

Katarzyna Zeug-Żebro

Uniwersytet Ekonomiczny w Katowicach

WPŁYW REDUKCJI POZIOMU SZUMU LOSOWEGO METODĄ NAJBLIŻSZYCH SĄSIADÓW NA DOKŁADNOŚĆ PROGNOZ FINANSOWYCH SZEREGÓW CZASOWYCH

Wstęp

Rzeczywisty szereg czasowy x_t jest zwykle zaburzony przez szum losowy. Jego źródłem mogą być błędy pomiaru oraz błędy zaokrągleń powstałe podczas analizy danych. Szum losowy może również reprezentować czynniki egzogeneiczne wpływające na dynamikę układu lub być konsekwencją statystycznego charakteru zjawisk, np. na które wpływ mają ludzkie decyzje. Zatem każdy rzeczywisty szereg czasowy x_t można podzielić na dwie części: deterministyczną \bar{x}_t i stochastyczną ξ_t , a następnie zapisać w postaci addytywnej:

$$x_t = \bar{x}_t + \xi_t, \quad (1)$$

gdzie ξ_t posiada szybko malejącą funkcję autokorelacji i jest nieskorelowana z \bar{x}_t .

W literaturze można spotkać wiele procedur redukujących poziom szumu, między innymi metodę największej wiarygodności, cieniowanie oraz metodę najbliższych sąsiadów. Jedną z zalet wynikającą z użycia tych metod jest poprawa możliwości prognozowania.

Celem artykułu będzie ocena zastosowania redukcji poziomu szumu, w wybranych szeregach finansowych, metodą najbliższych sąsiadów oraz wpływu tej filtracji na dokładność prognoz. W badaniach zostaną wykorzystane szeregi utworzone z cen zamknięcia trzech spółek notowanych na GPW w Warszawie, tj. Bytom, INGBSK, Żywiec oraz dziennych kursów franka szwajcarskiego i dolara amerykańskiego. Dane obejmą okres od 05.01.1995 do 14.10.2011. Obliczenia zostaną przeprowadzone z użyciem programów napisanych przez autorkę w języku programowania Delphi, pakietu Microsoft Excel, Gretl oraz TISEAN.

1. Redukcja poziomu szumu metodą najbliższych sąsiadów

Metodą należącą do grupy procedur redukujących poziom szumu w stanach układu jest metoda najbliższych sąsiadów [3]. Opiera się ona na tzw. wektorach opóźnień, które są wynikiem rekonstrukcji przestrzeni stanów układu dynamicznego (metoda opóźnień [8]). Wektory te zwane również d -historiami przyjmują postać:

$$\hat{x}_t^d = (x_t, x_{t+\tau}, \dots, x_{t+(d-1)\tau}), \quad (2)$$

gdzie d jest wymiarem zanurzenia, τ jest pewną liczbą naturalną nazwaną opóźnieniem czasowym, natomiast zmienna t przyjmuje $n = N - (d-1)\tau$ wartości (N jest długością szeregu czasowego).

Aby wyznaczyć wartość \bar{x}_l , dla ustalonego l , trzeba rozważyć wektor opóźnień, w przypadku gdy czas opóźnień przyjmuje wartość jeden, $\tau = 1$:

$$\hat{x}_t^d = (x_t, x_{t+1}, \dots, x_{t+(d-1)}),$$

którego jedną ze środkowych współrzędnych jest filtrowana obserwacja x_l (np.

wektor $\hat{x}_{l-\frac{d}{2}}^d$ dla parzystej wartości wymiaru zanurzenia lub $\hat{x}_{l-\frac{d+1}{2}}^d$ dla nieparzystej wartości d). Ustalono k najbliższych sąsiadów wektora $\hat{x}_{l-\frac{d}{2}}^d$:

$$\hat{x}_{v_1-\frac{d}{2}}^d, \hat{x}_{v_2-\frac{d}{2}}^d, \dots, \hat{x}_{v_k-\frac{d}{2}}^d. \quad (3)$$

Opierając się na wyznaczonych najbliższych sąsiadach, wartość deterministyczną \bar{x}_l należy wyznaczyć ze wzoru:

$$\bar{x}_l = \frac{1}{k} \sum_{i=1}^k x_{v_i}. \quad (4)$$

1.1. Miara *NRL*

Skuteczność stosowanych metod można badać za pomocą współczynnika poziomu redukcji szumu *NRL* [4]. Bada on zależność pomiędzy siłą szumu dostawanego do układu a strukturą geometryczną jego atraktora [9]. Zależność ta polega na „pogrubianiu” atraktora i oddalaniu się bliskich sobie stanów w stopniu proporcjonalnym do siły szumu. Współczynnik ten wyraża się wzorem:

$$NRL(d) = \frac{1}{n} \sum_{i=1}^n m_i \bigg/ \frac{1}{n} \sum_{i=1}^n M_i, \quad (5)$$

gdzie m_i i M_i oznaczają odległości od i -tego stanu (d -historii) do jego najbliższego i najdalszego sąsiada. Korzystając z powyższego kryterium, należy wybrać spośród otrzymanych szeregów taki, dla którego współczynnik NRL przyjmuje najmniejszą wartość.

