

Michał Stachura

Akademia Świętokrzyska w Kielcach

PROBLEMY W MODELOWANIU PRZY UŻYCIU ROZKŁADÓW α -STABILNYCH

1. Wstęp

Klasa rozkładów α -stabilnych obejmuje rozkłady o różnej skośności, skupieniu, położeniu i, co bardzo ważne, różnej grubości ogonów. Dzięki temu rozkłady α -stabilne znajdują szerokie zastosowania w opisie danych, w których wartości znacznie odbiegające od większości obserwacji występują stosunkowo często. Przede wszystkim więc rozkłady α -stabilne mogą posłużyć do modelowania szeregów finansowych czy ubezpieczeniowych.

Za modelowaniem przy użyciu tego typu rozkładów przemawiają także inne argumenty. Po pierwsze, suma α -stabilnych zmiennych losowych z tym samym indeksem stabilności ma rozkład α -stabilny. Można więc dokonywać agregacji kilku szeregów, co pozwala np. na przechodzenie od analizy pojedynczych walorów do analizy portfeli czy też od analizy spółek do analizy sektorów itd. Po drugie, ogólne twierdzenia graniczne, dotyczące odpowiednio centrowanych i skalowanych zmiennych losowych, wskazują, że rozkłady α -stabilne jako rozkłady graniczne mogą aproksymować inne rozkłady.

Wymienione powody mogą stać się nazbyt zachęcające i łatwo w tej sytuacji ulec wrażeniu, że α -stabilność szeregów empirycznych jest zjawiskiem powszechnym. W związku z tym niniejsze opracowanie ma na celu pokazanie, że do modelowania za pomocą tego typu rozkładów trzeba podchodzić z pewną dozą ostrożności. Zwrócono zatem uwagę na problemy i pułapki dotyczące szeroko rozumianych modelowania i estymacji, które szczególnie silnie ujawniają się w przypadku rozkładów gruboogonowych.

Po pierwsze, omówiono zagadnienie tzw. wartości odstających i związaną z tym kwestię niejednorodności procesu stochastycznego, generującego badany szereg

czasowy (por. [1]). W tym kontekście dla rozkładów α -stabilnych odróżnienie niejednorodnej próby statystycznej z wartościami odstającymi od próby jednorodnej z wartościami ekstremalnymi jest wręcz niemożliwe. Ponadto uznanie któregośkolwiek z przypadków za właściwy, ze względu na dużą wrażliwość estymatorów, może poskutkować obciążeniem oszacowań parametrów.

Po drugie, pokazano problem niespełnienia założeń modelowych przyjmowanych w estymacji. Zwrócono tu uwagę na komplikacje dotyczące niemożności wnioskowania co do charakteru procesu generującego dane.

2. Rozkłady α -stabilne i ich estymacja

Przyjmuje się, że zmienna losowa X_θ ma rozkład α -stabilny z parametrem θ , gdy jej funkcja charakterystyczna dana jest wzorem:

$$\phi_\theta(t) = \mathbf{E} \exp(itX_\theta) = \exp\left[i\mu t - |\sigma t|^\alpha (1 + i\beta \cdot \omega(\alpha, t) \cdot \operatorname{sgn} t)\right].$$

Jednowymiarowe składowe parametru $\theta = (\alpha, \beta, \mu, \sigma) \in \langle 0, 2 \rangle \times \langle -1, 1 \rangle \times \mathbb{R} \times (0, +\infty)$ to: α – indeks stabilności, β – miara skośności, μ – parametr położenia, σ – miara rozproszenia. Z kolei $\omega(\alpha, t) = \tan(\pi\alpha/2)$ dla $\alpha \neq 1$ i $\omega(\alpha, t) = (2/\pi) \ln|t|$ dla $\alpha = 1$.

