



Maria Szmuksta-Zawadzka

Zachodniopomorski Uniwersytet Technologiczny
w Szczecinie
Studium Matematyki

Jan Zawadzki

Zachodniopomorski Uniwersytet Technologiczny
w Szczecinie
Katedra Zastosowań Matematyki w Ekonomii
jan.zawadzki@zut.edu.pl

PROGNOZOWANIE BRAKUJĄCYCH DANYCH DLA SZEREGÓW O WYSOKIEJ CZĘSTOTLIWOŚCI OCZYSZCZONYCH Z SEZONOWOŚCI

Streszczenie: W pracy przedstawione zostało wykorzystanie wybranych modeli adaptacyjnych w prognozowaniu zmiennych o bardzo wysokiej częstotliwości obserwowania, na podstawie szeregów z lukami niesystematycznymi, z których wyeliminowano dwa lub trzy rodzaje sezonowości. Egzemplifikacją rozważań teoretycznych stanowi przykład empiryczny, dotyczący kształtowania się zapotrzebowania na moc energetyczną w okresach godzinnych w aglomeracji A.

Słowa kluczowe: prognozowanie, dane o wysokiej częstotliwości, złożona sezonowość, wyrównywanie wykładnicze, luki w danych.

Wprowadzenie

Z rozważań przedstawionych w pracy [Szmuksta-Zawadzka, Zawadzki, 2015] wynika, że modele adaptacyjne mogą być efektywnym narzędziem prognozowania na podstawie kompletnych szeregów czasowych zmiennych ekonomicznych o wysokiej częstotliwości, oczyszczonych z dwóch lub trzech rodzajów sezonowości. Uzyskanie w tym względzie zachęcających wyników dla pełnych danych stało się asumptem do rozszerzenia rozważań na szeregi z lukami niesystematycznymi, tzn. takimi, gdy dysponuje się przynajmniej jedną informacją o każdym podokresie wahań składowych złożonej sezonowości. Procedura modelowania i prognozowania polega na wykorzystaniu do budowy prognoz modeli Holta-Wintersa dla danych oczyszczonych z sezonowości rocznej i tygodniowej. Natomiast w przypadku wyeliminowania także wahań o cyklu

dobowym stosuje się modele Browna i Holta. Prognozy wyjściowe otrzymuje się na podstawie predyktorów dla danych oczyszczonych. Natomiast prognozy końcowe są sumami (modele addytywne) lub iloczynami (modele multiplikatywne) prognoz wyjściowych i składników albo wskaźników sezonowości. Modele te są przystosowane dla danych oczyszczonych o wysokiej częstotliwości, są znanymi w literaturze modelami dla wahań o niezłożonej sezonowości oraz dla zmiennych, w których wahania te nie występują. Ich zapisy można znaleźć w pracach badacza [Dittman, 2006; Pawłowski 1973; Zeliaś, Pawełek, Wanat 2003].

Z uwagi na to, że zarówno zapisy ogólne modeli w postaci addytywnej oraz multiplikatywnej, jak i ich zapisy analityczne podane zostały w pracy [Szmuksta-Zawadzka, Zawadzki, 2015], ograniczymy się do prezentacji modeli w postaci addytywnej.

Zapis ogólny modelu addytywnego ze złożoną sezonowością jest następujący:

$$Y_{t(a)} = P^{(a)}(t) + M^{(a)}(t) + D^{(a)}(t) + G^{(a)}(t) + V_{t(a)} \quad (1)$$

gdzie:

$P^{(a)}(t)$ – trend

$M^{(a)}(t)$ – składniki sezonowości o cyklu 12-miesięcznym

$D^{(a)}(t)$ – składniki sezonowości o cyklu 7-dniowym

$G^{(a)}(t)$ – składniki sezonowości o cyklu 24-godzinnym.

