \left(\frac{\theta}{y_{i:n}} \right)^a \ln \frac{\theta}{y_{i:n}} = 0, \\ \sum_{i=1}^n \frac{1}{i(n-i+1)} \left[1 - \left(\frac{\theta}{y_{i:n}} \right)^a - \frac{i}{n+1} \right] \left(\frac{1}{y_{i:n}} \right)^{a-2} = 0. \end{cases} \quad (3.55)$$

3.9. Metody D'Addario

Metody D'Addario (por. D'Addario 1934; 1939) wykorzystuje się m.in. do estymacji parametrów funkcji Zengi. Zenga i in. (2010a) zastosowali pięć metod D'Addario.

Założmy, że zmienna losowa Y ma rozkład o dystrybuancie $F(y; \mu, \alpha, \theta)$, gdzie μ jest parametrem położenia.

Pierwsza z metod estymacji polega na rozwiązaniu układu równań:

$$\begin{cases} \mu = \bar{y}, \\ F(Me; \mu, \alpha, \theta) = 0,5, \\ F(\bar{y}; \mu, \alpha, \theta) = \hat{F}(\bar{y}), \end{cases} \quad (3.56)$$

gdzie \bar{y} , Me są, odpowiednio, wartościami średniej arytmetycznej i mediany z n -elementowej próby losowej Y_1, Y_2, \dots, Y_n oraz $\hat{F}(\bar{y})$ empiryczną dystrybuantą obliczoną dla wartości średniej arytmetycznej z próby.

Druga metoda D'Addario pozwala uzyskać oszacowania parametrów rozkładu poprzez rozwiązanie układu równań:

$$\begin{cases} \mu = \bar{y}, \\ F(Me; \mu, \alpha, \theta) = 0,5, \\ P(\mu, \alpha, \theta) = \widehat{SP}, \end{cases} \quad (3.57)$$

gdzie $P(\mu, \alpha, \theta) = 2F(1; 1, \alpha, \theta) - 1$ jest indeksem Pietry oraz

$\widehat{SP} = \frac{1}{2\bar{y}n} \sum_{i=1}^n |y_i - \bar{y}|$ empirycznym indeksem Pietry (por. Zenga i in. (2010a)).

Kolejne metody estymacji wykorzystują wariancje rozkładu. Estymatory parametrów możemy otrzymać poprzez rozwiązanie układu równań:

$$\begin{cases} \mu = \bar{y}, \\ F(Me; \mu, \alpha, \theta) = 0,5, \\ D^2(\mu, \alpha, \theta) = m_2 \end{cases} \quad (3.58)$$

lub układu:

$$\begin{cases} \mu = \bar{y}, \\ F(\bar{y}; \mu, \alpha, \theta) = \hat{F}(\bar{y}), \\ D^2(\mu, \alpha, \theta) = m_2, \end{cases} \quad (3.59)$$

gdzie m_2 jest wariancją z próby, zaś $D^2(\bar{y}; \alpha, \theta)$ jest wariancją rozkładu teoretycznego.

W estymacji parametrów można również wykorzystać indeks nierównomierności Zengi:

$$\begin{cases} \mu = \bar{y}, \\ F(Me; \mu, \alpha, \theta) = 0,5, \\ A(\mu, \alpha, \theta) = \hat{A}(\bar{y}), \end{cases} \quad (3.60)$$

gdzie $A(\mu, \alpha, \theta) = 1 - \left\{ \frac{1 - F(1; 1, \alpha, \theta)}{F(1; 1, \alpha, \theta)} \right\}^2$ jest indeksem punktowym Zengi

$$(2007) \text{ dla } \bar{y} \text{ oraz } \hat{A}(\bar{y}) = \frac{\sum_{y_i \leq \bar{y}} y_i}{\hat{F}(\bar{y})} \cdot \frac{1 - \hat{F}(\bar{y})}{\sum_{y_i > \bar{y}} y_i}.$$

ROZDZIAŁ 4

Estymacja rozkładów dochodów gospodarstw domowych w Polsce

4.1. Uwagi wstępne

Badania budżetów gospodarstw domowych są podstawowym źródłem informacji o dochodach, wydatkach, spożyciu oraz innych aspektach warunków życia określonych grup ludności. Pozwalają one na analizę wpływu różnych czynników na kształtowanie się poziomu i zróżnicowania dochodów podstawowych grup gospodarstw domowych. Badanie to prowadzone jest metodą reprezentacyjną, która daje możliwość uogólnienia, z określonym błędem, uzyskanych wyników na wszystkie gospodarstwa domowe w Polsce i poszczególne grupy gospodarstw domowych wyróżnianych ze względu na położenie geograficzne, źródła utrzymania, liczbę osób w gospodarstwie domowym lub liczbę dzieci.

Na podstawie prób uzyskanych z badań budżetów gospodarstw domowych w roku 2016 dokonano aproksymacji rozkładów dochodów gospodarstw domowych w różnych regionach w Polsce za pomocą wybranych rozkładów teoretycznych zaprezentowanych w rozdziale 1. Ponadto odrębnie przeanalizowano grupę gospodarstw najbogatszych, tzn. o dochodach przekraczających pewną, ustaloną dla całego kraju, granicę zamożności. Do modelowania rozkładów dochodów tej grupy wykorzystano rozkład Pareto.

Dobroć dopasowania rozważanych rozkładów oceniona została za pomocą testów statystycznych: Kołmogorowa-Smirnowa i Andersona-Darlinga oraz współczynników zgodności: wskaźnika podobieństwa struktur (Wps), indeksu Mortara (A_1), indeksu Pearsona kwadratów

różnic częstości względnych (A_2) oraz indeksu uwzględniającego różnice między wartością średnią rozkładu empirycznego i wartością oczekiwaną rozkładu teoretycznego (A_3).

Współczynnik podobieństwa struktur obliczano, stosując wzór:

$$Wps = \sum_{j=1}^s \min(w_j, \hat{w}_j), \quad (4.1)$$

gdzie s oznacza liczbę przedziałów, na które zostały pogrupowane uporządkowane dane, natomiast w_j oraz \hat{w}_j to, odpowiednio, empiryczne oraz teoretyczne wskaźniki struktury w poszczególnych przedziałach.

Pozostałe indeksy mają postaci:

$$A_1 = \frac{1}{n} \sum_{j=1}^s |n_j - \hat{n}_j|, \quad (4.2)$$

$$A_2 = \sqrt{\frac{1}{n} \sum_{j=1}^s \frac{(n_j - \hat{n}_j)^2}{\hat{n}_j}}, \quad (4.3)$$

gdzie n_j oraz \hat{n}_j oznaczają, odpowiednio, empiryczne oraz teoretyczne częstości w poszczególnych przedziałach.

Zgodność rozkładów empirycznych z wybranym modelem teoretycznym bada się, także porównując najważniejsze charakterystyki statystyczne otrzymane z rozkładu teoretycznego z odpowiednimi wartościami empirycznymi – mogą to być średnie, odchylenia standardowe lub współczynniki nierównomierności. Przyjmuje się, że w przypadku bardzo dobrego dopasowania różnice między empirycznymi i teoretycznymi charakterystykami rozkładu powinny być mniejsze niż 5%. W przeprowadzonych analizach uwzględniono współczynnik oparty na względnej różnicy wartości średniej rozkładu empirycznego i wartości oczekiwanej rozkładu teoretycznego:

$$A_3 = \frac{|\bar{Y} - EY|}{EY} \cdot 100\%, \quad (4.4)$$

gdzie \bar{Y} oznacza średnie dochody, zaś EY jest wartością oczekiwaną rozkładu teoretycznego.

Inne miary badające „dobroć” dopasowania rozkładu teoretycznego zaprezentowane są np. w pracach Kordosa (1973) i Ostasiewicz (2013).

Dobre dopasowanie rozkładu teoretycznego do danych empirycznych może prowadzić do alternatywnego sposobu wyznaczania parametrów charakteryzujących badaną populację, m.in. można szacować miary nierównomierności, najpierw aproksymując rozkład empiryczny rozkładem teoretycznym, a następnie wyznaczyć odpowiednie indeksy w oparciu o oszacowane parametry funkcji gęstości rozkładu teoretycznego. Na przykład w celu oszacowania indeksu Giniego aproksymacja rozkładem Daguma pozwala skorzystać z wzoru (1.84), rozkładem Singha-Maddali – z wzoru (1.98), zaś gamma – z wzoru (1.70).

Wyniki badań empirycznych związanych z dopasowaniem różnych rozkładów teoretycznych do rozkładów dochodów gospodarstw domowych w regionach Polski, z uwzględnieniem różnych metod estymacji parametrów, stanowiły inspirację do uogólnienia otrzymanych rezultatów. W tym celu przeprowadzono badania symulacyjne dotyczące efektywności metod estymacji parametrów rozkładu Daguma i Pareto, które polegało na ocenie błędów stricte losowych i losowych błędów systematycznych. W tym celu w pierwszej kolejności oszacowano empiryczne obciążenie rozważanych estymatorów, jako miarę błędu systematycznego, a następnie za pomocą błędów średniokwadratowych oceniono błąd całkowity. Otrzymane wyniki zostały zaprezentowane w ostatnich podrozdziałach.

4.2. Charakterystyka zbioru danych

Przeprowadzona w pracy analiza empiryczna została oparta na próbie losowej pochodzącej z badania Budżetów Gospodarstw Domowych przeprowadzonego przez Główny Urząd Statystyczny w roku 2016. Do obliczeń wykorzystano mikrodane, czyli dane na poziomie jednostkowym. Jednostką obserwacji jest gospodarstwo domowe jedno- lub wieloosobowe. Całkowita liczebność próby otrzymanej w wyniku badania budżetów gospodarstw domowych wynosiła 36 616. Stanowi to około 0,3% ogólnej liczby gospodarstw domowych w Polsce.

W celu wylosowania próby stosowano złożony schemat losowania – było to losowanie dwustopniowe, warstwowe, z różnymi prawdopodobieństwami wyboru na I stopniu. Jednostkami losowania pierwszego stopnia (jps) były terenowe punkty badań, zaś jednostkami losowania

drugiego stopnia (jds) były mieszkania. Jps zostały powarstwowane według 6 klas miejscowości, przy czym duże miasta stanowiły odrębne warstwy. Małe miasta i tereny wiejskie zostały powarstwowane według podregionów (NUTS 3). Łącznie utworzono 191 warstw, w tym 58 wiejskich. Jednostki pierwszego stopnia losowane były oddzielnie w każdej warstwie, z wykorzystaniem doboru systematycznego po uprzednim losowym uporządkowaniu (tzw. schemat Hartleya-Rao). W 2016 roku w badaniu budżetów gospodarstw domowych wykorzystywane były dwie próbkę liczące po 783 jps: próbkę 1 – wylosowana w 2015 roku do badania na lata 2016–2017 oraz próbkę 2 – wylosowana w 2014 roku do badania na lata 2015–2016. Próbkę wylosowane zostały według tego samego schematu losowania.

Przy wyborze schematu losowania próbkę kierowano się założeniem, aby wylosowana próbkę była próbkę automatycznie wyważoną. Oznacza to, że każde gospodarstwo domowe powinno mieć w przybliżeniu jednakowe prawdopodobieństwo wyboru. W związku z tym przyjęto, że lokalizacja jps między warstwy będzie proporcjonalna do szacunkowej liczby mieszkań w warstwie. Jednocześnie jednostki pierwszego stopnia losowane były z prawdopodobieństwami proporcjonalnymi do szacunkowej liczby jds w terenowym punkcie badań, zaś z każdego terenowego punktu badań losowana jest taka sama liczba mieszkań.

Kategorię dochodu, która stanowi przedmiot przeprowadzonej aproksymacji oraz analizy nierównomierności, jest tzw. dochód rozporządzalny gospodarstwa domowego. Według definicji przyjętej przez GUS (2016):

Dochód rozporządzalny jest to suma bieżących dochodów gospodarstwa domowego z poszczególnych źródeł pomniejszona o zaliczki na podatek dochodowy od osób fizycznych płacone przez płatnika w imieniu podatnika (od dochodów z pracy najemnej oraz od niektórych świadczeń z ubezpieczenia społecznego i świadczeń pomocy społecznej), o podatki od dochodów i własności płacone przez osoby pracujące na własny rachunek, w tym przedstawicieli wolnych zawodów i osób użytkujących gospodarstwo indywidualne w rolnictwie oraz o składki na ubezpieczenia społeczne i zdrowotne. W skład dochodu rozporządzalnego wchodzi dochody pieniężne i niepieniężne, w tym spożycie naturalne (towary i usługi konsumpcyjne pobrane na potrzeby gospodarstwa domowego z gospodarstwa indywidualnego w rolnictwie bądź działalności gospodarczej na własny rachunek)

oraz towary i usługi otrzymane nieodpłatnie. Dochód rozporządzalny przeznaczony jest na wydatki oraz przyrost oszczędności.

W celu uwzględnienia wpływu na koszty utrzymania gospodarstwa domowego czynnika demograficznego, w badaniu przekształcono kategorię dochodu rozporządzalnego gospodarstwa domowego w tzw. dochód ekwiwalentny. Jako skalę ekwiwalentności przyjęto skalę stosowaną przez OECD, opartą na pierwiastku kwadratowym z liczby osób w gospodarstwie domowym. Oznacza to, że dochód każdego gospodarstwa domowego został przekształcony w dochód ekwiwalentny poprzez podzielenie go przez pierwiastek z liczby osób w gospodarstwie.

Przeprowadzone badania dotyczyły aproksymacji rozkładów dochodów ekwiwalentnych wszystkich gospodarstw domowych oraz grup gospodarstw domowych o największych dochodach w Polsce ogółem i w sześciu wyróżnionych regionach. Według klasyfikacji NUTS obszar Polski do roku 2018 podzielony był na 6 regionów NUTS1 (rys. 4.1), obecnie jest to 7 makroregionów, dodatkowo wyróżniono makroregion obejmujący województwo mazowieckie i zmieniły się granice makroregionu centralnego i wschodniego. Makroregion centralny obejmuje województwo łódzkie i świętokrzyskie, a wschodni – województwa lubelskie, podkarpackie i podlaskie.



Rysunek 4.1. Podział Polski na jednostki NUTS 1 w latach 2014–2016

Źródło: <https://stat.gov.pl/statystyka-regionalna/jednostki-terytorialne/klasyfikacja-nuts/rewizja-nuts-2013> (dostęp: 17.09.2019)

4.3. Aproksymacja rozkładów dochodów gospodarstw domowych w Polsce w podziale na regiony

Na podstawie próby losowej pochodzącej z Badania Budżetów Gospodarstw Domowych, liczącej 36 616 gospodarstw domowych o dodatnich dochodach, wyznaczono podstawowe charakterystyki liczbowe empirycznych rozkładów dochodów ekwiwalentnych w Polsce i w poszczególnych regionach Polski w 2016 roku. Jako skalę ekwiwalentności przyjęto pierwiastek kwadratowy z liczby osób w gospodarstwie domowym. Otrzymane wyniki przedstawione zostały w tab. 4.1.

Tabela 4.1. Podstawowe charakterystyki dochodów gospodarstw domowych w ujęciu regionalnym

Region	Liczba gospodarstw domowych	Min. (PLN)	Max. (PLN)	Średnia (PLN)	Odchylenie standardowe (PLN)
centralny	7971	18,00	32360,00	2867,69	2015,18
południowy	7397	21,21	20000,00	2511,73	1318,15
wschodni	6233	8,94	21743,18	2252,71	1310,14
północno-zachodni	5552	14,14	19391,79	2510,23	1282,63
południowo-zachodni	3856	20,21	20590,00	2565,84	1415,40
północny	5607	1,27	333978,50	2509,23	1712,00
ogółem	36616	1,27	333978,50	2552,63	1572,81

Źródło: opracowanie własne

Łatwo zauważyć, że średni dochód ekwiwalentny jest największy dla gospodarstw domowych z regionu centralnego (2867,69 PLN), a najmniejszy dla gospodarstw z regionu wschodniego (2252,71 PLN). W najbogatszym regionie centralnym obserwuje się również największe zróżnicowanie dochodów, gdyż odchylenie standardowe stanowi tu 70% średniej; natomiast najmniejszą dyspersją dochodów charakteryzują się regiony południowy i północno-zachodni. W regionach tych odchylenie standardowe stanowi około 51–52% wartości średniej, co w przypadku rozkładu dochodów jest stosunkowo niewielką wartością.

Analiza rozkładów dochodów to nie tylko badanie średniego poziomu dochodu i jego zróżnicowania, lecz także, a nawet przede wszystkim, analiza nierównomierności, czyli koncentracji. W literaturze istnieje wiele miar koncentracji, m.in.: Herfindahla (Hart, 1975), Rosenblutha (Wyżnikiewicz, 1987), Giniego (Gini, 1912), Bonferroniego (Bonferroni, 1930), Theila (Theil, 1967) czy Zengi (Zenga, 1990; 2007). Indeks Giniego jest miarą najczęściej wykorzystywaną w praktyce, dlatego zastosowano go w przeprowadzonych analizach do badania nierównomierności. Jednym z estymatorów tego współczynnika jest statystyka zaproponowana przez Fei i in. (1979) postaci:

$$\hat{G} = \frac{2 \sum_{i=1}^n \left(w_i y_{(i)} \sum_{j=1}^i w_j \right) - \sum_{i=1}^n w_i y_{(i)}}{\left(\sum_{i=1}^n w_i \right) \sum_{i=1}^n w_i y_{(i)}} - 1, \quad (4.5)$$

gdzie: $y_{(i)}$ jest dochodem i -tego gospodarstwa domowego z n -elementowej próby uporządkowanej w sposób rosnący, w_i – wagą i -tej jednostki, natomiast $\sum_{j=1}^i w_j$ – rangą i -tej jednostki.

Oszacowania współczynnika Giniego dla Polski i poszczególnych regionów podane zostały w tab. 4.2 i świadczą o największych nierównomiernościach rozkładu ekwiwalentnych dochodów gospodarstw domowych w regionie centralnym. Regiony centralny i północny charakteryzują się większymi wartościami indeksu Giniego w stosunku do wartości tego indeksu dla całej Polski.

Tabela 4.2. Oszacowania współczynnika Giniego

Region	centralny	południowy	wschodni	północno-zachodni	południowo-zachodni	północny	ogółem
\hat{G}	0,318	0,255	0,275	0,252	0,265	0,293	0,265

Źródło: opracowanie własne

Oczywiście na podstawie prób losowych można szacować także inne miary statystyczne, zarówno nierównomierności, jak i miary asymetrii i spłaszczenia, a następnie w oparciu o nie opisywać zbiorowości statystyczne; jednak przedmiotem przeprowadzanych w niniejszej pracy analiz jest modelowanie empirycznych rozkładów dochodów rozkładami teoretycznymi, badanie zgodności dopasowania i ewentualnie wnioskowanie o ich własnościach na podstawie otrzymanych parametrów.

