

Witold ORZESZKO¹

Wybrane aspekty nieparametrycznego prognozowania nieliniowych szeregów czasowych²

1. WPROWADZENIE

Jednym z podstawowych zadań ekonometrii jest budowa modeli regresji, opisujących mechanizmy kształtujące systemy gospodarcze. Ogólnie modele te mają postać:

$$Y = m(\mathbf{x}) + \varepsilon, \quad (1)$$

gdzie $m(\mathbf{x})$ jest wartością oczekiwaną zmiennej objaśnianej Y pod warunkiem wektora $\mathbf{x} = (x_1, x_2, \dots, x_d)$, będącego realizacją d -wymiarowego wektora zmiennych objaśniających $\mathbf{X} = (X_1, X_2, \dots, X_d)$ (tzn. $m(\mathbf{x}) = E(Y|\mathbf{X} = \mathbf{x})$), natomiast ε jest składnikiem losowym o zerowej warunkowej wartości oczekiwanej (tzn. $E(\varepsilon|\mathbf{X} = \mathbf{x}) = 0$) i wariancji $Var(\varepsilon|\mathbf{X} = \mathbf{x}) = \sigma^2(\mathbf{x})$ (Pagan, Ullah, 1999, s. 79–80).

W szczególnym przypadku zmienna Y może być objaśniana jej opóźnionymi wartościami, co prowadzi do modelu autoregresyjnego, postaci:

$$Y_t = m(y_{t-1}, y_{t-2}, \dots, y_{t-d}) + \varepsilon_t \quad (2)$$

gdzie d jest rzędem autoregresji.

Generalnie rzecz ujmując, celem prognozowania na podstawie modelu regresji jest oszacowanie rozkładu predyktywnej zmiennej Y , określonego funkcją gęstości:

$$f(y|x) = \frac{g(x,y)}{h(x)}, \quad (3)$$

gdzie $g(x,y)$ oraz $h(x)$ są funkcjami gęstości, odpowiednio, wektorów (\mathbf{X}, Y) i \mathbf{X} (zob. np. Hyndman i inni, 1996; Racine, 2008; Kosiorowski, 2015). Zgod-

¹ Uniwersytet Mikołaja Kopernika w Toruniu, Wydział Nauk Ekonomicznych i Zarządzania, Katedra Zastosowań Informatyki i Matematyki w Ekonomii, ul. Gagarina 13a, 87–100 Toruń, Polska, e-mail: witold.orzeszko@umk.pl.

² Projekt został sfinansowany ze środków Narodowego Centrum Nauki przyznanych na podstawie decyzji numer DEC-2013/11/B/HS4/00578.

nie ze wzorem (3) funkcję f szacuje się jako iloraz oszacowanych funkcji g i h . Jednak w zastosowaniach praktycznych przedmiotem zainteresowania nie musi być pełen rozkład predykcyjny, lecz jego wybrane charakterystyki, np. wartość oczekiwana lub wariancja. W takiej sytuacji konstrukcja modelu regresyjnego sprowadza się więc do oszacowania funkcji $m(\mathbf{x})$, a jeśli celem modelowania jest analiza zmienności – również funkcji $\sigma(\mathbf{x})$ (bowiem $Var(Y|\mathbf{X} = \mathbf{x}) = Var(\varepsilon|\mathbf{X} = \mathbf{x})$).

Analizę regresji można uprościć, przyjmując *a priori* określoną postać analityczną (np. liniową) funkcji $m(\mathbf{x})$. Konstrukcja modelu sprowadza się wówczas do oszacowania jego parametrów, co oznacza, że model staje się parametryczny. Formalnie rzecz ujmując, model nazywa się parametrycznym, jeśli zależność między jego zmiennymi wyrażona jest w postaci funkcji o określonej postaci analitycznej, a ponadto przestrzeń parametrów charakteryzujących tę zależność jest skończenie wymiarowa. Oznacza to, że modele nieparametryczne nie mają określonej postaci analitycznej lub ich parametry strukturalne należą do przestrzeni nieskończenie wymiarowej (Fan, Yao, 2005, s. 9).

W regresji nieparametrycznej wyznaczenie funkcji $m(\mathbf{x})$ polega na jej aproksymacji funkcjami określonego typu, na tyle elastycznymi, aby dokładność aproksymacji poprawiała się wraz z liczbą obserwacji w próbie (Härdle i inni, 1997). Wyróżnia się w tym zakresie dwa podejścia: lokalne oraz globalne. Podejście lokalne polega na wyznaczeniu parametrów funkcji aproksymującej oddzielnie dla każdego argumentu \mathbf{x} , przy czym podczas estymacji wykorzystuje się wyłącznie wektory obserwacji z jego określonego sąsiedztwa. Promień analizowanego sąsiedztwa zależy od liczby obserwacji, tzn. im więcej obserwacji, tym mniejszy promień. Z kolei w metodach globalnych wykorzystuje się funkcje określonego typu, w których, w zależności od liczby posiadanych obserwacji, można zmieniać liczbę szacowanych parametrów, a w konsekwencji – dokładność aproksymacji. W metodach globalnych szacuje się więc jedną funkcję – wspólną dla wszystkich argumentów \mathbf{x} , a estymacja odbywa się przy wykorzystaniu wszystkich dostępnych wektorów obserwacji.

Zaletą regresji nieparametrycznej jest brak konieczności przyjmowania hipotetycznej postaci analitycznej modelu, przez co pozwala się danym „mówić same za siebie”, a otrzymane modele dzięki swojej elastyczności mogą dobrze dopasowywać się do danych. Ze względu na bogactwo i zróżnicowanie klasy funkcji nieliniowych własność ta wydaje się bardzo atrakcyjna szczególnie w przypadku modelowania zależności nieliniowych. Warto jednak dodać, że regresja nieparametryczna stanowi alternatywę dla regresji parametrycznej nie tylko w sytuacji modelowania systemów nieliniowych lub braku teorii ekonomicznej, wskazującej

określoną postać modelu parametrycznego, lecz także w przypadku (zob. np. Stelmach, 2014):

- obecności zakłóceń i błędów pomiarowych,
- występowania obserwacji wpływowych i odstających³,
- skorelowania zmiennych objaśniających,
- rozbieżności rozkładów zmiennych od rozkładu normalnego,
- zmiennej objaśnianej mierzonej na słabej skali.

Należy jednak dodać, że z opublikowanych w literaturze przedmiotu badań wynika, że modele nieparametryczne mimo dobrego dopasowania w próbie nie zawsze są dobrymi predyktorami (zob. np. Ramsey, 1996; Granger, Teräsvirta, 1992). Ponadto w porównaniu z modelami parametrycznymi są trudniejsze do interpretacji i formułowania wniosków na temat własności badanych systemów gospodarczych. Wadą regresji nieparametrycznej jest także konieczność posiadania znacznie większej liczby obserwacji. Zebrane dane muszą bowiem posłużyć do wyznaczenia funkcji aproksymacyjnej, a nie tylko do oszacowania jej parametrów. Stone (1982) wykazał, że dla ustalonego stopnia precyzji ε , liczba obserwacji n musi spełniać warunek:

$$n^{-\frac{1}{z+d}} < C\varepsilon, \quad (4)$$

gdzie C jest pewną stałą, zależną od typu rozkładu cechy i rodzaju zastosowanej metody (Stinchcombe, Drukker, 2013). Ze wzoru (4) wynika, że wzrost wymiaru d (tj. liczby modelowanych zmiennych) powoduje wykładniczy wzrost liczby obserwacji, potrzebnej do oszacowania funkcji regresji. Problem ten, noszący nazwę przekleństwa wymiarowości (ang. *curse of dimensionality*), jest tym większy, im ogólniejsza jest forma zależności w modelu. Z tego powodu rozważa się w literaturze przedmiotu modele nieparametryczne o postaci mniej ogólnej niż (1), które wypełniają lukę między modelem (1) a modelami parametrycznymi. Modele te, choć są mniej uniwersalne i elastyczne, mają jednak dużo mniejsze wymagania w zakresie liczby potrzebnych obserwacji. Alternatywnie, proponuje się również modele nazywane semiparametrycznymi, będące pewną kombinacją (np. sumą) ogólnego modelu nieparametrycznego (1) i określonych specyfikacji parametrycznych.

