

MARIA SZMUKSTA-ZAWADZKA, JAN ZAWADZKI

## Z BADAŃ NAD METODAMI PROGNOZOWANIA NA PODSTAWIE NIEKOMPLETNYCH SZERGÓW CZASOWYCH Z WAHANIAMI OKRESOWYMI (SEZONOWYMI)

### 1. SYNTEZA WYNIKÓW BADAŃ NAD METODAMI PROGNOZOWANIA BRAKUJĄCYCH DANYCH SEZONOWYCH

Przystępując do modelowania zmiennych ekonomicznych w wielu przypadkach spotykamy się z sytuacją, gdy w szeregach czasowych, zwłaszcza dla danych o krótszych niż rok okresach jednostkowych, występują luki. W świetle jednego z klasycznych założeń dotyczących modelowania ekonometrycznego byłoby to jednoznaczne z rezygnacją z jego przeprowadzenia. Jednak jak wykazują wyniki wieloletnich badań autorów zawarte m.in. w pracach Szmuksta-Zawadzka, Zawadzki (2002a, 2002b, 2002c, 2003) oraz jako rozdziały w monografiach pod redakcją Zawadzkiego (1999, 2003), przy spełnieniu określonych założeń dotyczących m.in. skali wahań sezonowych, odsetka brakujących danych czy sekwencyjności luk, możliwe jest modelowanie i prognozowanie zmiennych w warunkach braku pełnej informacji pozwalające na otrzymanie prognoz dopuszczalnych. Pojęcie prognozy dopuszczalnej może być rozpatrywane w różnorodny sposób<sup>1</sup>. W pracy za dopuszczalne uważać będziemy takie prognozy, których spodziewana dokładność wyrażona w liczbach absolutnych lub względnych dla założonego horyzontu prognozy jest akceptowalna przez użytkownika.

Do budowy prognoz zmiennych ekonomicznych wykazujących wahania sezonowe mogą być wykorzystywane zarówno metody dla danych oryginalnych (z sezonowością) jak i danych oczyszczonych z sezonowości. Poniżej przedstawione zostaną metody wykorzystujące te dwa rodzaje danych. Z uwagi na to, że większość niżej wymienionych metod prognozowania na podstawie kompletnych szeregów czasowych posiada bogatą literaturę, ograniczymy się jedynie do wskazania w kolejności alfabetycznej autorów ważniejszych prac z tego zakresu. Więcej uwagi poświęcimy natomiast tym metodom,

---

<sup>1</sup> Zdaniem Zeliasia (1997) sposoby definiowania kryteriów dopuszczalności zależą m.in. od charakteru zmiennej prognozowanej i warunków predykcji oraz stopnia pewności prognozy. W pracy Dittmann i inni (2009) stopień niepewności (dopuszczalności) należy rozpatrywać z punktu widzenia: błędów *ex ante*, wiarygodności prognozy, błędów *ex post* oraz oceny słownej. Najobszerniej na ten temat wypowiada się Guzik (2003) poświęcając cały rozdział swojej pracy. Formułuje dwie zasadnicze klasy kryteriów:  $\mu$ -kryterium oraz  $\omega$  – kryterium. Pierwsze z nich odnosi się do miernika niedokładności prognozy a drugie do wiarygodności. Następnie rozpatruje to pojęcie w aspekcie: miernika błędu prognozy, przedziału ufności, funkcji sygnałnych oraz horyzontu prognozy. Autorzy przywołanych wyżej prac na pierwszy plan wysuwają spełnianie przez nie kryterium dokładności ustalonego przez użytkownika lub prognostyka.

których stosowanie łączy się ze znacznym stopniem komplikacji w przebiegu procesu modelowania. Odnosić się to będzie zwłaszcza do przypadku występowania systematycznych luk w danych. Będziemy odwoływać się także do wybranych prac, w których prezentowane są wyniki modelowania i prognozowania zmiennych wykazujących wahania sezonowe warunkach braku pełnej informacji.

Do pierwszej grupy metod, wykorzystujących dane oryginalne, można zaliczyć:

- a) klasyczne modele szeregu czasowego z wahaniami sezonowymi (Czerwiński, Guzik, 1980; Dittmann, 2008; Zawadzki (red.), 1999),
- b) modele hierarchiczne szeregu czasowego (Zawadzki (red.), 2003; Szmuksta- Zawadzka, Zawadzki, 2000, 2002a, 2002b, 2002c, 2003),
- c) modele adaptacyjne uwzględniające wahania sezonowe (np. Holta-Wintersa) (Dittmann, 2008; Pawłowski, 1973, 1982; Zeliaś, 1997; Zeliaś, Pawełek, Wanat, 2003)
- d) modele sieci neuronowych (podstawy metodologiczne metody sztucznych sieci neuronowych oraz przykłady ich zastosowań można znaleźć m.in. w pracach: Dittmanna, 2008; Gajdy, 2001; Luli, 1999; Markiewicz, Zawadzkiego, 2005; Witkowskiej, 2002)

Przebieg procesu modelowania związany z możliwościami wykorzystania metod wymienionych w punkcie a) i b) zależy przede wszystkim od rodzaju luk. W szeregach z wahaniami sezonowymi występować mogą luki niesystematyczne lub systematyczne. Z lukami niesystematycznymi mamy do czynienia wtedy, gdy dysponujemy przynajmniej jedną informacją dla każdego podokresu. Do prognozowania brakujących danych w warunkach występowania luk niesystematycznych można wykorzystać w zasadzie wszystkie wymienione wyżej modele. Jedynie w przypadku modeli wyrównywania wykładniczego konieczne jest dysponowanie określoną liczbą danych początkowych.

Natomiast luki systematyczne występują wtedy, gdy nie mamy informacji przynajmniej dla jednego podokresu (miesiąca, dekady, dnia itp.) we wszystkich latach okresu estymacyjnego. Następstwem występowania tego rodzaju luk jest niemożność użycia części modeli, a w przypadku innych łączy się to ze znaczną komplikacją przebiegu procesu modelowania i prognozowania, polegającą na szacowaniu dużej liczby równań o tych samych własnościach predyktywnych.

