

*Maciej Oesterreich*

## KOMBINACJE LINIOWE SKŁADOWYCH HARMONICZNYCH A DOKŁADNOŚĆ PROGNOZ W SEZONOWYCH SZEREGACH CZASOWYCH Z LUKAMI SYSTEMATYCZNYMI

### LINEAR COMBINATIONS OF HARMONIC COMPONENTS AND ACCURACY OF FORECASTS IN THE SEASONAL TIME SERIES WITH SYSTEMATIC GAPS

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 the paper were presented results of application of simulation methods in analysis of accuracy of inter- and extrapolative forecasts for systematic gaps in the data in situation of occurrence of linear combinations of harmonic components. In analyze were used the number of tourists accommodated in accommodation establishments by month in Poland in 2003–2011. The forecasts were constructed using the classical time series models with a linear trend and seasonality described by the trigonometric polynomial. Calculations were made using the R package and Statistica 10.

**Słowa kluczowe:** liniowe kombinacje składowych, luki systematyczne, modele z wielomianem trygonometrycznym.

**Key words:** linear combinations of harmonic components, systematic gaps, models with trigonometric polynomial.

#### WPROWADZENIE TEORETYCZNE

W prognozowaniu na podstawie danych w postaci niekompletnych szeregów czasowych z wahaniami sezonowymi możemy mieć do czynienia z dwoma rodzajami luk:

1. lukami niesystematycznymi,
2. lukami systematycznymi.

Luki niesystematyczne występują wtedy, gdy dostępne są dane o każdym z podokresów cyklu wahań sezonowych. Jeżeli nie są one dostępne przynajmniej dla jednego podokresu tego cyklu, oznacza to wystąpienie luk systematycznych. W modelowaniu dla luk systematycznych wykorzystywane są przede wszystkim modele, w których sezonowość opisywana jest za pomocą składowych sinuso- i cosinusoidalnych. Zapis takiego modelu z trendem liniowym i periodycznym składnikiem sezonowym jest następujący (Zieliński 1969, Zeliaś i in. 2003, s. 89):

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

gdzie:

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

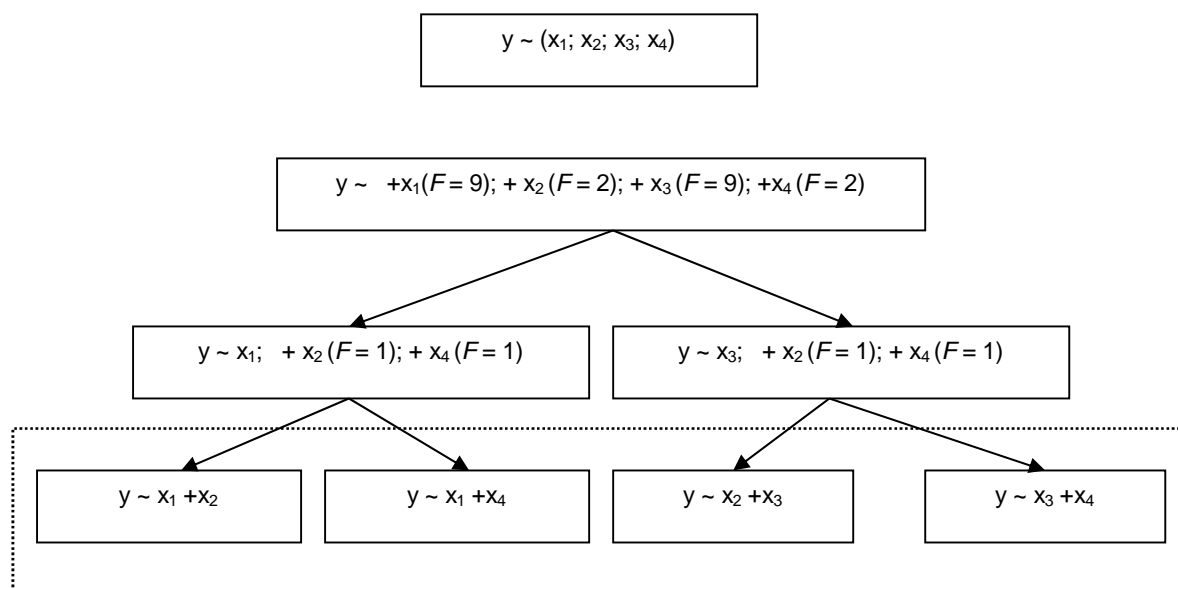
$m$  – liczba podokresów cyklu sezonowego,  
 $\varepsilon_t$  – składnik losowy.

