

WITOLD ORZESZKO

WPLYW REDUKCJI SZUMU LOSOWEGO METODĄ SCHREIBERA NA
IDENTYFIKACJĘ SYSTEMU GENERUJĄCEGO DANE. ANALIZA
SYMULACYJNA¹

1. WSTĘP

Wszelkie rzeczywiste, w tym i ekonomiczne szeregi czasowe cechują się obecnością szumu losowego. Oczywiście fakt ten może znacząco utrudnić trafne identyfikowanie analizowanych zależności, co uzasadnia stosowanie metod redukcji szumu. Jednak z drugiej strony należy podkreślić, że wszelkie filtrowanie może prowadzić do zniszczenia bądź zniekształcenia zależności obecnych w analizowanych danych (por. Mees, Judd [17]).

Wiele spośród znanych metod identyfikacji nieliniowości i chaosu jest wrażliwych na obecność szumu losowego (por. np. Castagli i in. [3], Kantz, Schreiber [12], Shintani, Linton [22], Zeng i in. [25]). Z tego powodu metody redukcji szumu wykorzystywane są do filtrowania rzeczywistych danych jako wstępny etap w procesie identyfikacji chaosu lub nieliniowości. I choć nie jest to powszechnie stosowana praktyka, to otrzymane wyniki wydają się obiecujące (np. Harrison i in. [8], Hassani i in. [9, 10], Kantz i in. [11], Leontitsis i in. [14], Liang i in. [15], Orzeszko [18], Perc [19], Schreiber [21], Sivakumar i in. [23]).

Jedną z metod nieliniowej redukcji szumu losowego jest procedura zaproponowana przez Schreibera [20]. Przeprowadzone symulacje wykazały, że metoda ta, mimo swojej prostoty, skutecznie redukuje szum z szeregów wygenerowanych przez deterministyczne systemy o dynamice chaotycznej (Orzeszko [18], Schreiber [20], Sivakumar i in. [23]). Jednak należy podkreślić, że w przypadku analizy rzeczywistych szeregów czasowych badacz zwykle pozbawiony jest wiedzy o naturze systemu generującego. W szczególności, brak jest pewności, czy spełnione jest założenie leżące u podstaw metod redukcji szumu, tzn. o deterministycznym charakterze systemu generującego. Co więcej, to właśnie poznanie natury systemu generującego może być zasadniczym celem przeprowadzanego badania. Z tego powodu istotną kwestią jest stwierdzenie, jak dana metoda redukcji szumu zachowuje się w zastosowaniu do danych stochastycznych. Celem symulacji przeprowadzonych w niniejszej pracy było zweryfikowanie, czy prze-

¹ Badanie zostało sfinansowane przez Uniwersytet Mikołaja Kopernika w Toruniu w ramach grantu UMK nr 397-E „Identyfikacja nieliniowych zależności w ekonomicznych szeregach czasowych”.

filtrowanie szeregów losowych metodą Schreibera nie wprowadza do nich zależności, które mogłyby prowadzić do błędnych wniosków o naturze systemu generującego.

Praca skonstruowana jest następująco: w rozdziale drugim przedstawiono istotę i przebieg procesu redukcji szumu losowego, w rozdziałach trzecim oraz czwartym scharakteryzowano metody identyfikacji zależności nieliniowych w szeregach czasowych, tj. odpowiednio: test BDS oraz miarę informacji wzajemnej, w rozdziale piątym zaprezentowano przebieg oraz wyniki przeprowadzonych symulacji.

2. REDUKCJA SZUMU LOSOWEGO

Analizując określony rzeczywisty szereg czasowy (x_t) należy liczyć się z obecnością w nim szumu losowego, reprezentującego szum obserwacyjny, systemowy lub pewną ich kombinację. W takiej sytuacji w szeregu (x_t) można wyróżnić część deterministyczną (y_t) i stochastyczną (ε_t), co po zapisaniu w postaci addytywnej oznacza istnienie dekompozycji:

$$x_t = y_t + \varepsilon_t, \quad (1)$$

gdzie od (ε_t) wymaga się co najmniej posiadania szybko malejącej funkcji autokorelacji i nieskorelowania z sygnałem (y_t) (Kantz, Schreiber [12]).

Redukcja szumu losowego ma na celu upodobnienie „zanieczyszczonego” szeregu (x_t) do oryginalnego sygnału (y_t). Bardzo prostą, lecz mimo to stosunkowo skuteczną metodą redukcji szumu jest metoda Najbliższych Sąsiadów – NS (por. Schreiber [20]). W celu wyznaczenia wartości y_i (dla dowolnego i) metodą Najbliższych Sąsiadów należy skonstruować wektory opóźnień $\hat{x}_i = (x_{i-K}, x_{i-K+1}, \dots, x_i, \dots, x_{i+L-1}, x_{i+L})$, gdzie K i L są ustalonymi liczbami naturalnymi. Niech $\hat{x}_{i_1}, \hat{x}_{i_2}, \dots, \hat{x}_{i_n}$ oznaczają n najbliższych (w sensie ustalonej $K + L + 1$ – wymiarowej metryki) sąsiadów wektora \hat{x}_i . W oparciu o wyznaczonych najbliższych sąsiadów, wartość \tilde{y}_i , będącą oszacowaniem y_i oblicza się ze wzoru:

$$\tilde{y}_i = \frac{1}{n} \sum_{i=1}^n x_{i_i}. \quad (2)$$

Dobór liczby najbliższych sąsiadów może odbywać się wprost (jak to zostało zaprezentowane w przedstawionym powyżej algorytmie) lub przez zadanie promienia sąsiedztwa r .

Metoda NS zazwyczaj prowadzi do różnych szeregów w zależności od przyjętych wartości parametrów K , L oraz r (lub n). Do wyboru właściwych wartości parametrów, a co za tym idzie – do wyboru odpowiedniego szeregu wynikowego, można zastosować wskaźnik poziomu redukcji szumu NRL (*Noise Reduction Level*) (Orzeszko [18]). Metoda ta opiera się na obserwowanej zależności pomiędzy strukturą geometryczną atraktora a siłą szumu dodawanego do systemu. Zależność ta polega na „pogrubianiu” atraktora, tzn., średnio rzecz biorąc, oddalaniu się bliskich sobie stanów wraz ze wzrostem siły szumu.

Punktem wyjścia metody jest rekonstrukcja atraktora systemu w oparciu o szereg (x_t) , polegająca na skonstruowaniu m -wymiarowych wektorów opóźnień, gdzie m jest ustaloną liczbą naturalną. Niech T oznacza liczbę wszystkich dostępnych wektorów opóźnień, natomiast d_i ($i = 1, 2, \dots, T$) – odległość i -tego oczyszczonego wektora opóźnień $(\tilde{y}_{i-K}, \tilde{y}_{i-K+1}, \dots, \tilde{y}_i, \dots, \tilde{y}_{i+L-1}, \tilde{y}_{i+L})$ do jego najbliższego sąsiada. Wówczas wielkość d_{\min} , określoną wzorem $d_{\min} = \frac{1}{T} \sum_{i=1}^T d_i$, można interpretować jako miarę grubości zrekonstruowanego atraktora systemu. Naczelną ideą proponowanej metody jest wskazanie spośród szeregów, będących wynikiem redukcji szumu, takiego, dla którego względny spadek wartości d_{\min} (w stosunku do zanieczyszczonego szeregu) jest największy. Należy jednak zauważyć, że uwzględnienie wyłącznie tego czynnika byłoby niewystarczające, gdyż prowadziłyby zawsze do wyboru szeregu złożonego z równych (lub prawie równych) wartości². Z tego powodu do wskaźnika NRL dodatkowo wprowadzono wartość $\left| \frac{diam - diam^0}{diam^0} \right|$, gdzie $diam^0$ i $diam$ oznaczają średnice zrekonstruowanego atraktora, odpowiednio, przed i po redukcji szumu. Rolą względnej zmiany średnicy jest zmierzenie poziomu deformacji oczyszczonego atraktora, spowodowanej przez redukcję szumu.

