

Jan PURCZYŃSKI, Kamila BEDNARZ

## **METODY ESTYMACJI PARAMETRÓW UOGÓLNIIONEGO ROZKŁADU GAUSSA**

### *Streszczenie*

*W pracy zaproponowano nową metodę estymacji parametru kształtu uogólnionego rozkładu Gaussa. Porównano właściwości estymatorów uzyskanych MNW, metodą Mallata [5] oraz proponowaną metodą. Jakość estymatora określano na podstawie wartości względnego błędu średnio kwadratowego. Wykonano symulacje komputerowe z wykorzystaniem generatorów liczb losowych dla parametru kształtu  $s = 0,5$ ,  $s=1,0$ ,  $s=2,0$  oraz  $s = 3,0$ .*

### **WSTĘP**

W pracy zostanie rozpatrzony uogólniony rozkład Gaussa:

$$f(x) = \frac{\lambda \cdot s}{2 \cdot \Gamma\left(\frac{1}{s}\right)} \cdot \exp(-\lambda^s |x - \mu|^s) \quad (1)$$

gdzie  $\Gamma(z)$  – funkcja gamma Eulera;

$s$  – parametr kształtu;

$\lambda$  - parametr skali.

Parametr skali  $\lambda$  spełnia zależność:

$$\lambda = \frac{1}{\sigma} \left[ \frac{\Gamma\left(\frac{3}{s}\right)}{\Gamma\left(\frac{1}{s}\right)} \right]^{\frac{1}{2}}$$

gdzie  $\sigma$  - odchylenie standardowe rozkładu.

Rozkład, którego gęstość opisuje wzór (1), nosi nazwę Generalized Gaussian Distribution (GGD) lub Generalized Error Distribution (GED).

Dla  $s = 1$ , rozkład GGD przechodzi w rozkład Laplace'a (dwuwykładniczy), natomiast dla  $s=2$  uzyskujemy rozkład Gaussa- stąd też pochodzi nazwa: uogólniony rozkład Gaussa.

Ze względu na zmieniającą się wartość parametru kształtu  $s$  (wzór(1)), rozkład ten umożliwia modelowanie szeregu zmiennych fizycznych oraz ekonomicznych. Rozkład GGD znalazł zastosowanie w rozpoznawaniu obrazów, modelowaniu zakłóceń sygnałów oraz

modelowaniu mowy [3]. Znajduje szerokie zastosowanie w kompresji obrazów, gdzie służy do modelowania rozkładu współczynników dyskretnej transformaty kosinusowej (DCT) [5]. Rozkład GED jest z powodzeniem stosowany do modelowania rozkładu stóp zwrotu indeksów giełdowych i spółek [8]. Rozkłady o tzw. "grubych ogonach" znalazły zastosowanie w modelowaniu zmieniającej się w czasie warunkowej wariancji, m.in. [2] a także [6], gdzie przy estymacji modelu GARCH, jako rozkład warunkowy, wykorzystano rozkład GED.

W celu uproszczenia rozważań, przyjmuje się, że na podstawie próby zostało wyznaczone oszacowanie parametru  $\mu$ :

$$\hat{\mu} = \frac{1}{N} \cdot \sum_{k=1}^N x_k \quad (2)$$

a następnie, ciąg wartości  $x_k$  został scentrowany poprzez odjęcie  $\hat{\mu}$ .

W związku z tym, rozważa się gęstość o postaci:

$$f(x) = \frac{\lambda \cdot s}{2 \cdot \Gamma\left(\frac{1}{s}\right)} \cdot \exp(-|\lambda \cdot x|^s) \quad (3)$$

Właściwości uogólnionego rozkładu błędów zostały po raz pierwszy opisane w pracy [7], natomiast z większym odzewem spotkała się praca [1], dzięki której rozkład GGD znalazł szerokie zastosowanie w zagadnieniu modelowania rozkładu. Jedną z bardziej popularnych metod estymacji parametrów rozkładu jest metoda Mallata zaprezentowana w pracy [5].

W niniejszej pracy zostanie zaproponowana nowa metoda estymacji parametrów rozkładu GGD.

