

EKONOMETRIA

26

Zastosowanie matematyki w ekonomii

Redaktor naukowy Janusz Łyko



**Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2009**

Spis treści

Wstęp	7
Beata Bal-Domańska , Ekonometryczna analiza sigma i beta konwergencji regionów Unii Europejskiej	9
Andrzej Bąk, Aneta Rybicka, Marcin Pelka , Modele efektów głównych i modele z interakcjami w <i>conjoint analysis</i> z zastosowaniem programu R	25
Katarzyna Budny , Kurtoza wektora losowego	44
Wiktor Ejsmont , Optymalna liczebność grupy studentów	55
Kamil Fijorek , Model regresji dla cechy przyjmującej wartości z przedziału $(0,1)$ – ujęcie bayesowskie	66
Paweł Hanczar , Wyznaczanie zapasu bezpieczeństwa w sieci logistycznej ...	77
Roman Huptas , Metody szacowania wewnątrzdziennej sezonowości w analizie danych finansowych pochodzących z pojedynczych transakcji	83
Aleksandra Iwanicka , Wpływ zewnętrznych czynników ryzyka na prawdopodobieństwo ruiny w skończonym horyzoncie czasowym w wieloklasowym modelu ryzyka.....	97
Agnieszka Lipieta , Stany równowagi na rynkach warunkowych	110
Krystyna Melich-Iwanek , Polski rynek pracy w świetle teorii histerezy.....	122
Rafał Piszczek , Zastosowanie modelu logit w modelowaniu upadłości	133
Marcin Salamaga , Próba weryfikacji teorii parytetu siły nabywczej na przykładzie kursów wybranych walut	149
Antoni Smoluk , O zasadzie dualności w programowaniu liniowym	160
Małgorzata Szulc-Janek , Influence of recommendations announcements on stock prices of fuel market	170
Jacek Welc , Regresja liniowa w szacowaniu fundamentalnych współczynników Beta na przykładzie spółek giełdowych z sektorów: budownictwa, informatyki oraz spożywczego	180
Andrzej Wilkowski , O współczynniku korelacji	191
Mirosław Wójciak , Klasyfikacja nowych technologii energetycznych ze względu na determinanty ich rozwoju.....	199
Andrzej Wójcik , Wykorzystanie modeli wektorowo-autoregresyjnych do modelowania gospodarki Polski.....	209
Katarzyna Zeug-Żebro , Rekonstrukcja przestrzeni stanów na podstawie wielowymiarowych szeregów czasowych.....	219

Summaries

Beata Bal-Domańska , Econometric analysis of sigma and beta convergence in the European Union regions	24
Andrzej Bąk, Aneta Rybicka, Marcin Pelka , Main effects models and main and interactions models in <i>conjoint analysis</i> with application of R software.....	43
Katarzyna Budny , Kurtosis of a random vector	53
Wiktor Ejsmont , Optimal class size of students	65
Kamil Fijorek , Regression model for data restricted to the interval (0,1) – Bayesian approach.....	76
Paweł Hanczar , Safety stock level calculation in a supply chain network.....	82
Roman Huptas , Estimation methods of intraday seasonality in transaction financial data analysis	96
Aleksandra Iwanicka , An impact of some outside risk factors on the finite-time ruin probability for a multi-classes risk model.....	109
Agnieszka Lipieta , States of contingent market equilibrium	121
Krystyna Melich-Iwanek , The Polish labour market in light of the hysteresis theory	132
Rafał Piszczek , Logit model applications for bankrupctcy modelling.....	148
Marcin Salamaga , Attempt to verify the purchasing power parity theory in the case of some foreign currencies.....	159
Antoni Smoluk , On dual principle of linear programming	168
Małgorzata Szulc-Janek , Analiza wpływu rekomendacji analityków na ceny akcji branży paliwowej (Analiza wpływu rekomendacji analityków na ceny akcji branży paliwowej).....	178
Jacek Welc , A linear regression in estimating fundamental betas in the case of the stock market companies from construction, it and food industries	190
Andrzej Wilkowski , About the coefficient of correlation	198
Mirosław Wójciak , Classification of new energy related technologies based on the determinants of their development	208
Andrzej Wójcik , Using vector-autoregressive models to modelling economy of Poland.....	218
Katarzyna Zeug-Żebro , State space reconstruction from multivariate time series	227

