

*Maciej OESTERREICH*

## O METODZIE BADANIA WPŁYWU LUK NA DOKŁADNOŚĆ PROGNOZ INTER- I EKSTRAPOLACYJNYCH

## THE METHOD OF ANALYSIS OF THE IMPACT OF OCCURRENCE OF GAPS ON ACCURACY OF INTER- AND EXTRAPOLATIVE FORECASTS

Katedra Zastosowań Matematyki w Ekonomii, Zachodniopomorski Uniwersytet Technologiczny w Szczecinie, ul. Klemensa Janickiego 31, 71-270 Szczecin, e-mail: moesterreich@zut.edu.pl

**Summary.** In paper of Oesterreich (2015) were presented the results of application of simulation methods in analysis of the impact of occurrence of unsystematic gaps on the levels of inter- and extrapolative forecasts errors in time series with seasonal fluctuations of the cycle of 12 months. In these studies were analysed distributions of forecast errors for 100 combination of gaps (out of ten thousand) characterized by the highest (99 th percentile) and the lowest (1st percentile) level of errors of inter- and extrapolative forecasts. However, it could not be clearly explained how combinations of gaps affect the level of errors of forecasts. The main goal of this paper is an attempt to solve this problem. In this paper was presented the method of testing the impact of occurrence of unsystematic gaps on the level of theoretical values and forecasts estimated for classic time series models with seasonal fluctuations.

**Słowa kluczowe:** dokładność prognoz, luki niesystematyczne, prognozowanie, układy luk.  
**Key words:** accuracy of forecast, combinations of gaps ,unsystematic gaps, forecasting.

### WSTĘP

W pracy Oesterreich (2015) przedstawiono wyniki zastosowania metod symulacyjnych do badania wpływu rozmieszczenia luk niesystematycznych na poziom błędów prognoz inter- i ekstrapolacyjnych w szeregach czasowych, z wahaniami sezonowymi, o cyklu 12-miesięcznym. Występowanie tego rodzaju luk oznacza, że dostępne są przynajmniej pojedyncze informacje o kształtowaniu się badanego zjawiska w każdym z podokresów (Zawadzki 1999, 2003; Oesterreich 2012). W badaniach tych szczegółowej analizie poddano, spośród 10 tysięcy pięcioelementowych układów luk, 100 układów charakteryzujących się odpowiednio najniższymi (percentyl pierwszy) i najwyższymi (percentyl 99) ocenami błędów prognoz inter- i ekstrapolacyjnych. Jednak w ich trakcie nie udało się jednoznacznie wyjaśnić, w jaki sposób układy luk wpływały na kształtowanie się wielkości błędów obu rodzajów prognoz. Celem niniejszej pracy jest próba rozwiązania tego problemu.

### METODA

Do identyfikacji układów luk charakteryzujących się przeciętnymi błędami prognoz inter- i ekstrapolacyjnych, należących do pierwszego i ostatniego percentyla, można zaproponować metodę polegającą na badaniu wpływu braku pojedynczych obserwacji na dokładność prognoz. Postępowanie związane z jej zastosowaniem obejmuje trzy etapy:

- Etap 1 – oszacowanie modelu dla pełnego szeregu (bez luk) oraz wyznaczenie na jego podstawie wartości teoretycznych w okresie próby ( $\hat{Y}_{teo_t}$ ) oraz prognoz ( $\hat{Y}_{prog_t}$ ) dla ustalonego horyzontu  $h$  ( $h = 1, \dots, r$ ).
- Etap 2 – budowa  $N$  modeli dla szeregów, z których eliminowano kolejno po jednej obserwacji, oraz wyznaczenie na ich podstawie wartości teoretycznych w okresie próby ( $\hat{Y}_{teo_{it}}^*$ ) oraz prognoz ekstrapolacyjnych ( $\hat{Y}_{prog_{it}}^*$ ) ( $i = 1, \dots, N$ ,  $N$  – liczba obserwacji w okresie próby).
- Etap 3 – porównanie wartości teoretycznych dla okresu próby oraz prognoz ekstrapolacyjnych obliczonych na podstawie pełnego szeregu oraz szeregu, z którego wyeliminowano pojedyncze obserwacje. W tym celu można wykorzystać mierniki dokładności wskazujące nie tylko wielkość przeciętnego błędu, ale i jego znak. Miernikami takimi dla wartości teoretycznych i prognoz mogą być na przykład błędy średnie dane wzorami (Gajda 2004; Cieślak 2005):

$$ME_{teo_i} = \frac{1}{N} \sum_{t=1}^N (\hat{Y}_{teo_t} - \hat{Y}_{teo_{it}}^*) \quad (1)$$

$$ME_{prog_i} = \frac{1}{h} \sum_{t=1}^N (\hat{Y}_{prog_t} - \hat{Y}_{prog_{it}}^*) \quad (2)$$

gdzie:

$\hat{Y}_{teo_t}$ ,  $\hat{Y}_{prog_t}$  – wartości teoretyczne i prognozy dla  $t$ -tej obserwacji, obliczone dla modelu oszacowanego na podstawie pełnego szeregu;

$\hat{Y}_{teo_{it}}^*$ ,  $\hat{Y}_{prog_{it}}^*$  – wartości teoretyczne i prognozy dla  $t$ -tej obserwacji, obliczone dla modelu oszacowanego po usunięciu  $i$ -tej obserwacji dla okresu próby.