Do wyboru statystycznie istotnych składowych harmoniczych, niezależnie od rodzaju luk, wykorzystuje się m.in. metodę selekcji a priori w ramach procedury regresji krokowej.

Dla luk systematycznych przebieg tego procesu ulega znacznej komplikacji na skutek (Zawadzki 2003, s. 26):

- współliniowości składowych harmoniczych,
- zerowych wariancji niektórych z nich,
- wystąpienia liniowych kombinacji niektórych składowych.

Sytuacje wymienione w punktach a i b są łatwe do wykrycia – wystarczy obliczenie współczynników korelacji liniowej Pearsona lub wariancji składowych. Natomiast wystąpienie liniowych kombinacji składowych można stwierdzić dopiero w trakcie wyboru składowych z wykorzystaniem procedury regresji krokowej opartej np. na statystykach  $F$  (Anderson i in. 2011, s. 739). Zjawisko to może wystąpić najwcześniej w drugiej iteracji, gdy przynajmniej dla dwóch składowych otrzyma się takie same, maksymalne bezwzględne wartości statystyki  $F$ . W tej sytuacji znajdzie konieczność oszacowania wersji modeli zawierających kolejno jedną z nich (Wiśniewski i Zieliński, s. 110). Schemat takiego postępowania przedstawiono na rysunku 1.



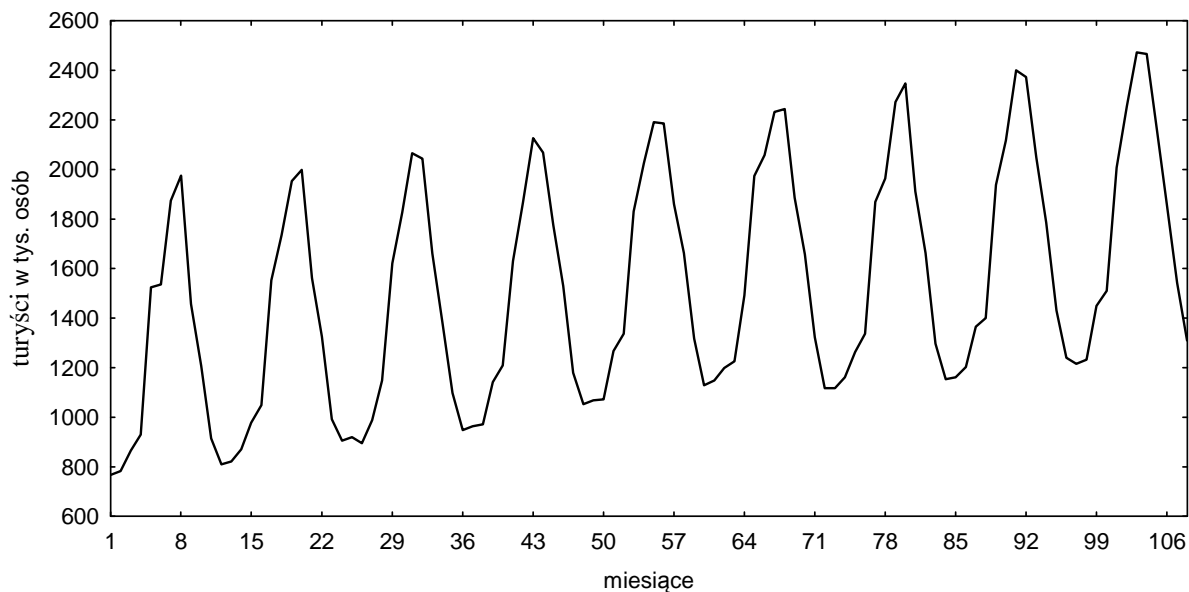
Rys. 1. Procedura budowy modeli na podstawie szeregu zawierającego luki systematyczne  
 Źródło: opracowanie własne.

Z rysunku wynika, że ostateczna liczba szacowanych wersji równa jest iloczynowi liczb składowych tworzących kombinacje liniowe w kolejnych iteracjach. Wszystkie szacowane wersje będą nierozróżnialne z punktu widzenia własności predykcyjnych (Zawadzki 1999, Oesterreich 2012). Różnić się będą one, niekiedy nawet bardzo, ocenami błędów prognoz inter- i ekstrapolacyjnych.

Celem niniejszego artykułu jest zastosowanie metod quasi-symulacyjnych do badania wpływu wystąpienia kombinacji liniowych składowych na dokładność prognoz inter- i ekstrapolacyjnych.