## 1. WYBRANE METODY ESTYMACJI PARAMETRÓW ROZKŁADU GED

### 1.1. Metoda największej wiarygodności (MNW)

Stosując MNW, wyznacza się logarytm funkcji wiarygodności :

$$\ln(L(\lambda, s)) = N \cdot \ln(\lambda) + N \cdot \ln\left(\frac{s}{2\Gamma\left(\frac{1}{s}\right)}\right) - \sum_{k=1}^N |\lambda x_k|^s$$

Z warunków

$$\frac{\partial \ln(L(\lambda, s))}{\partial \lambda} = 0 \quad ; \quad \frac{\partial \ln(L(\lambda, s))}{\partial s} = 0$$

otrzymuje się

$$\lambda = \left( \frac{N}{s \sum_{k=1}^N |x_k|^s} \right)^{\frac{1}{s}} \quad (4)$$

oraz

$$gW(s) = s + \Psi\left(\frac{1}{s}\right) + \ln\left(\frac{s}{N} \sum_{k=1}^N |x_k|^s\right) - \frac{\sum_{k=1}^N |x_k|^s \ln|x_k|}{\sum_{k=1}^N |x_k|^s} = 0 \quad (5)$$

gdzie:  $\Psi(z) = \frac{d}{dz} [\ln \Gamma(z)]$

Z równania (5) wyznacza się parametr kształtu  $s$ , a następnie ze wzoru (4) - parametr  $\lambda$ .

## 1.2. Proponowana metoda

Proponowana metoda estymacji parametrów rozkładu GED bazuje na momentach absolutnych.

Moment absolutny rzędu  $m$  wyraża się wzorem:

$$E_m = \int_{-\infty}^{\infty} |x|^m \cdot f(x) \cdot dx \quad (6)$$

Ze wzorów (3) i (6), otrzymuje się :

$$E_m = \frac{\Gamma\left(\frac{m+1}{s}\right)}{\lambda^m \cdot \Gamma\left(\frac{1}{s}\right)} \quad (7)$$

Estymator momentu  $E_m$  ma postać:

$$\hat{E}_m = \frac{1}{N} \cdot \sum_{k=1}^N |x_k|^m \quad (8)$$

Przyjmując dwie różne wartości momentów  $m_1$  i  $m_2$  we wzorze (7) i eliminując parametr  $\lambda$ , otrzymuje się:

$$G(s) = \frac{E_{m_1}}{(E_{m_2})^{\frac{m_1}{m_2}}} = \frac{\Gamma\left(\frac{m_1+1}{s}\right)}{\left[\Gamma\left(\frac{1}{s}\right)\right]^{1-\frac{m_1}{m_2}} \left[\Gamma\left(\frac{m_2+1}{s}\right)\right]^{\frac{m_1}{m_2}}} \quad (9)$$

Przyjmując we wzorze (10)  $m_2 = 2 \cdot m_1$ , uzyskuje się:

$$g = \frac{E_{2m_1}}{(E_{m_1})^2} = \frac{\Gamma\left(\frac{2m_1+1}{s}\right) \cdot \Gamma\left(\frac{1}{s}\right)}{\left[\Gamma\left(\frac{m_1+1}{s}\right)\right]^2} \quad (10)$$

Funkcja odwrotna do funkcji  $g$  (wzór (10)) przyjmuje następującą postać, zależną od wartości momentów  $m_2 = 2 \cdot m_1$  :

$$m_1 = 0,25; \quad m_2 = 0,5 \quad \hat{s}_0 = \begin{cases} 55g^{-70} + 0,73 & \text{dla } g \leq 1,079 \\ 5,7g^{-28} + 0,315 & \text{dla } 1,079 < g < 1,132 \\ 2,05g^{-15} + 0,18 & \text{dla } g \geq 1,132 \end{cases} \quad (11a)$$

$$(11b) \quad m_1 = 0,5; \quad m_2 = 1,0 \quad \hat{s}_1 = \begin{cases} 26g^{-18} + 0,67 & \text{dla } g \leq 1,27 \\ 5,5g^{-9} + 0,365 & \text{dla } g > 1,27 \end{cases}$$

$$m_1 = 1; \quad m_2 = 2,0 \quad \hat{s}_2 = \begin{cases} 29g^{-7} + 0,8 & \text{dla } g \leq 2 \\ 5g^{-3} + 0,37 & \text{dla } g > 2 \end{cases} \quad (11c)$$

$$m_1 = 2 \quad m_2 = 4 \quad \hat{s}_3 = \begin{cases} 12g^{-1,98} + 0,64 & \text{dla } g \leq 6 \\ 6g^{-1,3} + 0,42 & \text{dla } g > 6 \end{cases} \quad (11d)$$

