Uniwersytet Ekonomiczny w Katowicach

Studia Ekonomiczne. Zeszyty Naukowe
Uniwersytetu Ekonomicznego w Katowicach
ISSN 2083-8611 Nr 227 · 2015

Monika Miśkiewicz-Nawrocka

Uniwersytet Ekonomiczny w Katowicach
Wydział Zarządzania
Katedra Matematyki
monika.miskiewicz@ue.katowice.pl

WPŁYW METODY REDUKCJI SZUMU LOSOWEGO NA DOKŁADNOŚĆ PROGNOZ EKONOMICZNYCH SZEREGÓW CZASOWYCH

Streszczenie: Od momentu pojawienia się w literaturze pojęcia deterministycznego chaosu można zaobserwować ogromny wzrost zainteresowania wielu badaczy teorią nieliniowych układów dynamicznych. Owo zainteresowanie zaowocowało pojawieniem się nowych metod predykcji szeregów czasowych, tj. metody największego wykładnika Lapunowa oraz metody najbliższych sąsiadów. Rzeczywiste szeregi czasowe są zwykle zaburzone przez szum losowy, który może komplikować problem ich prognozowania. Obecność szumu w danych może znacząco wpływać na jakość otrzymanych prognoz, dlatego głównym celem artykułu będzie ocena dokładności prognozowania szeregów czasowych poddanych procesowi redukcji szumu losowego oraz ocena efektywoności wybranej metody redukcji.

Słowa kluczowe: redukcja szumu losowego, współczynnik NRL, największy wykładnik Lapunowa, prognozowanie za pomocą największego wykładnika Lapunowa, metoda najbliższych sąsiadów.

Wprowadzenie

Rzeczywiste szeregi czasowe (s_t) składają się z części deterministycznej szeregu (y_t) oraz części stochastycznej szeregu (ε_t), która wyraża poziom szumu losowego. Redukcja szumu losowego pozwala poznać własności szeregu (y_t) na podstawie analizy szeregu obserwacji (s_t). W literaturze można znaleźć kilka metod służących do redukcji poziomu szumu losowego w układach dynamicznych, a podstawowym atutem stosowania tych metod wydaje się poprawa możliwości prognozowania szeregów czasowych [Miśkiewicz-Nawrocka, 2013a, 2013b]. Jedną z nich jest metoda najbliższych sąsiadów, która wywodzi się z teorii nieliniowych układów dynamicznych i została stworzona do prognozowania przyszłych wartości szeregów czasowych, jednak może być również stosowana do redukcji szumu losowego w szeregach czasowych. Efektywność przeprowadzonej filtracji szeregów czasowych można ocenić w dwojaki sposób, wykorzystując współczynnik redukcji szumu losowego *NRL* [Orzeszko, 2005] oraz jego zmodyfikowaną wersję *MNRL* [Orzeszko, 2008]. Wskaźniki te pozwalają spośród przefiltrowanych szeregów czasowych wybrać ten o najmniejszym poziomie szumu losowego.

Celem artykułu było zbadanie wpływu wyboru metody redukcji szumu losowego (odpowiedniego wskaźnika *NRL*) metodą najbliższych sąsiadów na dokładność prognoz otrzymanych w wyniku zastosowania metody największego wykładnika Lapunowa oraz metody najbliższych sąsiadów. Badania empiryczne przeprowadzono na podstawie rzeczywistych danych natury ekonomicznej – szeregów finansowych utworzonych z logarytmów dziennych stóp zwrotu cen zamknięcia wybranych indeksów giełd światowych. Dane obejmują okres od 03.01.2000 do 30.04.2014. Do przeprowadzenia niezbędnych obliczeń wykorzystano program napisany przez autora w języku Delhi, arkusz kalkulacyjny Excel oraz program TISEAN[1].

1. Rzeczywiste szeregi czasowe

Rzeczywiste szeregi czasowe można zdefiniować jako układy dynamiczne (X, f) opisane za pomocą następujących równań rekurencyjnych [Nowiński, 2007]:

$$x_{t+1} = f(x_t + \eta_t), \tag{1}$$

$$s_{t+1} = h(x_{t+1}) + \xi_t, \; t = 0,1,2,... \tag{2}$$

gdzie:

$X \subset R^m$, X – przestrzeń stanów,

$f : X \to X$ – funkcja opisująca rzeczywistą dynamikę układu,

$h : X \to R$ – funkcja pomiarowa generująca szereg czasowy obserwacji s_t układu dynamicznego,

$x_t, x_{t+1} \in X$ – stan nieznanego, pierwotnego układu wielowymiarowego odpowiednio w chwilach t, $t+1$,

[1] Darmowy program autorstwa H. Kantza i T. Schreibera.

s_{t+1} – obserwacja szeregu czasowego w chwili $t+1$,

η_t – szum dynamiczny wewnątrz układu,

ξ_t – szum pomiarowy.

W skrócie, rzeczywisty szereg czasowy można zatem wyrazić jako:

$$s_t = y_t + \varepsilon_t, \tag{3}$$

gdzie:

s_t – obserwacja szeregu czasowego w momencie *t*,

y_t – część deterministyczna szeregu czasowego,

ε_t – część stochastyczna szeregu czasowego (szum losowy składający się z szumu obserwacyjnego, systemowego lub ich kombinacji).

Główną przyczyną występowania szumu obserwacyjnego w szeregach czasowych są błędy pomiaru oraz błędy zaokrągleń, natomiast szumu systemowego – czynniki egzogeniczne wpływające na dynamikę układu, których identyfikacja jest niemożliwa [Stawicki, 1993].