Rozkład α -stabilny dany jest w sposób jawny poprzez gęstość tylko w czterech przypadkach: 1) $\alpha = 2$ – rozkład normalny, 2) $\alpha = 1$, $\beta = \mu = 0$ – rozkład Cauchy'ego, 3) $\alpha = \frac{1}{2}$, $\beta = 1$, $\mu = 0$ – rozkład Gaussowski odwrotny, 4) $\alpha = 0$ – rozkład jednopunktowy. Poza tymi przypadkami gęstości nie da się wyrazić wzorem analitycznym i właśnie to sprawia, że estymacja parametru θ jest niebagatelnym zadaniem.

W literaturze znaleźć można wiele ujęć zagadnienia estymacji. Wśród ważniejszych rodzajów metod estymacji wymienić należy (por. [3]):

- (i) numeryczne metody największej wiarygodności¹,
- (ii) metody kwantyli próbkowych (m.in. metoda McCullocha),
- (iii) metody dopasowania próbkowej funkcji charakterystycznej do teoretycznej,
- (iv) metody regresji dla próbkowej funkcji charakterystycznej (m.in. metoda Koutrouvelisa).

Do ostatniej grupy należy metoda wykorzystana w dalszym ciągu opracowania. Jest ona bezpośrednią modyfikacją metody Koutrouvelisa, a opiera się na następujących przekształceniach funkcji charakterystycznej:

$$\psi_\theta(t) = \ln(-\operatorname{Re}(\ln(\phi_\theta(t)))) = \alpha \ln \sigma + \alpha \ln|t|,$$

$$\xi_\theta(t) = \operatorname{Im}(\ln(\phi_\theta(t))) = \mu t + |\sigma t|^\alpha \beta \cdot \omega(\alpha, t) \cdot \operatorname{sgn} t.$$

¹ Numerycznie uzyskuje się gęstość z funkcji charakterystycznej.

Zastosowanie tej metody sprowadza się kolejno do: 1) wyznaczenia analogicznych przekształceń $\hat{\psi}_\theta$, $\hat{\xi}_\theta$ empirycznej funkcji charakterystycznej $\hat{\phi}_\theta$ i przyjęcia ich za estymatory funkcji ψ_θ i ξ_θ ; 2) wyznaczenia liniowej regresji $\hat{\psi}_\theta(t)$ względem $\ln t$ i wyliczenia na tej podstawie $\hat{\alpha}$ i $\hat{\sigma}$; 3) wyznaczenia $\hat{\mu}$ i $\hat{\beta}$ z nieliniowej regresji $\hat{\xi}_\theta(t)$ względem t z ustalonymi, uprzednio przyjętymi $\hat{\alpha}$ i $\hat{\sigma}$. Metoda pozwala ponadto estymować parametr samopodobieństwa H (tzw. wykładnik Hursta) w sytuacji, gdy analizowany szereg empiryczny jest realizacją α -stabilnego procesu H -samopodobnego (por. [2]). Estymacji tej dokonuje się następująco.

Najpierw ustala się kilka liczb naturalnych (najwygodniej potęg 2, czyli liczb ze zbioru: $\{2^k : k \in \{0, \dots, K\}\}$), które stanowią skalę agregacji i dla każdego k dokonuje się sumowania danych po 2^k obserwacji. Następnie dla każdego k estymuje się α_k i σ_k . Wtedy, jeśli szereg empiryczny jest α -stabilny i H -samopodobny, to wszystkie α_k są sobie równe, σ_k zaś są związane relacją: $\sigma_k = (2^k)^H \cdot \sigma_0$, w której znajduje się estymator $\hat{H}$ (dzięki regresji: $\log_2 \sigma_k = H \cdot k + \log_2 \sigma_0$ zmiennej k).

3. Wpływ pojedynczych rekordów na estymatory

Pierwsza z omawianych kwestii dotyczy zasadności wnioskowania statystycznego na podstawie posiadanych danych empirycznych. Problem ten jest niezmiernie istotny, ponieważ w przypadku danych ekonomicznych dysponujemy najczęściej tylko i wyłącznie pojedynczą próbką obserwacji analizowanego zjawiska. Nie ma mowy o możliwości powtórnego przeprowadzenia „eksperymentu”, tak jak się to dzieje np. w naukach przyrodniczych. Ponadto szeregi empiryczne często są dość krótkie. Z kolei w sytuacjach, gdy długość szeregu jest znaczna i przekracza np. 1000 rekordów, mogą wystąpić uzasadnione przesłanki, że szereg nie jest jednolity w tym sensie, że nie można go uznać *en bloc* za realizację jednego procesu stochastycznego.

