PRACE NAUKOWE
Uniwersytetu Ekonomicznego we Wrocławiu

RESEARCH PAPERS
of Wrocław University of Economics

Nr 427

Taksonomia 27
Klasyfikacja i analiza danych – teoria i zastosowania

Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2016

Redaktor Wydawnictwa: Agnieszka Flasińska

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Beata Mazur

Projekt okładki: Beata Dębska

Tytuł dofinansowany ze środków Narodowego Banku Polskiego
oraz ze środków Sekcji Klasyfikacji i Analizy Danych PTS

Informacje o naborze artykułów i zasadach recenzowania
znajdują się na stronach internetowych
www.pracenaukowe.ue.wroc.pl
www.wydawnictwo.ue.wroc.pl

ISSN 1899-3192 (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)
e-ISSN 2392-0041
ISSN 1505-9332 (Taksonomia)

Wersja pierwotna: publikacja drukowana

Zamówienia na opublikowane prace należy składać na adres:
Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
ul. Komandorska 118/120, 53-345 Wrocław
tel./fax 71 36 80 602; e-mail:econbook@ue.wroc.pl
www.ksiegarnia.ue.wroc.pl

Druk i oprawa: TOTEM

Spis treści

Wstęp

W dniach 14–16 września 2015 r. w Hotelu Novotel Gdańsk Marina w Gdańsku odbyła się XXIV Konferencja Naukowa Sekcji Klasyfikacji i Analizy Danych PTS (XXIX Konferencja Taksonomiczna) „Klasyfikacja i analiza danych – teoria i zastosowania", zorganizowana przez Sekcję Klasyfikacji i Analizy Danych Polskiego Towarzystwa Statystycznego oraz Katedrę Statystyki Wydziału Zarządzania Uniwersytetu Gdańskiego.

W trakcie dwóch sesji plenarnych oraz 13 sesji równoległych wygłoszono 58 referatów poświęconych aspektom teoretycznym i aplikacyjnym zagadnienia klasyfikacji i analizy danych. Odbyła się również sesja plakatowa, na której zaprezentowano 14 plakatów.

Teksty 24 recenzowanych artykułów naukowych stanowią zawartość prezentowanej publikacji z serii Taksonomia nr 27. Teksty 25 recenzowanych artykułów naukowych znajdują się w Taksonomii nr 26.

Krzysztof Jajuga, Marek Walesiak

PRACE NAUKOWE UNIWERSYTETU EKONOMICZNEGO WE WROCŁAWIU
RESEARCH PAPERS OF WROCŁAW UNIVERSITY OF ECONOMICS nr 427 • 2016

Taksonomia 27 ISSN 1899-3192
Klasyfikacja i analiza danych – teoria i zastosowania e-ISSN 2392-0041

Michał Stachura, Barbara Wodecka
Uniwersytet Jana Kochanowskiego w Kielcach
e-mails: {michał.stachura; barbara.wodecka}@ujk.edu.pl

WYBRANE ASPEKTY I ZASTOSOWANIA MODELI ZDARZEŃ EKSTREMALNYCH

SELECTED FACETS AND APPLICATION OF MODELS OF EXTREMAL EVENTS

DOI: 10.15611/pn.2016.427.21

Streszczenie: W niniejszym artykule nakreślono, w jaki sposób można dzięki użyciu wartości rekordowych dokonać estymacji wartość indeksu ekstremalnego, a następnie na jego podstawie innych parametrów modelowanego rozkładu (np. indeksu stabilności dla rozkładów α-stabilnych dla $\alpha < 2$). Wobec powyższego, celem opracowania jest porównanie modelowania zdarzeń ekstremalnych opartego na wartościach rekordowych k-tego rzędu z podejściem bazującym na statystykach pozycyjnych. Czynione jest to z perspektywy arbitralnie wybranych trzech estymatorów: Berreda, Hilla i Pickandsa poprzez przeprowadzenie badań symulacyjnych. Dodatkowo, w artykule zaprezentowano ilustrację proponowanej metodologii dla przykładowych danych empirycznych, zaczerpniętych ze skandynawskiego rynku energii elektrycznej (Nord Pool Spot), dotyczących notowanych co godzinę cen rynkowych na obszarze Finlandii (z segmentu *regulating power market*).

Słowa kluczowe: indeks ekstremalny, rozkład Pareta, rozkład stabilny, wartości rekordowe k-tego rzędu.

