



Maciej Oesterreich

Zachodniopomorski Uniwersytet Technologiczny w Szczecinie
Wydział Ekonomiczny
Katedra Zastosowań Matematyki w Ekonomii
maciej.oesterreich@zut.edu.pl

SYMULACYJNE BADANIE EFEKTYWNOŚCI WYKORZYSTANIA METOD NUMERYCZNYCH W PROGNOZOWANIU ZMIENNEJ ZAWIERAJĄCEJ LUKI NIESYSTEMATYCZNE

Streszczenie: W artykule przedstawiono symulacyjną analizę efektywności wykorzystania metod numerycznych w prognozowaniu zmiennej ekonomicznej dla luk niesystematycznych. Do budowy prognoz inter- i ekstrapolacyjnych, na podstawie szeregów oczyszczonych z sezonowości, zostały wykorzystane metody: odcinkowa, luków oraz Lagrange'a dla węzłów interpolacyjnych rozmieszczonych proporcjonalnie. Rozpatrywane były trzy warianty luk, różniące się odsetkami brakujących danych. Przeprowadzono również analizę porównawczą dokładności błędów prognoz inter- i ekstrapolacyjnych względem klasycznych modeli szeregu czasowego z trendem liniowym oraz periodycznym składnikiem sezonowym, jak również z trendem wykładniczym z relatywnie stałą sezonowością. Obliczenia zostały wykonane z wykorzystaniem pakietu R oraz Statistica 12.

Słowa kluczowe: metody numeryczne, prognozowanie, luki niesystematyczne, metody symulacyjne.

JEL Classification: C63.

Wprowadzenie

W badaniach prowadzonych dotychczas na temat wpływu liczby luk na dokładność prognoz inter- i ekstrapolacyjnych wykorzystywane były przede wszystkim klasyczne modele szeregu czasowego z wahaniami sezonowymi, w których sezonowość opisywana była za pomocą zmiennych zero-jedynkowych

lub wielomianu trygonometrycznego, a także modele hierarchiczne oraz modele wyrównywania wykładniczego [Oesterreich, 2012; Szmuksta-Zawadzka i Zawadzki, 2014]. W tym ostatnim przypadku były to bądź modele Holta-Wintersa dla danych oryginalnych (z sezonowością), bądź modele Browna lub Holta dla danych oczyszczonych z sezonowości.

Wykorzystanie tego rodzaju modeli nie zawsze pozwala na otrzymanie prognoz dopuszczalnych. Odnosi się to zwłaszcza do zmiennych ekonomicznych o silnym lub bardzo silnym natężeniu wahań sezonowych. Dlatego też nasuwa się pytanie o efektywność predyktorów, opartych na metodach numerycznych.

Celem pracy jest ocena przydatności wybranych metod numerycznych, w prognozowaniu zmiennej o silnym natężeniu sezonowości, w przypadku występowania luk niesystematycznych z wykorzystaniem metod symulacyjnych.

W dotychczasowych zastosowaniach metod numerycznych w prognozowaniu brakujących danych zmiennych z wahaniami sezonowymi rozpatrywanych było dla kilku założonych wariantów luk kilka lub, co najwyżej, kilkanaście układów luk [Cheba, 2004; Oesterreich, 2010]. Ograniczało to uogólnianie wyników na dowolne układy luk. Taką możliwość daje zastosowanie metod symulacyjnych.

W pracy została przeprowadzona także analiza porównawcza dokładności wybranych metod numerycznych w prognozowaniu inter- i ekstrapolacyjnym z prognozami otrzymanymi na podstawie predyktorów, opartych na klasycznych modelach szeregu czasowego z wahaniami sezonowymi, a także na modelach hierarchicznych.

1. Metody badawcze

Z uwagi na to, że zarówno metody numeryczne, jak i klasyczne modele szeregu czasowego oraz modele hierarchiczne zostały szczegółowo opisane w literaturze ekonometrycznej z zakresu metod prognozowania, ograniczono się jedynie do syntetycznej prezentacji analitycznych zapisów modeli, wykorzystanych do budowy prognoz.

W celu budowy prognoz brakujących danych zostały wykorzystane następujące metody numeryczne:

- a) metoda interpolacji wielomianowej Lagrange'a,
- b) metoda odcinkowa,
- c) metoda łuków.

Metoda interpolacji wielomianowej Lagrange'a oparta jest na twierdzeniu, że jakkolwiek jest dana funkcja $f(x)$ i jakkolwiek są wybrane węzły interpolacji x_0, x_1, \dots, x_n , to istnieje dokładnie jeden wielomian interpolacyjny $P(x_i)$ stopnia n , który w punktach x_0, x_1, \dots, x_n przyjmuje te same wartości, co dana funkcja $f(x)$ [Fortuna, Macukow i Wąsowski, 1993]. Wzór interpolacyjny Lagrange'a można zapisać w sposób następujący [Stoer, 1979]:

$$P(x) = \sum_{i=0}^n f_i L_i(x) = \sum_{i=0}^n f_i \prod_{\substack{k=0 \\ k \neq i}}^n \frac{x - x_k}{x_i - x_k} \quad (1)$$