Oszacowanie końcowe  $\hat{s}_p$  wyznacza się z następującej zależności:

$$\hat{s}_p = \begin{cases} \hat{s}_3 & \text{dla } \hat{s}_1 > 1,9 \\ \hat{s}_2 & \text{dla } 1,4 < \hat{s}_1 \leq 1,9 \\ \hat{s}_1 & \text{dla } 0,53 < \hat{s}_1 \leq 1,4 \\ \hat{s}_0 & \text{dla } \hat{s}_1 \leq 0,53 \end{cases} \quad (12)$$

Na koniec, na podstawie oszacowania  $\hat{s}_p$  wyznacza się oszacowanie parametru  $\hat{\lambda}$ :

$$\hat{\lambda} = \left[ \frac{\Gamma\left(\frac{m+1}{\hat{s}_p}\right)}{E_m \cdot \Gamma\left(\frac{1}{\hat{s}_p}\right)} \right]^{\frac{1}{m}} \quad (13)$$

gdzie  $m = m_1$  lub  $m = m_2$ .

Proponowana metoda stanowi uogólnienie metody Mallata [5], która bazuje na dwóch momentach absolutnych rzędu  $m_1 = 1$ ;  $m_2 = 2$ . Oznacza to, że w metodzie Mallata jako wynik końcowy przybliżenia  $\hat{s}$  przyjmuje się wartość  $\hat{s}_2$  określoną wzorem (11c).

## 2. WYNIKI SYMULACJI KOMPUTEROWYCH

W celu oceny jakości poszczególnych estymatorów wykonano eksperyment numeryczny z wykorzystaniem generatora liczb losowych o rozkładzie GED dla wybranych wartości parametru kształtu:  $s=0,5$ ,  $s=1$  (rozkład Laplace'a),  $s=2$  (rozkład Gaussa),  $s=3$ . Symulacje komputerowe polegały na wykonaniu  $K=2000$  powtórzeń i wyznaczeniu na tej podstawie błędu estymacji RRMSE (*Relative Root Mean-Squared Error*):

$$RRMSE = \sqrt{\frac{1}{K} \sum_{k=1}^K \left( \frac{\hat{s}_k - s_d}{s_d} \right)^2} \quad (14)$$

gdzie:  $\hat{s}_k$  - oszacowanie wartości parametru  $s$  w  $k$ -tej symulacji;  
 $s_d$  - wartość dokładna parametru kształtu  $s$ .

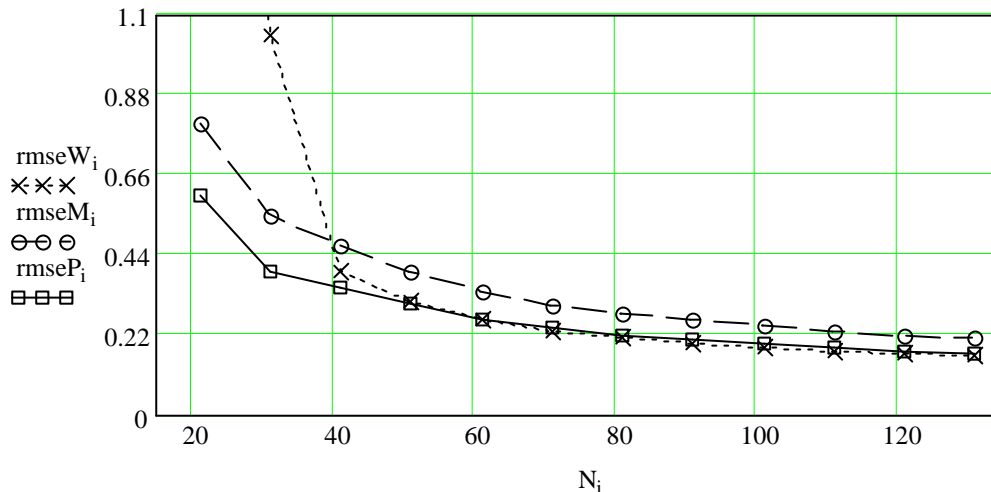
Zamieszczając wyniki symulacji komputerowych zastosowano oznaczenie *rmse* w miejsce *RRMSE*.

