

*Marcin Salamaga*

Katedra Statystyki

Uniwersytet Ekonomiczny w Krakowie

# **Badanie wpływu metody estymacji teoretycznych modeli rozkładu dochodów na jakość aproksymacji rozkładu dochodów mieszkańców Krakowa\***

## **Streszczenie**

Celem artykułu jest próba wskazania modeli teoretycznych oraz metod ich estymacji, które możliwie najdokładniej odzwierciedlają empiryczny rozkład dochodów mieszkańców Krakowa. W badaniach brano pod uwagę rozkłady teoretyczne najczęściej wykorzystywane w analizie płac i dochodów, jak: rozkład Singha-Maddali, Fiska, log-normalny czy gamma. Przedmiotowe rozkłady badano pod względem precyzji oszacowania miar położenia, zmienności i nierówności dochodowych w zależności od przyjętej metody estymacji parametrów. W obliczeniach wykorzystano dane o indywidualnych dochodach krakowian w 2013 r., pochodzące z badania panelowego przeprowadzonego w ramach „Diagnozy społecznej”. Wyniki badań wskazują, że najbardziej dokładne oszacowanie rozkładu dochodów i jego charakterystyk liczbowych można osiągnąć, stosując model Singha-Maddali, a rekomendowaną metodą szacowania parametrów teoretycznych modeli jest metoda największej wiarygodności.

---

\* Artykuł powstał w ramach realizacji projektu badawczego sfinansowanego ze środków przyznanych Wydziałowi Zarządzania Uniwersytetu Ekonomicznego w Krakowie w ramach dotacji na utrzymanie potencjału badawczego.

**Słowa kluczowe:** rozkład dochodów, metoda największej wiarygodności, współczynnik Giniego, model Singha-Maddali.

## 1. Wprowadzenie

Badania nad rozkładem dochodów oraz płac w społeczeństwie od dawna są podejmowane w literaturze przedmiotu. Celem badaczy jest często próba poszukiwania rozkładów teoretycznych, które możliwie najlepiej odzwierciedlają empiryczny rozkład dochodów. Modele teoretyczne mogą stanowić skuteczne narzędzie w identyfikacji niektórych własności rozkładów empirycznych i pozwalają na wyrażenie charakterystyk liczbowych, takich jak: średnia, modalna, odchylenie standardowe, współczynnik Giniego, Atkinsona czy miary dobrobytu społecznego za pomocą parametrów tych modeli. Dotychczasowe badania dowodzą, że najczęściej empiryczne rozkłady dochodów wyróżniają się jednomodalnością, prawostronną asymetrią i dodatnią kurtozą [Kot 1999a]. Z tego powodu do ich modelowania wykorzystuje się takie rozkłady teoretyczne, które przy odpowiedniej parametryzacji prawidłowo odzwierciedlają te właściwości. Badacze najczęściej odwołują się do rozkładów Pareto, Weibulla, Singha-Maddali, Daguma, Fiska, log-normalnego, gamma i innych. Wyniki modelowania rozkładów dochodów można znaleźć zarówno w literaturze krajowej [Kot 1999a, 2000, Ulman 2011, Ostasiewicz 2013], jak i zagranicznej [McDonald i Ransom 1979, Kleiber i Kotz 2003, Clementi i in. 2010]. Ważnym elementem poruszanego zagadnienia jest również badanie metod szacowania parametrów przedmiotowych rozkładów. Oprócz powszechnie stosowanej metody największej wiarygodności analizuje się też efektywność innych metod, takich jak metoda najmniejszych kwadratów czy minimalizacji statystyki chi-kwadrat Pearsona [McDonald i Ransom 1979, Ostasiewicz 2013].

Niniejszy artykuł również wpisuje się w ten nurt badań. Jego celem jest porównanie wybranych modeli teoretycznych oraz metod ich estymacji pod kątem dokładności aproksymacji rozkładu dochodów mieszkańców Krakowa. W badaniach przedstawionych w polskiej literaturze przedmiotu często wskazuje się na modele klasy Burra jako modele szczególnie dobrze dopasowane do danych empirycznych. Analiza prowadzona w niniejszym artykule pozwoli ten wniosek zweryfikować zarówno ze względu na stosowaną metodę estymacji, jak i precyzję oszacowania charakterystyk liczbowych empirycznego rozkładu dochodów. Pod uwagę wzięto teoretyczne rozkłady najczęściej wykorzystywane w modelowaniu dochodów i posłużono się nimi do oszacowania wybranych miar położenia, zmienności i nierówności rozkładu dochodów. Estymację modeli przeprowadzono, wykorzystując dane o dochodach indywidualnych mieszkańców Krakowa

w 2013 r., które zaczerpnięto z zasobów „Diagnozy społecznej”, [www.diagnoza.com](http://www.diagnoza.com) (dostęp: 2.10.2015). W obliczeniach korzystano z programów komputerowych EasyFit oraz R. Otrzymane rozkłady dochodów cechują się jednomodalnością i prawostronną asymetrią. Oznacza to, że większość osób ma niższy poziom dochodów niż wynosi średni poziom dochodów w badanej zbiorowości. Osoby o dochodach powyżej średniej stanowią mniejszość. Taki niesymetryczny rozkład dochodów jest konsekwencją rozwarstwienia zamożności mieszkańców Krakowa, przy czym zjawisko to jest obserwowane w rozkładach dochodów całego społeczeństwa Polski, jak i z różnym natężeniem w innych krajach.

## 2. Wybrane modele rozkładu dochodów

W niniejszych badaniach wykorzystano popularne modele teoretyczne stosowane w aproksymacji rozkładu dochodów: Singha-Maddali (Burra typu XII), Fiska (logarytmiczno-logistyczny), log-normalny i gamma. Do oceny stopnia dopasowania rozkładu teoretycznego do empirycznego rozkładu dochodów zastosowano wartości statystyk Kołmogorowa-Smirnowa  $D_n$  i Andersona-Darlinga  $A^2$  [Anderson 1962]. Dodatkowo zgodność rozkładu teoretycznego i empirycznego zmierzono za pomocą kwadratu współczynnika korelacji  $R^2$  pomiędzy kwantylami teoretycznymi i kwantylami empirycznymi [D’Agostino i Stephens 1986, Kot 1999b].