W pracy [Grabiński, Wydymus i Zeliaś, 1979] zostały zaprezentowane inne metody interpolacji brakujących danych w ekonomicznych szeregach czasowych, oparte na metodach numerycznych:

a) metodzie odcinkowej:

$$y_0 = \frac{t_0 - t_2}{t_1 - t_2} y_1 + \frac{t_0 - t_1}{t_2 - t_1} y_2, \quad (2)$$

$$t_1 < t_0 < t_2$$

gdzie:

$(y_1, t_1), (y_2, t_2)$ – współrzędne dwóch sąsiednich punktów, na podstawie których określa się równanie kolejnego odcinka,

b) metodzie łuków I:

$$y_1^* = \frac{(t_1 - t_2)(t_1 - t_3)}{(t_0 - t_2)(t_0 - t_3)} y_0 + \frac{(t_1 - t_0)(t_1 - t_3)}{(t_2 - t_0)(t_2 - t_3)} y_2 + \frac{(t_1 - t_0)(t_1 - t_2)}{(t_3 - t_0)(t_3 - t_2)} y_3, \quad (3)$$

$$t_0 < t_1 < t_2 < t_3$$

c) metodzie łuków II:

$$y_2^* = \frac{(t_2 - t_1)(t_2 - t_3)}{(t_0 - t_1)(t_0 - t_3)} y_0 + \frac{(t_2 - t_0)(t_2 - t_3)}{(t_1 - t_0)(t_1 - t_3)} y_1 + \frac{(t_2 - t_0)(t_2 - t_2)}{(t_3 - t_0)(t_3 - t_2)} y_3, \quad (4)$$

$$t_0 < t_1 < t_2 < t_3$$

Z zapisów analitycznych powyższych metod wynika, że wymagają one informacji dla pewnych obserwacji szeregu. W tabeli 1 przedstawiono warunki stosowalności powyższych metod.

Tabela 1. Warunki stosowalności wybrane metod numerycznych

Metoda	Warunki stosowalności
Interpolacji wielomianowej Lagrange'a	Dostępność danych w miejscach występowania węzłów interpolacyjnych
Odcinkowa	Dostępna pierwsza i ostatnia obserwacja szeregu
Łuków I	Dostępna pierwsza i dwie ostatnie obserwacje szeregu
Łuków II	Dostępne pierwsze dwie i ostatnia obserwacja szeregu

Źródło: Szmuksta-Zawadzka i Zawadzki [2012].

Metody: odcinkowa, łuków I i łuków II mogą służyć wyłącznie do celów interpolacji, czyli szacowania brakujących wyrazów szeregu czasowego, leżących między wyrazami znanymi. Ich wadą jest także brak możliwości określenia *a priori* błędów dokonanych szacunków [Dittmann, 2006]. Prognozy ekstrapolacyjne buduje się w sposób pośredni, przez zastosowanie innych metod dla szeregu uzupełnionego o prognozy interpolacyjne, np. ekstrapolacji trendu czy wyrównywania wykładniczego.

W celach porównawczych prognozy inter- i ekstrapolacyjne zbudowano także z wykorzystaniem klasycznych modeli szeregu czasowego wahaniami sezonowymi, różniące się sposobami opisu tych wahań.

Zapisy analityczne modeli, w których stałe (w.tryg_SS) (5) i relatywnie stałe (w.tryg_RSS) (6) wahania sezonowe opisane zostały za pomocą sino- i cosinusoidalnych składowych harmonicznnych:

$$Y_t = \alpha_0 + \alpha_1 t + \sum_{i=1}^{\frac{12}{2}} (a_{ji} \sin \omega_i t + b_{ji} \cos \omega_i t) + \varepsilon_t, \quad (5)$$

$$\ln Y_t = \alpha_0 + \alpha_1 t + \sum_{i=1}^{\frac{12}{2}} (a_{ji} \sin \omega_i t + b_{ji} \cos \omega_i t) + \varepsilon_t, \quad (6)$$

gdzie:

$$\omega_i = \frac{2\pi \cdot i}{m} \quad (7)$$

Drugi, równoważny sposób zapisu polega na wykorzystaniu zmiennych zero-jedynkowych. Modele ze stałymi (0/1_SS) (8) i relatywnie stałymi (0/1_RSS) (9) wahaniami sezonowymi wyrażają się wzorami:

$$Y = \beta_0 + \beta_1 t + \sum_{k=1}^{12} d_{0k} Q_{kt} + \varepsilon_t, \quad (8)$$

$$\ln Y = \beta_0 + \beta_1 t + \sum_{k=1}^{12} d_{0k} Q_{kt} + \varepsilon_t \quad (9)$$

Przy czym:

$$\sum_{k=1}^{12} d_{0k} = 0, \quad (10)$$

gdzie:

- $f(t)$ – funkcja trendu,
- d_{0k}, d_{1k} – parametry stojące przy zmiennych opisujących wahania sezonowe,
- Q_{kt} – zmienne zero-jedynkowe, przyjmujące wartość 1 w podokresie k oraz 0 w pozostałych.

Prognozy inter- i ekstrapolacyjne zostały wyznaczone także na podstawie predyktorów opartych na modelach hierarchicznych szeregu czasowego. Dla wahań sezonowych o cyklu 12-miesięcznym wyróżnia się cztery modele 2-stopniowe i trzy modele 3-stopniowe.