Oznaczając $Y_{t(a)}^{***}$ przez zmienną prognozowaną, z której wyeliminowano składniki sezonowości o cyklu rocznym, tygodniowym i dobowym możemy zapisać analitycznie modele Browna i Holta.

Równanie prostego modelu Browna (a_B) wyraża się wzorem:

$$m_t^{***} = \alpha Y_{t(a)}^{***} + (1 - \alpha) m_{t-1}^{***} \quad (2)$$

Predyktor oparty na tym modelu przyjmuje postać:

$$\Pi_{a_B}^{***} = m_{t_0}^{***} \quad (3)$$

gdzie t_0 jest okresem wyjściowym budowy prognoz.

Prognozę końcową otrzymuje się na podstawie predyktora:

$$\Pi_{a_B} = \Pi_{a_B}^{***} + M^{(a)}(t) + D^{(a)}(t) + G^{(a)}(t) \quad (4)$$

Równania liniowego modelu Holta (a_H) są następujące:

$$m_t^{***} = \alpha Y_{t(a)}^{***} + (1 - \alpha)(m_{t-1}^{***} + \delta_{1t-1}^{***}) \quad (5)$$

$$\delta_{1t}^{***} = \beta(m_t^{***} - m_{t-1}^{***}) + (1 - \beta)\delta_{1t-1}^{***} \quad (6)$$

Predyktory: wyjściowy dla horyzontu prognozy $h = 1, 2, \dots$ oraz końcowy przyjmują postać:

$$\Pi_{a_H}^{***} = m_{t_0}^{***} + \delta_{1t_0}^{***} h \quad (7)$$

$$\Pi_{a_H} = \Pi_{a_H}^{***} + M^{(a)}(t) + D^{(a)}(t) + G^{(a)}(t) \quad (8)$$

Natomiast w przypadku modelu Holta-Wintersa eliminuje się dwa rodzaje wahań o cyklu rocznym i tygodniowym. Zmienną, z której wyeliminowano składniki sezonowości o cyklach wskazanych wyżej oznaczmy przez $Y_{t(a)}^{**}$.

Równania addytywnego modelu Holta-Wintersa, uwzględniające występowanie wahań o cyklu dobowym o długości m , są następujące:

$$m_t^{**} = \alpha(Y_{t(a)}^{**} - C_{t-m}^*) + (1 - \alpha)m_{t-1}^{**} \quad (9)$$

$$\delta_{1t}^{**} = \beta(m_t^{**} - m_{t-1}^{**}) + (1 - \beta)\delta_{1t-1}^{**} \quad (10)$$

$$C_t^{**} = \delta(Y_{t(a)}^{**} - m_t^{**}) + (1 - \delta)C_{t-m}^{**} \quad (11)$$

Predyktor wyjściowy dla horyzontu prognozy $h = 1, 2, \dots$ wyraża się wzorem:

$$\Pi_{a_HW}^{**} = m_{t_0}^* + \delta_{1t_0}^* h + C_{t_0, 1+h}^* \quad (12)$$

Prognozę końcową wyznacza się na podstawie predyktora:

$$\Pi_{a_HW} = \Pi_{a_HW}^{**} + M^{(a)}(t) + D^{(a)}(t) \quad (13)$$

Jednym z ważnych zagadnień, wymagających rozstrzygnięcia jest wybór modelu adaptacyjnego dla celów prognozowania, charakteryzującego się „optymalnymi” wartościami stałych wygładzania (wyrównywania). W przypadku pełnych danych, wyboru tego można dokonać posługując się jednym z dwóch kryteriów (mierników), polegających na minimalizacji:

- przeciętnych błędów względnych wartości wyrównanych obliczanych dla dostępnych danych w przedziale czasowym „próby” (K_{WW}),
- przeciętnych względnych ocen błędów prognoz ekstrapolacyjnych, obliczanych dla okresu empirycznej weryfikacji prognoz (K_{EX}).

