Grzegorz Kończak
Michał Miłek
Uniwersytet Ekonomiczny w Katowicach

WYKORZYSTANIE METODY MOVING BLOCK BOOTSTRAP W PROGNOZOWANIU SZEREGÓW CZASOWYCH Z WAHANIAMI OKRESOWYMI*

Wprowadzenie

Celem analizy szeregu czasowego jest między innymi umożliwienie formułowania prognoz. Podstawą wnioskowania o przyszłych wartościach zmiennych są ich przeszłe realizacje. Do prognozowania szeregów czasowych wykorzystuje się różne metody, jak np. metoda naiwna, modele Holta i Wintersa, modele ARiMA, a także różne metody symulacyjne.

Do najczęściej stosowanych metod symulacyjnych w badaniach statystycznych należy metoda bootstrap. Została ona zaproponowana przez Efrona (1979). Metoda bootstrap jest najczęściej wykorzystywana do szacowania wariancji estymatorów oraz testowania hipotez w przypadkach, gdy nie jest znany rozkład statystyki testowej. Metoda ta nie wymaga, by pobierane próby pochodziły z populacji o rozkładzie normalnym. Ze względu na konstrukcję prób losowych w tej metodzie zazwyczaj nie ma możliwości bezpośredniego jej zastosowania w analizach szeregów czasowych. W literaturze są rozważane pewne modyfikacje metody bootstrap prowadzące do możliwości wykonania takich analiz. Jedną z takich metod jest *moving block bootstrap*. Propozycja wykorzystania tej metody do konstrukcji przedziałów predykcji została przedstawiona w artykule dla szeregów czasowych z wahaniami okresowymi. Wyniki otrzymywane za pomocą tej metody zostały porównane z ocenami otrzymanymi za pomocą klasycznej konstrukcji przedziałów predykcji oraz metody ARIMA.

* Część opracowania zrealizowana przez pierwszego z autorów została sfinansowana ze środków Narodowego Centrum Nauki przyznanych na podstawie decyzji numer DEC-2011/03/B/HS4/05630.

1. Przedziałowa predykcja

Podczas przeprowadzania analiz szeregów czasowych bardzo często są konstruowane prognozy punktowe, które określają przyszłą wartość badanej zmiennej poprzez ekstrapolację modelu opisującego badane zjawisko. Ten sposób prognozowania nie niesie jednak pełnej informacji na temat przyszłej wartości zmiennej – nie wiadomo, z jakim prawdopodobieństwem przyszła realizacja znajdzie się w zadanym otoczeniu prognozy. W wielu sytuacjach posiadanie takiej informacji mogłoby prowadzić do podjęcia zupełnie innych decyzji. Po wyznaczeniu prognozy może się okazać, że przedział, który z określonym prawdopodobieństwem obejmuje przyszłe realizacje szeregu czasowego, jest na tyle szeroki, że w ogóle nie warto podejmować ryzyka związanego z działaniem na podstawie takich przewidywań. Z tego powodu coraz częściej konstruuje się prognozy przedziałowe, które uściślają dokładność predykcji. Dzięki temu można określić niepewność prognozy, rozważyć różne strategie, a także porównać prognozy uzyskane na podstawie kilku metod (Chatfield, 1993). Prognoza przedziałowa składa się z dolnej i górnej granicy przedziału, w którym znajdzie się prognozowana wartość z pewnym określonym prawdopodobieństwem. W literaturze przedział ten jest różnie nazywany: zakresem przewidywania (Brockwell, 1987), przedziałem ufności (Granger, Newbold, 1986), a granice przedziału nazywa się limitami prognozy (Wei, 1990). Najczęściej pojawia się jednak sformułowanie: przedział prognozy. Istnieje wiele metod wyznaczania przedziałów prognoz, nie ma jednak jednej, którą można by zastosować we wszystkich przypadkach. W szczególności przedziały teoretyczne nie zawsze są możliwe do zastosowania ze względu np. na złożoność modeli, zwłaszcza wielorównaniowych. W dalszych rozważaniach zostaną przyjęte przedstawione poniżej oznaczenia. Niech $y_1, y_2, ... y_n$ będzie ciągiem obserwacji analizowanego szeregu czasowego, gdzie n oznacza liczbę obserwacji lub, inaczej mówiąc, liczbę rozpatrywanych okresów. Wspomniany szereg obserwacji jest jedną z możliwych realizacji dyskretnego procesu stochastycznego, będącego zmieniającym się w czasie zjawiskiem statystycznym zgodnie z pewnym rozkładem prawdopodobieństwa. Proces stochastyczny może być zapisany następująco $\{Y_t\}_{t=1,2,...,n} = (Y_1, Y_2, ..., Y_n)$. Niech celem analizy będzie postawienie prognozy na k = 1, 2, …, S okresów (przez S będzie oznaczany horyzont prognozy), wówczas prognoza punktowa będzie oznaczona jako $\hat{Y}_n(1), \hat{Y}_n(2), ..., \hat{Y}_n(S)$, natomiast wartości obserwowane w okresie prognozy jako $\hat{y}_n(1), \hat{y}_n(2), ..., \hat{y}_n(S)$. Błąd prognozy w odniesieniu do $\hat{y}_n(t)$ może być zapisany następująco $e_n(t) = \hat{Y}_n(t) - \hat{y}_n(t)$. Jedną z najczęściej stosowanych metod wyznaczania przedziałów prognozy jest obliczenie dolnej i górnej granicy przedziału na podstawie założonego rozkładu (najczęściej rozkładu normalnego). Wówczas do wyznaczania prognozy przedziałowej wykorzystuje się następującą formułę (Box, Jenkins, 1983):