Istota modeli hierarchicznych zawiera się w tym, że ich parametry są średnimi z parametrów d_{0k} z klasycznych modeli szeregu czasowego, a ich liczba nie przekracza połowy długości cyklu wahań.

Przykładowe zapisy modelu 2-stopniowego (H62), ze stałą (11) i relatywnie stałą sezonowością (12), są następujące [Zawadzki (red.), 2003]:

$$Y_{srt} = \alpha_0 + \alpha_1 t + \sum_{s=1}^6 b_{0s} Q_{st} + \sum_{r=1}^2 b_{0sr} Q_{str} + \varepsilon_t \quad (11)$$

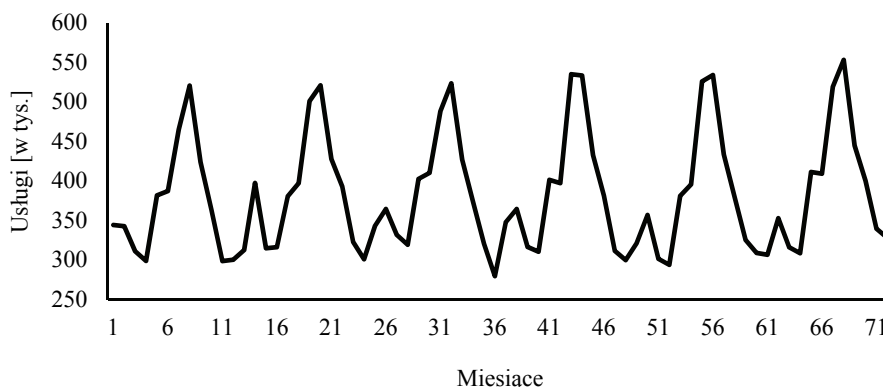
$$\ln Y_{srt} = \alpha_0 + \alpha_1 t + \sum_{s=1}^6 b_{0s} Q_{st} + \sum_{r=1}^2 b_{0sr} Q_{str} + \varepsilon_t \quad (12)$$

Przy czym:

$$\sum_{s=1}^6 b_{0s} = \sum_{r=1}^2 b_{0sr} \quad (13)$$

2. Charakterystyka zmiennej prognozowanej

Modelowaniu i prognozowaniu poddano liczbę udzielonych noclegów w obiektach zbiorowego zakwaterowania w województwie śląskim w latach 2009-2014 według miesięcy. Dane statystyczne zaczerpnięto z ukazujących się w cyklu rocznym opracowań GUS pt. *Turystyka*. Dane z lat 2009-2013 posłużyły do estymacji parametrów modeli, natomiast rok 2014 był okresem empirycznej weryfikacji prognoz. Kształtowanie się zmiennej prognozowanej przedstawiono na rys. 1.



Rys. 1. Udzielone noclegi w obiektach zbiorowego zakwaterowania w województwie śląskim według miesięcy w latach 2009-2014

Źródło: Na podstawie: GUS [2009-2014].

W tabeli 2 zostały przedstawione zostały oceny wskaźników sezonowości zmiennej prognozowanej.

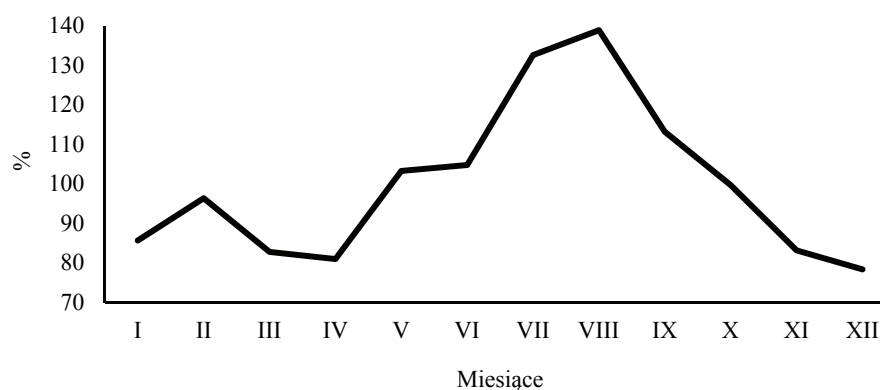
Tabela 2. Kształtowanie się ocen wskaźników sezonowości według miesięcy [w %]

Miesiące	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Wskaźniki sezonowości	85,7	96,4	82,8	81,0	103,3	104,8	132,6	138,9	113,2	99,7	83,2	78,4

Źródło: Na podstawie: GUS [2009-2014].

Z analizy informacji zawartych w tab. 2 wynika, że zmienna prognozowana charakteryzowała się dwoma „szczytami” sezonowymi liczby udzielonych noclegów w ciągu roku. Przypadały one w okresie zimowym na luty (96,4%), a w letnim na sierpień (138,9%). Miesiące te odpowiadały terminom ferii zimowych oraz wakacji. Minimalną ocenę wskaźnik przyjął w grudniu (78,4%). Różnica pomiędzy skrajnymi wartościami wskaźników sezonowych wynosiła 54,2 p.p., co wskazuje na silne wahania analizowanej zmiennej.

Kształtowanie się wskaźników sezonowości w sposób graficzny zostało przedstawione na rys. 2.



