

JAN PURCZYŃSKI

Uniwersytet Szczeciński

**OCENA JAKOŚCI ESTYMATORÓW
PARAMETRÓW ROZKŁADU GED
DLA WYBRANYCH METOD ESTYMACJI**

Streszczenie

W pracy dokonano oceny jakości estymatora parametru kształtu rozkładu GED. Rozpatrzono następujące estymatory: uzyskany metodą największej wiarygodności (MNV), metodą momentów (MM) oraz metodami uproszczonymi. Jakość estymatora określano na podstawie wartości obciążenia oraz względnego błędu średniokwadratowego. Wykonano symulacje komputerowe z wykorzystaniem generatorów liczb losowych o rozkładzie Gaussa oraz Laplace'a.

Słowa kluczowe: estymacja parametrów rozkładu GED, ocena jakości estymatorów

Wprowadzenie

Przy wyznaczaniu ryzyka z akcji dokonuje się analizy zmiennej losowej, jaką jest stopa zwrotu z akcji. Wyznaczając szereg parametrów opisujących tę zmienną, dokonuje się pomiaru ryzyka inwestycji w akcję, szczegółowo opisanego w pracy Tarczyńskiego i Mojsiewicza [2001: 61–84].

Najczęściej zakłada się, że rozkład stóp zwrotu ma rozkład normalny. Założenie to nie zawsze jest słuszne. W cytowanej pracy analizowano tygodniowe stopy zwrotu indeksu WIG za okres kwiecień 1991 – grudzień 2000, uzyskując negatywny wynik odnośnie do rozkładu Gaussa. Wynika stąd potrzeba modelowania empirycznych stóp zwrotu za pomocą innych rozkładów, np. rozkładów o tzw. grubych ogonach, do których zalicza się rozkład GED.

W niniejszej pracy porównano wybrane metody estymacji parametrów rozkładu GED (Generalized Error Distribution) ze względu na jakość uzyskanych estymatorów. Przyjęto następujące oznaczenia dla gęstości rozkładu GED:

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

gdzie $G(z)$ – funkcja gamma Eulera.

Dla $s = 1$, rozkład GED przechodzi w rozkład Laplace'a (dwuwkładniczy):

$$f(x) = \frac{\lambda}{2} \cdot \exp(-\lambda |x - \mu|) \quad (2)$$

W przypadku $s = 2$, otrzymuje się rozkład normalny:

$$f(x) = \frac{\lambda}{\sqrt{\pi}} \cdot \exp(-\lambda^2 (x - \mu)^2) \quad (3)$$

W celu uproszczenia rozważań przyjmuje się, że na podstawie próby zostało wyznaczone oszacowanie parametru μ

$$\hat{\mu} = \frac{1}{N} \cdot \sum_{k=1}^N x_k \quad (4)$$

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

W związku z tym, rozważa się gęstość o postaci:

$$f(x) = \frac{\lambda \cdot s}{2 \cdot \Gamma\left(\frac{1}{s}\right)} \cdot \exp(-|\lambda \cdot x|^s) \quad (5)$$

1. Wybrane metody estymacji 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

$$\frac{\partial \ln(L(\lambda, s))}{\partial \lambda} = 0 ; \quad \frac{\partial \ln(L(\lambda, s))}{\partial s} = 0 \quad (6)$$

otrzymuje się

$$\lambda = \left(\frac{N}{s \sum_{k=1}^N |x_k|^s} \right)^{\frac{1}{s}} \quad (7)$$

oraz

$$gW(s) = s + \Psi\left(\frac{1}{s}\right) + \ln \left(\frac{s}{N} \sum_{k=1}^N |x_k|^s \right) - \frac{\sum_{k=1}^N |x_k|^s \ln |x_k|}{\sum_{k=1}^N |x_k|^s} = 0 \quad (8)$$

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

Z równania (8) wyznacza się parametr kształtu s , a następnie ze wzoru (7) – parametr λ .

1.2. Metoda momentów

W pracy [Purczyński 2002: 71–72] zaproponowano metodę estymacji parametrów rozkładu GED bazującą na momentach absolutnych.