## PRZYKŁAD EMPIRYCZNY

Badanie wpływu wystąpienia kombinacji liniowej składowych na dokładność prognoz inter- i ekstrapolacyjnych zostanie przeprowadzona na przykładzie kształtowania się liczby turystów korzystających z obiektów zbiorowego zakwaterowania według miesięcy w Polsce. Dane statystyczne za lata 2003–2010 zostały zaczerpnięte z wydawanych corocznie przez Główny Urząd Statystyczny opracowań pt. „Turystyka”. Dane za lata 2003–2010 posłużyły do szacowania parametrów modeli, natomiast 2011 rok był okresem empirycznej weryfikacji prognoz. Kształtowanie się badanej zmiennej przedstawiono na rysunku 2.



Rys. 2. Turyści korzystający z obiektów zbiorowego zakwaterowania według miesięcy w Polsce w latach 2003–2011

Źródło: Turystyka 2003–2011.

Z rysunku wynika, że zmienna ta charakteryzuje się tendencją wzrostową i stałymi wahaniami sezonowymi.

W celu ustalenia struktury harmonicznej analizowanej zmiennej najpierw dokonano eliminacji trendu linowego, a następnie obliczono udziały poszczególnych składowych sinuso- i cosinusoidalnych w wyjaśnianiu łącznej wariancji sezonowej. Udziały procentowe składowych zestawiono w tabeli 1.

Tabela 1. Udziały poszczególnych składowych harmonicznych w wyjaśnianiu wariancji [w %]

Składowe	Udziały	Składowe	Udziały	Suma
s1	33,93	c1	61,88	95,81
s2	2,64	c2	0,02	2,66
s3	0,03	c3	0,01	0,04
s4	0,21	c4	0,25	0,46
s5	0,75	c5	0,18	0,93
–	–	c6	0,10	0,10
Suma	37,56	–	62,44	100,00

\*  $s_i = \sin \omega_i$ ,  $c_i = \cos \omega_i$

Źródło: opracowanie własne.

Z analizy informacji zawartych w tabeli 1 wynika, że zmienna prognozowana charakteryzuje się przede wszystkim sezonowością o cyklu rocznym (udział składowych s1 i c1 wynosi blisko 96%). Udział powyżej 1% posiadają także wahania o cyklu półrocznym. Wariancja sezonowa w największym stopniu wyjaśniana jest przez wahania cosinusoidalne (62,44%).

Korzystając ze skonstruowanego algorytmu dla środowiska R wygenerowano, wszystkie możliwe, 924 kombinacje układów luk systematycznych dla przypadku, gdy liczba miesięcy z lukami w każdym roku wynosiła 6. Następnie za pomocą kolejnego algorytmu oszacowano parametry modeli i zbudowano prognozy inter- i ekstrapolacyjne, a także obliczono średnie względne błędy prognoz (MAPE).

Ze względu na wystąpienie kombinacji linowych składowych oraz ewentualnie ich współliniowości oszacowano 4253 modele. Przykładowo, w tabeli 2 zestawiono wersje modeli dla kombinacji nr 638 wraz z informacją o własnościach predykcyjnych oraz ocenach średnich względnych błędów prognoz inter- i ekstrapolacyjnych. Luki w analizowanej kombinacji przypadały na: luty, kwiecień, czerwiec, sierpień, październik oraz grudzień.

Tabela 2. Zmienne objaśniające, parametry struktury stochastycznej i dobroci dopasowania oraz oceny średnich względnych błędów prognoz inter- i ekstrapolacyjnych dla wersji modeli obliczonych dla kombinacji nr 638