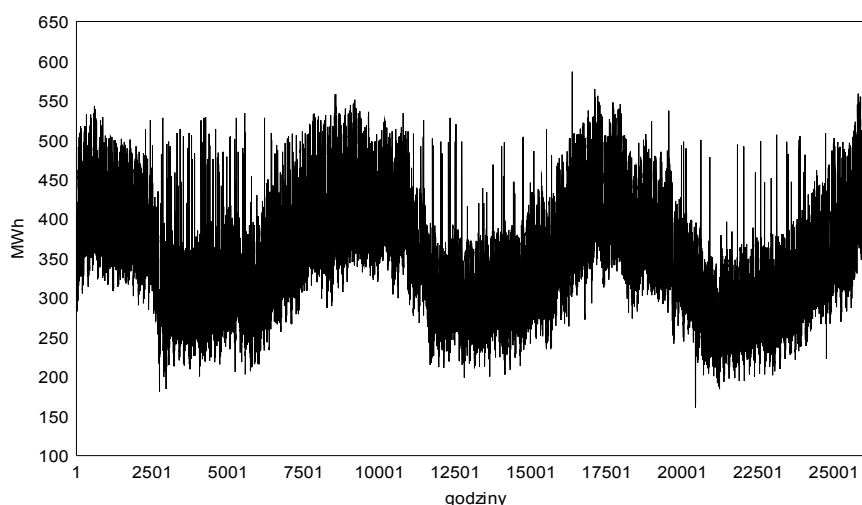
W przypadku szeregów z lukami w grę wchodzi także trzecie kryterium, minimalizujące przeciętne względne oceny błędów prognoz interpolacyjnych obliczonych dla podokresów, w których występowały luki (K_{IN}).

Z uwagi na to, że modele tego samego rodzaju, charakteryzujące się minimalnymi ocenami mierników dla poszczególnych kryteriów dość często mogą różnić się stałymi wygładzania. Konsekwencją tego faktu będzie występowanie dość znacznych różnic błędów prognoz, zwłaszcza ekstrapolacyjnych. W związku z tym, dla każdego z mierników należy obliczyć dwa pozostałe rodzaje mierników, odnoszące się do wartości wyrównanych i/lub przynajmniej jednego rodzaju prognoz. Będą one oznaczone dodatkowymi indeksami. Na przykład dla kryterium K_{WW} będzie to K_{WW_IN} i K_{WW_EX} .

1. Przykład empiryczny – syntetyczna charakterystyka zmiennej prognozowanej

Egzemplifikacją rozważań teoretycznych będzie przykład empiryczny, dotyczący modelowania i prognozowania zapotrzebowania na moc energetyczną w aglomeracji A, na podstawie godzinnych szeregów czasowych oczyszczone z dwóch lub trzech rodzajów wahań sezonowości.

Na rys. 1 przedstawione zostało kształtowanie się zmiennej prognozowanej za okres trzech lat, przy czym rok trzeci będzie okresem empirycznej weryfikacji prognoz.



Rys. 1. Kształtowanie się zapotrzebowania na moc energetyczną (w MWh)

Źródło: Bank Danych Katedry.

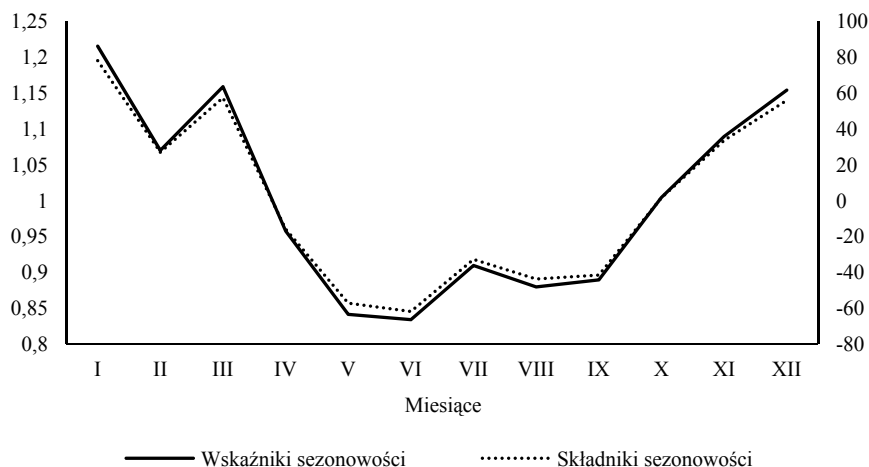
Poniżej zostanie przedstawione w sposób graficzny kształtowanie się ocen wskaźników oraz składników sezonowości o cyklu rocznym, tygodniowym i dobowym (rys. 2-4).