2. JĄDROWE ESTYMATORY FUNKCJI REGRESJI

Często stosowanym narzędziem estymacji modeli regresji nieparametrycznej jest Ważona Metoda Najmniejszych Kwadratów – WMNK, stanowiąca modyfikację Metody Najmniejszych Kwadratów. W WMNK minimalizacji podlega wyrażenie:

³ W dwóch pierwszych wymienionych przypadkach kluczowe znaczenie ma odporność wielu nieparametrycznych estymatorów funkcji regresji. Należy jednak podkreślić, że pojęcia regresji nieparametrycznej i regresji odpornej nie są tożsame. Dziękuję Recenzentowi za zwrócenie uwagi na tę kwestię.

$$\sum_{i=1}^n w_i(\mathbf{x})(y_i - m(\mathbf{x}))^2, \quad (5)$$

gdzie n jest liczbą obserwacji, y_i oraz \mathbf{x}_i (dla $i = 1, 2, \dots, n$) są realizacjami zmiennej Y oraz wektora losowego \mathbf{X} , natomiast $w_i(\mathbf{x}) = w(\mathbf{x}_i, \mathbf{x})$ są pewnymi wagami. Wagi konstruuje się w ten sposób by były nieujemne, sumowały się do jedności i były tym większe, im mniejsza jest odległość wektora obserwacji \mathbf{x}_i od argumentu \mathbf{x} (Pagan, Ullah, 1999, s. 84–85). Rozwiązaniem zadania minimalizacji (5) jest estymator:

$$\hat{m}(\mathbf{x}) = \sum_{i=1}^n w_i(\mathbf{x})y_i, \quad (6)$$

co oznacza, że oszacowana warunkowa wartość oczekiwana zmiennej Y jest średnią ważoną obserwacji $\{y_i\}$ z przyjętymi wagami $w_i(\mathbf{x})$ (por. Fan, Gijbels, 1992). Ponadto warto zauważyć, że estymator określony wzorem (6) jest lokalny, gdyż dla każdej wartości \mathbf{x} generowany jest inny zestaw wag i w konsekwencji otrzymuje się inny wektor oszacowań parametrów funkcji regresji.

Istotną grupą nieparametrycznych estymatorów funkcji regresji są estymatory jądrowe. Najpopularniejszym z nich jest estymator Nadarai-Watsona – ozn. N-W (Nadaraya, 1964; Watson, 1964). W przypadku jednowymiarowym, tzn. dla jednej zmiennej objaśniającej, estymator ten otrzymuje się w wyniku zastosowania Ważonej Metody Najmniejszych Kwadratów (5) dla wag:

$$w_i(x) = \frac{K\left(\frac{x - x_i}{h}\right)}{\sum_{i=1}^n K\left(\frac{x - x_i}{h}\right)} \quad (7)$$

i wyraża się wzorem:

$$\hat{m}(x) = \frac{\sum_{i=1}^n K\left(\frac{x - x_i}{h}\right)y_i}{\sum_{i=1}^n K\left(\frac{x - x_i}{h}\right)}, \quad (8)$$

gdzie K jest funkcją jądrową, natomiast h – jej parametrem wygładzania. Funkcją jądrową jest dowolna nieujemna funkcja rzeczywista, spełniająca warunki (np. Racine, 2008):

$$1) \int_{-\infty}^{+\infty} K(x)dx = 1, \quad (9)$$

$$2) \int_{-\infty}^{+\infty} xK(x)dx = 0, \quad (10)$$

$$3) K(x) = K(-x) \text{ dla każdego } x \in \mathbb{R} \text{ (tzn. jest parzysta)}. \quad (11)$$

W praktyce za jądro przyjmuje się funkcję, która spełnia dodatkowy warunek o posiadaniu w $x_0 = 0$ maksimum globalnego (tzn. $K(0) \geq K(x)$ dla każdego $x \in \mathbb{R}$) (Śliwicki, 2016, s. 35). Dzięki temu największy wpływ na wartość $\hat{m}(x)$ mają te obserwacje y_i , dla których powiązane z nimi x_i są najmniej odległe od argumentu x . W praktyce najczęściej stosowanymi przykładami jąder są: gaussowskie, Epanecznikowa, dwuwagowe, trójkątne i Cauchy'ego (zob. np. Kulczycki, 2005, s. 65–67, 138; Orzeszko, 2016, s. 67–68).

W przypadku wielowymiarowym konstruuje się jądro produktowe κ , postaci (por. Śliwicki, 2016, s. 91):

$$\kappa(x) = \prod_{i=1}^d K(x_i), \quad (12)$$

co prowadzi do estymatora:

$$\hat{m}(\mathbf{x}) = \frac{\sum_{i=1}^n K\left(\frac{x_1 - x_{1i}}{h_1}\right) \cdot \dots \cdot K\left(\frac{x_d - x_{di}}{h_d}\right) y_i}{\sum_{i=1}^n K\left(\frac{x_1 - x_{1i}}{h_1}\right) \cdot \dots \cdot K\left(\frac{x_d - x_{di}}{h_d}\right)}. \quad (13)$$

Można również zastosować jądro radialne (por. Śliwicki, 2016, s. 91–92):

$$\kappa(\mathbf{x}) = c \cdot K\left(\sqrt{\sum_{i=1}^d x_i^2}\right), \quad (14)$$

co w efekcie daje:

$$\hat{m}(\mathbf{x}) = \frac{\sum_{i=1}^n K\left(\sqrt{\left(\frac{x_1 - x_{1i}}{h_1}\right)^2 + \dots + \left(\frac{x_d - x_{di}}{h_d}\right)^2}\right) y_i}{\sum_{i=1}^n K\left(\sqrt{\left(\frac{x_1 - x_{1i}}{h_1}\right)^2 + \dots + \left(\frac{x_d - x_{di}}{h_d}\right)^2}\right)}. \quad (15)$$

Szczególnym przypadkiem regresji jądrowej jest Lokalna Jądrowa Regresja Liniowa (ozn. LJRL). Stanowi ona połączenie lokalnej aproksymacji liniowej i estymacji jądrowej, tzn. funkcję regresji $m(\mathbf{x})$ aproksymuje się lokalnie wielomianami stopnia pierwszego, których parametry szacuje się za pomocą Ważonej Metody Najmniejszych Kwadratów, przy czym wagi wyrażone są funkcją jądrową K . W przypadku jednowymiarowym wektor oszacowań parametrów $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)$ wielomianu aproksymującego znajduje się poprzez minimalizację wyrażenia:

$$\sum_{i=1}^n K\left(\frac{x-x_i}{h}\right) (y_i - \beta_0 - \beta_1(x_i - x))^2. \quad (16)$$