Ostatecznie wskaźnik poziomu redukcji szumu NRL wyraża się wzorem:

$$NRL = \frac{d_{\min} - d_{\min}^0}{d_{\min}^0} + \left| \frac{diam - diam^0}{diam^0} \right|. \quad (3)$$

Spośród szeregów otrzymanych w wyniku redukcji szumu należy wybrać taki, który cechuje się najniższym wskaźnikiem NRL.

Z samej konstrukcji wartość NRL zależy od przyjętego wymiaru zanurzenia m . Zgodnie z kryterium Takensa [1981], w procesie rekonstrukcji przestrzeni stanów należy uwzględnić wektory opóźnień, których wymiar spełnia warunek $m > 2d$, gdzie d jest wymiarem przestrzeni stanów.

3. TEST BDS

BDS (Brock i in. [1]) jest nieparametrycznym testem weryfikującym hipotezę zerową, że badany szereg czasowy jest realizacją procesu i.i.d. (tzn. procesu niezależnych zmiennych losowych o jednakowym rozkładzie). W swojej konstrukcji odwołuje się do całki korelacyjnej, zadanej wzorem:

$$C_m^T(\varepsilon) = \lim_{T \rightarrow \infty} \frac{\left| \{(i, j); \|\hat{x}_i - \hat{x}_j\| < \varepsilon, 1 \leq i, j \leq T\} \right|}{T^2}, \quad (4)$$

² W przypadku metody NS szereg taki otrzymuje się dla odpowiednio dużej wartości parametru r (lub n).

gdzie symbol $\|$ oznacza moc zbioru, natomiast T jest liczbą wszystkich m -wymiarowych wektorów opóźnień.

Teoretyczną podstawą testu BDS jest asymptotyczna zależność pomiędzy całkami korelacyjnymi $C_m^T(\varepsilon)$ i $C_1^T(\varepsilon)$. Udowodniono, że dla każdego $m > 1$ oraz $\varepsilon > 0$ wartość $C_m^T(\varepsilon)$ szeregu i.i.d. zbiega do $(C_1^T(\varepsilon))^m$ przy $T \rightarrow \infty$. Mająca tu zastosowanie statystyka W określona jest wzorem³:

$$W_{T,m}(\varepsilon) = \frac{\sqrt{T} \left(C_m^T(\varepsilon) - (C_1^T(\varepsilon))^m \right)}{\sigma_{T,m}(\varepsilon)}. \quad (5)$$

Przy założeniu prawdziwości hipotezy zerowej, że szereg jest realizacją procesu i.i.d., statystyka W jest zbieżna według rozkładu do $N(0, 1)$ przy $T \rightarrow \infty$ (por. Kanzler [13]). BDS jest testem dwustronnym, w którym weryfikacji podlega hipoteza zerowa o nieistotności statystyki W .

Należy podkreślić, że test BDS identyfikuje zależności zarówno liniowe, jak i nieliniowe, więc w celu identyfikacji nieliniowości należy badać szereg uprzednio prefiltrowany odpowiednim modelem ARMA. Zaletą tego testu jest fakt, że identyfikuje nieliniowości bardzo różnej natury.

Test BDS jest wrażliwy na liczbę obserwacji. Brock i in. [2] w oparciu o przeprowadzone symulacje stwierdzili, że wiarygodne wyniki otrzymuje się dla szeregów liczących co najmniej 250 obserwacji. Dodatkowo, w przypadku krótkich szeregów (do 500 obserwacji) w celu wyznaczenia wartości krytycznych bezpieczniej jest stosować procedurę *bootstrap* niż tablice rozkładu normalnego (por. Brock i in. [2], Kanzler [13]).

W oparciu o dokonane symulacje Brock i in. [2] proponują, aby parametr m przyjmował w teście wartości 2, 3, 4, 5 dla krótkich szeregów (do 500 obserwacji) i wartości 2, 3, ..., 10 – dla długich (co najmniej 2000 obserwacji). Ponadto, z symulacji tych wynika postulat, aby za ε przyjmować wartości pomiędzy $0,5 \cdot \sigma$ i $1,5 \cdot \sigma$, gdzie σ jest odchyleniem standardowym badanego szeregu. Z kolei z badań przeprowadzonych przez L. Kanzlera [13] wynika, że przyjmowana w teście BDS wartość parametru ε powinna zależeć od typu rozkładu analizowanego szeregu. Przykładowo w przypadku szeregów o rozkładzie normalnym lub do niego zbliżonego, należy przyjąć $\varepsilon = 1,5 \cdot \sigma$ lub $\varepsilon = 2 \cdot \sigma$.

4. MIARA INFORMACJI WZAJEMNEJ

Miara informacji wzajemnej (ang. Mutual Information – MI) jest jedną z najważniejszych metod pomiaru nieliniowych zależności w szeregach czasowych. Określona jest ona wzorem:

³ $\sigma_{T,m}(\varepsilon)$ jest czynnikiem normalizującym, którego wzór przedstawiony jest np. w Brock i in. [1991].

$$I(X, Y) = \iint p(x, y) \log \left(\frac{p(x, y)}{p_1(x)p_2(y)} \right) dx dy, \quad (6)$$

gdzie $p(x, y)$ jest funkcją gęstości rozkładu łącznego, natomiast $p_1(x)$ oraz $p_2(y)$ są gęstościami brzegowymi zmiennych X i Y .

Można udowodnić, że $I(X, Y)$ przyjmuje zawsze wartości nieujemne oraz $I(X, Y) = 0$ wtedy i tylko wtedy, gdy X i Y są niezależne (np. Granger, Lin [7]). Oznacza to, że test niezależności bazujący na mierze informacji wzajemnej polega na weryfikacji hipotezy $H_0: I(X, Y) = 0$ przeciw $H_1: I(X, Y) > 0$.

W literaturze przedmiotu istnieją różne propozycje szacowania wartości $I(X, Y)$. Zasadniczo, sprowadzają się one do oszacowania funkcji gęstości $p(x, y)$, $p_1(x)$ oraz $p_2(y)$ (por. wzór 6). Proponowane metody estymacji można podzielić na trzy główne grupy (por. Dionisio, Menezes, Mendes [4]):

- odwołujące się do histogramu,
- oparte na estymatorach jądrowych,
- metody parametryczne.

Miara MI identyfikuje zależności zarówno liniowe, jak i nieliniowe, więc podobnie jak w przypadku testu BDS, aby zidentyfikować zależności nieliniowe, badany szereg trzeba przefiltrować modelem typu ARMA. Miarę tę można wykorzystać również do pomiaru autozależności w pojedynczym szeregu czasowym (x_t). W tym celu za (y_t) należy przyjąć szereg opóźnionych wartości (x_t).

5. ANALIZA SYMULACYJNA

5.1 PRZEBIEG BADANIA

Celem przeprowadzonych symulacji było zweryfikowanie, czy dokonanie redukcji szumu losowego metodą Najbliższych Sąsiadów może zakłócić identyfikację zależności nieliniowych w szeregach czasowych, a dokładniej mówiąc: czy metoda ta może wprowadzić do szeregów losowych zależności o charakterze nieliniowym.