Funkcja gęstości rozkładu Singha-Maddali [Singh i Maddala 1976] ma następującą postać:

$$f_{S-M}(x) = \alpha\beta^{-1}k \left(\frac{x}{\beta}\right)^{\alpha-1} \left(1 + \left(\frac{x}{\beta}\right)^{\alpha}\right)^{-k-1} \quad \text{dla } x > 0, \quad (1)$$

gdzie:  $\alpha, \beta, k$  – parametry rozkładu.

Wartość średnia i odchylenie standardowe w rozkładzie Singha-Maddali dane są wzorami:

$$\mu_{S-M} = \beta\Gamma\left(\frac{1+\alpha}{\alpha}\right)\Gamma\left(\frac{k\alpha-1}{\alpha}\right)\Gamma^{-1}(k), \quad (2)$$

$$\sigma_{S-M} = \frac{\beta}{\Gamma(k)} \sqrt{\Gamma\left(\frac{\alpha+2}{\alpha}\right)\Gamma(k)\Gamma\left(\frac{k\alpha-2}{\alpha}\right) - \Gamma^2\left(\frac{\alpha+1}{\alpha}\right)\Gamma^2\left(\frac{k\alpha-1}{\alpha}\right)}, \quad (3)$$

gdzie:  $\Gamma$  – funkcja gamma.

Modalną w przedmiotowym rozkładzie można zapisać następująco:

$$Mo_{S-M} = \beta \left( \frac{\alpha - 1}{k\alpha + 1} \right)^{\frac{1}{\alpha}}. \quad (4)$$

Należy zauważyć, że średnia (2) istnieje wówczas, gdy spełniony jest warunek  $k\alpha > 1$ , natomiast odchylenie standardowe (3) istnieje, o ile  $k\alpha > 2$ .

Gdy w modelu Singha-Maddali parametr  $k = 1$ , wówczas przyjmie on postać modelu Fiska [Fisk 1961] z parametrem położenia  $\beta$  i parametrem kształtu  $\alpha$ .

Kolejnym z rozważanych modeli jest rozkład log-normalny [Aitchison i Brown 1957]. Tutaj funkcja gęstości, wartość średnia, odchylenie standardowe, modalna dane są wzorami odpowiednio (5) – (8):

$$f_{L-N}(x) = \frac{1}{\sigma x \sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right) \text{ dla } x > 0, \quad (5)$$

$$\mu_{L-N} = \exp(\mu + 0,5\sigma^2), \quad (6)$$

$$\sigma_{L-N} = \sqrt{\exp(2\mu + \sigma^2)[\exp(\sigma^2) - 1]}, \quad (7)$$

$$Mo_{L-N} = \exp(\mu - \sigma^2), \quad (8)$$

gdzie:  $\mu, \sigma$  – parametry rozkładu ( $\mu$  – parametr położenia,  $\sigma$  – parametr kształtu).

W rozkładzie gamma funkcję gęstości można przedstawić następująco:

$$f_G(x) = \beta^{-\alpha} x^{\alpha-1} \exp(-x\beta^{-1}) \Gamma^{-1}(\alpha) \text{ dla } x > 0, \quad (9)$$

gdzie:  $\alpha, \beta$  – parametry rozkładu ( $\alpha$  – parametr kształtu,  $\beta$  – parametr położenia).

Średnia i odchylenie standardowe w tym rozkładzie dane są wzorami odpowiednio:

$$\mu_G = \alpha\beta, \quad (10)$$

$$\sigma_G = \beta\sqrt{\alpha}, \quad (11)$$

$$Mo_G = \beta(\alpha - 1). \quad (12)$$

Estymację parametrów omawianych modeli przeprowadzono za pomocą metody największej wiarygodności (MNW), metody najmniejszych kwadratów (MNK) i minimalizacji wartości statystyki chi-kwadrat Pearsona ( $\min \chi^2$ ). Posłużono się w tym celu danymi pogrupowanymi w szeregach rozdzielczych przedziałowych.

Jeśli szereg rozdzielczy składa się z  $k$  przedziałów klasowych, a liczebność w  $j$ -tej klasie wynosi  $n_j$ , natomiast prawdopodobieństwo tego, że dochód przyjmie wartość z klasy  $j$  wynosi  $p_j(\theta)$ , to funkcję wiarygodności wykorzystywaną w estymacji można przedstawić następująco [Kot 1999a]:

$$L(\theta) = \frac{n!}{n_1! n_2! \dots n_k!} \prod_{j=1}^k p_j^{n_j}(\theta), \quad (13)$$

gdzie:

$$n = n_1 + n_2 + \dots + n_k,$$

$\theta$  – wektor szacowanych parametrów modelu.

Funkcja (13) podlega maksymalizacji ze względu na wektor parametrów  $\theta$ , co pozwala wyznaczyć estymatory tych parametrów.

W metodzie najmniejszych kwadratów (MNK) parametry wektora  $\theta$  są dobierane w ten sposób, aby minimalizować funkcję [McDonald i Ransom 1979]:

$$K(\theta) = \sum_{j=1}^k \left( \frac{n_j}{n} - p_j(\theta) \right)^2. \quad (14)$$

Z kolei istota metody minimalizacji wartości statystyki  $\chi^2$  Pearsona polega na minimalizacji funkcji [McDonald i Ransom 1979]:

$$CH(\theta) = n \sum_{i=1}^k \frac{\left( \frac{n_j}{n} - p_j(\theta) \right)^2}{p_j(\theta)}. \quad (15)$$

Powyższe metody estymacji zostaną porównane w dalszej części opracowania pod względem jakości dopasowania czterech przedstawionych modeli teoretycznych, jak również pod względem precyzji oszacowania wybranych charakterystyk liczbowych rzeczywistego rozkładu dochodów.

### 3. Miary nierówności dochodowych i indeksy dobrobytu społecznego

Rozkłady Singha-Maddali, Fiska, log-normalny i gamma porównywano pod względem dokładności oszacowań miar nierówności dochodowych i indeksów dobrobytu społecznego. W praktyce najczęściej wykorzystywanym wskaźnikiem nierówności dochodowych jest miara Giniego  $G$ . Współczynnik ten stanowi iloraz przeciętnej absolutnej różnicy pomiędzy dochodami dwóch dowolnych jednostek i podwojonego dochodu przeciętnego [Kakwani 1980, Kot 2000]:

$$G = \frac{E|X - Y|}{2\mu}, \quad (16)$$

gdzie:

$X, Y$  – dochody dwóch różnych jednostek,

$\mu$  – przeciętny dochód.