2. Redukcja szumu losowego w rzeczywistych szeregach czasowych

Jedną ze stosowanych metod redukcji szumu losowego jest metoda najbliższych sąsiadów, której podstawą jest rekonstrukcja przestrzeni stanów [Takens, 1981]. Rekonstrukcja pozwala na podstawie jednowymiarowego szeregu czasowego obserwacji odtworzyć przestrzeń stanów układu dynamicznego. Jedną z najpopularniejszych[2] metod rekonstrukcji jest metoda opóźnień, która została wprowadzona niezależnie przez N.H. Packarda [Packard i in., 1980] oraz F. Takensa [1981]. Elementami zrekonstruowanej przestrzeni stanów są wektory opóźnień, tzw. *d*-historie postaci [np. Zawadzki, 1996]:

$$s_t^d = \left(s_t, s_{t-\tau}, \ldots, s_{t-(d-1)\tau}\right), \tag{4}$$

gdzie:

s_t – obserwacja szeregu czasowego w momencie *t*,

d – wymiar zanurzenia,

τ – opóźnienie czasowe, $(d-1)\tau + 1 \le t \le N$.

---

[2] Wśród innych metod rekonstrukcji można wyróżnić analizę czynnikową, wprowadzoną przez D.S. Broomheada i P. Kinga [Broomhead, King, 1986], oraz metodę pochodnych [Packard i in., 1980].

Wymiar zanurzenia d oraz opóźnienie czasowe τ mogą być szacowane na kilka sposobów, ponieważ w literaturze nie istnieje jedna jednoznaczna metoda wyznaczenia optymalnego zestawu tych parametrów, która byłaby odpowiednia dla wszystkich celów [Nowiński, 2007; Orzeszko, 2005].

Do szacowania wielkości opóźnienia czasowego najczęściej stosuje się algorytmy oparte na funkcji autokorelacji [np. Nowiński, 2007, s. 88]:

$$ACF(\tau) = \frac{1}{N-\tau-1}\sum_{t=0}^{N-\tau-1} x_t x_{t+\tau} = \langle x_t x_{t+\tau} \rangle, \tag{5}$$

gdzie:

$x_t, x_{t+\tau} \in X$ – stany układu odpowiednio w chwilach t, $t+\tau$.

Jako optymalną wartość opóźnienia wybiera się tę wartość τ, dla której funkcja $ACF(\tau)$ po raz pierwszy przyjmuje wartość zero [Ramsey i in., 1990, s. 999].

Wartość wymiaru zanurzenia powszechnie szacuje się metodą pozornych fałszywych najbliższych sąsiadów FNN, wprowadzoną przez M.B. Kennela, R. Browna, H.D.I. Abarbanela [1992; Cao, 2001]. Wektory są fałszywymi najbliższymi sąsiadami, gdy przestają ze sobą sąsiadować po zwiększeniu wymiaru przestrzeni. Jako optymalną wartość wymiaru zanurzenia przyjmuje się wartość d, dla której przy zwiększeniu wymiaru nie pojawią się żadni nowi fałszywi sąsiedzi [Abarbanel, 1996; Small, 2005].

Redukcja szumu losowego metodą najbliższych sąsiadów, polegająca na wyznaczaniu wartości y_n, $1 < n < N$ szeregu czasowego $(s_1, s_2, ..., s_N)$, przebiega według następującego algorytmu [Kantz, Schreiber, 2004]:

1. Dla oszacowanego wymiaru zanurzenia d oraz opóźnienia czasowego $\tau = 1$ tworzymy wektor opóźnień postaci:

$$s_t^d = (s_t, s_{t-1}, ..., s_{t-(d-1)}), \tag{6}$$

   tak aby filtrowana obserwacja s_n była jedną ze środkowych współrzędnych wektora s_t^d.

2. Wyznaczamy k najbliższych sąsiadów (w sensie odległości euklidesowej) wektora s_t^d, postaci:

$$s_{l(1)}^d, s_{l(2)}^d, ..., s_{l(k)}^d. \tag{7}$$

3. Na podstawie wyznaczonych sąsiadów, obliczamy wartość y_n jako średnią arytmetyczną pierwszych współrzędnych najbliższych sąsiadów:

$$y_n = \frac{1}{k}\sum_{i=1}^{k} s_{l(i)} \,. \tag{8}$$

Do oceny skuteczności stosowanej metody filtracji (redukcji szumu) można zastosować współczynnik poziomu redukcji szumu losowego *NRL*, który bada zależność pomiędzy siłą szumu dodawanego do układu a strukturą geometryczną jego atraktora. Współczynnik *NRL* dany jest wzorem [Orzeszko, 2005]:

$$NRL(d) = \frac{\sum_{i=1}^{T} d_i}{\sum_{i=1}^{T} D_i}, \tag{9}$$

gdzie:
d_i i D_i oznaczają odległości *i*-tego wektora opóźnień (*d*-historii) odpowiednio od jego najbliższego i najdalszego sąsiada.

Inną miarą służącą do oceny efektywności redukcji szumu losowego jest zmodyfikowana wersja powyższego wskaźnika – *MNRL* zaproponowana przez W. Orzeszko [2008]. Wskaźnik *MNRL* dany jest wzorem:

$$MNRL(d) = \frac{d_{\min} - d^0_{\min}}{d^0_{\min}} + \left|\frac{diam - diam^0}{diam^0}\right|, \tag{10}$$

gdzie:
$d_{\min} = \frac{1}{n}\sum_{i=1}^{n} d_i$,
d_i – odległość *i*-tego wektora opóźnień (*d*-historii) od jego najbliższego sąsiada,
diam – maksymalna odległość pomiędzy wektorami opóźnień przed redukcją szumu,
$diam^0$ – maksymalna odległość pomiędzy wektorami opóźnień po redukcji szumu.

Stosowanie powyższych kryteriów pozwala na wybór optymalnych szeregów czasowych, tj. szeregów o najniższym poziomie szumu losowego. W celu ustalenia takiego optymalnego szeregu przeprowadza się redukcję szumu losowego oryginalnego szeregu dla różnych wymiarów zanurzenia oraz różnej ilości najbliższych sąsiadów wektora s_t^d .