Moment absolutny rzędu m wyraża się wzorem:

$$E_m = \int_{-\infty}^{\infty} |x|^m \cdot f(x) \cdot dx \quad (9)$$

Ze wzorów (5) i (9), otrzymuje się:

$$E_m = \frac{\Gamma\left(\frac{m+1}{s}\right)}{\lambda^m \cdot \Gamma\left(\frac{1}{s}\right)} \quad (10)$$

Estymator momentu E_m ma postać:

$$\hat{E}_m = \frac{1}{N} \cdot \sum_{k=1}^N |x_k|^m \quad (11)$$

Przyjmując dwie różne wartości m_1 i m_2 we wzorze (10) i rozwiązując powstały układ równań, uzyskuje się [Purczyński 2002: 72]:

$$\frac{\Gamma\left(\frac{m_1+1}{s}\right)^{\frac{1}{m_1}}}{\Gamma\left(\frac{m_2+1}{s}\right)^{\frac{1}{m_2}}} \cdot \Gamma\left(\frac{1}{s}\right)^{\frac{1}{m_2} - \frac{1}{m_1}} = \frac{(E_{m_1})^{\frac{1}{m_1}}}{(E_{m_2})^{\frac{1}{m_2}}} \quad (12)$$

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

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

Rozwiązując równanie (12), wyznacza się parametr kształtu s , a kolejno ze wzoru (13), parametr λ .

W cytowanej pracy zwraca się uwagę, że przyjmując we wzorach (9–13) wartość rzędu momentu m równą parametrowi kształtu s , uzyskuje się wyniki identyczne jak dla MNW.

1.3. Metoda uproszczona estymacji parametrów rozkładu GED

Metoda bazuje na zależnościach zamieszczonych w pracy Krupińskiego i Purczyńskiego [2006] i została szczegółowo omówiona w pracy [Purczyński 2011], w której dokonano weryfikacji przydatności tej metody na przykładzie modelowania stóp zwrotu indeksu giełdowego WIG20 oraz SP500.

W pracy [Purczyński 2011] zaproponowano poniższe wzory służące do estymacji wartości parametru s :

$$\hat{s}_1 = \left[\frac{0,2217}{\ln \left(\frac{\sqrt{\hat{E}_2}}{\hat{E}_1} \right) - 0,125} \right]^{0,879} \quad \text{gdzie } \hat{E}_1 = \frac{1}{N} \cdot \sum_{k=1}^N |x_k|; \quad \hat{E}_2 = \frac{1}{N} \cdot \sum_{k=1}^N |x_k|^2$$

$$\hat{s}_2 = \left[\frac{0,0747}{\ln \left(\frac{\sqrt{\hat{E}_1}}{\hat{E}_{0,5}} \right) - 0,0461} \right]^{0,935} \quad \text{gdzie } \hat{E}_{0,5} = \frac{1}{N} \cdot \sum_{k=1}^N |x_k|^{0,5} \quad (14)$$

Natomiast oszacowanie parametru λ wyznacza się z zależności, która powstaje ze wzoru (13) dla $m=1$:

$$\hat{\lambda} = \frac{\Gamma \left(\frac{2}{\hat{s}} \right)}{\hat{E}_1 \cdot \Gamma \left(\frac{1}{\hat{s}} \right)} \quad (15)$$

Na podstawie wzoru (14) uzyskuje się dwie wartości oszacowania parametru s : \hat{s}_1 oraz \hat{s}_2 .

2. Wyniki symulacji komputerowych

2.1. Opis eksperymentu numerycznego

W celu oceny jakości poszczególnych estymatorów wykonano eksperyment numeryczny z wykorzystaniem generatorów liczb losowych o rozkładzie Gaussa oraz Laplace'a. Symulacje komputerowe polegały na wykonaniu $K=1000$ powtórzeń i wyznaczeniu na tej podstawie błędu estymacji. Jako pierwszy wyznaczany był błąd średniokwadratowy MSE (Mean-Squared Error):

$$MSE = \frac{1}{K} \sum_{k=1}^K (\hat{s}_k - s_d)^2 \quad (16)$$

gdzie:

\hat{s}_k – oszacowanie wartości parametru s w k -tej symulacji;

s_d – wartość dokładna parametru kształtu s .

W przypadku rozkładu normalnego $s_d = 2$, natomiast dla rozkładu Laplace'a $s_d = 1$.

Błąd średniokwadratowy jest o tyle istotny, że ujmuje łącznie błąd obciążenia estymatora, jak również jego wariancję [Krzyśko 1997: 17]:

$$MSE = V(\hat{s}) + B(\hat{s})^2 \quad (17)$$

gdzie:

$V(\hat{s})$ – wariancja estymatora

$B(\hat{s})$ – obciążenie estymatora.