Jak wynika z wielu badań empirycznych oraz analiz teoretycznych, w tym rozważań przeprowadzonych w rozdziale 1, rozkłady dochodów są najczęściej jednomodalne i charakteryzują się prawostronną asymetrią oraz dodatnią kurtozą. Po rozważeniu własności wielu rozkładów teoretycznych można przyjąć hipotezę, że rozkładami, które mogą dobrze modelować rozkłady dochodów w Polsce w różnych przekrojach, są:

- 1) rozkład Daguma;
- 2) rozkład logarytmiczno-normalny;
- 3) rozkład Singha-Maddali.

Parametry tych rozkładów można szacować metodami zaprezentowanymi w rozdziale 3. Istotne jest, aby wybrać procedurę estymacji dającą najdokładniejszą z możliwych aproksymację.

W przeprowadzonych badaniach parametry wyżej wymienionych rozkładów szacowano czterema metodami:

- metodą największej wiarygodności (MNW);
- metodą momentów (MM);
- metodą najmniejszych kwadratów (MNK);
- ważoną metodą najmniejszych kwadratów (WMNK);

analizując wpływ procedury estymacji parametrów na dopasowanie rozkładu teoretycznego do danych empirycznych.

Analiza dopasowania rozkładów empirycznych rozkładem Daguma

Wyniki oszacowań parametrów rozkładu Daguma aproksymującego rozkłady ekwiwalentnych dochodów gospodarstw domowych w Polsce i jej regionach przedstawione zostały w tab. 4.3.

Tabela 4.3. Oszacowania parametrów rozkładu Daguma aproksymującego rozkład dochodów gospodarstw domowych w regionach w Polsce

Region	Metoda estymacji	Parametry rozkładu Daguma		
		a	ν	b
centralny	MNW	0,8702	3,2396	2540,1511
	MM	0,4151	3,8142	3638,1094
	MNK	1,0300	3,0618	2345,0914
	WMNK	1,0205	3,0950	2355,3962
południowy	MNW	0,8704	4,0501	2371,5204
	MM	0,6334	4,3842	2685,6778
	MNK	0,8758	3,9934	2367,1807
	WMNK	0,8988	3,9957	2343,1626
wschodni	MNW	0,7761	3,8891	2183,1217
	MM	0,6003	4,0472	2437,0188
	MNK	0,8856	3,7139	2072,7903
	WMNK	0,8794	3,7460	2077,8685
północno-zachodni	MNW	0,7828	4,2070	2472,6570
	MM	0,6177	4,4921	2712,1830
	MNK	0,8575	4,0421	2393,0183
	WMNK	0,8728	4,0515	2376,2933
południowo-zachodni	MNW	0,8126	3,9602	2466,7193
	MM	0,6174	4,2060	2758,7396
	MNK	0,8794	3,8115	2393,8877
	WMNK	0,8765	3,8528	2395,5471
północny	MNW	0,7858	3,6227	2367,5657
	MM	0,5593	3,5122	2780,7257
	MNK	1,0651	3,3634	2078,5533
	WMNK	1,0550	3,3736	2087,3699
ogółem	MNW	0,8335	3,7102	2384,1800
	MM	0,6232	3,8140	2709,3659
	MNK	0,9548	3,5543	2254,3792
	WMNK	0,9574	3,5704	2251,4294

Źródło: obliczenia własne przy użyciu pakietu Mathematica

Wartości oszacowanych parametrów wykorzystano jako wartości teoretyczne przy badaniu zgodności rozkładów za pomocą testów statystycznych

Kołmogorowa-Smirnowa i Andersona-Darlinga. Poza przypadkami szacowania parametrów metodą momentów oraz metodą największej wiarygodności dla dochodów ogółem, przeprowadzone testy pozwoliły podjąć decyzję o braku podstaw do odrzucenia hipotezy zerowej o zgodności rozkładów dochodów z rozkładem Daguma z przyjętymi parametrami (por. tab. 4.4 – wartości p -value większe od 0,05 przynajmniej dla jednego ze stosowanych testów, a w większości dla obydwu testów). Oznacza to, że metody momentów raczej nie powinno się stosować do szacowania parametrów rozkładu Daguma.

Potwierdzają to obliczone miary dopasowania przedstawione w tab. 4.5. Miary te są dobrym narzędziem do oceny zgodności rozkładów empirycznych z teoretycznymi, szczególnie gdy liczebność prób jest bardzo duża, jak ma to miejsce w przypadku badania całej populacji lub podziału całej populacji na makroregiony (NUTS1). W takiej sytuacji stosowanie niektórych testów zgodności, np. testu zgodności χ^2 , może prowadzić do odrzucenia hipotezy zerowej nawet dla minimalnych odchyień rozkładu teoretycznego od rozkładu empirycznego (por. np. Jędrzejczak, 2011). Wartość współczynnika podobieństwa struktur powinna być większa niż 0,95 i im jest ona bliższa 1, tym dopasowanie rozkładu teoretycznego jest lepsze. W przypadku pozostałych indeksów wartość bliższa 0 świadczy o lepszym dopasowaniu rozkładu empirycznego do teoretycznego. Zastosowanie metody momentów przy modelowaniu empirycznych rozkładów dochodów rozkładem Daguma daje najgorsze dopasowania, a szacowanie parametrów metodami najmniejszych kwadratów (MNK, WMNK) pozwala zwykle otrzymać najlepsze dopasowania, porównując wskaźniki podobieństwa struktur, indeksy Mortata i Pearsona, chociaż wskaźnik oparty na porównaniu średnich empirycznych i wartości oczekiwanych nie zawsze to potwierdza.

Tabela 4.4. Wartości p -value dla testów Kołmogorowa-Smirnowa i Andersona-Darlinga weryfikujących hipotezy o zgodności rozkładów dochodów z rozkładem Daguma

Region	Test	Metoda estymacji			
		MNW	MM	MNK	WMNK
centralny	K-S	0,098	0,000	0,260	0,184
	A-D	0,060	0,000	0,150	0,170
południowy	K-S	0,232	0,000	0,244	0,249
	A-D	0,316	0,000	0,267	0,332

Region	Test	Metoda estymacji			
		MNW	MM	MNK	WMNK
wschodni	K-S	0,458	0,000	0,823	0,754
	A-D	0,258	0,000	0,545	0,586
północno-zachodni	K-S	0,391	0,002	0,066	0,709
	A-D	0,344	0,001	0,482	0,548
południowo-zachodni	K-S	0,809	0,004	0,948	0,819
	A-D	0,573	0,001	0,607	0,671
północny	K-S	0,019	0,000	0,972	0,977
	A-D	0,006	0,000	0,203	0,209
ogółem	K-S	0,000	0,000	0,210	0,098
	A-D	0,000	0,000	0,012	0,015

Źródło: obliczenia własne przy użyciu pakietu Mathematica

Tabela 4.5. Miary dopasowania empirycznych rozkładów dochodów do rozkładu Daguma

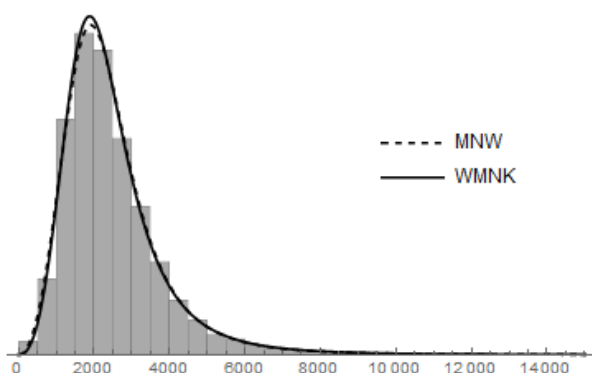
Region	Metoda estymacji	Wps	A_1	A_2	A_3
centralny	MNW	0,9811	0,0376	0,0452	1,8616
	MM	0,8917	0,2165	0,2748	1,4938
	MNK	0,9829	0,0340	0,0410	0,6235
	WMNK	0,9840	0,0317	0,0392	1,0025
południowy	MNW	0,9688	0,0623	0,1201	0,4086
	MM	0,9556	0,0886	0,1244	0,4799
	MNK	0,9888	0,0621	0,1185	0,1253
	WMNK	0,9745	0,0508	0,0711	0,2159
wschodni	MNW	0,9788	0,0424	0,0527	1,4472
	MM	0,9516	0,0966	0,1329	1,0217
	MNK	0,9829	0,0341	0,0488	0,7052
	WMNK	0,9887	0,0225	0,0304	0,9120
północno-zachodni	MNW	0,9840	0,0319	0,0525	0,5473
	MM	0,9654	0,0691	0,1013	0,4700
	MNK	0,9835	0,0328	0,0515	0,0474
	WMNK	0,9823	0,0353	0,0524	0,0536

Tabela 4.5 (cd.)

Region	Metoda estymacji	Wps	A_1	A_2	A_3
południowo-zachodni	MNW	0,9728	0,0543	0,0610	0,7693
	MM	0,9547	0,0903	0,1178	0,7119
	MNK	0,9717	0,0563	0,0637	0,1148
	WMNK	0,9722	0,0554	0,0614	0,3799
północny	MNW	0,9685	0,0630	0,0811	2,4617
	MM	0,8717	0,2565	0,3491	0,7055
	MNK	0,9858	0,0283	0,0722	1,3160
	WMNK	0,9861	0,0278	0,0709	1,3726
ogółem	MNW	0,9828	0,0343	0,0514	1,5556
	MM	0,9430	0,1140	0,1515	0,9584
	MNK	0,9854	0,0292	0,0452	0,8501
	WMNK	0,9846	0,0307	0,0451	0,9933

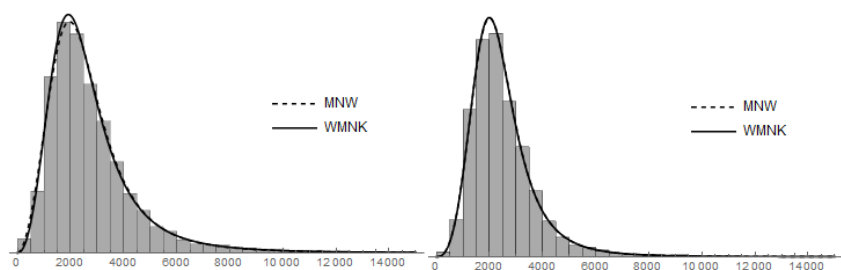
Źródło: obliczenia własne przy użyciu pakietu Mathematica

Na rys. 4.2–4.3 zaprezentowano graficznie wyniki aproksymacji rozkładów dochodów gospodarstw domowych w Polsce, otrzymane najbardziej popularną metodą szacowania parametrów, czyli metodą największej wiarygodności (MNW), oraz ważoną metodą najmniejszych kwadratów (WMNK), która wydaje się bardziej efektywna.



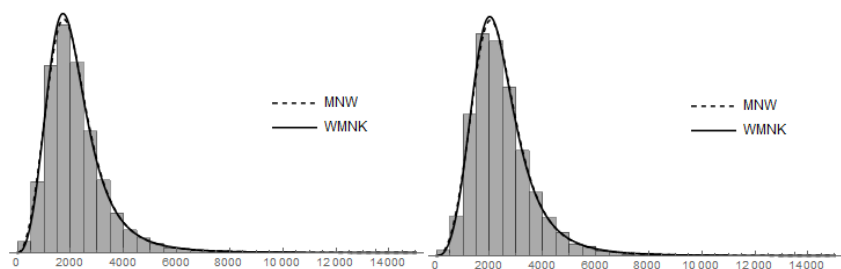
Rysunek 4.2. Aproksymacja rozkładu dochodów w Polsce rozkładem Daguma z wybranymi metodami estymacji parametrów

Źródło: opracowanie własne



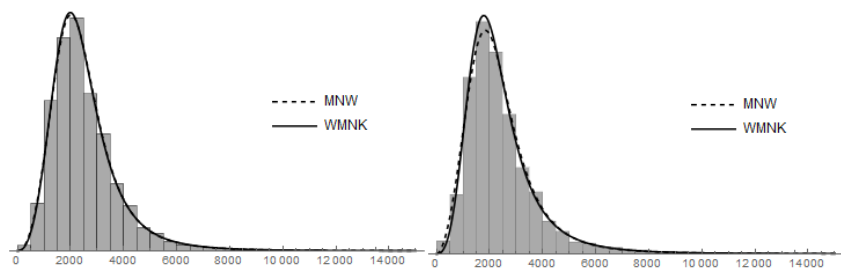
a) region centralny

b) region południowy



c) region wschodni

d) region północno-zachodni



e) region południowo-zachodni

f) region północny

Rysunek 4.3. Aproksymacja rozkładu dochodów w regionach rozkładem Daguma z wybranymi metodami estymacji parametrów

Źródło: opracowanie własne

Rys. 4.2–4.3 potwierdzają różnice w dopasowaniu rozkładów teoretycznych do empirycznych, przy stosowaniu różnych metod estymacji parametrów. Największe różnice zauważa się dla regionu północnego, dla którego ważona metoda najmniejszych kwadratów pozwala precyzyjniej modelować empiryczny rozkład ekwiwalentnych dochodów gospodarstw domowych. Wartości miar dopasowania: współczynnik podobieństwa struktur jest większy dla ważonej metody najmniejszych kwadratów (dla MNW – 0,9799, zaś dla WMNK – 0,9919), natomiast indeks Mortara jest ponad dwukrotnie mniejszy (dla MNW – 0,0402, natomiast dla WMNK – 0,0160). Również indeks A_3 jest mniejszy (dla MNW – 1,7398, dla WMNK – 0,6600) i świadczy o mniejszych różnicach między wartościami średnimi – średnią z próby i wartością oczekiwaną rozkładu Daguma.

Analiza dopasowania rozkładów empirycznych rozkładem logarytmiczno-normalnym

W analizach rozważano również rozkład logarytmiczno-normalny, który w pewnych sytuacjach może być przydatny do modelowania rozkładów dochodów, szczególnie w krajach w okresie transformacji ustrojowej. Podobnie jak w przypadku rozkładu Daguma, parametry rozkładu logarytmiczno-normalnego szacowano czterema metodami i badano dopasowanie rozkładów z oszacowanymi na podstawie prób wartościami parametrów do rozkładów empirycznych.

W tab. 4.6 przedstawiono wartości oszacowanych parametrów, zaś w tab. 4.7 wartości p -value dla testów statystycznych Kołmogorowa-Smirnowa i Andersona-Darlinga stosowanych do weryfikacji hipotez o zgodności rozkładu empirycznego z logarytmiczno-normalnym. Na poziomie istotności 0,05 dla każdego regionu przynajmniej jeden z testów, a często obydwa, odrzucały hipotezę o zgodności rozkładu dochodów gospodarstw domowych z rozkładem logarytmiczno-normalnym o parametrach uzyskanych jedną z czterech rozważanych metod estymacji.

Tabela 4.6. Estymatory parametrów rozkładu logarytmiczno-normalnego dla rozkładów dochodów w Polsce według regionów

Region	Metoda estymacji	Parametry rozkładu logarytmiczno-normalnego	
		μ	σ
centralny	MNW	7,7679	0,6065
	MM	7,7432	0,6376
	MNK	7,7741	0,5372
	WMNK	7,7744	0,5433
południowy	MNW	7,7127	0,4767
	MM	7,7038	0,4901
	MNK	7,7195	0,4361
	WMNK	7,7176	0,4414
wschodni	MNW	7,5750	0,5365
	MM	7,5631	0,5416
	MNK	7,5876	0,4665
	WMNK	7,5859	0,4736
północno-zachodni	MNW	7,7114	0,4916
	MM	7,7078	0,4809
	MNK	7,7228	0,4342
	WMNK	7,7210	0,4390
południowo-zachodni	MNW	7,7198	0,5099
	MM	7,7099	0,5158
	MNK	7,7300	0,4562
	WMNK	7,7282	0,4619
północny	MNW	7,6562	0,6068
	MM	7,6128	0,6447
	MNK	7,6667	0,4846
	WMNK	7,6674	0,4934
ogółem	MNW	7,6932	0,5471
	MM	7,6713	0,5728
	MNK	7,7014	0,4750
	WMNK	7,7007	0,4820

Źródło: obliczenia własne przy użyciu pakietu Mathematica

Tabela 4.7. Wartości p -value dla testów Kołmogorowa-Smirnowa i Andersona-Darlinga dla zgodności rozkładu dochodów z rozkładem logarytmiczno-normalnym

Region	Test	Metoda estymacji			
		MNW	MM	MNK	WMNK
centralny	K-S	0,000	0,000	0,164	0,234
	A-D	0,000	0,000	0,000	0,000
południowy	K-S	0,000	0,000	0,299	0,174
	A-D	0,000	0,000	0,003	0,006
wschodni	K-S	0,000	0,000	0,148	0,241
	A-D	0,000	0,000	0,000	0,001
północno-zachodni	K-S	0,000	0,000	0,251	0,284
	A-D	0,000	0,000	0,004	0,008
południowo-zachodni	K-S	0,000	0,000	0,357	0,261
	A-D	0,000	0,000	0,030	0,045
północny	K-S	0,000	0,000	0,149	0,275
	A-D	0,000	0,000	0,000	0,000
ogółem	K-S	0,000	0,000	0,000	0,000
	A-D	0,000	0,000	0,000	0,000

Źródło: obliczenia własne przy użyciu pakietu Mathematica

Ze względu na bardzo duże liczebności prób, które mogą powodować, że hipoteza zerowa jest odrzucana nawet przy minimalnych odchyleniach porównywanych rozkładów, w tab. 4.8 wyznaczono dodatkowo indeksy określające stopień dopasowania rozkładu logarytmiczno-normalnego do rozkładów empirycznych.

W tym przypadku, w odróżnieniu od poprzednich wyników dotyczących rozkładu Daguma, współczynniki podobieństwa struktur są mniejsze od 0,95 nawet dla metody największej wiarygodności, a dla metody momentów są dużo mniejsze niż w przypadku dopasowania rozkładem Daguma. Zastosowanie pozostałych dwóch metod estymacji wpływa na zwiększenie wartości współczynnika podobieństwa struktur (przekracza on 0,96) i zmniejszenie wartości indeksu Mortara, ale należy ostrożnie podejść do formułowania wniosków, gdyż w tych przypadkach konstrukcja estymatorów ma wpływ na wartości tych współczynników i należy wziąć pod uwagę wskaźnik A_3 , którego

wartość jest większa w porównaniu z MNW i świadczy o dużo większych rozbieżnościach między średnimi. Różnice w aproksymacjach rozkładów dochodów w Polsce i w podziale na sześć makroregionów przedstawione są na rys. 4.4 i 4.5.

