

**Maciej Oesterreich**

Zachodniopomorski Uniwersytet Technologiczny w Szczecinie

---

## **SYMULACYJNE BADANIE WPLYWU CZĘSTOŚCI WYSTĘPOWANIA LUK NIESYSTEMATYCZNYCH W SZEREGACH CZASOWYCH NA DOKŁADNOŚĆ PROGNOZ**

---

**Streszczenie:** W pracy przedstawione zostały wyniki wykorzystania metod symulacyjnych do badania wpływu częstości występowania luk niesystematycznych na dokładność prognoz inter- oraz ekstrapolacyjnych w szeregu czasowym z wahaniami sezonowymi o cyklu rocznym (12-miesięcznym). Analizie poddano kształtowanie się produkcji energii elektrycznej w Polsce w latach 2002-2011. Do budowy prognoz wykorzystano modele szeregu czasowego z periodycznym składnikiem sezonowym, w którym sezonowość była opisywana za pomocą zmiennej zero-jedynkowej lub wielomianu trygonometrycznego, a także modele hierarchiczne. Obliczenia zostały wykonane z wykorzystaniem pakietu R oraz Statistica 9.0.

**Słowa kluczowe:** symulacje, luki niesystematyczne, prognozowanie.

### **1. Wstęp**

W dotychczasowych badaniach nad dokładnością prognozowania w warunkach niepełnej informacji na podstawie ekonomicznych szeregów czasowych wykazujących wahania sezonowe (por. [Markiewicz, Zawadzki 2003; Szmuksta-Zawadzka, Zawadzki 1995, 2002; Zawadzki 1996; 1999; 2003]) zazwyczaj ograniczano się do kilku wariantów luk w danych, różniących się liczbą i rozmieszczeniem oraz natężeniem sezonowości. Dlatego też uogólniające wnioski wynikające z tych badań formułowane były w sposób dość ostrożny. Zastosowanie metod symulacyjnych w prognozowaniu brakujących informacji daje lepsze podstawy wyprowadzenia tego rodzaju wniosków, ponieważ analizuje się dużą liczbę losowych wariantów luk dla zmieniających się założeń dotyczących częstości ich występowania. Hipotezę badawczą, która zostanie poddana weryfikacji, sformułować można następująco: zastosowanie metod symulacyjnych do prognozowania szeregów w warunkach braku pełnej informacji daje lepsze podstawy wyprowadzenia wniosków uogólniających dotyczących wpływu częstości występowania luk na dokładność prognoz.

## 2. Metodyka badań

Na potrzeby pracy zbudowane zostały w środowisku R dwa algorytmy. Zadaniem pierwszego z nich było generowanie losowych luk niesystematycznych. Algorytm ten składał się z następujących etapów:

1. Generowanie losowego ciągu  $p$  liczb całkowitych zawierających się w przedziale  $(1; n - r)$ , gdzie  $n$  – długość analizowanego szeregu czasowego,  $r$  – horyzont prognozy. Liczby wchodzące w skład ciągu określają miejsca w badanym szeregu czasowym, w których znajdować się będą luki.

2. Sprawdzenie, czy wygenerowane luki nie mają charakteru luk systematycznych, tzn. występują w tym samym podokresie (miejscu) w całym szeregu czasowym.

Zadaniem drugiego algorytmu było wprowadzenie wszystkich wygenerowanych wariantów luk do badanej zmiennej, szacowanie parametrów i budowa prognoz oraz obliczanie średnich względnych błędów prognoz (MAPE) inter- oraz ekstrapolacyjnych.

W prognozowaniu w warunkach braku pełnych informacji zostały wykorzystane klasyczne modele szeregu czasowego z sezonowością periodyczną opisane za pomocą [Zawadzki 1999]:

- zmiennych zero-jedynkowych (0/1),
- wielomianu trygonometrycznego (w.tryg),

a także dwu- oraz trzystopniowe regularne modele hierarchicznych dla danych miesięcznych [Zawadzki 2003].