Wyższe wartości współczynnika (16) świadczą o większych nierównościach dochodowych w społeczeństwie. Współczynnik Giniego w modelach Singha-Maddali,

Fiska, log-normalnym i gamma można wyrazić następującymi wzorami [Kot 1999b, Kot 2000]:

$$G_{S-M} = 1 - \Gamma\left(2k - \frac{1}{\alpha}\right) \Gamma(k) \Gamma^{-1}\left(k - \frac{1}{\alpha}\right) \Gamma^{-1}(2k), \quad (17)$$

$$G_F = \frac{1}{\alpha}, \quad (18)$$

$$G_{L-N} = 2\Phi\left(\frac{\sigma}{\sqrt{2}}\right) - 1, \quad (19)$$

$$G_G = \sqrt{\pi} \Gamma(\alpha + 0,5) \Gamma^{-1}(\alpha + 1). \quad (20)$$

gdzie:  $\Phi$  – dystrybuanta rozkładu normalnego standaryzowanego.

Kolejną analizowaną miarą nierówności jest współczynnik Atkinsona [Atkinson 1970] odwołujący się do tzw. ekwiwalentnego dochodu (czyli dochodu, który rozdzielony pomiędzy każdą jednostkę zapewniłby taki sam łączny dobrobyt, jak rozkład referencyjny). Wskaźnik ten można obliczyć według wzoru:

$$At = 1 - \frac{\left(E(X^{1-\varepsilon})\right)^{\frac{1}{1-\varepsilon}}}{E(X)}, \quad (21)$$

gdzie:

$\varepsilon$  – parametr wyrażający wagę przypisywaną transferom dochodów pomiędzy różnymi jednostkami,

$E(X)$  – wartość oczekiwana rozkładu dochodów.

Wraz ze wzrostem parametru  $\varepsilon$  rośnie znaczenie dochodów transferowanych do najuboższych jednostek. W szczególnej sytuacji, gdy  $\varepsilon = 0$ , to wszystkim transferom dochodów nadawana jest jednakowa waga.

Przedmiotem oszacowania w niniejszym artykule są również indeksy Sena [1973] i Kakwaniego [1980], które obliczamy według wzorów odpowiednio (22) i (23):

$$IS = \mu(1 - G), \quad (22)$$

$$IK = \frac{\mu}{1 + G}. \quad (23)$$

Wartości obu indeksów umożliwiają porównanie rozkładów dochodów ze względu na przeciętny poziom dochodów, jak i rozmiary nierówności dochodowych. Przyjmuje się, że jeśli współczynnik Giniego jest mniejszy od 0,5, to wskaźnik Sena (22) jest dominowany przez przeciętny dochód  $\mu$ . Natomiast w sytuacji, gdy  $G \geq 0,5$ , to wówczas indeks Giniego  $G$  dominuje wskaźnik Sena. Z kolei miernik Kakwaniego (23) jest zawsze bardziej wrażliwy na przeciętne dochody niż na wskaźnik Giniego [Kot 1999b].

#### 4. Wyniki badań empirycznych

Liczebność próby badawczej pochodzącej z „Diagnozy społecznej” zwiększono dodatkowo do 5000 obserwacji poprzez zastosowanie losowania bootstrapowego z odpowiedniego panelu gospodarstw domowych. Aby zapewnić właściwą reprezentatywność próby, zastosowano taki system ważenia dochodów, który uwzględnił różnicę między częstościami udziałów zbadanych osób w łącznej klasyfikacji według miejsca zamieszkania, wieku i płci, a faktyczną strukturą tych udziałów w populacji [Czapiński i Panek 2007]. Aproksymację teoretycznego rozkładu dochodów wykonano za pomocą trzech omówionych wcześniej metod: MNW, MNK i  $\min \chi^2$  Pearsona.

Wyniki oszacowania modeli Singha-Maddali, Fiska, log-normalnego i gamma dopasowanych do rozkładu dochodów mieszkańców Krakowa zawierają odpowiednio tabele 1–4. Biorąc pod uwagę wartości statystyk Kołmogorowa-Smirnowa  $D_n$ , Andersona-Darlinga  $A^2$ , a także  $R^2$ , trudno wskazać jedną metodę aproksymacji, która we wszystkich modelach zapewni najlepsze oszacowania parametrów. W modelach Singha-Maddali oraz Fiska najdokładniejsze przybliżenia rzeczywistych rozkładów dochodów uzyskano za pomocą metody MNW, natomiast w modelach log-normalnym i gamma najlepszą aproksymację empirycznych rozkładów daje metoda  $\min \chi^2$ . Powyższy wniosek znajduje odzwierciedlenie w wynikach każdej z trzech statystyk:  $D_n$ ,  $A^2$  oraz  $R^2$ .

Najgorsze dopasowania wszystkich czterech modeli najczęściej otrzymano, stosując aproksymację metodą MNK. Stwierdzono jedynie dwa odstępstwa od tej zasady: zgodnie z wartościami statystyki  $A^2$  najsłabsze dopasowanie modelu Fiska do danych empirycznych uzyskano za pomocą metody  $\min \chi^2$ , a według wartości  $R^2$  najgorsze przybliżenie rzeczywistego rozkładu dochodu za pomocą modelu gamma uzyskano stosując MNW.

Aby szczegółowo porównać modele teoretyczne szacowane różnymi metodami pod względem stopnia dopasowania do danych empirycznych, sporządzono ich rankingi według wartości każdej z trzech statystyk:  $D_n$ ,  $A^2$  oraz  $R^2$ . Uporządkowania modeli okazały się bardzo podobne, ale nie identyczne (por. tabelę 5). Potwierdzają to współczynniki korelacji rang Spearmana obliczone dla wszystkich par rankingów (tabela 6). Wszystkie współczynniki są dodatnie i na poziomie istotności 0,05 statystycznie istotne. Ich stosunkowo wysokie wartości wskazują na dużą zgodność uporządkowań modeli według wartości statystyk  $D_n$ ,  $A^2$ ,  $R^2$ .