2. Prognoza szeregów czasowych – metoda najbliższych sąsiadów*

Od wielu lat jest widoczny ogromny wzrost zainteresowania teorią nieliniowych układów dynamicznych, który zaowocował pojawieniem się nowych metod predykcji wykorzystujących pojęcia i metody związane z tymi układami. Jedną z nich jest metoda najbliższych sąsiadów *MNS*. Jej celem jest wyznaczenie prognozy elementu o numerze $N + 1$ na podstawie szeregu czasowego złożonego z N obserwacji $\{x_1, x_2, \dots, x_N\}$.

Algorytm prognozowania jest następujący:

1. Wyznaczamy liczbę $K = 2(d + 1)$ najbliższych sąsiadów punktu \hat{x}_n^d , $n = N - (d - 1)\tau$, w sensie odległości euklidesowej, w d -wymiarowej zrekonstruowanej przestrzeni stanów. Element x_n d -wymiarowej przestrzeni stanów jest związany z punktem:

$$\hat{x}_n^d = (x_n, x_{n+\tau}, x_{n+2\tau}, \dots, x_{n+(d-1)\tau}). \quad (6)$$

2. Obliczamy sumę:

$$r_{TOT} = \sum_{i=1}^K r_i, \quad (7)$$

gdzie $r_i = r(x_i^d, \hat{x}_n^d)$ oznacza odległość między punktem \hat{x}_n^d i x_i^d , $i = 1, 2, \dots, K$.

3. Wyznaczamy wagę i -tego sąsiada według następującego wzoru:

$$w_i = \frac{1}{K - 1} \left(1 - \frac{r_i}{r_{TOT}} \right). \quad (8)$$

* [2, s. 315-332].

4. Wybieramy pierwsze współrzędne x_i^d K najbliższych sąsiadów punktu \hat{x}_n^d i na ich podstawie określamy pierwsze współrzędne ich następników x_{i+1}^d , $i = 1, 2, \dots, K$.
5. Obliczamy prognozę $N + 1$ elementu jako sumę ważoną następników pierwszych współrzędnych najbliższych sąsiadów:

$$x_{N+1} = \sum_{i=1}^K w_i x_{i+1}. \quad (9)$$

3. Przedmiot i przebieg badania

Interesujące rezultaty otrzymano dla szeregów finansowych pochodzących z rynków papierów wartościowych oraz walutowych. Ich długość pozwala na otrzymanie wiarygodnych rezultatów. Badaniu poddano szeregi finansowe* utworzone z cen zamknięcia trzech spółek notowanych na GPW w Warszawie, tj. Bytom, INGBSK, Żywiec oraz dziennych kursów franka szwajcarskiego i dolara amerykańskiego. Przeanalizowano obserwacje, które dla kursów walut były logarytmicznymi stopami zwrotu, zaś dla spółek – nadwyżkowymi (anormalnymi) stopami zwrotu [7, s. 533-553]:

$$AR_{i,t} = R_{i,t} - (\alpha_i + \beta_i R_{M,t}), \quad (10)$$

gdzie α_i, β_i są parametrami szacowanymi w okresie estymacji za pomocą klasycznej metody najmniejszych kwadratów, $R_{i,t}$ jest logarytmiczną stopą zwrotu i -tej spółki, $R_{M,t}$ – logarytmiczną stopą zwrotu portfela rynkowego (jako przybliżenie zmian wartości portfela rynkowego wykorzystano indeks WIG). W niniejszym artykule okres estymacji bezpośrednio poprzedza okres obserwacji. Parametry modelu rynkowego są szacowane na podstawie 246 sesji, co w przybliżeniu stanowi okres jednego roku kalendarzowego.

Analiza wymienionych wyżej szeregów czasowych będzie przebiegała w pięciu etapach:

1. Rekonstrukcja przestrzeni stanów metodą opóźnień.
2. Redukcja poziomu szumu metodą najbliższych sąsiadów.
3. Obliczenie współczynnika poziomu redukcji szumu NRL .
4. Prognozowanie:
 - metodą najbliższych sąsiadów MNS ,
 - z wykorzystaniem modeli $ARMA$.

* Dane pochodzą z archiwum plików programu Omega dostępnych na stronie internetowej www.bossa.pl.

5. Ocena jakości wyznaczonych prognoz wybranymi miernikami: średnim absolutnym błędem prognozy MAE , błędem średniokwadratowym MSE , pierwiastkiem błędu średniokwadratowego $RMSE$ oraz współczynnikami Theila.