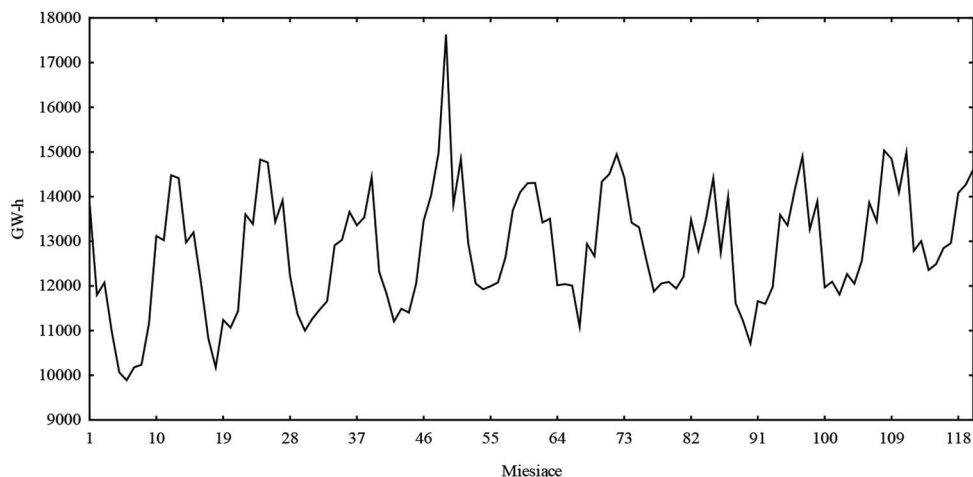
## 3. Weryfikacja hipotezy badawczej

Korzystając z algorytmu opisanego we wstępie, dla każdego z sześciu wariantów różniących się odsetkiem brakujących danych wygenerowano po 2000 zestawów luk, a więc także prognoz inter- i ekstrapolacyjnych. Udziały luk w danych w poszczególnych wariantach wynosiły:

- W1 – 5 brakujących ze 108 obserwacji  $\approx 5\%$ ,
- W2 – 11 brakujących ze 108 obserwacji  $\approx 10\%$ ,
- W3 – 16 brakujących ze 108 obserwacji  $\approx 15\%$ ,
- W4 – 22 brakujących ze 108 obserwacji  $\approx 20\%$ ,
- W5 – 27 brakujących ze 108 obserwacji  $\approx 25\%$ ,
- W6 – 32 brakujących ze 108 obserwacji  $\approx 30\%$ .

Luki otrzymano przez „wymazanie” określonej dla każdego w wariantu liczby obserwacji.

Analizę dokładności prognozowania inter- i ekstrapolacyjnego przeprowadzono na przykładzie kształtowania się produkcji energii elektrycznej w Polsce w latach 2002-2011. Dane pochodzą z Biuletynów Statystycznych GUS. Dane z lat 2002-2010 wchodziły w skład okresu estymacyjnego, natomiast rok 2011 posłużył do empirycznej weryfikacji prognoz. Kształtowanie się badanej zmiennej przedstawiono na rys. 1.



**Rys. 1.** Produkcja energii elektrycznej w Polsce w latach 2002-2011

Źródło: Biuletyny Statystyczne GUS.

W tabeli 1 przedstawiono kształtowanie się przeciętnych błędów prognoz interpolacyjnych według udziału luk w danych. Błędy te obliczono na podstawie rozkładów średnich względnych błędów prognoz interpolacyjnych (MAPE).

Z analizy tab. 1 wynika, że najwyższe przeciętne oceny współczynników determinacji kształtujące się na poziomie ok. 0,80 otrzymano dla modeli klasycznych. Dla czterech modeli hierarchicznych (TH34, TH43, TH62, TH322) przyjęły one wartości z przedziału 0,511 do 0,725. Zdecydowanie najniższe oceny współczynników determinacji nieprzekraczające 0,1 otrzymano dla trzech pozostałych modeli hierarchicznych: TH26, TH232, TH223. Wyjaśniły one 5-8% wariacji zmiennej objaśnianej.

Najmniej obciążone prognozy interpolacyjne uzyskano za pomocą modelu, w którym sezonowości opisano za pomocą wielomianu trygonometrycznego. Tylko nieznacznie gorsze okazały się prognozy zbudowane za pomocą klasycznego modelu ze zmienną zero-jedynkową. Przeciętne błędy interpolacyjne zawarły się w ich przypadku w przedziale od 4,09 do 4,21 punktu procentowego. Najwyższe oceny błędów prognoz interpolacyjnych uzyskano dla modelu hierarchicznego TH26, TH223 oraz TH223. Ich przeciętne oceny błędów kształtowały się od ok. 9,2 do 9,6%. Oznacza to, że w dalszych analizach te trzy modele zostały pominięte.