$$P\left(\hat{y}_n(t) + z_{\frac{\alpha}{2}}\sqrt{\mathrm{var}(e_n(t))} < y_n(t) < \hat{y}_n(t) + z_{1-\frac{\alpha}{2}}\sqrt{\mathrm{var}(e_n(t))}\right) = 1 - \alpha \quad (1)$$

gdzie $z_{\frac{\alpha}{2}}$ i $z_{1-\frac{\alpha}{2}}$ są kwantylami standardowego rozkładu normalnego rzędu $\frac{\alpha}{2}$ i $1-\frac{\alpha}{2}$, a $1-\alpha$ to prawdopodobieństwo określające wiarygodność prognozy.

Wzór (1) zakłada, że błędy prognozy mają rozkład normalny oraz że prognoza jest nieobciążona, wówczas prawdziwa jest następująca równość:

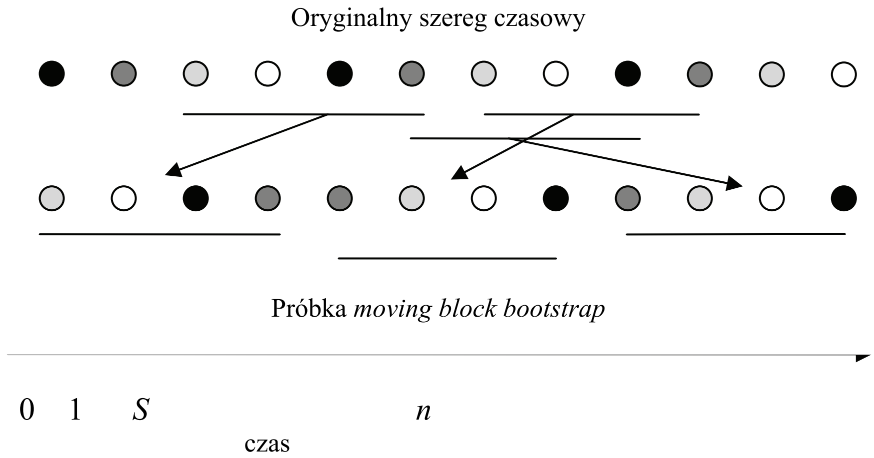
$$E\left(e_n^2(t)\right) = \mathrm{var}\left(e_n(t)\right) \quad (2)$$