Przeprowadzone badania empiryczne pozwoliły za pomocą metody opóźnień zrekonstruować przestrzeń stanów. Poddając analizie funkcję autokorelacji ACF [6], oszacowano czas opóźnień τ . Następnie za pomocą metody najbliższego pozornego sąsiada $MNPS$ [1, s. 3404-3411] obliczono wymiar zanurzenia d . Wyniki analizy przedstawiono w tabeli 1.

Tabela 1

Wartości parametrów rekonstrukcji przestrzeni stanów dla wybranych szeregów finansowych

Szereg	Parametry
Bytom	$\tau = 4, d = 7$
INGBSK	$\tau = 2, d = 8$
Żywiec	$\tau = 3, d = 10$
CHF	$\tau = 2, d = 10$
USD	$\tau = 2, d = 6$

W kolejnym kroku badań zastosowano redukcję poziomu szumu metodą najbliższych sąsiadów*. Aby dokonać filtracji, ustalono wartość czasu opóźnienia $\tau = 1$ oraz wartości dwóch parametrów:

- wymiar zanurzenia $d = 2, 3, 4, 5, 8, 10, 15, 20$,
- promień otoczenia $\rho = 0,001; 0,01; 0,1$.

W celu oceny redukcji poziomu szumu metodą najbliższych sąsiadów wykorzystano miarę $NRL(i)^{**}$ dla $i = 2, 3, \dots, 10$. Tabela 2 zawiera najmniejszą wartość współczynnika NRL obliczoną dla wybranych szeregów finansowych oraz wartości wymiaru zanurzenia i promienia otoczenia, dla których wartość NRL była najniższa.

Tabela 2

Wartości miary NRL dla przefiltrowanych szeregów finansowych

Nazwa szeregu	Parametry filtracji		Miara NRL
	d	ρ	
Bytom	2	0,1	0,002119
INGBSK	2	0,1	0,001542

* Redukcję szumu przeprowadzono z wykorzystaniem darmowego programu TISEAN autorstwa H. Kantza i T. Schreiberera.

** W celu obliczenia współczynnika NRL posłużono się programem autora napisanym w języku programowania Delphi.

cd. tabeli 2

Żywiec	2	0,1	0,000774
CHF	4	0,1	0,000253
USD	3	0,1	0,000306

Można zauważyć, że prawie we wszystkich przypadkach wartości wymiaru zaburzenia i promienia otoczenia były takie same. Zmiana wystąpiła tylko w przypadku kursów walut.

Następnie w celu wyznaczenia prognozy opartej na metodzie *MNS* wykorzystano $K = 2(d + 1)$ najbliższych sąsiadów punktu \hat{x}_n^d (posłużono się metryką euklidesową) oraz odpowiednie wartości parametrów rekonstrukcji przestrzeni stanów (tabela 1). W celu porównania wyników badanie przeprowadzono dwukrotnie dla szeregów przed i po filtracji (tzn. dla szeregów otrzymanych dla parametrów d i ρ zamieszczonych w tabeli 2).

Do ustalenia postaci modelu ARMA [5] posłużono się kryterium informacyjnym Schwarza, a jako możliwą maksymalną wartość parametrów p i q przyjęto arbitralnie 4. Parametry były ustalane za pomocą metody największej wiarygodności z wykorzystaniem programu ekonometrycznego Gretl. Prognozy dla modeli ARMA były ustalane za pomocą prognozowania dynamicznego. W tabeli 3 przedstawiono zestawienie błędów predykcji.

Tabela 3

Błędy prognoz otrzymanych metodą najbliższych sąsiadów dla szeregów finansowych

Szereg		Błąd						
		<i>ME</i>	<i>MAE</i>	<i>MSE</i>	<i>RMSE</i>	I_1^2	I_2^2	I_3^2
Bytom	Bytom	-0,01522	0,02519	0,00159	0,03989	0,12794	0,34280	0,40792
	Bytom_red	-0,00489	0,00831	0,00030	0,01720	0,07466	0,88877	0,29613
	Bytom_MA(2)	-0,01555	0,02560	0,00137	0,03696	0,13356	0,50567	0,11511
INGBSK	ING	0,00201	0,00802	0,00010	0,00998	0,07601	0,13830	1,66834
	ING_red	0,00003	0,00024	0,00000	0,00030	0,01493	0,56470	0,76632
	ING_MA(34)	0,00149	0,00623	0,00006	0,00763	0,04223	0,62505	0,43425
Żywiec	Żywiec	-0,00021	0,00391	0,00002	0,00478	0,00213	0,28536	0,84706
	Żywiec_red	-0,00002	0,00039	0,00000	0,00151	0,00002	0,71082	0,42996
	Żywiec_MA(1)	-0,00163	0,00326	0,00002	0,00432	0,13181	0,63571	0,16111
CHF	CHF	0,0023	0,0104	0,0002	0,0123	0,0373	0,4552	0,6153
	CHF_red	0,0016	0,0097	0,0001	0,0117	0,0188	0,8701	0,1106
	CHF_AR(3, 1)	-0,00488	0,00805	0,00008	0,00910	0,28297	0,56116	0,14211
USD	USD	0,0005	0,0142	0,0003	0,0159	0,0009	0,3384	0,7388
	USD_red	-0,0022	0,0130	0,0002	0,0153	0,0198	0,6726	0,3372
	USD_AR(1, 1)	-0,00468	0,00975	0,00015	0,01234	0,13732	0,75661	0,06303

