

## WYBRANE ZASTOSOWANIA GŁĘBI STUDENTA W ODPORNEJ ANALIZIE STATYSTYCZNEJ\*

DANIEL KOSIOROWSKI

Katedra Statystyki  
Uniwersytet Ekonomiczny w Krakowie  
PL 31-510 Kraków, ul. Rakowicka 27  
e-mail: daniel.kosiorowski@uek.krakow.pl

Praca była przedstawiona przez autora na posiedzeniu Komisji Nauk Ekonomicznych i Statystyki Oddziału PAN w Krakowie, 17 grudnia 2010 r.

### ABSTRACT

D. Kosiorowski. *Certain Applications of Student Depth in Robust Economic Analysis*. Folia Oeconomica Cracoviensia 2010, 51: 27–55.

In this paper we present selected application of Mizera & Muller location — scale depth. We focus our attention on a Student depth function and propose several statistical procedures based on that statistical depth function.

### KEY WORDS — SŁOWA KLUCZOWE

location — scale depth, student median, robustness  
głębia położenia — rozrzutu, mediana studenta, odporność

### 1. WPROWADZENIE

Z formalnego punktu widzenia procedury odporne rozpatruje się obecnie jako funkcjonały statystyczne, definiowane na pewnej przestrzeni funkcji rozkładu (por. Huber i Ronchetti 2009). Statystyk analizuje zachowanie się procedury w pewnym otoczeniu zakładanego przez nią rozkładu. Rozpatrywane w badaniu rozkłady precyzują wiedzę statystyka na temat mechanizmu losowego,

---

\* Niniejsza praca powstała dzięki częściowemu wsparciu finansowemu, udzielonemu ze strony Ministerstwa Nauki i Szkolnictwa Wyższego RP w postaci grantu nr NN 111 193036.

rzządzającego zjawiskiem. Otoczenie zakładanego przez procedurę rozkładu ujmuje możliwe odstępstwa od przyjmowanych założeń odnośnie do zjawiska. Odstępstwa mogą dotyczyć występowania pośród danych: obserwacji obciążonych sporym błędem, błędną specyfikacją rozpatrywanego zjawiska itd. Otoczenia konstruowane są z wykorzystaniem stosownej odległości pomiędzy rozkładami prawdopodobieństwa. Wykorzystuje się w tym celu między innymi odległość Kołmogorowa, odległość Prohorowa (por. Jureckova i Picek 2006).

Istnieje co najmniej kilka, po części alternatywnych a po części komplementarnych, podejść do pomiaru odporności procedury statystycznej. Znane są też podejścia mające charakter wyłącznie jakościowy. Historycznie pierwsza własność procedury, która została wykorzystana do pomiaru jej odporności, wiązała się z pojęciem efektywności względnej estymatora w pewnym zakresie rozpatrywanych modeli. Obecnie centralną rolę odgrywają pojęcia: funkcji wpływu Hampela (patrz Hampel i in. 1986) oraz punktu załamania i punktu załamania próby skończonej Donoho i Hubera (Donoho i Huber 1983).

Przypuśćmy, że zastanawiamy się nad wpływem na wartość pewnej statystyki  $T_{n-1} = T(x_1, \dots, x_{n-1})$ , zmieszania zbioru obserwacji  $x_1, \dots, x_{n-1}$  z obserwacją odstającą  $\tilde{x}$ . Oznaczmy taki zmieszany zbiór danych jako  $x_1, \dots, x_{n-1}, \tilde{x}$  oraz wartość statystyki przy takim zmieszaniu  $T_n = T(x_1, \dots, x_{n-1}, \tilde{x})$ . Wpływ na wartość statystyki w przedstawionej sytuacji można mierzyć za pomocą zaproponowanej przez Tukey'a krzywej wrażliwości.

Krzywą wrażliwości (ang. *sensitivity curve*) nazywamy:

$$SC_n(\tilde{x}) = n(T_n - T_{n-1}). \quad (1)$$

Z definicji wynika natychmiast, że:

$$T_n = T_{n-1} + 1/n SC_n(\tilde{x}). \quad (2)$$

Niezmiernie popularna obecnie funkcja wpływu jest wersją krzywej wrażliwości w przypadku populacji. Funkcja wpływu została zaproponowana przez F. Hampela (Hampel 1975).

Rozważmy mieszaninę dwóch rozkładów  $F_\epsilon = (1 - \epsilon)F + \epsilon\delta_x$ , gdzie  $\delta_x$  oznacza rozkład skoncentrowany w punkcie. Możemy określić jakościowo odporność procedury porównując  $T(F)$  i  $T(F_\epsilon)$  w sytuacji, gdy  $\epsilon \rightarrow 0$ . Aby ująć odporność ilościowo, wygodnie jest posłużyć się funkcją wpływu (ang. *influence function*) definiowaną:

$$IF(x; T, F) = \lim_{\epsilon \rightarrow 0} \frac{T(F_\epsilon) - T(F)}{\epsilon}. \quad (3)$$

Funkcja wpływu jest jedną z najważniejszych charakterystyk funkcjonału statystycznego, estymatora. Wartość  $IF(x; T, P)$  mierzy efekt zakłócenia funkcjonału  $T$  poprzez pojedynczą wartość  $x$ . Odporny funkcjonał  $T$  powinien mieć ograniczoną funkcję wpływu.

Bez wątpienia najciekawszą (Davies i Gather 2005, Davies 2002) z praktycznego punktu widzenia miarą odporności procedury statystycznej jest wersja koncepcji punktu załamania odnosząca się do prób skończonych, a mianowicie tzw. punkt załamania próby skończonej, wprowadzony przez Donoho i Hubera<sup>1</sup> (Donoho i Huber 1983).

Przypuśćmy, że dysponujemy próbą  $\mathbb{X}^n = \{x_1, \dots, x_n\}$ , złożoną z  $n$  obserwacji generowanych przez zakładany model oraz niech  $\mathbb{Y}^m = \{y_1, \dots, y_m\}$  oznacza  $m$  dowolnych (być może szczególnie odstających) obserwacji. Oznaczmy przez  $Z^{n+m} = \mathbb{X}^n \cup \mathbb{Y}^m$  próbę powstałą z połączenia powyższych zbiorów obserwacji. Określmy ją mianem  $\varepsilon_m$  zmieszanej, próby gdzie  $\varepsilon_m = \frac{m}{n+m}$ .

Wielkość  $|T(\mathbb{X}^n \cup \mathbb{Y}^m) - T(\mathbb{X}^n)|$  oznacza obciążenie statystyki, natomiast maksymalne obciążenie statystyki  $T$ , powstające przy  $\varepsilon_m$  zmieszaniu, oznaczmy jako:

$$B(\varepsilon_m, T, \mathbb{X}^n) = \sup_{\mathbb{Y}^m} |T(\mathbb{X}^n \cup \mathbb{Y}^m) - T(\mathbb{X}^n)|. \quad (4)$$

Punkt załamania próby skończonej (Donoho i Huber 1983) definiujemy jako:

$$BP(T, \mathbb{X}^n) = \inf \{ \varepsilon_m : B(\varepsilon_m, T, \mathbb{X}^n) = \infty \}. \quad (5)$$

Punkt załamania próby skończonej posiada odpowiednik w populacji. Przypuśćmy, że otoczenia rozkładu generującego dane definiujemy z wykorzystaniem odległości mieszanki dwóch rozkładów. Niech  $F$  oznacza zakładany rozkład, natomiast  $H$  oznacza rozkład reprezentujący błąd, zaburzenie (odstępstwo od modelu). Rozważamy model mieszanki postaci:  $F_\varepsilon = (1 - \varepsilon)F + \varepsilon H$ . Wprowadzając pojęcie maksymalnego obciążenia przy tego rodzaju  $\varepsilon$  — zmieszaniu, tzn.:

$$B(\varepsilon, T, F) = \sup |T(F_\varepsilon) - T(F)|, \quad (6)$$

otrzymamy definicję punktu załamania procedury statystycznej F. Hampela (Hampel 1968):

$$\varepsilon^*(T, F) = \inf \{ \varepsilon : B(\varepsilon, T, F) = \infty \}. \quad (7)$$

Punkt załamania próby skończonej wskazuje na maksymalną frakcję obserwacji odstających w próbie, która nie sprawia, że procedura statystyczna „łamie się” — np. obciążenie wskazania estymatora staje się nieakceptowalne. Koncepcja punktu załamania zależy od odległości wykorzystywanej do konstruowania otoczeń zakładanego rozkładu generującego obserwacje. Zależy

---

<sup>1</sup> Należy podkreślić, że koncepcja punktu załamania ma wiele, często istotnie różniących się wariantów. Mamy tutaj m.in. zastosowanie pojęcia w przypadku prób zależnych, szeregów czasowych itd. (por. Genton i Lucas 2003).

także od zagadnienia, do którego się stosuje. Czymś odmiennym jest „załamanie się” estymatora położenia centrum, estymatora wielowymiarowego rozrzutu czy estymatora parametrów funkcji regresji (por. Davies 2002).

Zagadnienia odporności wiążą się niemalże z każdą procedurą statystyczną wykorzystywaną w ekonomii. Można mówić o odpornej, bankowej procedurze scoringowej, odpornej prognozie inflacji bądź odpornym szacowaniu ryzyka ubezpieczeniowego. Warto zauważyć, że statystyka pojawia się w ekonomii nie tylko na poziomie szacowania, weryfikacji pewnego modelu, ale także na poziomie pojęć, którymi posługuje się ekonomista. Produkt narodowy brutto, tempo wzrostu gospodarczego, dobrobyt, sprawiedliwość społeczna, dyskryminacja na rynku pracy — to agregaty statystyczne. Można zadać pytanie: czy agregaty te są odporne?

Zdaniem autora zasadą jest wyodrębnienie ze zbioru postępowań, wykorzystujących statystykę w ekonomii, tzw. odpornej analizy statystycznej. Analizę, którą można opisać jako ciągle stosowanie wysoce efektywnych narzędzi statystycznych w procesie decyzyjnym, z naciskiem na poszukiwanie tendencji wyznaczonej przez większość rozpatrywanych obiektów. Mamy tu na uwadze analizę opierającą się na ustawicznie aktualizowanych danych (co godzinę, co dzień itd.), dotyczących np. sytuacji na rynku płodów rolnych, na rynku nieruchomości itd. Prezentowane w dalszej części pracy odporne procedury indukowane przez uogólnienia statystycznej funkcji głębi Tukey’a mogą — zdaniem autora — znaleźć szereg zastosowań w tak rozumianej odpornej analizie statystycznej. Autor żywi nadzieję, że prezentowane pojęcia w przyszłości znajdą też zastosowanie w ekonometrycznym modelowaniu zjawisk ekonomicznych, stając się alternatywą np. dla procesów GARCH, SV. Zagadnienia te wymagają jednakże dalszych studiów.