Summary: The paper describes how to estimate extreme value index with the use of records values, and next on this base other parameters of the model distribution (e.g. stability index of α-stabile distribution with $\alpha < 2$). Therefore, the main goal of this article is to compare two approaches to modeling of extreme values—one based on k-th record values, and the other based on order statistics. This idea is realised from a perspective of three arbitrarily chosen estimators: Berred's, Hill's, and Pickands'. Furthermore, an empirical illustration of the proposed methodology is presented.

Keywords: extremal value index, k-th record values, Pareto distribution, stable distribution.

1. Wstęp

Rozważane w opracowaniu zagadnienia znajdują się na pograniczu dwu teorii: teorii wartości rekordowych i teorii wartości ekstremalnych.

Intuicje leżące u podstaw pojęcia **wartości rekordowych** wywodzą się z obserwacji z życia codziennego. Bardzo często spotykamy się z określeniami typu „rekordowe temperatury" czy że ktoś odniósł rekordowy sukces. Coraz więcej danych jest zbieranych w ten sposób, że notowane są tylko wartości największe bądź najmniejsze. Podobnie nie tylko sporządzane są rejestry np. rekordów w zawodach lekkoatletycznych, ale też zbierane są niektóre dane z takich obszarów, jak finanse, ubezpieczenia, meteorologia, hydrologia.

W celu zdefiniowania wprowadzanych w dalszym ciągu pojęć przyjęte są następujące założenia i oznaczenia. Niech $X_1, X_2, X_3, \ldots$ będzie nieskończonym ciągiem niezależnych zmiennych losowych o tym samym rozkładzie zadanym niezdegenerowaną dystrybuantą F. Ponadto niech $X_{1:n} \le X_{2:n}, \le \ldots \le X_{n:n}$ oznaczają statystyki pozycyjne z próby $X_1, X_2, \ldots, X_n$ wybranej z rozważanego ciągu.

Wówczas ciąg czasów rekordowych $\{T_n\}$ i ciąg wartości rekordowych $\{R_n\}$ można zdefiniować następująco (zob. [Arnold i in. 1998])

$$T_1 = 1, \qquad T_n = \min\{j : X_j > X_{T_{n-1}}\} \text{ dla } n \ge 2, \tag{1}$$

$$R_n = X_{T_n}, \qquad n \in \mathbb{N}_+. \tag{2}$$

Istnieje wiele sytuacji, w których oprócz wartości największych (najmniejszych), w kręgu zainteresowań są wartości drugie bądź trzecie w kolejności od największej do najmniejszej (lub od najmniejszej do największej). Wobec tego w pełni naturalnym stało się rozszerzenie teorii wartości rekordowych na wartości rekordowe k-tego rzędu – ozn. $R_n^{(k)}$ – definiowane w następujący sposób poprzez czasy rekordowych k-tego rzędu – ozn. $T_n^{(k)}$ (zob. [Dziubdziela, Kopociński 1976])

$$T_1^{(k)} = k, \;\; T_n^{(k)} = \min\{j : j > T_{n-1}^{(k)}, \; X_j > X_{T_{n-1}^{(k)}-k+1 \,:\, T_{n-1}^{(k)}}\} \text{ dla } n \ge 2, \tag{3}$$

$$R_n^{(k)} = X_{T_n^{(k)}-k+1 \,:\, T_n^{(k)}}, \tag{4}$$

gdzie $k \ge 1$ jest ustaloną liczbą naturalną.

Przypomnijmy, że główne twierdzenie **teorii wartości ekstremalnych** brzmi następująco. Jeżeli istnieją stałe $a_n > 0$, b_n dla $n \in \mathbb{N}_+$ oraz pewna niezdegenerowana dystrybuanta G takie, że dla wszystkich $x \in \mathbb{R}$ zachodzi

$$\lim_{n \to \infty} \mathbb{P}\left(\frac{X_{n:n} - b_n}{a_n} \le x\right) = G(x), \tag{5}$$

to z dokładnością do liniowej zmiany argumentu dystrybuanta G jest postaci

$$G(x) = G_\gamma(x) = \begin{cases} \exp(-(1+\gamma x)^{-1/\gamma}) & 1+\gamma x > 0 \quad \gamma \neq 0 \\ \exp(-e^{-x}) & x \in \mathbb{R} \quad \gamma = 0 \end{cases} \tag{6}$$

z parametrem $\gamma \in \mathbb{R}$, który nazywany jest indeksem ekstremalnym (zob. [Beirlant i in. 2004] lub [De Haan, Ferreira 2006]).