Kierując się wartościami statystyki  $D_n$ , stwierdzamy, że najlepiej dopasowanymi modelami do empirycznego rozkładu dochodów mieszkańców Krakowa są kolejno: rozkład log-normalny (szacowany metodą  $\min \chi^2$ ), Singha-Maddali (MNW) i Singha-Maddali ( $\min \chi^2$ ). Natomiast według wartości  $D_n$  najsłabiej dopasowanymi modelami są: rozkład gamma (MNK), rozkład Fiska (MNK) i rozkład

Tabela 1. Wyniki estymacji parametrów modelu Singha-Maddali aproksymującego rozkład dochodów mieszkańców Krakowa w 2013 r.

| Metoda estymacji | Parametry modelu |         |         | Miary dopasowania do danych empirycznych |        |       |
|------------------|------------------|---------|---------|--|--------|-------|
|                  | $\alpha$         | $\beta$ | $k$     | $D_n$                                    | $A^2$  | $R^2$ |
| MNW              | 3,2472           | 1747    | 0,89533 | 0,04961                                  | 7,905  | 0,942 |
| MNK              | 3,1041           | 1809,6  | 1,00459 | 0,06872                                  | 13,925 | 0,884 |
| min $\chi^2$     | 3,2358           | 1756,8  | 0,90001 | 0,05054                                  | 10,118 | 0,918 |

Źródło: obliczenia własne na podstawie danych z badania „Diagnoza społeczna” 2013, www.diaгноza.com (dostęp: 2.10.2014).

Tabela 2. Wyniki estymacji parametrów modelu Fiska aproksymującego rozkład dochodów mieszkańców Krakowa w 2013 r.

| Metoda estymacji | Parametry modelu |         | Miary dopasowania do danych empirycznych |        |       |
|------------------|------------------|---------|--|--------|-------|
|                  | $\alpha$         | $\beta$ | $D_n$                                    | $A^2$  | $R^2$ |
| MNW              | 3,1334           | 1851,7  | 0,05587                                  | 8,769  | 0,908 |
| MNK              | 2,9249           | 1902,7  | 0,13143                                  | 12,601 | 0,831 |
| min $\chi^2$     | 3,0187           | 1886,4  | 0,07048                                  | 16,438 | 0,885 |

Źródło: jak do tabeli 1.

Tabela 3. Wyniki estymacji parametrów modelu log-normalnego aproksymującego rozkład dochodów mieszkańców Krakowa w 2013 r.

| Metoda estymacji | Parametry modelu |          | Miary dopasowania do danych empirycznych |        |       |
|------------------|------------------|----------|--|--------|-------|
|                  | $\mu$            | $\sigma$ | $D_n$                                    | $A^2$  | $R^2$ |
| MNW              | 7,5075           | 0,57965  | 0,05647                                  | 17,702 | 0,914 |
| MNK              | 7,6201           | 0,54021  | 0,11254                                  | 23,514 | 0,843 |
| min $\chi^2$     | 7,5242           | 0,57506  | 0,04924                                  | 13,092 | 0,934 |

Źródło: jak do tabeli 1.

Tabela 4. Wyniki estymacji parametrów modelu gamma aproksymującego rozkład dochodów mieszkańców Krakowa w 2013 r.

| Metoda estymacji | Parametry modelu |          | Miary dopasowania do danych empirycznych |         |       |
|------------------|------------------|----------|--|---------|-------|
|                  | $\alpha$         | $\beta$  | $D_n$                                    | $A^2$   | $R^2$ |
| MNW              | 2,1548           | 1046,215 | 0,12541                                  | 102,764 | 0,815 |
| MNK              | 2,3457           | 871,38   | 0,19312                                  | 161,882 | 0,835 |
| min $\chi^2$     | 2,2937           | 980,16   | 0,09616                                  | 89,738  | 0,864 |

Źródło: jak do tabeli 1.



Tabela 5. Rankingi modeli teoretycznych i metod estymacji według jakości dopasowania do danych empirycznych

| Rozkład (metoda estymacji)     | Miejsce w rankingu według wartości statystyki |       |       |
|--------------------------------|---|-------|-------|
|                                | $D_n$   | $A^2$ | $R^2$ |
| Singha-Maddali (MNW)           | 2   | 1     | 1     |
| Singha-Maddali (MNK)           | 6   | 6     | 7     |
| Singha-Maddali (min $\chi^2$ ) | 3   | 3     | 3     |
| Fiska (MNW)                    | 4   | 2     | 5     |
| Fiska (MNK)                    | 11  | 4     | 11    |
| Fiska (min $\chi^2$ )          | 7   | 7     | 6     |
| Log-normalny (MNW)             | 5   | 8     | 4     |
| Log-normalny (MNK)             | 9   | 9     | 9     |
| Log-normalny (min $\chi^2$ )   | 1   | 5     | 2     |
| Gamma (MNW)                    | 10  | 11    | 12    |
| Gamma                          | 12  | 12    | 10    |
| Gamma (min $\chi^2$ )          | 8   | 10    | 8     |

Źródło: jak do tabeli 1.

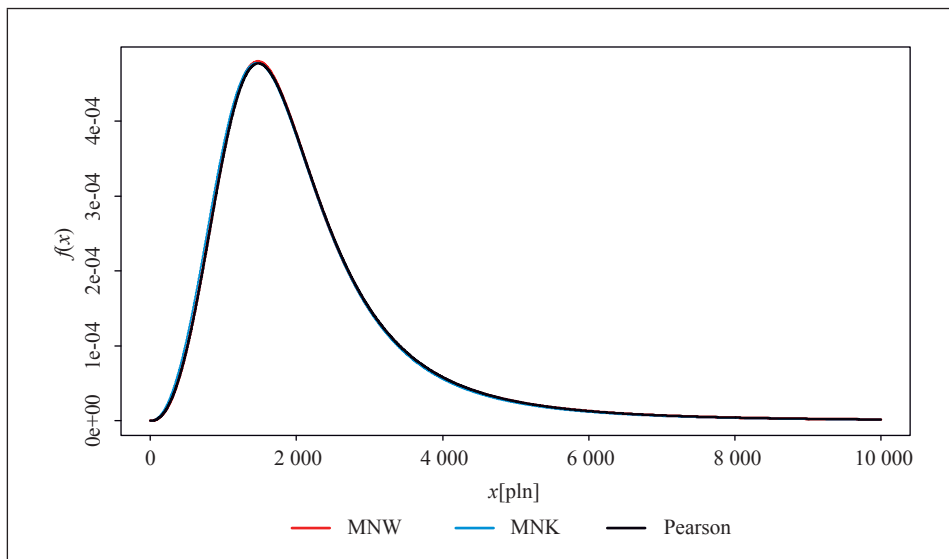
Tabela 6. Współczynniki korelacji rang Spearmana dla uporządkowań modeli według jakości dopasowania do danych empirycznych

| Pary rankingów modeli | $r_s$ | $t$   | $p$   |
|-----------------------|-------|-------|-------|
| $D_n$ & $A^2$         | 0,706 | 3,155 | 0,010 |
| $D_n$ & $R^2$         | 0,951 | 9,732 | 0,000 |
| $A^2$ & $R^2$         | 0,671 | 2,864 | 0,017 |

Źródło: jak do tabeli 1.