Rys. 2. Kształtowanie się wskaźników sezonowości dla liczby udzielonych noclegów w województwie śląskim w obiektach zbiorowego zakwaterowania według miesięcy w latach 2009-2014

Źródło: Na podstawie: GUS [2009-2014].

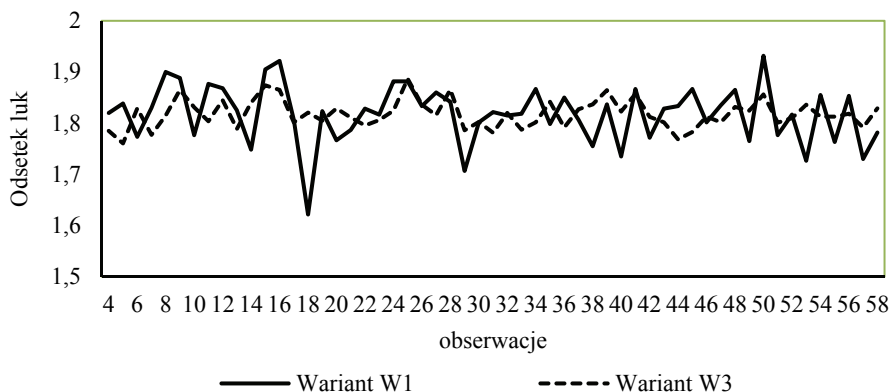
3. Przebieg badania

Proces badania efektywności numerycznych metod prognozowania brakujących danych przebiegał w kilku etapach.

W pierwszym etapie zostały wyeliminowane wahania sezonowe, ponieważ proces prognozowania w przypadku metod numerycznych oparty jest na danych oczyszczonych. Następnie, korzystając z pakietu statystycznego R, wylosowano po 10 tys. układów luk niesystematycznych dla trzech wariantów, zawierających odpowiednio:

- W1 – 6 luk (10% długości okresu próby),
- W2 – 12 luk (20% długości okresu próby),
- W3 – 30 luk (30% długości okresu próby).

Na rysunku 3 przedstawiono odsetki luk, przypadające na poszczególne obserwacje szeregu, według wariantów W1 i W3. Ze względu na wymogi stosowalności metod numerycznych, luki nie były losowane dla obserwacji nr 1-3 oraz 59-60.



Rys. 3. Odsetki luk przypadające na obserwacje 4-58 dla wariantów luk W1 i W3

Z rysunku 3 wynika, że odsetki luk przypadające na kolejne obserwacje szeregu „oscylowały” wokół 1,8% dla obu analizowanych wariantów. Nieco większe odchylenia od tej wartości odnotowano dla wariantu W1 dla obserwacji nr 18, 29, 50. Wzrost liczby luk w wariacie W3 spowodował, że częstości luk tylko nieznacznie odchylają się od wartości 1,8%.

W trzecim etapie dla każdego spośród 10 tys. układów luk dla wariantów W1-W3, korzystając z metod numerycznych, zbudowano wyjściowe prognozy inter- i ekstrapolacyjne. Prognozy ostateczne otrzymano po przemnożeniu prognoz wyjściowych przez wskaźniki sezonowości. W przypadku klasycznych modeli szeregu czasowego najpierw szacowano parametry modeli o równaniach 5-6 i 8-9 dla danych oryginalnych (z sezonowością), a następnie wyznaczono prognozy inter- i ekstrapolacyjne. Proces ten przebiegał podobnie dla modeli hierarchicznych.

Etapem końcowym było obliczenie średnich błędów względnych obu rodzajów prognoz.

4. Wyniki badań

W tabeli 3 zestawione zostały przeciętne oceny błędów prognoz interpolacyjnych dla poszczególnych wariantów luk, dla wszystkich metod numerycznych i modeli klastycznych oraz najlepszego modelu hierarchicznego.

Tabela 3. Przeciętne względne błędy (MAPE) prognoz interpolacyjnych według wariantów luk i metod

Metoda	Błędy (MAPE) prognoz interpolacyjnych [w %]		
	Wariant W1	Wariant W2	Wariant W3
L_WP3	4,223	4,218	4,202
L_WP4	7,032	7,070	7,067
ODC	2,796	2,754	2,728
LUK1	3,618	3,565	3,553
LUK2	3,648	3,639	3,691
0/1_SS	3,084	3,159	3,226
0/1_RSS	3,018	3,092	3,160
w.tryg_SS	3,110	3,184	3,242
w.tryg_RSS	3,119	3,188	3,272
H62_SS	4,534	4,603	4,624
H62_RSS	4,541	4,601	4,617

Z informacji zawartych w tab. 3 wynika, że przeciętne oceny błędów prognoz interpolacyjnych, wyznaczonych za pomocą metod numerycznych wraz ze wzrostem liczby luk, wykazywały tendencję malejącą. Natomiast w przypadku metod klasycznych oceny błędów tego rodzaju prognoz ma miejsce niewielka tendencja wzrostowa.

Najniższe przeciętne oceny błędów, niezależnie od wariantu luk, uzyskano dla metody odcinkowej. Zawarły się one w przedziale od 2,728% (wariant W3) do 2,796% (wariant W1).

Spośród pozostałych metod numerycznych błędy prognoz wyższe o ok. 0,8 p.p. otrzymano dla metody luków I. Były one jednocześnie od 0,030 do 0,138 p.p. niższe niż dla metody luków II.