Największy przyrost przeciętnych błędów interpolacyjnych między pięćdziesiątą a trzydziestoprocentowym udziałem luk w danych zanotowano dla klasycznego modelu ze zmienną zero-jedynkową. Wyniósł on prawie 0,2 punktu procentowego. Należy jednak zauważyć, że przyrosty w przypadku wszystkich analizowanych metod były stosunkowo niewielkie.

**Tabela 1.** Kształtowanie się przeciętnych błędów interpolacyjnych (w %) oraz przeciętnego współczynnika determinacji według metod oraz wariantów udziału luk w danych

Model		Wariant udziału luk					
		W1	W2	W3	W4	W5	W6
w.tryg	przeciętne R <sup>2</sup>	0,800	0,802	0,804	0,806	0,808	0,810
	przeciętny błąd	4,098	4,103	4,127	4,108	4,143	4,147
0/1	przeciętne R <sup>2</sup>	0,804	0,806	0,808	0,810	0,812	0,814
	przeciętny błąd	4,091	4,109	4,159	4,152	4,196	4,209
TH26	przeciętne R <sup>2</sup>	0,077	0,082	0,085	0,091	0,097	0,102
	przeciętny błąd	9,512	9,476	9,518	9,523	9,586	9,557
TH34	przeciętne R <sup>2</sup>	0,725	0,726	0,727	0,729	0,730	0,732
	przeciętny błąd	4,488	4,502	4,542	4,531	4,577	4,580
TH43	przeciętne R <sup>2</sup>	0,511	0,513	0,514	0,519	0,519	0,522
	przeciętny błąd	6,067	5,928	5,987	6,002	6,008	6,033
TH62	przeciętne R <sup>2</sup>	0,622	0,623	0,625	0,628	0,629	0,632
	przeciętny błąd	5,553	5,462	5,522	5,507	5,542	5,556
TH223	przeciętne R <sup>2</sup>	0,074	0,077	0,079	0,084	0,087	0,090
	przeciętny błąd	9,318	9,262	9,305	9,315	9,371	9,345
TH232	przeciętne R <sup>2</sup>	0,054	0,057	0,060	0,064	0,067	0,071
	przeciętny błąd	9,373	9,347	9,380	9,353	9,408	9,393
TH322	przeciętne R <sup>2</sup>	0,511	0,512	0,513	0,517	0,517	0,519
	przeciętny błąd	6,006	5,868	5,926	5,937	5,941	5,962

Źródło: opracowanie własne.

W przypadku modeli hierarchicznych zanotowano spadki przeciętnego poziomu błędów prognoz interpolacyjnych. Zjawisko to prawdopodobnie było spowodowane tym, że obserwacje, na które przypadają luki, miały charakter obserwacji odstających. „Wymazanie” z analizowanej zmiennej tych informacji umożliwiło budowę lepszych jakościowo predyktorów. Potwierdza to wartość przeciętnego współczynnika determinacji obliczonego dla poszczególnych metod, który rośnie wraz ze wzrostem udziałów luk w danych (por. tab. 1). Możliwe jest również, że wyniki uległyby zmianie, gdyby przeanalizowano większą liczbę kombinacji luk. W przypadku modeli hierarchicznych fluktuacje przeciętnego poziomu błędów można tłumaczyć także jako niedostosowanie struktury harmonicznej predyktora do struktury harmonicznej wyjaśnianej zmiennej.

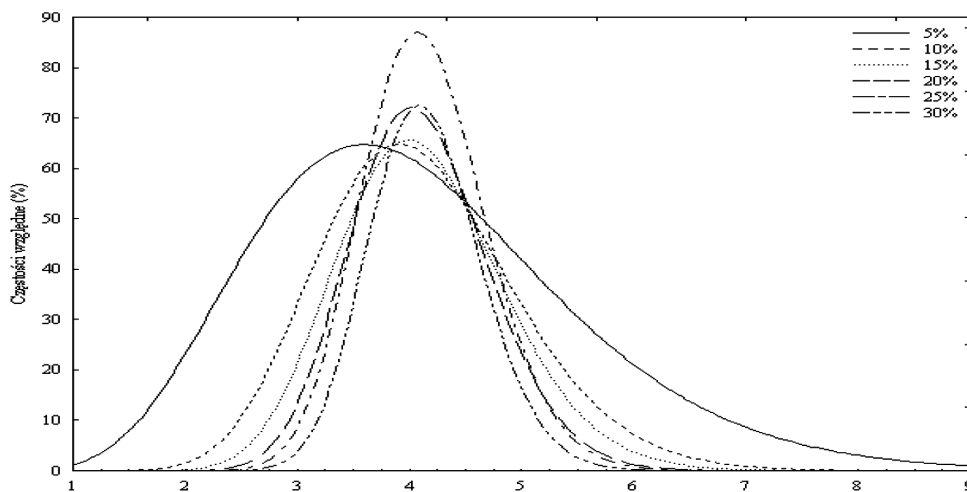
W tabeli 2 przedstawiono kształtowanie się pozycyjnych miar tendencji centralnej uzupełnionych o informacje na temat odchylenia standardowego, współczynnika zmienności, wartości minimalnych i maksymalnych, rozstępu oraz współczynnika skośności obliczonych dla modeli, w których sezonowość opisana była za pomocą zmiennej zero-jedynkowej, wielomianu trygonometrycznego oraz modeli hierarchicznych TH34, TH43, TH62, TH322. Modele te wybrano ze względu na zbliżoną wartość przeciętnych błędów interpolacyjnych oraz wartości przeciętnych współczynników determinacji (por. tab. 1).