## CHARAKTERYSTYKA ZMIENNEJ PROGNOZOWANEJ

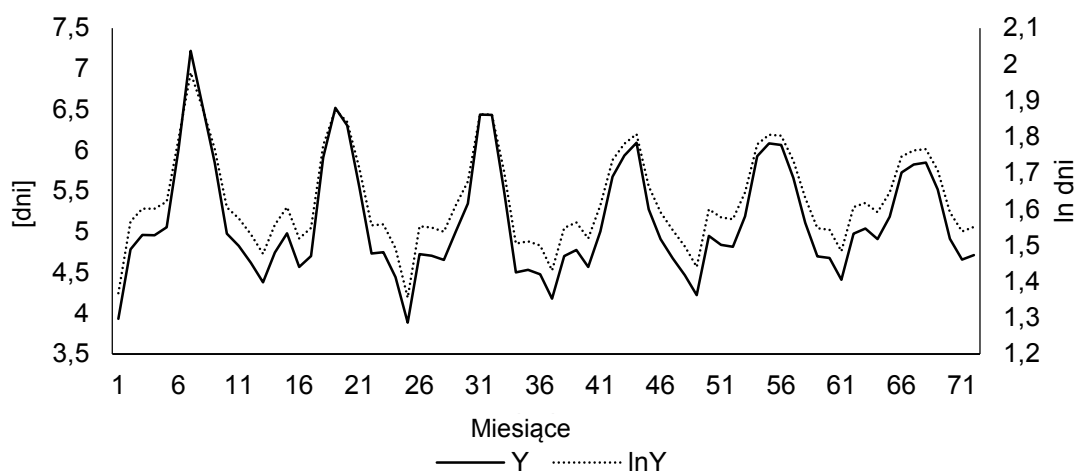
W artykule Oesterreich (2015) modelowaniu i prognozowaniu poddano kształtowanie się średniej długości okresu pobytu turystów w obiektach zbiorowego zakwaterowania w województwie zachodniopomorskim, w latach 2008–2013, według miesięcy, przy czym wartości były ilorazami liczby udzielonych noclegów w obiektach zbiorowego zakwaterowania oraz liczby turystów. Dane statystyczne zaczerpnięto z ukazujących się w cyklu rocznym opracowań GUS (Turystyka 2008, 2009, 2010, 2011, 2012, 2013). Dane z lat 2008–2012 posłużyły do szacowania modeli, natomiast rok 2013 był okresem empirycznej weryfikacji prognoz. Kształtowanie się zmiennej prognozowanej w wielkościach absolutnych (w dniach) oraz jej logarytmów naturalnych przedstawiono na rys. 1.

W celach porównawczych z wynikami zamieszczonymi w artykule Oesterreich (2015) do budowy prognoz inter- i ekstrapolacyjnych wykorzystano klasyczne predyktory oparte na modelach szeregu czasowego z trendem wykładniczym oraz z relatywnie stałą sezonowością opisaną za pomocą wielomianu trygonometrycznego (Zeliaś i in. 2003; Dittman i in. 2009):

$$Y_t = e^{\alpha_0 + \alpha_1 t + \sum_{j=1}^6 (a_{j1} \sin \omega_j t + b_{j1} \cos \omega_j t)} + \varepsilon_t \quad (3)$$

gdzie:

$$\omega_j = \frac{2\pi \cdot j}{m}, \quad (t = 1, 2, \dots, 60; \quad j = 1, 2, \dots, 6; \quad m = 12)$$



Rys. 1. Kształtowanie się długości przeciętnej okresu pobytu turystów w obiektach zbiorowego zakwaterowania w województwie zachodniopomorskim, według miesięcy  
 Źródło: opracowano na podstawie opracowań Głównego Urzędu Statystycznego (2008, 2009, 2010, 2011, 2012, 2013).

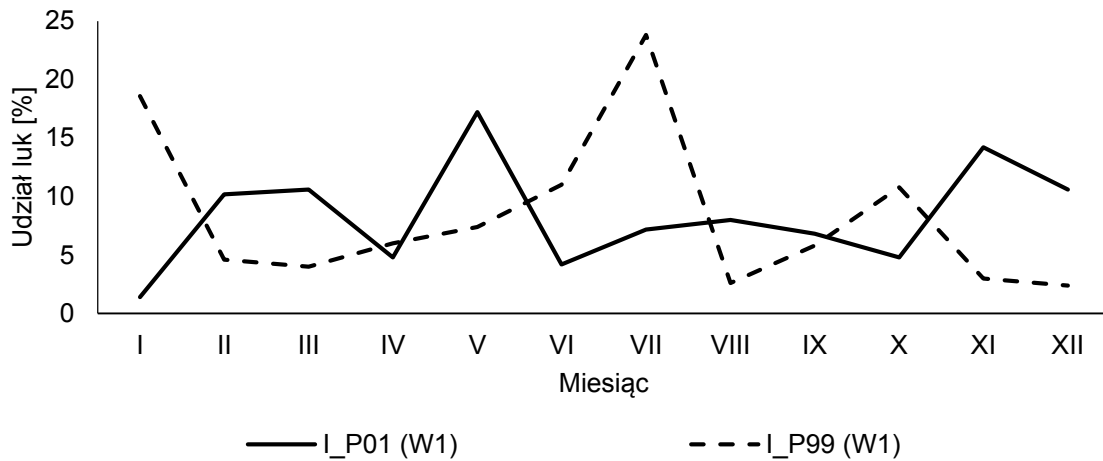
Model ten charakteryzował się lepszymi ocenami parametrów struktury stochastycznej i dobroci dopasowania niż inne modele, w tym model z linowym trendem oraz stałą sezonowością.