W przypadku wahań o cyklu 12-miesięcznym (rys. 2) dodatnie odchylenia sezonowe ocen wskaźników większe od jedności otrzymano dla I i IV kwartału. Maksimum sezonowe przypada w styczniu. Następnymi miesiącami w kolejności są marzec oraz grudzień. Miesiącami o minimalnym zapotrzebowaniu na moc są czerwiec i maj. Amplituda wskaźników sezonowości wynosi ok. 40 p.p. (punktów procentowych), a składników sezonowości – ok. 140 MWh.

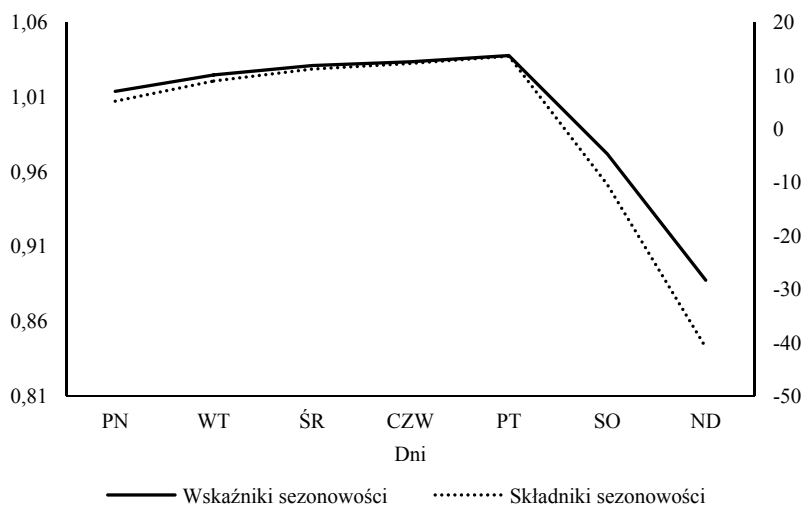
Dodatnie oceny odchyleń sezonowych wyznaczone dla cyklu tygodniowego (rys. 3) otrzymano dla dni roboczych – od poniedziałku do piątku. Maksimum sezonowym, wynoszącym ok. 14 MWh, charakteryzuje się piątek. Widoczna jest ro-

snąca tendencja ocen dodatnich w kolejnych dniach. Zdecydowanie najniższą ocenę składnika sezonowości, przekraczającą -40 MWh, otrzymano dla niedzieli.

Rozstęp wskaźników sezonowości wyniósł ok. 15 p.p. i był zdecydowanie niższy niż dla cyklu rocznego. Różnica skrajnych ocen składników sezonowości wynosiła ok. 55 MWh.