3. Prognozowanie szeregów czasowych

3.1. Metoda największego wykładnika Lapunowa LEM

Wykładniki Lapunowa są miarą wrażliwości układu dynamicznego na zmianę warunków początkowych. Zgodnie z definicją Devaneya [1987] i Wigginsa [1990] układ dynamiczny (X, f) jest wrażliwy na zmianę warunków początkowych, jeżeli istnieje liczba $\varepsilon > 0$, spełniająca warunek, że dla każdego $x \in X$ oraz dla każdego otoczenia U punktu x istnieją $y \in U$ oraz $n \geq 1$, takie że:

$$\left\| f^n(x) - f^n(y) \right\| > \varepsilon, \tag{11}$$

gdzie: f^n jest n-krotnym złożeniem odwzorowania f.

Wykładniki Lapunowa definiuje się jako granice [Zawadzki, 1996]:

$$\lambda_i(x_0) = \lim_{n \to \infty} \frac{1}{n} \ln\left|\mu_i(n, x_0)\right|,\ i = 1, ..., m, \text{ dla } m \geq 1, \tag{12}$$

gdzie:

$\mu_i(n, x_0)$ – są wartościami własnymi macierzy Jacobiego odwzorowania f^n,

f^n – n-krotne złożenie funkcji f,

f – funkcja generująca układ dynamiczny.

n-wymiarowy układ dynamiczny posiada n wykładników Lapunowa, jednak z praktycznego punktu widzenia najważniejszy jest największy z wykładników. Pozwala on na identyfikację chaosu w układach dynamicznych oraz na prognozowanie przyszłych stanów układów dynamicznych [Miśkiewicz-Nawrocka, 2012].

W praktyce, dla rzeczywistych szeregów czasowych, gdy nie jest znana postać funkcji generującej f, największy wykładnik Lapunowa szacuje się na podstawie zależności [Kantz, Schreiber, 2004]:

$$\Delta_n = \Delta_0 \cdot e^{n\lambda_{\max}}, \tag{13}$$

jako współczynnik kierunkowy równania regresji:

$$\ln\Delta_n = \ln\Delta_0 + \lambda_{\max} n, \tag{14}$$

gdzie:

Δ_0 – początkowa odległość pomiędzy dwoma początkowo bliskimi (w sensie metryki euklidesowej) punktami zrekonstruowanej przestrzeni stanów,

Δ_n – odległość pomiędzy tymi punktami po n iteracjach,

$\lambda_{\max}$ – największy wykładnik Lapunowa.

Lokalne wykładniki Lapunowa mierzą lokalne tempo rozbieżności (lub zbieżności) sąsiednich trajektorii, czyli lokalną chaotyczność układu dynamicznego. Największy lokalny wykładnik Lapunowa pozwala określić jak bardzo zmienia się (zwiększa się lub zmniejsza się) odległość pomiędzy bieżącym stanem x_N układu a jego najbliższym sąsiadem x_i podczas ewolucji układu, a także oszacować odległość pomiędzy wektorami x_{N+1} i x_{i+1}. Na podstawie tej odległości wyznacza się wartość prognozy $\hat{x}_{N+1}$ [Guégan, Leroux, 2009; Zhang i in., 2004].

W przypadku rzeczywistych szeregów czasowych, gdy dysponujemy jedynie jednowymiarowym szeregiem czasowym złożonym z N obserwacji $(s_1, s_2, ..., s_N)$, wyznaczenie prognozy za pomocą największego wykładnika Lapunowa wymaga wcześniejszego przeprowadzenia rekonstrukcji przestrzeni stanów. Spośród wszystkich wektorów s_t^d zrekonstruowanej przestrzeni stanów wybieramy wektor najbliższy (w sensie odległości euklidesowej) wektorowi s_N^d i oznaczamy przez $s_{\min}^d$. Jeśli $\Delta_{\min}$ oznacza odległość pomiędzy wektorami s_N^d i $s_{\min}^d$, a Δ_1 – odległość pomiędzy wektorami s_{N+1}^d i $s_{\min+1}^d$, oraz przyjmując założenie, że $\Delta_1 / \Delta_{\min}$ ulega małym zmianom podczas ewolucji układu, to odległość między wektorami s_{N+1}^d i $s_{\min+1}^d$ wyraża się wzorem [Guégan, Leroux, 2009]:

$$\Delta_1 \approx \Delta_{\min} \cdot e^{\lambda_{\max}}, \tag{15}$$

gdzie: $\lambda_{\max}$ jest wykładnikiem Lapunowa.

Ponieważ:

$$s_{N+1}^d = \left(s_{N+1}, s_{N-\tau+1}, ..., s_{N-(d-1)\tau+1}\right), \tag{16}$$

prognozowaną wartość s_{N+1} można wyznaczyć z równania (15) jako rozwiązanie równania postaci:

$$\left(z - s_{i+1}\right)^2 + \left(s_N - s_i\right)^2 + ... + \left(s_{N-(d-1)\tau+1} - s_{i-(d-1)\tau+1}\right)^2 - \left(\Delta_{\min} e^{\lambda_{\max}}\right)^2 = 0. \tag{17}$$

Prognoza $\hat{s}_{N+1}$ może więc przyjmować dwie wartości [Miśkiewicz-Nawrocka, 2012]: $\hat{s}_{N+1}^+$ oraz $\hat{s}_{N+1}^-$, będące odpowiednio przeszacowaną i niedoszacowaną wartością rzeczywistego s_{N+1}.

Kolejne prognozy $\hat{s}_{N+T}$, dla horyzontu prognozy T = 2, 3, … można wyznaczyć bezpośrednio z zależności:

$$\Delta_T \approx \Delta_{\min} \cdot e^{\lambda_{\max} \cdot T}, \tag{18}$$

gdzie Δ_T oznacza odległość pomiędzy wektorami s_N^d i $s_{\min}^d$ po T krokach iteracji, czyli pomiędzy s_{N+T}^d i $s_{\min+T}^d$, lub metodą iteracyjną, stosując opisaną powyżej procedurę dla wektora s_{N+1}^d [Zhang i in., 2004; Guégan, Leroux, 2009].