Katarzyna Zeug-Żebro

Akademia Ekonomiczna w Katowicach

REKONSTRUKCJA PRZESTRZENI STANÓW NA PODSTAWIE WIELOWYMIAROWYCH SZEREGÓW CZASOWYCH

Streszczenie: Jednym z narzędzi matematycznych potrzebnych do opisu szeregów czasowych jest metoda rekonstrukcji. Umożliwia ona rekonstrukcję przestrzeni stanów wielowymiarowych systemów dynamicznych na podstawie jednowymiarowego szeregu obserwacji. Otrzymana w ten sposób przestrzeń będzie w pewnym sensie równoważna z „oryginalną” przestrzenią. W ostatnich latach pojawiło się wiele prac przedstawiających różne metody rekonstrukcji przestrzeni stanów, m.in. metodę opóźnień, która bazuje na twierdzeniu Takensa o zanurzeniu.

Celem pracy jest przedstawienie jednej z metod wyznaczania wymiaru zanurzenia, tj. metody najbliższego fałszywego sąsiada oraz jej zmodyfikowanej postaci w przypadku wielowymiarowych szeregów czasowych. Testowano trzy metody wyznaczania optymalnego wymiaru zanurzenia, korzystając z symulacji otrzymanych za pomocą metody Monte Carlo.

Słowa kluczowe: rekonstrukcja przestrzeni stanów, szeregi czasowe, atraktor, wymiar zanurzenia.

1. Wstęp

Ważną rolę w prognozie nieliniowych systemów dynamicznych odgrywa rekonstrukcja przestrzeni stanów. Pozwala ona na konstrukcję przestrzeni, która będzie w pewnym sensie równoważna z „oryginalną” przestrzenią. Jedną z metod rekonstrukcji, zwana metodą opóźnień, jest oparta na twierdzeniu o zanurzeniu F. Takensa [1981]. Stosując tę metodę, można skonstruować zbiór d zmiennych za pomocą jednowymiarowego szeregu czasowego. Zmienne te otrzymujemy, przesuwając „oryginalny” szereg czasowy o stałe opóźnienie τ . Niestety głównym problemem rekonstrukcji przestrzeni stanów jest właściwy dobór parametrów: d – wymiaru zanurzenia [Abarbanel, Brown, Kennel 1992] i τ – czasu opóźnienia [Kim, Eykholt, Salas 1999].

Celem poniższej pracy jest przedstawienie jednej z metod wyznaczania wymiaru zanurzenia, tj. metody najbliższego fałszywego sąsiada oraz jej zmodyfikowanej postaci w przypadku wielowymiarowych szeregów czasowych. Metody te posłużyły do analizy wybranych nieliniowych systemów dynamicznych: odwzorowania

Henona i odwzorowania Ikedy. Badania przeprowadzono przy użyciu programów napisanych przez autorkę w języku programowania Visual Basic oraz programu Statistica.

2. Rekonstrukcja przestrzeni stanów – metoda opóźnień

Niech stan pewnego systemu dynamicznego w chwili t będzie przedstawiony przez wektor $y_t \in Y \subset R^m$ oraz związek pomiędzy stanami w kolejnych momentach czasu opisuje równanie różnicowe pierwszego rzędu

$$y_{t+1} = f(y_t), \quad t = 0, 1, 2, 3, \dots, \quad (1)$$

gdzie $y_t = (y_{t1}, y_{t2}, \dots, y_{tm})$ oraz $f: Y \rightarrow Y$ jest nieznaną różniczkowalną funkcją nieliniową. Przypuśćmy również, że jedyną informacją, jaką mamy o danym systemie, jest jednowymiarowy ciąg obserwacji x_t , gdzie $x_t = h(y_t)$ i $h: R^m \rightarrow R$ jest nieznanym odwzorowaniem. Mimo że mamy tak niewiele informacji na temat badanego układu dynamicznego, możemy scharakteryzować jego stan w przestrzeni wielowymiarowej, tj. zrekonstruować atraktor badanego systemu, obliczyć jego wymiar, oszacować wykładniki Lapunowa czy też entropie Kołmogorowa.

W 1981 r. F. Takens przedstawił twierdzenie o zanurzaniu, z którego wynika, że atraktor może być zrekonstruowany bez znajomości natury jego zmiennych czy postaci równania różnicowego.