Podsumowując, można stwierdzić, że porównanie wyników testów statystycznych (tab. 4.4 i 4.7) i wartości miar dopasowania z tab. 4.5 i 4.8 świadczą o zdecydowanie lepszym dopasowaniu rozkładów empirycznych za pomocą teoretycznego rozkładu Daguma. W przypadku rozkładu logarytmiczno-normalnego dla dwóch pierwszych metod estymacji, czyli MNW i MM, wartości p -value mniejsze niż 0,001 świadczą o tym, że dla wszystkich przyjmowanych poziomów istotności odrzucamy hipotezę o zgodności wszystkich rozważanych rozkładów z rozkładem logarytmiczno-normalnym. Nieakceptowalne wyniki uzyskano także, stosując miary podobieństwa rozkładów (tab. 4.8).

Dopasowanie rozkładu logarytmiczno-normalnego do rozważanych rozkładów empirycznych dla wybranych metod estymacji parametrów przedstawiono na rysunkach 4.4–4.5.

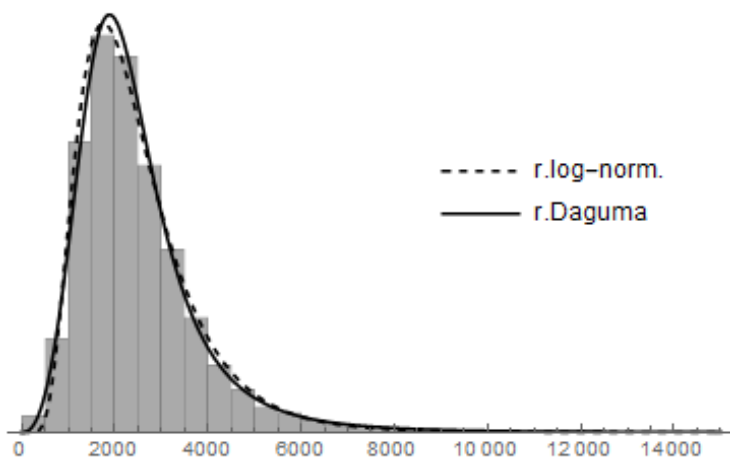
Tabela 4.8. Miary dopasowania empirycznych rozkładów dochodów do rozkładu logarytmiczno-normalnego

Region	Metoda estymacji	Wps	A_1	A_2	A_3
centralny	MNW	0,9345	0,1309	0,1487	0,9486
	MM	0,9134	0,1732	0,1948	1,4905
	MNK	0,9710	0,0579	0,1551	4,3802
	WMNK	0,9687	0,0626	0,1449	4,0056
południowy	MNW	0,9438	0,1124	3,3495	0,2405
	MM	0,9374	0,1251	0,1768	0,4840
	MNK	0,9607	0,0786	1,5579	1,4236
	WMNK	0,9682	0,0636	0,2450	1,3804
wschodni	MNW	0,9221	0,1558	0,1916	0,0974
	MM	0,9169	0,1661	0,1952	1,0175
	MNK	0,9748	0,0504	0,2981	2,3756
	WMNK	0,9700	0,0600	0,2181	2,2081

Tabela 4.8 (cd.)

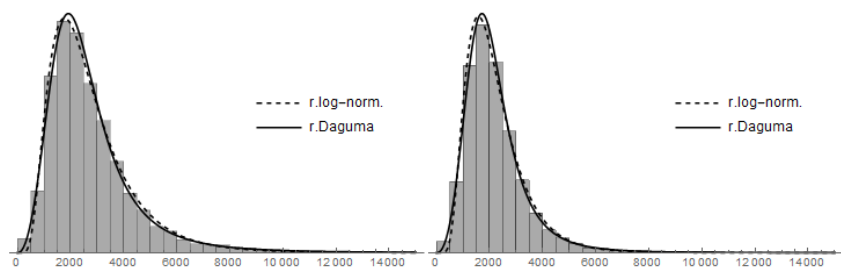
Region	Metoda estymacji	Wps	A_1	A_2	A_3
północno-zachodni	MNW	0,9379	0,1241	0,1964	0,4097
	MM	0,9447	0,1106	0,1986	0,4708
	MNK	0,9727	0,0547	0,3587	1,1126
	WMNK	0,9706	0,0587	0,3285	1,0827
południowo-zachodni	MNW	0,9377	0,1247	0,1703	0,0242
	MM	0,9320	0,1360	0,1758	0,7142
	MNK	0,9621	0,0757	0,2376	1,6110
	WMNK	0,9625	0,0751	0,2176	1,5281
północny	MNW	0,9014	0,1972	0,2429	1,2493
	MM	0,8702	0,2597	0,3061	0,7138
	MNK	0,9634	0,0732	0,2523	4,4578
	WMNK	0,9595	0,0810	0,2278	3,9364
ogółem	MNW	0,9270	0,1460	0,1750	0,2022
	MM	0,9063	0,1873	0,2131	0,9575
	MNK	0,9756	0,0488	0,2268	3,1142
	WMNK	0,9740	0,0520	0,2062	2,4814

Źródło: obliczenia własne przy użyciu pakietu Mathematica



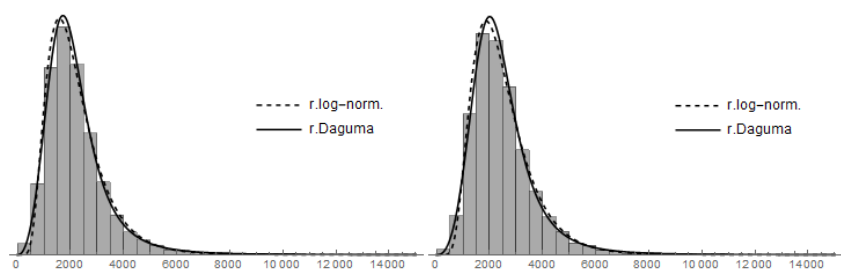
Rysunek 4.4. Aproksymacja rozkładu dochodów w Polsce przy zastosowaniu rozkładu logarytmiczno-normalnego i Daguma

Źródło: opracowanie własne



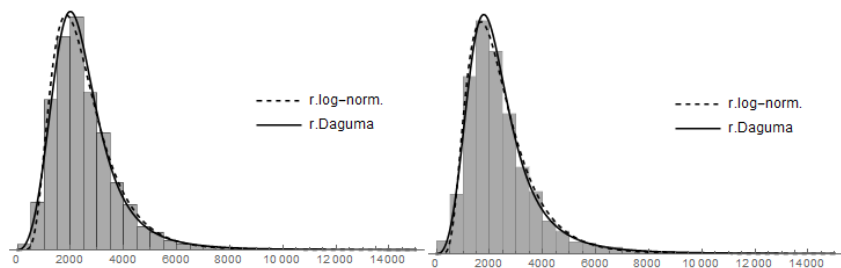
a) region centralny

b) region południowy



c) region wschodni

d) region północno-zachodni



e) region południowo-zachodni

f) region północny

Rysunek 4.5. Aproksymacja rozkładu dochodów w regionach przy zastosowaniu rozkładu logarytmiczno-normalnego i Daguma

Źródło: opracowanie własne

Analiza dopasowania rozkładów empirycznych rozkładem Singha-Maddali

Kolejnym analizowanym rozkładem był rozkład Singha-Maddali, który podobnie jak rozkład Daguma należy do rodziny rozkładów Burra (por. podrozdział 2.5) i jednocześnie stanowi szczególny przypadek rozkładu GB2. Rozkład ten wykorzystano do estymacji rozkładów dochodów na podstawie tych samych prób pochodzących z badania budżetów gospodarstw domowych. Podobnie jak w przypadku poprzednio analizowanych rozkładów teoretycznych, za pomocą czterech metod estymacji wyznaczone zostały oszacowania parametrów rozkładu Singha-Maddali $S - M(q, a, b)$, które przedstawiono w tab. 4.9. Następnie zbadano dobroć dopasowania, wykorzystując testy statystyczne i wybrane miary zgodności. Otrzymane wyniki przedstawione zostały w tab. 4.10 oraz 4.11.

Tabela 4.9. Oszacowania parametrów rozkładu Singha-Maddali aproksymujące rozkład dochodów gospodarstw domowych w regionach w Polsce

Region	Metoda estymacji	Parametry rozkładu Singha-Maddali		
		q	a	b
centralny	MNW	1,1026	2,9798	2495,7670
	MM	1,6961	2,8142	3207,8565
	MNK	1,0143	3,0770	2394,3791
	WMNK	1,0014	3,1148	2380,2691
południowy	MNW	1,1632	3,6624	2388,8246
	MM	1,3211	3,4053	2526,8904
	MNK	1,1919	3,6163	2412,8741
	WMNK	1,5553	3,6760	2382,1096
wschodni	MNW	1,2500	3,2867	2172,1795
	MM	1,3437	3,0801	2248,2636
	MNK	1,1690	3,3959	2107,3330
	WMNK	1,1614	3,4138	2101,2128
północno-zachodni	MNW	1,2821	3,5420	2494,6570
	MM	1,3525	3,4241	2557,2069
	MNK	1,2142	3,6134	2438,2229
	WMNK	1,1864	3,6577	2415,1425

Region	Metoda estymacji	Parametry rozkładu Singha-Maddali		
		q	a	b
południowo-zachodni	MNW	1,2397	3,4191	2487,4586
	MM	1,3315	3,2357	2571,7396
	MNK	1,1806	3,4658	2393,8877
	WMNK	1,1876	3,4825	2395,5471
północny	MNW	1,1344	3,1721	2263,3916
	MM	1,4376	2,6377	2452,3585
	MNK	0,9466	3,4952	2086,7719
	WMNK	0,9462	3,4993	2086,4644
ogółem	MNW	1,1426	3,3176	2343,2951
	MM	1,2927	2,9948	2481,0953
	MNK	1,0746	3,4211	2279,6475
	WMNK	1,0607	3,4508	2266,9574

Źródło: obliczenia własne przy użyciu pakietu Mathematica

Tabela 4.10. Wartości p -value dla testów Kołmogorowa-Smirnowa i Andersona-Darlinga weryfikujących hipotezy o zgodności rozkładów dochodów z rozkładem Singha-Maddali

Region	Test	Metoda estymacji			
		MNW	MM	MNK	WMNK
centralny	K-S	0,158	0,000	0,308	0,216
	A-D	0,095	0,000	0,157	0,174
południowy	K-S	0,416	0,061	0,561	0,374
	A-D	0,476	0,016	0,417	0,477
wschodni	K-S	0,640	0,006	0,828	0,800
	A-D	0,336	0,001	0,638	0,647
północno-zachodni	K-S	0,473	0,114	0,730	0,849
	A-D	0,525	0,115	0,679	0,716
południowo-zachodni	K-S	0,914	0,205	0,968	0,918
	A-D	0,769	0,091	0,785	0,821
północny	K-S	0,077	0,000	0,983	0,988
	A-D	0,015	0,000	0,225	0,225
ogółem	K-S	0,002	0,0000	0,152	0,139
	A-D	0,001	0,0000	0,016	0,019

Źródło: obliczenia własne przy użyciu pakietu Mathematica

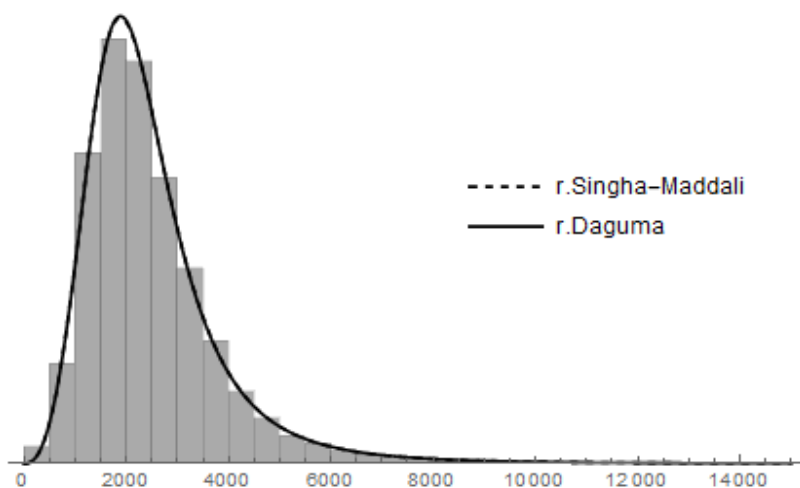
Tabela 4.11. Miary dopasowania empirycznych rozkładów dochodów do rozkładu Singha-Maddali

Region	Metoda estymacji	Wps	A_1	A_2	A_3
centralny	MNW	0,9834	0,03289	0,0414	1,6384
	MM	0,9292	0,1416	0,1698	2,1493
	MNK	0,9839	0,0319	0,0380	1,0326
	WMNK	0,9844	0,0309	0,0375	1,1647
południowy	MNW	0,9698	0,0604	0,1260	0,5615
	MM	0,9747	0,0504	0,0837	0,4807
	MNK	0,9697	0,0605	0,1226	0,6073
	WMNK	0,8870	0,2260	0,3225	15,4661
wschodni	MNW	0,9833	0,0333	0,0510	1,3298
	MM	0,9718	0,0562	0,0817	1,0224
	MNK	0,9851	0,0297	0,0499	1,1968
	WMNK	0,9903	0,0194	0,0325	1,2234
północno-zachodni	MNW	0,9890	0,0219	0,0460	0,6154
	MM	0,9846	0,0306	0,05405	0,4714
	MNK	0,9870	0,0260	0,0469	0,4271
	WMNK	0,9860	0,0279	0,0488	0,3846
południowo-zachodni	MNW	0,9745	0,0508	0,0553	0,9257
	MM	0,9721	0,0557	0,0749	0,6987
	MNK	0,9698	0,0602	0,0707	2,3854
	WMNK	0,9700	0,0598	0,0717	2,7216
północny	MNW	0,9726	0,0548	0,0730	1,6257
	MM	0,9212	0,1576	0,1920	5,2163
	MNK	0,9866	0,0268	0,0677	1,2029
	WMNK	0,9865	0,0269	0,0680	1,2298
ogółem	MNW	0,9848	0,0303	0,0466	1,3197
	MM	0,9609	0,0781	0,1031	0,9614
	MNK	0,9864	0,0273	0,0439	1,1862
	WMNK	0,9857	0,0286	0,0444	1,1930

Źródło: obliczenia własne przy użyciu pakietu Mathematica

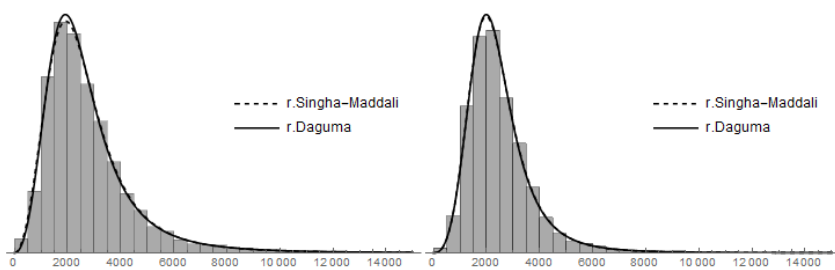
Na podstawie obliczonych wartości p -value oraz miar dopasowania można stwierdzić, że rozkład Singha-Maddali może być również stosowany do estymacji empirycznych rozkładów dochodów. Wybór metody szacowania parametrów rozkładu Singha-Maddali (spośród metod: MNW, MM, MNK i WMNK) w mniejszym stopniu, w porównaniu z poprzednimi rozkładami, wpływa na wynik aproksymacji rozkładu empirycznego rozkładem Singha-Maddali, chociaż w niektórych przypadkach (szczególnie regiony północny i południowy) różnice w miarach są zauważalne. W przypadku regionów centralnego i północnego wykorzystanie metody momentów do oszacowania parametrów prowadziło do otrzymania teoretycznego rozkładu Singha-Maddali, niedopasowanego do danych empirycznych (Wps dla obydwu regionów – około 0,92 i $A_3 = 5,2163$ dla regionu północnego), natomiast dla regionu południowego ważona metoda najmniejszych kwadratów nie dała dobrego dopasowania ($Wps = 0,887$ oraz $A_3 = 15,4661$).

Porównując dopasowanie rozkładu Daguma i Singha-Maddali do rozkładów dochodów gospodarstw domowych w Polsce i regionach, poza regionem centralnym, odnotowano brak widocznych różnic (por. rys. 4.6, 4.7).



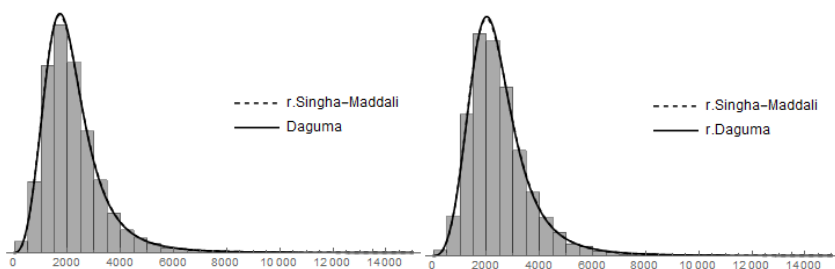
Rysunek 4.6. Aproksymacja rozkładu dochodów w Polsce przy zastosowaniu rozkładu Daguma i Singha-Maddali

Źródło: opracowanie własne



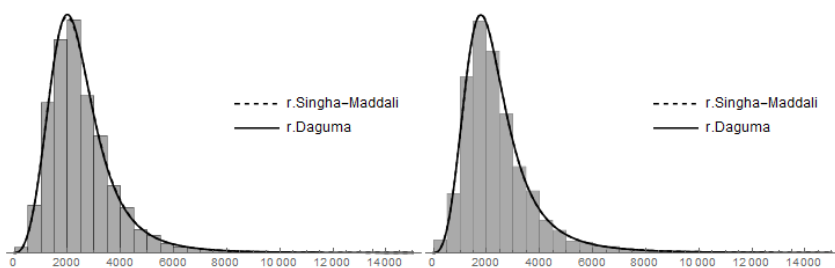
a) region centralny

b) region południowy



c) region wschodni

d) region północno-zachodni



e) region południowo-zachodni

f) region północny

Rysunek 4.7. Aproksymacja rozkładu dochodów w regionach przy zastosowaniu rozkładu Daguma i Singha-Maddali

Źródło: opracowanie własne

Podsumowanie uzyskanych wyników

W tab. 4.12 przedstawiono ranking dopasowania rozkładów dochodów w regionach do jednego z dwóch rozkładów przy różnych metodach szacowania parametrów rozkładu (wartość 1 oznacza najlepsze dopasowanie). Rankingi według współczynnika podobieństwa struktur i współczynnika Mortara są bardzo zbliżone, ale różnią się od rankingu według indeksu Pearsona oraz współczynnika uwzględniającego różnice między średnimi empirycznymi i teoretycznymi.