W skład drugiej grupy metod wchodzi:

- a) klasyczne modele trendu (Cieślak, 1997; Czerwiński, Guzik, 1980; Zeliaś, 1997),
- b) modele adaptacyjne bez wahań sezonowych (np. Browna i Holta) (Dittmann, 2008; Pawłowski, 1973, 1982; Zeliaś, 1997; Zeliaś, Pawełek, Wanat, 2003),
- c) metody numeryczne (np. odcinkowa, łuków, wielomianowa Lagrange'a) (Fortuna, 2001; Grabiński, Wydymus, Zeliaś, 1979; Zboś, 1995).

Modelowanie i prognozowanie dla danych oczyszczonych z sezonowości przebiega w trzech etapach. W etapie pierwszym dokonuje się eliminacji wahań sezonowych ze zmiennej prognozowanej. Na podstawie tak otrzymanych danych przeprowadzane jest szacowanie parametrów odpowiednich modeli na podstawie, których następnie wyznaczane są prognozy inter- i ekstrapolacyjne. Końcowym etapem jest ich przemnożenie przez wskaźniki sezonowości.

Za pomocą niektórych metod numerycznych (np. odcinkowej i łuków) można wyznaczać tylko prognozy interpolacyjne. Prognozy ekstrapolacyjne wyznacza się dla „pełnego” szeregu (uzupełnionego o prognozy interpolacyjne) za pomocą np. modeli trendu lub modeli adaptacyjnych.

Z wieloletnich badań empirycznych zawartych m.in. w przywołanych wyżej pracach autorów, związanych ze stosowaniem różnych metod wynika, że dokładność prognoz otrzymywanych na podstawie szeregów z wahaniami sezonowymi zawierających luki w danych zależy od:

- rodzaju i skali wahań sezonowych,
- struktury harmonicznej zmiennych,
- rodzaju luk w danych,
- regularności w przebiegu zmiennych,
- liczby i rozmieszczenia luk względem minimów i maksimów sezonowych oraz od sekwencyjności luk,
- długości okresu jednostkowego,
- horyzontu prognozy.

Z doświadczeń, o których mowa, wynika także, że prognozy otrzymane na podstawie modeli o najlepszych własnościach predykcyjnych dość często nie charakteryzują się minimalnymi ocenami błędów. Ponadto minimalne oceny błędów prognoz interpolacyjnych otrzymuje się na podstawie innych modeli niż prognoz ekstrapolacyjnych.

W tabeli poniżej przedstawiona została syntetyczna charakterystyka wymienionych wyżej metod w aspekcie prognozowania brakujących danych. Dokonując tabelarycznego zestawienia metod wykorzystywanych w prognozowaniu w warunkach braku pełnej informacji, brano pod uwagę: rodzaj luk, możliwość wyznaczania albo obu rodzajów prognoz albo tylko prognoz interpolacyjnych, liczbę obserwacji początkowych oraz liczbę potencjalnych modeli, spośród których dokonuje się wyboru dla celów prognozowania.

## 2. PRZYKŁAD EMPIRYCZNY MODELOWANIE I PROGNOZOWANIA ZMIENNEJ EKONOMICZNEJ

Ilustracją przeprowadzonych powyżej rozważań wynikających z wieloletnich badań nad wykorzystaniem metod prognozowania brakujących informacji będzie przykład empiryczny dotyczący prognozowania miesięcznych wielkości skupu mleka (w tys. l) dla luk systematycznych występujących w II i IV kwartale każdego roku.

Przedział czasowy „próby” obejmował lata 1999-2004, przy czym rok 2004 będzie rokiem empirycznej weryfikacji prognoz. Luki w danych otrzymano przez „wymazanie” odpowiednich danych z pełnego szeregu.

Na rysunku 1 przedstawiono kształtowanie się wielkości skupu mleka według miesięcy.

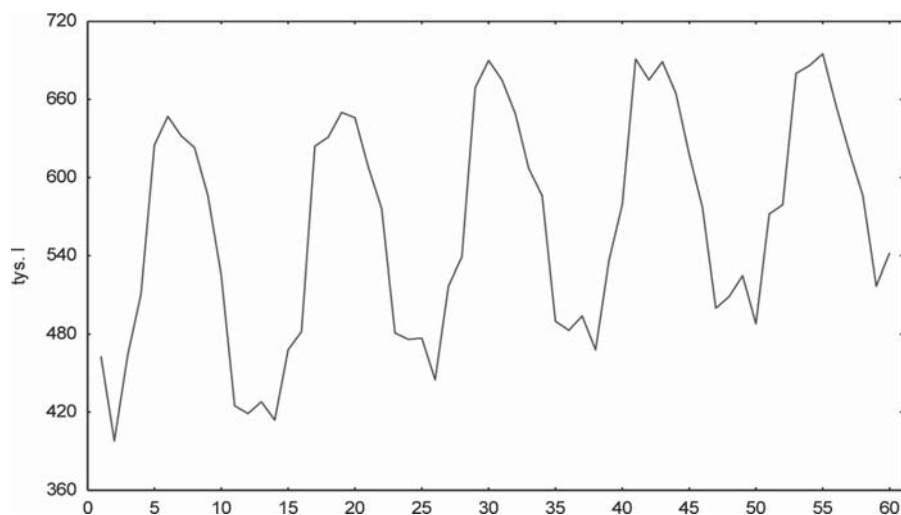
Tabela 1.