Badania oparto na wygenerowanych w programie Matlab 6.5 szeregach liczb pseudolosowych o rozkładzie $N(0, 1)$: 1000 szeregów krótkich (300 obserwacji) oraz 1000 szeregów długich (2000 obserwacji).

W pierwszej kolejności, każdy z wygenerowanych szeregów poddano redukcji szumu losowego przy zastosowaniu metody NS. Następnie przefiltrowane szeregi przetestowano pod kątem ich niezależności. W tym celu zastosowano procedurę opartą na teście BDS, miarę informacji wzajemnej oraz współczynnik korelacji liniowej Pearsona.

5.2 REDUKCJA SZUMU LOSOWEGO

W redukcji szumu losowego zastosowano metodę Najbliższych Sąsiadów.⁴ W badaniu arbitralnie przyjęto wartości parametrów $K = 2$ oraz $L = 2$, natomiast promień otoczenia r wyznaczono przy użyciu wskaźnika NRL.⁵

Przy wyznaczaniu wskaźnika NRL analizie poddano po jednym z wygenerowanych długich i krótkich szeregów. W badaniu rozważono osiem wartości promienia otoczenia: $r = 0,2; 0,3; 0,4; 0,5; 0,7; 1; 1,5; 2$. W celu wyznaczenia optymalnego promienia obliczono wartości wskaźnika $NRL(m)$, kolejno, dla $m = 1, 2, 3, 5, 7, 10$. Obliczone wartości NRL zaprezentowano w Tabelach 1-2. Dla każdego m pogrubioną czcionką zaznaczono najmniejszą wartość wskaźnika $NRL(m)$. Dodatkowo w każdej komórce obydwu tych tabel przedstawiono w nawiasach wartości drugiej składowej wskaźnika NRL, tzn. względnej zmiany średnicy – $\Delta diam$ [%], umożliwiające ocenę stopnia deformacji przefiltrowanego szeregu.

Tabela 1

Wartości NRL dla szeregu losowego $N(0,1)$, złożonego z 300 obserwacji

$r \backslash$ NRL	NRL(1)	NRL(2)	NRL(3)	NRL(5)	NRL(7)	NRL(10)
0,2	0,0000 /0,00/	0,3642 /0,00/	0,0417 /0,00/	0,1551 /0,00/	-0,0021 /0,00/	-0,0351 /0,00/
0,3	0,0000 /0,00/	1,6894 /0,00/	0,0509 /0,00/	1,0959 /0,00/	0,2353 /0,00/	0,2390 /0,00/
0,4	0,0319 /0,00/	-0,0749 /0,00/	0,9184 /0,00/	0,7479 /0,00/	1,1614 /0,00/	0,5870 /0,00/
0,5	0,0427 /0,00/	-3,8614 /0,00/	-0,3144 /0,00/	0,5066 /0,00/	1,5307 /0,02/	2,8491 /0,43/
0,7	3,9391 /0,00/	-8,2530 /0,00/	-9,0608 /0,00/	-10,7853 /0,00/	-7,2211 /0,39/	-1,1949 /5,07/
1	6,7304 /4,16/	-14,9189 /3,28/	-13,9593 /7,46/	-14,8774 /10,88/	-16,0313 /10,55/	-9,8927 /13,39/
1,5	-1,6249 /38,82/	-7,3907 /39,18/	-7,9031 /41,91/	-10,7387 /39,20/	-10,9918 /39,51/	-5,1402 /44,98/
2	-5,5032 /52,41/	-12,5752 /52,74/	-12,9643 /55,27/	-9,6012 /57,76/	-10,1073 /57,31/	-5,8164 /61,35/

Źródło: obliczenia własne

Biorąc pod uwagę otrzymane rezultaty, do metody redukcji szumu losowego przyjęto wartość $r = 1$ dla szeregów krótkich oraz $r = 0,7$ dla szeregów długich. W przypadku szeregów krótkich, wybór $r = 1$ nie pozostawiał wątpliwości; jedynie dla $m = 1$ ten poziom parametru r nie prowadził do najmniejszej wartości wskaźnika $NRL(m)$. Nieco mniej jednoznaczne wskazanie otrzymano dla szeregów długich. Jed-

⁴ Obliczenia wykonano w programie Matlab 6.5 w oparciu o procedurę napisaną przez A. Leontitsisa.

⁵ Obliczenia wykonano w programie Excel w oparciu o procedurę własną.

Tabela 2

Wartości NRL dla szeregu losowego $N(0,1)$, złożonego z 2000 obserwacji

$r \backslash$ NRL	NRL(1)	NRL(2)	NRL(3)	NRL(5)	NRL(7)	NRL(10)
0,2	0,0340 /0,00/	-0,2248 /0,00/	-0,2762 /0,00/	-0,1610 /0,00/	0,0440 /0,00/	0,0965 /0,00/
0,3	-1,2391 /0,00/	-0,5434 /0,00/	-1,7409 /0,00/	-1,0148 /0,00/	-0,0353 /0,00/	0,4910 /0,45/
0,4	1,1279 /0,00/	0,7097 /0,00/	-2,5220 /0,00/	-5,4955 /0,00/	-1,1727 /0,88/	-1,1857 /0,19/
0,5	-1,6668 /0,00/	-1,9316 /0,00/	-4,6991 /0,84/	-9,3423 /0,00/	-3,9642 /1,67/	-3,6049 /1,70/
0,7	8,0467 /4,19/	-7,4150 /0,00/	-12,3525 /0,84/	-13,1207 /4,94/	-13,3316 /2,18/	-8,0552 /5,85/
1	1,9431 /10,79/	-5,5266 /13,37/	-11,2943 /14,95/	-8,4624 /19,54/	-7,0748 /20,89/	-7,4781 /20,17/
1,5	-3,4391 /35,26/	-6,1141 /36,96/	-7,0012 /39,91/	-6,2428 /41,95/	-5,0225 /43,22/	-6,9233 /41,44/
2	-2,1241 /50,58/	-9,6498 /52,79/	-10,4288 /55,29/	-9,6989 /57,58/	-8,1550 /59,32/	-9,3679 /58,31/

Źródło: obliczenia własne

nak i w tej sytuacji zauważalna jest wyższość parametru $r = 0,7$ nad pozostałymi analizowanymi poziomami.

Metoda NS pozostawia nieprzefiltrowane K pierwszych oraz L ostatnich obserwacji, co skutkuje tym, że szeregi krótkie zostały skrócone do 296 a długie – do 1996 obserwacji.

5.3 TEST BDS

Dla każdego z przefiltrowanych szeregów wyznaczono statystykę W , przyjmując kolejno wartości wymiaru zanurzenia $m = 2, 3, 4, 5$.⁶ Wartość ε ustalono na poziomie $\varepsilon = 1,5\sigma$, gdzie σ jest odchyleniem standardowym analizowanego szeregu.