Aby wskazać na niebagatelny wpływ wyżej nakreślonych zagrożeń na jakość estymacji rozkładów α -stabilnych, dokonajmy prostej analizy wrażliwości estymatorów indeksów stabilności i rozproszenia dla symulowanych szeregów długości 1000. Do symulacji wykorzystana jest następująca procedura (por. [3; 5]).

Na wstępie generowane są dwie niezależne zmienne Z i Y odpowiednio o rozkładzie jednostajnym na przedziale $(-\frac{\pi}{2}, \frac{\pi}{2})$ oraz wykładniczym o wartości oczekiwanej 1^2 . Niech określona będzie także stała $d = \beta \tan(\pi\alpha/2)$. Na tej podstawie dla $\alpha \neq 1$ definiuje się:

² Np. $Z = \pi(U - \frac{1}{2})$, $Y = -\ln V$, gdzie U i V mają rozkład jednostajny na przedziale $(0, 1)$.

$$X = -d + \frac{\sin \alpha Z + d \cos \alpha Z}{(\cos Z)^{1/\alpha} Y^{(1-\alpha)/\alpha}} (\cos(1-\alpha)Z + d \sin(1-\alpha)Z)^{(1-\alpha)/\alpha},$$

a dla $\alpha = 1$:

$$X = \left(1 + \frac{2}{\pi} \beta Z\right) \tan Z - \beta \ln \left(\frac{2Y \cos Z}{\pi + 2\beta Z}\right).$$

Wówczas dla $\alpha \neq 1$ zmienna $\sigma X + \mu$, a dla $\alpha = 1$ zmienna $\sigma X + \mu + \frac{2}{\pi} \beta \sigma \ln \sigma$ ma rozkład α -stabilny z parametrem $\theta = (\alpha, \beta, \mu, \sigma)$.

Za pomocą opisanej procedury w środowisku SAS 9.1 wygenerowano 40 serii po 1000 realizacji zmiennych o parametrach: $\alpha = 1,8$; $\sigma = 10$; $\mu = 0$; $\beta = 1$. Zgodnie ze zmodyfikowaną metodą Koutrouvelisa dokonano estymacji parametrów α i σ dla wygenerowanych szeregów. Średnie 40 estymatorów parametru α i 40 estymatorów parametru σ podano w tab. 1, w której dodatkowo umieszczone są próbkowe odchylenia standardowe i próbkowe współczynniki zmienności tych średnich.

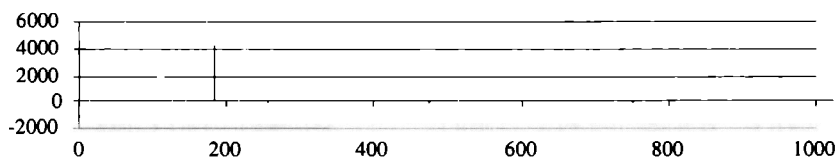
Tabela 1. Estymatory parametrów symulowanych rozkładów

	α	σ
Średnia	1,79845	10,0385
Odchylenie standardowe	0,04774	0,32647
Współczynnik zmienności	0,02655	0,03252

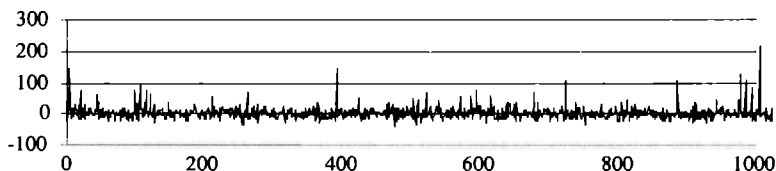
Źródło: opracowanie własne.