Rozwiązanie powyższego problemu minimalizującego ma postać (zob. Stone, 1977; Fan, Gijbels, 1992; Gajek, Kałuszka, 1996):

$$\hat{m}(x) = \sum_{i=1}^n w_i(x) y_i, \quad (17)$$

dla wag:

$$w_i(x) = \frac{K\left(\frac{x-x_i}{h}\right) (\hat{s}_2(x) - (x-x_i)\hat{s}_1(x))}{\sum_{i=1}^n K\left(\frac{x-x_i}{h}\right) (\hat{s}_2(x) - (x-x_i)\hat{s}_1(x))}, \quad (18)$$

gdzie

$$\hat{s}_r(x) = \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right) (x-x_i)^r \quad \text{dla } r = 1, 2. \quad (19)$$

Estymator $\hat{m}(x)$ wyrażony wzorami (17)–(19) nosi nazwę estymatora Stone'a-Fana (np. Gajek, Kałuszka, 1996). Warto dodać, że estymator Nadarai-Watsona jest szczególnym przypadkiem estymatora Stone'a-Fana, gdyż można go potraktować jako rozwiązanie problemu minimalizacji:

$$\sum_{i=1}^n K\left(\frac{x-x_i}{h}\right) (y_i - \beta_0)^2, \quad (20)$$

czyli problemu mniej ogólnego w stosunku do (16).

W przypadku estymatorów Nadarai-Watsona, Stone'a-Fana i innych lokalnie liniowych funkcji wygładzających (ang. *local linear smoothers*) postaci (6), dla punktów należących do wnętrza nośnika danych (tj. oddalonych od jego brzegu) obciążenie estymatora jest proporcjonalne do h^2 , natomiast wariancja, opisująca tempo zbieżności do teoretycznej funkcji regresji, jest rzędu $(nh^d)^{-1}$ (zob. np. Stinchcombe, Drukker, 2013), tzn.:

$$E(\hat{m}(x)) - m(x) = O(h^d) \quad (21)$$

oraz:

$$\text{Var}(\hat{m}(x)) = O((nh^d)^{-1}). \quad (22)$$

Oznacza to, że przy wyborze parametru h konieczny jest wybór (ang. *trade-off*) między obciążeniem i wariancją.

3. SYMULACJE MONTE CARLO

Celem symulacji Monte Carlo jest ocena efektywności jądrowych estymatorów modeli regresji, tj. estymatora Nadarai-Watsona oraz Lokalnej Jądrowej Regresji Liniowej jako narzędzi prognozowania nieliniowych szeregów czasowych oraz porównanie ich z innymi metodami prognozowania. W badaniu rozważono 20 autoregresyjnych procesów generujących dane (ang. *data generating processes* – DGP)⁴:

$$\text{DGP 1: } Y_t = 4Y_{t-1}(1 - Y_{t-1});$$

$$\text{DGP 2: } Y_t = X_t + \sigma\varepsilon_t; \text{ gdzie } X_t \text{ został wygenerowany z DGP 1, } \sigma = 0,5\sigma_Y;$$

$$\text{DGP 3: } Y_t = X_t + \sigma\varepsilon_t; \text{ gdzie } X_t \text{ został wygenerowany z DGP 1, } \sigma = \sigma_Y;$$

$$\text{DGP 4: } Y_t = 1 + 0,3Y_{t-2} - 1,4Y_{t-1}^2;$$

$$\text{DGP 5: } Y_t = X_t + \sigma\varepsilon_t; \text{ gdzie } X_t \text{ został wygenerowany z DGP 4, } \sigma = 0,5\sigma_Y;$$

$$\text{DGP 6: } Y_t = X_t + \sigma\varepsilon_t; \text{ gdzie } X_t \text{ został wygenerowany z DGP 4, } \sigma = \sigma_Y;$$

$$\text{DGP 7: } Y_t = 0,3Y_{t-1} + \varepsilon_t;$$

$$\text{DGP 8: } Y_t = 0,8Y_{t-1} + 0,15Y_{t-2} + \varepsilon_t + 0,3\varepsilon_{t-1};$$

$$\text{DGP 9: } Y_t = \varepsilon_t + 0,8\varepsilon_{t-1}^2;$$

$$\text{DGP 10: } Y_t = \varepsilon_t + 0,6\varepsilon_{t-1}^2 + 0,6\varepsilon_{t-2}^2;$$

$$\text{DGP 11: } Y_t = \varepsilon_t + 0,8\varepsilon_{t-1}\varepsilon_{t-2};$$

$$\text{DGP 12: } Y_t = 0,8|Y_{t-1}|^{0,5} + \varepsilon_t;$$

$$\text{DGP 13: } Y_t = \text{sign}(Y_{t-1}) + \varepsilon_t, \text{ gdzie } \text{sign} \text{ jest funkcją znaku};$$

$$\text{DGP 14: } Y_t = 0,6\varepsilon_{t-1}Y_{t-2} + \varepsilon_t;$$

$$\text{DGP 15: } Y_t = -0,015 - 0,3Y_{t-1} + 0,01Y_{t-1}(1 + \exp(-3Y_{t-1}^2)) + \varepsilon_t;$$

$$\text{DGP 16: } Y_t = \begin{cases} -0,5Y_{t-1} + \varepsilon_t, & Y_{t-1} < 1, \\ 0,4Y_{t-1} + \varepsilon_t, & Y_{t-1} \geq 1 \end{cases}$$

$$\text{DGP 17: } Y_t = \sqrt{h_t}\varepsilon_t, \quad h_t = 1 + 0,4Y_{t-1}^2;$$

$$\text{DGP 18: } Y_t = \sqrt{h_t}\varepsilon_t, \quad h_t = 0,01 + 0,8Y_{t-1} + 0,15Y_{t-1}^2;$$

$$\text{DGP 19: } Y_t = \sqrt{h_t}\varepsilon_t, \quad \ln h_t = 0,1 + 0,95\ln h_{t-1} + \sqrt{0,009}\eta_t; \quad \eta_t \sim \text{i.i.d. } N(0,1);$$

$$\text{DGP 20: } Y_t = 1 + 10h_t + \varepsilon_t, \quad \varepsilon_t = \sqrt{h_t}Z_t, \quad h_t = 0,001 + 0,85h_{t-1} + 0,1\varepsilon_{t-1}^2.$$

Zastosowane procesy były przedmiotem badań symulacyjnych w pracach: Granger, Lin, 1994; Brock i inni, 1996; Granger i inni, 2004; Hong, White, 2005; Osińska, Górka, 2006; Diks, Panchenko, 2007; Orzeszko, 2016; Orzeszko, 2017. Należą one do klas modeli o dużym znaczeniu w analizie procesów ekonomicznych i finansowych (zob. np. Granger, Teräsvirta, 1993; LeBaron, 1994; Markellos, 2002; Orzeszko, 2005; Bruzda, 2007; Morley, 2009; Orzeszko, 2016 i zawarte w nich odnośniki do literatury).