Twierdzenie (o zanurzeniu) [Zawadzki 1996]: Niech M będzie zwartą, a m wymiarową rozmaitością różniczkową. Dla par (f, h) , $f \in \text{Diff}^2(M, M)$, $h \in C^2(M, R)$ jest własnością generyczną taką, że odwzorowanie $\Phi: M \rightarrow R^{2m+1}$ określone wzorem

$$\Phi_{(f,h)} = [h(y), h(f(y)), \dots, h(f^{2m}(y))] \quad (2)$$

jest zanurzeniem, tj. dyfeomorfizmem klasy C^1 odwzorowującym M na $\Phi_{(f,h)}(M)$.

Przestrzeń zawierającą obraz odwzorowania $\Phi_{(f,h)}$ będziemy nazywać przestrzenią zanurzenia R^d (*embedding space*), a jej wymiar d wymiarem zanurzenia.

F. Takens udowodnił, że dla

$$d \geq 2m + 1, \quad (3)$$

gdzie m jest wymiarem atraktora, przestrzeń stanów rozpięta przez zbiór d zmiennych będzie topologicznie równoważna z „oryginalną” przestrzenią.

Stosując metodę opóźnień, można skonstruować zbiór d zmiennych za pomocą jednowymiarowego szeregu czasowego. Zmienne te otrzymuje się, przesuwając

dany szereg czasowy o stałe opóźnienie $\tau, \tau \in N$, w wyniku czego rekonstrukcja przestrzeni stanów wygląda następująco:

$$x_t = (x_{t-(d-1)\tau}, x_{t-(d-2)\tau}, \dots, x_t). \quad (4)$$

3. Metoda najbliższego fałszywego sąsiada (MNFS)

[Abarbanel 1996; Nowiński 2007]

Założmy, że x_t^r będzie r -tym najbliższym sąsiadem punktu $x_t = (x_t, x_{t+\tau}, \dots, x_{t+(d-1)\tau})$ w d -wymiarowej przestrzeni.

1. Obliczmy odległość pomiędzy punktami x_t^r i x_t

$$R_d(t, r) = \sqrt{\sum_{k=0}^{d-1} [x_{t+k\tau} - x_{t+k\tau}^r]^2}, \quad (5)$$

gdzie $R_d(t, r)$ jest odległością euklidesową.

2. W kolejnym kroku obliczmy odległość $R_{d+1}(t, r)$ w $d + 1$ -wymiarowej przestrzeni

$$R_{d+1}^2(t, r) = R_d^2(t, r) + [x_{t+d\tau} - x_{t+d\tau}^r]^2. \quad (6)$$

3. Jeśli $R_{d+1}(t, r)$ znacznie przewyższa $R_d(t, r)$, wtedy punkty x_t^r i x_t nie są najbliższymi sąsiadami i są tzw. fałszywymi sąsiadami.

4. Istnieją dwa kryteria wyznaczania wymiaru zanurzenia. Po pierwsze dobieramy wartość progową R_T (zwykle przyjmuje się wartość $R_T = 15$ [Abarbanel 1996]) i wykrywamy fałszywego sąsiada pewnego punktu przez sprawdzenie warunku

$$\left[\frac{R_{d+1}^2(t, r) - R_d^2(t, r)}{R_d^2(t, r)} \right]^{1/2} = \frac{|x_{t+d\tau} - x_{t+d\tau}^r|}{R_d(t, r)} > R_T. \quad (7)$$

5. Powyższe kryterium nie jest wystarczające, aby jednoznacznie wyznaczyć właściwy wymiar zanurzenia. Dla małych, zaszumionych zbiorów danych najbliżsi sąsiedzi danego punktu mogą wcale nie znajdować się blisko w przestrzeni stanów, dlatego drugim warunkiem, jaki musi być spełniony, jest nierówność

$$\frac{R_{d+1}(t)}{R_A} > A_T, \quad (8)$$

gdzie

$$R_A = \frac{1}{n} \sum_{t=1}^n [x(t) - \bar{x}]^2,$$

$$\bar{x} = \frac{1}{n} \sum_{t=1}^n x(t),$$

A_T – pewne ograniczenie (zwykle przyjmuje się wartość $A_T = 2$ [Abarbanel 1996].

6. Jeśli zachodzą kryteria z punktów 4 i 5, wtedy punkty x_t'' i x_t są fałszywymi sąsiadami.

4. Wielowymiarowa metoda opóźnień oraz metoda najbliższego fałszywego sąsiada w przypadku wielowymiarowym [Vlachos, Kugiumtzis 2008]

Załóżmy, że danych jest n szeregów czasowych x_t^i , gdzie $i = 1, 2, \dots, n$, otrzymanych w wyniku symulacji tego samego systemu dynamicznego $y_t \in Y \subset R^m$, gdzie $x_t^i = h_i(y_t)$ i $h_i : R^m \rightarrow R$ są nieznanymi odwzorowaniami.