Wartości w tabeli wskazują, że już przy niewielkiej liczbie powtórzeń (równej 40) uzyskuje się bardzo dobre oszacowania parametrów, co zresztą nie powinno dziwić, ponieważ estymatory Koutrouvelisa, obok estymatorów McCullocha, uznawane są w literaturze za najlepsze (por. [3]). Niemniej analiza danych rzeczywistych, jak już nadmieniono, nie może odwoływać do wielokrotnego powtarzania eksperymentu. W celu pokazania pułapki, polegającej na możliwości wyciągnięcia błędnych wniosków na podstawie pojedynczej próby oraz poczynionych założeń co do jej jednorodności lub niejednorodności, dokonano prezentacji bardzo prostego przykładu. Mianowicie, opisaną procedurę generowania kontynuowano do momentu uzyskania pierwszego szeregu, w którym pojawił się choć jeden rekord przekraczający umowną wartość 4000. Do dalszych rozważań wykorzystano właśnie ten szereg, a jego modyfikacja powstała poprzez zastąpienie rekordu ekstremalnego średnią arytmetyczną wartości sąsiednich, przy czym szereg niezmodyfikowany reprezentuje przypadek założenia jednorodności, a zmodyfikowany – niejednorodności próby statystycznej.

Kształtowanie się obu szeregów prezentują rys 1 i 2. Na rysunku 1 dominująca wartość ekstremalna jest tak duża, że wręcz uniemożliwia dostrzeżenie pozostałych wartości. Można je jednak zaobserwować na rys. 2.



Rys. 1. Wygenerowana trajektoria z dominującą wartością ekstremalną
Źródło: opracowanie własne.



Rys. 2. Wygenerowana trajektoria z usuniętą wartością ekstremalną
Źródło: opracowanie własne.

Zarówno dla wybranego szeregu z pojedynczą wartością ekstremalną, jak i dla jego modyfikacji dokonano estymacji parametrów α i σ . W pierwszym przypadku uzyskano $\hat{\alpha} = 1,6649$ i $\hat{\sigma} = 9,7874$, w drugim zaś: $\hat{\alpha} = 1,9066$ i $\hat{\sigma} = 2,2612$. Są to wartości znacznie różniące się parami między sobą, a także odbiegające od uzyskanych wcześniej średnich wartości estymatorów odpowiednich parametrów. Widać więc, że zarówno pojawienie się tylko jednej silnie dominującej wartości, jak i jej brak powoduje diametralne zmiany w wartościach estymatorów. Co więcej, analiza pojedynczych trajektorii prowadzi do odmiennych wniosków niż równoczesna analiza wielu trajektorii.

4. Badania empiryczne

W celu pokazania drugiej pułapki związanej z modelowaniem przy użyciu rozkładów α -stabilnych zaprezentowane zostaną przykładowe badania, dotyczące wolumenów obrotu akcjami kilku wybranych banków notowanych na GPW w Warszawie.

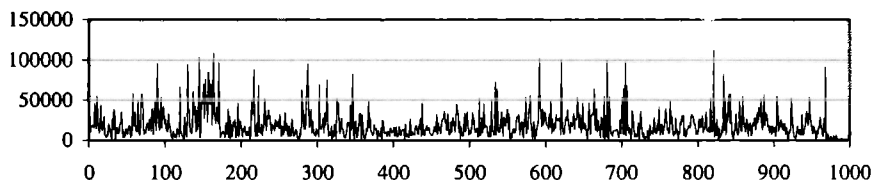
Przesłanki przemawiające za wyborem wolumenów obrotu, jako być może zmiennych o rozkładach α -stabilnych, są następujące: 1) wolumen obrotu przyjmuje tylko wartości nieujemne, co wskazuje na całkowitą skośność rozkładu, 2) wolumen obrotu przyjmuje często bardzo zróżnicowane wartości, a wartości ekstremalne nie są rzadkością, co z kolei sugeruje występowanie grubych ogonów³.

Analizowane dane dotyczą lat 2002-2006 i obejmują po 1024 rekordy⁴ dla każdego z banków. Ostatnim dniem obserwacji jest 31 stycznia 2006 r., daty zaś po-

³ Dodatkowo za wolumenem obrotów przemawia podobieństwo tych szeregów do szeregów rozważanych w [2] dotyczących ilości przesyłanych danych w sieciach komputerowych.