Wykorzystując powyższy predyktor, oszacowano jego parametry modeli dla 10 tysięcy losowych kombinacji zawierających po 5 luk oraz zbudowano prognozy interpolacyjne dla podokresów (miesiące), w których występowały luki, oraz prognozy ekstrapolacyjne dla okresu wynoszącego 12 miesięcy. Liczba i rozmieszczenie luk odpowiadały wariantowi W1 przedstawionemu w artykule Oesterreich (2015).

## WYNIKI ANALIZY

Z przedstawionych w artykule Oesterreich (2015) wyników badań symulacyjnych, obejmujących 10 tysięcy pięcioelementowych układów luk (wariant W1), wynikało, że w przypadku 100 losowych układów luk, o najniższych ocenach przeciętnych względnych błędów prognoz interpolacyjnych (pierwszy percentyl –  $I_{P01}$ ), luki najczęściej występowały w: lutym, marcu, maju, listopadzie i grudniu. Ich udział w poszczególnych miesiącach wynosił odpowiednio: 10,2, 10,6, 17,2, 14,2 i 10,6%. Natomiast dla percentyla 99 ( $I_{P99}$ ), a więc 100 kombinacji, dla których otrzymano błędy największe, miesiącami tymi najczęściej były: styczeń, czerwiec, lipiec oraz październik. Dla wymienionych miesięcy odsetki wynosiły odpowiednio: 18,6, 11,0, 23,8 i 10,8%.

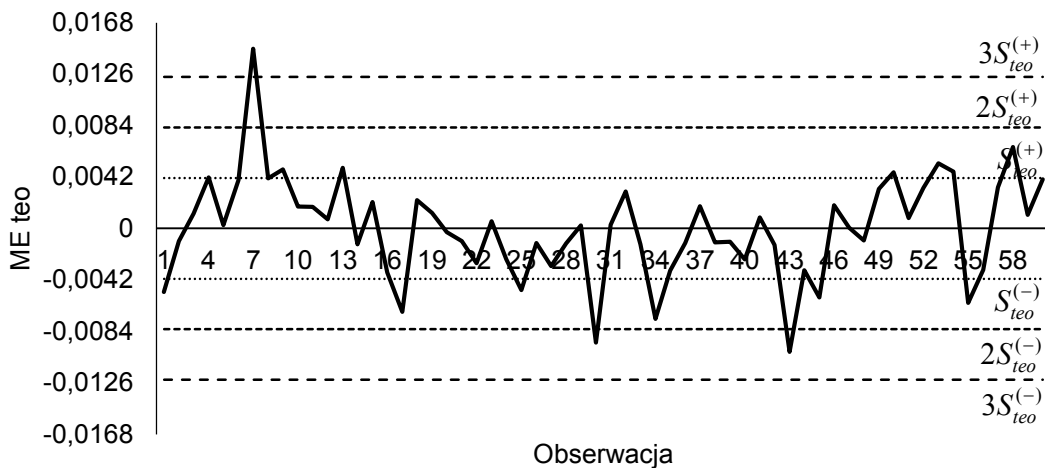
Na rysunku 2 przedstawiono procentowy udział luk w poszczególnych miesiącach dla pierwszego i ostatniego percentyla błędów (MAPE) prognoz interpolacyjnych w wariacie W1.



Rys. 2. Udział luk dla pierwszego i ostatniego percentyla błędów (MAPE) prognoz interpolacyjnych w wariancie W1, według miesięcy

Źródło: opracowano na podstawie Oesterreich (2015).

W celu wyjaśnienia wpływu udziału luk, w poszczególnych miesiącach, na kształtowanie się błędów prognoz inter- i ekstrapolacyjnych dla percentyli I\_P01 i I\_P99 zastosowano metodę opisaną w rozdziale 1. W wyniku jej zastosowania otrzymano szereg składający się z 60 ocen przeciętnych błędów wartości teoretycznych ( $ME_{teo_i}$ ), obliczonych według wzoru (1) w następstwie usunięcia jednej obserwacji z szeregu wyjściowego. Oceny tych błędów zostały przedstawione na rys. 3.



Rys. 3. Przeciętne błędy wartości teoretycznych w przedziale czasowym próby w następstwie usunięcia  $t$ -tej obserwacji z szeregu wyjściowego

Linie poziome na rys. 3 oznaczają przeciętne odchylenia absolutne ( $ME_{teo_i}$ ) wynoszące odpowiednio *in plus* lub *in minus*: jedno, dwa lub trzy odchylenia standardowe ( $S_{teo}^{(+)} / S_{teo}^{(-)}$ ). Z wykresu wynika, że w następstwie usunięcia obserwacji nr 7 (lipiec) otrzymano błąd *in plus*

przekraczający trzykrotną wartość odchylenia standardowego błędu średniego  $ME_{teo_i}$ . Przeciętne błędy dla obserwacji nr 30 (czerwiec) i nr 43 (lipiec) były większe *in minus* od dwóch odchyłeń standardowych. Informacje o liczbie przypadków, w których wartości  $ME_{teo_i}$  zawierały się w przedziałach  $(2S_{teo}^{(-)}; S_{teo}^{(-)})$  oraz  $(S_{teo}^{(+)}; 2S_{teo}^{(+)})$ , przedstawiono w tab. 1.