Niezmiernie istotną kwestią jest właściwa estymacja indeksu ekstremalnego. Najpopularniejszymi estymatorami są estymatory Hilla i Pickandsa, zdefiniowane odpowiednio wzorami

$$\hat{\gamma}_H^k = \frac{1}{k}\sum_{i=0}^{k-1} \ln X_{n-i:n} - \ln X_{n-k:n}, \qquad \hat{\gamma}_P^k = \log_2 \frac{X_{n-k:n} - X_{n-2k:n}}{X_{n-2k:n} - X_{n-4k:n}}, \tag{7}$$

$$\hat{\gamma}_P^k = \log_2 \frac{X_{n-k:n} - X_{n-2k:n}}{X_{n-2k:n} - X_{n-4k:n}}. \tag{8}$$

Estymatory te konstruowane są na podstawie statystyk pozycyjnych z próby. Możliwe jest jednak odrzucenie podejścia opartego na statystykach pozycyjnych na rzecz wartości rekordowych. Taki estymator zaproponował Berred (zob. [Berred 1995]), nadając mu postać

$$\hat{\gamma}_B = \ln \frac{R_{N(k,n)}^{(k)} - R_{N(k,n)-k}^{(k)}}{R_{N(k,n)-k}^{(k)} - R_{N(k,n)-2k}^{(k)}}, \tag{9}$$

gdzie $N(k, n)$ oznacza liczbę wartości rekordowych k-tego rzędu w skończonej n-elementowej próbie.

Wobec powyższego celem opracowania jest porównanie modelowania zdarzeń ekstremalnych w oparciu o wartości rekordowe k-tego rzędu z podejściem bazującym na statystykach pozycyjnych. Czynione jest to z perspektywy arbitralnie wybranych trzech estymatorów: Berreda, Hilla i Pickandsa.[1]

Warto nadmienić, że estymacja indeksu ekstremalnego umożliwia diagnostykę asymptotyki i ocenę grubości ogona rozkładu danego przez dystrybuantę F, co pozwala m.in. na właściwe szacowanie prawdopodobieństwa zdarzeń ekstremalnych. Co więcej, w przypadku niektórych rozkładów indeks ekstremalny γ przekłada się bezpośrednio na wybrane parametry dystrybuanty F. Na przykład dla rozkładów α-stabilnych między indeksem stabilności a indeksem ekstremalnym zachodzi następująca zależność $\alpha = \gamma^{-1}$ dla $\alpha \in (0, 2)$, która nie obowiązuje w przypadku szczególnym – rozkładu normalnego, gdy $\alpha = 2$, a $\gamma = 0$ (zob. [Samorodnitsky Taqqu 1994]). Z kolei dla uogólnionego rozkładu Pareta parametr kształtu ξ jest wprost indeksem ekstremalnym, tzn. $\xi = \gamma$ (zob. [Beirlant i in. 2004]). Wobec tego w dalszej części opracowania będzie używany symbol γ zamiast ξ.

[1] Wybór padł akurat na te estymatory, ponieważ: pionierski estymator Berreda wciąż pozostaje jedynym bazującym na k-tych wartościach rekordowych, estymator Pickandsa jest analogonem i pierwowzorem estymatora Berreda, a estymator Hilla jest wyjątkowo powszechnie stosowany.

2. Badania symulacyjne

Aby porównać jakość i dokładność oszacowań uzyskiwanych na podstawie estymatorów opartych na statystykach pozycyjnych z estymatorami opartymi na wartościach rekordowych k-tego rzędu przeprowadzono badania symulacyjne[2] dla rozkładów: uogólnionego Pareta z parametrami $\sigma = 1$, $\gamma \in \{0{,}05;\ 0{,}1;\ 0{,}2;\ 0{,}5;\ 1;\ 2;\ 5;\ 10;\ 20\}$ oraz symetrycznego (tzn. dla $\beta = 0$, $\mu = 0$) α-stabilnego z parametrami $\sigma = 1$, $\alpha \in \{0{,}1;\ 0{,}2;\ 0{,}3;\ 0{,}4;\ \ldots;\ 1{,}8;\ 1{,}9\}$.[3]

Dla każdej wartości parametru, oddzielnie dla γ oraz α, wygenerowano niezależną próbę pseudolosową liczebności $n = 8000$.[4] Na jej podstawie wyznaczono ciągi wartości rekordowych oraz statystyk pozycyjnych dla wszystkich $k \in \{1, 2, \ldots, 650\}$. Dzięki nim wyznaczono wartości estymatorów $\hat{\gamma}_B^k$, $\hat{\gamma}_H^k$, $\hat{\gamma}_P^k$, $\hat{\alpha}_B^k$, $\hat{\alpha}_H^k$, $\hat{\alpha}_P^k$.