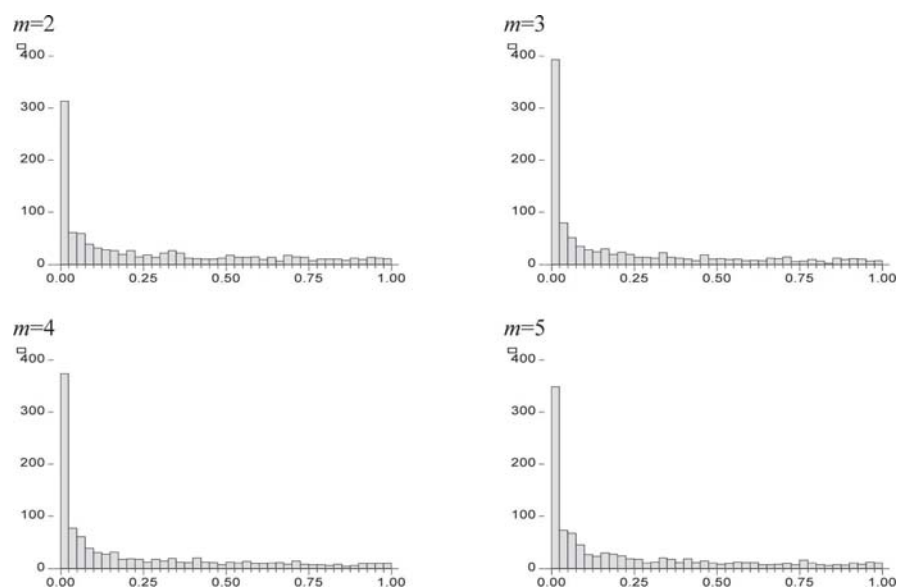
Dla poszczególnych analizowanych szeregów, w odniesieniu do kolejnych rozważanych wartości parametru m , wyznaczono wartość empirycznego poziomu istotności (ang. p -value), umożliwiającą zweryfikowanie hipotezy zerowej o nieistotności statystyki W (tzn. o braku zależności w szeregu). Wartość ta została wyznaczona w oparciu o procedurę bootstrap (1000 iteracji).⁷

Na Rysunkach 1-2 przedstawiono histogramy empirycznych poziomów istotności obliczonych dla różnych wartości parametru m . Ponadto, w oparciu o otrzymane hi-

⁶ Obliczenia wykonano w programie Matlab 6.5 w oparciu o procedurę napisaną przez L. Kanzlera.

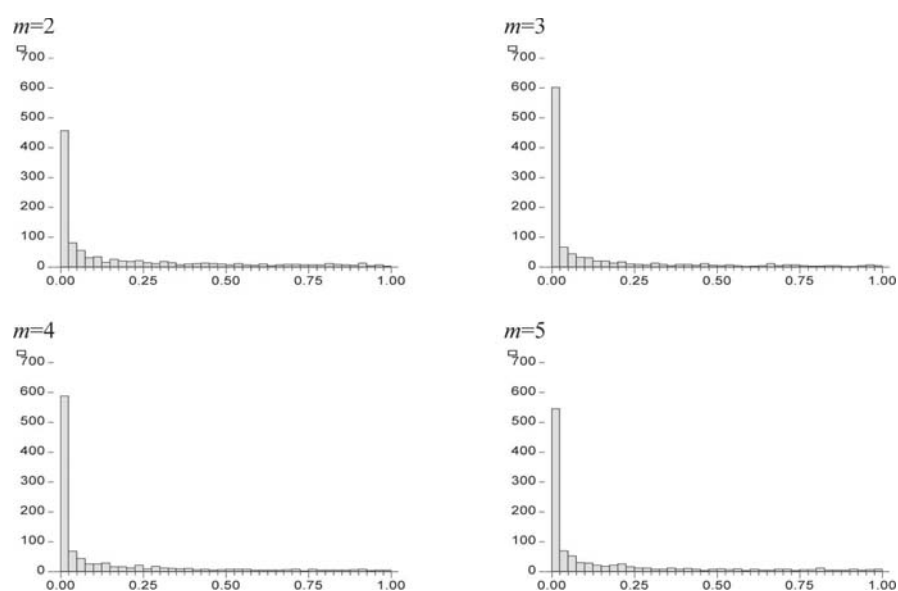
⁷ Zastosowano procedurę bootstrap bez powtarzania. Wyznaczone empiryczne poziomy istotności odpowiadają testowi dwustronnemu (tzn. są dwukrotnością grubości odpowiedniego ogona rozkładu empirycznego).

stogramy wyznaczono odsetek szeregów, dla których odrzucono hipotezę zerową, przy wybranych poziomach istotności $\alpha = 1\%$, 5% , 10% . Wyniki te zostały przedstawione w Tabelach 3-4.



Rysunek 1. Histogramy empirycznych poziomów istotności statystyki W dla krótkich szeregów

Źródło: obliczenia własne



Rysunek 2. Histogramy empirycznych poziomów istotności statystyki W dla długich szeregów

Źródło: obliczenia własne

Tabela 3

Odsetek krótkich szeregów, dla których odrzucono H_0

	$m=2$	$m=3$	$m=4$	$m=5$
$\alpha = 1\%$	23,6%	31,4%	29,5%	26,7%
$\alpha = 5\%$	38,0%	47,4%	45,7%	42,8%
$\alpha = 10\%$	47,2%	56,0%	55,4%	53,6%

Źródło: obliczenia własne

Tabela 4

Odsetek długich szeregów, dla których odrzucono H_0

	$m=2$	$m=3$	$m=4$	$m=5$
$\alpha = 1\%$	37,8%	52,8%	50,6%	46,9%
$\alpha = 5\%$	54,0%	67,1%	65,8%	61,6%
$\alpha = 10\%$	62,2%	74,4%	72,4%	69,4%

Źródło: obliczenia własne

W otrzymanych rozkładach empirycznych zdecydowanie dominują wartości bliskie zeru. Potwierdzają to wyniki zaprezentowane na Rysunkach 1-2 oraz w Tabelach 3-4. Jak widać na wykresach, w przypadku szeregów krótkich – ponad 30%, a w przypadku szeregów długich – nawet ponad 45% obliczonych empirycznych poziomów istotności nie przekracza wartości 0,025. Z kolei obliczenia zestawione w Tabelach 3-4 pokazują, że odsetek szeregów, dla których odrzucono hipotezę zerową znacznie przewyższa przyjęty krytyczny poziom istotności α . Z otrzymanych rezultatów testu BDS wynika więc, że przefiltrowanie szeregów metodą NS powoduje pojawienie się w wielu z nich autozależności pomiędzy obserwacjami. Naturalnie na tym etapie badania nie można wyciągnąć wniosków dotyczących charakteru tych zależności.⁸

5.4 MIARA INFORMACJI WZAJEMNEJ

Dla każdego z przefiltrowanych szeregów oszacowano wartości miary informacji wzajemnej dla opóźnień czasowych $k = 1, 2, \dots, 10$.⁹ Do oszacowania miary MI wykorzystano metodę zaproponowaną przez Frasera, Swinneya [5]. Jest to metoda oparta na analizie dwuwymiarowego histogramu. Mówiąc w uproszczeniu, polega ona na podziale przestrzeni dwuwymiarowej zawierającej pary (x_t, y_t) na prostokątne komórki (według określonej zasady) i obliczeniu, jaki odsetek naniesionych punktów znajduje się w każdej z komórek. Następnie znajduje zastosowanie wzór (6), przy czym ob-

⁸ Zależności te mogą być zarówno liniowe, jak i nieliniowe. Kwestii tej został poświęcony podrozdział 5.5.

⁹ Obliczenia wykonano w programie Matlab 6.5 w oparciu o procedurę napisaną przez A. Leontitsisa.

liczone odsetki są oszacowaniem funkcji gęstości, natomiast całkowanie odbywa się numerycznie.

Tak jak w przypadku testu BDS, w celu zweryfikowania hipotezy o nieistotności miary MI (tzn. o braku zależności w szeregu) obliczono empiryczne poziomy istotności. Wartości te zostały wyznaczone w oparciu o procedurę bootstrap (10 000 iteracji).¹⁰ Dla każdego z szeregów otrzymano w ten sposób 10 000 wartości MI(1), których rozkład był podstawą do wyznaczenia empirycznych poziomów istotności dla każdego kolejnego $k = 1, 2, \dots, 10$.