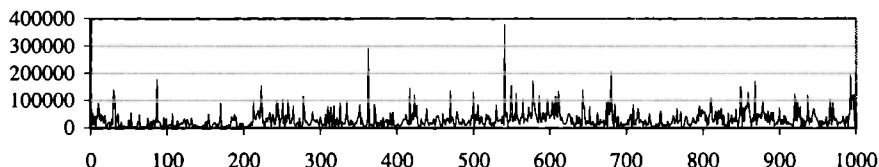
⁴ Wartość 1024 wybrana jest dla wygody, gdyż jako potęga 2 dzieli się przez 16.

czątkowe są różne ze względu na to, że w niektóre dni nie wszystkie banki były notowane. Kształtowanie się danych przedstawiają wykresy na rys. 3-9.



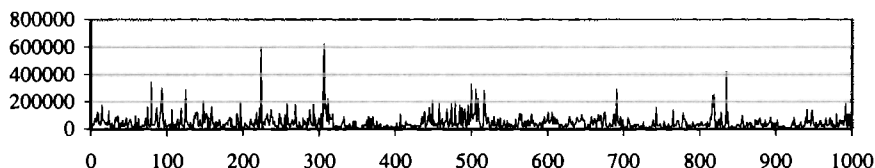
Rys. 3. Wolumen obrotu, BPHPBK

Źródło: opracowanie własne.



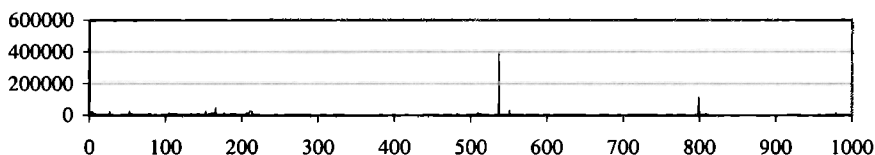
Rys. 4. Wolumen obrotu, BRE

Źródło: opracowanie własne.



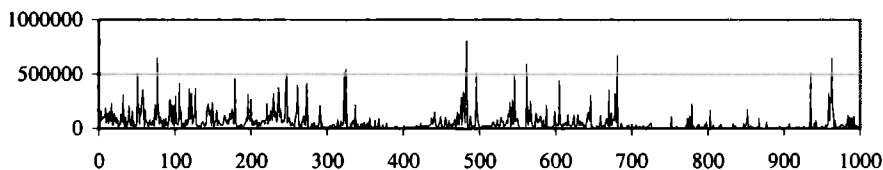
Rys. 5. Wolumen obrotu, BZWBK

Źródło: opracowanie własne.



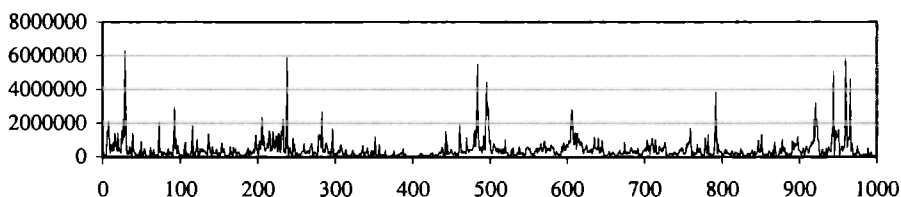
Rys. 6. Wolumen obrotu, INGBSK

Źródło: opracowanie własne.

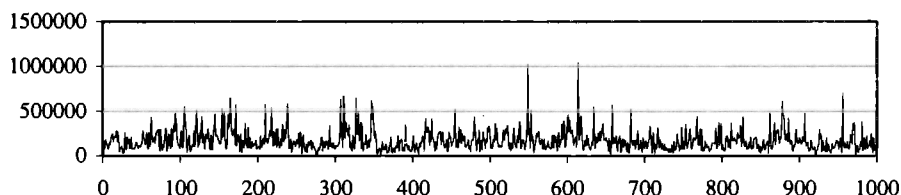


Rys. 7. Wolumen obrotu, KREDYT

Źródło: opracowanie własne.