Tabela 1. Oszacowania parametru γ dla rozkładu Pareta

γ	0,05	0,1	0,2	0,5	1	2	5	10	20
$\hat{\gamma}_B$	0,054	0,104	0,202	0,503	1,004	2,003	4,992	10,024	20,027
$\hat{\gamma}_H$	0,274	0,303	0,363	0,580	1,017	1,999	4,978	10,009	19,955
$\hat{\gamma}_P$	0,052	0,103	0,201	0,500	1,000	2,004	4,991	10,011	20,005

Źródło: opracowanie własne.

Tabela 2. Oszacowania parametru α dla rozkładu α-stabilnego[5]

α	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1
$\hat{\alpha}_B$	0,092	0,184	0,277	0,374	0,473	0,578	0,691	0,807	0,945	1,100
$\hat{\alpha}_H$	0,099	0,196	0,294	0,391	0,491	0,589	0,686	0,789	0,892	0,996
$\hat{\alpha}_P$	0,094	0,187	0,281	0,380	0,479	0,585	0,689	0,802	0,918	1,042

α	1,1	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9	
$\hat{\alpha}_B$	1,309	1,618	2,032	2,219	2,039	1,612	–3,161	–3,466	–3,352	
$\hat{\alpha}_H$	1,109	1,230	1,359	1,510	1,695	1,935	2,286	2,754	3,481	
$\hat{\alpha}_P$	1,199	1,401	1,633	2,068	2,819	2,718	1,644	–4,581	–4,210	

Źródło: opracowanie własne.

[2] W literaturze znaleźć można opracowania traktujące o analogicznych, symulacyjnych porównaniach jakości estymatorów indeksu ekstremalnego (zob. np. [Mojsiewicz, Guzowska, Purczyński 2003]), jednak autorzy nie natknęli się na opracowania uwzględniające porównanie estymatorów bazujących na statystykach pozycyjnych i estymatorów bazujących na wartościach rekordowych.

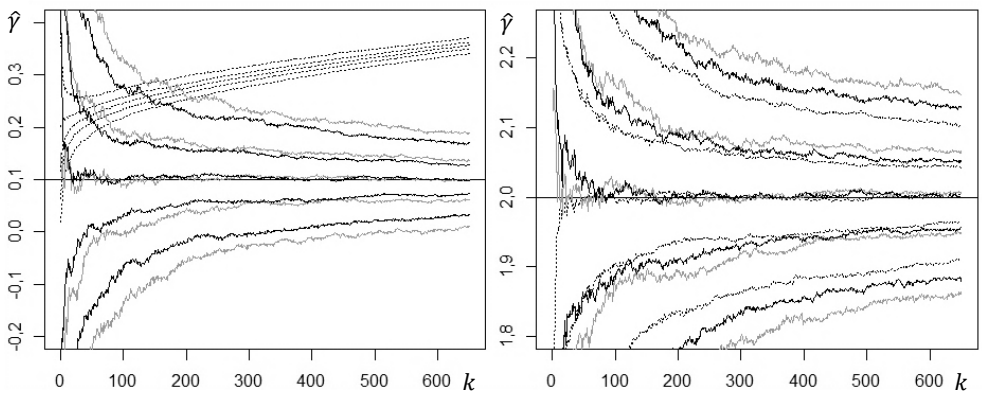
[3] Wszystkie obliczenia i zamieszczone w pracy wykresy zostały wykonane w programie R (zob. [R Core Team 2012]) z użyciem pakietów `gPdtest`, `stabledist`, `Records` (zob. [Estrada, Villasenor Alva 2012; Wuertz, Maechler 2013; Chrapek 2012]).

[4] Skorzystano z domyślnych ustawień generatorów liczb pseudolosowych.

[5] Narastające wraz ze wzrostem α obciążenie estymatorów (w tym uzyskane wartości ujemne) są zdaniem autorów najpewniej skutkiem nadmienionej we wstępie nieciągłości w zależności funkcyjnej między parametrami α oraz γ dla $\alpha = 2$.