Na Rysunkach 3-4 przedstawiono histogramy empirycznych poziomów istotności obliczonych dla różnych wartości parametru k . Ponadto, w oparciu o otrzymane histogramy wyznaczono odsetek szeregów, dla których odrzucono hipotezę zerową, przy wybranych poziomach istotności $\alpha = 1\%, 5\%, 10\%$. Wyniki te zostały przedstawione w Tabelach 5-6.

Tabela 5

Odsetek krótkich szeregów, dla których odrzucono H_0

	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$	$k=9$	$k=10$
$\alpha = 1\%$	44,8%	41,3%	1,1%	1,3%	1,1%	1,0%	1,8%	1,5%	1,6%	1,8%
$\alpha = 5\%$	66,5%	61,6%	6,3%	5,7%	6,1%	6,1%	5,5%	7,5%	6,0%	6,8%
$\alpha = 10\%$	76,4%	72,1%	12,1%	11,6%	11,2%	10,6%	11,2%	13,6%	10,8%	14,5%

Źródło: obliczenia własne

Tabela 6

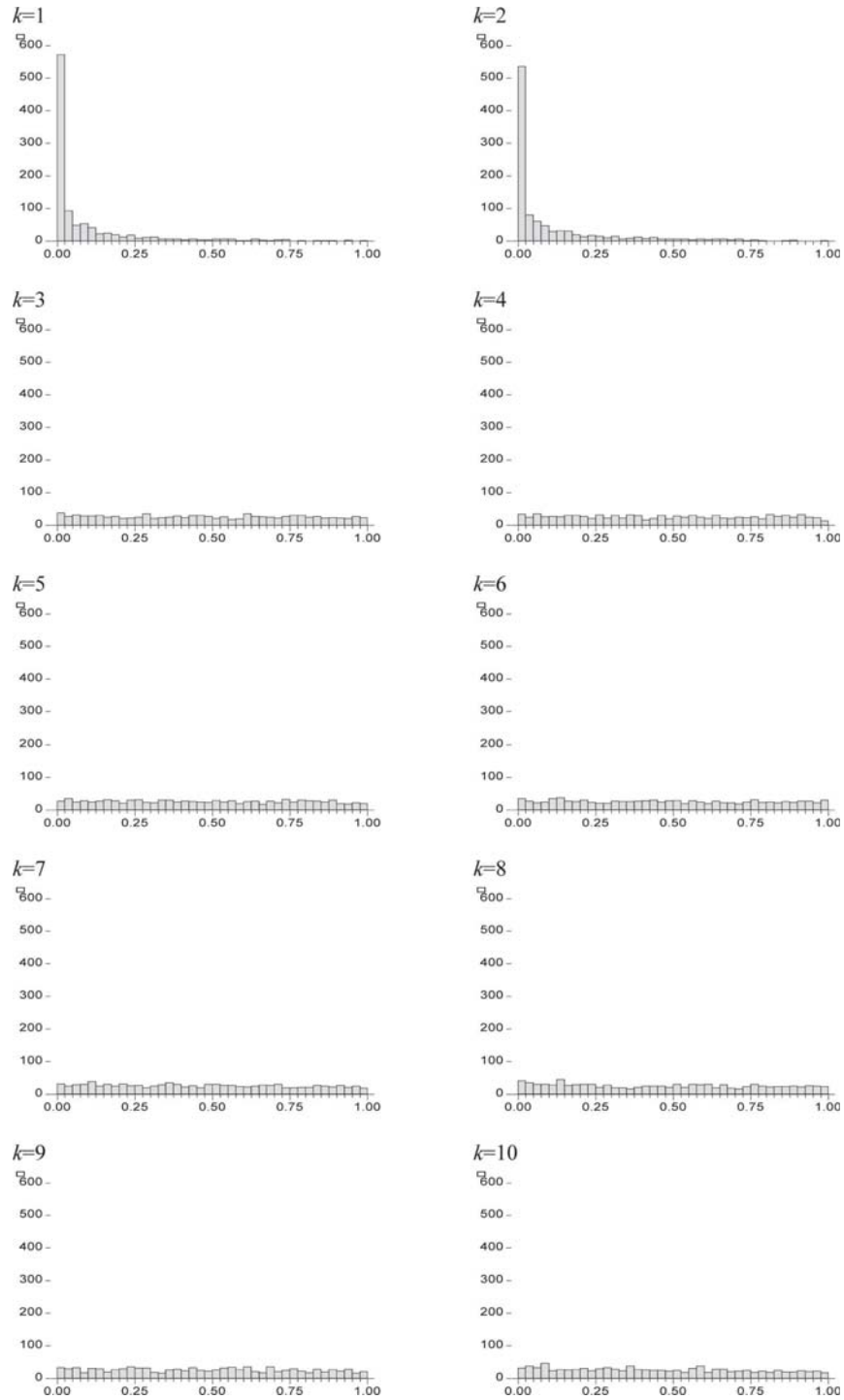
Odsetek długich szeregów, dla których odrzucono H_0

	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$	$k=9$	$k=10$
$\alpha = 1\%$	73,7%	66,1%	0,8%	0,8%	0,8%	0,7%	1,1%	1,0%	1,3%	1,3%
$\alpha = 5\%$	88,6%	81,4%	5,2%	5,3%	4,5%	5,0%	4,9%	5,9%	6,1%	6,1%
$\alpha = 10\%$	93,7%	89,8%	9,7%	9,2%	9,1%	10,3%	10,3%	10,9%	10,2%	11,6%

