

Pakiet *DepthProc 1.0* w eksploracyjnej analizie wielowymiarowego szeregu czasowego

Daniel Kosiorowski¹ Mateusz Bocian² Anna Węgrzynkiewicz²
Zygmunt Zawadzki²

¹Katedra Statystyki
Uniwersytetu Ekonomicznego w Krakowie

²Studenci IE i Finansów
Uniwersytetu Ekonomicznego w Krakowie

Konferencja Wielowymiarowa Analiza Statystyczna 2012

Dziękujemy za wsparcie finansowe w postaci grantu NCN DEC-011/03/B/HS4/01138

- 1 Wprowadzenie
- 2 Procedury oferowane przez pakiet
- 3 Przykłady zastosowań pakietu

Motywacje

- Upowszechnienie na szerszą skalę wielowymiarowych procedur statystycznych wykorzystujących wielowymiarowe statystyki porządkowe, rangi i miary odstawania.
- Konieczność prowadzenia odpornej analizy statystycznej za pomocą prostych, przyjaznych dla użytkownika narzędzi statystycznych.
- Intrygujące związki pomiędzy odpornością procedury statystycznej a zachowaniami konsumentów, inwestorów, podmiotów gospodarczych – Próba odpowiedzi na pytanie – Jak rozumieć "wpływową większość" na pewnym rynku, wśród inwestorów, konsumentów...

Możliwości pakietu

- Wartości głębokości projekcyjnej, Mahalanobisa, Tukeya, strefowej (ang. zonoid depth), regresyjnej, głębokości Lopez-Pintado & Romo i Fraiman & Muniz (dla danych funkcjonalnych), niebawem głębokości położenia-rozrzutu Mizery & Müller.
- Dobrej jakości wykresy konturowe i perspektywiczne.
- Wykresy głębokości vs. głębokości (uogólnienie wykresu kwantyl – kwantyl).
- Krzywa rozrzutu.
- Krzywa asymetrii.
- Estymator maksymalnej głębokości regresyjnej.

Przykładowe komendy

- `depthContour(X,method = "Projection",plot.title = paste(method,"depth"),...)`
- `depthPersp(X,method = "Projection",plot.method = "rgl",xlim = extendrange(X[,1],f=0.1),...)`
- `ddPlot(x, y=NULL, distribution = c("mvnorm", "t", "smvnorm", "st"), method = "Projection", scale = FALSE,...)`
- `ddmvnorm(x, method = "Projection", robust=FALSE,...)`
- `scalecurve(X, Y = NULL, alpha = seq(0,1,0.01), method = "Projection", draw = TRUE, nameX = "X", nameY = "Y", ...)`
- `asymmetrycurvej-function(X,Y = NULL, alpha = seq(0,1,0.01), method = "Projection", movingmedian = FALSE,draw = TRUE,nameX = "X", nameY = "Y",...)`
- `deepreg2d(...)`
- `trimprojreg2d(...)`

Pakiet korzysta z pakietów `{rgl}`, `{geometry}`, `{ggplot2}`, `{lattice}`, `{MNM}`, `{sn}`, `{MASS}`, `{robustbase}`.

Co to jest koncepcja głębi danych

- W przypadku jednowymiarowym dysponujemy liniowym porządkiem obserwacji, które przyjmują wartości w zbiorze liczb rzeczywistych. Posługując się tym porządkiem definiujemy statystyki porządkowe oraz rangi obserwacji. Zaznaczmy, że ów porządek obserwacji indukowany przez zbiór liczb rzeczywistych nie ma naturalnego uogólnienia na \mathbb{R}^d , $d \geq 2$.
- **Koncepcja głębi danych** uogólnia na przypadek wielowymiarowy jednowymiarowe metody statystyczne wykorzystujące statystyki porządkowe, kwantyle, rangowanie i miary odstawania (patrz Zuo i Serfling, 2000). Podkreślmy, że mamy tu na uwadze wielowymiarowe statystyki porządkowe (aby uwzględnić wielowymiarową geometrię zbioru danych rozważa się wielowymiarową medianę zamiast wektora jednowymiarowych median).

Co to jest koncepcja głębi danych cd

- Za źródło inspiracji prowadzących do powstania koncepcji głębi danych można przyjąć prace Hotellinga dotyczące ekonomicznej teorii gier oraz test znaków zaproponowany przez Hodgesa (patrz Serfling, 2006). Burzliwy rozwój koncepcji został zainicjowany przez wybitnego matematyka i statystyka amerykańskiego Johna W. Tukey'a, który na kongresie ICM w 1974 r. postawił problem obisu wielowymiarowej chmury danych za pomocą wielowymiarowych statystyk porządkowych.
- Tukey (por. Tukey, 1975) zaproponował *głębnię domkniętej półprzestrzeni* do generowania konturów w wielu wymiarach, które miały spełniać rolę analogiczną do rang i statystyk porządkowych w przypadku jednowymiarowym.