Tabela 4.12. Ranking dopasowania rozkładów dochodów w regionach Polski według współczynnika podobieństwa struktur i innych miar

Estymacja		Region						ogółem
		centralny	południowy	wschodni	północno-zachodni	południowo-zachodni	północny	
Współczynnik podobieństwa struktur								
Daguma	MNW	6	6	6	4	2	6	6
	MM	8	7	8	8	8	8	8
	MNK	5	1	5	5	6	4	3
	WMNK	2	3	2	7	3	3	5
Singha-Maddali	MNW	4	4	4	1	1	5	4
	MM	7	2	7	6	5	7	7
	MNK	3	5	3	2	7	1	1
	WMNK	1	8	1	3	4	2	2
Współczynnik Mortara								
Daguma	MNW	6	6	6	5	2	6	6
	MM	8	7	7	8	8	8	8
	MNK	1	5	5	6	5	4	3
	WMNK	2	2	2	7	3	3	5
Singha-Maddali	MNW	4	3	4	1	1	5	4
	MM	7	1	8	4	4	7	7
	MNK	3	4	3	2	7	1	1
	WMNK	1	8	1	3	6	2	2

Tabela 4.12 (cd.)

Estymacja		Region						ogółem
		centralny	południowy	wschodni	północno- -zachodni	południowo- -zachodni	północny	
Współczynnik Pearsona								
Daguma	MNW	6	4	6	6	2	6	6
	MM	8	6	8	8	8	8	8
	MNK	4	3	3	4	4	4	4
	WMNK	3	1	1	5	3	3	3
Singha- -Maddali	MNW	5	7	5	1	1	5	5
	MM	7	2	7	7	7	7	7
	MNK	2	5	4	2	5	1	3
	WMNK	1	8	2	3	6	2	1
Współczynnik A_3								
Daguma	MNW	7	3	8	7	5	7	8
	MM	5	4	3	5	4	1	2
	MNK	1	1	1	1	1	4	1
	WMNK	2	2	2	2	2	5	4
Singha- -Maddali	MNW	6	6	7	8	6	6	7
	MM	8	5	4	6	3	8	3
	MNK	3	7	5	4	7	2	5
	WMNK	4	8	6	3	8	3	6

Źródło: opracowanie własne

W tab. 4.13 przedstawiono dwa najlepsze, według współczynnika A_3 , dopasowania rozkładów teoretycznych. Podano ich wartości oczekiwane, odchylenia standardowe i współczynniki Giniego, wykorzystywane do porównywania regionów pod względem nierównomierności rozkładów.

Okazuje się, że najbardziej nierównomierny rozkład obserwuje się w regionie centralnym, w którym jest największe zróżnicowanie dochodów (2212,307 PLN) i największy współczynnik Giniego (0,324). Oszacowania tych wielkości z próby również wskazują na największe

nierównomierności rozkładu ekwiwalentnych dochodów gospodarstw domowych (por. tab. 4.2, $G = 0,318$).

Według pozostałych miar dopasowania w regionach północno-zachodnim, południowo-zachodnim i północnym rozkłady dochodów powinny być aproksymowane rozkładem Singha-Maddali. Również w regionie centralnym, według współczynnika podobieństwa struktur i współczynnika Pearsona, rozkład Singha-Maddali dobrze aproksymuje rozkład empiryczny dochodów.

Tabela 4.13. Teoretyczne rozkłady aproksymujące rozkłady dochodów w regionach i ich podstawowe parametry

Region	Rozkład	Wartość oczekiwana (PLN)	Odchylenie standardowe (PLN)	Indeks Giniego
centralny	D(1,03; 3,062; 2345,091)	2849,921	2212,307	0,324
	D(1,021; 3,095; 2355,396)	2839,226	2166,649	0,322
południowy	D(0,876; 3,996; 2367,181)	2508,588	1351,500	0,259
	D(0,899; 3,889; 2343,163)	2506,320	1341,566	0,257
wschodni	D(0,886; 3,714; 2072,790)	2236,935	1326,074	0,278
	D(0,879; 3,746; 2077,869)	2232,351	1309,639	0,276
północno-zachodni	D(0,858; 4,042; 2393,018)	2511,420	1338,080	0,258
	D(0,873; 4,052; 2376,293)	2508,886	1327,361	0,256
południowo-zachodni	D(0,879; 3,812; 2393,888)	2562,898	1468,277	0,271
	D(0,877; 3,853; 2395,547)	2556,128	1444,206	0,269
północny	D(0,559; 3,512; 2780,726)	2491,651	1788,741	0,336
	S-M(0,947; 3,495; 2086,772)	2479,406	1661,684	0,294
ogółem	D(0,955; 3,554; 2254,379)	2531,113	1572,932	0,285
	D(0,623; 3,814; 2709,366)	2528,398	1575,633	0,301

Źródło: opracowanie własne

Z przeprowadzonych analiz w sposób jednoznaczny wynika, że rozkład Daguma najlepiej odzwierciedla rozkłady dochodów ekwiwalentnych gospodarstw domowych w regionie południowym, wschodnim oraz w całym kraju. Należy podkreślić istotne znaczenie na dopasowanie rozkładu teoretycznego metody estymacji jego parametrów. W tych przypadkach metoda najmniejszych kwadratów oraz ważona metoda najmniejszych kwadratów okazały się najefektywniejsze.

Wyznaczone z rozkładów teoretycznych wartości średnie są zwykle mniejsze od wartości wyznaczonych na podstawie prób losowych pochodzących z badania budżetów gospodarstw domowych, a współczynniki Giniego są nieznacznie większe.

Najslabiej dopasowanym do rozkładów empirycznych okazał się rozkład logarytmiczno-normalny, co potwierdziły stosowane testy statystyczne i miary „dobroci” dopasowania.

Analizując dochody w Polsce, nie tylko bada się ich rozkłady oraz stopień nierównomierności w różnych regionach, lecz także w różnych grupach społeczno-ekonomicznych czy typach rodzin. Wyniki takich badań, w tym modelowania rozkładów dochodów ekwiwalentnych gospodarstw domowych – *pracowników, rolników, pracujących na własny rachunek, emerytów i rencistów*, utrzymujących się z *niezarobkowych źródeł* – przedstawione są w pracy Pekasiewicz, Jędrzejczak (2017). Okazuje się, że rozkład Daguma może być wykorzystany również do estymacji parametrów w grupach społeczno-ekonomicznych, gdyż współczynniki podobieństwa struktur wyznaczone dla rozkładów empirycznych i teoretycznych są większe od 0,97, poza grupą rolników, dla której współczynnik ten wyniósł 0,949 (por. Jędrzejczak, Pekasiewicz, 2018). Dobór modeli teoretycznych odzwierciedlających rozkłady dochodów w Polsce był przedmiotem rozważań m.in. w pracach Salamagi (2016a; 2016b). W artykułach tych przedstawione są rezultaty modelowania rozkładów dochodów kobiet i mężczyzn w województwie małopolskim oraz rozkładów dochodów mieszkańców Krakowa w roku 2013 za pomocą rozkładów Singha-Maddali, Fiska, logarytmiczno-normalnego i gamma. Stosowane metody estymacji parametrów rozkładów to metoda największej wiarygodności, metoda najmniejszych kwadratów oraz minimalizacji statystyki chi-kwadrat Pearsona.

4.4. Aproksymacja rozkładów dochodów zamożnych gospodarstw domowych w Polsce

Przy całościowej analizie dochodów gospodarstw domowych wykorzystuje się rozkłady jednomodalne, dodatnio asymetryczne. W przypadku badań dotyczących gospodarstw domowych o wysokich dochodach,

przekraczających ustaloną wartość, do aproksymacji rozkładów stosuje się rozkłady zeromodalne, w tym rozkład Pareto.

Przeprowadzone analizy dotyczyły grup gospodarstw domowych nazywanych zamożnymi, czyli o dochodach przekraczających tak zwaną granicę bogactwa. Oczywiście na zamożność nie tylko wpływają dochody. Duże znaczenie mają też wielkości dochodów niewydatkowanych na konsumpcję i kumulowanych w postaci oszczędności rzeczowych oraz finansowych i generujących posiadanie aktywów, ale przeprowadzone analizy dotyczą wyłącznie rozkładów dochodów tej grupy gospodarstw domowych.

Wartość granicy (linii) zamożności (bogactwa) ustalana jest w różny sposób, m.in. na poziomie dwóch lub trzech median rozkładu dochodów ekwiwalentnych lub wybranego kwantyla rozkładu dochodów. W badaniach przyjęto dwie granice zamożności, na poziomie:

- dwóch median, czyli $y_{r1}^* = 2Me$;
- trzech median, czyli $y_{r2}^* = 3Me$.

Do analiz wykorzystano dane pochodzące z Badania Budżetów Gospodarstw Domowych z roku 2016. Oszacowane granice zamożności dla tego roku wynoszą:

- $y_{r1}^* = 4477,12$ PLN;
- $y_{r2}^* = 6715,68$ PLN.

Liczba gospodarstw o dochodach powyżej dwóch median wynosi $N_1 = 2686$, co stanowi około 7,3% badanych gospodarstw domowych, a powyżej trzech median $N_2 = 698$, czyli około 1,9% badanych gospodarstw. Podstawowe charakterystyki przedstawione zostały w tab. 4.14 i 4.15.

Tabela 4.14. Charakterystyki rozkładów dochodów gospodarstw domowych przekraczających granicę y_{r1}^* w ujęciu regionalnym

Region	Liczba gospodarstw zamożnych	Odsetek gospodarstw zamożnych	Średnia	Odchylenie standardowe
centralny	964	12,09	6706,09	3094,07
południowy	429	5,80	6072,70	2000,84
wschodni	286	4,59	6192,45	2392,10

Tabela 4.14 (cd.)

Region	Liczba gospodarstw zamożnych	Odsetek gospodarstw zamożnych	Średnia	Odchylenie standardowe
północno-zachodni	333	6,00	5894,43	1869,08
południowo-zachodni	258	6,69	6116,75	2269,18
północny	416	7,40	6734,47	3763,43
ogółem	2686	7,33	6396,56	2810,90

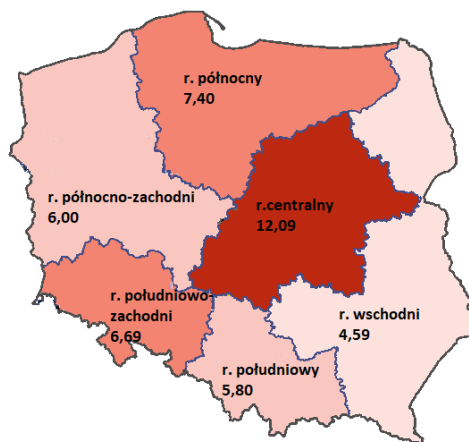
Źródło: obliczenia własne

Tabela 4.15. Charakterystyki rozkładów dochodów gospodarstw domowych przekraczających granicę $y_{r_2}^*$ w ujęciu regionalnym

Region	Liczba gospodarstw zamożnych	Odsetek gospodarstw zamożnych	Średnia	Odchylenie standardowe
centralny	308	3,86	9779,81	3916,22
południowy	95	1,28	8898,49	2569,05
wschodni	64	1,03	9521,49	3178,24
północno-zachodni	61	1,10	8902,85	2546,95
południowo-zachodni	52	1,35	9504,19	3155,37
północny	118	2,10	10341,90	5545,54
ogółem	698	1,91	9634,02	3902,89

Źródło: obliczenia własne

Największym odsetkiem gospodarstw zamożnych charakteryzuje się region centralny i tam jest też wysoki średni dochód, natomiast najniższym odsetkiem – region wschodni (por. rys. 4.8). Najbardziej zróżnicowany pod względem wielkości wysokości dochodów jest region północny, odchylenie standardowe stanowi prawie 56% wartości średniej.



Rysunek 4.8. Odsetek zamożnych w poszczególnych regionach

Źródło: opracowanie własne

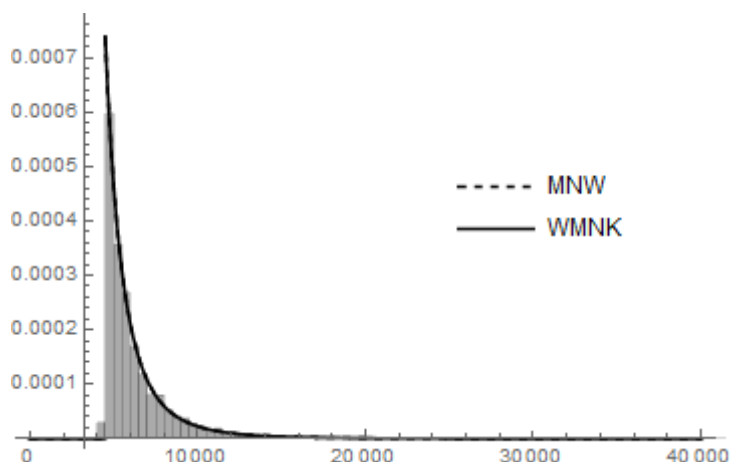
W celu aproksymacji rozkładów empirycznych rozkładem Pareto jego parametry oszacowano, podobnie jak rozkładów Daguma, logarytmiczno-normalnego i Singha-Maddali, metodami: największej wiarygodności, momentów, najmniejszych kwadratów i ważoną metodą najmniejszych kwadratów. Uzyskane rezultaty dla gospodarstw domowych w Polsce o dochodach przekraczających ustalone granice zamożności zostały przedstawione w tab. 4.16.

Tabela 4.16. Oszacowania parametrów rozkładu Pareto aproksymującego rozkład dochodów zamożnych gospodarstw domowych w Polsce

Granica zamożności	Metoda estymacji	Parametry rozkładu Pareto	
		θ	a
$y_{r1}^* = 2Me$	MNW	4477,1954	3,3345
	MM	4561,6577	3,4860
	MNK	4460,3429	3,3097
	WMNK	4472,6394	3,3480
$y_{r2}^* = 3Me$	MNW	6717,4791	3,2524
	MM	7005,7461	3,6655
	MNK	6730,0090	3,2334
	WMNK	6713,1315	3,2034

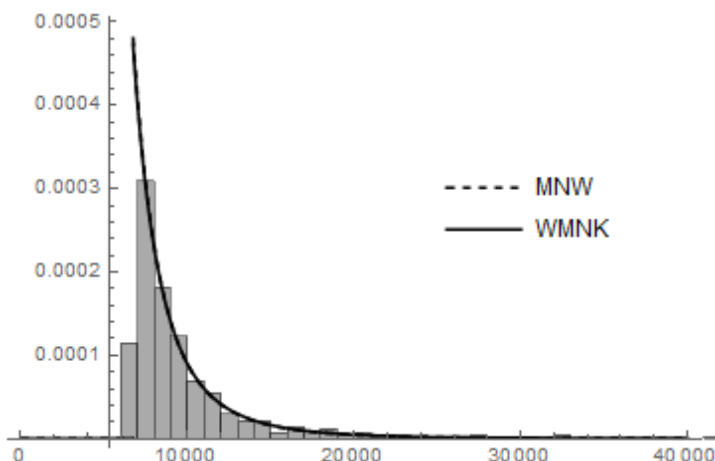
Źródło: obliczenia własne przy użyciu pakietu Mathematica

Oszacowania parametrów rozkładu Pareto zawarte w tab. 4.16 pokazują, że zastosowanie metod MNW i WMNK prowadzi do uzyskania bardzo podobnych wartości estymatorów, a tym samym prawie identycznych dopasowań rozkładu empirycznego do rozkładu Pareto, co potwierdzają także rys. 4.9–4.10. Prawdliwość ta zauważalna jest dla obu przyjętych progów bogactwa – dwukrotności oraz trzykrotności mediany dochodów ekwiwalentnych.



Rysunek 4.9. Aproksymacja rozkładów dochodów zamożnych gospodarstw domowych dla $y_{r1}^* = 2Me$

Źródło: opracowanie własne



Rysunek 4.10. Aproksymacja rozkładów dochodów zamożnych gospodarstw domowych dla $y_{r2}^* = 3Me$

Źródło: opracowanie własne

W celu zbadania zgodności empirycznych rozkładów dochodów zamożnych gospodarstw w Polsce z rozkładem Pareto o parametrach ustalonych na podstawie rezultatów stosowania różnych metod estymacji wykorzystano testy Kołmogorowa-Smirnowa i Andersona-Darlinga oraz obliczono miary dopasowań: współczynnik podobieństwa struktur, współczynnik Mortara, współczynnik Pearsona oraz współczynnik oparty na porównaniu średnich empirycznych i teoretycznych (por. tab. 4.17, 4.18).

Tabela 4.17. Wartości p -value dla testów Kołmogorowa-Smirnowa i Andersona-Darlinga

Granica zamożności	Test	Metoda estymacji			
		MNW	MM	MNK	WMNK
$y_{r1}^* = 2Me$	K-S	0,223	0,000	0,454	0,438
	A-D	0,049	0,000	0,362	0,348
$y_{r2}^* = 3Me$	K-S	0,528	0,000	0,777	0,630
	A-D	0,567	0,000	0,000	0,814

Źródło: obliczenia własne przy użyciu pakietu Mathematica

Tabela 4.18. Miary dopasowania empirycznych rozkładów dochodów zamożnych gospodarstw domowych w Polsce do rozkładu Pareto

Granica zamożności	Metoda estymacji	Wps	A_1	A_2	A_3
$y_{r1}^* = 2Me$	MNW	0,9816	0,0362	0,1132	0,0239
	MM	0,9724	0,0547	0,1338	0,0006
	MNK	0,9710	0,0449	0,1123	0,0795
	WMNK	0,9796	0,0369	0,1150	0,2987
$y_{r2}^* = 3Me$	MNW	0,9775	0,0423	0,1522	6,6191
	MM	0,9659	0,0666	0,1988	7,3472
	MNK	0,9796	0,0379	0,1494	6,1430
	WMNK	0,9789	0,0392	0,1472	5,9635

Źródło: obliczenia własne przy użyciu pakietu Mathematica

Zarówno wartości p -value stosowanych testów statystycznych, jak i miary dopasowania świadczą o zgodności rozkładów dochodów zamożnych gospodarstw domowych z rozkładem Pareto, gdy za

wartości parametrów przyjmie się oszacowania z próby uzyskane metodą największej wiarygodności lub ważoną metodą najmniejszych kwadratów. Metody momentów, podobnie jak w przypadku estymacji parametrów wcześniej rozważanych rozkładów teoretycznych (Daguma, logarytmiczno-normalnego, Singha-Maddali), nie powinno się stosować.