Syntetyczne zestawienie własności metod prognozowania brakujących danych sezonowych

Lp	Metody(modeli)	Rodaj luk	Rodzaj prog.	L.obs-początkowych	Liczba modeli
I	Dane oryginalne( z sezonowością)				
1	Klasyczne ze zmiennymi zerjedynkowymi	NS	I+E		1
2	Klasyczne z wielomianem trygonometrycznym.	NS+S	I+E		NS -1 S-zależy od układu luk i str. harmonicznej
3	Hierarchiczne	NS+S	I+E		NS -7 S-zależy od układu luk i str. harmonicznej.
4	Holta - Wintersa	NS	I+E	13	Kilkanaście wybranych z 729 (dla param. wygładzania zmieniających się co 0,1)
5	Sieci neuronowych	NS+S	I+E		Kilka (kilkanaście) wybranych z kilkuset
II	Dane oczyszczone z sezonowości				
	A. wyrównywania wykładniczego				
6	Browna -prosty	NS	I+E	1	Kilka wybranych z 9 (dla parametrów wygładzania zmieniających się co 0,1)
7	Browna -liniowy	NS	I+E	2	Kilka wybranych z 9 (dla parametrów wygładzania zmieniających się co 0,1)
8	Browna -kwadratowy	NS	I+E	3	Kilka wybranych z 9 (dla parametrów wygładzania zmieniających się co 0,1)
9	Holta - liniowy	NS	I+E	2	Kilkanaście wybr. z 81 (dla parametrów wygładzania zmieniających się co 0,1)
	B. Metody numeryczne				
10	Odcinkowa/MNK(WW)	NS+S	I	1	1
11	Łuków I/MNK(WW)	NS+S	I	2	1
12	Łuków II/MNK(WW)	NS+S	I	1	1
13	Wielomianowa Lagrange'a	NS+S	I+E	1	1
14	Wielomianowa Lagrange'a z opt. Czebyszewa	NS+S	I+E	1	1
15	MNK - metoda trendu	NS+S	I+E		Zależy od liczby postaci anal. trendu

Objaśnienia skrótów: NS - luki niesystematyczne; S-luki systematyczne; I- prognozy interpolacyjne; E - prognozy ekstrapolacyjne.

Źródło: opracowanie własne

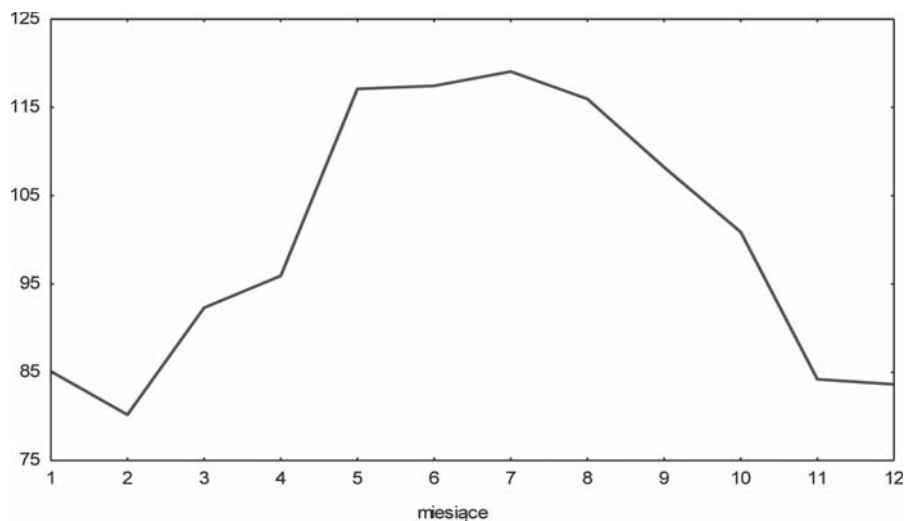


Rysunek 1. Skup mleka w latach 1999-2004

Źródło: opracowanie własne

Z powyższego rysunku wynika, że badana zmienna charakteryzuje się występowaniem dość silnych wahań sezonowych.

Natomiast na rysunku 2 przedstawiono kształtowanie się wskaźników sezonowości. Informacje zawarte na rysunku wskazują, że zmienna prognozowana charakteryzuje się dość silnym natężeniem wahań sezonowych, ponieważ amplituda ocen wskaźników sezonowości wynosi ok. 40 punktów procentowych.



Rysunek 2. Kształtowanie się wskaźników sezonowości

Źródło: opracowanie własne

W celu obliczenia udziałów poszczególnych składowych harmonicznych w wyjaśnieniu wariacji sezonowej najpierw został oszacowany model szeregu czasowego z liniowym trendem i periodycznym składnikiem sezonowym a następnie dokonano eliminacji trendu.

W tabeli 1 zestawione zostały udziały procentowych składowych harmonicznych oraz harmonik w wyjaśnianiu wariacji sezonowej zmiennej prognozowanej (składowe sinusoidalne oznaczono przez  $S_i$  a kosinusoidalne przez  $C_i$ ).

Tabela 2.

Procentowe udziały składowych harmonicznych sinuso- i kosinusoidalnych oraz harmonik w wyjaśnianiu wariacji sezonowej (w %)

Składowe sinusoidalne	Odsetki	Składowe kosinusoidalne	Odsetki	Suma
$S_1$	18,17	$C_1$	75,89	94,05
$S_2$	0	$C_2$	0,8	0,8
$S_3$	2,06	$C_3$	0,10	2,17
$S_4$	0,09	$C_4$	0,03	0,12
$S_5$	2,20	$C_5$	0,28	2,48
–	–	$C_6$	0,38	0,38
Suma	22,53		77,47	

Źródło: obliczenia własne

Z informacji zawartych w tabeli wynika, że najwyższym udziałem w wyjaśnianiu ogólnej wariacji sezonowej charakteryzuje się składowa kosinusoidalna o cyklu 12 miesięcznym (75,89%) a następnie składowa sinusoidalna o takiej samej długości cyklu (18,17%). Pierwsza harmonika wyjaśnia zatem w ponad 94 procentach całkowitą zmienność sezonową. Ponadto wskaźniki udziałów jeszcze tylko dwóch składowych sinusoidalnych ( $S_3$  i  $S_5$ ) o cyklach wynoszących odpowiednio: 4 oraz 2,4 miesiąca przekraczają 2 procent. Łączna wielkość udziałów pozostałych 7 składowych wynosi 1,69%, przy czym dla wszystkich składowych wskaźniki udziałów są niższe od 1 procenta.