Istnieją jednak metody pozwalające na pominięcie założenia o normalności rozkładu i nieobciążoności błędów prognozy (Chatfield, 1993). W tym celu często stosuje się metody bootstrapowe. Ich zaletą jest to, że mogą być stosowane w sytuacji, gdy nie jest znany rozkład badanej zmiennej albo gdy próbka jest zbyt mała, by możliwe było wykorzystanie twierdzeń granicznych. Metoda bootstrap jest odmianą metody Monte Carlo, polegającą na wielokrotnym losowaniu próby z próbki pierwotnej (Kopczewska, Kopczewski, Wójcik, 2009). Z wejściowego szeregu danych jest losowana B-krotnie n-elementowa próbka ze zwracaniem. Na podstawie każdej z próbek jest wyznaczana prognoza, z których jest wyliczana średnia (estymator bootstrapowy) i w efekcie jest otrzymywana prognoza punktowa. Uzyskany zbiór wartości pozwala skonstruować także przedziały prognoz. Dolna i górna granica przedziału jest wyznaczana na podstawie odpowiednich kwantyli wartości uzyskanych w procedurze bootstrap.

Jednak bezpośrednie zastosowanie metody bootstrap w analizie szeregów czasowych z wahaniami okresowymi nie jest możliwe, ponieważ takie rozwiązanie nie uwzględnia okresowości badanego zjawiska. W dalszych rozważaniach dotyczących prognozowania dla wspomnianych szeregów czasowych zostanie wykorzystana modyfikacja metody *moving block bootstrap*.

2. Moving block bootstrap

Efron i Tibshirani (1993) przedstawiają możliwości wykorzystania metody bootstrap w zastosowaniu dla szeregu czasowego. Opisywana metoda *moving block bootstrap* (*MBB*) jest wykorzystywana do estymacji parametrów modelu autoregresyjnego. Idea *MBB* przedstawiona przez Efrona i Tibshirani (1993) została zaprezentowana schematycznie na rysunku 1.



Rys. 1. Idea zastosowania metody *moving block bootstrap*

Metoda ta prowadzi do zwrotnego pobierania próbki pełnych bloków o długości G obserwacji i wstawiania ich łącznie do szeregu czasowego. Na ilustracji przedstawionej na rysunku 1 przyjęto $G = 4$. Możliwości zastosowania metody *moving block bootstrap* w mniejszym stopniu zależą od modelu szeregu czasowego niż w przypadku klasycznej metody bootstrap. Może być zastosowana do analizy szeregów czasowych z występującą autokorelacją, a w dalszych rozważaniach zostanie wykorzystana do analizy szeregów czasowych z wahaniami okresowymi.

3. Wykorzystanie metody moving block bootstrap w analizie szeregów czasowych z wahaniami okresowymi

Montgomery i in. (2008) rozważają dwa modele szeregów czasowych okresowych: model addytywny i model multiplikatywny. W poniższych analizach będzie rozpatrywany model addytywny. Model taki może być zapisany następująco (Montgomery i in., 2008; Zeliaś i in., 2002):

$$Y_t = f(t) + w_t^+ + \varepsilon_t \qquad (3)$$

gdzie:
$t = 1, 2, \ldots n$,
Y_t – wartość szeregu w okresie t,
$f(t)$ – funkcja trendu,
w_t^+ – składniki sezonowe dla modelu addytywnego $w_t^+ = w_{t+S}^+ = \ldots = w_{t+(k-1)S}^+$ spełniające warunek $\sum_{t=1}^{G} w_t^+ = 0$, gdzie t = 1, 2, …, S, natomiast S jest liczbą wyróżnionych okresów w szeregu czasowym, czyli $k = \frac{n}{S}$,
ε_t – składnik resztowy o własnościach $E(\varepsilon_t) = 0$ i $D^2(\varepsilon_t) = \sigma_\varepsilon^2$.

Dla wyróżnionych k pełnych okresów jest spełniony warunek $n = kS$.