Przeprowadzono również analizę rozkładów dochodów gospodarstw zamożnych w poszczególnych regionach Polski przy granicy zamożności równej dwóm medianom. Przypadek granicy zamożności na poziomie trzech median nie był rozważany ze względu na zbyt mało liczne próby dla poszczególnych regionów. Wartości uzyskanych oszacowań parametrów rozkładu Pareto zostały przedstawione w tab. 4.19.

Tabela 4.19. Oszacowania parametrów rozkładu Pareto aproksymującego rozkłady dochodów zamożnych gospodarstw domowych w regionach

Region	Metoda estymacji	Parametry rozkładu Pareto	
		θ	a
centralny	MNW	4477,1954	2,9580
	MM	4726,7187	3,3880
	MNK	4449,0051	2,8610
	WMNK	4467,1989	2,9179
południowy	MNW	4496,0678	3,8023
	MM	4626,4553	4,1989
	MNK	4471,8797	3,6082
	WMNK	4478,3264	3,6718
wschodni	MNW	4480.0000	3,6356
	MM	4553,9624	4,1989
	MNK	4452,4410	3,6450
	WMNK	4466,6171	3,6702
północno-zachodni	MNW	4483,5925	4,1815
	MM	4527,7362	4,3129
	MNK	4452,5977	4,0304
	WMNK	4468,8165	4,1287

Region	Metoda estymacji	Parametry rozkładu Pareto	
		θ	a
południowo-zachodni	MNW	4490,8500	3,7818
	MM	4540,2680	3,8800
	MNK	4450,3621	4,0364
	WMNK	4509,6612	3,9288
północny	MNW	4477,8363	3,0300
	MM	4527,7499	3,0518
	MNK	4456,2472	3,0171
	WMNK	4468,2683	3,0525

Źródło: obliczenia własne przy użyciu pakietu Mathematica

Podobnie jak przy szacowaniu parametrów rozkładów charakteryzujących dochody wszystkich grup dochodowych ogółem i w regionach (rozkład Daguma, logarytmiczno-normalny, Singha-Maddali), najgorsze rezultaty otrzymano, stosując metodę momentów. W przypadku tej metody, wykorzystując zarówno test Kołmogorowa-Smirnowa, jak i Andersona-Darlinga, otrzymano decyzje o odrzuceniu hipotezy o zgodności rozkładów dochodów gospodarstw zamożnych z rozkładem Pareto o wartościach oszacowanych na podstawie próby metodą momentów, na korzyść hipotezy alternatywnej mówiącej, że nie jest to rozkład Pareto o zadanych wartościach parametrów. Pozostanie przy tej samej klasie rozkładów i zmiana wartości parametrów na inne (uzyskane z oszacowania innymi metodami estymacji) pozwoliła uzyskać dobre dopasowanie rozkładu Pareto do danych empirycznych (por. tab. 4.20, 4.21).

Tabela 4.20. Wartości p -value dla testów weryfikujących hipotezy o zgodności rozkładów dochodów grup zamożnych z rozkładem Pareto

Region	Test	Metoda estymacji			
		MNW	MM	MNK	WMNK
centralny	K-S	0,720	0,000	0,817	0,912
	A-D	0,109	0,000	0,684	0,726
południowy	K-S	0,537	0,000	0,900	0,893
	A-D	0,105	0,000	0,701	0,652

Tabela 4.20 (cd.)

Region	Test	Metoda estymacji			
		MNW	MM	MNK	WMNK
wschodni	K-S	0,816	0,032	0,947	0,940
	A-D	0,613	0,000	0,856	0,927
północno-zachodni	K-S	0,505	0,033	0,840	0,750
	A-D	0,371	0,000	0,787	0,792
południowo-zachodni	K-S	0,534	0,224	0,885	0,890
	A-D	0,173	0,000	0,000	0,000
północny	K-S	0,710	0,956	0,484	0,902
	A-D	0,220	0,000	0,950	0,982

Źródło: obliczenia własne

Tabela 4.21. Miary dopasowania empirycznych rozkładów dochodów zamożnych gospodarstw domowych do rozkładu Pareto

Region	Metoda estymacji	Wps	A_1	A_2	A_3
centralny	MNW	0,9796	0,0378	0,1325	0,8534
	MM	0,9406	0,1173	0,2086	0,0001
	MNK	0,9618	0,0551	0,1272	1,9528
	WMNK	0,9730	0,0444	0,1288	1,3290
południowy	MNW	0,9732	0,0502	0,1690	0,4555
	MM	0,9680	0,0619	0,2039	0,0003
	MNK	0,9685	0,0542	0,1560	1,8383
	WMNK	0,9748	0,0463	0,1590	1,3287
wschodni	MNW	0,9745	0,0478	0,1898	0,2047
	MM	0,9706	0,0561	0,2076	3,5949
	MNK	0,9597	0,0575	0,1947	0,9235
	WMNK	0,9703	0,0478	0,1967	0,8544
północno-zachodni	MNW	0,9670	0,0637	0,1812	0,0266
	MM	0,9653	0,0676	0,1931	0,0000
	MNK	0,9424	0,0907	0,1750	0,4640
	WMNK	0,9593	0,0714	0,1778	0,0461
południowo-zachodni	MNW	0,9652	0,0665	0,1859	0,1889
	MM	0,9616	0,0740	0,1972	0,0000
	MNK	0,9669	0,0403	0,2294	2,3928
	WMNK	0,9715	0,0545	0,2036	1,1129

Region	Metoda estymacji	Wps	A_1	A_2	A_3
północny	MNW	0,9826	0,0335	0,1638	0,7601
	MM	0,9799	0,0391	0,1659	0,0000
	MNK	0,9696	0,0457	0,1641	1,0350
	WMNK	0,9793	0,0343	0,1677	1,3425

Źródło: obliczenia własne

Obliczone współczynniki pozwoliły dokonać uporządkowania zgodności aproksymowanych rozkładów z rozkładem Pareto dla poszczególnych regionów (tab. 4.22). Gdy weźmie się pod uwagę współczynnik podobieństwa struktur, najczęściej najlepiej dopasowany rozkład uzyskuje się, szacując parametry rozkładu Pareto metodą największej wiarygodności. Prawie analogiczny jest ranking według współczynnika Mortara, czego należało się spodziewać ze względu na podobną konstrukcję tej miary. Współczynniki Pearsona i A_3 inaczej porządkują efektywność metod estymacji parametrów rozkładu Pareto, ale należy zauważyć, że wyznaczanie estymatorów metodą momentów związane jest z porównywaniem średniej: empirycznej z teoretyczną, czyli mała wartość A_3 jest naturalna i budzi wątpliwości co do jej przydatności jako miary „dobroci” dopasowania.

Tabela 4.22. Ranking dopasowania rozkładów dochodów gospodarstw domowych o wysokich dochodach w regionach Polski

Estymacja	Region						ogółem
	centralny	południowy	wschodni	północno- zachodni	południowo- zachodni	północny	
Współczynnik podobieństwa struktur							
MNW	1	2	1	1	3	1	1
MM	4	4	2	2	4	2	3
MNK	3	3	4	4	2	4	4
WMNK	2	1	3	3	1	3	2

Tabela 4.22 (cd.)

Estymacja	Region						ogółem
	centralny	południowy	wschodni	północno- -zachodni	południowo- -zachodni	północny	
Współczynnik Mortara							
MNW	1	2	1	1	3	1	1
MM	4	4	3	2	4	3	4
MNK	3	3	4	4	1	4	3
WMNK	2	1	2	3	2	2	2
Współczynnik Pearsona							
MNW	3	3	1	3	1	1	2
MM	4	4	4	4	2	3	4
MNK	1	1	2	1	4	2	1
WMNK	2	2	3	2	3	4	3
Współczynnik A_3							
MNW	2	2	1	2	3	2	2
MM	1	1	4	1	1	1	1
MNK	4	3	3	3	2	3	3
WMNK	3	4	2	4	4	4	4

Źródło: obliczenia własne

W tab. 4.23 zostały podane parametry rozkładu Pareto $Pa(\theta, a)$ aproksymującego rozkłady dochodów w tych grupach. Wybrano te parametry dla których dopasowanie, według współczynnika podobieństwa struktur, jest najlepsze, czyli dla regionów: centralnego, wschodniego, północno-zachodniego, północnego i ogółu zamożnych gospodarstw domowych – parametry uzyskane metodą największej wiarygodności, a dla regionu południowego i południowo-zachodniego – ważoną metodą najmniejszych kwadratów.

Oszacowane wartości parametru a rozkładu $Pa(\theta, a)$ są większe niż 2, co oznacza, że istnieje zarówno wartość oczekiwana, jak i odchylenie standardowe zmiennych losowych o rozkładach Pareto. Można je oszacować z wzorów (1.7) i (1.8), wykorzystując wartości estymatorów pa-

rametrów a oraz θ . Ponadto, na podstawie oszacowanych parametrów rozkładu Pareto, można wyznaczyć także miary nierównomierności – współczynniki Giniego i Pietry (por. tab. 4.23). Współczynniki Giniego i Pietry są największe dla regionu centralnego i północnego.

Tabela 4.23. Parametry rozkładów aproksymujących rozkłady dochodów zamożnych gospodarstw domowych w regionach

Region	Parametry rozkładu		Wartość oczekiwana	Odchylenie standardowe	Współczynnik Giniego	Indeks Pietry
	θ	a				
centralny	4477,20	2,96	6761,49	4011,07	0,2033	0,7048
południowy	4478,33	3,67	6155,61	2486,45	0,1577	0,6889
wschodni	4480,00	3,64	6176,97	2528,15	0,1592	0,6894
północno-zachodni	4483,59	4,18	5893,52	1952,35	0,1359	0,6811
południowo-zachodni	4509,66	3,93	6048,79	2196,31	0,1458	0,6846
północny	4477,84	3,03	6683,67	3783,39	0,1976	0,7029
ogółem	4477,20	3,33	6398,75	3040,51	0,1767	0,6955

Źródło: obliczenia własne

Na podstawie wartości współczynnika a można wnioskować też o innych cechach rozkładu, np. z wzoru (1.15) wynika, że w regionie centralnym 10% najbogatszych gospodarstw domowych skupia aż 22,01% dochodów wszystkich gospodarstw domowych uznanych za zamożne (dochodach powyżej dwóch median), zaś dla porównania w regionie północno-zachodnim odsetek ten wynosi 17,34%.

4.5. Analiza symulacyjna własności estymatorów parametrów rozkładu Daguma

Wyniki badań empirycznych uzyskane w podrozdziale 4.3 mogą sugerować, że metoda najmniejszych kwadratów (MNK) oraz ważona metoda najmniejszych kwadratów (WMNK) dają dokładniejsze, a ostrożniej formułując

hipotezę badawczą – nie gorsze oszacowania parametrów niż inne stosowane metody estymacji, czyli metoda największej wiarygodności (MNW) i metoda momentów (MM). Aby zweryfikować tak sformułowaną hipotezę, przeprowadzono badanie symulacyjne dotyczące własności estymatorów parametrów rozkładu Daguma otrzymanych różnymi metodami estymacji.

Do szacowania parametrów rozkładu Daguma $D(a, v, b)$ wykorzystano cztery metody estymacji:

- metodę największej wiarygodności;
- metodę momentów ważonych prawdopodobieństwami;
- metodę najmniejszych kwadratów;
- ważoną metodę najmniejszych kwadratów.

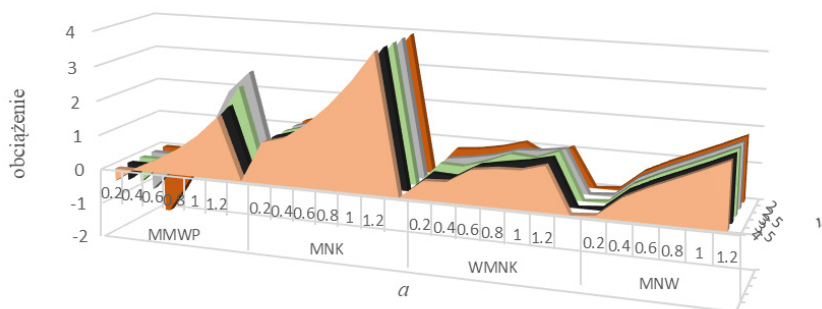
Metoda momentów została pominięta ze względu na fakt, że parametr ϑ rozkładu Daguma aproksymującego rozkłady dochodów gospodarstw domowych w Polsce nie zawsze jest większy od 3, czyli nie zawsze istnieją trzy momenty zwykle niezbędne do oszacowania trzech parametrów rozkładu. Poza tym wyniki przeprowadzonych badań empirycznych wskazywały, że wartości estymatorów uzyskanych tą metodą prowadziły do najmniej dokładnych dopasowań rozkładów teoretycznych do empirycznych. Nie uwzględniono również metody percentyli i kwantylowej metody najmniejszych kwadratów ze względu na konieczność ustalania rzędów wykorzystywanych kwantyli, co w przypadku rozkładów z trzema parametrami nie jest łatwe bez wykonania dodatkowych skomplikowanych analiz.

W ramach eksperymentu symulacyjnego parametry rozkładu szacowane były na podstawie n -elementowych prób prostych dla $n = 1000$ oraz $n = 2000$. Należy podkreślić, że dla rozkładu Daguma próby muszą być bardzo duże, co najmniej 1000-elementowe, co wykazano w pracy Domańskiego i Jędrzejczak (1998). Procedurę estymacji powtarzano $N = 1000$ razy i wyznaczano względne obciążenia oraz względne błędy średniokwadratowe estymatorów (RMSE), korzystając z wzorów:

$$B(\hat{\theta}) = \frac{\frac{1}{N} \sum_{i=1}^N (\hat{\theta}_i - \theta)}{\theta} \cdot 100\% \quad (4.6)$$

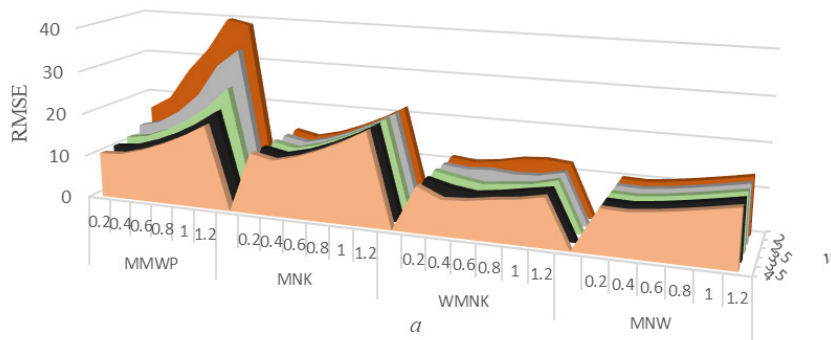
$$RMSE(\hat{\theta}) = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{\theta}_i - \theta)^2}}{\theta} \cdot 100\% \quad (4.7)$$

Relatywne obciążenia i błędy średniokwadratowe dla estymatorów parametrów kształtu a i v dla wybranych rozkładów zostały przedstawione na rys. 4.11–4.14, a dokładne wartości oszacowań parametrów rozkładu Daguma zawarto w tab. 4.24.



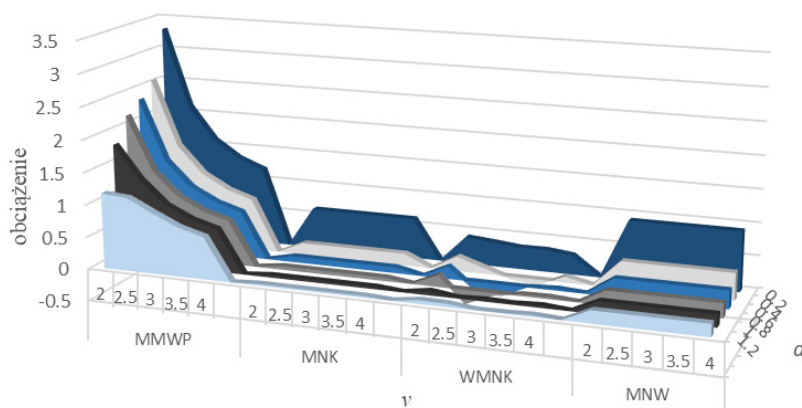
Rysunek 4.11. Oszacowane względne obciążenia różnych estymatorów parametru a

Źródło: Jędrzejczak, Pekasiewicz, Zieliński (2018)



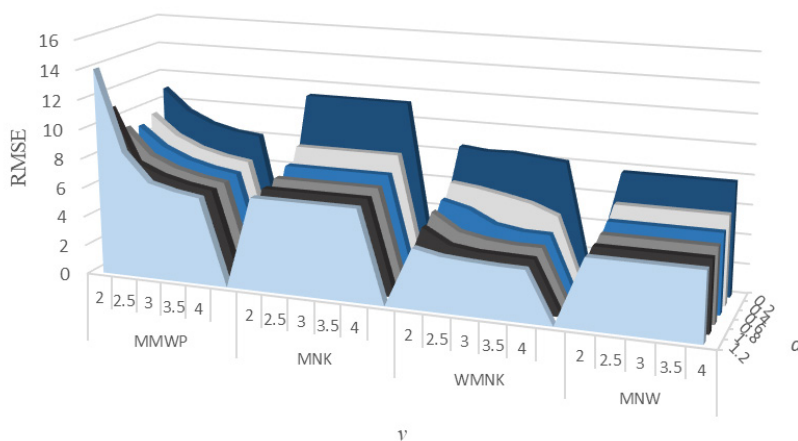
Rysunek 4.12. Oszacowane relatywne błędy średniokwadratowe różnych estymatorów parametru a

Źródło: Jędrzejczak, Pekasiewicz, Zieliński (2018)



Rysunek 4.13. Oszacowane względne obciążenia różnych estymatorów parametru ν

Źródło: Jędrzejczak, Pekasiewicz, Zieliński (2018)



Rysunek 4.14. Oszacowane relatywne błędy średniokwadratowe różnych estymatorów parametru ν

Źródło: Jędrzejczak, Pekasiewicz, Zieliński (2018)

Z analiz wynika, że estymatory parametru a uzyskane metodą WMNK lub MNW charakteryzują się najmniejszym relatywnym błędem średniokwadratowym. Wśród rozważanych estymatorów parametru ν najmniejsze obciążenia są obserwowane dla estymatorów otrzymanych metodą WMNK, a najmniejsze błędy średniokwadratowe – dla estymatorów uzyskanych metodą MNW lub WMNK. Można zauważyć, że estymato-

ry *MMWP* charakteryzują się najmniejszą precyzją (największa wartość RMSE dla estymatora parametru a i największa wartość RMSE dla estymatorów parametru v , gdy $v < 2,5$) spośród wszystkich statystyk uwzględnionych w badaniu. Precyzja estymatorów dla $n = 1000$ jest bardzo słaba, dopiero dla $n = 2000$ RMSE nie przekracza na ogół 10% szacowanych parametrów. Rekomendowane byłyby więc liczebności prób $n > 2000$.