## 2. UOGÓLNIENIA GŁĘBI DOMKNIĘTEJ PÓLPRZESTRZENI TUKEY’A

Rozwijany obecnie nurt odpornej, wielowymiarowej analizy danych, określany mianem koncepcji głębi danych, został zapoczątkowany przez propozycje Johna Tukey’a — miał on na celu rozszerzenie na przypadek wielowymiarowy jednowymiarowych procedur statystycznych, opierających się o statystyki porządkowe i rangi (Tukey 1975). Statystyczna funkcja głębi punktu  $\mathbf{x} \in \mathbb{R}^d$  względem rozkładu prawdopodobieństwa  $F$  określonego na  $\mathbb{R}^d$  przyporządkowująca punktowi  $\mathbf{x}$  najmniejsze prawdopodobieństwo zgromadzone na domkniętej półprzestrzeni, do której brzegu należy  $\mathbf{x}$ , nazywamy głębią domkniętej półprzestrzeni bądź głębią Tukey’a.

$$TD(\mathbf{x}; F) = \inf_H \{P(H) : \mathbf{x} \in H \text{ i } H \text{ jest domkniętą półprzestrzenią } \mathbb{R}^d\}. \quad (8)$$

W przypadku próby  $X^n = \{x_1, \dots, x_n\}$  rozkład prawdopodobieństwa  $F$  zastępujemy rozkładem empirycznym  $F_n$ . W przypadku próby definicję można zapisać równoważnie:

$$TD(\mathbf{y}; X^n) = \frac{1}{n} \cdot \min_{\|\mathbf{u}\|=1} \#\{i : \mathbf{u}'x_i \leq \mathbf{u}'\mathbf{y}\}, \quad (9)$$

gdzie,  $\mathbf{u}$  przebiega wszystkie wektory w  $\mathbb{R}^d$  z  $\|\mathbf{u}\| = 1$ .

Głębia Tukey'a umożliwia porządkowanie punktów  $\mathbf{x} \in \mathbb{R}^d$  względem rozkładu prawdopodobieństwa  $F$  (rozkładu z próby  $F_n$ ) na zasadzie odstawiania od centrum — punktu, w którym statystyczna funkcja głębi przyjmuje wartość maksymalną, określanego medianą Tukey'a.

Głębia Tukey'a w punkcie  $\mathbf{x} \in \mathbb{R}^d$  przyjmuje wartości z przedziału  $[0, 1]$ . Niewielkie wartości odpowiadają peryferiom rozkładu, wartości bliższe jedności odpowiadają centrum rozkładu. Zbiór punktów:

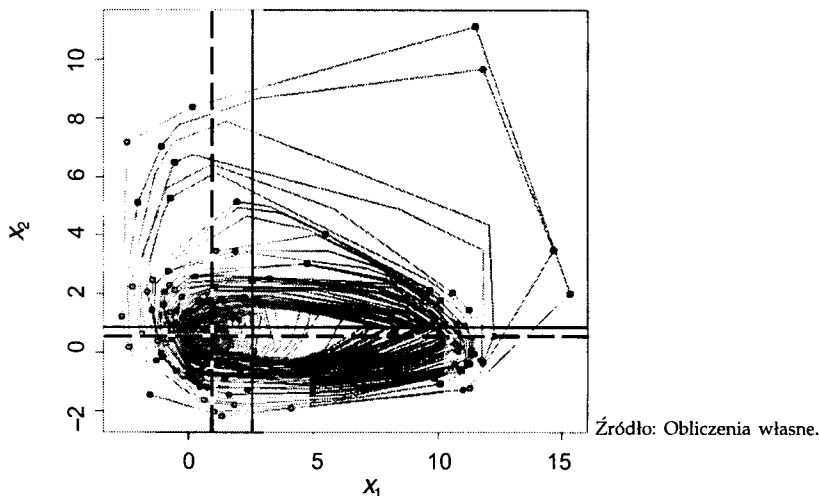
$$D_\alpha = \{\mathbf{x} \in \mathbb{R}^d : D(\mathbf{x}, F) \geq \alpha\},$$

nazywamy  $\alpha$  — obszarem centralnym, zbiór punktów:

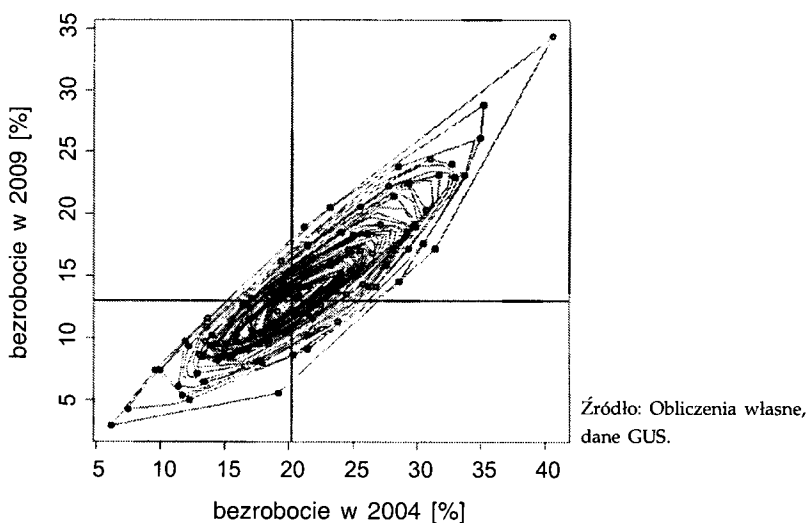
$$C_\alpha = \{\mathbf{x} \in \mathbb{R}^d : D(\mathbf{x}, F) \geq \alpha\}$$

nazywamy konturem rzędu. Obszary centralne tworzą zagnieżdżoną rodzinę wypukłych obszarów. Rozkład empiryczny dowolnego zbioru danych  $X^n \in \mathbb{R}^d$  jest jednoznacznie wyznaczony przez jego empiryczną głębię Tukey'a, tzn. listę konturów z próby. Głębia ta jest afinicznie niezmiennicza. Niektóre ze statystycznych głębi Tukey'a zostały pokazane przez Donoho i Gasko (1992) oraz Masse i Theoredescu (1994). Mamy tu na uwadze między innymi zbieżność konturów z próby do konturów w populacji, słabą zbieżność głębi z próby w punkcie do odpowiednika w populacji, zbieżność prawie na pewno mediany Tukey'a z próby do jej odpowiedniczki w populacji. O ile głębia Tukey'a w punkcie nie jest zbyt odporna, to mediana Tukey'a ma BP > 20% oraz ograniczoną funkcję wpływu. To wystarcza dla wielu zastosowań praktycznych. Głębia Tukey'a nie wykorzystuje informacji metrycznych zawartych w próbie. To jej zaleta w kontekście zastosowań w badaniach zmiennych o wartościach na słabszych aniżeli ilorazowa skalach.

Na rycinie 1 przedstawiono wykres konturowy głębi Tukey'a dla próby 200 obserwacji, wygenerowanych z mieszaniny dwuwymiarowego rozkładu normalnego o udziale 80% i dwuwymiarowego rozkładu T-Studenta o trzech stopniach swobody i udziale 20%. Przecięcie linii przerywanej reprezentuje medianę Tukey'a z próby, przecięcie linii koloru czarnego reprezentuje wektor średnich. Na rycinie 2 przedstawiono wykres konturowy głębi Tukey'a dla próby powiatów województw łódzkiego, dolnośląskiego, mazowieckiego i małopolskiego, rozpatrywanych ze względu na stopę bezrobocia rejestrowanego w latach 2004 i 2009. Przecięcie linii reprezentuje medianę Tukey'a.



Ryc. 1. Kontury głębi Tukey'a dla próby z mieszaniny rozkładu normalnego (80%) i rozkładu T-Studenta o trzech stopniach swobody (20%)



Rys. 2. Kontury głębi Tukey'a dla próby powiatów rozpatrywanych co do stopy bezrobocia rejestr. w latach 2004 i 2009

Głębia Tukey'a posiada centralną pozycję w koncepcji głębi danych. Znanych jest wiele sposobów jej uogólnienia, których najbardziej znane jest podejście Zhanga, opierające się na metodologii *projection pursuit* (Zhang 2002) oraz podejście Mizery (Mizera 2002; Mizera i Muller 2004). Niniejsza praca ma na celu zaprezentowanie wybranych możliwości zastosowania podejścia Mizery i Mullera w odpornej analizie ekonomicznej.

### 3. GŁĘBIA POŁOŻENIA–ROZRZUTU MIZERY I MULLER

Przypuśćmy, że staramy się znaleźć dopasowanie (punkt wiernie oddający położenie, postać zależności itp.), element pewnego zbioru parametrów  $\theta \in \Theta$  do obserwacji  $\mathbf{Z}^n = \{\mathbf{z}_1, \dots, \mathbf{z}_n\}$  generowanych przez rozkład wektora losowego  $\mathbf{Z}$ . W celu wskazania optymalnego dopasowania przyjmuje się pewne kryterium optymalności, np. pewną funkcję obserwacji  $F$ . Oznaczmy wartość funkcji kryterium w punkcie  $\mathbf{z}_i$  jako  $F_i$ . Można przyjąć, że im niższa wartość  $F_i$  tym lepiej  $\theta$  odzwierciedla  $\mathbf{z}_i$ . Na ogół nie istnieje dopasowanie jednostajnie najlepsze dla wszystkich obserwacji  $\mathbf{z}_i$ , dlatego też decydujemy się na pewien kompromis, przykładowo na sumę  $F_i$ .

Głębia w ujęciu Mizery odzwierciedla pewien stopień dopuszczalności dopasowania, zważywszy na rozpatrywany zbiór danych. Jest rozwinięciem ogólnej definicji głębi przedstawionej w artykule Rousseeuw i Hubert (1999), a wywodzącej się od głębi Tukey'a. Głębia dopasowania  $\theta$  oznacza najmniejszą liczbę obserwacji, których usunięcie z próby sprawia, że  $\theta$  przestaje być dopasowaniem, to znaczy staje się wartością parametru, która zważywszy na rozpatrywany zbiór danych jest niedopuszczalna — wartość głębi równa się wówczas zero. Mizera proponuje mianowicie: niech  $\mathbf{Z}^n = \{\mathbf{z}_1, \dots, \mathbf{z}_n\}$  oznacza  $n$ -elementową próbę z  $\mathbf{Z}$ , niech  $N = \{1, \dots, n\}$  oznacza zbiór indeksów, niech  $A \subseteq N$  oznacza pewien podzbiór zbioru indeksów. Załóżmy, że funkcja kryterium  $F$  działa ze zbioru parametrów  $\Theta$  w zbiór  $[0, \infty)$  oraz, że jest określona dla każdej obserwacji  $\mathbf{z}_i \in \mathbf{Z}^n$ ,  $F_{\mathbf{z}_i}(\theta) \equiv F_i$ .