Pierwszych sześć procesów generujących to odwzorowania chaotyczne: logistyczne i Henona – bez szumu i z dodanym szumem obserwacyjnym. Systemy chaotyczne są przykładem deterministycznych systemów nieliniowych o skompli-

⁴ W każdym z poniższych modeli $\{\varepsilon_t\}$ jest procesem i.i.d. o rozkładzie $N(0,1)$.

kowej, pozornie losowej dynamice (por. np. Orzeszko, 2005). DGP 1 i DGP 4 to, odpowiednio, odwzorowanie logistyczne i Henona. Procesy 2, 3, 5, 6, to przykłady chaosu z szumem, otrzymane w wyniku dodania do szeregów chaotycznych, wygenerowanych z DGP 1 i DGP 4 obserwacyjnego szumu losowego o dwóch poziomach.⁵ Procesy DGP 7 i DGP 8 są liniowe – DGP 7 to model autoregresyjny (AR), a DGP 8 to autoregresyjny model średniej ruchomej (ARMA). DGP 9, 10, 11 są przykładami nieliniowych modeli średniej ruchomej (NMA), natomiast DGP 12 i 13 to nieliniowe modele autoregresyjne (NAR). DGP 14 jest modelem dwuliniowym (BL), DGP 15 – wykładniczym modelem autoregresyjnym (EXPAR), a DGP 16 – progowym modelem autoregresyjnym (SETAR). Kolejne trzy procesy są nieliniowe w wariancji: DGP 17 to model autoregresyjny z warunkową heteroskedastycznością (ARCH), DGP 18 to uogólniony model ARCH (GARCH), a DGP 19 to model zmienności stochastycznej (SV). Z kolei DGP 20 to model GARCH-M, będący przykładem procesu nieliniowego zarówno w zakresie średniej, jak i wariancji.

Z każdego rozważonego procesu wygenerowano po 1000 replikacji złożonych z $n = 50$ obserwacji („szeregi krótkie”), $n = 300$ obserwacji („szeregi średnie”) oraz $n = 1000$ obserwacji („szeregi długie”).⁶ Dla każdej replikacji prognozowaniu poddano ostatnią obserwację y_n . Do oceny trafności prognoz zastosowano średni bezwzględny błąd prognozy (MAE), obliczony na podstawie wszystkich 1000 replikacji.

Do wyznaczenia prognoz zastosowano siedem metod prognozowania. Trzy z nich to modele regresji: estymator N-W, LJRL oraz model regresji liniowej. W modelach tych prognoza y_n^* wyznaczana była na podstawie poprzedzającej obserwacji, zgodnie z modelem (2), gdzie $d = 1$, tj. $y_n^* = \hat{m}(y_{n-1})$, przy czym do estymacji funkcji m wykorzystano obserwacje ze wszystkich okresów poprzedzających, tj. y_1, y_2, \dots, y_{n-1} . W obu metodach jądrowych zastosowano jądro gaussowskie, postaci:

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad (23)$$

a parametr wygładzania wyznaczono z formuły:

$$h = \left(\frac{4}{3(n-1)} \right)^{0,2} \frac{Me(|y_i - Me(y)|)}{0,6745}, \quad (24)$$

gdzie Me oznacza medianę z próby (Bowman, Azzalini, 1997, s. 31).

⁵ W przypadku DGP 2 i DGP 5 wartość wskaźnika SNR (ang. *signal-to-noise ratio*) wynosi 2, natomiast w przypadku DGP 3 i DGP 6 wartość ta wynosi 1. W obu sytuacjach dodany szum losowy należy uznać za silny (por. Rosenstein i inni, 1993).

⁶ W celu wyeliminowania wpływu wartości początkowych wygenerowano szeregi złożone z, odpowiednio, 750, 1000 i 1700 obserwacji, a następnie odrzucono pierwszych 700 obserwacji. Szeregi krótkie stanowią początek szeregów średnich, a te z kolei – początek szeregów długich.

Wśród pozostałych czterech metod prognozowania znalazły się dwie metody naiwne. W pierwszej z nich (ozn. naiwna1) prognozę wyznaczono ze wzoru:

$$y_n^* = \frac{\sum_{t=1}^{n-1} y_t}{n-1}, \quad (25)$$

natomiast w drugiej (ozn. naiwna2) – zgodnie z formułą:

$$y_n^* = y_{n-1}. \quad (26)$$

Pierwsza z zastosowanych metod naiwnych jest zgodna z modelem białego szumu, natomiast druga – z modelem błędzenia przypadkowego.

Ponadto zastosowano dwie metody cechujące się wysoką efektywnością prognozowania chaotycznych szeregów czasowych (zob. Orzeszko, 2004a) – metodę Najbliższych Sąsiadów oraz Lokalną Aproksymację Liniową. Punktem wyjścia obu metod jest rekonstrukcja przestrzeni stanów, która, zgodnie z tw. Takensa o zanurzaniu (Takens, 1981), polega na konstrukcji p -wymiarowych wektorów opóźnień postaci:

$$\hat{y}_t^p = (y_t, y_{t-1}, \dots, y_{t-(p-1)}), \quad (27)$$

gdzie parametr p nazywany jest wymiarem zanurzenia. W metodzie Najbliższych Sąsiadów (ozn. NS) wyznacza się spośród wszystkich wektorów opóźnień k najbliższych (w sensie zadanej p -wymiarowej metryki) sąsiadów wektora \hat{y}_{n-1}^p . Niech $\hat{y}_{t_1}^p, \hat{y}_{t_2}^p, \dots, \hat{y}_{t_k}^p$ będą wyznaczonymi najbliższymi sąsiadami. Według metody, prognozą y_n^* jest średnia ważona obserwacji $\{y_{t_1+1}, y_{t_2+1}, \dots, y_{t_k+1}\}$, z wagami dobranymi tak, by bliżsi sąsiedzi mieli na nią większy wpływ:

$$y_n^* = \sum_{i=1}^k w(\|\hat{y}_{n-1}^p - \hat{y}_{t_i}^p\|) \cdot y_{t_i+1}, \quad (28)$$

gdzie $\|\cdot\|$ oznacza dowolną, przyjętą normę w \mathbb{R}^p , natomiast $w: \mathbb{R} \rightarrow \mathbb{R}$ jest dowolną funkcją malejącą spełniającą warunki:

$$w(\|\hat{y}_{n-1}^p - \hat{y}_{t_i}^p\|) > 0, \text{ dla każdego } i = 1, 2, \dots, k, \quad (29)$$

$$\sum_{i=1}^k w(\|\hat{y}_{n-1}^p - \hat{y}_{t_i}^p\|) = 1. \quad (30)$$

W przeprowadzonym badaniu przyjęto wagi wykładnicze (Finkenstädt, Kuhbier, 1995) postaci:

$$w(\|\hat{y}_{n-1}^p - \hat{y}_{t_i}^p\|) = \frac{e^{-\|\hat{y}_{n-1}^p - \hat{y}_{t_i}^p\|}}{\sum_{i=1}^k e^{-\|\hat{y}_{n-1}^p - \hat{y}_{t_i}^p\|}} \quad (31)$$

W metodzie Lokalnej Aproksymacji Liniowej (ozn. LA) prognozę wyznacza się ze wzoru:

$$y_n^* = g(\hat{y}_{n-1}^p), \quad (32)$$

przy czym funkcję g aproksymuje się funkcją liniową, tj.:

$$g(\hat{y}_{n-1}^p) \approx \alpha_0 + \alpha_1 y_{n-1} + \alpha_2 y_{n-2} + \dots + \alpha_p y_{n-p}. \quad (33)$$

„Lokalność” metody polega na tym, że estymacja parametrów strukturalnych α_i dokonywana jest Metodą Najmniejszych Kwadratów na podstawie jedynie najbliższych sąsiadów \hat{y}_{n-1}^p , czyli wektorów $\hat{y}_{t_1}^p, \hat{y}_{t_2}^p, \dots, \hat{y}_{t_k}^p$.