Przy rozwiązaniu równania (5) stosowana była metoda bisekcji (połowienia).

Obliczenia wykonano dla zmieniającej się długości próbki (liczby obserwacji)  $N=21,31,41,\dots,131$ .

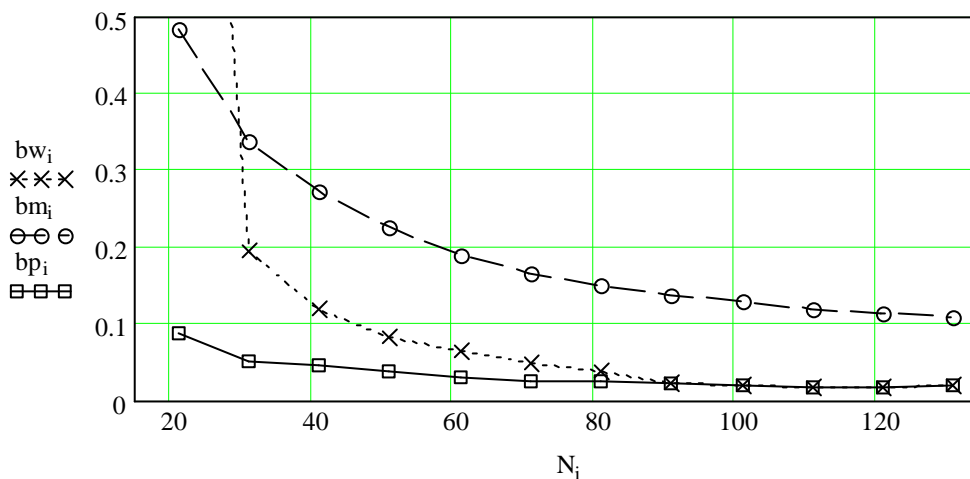
Na rysunku 1 przedstawiono wartości błędu dla generatora ze współczynnikiem kształtu  $s=0,5$ . Na uwagę zasługuje bardzo duży błąd względny  $rmseW=2,934$  dla próbki z  $N=21$

elementami. Z rysunku 1 wynika, że najmniejszy błąd średniokwadratowy zapewnia estymator  $\hat{s}_p$  - linia *rmseP*. Estymator uzyskany MNW obarczony jest dużym błędem dla  $N=21$  oraz  $N=31$ . Dla  $N>50$  błąd estymatora MNW jest porównywalny z błędem estymatora uzyskanego proponowaną metodą. W całym przedziale zmienności  $21 < N < 131$  błąd estymatora Mallata jest większy od błędu estymatora  $\hat{s}_p$ .



**Rys. 1.** Wartości rmse uzyskane dla generatora rozkładu GED ze współczynnikiem kształtu  $s=0.5$ . Linia kropkowaną z xkami *rmseW* oznaczono wartości uzyskane MNW. Linia przerywana z kółkami *rmseM* odpowiada metodzie Mallata (wzory (10), (11c)). Linia ciągłą z prostokątami *rmseP* zaznaczono wartości błędów uzyskane proponowaną metodą (wzory(10),(11),(12)).