Dla metody Lagrange'a wyraźnie lepsze efekty dało zastosowanie trzech węzłów (L_WP3) niż czterech węzłów rozmieszczonych proporcjonalnie (L_WP4). Błędy uzyskane za pomocą pierwszego wariantu tej metody zawierały się w przedziale od 4,202 (W3) do 4,223 (W1) i były aż o ok. 2,8 p.p. niższe od wariantu drugiego.

Spośród modeli klasycznych najniższe oceny błędów, zawierające się w przedziale od 3,018% dla wariantu W1 do 3,160% dla wariantu W3, otrzymano za pomocą modelu, w którym relatywnie stała sezonowość została opisana poprzez zmienne zerowyjedynekowe (0/1_RSS). Były one o 0,222-0,432 p.p. wyższe od otrzymanych metodą odcinkową i tylko o ok. 0,06 p.p. niższe od błędów prognoz dla modelu ze stałą sezonowością (0/1_SS), a także ok. 0,09 p.p. niższe niż dla modelu, w którym stałą sezonowość opisano za pomocą wielomianu trygonometrycznego (w.tryg_SS).

Błędy prognoz interpolacyjnych dla najlepszego modelu hierarchicznego (H62) były wyższe aż od 1,738 p.p. (W1 – SS) do 1,890 p.p. (W3 – RSS) w porównaniu z najlepszą metodą numeryczną.

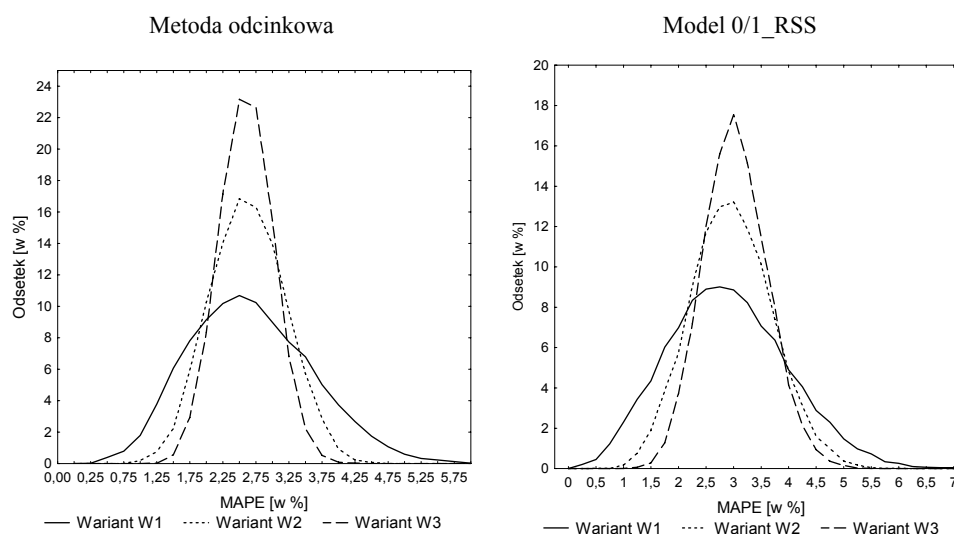
W tabeli 4 zestawione zostały statystyki opisowe rozkładów średnich względnych błędów (MAPE) prognoz interpolacyjnych dla metody odcinkowej (ODC) modelu klasycznego ze zmiennymi zerowymi i relatywnie stałą sezonowością (0/1_RSS) oraz modelu hierarchicznego H62 (H62_RSS).

Tabela 4. Statystyki opisowe rozkładów średnich względnych błędów (MAPE) prognoz interpolacyjnych dla wariantów luk W1-W3 i wybranych metod

Metoda	Wariant	Q1	Mediana	Q3	S(x)	Skośność	Min	Max	R
ODC	W1	2,120	2,733	3,409	0,923	0,348	0,266	6,271	6,005
	W2	2,355	2,747	3,142	0,561	0,061	1,029	4,723	3,694
	W3	2,449	2,727	3,003	0,403	0,068	1,463	4,400	2,937
0/1_RSS	W1	2,250	2,958	3,734	1,083	0,304	0,338	8,141	7,803
	W2	2,576	3,067	3,582	0,735	0,177	0,626	6,002	5,376
	W3	2,759	3,136	3,540	0,583	0,212	1,129	5,545	4,416
H62_RSS	W1	3,547	4,441	5,423	1,373	0,413	0,923	10,037	9,114
	W2	3,943	4,563	5,213	0,927	0,260	1,997	8,530	6,533
	W3	4,116	4,599	5,094	0,722	0,143	2,004	7,527	5,523

Z informacji zawartych w tab. 4 wynika, że wraz ze wzrostem liczby luk wzrastały oceny błędów pierwszych kwartyli dla wszystkich metod. Dla metody odcinkowej mediany błędów były bardzo zbliżone i nie wykazywały jednego kierunku zmian. Natomiast zarówno dla modelu 0/1_RSS, jak i H62_RSS wzrastały wraz ze wzrostem odsetek luk. Dla wszystkich metod oceny trzecich kwartyli, jak również odchylen standardowych błędów prognoz, wykazywały tendencję malejącą. Malejące oceny odchylenia standardowego świadczą o wysmuklaniu rozkładów błędów prognoz. Bez względu na rodzaj metody mały rozstęp, jak i rozstęp międzykwartylowe. Rozkłady błędów dla omawianych metod charakteryzowały się prawostronną asymetrią, której siła malała wraz ze wzrostem liczby luk.