3.2. Metoda najbliższych sąsiadów NNM

Podstawą teoretyczną metody najbliższych sąsiadów jest fakt, że stany układów deterministycznych ewoluują w czasie w podobny sposób. W przypadku rzeczywistych szeregów czasowych, gdy nie znamy funkcji f opisującej dynamikę układu i dysponujemy tylko jednowymiarowym szeregiem obserwacji $(s_1,...,s_N)$, należy w pierwszej kolejności przeprowadzić rekonstrukcję przestrzeni stanów. Jeśli $s_{t_0}^d$ jest najbliższym sąsiadem punktu s_N^d, to również $f_T(s_N^d) \approx f_T(s_{t_0}^d)$, a stąd wynika, że $s_{N+T} \approx s_{t_0+T}$. Wartość s_{t_0+T} można zatem przyjąć jako prognozę obserwacji s_{N+T} analizowanego szeregu czasowego [np. Nowiński, 2007].

W metodzie najbliższych sąsiadów prognozę dla $N+1$ elementu $\hat{s}_{N+1}$ szacuje się jako średnią ważoną obserwacji s_{i+1}, gdzie wektory s_i^d są k najbliższymi sąsiadami wektora s_N^d w zrekonstruowanej d-wymiarowej przestrzeni stanów [Kantz, Schreiber, 2004]:

$$\hat{s}_{N+1} = \sum_{i=1}^{k} w_i s_{i+1}, \tag{19}$$

gdzie:

w_i – waga i-tego najbliższego sąsiada,

s_{i+1} – pierwsza współrzędna wektora s_{i+1}^d w zrekonstruowanej przestrzeni stanów.

Wagi są dobierane w ten sposób, aby bliżsi sąsiedzi mieli większy wpływ na otrzymaną prognozę. Wagę i-tego sąsiada można więc wyznaczyć według wzoru [Orzeszko, 2005; Chun, Kim, Kim, 2002]:

$$w_i = \frac{1}{k-1}\left(1 - \frac{d_i}{d_{TOT}}\right), \tag{20}$$

gdzie:

$d_i = \left\| s_N^d - s_i^d \right\|$ oznacza odległość między wektorami s_N^d i s_i^d,

$d_{TOT} = \sum_{i=1}^{k} d_i$, $i = 1, 2, ..., k$.

4. Wpływ redukcji szumu losowego na wyniki prognoz – badanie empiryczne

Przedmiotem badania były logarytmy dziennych stóp zwrotu indeksów giełd światowych: AEX – indeks giełdy w Amsterdamie, ATH – indeks na Giełdzie Papierów Wartościowych w Atenach, BEL 20 – indeks na giełdzie Euronext w Brukseli, CAC40 – indeks na Giełdzie Papierów Wartościowych w Paryżu (CAC), DAX – niemiecki indeks giełdowy, FTSE250 – indeks na Giełdzie Papierów Wartościowych w Londynie (FTM), IBEX – indeks na Giełdzie Papierów Wartościowych w Madrycie, PX – indeks na Giełdzie Papierów Wartościowych w Pradze, SAX – indeks na Giełdzie Papierów Wartościowych w Bratysławie oraz WIG20 – indeks na Giełdzie Papierów Wartościowych w Warszawie, postaci:

$$x_t = \ln s_t - \ln s_{t-1}, \tag{21}$$

gdzie: s_t – obserwacja szeregu, notowane w okresie 3.01.2000–30.04.2014[3].

W pierwszym etapie badania, dla wybranych szeregów czasowych oszacowano parametry rekonstrukcji przestrzeni stanów. Stosując funkcję autokorelacji – *ACF*, oszacowano czas opóźnień τ, natomiast za pomocą metody najbliższego pozornego sąsiada – *FNN*, obliczono wymiar zanurzenia *d* (tab. 2). Następnie analizowane szeregi czasowe poddano procesowi redukcji szumu losowego metodą najbliższych sąsiadów dla opóźnienia czasowego $\tau = 1$. W celu ustalenia optymalnych parametrów, tj. parametrów, dla których poziom szumu losowego jest najniższy, pod uwagę wzięto następujące wartości wymiaru zanurzenia $d = 2, 3, 4, 5, 6, 8, 10, 15$, natomiast jako najbliższych sąsiadów wektora x_t^d ustalono wektory z jego otoczenia o promieniu $\rho = 0{,}001; 0{,}01; 0{,}1$[4]. Do oceny poziomu szumu w przefiltrowanych szeregach zastosowano współczynniki *NRL* oraz *MNRL*. W tab. 1 przedstawiono najniższe wartości miar *NRL* i *MNRL* obliczone dla analizowanych szeregów czasowych wraz z odpowiadającymi im wartościami wymiaru zanurzenia *d* i promienia otoczenia wektora x_t^d.

---

[3] Dane pochodzą z archiwum plików strony internetowej stooq.com.

[4] Redukcję szumu przeprowadzono przy wykorzystaniu darmowego programu TISEAN autorstwa H. Kantza i T. Schreibera.

Tab. 1. Wartości współczynników *NRL* i *MNRL* dla analizowanych szeregów czasowych

Szereg	Parametry redukcji		*NRL*	Parametry redukcji		*MNRL*
	d	ρ		d	ρ	
AEX	2	0,1	0,00047	3	0,1	–1,79322
ATH	3	0,1	0,00112	2	0,1	–1,63919
BEL 20	2	0,1	0,00038	2	0,1	–1,83709
CAC 40	2	0,1	0,00056	2	0,1	–1,79705
DAX	2	0,1	0,00068	2	0,1	–1,84155
FTM	3	0,1	0,00053	2	0,1	–1,93283
IBEX	3	0,1	0,00086	3	0,1	–1,69669
PX	3	0,1	0,00077	2	0,1	–1,50632
SAX	4	0,1	0,00081	3	0,1	–1,59959
WIG20	3	0,1	0,00042	3	0,1	–1,84161

Źródło: [Miśkiewicz-Nawrocka, Zeug-Żebro, 2014].