Jak widać, zarówno metoda Najbliższych Sąsiadów, jak i Lokalna Aproksymacja Liniowa wymagają przyjęcia *a priori* wartości dwóch parametrów – wymiaru zanurzenia p i liczby najbliższych sąsiadów k . Na podstawie wyników zaprezentowanych przez Orzeszko (2004b) do wyboru odpowiednich wartości tych parametrów zastosowano w badaniu procedurę, polegającą na analizie dokładności prognoz wygasłych. W tym celu dla każdego procesu generującego wybrano po jednym z szeregów krótkich, średnich oraz długich. Następnie dla każdego szeregu obliczono średni bezwzględny błąd prognozy (MAE) dla pięciu obserwacji poprzedzających okres prognozowany (tj. dla y_{n-1}, \dots, y_{n-5}), przyjmując kolejno wartości $p = 1, 2, \dots, 15$ oraz $k = 1, 2, \dots, n - p + 1$. Ostatecznie dla każdego szeregu czasowego wybrano wartości parametrów minimalizujące błąd prognozy. Wybrane wartości zestawiono w tabeli 1.

Tabela 1. WARTOŚCI PARAMETRÓW MINIMALIZUJĄCE BŁĄD PROGNOZY W METODACH NS i LA

DGP	Długość szeregu	NS		LA	
		p	k	p	k
DGP 1 logist	krótkie	1	2	1	3
	średnie	1	7	2	4
	długie	1	3	2	4
DGP 2 logist+0,5	krótkie	14	1	2	6
	średnie	3	1	9	14
	długie	3	2	14	21
DGP 3 logist+1	krótkie	9	3	5	15
	średnie	1	2	7	10
	długie	5	1	14	52
DGP 4 Henon	krótkie	2	2	3	5
	średnie	2	3	3	5
	długie	4	2	4	6

Tabela 1. WARTOŚCI PARAMETRÓW MINIMALIZUJĄCE BŁĄD PROGNOZY
W METODACH NS I LA (dok.)

DGP	Długość szeregu	NS		LA	
		p	k	p	k
DGP 5 Henon+0,5	krótkie	4	39	6	12
	średnie	3	5	6	17
	długie	6	5	14	25
DGP 6 Henon+1	krótkie	3	1	5	15
	średnie	2	11	2	5
	długie	10	3	8	19
DGP 7 AR	krótkie	1	2	1	3
	średnie	2	3	2	21
	długie	2	3	7	17
DGP 8 ARMA	krótkie	1	2	4	8
	średnie	14	3	12	35
	długie	1	5	8	22
DGP 9 NMA	krótkie	15	1	14	24
	średnie	4	23	2	23
	długie	8	1	15	19
DGP 10 NMA	krótkie	5	7	5	7
	średnie	4	1	2	22
	długie	2	1	5	43
DGP 11 NMA	krótkie	5	1	7	10
	średnie	1	43	1	42
	długie	4	1	5	11
DGP 12 NAR	krótkie	6	2	2	4
	średnie	3	10	9	12
	długie	1	2	7	11
DGP 13 NAR	krótkie	11	5	7	24
	średnie	2	9	1	62
	długie	2	3	2	10
DGP 14 BL	krótkie	5	1	7	11
	średnie	2	15	3	33
	długie	6	3	5	13
DGP 15 EXPAR	krótkie	1	3	1	3
	średnie	7	1	2	63
	długie	8	2	12	16
DGP 16 SETAR	krótkie	14	3	11	14
	średnie	2	56	3	15
	długie	5	4	7	11
DGP 17 ARCH	krótkie	13	2	6	9
	średnie	6	2	2	59
	długie	5	7	2	7
DGP 18 GARCH	krótkie	4	3	7	10
	średnie	6	2	1	106
	długie	3	2	2	6
DGP 19 SV	krótkie	1	3	15	20
	średnie	2	1	11	78
	długie	9	5	1	13
DGP 20 GARCH-M	krótkie	11	1	7	33
	średnie	6	12	6	12
	długie	11	4	10	12

Źródło: obliczenia własne.

W tabeli 2 zaprezentowano wyniki prognozowania rozważonych procesów generujących z podziałem na szeregi o różnych długościach.⁷ W każdej komórce podano błąd prognozy, wyróżniając pogrubioną czcionką wartość najmniejszą.

Tabela 2. BEZWZGLĘDNE BŁĘDY PROGNOZ DLA BADANYCH DGP

DGP	Długość szeregu	Naiwna1	Naiwna2	Regresja liniowa	N-W	LJRL	NS	LA
DGP 1 logist	krótkie	0,3165	0,4013	0,3192	0,2079	0,1230	0,0212	0,0021
	średnie	0,3109	0,4057	0,3116	0,1523	0,0750	0,0050	7E-06
	dłgie	0,3171	0,4192	0,3173	0,1170	0,0520	0,0012	2E-07
DGP 2 logist+0,5	krótkie	0,9530	1,3044	0,9611	0,8126	0,7955	1,2858	0,9504
	średnie	0,9608	1,3132	0,9619	0,7973	0,7840	1,0020	1,4343
	dłgie	0,9490	1,2960	0,9499	0,7656	0,7581	0,8690	1,4708
DGP 3 logist+1	krótkie	1,1550	1,6270	1,1684	1,1460	1,1851	1,3342	1,5208
	średnie	1,1858	1,6573	1,1883	1,1630	1,1645	1,3723	2,0052
	dłgie	1,1343	1,6140	1,1351	1,1081	1,1106	1,5383	1,3493
DGP 4 Henon	krótkie	0,6235	1,0252	0,5920	0,3272	0,2144	0,1030	0,0743
	średnie	0,6129	0,9730	0,5883	0,2512	0,1918	0,0266	0,0017
	dłgie	0,6069	0,9575	0,5839	0,2214	0,1806	0,0171	0,0002
DGP 5 Henon+0,5	krótkie	0,9363	1,5231	0,9058	0,7785	0,7638	0,8412	1,1190
	średnie	0,9116	1,4210	0,8933	0,7249	0,7178	0,6868	0,7920
	dłgie	0,8837	1,4225	0,8628	0,7232	0,7246	0,6528	1,0463
DGP 6 Henon+1	krótkie	1,1548	1,7930	1,1563	1,1453	1,1738	1,4864	1,4531
	średnie	1,1145	1,6948	1,1168	1,0844	1,0835	1,1090	1,3623
	dłgie	1,0906	1,7079	1,0804	1,0632	1,0667	1,2148	1,3111
DGP 7 AR	krótkie	0,8615	0,9910	0,8266	0,8340	0,8686	0,9515	1,3065
	średnie	0,8534	1,0097	0,8145	0,8200	0,8299	0,9547	0,8445
	dłgie	0,8468	0,9996	0,7967	0,8006	0,8012	0,9401	0,9934
DGP 8 ARMA	krótkie	2,6490	0,8248	0,8544	1,0208	0,9471	1,0341	1,4033
	średnie	3,0816	0,8295	0,8253	0,8999	0,8447	1,0788	0,9994
	dłgie	3,1700	0,8076	0,8061	0,8640	0,8102	0,8973	0,9796
DGP 9 NMA	krótkie	1,1965	1,5839	1,2052	1,1391	1,2494	1,7248	2,1152
	średnie	1,1613	1,5154	1,1637	1,0479	1,0574	1,0923	1,0623
	dłgie	1,1336	1,5236	1,1338	1,0350	1,0538	1,2230	2,9258
DGP 10 NMA	krótkie	1,2731	1,4127	1,2310	1,1736	1,2455	1,2666	2,9759
	średnie	1,2128	1,3942	1,1904	1,1011	1,0625	1,4653	1,0512
	dłgie	1,1426	1,3589	1,1002	1,0190	1,0246	1,4097	1,0255
DGP 11 NMA	krótkie	0,9978	1,3640	1,0177	1,0295	1,1040	1,2971	2,5605
	średnie	1,0090	1,3620	1,0090	1,0225	1,1089	1,0202	1,0320
	dłgie	1,0276	1,4263	1,0255	1,0398	1,0514	1,2864	1,1522
DGP 12 NAR	krótkie	0,8596	1,0379	0,8458	0,8338	0,8522	1,0102	1,5111
	średnie	0,8571	1,0506	0,8360	0,8222	0,8280	0,8568	2,1773
	dłgie	0,8321	1,0538	0,8151	0,8032	0,8040	0,9716	1,3517
DGP 13 NAR	krótkie	1,1840	1,0300	0,9484	0,8980	0,9325	1,0576	1,1467
	średnie	1,1685	1,0223	0,9125	0,8525	0,8525	0,8814	0,8390
	dłgie	1,1873	1,0389	0,9154	0,8178	0,8202	0,9324	0,9072