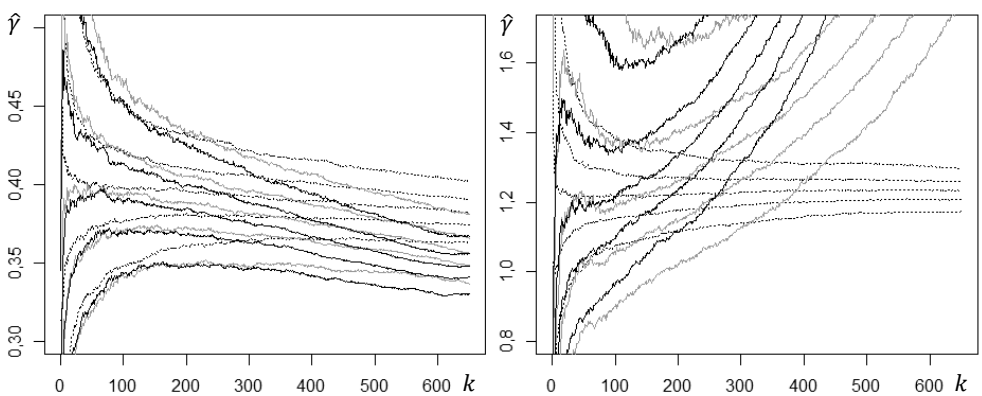
Procedura ta była replikowana J =1000-krotnie. Następnie na podstawie każdego kompletu J replikacji wyznaczono ciągi (względem k) kwantyli rzędów 0,1; 0,3; 0,5; 0,7; 0,9. Jako ostateczne oszacowania parametrów γ oraz α przyjęto mediany z wyznaczonych uprzednio ciągów median (wyniki zebrane zostały w tab. 1 i 2).

Na rysunkach 1 i 2 przedstawiono ciągi linii kwantylowych (czarne linie ciągłe – estymator Berreda, czarne linie kropkowane – Hilla, szare linie ciągłe – Pickandsa) dla wybranych teoretycznych wartości parametrów γ oraz α (poziome linie ciągłe).



Rys. 1. Linie kwantylowe estymatorów Berreda, Hilla, Pickandsa dla rozkładu Pareta z parametrem $\gamma = 0{,}1$ (lewy wykres) i z parametrem $\gamma = 2$ (prawy wykres)

Źródło: opracowanie własne.



Rys. 2. Linie kwantylowe estymatorów oparte na estymatorach Berreda, Hilla, Pickandsa dla rozkładu α-stabilnego z parametrem $\alpha = 0{,}4$ (lewy wykres) i z parametrem $\alpha = 1{,}2$ (prawy wykres)

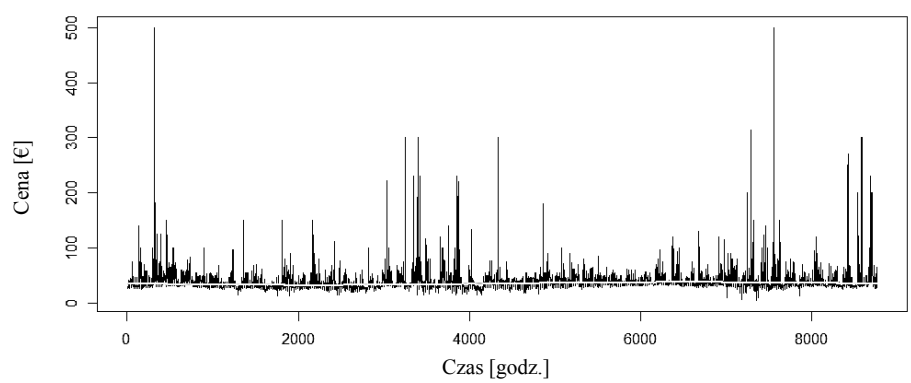
Źródło: opracowanie własne.

Z przeprowadzonych badań symulacyjnych można wyciągnąć następujące wnioski dotyczące zarówno obciążenia, jak i rozproszenia uzyskiwanych oszacowań. Otóż, dla **rozkładu Pareta** wartości estymatora Hilla są zawyżane dla $\gamma < 0{,}5$ i tym bardziej jest to widoczne, im parametr γ jest bliższy 0.[6] Natomiast dla $\gamma \geq 0{,}5$ wartości są zbliżone do szacowanej wartości parametru. Estymatory Berreda i Pickandsa dają podobne rezultaty. Natomiast dla **rozkładu α-stabilnego** wartości estymatorów bazujących na estymatorach Berreda, Hilla i Pickandsa są niedoszacowane dla $\alpha < 1$, z kolei dla $\alpha > 1$ wartości estymatorów są przeszacowane.