Tabela 4.24. Relatywne błędy średniokwadratowe estymatorów parametrów rozkładu Daguma

Rozkład	Metoda estymacji	$RMSE(\hat{a})$		$RMSE(\hat{v})$		$RMSE(\hat{b})$	
		$n = 1000$	$n = 2000$	$n = 1000$	$n = 2000$	$n = 1000$	$n = 2000$
$D(0,2;2;1)$	MNW	9,993	6,982	7,703	5,240	8,354	6,065
	MMWP	16,510	12,574	10,944	8,097	11,971	9,037
	MNK	14,054	10,022	11,210	7,860	9,959	7,390
	WMNK	11,088	7,883	8,522	5,976	8,848	6,544
$D(0,6;2,5;1)$	MNW	10,438	7,243	5,412	3,743	6,403	4,590
	MMWP	17,572	12,359	7,659	5,572	9,727	7,204
	MNK	14,306	9,892	7,110	4,920	8,314	5,990
	WMNK	10,684	7,279	5,312	3,804	6,462	4,541
$D(0,8;3;1)$	MNW	11,288	7,800	5,066	3,517	5,643	4,013
	MMWP	17,261	11,820	6,742	4,825	8,053	5,840
	MNK	16,328	11,120	6,689	4,623	7,717	5,487
	WMNK	9,601	6,400	3,942	2,800	4,635	3,239
$D(1,2;4;1)$	MNW	13,313	9,119	4,691	3,277	4,786	3,373
	MMWP	19,513	12,884	4,180	4,825	6,552	4,659
	MNK	21,702	14,244	6,302	4,356	7,109	4,966
	WMNK	11,131	7,433	3,475	2,497	3,797	2,680

Źródło: Jędrzejczak, Pekasiewicz, Zieliński (2018)

Wyniki symulacji pokazują ponadto, że zarówno obciążenia, jak i wariancje estymatorów parametru a są większe niż dla estymatorów parametru v dla wszystkich rozważanych przypadków i metod estymacji, co przekłada się na wartości błędów całkowitych estymacji mierzonych za pomocą względnego błędu średniokwadratowego. Dla rozkładów Daguma, opisujących dochody o mniejszej nierównomierności (np. $v = 4$), ważona metoda

najmniejszych kwadratów wydaje się najodpowiedniejsza do estymacji parametrów, jednak w przypadku większej koncentracji (mniejsze wartości parametrów a i v) nie uzyskujemy właściwie żadnej korzyści ze stosowania tej metody w porównaniu z metodą największej wiarygodności.

Można również zauważyć, że precyzja estymatorów parametrów v i a rozkładu Daguma wzrasta wraz ze wzrastającymi wartościami parametru v , co odpowiada coraz cieńszemu prawemu ogonowi i mniejszej asymetrii rozkładu, i wraz z malejącymi wartościami parametru a , co odpowiada malejącej dyspersji rozkładu (por. rys. 1.9, 1.10). Wyniki takie są zgodne z oczekiwaniami, gdyż parametr v rozkładu Daguma, określający liczbę skończonych momentów rozkładu, odpowiada za grubość tzw. prawego ogona funkcji gęstości, natomiast parametr a jest dodatnio skorelowany z nierównością rozkładu, więc jego wysokie wartości oznaczają rozkłady silnie rozproszone. Co interesujące, dla małych wartości a obserwujemy znaczny wzrost obciążenia parametru v , natomiast zmiany v nie mają wpływu na obciążenie estymatorów parametru a . Należy zauważyć, że zmiany wartości parametru a mają silniejszy wpływ na obciążenie i wariancję rozważanych estymatorów niż zmiany wartości parametru v .

4.6. Analiza symulacyjna własności estymatorów parametrów rozkładu Pareto

Kolejnym analizowanym rozkładem jest rozkład Pareto $Pa(\theta, a)$, który był wykorzystywany do aproksymacji rozkładów grup gospodarstw domowych o wysokich dochodach. Podobnie jak w przypadku rozkładu Daguma, metodą symulacyjną badano własności estymatorów parametrów θ i a . Dodatkowo, w badaniach pominięto metodę momentów, zastępując ją metodą momentów ważonych prawdopodobieństwami. Dla liczebności prób $n = 300$ i $n = 600$ oraz wybranych rozkładów wyznaczone relatywne błędy średniokwadratowe zawarte są w tab. 4.25, a na rysunkach 4.15–4.18 przedstawiono relatywne obciążenia i błędy średniokwadratowe.

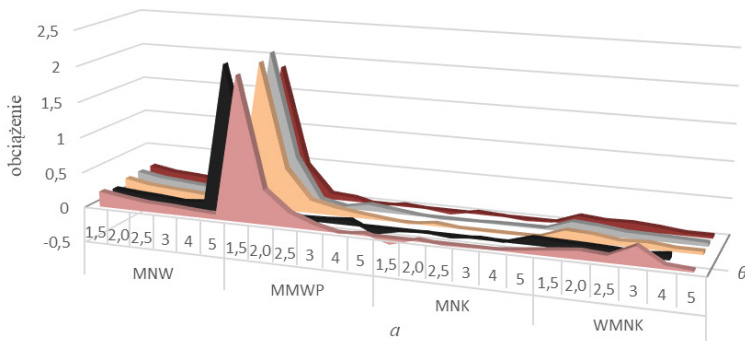
Względne błędy średniokwadratowe estymatorów są najmniejsze dla estymatorów uzyskanych metodą największej wiarygodności, natomiast największe metodą momentów ważonych prawdopodobieństwami.

Ponadto obciążenia estymatorów otrzymanych metodą momentów ważonych prawdopodobieństwami są największe.

Tabela 4.25. Relatywne błędy średniokwadratowe estymatorów parametrów rozkładu Pareto

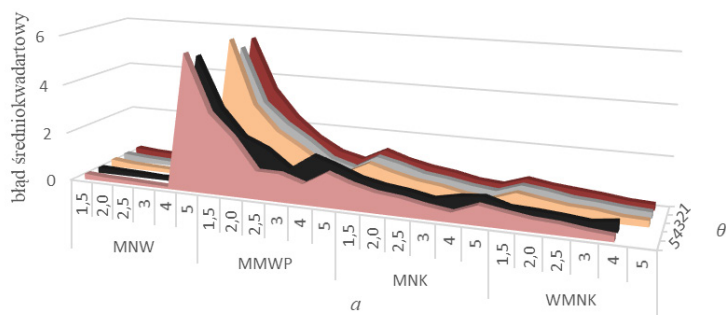
Rozkład	Metoda estymacji	$RMSE(\hat{\theta})$		$RMSE(\hat{a})$	
		$n = 300$	$n = 600$	$n = 300$	$n = 600$
$Pa(4; 1,5)$	MNW	0,299	0,156	5,787	4,211
	MMWP	5,314	4,528	11,371	9,220
	MNK	1,390	0,916	6,467	4,691
	WMNK	0,800	0,447	6,104	4,468
$Pa(3; 2)$	MNW	0,214	0,079	6,010	4,248
	MMWP	3,262	1,144	10,115	5,780
	MNK	1,036	0,476	6,801	4,860
	WMNK	0,637	0,251	6,366	4,572
$Pa(2; 3)$	MNW	0,161	0,077	5,916	4,152
	MMWP	1,739	1,275	8,595	5,999
	MNK	0,700	0,458	6,545	4,872
	WMNK	0,410	0,240	6,183	4,532
$Pa(1; 4)$	MNW	0,115	0,057	5,683	4,217
	MMWP	1,093	0,791	7,445	5,514
	MNK	0,490	0,345	6,567	4,788
	WMNK	0,2805	0,178	6,122	4,496

Źródło: opracowanie własne



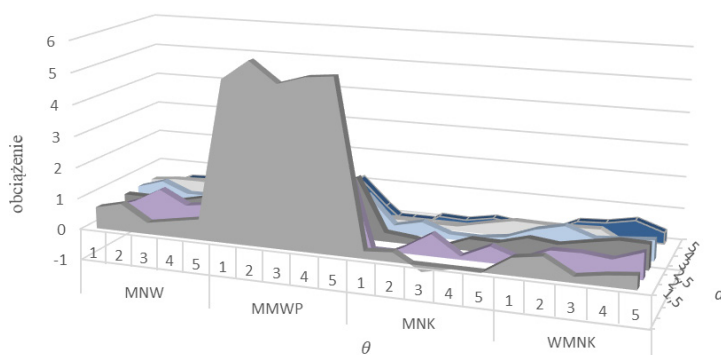
Rysunek 4.15. Oszacowane względne obciążenia różnych estymatorów parametru α

Źródło: opracowanie własne



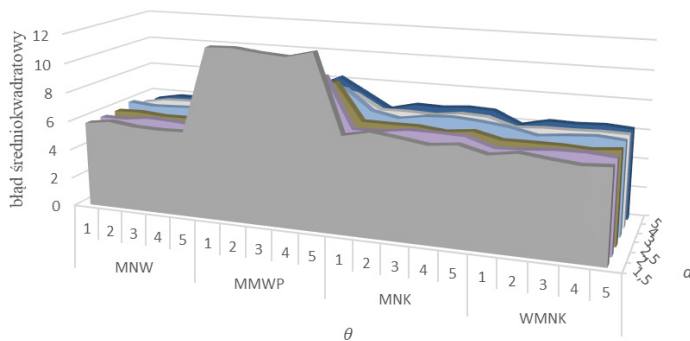
Rysunek 4.16. Oszacowane względne błędy średniokwadratowe różnych estymatorów parametru a

Źródło: opracowanie własne



Rysunek 4.17. Oszacowane względne obciążenia różnych estymatorów parametru θ

Źródło: opracowanie własne



Rysunek 4.18. Oszacowane relatywne błędy średniokwadratowe różnych estymatorów parametru θ

Źródło: opracowanie własne

Zmiany wartości względnego obciążenia i względnego błędu średniokwadratowego dla poszczególnych wartości estymatorów parametrów θ i a można przeanalizować na podstawie rys. 4.15–4.18. Można zauważyć, że poza metodą MMWP wartości tych charakterystyk są zauważalnie mniejsze niż w przypadku rozkładu Daguma, mimo znacznie mniejszych liczebności prób. Już dla $n = 600$ RMSE dla MNW, MNK i WMNK nie przekraczają 5% szacowanych parametrów. Wraz ze wzrostem wartości parametru a maleje dyspersja rozkładu, czego skutkiem jest wzrost precyzji szacunku parametru rozkładu Pareto (spadek RMSE) dla wszystkich rozważanych metod estymacji.

Zakończenie

W publikacji przedstawiono analizę własności rozkładów prawdopodobieństwa, które mogą być stosowane do analizy rozkładów płac i dochodów ludności. Biorąc pod uwagę szereg własności, które powinien posiadać idealny rozkład teoretyczny opisujący tego rodzaju zjawiska, wybrano kilka rozkładów, które poddane zostały analizie. Wśród nich znalazły się trójparametrowe rozkłady Daguma i Singha-Maddali, które są szczególnymi przypadkami czteroparametrowej rodziny GB2. Oba te rozkłady, ze względu na swoje własności statystyczne, w tym znaczną elastyczność, można traktować jako modele teoretyczne dobrze opisujące rozkłady dochodów ogółem i w różnych przekrojach. Za szczególnie interesujący został także uznany dość skomplikowany trójparametrowy rozkład Zengi, którego własności pozwalają na wykorzystanie go do bardziej zaawansowanych analiz. Obok tych dwóch rozkładów, które mają charakter uniwersalny, wyróżniono także dwa dwuparametrowe rozkłady – Pareto i logarytmiczno-normalny, które z racji swoich własności mogą być wykorzystywane w szczególnych przypadkach. Rozkład Pareto, odpowiadający prawemu ogonowi pełnego rozkładu dochodów, doskonale opisuje rozkłady osób najbogatszych, natomiast rozkład logarytmiczno-normalny, ze względu na prostotę, łatwą interpretację parametrów i cienki ogon, może być stosowany do aproksymacji rozkładów o niewielkiej koncentracji, np. rozkładów niektórych grup społeczno-ekonomicznych.

Podejście systemowe pozwoliło na pogłębioną analizę własności rozkładów teoretycznych, gdyż umożliwiło zauważenie podobieństw dla rozkładów wchodzących w skład tego samego systemu, a także wyjaśnienie różnic między rozkładami z różnych systemów. W przypadku systemów generujących rozkłady analiza ich szczególnych przypadków, związanych z różnymi funkcjami generującymi i transformującymi, pozwoliła lepiej zrozumieć proces kształtowania się rozkładu dochodów.

Wybrane rozkłady teoretyczne: logarytmiczno-normalny, Daguma i Singha-Maddali, zostały następnie poddane pogłębionej analizie, która dotyczyła przede wszystkim oceny ich zgodności z aktualnymi

rozkładami dochodów w Polsce ogółem i według województw oraz oceny własności metod estymacji ich parametrów. Zgodność empirycznych rozkładów dochodów zamożnych gospodarstw domowych w Polsce aproksymowano za pomocą rozkładu Pareto. Do oceny dobroci dopasowania rozkładów empirycznych i teoretycznych zastosowano szereg metod, w tym dwa formalne testy statystyczne Kołmogorowa-Smirnowa i Andersona-Darlinga, obliczono także współczynnik podobieństwa struktur, współczynnik Mortara, współczynnik Pearsona oraz współczynnik oparty na porównaniu średnich empirycznych i teoretycznych. Zauważono, że zgodność rozkładów empirycznych z teoretycznymi była uzależniona od przyjętej metody estymacji parametrów rozkładu. Na podstawie wyników różnych metod estymacji można jednoznacznie stwierdzić, że rozkłady Daguma i Singha-Maddali charakteryzują się bardzo wysoką zgodnością we wszystkich przypadkach. Najlepsze wyniki uzyskano, szacując parametry tych rozkładów metodą największej wiarygodności oraz uogólnioną metodą najmniejszych kwadratów. MNW okazała się też najlepsza do szacowania parametrów rozkładu Pareto, który wykazał się wysoką zgodnością z rozkładem najwyższych grup dochodowych w Polsce. Można więc potwierdzić hipotezę, że w przypadku rozkładów dochodów w Polsce wystarczy posługiwać się rozkładami co najwyżej trójparametrowymi. Z dwóch rozkładów, które są bardzo dobrze dopasowane – Daguma i Singha-Maddali – ten pierwszy wydaje się mieć lepsze własności, takie jak podwaliny empiryczne i stochastyczne, interpretacje parametrów i elastyczność, co wiąże się z dodatkowym parametrem charakteryzującym lewy ogon, który nie występuje w drugim z rozkładów.

Dla rozkładów Daguma i Pareto przeprowadzone zostały analizy symulacyjne mające na celu wskazanie najlepszych metod szacowania parametrów tych rozkładów. Z otrzymanych analiz własności estymatorów parametrów rozkładu Daguma wynika, że estymatory parametru a uzyskane metodą WMNK lub MNW charakteryzują się najmniejszym relatywnym błędem średniokwadratowym. Wśród rozważanych estymatorów parametru v najmniejsze obciążenia są obserwowane dla estymatorów otrzymanych metodą WMNK, a najmniejsze błędy średniokwadratowe – dla estymatorów uzyskanych metodą MNW lub WMNK. Można również zauważyć, że precyzja estymatorów parametrów v i a rozkładu

Daguma wzrasta wraz ze wzrastającymi wartościami parametru ν , co odpowiada coraz cieńszemu prawemu ogonowi i mniejszej asymetrii rozkładu, wraz z malejącymi wartościami parametru a , co odpowiada malejącej dyspersji rozkładu. Wyniki takie są zgodne z oczekiwaniami, gdyż parametr ν rozkładu Daguma, określający liczbę skończonych momentów rozkładu, odpowiada za grubość tzw. prawego ogona funkcji gęstości, natomiast parametr a jest dodatnio skorelowany z nierównością rozkładu, więc jego wysokie wartości oznaczają rozkłady silnie rozproszone. Co interesujące, dla małych wartości a zaobserwowano znaczny wzrost obciążenia parametru ν , natomiast zmiany ν nie miały wpływu na obciążenie estymatorów parametru a . Należy zauważyć, że zmiany wartości parametru a mają silniejszy wpływ na obciążenie i wariancję rozważanych estymatorów niż zmiany wartości parametru ν .

Wnioski płynące z przeprowadzonych analiz mogą być wykorzystane do pogłębionej analizy rozkładu płac, dochodów i ich nierównomierności w różnych przekrojach oraz do oceny zmian, jakie zachodzą w tych rozkładach w czasie. Zaproponowane narzędzia pozwalają m.in. na precyzyjną ocenę skutków prowadzonej polityki gospodarczej i społecznej.

Literatura

- Aitchison J., Brown J. A. C. (1957), *The Lognormal Distribution*, Cambridge University Press, Cambridge.
- Amoroso L. (1925), Ricerche Intorno alla Curva dei Redditi, *Annali di Matematica Pura ed Applicata*, seria 4–21, II, s. 123–157.
- Bartoszewicz J. (1996), *Wykłady ze statystyki matematycznej*, PWN, Warszawa.
- Benini R. (1906), *Principi di Statistica Metodologica*, UTET, Torino.
- Bonferroni C. E. (1930), *Elementi di Statistica Generale*, Liberia Seber, Firenze.
- Champernowne D. G. (1937), Theory of Income Distribution, *Econometrica*, 5, s. 379–381.
- Champernowne D. G. (1952), The Graduation of Income Distributions, *Econometrica*, 20, s. 591–615.
- Champernowne D. G. (1953), A Model of Income Distribution, *Economic Journal*, 53, s. 318–351.
- Chaudhry M. A., Ahmad M. (1993), On a Probability Function Useful in Size Modeling, *Canadian Journal of Forest Research*, 23 (8), s. 1679–1683.
- Cobb L., Koppstein P., Chen N. H. (1983), Estimation and Moment Recursion Relations for Multimodal Distributions of the Exponential Family, *Journal of the American Statistical Association*, 78 (381), s. 124–130.
- Chotikapanich D. (ed.) (2008), *Modeling Income Distributions and Lorenz Curves*, Springer, New York.
- Cramer H. (1958), *Metody matematyczne w statystyce*, PWN, Warszawa.
- 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'Universita di Bari*, XVI, s. 3–56.
- D'Addario R. (1949), Ricerche sulla Curva dei Redditi, *Giornale Degli Economisti ed Annali de Economia*, 8, s. 91–114.
- Dagum C. (1977), A New Model of Personal Income Distribution: Specification and Estimation, *Economie Appliquee*, 30, s. 413–437.
- Dagum C. (1980), Generating Systems and Properties of Income Distribution Models, *Metron*, XXVIII (4), s. 3–26.
- Dagum C. (1985), Analyses of Income Distribution and Inequality by Education and Sex in Canada, [w:] R. L. Basmann, G. F. Rhodes, Jr. (eds.), *Advances in Econometrics*, vol. IV, JAI Press, Greenwich, s. 167–227.