**Źródło:** Opracowanie własne



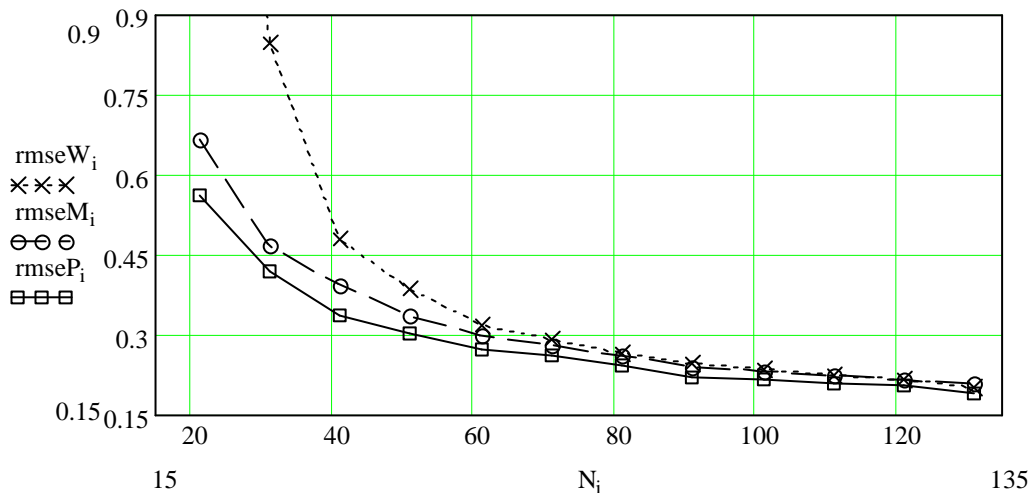
**Rys. 2.** Wartości względnego obciążenia uzyskane dla generatora rozkładu GED ze współczynnikiem kształtu  $s=0.5$ . Linia kropkowaną z xkami *bw* oznaczono wartości uzyskane MNW. Linia przerywana z kółkami *bm* odpowiada metodzie Mallata.. Linia ciągłą z prostokątami *bp* zaznaczono wartości błędów uzyskane proponowaną metodą.

**Źródło:** Opracowanie własne

Na rysunku 2 zamieszczono wartości względnego obciążenia uzyskane dla generatora ze współczynnikiem kształtu  $s=0,5$ . Zrezygnowano z zaznaczenia wartości obciążenia dla MNW, które dla  $N=21$  wyniosło  $bw=1,409$ . Dla  $N>90$  wartości obciążenia estymatora

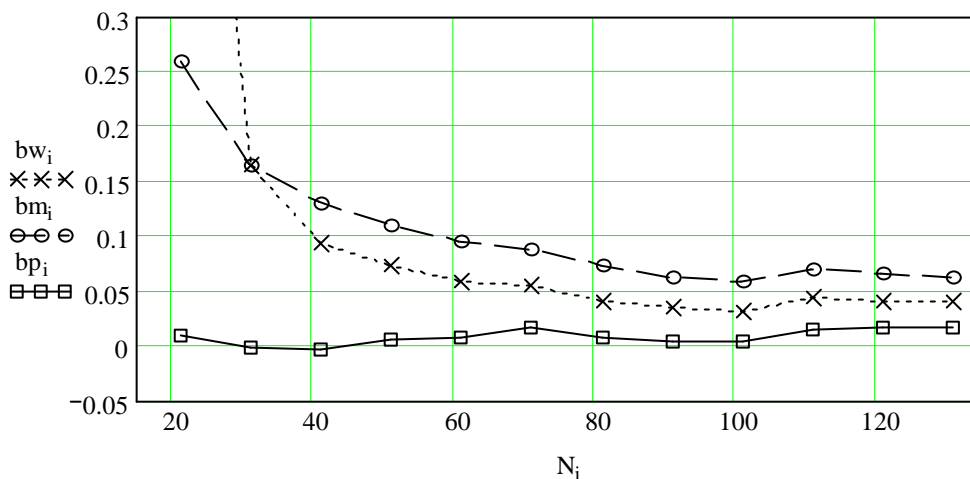
MNW i  $\hat{s}_p$  przyjmują jednakowe wartości. Obciążenie estymatora Mallata jest kilkakrotnie większe niż obciążenie estymatora  $\hat{s}_p$ .

Na rysunku 3 przedstawiono wartości błędów dla generatora ze współczynnikiem kształtu  $s=1$ . Na uwagę zasługuje bardzo duży błąd względny  $rmseW=2,153$  uzyskany dla próbki z  $N=21$  elementami. Z rysunku 3 wynika, że najmniejszy błąd średniokwadratowy zapewnia estymator  $\hat{s}_p$  - linia  $rmseP$ . Estymator uzyskany MNW obarczony jest dużym błędem dla  $N=21$  oraz  $N=31$ . Dla  $N>60$  błąd estymatora MNW jest porównywalny z błędem estymatora uzyskanego metodą Mallata. W całym przedziale zmienności  $21<N<131$  najmniejszy błąd zapewnia estymator  $\hat{s}_p$ .