Analizując linie kwantylowe estymatorów, można zauważyć, że – w przypadku rozkładów gruboogonowych ($0 < \alpha < 1$, $\gamma > 0$), dla których estymatory zdają się nieobciążone – wartości estymatorów opartych na k-tych wartościach rekordowych charakteryzują się mniejszym rozproszeniem niż te bazujące na statystykach pozycyjnych. Ponadto wraz ze wzrostem k maleje rozproszenie, a rośnie obciążenie rozważanych estymatorów[7].

3. Przykład empiryczny

W badaniach empirycznych wykorzystano ceny rynkowe (w euro) energii elektrycznej w Finlandii (z segmentu *regulating power market*) za okres od godz. 00:00 1 stycznia 2014 r. do 24:00 31 grudnia 2014 r., rejestrowane w odstępach godzinnych (dane ze Skandynawskiego rynku energii, zob. [Nord Pool Spot]). Badany szereg empiryczny zawiera $n = 8761$ obserwacji.



Rys. 3. Szereg empiryczny i wygładzenie funkcją sinus

Źródło: opracowanie własne.

[6] Wynika to najpewniej z tego, że estymator Hilla jest zdefiniowany dla przypadku, gdy $\gamma > 0$.

[7] Stąd konieczność znajdowania właściwego zakresu k, dla którego oszacowania można uznawać za wiarygodne. Z tego powodu w dalszym ciągu (część 3) odwołano się do zakresu typowanego zgodnie ze wskazówkami z [Chrapek, Stachura, Wodecka 2012].

Specyfika badanej zmiennej sprawia, że w szeregu empirycznym występują trzy okresy zmienności sezonowej: dobowy, tygodniowy i roczny. Sezonowość roczna jest traktowana jako deterministyczna i modelowana za pomocą odpowiednio skalowanej funkcji sinus (zob. [Weron 2005]). Względem pozostałej składowej szeregu danych zaproponowano opisy w postaci sezonowych modeli autoregresyjnych ze średnią ruchomą postaci (a) SARMA$(p,q)\times(0,1)_{s_1}$, (b) SARMA$(p,q)\times(0,1)_{s_2}$, (c) SARMA$(p,q)\times(0,1)_{s_1}\times(0,1)_{s_2}$ dla opóźnień sezonowych $s_1 = 24$ i $s_2 = 168$ (zob. [Nazarko, Chrabołowska, Rybaczuk 2004]).

Aby wybrać ostateczną postać modelu, dla wszystkich wartości par $(p, q) \in \{(1, 1), (1, 2), (2, 1), (2, 2), (2, 3), (3, 2), (3, 3)\}$ i wszystkich trzech postaci (a)–(c) wyestymowano stosowne modele, a ponadto wyliczono przebiegi funkcji ACF i PACF, wartości logarytmicznej funkcji wiarogodności (dalej: LogLik), wartości krytyczne testu Ljunga-Boxa. Okazało się, że wszystkie wyestymowane modele postaci (a) i (c) charakteryzują się występowaniem silnej autokorelacji, przez co zostały odrzucone jako opisy modelowe. Z kolei wśród modeli postaci (b) brakiem autokorelacji wykazał się jedynie model z parametrami $(p, q) = (3, 3)$ (zob. tabela 3), dlatego właśnie ten model wybrano jako model analizowanego szeregu cen[8].

Wobec powyższego dla reszt tego modelu wyznaczono ciągi estymatorów Hilla, Pickandsa i Berreda, a następnie wyznaczono mediany ($\hat{\gamma}_B^m$, $\hat{\gamma}_H^m$, $\hat{\gamma}_P^m$) i uśredniania ($\hat{\gamma}_B^s$, $\hat{\gamma}_H^s$, $\hat{\gamma}_P^s$) po stosowanie dobranym zakresie k ($80 \le k \le 250$, $80 \approx n^{0,48}$, $250 \approx n^{0,61}$ – zob. przyp. 7. Analogiczne ciągi estymatorów oraz ich mediany i uśrednienia wyliczono dla pozostałych modeli postaci (b), mimo odrzucenia tych modeli jako właściwego opisu szeregu cen. Wszystkie wyniki dotyczące modeli postaci (b) zamieszczono tab. 3. W kolumnach oznaczonych ACF, PACF i Ljung-Box symbol „+" oznacza wykrycie występowania autokorelacji, „–" – jej brak, „+/–"zaś – sytuację dyskusyjną, zależną od poziomu istotności.