**Tabela 2.** Wybrane statystyki opisowe empirycznych rozkładów średnich względnych błędów prognoz interpolacyjnych dla wybranych metod oraz wariantów udziałów luk

Metoda	Wariant	Odch.std	Wsp.zmn. (w %)	Skośność	Q1	Mediana	Q3	Minimum	Maksimum	Rozstęp
w.tryg	W1	1,403	34,229	0,569	3,056	4,010	4,966	0,441	9,807	9,365
	W2	0,879	21,414	0,388	3,475	4,039	4,654	1,850	7,424	5,574
	W3	0,736	17,841	0,246	3,619	4,103	4,610	2,171	6,700	4,529
	W4	0,615	14,964	0,266	3,674	4,073	4,492	2,137	6,364	4,228
	W5	0,554	13,383	0,260	3,758	4,135	4,508	2,558	7,084	4,526
	W6	0,499	12,033	0,227	3,791	4,135	4,473	2,589	5,977	3,388
0/1	W1	1,416	34,612	0,628	3,025	3,959	4,976	0,712	10,182	9,470
	W2	0,892	21,705	0,357	3,499	4,045	4,679	1,696	7,365	5,668
	W3	0,740	17,781	0,230	3,653	4,136	4,636	2,110	6,706	4,596
	W4	0,621	14,954	0,257	3,710	4,128	4,536	2,196	6,427	4,231
	W5	0,556	13,241	0,272	3,805	4,184	4,562	2,721	7,058	4,337
	W6	0,501	11,906	0,199	3,851	4,205	4,551	2,903	6,040	3,137
TH34	W1	1,623	36,172	0,785	3,291	4,273	5,438	0,664	12,532	11,868
	W2	1,020	22,666	0,417	3,762	4,409	5,166	1,806	8,300	6,493
	W3	0,846	18,635	0,364	3,937	4,507	5,084	2,271	7,922	5,651
	W4	0,710	15,666	0,189	4,045	4,493	5,010	2,184	7,415	5,231
	W5	0,622	13,594	0,135	4,149	4,552	4,987	2,574	6,885	4,311
	W6	0,561	12,238	0,158	4,201	4,583	4,940	2,725	6,919	4,194
TH62	W1	1,834	33,022	0,415	4,281	5,415	6,748	0,886	13,151	12,265
	W2	1,205	22,053	0,243	4,630	5,391	6,217	2,168	9,466	7,298
	W3	0,983	17,801	0,181	4,834	5,491	6,193	2,634	8,741	6,107
	W4	0,822	14,922	0,281	4,925	5,458	6,041	3,059	8,534	5,475
	W5	0,704	12,700	0,089	5,066	5,523	6,003	3,147	8,475	5,328
	W6	0,628	11,309	0,143	5,119	5,555	5,976	3,772	8,611	4,839
TH43	W1	2,216	36,523	0,372	4,480	5,875	7,545	0,839	13,914	13,075
	W2	1,474	24,857	0,270	4,879	5,852	6,883	1,927	11,227	9,300
	W3	1,164	19,449	0,112	5,156	5,970	6,740	2,318	9,899	7,581
	W4	1,009	16,805	0,262	5,276	5,962	6,667	2,623	9,898	7,275
	W5	0,851	14,158	0,138	5,415	5,983	6,583	3,278	8,906	5,628
	W6	0,764	12,662	0,188	5,495	6,035	6,533	3,873	8,994	6,021
TH322	W1	2,204	36,695	0,366	4,419	5,811	7,491	0,635	13,874	13,239
	W2	1,468	25,015	0,270	4,809	5,809	6,824	1,836	10,927	9,091
	W3	1,155	19,490	0,117	5,117	5,909	6,686	2,355	9,807	7,453
	W4	0,998	16,808	0,245	5,225	5,921	6,592	2,621	9,716	7,095
	W5	0,841	14,155	0,118	5,348	5,916	6,504	3,277	8,908	5,631
	W6	0,752	12,604	0,154	5,438	5,975	6,468	3,820	9,707	5,886