- Dagum C. (1990), Generation and Properties of Income Distribution Functions, [w:] C. Dagum, M. Zenga (eds.), *Income and Wealth Distribution, Inequality and Poverty*, Springer-Verlag, Berlin, s. 1–18.
- Dagum C., Lemmi A. (1989), A Contribution to the Analysis of Income Distribution and Income Inequality and a Case Study: Italy, [w:] D. I. Slottjee, *Advances in Econometrics*, Jai Press, Greenwich, s. 1–17.
- Dalton H. (1920), The Measurement of Inequality of Income, *Economic Journal*, 30, s. 348–361.
- Davis H. T. (1941), *The Analysis of Economic Time Series*, The Principia Press, Bloomington.
- Domma F., Giordano S., Zenga. M. M. (2011), Maximum Likelihood Estimation in Dagum Distribution with Censored Sample, *Journal of Applied Statistics*, 38, s. 2971–2985.
- Domański Cz. (1973), Struktura płac pracowników w województwach grodzkich według działów gospodarki narodowej, *Studia Prawno-Ekonomiczne*, 11, s. 1–20.
- Domański Cz., Jędrzejczak A. (1998), Maximum Likelihood Estimation of the Dagum Model Parameters, *International Advances in Economic Research*, 4, s. 243–252.
- Domański Cz., Michałkiewicz Z. (1974), *Rozkład dochodów gospodarstw domowych emerytów*, Prace Instytutu Ekonometrii i Statystyki UŁ, nr 10.
- Edgeworth F. Y. (1898), On the Representation of Statistics by Mathematical Formulae, *Journal of the Royal Statistical Society*, 1, s. 670–700.
- Fattorini L., Lemmi A. (1979), Least Squares of the Lorenz Curve and the Gini Index from Grouped Data, [w:] *Research on Economic Inequality*, 6, Jai Press, Greenwich, s. 202–223.
- Fei J., Ranis G., Kuo S. (1979), Growth and the Family Distribution of Income by Factor Components, *Quarterly Journal of Economics*, 92, s. 17–53.
- Fisk P. R. (1961), The Graduation of Income Distributions, *Econometrica*, 29, s. 171–185.
- Fisz M. (1967), *Rachunek prawdopodobieństwa i statystyka matematyczna*, PWN, Warszawa.
- Friedman M. (1953), Choice, Chance and the Personal Distribution of Income, *Journal of Public Economy*, 61, s. 277–290.
- Furrer R., Naveau P. (2007), Probability Weighted Moments Properties for Small Samples, *Statistics & Probability Letters*, 77 (2), s. 190–195.
- Garvey G. (1952), *Inequality of Income, Causes and Measurement*, *Studies in Income and Wealth*, vol. 15, National Bureau of Economic Research, New York.
- George F., Ramachandran K. M. (2011), Estimation of Parameters of Johnson's System of Distributions, *Journal of Modern Applied Statistical Methods*, 10 (2), s. 494–504.

- Gibrat R. (1931), *Les inegalities economiques*, Sirely, Paris.
- Gini C. (1909), II Diverso Accrescimento delle Classi Sociali e Concentrazione della Ricchezza, *Giornale degli Economisti*, seria II, 37.
- Gini C. (1911), Considerazioni sulle probabilità a posteriori e applicazioni al rapporto dei sessi nelle nascite umane, *Studi Economico-Giuridici della Università di Cagliari*, Anno III, s. 133–171.
- Gini C. (1912), *Variabilità e Mutabilità*, Studi Economico-Giuridici Anno III(2), Facoltà di giurisprudenza della Regia Università di Cagliari, Cuppini, Bologna.
- Gilchrist W. G. (2000), *Statistical Modelling with Quantile Functions*, Chapman & Hall/CRT, Boca Raton.
- Hart P. E. (1975), Moment Distributions in Economic: an Exposition, *Journal of The Royal Statistical Society, Series A*, 138 (3), s. 423–434.
- Hayek F. (1981), *Recht Gesetzgebung und Freiheit*, Verlag Moderne Industrie, Monachium.
- Jędrzejczak A. (1990), Uwagi o zastosowaniu rozkładu Daguma do badania rozkładów płac, *Wiadomości Statystyczne*, 7, s. 23–25.
- Jędrzejczak A. (1993), Application of the Dagum Distribution in the Analysis of Income Distribution in Poland, *Acta Universitatis Lodzensis. Folia Oeconomica*, 131, s. 103–111.
- Jędrzejczak A. (2006). The Characteristic of Theoretical Income Distributions and their Application to the Analysis of Wage Distributions in Poland by Regions, *Acta Universitatis Lodzensis. Folia Oeconomica*, 196, s. 183–198.
- Jędrzejczak A. (2011), *Metody analizy rozkładów dochodów i ich koncentracji*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź.
- Jędrzejczak A. (2015), Regional Income Inequalities in Poland and Italy, *Comparative Economic Research*, 18 (4), s. 27–45.
- Jędrzejczak A., Pekasiewicz D. (2018), Differentiation of Income Distribution of Farmers' Households in the Polish Macro-regions, *Zagadnienia Ekonomiki Rolnej*, 3 (356), s. 150–167.
- Jędrzejczak A., Pekasiewicz D., Zieliński W. (2018), Comparison of Parameter Estimators for the Dagum of Distribution, *The 12th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena. Conference Proceedings*, s. 180–189.
- Johnson N. L., Kotz S., Balakrishnan N. (1994), *Continuous Univariate Distributions*, Wiley, New York.
- Kalecki M. (1945), On the Gibrat distribution, *Econometrica*, 13, s. 161–170.
- Kleiber C. (1996), Dagum vs. Singh-Maddala Income Distributions, *Economics Letter*, 53, s. 265–268.
- Kleiber C., Kotz S. (2003), *Statistical Size Distributions in Economics and Actuarial Sciences*, Wiley, Hoboken.

- Kloek T., van Dijk H. K. (1977), Further Results on Efficient Estimation of Income Distribution Parameters, *Economie Appliquee*, 30 (3), s. 439–459.
- Kordos J. (1968), *Metody matematyczne badania i analizy rozkładu dochodów ludności*, GUS, Warszawa.
- Kordos J. (1973), *Metody analizy i prognozowania rozkładów płac i dochodów ludności*, PWE, Warszawa.
- Kordos J., Strońska B. (1971), *Statystyczne metody analizy rozkładu płac i dochodów ludności*, Prace Zakładu Badań Statystyczno-Ekonomicznych, 62, GUS, Warszawa.
- Kot S. M. (1995), A Probabilistic Model of Size Distribution of Income with Covariates, *Przegląd Statystyczny*, 42 (2), s. 155–180.
- Kumor P., Sztudynier J. J. (2007), Optymalne zróżnicowanie płac w Polsce – analiza ekonometryczna, *Ekonomista*, 1, s. 45–58.
- Lahcene B. (2013), On Pearson Families of Distributions and its Applications, *African Journal of Mathematics and Computer Science Research*, 6 (5), s. 108–117.
- Lange O. (1967), *Wstęp do ekonometrii*, PWN, Warszawa.
- Lorenz M. O. (1905), Methods of Measuring the Concentration of Wealth, *Publications of the ASA*, 70 (9), s. 209–219.
- Lydall H. F. (1968), *The Structure of Earnings*, Oxford University Press, Oxford.
- Macaulay F. R. (1922), The Personal Distribution of Income in the United States, [w:] W. C. Mitchell, *Income in the United States: Its Amount and Distribution 1909–1919*, Harcourt, Brace and Co., New York.
- Mandelbrot B. (1963), New Methods in Statistical Economics, *Journal of Political Economy*, 5, s. 79–106.
- March L. (1898), Quelques Exemples de Distribution des Salaires, *Journal de la Societe Statistique de Paris*, s. 193–206.
- Mc Donald J. B. (1984), Some Generalized Functions for the Distributions of Income, *Econometrica*, 52, s. 647–663.
- Metcalf E. (1969), The Size Distribution of Personal Income during the Business Cycle, *American Economic Review*, 57, s. 657–669.
- Ostasiewicz K. (2013), Adekwatność wybranych rozkładów teoretycznych dochodów w zależności od metody estymacji, *Przegląd Statystyczny*, 60 (4), s. 499–521.
- Pareto V. (1895), La legge della domanda, *Giornale Degli Economisti*, 10, s. 59–68.
- Pareto V. (1897), *Cours d'économie politique*, Rouge, Lausanne.
- Pawłowski Z. (1960), Badania elastyczności popytu a rozkład dochodów, *Przegląd Statystyczny*, 7 (4), s. 445–465.
- Pearson K. (1895), Contributions to the Mathematical Theory of Evolutions II: Skew-variation in Homogeneous Material, *Philosophical Transactions of the Royal Society of London*, 186, s. 343–414.

- Pearson K. (1948), *Karl Pearson's Early Statistical Papers*, Cambridge University Press, Cambridge.
- Pekasiewicz D. (2012), Zastosowanie metod symulacyjnych do badania własności estymatorów otrzymanych metodą kwantyli, [w:] Z. E. Zieliński (red.), *Rola informatyki w naukach ekonomicznych i społecznych. Innowacje i implikacje interdyscyplinarne*, Wydawnictwo Wyższej Szkoły Handlowej, Kielce, s. 236–244.
- Pekasiewicz D. (2015), *Statystyki pozycyjne w procedurach estymacji i ich zastosowania w badaniach społeczno-ekonomicznych*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź.
- Pekasiewicz D., Jędrzejczak A. (2017), Application of the Measures Based on Quantiles to the Analysis of Income Inequality and Poverty in Poland by Socio-economic Group, *International Conference „Mathematical Methods in Economic”, Conference Proceeding*, s. 532–537.
- Piketty T. (2014), *Capital in the Twenty-First Century*, Harvard University Press, Cambridge.
- Piltz F. (1929), *Dochody mieszkańców Warszawy*, Instytut Gospodarstwa Społecznego, Warszawa.
- Pigou A. C. (1920), *The Economics of Welfare*, Macmillan, London.
- Polisicchio M. (2008), The Continuous Random Variable with Uniform Point Inequality Measure $I(p)$, *Statistica & Applicazioni*, VI (2), s. 137–151.
- Rasool M. T., Arshad M., Ahmad M. I. (2002), Estimation of Generalized Logistic Distribution by Probability Weight Moments, *Pakistan Journal of Applied Sciences*, 2 (4), s. 485–487.
- Roy L. K. (1971), An Extension of the Pearson System of Frequency Curves, *Trabajos de estadística y de investigación operativa*, 22 (1–2), s. 113–123.
- Rutherford R. (1955), Income Distributions: A. A. New Model, *Econometrica*, 3, s. 277–294.
- Salamaga M. (2016a), Modelowanie rozkładów dochodów kobiet i mężczyzn w województwie małopolskim, *Wiadomości Statystyczne*, 8, s. 32–44.
- Salamaga M. (2016b), Badanie wpływu metody estymacji teoretycznych modeli rozkładu dochodów na jakość aproksymacji rozkładu dochodów mieszkańców Krakowa, *Zeszyty Naukowe Uniwersytetu Ekonomicznego w Krakowie*, 3 (951), s. 63–79.
- Salem A., Mount T. D. (1974), A Convenient Descriptive Model of Income Distribution of Gamma Density, *Econometrica*, 42, s. 1115–1127.
- Sen A. (1973), *On Economic Inequality*, Oxford, Clarendon Press.
- Sen A. (1976), Poverty: on Ordinal Approach to Measurement, *Econometrica*, 44, s. 219–231.
- Shakil M., Golam Kibria B. M., Singh J. N. (2016), Review on Generalized Pearson System of Probability Distributions, *Journal of Mathematical Sciences & Mathematics Education*, 11 (2), s. 13–33.

- Singh S., Maddala G. (1976), A Function for Size Distribution of Income, *Econometrica*, 44, s. 963–970.
- Steyn H. S. (1966), On the Departure from the Logarithmic Normal Distribution of Income, *South African Journal of Economics*, 34 (3), s. 225–232.
- Stacy E. W. (1962), A Generalization of the Gamma Distribution, *The Annals of Mathematical Statistics*, 33, s. 1187–1192.
- Stoppa G. (1990), A New Model for Income Size Distribution, [w:] C. Dagum, M. Zenga (eds.), *Income and Wealth Distribution, Inequality and Poverty*, Springer-Verlag, Berlin, s. 33–41.
- Swain J., Venkatraman S., Wilson J. (1988), Least Squares Estimation of Distribution Function in Johnson's Translation System, *Journal of Statistical Computation and Simulation*, 29, s. 271–297.
- Tadikamalla P. R. (1980), A Look at the Burr and Related Distributions, *International Statistical Review*, 48, s. 337–344.
- Taussig F. W. (1915), *Principles of Economics*, The Macmillan Company, New York.
- Tinbergen J. (1956), On the Theory of Income Distribution, *Weltwirtschaftliches Archiv*, 77, s. 155–175.
- Tinbergen J. (1972), Impact of Education on Income Distribution, *Review of Income and Wealth*, 18, s. 255–265.
- Theil H. (1967), *Economics and Information Theory*, North-Holland, Publishing Company, Amsterdam.
- Thurow L. C. (1970), Analyzing the American Income Distribution, *American Economic Review*, 60, s. 261–269.
- Wąsik B. (1967), Dwu- i trójparymetrowy rozkład logarytmiczno-normalny jako aproksymacja rozkładu zarobków pracowników gospodarki społecznej 1955–1965, *Przegląd Statystyczny*, 14 (4), s. 409–424.
- Weibull W. (1951), A Statistical Distribution Function of Wide Applicability, *Journal of Applied Mechanics*, 16, s. 293–297.
- Wiśniewski J. (1934), *Rozkład dochodów według wysokości*, Instytut Badania Konjunktur Gospodarczych i Cen, Warszawa.
- Wywiał J. (2004), *Wprowadzenie do wnioskowania statystycznego*, Wydawnictwo Akademii Ekonomicznej, Katowice.
- Wyżnikiewicz B. (1987), *Zmiany strukturalne w gospodarce. Prawidłowości i ograniczenia*, PWE, Warszawa.
- Vielrose E. (1960), *Rozkład dochodów według wielkości*, Polskie Wydawnictwo Gospodarcze, Warszawa.
- Vinci F. (1921), Nuovoi contribute allo studio della distribuzione dei redditi, *Giornale degli Economisti ed Rivista di Statistica*, 61, s. 365–369.

- Zenga M. (1990), Concentration Curves and Concentration Indices Derived from Them, [w:] C. Dagum, M. Zenga (eds.), *Income and Wealth Distribution, Inequality and Poverty*, Springer-Verlag, Berlin, s. 94–110.
- Zenga M. (2007), Inequality Curve and Inequality Index Based on the Ratios between Lower and Upper Arithmetic Means, *Statistica & Applicazioni*, V (1), s. 3–27.
- Zenga M. M. (2010), Mixture of Poliscicchio's Truncated Pareto Distributions with Beta Weights, *Statistica & Applicazioni*, VIII (1), s. 3–25.
- Zenga M. M., Pasquazzi L., Zenga Ma. (2010a), First Applications of a New Three Parameter Distribution for Non-Negative Variables, Dip. to Metodi Quantitativi per le Scienze Economiche ed Aziendali – Università degli Studi di Milano Bicocca, *Rapporto di Ricerca*, N. 188.
- Zenga M. M., Pasquazzi L., Poliscicchio M., Zenga Ma. (2010b), More on M. M. Zenga's New Three-Parameter Distribution for Non-Negative Variables, *Statistica & Applicazioni*, IX (1), s. 5–33.
- Zenga M. M., Porro F., Arcagni A. (2010c), Method of moments for Zenga's distribution, Dip. to Metodi Quantitativi per le Scienze Economiche ed Aziendali – Università degli Studi di Milano Bicocca, *Rapporto di Ricerca*, N. 193.