Wartość parametru  $\theta \in \tilde{\Theta} \subseteq \Theta$  nazwiemy słabo optymalną w  $\tilde{\Theta}$  względem  $A \subseteq N$  (względem obserwacji o indeksach z  $A$ ), jeżeli  $A \neq \emptyset$  i nie istnieje  $\theta \in \Theta$  taki, że  $F_i(\tilde{\theta}) < F_i(\theta)$  dla wszystkich  $i \in A$ .

Definiujemy globalną głębię dopasowania  $\tilde{\theta} \in \tilde{\Theta}$  jako:

$$DG(\theta, \mathbf{Z}^n) = \min\{A \subseteq N : \theta \text{ nie jest słabo optymalny w } \Theta \text{ względem } A'\}, \quad (10)$$

gdzie  $A'$  oznacza dopełnienie zbioru  $A$ . (Ile wynosi minimalna liczba obserwacji, których usunięcie sprawia, że dopasowanie przestaje być optymalne względem obserwacji, które zostały.)

Typowe funkcje kryterium w wielowymiarowym zagadnieniu położenia mają postać:  $F_i(\mathbf{z}_i) = \|\mathbf{z}_i - \theta\|$  bądź  $F_i(\mathbf{z}_i) = \|\mathbf{z}_i - \theta\|^2$ . W zagadnieniu regresji liniowej: gdzie  $\mathbf{z}_i = (x_i, y_i)$  powszechnie wykorzystuje się funkcje kryterium postaci:  $F_i(\theta) = (y_i - x_i' \theta)^2$  albo  $F_i(\theta) = |y_i - x_i' \theta|$ .

Biorąc pochodne w zagadnieniu optymalizacyjnym funkcji kryterium Mizery definiuje głębię styczną dopasowania  $\theta$  jako:

$$TD(\theta; \mathbf{Z}^n) = \frac{1}{n} \inf_{\mathbf{u} \neq 0} \#\{i : \mathbf{u}' \nabla F_i(\theta) \geq 0\}, \quad (10)$$

gdzie  $\nabla F_i(\theta)$  to gradient funkcji kryterium dla dopasowania  $\theta$  w ustalonym punkcie  $\mathbf{z}_i$ .

W wielowymiarowym zagadnieniu położenia gradient funkcji kryterium ma postać:  $\nabla F_i(\theta) = \theta - z$ , dlatego też:

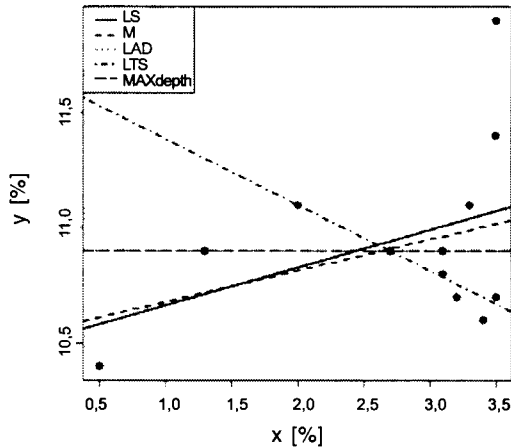
$$TD(\theta; Z^n) = \min_{\substack{\# \\ u \neq 0}} \{i : u'(\theta - z_i) \geq 0\} = \min_{\substack{\# \\ |u|=1}} \{i : z_i \in H_{\theta, u}\}. \quad (11)$$

Powyższe sformułowanie jest równoważne z oryginalną definicją głębi domkniętej półprzestrzeni Tukey'a (1975) oraz Donoho i Gasko (1992).

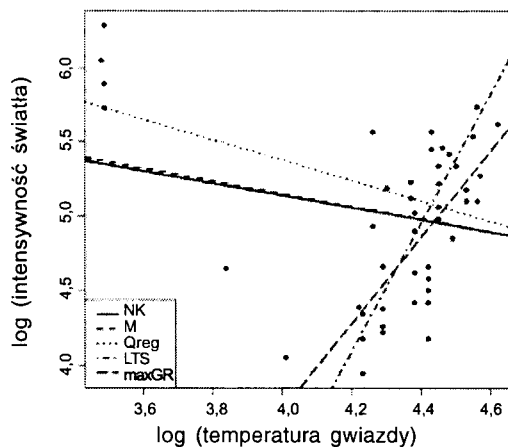
W przypadku regresji liniowej dla kryterium  $F_i(\theta) = 1/2(y_i - x_i'\theta)^2$ :

$$TD(\theta; Z^n) = \min \{i : -u'x_i(y_i - x_i'\theta) \geq 0\} = \min \{i : \text{sgn}(u'x_i)\text{sgn}(y_i - \theta'x_i) \geq 0\}. \quad (12)$$

Łatwo zauważyć tu oryginalną definicję głębi regresyjnej Rousseeuw i Hubert (1999).



Ryc. 3. Miesięczna stopa inflacji (x) vs. miesięczna stopa bezrobocia w Polsce w 2009 r. Dopasowania uzyskane za pomocą pięciu kryteriów



Ryc. 4. Liniowa funkcja regresji dopasowana do zbioru danych CYGOB1 za pomocą pięciu alternatywnych kryteriów



Na rycinach 3 i 4 pokazano przykłady zastosowania estymatora maksymalnej głębi regresyjnej w porównaniu z estymatorami NK, M-, LTS (najmniejszych przyciętych kwadratów) i LMS (najmniejszej mediany kwadratów). Estymator maksymalnej głębi regresyjnej Rousseeuw i Hubert odznacza się punktem załamania bliskim 30%, dobrą szybkością zbieżności oraz efektywnością. O modelu generującym dane zakładamy jedynie, że warunkowa mediana odpowiedzi jest liniowa względem zmiennych objaśniających. Głębina dopasowania (prostej) równa jest minimalnej liczbie obserwacji, które napotykamy obracając dopasowanie do pozycji pionowej.

Mizera rozszerza swoje rozważania na przypadek populacji. Wprowadza pole losowe na rozważanej w danym zagadnieniu borelowskiej  $\sigma$  algebrze zbiorów, pole reprezentuje możliwe populacje generujące dane. Oznaczając przez  $\Phi_{\theta}^{-1}(E) = \{z : \Phi_{\theta} \in E\}$ , przeciwobraz zbioru  $E$  przy funkcji  $\Phi$ , przyjmując, jak poprzednio:  $\Phi_{\theta}(Z) = \nabla F_z(\theta)$  oraz przyjmując, że  $P_{\theta} = P \circ \Phi_{\theta}^{-1}$  definiujemy głębię styczną parametru  $\theta$  przy rozkładzie prawdopodobieństwa  $P$  jako:

$$TD(\theta, P) = D(P_{\theta}) = D(P \circ \Phi_{\theta}^{-1}) = \inf_{u \neq 0} P(\Phi_{\theta}^{-1}(H_u)). \quad (13)$$

Mizera pokazuje zastosowanie powyższe definicji między innymi w przypadku parametru położenia i regresji liniowej, uzyskując dolne ograniczenie dla BP procedury indukowanej przez głębię.

Zdaniem autora warto zwrócić szczególną uwagę na rozwinięcie podejścia Mizery wiążące się zastosowaniem funkcji kryterium wywodzących się z zasady największej wiarygodności. Niech  $y_i$  oznaczają realizacje niezależnych zmiennych losowych o tym samym rozkładzie, o gęstości  $f$ , wyznaczone z dokładnością do parametru położenia  $\mu$  i rozrzutu  $\sigma$ . W takiej sytuacji logarytm wiarygodności próby przyjmuje postać:

$$\sum_{i=1}^n \left( -\log f \left( \frac{y_i - \mu}{\sigma} \right) + \log \sigma \right). \quad (14)$$

Mizera i Muller proponują, aby jako funkcję kryterium przyjąć:

$$F_i(\mu, \sigma) = -\log f \left( \frac{y_i - \mu}{\sigma} \right) + \log \sigma, \quad (15)$$

oraz jako definicję głębi — formułę głębi stycznej zaproponowaną przez Mizere (2002):

$$D(\theta, Y^n) = \inf_{u \neq 0} \# \{i : -u' \nabla F_i(\theta) \geq 0\}.$$

Uzyskaną w ten sposób rodzinę funkcji głębi, zależną od przyjętej funkcji gęstości, nazywają głębiami położenia-rozrzutu.

**DEFINICJA 1.** Głębina położenia i rozrzutu Mizery i Muller punktu  $(\mu, \sigma) \in \mathbb{R} \times [0, \infty)$  względem próby  $Y^n = \{y_1, \dots, y_n\}$  określona jest wyrażeniem:

$$D((\mu, \sigma, Y^n) = \inf_{u \neq 0} \# \left\{ i : (u_1, u_2) \left( \frac{\psi(\tau_i)}{\chi(\tau_i) - 1} \right) \geq 0 \right\}, \quad \sigma > 0$$

$$= \# \{ i : y_i = \mu \}, \quad \text{dla } \sigma = 0, \quad (16)$$

gdzie  $\tau_i$  jest skrótem dla  $(y_i - \mu)/\sigma$  oraz  $\psi, \chi$  zależą od gęstości  $f$   $\psi(\tau) = (-\log f(\tau))' = -f'(\tau)/f(\tau)$  oraz  $\chi(\tau) = \tau \psi(\tau)$ .

Ustalając jeden z parametrów, otrzymujemy definicję głębi położenia bądź głębi rozrzutu. Definicja 1 daje wiele możliwości — umożliwia wprowadzenie całej rodziny głębi zależnych od gęstości, np. gęstości  $f$  rozkładu Studenta z stopniami swobody:

$$D((\mu, \sigma, Y^n) = \inf_{u \neq 0} \# \left\{ i : (u_1, u_2) \left( \frac{n \tau_i}{n+1} (\tau_i^2 - 1) \right) \geq 0 \right\}. \quad (17)$$

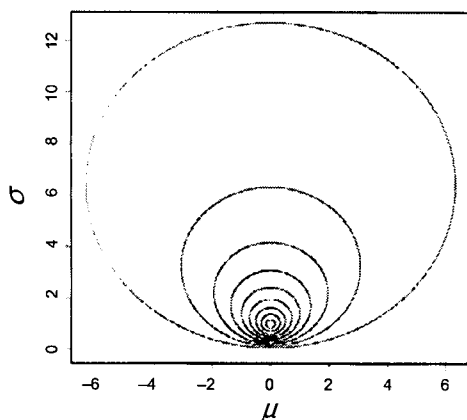
DEFINICJA 2. Głębia Studenta położenia i rozrzutu punktu  $(\mu, \sigma) \in \mathbb{R} \times [0, \infty)$  względem rozkładu prawdopodobieństwa  $R$  na  $\mathbb{R}$  określona jest wyrażeniem:

$$D((\mu, \sigma), P) = \inf_{(u_1, u_2) \neq 0} P \{ y : u_1(y - \mu) + u_2((y - \mu)^2 - \sigma^2) \geq 0 \}. \quad (18)$$