Źródło: opracowanie własne.

W przypadku analizowanych metod zauważono, że wraz ze wzrostem udziału luk w analizowanej zmiennej spada odchylenie standardowe, a także rozstęp między maksymalnymi a minimalnymi błędami MAPE. Aby znaleźć przyczynę tej sytuacji, należy dokładnie przeanalizować rozkłady błędów MAPE uzyskanych dla poszczególnych metod. Na rysunku 2 przedstawiono rozkłady empiryczne średnich względnych błędów prognoz interpolacyjnych dla poszczególnych wariantów udziałów luk dla modelu z wielomianem trygonometrycznym.



**Rys. 2.** Rozkłady empiryczne średnich względnych błędów prognoz interpolacyjnych dla modelu z wielomianem trygonometrycznym w zależności od wariantu udziału luk w danych

Źródło: opracowanie własne.

Z analizy rys. 2 oraz informacji zawartych w tab. 2 wynika, że wraz ze wzrostem udziału luk interpolacyjnych w zmiennej rozkłady błędów MAPE ulegają wysmukleniu. Zauważalne jest również zwiększanie odsetka obserwacji w przedziale dominanty. Zmiany w poziomie współczynnika skośności przedstawionego w tab. 2, którego wartości zmniejszają się wraz ze wzrostem udziałów luk w analizowanej zmiennej, oznaczają, że rozkłady empiryczne błędów prognoz interpolacyjnych zmieniają swój kształt z silnie prawostronnie asymetrycznych w kierunku bardziej symetrycznych. Zwiększanie liczby luk w danych powoduje więc w pewnym stopniu „uśrednienie” średnich względnych błędów prognoz interpolacyjnych. Podobną sytuację zaobserwowano w przypadku pozostałych analizowanych metod. Odpowiedzialny za ten stan rzeczy może być wewnętrzny układ luk, ponieważ zwiększanie liczby luk powoduje wzrost liczby kombinacji rozmieszczenia luk w badanej zmiennej.

Nie oznacza to oczywiście, że duża liczba luk interpolacyjnych jest pożądana. Wręcz przeciwnie – im mniejszy udział luk, tym większa szansa, że prognozy in-

terpolacyjne zbudowane na podstawie takiego szeregu czasowego będą obciążone mniejszym błędem. Ma to tym większe znaczenie w przypadku budowy prognoz ekstrapolacyjnych.

W tabeli 3 przedstawiono kształtowanie się przeciętnych błędów prognoz ekstrapolacyjnych według wariantów udziału luk w danych. Błędy te obliczono na podstawie rozkładów średnich względnych błędów prognoz (MAPE) ekstrapolacyjnych. W tabeli zamieszczono również informację o średnim względnym błędzie prognoz ekstrapolacyjnych obliczonych na podstawie pełnego szeregu.

**Tabela 3.** Kształtowanie średnich względnych błędów prognoz ekstrapolacyjnych dla pełnego szeregu oraz przeciętnych błędów obliczonych na podstawie rozkładów ekstrapolacyjnych błędów MAPE według wariantów udziału luk w danych (w %)

Model	Pełny szereg	Wariant udziału luk					
		W1	W1	W1	W1	W1	W1
w.tryg	3,283	3,298	3,308	3,321	3,332	3,357	3,379
0/1	3,276	3,279	3,289	3,301	3,313	3,339	3,362
TH34	2,938	2,974	3,002	3,028	3,051	3,076	3,117
TH43	4,890	4,889	4,888	4,890	4,898	4,910	4,935
TH62	4,345	4,345	4,347	4,350	4,363	4,372	4,390
TH322	4,941	4,939	4,940	4,939	4,947	4,961	4,980

Źródło: opracowanie własne.