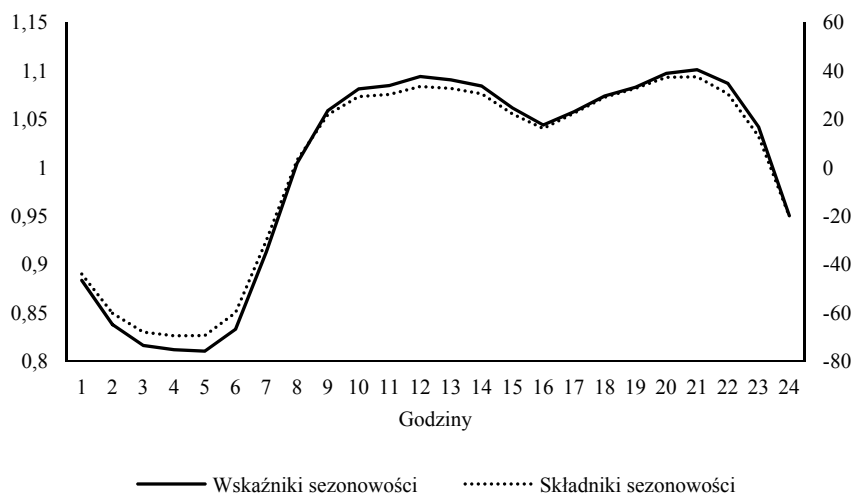
Rys. 2. Oceny składników i wskaźników sezonowości o cyklu rocznym



Rys. 3. Oceny składników i wskaźników sezonowości o cyklu tygodniowym

Z kształtowania się ocen składników sezonowości dla cyklu 24-godzinnego wynika, że oceny ujemne otrzymano dla godzin od 24 do 7 rano (rys. 4). Minimum sezonowe przyjmuje wartość ujemną wynoszącą ok. 70 MWh i przypada na godz. 4 oraz 5.

Z rysunku wynika także, że w okresie zwiększonego zapotrzebowania na moc energetyczną, przypadającego na godziny od 8 do 23, widoczne są dwa cykle związane odpowiednio: ze szczytem ranno-przedpołudniowym oraz popołudniowo-wieczornym. Maksymalne odchylenie sezonowe w pierwszym przypadku, wynoszące ok. 34 MWh, przypada na godzinę 12 (9,4 p.p. powyżej trendu). Szczyt zapotrzebowania na moc w drugim przypadku przypada na godzinę 20 oraz 21. Oceny składników sezonowości wynoszą ok. 37 MWh, tj. ok. 10 p.p. powyżej trendu. Rozstęp ocen wskaźników sezonowości wynosi 29 p.p. i jest blisko dwukrotnie wyższy niż dla cyklu tygodniowego, a także ok. 9 p.p. niższy niż dla cyklu rocznego.



Rys. 4. Oceny składników i wskaźników sezonowości o cyklu 24-godzinnym

3. Wyniki modelowania i prognozowania

W modelowaniu i prognozowaniu adaptacyjnym zapotrzebowania na moc energetyczną w aglomeracji A zostały wykorzystane godzinne szeregi czasowe, oczyszczone z dwóch lub trzech rodzajów wahań sezonowości. Rozpatrywany był jeden wariant luki niesystematycznych. Przyjęto założenie, iż luki występują w godzinach: 2, 6, 8, 10, 14, 16, 18 i 22. 22, w okresie od 1 lipca roku pierwszego do 30 czerwca roku drugiego.

Liczba brakujących danych wynosiła 2920, co stanowiło 1/6 długości szeregu (16,64% z 17520 obserwacji).

Do budowy prognoz na podstawie modeli Holta-Wintersa w postaci addytywnej i multiplikatywnej wykorzystane zostały szeregi czasowe, oczyszczone z wahań o cyklu rocznym oraz tygodniowym (Y_t^{**}). Natomiast prognozy na podstawie prostych modeli Browna i modeli Holta były budowane dla zmiennej, z której zostały wyeliminowane także wahania o cyklu 24-godzinnym (Y_t^{***}).

W tab. 1 zostały zestawione błędy prognoz inter- i ekstrapolacyjnych, otrzymanych na podstawie predyktorów opartych na modelach Browna, Holta i Holta-Wintersa w postaci addytywnej oraz multiplikatywnej o optymalnych wartościach stałych wygładzania, minimalizujących poszczególne kryteria. Kolumna druga zawiera skrótowe nazwy modeli poprzedzone symbolem „a_” dla modeli addytywnych oraz „m_” dla modeli multiplikatywnych. W celach porównawczych w wierszach poniżej podane zostały wielkości odpowiednich błędów, otrzymanych na podstawie modeli dla pełnych danych (bez luk). Zostały one oznaczone występującymi po nazwach modeli dolnymi indeksami „p”.