Co to jest koncepcja głębi danych cd

- Statystyczna funkcja głębi związana z danym rozkładem prawdopodobieństwa P na \mathbb{R}^d umożliwia porządkowanie punktów $\mathbf{x} \in \mathbb{R}^d$ na zasadzie odstawania od centrum rozkładu P .

Co to jest koncepcja głębi danych cd

- Statystyczna funkcja głębi związana z danym rozkładem prawdopodobieństwa P na \mathbb{R}^d umożliwia porządkowanie punktów $\mathbf{x} \in \mathbb{R}^d$ na zasadzie odstawania od centrum rozkładu P .
- W przypadku punktów położonych w pobliżu centrum rozkładu funkcja głębi przyjmuje wartości bliskie jedności, wartości funkcji głębi bliskie zeru odpowiadają peryferiom rozkładu.

Co to jest koncepcja głębi danych cd

- Statystyczna funkcja głębi związana z danym rozkładem prawdopodobieństwa P na \mathbb{R}^d umożliwia porządkowanie punktów $\mathbf{x} \in \mathbb{R}^d$ na zasadzie odstawania od centrum rozkładu P .
- W przypadku punktów położonych w pobliżu centrum rozkładu funkcja głębi przyjmuje wartości bliskie jedności, wartości funkcji głębi bliskie zeru odpowiadają peryferiom rozkładu.
- Wykorzystując funkcję głębi można określić centrum rozkładu jako zbiór punktów, które globalnie maksymalizują wartości funkcji głębi. Wielomodalność rozkładu P na ogół jest pomijana.

Co to jest koncepcja głębi danych cd

- Statystyczna funkcja głębi związana z danym rozkładem prawdopodobieństwa P na \mathbb{R}^d umożliwia porządkowanie punktów $\mathbf{x} \in \mathbb{R}^d$ na zasadzie odstawania od centrum rozkładu P .
- W przypadku punktów położonych w pobliżu centrum rozkładu funkcja głębi przyjmuje wartości bliskie jedności, wartości funkcji głębi bliskie zeru odpowiadają peryferiom rozkładu.
- Wykorzystując funkcję głębi można określić centrum rozkładu jako zbiór punktów, które globalnie maksymalizują wartości funkcji głębi. Wielomodalność rozkładu P na ogół jest pomijana.
- Punkt (zbiór punktów), w którym funkcja głębi przyjmuje wartość maksymalną określa się mianem wielowymiarowej mediany indukowanej przez tę funkcję głębi.

Co to jest koncepcja głębi danych cd

- Statystyczna funkcja głębi związana z danym rozkładem prawdopodobieństwa P na \mathbb{R}^d umożliwia porządkowanie punktów $\mathbf{x} \in \mathbb{R}^d$ na zasadzie odstawania od centrum rozkładu P .
- W przypadku punktów położonych w pobliżu centrum rozkładu funkcja głębi przyjmuje wartości bliskie jedności, wartości funkcji głębi bliskie zeru odpowiadają peryferiom rozkładu.
- Wykorzystując funkcję głębi można określić centrum rozkładu jako zbiór punktów, które globalnie maksymalizują wartości funkcji głębi. Wielomodalność rozkładu P na ogół jest pomijana.
- Punkt (zbiór punktów), w którym funkcja głębi przyjmuje wartość maksymalną określa się mianem wielowymiarowej mediany indukowanej przez tę funkcję głębi.
- Można rozważać głębie na zbiorach liczbowych, na przestrzeniach funkcji, pierścieniach macierzy itd.

statystyczna funkcja głębi

DEFINICJA (Zuo i Serfling, 2000): Niech \mathcal{P} oznacza rodzinę rozkładów prawdopodobieństwa na \mathbb{R}^d . Odwzorowanie $D(\cdot, \cdot) : \mathbb{R}^d \times \mathcal{P} \rightarrow [0, 1]$, które jest ograniczone, nieujemne oraz spełnia poniższe warunki ZS1 – ZS4:

- ZS1: $D(\mathbf{Ax} + \mathbf{b}, P_{\mathbf{Ax}+\mathbf{b}}) = D(\mathbf{x}, P_{\mathbf{x}})$ dla dowolnego wektora losowego \mathbf{X} o wartościach w \mathbb{R}^d , dowolnych $d \times d$ nieosobliwej macierzy \mathbf{A} , i d wektora \mathbf{b} .
- ZS2: $D(\mathbf{m}, P) = \sup_{\mathbf{x} \in \mathbb{R}^d} D(\mathbf{x}, P)$ zachodzi dla każdego $P \in \mathcal{P}$ mającego centrum \mathbf{m} .
- ZS3: Dla dowolnego $P \in \mathcal{P}$ mającego punkt o największej głębi \mathbf{m} , zachodzi $D(\mathbf{x}, P) \leq D(\mathbf{m} + \alpha(\mathbf{x} - \mathbf{m}), P)$, $\alpha \in [0, 1]$.
- ZS4: $D(\mathbf{x}, P) \rightarrow 0$, gdy $\|\mathbf{x}\| \rightarrow \infty$ dla każdego $\mathbf{x} \in \mathbb{R}^d$.

nazywamy statystyczną funkcją głębi.

Co możemy policzyć dysponując funkcją głębi

- Kontury równego odstawania - zbiory dla których funkcja głębi przyjmuje stałą wartość.
- Indukowane przez głębie statystyki porządkowe. Możemy porządkować dane ze względu na wartość głębi - względem centrum.
- Ważone głębią funkcjonały położenia.
- Ważone głębią macierze rozrzutu.
- Krzywe rozrzutu - objętość obszaru centralnego względem $\alpha \in [0, 1]$ (parametr odstawania od centrum).
- Funkcjonały skośności. Skalowana różnica pomiędzy dwoma funkcjonałami położenia.
- Funkcjonały kurtozy.

Procedury statystyczne indukowane przez głębie

- Uogólnienia wykresów ramka wąsy - np. dla danych funkcjonalnych.
- Porównanie dwóch prób (populacji) za pomocą rysunku 2D - uogólnienia wykresu kwantyl - kwantyl.
- Nieparametryczny opis wielowymiarowego rozkładu. Alternatywa dla dystrybuanty.
- Testy wielowymiarowej symetrii, testy normalności.
- Procedury statystycznej kontroli jakości.
- Estymacja wielowymiarowej gęstości prawdopodobieństwa.
- Klasyfikacja obiektów za pomocą głębi.
- Analiza skupisk za pomocą głębi.
- Odporne rozpoznawanie obrazów.

Przykłady statystycznych funkcji głębi

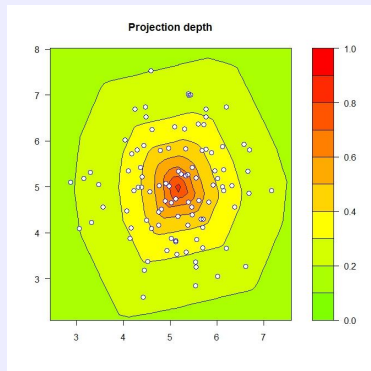
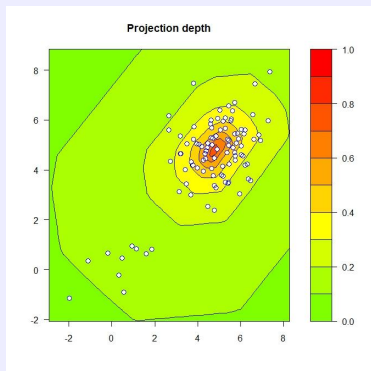
Symetryczna głębia projekcyjna punktu $\mathbf{x} \in \mathbb{R}^d$ wygenerowanego przez d wymiarowy wektor losowy \mathbf{X} o rozkładzie F , $D(\mathbf{x}, F)$ definiowana jest jako

$$D(\mathbf{x}, F) = \left[1 + \sup_{\|\mathbf{u}\|=1} \frac{|\mathbf{u}^T \mathbf{x} - \text{Med}(\mathbf{u}^T \mathbf{X})|}{\text{MAD}(\mathbf{u}^T \mathbf{X})} \right]^{-1}, \quad (1)$$

gdzie Med oznacza jednowymiarową medianę, $\text{MAD}(Z) = \text{Med}(|Z - \text{Med}(Z)|)$, wersję z próby oznaczaną $D(\mathbf{x}, F_n)$ lub $PD(\mathbf{x}, \mathbf{X}^n)$ uzyskamy zastępując rozkład F jego empirycznym odpowiednikiem F_n obliczanym na podstawie próby \mathbf{X}^n .

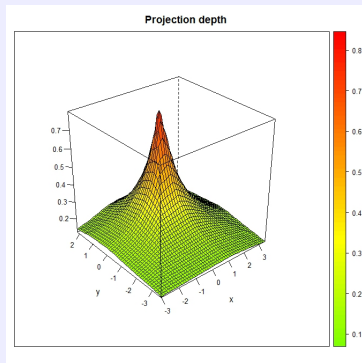
To jedna z najlepszych znanych obecnie głębi.