Rys. 8. Wolumen obrotu, MILLENNIUM
Źródło: opracowanie własne.



Rys. 9. Wolumen obrotu, PEKAO
Źródło: opracowanie własne.

Także wzrokowa analiza zaprezentowanych wykresów sugeruje możliwość wykorzystania rozkładów α -stabilnych. By stwierdzić, czy tak rzeczywiście jest, dla każdego z szeregów empirycznych dokonujemy kolejno: 1) estymacji parametrów α i σ , 2) estymacji wykładnika Hursta H , 3) próby oszacowania μ i β .

Zastosowanie zmodyfikowanej metody Koutrouvelisa i uprzednio opisanej regresji, służącej estymacji wykładnika H dla skal agregacji wynoszących 1, 2, 4, 8, 16, prowadzi do następujących estymatorów $\hat{\alpha}$, $\hat{\sigma}$, $\hat{H}$, umieszczonych w tab. 2.

Tabela 2. Estymatory parametrów rozkładów empirycznych

	$\hat{\alpha}$	$\hat{\sigma}$	$\hat{H}$	odch. $\hat{\alpha}$	zm. $\hat{\alpha}$	rozstęp $\hat{\alpha}$
BPHPBK	1,99298	11656,3	0,706812	0,00383	0,001923	0,00847
BRE	1,98719	20890,4	0,72036	0,00545	0,002744	0,0133
BZWBK	1,97787	34804,6	0,672363	0,00332	0,001679	0,00777
INGBSK	1,90855	5504,78	0,596582	0,01045	0,005477	0,02496
KREDYTB	1,9848	58904,1	0,719619	0,00248	0,001251	0,00637
MILLENNIUM	1,96643	376570	0,726159	0,00233	0,001184	0,00622
PEKAO	1,98611	78479,9	0,699316	0,00386	0,001946	0,00898

Źródło: opracowanie własne.

Otrzymane wyniki są bardzo zachęcające, bowiem nie dość, że we wszystkich regresjach pomocniczych współczynniki determinacji są większe niż 0,99⁵, to jeszcze bez względu na skale agregacji uzyskiwane dla każdego z banków wartości $\hat{\alpha}$

⁵ Co więcej, statystycznie istotne są parametry. Jednak w przypadku tego typu regresji wątpliwa jest normalność składników losowych a przez to i sensowność wnioskowania o istotności parametrów.

różnią się między sobą w nieznacznym stopniu⁶ (próbkowe odchylenia standardowe, współczynniki zmienności oraz rozstępy pomiędzy największą i najmniejszą wartością $\hat{\alpha}$ umieszczono w trzech ostatnich kolumnach tab. 2). Zauważmy także, że najniższą wartość $\hat{\alpha}$ (równą ok. 1,91) uzyskano dla INGBSK i jest to w pełni zgodne z tym, iż na wykresie wolumenu obrotu tego banku pojawiają się najsilniej dominujące wartości ekstremalne.

Można by zatem sądzić, iż rozpatrywane wolumeny obrotów mają rozkłady α -stabilne oraz że szeregi te charakteryzują się długoterminową zależnością, jako że w każdym przypadku: $\hat{H} > 1/\hat{\alpha}$. Jednak próba oszacowania parametrów μ i β , dokonana poprzez wyznaczenie empirycznej funkcji $\hat{\xi}_\theta = \text{Im}(\ln(\hat{\phi}_\theta))$, podważa te przypuszczenia. Dzieje się tak dlatego, że w przypadku każdego z szeregów empirycznych funkcja $\hat{\xi}_\theta$ w ogóle nie przypomina swoim przebiegiem funkcji postaci $t \rightarrow at + bt^c$, z wyjątkiem bardzo małego przedziału $\langle 0, \varepsilon \rangle$ ⁷, lecz nawet w tym przedziale dopasowanie funkcji postaci $t \rightarrow at + bt^c$ do każdego z szeregów jest dość słabe, a co gorsza – uzyskane w ten sposób estymatory $\hat{\beta}$ są większe od 1 (najmniejszy równy ponad 2).