Szereg Y_t pozbawiony trendu przyjmuje następującą postać:

$$X_t = w_t^+ + \varepsilon_t \qquad (t = 1, 2, \ldots, n) \tag{4}$$

W otrzymanym szeregu X_t kolejne realizacje $x_1, x_2, \ldots, x_n$ mogą zostać zapisane według następującego schematu:

$$x_1, x_2, \ldots, x_S, x_{S+1}, x_{S+2}, \ldots, x_{2S}, \ldots, x_{(k-1)S+1}, x_{(k-1)S+2}, \ldots, x_{kS} \tag{5}$$

Uwzględniając okresowość zjawiska, można powyższy szereg przedstawić z wyróżnieniem bloków okresowych w następującej postaci:

$$(x_1, x_2, \ldots, x_S), (x_{S+1}, x_{S+2}, \ldots, x_{2S}), \ldots, (x_{(k-1)S+1}, x_{(k-1)S+2}, \ldots, x_{kS}) \tag{6}$$

Niech i-ty (i = 1,. 2, …, k) blok będzie oznaczony przez $\ddot{\mathbf{x}}_i$, czyli:

$$\ddot{\mathbf{x}}_i = (x_{(i-1)S+1}, x_{(i-1)S+2}, \ldots, x_{iS}) \tag{7}$$

Prowadzi to do otrzymania ciągu bloków realizacji szeregu czasowego X_t:

$$\ddot{\mathbf{x}}_1, \ddot{\mathbf{x}}_2, \ldots, \ddot{\mathbf{x}}_k \tag{8}$$

Podobnie jak poprzednio, można określić blok o długości S wartości funkcji trendu:

$$\ddot{\mathbf{f}}(t) = (f(t), f(t+1), \ldots, f(t+S-1)) \tag{9}$$

Przyjmując powyższe oznaczenie, kolejne wartości funkcji trendu mogą być zapisane w postaci ciągu bloków:

$$\ddot{\mathbf{f}}(1), \ddot{\mathbf{f}}(S+1), \ldots, \ddot{\mathbf{f}}((k-1)S+1) \tag{10}$$

Uwzględniając powyższe oznaczenia – (7) i (10) – szereg czasowy Y_t może być zapisany następująco:

$$\mathbf{Y}_t = (\dddot{\mathbf{f}}(1), \dddot{\mathbf{f}}(S+1), \ldots, \dddot{\mathbf{f}}((k-1)S+1)) + (\ddot{\mathbf{x}}_1, \ddot{\mathbf{x}}_2, \ldots, \ddot{\mathbf{x}}_k) \tag{11}$$

Wprowadźmy oznaczenie próby *MBB*:

$$(\dddot{\mathbf{x}}_1^*, \dddot{\mathbf{x}}_2^*, \ldots, \dddot{\mathbf{x}}_k^*)$$

gdzie $\ddot{\mathbf{x}}_i^* = (x_{(j-1)S+1}, x_{(j-1)S+2}, \ldots, x_{jS})$ oraz $i, j \in \{1, 2, \ldots, k\}$. Wykorzystując taką próbę, można określić szereg bootstrapowy, w którym trend jest ustalony, zgodny z otrzymaną oceną uzyskaną na podstawie obliczeń z wyjściowego szeregu czasowego, a składniki sezonowe i odchylenia losowe są dodawane na podstawie próbkowania bootstrap. Tak otrzymany szereg można zapisać następująco:

$$\mathbf{Y}_t^* = (\dddot{\mathbf{f}}(1), \dddot{\mathbf{f}}(G+1), \ldots, \dddot{\mathbf{f}}((k-1)G+1)) + (\ddot{\mathbf{x}}_1^*, \ddot{\mathbf{x}}_2^*, \ldots, \ddot{\mathbf{x}}_k^*) \tag{12}$$

Prognozę na kolejne S okresów czasowych można zapisać następująco:

$$(\hat{Y}_{\mathbf{n}}(1), \hat{Y}_{\mathbf{n}}(2), ..., \hat{Y}_{\mathbf{n}}(S)) = \dddot{\mathbf{f}}(n+1) + \mathbf{x}_i^* \tag{13}$$

gdzie i = 1, 2,…, k. Pobierając w ten sposób B-krotnie próbki bootstrapowe, otrzymuje się B szeregów czasowych:

$$\mathbf{Y}_t^{*(1)}, \mathbf{Y}_t^{*(2)}, \ldots, \mathbf{Y}_t^{*(B)}$$

Dla każdego z szeregów czasowych (12) jest wyznaczana prognoza. Na podstawie B prognoz dla ustalonego okresu $n + v$ (v = 1, 2, …, S) wyznacza się kwantyle rzędu $\frac{\alpha}{2}$ i $1-\frac{\alpha}{2}$. Są to odpowiednio dolna i górna granica przedziału predykcji wyznaczonego metodą *MBB*.