Źródło: obliczenia własne

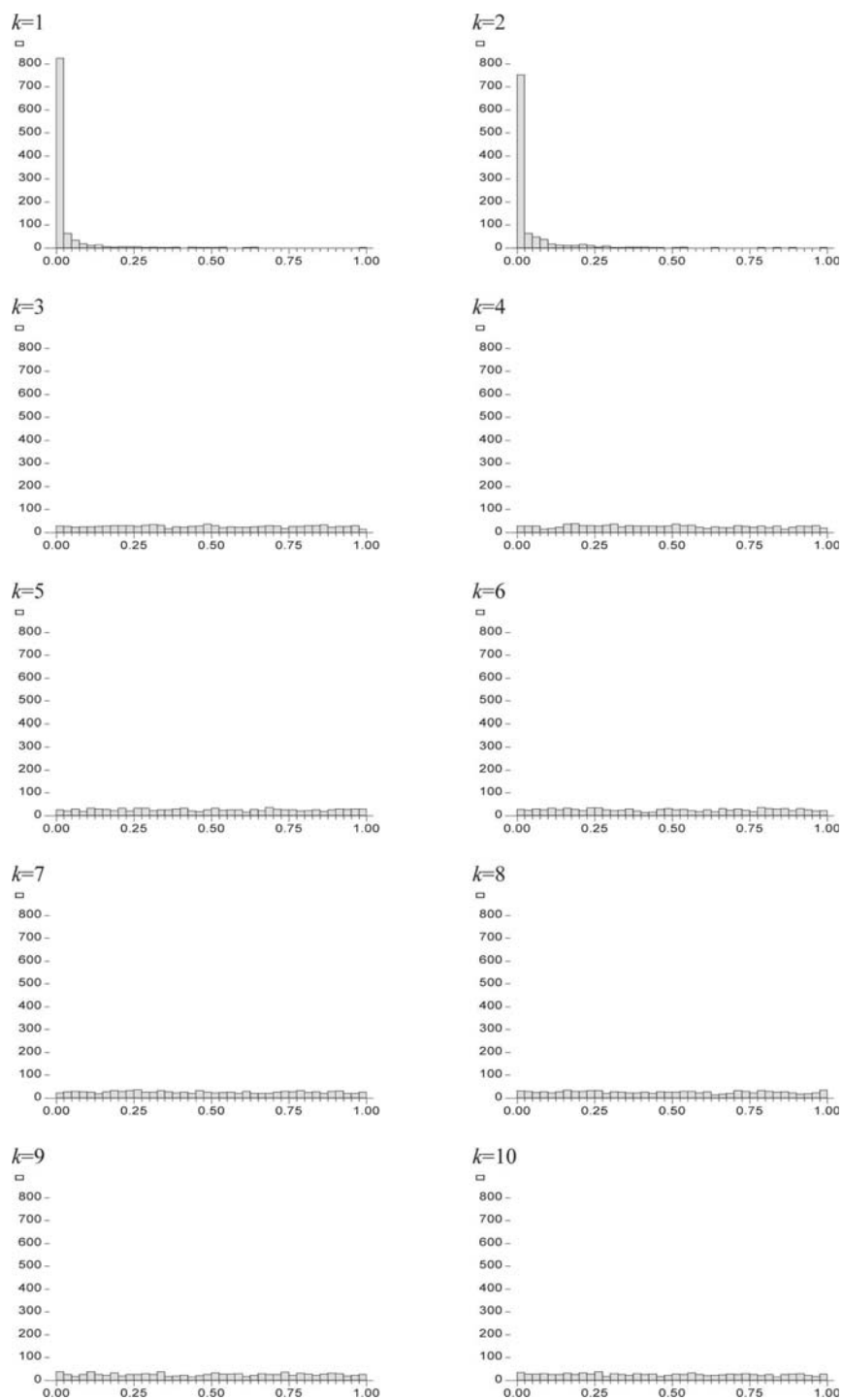
Z Rysunków 3-4 dla $k = 1$ oraz $k = 2$ wynika, że w przypadku szeregów krótkich – ponad 50%, a w przypadku szeregów długich – nawet ponad 75% wyznaczonych empirycznych poziomów istotności nie przekracza wartości 0,025. Istnienie autozależności potwierdzają również obliczenia zaprezentowane w Tabelach 5-6. Pokazują one, że dla $k = 1$ oraz $k = 2$ odsetek szeregów, dla których odrzucono hipotezę zerową znacznie przewyższa przyjęty krytyczny poziom istotności α . Z kolei w przypadku wyższych opóźnień (tzn. $k > 2$), histogramy są płaskie, a widoczne w Tabelach 5-6

¹⁰ Zastosowano procedurę bootstrap bez powtarzania. Wyznaczone empiryczne poziomy istotności odpowiadają testowi prawostronnemu (tzn. są grubością prawego ogona rozkładu empirycznego).



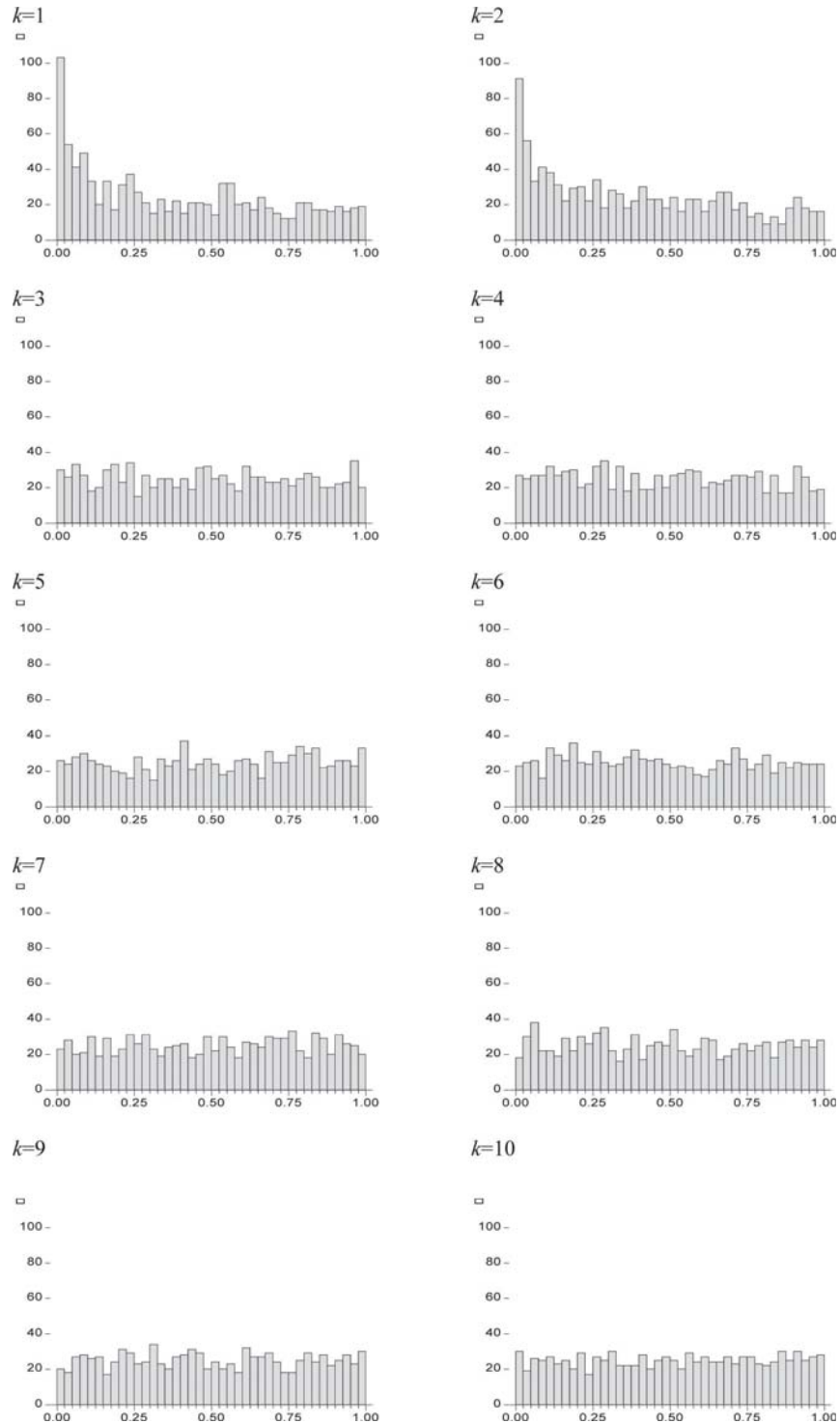
Rysunek 3. Histogramy empirycznych poziomów istotności miary MI dla krótkich szeregów