Rozpatrywany będzie jeden wariant luk systematycznych, będą one występować w II i IV kwartale każdego roku. Luki te otrzymano przez „wymazanie” odpowiednich danych z pełnego szeregu.

W prognozowaniu brakujących danych zostaną wykorzystane następujące metody:

I. Dla danych oryginalnych (z sezonowością):

- 1) klasyczny model szeregu czasowego z liniowym trendem i periodycznym składnikiem sezonowym opisanym za pomocą wielomianu trygonometrycznego,
- 2) modele hierarchiczne,
- 3) model Holta –Wintersa,
- 4) sieci neuronowe,

II. Dla danych oczyszczonych:

- 5) metoda trendu(MNK),
- 6) model prosty Browna,
- 7) model liniowy Browna,
- 8) model kwadratowy Browna,
- 9) model liniowy Holta (dwuparametryczny),
- 10) metoda odcinkowa/MNK
- 11) metoda łuków I/MNK,
- 12) metoda łuków II/MNK,
- 13) metoda Lagrange'a dla 3 i 4 węzłów interpolacyjnych rozmieszczonych w równych odległościach ,
- 14) metoda Lagrange'a z optymalizacją Czebyszewa z 3 i 4 węzłami.

Zanim przejdziemy do zestawienia tabelarycznego wyników prognozowania inter- i ekstrapolacyjnego dla najlepszych predyktorów, przedstawimy kolejno wyniki modelowania oraz prognozowania dla metod wymienionych w punktach: 1,3,6-9 oraz 11-14.

W przypadku metody klasycznej z wielomianem trygonometrycznym, występowanie luk systematycznych oznacza m.in., że macierz współczynników korelacji składowych harmonicznych przestaje być macierzą jednostkową. W naszym przypadku przyjęła ona postać:

Tabela 3.

Oceny współczynników korelacji liniowej składowych harmonicznych dla rozpatrywanego wariantu luk w danych

	S <sub>1</sub>	S <sub>2</sub>	S <sub>3</sub>	S <sub>4</sub>	S <sub>5</sub>	C <sub>1</sub>	C <sub>2</sub>	C <sub>3</sub>	C <sub>4</sub>	C <sub>5</sub>
S <sub>1</sub>	1,00	0,00	0,50	0,00	-0,50	-0,61	0,00	0,61	0,00	0,00
S <sub>2</sub>	0,00	1,00	0,00	0,00	0,00	0,00	0,76	0,00	1,00	0,00
S <sub>3</sub>	0,50	0,00	1,00	0,00	0,50	-0,61	0,00	0,00	0,00	0,61
S <sub>4</sub>	0,00	0,00	0,00	1,00	0,00	0,00	-0,65	0,00	0,00	0,00
S <sub>5</sub>	-0,50	0,00	0,50	0,00	1,00	0,00	0,00	-0,61	0,00	0,61
C <sub>1</sub>	-0,61	0,00	-0,61	0,00	0,00	1,00	0,00	0,25	0,00	0,25
C <sub>2</sub>	0,00	0,76	0,00	-0,65	0,00	0,00	1,00	0,00	0,76	0,00
C <sub>3</sub>	0,61	0,00	0,00	0,00	-0,61	0,25	0,00	1,00	0,00	0,25
C <sub>4</sub>	0,00	1,00	0,00	0,00	0,00	0,00	0,76	0,00	1,00	0,00
C <sub>5</sub>	0,00	0,00	0,61	0,00	0,61	0,25	0,00	0,25	0,00	1,00

Źródło: obliczenia własne

Z powyższej tabeli wynika, że 14 elementów leżących powyżej i poniżej głównej przekątnej jest różnych od zera, przy czym cztery z nich są ujemne. Otrzymanie elementów różnych od zera może oznaczać wystąpienie liniowych kombinacji składowych.

Wyboru statystycznie istotnych składowych harmonicznych dokonano za pomocą metody *selekcji a priori*, będącej jedną z procedur regresji krokowej. Polega ona, najogólniej rzecz biorąc, na wprowadzaniu do modelu kolejnej składowej (oprócz pierwszej) najsilniej skorelowanej z resztami modelu otrzymanego w poprzednim etapie. Opis tej metody można znaleźć m.in. w pracy Drapera, Smitha (1973).

W trakcie wyboru składowych okazało się, że kombinacje liniowe tworzą: para składowych:  $S_3$  i  $C_1$  oraz trójka:  $S_5$ ,  $C_3$  oraz  $C_5$ . Oznacza to, że do modelu można wprowadzić tylko po jednej zmiennej wchodzącej w skład pary lub trójki. Tym samym więc szacowanych będzie 6 wersji modelu – zostały one zestawione w poniższej tabeli. Zawiera ona także oceny błędów prognoz interpolacyjnych oraz prognoz ekstrapolacyjnych dla horyzontu  $h=3, 6, 9$  i  $12$  miesięcy.

Tabela 4.

Oceny parametrów struktury stochastycznej modeli z wielomianem trygonometrycznym oraz błędy prognoz inter- i ekstrapolacyjnych

Lp	Składowe	$R^2$	$S_e$	$V_{Se}(\%)$	MAPE [%]				
			(tys.l)		inter.	h=3	h=6	h=9	h=12
1	S145C1	0,98	13,05	2,37	5,42	3,65	3,53	3,66	4,22
2	S14C13	0,98	13,05	2,37	6,97	3,65	5,17	4,76	4,53
3	S14C15	0,98	13,05	2,37	6,38	3,65	5,17	4,76	5,32
4	S1345	0,98	13,05	2,37	22,48	3,65	10,85	8,54	12,98
5	S134C3	0,98	13,05	2,37	21,47	3,65	10,18	8,1	12,29
6	S134C5	0,98	13,05	2,37	22,17	3,65	10,64	8,4	12,81

Źródło: obliczenia własne

W kolumnie drugiej podane zostały zestawy składowych wchodzących w skład poszczególnych wersji modelu. Z informacji zawartych w kolejnych trzech kolumnach wynika, że wersje te charakteryzują się identycznymi ocenami: współczynników determinacji ( $R^2$ ), odchyłeń standardowych składników losowych ( $S_e$ ) oraz współczynników zmienności losowej ( $V_{Se}$ ). Oznacza to, że są one nierozróżnialne z punktu widzenia własności predykcyjnych. Mogą się natomiast różnić ocenami błędów prognoz inter- lub/i ekstrapolacyjnych. Oceny względnych błędów prognoz (MAPE) podane zostały w ostatnich pięciu kolumnach.