Tabela 1. Liczba przeciętnych błędów wartości teoretycznych przekraczających *in plus* lub *in minus* wartość jednego odchylenia standardowego, według miesięcy

	Miesiąc											
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
$S_{teo}^{(+)}$	2	0	0	0	1	0	1	0	1	1	0	0
$S_{teo}^{(-)}$	1	1	0	1	1	1	0	0	1	1	0	0
Razem	3	1	0	1	2	1	1	0	2	2	0	0

Z informacji zawartych w tab. 1 wynika, że odległość jednego odchylenia standardowego została przekroczona trzykrotnie w styczniu (obserwacje nr 1, 13, 25). Dwukrotnie granica ta została przekroczona w maju (obserwacje nr 17 i nr 53), we wrześniu (obserwacje nr 9 i nr 45) oraz w październiku (obserwacje nr 34 i nr 58). Natomiast w lutym (obserwacja nr 50), kwietniu (obserwacja nr 4), czerwcu (obserwacja nr 54) oraz lipcu (obserwacja nr 55) notowano pojedyncze przekroczenia wartości  $S_{teo}$ .

Z porównania informacji z rys. 2 i rys. 3 oraz z tab. 1 wynika, że następstwem usunięcia z szeregu obserwacji, odpowiadających: styczniowi, czerwcowi oraz lipcowi, było wystąpienie dużych przeciętnych błędów wartości teoretycznych. Były to miesiące, w których wystąpiło minimum sezonowe (styczeń) bądź maksimum sezonowe (lipiec). Charakteryzowały się one także największym udziałem luk w kombinacjach w ostatnim percentylu błędów prognoz interpolacyjnych (I\_P99 – rys. 2).

Natomiast dla luk występujących w: lutym, marcu, maju, listopadzie i grudniu otrzymano wartości bezwzględne  $ME_{teo_i}$  mniejsze od jednego odchylenia standardowego. Miesiące te charakteryzowały się największym udziałem liczby luk w kombinacjach w pierwszym percentylu błędów prognoz interpolacyjnych (I\_P01 – rys. 2). Obserwacje w tych miesiącach były również względnie „neutralne”, tzn. nie występowały w nich minima czy maksima sezonowe.

Kolejnym etapem było porównanie charakterystyk opisowych oszacowanych modeli z 5-elementowymi układami luk utworzonymi z obserwacji spełniających nierówność  $|ME_{teo_i}| > S_{teo}$  oraz z błędów prognoz interpolacyjnych otrzymanych na ich podstawie.

Porównanie zostało przeprowadzone w następujących układach:

- I\_  $S^{(+)}$  dla  $ME_{teo_i} > S_{teo}^{(+)}$ ,
- I\_  $S^{(-)}$  dla  $ME_{teo_i} < S_{teo}^{(-)}$ ,
- I\_  $S^{(-/+)}$  dla  $|ME_{teo_i}| > S_{teo}$ .

Liczba luk w kombinacji została wybrana w sposób dający możliwość porównania wyników z wynikami zamieszczonymi w artykule Oesterreich (2015). Do powyższego porównania zostały dołączone informacje na temat charakterystyk modeli oraz przeciętnych błędów (MAPE) w pierwszym (I\_P01) i ostatnim (I\_P99) percentylu błędów prognoz interpolacyjnych.

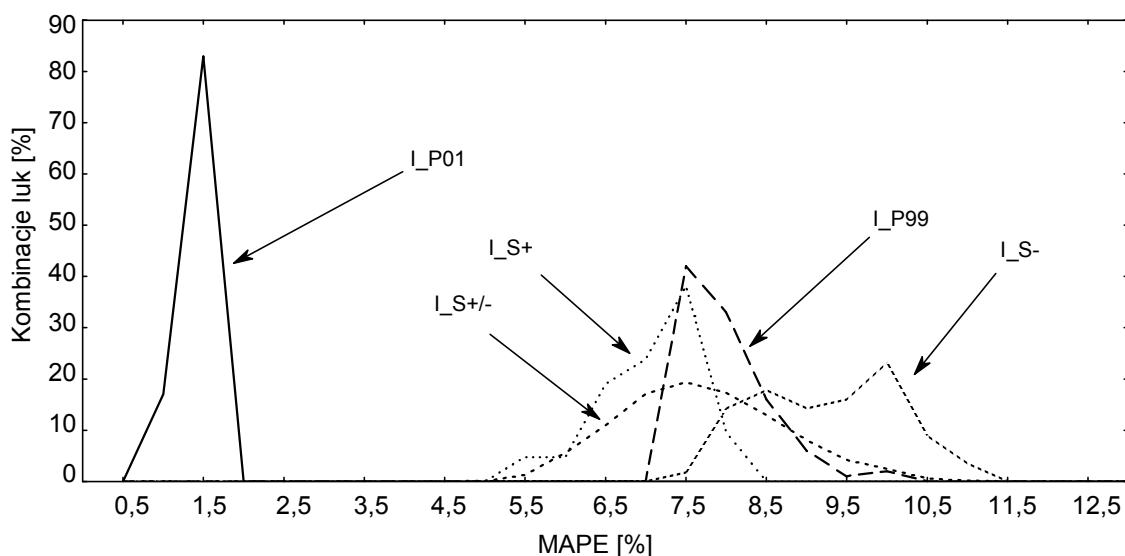
W tabeli 2 zamieszczono informacje o numerach usuniętych obserwacji, liczbie odpowiadającym im 5-elementowym kombinacji oraz przeciętnych wartościach: współczynników determinacji ( $\overline{R^2}$ ), odchylenia standardowego składnika losowego ( $\overline{SE}$ ), a także przeciętnych błędach ( $\overline{MAPE}$ ) prognoz interpolacyjnych.