Z informacji zawartych w tab. 3 wynika, że najniższymi przeciętnymi błędami, zawierającymi się w przedziale od 2,97 do 3,12%, charakteryzował się model TH34. Był on nieznacznie lepszy od modeli klasycznych: z wielomianem trygonometrycznym oraz zmienną zero-jedynkową, których przeciętne błędy wynosiły między 3,28 a 3,78%.

Największy spadek dokładności prognoz ekstrapolacyjnych zanotowano dla modelu TH34, aczkolwiek wyniósł on jedynie 0,179 punktu procentowego. W przypadku prognoz interpolacyjnych usunięcie części obserwacji poprawiało czasami dokładność prognoz, jednak w przypadku prognoz ekstrapolacyjnych obserwujemy jedynie negatywne skutki wzrostu udziału luk w badanej zmiennej. Na podstawie analizy tab. 3 stwierdzić można również, że podobnie jak w przypadku prognoz interpolacyjnych wpływ wzrostu ilości brakujących informacji w zmiennej objaśnianej na dokładność prognoz był różny dla różnych metod.

Analiza pozostałych statystyk opisowych rozkładów empirycznych średnich względnych błędów prognoz ekstrapolacyjnych zawartych w tab. 4 wskazuje, że ulegają one zmianom w wyniku zwiększenia się udziału luk w danych. Na podstawie współczynnika skośności możemy wnioskować, że badane rozkłady zmieniają się, dążąc do prawostronnie asymetrycznych. Dostrzegalne jest również ich „wypląszczanie”, co oznacza rosnące wraz ze zwiększaniem udziałów luk: odchylenie

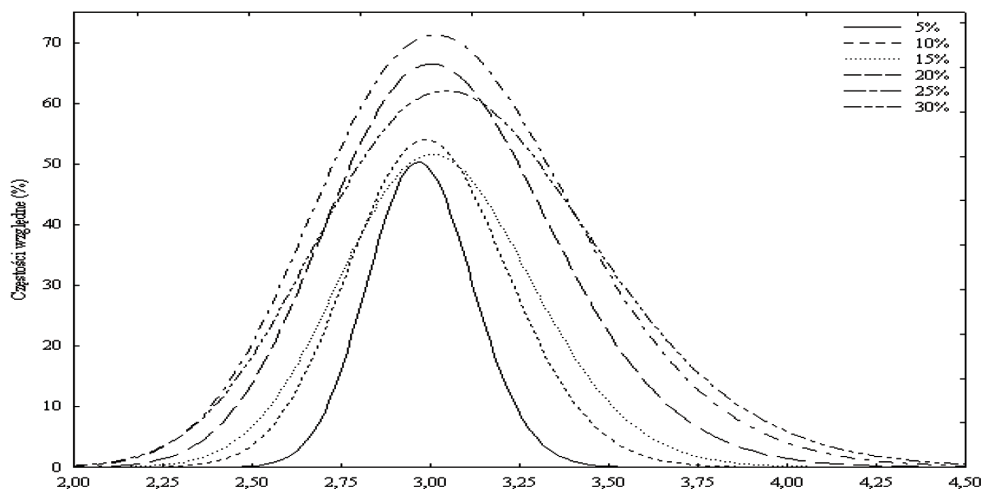
**Tabela 4.** Wybrane statystyki opisowe empirycznych rozkładów średnich względnych prognoz ekstrapolacyjnych dla wybranych metod oraz udziałów luk