**Rys. 3.** Wartości  $rmse$  uzyskane dla generatora rozkładu GED ze współczynnikiem kształtu  $s=1$  (rozkład Laplace'a). Zastosowano oznaczenia jak na rysunku 1.

**Źródło:** Opracowanie własne

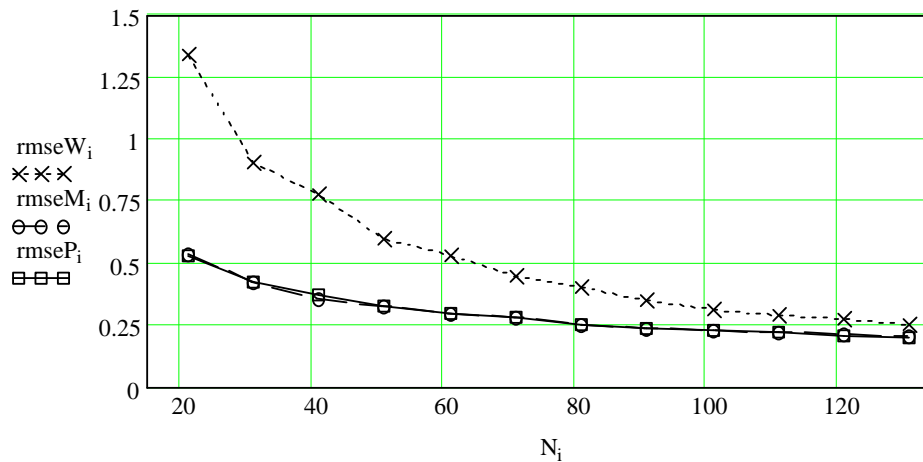


**Rys. 4.** Wartości względnego obciążenia uzyskane dla generatora rozkładu GED ze współczynnikiem kształtu  $s=1$ . Zastosowano oznaczenia jak na rysunku 2.

**Źródło:** Opracowanie własne

Na rysunku 4 zamieszczono wartości względnego obciążenia uzyskane dla generatora ze współczynnikiem kształtu  $s=1$ . Zrezygnowano z zaznaczenia wartości obciążenia dla MNW, które dla  $N=21$  wyniosło  $bw=0,803$ . Dla  $N>30$  wartości obciążenia estymatora

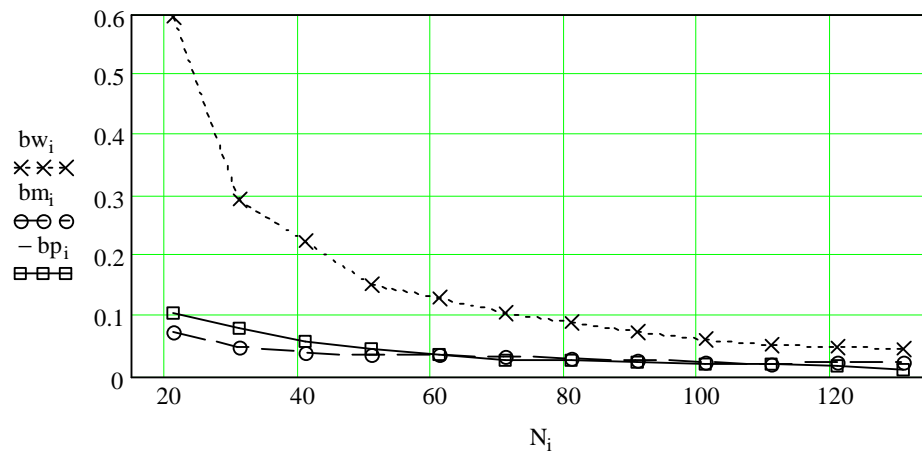
MNW są mniejsze niż dla estymatora Mallata. Do najmniejszych wartości obciążenie względne prowadzi proponowana metoda (bp).



**Rys. 5.** Wartości *rmse* uzyskane dla generatora rozkładu GED ze współczynnikiem kształtu  $s=2$  (rozkład normalny). Zastosowano oznaczenia jak na rysunku 1.