Przykłady statystycznych funkcji głębi cd

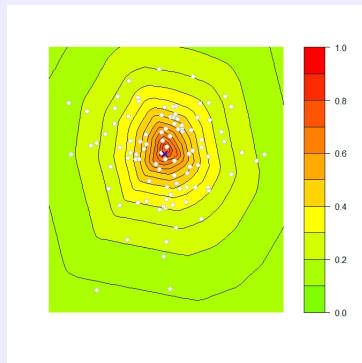


Rysunek : Wykres konturowy dla głębi projekcyjnej z próby w przypadku mieszania dwóch rozkładów normalnych 2D (lewa strona) i rozkładu normalnego 2D (prawa strona).

Wykresy perspektywiczny i konturowy.



Rysunek : Głębina projekcyjna z próby
- wykres perspektywiczny.



Rysunek : Głębina projekcyjna z próby
- wykres konturowy.

- Bezpośrednie obliczanie wartości funkcji gęblei jest w wielu przypadkach zadaniem niezmiernie złożonym pod względem obliczeniowym. Wobec zapotrzebowania na procedury statystyczne indukowane przez funkcje gęblei szczególnego znaczenia nabierają propozycje przybliżonego obliczania wartości funkcji gęblei z próby.
- Pakiet DepthProc wykorzystuje algorytm Dyckerhoffa (2004) do przybliżonego obliczania gęblei położenia, bezpośredni algorytm Rousseeuw i Hubert (1998) do obliczania estymatora maksymalnej gęblei regresyjnej 2D, definicję Marony i in. (2006) gęblei regresyjnej regresji wielorakiej, bezpośredni algorytm Lopez-Pintado i Romo (2009) gęblei dla danych funkcjonalnych.
- Należy zauważyć całkowicie nowe propozycje przybliżonego obliczania gęblei przedstawiane przez Zuo (2012) odwołujące się do MCMC oraz propozycje Atchison i in. (2012) opierające się o zasadę "forward search" (por Hastie i in. 2009). Zagadnienie jest przedmiotem naszych intensywnych studiów.

przybliżone obliczanie wartości głębi

- Podejście Dyckerhoffa można wyrazić w uproszczeniu w następujący sposób: **jeżeli przyjmujemy, że punkt jest centralny względem pewnego wielowymiarowego rozkładu prawdopodobieństwa, gdy wszystkie jego jednowymiarowe projekcje są centralne względem jednowymiarowej projekcji rozkładu to możemy definiować wielowymiarowe głębie punktu jako minimum jednowymiarowych głębi wszystkich takich jednowymiarowych projekcji.**
- Dysponując zatem jednowymiarową głębią możemy aproksymować wielowymiarową głębią losując pewną liczbę punktów z d -wymiarowej sfery i wykorzystując twierdzenie Dyckerhoffa (2004). Podejście takie można zastosować do pewnej klasy funkcji głębi odznaczających się tzw. mocną własnością projekcji (stosujemy je dla głębi projekcyjnej, Tukey'a, zonoidalnej).

szczegóły naszego podejścia

Niech $D(\mathbf{x}, \mathbf{Z}^n)$ oznacza funkcję głębzi z próby w punkcie $\mathbf{x} \in \mathbb{R}^d$, $d \geq 2$, $\mathbf{Z}^n = \{\mathbf{z}_1, \dots, \mathbf{z}_n\} \subset \mathbb{R}^d$. Przyjmijmy, że wartość głębzi punktu $\mathbf{x} \in \mathbb{R}^d$ względem $\mathbf{Z}^n = \{\mathbf{z}_1, \dots, \mathbf{z}_n\} \subset \mathbb{R}^d$ to minimalna głębza projekcji 1D punktu $\mathbf{u}^T \mathbf{x}$ względem $\mathbf{u}^T \mathbf{Z}^n = \{\mathbf{u}^T \mathbf{z}_1, \dots, \mathbf{u}^T \mathbf{z}_n\} \subset \mathbb{R}$, $\mathbf{u}^T \in \mathbb{R}^d$, $\|\mathbf{u}\| = 1$.

Jak może wyglądać taka jednowymiarowa głębza?

Oznaczmy ją $D^1(y, Y^n)$, $y \in \mathbb{R}$, $Y^n = \{y_1, \dots, y_n\}$.

- „Głębza symplecjalna”: $D^1(y, Y^n) = F_n(y)(1 - F_n(y))$,
gdzie F_n – oznacza dystrybuantę empiryczną wyznaczoną na podstawie próby $Y^n = \{y_1, \dots, y_n\}$.
- „Głębza Tukey’a”: $D^1(y, Y^n) = \min\{F_n(y), 1 - F_n(y)\}$,
gdzie F_n – oznacza dystrybuantę empiryczną wyznaczoną na podstawie próby $Y^n = \{y_1, \dots, y_n\}$.