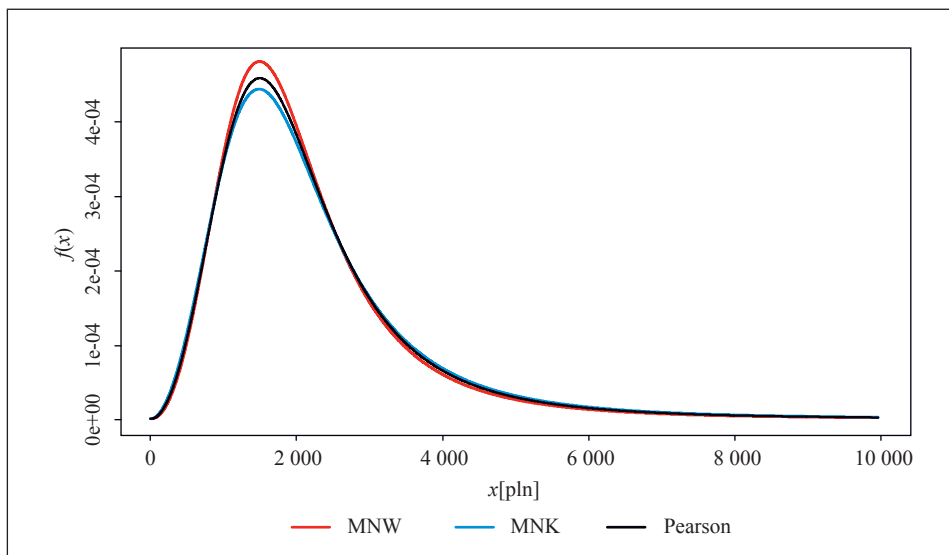
gamma (MNW). Jeśli kierować się wartościami statystyki  $A^2$ , to najlepszą aproksymację dochodów otrzymujemy za pomocą rozkładu Singha-Maddali (MNW), Fiska (MNW) oraz Singha-Maddali (min  $\chi^2$ ). Najślabiej dopasowanymi modelami w sensie  $A^2$  są modele gamma (MNK), gamma (MNW), gamma (min  $\chi^2$ ). Biorąc pod uwagę wartości kwadratu współczynnika korelacji pomiędzy kwantylami rozkładu teoretycznego i empirycznego, należy zauważyć, że najlepiej dopasowanym modelami do danych empirycznych są kolejno: rozkład Singha-Maddali (MNW), log-normalny (min  $\chi^2$ ) i Singha-Maddali (min  $\chi^2$ ), a najgorzej dopasowane są odpowiednio: rozkład gamma (MNW), Fiska (MNK) i gamma (MNK).

Przebieg funkcji gęstości czterech rozważanych rozkładów teoretycznych szacowanych za pomocą metod MNW, MNK i min  $\chi^2$  Pearsona przedstawiono na rys. 1–4. Analizując wykresy funkcji gęstości, można stwierdzić, że stosując trzy



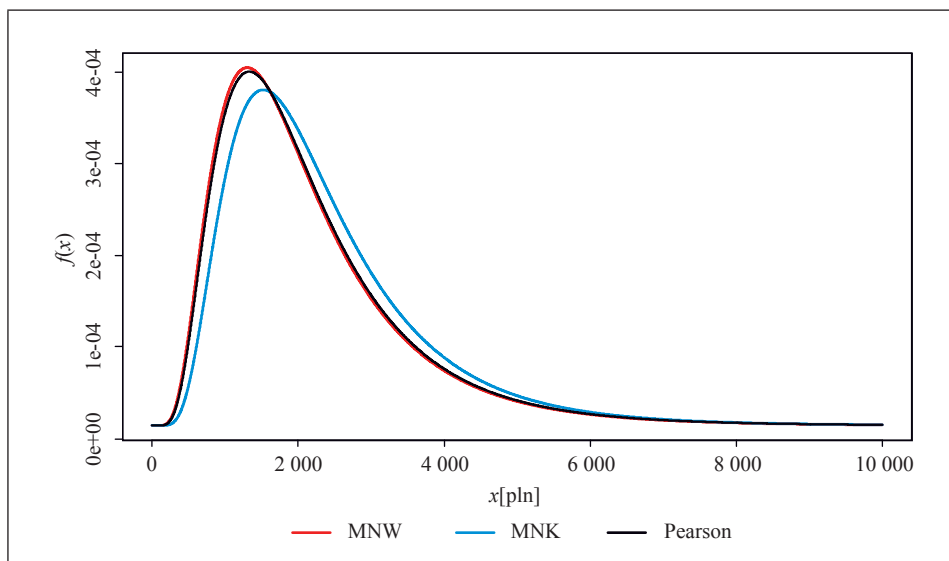
Rys. 1. Funkcja gęstości modelu Singha-Maddali aproksymującego empiryczny rozkład dochodów mieszkańców Krakowa w 2013 r.

Źródło: opracowanie własne na podstawie danych z badania „Diagnoza społeczna” 2013, [www.diagnoza.com](http://www.diagnoza.com) (dostęp: 2.10.2014).