Tabela 2. Liczba układów luk oraz przeciętne wartości współczynników determinacji, odchyłeń standardowych składników losowych oraz błędów prognoz interpolacyjnych

Grupy	Numery obserwacji	Liczba układów luk	$\overline{R^2}$	$\overline{SE}$	Przeciętne błędy prognoz interpolacyjnych
					$\overline{MAPE}$
Szereg pełny	–	–	0,929	0,041	–
I_S <sup>(+)</sup>	7, 9, 13, 50, 53, 54, 58	21	0,943	0,036	6,879
I_S <sup>(-)</sup>	1, 17, 25, 30, 34, 43, 45, 55	56	0,946	0,034	9,019
I_S <sup>(+/-)</sup>	1, 7, 9, 13, 17, 25, 30, 34, 43, 45, 50, 53, 54, 55, 58	3003	0,942	0,035	7,448
I_P01	–	100	0,925	0,042	1,159
I_P99	–	100	0,939	0,035	7,749

Źródło: opracowano podstawie Oesterreich (2015).

Na rysunku 4 przedstawiano kształtowanie się rozkładów błędów prognoz interpolacyjnych, otrzymanych w toku badań symulacyjnych dla poszczególnych wyodrębnionych grup obserwacji (I\_S<sup>(-)</sup>; I\_S<sup>(+)</sup>; I\_S<sup>(+/-)</sup>), na tle rozkładów dla pierwszego (I\_P01) i ostatniego (I\_P99) percentyla.



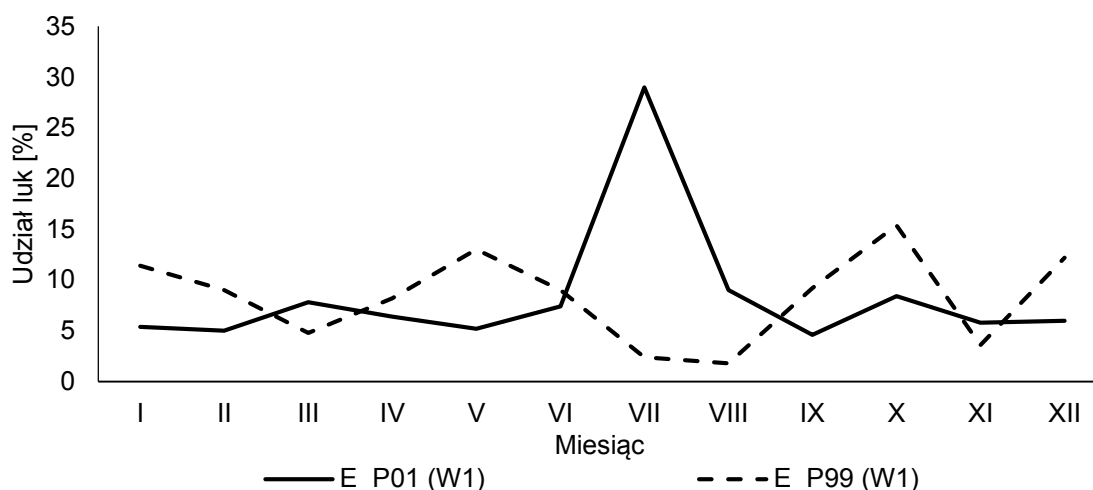
Rys. 4. Rozkład błędów (MAPE) prognoz interpolacyjnych dla grup I\_S<sup>(+)</sup>, I\_S<sup>(-)</sup>, I\_S<sup>(+/-)</sup> oraz pierwszego i ostatniego percentyla

Źródło: opracowano na podstawie Oesterreich (2015).

Z informacji zawartych w tab. 2 oraz na rys. 5 wynika, że usunięcie 5-elementowych obserwacji, wyraźnie wpływających na wartości teoretyczne, poprawiło dopasowanie modeli i wpłynęło na przeciętne oceny odchylenia standardowego składnika losowego.

Duże błędy prognoz interpolacyjnych uzyskano dla modeli, z których usunięto układy luk zawierające obserwacje należące odpowiednio do grup:  $I\_S^{(-)}$ ;  $I\_S^{(+)}$ ;  $I\_S^{(-+)}$ . Usunięcie obserwacji o numerach podanych w tab. 2, tzn. o wartościach znacznie większych lub mniejszych od obserwacji pozostałych, wpływało na poprawę charakterystyk opisowych szacowanych modelu. Wartości teoretyczne wyznaczone dla obserwacji usuniętych, będących prognozami interpolacyjnymi, będą znacznie odbiegać od wartości rzeczywistych. Dlatego rozkłady błędów prognoz interpolacyjnych, wyznaczonych dla kombinacji luk dla grup:  $I\_S^{(-)}$ ;  $I\_S^{(+)}$ ;  $I\_S^{(-+)}$ , przedstawione graficznie na rys. 4, znajdują się w tej samej jego części co dla ostatniego percentyla błędów ( $I\_P99$ ).