Analizując dane zawarte w tab. 1, można zauważyć, że najniższym poziomem szumu losowego charakteryzują się szeregi, dla których w procesie redukcji szumu pod uwagę wzięto najbliższych sąsiadów z otoczenia wektora x_t^d o promieniu równym 0,1 oraz wymiarze zanurzenia 2 lub 3, a w przypadku szeregu SAX – wymiarze 4.

W dalszej analizie przefiltrowane szeregi czasowe oznaczono symbolami *NazwaSzeregu*_red1 oraz *NazwaSzeregu*_red2 dla szeregów uzyskanych odpowiednio za pomocą współczynnika *NRL* oraz *MNRL*. Symbol *NazwaSzeregu_red* oznacza natomiast przefiltrowane szeregi, dla których stosując miary *NRL* i *MNRL* uzyskano dokładnie te same szeregi. Dla przefiltrowanych szeregów również przeprowadzono rekonstrukcję przestrzeni stanów. Tab. 2 zawiera parametry rekonstrukcji d i τ dla analizowanych szeregów czasowych przed i po procesie redukcji szumu.

Tab. 2. Parametry rekonstrukcji przestrzeni stanów dla badanych szeregów czasowych

Szereg	Czas opóźnień τ	Wymiar zanurzenia d	Szereg	Czas opóźnień τ	Wymiar zanurzenia d
AEX	2	6	FTM	9	7
AEX_red1	10	2	FTM_red1	2	2
AEX_red2	10	2	FTM_red2	1	2
ATH	20	8	IBEX	19	7
ATH_red1	2	2	IBEX_red	4	2
ATH_red2	2	6	PX	11	7
BEL20	24	6	PX_red1	4	2
BEL20_red	4	3	PX_red2	4	7
CAC40	20	8	SAX	13	12
CAC40_red	1	8	SAX_red1	4	2
DAX	13	7	SAX_red2	4	2
DAX_red	8	7	WIG20	3	7
			WIG20_red	2	2

Z oszacowanych parametrów d i τ (tab. 2) wynika, że w większości badanych szeregów proces redukcji szumu pozwala zmniejszyć wymiar zanurzenia d zrekonstruowanej przestrzeni stanów dla obu współczynników *NRL*. Wyjątek stanowią szeregi CAC40, DAX i PX, dla których wymiar zanurzenia przefiltrowanych szeregów pozostał na tym samym poziomie.

W celu wyznaczenia prognoz za pomocą największego wykładnika Lapunowa, oszacowano wartości największego wykładnika Lapunowa dla badanych szeregów czasowych. Oszacowane wartości wraz z równaniem regresji przedstawiono w tab. 3. Znakiem „–” oznaczono sytuacje, w których wartość współczynnika determinacji ($R^2 < 0{,}3$) nie pozwoliła traktować w wiarygodny sposób oszacowanego współczynnika kierunkowego równania regresji jako wartości największego wykładnika Lapunowa.

Tab. 3. Wykładniki Lapunowa dla badanych szeregów czasowych

Szereg	Równanie regresji	Największy wykładnik Lapunowa
1	2	3
AEX	$y = 0{,}0021x - 4{,}2778$ $R^2 = 0{,}7598$	0,0021
AEX_red1	$y = 0{,}0079x - 9{,}0468$ $R^2 = 0{,}1642$	–
AEX_red2	$y = 0{,}0043x - 9{,}063$ $R^2 = 0{,}3318$	0,0043
ATH	$y = 0{,}0029x - 4{,}0993$ $R^2 = 0{,}6387$	0,0029
ATH_red1	$y = 0{,}0298x - 8{,}4577$ $R^2 = 0{,}4371$	0,0298
ATH_red2	$y = 0{,}0435x - 8{,}5471$ $R^2 = 0{,}3138$	0,0435
BEL20	$y = 0{,}0012x - 4{,}4082$ $R^2 = 0{,}5347$	0,0012
BEL20_red	$y = 0{,}0389x - 10{,}103$ $R^2 = 0{,}5304$	0,0389
CAC40	$y = 0{,}0011x - 4{,}2342$ $R^2 = 0{,}0936$	–
CAC40_red	$y = 0{,}0201x - 9{,}2654$ $R^2 = 0{,}8053$	0,0201
DAX	$y = -0{,}0024x - 4{,}1855$ $R^2 = 0{,}4135$	–0,0024
DAX_red	$y = 0{,}0097x - 9{,}2418$ $R^2 = 0{,}4685$	0,0097
FTM	$y = 0{,}0031x - 4{,}5538$ $R^2 = 0{,}3947$	0,0031
FTM_red1	$y = -0{,}0034x - 12{,}983$ $R^2 = 0{,}601$	–0,0034
FTM_red2	$y = -0{,}0026x - 13{,}143$ $R^2 = 0{,}4971$	–0,0026
IBEX	$y = 0{,}0015x - 4{,}2141$ $R^2 = 0{,}2441$	0,0015

cd. tab. 3

1	2	3
IBEX_red	$y = -0{,}0259x - 9{,}0399$ $R^2 = 0{,}6455$	–0,0259
PX	$y = 0{,}0012x - 4{,}2872$ $R^2 = 0{,}568$	0,0012
PX_red1	$y = 0{,}0201x - 8{,}5591$ $R^2 = 0{,}5531$	0,0201
PX_red2	$y = 0{,}0094x - 8{,}556$ $R^2 = 0{,}1033$	0,0094
SAX	$y = 0{,}0036x - 4{,}5889$ $R^2 = 0{,}4802$	0,0036
SAX_red1	$y = 0{,}0437x - 9{,}3449$ $R^2 = 0{,}4745$	0,4745
SAX_red2	$y = 0{,}0649x - 9{,}507$ $R^2 = 0{,}4653$	0,0649
WIG20	$y = 0{,}001x - 4{,}1345$ $R^2 = 0{,}3052$	0,001
WIG20_red	$y = 0{,}0674x - 9{,}9651$ $R^2 = 0{,}4128$	0,0674