Na rysunku 4 przedstawiano kształtowanie się rozkładów empirycznych średnich względnych błędów prognoz interpolacyjnych dla metody odcinkowej oraz modelu 0/1_RSS.



Rys. 4. Empiryczne rozkłady średnich względnych błędów prognoz interpolacyjnych dla metody odcinkowej oraz modelu 0/1_RSS według wariantów luk

W tabeli 5 zestawione zostały przeciętne oceny błędów prognoz ekstrapolacyjnych dla poszczególnych wariantów luk oraz metod

Tabela 5. Przeciętne względne błędy (MAPE) prognoz ekstrapolacyjnych według wariantów luk i metod

Metoda	Błędy prognoz ekstrapolacyjnych [w %]		
	Wariant W1	Wariant W2	Wariant W3
L WP3	5,966	5,966	5,966
L WP4	84,627	84,627	84,627
ODC*	2,841	2,964	3,114
LUK1*	2,912	3,131	3,420
LUK2*	2,936	3,205	3,574
0/1 SS	4,013	4,016	4,042
0/1 RSS	3,951	3,949	3,968
w.tryg SS	4,033	4,057	4,105
w.tryg RSS	3,982	4,022	4,081
H62 SS	4,558	4,615	4,681
H62 RSS	4,537	4,604	4,672

* Wyjściowe prognozy ekstrapolacyjne wyznaczone zostały za pomocą modeli trendu na podstawie szeregów uzupełnionych o prognozy interpolacyjne.

Z informacji zwartych w tab. 5 wynika, że wraz ze wzrostem liczby luk przeciętne oceny błędów, poza metodą Lagrange'a, wykazywały zróżnicowaną co do tempa tendencję rosnącą. Najszybciej, bo o ok. 0,2-0,3 p.p. wzrastały błędy prognoz otrzymanych za pomocą metody łuków II, a najwolniej dla klasycz-

nych modeli szeregu czasowego 0/1_RSS i 0/1_SS. Najniższymi przeciętnymi błędami prognoz ekstrapolacyjnych, podobnie jak w przypadku prognoz interpolacyjnych, charakteryzowała się metoda odcinkowa. Oceny te zawierały się w przedziale od 2,841% (W1) do 3,114% (W3). Tylko nieznacznie wyższe przeciętne błędy – od 0,072 p.p. (W1) do 0,306 p.p. (W3) – otrzymano dla metody łuków I.

Najwyższymi przeciętnymi ocenami prognoz ekstrapolacyjnych charakteryzowała się metoda Lagrange'a. Ze względu na charakter tej metody, wielkość błędu nie zależała od wariantu łuk – wielomian na podstawie, którego zbudowano prognozy, był zawsze taki sam, niezależnie od kombinacji czy też liczby łuk. Wyraźne, bo kilkunastokrotnie niższe ich oceny otrzymano dla wersji z trzema węzłami rozmieszczonymi proporcjonalnie. Wysokie oceny błędów dla wersji z czterema węzłami były następstwem działania efektu Rungego [Ralson i Rabinowitz, 1978, s. 65-66], objawiającego się spadkiem jakości zarówno interpolacji, jak i ekstrapolacji, wraz ze wzrostem ilości węzłów interpolacyjnych rozmieszczonych proporcjonalnie.

Spośród metod klasycznych najniższymi przeciętnymi błędami prognoz ekstrapolacyjnych, zawierającymi się w przedziale od 3,949% (W1) do 3,968 (W3), charakteryzował się model, w których relatywnie stałą sezonowość opisano za pomocą zmiennych zero-jedynkowych. Tylko nieznacznie wyższymi ocenami błędów – od 0,031 p.p. (W1) do 0,113 p.p. (W3) – charakteryzował się model z wielomianem trygonometrycznym i stałą sezonowością.

Stosunkowo wyższe oceny błędów otrzymano dla najlepszych modeli hierarchicznych – były one od 1,558 p.p. dla wariantu W3 do 1,696 p.p. dla wariantu W1 wyższe od błędów uzyskanych za pomocą metody odcinkowej. Wynikało to z faktu, że oceny parametrów tych modeli są uśrednionymi ocenami parametrów d_{0k} modeli klasycznych szeregu czasowego z wahaniami sezonowymi. Oznacza to, że ich wykorzystanie w prognozowaniu brakujących danych, w warunkach silnego natężenia wahań sezonowych, jest ograniczone.

W tabeli 6 zestawiono statystyki opisowe rozkładów średnich względnych błędów (MAPE) prognoz interpolacyjnych dla metody odcinkowej (ODC), modelu klasycznego ze zmiennymi zero-jedynkowymi i relatywnie stałą sezonowością (0/1_RSS) oraz modelu hierarchicznego H62 (H62_RSS).