Zatem

$$D(\mathbf{x}, \mathbf{X}^n) = \min_{\|\mathbf{u}\|=1} D^1(\mathbf{u}^T \mathbf{x}, \mathbf{u}^T \mathbf{X}^n)$$

szczegóły naszego podejścia cd

Niech $\alpha \in \{0.01, 0.02, \dots, 0.99, 1\}$, niech $Q(\alpha)$ oznacza kwantyl rzędu α

Głębina projekcyjna

$$\begin{aligned} D^1(y, Z^n) &= \\ &= \min_{\alpha} \left\{ \alpha : y \in \left[\text{Med}(Z^n) - \frac{1-\alpha}{\alpha} \text{MAD}(Z^n), \text{Med}(Z^n) + \frac{1-\alpha}{\alpha} \text{MAD}(Z^n) \right] \right\} \end{aligned} \quad (2)$$

Głębina strefowa (ang. zonoid depth)

$$D^1(y, Z^n) = \min_{\alpha} \left\{ \alpha : y \in \left[\frac{1}{\beta} \sum_{\beta \leq \alpha} Q(\beta), \frac{1}{\beta} \sum_{\beta \leq \alpha} Q(1 - \beta) \right] \right\} \quad (3)$$

wykres głębia vs. głębia

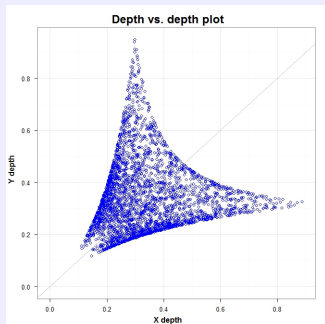
Dla dwóch rozkładów prawdopodobieństwa F i G , obydwu na \mathbb{R}^d , wykres głębia vs. głębia definiujemy

$$DD(F, G) = \{(D(\mathbf{z}, F), D(\mathbf{z}, G)), \mathbf{z} \in \mathbb{R}^d\} \quad (4)$$

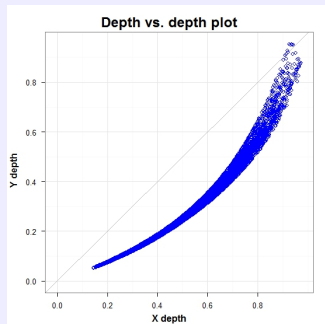
Dla dwóch prób $\mathbf{X}^n = \{X_1, \dots, X_n\}$ z F , oraz $\mathbf{Y}^m = \{Y_1, \dots, Y_m\}$ z G , wykres głębia vs. głębia definiowany jest jako

$$DD(F_n, G_m) = \{(D(\mathbf{z}, F_n), D(\mathbf{z}, G_m)), \mathbf{z} \in \{\mathbf{X}^n \cup \mathbf{Y}^m\}\} \quad (5)$$

Wykresy głębia vs. głębia



Rysunek : Wykres głębia vs. głębia - różnice co do położenia.



Rysunek : Wykres głębia vs. głębia - różnice co do rozrzutu.

głębia regresyjna

Niech $\mathbf{Z}^n = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\} \subset \mathbb{R}^d$ oznacza próbę rozpatrywaną z punktu widzenia następującego modelu semiparametrycznego:

$$y_l = a_0 + a_1 x_{1l} + \dots + a_{(d-1)l} x_{(d-1)l} + \varepsilon_l, l = 1, \dots, n,$$

obliczamy głębnię konkretnego dopasowania $\alpha = (a_0, \dots, a_{d-1})$ za pomocą

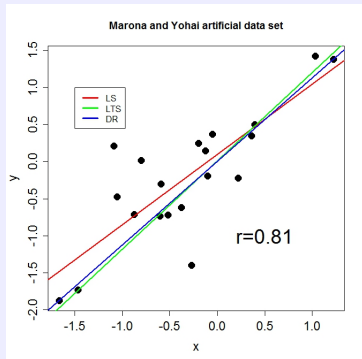
$$RD(\alpha, \mathbf{Z}^n) = \min_{\mathbf{u} \neq \mathbf{0}} \# \left\{ \frac{r_l(\alpha)}{\mathbf{u}^T \mathbf{x}_l} < 0, l = 1, \dots, n \right\}, \quad (6)$$

gdzie $r(\cdot)$ oznacza resztę regresji, $\alpha = (a_0, \dots, a_{d-1})$, $\mathbf{u}^T \mathbf{x}_l \neq 0$.