W kształtowaniu się ocen błędów prognoz interpolacyjnych w sposób zdecydowany wyodrębnić można dwie grupy predyktorów. Do pierwszej grupy należą predyktory wymienione w wierszach od drugiego do czwartego. Charakteryzują się one tym, że zawierają jednocześnie składowe  $C_1$  oraz  $S_1$  tj. składowe o największych udziałach w wyjaśnianiu wariacji sezonowej. Oceny błędów prognoz kształtują się w granicach od 5,42 do 6,97 procent. Oceny błędów dla pozostałych trzech predyktorów, zaliczonych do grupy drugiej, przyjmują wartości znacznie wyższe i zawierają się w przedziale



21,47-22,48 procent. Spośród pary składowych o najwyższych udziałach występuje tylko składowa  $S_1$ .

W przypadku prognoz ekstrapolacyjnych wyodrębnienie tych grup jest możliwe dla horyzontu prognozy wynoszącego co najmniej 6 miesięcy, ponieważ dla  $h=3$  otrzymano identyczne oceny błędów wynoszące 3,65 procent. Dokładność predyktorów należących do grupy pierwszej jest od około 4 do około 8 punktów procentowych wyższa od dokładności predyktorów należących do grupy drugiej.

Większa dokładność predyktorów zaliczonych do grupy pierwszej wynika, podobnie jak to miało miejsce w przypadku prognoz interpolacyjnych, z faktu, że zawierają one składową kosinusoidalną o cyklu dwunastomiesięcznym ( $C_1$ ). Udział jej w wyjaśnieniu wariacji sezonowej dla pełnych danych, o czym była mowa wyżej, przekracza 75 procent.

W tabeli 4 zestawiono oceny błędów prognoz otrzymanych za pomocą metod numerycznych. Z uwagi na fakt, że za pomocą metody odcinkowej oraz metody łuków można wyznaczyć tylko prognozy interpolacyjne, prognozy ekstrapolacyjne wyznaczono MNK szacując trendy liniowe na podstawie szeregów z lukami uzupełnionych o prognozy interpolacyjne.

Tabela 5.

Oceny błędów prognoz inter- i ekstrapolacyjnych otrzymanych za pomocą metod numerycznych

Metoda	MAPE [%]				
	interpol.	h=3	h=6	h=9	h=12
Odcinkowa/MNK	1,88	5,12	3,45	2,68	4,13
Łuków I/MNK	1,41	5,99	4,73	4,18	6,18
Łuków II/MNK	2,04	5,34	3,75	2,94	4,51
Lagrange'a WP 3	3,29	2,26	5,57	8,48	8,66
Lagrange'a WP 4	4,75	33,87	40,09	48,2	58,34
Lagrange'a z opt. Czebyszewa 3	3,89	7,8	5,71	4,24	4,99
Lagrange'a z opt. Czebyszewa 4	3,04	15,46	16,97	19,62	24,63

Źródło: obliczenia własne

W przypadku metody Lagrange'a z 4 węzłami interpolacyjnymi rozmieszczonymi w równych odległościach wyraźnie widoczny jest wynik oddziaływania tzw. efektu Rungego, polegającego na pogorszeniu się jakości interpolacji wielomianowej wraz ze zwiększeniem liczby węzłów. W naszym przypadku odnosi się to do prognoz ekstrapolacyjnych.

Dla metody Lagrange'a z optymalizacją Czebyszewa dla 4 węzłów efekt ten występuje w skali znacznie słabszej, ale i tak oceny błędów są zdecydowanie wyższe niż ma to miejsce dla 3 węzłów.

W tabeli 5 zestawione zostały modele (wersje modeli) charakteryzujące się najniższymi wartościami ocen błędów prognoz inter- i ekstrapolacyjnych. Dlatego też niektóre metody (modele) zwłaszcza adaptacyjne są reprezentowane przez dwa a nawet trzy predyktory, różniące się stałymi wygładzenia lub w przypadku modelu ze składowymi harmonicznymi zestawem składowych harmonicznymi. W kolumnie drugiej podane zostały liczby modeli, spośród których zostały one wybrane. „Tłustym” drukiem zaznaczone zostały predyktory o najniższych ocenach: mierników wartości predyktywnych, błędów prognoz interpolacyjnych lub ekstrapolacyjnych. Ponadto w dwóch ostatnich kolumnach zestawione zostały miejsca (rangi) poszczególnych predyktorów uszeregowane według rosnących ocen błędów prognoz inter- i ekstrapolacyjnych.

W pierwszych sześciu wierszach zestawione zostały predyktory na podstawie, których wyznaczano prognozy dla danych oryginalnych (z sezonowością). Z analizy mierników charakteryzujących własności predyktywne wynika, że najmniej korzystne ich oceny otrzymano dla sieci neuronowych.

Oceny błędów prognoz interpolacyjnych dla pięciu spośród sześciu predyktorów przyjęły wartości z przedziału 5,42 do 7,08 procent. Znacznie *in minus* od nich odbiega ocena tego błędu dla predyktora hierarchicznego – jest ona ponad dwukrotnie wyższa. Zróżnicowanie ocen błędów prognoz ekstrapolacyjnych jest znacznie niższe. Dla pięciu predyktorów zawierają się one w przedziale od 3,27 do 6,12 procent. Jedynie w przypadku modelu hierarchicznego błąd ten przekracza 8 procent. Oznacza to tym samym, że sekwencyjny układ systematycznych luk w danych, obejmujący miesiące wchodzące w skład 2 i 4 kwartału, najsilniej oddziaływał w przypadku modeli hierarchicznych. Najbardziej stabilny z punktu widzenia względnych ocen obu rodzajów błędów okazał się model klasyczny z wielomianem trygonometrycznym zajmując odpowiednio 1 i 2 miejsce wśród modeli szacowanych dla danych oryginalnych (sezonowością). Następnym w kolejności okazał się model Holta-Wintersa.