Tabela 6. Statystyki opisowe rozkładów średnich względnych błędów (MAPE) prognoz ekstrapolacyjnych dla wariantów łuk W1-W3 i wybranych metod

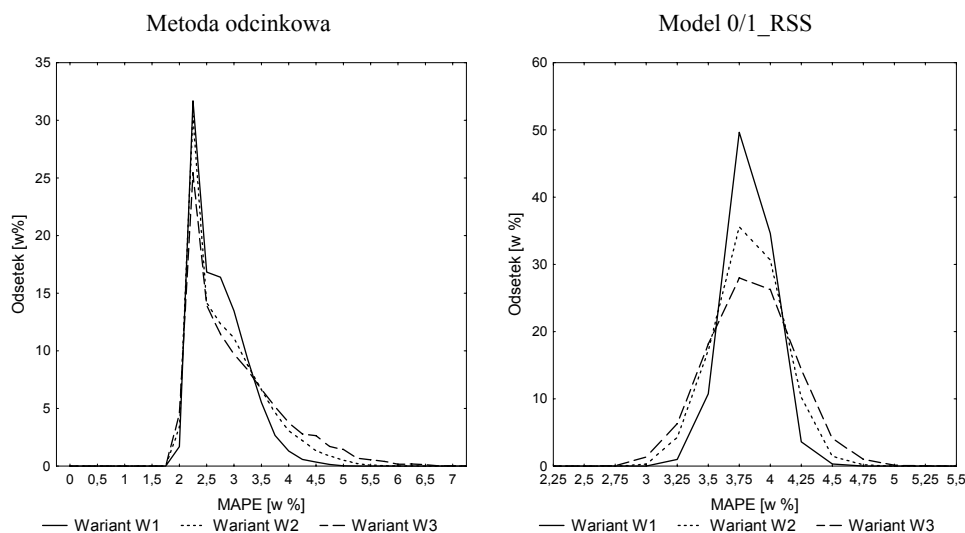
Metoda	Wariant	Q1	Mediana	Q3	S(x)	Skośność	Min	Max	R
1	2	3	4	5	6	7	8	9	10
ODC	W1	2,409	2,747	3,146	0,493	0,981	2,125	5,868	3,743
	W2	2,414	2,791	3,349	0,662	1,149	2,058	6,417	4,359
	W3	2,445	2,876	3,558	0,835	1,288	1,994	7,786	5,792

cd. tabeli 6

1	2	3	4	5	6	7	8	9	10
0/1_RSS	W1	3,847	3,955	4,059	0,177	-0,067	3,256	4,819	1,563
	W2	3,777	3,952	4,120	0,260	0,001	3,018	5,003	1,985
	W3	3,740	3,966	4,192	0,333	0,107	2,950	5,560	2,609
H62_RSS	W1	4,435	4,527	4,631	0,154	0,390	4,007	5,312	1,305
	W2	4,443	4,590	4,748	0,234	0,378	3,785	5,747	1,962
	W3	4,453	4,657	4,861	0,315	0,436	3,696	6,496	2,800

Z informacji zawartych w tab. 6 wynika, że wraz ze wzrostem odsetek luk wzrastały również oceny median błędów prognoz. Jedynie w przypadku modelu klasycznego ze zmiennymi zero-jedynkowymi i relatywnie stałą sezonowością (0/1_RSS) w wariancie W2 nastąpił jej nieznaczny spadek w porównaniu z wariantem W1. W odróżnieniu od błędów prognoz interpolacyjnych, zarówno rozstęp, jak i rozstęp międzykwartylowy ulegały zwiększeniu – wzrastały oceny tak kwartyła pierwszego, jak i kwartyła trzeciego. Analizowane rozkłady błędów charakteryzowały się asymetrią prawostronną, jednak poszczególne metody wyraźnie różniły się jej siłą. Metoda odcinkowa charakteryzowała się bardzo wysokimi ocenami wskaźników skośności – dla wariantów W2 i W3 przekraczającymi 1. Natomiast dla modelu 0/1_RSS ocena tego wskaźnika skośności była bliska 0.

Na rysunku 5 przedstawiano kształtowanie się rozkładów empirycznych średnich względnych błędów prognoz ekstrapolacyjnych, otrzymanych za pomocą metody odcinkowej oraz modelu 0/1_RSS.



Rys. 5. Empiryczne rozkłady średnich względnych błędów prognoz ekstrapolacyjnych dla metody odcinkowej oraz modelu 0/1_RSS według wariantów luk

Z rysunku 6 wynika, że kształty rozkładów błędów prognoz ekstrapolacyjnych dla obu przedstawionych metod wyraźnie się różniły. Dla metody odcinkowej wyraźnie dostrzegalna jest silna asymetria prawostronna, podczas gdy dla modelu klasycznego asymetria praktycznie nie występuje. Może to wynikać z pośredniego sposobu budowy prognozy ekstrapolacyjnych dla metod numerycznych. Jednak pomimo różnic, kierunek zmian spowodowany przez rosnącą liczbę luk był zbliżony dla obu analizowanych metod. Potwierdza to analizę statystyk opisowych rozkładów średnich względnych błędów prognoz ekstrapolacyjnych.