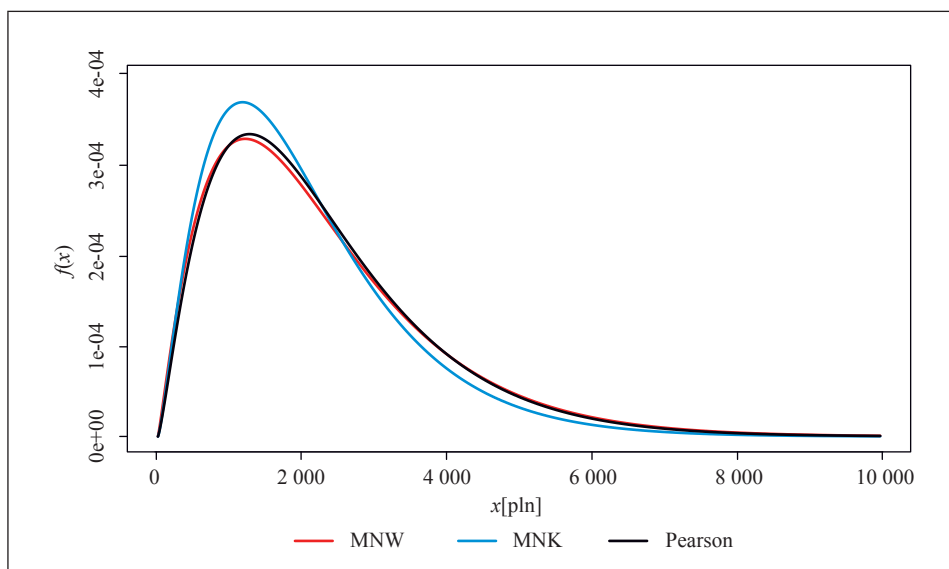
Rys. 2. Funkcja gęstości modelu Fiska aproksymującego empiryczny rozkład dochodów mieszkańców Krakowa w 2013 r.

Źródło: jak do rys. 1.



Rys. 3. Funkcja gęstości modelu log-normalnego aproksymującego empiryczny rozkład dochodów mieszkańców Krakowa w 2013 r.

Źródło: jak do rys. 1.



Rys. 4. Funkcja gęstości modelu gamma aproksymującego empiryczny rozkład dochodów mieszkańców Krakowa w 2013 r.

Źródło: jak do rys. 1.

różne metody estymacji parametrów, najbardziej zbliżone aproksymacje rozkładu empirycznego otrzymano w modelu Singha-Maddali.

Posługując się oszacowanymi modelami teoretycznymi, obliczono wybrane charakterystyki liczbowe rozkładów, tj. średnią, modalną, odchylenie standardowe i współczynnik zmienności, a otrzymane wyniki porównano z charakterystykami obliczonymi dla rzeczywistego rozkładu dochodów mieszkańców Krakowa (tabela 7).

Na podstawie przedstawionych rezultatów można stwierdzić, że modele teoretyczne Singha-Maddali, Fiska i gamma szacowane metodą MNW przeszacowują średnią empiryczną wysokość dochodów, a także przeszacowują odchylenie standardowe i współczynnik zmienności. Natomiast wykorzystując rozkład log-normalny szacowany metodą MNW, otrzymamy zaniżone oceny tych charakterystyk.

Niedoszacowanie średniej i modalnej rzeczywistych dochodów mieszkańców Krakowa ma również miejsce, gdy stosujemy modele Singha-Maddali i gamma w połączeniu z metodą MNK, a przeszacowanie tych charakterystyk otrzymujemy dla modeli Fiska i log-normalnego, gdy są one szacowane metodą MNK. Większość rozkładów teoretycznych (z wyjątkiem modelu Fiska) szacowanych metodą  $\min \chi^2$  zaniża wartość modalną empirycznego rozkładu dochodów i jednocześnie zawyża wartość odchylenia standardowego i współczynnika zmienności (z wyjątkiem rozkładu log-normalnego). Ponadto rozkłady Fiska i gamma szacowane metodą  $\min \chi^2$  przeszacowują średnie dochody, a pozostałe dwa rozkłady estymowane tą metodą niedoszacowują średniej. W większości przypadków to, czy model teoretyczny niedoszacowuje, czy też przeszacowuje miary położenia lub zmienności, zależy od wyboru metody estymacji. Jedynie dwa modele: Singha-Maddali i Fiska zawsze przeszacowywały odchylenie standardowe i współczynnik zmienności (niezależnie od metody estymacji), a model log-normalny zaniżał wartości tych charakterystyk.

Najmniejszy względny błąd oszacowania średniego dochodu popełnimy, stosując rozkład Fiska szacowany metodą MNW, a największy – gdy posłużymy się rozkładem log-normalnym szacowanym metodą MNK. Najlepsze oszacowanie modalnej rozkładu dochodów otrzymamy, stosując model Singha-Maddali estymowany metodą  $\min \chi^2$ , a najgorsze – gdy zastosujemy rozkład gamma w połączeniu z metodą MNK. Najdokładniejsze oszacowanie zarówno odchylenia standardowego, jak i współczynnika zmienności uzyskano, posługując się modelem gamma i metodą  $\min \chi^2$ , a najmniej dokładne, gdy zastosowano model Fiska i metodą MNK. Uwzględniając średni względny błąd oszacowania dla poszczególnych charakterystyk liczbowych i biorąc pod uwagę wszystkie metody estymacji, można stwierdzić, że przeciętnie najdokładniejsze oszacowania średniej i modalnej otrzymano za pomocą modelu Singha-Maddali, natomiast najdo-

kładniejsze oszacowania odchylenia standardowego i współczynnika zmienności uzyskano za pomocą modelu gamma.

Tabela 7. Wybrane miary położenia i zmienności rozkładu dochodów mieszkańców Krakowa w 2013 r.

| Rozkład        | Metoda estymacji | Średnia  | Modalna  | Odchylenie standardowe | Współczynnik zmienności |
|----------------|------------------|----------|----------|------------------------|-------------------------|
| Singha-Maddali | MNW              | 2217,708 | 1473,364 | 1815,807               | 0,819                   |
|                | MNK              | 2152,933 | 1457,552 | 1630,907               | 0,758                   |
|                | min $\chi^2$     | 2225,019 | 1477,838 | 1819,614               | 0,818                   |
| Fiska          | MNW              | 2202,615 | 1499,346 | 1651,691               | 0,750                   |
|                | MNK              | 2324,577 | 1491,358 | 1969,224               | 0,847                   |
|                | min $\chi^2$     | 2275,474 | 1501,689 | 1819,867               | 0,800                   |
| Log-normalny   | MNW              | 2154,895 | 1301,803 | 1361,734               | 0,632                   |
|                | MNK              | 2359,095 | 1522,78  | 1373,292               | 0,582                   |
|                | min $\chi^2$     | 2185,384 | 1330,761 | 1368,151               | 0,626                   |
| Gamma          | MNW              | 2254,384 | 1208,169 | 1535,764               | 0,681                   |
|                | MNK              | 2043,996 | 1172,616 | 1334,578               | 0,653                   |
|                | min $\chi^2$     | 2248,193 | 1268,033 | 1484,449               | 0,660                   |
| Empiryczny     |                  | 2196,116 | 1482,721 | 1459,503               | 0,665                   |

Źródło: jak do tabeli 1.