Z analizy rang metod dla błędów prognoz interpolacyjnych wynika, że modele dla danych oryginalnych zajmują 6 spośród 7 miejsc w dole tabeli – zostały one jedynie przedzielone liniowym modelem Holta sklasyfikowanym na 20 miejscu.

Natomiast w przypadku prognoz ekstrapolacyjnych mamy do czynienia z bardzo dużym zróżnicowaniem miejsc (rang). Najlepszą metodą spośród 22 analizowanych predyktorów okazała się sieć neuronowa z 13 warstwami ukrytymi. W przypadku modeli sieci neuronowych (S3MLP) zaobserwowano największe zróżnicowanie rang ocen błędów. Otrzymano dla niego najniższą oceną błędu prognoz ekstrapolacyjnych i najwyższą oceną błędu prognoz interpolacyjnych. W przypadku modelu MLP1 różnica rang była niewielka, ale został on sklasyfikowany na odpowiednio: na 18 i 16 miejscu. Prezentowane w tabeli modele sieci neuronowych, wybrane spośród 400 wcześniej oszacowanych, charakteryzowały się względnie stabilną relacją błędów prognoz inter- i ekstrapolacyjnych. W wielu przypadkach modele o niższych ocenach błędów prognoz interpolacyjnych dawały bardzo duże błędy prognoz ekstrapolacyjnych i odwrotnie. Predyktor klasyczny z wielomianem trygonometrycznym został sklasyfikowany na 8 miejscu. Miejsce 15 zajął model Holta – Wintersa charakteryzujący się najniższą oceną

Tabela 6.

Zestawienie wyników modelowania i prognozowania inter- i ekstrapolacyjnego oraz rangowania metod

Metody(modele)	L. mod.	Stałe wygładzania			MAPE wart.wyr. lub $V_{se}^{**}$	Błędy prognoz		Ranga metody prognozowania	
		$\alpha$	$\beta$	$\gamma$		interpol.	Ekstrapol. h=12	inter.	ekstr.
Klasyczne z wiel. tryg.(s145c1)	6				2,37**	5,42	4,22	16	8
Hierarchiczny(H43)*	14				2,34**	15,02	8,17	22	18
Holta Wintersa	13	0,1	0,1	0,5	<b>2,26</b>	<b>5,56</b>	5,51	17	15
Holta Wintersa	13	0,3	0,3	0,8	4,38	6,58	<b>4,96</b>	19	12
Sieci neuronowe (S3-MLP3)*	6				<b>5,61**</b>	7,08	<b>3,27</b>	21	1
Sieci neuronowe (S2-MLP1)*	6				6,58**	<b>5,8</b>	6,12	18	16
Browna -prosty	9	0,9			<b>2,21</b>	3,63	5,33	11	14
Browna -prosty	9	0,4			2,66	<b>3,32</b>	4,51	9	10
Browna -prosty	9	0,3			2,78	3,4	<b>4,42</b>	10	9
Browna -liniowy	9	0,2			<b>2,82</b>	<b>2,95</b>	3,75	7	6
Browna -liniowy	9	0,1			2,97	3,66	<b>3,56</b>	12	2
Browna -kwadratowy	8	0,1			<b>3,09</b>	4,14	<b>3,7</b>	15	5
Browna -kwadratowy	8	0,2			3,59	<b>2,88</b>	10,9	4	21
Holta-liniowy	18	0,9	0,9		<b>3,64</b>	<b>4,11</b>	13,03	14	22
Holta-liniowy	18	0,3	0,4		8,27	6,83	<b>8,79</b>	20	20
Odcinkowa/MNK	1					1,88	4,13	2	7
Łuków I/MNK	1					1,41	6,18	1	17
Łuków II/MNK	1					2,04	4,51	3	11
Lagrange'a WP 3	2					3,29	8,66	8	19
Lagrange'a z opt. Czebyszewa 3	2					3,89	4,99	13	13
MNK-trend liniowy	1				2,78**	2,95	3,66	6	3
MNK-trend wykładn. o stałej stopie wzrostu	1				2,82**	2,95	3,69	5	4

Źródło: opracowanie własne oraz (\*) Markiewicz, Zawadzki, (2005)  
 (\*\*) współczynnik zmienności losowej

błędu interpolacyjnego. Natomiast predyktor hierarchiczny sklasyfikowany został na 18 miejscu.

Obecnie przechodzimy do analizy własności predykcyjnych mierników dokładności prognoz dla danych oczyszczonych z sezonowości. Najpierw analizie poddane zostaną predyktory oparte na modelach Browna i liniowym modelu Holta, a następnie

predyktory wykorzystujące metody numeryczne. Dla trzech modeli wyrównania wykładniczego otrzymano po 2 predyktory charakteryzujące się minimalnymi ocenami własności predykcyjnych i jednego rodzaju prognoz. Dla prostego modelu Holta były to nawet 3 wersje modelu różniące się stałymi wygładzenia.

Zróżnicowanie błędów wartości wyrównanych dla 8 z 9 modeli było niewielkie – oceny te mieściły się w przedziale od 2,21 do 3,64 procent. Znacznie wyższą ocenę otrzymano dla liniowego modelu Holta o stałych wygładzenia  $\alpha$  i  $\beta$  wynoszących odpowiednio: 0,3 i 0,4. Dla modelu tego otrzymano minimalną ocenę błędu prognoz ekstrapolacyjnych. W przypadku prognoz interpolacyjnych dla 8 wersji modeli oceny te mieściły się w przedziale 2,95 do 4,14 procent. Natomiast dla wymienionego wyżej modelu błąd ten był wyższy od najwyższego o blisko 2,7 punktu procentowego.