Zmienne objaśniające										Parametry struktury stochastycznej i dobroci dopasowania			MAPE		Udział składowych w wyjaśnianiu wariancji (%)
t	s1	s2	s3	s4	s5	c1	c2	c4	c5	R <sup>2</sup>	Se	Vs (%)	I	E (h = 12)	
1	1	1	1	0	0	1	1	0	0	0,99	48041,96	3,21	6,44	4,90	98,68
1	1	1	1	0	0	1	0	1	0	0,99	48041,96	3,21	6,24	4,23	98,73
1	1	0	1	1	0	1	1	0	0	0,99	48041,96	3,21	10,39	5,08	96,07
1	1	0	1	1	0	1	0	1	0	0,99	48041,96	3,21	9,83	4,50	96,30
1	0	1	1	0	1	1	1	0	0	0,99	48041,96	3,21	30,52	14,92	65,32
1	0	1	1	0	1	1	0	1	0	0,99	48041,96	3,21	30,55	14,25	65,55
1	0	0	1	1	1	1	1	0	0	0,99	48041,96	3,21	29,75	14,63	62,89
1	0	0	1	1	1	1	0	1	0	0,99	48041,96	3,21	29,78	13,96	63,12
1	1	1	1	0	0	0	1	0	1	0,99	48041,96	3,21	45,77	20,79	36,80
1	1	1	1	0	0	0	0	1	1	0,99	48041,96	3,21	45,41	20,66	37,03
1	1	0	1	1	0	0	1	0	1	0,99	48041,96	3,21	44,35	20,25	34,37
1	1	0	1	1	0	0	0	1	1	0,99	48041,96	3,21	43,99	20,12	34,60
1	0	1	1	0	1	0	1	0	1	0,99	48041,96	3,21	52,34	24,18	3,62
1	0	1	1	0	1	0	0	1	1	0,99	48041,96	3,21	52,30	24,14	3,62
1	0	0	1	1	1	0	1	0	1	0,99	48041,96	3,21	52,53	23,97	1,19
1	0	0	1	1	1	0	0	1	1	0,99	48041,96	3,21	53,22	24,32	1,19

Źródło: opracowanie własne.

Z tabeli 2 wynika, że dla wszystkich wersji modeli otrzymano identyczne oceny parametrów struktury stochastycznej i dobroci dopasowania: Se, Vs oraz R<sup>2</sup>. Różnią się natomiast bardzo wyraźnie ocenami zarówno prognoz inter- i ekstrapolacyjnych. W przypadku prognoz interpolacyjnych błędy zawierają się w przedziale od 6,24% do 52,53%, a błędy prognoz ekstrapolacyjnych od 4,23% do 24,32%.

W ostatniej kolumnie tabeli 2 podane zostały dla poszczególnych wersji także sumaryczne wielkości udziałów składowych w wyjaśnianiu wariancji sezonowej dla pełnych danych. Z porównania ich z ocenami błędów obu rodzajów prognoz wynika, że istnieje ścisła zależność odwrot-

na (ujemna) między udziałami poszczególnych składowych harmonicznych a wielkościami błędów. Najniższe oceny błędów prognoz otrzymano dla tych kombinacji, w których występowały składowe  $s_1$  i  $c_1$ , a najwyższe, jeżeli te składowe nie występowały.

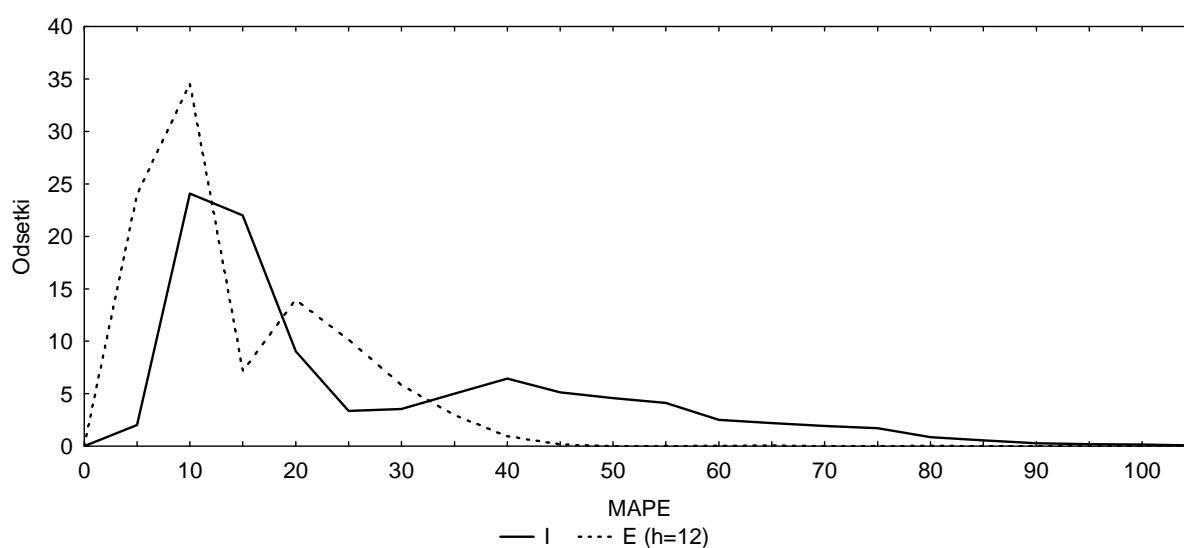
Otrzymanie bardzo dużej liczby szacowanych modeli pozwala na przeprowadzenie analizy porównawczej rozkładów błędów obu rodzajów prognoz. W tabeli 3 przedstawiono kształtowanie się wybranych statystyk opisowych rozkładów średnich względnych błędów prognoz (MAPE) inter- i ekstrapolacyjnych.