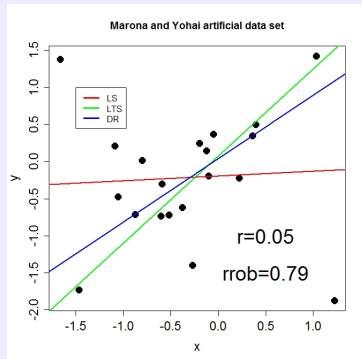
Estymator maksymalnej głębni regresyjnej (ang. deepest regression)
 $DR(\alpha, \mathbf{Z}^n)$ definiujemy

$$DR(\alpha, \mathbf{Z}^n) = \arg \max_{\alpha \neq \mathbf{0}} RD(\alpha, \mathbf{Z}^n). \quad (7)$$

estymator maksymalnej głębi regresyjnej 2D



Rysunek : Sztuczny zbiór danych Marona i Yohai'a - brak obserwacji odstających, {robustbase}.



Rysunek : Sztuczny zbiór danych Marona i Yohai'a - 10% obserwacji odstających, {robustbase}.

eksploracja wielowymiarowego szeregu czasowego

- diagramy rozrzutu obserwacja w chwili t vs. w chwili $(t-1)$, $(t-2)$,... itd. – wybór modelu procesu i jego specyfikacji.
- ruchomy wykres głębienia vs. głębienia - monitorowanie położenia, rozrzutu, skośności procesu (także ruchoma mediana, ruchoma miara rozrzutu...).
- ruchoma krzywa asymetrii.
- ruchoma krzywa rozrzutu.
- odporne szacowania charakterystyk, rozkładów warunkowych.
- przewidywanie kolejnej wartości procesu za pomocą estymatora maksymalnej głębokości regresyjnej.
- filtrowanie, wygładzanie, dekomponowanie szeregu.

przykłady zastosowań pakietu

Wygenerowano 3500 obserwacji z dwuwymiarowego procesu VAR(1).

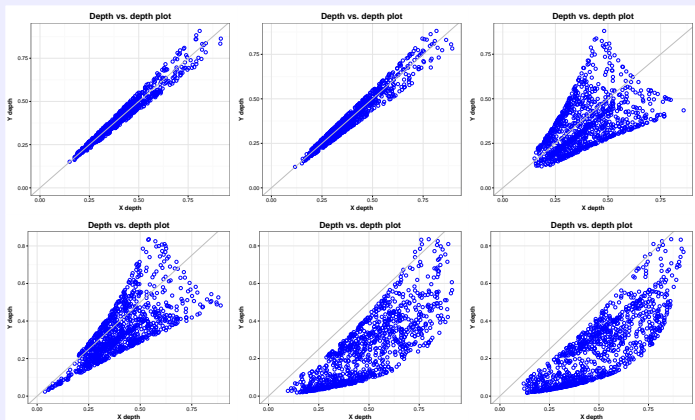
$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} 0.2 & -0.6 \\ 0.3 & 1.1 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}$$

Zaburzenie 2d generuje niezależnie izotropowy rozkład normalny 2d. Od 1401 do 2450 obserwacji nastąpiło przesunięcie danych w stosunku trajektorii wynikającej z modelu. Od 2451 do 3500 obserwacji nastąpiła zmiana rozproszenia w stosunku do trajektorii wygenerowanej przez proces VAR(1).

Szerokość okna to 500 obserwacji. Oknem referencyjnym jest pierwszy przedział, czyli obserwacje 1:500.

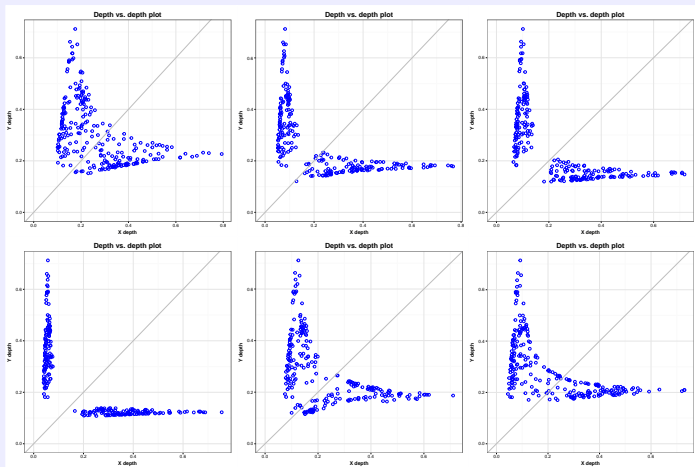
Dane empiryczne: Zbiór danych obejmuje kursy otwarcia i zamknięcia dla indeksu WIG20 od 1 stycznia 2009 (wykorzystano 861 obserwacji) do 30 czerwca 2012 roku. Na wykresach porównano kolejne półrocza. Okresem referencyjnym było pierwsze półrocze 2009 roku.