⁷ Obliczenia wykonano przy zastosowaniu własnych kodów komputerowych, napisanych w środowisku Matlab oraz w Visual Basic for Applications. W przypadku metod jądrowych wykorzystano funkcje ksr i ksrLin, stworzone przez Yi Cao (Cranfield University).

Tabela 2. BEZWZGLĘDNE BŁĘDY PROGNOZ DLA BADANYCH DGP (dok.)

DGP	Długość szeregu	Naiwna1	Naiwna2	Regresja liniowa	N-W	LJRL	NS	LA
DGP 14 BL	krótkie	0,9764	1,3270	0,9929	1,0068	1,0893	1,2810	2,0827
	średnie	0,9933	1,3512	0,9932	0,9969	1,0278	0,9479	0,9189
	długie	0,9910	1,4054	0,9916	0,9989	0,9970	1,0284	1,0051
DGP 15 EXPAR	krótkie	0,8516	1,3691	0,8284	0,8236	0,8489	0,9184	1,2809
	średnie	0,8477	1,3190	0,8123	0,8184	0,8261	1,1054	0,8221
	długie	0,8358	1,3686	0,7968	0,8014	0,8065	0,9571	1,9485
DGP 16 SETAR	krótkie	0,9089	1,2386	0,9196	0,8667	0,8800	1,0642	2,9509
	średnie	0,8887	1,2195	0,8920	0,8234	0,8307	0,8460	0,9042
	długie	0,8704	1,2434	0,8705	0,7983	0,8017	0,8900	1,4156
DGP 17 ARCH	krótkie	1,0289	1,3712	1,0466	1,0571	1,1825	1,2596	2,5127
	średnie	1,0102	1,3605	1,0144	1,0270	1,0724	1,1731	1,0333
	długie	0,9922	1,3736	0,9906	0,9997	1,1114	1,0650	1,0843
DGP 18 GARCH	krótkie	0,3412	0,4720	0,3455	0,3476	0,3680	0,3808	0,9093
	średnie	0,3389	0,4631	0,3399	0,3458	0,3488	0,3956	0,3414
	długie	0,3483	0,4897	0,3487	0,3552	0,3610	0,4428	0,4211
DGP 19 SV	krótkie	2,0250	2,8896	2,0661	2,0780	2,3114	2,1974	6,0470
	średnie	2,3361	3,2802	2,3437	2,3626	2,3786	3,1478	2,5761
	długie	2,2225	3,1276	2,2277	2,2420	2,2566	2,4853	2,3225
DGP 20 GARCH-M	krótkie	0,1302	0,1564	0,1261	0,1236	0,1277	0,1637	0,1362
	średnie	0,1302	0,1531	0,1262	0,1231	0,1241	0,1205	0,1647
	długie	0,1320	0,1576	0,1248	0,1217	0,1244	0,1302	0,4135

Źródło: obliczenia własne.

W przypadku procesów chaotycznych (DGP 1 i DGP 4) najlepsze rezultaty otrzymano za pomocą metod LA oraz NS, przy czym metoda LA doprowadziła do prognoz parę rzędów dokładniejszych niż NS. Z kolei metoda NS dała dużo dokładniejsze prognozy niż metody jądrowe, spośród których lepsza okazała się LJRL. Model regresji liniowej okazał się słabym narzędziem prognozowania szeregów chaotycznych zarówno w przypadku systemów bez szumu, jak i z szumem. Dodanie do procesu generującego szumu (DGP 2, DGP 3, DGP 5, DGP 6) znacząco pogorszyło efektywność metody LA oraz NS. Zarówno w przypadku odwzorowania logistycznego, jak i systemu Henona, metoda LA stała się gorsza od naiwnej już w przypadku szumu na poziomie $0,5\sigma_Y$. Z kolei metoda NS dała dokładne prognozy jedynie dla systemu Henona ze słabszym szumem (DGP 5), ale zawiodła dla systemu z szumem silniejszym (DGP 6). Dodanie szumu do szeregów chaotycznych w relatywnie dużo mniejszym stopniu pogorszyło dokładność estymatorów jądrowych. W efekcie, za wyjątkiem szeregów średnich i długich dla procesu DGP 5, metody te okazały się najskuteczniejszymi narzędziami prognozowania chaosu z szumem, przy czym w sytuacji słabszego szumu LJRL miała przewagę nad estymatorem N-W, natomiast dla silnego szumu można zaobserwować sytuację odwrotną.

Zgodnie z oczekiwaniami najlepszym narzędziem prognozowania procesów liniowych (tj. DGP 7 i DGP 8) okazał się model regresji liniowej, choć tylko nieznacznie przewyższył on metodę naiwną². Wszystkie pozostałe metody dały prognozy mniej dokładne niż metoda naiwna. Dla procesu AR estymator N-W doprowadził do lepszych prognoz niż LJRL, natomiast dla ARMA dokładniejsza okazała się metoda LJRL. Najgorsze prognozy otrzymano przy zastosowaniu metod NS i LA.

Spośród pozostałych procesów z nieliniowością w średniej w pierwszej kolejności zbadano szczególne procesy klasy NARMA: NAR i NMA (DGP 9-DGP 13). Zastosowane narzędzia zawiodły w przypadku procesu DGP 11, dla którego jedynie model liniowy dał prognozy nieco dokładniejsze niż metoda naiwna¹. Dla pozostałych procesów można uznać, że najskuteczniejszymi narzędziami prognozowania okazały się metody jądrowe, przy czym estymator N-W zwykle prowadził do prognoz nieco dokładniejszych niż LJRL. Ponadto przeprowadzone badanie wykazało, że model liniowy jest mało skutecznym narzędziem prognozowania tego typu procesów. Jego wyższość nad metodą naiwną zaobserwowano jedynie w przypadku procesów DGP 13 i – w niewielkim stopniu – DGP 11. Relatywnie słabo wypadły również metody NS i LA, choć w przypadku tej drugiej można zauważyć, że dawała ona bardzo skrajne rezultaty, będąc w niektórych sytuacjach metodą najlepszą, a w niektórych – najgorszą.