Tabela 3. Wybrane statystyki opisowe rozkładów średnich względnych błędów prognoz inter- i ekstrapolacyjnych

Statystyki opisowe	Inter- (I)	Ekstra- (h = 12)
Średnia	25,93	12,00
Odchylenie standardowe	21,55	9,14
Współczynnik zmienności (%)	83,10	76,17
Skośność	1,81	1,62
Q1	9,79	5,07
Mediana	15,64	7,23
Q3	39,82	17,97
Minimum	3,64	2,11
Maksimum	310,62	124,59
Rozstęp	306,97	122,48

Źródło: opracowanie własne.

Informacje zawarte w tabeli 3 wskazują, że niższe o około 14 p.p. przeciętne oceny błędów otrzymano dla prognoz ekstrapolacyjnych. Rozkłady błędów prognoz ekstrapolacyjnych charakteryzowały się także niższą dyspersją oraz wyraźnie niższym rozstępem pomiędzy ocenami minimalnymi i maksymalnymi. Na rysunku 3 przedstawiono kształtowanie się empirycznych rozkładów średnich względnych błędów prognoz inter- i ekstrapolacyjnych.



Rys. 3. Empiryczne rozkłady średnich względnych błędów prognoz inter- (I) i ekstrapolacyjnych (E) dla modelu ze stałą sezonowością opisywaną za pomocą wielomianu trygonometrycznego

Źródło: opracowanie własne.

Z rysunku wynika, że zarówno rozkłady błędów prognoz inter- jak i ekstrapolacyjnych charakteryzują się dużą nieregularnością. Spowodowana ona jest występowaniem wielomodalności oraz tzw. długich ogonów, wskazujących na pojawianie się bardzo wysokich ocen błędów obu rodzajów prognoz.

Ze względu na wystąpienie dużych różnic pomiędzy błędami minimalnymi a maksymalnymi wyodrębnione zostały dwie grupy różniące się liczbą szacowanych modeli. W pierwszej grupie (1) znajdowały się te modele, w których nie występowały kombinacje liniowe składowych. Do grupy drugiej (2) zaliczono modele, dla których w następstwie kombinacji liniowych szacowano przynajmniej dwie wersje. W tabeli 4 zestawiono statystyki opisowe rozkładów średnich względnych błędów prognoz inter- i ekstrapolacyjnych dla wyodrębnionych grup modeli.

Tabela 4. Wybrane statystyki opisowe rozkładów średnich względnych błędów prognoz inter- i ekstrapolacyjnych dla grup 1. i 2.

Statystyki opisowe	Inter- (I)		Ekstra- (E, h = 12)	
	1	2	1	2
Liczba kombinacji układów luk	442	482	442	482
Liczba szacowanych modeli	442	3811	442	3811
Średnia	18,19	26,83	8,83	12,37
Odchylenie standardowe	18,88	21,66	8,04	9,19
Współczynnik zmienności (%)	103,79	80,74	90,98	74,31
Skośność	2,10	1,800	2,04	1,59
Q1	6,63	10,17	4,11	5,20
Mediana	11,00	16,59	5,51	7,66
Q3	17,22	40,37	8,36	18,16
Minimum	4,22	3,64	2,81	2,11
Maksimum	121,59	310,62	51,26	124,59
Rozstęp	117,37	306,97	48,46	122,48

Źródło: opracowanie własne.

Z tabeli 4 wynika, że dla 442 układów luk szacowano tylko po jednym modelu. Natomiast dla 482 wariantów układów układy luk z kombinacjami liniowymi dokonano estymacji parametrów 3811 wersji modeli. Oznacza to, że przeciętnie dla każdego wariantu szacowano około ośmiu wersji modeli.

Z analizy kształtowania się przeciętnych ocen błędów prognoz dla grupy 1. wynika, że otrzymano wyraźnie niższe oceny niż dla grupy 2. Różnica między nimi wynosiła około 8,5 p.p. dla błędów prognoz interpolacyjnych i około 3,5 p.p. dla błędów prognoz ekstrapolacyjnych.