Źródło: obliczenia własne



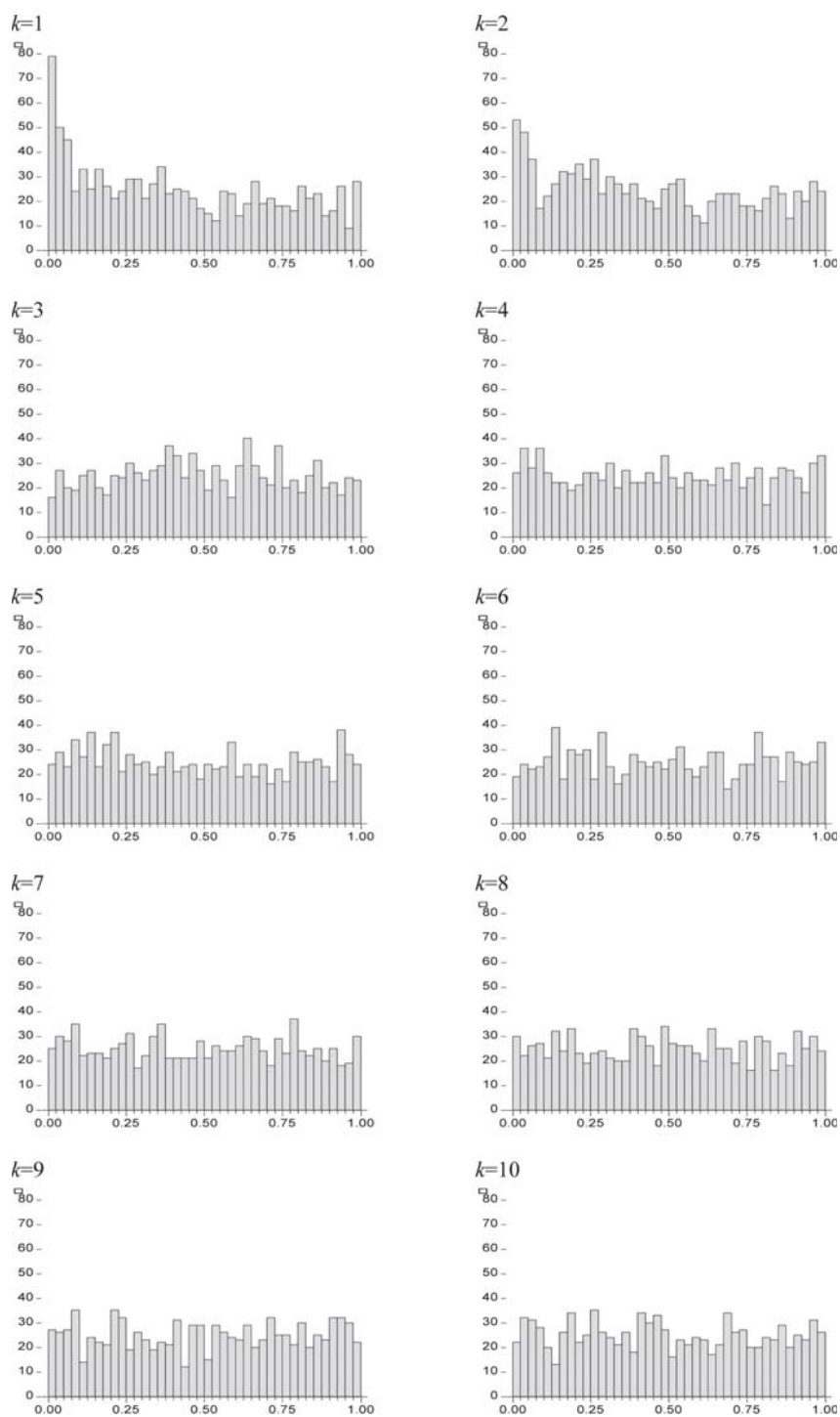
Rysunek 4. Histogramy empirycznych poziomów istotności miary MI dla długich szeregów

Źródło: obliczenia własne



Rysunek 5. Histogramy empirycznych poziomów istotności współczynnika AC dla krótkich szeregów

Źródło: obliczenia własne



Rysunek 6. Histogramy empirycznych poziomów istotności współczynnika AC dla długich szeregów

Źródło: obliczenia własne

odsetki szeregów, dla których odrzucono hipotezę zerową, są zbliżone do przyjętych poziomów istotności. Otrzymane rezultaty wyraźnie wskazują, że w analizowanych szeregach pojawiły się autozależności pierwszego oraz drugiego rzędu. Istnienie w szeregach zależności do drugiego rzędu włącznie wydaje się być konsekwencją ustalenia wartości parametrów $K = 2$ i $L = 2$ w metodzie NS.

5.5 ANALIZA LINIOWOŚCI

Zidentyfikowane przez test BDS oraz miarę MI zależności w przefiltrowanych szeregach mogą być różnej natury. Celem niniejszego podrozdziału jest odpowiedź na pytanie, czy mają one charakter nieliniowy.

W tym celu dla każdego z analizowanych szeregów obliczono współczynnik autokorelacji liniowej Pearsona, dla opóźnień czasowych $k = 1, 2, \dots, 10$.¹¹

Na Rysunkach 5-6 przedstawiono histogramy empirycznych poziomów istotności obliczonych dla różnych wartości parametru k . Ponadto, w oparciu o otrzymane histogramy wyznaczono odsetek szeregów, dla których odrzucono hipotezę zerową o nieistotności współczynnika, przy wybranych poziomach istotności $\alpha = 1\%$, 5% , 10% . Wyniki te zostały przedstawione w Tabelach 7-8.

Tabela 7

Odsetek krótkich szeregów, dla których odrzucono H_0

	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$	$k=9$	$k=10$
$\alpha = 1\%$	5,9%	5,5%	1,3%	1,0%	1,1%	1,0%	1,0%	0,7%	0,6%	1,2%
$\alpha = 5\%$	15,7%	14,7%	5,6%	5,2%	5,0%	4,8%	5,1%	4,8%	3,8%	4,9%
$\alpha = 10\%$	24,7%	22,1%	11,6%	10,6%	10,8%	9,0%	9,2%	10,8%	9,3%	10,0%

Źródło: obliczenia własne

Tabela 8

Odsetek długich szeregów, dla których odrzucono H_0

	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$	$k=9$	$k=10$
$\alpha = 1\%$	4,3%	2,9%	0,6%	0,6%	0,5%	1,0%	0,8%	1,5%	0,7%	0,8%
$\alpha = 5\%$	12,9%	10,1%	4,3%	6,2%	5,3%	4,3%	5,5%	5,2%	5,3%	5,4%
$\alpha = 10\%$	19,8%	15,5%	8,2%	12,6%	11,0%	8,8%	11,8%	10,5%	11,5%	11,3%

Źródło: obliczenia własne

Podobnie jak w przypadku miary MI otrzymane rezultaty wskazują na obecność autozależności pierwszego i drugiego rzędu. Jednak w tym wypadku, odsetek szeregów ze zidentyfikowanymi zależnościami jest dla każdego zaprezentowanego poziomu

¹¹ Do tego celu wykorzystano funkcję `corrcoef` w programie Matlab 6.5.

istotności dużo niższy niż w przypadku metody MI. Prowadzi to do wniosku, że tylko niewielką część zidentyfikowanych przez MI zależności można tłumaczyć obecnością autokorelacji liniowej. Oznacza to, że w przypadku większości przefiltrowanych szeregów zidentyfikowane autozależności mają charakter nieliniowy.

5.6 PODSUMOWANIE WYNIKÓW SYMULACJI

Przeprowadzone symulacje wskazały, że szeregi losowe po dokonaniu redukcji szumu metodą NS mogą wykazywać obecność zależności nieliniowych. Naturalnie wniosek ten zasadniczo podważa zasadność stosowania analizowanej metody w procesie identyfikacji nieliniowości w rzeczywistych szeregach czasowych. Oznacza on bowiem, że zidentyfikowane zależności mogą nie pochodzić z systemu generującego lecz być efektem filtrowania danych. Z otrzymanego wniosku wynika, że w procesie analizy rzeczywistych szeregów czasowych metodą Schreibera należy stosować z dużą ostrożnością. Przede wszystkim nie należy stosować redukcji szumu, gdy brak jest wyraźnych podstaw do stwierdzenia, że analizowane dane pochodzą z systemu deterministycznego. Z otrzymanych wyników płynie też wniosek ogólniejszej natury – o potrzebie przebadania innych metod redukcji szumu pod kątem ich zachowania w zastosowaniu do szeregów stochastycznych.