Analizując otrzymane rezultaty, można zauważyć, że najniższe błędy *ex post* uzyskano dla prognoz wyznaczonych metodą najbliższych sąsiadów dla szeregów przefiltrowanych, jednakże w tym przypadku wyznaczanie prognoz było najbardziej czasochłonne. Nieco gorsze, choć również dobre rezultaty uzyskano stosując model ARMA. Najgorsze wyniki otrzymano korzystając z metody *MNS* dla szeregów, w których nie przeprowadzono redukcji poziomu szumu.

Podsumowanie

Opierając się na wybranych szeregach finansowych, porównano dwie metody prognozowania: najbliższych sąsiadów oraz modele typu ARMA. Przeprowadzone badania pokazały, że modele ARMA dają dokładniejsze prognozy w przypadku szeregów nieprzefiltrowanych. Natomiast w sytuacji, gdy szeregi te uprzednio poddano redukcji poziomu szumu, lepsze wyniki otrzymano korzystając z metody najbliższych sąsiadów. Zatem zastosowanie metod wywodzących się z teorii nieliniowych układów dynamicznych może stanowić interesującą alternatywę dla modeli ARMA.

Warto zauważyć, że wartości prognoz wyznaczonych metodą najbliższych sąsiadów w dużej mierze zależą od przyjętej metryki, wartości parametrów zrekonstruowanej przestrzeni stanów oraz liczby najbliższych sąsiadów K . Zatem wydaje się, że w celu poprawy jakości tych prognoz można przeprowadzić dodatkowe obliczenia dla różnej liczby najbliższych sąsiadów oraz różnych wartości wymiaru zanurzenia i czasu opóźnienia.

Literatura

1. Abarbanel H.D., Brown R., Kennel M.B., *Determining Embedding Dimension for Phase Space Reconstruction Using a Geometrical Construction*, „Physical Review A” 1992, Vol. 45(6), s. 3404-3411.
2. Diebold F.X., Nason J.A., *Nonparametric Exchange Rate Prediction?* „Journal of International Economics” 1990, Vol. 28, s. 315-332.
3. Kantz H., Schreiber T., *Nonlinear Time Series Analysis*, Cambridge University Press, Cambridge 1997.
4. Orzeszko W., *Identyfikacja i prognozowanie chaosu deterministycznego w ekonomicznych szeregach czasowych*, Polskie Towarzystwo Ekonomiczne, Warszawa 2005.
5. Osińska M., *Ekonometria finansowa*, PWE, Warszawa 2006.

6. Ramsey J.B., Sayers C.L., Rothman P., *The Statistical Properties of Dimension Calculations Using Small Data Sets: Some Economic Applications*, „International Economic Review” 1990, Vol. 31, No. 4.
7. Strong N., *Modelling Abnormal Returns: A Review Article*, „Journal of Business Finance & Accounting” 1992, Vol. 19, Iss. 1, s. 533-553.
8. Takens F., *Detecting Strange Attractors in Turbulence*, w: *Lecture Notes in Mathematics*, red. D.A. Rand, L.S. Young, Springer, Berlin 1981, s. 366-381.
9. Zawadzki H., *Chaotyczne systemy dynamiczne*, Akademia Ekonomiczna, Katowice 1996.

EFFECT OF REDUCTION OF RANDOM NOISE BY METHOD THE NEAREST NEIGHBORS ON THE ACCURACY OF FORECASTS OF THE FINANCIAL TIME SERIES

Summary

The real time series x_t is usually disturbed by random noise. His source may be errors of measurement and errors of rounding made during data analysis. The random noise may also represent an exogenous factors affecting the dynamics of the system or be a consequence of the statistical nature of the phenomena, e.g. which are affected by human decisions.

Since the presence of noise in the data can significantly affect the quality of the forecasts, the aim of the article will be to evaluate the accuracy of predicting the time series filtered using the method of nearest neighbors. The test will be conducted on the basis of the financial time series, which consist of closing prices of companies listed on the Warsaw Stock Exchange and the daily exchange rates.