Teoretyczne modele rozkładu dochodów wykorzystano następnie do obliczenia wybranych miar nierówności dochodowych oraz indeksów dobrobytu społecznego, a otrzymane wyniki porównano z rezultatami obliczonymi dla empirycznych rozkładów dochodów. Wyniki współczynników Giniego, Atkinsona ( $\varepsilon = 0,5$ ), wskaźnika Sena i Kakwaniego przedstawiono w tabeli 8.

Z tabeli 8 wynika, że większość modeli przeszacowuje zarówno empiryczny wskaźnik Giniego, jak i wskaźnik Atkinsona. Współczynnik Giniego jest zawsze przeszacowywany w modelach Singha-Maddali i gamma (niezależnie od stosowanej metody estymacji), a także w modelu Fiska w połączeniu z metodami MNK oraz min  $\chi^2$ . Natomiast w modelu log-normalnym przedmiotowy wskaźnik jest zawsze zaniżony przy każdej z trzech metod estymacji. Wskaźnik Atkinsona jest z kolei przeszacowywany, gdy stosujemy modele Singha-Maddali, Fiska i gamma oraz niedoszacowany przez model log-normalny (bez względu na metodę estymacji parametrów modeli). Ponadto należy zwrócić uwagę, że modele Singha-Maddali i Fiska najczęściej przeszacowują wartości indeksów dobrobytu społecznego (w zależności od przyjętej metody estymacji), a pozostałe modele przeważnie ich niedoszacowują. Najbardziej precyzyjne oszacowanie wskaź-

Tabela 8. Wybrane miary nierówności dochodowych i indeksy dobrobytu społecznego

| Rozkład        | Metoda estymacji | Współczynnik Giniego | Współczynnik Atkinsona ( $\epsilon = 0,5$ ) | Wskaźnik Sena | Wskaźnik Kakwaniego |
|----------------|------------------|----------------------|---|---------------|---------------------|
| Singha-Maddali | MNW              | 0,327                | 0,111                                       | 1493,101      | 1671,551            |
|                | MNK              | 0,321                | 0,086                                       | 1460,976      | 1629,279            |
|                | min $\chi^2$     | 0,327                | 0,090                                       | 1497,565      | 1676,801            |
| Fiska          | MNW              | 0,319                | 0,085                                       | 1499,668      | 1669,733            |
|                | MNK              | 0,342                | 0,098                                       | 1529,823      | 1732,313            |
|                | min $\chi^2$     | 0,331                | 0,092                                       | 1521,682      | 1709,253            |
| Log-normalny   | MNW              | 0,318                | 0,081                                       | 1469,419      | 1634,847            |
|                | MNK              | 0,298                | 0,070                                       | 1657,198      | 1818,146            |
|                | min $\chi^2$     | 0,316                | 0,079                                       | 1495,417      | 1660,981            |
| Gamma          | MNW              | 0,363                | 0,109                                       | 1436,354      | 1654,154            |
|                | MNK              | 0,349                | 0,100                                       | 1329,840      | 1514,753            |
|                | min $\chi^2$     | 0,353                | 0,103                                       | 1454,778      | 1661,743            |
| Empiryczny     |                  | 0,319                | 0,083                                       | 1495,241      | 1664,805            |

Źródło: jak do tabeli 1.

nika Giniego rozkładu dochodów mieszkańców Krakowa uzyskano za pomocą modelu Fiska w połączeniu z metodą MNW, a najmniej precyzyjne oszacowanie tego wskaźnika otrzymano za pomocą modelu gamma i metody MNW. Najlepszą ocenę wskaźnika Atkinsona również otrzymano przy wykorzystaniu modelu Fiska i metody MNW, a najgorszej oceny tego wskaźnika dostarczył model Singha-Maddali z metodą MNW. Model log-normalny (min  $\chi^2$ ) najbardziej precyzyjnie odzwierciedlił wskaźnik Sena, a najgorszą ocenę tego wskaźnika otrzymano za pomocą modelu gamma (MNK). Najlepsze oszacowanie wskaźnika Kakwaniego uzyskano za pomocą modelu gamma w połączeniu z metodą min  $\chi^2$ , a najgorszą ocenę tego wskaźnika dostarczył model log-normalny (MNK). Posługując się średnim względnym błędem oszacowania obliczonym dla poszczególnych miar nierówności dochodowych oraz indeksów dobrobytu społecznego (przy uwzględnieniu wszystkich rozpatrywanych metod estymacji), można stwierdzić, że przeciętnie najlepsze oszacowanie współczynnika Giniego, a także wskaźników Sena i Kakwaniego zapewnia model Singha-Maddali, natomiast przeciętnie najdokładniejsze oszacowanie wskaźnika Atkinsona można uzyskać za pomocą modelu log-normalnego.

## 5. Podsumowanie

Na podstawie wyników modelowania rozkładu dochodów mieszkańców Krakowa można stwierdzić, że dokładność oszacowania empirycznych charakterystyk tego rozkładu, jak i sposób tego oszacowania (przeszacowanie, niedoszacowanie) zależą w wielu przypadkach nie tylko od przyjętego modelu teoretycznego, ale również od stosowanej metody estymacji parametrów. Analizując przeciętne względne błędy oszacowań charakterystyk rozkładów dla poszczególnych metod estymacji, należy przede wszystkim rekomendować metodę MNW. Modele szacowane za pomocą tej metody przeciętnie najlepiej aproksymowały średnią, miary zróżnicowania, współczynnik Giniego i współczynnik Kakwaniego. Do szacowania modalnej rozkładu dochodów należałoby rekomendować metodę MNK, a do aproksymacji współczynników Atkinsona i Sena – metodę  $\min \chi^2$  Pearsona.

W niniejszym artykule wykazano, że najlepsze dopasowanie do danych empirycznych ma model Singha-Maddali. Jego średnia pozycja w rankingach utworzonych względem różnych miar dobroci dopasowania jest wyraźnie wyższa od pozostałych modeli. Można też stwierdzić, że model ten dostarcza przeciętnie najdokładniejszych oszacowań niektórych charakterystyk liczbowych rozkładu dochodów. Dotyczy to zwłaszcza średniej, modalnej, współczynnika Giniego, indeksów Sena i Kakwaniego. Wniosek jest zgodny z wynikami podobnych badań prowadzonych wcześniej nad rozkładami dochodów [McDonald i Ransom 1979, Kot 2000, Ostasiewicz 2013].