Empiryczną głębię Studenta otrzymamy zastępując rozkład  $P$  jego empirycznym odpowiednikiem  $P_n$ . Głębia Studenta jest dwuwymiarową głębią Tukey'a na płaszczyźnie Poincaré'a, będącej modelem hiperbolicznej geometrii Łobaczewskiego. Dla dowolnego rozkładu prawdopodobieństwa o ciągłej dystrybucji i o spójnym nośniku  $D((\mu, \sigma), P) < 1/2$ , istnieje punkt o maksymalnej głębi i jest on wyznaczony jednoznacznie — istnieje punkt  $(\tilde{\mu}, \tilde{\sigma})$  taki, że  $D((\tilde{\mu}, \tilde{\sigma}), P) \geq 1/3$ . Można zdefiniować symplcjonalną głębię położenia-rozrzutu, która jest niezależna od własności metrycznych próby. Głębia położenia-rozrzutu jest ekwiwariantna względem położenia i rozrzutu: jeżeli punkty próby przekształcimy za pomocą  $g(y) = ay + b$ , wówczas głębia przekształconego parametru  $(a\mu + b, a\sigma)$  jest taka sama jak  $(\mu, \sigma)$ . Głębia Studenta jest niezmiennicza względem grupy Möbiusa. Dla symetrycznych rozkładów  $\mu$  mediany Studenta leży blisko mediany z próby, dla rozkładów asymetrycznych jednodalnych  $\mu$  mediany Studenta — leży bliżej modalnej niż mediana,  $\sigma$  mediany Studenta jest na ogół mniejsze niż MAD. Głębia Studenta dla dowolnego rozkładu  $P$  zbiega jednostajnie względem  $(\mu, \sigma)$  prawie na pewno  $D((\mu, \sigma), P_n) \rightarrow D((\mu, \sigma), P)$ . Można pokazać zbieżność konturów z próby do konturów w populacji oraz zbieżność prawie na pewno estymatora maksymalnej głębi — mediany Studenta. Postać rozkładu asymptotycznego mediany Studenta nie jest znana. Szybkość zbieżności rzędu  $\sqrt{n}$  wynika z ogólnej teorii, wydaje się, że jest lepsza. Punkt załamania BP mediany Studenta wynosi 33%. Pozostaje otwarty problem: czy głębia Studenta charakteryzuje rozkład jednoznacznie? Głębia Studenta dobrze wychwytuje asymetrię oraz tłuste ogony populacji. Mediana Stu-

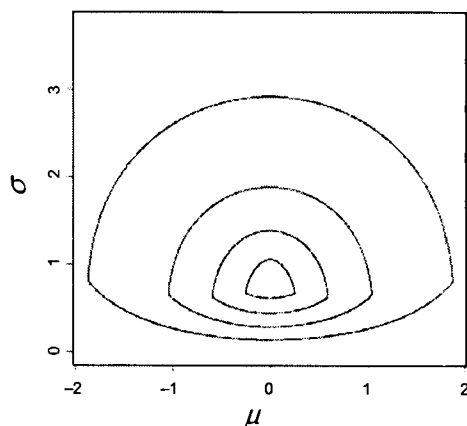
denta jest zgodnym estymatorem centrum symetrii populacji. W kontekście zastosowań głębi położenia–rozrzutu wskaźmy przykładowo, że ocena własnej sytuacji w grupie przez jednostkę zależy od jej odległości do centrum (położenie), panującego w grupie zróżnicowania pozycji (rozrzut) oraz przeświadczenia o mechanizmie losowym, generującym poszczególne pozycje (gęstość).

Ryciny 5–14 przedstawiają wykresy konturowe głębi Studenta dla wybranych a często wykorzystywanych w ekonomii, rozkładów oraz dla mieszanin rozkładów. Z rysunków jasno wynika, że głębia Studenta jest wrażliwa na typ rozkładu. Wykres konturowy może zostać wykorzystany jako alternatywa dla wykresu kwantyl–kwantyl. Jego zaletą jest to, że nie musimy skalować danych co jest konieczne w przypadku wykresu kwanty–kwantyl. Z rycin wynika, że wykres konturowy bardzo dobrze wychwytuje asymetrię rozkładu oraz zmieszanie rozkładu z rozkładem reprezentującym zaburzenie.



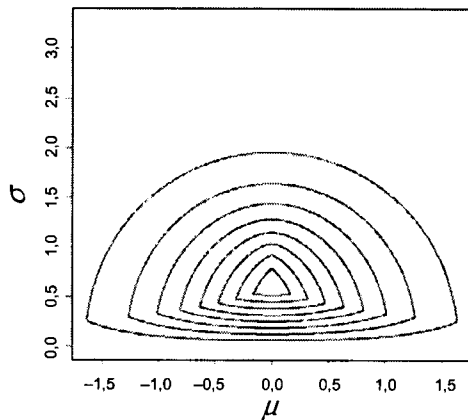
Źródło: obliczenia własne.

Ryc. 5. Kontury głębi Studenta dla rozkładu Cauchy'ego



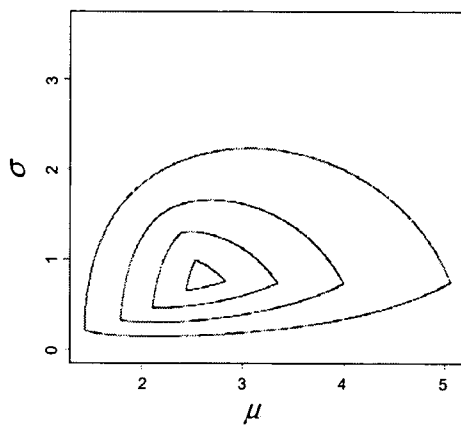
Źródło: obliczenia własne.

Ryc. 6. Kontury głębi Studenta dla rozkładu Studenta  $t$  o dwóch stopniach swobody



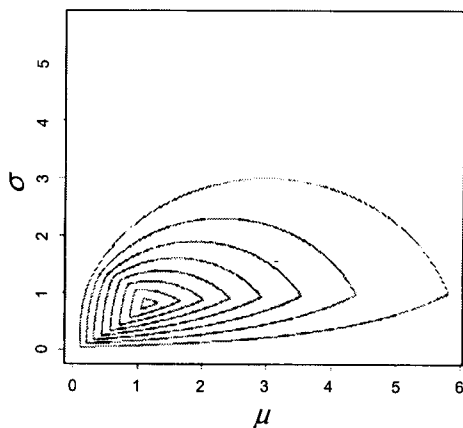
Źródło: obliczenia własne.

Ryc. 7. Kontury głębi Studenta dla rozkładu  $N(0,1)$



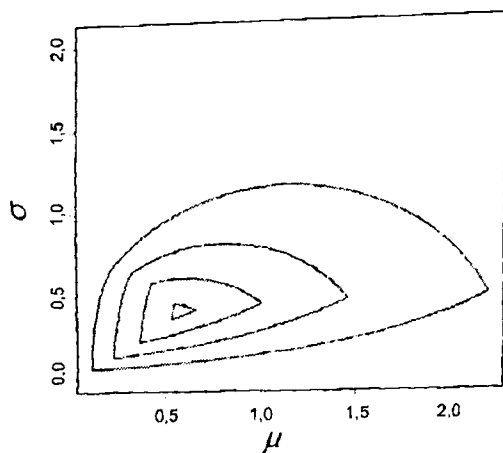
Źródło: obliczenia własne.

Ryc. 8. Kontury głębi Studenta dla rozkładu lognormalnego o parametrach 1 i 0,5



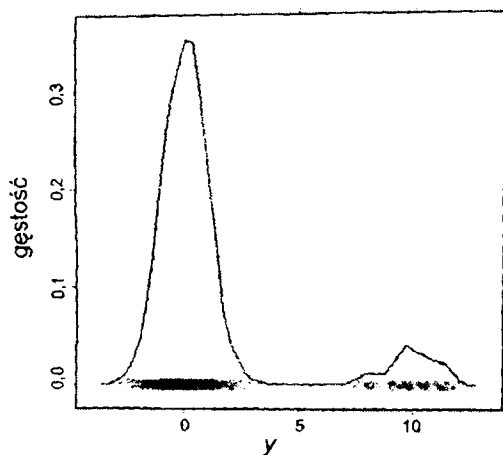
Źródło: obliczenia własne.

Ryc. 9. Kontury głębi Studenta dla rozkładu  $\chi^2$  o dwóch stopniach swobody



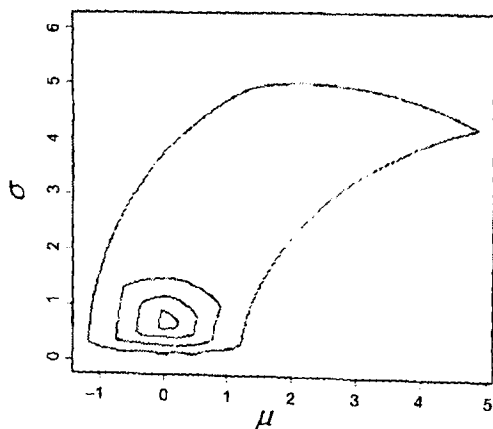
Źródło: obliczenia własne.

Ryc. 10. Kontury gębi Studenta dla rozkładu wykładniczego o wartości parametru równej 1



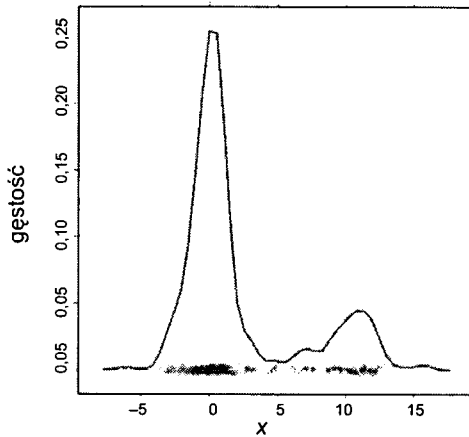
Źródło: obliczenia własne.

Ryc. 11. Oszacowanie gęstości mieszanki rozkładu  $N(0,1)$  (udział 90%) i rozkładu  $N(10,1)$  (udział 10%)



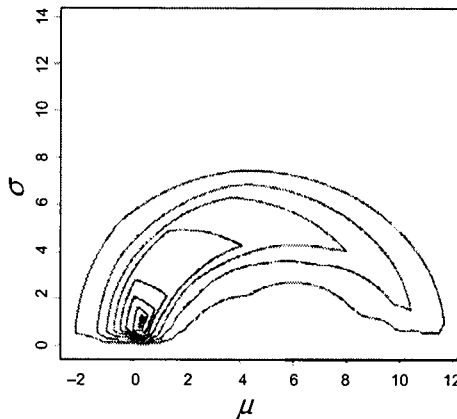
Źródło: obliczenia własne.

Ryc. 12. Kontury gębi Studenta dla mieszanki rozkładu  $N(0,1)$  (udział 90%) i rozkładu  $N(10,1)$  (udział 10%)



Źródło: obliczenia własne.