4. Analiza symulacyjna

Celem przeprowadzonych symulacji było porównanie wyników otrzymywania prognoz przedziałowych z wykorzystaniem proponowanej metody wykorzystującej *MBB* oraz klasycznej metody predykcji i metody ARIMA. Do takich porównań zwykle wykorzystuje się dane pochodzące z generatorów liczb losowych. W poniższych rozważaniach zdecydowano się na odwołanie do danych o cenach energii w dniach 1.06.2012-31.12.2012 pochodzących z rynku energii dnia następnego. Dane o cenach były rejestrowane w systemie godzinnym. Symulacje przeprowadzono z wykorzystaniem procedur opracowanych w języku R. We wszystkich symulacjach przyjęto poziom wiarygodności prognozy $1 - \alpha = 0{,}95$. Przebieg procedury symulacyjnej był następujący:

1. Na podstawie obserwacji z 13 pełnych tygodni (13 x 7 x 24 = 2184 obserwacji) wyznaczano prognozy przedziałowe trzema metodami: *MBB*, klasyczną i ARIMA na *k* okresów (godzin) do przodu dla v = 1, 5, 9, 13, 17, 21 oraz na godz. 1.00 kolejnych dni tygodnia dla v = 25, 49, 73, 97, 121, 145.
2. Dla wszystkich wymienionych okresów sprawdzano, czy obserwowana wartość znalazła się w wyznaczonym przedziale predykcji. W każdym przypadku rejestrowano długość przedziału predykcji.
3. Kroki 1 i 2 były przeprowadzone po przesunięciu okna obserwacji związanym z wykreśleniem pierwszych t = 1, 2, …, 1000 obserwacji, co prowadziło do wykonania 1000 powtórzeń wyznaczenia predykcji dla ustalonych wartości v.
4. Każdorazowo pobierano B = 200 próbek bootstrapowych. Otrzymane wyniki zostały uśrednione.

Wyniki symulacji przedstawiono w tabelach 1 i 2. Rezultaty symulacji komputerowych przedstawiono również na rysunku 2. Wyniki uzyskane metodami *MBB* i ARIMA są zbliżone i bliskie założonemu poziomowi wiarygodności prognozy. Znacznie mniejsze częstości pokrycia wartości prognozowanych uzyskano z zastosowaniem metody klasycznej.

Tabela 1

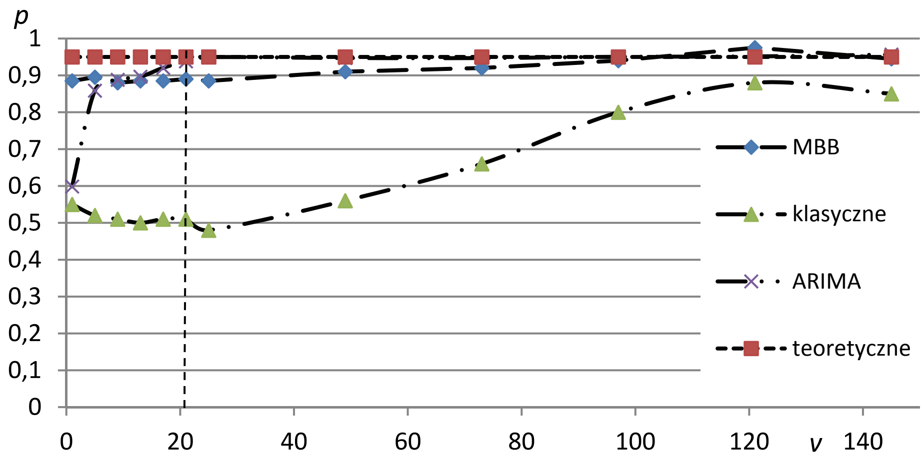
Ocena pokrycia wartości prognozowanych w następnym dniu

v	Ocena prawdopodobieństwa pokrycia		
	MBB	Prognoza klasyczna	ARIMA
1	0,885	0,55	0,599
5	0,895	0,52	0,859
9	0,880	0,51	0,888
13	0,885	0,50	0,897
17	0,885	0,51	0,921
21	0,890	0,51	0,938

Tabela 2

Ocena pokrycia wartości prognozowanych w kolejnych dniach (o godz. 1.00)

v	Ocena pokrycia		
	MBB	Prognoza klasyczna	ARIMA
25	0,885	0,48	0,950
49	0,910	0,56	0,947
73	0,920	0,66	0,947
97	0,940	0,80	0,952
121	0,975	0,88	0,952
145	0,945	0,85	0,957



Rys. 2. Ocena pokrycia wartości prognozowanych trzema metodami

Źródło: Tabela 1 i tabela 2.