Kolejno wyznaczano wartość błędu RMSE (Root Mean-Squared Error):

$$RMSE = \sqrt{MSE} \quad (18)$$

Jako ostatni wyznaczano błąd RRMSE (Relative Root Mean-Squared Error):

$$RRMSE = \sqrt{\frac{1}{K} \sum_{k=1}^K \left(\frac{\hat{s}_k - s_d}{s_d} \right)^2} \quad (19)$$

gdzie oznaczenia jak we wzorze (16).

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

Przy rozwiązaniu równań (8) i (12) stosowano metodę bisekcji (połowienia).

Obliczenia wykonano dla zmieniającej się długości próbki (liczby obserwacji) $N=31, 41, 51, \dots, 121$.

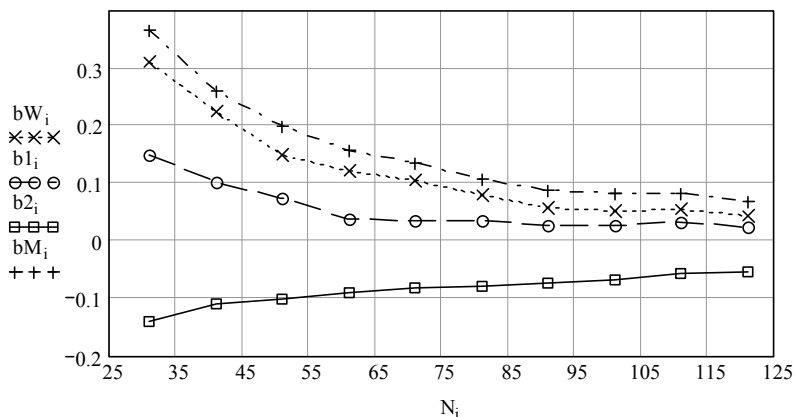
2.2. Wyniki symulacji komputerowych dla rozkładu normalnego

Na rysunku 1 przedstawiono wartości obciążenia względnego b estymatora parametru s :

$$b = \frac{E(\hat{s}) - s_d}{s_d} ; \text{gdzie: } E(\hat{s}) = \frac{1}{K} \sum_{k=1}^K \hat{s}_k \quad (20)$$

uzyskane dla generatora liczb losowych o rozkładzie Gaussa ($s=2$).

Zastosowano następujące oznaczenia dla obciążenia względnego (wzór (20)) poszczególnych estymatorów: $b1$ – wzór (14) (estymator \hat{s}_1); $b2$ – wzór (14) (estymator \hat{s}_2); bW – wzór (8) (estymator MNW); bM – wzór (12) (estymator MM).



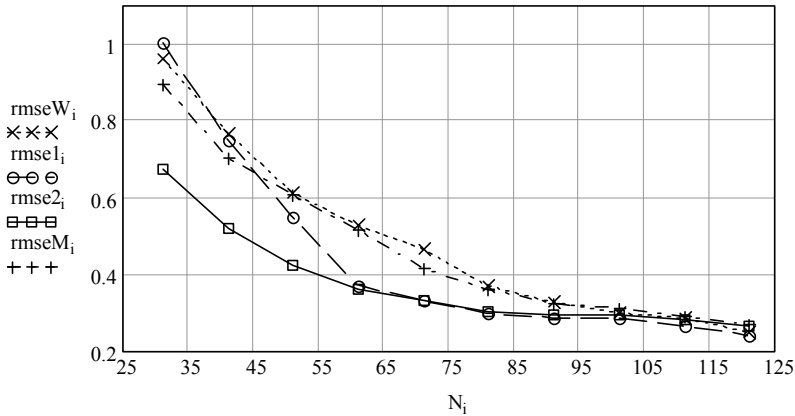
Zastosowano następujące oznaczenia:
 linia kropkowana z xkami bW – obciążenie estymatora $s\hat{W}$
 linia przerywana z kółkami $b1$ – obciążenie estymatora \hat{s}_1
 linia ciągła z prostokątami $b2$ – obciążenie estymatora \hat{s}_2
 linia kropkowo-przerywana bM – obciążeni estymatora $s\hat{M}$

Rysunek 1. Wartości obciążenia względnego b estymatora parametru s

Źródło: opracowanie własne.

Z rysunku 1 wynika, że najmniejszym obciążeniem jest obarczony estymator \hat{s}_1 , nieco większym obciążeniem charakteryzuje się estymator \hat{s}_2 . Kolejnym estymatorem ze względu na wielkość obciążenia jest estymator $s\hat{W}$ (MNW). Największą wartością obciążenia odznacza się estymator metody momentów $s\hat{M}$.