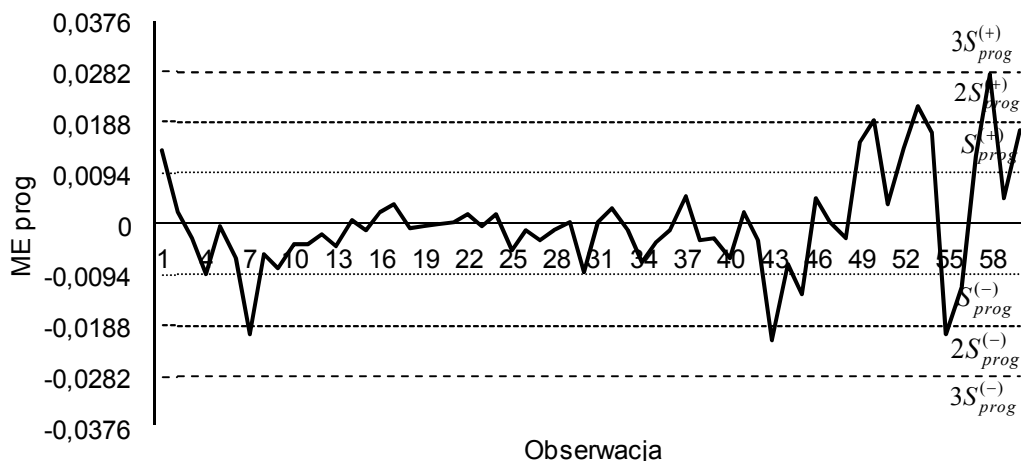
Na rysunku 5 przedstawiano udział luk w poszczególnych miesiącach dla pierwszego ( $E\_P01$ ) oraz ostatniego ( $E\_P99$ ) percentyla błędów prognoz ekstrapolacyjnych w wariancie W1 (por. Oesterreich 2015).



Rys. 5. Udział luk w miesiącach dla pierwszego i ostatniego percentyla błędów (MAPE) prognoz ekstrapolacyjnych w wariancie W1  
Źródło: opracowano na podstawie Oesterreich (2015).

Z powyższego rysunku wynika, że najniższe oceny błędów prognoz ekstrapolacyjnych otrzymano dla kombinacji luk, w których występowały one przede wszystkim w lipcu. Natomiast najwyższe oceny błędów prognoz były wynikiem usunięcia z szeregu obserwacji przypadających na styczeń, maj, październik oraz grudzień.

Na rysunku 6 przedstawiano informacje nt. przeciętnych błędów prognoz ekstrapolacyjnych ( $ME_{porg_i}$ ), wyznaczonych dla 12 miesięcy, na podstawie szeregu, z którego usuwano kolejne obserwacje (przedział czasowy „próby”) – zob. wzór 2.



Rys. 6. Przeciętne błędy prognoz ekstrapolacyjnych w okresie prognozowanym w następstwie usunięcia  $t$ -tej obserwacji z szeregu wyjściowego

Podobnie jak w na rys. 3 poziome linie oznaczają odchylenia absolutne  $ME_{prog_i}$ , wynoszące odpowiednio:  $\pm$  jedno odchylenie standardowe, dwa i trzy odchylenia standardowe ( $S_{prog}^{(+)}$  /  $S_{prog}^{(-)}$ ). Z porównania średnich oraz odchyłeń standardowych dla szeregów zmian wartości teoretycznych ( $ME_{teo_i}$ ) i wartości prognoz ( $ME_{prog_i}$ ) wynika, że wartości obu parametrów opisowych są nieznacznie większe dla drugiego z szeregów.

Z analizy rys. 7 wynika, że granica  $2S_{prog}^{(+)}$  została przekroczona dla obserwacji nr 50 (luty), nr 53 (maj) oraz nr 58 (październik). Natomiast wartości mniejsze niż  $-2S_{prog}^{(-)}$  otrzymano dla obserwacji nr 7 (lipiec), nr 43 (lipiec) oraz nr 55 (lipiec). Informacje o liczbie przypadków, gdy wartości  $ME_{prog_i}$  zawierały się w przedziale  $(2S_{prog}^{(-)}, S_{prog}^{(-)})$  oraz  $(S_{teo}^{(+)}, 2S_{teo}^{(+)})$ , przedstawiono w tab. 3.

Tabela 3. Liczba zmian wartości teoretycznych przekraczających wartość  $\pm$  jednego odchylenia standardowego od średniej (okres prognozowany), według miesięcy

	Miesiąc											
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
$S_{prog}^{(+)}$	2	0	0	1	0	1	0	0	1	0	0	1
$S_{prog}^{(-)}$	0	0	0	0	0	0	0	1	1	0	0	0
Razem	2	0	0	1	0	1	0	2	2	0	0	1

Z tabeli 3 wynika, że najczęściej (sześciokrotnie) przekroczono granicę  $S_{prog}^{(+)}$ . Dotyczyło to obserwacji nr 1 (styczeń), nr 49 (styczeń), nr 52 (kwiecień), nr 54 (czerwiec), nr 57 (wrzesień) oraz nr 60 (grudzień). Wartości zmian mniejsze niż  $S_{prog}^{(-)}$  zanotowano tylko w przypadku dwóch obserwacji – nr 45 (wrzesień) oraz nr 56 (sierpień).