Tabela 5. Oceny statystyk testowych testu dla grupy 1. i 2. błędów prognoz inter- i ekstrapolacyjnych

Rodzaj prognoz	Test	Poziom statystyki testowej	P-value
Inter-	Serii Walda-Wolfowitza	$Z = -3,4702$	0,0005
	U Mana-Whitneya	$Z = -10,8589$	0,0000
	Kołmogorowa-Smirnowa	–	$p < 001$
Ekstra-	Serii Walda-Wolfowitza	$Z = -4,7881$	0,0000
	U Mana-Whitneya	$Z = -10,0485$	0,0000
	Kołmogorowa-Smirnowa	–	$p < 001$

Źródło: opracowanie własne.

W celu zbadania czy rozkłady błędów prognoz pochodzą z tej samej populacji zastosowano trzy testy nieparametryczne (zob. Domański 1990):

- U Manna-Whitneya,
- serii Walda-Wolfowitza,
- Kołmogorowa-Smirnowa.

Wartości statystyk testowych wraz z ocenami granicznych poziomów istotności przedstawiono w tabeli 5.

Z tabeli 5 wynika, że w przypadku wszystkich testów należy odrzucić hipotezę zerową, która wskazuje na brak statystycznych różnic pomiędzy rozkładami błędów prognoz w obu grupach.

Kolejnym etapem badań było testowanie hipotezy o zgodności rozkładów błędów pomiędzy grupą 1. a wyodrębnioną z grupy 2. podgrupę 3., w skład której wchodziły wersje modeli charakteryzujące się najniższymi ocenami błędów prognoz dla każdego z 482 układów luk z kombinacjami liniowymi składowych.

W tabeli 6 przedstawiano charakterystyki opisowe rozkładów błędów prognoz inter- i ekstrapolacyjnych dla grup 1. i 3.

Tabela 6. Wybrane statystyki opisowe rozkładów średnich względnych błędów prognoz inter- i ekstrapolacyjnych dla grup 1. i 3.

Statystyki opisowe	Inter- (I)		Ekstra- (E, h = 12)	
	1	3	1	3
N	442	482	442	482
Średnia	18,19	17,47	8,83	8,50
Odchylenie standardowe	18,88	17,88	8,04	7,69
Współczynnik zmienności (%)	103,79	102,37	90,98	90,54
Skośność	2,10	1,70	2,04	1,62
Q1	6,63	6,16	4,11	3,78
Mediana	11,00	9,49	5,51	4,90
Q3	17,22	17,87	8,36	8,74
Minimum	4,22	3,64	2,81	2,11
Maksimum	121,59	88,24	51,26	37,24
Rozstęp	117,37	84,60	48,46	35,13

Źródło: opracowanie własne.

Z porównania wielkości przedstawionych w tabeli 6 wynika, że oceny statystyk opisowych obliczone dla obu grup są zbliżone. Nieco niższe wartości parametrów opisujących rozkłady błędów prognoz inter- i ekstrapolacyjnych otrzymano dla grupy 3. Tylko w przypadku kwartyli Q3 niższą ocenę uzyskano dla grupy 1.

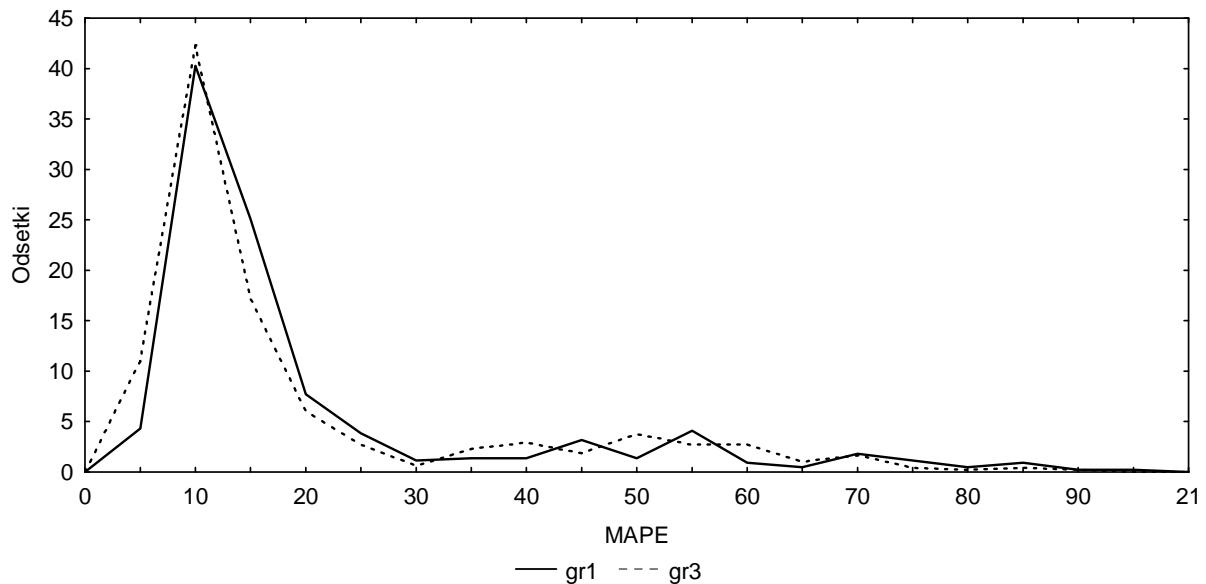
Ze względu na zbliżone oceny statystyk opisowych rozkładów testowano zgodność rozkładów za pomocą trzech wymienionych wcześniej testów statystycznych. W tabeli 7 przedstawiono informację o poziomach statystyk testowych oraz granicznych poziomach istotności.