Ryc. 13. Oszacowanie gęstości mieszaniny rozkładu  $t(3)$  (udział 80%) i rozkładu  $N(10,2)$  (udział 20%)



Źródło: obliczenia własne.

Ryc. 14. Kontury głębi Studenta dla mieszaniny rozkładu  $t(3)$  (udział 80%) i rozkładu  $N(10,2)$  (udział 20%)

Procedury statystyczne indukowane przez statystyczne funkcje głębi często traktowane są jako niezależne, które nie mają związku z procedurami statystycznymi, rozwijanymi w ramach głównego nurtu statystyki odpornej. Podejścia Mizery i Mizery i Mullera pokazują, że istnieje ścisły związek pomiędzy koncepcją głębi danych a dorobkiem klasycznej statystyki matematycznej. Dla przykładu poniżej wskażemy na związki estymatorów maksymalnej głębi z szeroko rozpowszechnioną klasą M-estymatorów (por. Maronna i in. 2006). M-estymatory wywodzące się z metody największej wiarygodności występują jako estymatory odporne w większości komercyjnych pakietów statystycznych.

Rozważmy jednowymiarowy zbiór danych  $\mathbb{X}^n = \{x_1, \dots, x_n\}$  o rozkładzie empirycznym  $F_{nX}$ . Niech  $\mu(\cdot)$  oraz  $s(\cdot)$  oznaczają jednowymiarowe, ekwiwariantne

funkcjonały położenia i rozrzutu na rodzinie jednowymiarowych rozkładów prawdopodobieństwa. Wówczas M-estymator położenia i rozrzutu definiujemy odpowiednio jako rozwiązanie  $(\beta, \sigma)$  równań:

$$\frac{1}{n} \sum_{i=1}^n \psi \left( \frac{x_i - \beta}{s(F_{nZ})} \right) = 0, \quad (19)$$

$$\frac{1}{n} \sum_{i=1}^n \chi \left( \frac{z_i - \mu(F_{nZ})}{\sigma} \right) = 0, \quad (20)$$

gdzie  $\psi$  oraz  $\chi$  oznaczają stosownie dobrane funkcje, odpowiednio nieparzystą i parzystą.

Zauważmy, że wielkość:

$$O_{\psi n} = \left| \sum_{i=1}^n \underbrace{\psi \left( \frac{x_i - b}{s(F_{nZ})} \right)}_{F_i} \right| / n, \quad (21)$$

można potraktować jako miarę odstawiania  $\beta$  względem zbioru danych  $\mathbb{X}^n$ ,  $F_i$  natomiast jako uogólnienie funkcji kryterium Mizery i Mullera.

Aby zdefiniować głębnię położenia można posłużyć się podejściem Mizery i określić ją jako:

$$D(\beta, \mathbb{X}^n) = \frac{1}{1 + O_{\psi n}(\beta, \mathbb{X}^n)}. \quad (22)$$

Podobnie wielkość:

$$O_{\chi n}(\sigma, \mathbb{X}^n) = \left| \sum_{i=1}^n \underbrace{\chi \left( \frac{z_i - \mu(F_{nZ})}{\sigma} \right)}_{G_i} \right| / n, \quad (23)$$

możemy potraktować jako miarę odstawiania charakterystyki rozrzutu  $\sigma$  względem  $\{|x_1 - \mu(F_{nZ})|, \dots, |x_n - \mu(F_{nZ})|\}$ ,  $G_i$  natomiast jako funkcję kryterium.

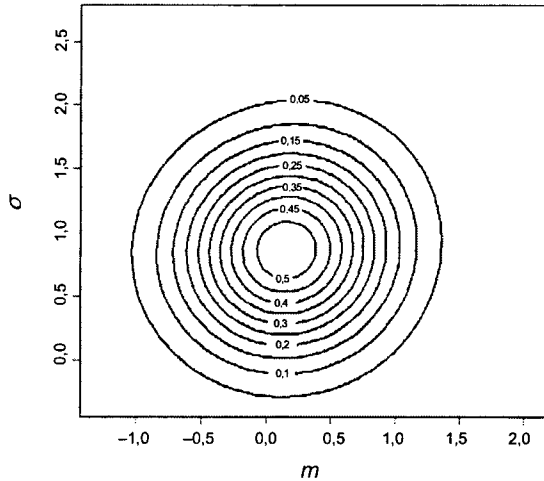
Odpowiednią funkcję głębni charakterystyki rozrzutu definiujemy jako:

$$D(\sigma, \mathbb{X}^n, \mu) = \frac{1}{1 + O_{\chi n}(\sigma, \mathbb{X}^n)}. \quad (24)$$

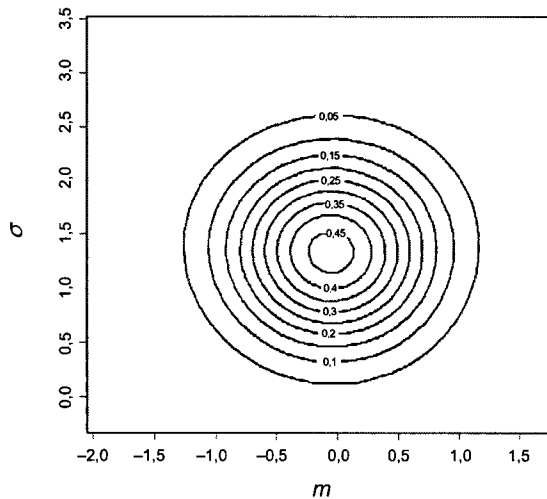
Zauważmy zatem, że M-estymatory położenia i rozrzutu są po prostu estymatorami maksymalnej głębni zdefiniowanych powyżej funkcji głębni 22 i 24 bądź estymatorami maksymalnej głębni stycznej Mizery.

## 4. ZASTOSOWANIA GŁĘBI STUDENTA

Łączne szacowanie charakterystyk położenia i rozrzutu zmiennej losowej pojawia się w wielu szczególnie istotnych z punktu widzenia zastosowań praktycznych zagadnieniach statystyki matematycznej. Dla przykładu wskaźmy na popularny test  $t$ -Studenta, potrzebę łącznego szacowania oczekiwanej stopy zwrotu i ryzyka portfela, szacowania składników mieszaniny generujących obserwacje w analizie skupisk. Głębia Studenta odznacza się dobrymi własnościami w próbach o wielkości 30–50 obserwacji. Ryciny 15–18 przedstawiają

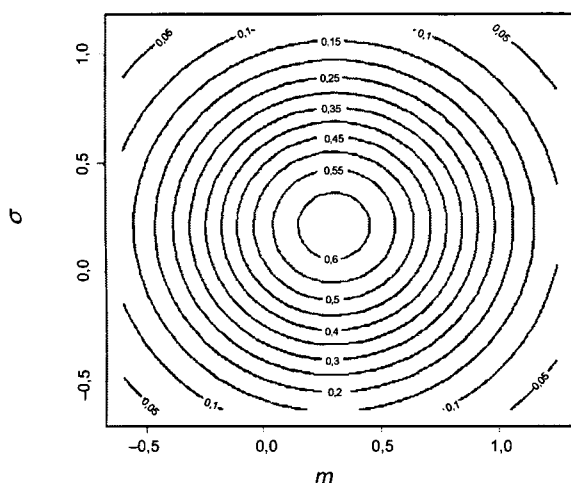


Ryc. 15. Oszacowanie gęstości mediany Studenta z próby  $N(0,1) \times 80\% + N(5,2) \times 20\%$

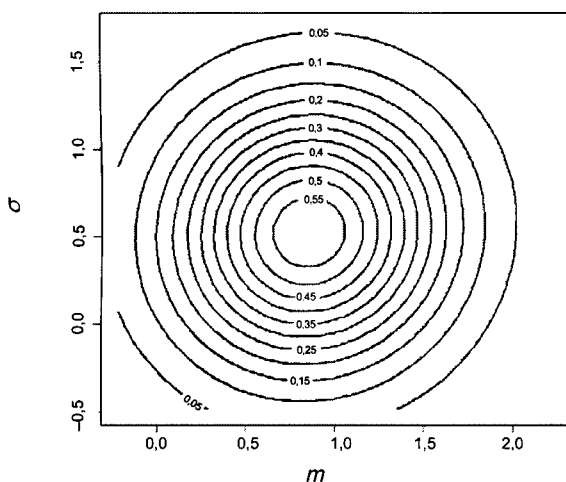


Ryc. 16. Oszacowanie gęstości mediany Studenta z próby  $t(3,-5) \times 15\% + t(3) \times 70\% + t(3,5) \times 15\%$





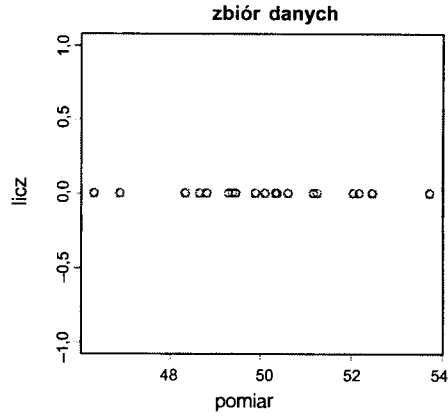
Ryc. 17. Oszacowanie gęstości mediany Studenta z próby rozkładu wykładniczego  $\lambda = 2$



Ryc. 18. Oszacowanie gęstości mediany Studenta z próby rozkładu log normalnego (0,1)

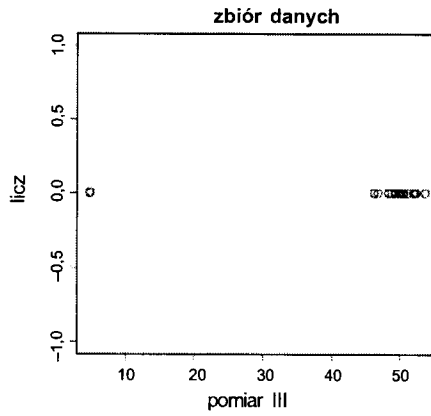
oszacowanie jądrowe gęstości mediany Studenta wykonane na podstawie 50 elementowych prób z rozkładów odpowiednio: mieszaniny  $N(0,1) \times 80\% + N(5,2) \times 20\%$ , mieszaniny  $t(3,-5) \times 15\% + t(3) \times 70\% + t(3,5) \times 15\%$ , rozkładu wykładniczego  $\lambda = 2$  i log normalnego o parametrach 0 i 1. Oszacowania wskazują, że mediana Studenta dobrze wskazuje centrum i rozrzut głównych składników mieszanin, w przypadkach rozkładów o asymetrycznych gęstościach położenie mediany Studenta jest bliższe modalnej, rozrzut mediany Studenta przyjmuje mniejszą wartość niż MAD.