Z porównania informacji z rys. 5 i rys. 6 oraz z tab. 2 wynika, że usunięte obserwacje, które zwiększały wartości teoretyczne w okresie prognozowanym, odpowiadają: styczniowi,



lutemu, kwietniowi, majowi, czerwcowi, wrześniowi, październikowi oraz grudniowi. Odpowiada to w pełni miesiącom o największym udziale luk w kombinacjach dla ostatniego percentyla błędów (rys. 5 – E\_P99).

W przypadku obserwacji, których usunięcie z szeregu powodowało zmniejszenie wartości teoretycznych w okresie prognozowanym, przeważają te, które odpowiadały lipcowi. Odchylenia *in minus* zanotowano również dla sierpnia i września. Miesiące te prawie w pełni odpowiadają miesiącom o największym udziale luk w kombinacjach dla pierwszego percentyla błędów prognoz ekstrapolacyjnych (rys. 5 – E\_P01).

Ostatnim etapem analizy było porównanie charakterystyk opisowych oszacowanych modeli z 5-elementowymi układami luk, utworzonymi z obserwacji spełniających nierówność

$|ME_{prog_i}| > S_{prog}$  w następujących układach:

–  $E\_S^{(+)}$  dla  $ME_{prog_i} > S_{prog}^{(+)}$ ,

–  $E\_S^{(-)}$  dla  $ME_{prog_i} < S_{prog}^{(-)}$ ,

–  $E\_S^{(-+)}$  dla  $|ME_{prog_i}| > S_{prog}$

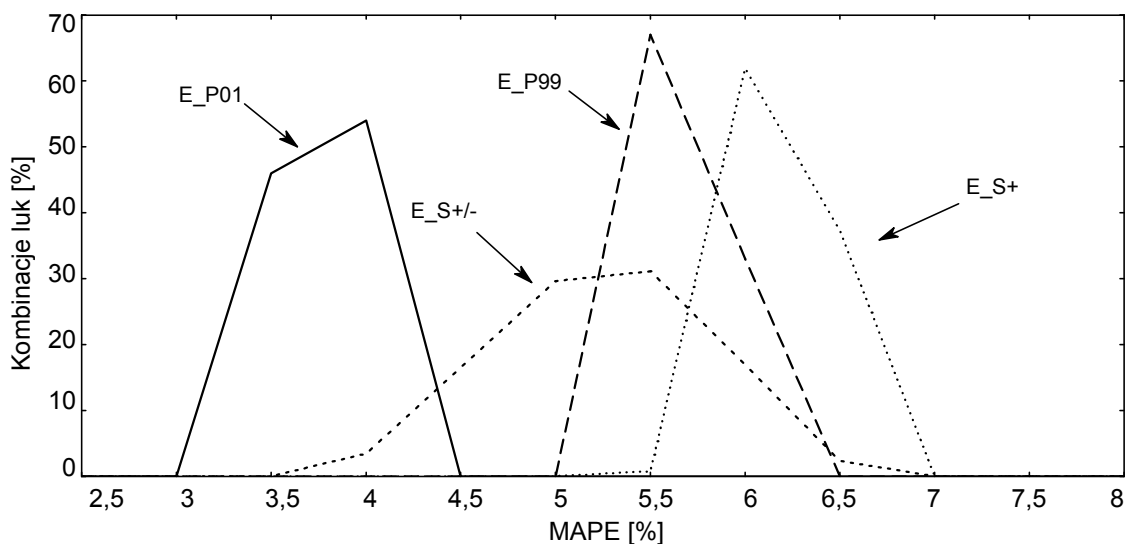
oraz z błędów prognoz ekstrapolacyjnych otrzymanych na ich podstawie. Do powyższego porównania dołączono informacje na temat charakterystyk modeli oraz rozkładów błędów ekstrapolacyjnych w pierwszym (E\_P01) i ostatnim (E\_P99) percentylu. Powyższe informacje przedstawiono w tab. 4.

Tabela 4. Liczba układów luk oraz przeciętne wartości współczynników determinacji, odchyłeń standardowych składników losowych oraz prognoz inter- i ekstrapolacyjnych (okres prognozowany)

Grupy	Numery obserwacji	Liczba układów luk	$\bar{R}^2$	$\overline{SE}$	Prognozy ekstrapolacyjne
					$\overline{MAPE}$
Szereg pełny	–	–	0,929	0,041	4,561
$E\_S^{(+)}$	1, 49, 50, 52, 53, 54, 57, 58, 60	126	0,937	0,038	5,937
$E\_S^{(-)}$	7, 43, 45, 55, 56	1	0,937	0,035	4,352
$E\_S^{(-+)}$	1, 7, 43, 45, 49, 50, 52, 53, 54, 55, 56, 57, 58, 60	2002	0,937	0,037	4,991
E_P01	–	100	0,933	0,036	3,493
E_P99	–	100	0,935	0,039	5,464

Na rysunku 7 przedstawiano kształtowanie się rozkładów błędów prognoz ekstrapolacyjnych w poszczególnych wyodrębnionych grupach obserwacji ( $E\_S^{(+)}$ ;  $E\_S^{(-)}$ ;  $E\_S^{(-+)}$ ) na tle rozkładów dla pierwszego (E\_P01) i ostatniego (E\_P99) percentyla.