Stosując metodę opóźnień, która została opisana powyżej, można skonstruować zbiór D zmiennych:

$$x_t = \left(x_{t-(d_1-1)\tau_1}^1, x_{t-(d_1-2)\tau_1}^1, \dots, x_t^1, x_{t-(d_2-1)\tau_2}^2, x_{t-(d_2-2)\tau_2}^2, \dots, x_t^2, \dots, x_t^n \right), \quad (9)$$

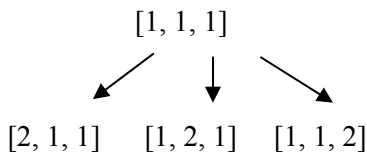
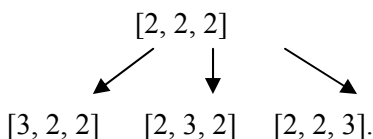
gdzie $D = \sum_{i=1}^n d_i$ jest całkowitym wymiarem zanurzenia, $d = [d_1, d_2, \dots, d_n]$ jest wektorem wymiaru zanurzenia i $\tau = [\tau_1, \tau_2, \dots, \tau_n]$ jest wektorem czasu opóźnienia odpowiadającym każdemu i -temu szeregowi.

Korzystając ze wspomnianego algorytmu wyboru właściwego wymiaru zanurzenia, tj. metody najbliższego fałszywego sąsiada, przedstawimy jej zmodyfikowaną postać w przypadku wielowymiarowym.

- MNFS1 – rozpatrzmy przypadek, w którym dla każdego szeregu x_t^i wymiar zanurzenia przyjmuje taką samą wartość, tj.

$$d = [d, d, \dots, d].$$

W celu sprawdzenia, który wymiar jest właściwy, sprawdzamy kryteria (7) i (8) dla następujących wektorów zanurzenia: $[d+1, d, \dots, d]$, $[d, d+1, \dots, d]$, \dots , $[d, d, \dots, d+1]$, np.

$d = 1$

 $d = 2$


- MNFS2 – rozpatrzmy przypadek, w którym dla każdego szeregu x_t^i wymiar zanurzenia przyjmuje różną wartość, tj.

$$d = [d_1, d_2, \dots, d_n].$$

Postępujemy jak w metodzie MNFS1, tzn. sprawdzamy kryteria wyboru wymiaru zanurzenia dla $[d_1 + 1, d_2, \dots, d_n]$, $[d_1, d_2 + 1, \dots, d_n]$, ..., $[d_1, d_2, \dots, d_n + 1]$,

gdzie $D = \sum_{i=1}^n d_i$, np.

 $D = 1$
 $[1, 0]$ $[0, 1]$
 $D = 2$
 $[2, 0]$ $[1, 1]$ $[0, 2]$
 $D = 3$
 $[3, 0]$ $[2, 1]$ $[1, 2]$ $[0, 3]$.

5. Znormalizowany błąd średniokwadratowy – *NRMSE*

Jednym ze wskaźników, który pozwala wybrać lepszą metodę wyznaczania wymiaru zanurzenia, tzn. wskazuje wymiar zanurzenia, dla którego (na podstawie rekonstrukcji przestrzeni stanów) otrzymamy lepszą prognozę, jest znormalizowany błąd średniokwadratowy:

$$NRMSE = \sqrt{\frac{\sum_t \|x_t^1 - \hat{x}_t^1, \dots, x_t^p - \hat{x}_t^p\|}{\sum_t \|x_t^1 - \bar{x}^1, \dots, x_t^p - \bar{x}^p\|}}, \quad (10)$$

gdzie \hat{x}_t^p jest wartością prognozowaną, a \bar{x}^i jest wartością średnią x_t^i , gdzie $i = 1, 2, \dots, n$. Im mniejsza wartość powyższego wskaźnika, tym wybrana metoda wyboru wymiaru opóźnienia jest lepsza.