Ryciny 19–21 przedstawiają trzy zbiory danych zaczerpnięte z książki Jurckowa i Picek (2006). Na rycinie 19 nie występują obserwacje odstające, na



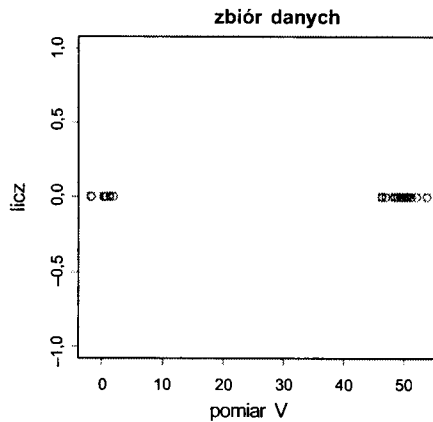
Źródło: Jureckowa i Picek (2006).

Ryc. 19. Przykładowy zbiór danych bez jednostek odstających



Źródło: Jureckowa i Picek (2006).

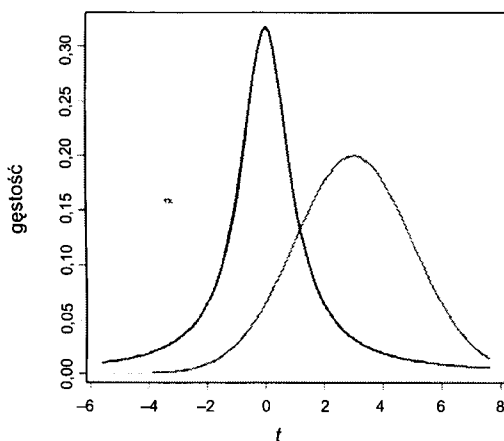
Ryc. 20. Przykładowy zbiór danych z jednostką odstającą



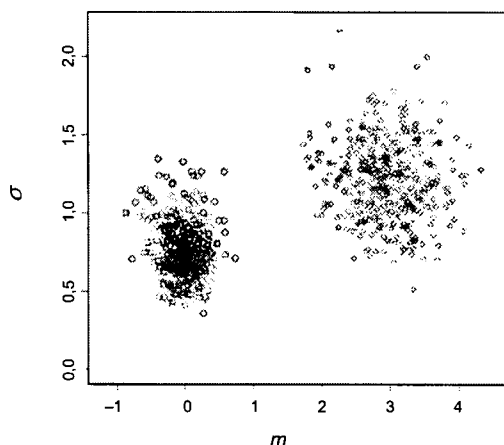
Źródło: Jureckowa i Picek (2006).

Ryc. 21. Przykładowy zbiór danych ze skupiskiem odstających

rycynie 20 dostrzegamy pojedynczą obserwację odstającą, na rycynie 21 łatwo zauważyć skupisko obserwacji odstających. Wykorzystując odpowiednie zbiory danych w celu skonstruowania 95% przedziałów ufności dla wartości oczekiwanych otrzymamy odpowiednio (49,19; 50,89), (43,045; 52,664), (27,195; 47,763) w przypadku zastosowania znanej formuły  $x \pm t_{\alpha/2}^S / \sqrt{n-1}$ , oraz otrzymamy przedziały (49,36; 50,28), (49,401; 50,494), (48,629; 50,188), stosując za każdym razem medianę Studenta, tzn. współrzędną położenia zamiast średniej, a współrzędną rozrzutu zamiast odchylenia standardowego. Porównanie przedziałów wskazuje na przewagę zastosowania mediany Studenta w przypadku występowania pośród danych jednostek odstających. Na rycynie 22 przedstawiono gęstości rozkładu  $t$ -Studenta o dwóch stopniach swobody i rozkładu  $N(3,2)$ . Na rycynie 23 pokazano mediany Studenta z prób 30-elementowych



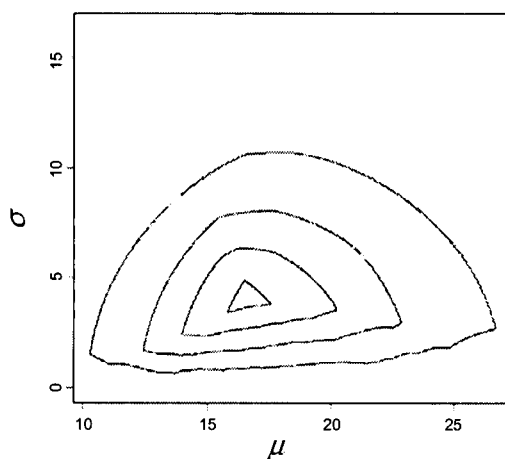
Ryc. 22. Gęstość rozkładu  $t(2)$  i  $N(3,2)$



Ryc. 23. Mediana Studenta z próby z  $t(2)$  — lewa strona i z  $N(3,2)$  prawa strona wykresu

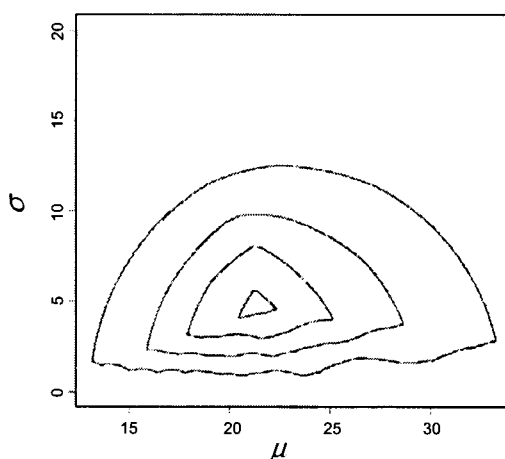
z tych rozkładów. Rycina 23 sugeruje spory potencjał mediany Studenta w kontekście proponowania testów istotności równości dwóch rozkładów.

Na rycinach 24–27 przedstawiono wykresy konturowe gębi Studenta dla stopy bezrobocia i przeciętnego wynagrodzenia brutto w powiatach Polski w latach 2000 i 2005. W przypadku stopy bezrobocia w roku 2000 mamy oszacowanie położenia i rozrzutu za pomocą mediany Studenta jako StudentMed = (16,6; 3,97), wobec oszacowania za pomocą średniej arytmetycznej i odchylenia standardowego  $\bar{x} = 17,75$ ,  $sd = 6,55$ . Sytuacja w roku 2005 przedstawia się jako StudentMed = (21,46; 4,89), wobec  $\bar{x} = 22,42$ ,  $sd = 7,72$ . W przypadku przeciętnego wynagrodzenia brutto w roku 2000 mamy StudentMed = (1613,5; 99,7), wobec  $\bar{x} = 1706,16$ ,  $sd = 251,2$ .



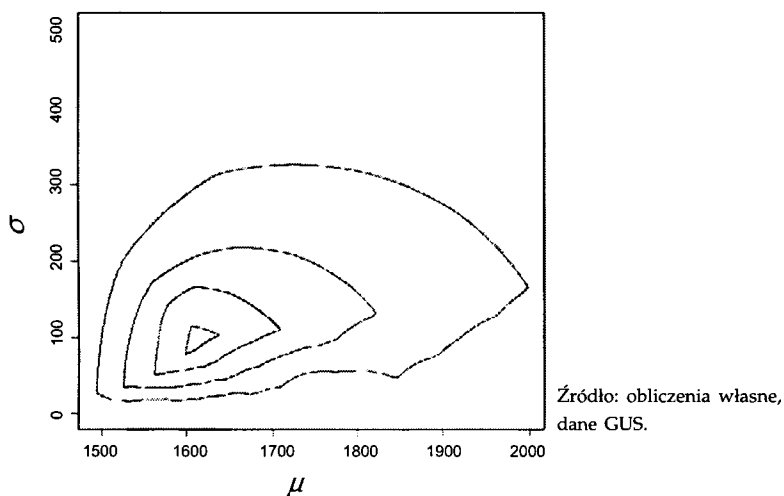
Źródło: obliczenia własne, dane GUS.

Ryc. 24. Kontury gębi Studenta — stopa bezrobocia w polskich powiatach w roku 2000

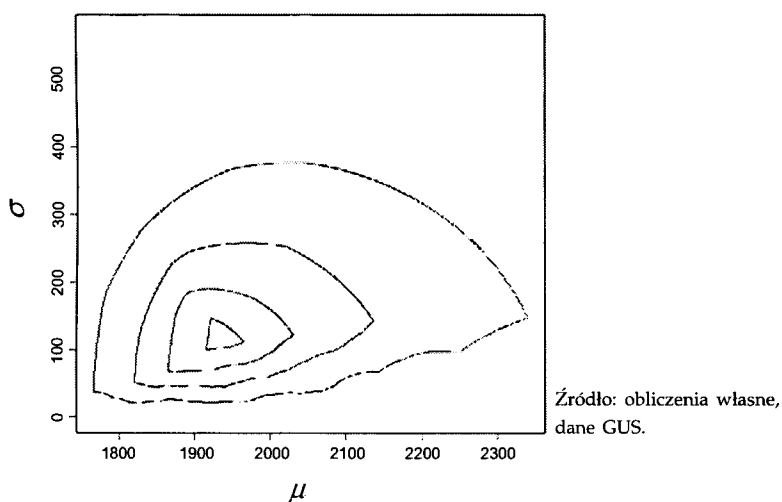


Źródło: obliczenia własne, dane GUS.

Ryc. 25. Kontury gębi Studenta — stopa bezrobocia w polskich powiatach w roku 2005



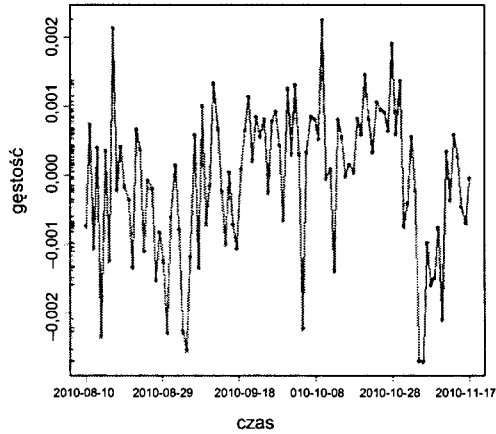
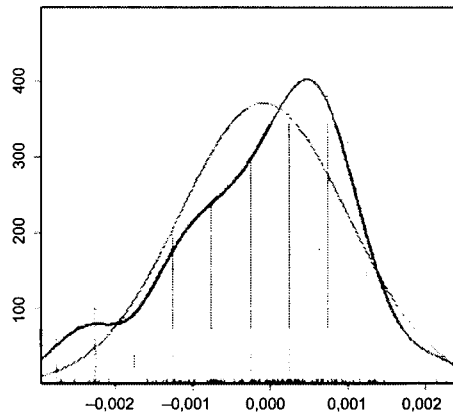
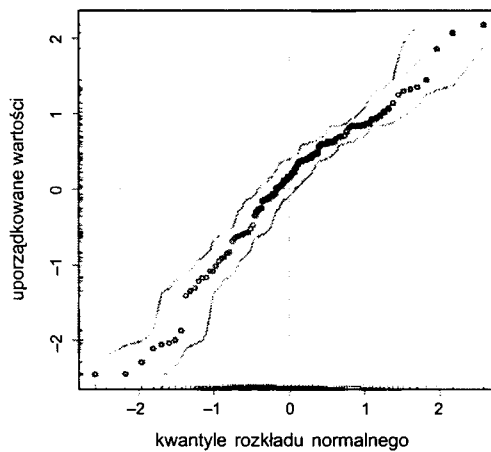
Ryc. 26. Kontury głębi Studenta — przeciętne wynagrodzenie w polskich powiatach w roku 2000