Z analizy informacji zawartych w tab. 4 oraz na rys. 7 wynika, że najmniejsze błędy prognoz ekstrapolacyjnych uzyskano w wyniku usunięcia obserwacji zmniejszających wartości prognozowane, tzn. spełniające nierówność  $ME_{prog_i} < S_{prog}^{(-)}$ . Uzyskany błąd prognoz był mniejszy o 0,209 p.p. (ok. 4,5%) w pełnym szeregu. Oznacza to, że prognozy zbudowane dla szeregu pełnego były zawyżone. Spostrzeżenia potwierdzają rozkład błędów prognoz ekstrapolacyjnych dla grupy  $E\_S^{(+)}$  i wyznaczona dla niego wartość przeciętna.



Rys. 7. Rozkłady błędów prognoz ekstrapolacyjnych dla grup  $E_{S^{(+)}}$ ,  $E_{S^{(-)}}$ ,  $E_{S^{(+/+)}}$  oraz dla pierwszego i ostatniego percentyla

Źródło: opracowano na podstawie Oesterreich (2015).

## PODSUMOWANIE

1. Oceny przeciętnych błędów  $ME_{teo_i}$  oraz  $ME_{prog_i}$  wskazują, że oddziaływanie rozmieszczenia luk na kształtowanie się wartości teoretycznych w okresie estymacyjnym oraz na prognozy *ex post* było odmienne. W pierwszym przypadku miały na to wpływ obserwacje w całym przedziale czasowym próby, a w drugim przypadku – obserwacje tylko w początkowym i końcowym fragmencie.
2. Wyższy stopień dopasowania modeli do danych rzeczywistych oraz niższe przeciętne oceny odchylenia standardowego składnika losowego otrzymano poprzez usunięcie z szeregu czasowego obserwacji z grupy  $I_{S^{(-)}}$ . Modele zbudowane na podstawie tak zmodyfikowanego szeregu charakteryzowały się jednak dużymi błędami prognoz interpolacyjnych.
3. Niższe oceny błędów prognoz ekstrapolacyjnych uzyskano po wyeliminowaniu pięciu obserwacji o numerach: 7, 43, 45, 55, 56 – o wartościach  $ME_{prog_i}$  przekraczających *in minus* jedno odchylenie standardowe ( $E_{S^{(-)}}$ ). Ocena błędu prognoz była o ok. 4,5% niższa od błędu prognoz wyznaczonego na podstawie szeregu pełnego.
4. Zaproponowana w pracy metoda może być uogólniona na przypadek większej liczby luk.

## PIŚMIENNICTWO

- Cieślak M. 2005. Prognozowanie gospodarcze. Metody i zastosowania. Warszawa, PWN, 50–53.
- Dittmann P., Dittmann I., Szabela-Pasierbińska E., Szpulak A. 2009. Prognozowanie w zarządzaniu przedsiębiorstwem. Kraków, Wolters Kluwer Polska.
- Gajda J.B. 2004. Ekonometria. Warszawa, Wydaw. C.H. Beck.

- Główny Urząd Statystyczny.** 2008. Turystyka w 2008. Informacje i opracowania statystyczne. Warszawa, GUS.
- Główny Urząd Statystyczny.** 2009. Turystyka w 2009. Informacje i opracowania statystyczne. Warszawa, GUS.
- Główny Urząd Statystyczny.** 2010. Turystyka w 2010. Informacje i opracowania statystyczne. Warszawa, GUS.
- Główny Urząd Statystyczny.** 2011. Turystyka w 2011. Informacje i opracowania statystyczne. Warszawa, GUS.
- Główny Urząd Statystyczny.** 2012. Turystyka w 2012. Informacje i opracowania statystyczne. Warszawa, GUS.
- Główny Urząd Statystyczny.** 2013. Turystyka w 2013. Informacje i opracowania statystyczne. Warszawa, GUS.
- Oesterreich M.** 2012. Symulacyjne badanie wpływu częstości występowania luk niesystematycznych w szeregach czasowych na dokładność prognoz. Wrocław, Wydaw. UE Wroc.
- Oesterreich M.** 2015. Symulacyjna analiza wpływu liczby i rozmieszczenia luk niesystematycznych na dokładność prognoz. Wrocław, Wydaw. UE Wroc. (w druku).
- Zawadzki J.** 1999. Ekonometryczne metody predykcji dla danych sezonowych w warunkach braku pełnej informacji. Szczecin, Wydaw. USzczec.
- Zawadzki J.** 2003. Zastosowanie hierarchicznych modeli szeregów czasowych w prognozowaniu zmiennych ekonomicznych z wahaniami sezonowymi. Szczecin, Wydaw. AR w Szczecinie.
- Zeliaś A., Pawełek B., Wanat S.** 2003. Prognozowanie ekonomiczne. Teoria, przykłady, zadania. Warszawa, PWN, 89.