Oceny błędów prognoz ekstrapolacyjnych otrzymanych na podstawie prostych i liniowych modeli Browna kształtują się w przedziale od 3,56 do 5,33 procent. Dla modelu Holta z minimalną oceną błędu tego rodzaju prognoz, ocena ta jest o około 3,5 punktu procentowego wyższa. Oceny błędów prognoz przekraczające 10 procent otrzymano dla modelu kwadratowego i modelu Holta, charakteryzujących się minimalnymi ocenami błędów prognoz interpolacyjnych.

Z kształtowania się rang błędów prognoz interpolacyjnych wynika, że mieszczą się one na ogół w grupie środkowej od 7 – 15 miejsca. Wysoką czwartą pozycję zajmuje kwadratowy model Browna dla stałej wygładzenia równej 0,2 a dwudziestą model Holta o minimalnej ocenie błędu prognoz ekstrapolacyjnych.

W przypadku rang ocen błędów prognoz ekstrapolacyjnych występuje większe zróżnicowanie – od 2, 5 i 6 miejsca dla liniowego i kwadratowego modelu Browna do 21 i 22 miejsca dla modeli Holta. Zwraca uwagę znaczne zróżnicowanie rang metod dla błędów prognoz inter- oraz ekstrapolacyjnych. Z największym zróżnicowaniem mamy do czynienia w przypadku modelu kwadratowego Browna sklasyfikowanym odpowiednio: na 4 i 21 miejscu. W przypadku liniowego i kwadratowego modelu Browna o minimalnych ocenach błędów prognoz różnica rang wynosi 10. Oznacza to, że inne predyktory powinny być wykorzystane w budowie prognoz interpolacyjnych oraz inne w przypadku prognozowania ekstrapolacyjnego.

Obecnie przechodzimy do analizy błędów prognoz oraz do rankingu metod numerycznych. W przypadku metody odcinkowej oraz łuków bezpośrednio za ich pomocą otrzymano prognozy interpolacyjne. Prognozy ekstrapolacyjne wyznaczono metodą MNK na podstawie liniowych równań trendów. W przypadku metody Lagrange'a wybrano metodę z trzema węzłami rozmieszczonymi proporcjonalnie lub zgodnie z metodą optymalizacji Czebyszewa. Prognozy ekstrapolacyjne wyznaczone na ich podstawie charakteryzują się brakiem omówionego wcześniej efektu Rungego.

Zróżnicowanie ocen błędów prognoz interpolacyjnych jest niewielkie i waha się od 1,42 do 3,89 procent, przy czym metody: odcinkowa i łuków zostały sklasyfikowane na trzech pierwszych miejscach. Natomiast oba warianty metody Lagrange'a odpowiednio: na 8 i 13 miejscu.

Oceny błędów prognoz ekstrapolacyjnych mieszczą się w przedziale od 4,13 dla metody odcinkowej/MNK do 8,66 dla metody Lagrange'a z 3 węzłami rozmieszczonymi proporcjonalnie. W klasyfikacji metod zajmują one miejsca: 7, 11, 13, 17 i 19. Z najmniejszą różnicą miejsc dla predyktorów o najniższym błędzie mamy do czynienia w przypadku metody odcinkowej/MNK. Natomiast najbardziej stabilny okazał się predyktor Lagrange'a z optymalizacją Czebyszewa, zajmując w obu klasyfikacjach 13 miejsce z ocenami błędów niższymi odpowiednio: od 4 i 5 procent.

Z informacji zawartych w dwóch ostatnich wierszach tabeli wynika, że oceny błędów prognoz otrzymanych na podstawie modeli trendu: liniowego i wykładniczego są identyczne dla prognoz interpolacyjnych oraz prawie takie same dla prognoz interpolacyjnych. W przypadku pierwszego rodzaju prognoz predyktory te zajmują miejsca: 6 i 5 a dla ekstrapolacyjnych: 3 i 4.

### 3. PODSUMOWANIE

Z przedstawionych wyżej analiz mierników charakteryzujących proces modelowania i prognozowania wynika, że różne metody (lub różne warianty (wersje)) charakteryzują się minimalnymi ocenami parametrów (własności predyktywne) oraz prognoz inter- i ekstrapolacyjnych.

Oznacza to, że proces prognozowania *ex ante* zmiennych charakteryzujących się występowaniem co najmniej dość silnym natężeniem skali wahań sezonowych, powinien być poprzedzony analizą błędów zarówno prognoz interpolacyjnych jak i przede wszystkim błędów prognoz *ex post*. Predyktory charakteryzujące się minimalnymi ocenami błędów tych ostatnich prognoz powinny być wykorzystane do budowy prognoz *ex ante*. Nie należy wykluczyć wykorzystania metod „łamanych” tzn. szacowania modeli wykorzystanych w budowie prognoz *ex post*, które zostaną następnie będą wykorzystane do budowy prognoz *ex ante*.

Dokonując próby podsumowania rozważań przeprowadzonych w pracy możemy stwierdzić, że metody statystyczno-ekonometryczne mogą być z powodzeniem wykorzystywane w prognozowaniu brakujących danych zmiennych ekonomicznych o dość silnym natężeniu wahań sezonowych tzn. charakteryzujących się rozstępem skrajnych ocen wskaźników sezonowości nie przekraczających na ogół 40-60 punktów procentowych.

Zachodniopomorski Uniwersytet Technologiczny w Szczecinie

### LITERATURA

- [1] Cieślak M.(red.), (1997), *Prognozowanie gospodarcze*, PWN, Warszawa.
- [2] Czerwiński Z., Guzik B., (1980), *Prognozowanie ekonometryczne*, PWE, Warszawa.
- [3] Dittmann P., (2008), *Prognozowanie w przedsiębiorstwie. Metody i ich zastosowanie*, Walters Kluwer Polska.