W przypadku pozostałych procesów z nieliniowością w średniej, tj. BL, EXPAR i SETAR zaobserwowano, że model liniowy oraz modele regresji jądrowej dały prognozy o zbliżonej dokładności, przy czym dla procesów BL i EXPAR model liniowy był nieco lepszy niż modele jądrowe, natomiast dla SETAR – nieco gorszy. Porównując obie metody jądrowe, ponownie można zauważyć niewielką wyższość estymatora N-W nad LJRL. Z otrzymanych rezultatów wynika również, że wszystkie zastosowane metody zawiodły w przypadku procesu BL. Jedynie dla szeregów średnich tego procesu znalazły się metody – LA i NS – lepsze niż metoda naiwna. Jednak dla pozostałych długości szeregów tego procesu oraz dla procesów EXPAR i SETAR metody LA i NS okazały się mało skutecznymi narzędziami prognozowania.

Analizując wyniki dla procesów z nieliniowością w wariancji (DGP 17-DGP 19), rzuca się w oczy niska efektywność zastosowanych narzędzi w porównaniu z metodą naiwną¹. Rezultat ten należy uznać za spodziewany, gdyż zastosowane metody zostały skonstruowane do prognozowania warunkowej wartości oczekiwanej. Dodatkowo, zwraca jednak uwagę wyższość modelu liniowego nad metodami jądrowymi, spośród których estymator N-W okazał się lepszy niż LJRL (co szczególnie wyraźne jest przypadku procesów ARCH i SV). Z kolei metody jądrowe okazały się mieć przewagę nad metodami NS i LA.

Inne rezultaty otrzymano dla procesu hybrydowego GARCH-M. Pojawienie się w procesie generującym nieliniowości w średniej spowodowało, że zastoso-

wane modele regresji okazały się lepsze niż metoda naiwna. Metody jądrowe doprowadziły do prognoz nieznacznie lepszych niż model liniowy, przy czym estymator N-W okazał się lepszy niż LJRL. Metoda LA ponownie zawiodła, natomiast NS dał zróżnicowane efekty, prowadząc dla szeregów o średniej długości do prognoz najlepszych, a dla szeregów długich – do prognoz gorszych niż metoda naiwna.

4. PODSUMOWANIE WYNIKÓW BADAŃ

Z przeprowadzonych symulacji Monte Carlo wynika, że w przypadku większości procesów generujących dokładność prognoz zastosowanych modeli regresji (tj. modelu liniowego, estymatora Nadarai-Watsona oraz Lokalnej Jądrowej Regresji Liniowej) rośnie wraz ze zwiększaniem się długości szeregów czasowych. Wyjątkiem są nieliczne sytuacje, gdy metody te w ogóle zawodzą, dając prognozy mniej dokładne niż metoda naiwna. W efekcie ranking zastosowanych modeli regresji jest podobny dla wszystkich rozważonych długości szeregów czasowych. Porównując obie metody jądrowe, można zauważyć, że LJRL jest bardziej wrażliwa na liczbę obserwacji, co powoduje, że w niektórych przypadkach jej przewaga np. nad modelem liniowym jest zauważalna dopiero dla szeregów średnich i długich.

Dla metody Najbliższych Sąsiadów, a zwłaszcza Lokalnej Aproksymacji Liniowej, nie da się sformułować prostej zależności między długością szeregu czasowego a dokładnością prognoz. Jedynie w przypadku procesów deterministycznych (tj. DGP 1 i DGP 4) dokładność prognoz rosła wraz ze wzrostem liczby obserwacji. W przypadku pozostałych procesów zależność ta często nie miała miejsca, co najprawdopodobniej jest konsekwencją wrażliwości obu metod na wartości parametrów (por. Orzeszko, 2004b).

Z otrzymanych rezultatów wynika, że nie ma jednej metody, która miałaby przewagę nad pozostałymi bez względu na proces generujący. Można jednak stwierdzić, że najgorszą z nich okazała się metoda naiwna², która jedynie w przypadku procesu DGP 8 doprowadziła do prognoz o relatywnie dużej dokładności. Dla procesów chaotycznych (bez szumu) najlepsze rezultaty otrzymano przy zastosowaniu metod LA oraz NS. W przypadku chaosu z szumem i większości innych stochastycznych procesów z nieliniowością w średniej najskuteczniejszymi narzędziami prognozowania okazały się metody jądrowe, z których estymator Nadarai-Watsona miał zwykle przewagę nad Lokalną Jądrową Regresją Liniową. Zgodnie z oczekiwaniami, w przypadku procesów liniowych model regresji liniowej wykazał swoją wyższość nad pozostałymi metodami. Przeprowadzone badanie potwierdziło również, że zastosowane metody nie są dobrym narzędziem prognozowania procesów z nieliniowością w wariancji.

LITERATURA

- Bowman A. W., Azzalini A., (1997), *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations*, Clarendon Press, Oxford.
- Brock W. A., Dechert W. D., Scheinkman J. A., LeBaron B., (1996), A Test for Independence Based on the Correlation Dimension, *Econometric Reviews*, 15 (3), 197–235.
- Bruzda J., (2007), *Procesy nieliniowe i zależności długookresowe w ekonomii. Analiza kointegracji nieliniowej*, Wydawnictwo Naukowe UMK, Toruń.
- Diks C., Panchenko V., (2007), Nonparametric Tests for Serial Independence Based on Quadratic Forms, *Statistica Sinica*, 17, 81–98.
- Fan J., Gijbels I., (1992), Variable Bandwidth and Local Linear Regression Smoothers, *Annals of Statistics*, 20 (4), 2008–2036.
- Fan J., Yao Q., (2005), *Nonlinear Time Series. Nonparametric and Parametric Methods*, Springer, New York.
- Finkenstädt B., Kuhbier P., (1995), Forecasting Nonlinear Economic Time Series: A Simple Test to Accompany the Nearest Neighbor Approach, *Empirical Economics*, 20, 243–263.
- Gajek L., Kaluszka M., (1996), *Wnioskowanie statystyczne. Modele i metody*, Wydawnictwa Naukowo-Techniczne, Warszawa.
- 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.
- Granger C. W. J., Maasoumi E., Racine J., (2004), A Dependence Metric for Possibly Nonlinear Processes, *Journal of Time Series Analysis*, 25 (5), 649–669.
- Granger C. W. J., Teräsvirta T., (1992), Experiments in Modeling Nonlinear Relationships Between Time Series, w: Castagli M., Eubank S., (red.), *Nonlinear Modeling and Forecasting*, Redwood City, Addison-Wesley, 189–197.
- Granger C. W. J., Teräsvirta T., (1993), *Modelling Nonlinear Economic Relationships*, Oxford University Press.
- Härdle W., Lütkepohl H., Chen R., (1997), A Review of Nonparametric Time Series Analysis, *International Statistical Review*, 65 (1), 49–72.
- Hong Y., White H., (2005), Asymptotic Distribution Theory for an Entropy-Based Measure of Serial Dependence, *Econometrica*, 73, 837–901.
- Hyndman R. J., Bashtannyk D. M., Grunwald G. K., (1996), Estimating and Visualizing Conditional Densities, *Journal of Computational and Graphical Statistics*, 5, 315–336.
- Jerzman B., Kicinski W., (2009), Kernel Estimation of Probability Density Function: Properties and Parameters Optimization, *Metrology and Measurement Systems*, 16 (1), 85–105.
- Kosiorowski D., (2015), Two Procedures for Robust Monitoring of Probability Distributions of Economic Data Streams, *Operations Research and Decisions*, 55–79.
- Kulczycki P., (2005), *Estymatory jądrowe w analizie systemowej*, Wydawnictwa Naukowo-Techniczne, Warszawa.
- LeBaron B., (1994), Chaos and Nonlinear Forecastability in Economics and Finance, *Philosophical Transactions of the Royal Society of London A*, 348 (1686), 397–404.
- Markellos R. N., (2002), Nonlinear Dynamics in Economics and Finance, *Working paper*, Athens University of Economics and Business.
- Morley J., (2009), Macroeconomics, Nonlinear Time Series in, w: Meyers R. A., (red.), *Encyclopedia of Complexity and System Science*, Springer Verlag, New York, 5325–5348.
- Nadaraya E. A., (1964), On Estimating Regression, *Theory of Probability and its Applications*, 9 (1), 141–142.