Spis rysunków

Rysunek 1.1. Gęstości rozkładów Pareto dla $\theta = 3$ i wybranych wartości a ...	24
Rysunek 1.2. Gęstości rozkładów Pareto dla $a = 2$ i wybranych wartości θ ...	24
Rysunek 1.3. Funkcje Lorenza dla rozkładów Pareto o wybranych wartościach parametru a	28
Rysunek 1.4. Gęstości rozkładów logarytmiczno-normalnych dla $\mu = 0$ i wybranych wartości σ	38
Rysunek 1.5. Gęstości rozkładów logarytmiczno-normalnych dla $\sigma = 0,5$ i wybranych wartości μ	39
Rysunek 1.6. Funkcje Lorenza dla rozkładów logarytmiczno-normalnych o wybranych wartościach parametru σ	41
Rysunek 1.7. Gęstości rozkładów gamma dla $\lambda = 1$ i wybranych wartości α	44
Rysunek 1.8. Gęstości rozkładów gamma dla $\alpha = 2$ i wybranych wartości λ	45
Rysunek 1.9. Gęstości rozkładów Daguma dla wybranych wartości a i ustalonych pozostałych parametrów	49
Rysunek 1.10. Gęstości rozkładów Daguma dla wybranych wartości v i ustalonych pozostałych parametrów	49
Rysunek 1.11. Gęstości rozkładów Daguma dla wybranych wartości b i ustalonych pozostałych parametrów	50
Rysunek 1.12. Funkcje Lorenza dla wybranych rozkładów Daguma ...	51
Rysunek 1.13. Gęstości rozkładów Singha-Maddali dla wybranych wartości q i ustalonych pozostałych parametrów	57
Rysunek 1.14. Gęstości rozkładów Singha-Maddali dla wybranych wartości a i ustalonych pozostałych parametrów	57
Rysunek 1.15. Gęstości rozkładów Singha-Maddali dla wybranych wartości b i ustalonych pozostałych parametrów	58
Rysunek 1.16. Gęstości rozkładów Fiska dla wartości $b = 1$ i wybranych wartości a	61
Rysunek 1.17. Gęstości rozkładów Fiska dla wartości $a = 3$ i wybranych wartości b	61
Rysunek 1.18. Gęstości rozkładów Weibulla dla wartości $\alpha = 2$ i wybranych wartości k	63
Rysunek 1.19. Gęstości rozkładów Weibulla dla wartości $k = 2$ i wybranych wartości α	63
Rysunek 1.20. Funkcje Lorenza dla wybranych rozkładów Weibulla ...	64
Rysunek 1.21. Gęstości rozkładu Zengi dla $\mu = 2$ oraz $\alpha = 2$ i wybranych wartości parametru θ	67

Rysunek 1.22. Gęstości rozkładu Zengi dla $\mu = 2$ oraz $\theta = 0,25$ i wybranych wartości parametru α	68
Rysunek 2.1. Gęstości rozkładów GB2 dla $b = 3, p = 2, q = 4$ i wybranych wartości a	87
Rysunek 2.2. Gęstości rozkładów GB2 dla $a = 2, b = 3, q = 4$ i wybranych wartości p	87
Rysunek 2.3. Gęstości rozkładów GB2 dla $a = 2, b = 3, p = 2$ i wybranych wartości q	88
Rysunek 2.4. Gęstości rozkładów GB2 dla $a = 2, b = 3$ i wybranych wartości p, q	88
Rysunek 4.1. Podział Polski na jednostki NUTS 1 w latach 2014–2016 ..	113
Rysunek 4.2. Aproksymacja rozkładu dochodów w Polsce rozkładem Daguma z wybranymi metodami estymacji parametrów	120
Rysunek 4.3. Aproksymacja rozkładu dochodów w regionach rozkładem Daguma z wybranymi metodami estymacji parametrów	121
Rysunek 4.4. Aproksymacja rozkładu dochodów w Polsce przy zastosowaniu rozkładu logarytmiczno-normalnego i Daguma	126
Rysunek 4.5. Aproksymacja rozkładu dochodów w regionach przy zastosowaniu rozkładu logarytmiczno-normalnego i Daguma	127
Rysunek 4.6. Aproksymacja rozkładu dochodów w Polsce przy zastosowaniu rozkładu Daguma i Singha-Maddali	131
Rysunek 4.7. Aproksymacja rozkładu dochodów w regionach przy zastosowaniu rozkładu Daguma i Singha-Maddali	132
Rysunek 4.8. Odsetek zamożnych w poszczególnych regionach	139
Rysunek 4.9. Aproksymacja rozkładów dochodów zamożnych gospodarstw domowych dla $y_{r1}^* = 2Me$	140
Rysunek 4.10. Aproksymacja rozkładów dochodów zamożnych gospodarstw domowych dla $y_{r2}^* = 3Me$	140
Rysunek 4.11. Oszacowane względne obciążenia różnych estymatorów parametru a	149
Rysunek 4.12. Oszacowane relatywne błędy średniokwadratowe różnych estymatorów parametru a	149
Rysunek 4.13. Oszacowane względne obciążenia różnych estymatorów parametru v	150
Rysunek 4.14. Oszacowane relatywne błędy średniokwadratowe różnych estymatorów parametru v	150
Rysunek 4.15. Oszacowane względne obciążenia różnych estymatorów parametru α	153
Rysunek 4.16. Oszacowane względne błędy średniokwadratowe różnych estymatorów parametru a	154
Rysunek 4.17. Oszacowane względne obciążenia różnych estymatorów parametru θ	154
Rysunek 4.18. Oszacowane relatywne błędy średniokwadratowe różnych estymatorów parametru θ	154

Spis tabel

Tabela 1.1. Wartości parametru a rozkładu Pareto.....	27
Tabela 1.2. Wartości współczynników Giniego i Pietry dla wybranych rozkładów Pareto	29
Tabela 1.3. Funkcje gęstości rozkładów Pareto I–IV	30
Tabela 1.4. Podstawowe charakterystyki rozkładów Pareto I–IV	31
Tabela 1.5. Wartości współczynników Giniego i Pietry dla wybranych rozkładów logarytmiczno-normalnych	40
Tabela 1.6. Wartości współczynnika Giniego dla wybranych rozkładów gamma	46
Tabela 1.7. Wartości współczynnika Giniego dla wybranych rozkładów Daguma	52
Tabela 1.8. Wartości współczynnika Giniego dla wybranych rozkładów Singha-Maddali	59
Tabela 1.9. Wartości współczynnika Giniego dla wybranych rozkładów Weibulla	64
Tabela 1.10. Wartości współczynnika Giniego dla wybranych rozkładów Zengi	70
Tabela 2.1. Rozkłady dochodów należące do systemu Daguma	73
Tabela 2.2. Funkcje gęstości krzywych Pearsona	76
Tabela 2.3. Wybrane rozkłady dochodów należące do systemu D’Addario ..	80
Tabela 2.4. Dystrybuanty i funkcje gęstości rozkładów Burra	81
Tabela 2.5. Funkcje transformujące i ich funkcje odwrotne dla rozkładów Johnsona	83
Tabela 2.6. Rozkłady należące do rodziny GB2	89
Tabela 4.1. Podstawowe charakterystyki dochodów gospodarstw domowych w ujęciu regionalnym	114
Tabela 4.2. Oszacowania współczynnika Giniego	115
Tabela 4.3. Oszacowania parametrów rozkładu Daguma aproksymującego rozkład dochodów gospodarstw domowych w regionach w Polsce ...	117
Tabela 4.4. Wartości p -value dla testów Kołmogorowa-Smirnowa i Andersona-Darlinga weryfikujących hipotezy o zgodności rozkładów dochodów z rozkładem Daguma	118
Tabela 4.5. Miary dopasowania empirycznych rozkładów dochodów do rozkładu Daguma	119
Tabela 4.6. Estymatory parametrów rozkładu logarytmiczno-normalnego dla rozkładów dochodów w Polsce według regionów	123
Tabela 4.7. Wartości p -value dla testów Kołmogorowa-Smirnowa i Andersona-Darlinga dla zgodności rozkładu dochodów z rozkładem logarytmiczno-normalnym	124

Tabela 4.8. Miary dopasowania empirycznych rozkładów dochodów do rozkładu logarytmiczno-normalnego	125
Tabela 4.9. Oszacowania parametrów rozkładu Singha-Maddali aproksymujące rozkład dochodów gospodarstw domowych w regionach w Polsce	128
Tabela 4.10. Wartości p -value dla testów Kołmogorowa-Smirnowa i Andersona-Darlinga weryfikujących hipotezy o zgodności rozkładów dochodów z rozkładem Singha-Maddali	129
Tabela 4.11. Miary dopasowania empirycznych rozkładów dochodów do rozkładu Singha-Maddali	130
Tabela 4.12. Ranking dopasowania rozkładów dochodów w regionach Polski według współczynnika podobieństwa struktur i innych miar	133
Tabela 4.13. Teoretyczne rozkłady aproksymujące rozkłady dochodów w regionach i ich podstawowe parametry	135
Tabela 4.14. Charakterystyki rozkładów dochodów gospodarstw domowych przekraczających granicę y_{r1}^* w ujęciu regionalnym	137
Tabela 4.15. Charakterystyki rozkładów dochodów gospodarstw domowych przekraczających granicę y_{r2}^* w ujęciu regionalnym	138
Tabela 4.16. Oszacowania parametrów rozkładu Pareto aproksymującego rozkład dochodów zamożnych gospodarstw domowych w Polsce ..	139
Tabela 4.17. Wartości p -value dla testów Kołmogorowa-Smirnowa i Andersona-Darlinga	141
Tabela 4.18. Miary dopasowania empirycznych rozkładów dochodów zamożnych gospodarstw domowych w Polsce do rozkładu Pareto ..	141
Tabela 4.19. Oszacowania parametrów rozkładu Pareto aproksymującego rozkłady dochodów zamożnych gospodarstw domowych w regionach	142
Tabela 4.20. Wartości p -value dla testów weryfikujących hipotezy o zgodności rozkładów dochodów grup zamożnych z rozkładem Pareto	143
Tabela 4.21. Miary dopasowania empirycznych rozkładów dochodów zamożnych gospodarstw domowych do rozkładu Pareto	144
Tabela 4.22. Ranking dopasowania rozkładów dochodów gospodarstw domowych o wysokich dochodach w regionach Polski	145
Tabela 4.23. Parametry rozkładów aproksymujących rozkłady dochodów zamożnych gospodarstw domowych w regionach	147
Tabela 4.24. Relatywne błędy średniokwadratowe estymatorów parametrów rozkładu Daguma	151
Tabela 4.25. Relatywne błędy średniokwadratowe estymatorów parametrów rozkładu Pareto	153

ANEKS

Wybrane charakterystyki
funkcyjne i liczbowe
podstawowych rozkładów
dochodów

Rozkład	Charakterystyki funkcyjne	Charakterystyki liczbowe
Daguma $D(a, v, b)$ $v, a, b > 0$	<p>Funkcja gęstości:</p> $f(y) = \begin{cases} \frac{av}{b} \left(\frac{y}{b}\right)^{av-1} \left[1 + \left(\frac{y}{b}\right)^v\right]^{-a-1} & \text{dla } y > 0 \\ 0 & \text{dla } y \leq 0 \end{cases}$ <p>Dystrybuanta:</p> $F(y) = \begin{cases} \left[1 + \left(\frac{y}{b}\right)^v\right]^{-a} & \text{dla } y > 0 \\ 0 & \text{dla } y \leq 0 \end{cases}$	<p>Wartość oczekiwana:</p> $EY = b \frac{\Gamma\left(1 - \frac{1}{v}\right) \Gamma\left(a + \frac{1}{v}\right)}{\Gamma(a)} \quad \text{dla } v > 1$ <p>Odchylenie standardowe:</p> $DY = b \sqrt{\frac{\Gamma\left(1 - \frac{2}{v}\right) \Gamma(a) \Gamma\left(a + \frac{2}{v}\right) - \Gamma^2\left(1 - \frac{1}{v}\right) \Gamma^2\left(a + \frac{1}{v}\right)}{\Gamma^2(a)}}$ <p>dla $v > 2$</p> <p>Funkcja Lorentza:</p> $L(p) = B \left(\frac{1}{p^a}; a + \frac{1}{v}, 1 - \frac{1}{v} \right) \quad \text{dla } v > 1$ <p>Współczynnik Giniego:</p> $G = -1 + \frac{B(a, a)}{B\left(a, a + \frac{1}{v}\right)} \quad \text{dla } v > 1$
Fiska $F(a, b)$ $a, b > 0$	<p>Funkcja gęstości:</p> $f(y) = \begin{cases} \frac{a \left(\frac{y}{b}\right)^{a-1}}{\left[b \left(1 + \left(\frac{y}{b}\right)^a\right)\right]^2} & \text{dla } y \geq 0 \\ 0 & \text{dla } y < 0 \end{cases}$	<p>Wartość oczekiwana:</p> $EY = \frac{b\pi}{a \sin\left(\frac{\pi}{a}\right)} \quad \text{dla } a > 1$ <p>Odchylenie standardowe:</p> $DY = b \sqrt{\frac{2\pi}{a \sin\left(\frac{2\pi}{a}\right)} \frac{\pi^2}{a^2 \sin^2\left(\frac{\pi}{a}\right)}} \quad \text{dla } a > 2$

	<p>Dystrybucja:</p> $F(y) = \begin{cases} 1 - \frac{1}{1 + \left(\frac{y}{b}\right)^a} & \text{dla } y \geq 0 \\ 0 & \text{dla } y < 0 \end{cases}$	<p>Funkcja Lorentza:</p> $L(p) = B \left(1; 1 + \frac{1}{a}, 1 - \frac{1}{a} \right) \quad \text{dla } a > 1$ <p>Współczynnik Giniego:</p> $G = \frac{1}{a}$
<p>gamma</p> $G(\alpha, \lambda)$ $\alpha, \lambda > 0$	<p>Funkcja gęstości:</p> $f(y) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\lambda y} & \text{dla } y > 0 \\ 0 & \text{dla } y \leq 0 \end{cases}$ <p>Dystrybucja:</p> $F(y) = \begin{cases} \frac{\Gamma\left(\frac{y}{\lambda}, \alpha\right)}{\Gamma(\alpha)} & \text{dla } y > 0 \\ 0 & \text{dla } y \leq 0 \end{cases}$	<p>Wartość oczekiwana:</p> $E(Y) = \frac{\alpha}{\lambda} \quad \text{dla } \alpha > 1$ <p>Odchylenie standardowe:</p> $D(Y) = \frac{\sqrt{\alpha}}{\lambda}$ <p>Współczynnik Giniego:</p> $G = \frac{\Gamma(\alpha+0,5)}{\sqrt{\pi}\Gamma(\alpha+1)}$

Rozkład	Charakterystyki funkcyjne	Charakterystyki liczbowe
logarytmiczno-normalny $LN(\mu, \sigma)$ $\mu \in R, \sigma > 0$	<p>Funkcja gęstości:</p> $f(y) = \frac{1}{y\sigma\sqrt{2\pi}} \exp\left(-\frac{\ln y - \mu}{2\sigma^2}\right) \text{ dla } y > 0$ <p>Dystrybuanta:</p> $F(y) = \frac{1}{2} + \frac{1}{2} \operatorname{erf}\left(\frac{\ln y - \mu}{\sigma\sqrt{2}}\right) \text{ dla } y > 0$	<p>Wartość oczekiwana:</p> $E(Y) = e^{\mu + 0,5\sigma^2}$ <p>Odchylenie standardowe:</p> $D(Y) = e^{\mu + 0,5\sigma^2} \sqrt{e^{\sigma^2} - 1}$ <p>Funkcja Lorentza:</p> $L(p) = \Phi(\Phi^{-1}(p) - \sigma)$ <p>Współczynnik Giniego:</p> $G = 2\Phi\left(\frac{\sigma}{\sqrt{2}}\right) - 1, \Phi(\cdot) \text{ - dystrybuant rozkładu normalnego standaryzowanego}$
Pareto $Pa(\theta, a)$ $\theta, a > 0$	<p>Funkcja gęstości:</p> $f(y) = \begin{cases} \frac{a\theta^a}{y^{a+1}} & \text{dla } y \geq \theta \\ 0 & \text{dla } y < \theta \end{cases}$ <p>Dystrybuanta:</p> $F(y) = \begin{cases} 1 - \left(\frac{y}{\theta}\right)^{-a} & \text{dla } y \geq \theta \\ 0 & \text{dla } y < \theta \end{cases}$	<p>Wartość oczekiwana:</p> $E(Y) = \frac{a\theta}{a-1} \text{ dla } a > 1$ <p>Odchylenie standardowe:</p> $D(Y) = \frac{\theta}{a-1} \sqrt{\frac{a}{a-2}} \text{ dla } a > 2$ <p>Funkcja Lorentza:</p> $L(p) = 1 - (1-p)^{\frac{1}{a}} \text{ dla } a > 1$ <p>Współczynnik Giniego:</p> $G = \frac{1}{2a-1} \text{ dla } a > 1$

Singha-Maddali $S - M(q, a, b)$ $q, a, b > 0$	<p>Funkcja gęstości:</p> $f(y) = \begin{cases} ab^{-a} q y^{a-1} \left[1 + \left(\frac{y}{b} \right)^a \right]^{-q-1} & \text{dla } y \geq 0 \\ 0 & \text{dla } y < 0 \end{cases}$ <p>Dystrybuanta:</p> $F(y) = \begin{cases} 1 - \frac{1}{\left[1 + \left(\frac{y}{b} \right)^a \right]^q} & \text{dla } y \geq 0 \\ 0 & \text{dla } y < 0 \end{cases}$	<p>Wartość oczekiwana:</p> $E(Y) = \frac{b \Gamma\left(1 + \frac{1}{a}\right) \Gamma\left(q - \frac{1}{a}\right)}{\Gamma(q)} \quad \text{dla } aq > 1$ <p>Odchylenie standardowe:</p> $D(Y) = \frac{b}{\Gamma(q)} \sqrt{\Gamma\left(1 + \frac{2}{a}\right) \Gamma\left(q - \frac{2}{a}\right) \Gamma(q) - \Gamma^2\left(1 + \frac{1}{a}\right) \Gamma^2\left(q - \frac{1}{a}\right)}$ <p>dla $aq > 2$</p> <p>Funkcja Lorentza:</p> $L(p) = B \left(1 - (1 - q)^{\frac{1}{q}}; 1 + \frac{1}{a}, q - \frac{1}{a} \right) \quad \text{dla } aq > 1$ <p>Współczynnik Giniego:</p> $G = 1 - \frac{\Gamma(q) \Gamma\left(2q - \frac{1}{a}\right)}{\Gamma\left(q - \frac{1}{a}\right) \Gamma(2q)} \quad \text{dla } aq > 1$
--	--	---

Rozkład	Charakterystyki funkcyjne	Charakterystyki liczbowe
Weibull $W(\alpha, k)$ $k, \alpha > 0$	<p>Funkcja gęstości:</p> $f(y) = \begin{cases} \frac{\alpha}{k} \left(\frac{y}{k}\right)^{\alpha-1} e^{-\left(\frac{y}{k}\right)^\alpha} & \text{dla } y \geq 0 \\ 0 & \text{dla } y < 0 \end{cases}$ <p>Dystrybuanta:</p> $F(y) = \begin{cases} 1 - e^{-\left(\frac{y}{k}\right)^\alpha} & \text{dla } y \geq 0 \\ 0 & \text{dla } y < 0 \end{cases}$	<p>Wartość oczekiwana:</p> $EY = k\Gamma\left(1 + \frac{1}{\alpha}\right)$ <p>Odchylenie standardowe:</p> $DY = k\sqrt{\Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma^2\left(1 + \frac{1}{\alpha}\right)}$ <p>Funkcja Lorentza:</p> $L(p) = 1 - \frac{\Gamma\left(-\ln(1-p), 1 + \frac{1}{\alpha}\right)}{\Gamma\left(1 + \frac{1}{\alpha}\right)}$ <p>Współczynnik Giniego:</p> $G = 1 - 2^{-1/\alpha}$

<p>Zengi $Z(\mu, \alpha, \theta)$ $\mu > 0, \alpha > 0,$ $\theta > 0$</p>	<p>Dystrybuanta:</p> $F(y) = \begin{cases} \frac{1}{B(\alpha, \theta)} \cdot \sum_{i=1}^{\infty} \left[B\left(\frac{y}{\mu}; \alpha + i - 1, \theta\right) - \left(\frac{\mu}{y}\right)^{0.5} B\left(\frac{y}{\mu}; \alpha + i - \frac{1}{2}, \theta\right) \right] & \text{gdy } 0 < y \leq \mu \\ 1 - \frac{1}{B(\alpha, \theta)} \cdot \sum_{i=1}^{\infty} \left[\left(\frac{\mu}{y}\right)^{0.5} B\left(\frac{\mu}{y}; \alpha + i - \frac{1}{2}, \theta\right) - B\left(\frac{\mu}{y}; \alpha + i, \theta\right) \right] & \text{gdy } y > \mu \end{cases}$	<p>Wartość oczekiwana: $EY = \mu$ Odchylenie standardowe: $DY = \mu \sqrt{\frac{1}{3B(\alpha, \theta)} \sum_{i=1}^3 B(\alpha - 2 + i, \theta) - 1}$</p> <p>Funkcja Lorenza i współczynnik Giniego nie mają jawnej postaci.</p>
---	---	---