6. Badania empiryczne

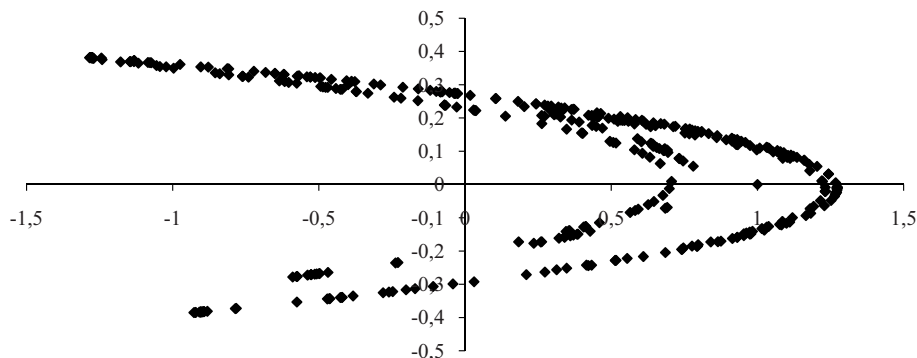
Interesującego i ważnego z praktycznego punktu widzenia materiału badawczego dostarcza nam analiza szeregów czasowych związanych z systemami dynamicznymi z „dziwnymi atraktorami”. W poniższych rozważaniach rozpatrywano dwa takie systemy: atraktor Henona i atraktor Ikedy.

W badaniach testujemy trzy metody wyznaczania optymalnego wymiaru zanurzenia – MNFS, MNFS1 i MNFS2 – korzystając z symulacji otrzymanych za pomocą metody Monte Carlo (Statistica). Spośród 100 różnych realizacji każdego układu dynamicznego – Ikeda, Henon – wybieramy wektory wymiaru zanurzenia i badamy te, które powtarzają się najczęściej. Dodatkowo dla każdej realizacji i wybranych wektorów zanurzenia, dla każdej metody, obliczamy znormalizowany błąd średniokwadratowy *NRMSE*.

W celu wyznaczenia wymiaru zanurzenia z wykorzystaniem MNFS i wektorów zanurzenia (metodami MNFS1 i MNFS2) wybieramy trzy czwarte początkowych elementów szeregu $N_1 = \frac{3}{4}N$, a dla pozostałych $(N - N_1)$ elementów obliczamy wskaźnik *NRMSE*. Tak zastosowany podział szeregu oparto na badaniach I. Vlachos i D. Kugiumtzisa. Wyniki przeprowadzonych analiz zostały przedstawione w tab. 1 i 2.

System dynamiczny opisany za pomocą równania Henona ma następującą postać (rys. 1):

$$\begin{cases} x_{t+1} = 1 - ax_t^2 + y_t \\ y_{t+1} = bx_t \end{cases} \quad (11)$$



Rys. 1. Atraktor Henona

Źródło: opracowanie własne.

Tabela 1. Wartość wskaźnika *NRMSE* – odwzorowanie Henona

<i>N</i>	Wymiary zanurzenia			<i>NRMSE</i>		
	MNFS	MNFS 1	MNFS 2	MNFS	MNFS 1	MNFS 2
500	2	[2,2]	[2,2] [1,2]	0,33	0,19	0,196
2000	2	[2,2]	[2,2]	0,187	0,093	0,093
8000	2	[2,2]	[2,2]	0,11	0,051	0,051

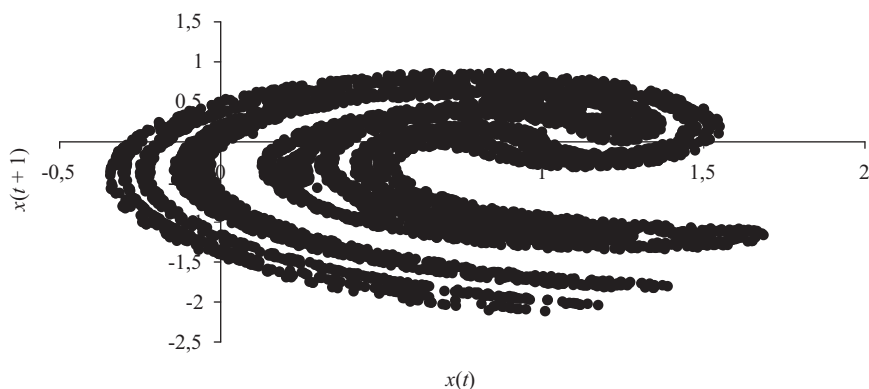
Źródło: opracowanie własne.

System dynamiczny opisany za pomocą równania Ikedy jest postaci (rys. 2):

$$\begin{cases} x_{n+1} = \gamma + \mu(x_n \cos \phi - y_n \sin \phi) \\ y_{n+1} = \mu(x_n \sin \phi - y_n \cos \phi) \end{cases}, \quad (12)$$

$$\phi = \beta - \frac{\alpha}{1 + x_n^2 + y_n^2},$$

gdzie $\alpha = 6$, $\beta = 0,4$, $\gamma = 1$, $\mu = 0,9$.