Tabela 7. Oceny statystyk testowych testu zgodności dla dwóch niezależnych próbek wykonanych dla grup 1. i 3. prognoz inter- i ekstrapolacyjnych

Rodzaj prognoz	Test	Poziom statystyki testowej	P-value
Inter-	Serii Walda-Wolfowitza	$Z = -1,7237$	0,0909
	U Mana-Whitneya	$Z = 2,1556$	0,0311
	Kołmogorowa-Smirnowa	–	$p < 05$
Ekstra-	Serii Walda-Wolfowitza	$Z = -2,7789$	0,0054
	U Mana-Whitneya	$Z = 2,8892$	0,0039
	Kołmogorowa-Smirnowa	–	$p < 005$

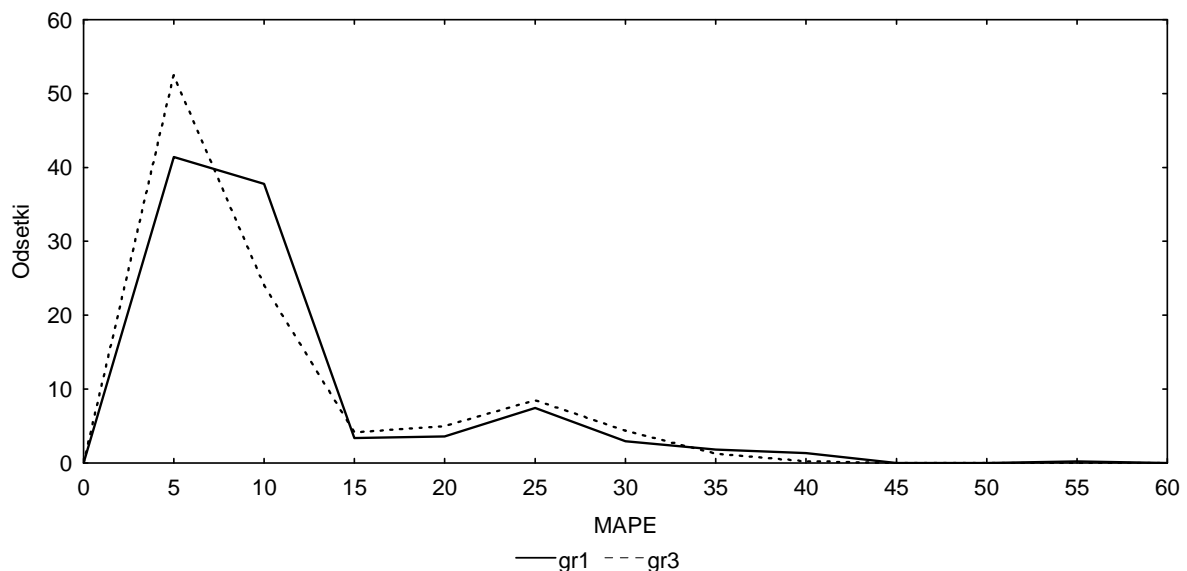
Źródło: opracowanie własne.

Wyniki uzyskane dla rozkładów błędów prognoz interpolacyjnych nie są jednoznaczne. Test serii Walda-Wolfowitza potwierdził zgodność rozkładów na poziomie istotności  $\alpha = 0,05$ , podczas gdy dwa pozostałe testy odrzuciły taką możliwość. Ze względu na niejednoznaczność wyników testowania zgodności rozkładów dla rozkładów błędów prognoz interpolacyjnych na rysunku 4 przedstawiano je w formie wieloboków liczebności.