Jednak stwierdzony charakter oszacowania tym modelem niektórych charakterystyk rozkładu dochodów jest sprzeczny z wynikami innych autorów. J.B. McDonald, M.R. Ransom i K. Ostasiewicz wskazują na tendencje do niedoszacowywania przez model Singha-Maddali wielu charakterystyk liczbowych, podczas gdy w artykule wykazano, że częściej przeszacowuje on te charakterystyki. Przyczyn rozbieżności można upatrywać w odmiennej próbie badawczej, czy innym okresie badania.

Aby ostatecznie potwierdzić zaobserwowane własności rozkładu Singha-Maddali, konieczne są dodatkowe badania przy zwiększonej liczebności próby badawczej i zastosowaniu ewentualnie innych metod estymacji parametrów rozkładu.

Znalezienie możliwie najlepszego modelu rozkładu dochodów oraz najlepszej metody jego estymacji pozwala potem na szczegółową i rzetelną ocenę właściwości rozkładu dochodów, dokładne zbadanie jego charakterystyk oraz porównanie go z innymi rozkładami dochodów. Ponadto znajomość teoretycznego rozkładu dochodów często jest konieczna do prowadzenia bardziej zaawansowanych analiz statystycznych i ekonometrycznych w zakresie struktury zamożności społeczeństwa, zróżnicowania oraz dysproporcji dochodów itp. Taka wiedza

może być pomocna np. w odpowiednim kształtowaniu polityki społecznej przez właściwe gremia rządowe i samorządowe. Przedmiotowe badania wzbogacają także teorie genezy i kształtowania się rozkładu dochodów w społeczeństwie.

## Literatura

- Aitchison J., Brown J.A.C. [1957], *The Lognormal Distribution with Special Reference to Its Use in Economics*, Cambridge University Press, Cambridge.
- Anderson T.W. [1962], *On the Distribution of the Two-Sample Cramer-von Mises Criterion*, „The Annals of Mathematical Statistics”, vol. 33(3), <http://dx.doi.org/10.1214/aoms/1177704477>.
- Atkinson A.B. [1970], *On Measurement of Inequality*, „Journal of Economic Theory”, vol. 2(3), [http://dx.doi.org/10.1016/0022-0531\(70\)90039-6](http://dx.doi.org/10.1016/0022-0531(70)90039-6).
- Clementi F., Gallegati M., Kaniadakis G. [2010], *A Model of Personal Income Distribution with Application to Italian Data*, „Empirical Economics”, vol. 39(2), <http://dx.doi.org/10.1007/s00181-009-0318-2>.
- Czapiński J., Panek T. [2007], *Diagnoza Społeczna 2000–2007. Warunki i jakość życia Polaków*, Instytut Psychologii Uniwersytetu Warszawskiego, Rada Monitoringu Społecznego przy Wyższej Szkole Psychologii i Zarządzania w Warszawie, <http://www.ads.org.pl/opis-szczeg.php?id=58>.
- D’Agostino R.B., Stephens M.A. [1986], *Goodness-of-fit Techniques*, Basel, Marcel Dekker Inc., New York.
- Fisk P.R. [1961], *The Graduation of Income Distributions*, „Econometrica”, vol. 29(2), <http://dx.doi.org/10.2307/1909287>.
- Kakwani N.C. [1980], *Income Inequality and Poverty: Methods of Estimation and Policy Applications*, Oxford University Press, New York–Oxford–London.
- Kleiber C., Kotz S. [2003], *Statistical Size Distributions in Economics and Actuarial Sciences*, John Wiley and Sons, Inc., Hoboken, New Jersey.
- Kot S.M. [1999a], *Empiryczna weryfikacja rozkładów płac w Polsce [w:] Analiza ekonometryczna kształtowania się płac w Polsce w okresie transformacji*, red. S.M. Kot, Wydawnictwo Naukowe PWN, Warszawa.
- Kot S.M. [1999b], *Teorie genezy rozkładów płac [w:] Analiza ekonometryczna kształtowania się płac w Polsce w okresie transformacji*, red. S.M. Kot, Wydawnictwo Naukowe PWN, Warszawa.
- Kot S.M. [2000], *Ekonometryczne modele dobrobytu*, Wydawnictwo Naukowe PWN, Warszawa–Kraków.
- McDonald J.B., Ransom M.R. [1979], *Functional Forms, Estimation Techniques and the Distribution of Income*, „Econometrica”, vol. 47(6), <http://dx.doi.org/10.2307/1914015>.
- Ostasiewicz K. [2013], *Adekwatność wybranych rozkładów teoretycznych dochodów w zależności od metody aproksymacji*, „Przegląd Statystyczny”, vol. 60(4).
- Sen A.K. [1973], *On Ignorance and Equal Distribution*, „American Economic Review”, vol. 63(5).
- Singh S.K., Maddala G.S. [1976], *A Function for Size Distribution of Incomes*, „Econometrica”, vol. 44(5), <http://dx.doi.org/10.2307/1911538>.



Ulman P. [2011], *Sytuacja osób niepełnosprawnych i ich gospodarstw domowych w Polsce*, Zeszyty Naukowe, Seria specjalna: Monografie, nr 199, Wydawnictwo Uniwersytetu Ekonomicznego w Krakowie, Kraków.

### **The Effects of Parameter Estimation Methods of Theoretical Models of Income Distribution on the Quality of Approximation of the Empirical Income Distribution of the Inhabitants of Cracow**

(Abstract)

The purpose of this article is to find theoretical models and the best estimation methods that as closely as possible reflect the income distribution of the inhabitants of Cracow. The study took into account models most commonly used in the analysis of wages and incomes – Singh-Maddala, Fisk, log-normal distribution, and gamma distribution. Selected distributions were tested for their precision in estimating the measures of position, dispersion, and inequality in empirical income distribution depending on the method of parameter estimation used. To estimate the models, data on the individual income of the inhabitants of Cracow in 2013 were used; the data were taken from the “Social Diagnosis” database. The results indicate that the most precise estimation of income distribution and its characteristics can be achieved using Singh-Maddala distribution, while the recommended parameter estimation methods is maximum likelihood method.

**Keywords:** income distribution, maximum likelihood method, Gini coefficient, Singh-Maddala distribution.