Metoda	Wariant	Odch.std	Wsp.zmn. (w %)	Skośność	Q1	Mediana	Q3	Minimum	Maksimum	Rozstęp
w.tryg	W1	0,134	4,061	-0,454	3,216	3,307	3,393	2,657	3,691	1,034
	W2	0,192	5,813	-0,246	3,185	3,320	3,436	2,689	3,862	1,174
	W3	0,238	7,165	-0,112	3,157	3,323	3,490	2,493	4,212	1,719
	W4	0,286	8,592	-0,101	3,139	3,337	3,525	2,386	4,259	1,873
	W5	0,338	10,064	0,017	3,131	3,364	3,571	2,141	4,645	2,504
	W6	0,362	10,720	0,037	3,134	3,393	3,616	2,151	4,747	2,596
0/1	W1	0,133	4,044	-0,365	3,195	3,287	3,372	2,702	3,680	0,978
	W2	0,200	6,072	-0,190	3,157	3,293	3,429	2,688	3,853	1,165
	W3	0,249	7,556	-0,069	3,138	3,301	3,474	2,400	4,230	1,830
	W4	0,297	8,962	-0,002	3,112	3,320	3,511	2,311	4,284	1,973
	W5	0,352	10,534	0,123	3,108	3,340	3,559	2,177	4,725	2,548
	W6	0,382	11,349	0,099	3,100	3,366	3,605	2,218	4,779	2,561
TH34	W1	0,148	4,965	0,091	2,879	2,973	3,066	2,520	3,586	1,066
	W2	0,223	7,434	0,261	2,849	2,995	3,149	2,425	3,789	1,364
	W3	0,268	8,852	0,353	2,837	3,017	3,202	2,303	4,087	1,785
	W4	0,319	10,449	0,423	2,816	3,022	3,257	2,374	4,266	1,892
	W5	0,367	11,944	0,513	2,805	3,032	3,313	2,261	4,726	2,465
	W6	0,399	12,811	0,587	2,818	3,071	3,364	2,253	4,660	2,407
TH62	W1	0,100	2,292	0,151	4,279	4,342	4,407	4,010	4,706	0,696
	W2	0,145	3,338	0,126	4,243	4,346	4,445	3,859	4,937	1,078
	W3	0,184	4,232	0,227	4,224	4,345	4,466	3,814	5,126	1,312
	W4	0,214	4,901	0,114	4,221	4,362	4,504	3,789	5,161	1,372
	W5	0,257	5,870	0,230	4,199	4,366	4,533	3,540	5,414	1,873
	W6	0,283	6,445	0,355	4,184	4,387	4,559	3,509	5,667	2,158
TH43	W1	0,206	4,222	-0,195	4,761	4,901	5,024	4,161	5,592	1,431
	W2	0,302	6,183	-0,088	4,682	4,890	5,101	3,925	5,846	1,921
	W3	0,381	7,786	0,014	4,640	4,891	5,145	3,645	6,246	2,601
	W4	0,452	9,229	0,006	4,591	4,898	5,212	3,723	6,347	2,625
	W5	0,518	10,539	0,182	4,551	4,904	5,261	3,416	6,963	3,547
	W6	0,569	11,525	0,206	4,537	4,919	5,321	3,515	7,054	3,539
TH322	W1	0,209	4,226	-0,387	4,815	4,957	5,073	4,142	5,569	1,427
	W2	0,304	6,150	-0,177	4,742	4,944	5,149	3,929	5,838	1,909
	W3	0,383	7,758	-0,055	4,687	4,954	5,203	3,665	6,281	2,616
	W4	0,453	9,166	-0,045	4,633	4,957	5,269	3,739	6,431	2,692
	W5	0,519	10,467	0,107	4,591	4,956	5,312	3,505	6,903	3,398
	W6	0,568	11,409	0,177	4,576	4,971	5,357	3,582	7,039	3,458

Źródło: opracowanie własne.



standardowe, współczynnik zmienności oraz rozstęp. Jako przykład na poniższym rysunku przedstawiono rozkłady empiryczne średnich względnych błędów prognoz ekstrapolacyjnych dla poszczególnych wariantów udziału luk w danych dla modelu TH34.



**Rys. 3.** Rozkłady empiryczne średnich względnych błędów prognoz ekstrapolacyjnych dla modelu TH34 w zależności od wariantu udziału luk w danych

Źródło: opracowanie własne.

Porównanie informacji zawartych w tab. 3 oraz pozycyjnych miar tendencji centralnej rozkładów z tab. 4 wskazuje, że budowa prognoz na podstawie szeregu zawierającego brakujące informacje nie zawsze powoduje zmniejszenie dokładności prognoz ekstrapolacyjnych. Dla większości analizowanych modeli oraz wariantów udziałów luk aż w ponad 25% przypadków uzyskano prognozy lepsze niż dla danych pełnych. W niektórych przypadkach odsetek ten był wyższy, przekraczając 50%. Wynika to z usunięcia z analizowanego szeregu obserwacji odstających oraz wpływowch, dzięki czemu poprawiły się własności predyktorów na nim zbudowanych.