Z oszacowanych danych przedstawionych w tab. 3 wyraźnie widać, że wartości największego wykładnika Lapunowa dla przefiltrowanych szergów są znacznie większe niż dla szeregów przed redukcją. Dodatnie wartości największego wykładnika Lapunowa wskazują na istnienie chaosu w badanych szeregach czasowych, a proces filtrowania szeregu powoduje zwiększenie poziomu chaosu. Wyjątek stanowią szeregi DAX, FTM i IBEX, dla których proces redukcji szumu spowodował zmianę znaku największego wykładnika Lapunowa, a zatem i charakteru badanego zjawiska. Dla szeregu DAX proces filtrowania szeregu okazał się korzystniejszy, ponieważ nastapiła zmiana znaku z „–“ na „+“, która ujawniła choatyczny charakter badanego zjawiska, co pozwala na wyznaczenie prognoz metodą opartą na największym wykładniku Lapunowa. W przypadku szeregów AEX_red1 i CAC40 oszacowany współczynnik kierunkowy równania regresji nie może być traktowany jako wartość największego wykładnika Lapunowa ze względu na niską wartość współczynnika determinacji.

W kolejnym etapie badań wyznaczono prognozy metodą opartą na wartości największego wykładnika Lapunowa LEM oraz metodą najbliższych sąsiadów NNM. W celu wyznaczenia prognoz metodą NNM wzięto pod uwagę $k = 2(d+1)$ najbliższych sąsiadów punktu x_n^d. Oceny dokładności wyznaczonych prognoz dokonano następującymi miernikami: *d* – średnim błędem prognozy *ME*, *q* – średnim absolutnym błędem prognozy *MAE*, σ – pierwiastkiem błędu średniokwadratowego *RMSE*, σ’– względnym błędem prognozy oraz *I* – współczynnikiem Theila.

W tab. 4-5 przedstawiono błędy predykcji w całym przedziale weryfikacji dla horyzontu prognozy równego 10, uzyskane metodą LEM dla przeszacowanych (LEM+) i niedoszacowanych (LEM–) prognoz. Znakiem „–” oznaczono sytuacje, w których wyznaczenie prognozy metodą LEM było niemożliwe ze względu na ujemną lub niewiarygodną wartość największego wykładnika Lapunowa.

Tab. 4. Błędy prognoz w całym przedziale weryfikacji uzyskanych metodą LEM+ dla badanych szeregów czasowych

Szereg	d	q	σ	σ'	I
AEX	–0,01709	0,01998	0,02265	1,50335	11,63428
AEX_red1	–	–	–	–	–
AEX_red2	0,00245	0,00058	0,00167	1,87774	1,02587
ATH	–0,00010	0,00010	0,00013	0,14823	0,47969
ATH_red1	0,01471	0,00318	0,00966	6,31582	1,00679
ATH_red2	–0,00097	0,00022	0,00025	0,17835	0,35050
BEL	–0,02815	0,02858	0,03152	2,40726	15,97075
BEL_red	–0,00021	0,00012	0,00013	0,20488	29,14826
CAC	–	–	–	–	–
CAC_red	–0,00057	0,00013	0,00014	0,17571	2,08460
DAX	–	–	–	–	–
DAX_red	–0,00047	0,00010	0,00012	0,19270	2,32322
FTM	–0,01981	0,02207	0,02658	2,40251	13,88767
FTM_red1	–	–	–	–	–
FTM_red2	–	–	–	–	–
IBEX	–0,02807	0,02807	0,03159	2,05311	10,43686
IBEX_red	–	–	–	–	–
PX	–0,03629	0,03629	0,03805	2,56992	35,06027
PX_red1	0,00023	0,00013	0,00028	0,10741	0,49588
PX_red2	–0,00042	0,00008	0,00010	0,04490	0,23653
SAX	–0,03999	0,04217	0,04785	3,93994	7,50442
SAX_red1	0,00485	0,00114	0,00326	2,44370	0,94125
SAX_red2	0,00475	0,00111	0,00326	2,45612	0,94081
WIG20	–0,02996	0,03533	0,03700	2,30921	19,48153
WIG20_red	–0,00295	0,00061	0,00165	3,03145	1,07620

Tab. 5. Błędy prognoz w całym przedziale weryfikacji uzyskanych metodą LEM– dla badanych szeregów czasowych

Szereg	d	q	σ	σ'	I
AEX	0,02237	0,02237	0,02513	1,66822	14,32594
AEX_red1	–	–	–	–	–
AEX_red2	0,00056	0,00058	0,00167	1,87733	1,02541
ATH	0,00013	0,00014	0,00018	0,20431	0,91125
ATH_red1	0,00311	0,00319	0,00981	6,41328	1,03810
ATH_red2	0,00019	0,00019	0,00024	0,16515	0,30055
BEL20	0,02658	0,02658	0,02864	2,18738	13,18636
BEL20_red	0,00015	0,00015	0,00019	0,29880	62,00107
CAC40	–	–	–	–	–
CAC40_red	0,00012	0,00013	0,00014	0,17207	1,99916
DAX	–	–	–	–	–
DAX_red	0,00016	0,00016	0,00018	0,29784	5,55002
FTM	0,01942	0,02033	0,02421	2,18850	11,52373
FTM_red1	–	–	–	–	–
FTM_red2	–	–	–	–	–
IBEX	0,02583	0,02583	0,02802	1,82082	8,20872
IBEX_red1	–	–	–	–	–
PX	0,02864	0,02864	0,03227	2,17931	25,21238
PX_red1	0,00010	0,00013	0,00029	0,11269	0,54584
PX_red2	0,00009	0,00009	0,00010	0,04373	0,22434
SAX	0,04147	0,04147	0,04507	3,71158	6,65970
SAX_red1	0,00110	0,00113	0,00326	2,44429	0,94170
SAX_red2	0,00105	0,00110	0,00326	2,45770	0,94202
WIG20	0,03011	0,03011	0,03254	2,03073	15,06601
WIG20_red	–0,00049	0,00055	0,00155	2,85466	0,95433