Podsumowanie

Przeprowadzone badania pozwoliły na potwierdzenie możliwości zastosowania metody *moving block bootstrap* do prognozowania szeregów czasowych z wahaniami okresowymi. Zaproponowane rozwiązanie nie jest z pewnością jedynym z możliwych i należy dalej poszukiwać metod nieparametrycznych pozwalających na osiągnięcie jeszcze lepszych wyników. W analizach symulacyjnych oparto się na danych rzeczywistych, dla których zwykle nie są spełnione założenia występujące w klasycznych metodach. Analizując uzyskane wyniki, można zauważyć, że częstości pokrycia przedziałem predykcji uzyskanym metodą *MBB* są bardzo zbliżone do wyników uzyskanych z wykorzystaniem metody ARIMA. Przewagą metody *MBB* nad metodą klasyczną jest możliwość jej zastosowania także w przypadku niespełnienia założenia o normalności reszt analizowanego szeregu. Wadą wszystkich analizowanych metod jest wrażliwość na niestabilność rozkładu w czasie.

Literatura

Box G.E.P., Jenkins G.M. (1983): *Analiza szeregów czasowych. Prognozowanie i sterowanie*. Państwowe Wydawnictwo Naukowe, Warszawa.

Brockwell P.J., Davis R.A. (1987): *Time Series: Theory and Methods*. Springer-Verlag, New York.

Chatfield C. (1993): *Calculating Interval Forecasts*. „Journal of Business & Economic Statistics”, Vol. 11, No. 2, s. 121-135, 128.

Efron B. (1979): *Bootstrap Methods: Another Look at the Jackknife*. „Annals Statistics 7”, s. 1-26.

Efron B., Tibshirani R. (1993): *An Introduction to the Bootstrap.* Science Business Media, Inc.

Granger C.W.J., Newbold P. (1986): *Forecasting Economic Time Series*. Academic Press, New York.

Kopczewska K., Kopczewski T., Wójcik P. (2009): *Metody ilościowe w R*. Warszawa.

Montgomery D.C., Jennings C.L., Kulahci M. (2008): *Introduction to Time Series Analysis and Forecasting*. John Wiley & Sons, Inc., New Jersey.

Wei W.W.S. (1990): *Time Series Analysis*. Addison-Wesley, Redwood City, CA.

Zeliaś A., Pawełek B., Wanat S. (2002): *Metody statystyczne. Zadania i sprawdziany*. Państwowe Wydawnictwo Ekonomiczne, Warszawa.

THE USE OF THE MOVING BLOCK BOOTSTRAP METHOD IN PERIODIC TIME SERIES FORECASTING

Summary

The aim of the analysis of the time series is, among others, to facilitate the formulation of prognosis. The basis for the inference of the future variables are their future realizations. There are various methods used in time series forecasting, such as for example naïve method, Holt-Winters models, ARIMA models and various simulation methods.

One of the most popular and widely used simulation method in statistical research is the bootstrap method proposed by B. Efron. It is usually applied in measuring the estimates of the variance and testing the hypotheses in cases when the distribution of the test statistic is unknown. This method does not require for the selected samples to be from the standard normal distribution population. Due to the construction of the random

samples in this method, there is usually no possibility to directly apply it in the analysis of the periodic time series. In the literature written on this subject, there are the proposals to introduce some modifications to the bootstrap method that would provide the possibility to conduct such analyses. One of such methods is the moving block bootstrap. In the present essay, we will present the proposal to apply this method to create the confidential intervals for the periodic time series forecasts. The results gathered by applying that method are compared with the results obtained via the classic construction of the confidential intervals for the forecasts and on the confidential intervals based on ARIMA models.