## 4. Podsumowanie

W pracy przedstawiono wyniki zastosowania metod symulacyjnych w badaniu wpływu rozmieszczenia luk niesystematycznych na dokładność prognozowania inter- i ekstrapolacyjnego na podstawie zmiennej z wahaniami sezonowymi. Z analizy załączonego przykładu wynika, że:

1. Wraz ze wzrostem udziału luk zmieniały się w niewielkim stopniu wielkości przeciętnych błędów prognoz obliczonych na podstawie rozkładów empirycznych

średnich względnych błędów prognoz inter- i ekstrapolacyjnych. Nie zawsze zwiększenie udziału luk oznaczało pogorszenie prognoz. Zależało to przede wszystkim od metody na podstawie, której zbudowano prognozy.

2. Natomiast wraz ze wzrostem udziału luk zmieniały się kształty rozkładów empirycznych. Rozkłady błędów interpolacyjnych zmieniały kształt z prawostronnie asymetrycznych w kierunku symetrycznych. Ulegały one również „wysmukleniu”. Świadczą o tym oceny klasycznych i pozycyjnych miar tendencji centralnej oraz dyspersji. W przypadku rozkładów średnich względnych błędów ekstrapolacyjnych zmiany te przebiegały w odwrotnym kierunku.

3. Na dokładność prognoz inter- i ekstrapolacyjnych wpływ miał rodzaj modelu. Widoczne to było w przypadku ocen błędów prognoz dla modeli hierarchicznych. Trzy spośród nich (TH26, TH223, TH232) zostały wyeliminowane ze względu na zbyt niskie oceny średniego współczynnika determinacji oraz znacznie wyższe błędy prognoz inter- i ekstrapolacyjnych. Oceny błędów bliskie modelom klasycznym otrzymano na podstawie modelu TH34. Rozkłady błędów dla prognoz interpolacyjnych dla tego modelu były przesunięte względem modeli klasycznych w prawo o ok. 0,5 punktu procentowego, natomiast rozkłady błędów prognoz ekstrapolacyjnych w lewo o ok. 0,3 punktu procentowego.

4. Nie zawsze budowa prognoz na podstawie szeregu zawierającego brakujące informacje powoduje zmniejszenie dokładności prognoz ekstrapolacyjnych. Prognozy otrzymane dla niektórych wariantów luk mogą być nieznacznie lepsze niż zbudowane na podstawie pełnego szeregu czasowego. Może to wynikać z faktu występowania w okresach objętych lukami obserwacji odstających oraz wpływowych.

## Literatura

- Markiewicz A., Zawadzki J., *Prognozowanie brakujących danych dekadowych dla luk systematycznych na podstawie hierarchicznych modeli szeregu czasowego*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 1001, AE, Wrocław 2003, s. 89-95.
- Szmuksta-Zawadzka M., Zawadzki J., *O metodzie prognozowania brakujących informacji dla danych sezonowych*, „Przegląd Statystyczny” 1995, nr 3-4, s. 377-385.
- Szmuksta-Zawadzka M., Zawadzki J., *Prognozowanie na podstawie hierarchicznych modeli szeregu czasowego z lukami w danych*, [w:] *Analiza szeregów czasowych na początku XXI wieku*, Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika, Toruń 2002, s. 93-102.
- Zawadzki J., *Prognozowanie brakujących informacji na podstawie sezonowych predyktorów przy czynowo-opisowych*, „Przegląd Statystyczny” 1996, nr 1-2, s. 15-24.
- Zawadzki J. (red.), *Ekonomiczne metody predykcji dla danych sezonowych w warunkach braku pełnej informacji*, Rozprawy i Studia T. (CDXVI) 342, Wydawnictwo Naukowe Uniwersytetu Szczecińskiego, Szczecin 1999.
- Zawadzki J. (red.), *Zastosowanie hierarchicznych modeli szeregów czasowych w prognozowaniu zmiennych ekonomicznych z wahaniami sezonowymi*, Wydawnictwo Akademii Rolniczej w Szczecinie, Szczecin 2003.

## **SIMULATION STUDY OF INFLUENCE OF FREQUENCY OF INCIDENCE OF NON-SYSTEMATIC GAPS IN TIME SERIES ON ACCURACY OF FORECASTS**

**Summary:** This paper presents the results of application of simulation methods to the analysis of frequency of incidence of non-systematic gaps on the accuracy of inter-and extrapolative forecasts for time series with seasonal fluctuations. In the analysis there was used variable which describes production of electricity in Poland in the years 2002-2011. In the process of modelling there were used time series models with the periodic seasonal component, in which seasonality was described by dummy variable or trigonometric polynomial, and hierarchical models. In the calculation process there were used packages R and Statistica 9.0.

**Keywords:** simulations, non-systematic gaps, forecasting.