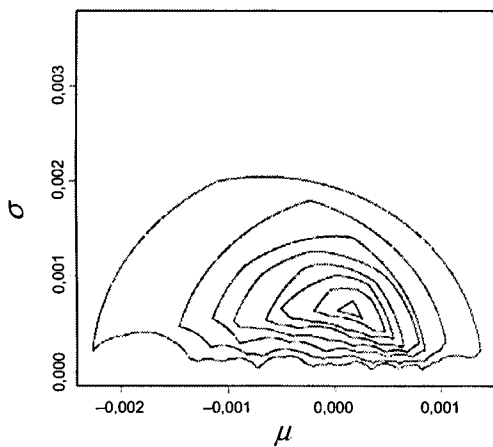


Ryc. 27. Kontury głębi Studenta — przeciętne wynagrodzenie w polskich powiatach w roku 2005

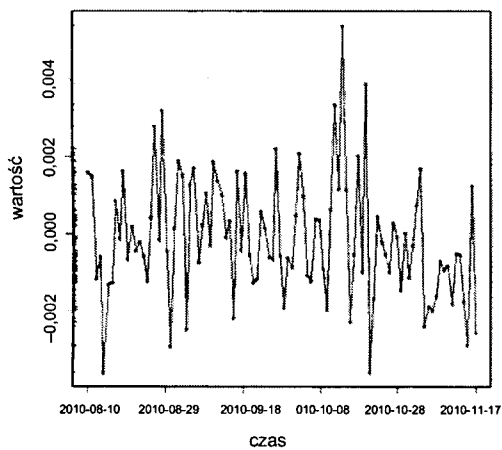
Dla tej samej cechy w roku 2005 StudentMed = (1931,43; 112,85), wobec  $\bar{x} = 2026,36$ ,  $sd = 301$ . Łatwo zauważyć niższe wartości wskaźnika położenia i rozrzutu za pomocą mediany Studenta w porównaniu ze średnią i odchyleniem standardowym.

Ryciny 28–31 przedstawiają odpowiednio sto obserwacji wygenerowanych z modelu AR(1) o parametrze  $\phi_1 = 0,5$ , wykres oszacowanej funkcji gęstości dla tych obserwacji, wykres kwantyl–kwantyl i konturowy wykres głębi Studenta. Konturowy wykres głębi Studenta sugeruje skośność rozkładu, jednakże podobnie jak wykresy kwanty–kwantyl i wykres gęstości wskazanie na auto-

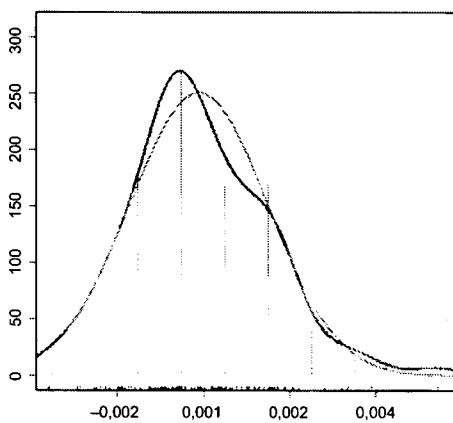
Ryc. 28. Sto obserwacji wygenerowanych z modelu AR(1):  $\phi_1 = 0,5$ Ryc. 29. Oszacowanie gęstości dla 100 obserwacji z modelu AR(1):  $\phi_1 = 0,5$ Ryc. 30. Wykres kwanty-kwanty dla obserwacji wygenerowanych z modelu AR(1):  $\phi_1 = 0,5$



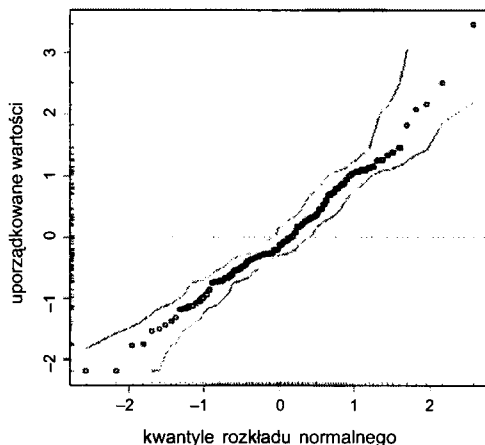
Ryc. 31. Wykres konturowy gęści Studenta dla obserwacji wygenerowanych z modelu AR(1):  $\phi_1 = 0,5$



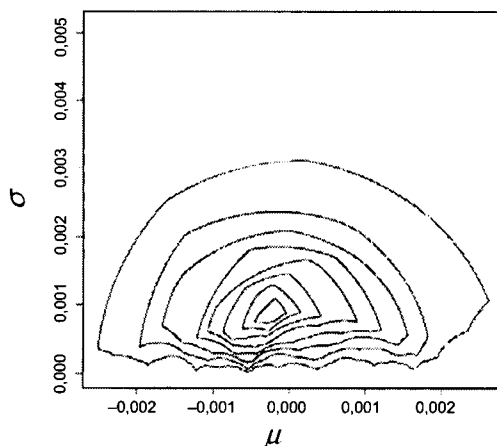
Ryc. 32. Sto obserwacji wygenerowanych z modelu GARCH(2,1) :  $\alpha_1 = 0,2$ ,  $\alpha_2 = 0,4$ ,  $\beta_1 = -0,6$



Ryc. 33. Oszacowanie gęści dla 100 obserwacji z modelu GARCH(2,1)



Ryc. 34. Wykres kwanty–kwanty dla obserwacji wygenerowanych z modelu GARCH(2,1)

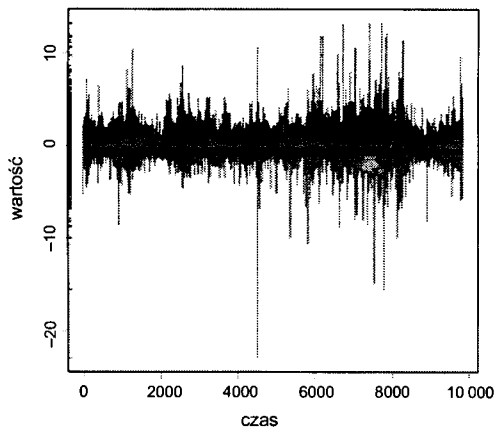


Ryc. 35. Wykres konturowy głębi Studenta dla obserwacji wygenerowanych z modelu GARCH(2,1) :  $\alpha_1 = 0,2$ ,  $\alpha_2 = 0,4$ ,  $\beta_1 = -0,6$

korelację nie jest wyraźne. Nieco lepiej przedstawia się sytuacja w przypadku modelu GARCH(2,1). Ryciny 32–35 przedstawiają analogiczne wykresy co ryciny 28–31 w przypadku modelu AR(1). Jednakże i w tym przypadku manifestacja efektu GARCH (ogólniej braku niezależności obserwacji, np. reszt regresji) w postaci wykresu konturowego głębi Studenta nie jest jednoznaczna. W kontekście dalszych studiów zagadnienia wydaje się zasadnym rozważenie wykresu ruchomej mediany Studenta w charakterze narzędzia wskazującego na występowanie efektu GARCH.

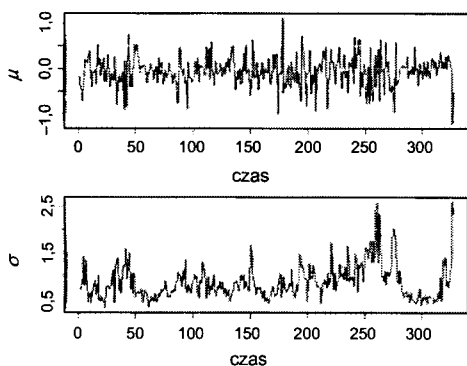
Rycina 36 przedstawia wykres dziennych prostych stóp zwrotu z akcji spółki IBM 01.02.1970–31.12.2008 roku, dane zaczerpnięto ze strony internetowej klasycznego podręcznika analizy finansowych szeregów czasowych Tsay (2010).





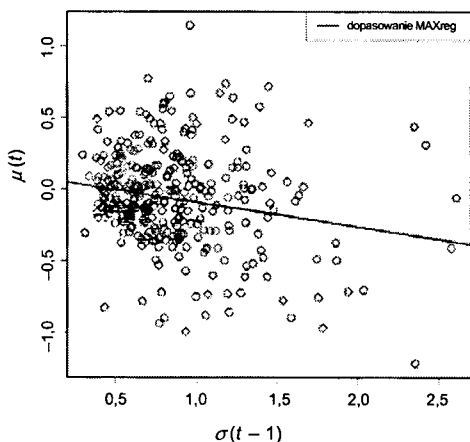
Źródło: obliczenia własne,  
dane Tsay (2010).

Ryc. 36. Dziennie proste stopy zwrotu z akcji spółki IBM 01.02.1970–31.12.2008 roku



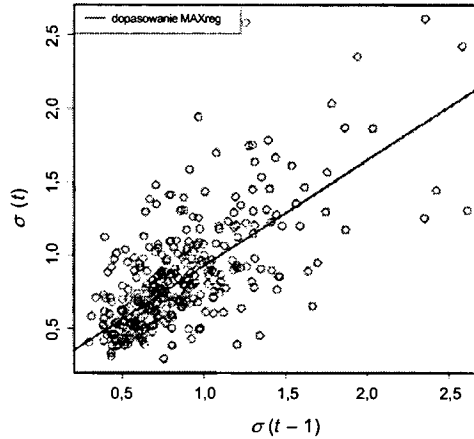
Źródło: obliczenia własne,  
dane Tsay (2010).

Ryc. 37. Miesięczne współrzędne położenia i rozrzutu mediany Studenta dla dziennych prostych stóp zwrotu z akcji spółki IBM 01.02.1970–31.12.2008 roku



Źródło: obliczenia własne,  
dane Tsay (2010).