Analizując błędy prognoz otrzymane metodą LEM+ oraz LEM– (tab. 4-5), należy zauważyć, że dla większości branych pod uwagę szeregów (AEX, BEL20, PX, SAX, WIG20) błędy prognoz uzyskane metodą LEM+ oraz LEM– dla szeregów przefiltrowanych są zdecydowanie mniejsze od błędów uzyskanych dla szeregów przed redukcją. Wyjątek stanowi szereg ATH, który charakteryzuje się mniejszymi błędami prognoz przed filtracją. Ze względu na niewiarygodne wartości największego wykładnika Lapunowa dla szeregów AEX_red1 i CAC40 oraz ujemne wartości największego wykładnika Lapunowa dla szeregów DAX, FTM_red1, FTM_red2 i IBEX_red nie można porównać dokładności prognoz tych szeregów przed i po redukcji.

Porównując wyniki błędów prognoz uzyskane metodami LEM+ oraz LEM– dla szeregów przefiltrowanych za pomocą współczynnika *NRL* oraz *MNRL* (szeregi ATH, PX, SAX w tab. 4-5), można stwierdzić, że dokładniejsze wyniki uzyskano dla szeregów filtrowanych z użyciem miary *MNRL*. Błędy prognozy dla szeregów *NazwaSzeregu*_red2 są zdecydowanie mniejsze niż dla szeregów *NazwaSzeregu*_red1, z wyjątkiem szeregów SAX_red1 i SAX_red2, dla których wartości błędów są bardzo zbliżone.

W tab. 6 przedstawiono błędy predykcji w całym przedziale weryfikacji dla horyzontu prognozy równego 10, uzyskane metodą NNM.

Tab. 6. Błędy prognoz w całym przedziale weryfikacji uzyskanych metodą NNM dla badanych szeregów czasowych

Szereg	d	q	σ	σ'	I
AEX	0,00300	0,00644	0,00719	0,47751	1,17378
AEX_red1	0,00012	0,00005	0,00006	0,06435	0,14745
AEX_red2	0,00265	0,00058	0,00169	1,89630	1,04624
ATH	–0,00214	0,00293	0,00364	4,09826	366,66330
ATH_red1	0,01522	0,00315	0,00968	6,32765	1,01057
ATH_red2	–0,00013	0,00012	0,00013	0,09442	0,09824
BEL20	0,00068	0,00673	0,00881	0,67281	1,24758
BEL20_red	–0,00008	0,00009	0,00016	0,24508	41,70855
CAC40	0,00077	0,00699	0,00806	0,52906	0,86057
CAC40_red	0,00006	0,00005	0,00005	0,06220	0,26119
DAX	0,00089	0,01090	0,01266	0,80512	1,01909
DAX_red	0,00015	0,00005	0,00005	0,08889	0,49428
FTM	0,00203	0,00705	0,00796	0,71904	1,24397
FTM_red1	–0,00251	0,00055	0,00166	8,22344	1,07690
FTM_red2	–0,00003	0,00001	0,00001	0,06547	0,00223
IBEX	0,00434	0,00919	0,01149	0,74686	1,38109
IBEX_red	0,00921	0,00662	0,00897	0,58306	1,06094
PX	0,00202	0,00414	0,00571	0,38546	0,78873
PX_red1	0,00038	0,00012	0,00026	0,10086	0,43725
PX_red2	0,00000	0,00002	0,00003	0,01392	0,02272
SAX	0,00323	0,01234	0,01784	1,46875	1,04288
SAX_red1	0,00692	0,00329	0,00546	4,08723	2,63309
SAX_red2	0,00879	0,00297	0,00533	4,01509	2,51415
WIG20	–0,00043	0,00863	0,01085	0,67736	1,67624
WIG20_red	–0,00104	0,00078	0,00166	3,06279	1,09856

Na podstawie oszacowanych błędów prognoz metodą NNM przedstawionych w tab. 5 można zauważyć, że we wszystkich rozważanych przypadkach błędy prognoz uzyskane dla szeregów po redukcji szumu są zdecydowanie mniejsze niż dla szeregów przed redukcją. Oceniając efektywność prognozowania ze względu na wpływ miar *NRL* i *MNRL*, należy stwierdzić, że dla prawie wszystkich badanych szeregów redukcja z wykorzystaniem współczynnika *MNRL* dała lepsze rezultaty, tj. dokładniejsze prognozy charakteryzujące się mniejszymi błędami. Wyjątek stanowi szereg AEX, dla którego najdokładniejsze prognozy uzyskano po filtracji za pomocą miary *NRL*. Oczywiście w przypadku szeregów BEL20, CAC40, DAX, IBEX i WIG20 należy wnioskować, że zarówno współczynnik *NRL*, jak i współczynnik *MNRL* dały tak samo dobre (te same) wyniki prognoz, ponieważ w obu przypadkach w procesie filtracji wyłoniono te same szeregi czasowe.

Podsumowanie

W opracowaniu zbadano wpływ metod redukcji szumu losowego ze względu na wybrany współczynnik *NRL* lub *MNRL* na dokładność prognoz wybranych finansowych szeregów czasowych. Na podstawie przeprowadzonych badań można stwierdzić, że dla prawie wszystkich badanych szeregów błędy prognoz *ex post* uzyskane dla szeregów, w których zastosowano redukcję szumu są zdecydowanie niższe od prognoz otrzymanych dla szeregów nieprzefiltrowanych. Porównując dokładność prognozowania ze względu na wpływ miar *NRL* i *MNRL*, należy stwierdzić, że dla prawie wszystkich badanych szeregów proces redukcji szumu z wykorzystaniem współczynnika *MNRL* przyniósł dokładniejsze prognozy, charakteryzujące się mniejszymi błędami.

Dodatkowo należy zaznaczyć, że w niektórych przypadkach redukcja szumu losowego na podstawie współczynnika *NRL* pozwoliła uzyskać dokładnie te same wyniki co proces redukcji z wykorzystaniem jego zmodyfikowanej wersji *MNRL*.