**Źródło:** Opracowanie własne

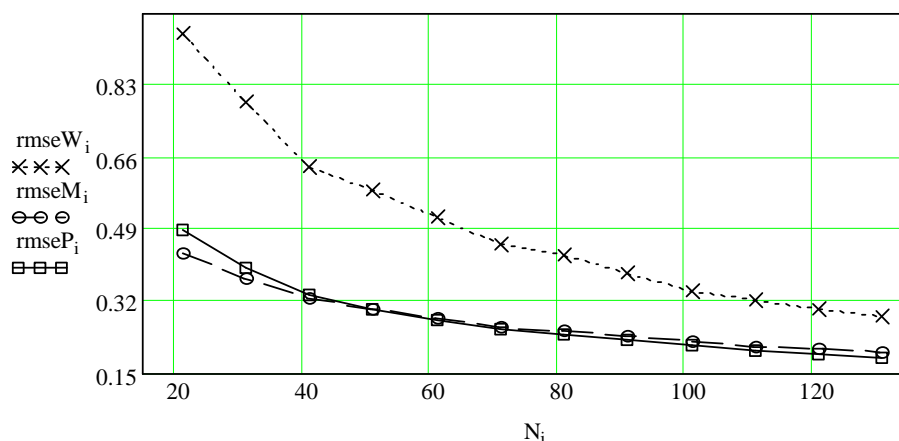
Na rysunku 5 przedstawiono wartości błędu dla generatora ze współczynnikiem kształtu  $s=2$  (rozkład Gaussa). Z rysunku 5 wynika, że metoda Mallata i proponowana metoda prowadzą do jednakowej wartości błędu *rmse*. MNW obarczone jest większym błędem *rmse*. Rysunek 6 ilustruje wartości względnego obciążenia estymatorów dla generatora z  $s=2$ . Dla rysunku 6 obowiązują te same wnioski, co dla rysunku 5.



**Rys. 6.** Wartości względnego obciążenia uzyskane dla generatora rozkładu GED ze współczynnikiem kształtu  $s=2$ . Zastosowano oznaczenia jak na rysunku 2.

**Źródło:** Opracowanie własne

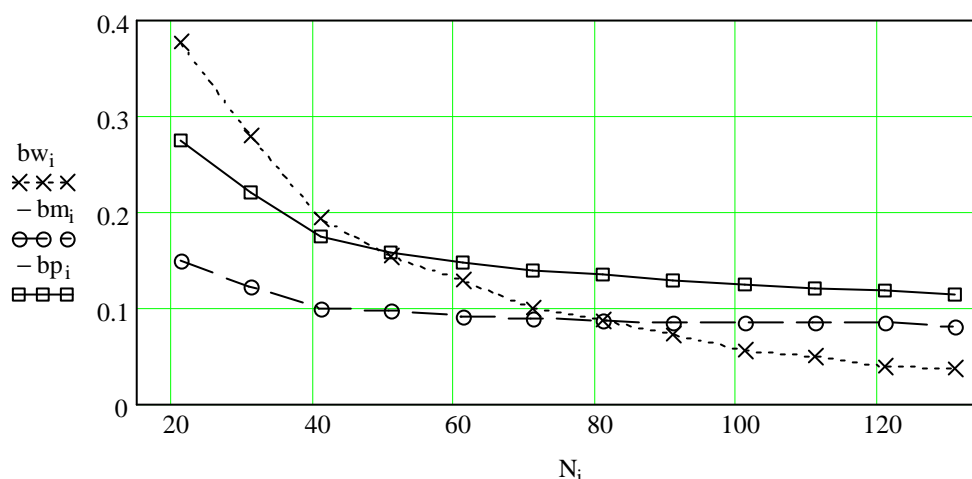
Na rysunku 7 zamieszczono wartości błędu *rmse* uzyskane dla generatora ze współczynnikiem kształtu  $s=3$ . Na uwagę zasługują duże wartości błędu estymatora MNW – dwukrotnie większe niż dla pozostałych metod. Metoda Mallata i proponowana metoda prowadzą do zbliżonych wartości błędu *rmse*.



**Rys. 7.** Wartości rmse uzyskane dla generatora rozkładu GED ze współczynnikiem kształtu  $s=3$ . Zastosowano oznaczenia jak na rysunku 1.

**Źródło:** Opracowanie własne

Rysunki 5 (parametr kształtu  $s=2$ ) i 7 (parametr kształtu  $s=3$ ) wykazują duże podobieństwo: największym błędem jest obarczona metoda MNW a dwie pozostałe metody wykazują zbliżone wartości błędu średniokwadratowego.