W tej sytuacji należałoby odrzucić hipotezę, iż badane szeregi są α -stabilne. Jednak wcale nie jest to aż tak oczywiste, jak wskazują na to nieudane próby estymacji parametrów μ i β . A to dlatego, że opisane w literaturze metody estymacji (w tym wykorzystana modyfikacja metody Koutrouvelisa) wymagają założenia niezależności zmiennych w badanym szeregu⁸. Z kolei niezależność stoi w sprzeczności z wykrytą długoterminową zależnością. Zatem w przypadku badanych wolumenów obrotu brak jest rozstrzygnięcia co do ich α -stabilności.

5. Wnioski

Przeprowadzone badania pokazują, po pierwsze, że w przypadku rozkładów α -stabilnych uznanie próby statystycznej z wartościami odbiegającymi od większości obserwacji za jednorodną bądź niejednorodną nie dość, że jest samo w sobie wysoce kłopotliwe, to jeszcze w konsekwencji może zaważyć na obciążoności estymatorów. Fakt ten znalazł odbicie w badaniach symulacyjnych, które wykryły dużą wrażliwość estymatorów parametrów rozkładu α -stabilnego na pojedyncze nawet wartości ekstremalne. Można zatem wnioskować, że metoda Koutrouvelisa nie jest metodą odporną.

⁶ Fakt ten można uznać za swego rodzaju test α -stabilności.

⁷ ε jest zależne od rozpatrywanego szeregu.

⁸ Na uchylenie założenia niezależności wrażliwe są właśnie estymatory parametrów μ i β .

Po drugie, uwidoczniło się niedoskonałość dostępnych metod estymacji rozkładów α -stabilnych. Modelowym, a równocześnie mało realistycznym założeniem jest niezależność zmiennych. W tym kontekście przeprowadzone przykładowe badania empiryczne wskazują na problem związany z możliwością wystąpienia braku jakichkolwiek przesłanek przemawiających na korzyść lub niekorzyść użycia rozkładów α -stabilnych.

Wymienione w opracowaniu problemy i pułapki nie mają na celu ani zniechęcenia, ani odwołania od wykorzystywania rozkładów α -stabilnych, lecz jedynie uwidocznienie, iż do tego rodzaju modelowania należy podchodzić ostrożnie, a przy tym koniecznie opierać się na dogłębnej analizie danych empirycznych i znajdować pozastatystyczne estymacyjne przesłanki, przemawiające za użyciem rozkładów α -stabilnych.

Literatura

- [1] Domański C., Pruska K., Wagner W., *Wnioskowanie statystyczne przy nieklasycznych założeniach*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź 1998.
- [2] Gallardo J.R., Makrakis D., Orozco-Barbosa L., *Use of α -Stable Self-Similar Stochastic Processes for Modeling Traffic in Broadband Networks*, Performance Evaluation 2000, s. 71-78.
- [3] Garcia R., Renault E., Veredas D., *Estimation of Stable Distributions by Indirect Inference*, <http://www.unc.edu/~renault/research/stableSept04rg.pdf>, 2004.
- [4] Magiera R., *Modele i metody statystyki matematycznej*, Oficyna Wydawnicza GiS, Wrocław 2002.
- [5] Samorodnitsky G., Taqqu M.S., *Stable Non-Gaussian Random Processes*, Chapman and Hall, New York 1994.

PROBLEMS WITH MODELLING WITH USE OF α -STABLE DISTRIBUTIONS

Summary

The paper treats of some dangers and obstacles that may emerge when modeling data with use of α -stable distributions. These problems are illustrated by a simple simulation example and by an analysis of properly chosen empirical data alike. The carried research shows strong sensitivity of α -stable distribution parameter estimates. It is also shown that empirical data analysis may lead to unsolvable questions unfortunately.

The goal of the paper is not to discourage for using α -stable distributions but to emphasize that α -stable modelling is under the necessity of special caution, careful data analysis and of giving non-statistic arguments that could additionally validate usage of α -stable distributions.