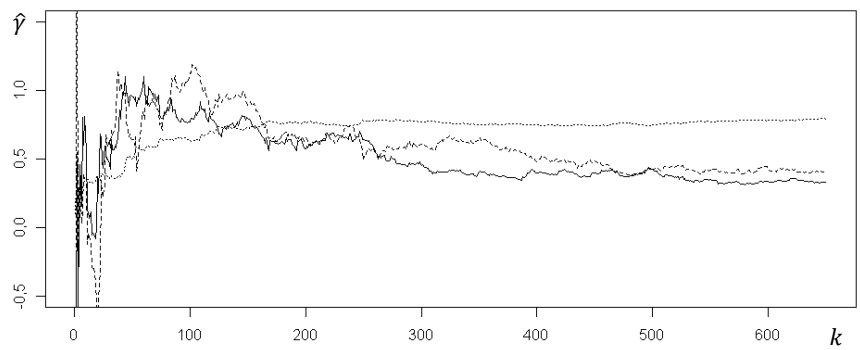
Tabela 3. Wyniki badań empirycznych

p	q	ACF	PACF	LogLik	Ljung-Box	$\hat{\gamma}_B^m$	$\hat{\gamma}_H^m$	$\hat{\gamma}_P^m$	$\hat{\gamma}_B^s$	$\hat{\gamma}_H^s$	$\hat{\gamma}_P^s$
1	1	+/–	+/–	–37987,1	+	0,681	0,743	0,743	0,707	0,727	0,811
1	2	+/–	+/–	–37983,6	+	0,664	0,748	0,706	0,691	0,733	0,780
2	1	+/–	+/–	–37986,9	+	0,679	0,744	0,738	0,710	0,728	0,806
2	2	+/–	+/–	–37981,5	+	0,698	0,744	0,722	0,718	0,724	0,819
2	3	+/–	+/–	–37916,2	+	0,668	0,751	0,691	0,688	0,728	0,770
3	2	+/–	+/–	–37973,3	+	0,719	0,740	0,775	0,719	0,725	0,814
3	3	–	–	–37909,7	–	0,675	0,746	0,703	0,694	0,725	0,774
Współczynniki zmienności						0,0260	0,0046	0,0371	0,0168	0,0039	0,0242

Źródło: opracowanie własne.

[8] Należy zwrócić uwagę, że podejmowanie decyzji o dobroci modeli przy nieklasycznych założeniach o typie rozkładu jest wysoce dyskusyjne – nie dość, że rozkład statystyki Ljunga-Boxa może być zupełnie inny, to teoretyczne odpowiedniki funkcji ACF i PACF mogą w ogóle nie istnieć.

Dla modelu postaci (b) z parametrami $(p, q) = (3, 3)$ wartości wszystkich estymatorów są bardzo zbliżone do siebie, co więcej w grupie median i w grupie uśrednień najniższe wartości oszacowań dają estymatory Berreda. Podobne prawidłowości zaobserwować można także w przypadku pozostałych modeli postaci (b) (zob. wiersze ostatnich sześciu kolumn tab. 3).



Rys. 4. Wartości estymatorów Berreda, Hilla, Pickandsa wyznaczone na podstawie reszt modelu SARMA $(3, 3) \times (0, 1)_{168}$

Źródło: opracowanie własne.

Do interesujących wniosków prowadzi analiza wartości oszacowań uzyskiwanych za pomocą ustalonego estymatora względem wszystkich rozważanych modeli postaci (b). Są one bowiem niezmiernie zbliżone do siebie[9].

Przeprowadzona analiza danych empirycznych nie pozwala jednoznacznie wnioskować, który z estymatorów dał lepsze oszacowania w przypadku badanego szeregu, tym bardziej że wskazania wszystkich estymatorów są bardzo zbliżone.

5. Podsumowanie

Przeprowadzone badania symulacyjne pozwalają stwierdzić, że estymatory oparte na estymatorze Berreda i Pickandsa dają bardzo zbliżone wartości, natomiast tylko w około połowie badanych przypadków estymatory oparte na estymatorze Hilla prowadzą do podobnych oszacowań jak pozostałe estymatory. Warto podkreślić, że pomimo jego dobrych własności teoretycznych, oszacowania estymatora Hilla dla γ bliskich 0 są symptomatycznie znacznie zawyżone.