Wykresy głębia vs. głębia - symulacje



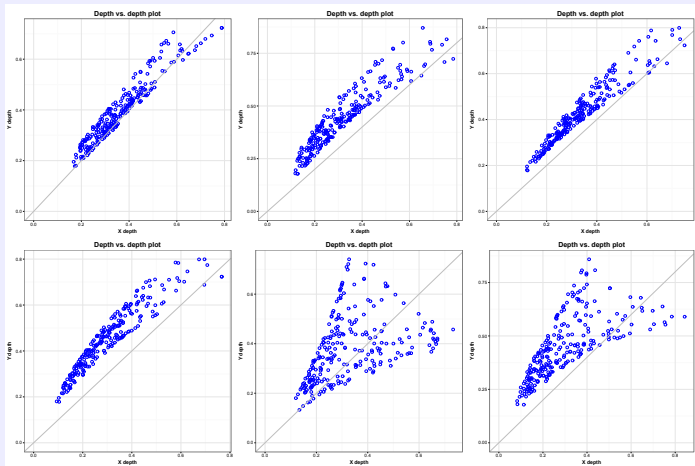
Rysunek : Pierwsze 500 obserwacji (próba x) porównujemy ją z obserwacjami 501:1000, 1001:1500, 1501:2000, 2001:2500, 2501:3000, 3001:3500 (próba y).

Wykresy głębia vs. głębia - przykład empiryczny



Rysunek : WIG 20 od 1 stycznia 2009 do 30 czerwca 2012 roku, otwarcie i zamknięcie – badamy położenie.

Wykresy głębia vs. głębia - przykład empiryczny



Rysunek : WIG 20 od 1 stycznia 2009 do 30 czerwca 2012 roku, otwarcie i zamknięcie – badamy rozrzut.

pomiar rozrzutu i asymetrii w wielu wymiarach

Niech $D(\mathbf{x}, \mathbf{Z}^n)$ oznacza funkcję głębi w punkcie $\mathbf{x} \in \mathbb{R}^d$, $d \geq 2$,
 $\mathbf{Z}^n = \{\mathbf{z}_1, \dots, \mathbf{z}_n\} \subset \mathbb{R}^d$.

Niech $D_\alpha(\mathbf{Z}^n) = \{\mathbf{z} \in \mathbb{R}^d : D(\mathbf{z}, \mathbf{Z}^n) \geq \alpha\}$ oznacza tzw. obszar centralny
rzędu $\alpha \in [0, 1]$.

Dla $\beta \in \{0.01, 0.02, \dots, 0.99, 1\}$ definiujemy **krzywą rozrzutu** z próby

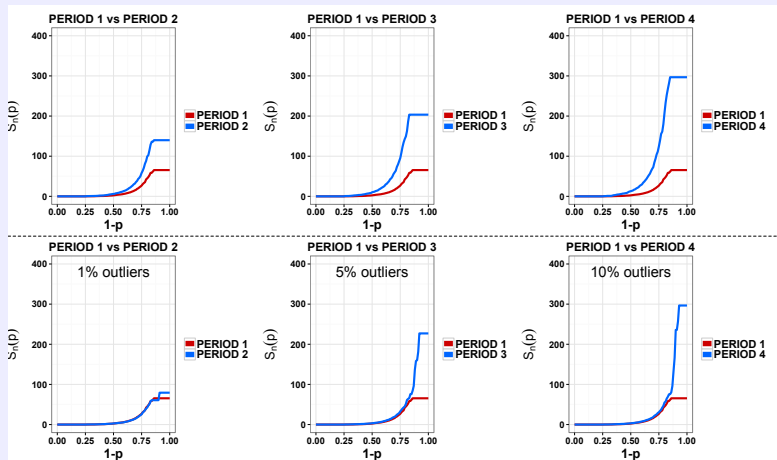
$$KS(\beta) = (\beta, \text{vol}(D_\beta(\mathbf{Z}^n))) \in \mathbb{R}^2.$$

Dla $\beta \in \{0.01, 0.02, \dots, 0.99, 1\}$ definiujemy **krzywą asymetrii** z próby

$$KS(\beta) = (\beta, \|c^{-1}(\{\bar{\mathbf{z}} - \text{med} | D_\beta(\mathbf{Z}^n)\})\|) \in \mathbb{R}^2,$$