Na rysunku 2 zamieszczono wartości błędu średniokwadratowego $rmse$ estymatora parametru kształtu s uzyskanego dla poszczególnych metod. Z rysunku 2 wynika, że najmniejszym błędem średniokwadratowym jest obarczony estymator \hat{s}_2 . Począwszy od $N=61$, nieco mniejszym błędem charakteryzuje się estymator \hat{s}_1 . Estymator uzyskany MNW (wzór (8)) ma błąd $rmseW$ zbliżony wartością do błędu $rmseM$ estymatora $s\hat{M}$ (metoda momentów).

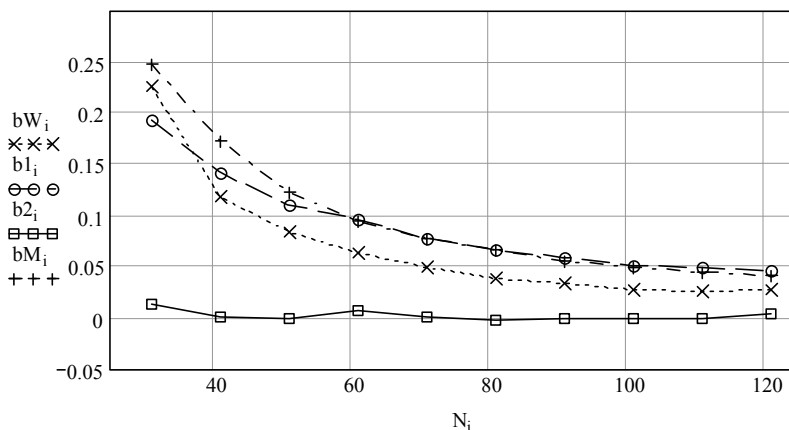


Zastosowano następujące oznaczenia:
 linia kropkowana z xkami $rmseW$ – błąd estymatora $\hat{s}W$
 linia przerywana z kółkami $rmse1$ – błąd estymatora \hat{s}_1
 linia ciągła z prostokątami $rmse2$ – błąd estymatora \hat{s}_2
 linia kropkowo-przerywana $rmseM$ – błąd estymatora $s\hat{M}$

Rysunek 2. Wartości względnego błędu średniokwadratowego $rmse$

Źródło: opracowanie własne.

2.3. Wyniki symulacji komputerowych dla rozkładu Laplace'a



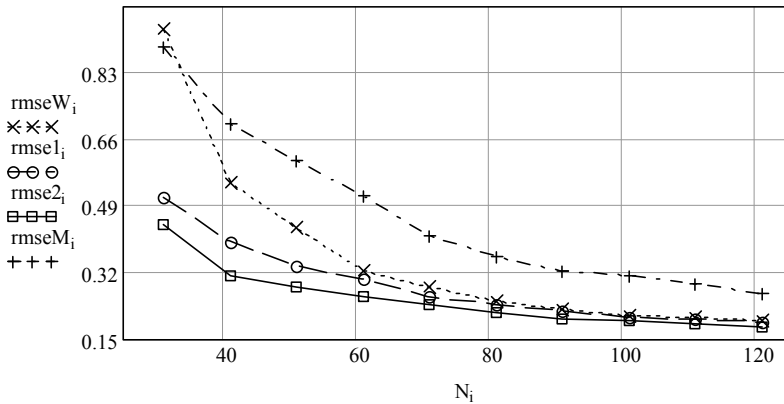
Zastosowano następujące oznaczenia:
 linia kropkowana z xkami bW – obciążenie estymatora $\hat{s}W$
 linia przerywana z kółkami $b1$ – obciążenie estymatora \hat{s}_1
 linia ciągła z prostokątami $b2$ – obciążenie estymatora \hat{s}_2
 linia kropkowo-przerywana bM – obciążeni estymatora $s\hat{M}$

Rysunek 3. Wartości obciążenia względnego b estymatora parametru s

Źródło: opracowanie własne.

Na rysunku 3 zamieszczono wartość obciążenia poszczególnych estymatorów uzyskanych dla generatora liczb losowych o rozkładzie Laplace'a ($s=1$). Zdecydowanie najmniejszym obciążeniem odznacza się estymator \hat{s}_2 – linia $b2$ oscyluje wokół zera.

Kolejnym estymatorem ze względu na wartość obciążenia jest $\hat{s}W$ (MNW). Zbliżone wartości, nieco większe niż obciążenie bW , wykazuje estymator \hat{s}_1 oraz $s\hat{M}$.