- Orzeszko W., (2004a), Krótkoterminowe prognozowanie chaotycznych szeregów czasowych, *Przeгляд Statystyczny*, 51 (3), 115–127.
- Orzeszko W., (2004b), How the Prediction Accuracy of Chaotic Time Series Depends on Methods of Determining the Parameters of Delay Vectors, *Dynamic Econometric Models*, 6, 231–239.
- Orzeszko W., (2005), *Identyfikacja i prognozowanie chaosu deterministycznego w ekonomicznych szeregach czasowych*, FPIAKE, PTE, Warszawa.
- Orzeszko W., (2016), *Nieparametryczna identyfikacja nieliniowości w finansowych i ekonomicznych szeregach czasowych*, Wydawnictwo UMK, Toruń.
- Orzeszko W., (2017), Nonparametric Testing for Serial Independence Using the NRL Statistic, *Communications in Statistics - Simulation and Computation*, 46 (7), 5151–5165.
- Osińska M., Górka J., (2006), Identification of Non-linearity in Economic Time Series, *Dynamic Econometric Models*, 7, 83-92.
- Pagan A., Ullah A., (1999), *Nonparametric Econometrics*, Cambridge University Press, Cambridge.
- Racine J. S., (2008), Nonparametric Econometrics: a Primer, *Foundations and Trends in Econometrics*, 3 (1), 1–88.
- Ramsey J. B., (1996), If Nonlinear Models Cannot Forecast, What Use are They?, *Studies in Nonlinear Dynamics & Econometrics*, 1 (2), 65–86.
- Rosenstein M. T., Collins J. J., De Luca C. J., (1993), A Practical Method for Calculating Largest Lyapunov Exponents from Small Data Sets, *Physica D*, 65, 117–134.
- Śliwicki D., (2016), *Estymacja jądrowa w analizie ekonometrycznej*, Wydawnictwo UMK, Toruń.
- Stelmach J., (2014), O interpretacji nieparametrycznych modeli regresyjnych, w: Barczak A. S., Miśkiewicz-Nawrocka M., (red.), *Studia Ekonomiczne*, 203, 154–162, Uniwersytet Ekonomiczny w Katowicach.
- Stinchcombe M. B., Drukker D. M., (2013), Regression Efficacy and the Curse of Dimensionality, w: Chen X., Swanson N., (red.), *Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis*, 527-549, Springer, New York.
- Stone C. J., (1977), Consistent Nonparametric Regression, *Annals of Statistics*, 5, 595–620.
- Stone C. J., (1982), Optimal Global Rates of Convergence for Nonparametric Regression, *Annals of Statistics*, 10, 1040–1053.
- Takens F., (1981), Detecting Strange Attractors in Turbulence, w: Rand D., Young L., (red.), *Dynamical Systems and Turbulence*, Springer-Verlag, 366–381.
- Wand M. P., Jones M. C., (1995), *Kernel Smoothing*, Chapman & Hall, London.
- Watson G. S., (1964), Smooth Regression Analysis, *Sankhya: The Indian Journal of Statistics (Series A)*, 26 (4), 359–372.

WYBRANE ASPEKTY NIEPARAMETRYCZNEGO PROGNOZOWANIA NIELINIOWYCH SZEREGÓW CZASOWYCH

Streszczenie

Regresja nieparametryczna stanowi obiecujące, lecz jednocześnie wciąż niedoceniane podejście do modelowania finansowych i ekonomicznych szeregów czasowych. Polega ona na konstrukcji modeli nieparametrycznych, w których zależność pomiędzy zmiennymi nie jest wyrażona w postaci funkcji o określonej postaci analitycznej lub parametry charakteryzujące tę zależność

należą do przestrzeni nieskończenie wymiarowej. W przeciwieństwie do modeli parametrycznych, modele nieparametryczne nie są ograniczone do określonej z góry postaci, lecz pozwalają „mówić danym samym za siebie”. Z tego względu wydają się interesującym narzędziem prognozowania zwłaszcza w przypadku nieliniowych szeregów czasowych.

W zakresie nieparametrycznych metod regresji na szczególną uwagę zasługują estymatory, które w swojej konstrukcji wykorzystują funkcje jądrowe. Spośród nich najczęściej stosowanym jest estymator Nadarai-Watsona, choć obecnie znane są pewne rozwinięcia tego podejścia. Jednym z nich jest Lokalna Jądrowa Regresja Liniowa, będąca połączeniem estymacji jądrowej i lokalnej aproksymacji liniowej. W pracy przeprowadzono symulacje Monte Carlo, mające na celu ocenę przydatności metod regresji jądrowej do prognozowania nieliniowych szeregów czasowych i porównanie ich z innymi metodami prognozowania.

Słowa kluczowe: regresja nieparametryczna, nieparametryczne prognozowanie, metody jądrowe, nieliniowe szeregi czasowe, analiza Monte Carlo

SEVERAL ASPECTS OF NONPARAMETRIC PREDICTION OF NONLINEAR TIME SERIES

Abstract

Nonparametric regression is an alternative to the parametric approach, which consists of applying parametric models, i.e. models of the certain functional form with a fixed number of parameters. As opposed to the parametric approach, nonparametric models have a general form, which can be approximated increasingly precisely when the sample size grows. Hereby they do not impose such restricted assumptions about the form of the modelling dependencies and in consequence, they are more flexible and let the data speak for themselves. That is why they are a promising tool for forecasting, especially in case of nonlinear time series.

One of the most popular nonparametric regression method is the Nadaraya-Watson kernel smoothing. Nowadays, there are a number of variations of this method, like the local-linear kernel estimator, which combines the local linear approximation and the kernel estimator. In the paper a Monte Carlo study is conducted in order to assess the usefulness of the kernel smoothers to nonlinear time series forecasting and to compare them with the other techniques of forecasting.

Keywords: nonparametric regression, nonparametric forecasting, kernel smoothers, nonlinear time series, Monte Carlo study