Ryc. 38. Diagram rozrzutu miesięczne współrzędne położenia w chwili  $t$  i rozrzutu w chwili  $t - 1$  mediany Studenta dla dziennych prostych stóp zwrotu z akcji spółki IBM

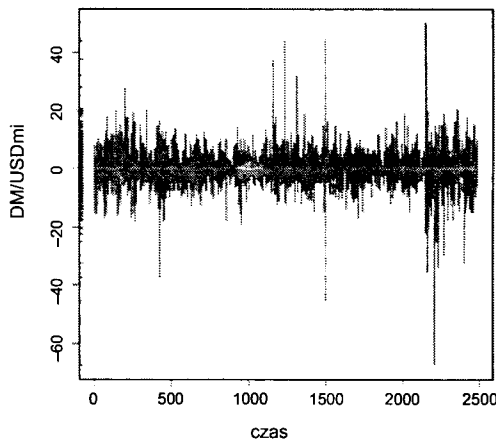


Źródło: obliczenia własne,  
dane Tsay (2010).

Ryc. 39. Diagram rozrzutu miesięczne współrzędne rozrzutu w chwili  $t$  i rozrzutu w chwili  $t - 1$  mediany Studenta dla dziennych prostych stóp zwrotu z akcji spółki IBM

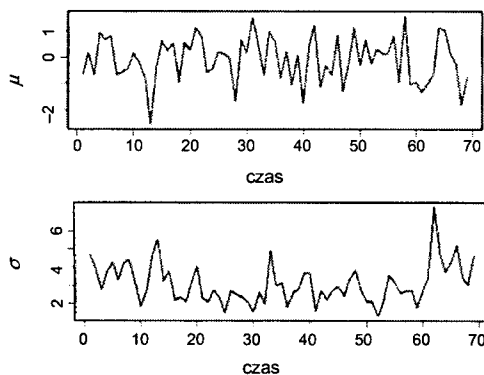
Opierając się na tym szeregu policzono miesięczne mediany Studenta, które przedstawiono na rycinie 37. Na rycinie 38 przedstawiono diagram rozrzutu współrzędna położenia mediany Studenta 30-dniowej w okresie  $t$  względem współrzędnej rozrzutu w okresie  $t - 1$  wraz z naniesionym dopasowaniem maksymalnej głębi regresyjnej. Z ryciny wynika istnienie odwrotnego związku pomiędzy średnią stopą zwrotu a zmiennością stopy zwrotu. Z ryciny 39 wynika natomiast istnienie dodatniego związku pomiędzy zmiennością stopy zwrotu w chwili  $t$  a zmiennością w chwili  $t - 1$ . Wniosek opieramy na postaci dopasowania współrzędnych rozrzutu median Studenta w chwili  $t$  i  $t - 1$ .

Rycina 40 przedstawia wykres 10-minutowych logarytmów stóp zwrotu, związanych z kursem DM/USD w czerwcu 1989, dane zaczerpnięto ze strony



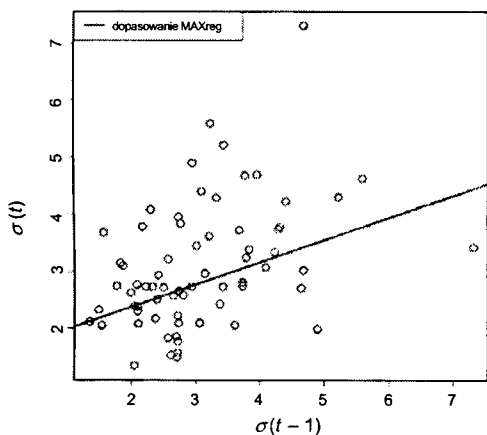
Źródło: obliczenia własne,  
dane Tsay (2010).

Ryc. 40. Logarytmy 10-minutowych stóp zwrotu związane z kursem DM/USD



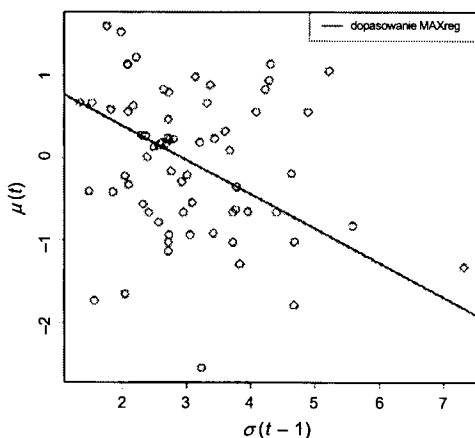
Źródło: obliczenia własne,  
dane Tsay (2010).

Ryc. 41. 6-godzinne współrzędne położenia i rozrzutu mediany Studenta dla logarytmów stóp zwrotu, związanych z kursem DM/USD



Źródło: obliczenia własne,  
dane Tsay (2010).

Ryc. 42. Diagram rozrzutu 6-godzinne współrzędne rozrzutu w chwili  $t$  i rozrzutu w chwili  $t - 1$  mediany Studenta dla logarytmów stóp zwrotu DM/USD



Źródło: obliczenia własne,  
dane Tsay (2010).

Ryc. 43. Diagram rozrzutu 6-godzinne współrzędne położenia w chwili  $t$  i rozrzutu w chwili  $t - 1$  mediany Studenta dla logarytmów stóp zwrotu DM/USD

internetowej klasycznego podręcznika analizy finansowych szeregów czasowych Tsay (2010). Opierając się na tym szeregu policzono miesięczne mediany Studenta, które przedstawiono na rycinie 41. Na rycinie 42 przedstawiono diagram rozrzutu współrzędna położenia mediany Studenta 6-godzinnej w okresie  $t$  względem współrzędnej rozrzutu w okresie  $t - 1$  wraz z naniesionym dopasowaniem maksymalnej głębi regresyjnej. Z ryciny wynika istnienie odwrotnego związku pomiędzy średnią stopą zwrotu a zmiennością stopy zwrotu. Z ryciny 43 wynika natomiast istnienie dodatniego związku pomiędzy zmiennością stopy zwrotu w chwili  $t$  a zmiennością w chwili  $t - 1$ . Wniosek opieramy na postaci dopasowania współrzędnych rozrzutu median Studenta w chwili  $t$  i  $t - 1$ .

## 5. PODSUMOWANIE

Procedury statystyczne indukowane przez statystyczne funkcje głębi cechuje odporność przy jednocześnie zadowalającej efektywności. Jednolite spojrzenie Mizery i Mizery i Mullera na funkcje głębi, wywodzące się od głębi domkniętej półprzestrzeni, z jednej strony pozwala dostrzec związki koncepcji z dorobkiem klasycznej statystyki matematycznej, z drugiej strony — zaproponować klasę parametrycznych funkcji głębi wywodzących się z zasady największej wiarygodności. Przykładem tej klasy parametrycznych funkcji głębi jest głębia Studenta. Głębia Studenta jest wrażliwa na typ rozkładu generującego obserwację, co daje nadzieję na jej wykorzystanie w weryfikacji hipotez dotyczących rozkładu. Estymator maksymalnej głębi Studenta — mediana Studenta odznacza się bardzo dobrymi statystycznymi własnościami. Mamy tutaj na uwadze szybkość zbieżności z próby, wysoki punkt załamania próby skończonej, zadowalającą efektywność. Mediana Studenta może zostać z powodzeniem wykorzystana w procesie weryfikacji hipotez dotyczących równości rozkładów, w analizie skupisk, wstępnej analizie szeregów czasowych. Warto podkreślić, że estymator ten odznacza się dobrymi własnościami w próbie o wielkości trzydziestu obserwacji. Fakt ten pozwala mieć nadzieję na szereg interesujących zastosowań głębi Studenta w odpornej analizie statystycznej.

## BIBLIOGRAFIA

- Davies P.L. 2002. *Statistical Procedures and Robust Statistics*. <http://wwwstat.mathematik.uni-essen.de/~davies/statproc.ps.gz>.
- Davies P.L., Gather U. 2005. *Breakdown and Groups (with discussion and rejoinder)*. *The Annals of Statistics*, 33, 977–1035.
- Donoho D.L., Gasko M. 1992. *Breakdown Properties of Location Estimates Based on Halfspace Depth and Projected Outlyingness*. *Annals of Statistics* 20, 1803–1827.

- Donoho D.L., Huber P.J. 1983. *The Notion of Breakdown Point*, [w:] *A Festschrift for Erich Lehmann*, red. P.J. Bickel, K.A. Doksum, J.L. Hodges, Wadsworth, Belmont, CA, 157–184.
- Genton M.G., Lucas A. 2003. *Comprehensive Definitions of Breakdown Points for Independent and Dependent Observations*, Journal of the Royal Statistical Society, Series B 65(1), 81–84.
- Hampel F.R., Ronchetti E.M., Rousseeuw P.J., Stahel W.A. 1986. *Robust Statistics: The approach based on influence functions*, New York.
- Huber P., Ronchetti E.M. 2009. *Robust Statistics*, New York.
- Jureckova J., Picek J. 2006. *Robust Statistical Methods with R*. Boca Raton, Chapman & Hall/CRC.
- Kosiorowski D. 2007. *O odpornej analizie regresji w ekonomii na przykładzie koncepcji głębi regresyjnej*, Przegląd Statystyczny 1, 109–121.
- Kosiorowski D. 2009. *Wybrane zagadnienia koncepcji głębi danych*, Folia Oeconomica Cracoviensia 49, 5–30.
- Kosiorowski D. 2010a. *Depth Based Procedures for Estimation ARMA and GARCH Models*, [w:] *Proceedings of COMPSTAT 2010*, red. Y. Lechevallier, G. Saporta, 19<sup>th</sup> International Conference on Computational Statistics, Physica-Verlag, Heidelberg, 1207–1214.
- Kosiorowski D. 2010b. *Wybrane zastosowania uogólnionej głębi Tukey'a w odpornej analizie ekonomicznej*, Konferencja Statystyka Matematyczna, Wisła 2010.
- Liu R.Y., Parelius J.M., Singh K. 1999. *Multivariate Analysis by Data Depth: Descriptive Statistics, Graphics and Inference (with discussion)*, The Annals of Statistics 27, 783–858.
- Maronna R.A., Martin R.D., Yohai V.J. 2006. *Robust Statistics — Theory and Methods*, Chichester.
- Mizera I. 2002. *On Depth and Depth Points: a Calculus*, The Annals of Statistics 30, 1681–1736.
- Mizera I., Muller C.H. 2004. *Location — Scale Depth (with Discussion and Rejoinder)*, Journal of the American Statistical Association 99(4), 981–989.
- Rousseeuw J.P., Hubert M. 1998. *Regression Depth*, Journal of The American Statistical Association 94, 388–433.
- Tsay R.S. 2010. *Analysis of Financial Time Series*, Wiley-Interscience, Hoboken, New-Yersey.
- Tukey J. 1975. *Mathematics and Picturing Data*, Proceedings of the International Congress of Mathematicians, red. R. James, Canadian Math. Congress 1974, 2, 523–531.
- Zuo Y., Serfling R. 2000a. *General Notions of Statistical Depth Function*, The Annals of Statistics 28, 461–482.
- Zuo Y., Serfling R. 2000b. *Structural Properties and Convergence Results for Contours of Sample Statistical Depth Functions*, The Annals of Statistics 28, 483–499.