Uniwersytet Mikołaja Kopernika

LITERATURA

- [1] Brock W.A., Dechert W.D., Scheinkman J.A. [1987], A test for independence based on the correlation dimension, SSRI Working Paper no. 8702, Department of Economics, University of Wisconsin, Madison.
- [2] Brock W.A., Hsieh D.A., LeBaron B. [1991], *Nonlinear Dynamics, Chaos, and Instability: Statistical Theory and Economic Evidence*, The MIT Press, Cambridge, Massachusetts, London.
- [3] Castagli M., Eubank S., Farmer J.D., Gibson J. [1991], State space reconstruction in the presence of noise, *Physica D*, 51, 52-98.
- [4] Dionisio A., Menezes R., Mendes D.A. [2003], Mutual Information: a dependence measure for nonlinear time series, working paper.
- [5] Fraser A.M., Swinney H.L. [1986], Independent coordinates for strange attractors from mutual information, *Physical Review A*, 33.2, 1134-1140.
- [6] Granger C. W. J., Terasvirta T. [1993], *Modelling nonlinear economic relationship*, Oxford University Press.
- [7] Granger C. W. J., Lin J-L. [1994], Using the mutual information coefficient to identify lags in nonlinear models, *Journal of Time Series Analysis*, 15, 371-384.
- [8] Harrison R.G., Yua D., Oxleyb L., Lua W., Georgec D. [1999], Non-linear noise reduction and detecting chaos: some evidence from the S&P Composite Price Index, *Mathematics and Computers in Simulation*, 48, 497-502.
- [9] Hassani H., Dionisio A., Ghodsi M. [2009a], The effect of noise reduction in measuring the linear and nonlinear dependency of financial markets, *Nonlinear Analysis: Real World Applications*, doi:10.1016/j.nonrwa.2009.01.004.

- [10] Hassani H., Zokaeib M., Rosenc D., Amiric S., Ghodsi M. [2009b], Does noise reduction matter for curve fitting in growth curve models?, *Computer methods and programs in biomedicine*, 96, 173-181.
- [11] Kantz H., Schreiber T., Hoffman I. [1993], Nonlinear noise reduction: A case study on experimental data, *Physical Review E*, 48.2, 1529-1538.
- [12] Kantz H., Schreiber T. [1997], *Nonlinear time series analysis*, Cambridge University Press
- [13] Kanzler L. [1999], Very Fast and Correctly Sized Estimation of the BDS Statistic, Department of Economics, Oxford University.
- [14] Leontitsis A., Bountis T., Pagge J. [2004], An adaptive way for improving noise reduction using local geometric projection, *Chaos* 14, 106-110.
- [15] Liang H., Lin Z., Yin F., [2005], Removal of ECG contamination from diaphragmatic EMG by nonlinear filtering, *Nonlinear Analysis*, 63, 745-753.
- [16] Maasoumi E., Racine J. [2002], Entropy and predictability of stock market returns, *Journal of Econometrics*, 107, 291-312.
- [17] Mees A. I., Judd K. [1993], Dangers of geometric filtering, *Physica D*, 68, 427-436.
- [18] Orzeszko W. [2008], The New Method of Measuring the Effects of Noise Reduction in Chaotic Data, *Chaos Solitons and Fractals*, 38, 1355-1368.
- [19] Perc M., [2005], Nonlinear time series analysis of the human electrocardiogram, *European Journal of Physics*, 26, 757-768.
- [20] Schreiber T. [1993], Extremely simple nonlinear noise-reduction method, *Physical Review E*, 47.4, 2401-2404.
- [21] Schreiber T. [1999], Interdisciplinary application of nonlinear time series methods, *Physics Reports*, 308, 1-64.
- [22] Shintani M., Linton O. [2001], Is There Chaos in the World Economy? A Nonparametric Test Using Consistent Standard Errors, working paper, Vanderbilt University Nashville.
- [23] Sivakumar B., Phoon K.-K., Liang S.-Y., Liaw C.-Y. [1999], A systematic approach to noise reduction in chaotic hydrological time series, *Journal of Hydrology*, 219, 103-135.
- [24] Takens F. [1981], Detecting Strange Attractors in Turbulence, in: *Dynamical Systems and Turbulence*, Rand D., Young L., eds., Springer-Verlag, 366-381.
- [25] Zeng X., Pielke R.A., Eykholt R. [1992], Extracting Lyapunov exponents from short time series of low precision, *Modern Physics Letters B*, 6, 55-75.

WPLYW REDUKCJI SZUMU LOSOWEGO METODĄ SCHREIBERA NA IDENTYFIKACJĘ SYSTEMU GENERUJĄCEGO DANE. ANALIZA SYMULACYJNA

Streszczenie

Jednym ze sposobów ograniczenia negatywnego wpływu obecności szumu losowego na analizę rzeczywistych szeregów czasowych jest stosowanie metod redukcji szumu. Prezentowane w literaturze przedmiotu rezultaty zastosowania takich procedur w procesie identyfikacji nieliniowości i chaosu są zachęcające. Jedną z metod redukcji szumu jest metoda Schreibera, która, jak wykazano, prowadzi do efektywnej redukcji szumu losowego dodanego do danych wygenerowanych z systemów deterministycznych o dynamice chaotycznej. Jednakże w przypadku danych rzeczywistych, badacz zwykle pozbawiony jest wiedzy, czy system generujący rzeczywiście jest deterministyczny. Istnieje więc ryzyko, że redukcji szumu zostaną wówczas poddane dane losowe. W niniejszym artykule wykazano, iż w sytuacji, gdy brak jest wyraźnych podstaw do stwierdzenia, że badany szereg pochodzi z systemu deterministycznego, metodę Schreibera należy stosować z dużą ostrożnością. Z przeprowadzonych symulacji, w których wykorzystano test BDS, miarę informacji wzajemnej oraz współczynnik korelacji liniowej Pearsona wynika bowiem,

że redukcja szumu może wprowadzić do analizowanych danych, zależności o charakterze nieliniowym. W efekcie szeregi losowe mogą zostać błędnie zidentyfikowane jako nieliniowe.

Słowa kluczowe: Identyfikacja nieliniowości, redukcja szumu losowego, metoda Schreibera, wskaźnik NRL, test BDS, miara informacji wzajemnej, bootstrap

IMPACT OF THE SCHREIBER NOISE REDUCTION METHOD ON DGP IDENTIFICATION. SIMULATION ANALYSIS

S u m m a r y

A presence of a noise is typical for real-world data. In order to avoid its negative impact on methods of time series analysis, noise reduction procedures may be used. The achieved results of an application of such procedures in identification of chaos or nonlinearity seem to be encouraging. One of the noise reduction methods is the Schreiber method, which, as it has been shown, is able to effectively reduce a noise added to time series generated by deterministic systems with chaotic dynamics. However, while analyzing real-world data, a researcher usually cannot be sure if the generating system is deterministic. Therefore, there is a risk that a noise reduction method will be applied to random data. In this paper, it has been shown that in situations where there is no clear evidence that investigated data are generated by a deterministic system, the Schreiber noise reduction method may negatively affect identification of time series. In the simulation carried out in this paper, the BDS test, the mutual information measure and the Pearson autocorrelation coefficient were used. The research has shown that an application of the Schreiber method may introduce spurious nonlinear dependencies to investigated data. As a result, random series may be misidentified as nonlinear.

Keywords: Nonlinear time series, noise reduction, Schreiber method, NRL quantity, BDS test, mutual information, bootstrap