Podsumowanie

Z przeprowadzonych w pracy badań wynika, że za wykorzystaniem wybranych metod numerycznych do budowy prognoz inter- i ekstrapolacyjnych, w przypadku wystąpienia luk niesystematycznych, przemawia dokładność uzyskanych za ich pomocą prognoz, zwłaszcza na tle klasycznych modeli szeregu czasowego. Należy mieć jednak na uwadze ograniczenia tych metod, a przede wszystkim możliwość ich zatasowania wyłącznie dla szeregów niewykazujących wahań sezonowych lub takich, z których zostały one usunięte. Z przeprowadzonych analiz można wyprowadzić następujące wnioski o charakterze szczegółowym:

1. Najniższe oceny błędów prognoz inter- i ekstrapolacyjnych otrzymano dla metody odcinkowej.
2. Spośród pozostałych metod numerycznych nieznacznie wyższe oceny uzyskano dla obu metod łuków. Błędy prognoz interpolacyjnych zbudowanych za ich pomocą były średnio o ok. 0,7-0,9 p.p. wyższe, a błędy prognoz ekstrapolacyjnych o ok. 0,1-0,3 p.p.
3. Dla metody Lagrange'a dla czterech węzłów interpolacyjnych rozmieszczonych proporcjonalnie oceny błędów prognoz ekstrapolacyjnych były ponad 14-krotnie wyższe niż dla 3 węzłów. Związane było to z wystąpieniem tzw. efektu Rungego.
4. Oceny błędów prognoz interpolacyjnych otrzymanych zarówno na podstawie klasycznych modeli, jak i hierarchicznych modeli szeregu czasowego były wyższe o ok. 0,2-0,4 p.p. od uzyskanych metodą odcinkową. Dla prognoz ekstrapolacyjnych oceny te były wyższe o ok. 0,8-1,1 p.p.

Z przeprowadzonych w pracy badań wynika również, że występowanie luk niesystematycznych w szeregu czasowym wykazującym wahania sezonowe nie musi uniemożliwiać budowy prognoz na jego podstawie. Zastosowanie przedstawionych metod i uzupełnienie szeregu daje badaczowi możliwość przeprowadzania analiz, które bez tej operacji nie mogłyby zostać wykonane.

Literatura

- Cheba K. (2004), *Zastosowanie metod prognozowania stanów terminowych depozytów bankowych w warunkach braku pełnej informacji*, „Folia Universitatis Agriculturae Stetinensis. Oeconomica”, nr 237(43), s. 339-344.
- Dittmann P. (2006), *Prognozowanie w przedsiębiorstwie. Metody i ich zastosowanie*, Wolters Kluwer, Kraków.
- Fortuna Z., Macukow B., Wąsowski J. (1993), *Metody numeryczne*, Wydawnictwo Naukowo-Techniczne, Warszawa.
- Grabiński T., Wydimus S., Zeliaś A. (1979), *Z badań nad metodami predykcji brakujących informacji*, „Zeszyty Naukowe Akademii Ekonomicznej w Krakowie”, nr 114, s. 31-59.
- GUS (2009-2014), *Turystyka*, informacje i opracowania statystyczne, Warszawa.
- Oesterreich M. (2010), *Wykorzystanie metod numerycznych w prognozowaniu brakujących danych w szeregach czasowych z sezonowością*, „Folia Universitatis Agriculturae Stetinensis. Oeconomica”, nr 280(59), s. 77-85.
- Oesterreich M. (2012), *Symulacyjne badanie wpływu częstości występowania luk niesystematycznych na dokładność prognoz*, „Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu. Ekonometria”, nr 4(38), s. 186-196.
- Ralson A., Rabinowitz F. (1978), *First Course of Numerical Analysis – Second Edition*, McGraw-Hill Inc., New York.
- Stoer J. (1979), *Wstęp do metod numerycznych. Tom pierwszy*, PWN, Warszawa.
- Szmuksta-Zawadzka M., Zawadzki J. (2012), *Z badań nad metodami prognozowania na podstawie niekompletnych szeregów czasowych z wahaniami okresowymi (sezonowymi)*, „Przegląd Statystyczny”, numer specjalny, nr 1, s. 140-154.
- Szmuksta-Zawadzka M., Zawadzki J. (2014), *Zastosowanie wybranych modeli adaptacyjnych w prognozowaniu brakujących danych w szeregach ze złożoną sezonowością dla luk niesystematycznych*, „Metody Ilościowe w Badaniach Ekonomicznych. Quantitative Methods in Economics”, t. 15, z. 4, s. 181-195.
- Zawadzki J. (red.) (2003), *Zastosowanie hierarchicznych modeli szeregów czasowych w prognozowaniu zmiennych ekonomicznych z wahaniami sezonowymi*, Wydawnictwo Akademii Rolniczej w Szczecinie, Szczecin.

THE SIMULATION EFFICIENCY ANALYSIS OF NUMERICAL METHODS IN FORECASTING VARIABLE WITH UNSYSTEMATIC GAPS

Summary: In the article was presented an efficiency analysis of numerical methods in forecasting economic variable with unsystematic gaps. To construction of inter- and extrapolative forecasts, based on the seasonal adjusted time series, were used: the segmental method, the arches method and the Lagrange interpolation method for nodes distributed proportionally. In analysis were considered three variants of gaps, differing in

the percent-age of the missing data. A comparative analysis of the accuracy of forecast errors of classical time series with linear trend and periodic seasonal component and exponential trend with relatively constant seasonality was also performed. Calculations were made using R environment and Statistica 12.

Keywords: numerical methods, forecasting, unsystematic gaps, simulation methods.