Zastosowano następujące oznaczenia:
 linia kropkowana z xkami $rmseW$ – obciążenie estymatora $\hat{s}W$
 linia przerywana z kółkami $rmse1$ – obciążenie estymatora \hat{s}_1
 linia ciągła z prostokątami $rmse2$ – obciążenie estymatora \hat{s}_2
 linia kropkowo-przerywana $rmseM$ – obciążeni estymatora $s\hat{M}$

Rysunek 4. Wartości względnego błędu średniokwadratowego $rmse$

Źródło: opracowanie własne.

Na rysunku 4 przedstawiono wartości błędu $rmse$ uzyskane dla generatora o rozkładzie Laplace'a. Z rysunku 4 wynika, że najmniejszym błędem względnym $rmse$ jest obciążony estymator \hat{s}_2 . Nieco mu ustępuje estymator \hat{s}_1 . Największym błędem $rmse$ charakteryzuje się estymator $s\hat{M}$. Błąd $rmseW$ estymatora MNW przyjmuje dla małej liczebności próby ($N=31$) duże wartości. Wraz ze wzrostem liczby elementów N wartości tego błędu zdążają asymptotycznie do błędów estymatora \hat{s}_1 i \hat{s}_2 .

Uwzględniając obciążenie estymatora oraz błąd $rmse$, w przypadku generatora liczb losowych o rozkładzie Laplace'a należy polecić stosowanie wzoru (14) w odniesieniu do estymatora \hat{s}_2 .

Wnioski końcowe

W pracy rozpatrzono jakość estymatora parametru kształtu rozkładu GED uzyskanego metodą największej wiarygodności (MNW), metodą momentów (MM) oraz metodami uproszczonymi. Jakość estymatora określano na podstawie wartości obciążenia oraz względnego błędu średniokwadratowego wyznaczonego w wyniku zastosowania generatorów liczb losowych o rozkładzie Gaussa oraz Laplace'a.

Podsumowując wyniki obliczeń zamieszczonych w podrozdziałach 2.2 i 2.3, należy stwierdzić, że najgorszą jakość wykazuje estymator uzyskany metodą momentów (MM). Najlepszą jakość uzyskano dla estymatorów wywodzących się z metod uproszczonych.

Pośrednie wartości obciążenia i błędu średniokwadratowego wykazuje estymator uzyskany metodą największej wiarygodności (MNW).

W trakcie eksperymentu numerycznego wykonano obliczenia dla generatora liczb losowych o rozkładzie GED z parametrem kształtu $s=0,5$. Jednakże nie zamieszczono wyników, ze względu na bardzo duże wartości błędu średniokwadratowego uzyskanego dla wszystkich rozpatrzonych estymatorów.

Łącząc wyniki symulacji uzyskane dla rozkładu Gaussa (parametr kształtu $s=2$) oraz rozkładu Laplace'a (parametr kształtu $s=1$), można stwierdzić przydatność estymatora \hat{s}_2 (wzór (14)) do estymacji wartości parametru kształtu s dla rozkładu GED.

Literatura

- Krupiński R., Purczyński J. [2006], *Approximated fast estimator for the shape parameter of generalized Gaussian distribution*, „Signal Processing”, Vol. 86, No. 4, 205–211.
- Krzyśko M. [1997], *Statystyka matematyczna*, cz. 2, Wydawnictwo UAM, Poznań.
- Purczyński J. [2002], *Estymacja parametrów rozkładu GED*, w: *Rynek kapitałowy – Skuteczne inwestowanie*, cz. 1, Wydawnictwo Naukowe Uniwersytetu Szczecińskiego, Szczecin.
- Purczyński J. [2011], *Uproszczona metoda estymacji parametrów rozkładu GED*, konferencja „Mikroekonometria w teorii i praktyce”, Międzyzdroje 4–6 września 2011.
- Tarczyński W., Mojsiewicz M. [2001], *Zarządzanie ryzykiem*, PWE, Warszawa.

**QUALITATIVE EVALUATION
OF GED DISTRIBUTION PARAMETER ESTIMATORS
FOR SELECTED METHODS OF ESTIMATION**

Summary

In this paper the quality of an estimator of GED distribution shape parameter has been evaluated. The following estimators were considered: the one obtained through the maximum likelihood method, the method of moments, and simplified methods. The quality of estimator was evaluated on the basis of the value of the bias and mean squared error. Computer simulations were conducted using random number generators with the Gaussian and Laplace distribution.

Keywords: estimation of GED distribution parameters, evaluation of estimator quality

Translated by Ewa Stefanowska