[9] Celowe rozważanie estymatorów uzyskiwanych na podstawie odrzuconych modeli wskazuje, że wątpliwości wyrażone w przypisie 8 i tak zdają się nie mieć znaczącego wpływu na wyniki estymacji grubości ogona rozkładu reszt. Ocena „bliskości" uzyskiwanych oszacowań poczyniona została na podstawie współczynników zmienności wyznaczonych dla kolumn tab. 3.

Z kolei analizując przykład empiryczny, można stwierdzić, że wartości oszacowań zaprezentowanych w pracy estymatorów parametru γ są bardzo zbliżone.

Pomimo tego, że pozornie nie można wskazać, który z zaprezentowanych w pracy estymatorów daje lepsze i dokładniejsze oszacowania należy zwrócić szczególną uwagę na następujący fakt.

Estymatory oparte na wartościach rekordowych *k*-tego rzędu mogą być użyte bez względu na to, czy znane są wszystkie wartości z próby, czy też tylko rekordy. Natomiast estymatory oparte na statystykach pozycyjnych nie są już tak uniwersalnym narzędziem, gdyż nie można wyznaczać odpowiednich wartości statystyk pozycyjnych na podstawie samych jedynie wartości rekordowych. Fakt ten jednoznacznie wskazuje na szersze pole zastosowań estymatora Berreda niż estymatorów opartych na statystykach pozycyjnych.

Dodać można jeszcze, że równolegle prowadzone przez autorów obiecujące badania pozwalają wskazać na kolejną przewagę estymatora Berreda. Otóż, skoro w przeciwieństwie do statystyk pozycyjnych, wartości rekordowe są zależne od kolejności obserwacji w próbie, to próbę – o ile jest ona niezależna – można dowolnie wiele razy permutować, a następnie dla każdej takiej permutacji wyznaczyć wartości estymatorów i je uśredniać. Zabieg ten skutkuje znaczącym spadkiem dyspersji uzyskiwanych oszacowań.

Literatura

Arnold B.C., Balakrishnan N., Nagaraja H.N., 1998, *Records*, Wiley, New York.

Beirlant J., Goegebeur Y., Segers J., Teugels J., 2004, *Statistics of Extremes. Theory and Applications*, Wiley Series in Probability and Statistics, Wiley & Sons, Chichester.

Berred M., 1995, *K-record values and the extreme-value index*, Journal of Statistical Planning and Inference, vol. 45, no. 1/2, s. 49–63.

Chrapek M., 2012, *Records: Record Values and Record Times*, R package version 1.0. http://CRAN.R-project.org/package=Records.

Chrapek M., Stachura M., Wodecka B., 2012, *Estymacja indeksu ekstremalnego w oparciu o k-te wartości rekordowe – sugestia poprawy jakości estymacji*, [w:] A.S. Bartczak, D. Iskra (red.), *Metody matematyczne, ekonometryczne i komputerowe w finansach i ubezpieczeniach 2010*, Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach, Katowice, s. 9–28.

De Haan L., Ferreira A., 2006, *Extreme Value Theory. An Introduction*, Springer, New York.

Dziubdziela W., Kopociński B., 1976, *Limiting properties of the k-th record values*, Zastosowania Matematyki, nr 15, s. 187–190.

Estrada E.G., Villasenor Alva J.A., 2012, *gPdtest: Bootstrap goodness-of-fit test for the generalized Pareto distribution*, R package version 0.4, http://CRAN.R-project.org/package=gPdtest.

Mojsiewicz M., Guzowska M., Purczyński J., 2003, *Ocena jakości estymatorów grubości ogona rozkładu w przypadku próby o małej liczebności*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu, nr 991, s. 412–422.

Nazarko J., Chrabołowska J., Rybaczuk M., 2004, *Zastosowanie wielosezonowego modelu ARIMA w prognozowaniu obciążeń mocą elektryczną*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu, nr 1022, Taksonomia 11, s. 173–182.

Nord Pool Spot, *Skandynawski rynek energii*, http://www.nordpoolspot.com/.

R Core Team, 2012, *R: A language and environment for statistical computing*, The R Foundation for Statistical Computing, Vienna, http://www.R-project.org/.

Samorodnitsky G., Taqqu M.S., 1994, *Stable Non-Gaussian Random Processes. Stochastic Models with Infinite Variance*, Chapman & Hall, New York–London.

Weron R., 2005, *Heavy tails and electricity prices*, Research Report HSC/05/2, Hugo Steinhaus Center, Wrocław University of Technology, Wrocław.

Wuertz D., Maechler M., 2013, *stabledist: Stable Distribution Functions*, R package version 0.6-6, http://CRAN.R-project.org/package=stabledist.