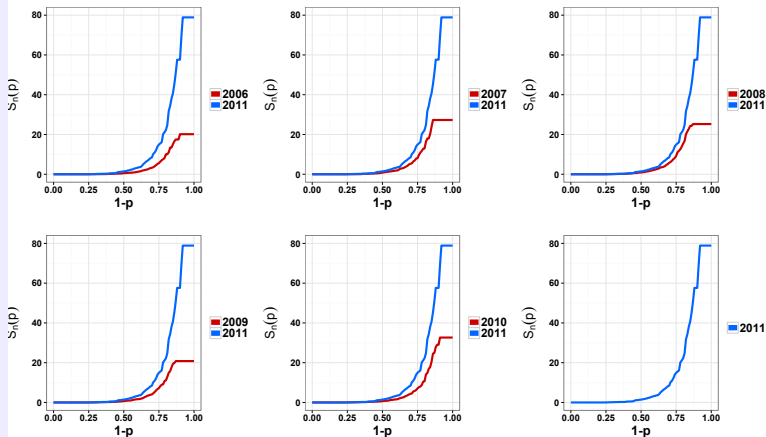
gdzie c —oznacza stałą, $\bar{\mathbf{z}}$ —oznacza wektor średnich, med —oznacza
medianę indukowaną przez funkcję głębi, vol —oznacza objętość.

Krzywe rozrzutu - symulacje



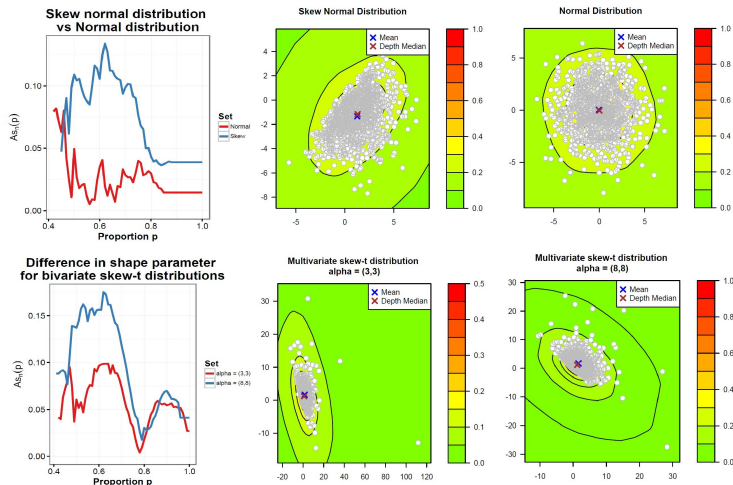
Rysunek : Krzywe rozrzutu - symulacje

Krzywe rozrzutu - dane empiryczne



Rysunek : Krzywe rozrzutu - ceny zamknięcia KGHM (KGH) oraz PKO BP (PKO) 2006 - 2011

Krzywe asymetrii



Rysunek : Krzywe asymetrii - skośne rozkłady Studenta 2D.

podsumowanie + plany na przyszłość

- Nasz pakiet już teraz oferuje szerokie spektrum procedur indukowanych przez funkcje głębokości. Oferuje dobrej jakości grafikę, cechującą go akceptowalnym czasem obliczeń.
- Pakiet jest całkowicie darmowy i w pełni profesjonalny. Można go ściągnąć wraz z opisem ze strony naszego projektu:
<https://r-forge.r-project.org/projects/depthproc>
- Pakiet ewoluuje, jest stale rozwijany, niebawem nowe, możliwości:
 - *M. Bocian*: procedury wykorzystujące głębokości funkcjonalne w analizie danych finansowych.
 - *A. Węgrzynkiewicz*: przewidywanie, wygładzanie szeregu czasowego m. in. za pomocą funkcji głębokości.
 - *Z. Zawadzki*: szybki algorytm do obliczania estymatora maksymalnej głębokości regresyjnej regresji wielu zmiennych.
 - wraz z nowymi możliwościami (części prac magisterskich współautorów) - pakiet zostanie zgłoszony do *Journal of Statistical Software*.

Wybrane pozycje literatury

- Dyckerhoff, R. (2004), Data depths satisfying the projection property. *Allgemeines Statistisches Archiv*. **88**, 163–190.
- Oja, H. (2010). *Multivariate nonparametric methods with R. An approach based on spatial signs and ranks.*, *Lecture Notes in Statistics*, **199**, Springer, New York.
- Rousseeuw, P. J., Hubert, M. (1999), Regression depth, *Journal of the American Statistical Association*, **94**, 388 – 433.
- Serfling, R. (2006). Depth functions in nonparametric multivariate inference, In: Liu R.Y., Serfling R., Souvaine D. L. (Eds.): *Series in Discrete Mathematics and Theoretical Computer Science*, AMS, **72**, 1 – 15.