Warto zauważyć, że wartości prognoz wyznaczonych wspomnianymi metodami w dużej mierze zależą od przyjętej metryki, wagi najbliższego sąsiada, wartości parametrów zrekonstruowanej przestrzeni stanów oraz liczby najbliższych sąsiadów. Wydaje się zatem, że w celu poprawy jakości tych prognoz można przeprowadzić dodatkowe badania, zmieniając wspomniane parametry.

Literatura

Abarbanel H.D. (1996), *Analysis of Observed Chaotic Data*, Springer Verlag, Berlin, Heidelberg, New York.

Broomhead D.S., King P. (1986), *Extracting Qualitative Dynamics from Experimental Data*, „Physica D", Vol. 20, s. 217-236.

Cao L. (2001), *Method of False Nearest Neighbors* [w:] A.S. Soofi, L. Cao, red., *Modeling and Forecasting Financial Data*, Kluwer, Boston.

Chun S.H., Kim K.J., Kim S.H. (2002), *Chaotic Analysis of Predictability versus Knowledge Discovery Techniques: Case Study of the Polish Stock Market*, „Expert Systems", Vol. 19, No. 5, s. 264-272.

Devaney R.L. (1987), *An Introduction to Chaotic Dynamical Systems*, Addison-Wesley Publishing Company, Redwood City.

Guégan D., Leroux J. (2009), *Forecasting Chaotic Systems: The Role of Local Lyapunov Exponents*, „Chaos, Solitons & Fractals", Vol. 41, s. 2401-2404.

Kantz H., Schreiber T. (2004), *Nonlinear Time Series Analysis*, Cambridge University Press, Cambridge.

Kennel M.B., Brown R., Abarbanel H.D.I. (1992), *Determining Embedding Dimension for Phase Space Reconstruction Using a Geometrical Construction*, „Physical Review A" 1992, Vol. 45(6), s. 3404-3411.

Miśkiewicz-Nawrocka M. (2012), *Zastosowanie wykładników Lapunowa do analizy ekonomicznych szeregów czasowych*, Wydawnictwo UE, Katowice.

Miśkiewicz-Nawrocka M. (2013a), *The Application of Random Noise Reduction by Nearest Neighbor Method of Forecasting of Economic Time Series*, "Folia Oeconomica Stetinensia", 13(21) 2013/2, Wydawnictwo Naukowe Uniwersytetu Szczecińskiego, s. 96-108.

Miśkiewicz-Nawrocka M. (2013b), *Wpływ redukcji szumu losowego metoda najbliższych sąsiadów na wyniki prognoz otrzymanych za pomocą największego wykładnika Lapunowa*, „Studia Ekonomiczne", nr 159, Wydawnictwo UE, Katowice, s. 82-98.

Miśkiewicz-Nawrocka M., Zeug-Żebro K. (2014), *The Effect of the NRL Indicator on the Accuracy of Financial Series Forecasts*, Conference Proceedings. 32nd International Conference Mathematical Methods in Economics MME 2014, Olomouc.

Nowiński M. (2007), *Nieliniowa dynamika szeregów czasowych*, Wydawnictwo AE, Wrocław.

Orzeszko W. (2005), *Identyfikacja i prognozowanie chaosu deterministycznego w ekonomicznych szeregach czasowych*, PTE, Warszawa.

Orzeszko W. (2008), *The New Method of Measuring the Effects of Noise Reduction in Chaotic Data*, „Chaos, Solitons and Fractals", 38, s. 1355-1368.

Packard N.H., Crutchfield J.P., Farmer J.D., Shaw R.S. (1980), *Geometry from a Time Series*, „Physical Review Letters", Vol. 45, s. 712-716.

Ramsey J.B., Sayers C.L., Rothman P. (1990), *The Statistical Properties of Dimension Calculations Using Small Data Sets: Some Economic Applications*, „International Economic Review", 31, No. 4.

Small M. (2005), *Applied Nonlinear Time Series Analysis. Applications in Physics, Physiology and Finance*, „World Scientific Series on Nonlinear Science", Series A, Vol. 52, World Scientific Publishing Co. Pte. Ltd., Singapore.

Stawicki J. (1993), *Metody filtracji w modelowaniu procesów ekonomicznych*, Wydawnictwo Uniwersytetu Mikołaja Kopernika, Toruń.

Takens F. (1981), *Detecting Strange Attractors in Turbulence* [w:] D.A. Rand, L.S. Young, *Lecture Notes in Mathematics*, Springer, Berlin, s. 366-381.

Wiggins S. (1990), *Introduction to Applied Nonlinear Dynamical Systems and Chaos*, Springer-Verlag, New York, Berlin, Heidelberg.

Zawadzki H. (1996), *Chaotyczne systemy dynamiczne*, Wydawnictwo AE, Katowice.

Zhang J., Lam K.C., Yan W.J., Gao H., Li Y. (2004), *Time Series Prediction Using Lyapunov Exponents in Embedding Phase Space*, „Computers and Electrical Engineering", 30, s. 1-15.

THE EFFECT OF RANDOM NOISE REDUCTION METHOD ON THE ACCURACY OF FORECASTING ECONOMIC TIME SERIES

Summary: Since the deterministic chaos appeared in the literature, we have observed a huge increase of interest in nonlinear dynamic systems theory among researchers. This interest has led to the creation of new methods of time series prediction, e.g. the method of the largest Lyapunov exponent and the nearest neighbors. Real time series are usually disturbed by random noise, which can complicate the problem of time series forecasting. As the presence of noise in the data can significantly affect the quality of forecasts, the aim of the article is to evaluate the accuracy of predicting the time series filtered using the nearest neighbor method and the effectiveness of the chosen method of reduction.

Keywords: random noise reduction, NRL indicator largest Lyapunov exponent, Lyapunov exponent method of prediction, nearest neighbor method, state space reconstruction.