**Rys. 8.** Wartości względnego obciążenia uzyskane dla generatora rozkładu GED ze współczynnikiem kształtu  $s=3$ . Zastosowano oznaczenia jak na rysunku 2.

**Źródło:** Opracowanie własne

Na rysunku 8 zaprezentowano wartości względnego obciążenia dla generatora z parametrem  $s=3$ . Dla  $N>80$  najmniejszym błędem charakteryzuje się MNW. Zarówno metoda Mallata, jak i proponowana metoda, prowadzi do ujemnych wartości błędu. W związku z czym, na rysunku 8 przedstawiono wartości  $-bm$  oraz  $-bp$ . Obciążenie estymatora Mallata jest mniejsze, niż obciążenie proponowanego estymatora.

## PODSUMOWANIE

W pracy zaproponowano nową metodę estymacji parametru kształtu uogólnionego rozkładu Gaussa. W celu oceny jakości proponowanego estymatora wykonano symulacje



komputerowe i porównano z właściwościami estymatorów uzyskanych MNW oraz metodą Mallata [5]. Jakość estymatora określano na podstawie wartości względnego błędu średnio kwadratowego *rmse* oraz względnego obciążenia. Eksperyment numeryczny wykonano z wykorzystaniem generatorów liczb losowych dla parametru kształtu  $s = 0,5, s=1,0, s=2,0$  oraz  $s=3,0$ .

Zasadniczym błędem, decydującym o jakości estymatora jest błąd średniokwadratowy (*rmse*). Biorąc to pod uwagę oraz wnioski wynikające z analizy wyników należy stwierdzić, że najmniejszym błędem jest obciążony estymator uzyskany zaproponowaną metodą. Odnosnie dwóch pozostałych metod, to dla  $s=1, s=2$  oraz  $s=3$  mniejsze wartości *rmse* zapewnia metoda Mallata. Natomiast, dla  $s=0,5$  dokładniejsza jest metoda MNW.

Zaproponowana metoda jest szczególnie przydatna dla małej liczebności próbki  $N \leq 100$ . W przypadku dużej liczebności próbki najmniejszym błędem charakteryzuje się estymator uzyskany MNW.

## METHODS OF ESTIMATING GENERALIZED GAUSSIAN DISTRIBUTION PARAMETERS

### *Abstract*

*In this paper a new method of estimating the shape parameter of generalized Gaussian distribution was proposed. Characteristics of the following estimators were compared: the one obtained through the maximum likelihood method, Mallat method [5] and the proposed method. The quality of estimator was evaluated on the basis of the value of the relative mean squared error. Computer simulations were conducted using random number generators for the shape parameter  $s=0.5, s=1.0, s=2.0, \text{ and } s=3.0$*

### BIBLIOGRAFIA

1. Box G.E.P., Tiao G.C.: *A further look at robustness via Bayes theorem*, Biometrika no. 49 (3/4), p. 419-432, 1962
2. Hsieh D.A.: *Testing for nonlinear dependence in daily foreign exchange rate changes*. Journal of Business, 62, p. 339-368, 1989
3. Kokkinakis K., Nandi A.K.: *Exponent parameter estimation for generalized Gaussian probability density functions with application to speech modeling*, Signal Processing 85 p. 1852-1858, 2005
4. Krupiński R., Purczyński J.: *Modeling the distribution of DCT coefficients for JPEG reconstruction*, Signal Processing: Image Communication vol. 22 issue 5, p. 435-441, 2007
5. Mallat S.G.: *A theory of multiresolution signal decomposition: the velvet representation* IEEE Trans. Pattern Anal. Machine Intell. 11 (7) July, p. 674-693, 1989
6. Nelson D.B.: *Conditional heteroskedasticity in asset returns: A new approach*, Econometrica, Vol.59, No.2, p. 347-370. 1991
7. Subbotin M.T.H.: *On the law of frequency of error*, Matematicheski Sbornik 31, p. 296-301, 1923
8. Weron A., Weron R.: *Inżynieria finansowa*, WNT, Warszawa 1998

### *Autorzy:*

**prof. dr hab. inż. Jan PURCZYŃSKI** - Uniwersytet Szczeciński  
**mgr Kamila BEDNARZ** - Uniwersytet Szczeciński