**Rys. 2.** Atraktor Ikedy

Źródło: opracowanie własne.

Tabela 2. Wartość wskaźnika *NRMSE* – odwzorowanie Ikedy

<i>N</i>	Wymiary zanurzenia			<i>NRMSE</i>		
	MNFS	MNFS 1	MNFS 2	MNFS	MNFS 1	MNFS 2
500	2	[1,1]	[1,1]	0,13	0,051	0,051
2000	2	[1,1]	[1,1]	0,1	0,028	0,028
8000	2	[1,1]	[1,1]	0,07	0,013	0,013

Źródło: opracowanie własne.

Powyższe wyniki wskazują na przydatność wykorzystanych metod (przypadek wielowymiarowy) – wartość wskaźnika *NRMSE* dla metod MNFS1 i MNFS2 jest zdecydowanie niższa niż w przypadku zastosowania tradycyjnej metody najbliższego fałszywego sąsiada. Można również zauważyć, że wartość *NRMSE* maleje w miarę wzrostu liczby elementów szeregu czasowego, a dla metody MNFS1 i MNFS2 przyjmuje te same wartości.

7. Podsumowanie

Wyboru wymiaru zanurzenia d zwykle dokonuje się metodą prób i błędów lub też za pomocą wymiaru korelacyjnego. Wspomniane metody niezbyt dokładnie wyznaczają wymiar d , dlatego w opracowaniu posłużono się metodą fałszywych najbliższych sąsiadów. Metoda ta polega na testowaniu wartości d podczas obserwacji zachowań sąsiednich punktów z przedziału $\langle d, d + 1 \rangle$.

Ze względu na to, że dla sygnałów losowych możemy otrzymać pewne nieprawidłowości przy szacowaniu wymiaru zanurzenia metodą fałszywych najbliższych sąsiadów, podjęto próby udoskonalenia algorytmu tej metody. Jedną z modyfikacji tej procedury polega na szacowaniu przeciętnych odległości między wektorami opóźnionymi czasowo [Cao 1997]. Inną zmianą jest szacowanie wymiaru zanurzenia w przypadku wielowymiarowym (MNFS1 i MNFS2). Wtedy rekonstrukcja przestrzeni stanów na podstawie metody opóźnień w przypadku wielowymiarowym daje lepsze wyniki przy kryterium opartym na mierze dopasowania (*NRMSE*).

Zważywszy na to, że w powyższych rozważaniach rozpatrywane były tylko dwa, dobrze znane z literatury, systemy dynamiczne, w dalszej części prac badawczych autorka przewiduje adaptację metody do zastosowań ekonomicznych (ekonomiczne szeregi czasowe).

Literatura

- Abarbanel H.D., *Analysis of Observed Chaotic Data*, Springer Verlag, Berlin, Heidelberg, New York 1996.
- 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, nr 6.
- Cao L., *Practical Method for Determining Minimum Embedding Dimension of a Scalar Time Series*, „Physica D” 1997 no 110.
- Kim H.S., Eykholt R., Salas J.D., *Nonlinear Dynamics, Delay Time, and Embedding Windows*, „Physica D” 1999 no 127.
- Nowiński M., *Nieliniowa dynamika szeregów czasowych w badaniach ekonomicznych*, AE, Wrocław 2007.
- Takens F., *Detecting Strange Attractors in Turbulence*, [w:] *Lecture Notes in Mathematics*, D.A. Rand and L.S. Young (eds.), Springer, Berlin 1981.
- Vlachos I., Kugiumtzis D., *State Space Reconstruction for Multivariate Time Series Prediction*, „Nonlinear Phenomena in Complex Systems” 2008 vol. 11, nr 2, s. 241-249.
- Zawadzki H., *Chaotyczne systemy dynamiczne*, AE, Katowice 1996.

STATE SPACE RECONSTRUCTION FROM MULTIVARIATE TIME SERIES

Summary: State space reconstruction is one of the tools necessary for the description of time series. This method allows state space reconstruction of multidimensional dynamic systems based on one-dimensional observation series. The space obtained will be, in a way, equivalent to the “original” space. In recent years, many papers have presented different methods of state space reconstruction among other things the delays method which is based on the Takens Embedding Theorem.

In this paper we focus on the state space reconstruction from multivariate time series. For this, we adjust well-known approaches used for univariate time series, i.e. the false nearest neighbour.