- [4] Dittmann P., Szabela-Pasierbińska E., Dittmann I., Szpulak A., (2011), *Prognozowanie w zarządzaniu sprzedażą i finansami przedsiębiorstwa*, Walters Kluwer Polska-Oficyna.
- [5] Draper N. R., Smith H., (1973), *Analiza regresji stosowana*, PWN, Warszawa.
- [6] Fortuna Z., (2001), *Metody numeryczne*, WNT, Warszawa.
- [7] Gajda J., (2001), *Prognozowanie i symulacje a decyzje gospodarcze*, C.H.Beck, Warszawa.
- [8] Grabiński T., Wydymus S., Zeliaś A., (1979), *Z badań nad metodami predykcji brakujących informacji*, Zeszyty Naukowe AE w Krakowie Nr 114, Kraków.
- [9] Grabiński T., Wydymus S., Zeliaś A., (1981), *Modele ekonometryczne w procesie prognozowania*, Akademia Ekonomiczna w Krakowie, Kraków, str. 74-75.
- [10] Guzik B., (2003), *Wstęp do teorii prognozowania i symulacji*, Wydawnictwo Akademii Ekonomicznej w Poznaniu, Poznań.
- [11] Lula P., (1999), *Jednokierunkowe sieci neuronowe w modelowaniu zjawisk ekonomicznych*, Wyd. AE w Krakowie, Kraków.
- [12] Markiewicz. A, Zawadzki J., (2005), *Dokładność prognoz skupu mleka dla systematycznych luk w danych*, Folia Universitatis Agriculturae Stetinensis, seria: Oeconomica 245, s. 411-416.
- [13] Pawłowski Z., (1973), *Prognozy ekonometryczne*, PWN, Warszawa.
- [14] Pawłowski Z., (1982), *Zasady predykcji ekonometrycznej*, PWN, Warszawa.
- [15] Szmuksta-Zawadzka M., Zawadzki J., (1995), *O metodzie prognozowania brakujących informacji dla danych sezonowych*, Przegląd Statystyczny, nr 3-4, str. 377-385.
- [16] Szmuksta-Zawadzka M., Zawadzki J., (2000), *On hierarchic models of time series with seasonal fluctuations*, Wyd. UMK, Toruń, str. 25-30.
- [17] Szmuksta-Zawadzka M., Zawadzki J., (2002a), *Forecasting on basis of time series hierarchic models with variable seasonality*, [w:] Dynamics Econometrics Models No. 5, Toruń.
- [18] Szmuksta-Zawadzka M., Zawadzki J., (2002b), *Prognozowanie na podstawie hierarchicznych modeli szeregu czasowego z lukami w danych*, w: Analiza szeregów czasowych na początku XXI wieku, Toruń, str. 93-102.
- [19] Szmuksta-Zawadzka M., Zawadzki J., (2002c), *Hierarchiczne modele szeregów czasowych z wahaniami sezonowymi. Budowa. Estymacja. Prognozowanie.*, [w:] Przestrzenno-czasowe modelowanie i prognozowanie zjawisk gospodarczych, Wyd. AE, Kraków, str. 193-204.
- [20] Szmuksta-Zawadzka M., Zawadzki J., (2003), *O modelach hierarchicznych dla danych dekadowych z wahaniami sezonowymi*, [w:] Dynamiczne modele ekonometryczne, Toruń.
- [21] Witkowska D., (2002), *Sztuczne sieci neuronowe i metody statystyczne*, Wyd. C.H. Beck, Warszawa.
- [22] Zawadzki J.(red.), (1999), *Ekonometryczne metody predykcji dla danych sezonowych w warunkach braku pełnej informacji*, Uniwersytet Szczeciński, Szczecin.
- [23] Zawadzki J.(red.), (2003), *Zastosowanie hierarchicznych modeli szeregów czasowych w prognozowaniu zmiennych ekonomicznych z wahaniami sezonowymi*, Akademia Rolnicza, Szczecin.
- [24] Zboś D., (1995), *Metody numeryczne*, Politechnika Krakowska, Kraków.
- [25] Zeliaś A., (1987), *Estymacja parametrów równania regresji liniowej w warunkach niepełnej informacji*, Zeszyty Naukowe AE w Krakowie, Nr 243, Kraków.
- [26] Zeliaś A., (1997), *Teoria prognozy*, PWE, Warszawa.
- [27] Zeliaś A., Pawełek B., Wanat S., (2003), *Prognozowanie ekonomiczne. Teoria, przykłady, zadania.*, PWN, Warszawa.

Z BADAŃ NAD METODAMI PROGNOZOWANIA NA PODSTAWIE NIEKOMPLETNYCH  
SZERGÓW CZASOWYCH Z WAHANIAMI OKRESOWYMI (SEZONOWYMI)

S t r e s z c z e n i e

Praca została poświęcona syntetycznemu omówieniu wyników wieloletnich badań autorów nad zastosowaniami metod prognozowania w warunkach braku pełnej informacji w szeregach czasowych z wahaniami sezonowymi. Rozważania odnoszą się do dwóch rodzajów luk w danych: systematycznych i niesystematycznych. Z lukami systematycznymi mamy do czynienia wtedy, gdy nie są dostępne informacje liczbowe przynajmniej o jednym podokresie w całym przedziale czasowym „próby”. Rozpatrywane będą metody prognozowania zarówno dla danych oryginalnych (z sezonowością) jak i danych, z których wyeliminowano wahania sezonowe. Egzemplifikacją rozważań o charakterze teoretycznym będzie przykład empiryczny.

**Słowa kluczowe:** szeregi czasowe, wahania sezonowe, brakujące dane, prognozowanie

STUDIES OF METHODS APPLIED TO FORECASTING INCOMPLETE DATA IN SEASONAL  
TIME SERIES

A b s t r a c t

This work presents discussion about results of long-term of authors research on applications of different forecasting methods in condition of lack of full information. There will be considered two types of gaps in data: systematic and unsystematic. The systematic gaps in data are only when we have not any information about at least one sub-period in the whole of analyzed data. There will be presented two types of methods applied to time series with and without seasonal component. Exemplification of theoretical considerations will be an empirical example.

**Key words:** time series, seasonal fluctuations, missing data, forecasting