Rys. 4. Empiryczne rozkłady średnich względnych błędów prognoz interpolacyjnych dla grup 1. i 3.  
Źródło: opracowanie własne.

Analiza rysunku potwierdza, że rozkłady błędów prognoz interpolacyjnych dla grup 1. i 3. mają bardzo zbliżony przebieg.



Rys. 5. Empiryczne rozkłady średnich względnych błędów prognoz ekstrapolacyjnych dla grup 1. i 3.  
Źródło: opracowanie własne.

W przypadku rozkładów błędów prognoz ekstrapolacyjnych wartości statystyk testowych wskazują, że przy przyjętym poziomie istotności należy odrzucić hipotezę zerową. Oznacza to,



że analizowane rozkłady błędów prognoz nie pochodzą z tej samej populacji. Na rysunku 5 przedstawiono empiryczne rozkłady błędów prognoz ekstrapolacyjnych dla grup 1. i 3.

## PODSUMOWANIE

Z analiz przeprowadzonych w artykule można wyprowadzić następujące wnioski:

1. Występowanie kombinacji liniowych składowych harmonicznych jest jednoznaczne z koniecznością szacowania liczby wersji modeli wielokrotnie przewyższającej liczbę układów luk systematycznych.
2. Najwyższe oceny średnich względnych błędów prognoz inter- i ekstrapolacyjnych otrzymano dla modeli zawierających składowe harmoniczne o najniższych udziałach w wyjaśnianiu wariacji sezonowej.
3. Przeciętne oceny błędów prognoz inter- i ekstrapolacyjnych uzyskane za pomocą modeli, dla modeli, których nie wystąpiły kombinacje liniowe składowych (grupa 1.) były niższe niż dla tych, w których je stwierdzono (grupa 2.). Rozkłady błędów obu rodzajów prognoz w sposób istotny różniły się od siebie.
4. Dla wersji modeli należących do grupy 2., charakteryzujących się najniższymi ocenami błędów (grupa 3.), otrzymano niższe przeciętne błędy obu rodzajów prognoz niż w przypadku modeli z grupy 1.

Zastosowane metody badawcze pozwoliły na osiągnięcie założonego celu pracy.

## PIŚMIENNICTWO

- Anderson D.R., Sweeney D.J., Williams T.A.** 2011. *Statistics for business and economics* 11e. South-Western Cengage Learning. ISBN 978-0-324-78325-4.
- Domański C.** 1990. *Testy statystyczne*. PWE. Warszawa. ISBN 83-208-0696-8.
- Oesterreich M.** 2012. Wykorzystanie programu R w prognozowaniu na podstawie modeli przyczynowo opisowych w warunkach braku pełnej informacji, *Folia Pomer. Univ. Technol. Stetin. Ser. Oecon.* 297(68), 55–65.
- Turystyka w latach:** 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011. 2003–2012. Warszawa. Informacje i opracowania statystyczne GUS. ISSN 1425–8846.
- Wiśniewski J.W., Zieliński Z.** 1996. *Elementy ekonometrii*. Toruń. Wydaw. Uniw. Mikołaja Kopernika. ISBN 83-231-0758-0.
- Zawadzki J. (red).** 1999. *Ekonometryczne metody predykcji dla danych sezonowych w warunkach braku pełnej informacji*. Szczecin. Wydaw. Uniw. Szczec. ISBN 83-7241-067-4.
- Zawadzki J. (red.).** 2003. *Zastosowanie hierarchicznych modeli szeregów czasowych w prognozowaniu zmiennych ekonomicznych z wahaniami sezonowymi*. Szczecin. Wydaw. Akad. Rol. Szczec. ISBN 83-7317-075-8.
- Zawadzki J.** 2012. Z badań nad metodami prognozowania na podstawie niekompletnych szeregów czasowych z wahaniami okresowymi (sezonowymi). *Prz. Stat.*, nr spec. 1. Warszawa, 140–154.
- Zeliś A., Pawełek B., Wanat S.** 2003. *Prognozowanie ekonomiczne. Teoria. Przykłady. Zadania*. Warszawa. Wydaw. Nauk. PWN. ISBN 978-83-01-14043-4.
- Zieliński Z.** 1969. *Ekonometryczne metody analizy wahań sezonowych*. Zesz. Nauk. Politech. Szczec., 112. Szczecin.