Ponadto w dwóch ostatnich wierszach dla danych z lukami i bez luk, dla postaci addytywnej oraz multiplikatywnej, podane zostały mierniki dokładności wartości teoretycznych (K_{WW}), prognoz ekstrapolacyjnych (K_{EX}), a dla danych z lukami także prognoz interpolacyjnych (K_{IN}), otrzymanych na podstawie klasycznych modeli szeregu czasowego z liniowym trendem i okresowymi składnikami sezonowymi (a_{Kl} , a_{Klp}) oraz trendem wykładniczym o stałej stopie wzrostu, a także relatywnie stałych wahaniami sezonowych (m_{Kl} , m_{Klp}).

Z informacji zawartych w kolumnie szóstej tab. 1 wynika, że minimalną oceną błędu wartości wyrównanych (K_{WW}), wynoszącą 4,308%, charakteryzuje się predyktor oparty na multiplikatywnym modelu Browna (m_B) dla stałej wygładzania $\alpha = 0,94$. Jest ona o 0,325 p.p. wyższa od oceny błędu otrzymanej dla modelu o tej samej postaci dla pełnych danych (m_{Bp}), ale dla stałej wygładzania $\alpha = 0,97$. Najniższą ocenę błędu prognoz interpolacyjnych dla tego kryterium ($K_{WW_IN} = 4,324\%$) otrzymano dla addytywnego modelu a_B dla $\alpha = 0,96$.

Uzyskana ocena błędu dla modelu a_B jest tylko o 0,011 p.p. niższa od błędu dla modelu multiplikatywnego m_B . Najniższą ocenę błędu prognoz ekstrapolacyjnych dla tego kryterium otrzymano także dla modelu m_B ($K_{WW_EX} = 7,564\%$). Zwracają uwagę bardzo wysokie, przekraczające 80%, wartości tego miernika otrzymane dla: addytywnego modelu Holta (a_B) oraz obu postaci modelu Holta-Wintersa (a_{HW} , m_{HW}).

Tabela 1. Oceny błędów wartości wyrównanych, prognoz interpolacyjnych i prognoz ekstrapolacyjnych dla najlepszych modeli adaptacyjnych

Model	Oznaczenie	Wartości wyrównane (WW)						Prognozy interpolacyjne (IN)						Prognozy ekstrapolacyjne (EX)					
		Stale wygładzania			MAPE (w %)			Stale wygładzania			MAPE (w %)			Stale wygładzania			MAPE (w %)		
		α	β	γ	K_{WW}	$K_{WW,IN}$	$K_{WW,EX}$	α	β	γ	K_{IN}	$K_{IN,WW}$	$K_{IN,EX}$	α	β	γ	K_{EX}	$K_{EX,WW}$	$K_{EX,IN}$
Modele addytywne																			
Browna	a_B	0,95	-	-	4,334	4,324	7,755	0,96	-	-	4,323	4,335	7,781	0,75	-	-	7,531	4,432	4,459
	a_B _p	0,97	-	-	4,007		7,808							0,75	-	-	7,531	4,132	
	a_H	0,20	0,01	-	5,568	5,550	91,758	0,20	0,01	-	5,550	5,568	91,758	0,18	0,04	-	7,536	5,740	5,720
	a_H _p	0,15	0,01	-	5,545		84,750							0,11	0,09	-	8,694	5,993	
Holta-Wintersa	a_HW	0,15	0,01	0,14	4,639	6,053	83,436	0,15	0,01	0,09	6,017	4,689	50,267	0,08	0,01	0,07	8,047	4,975	6,267
	a_HW _p	0,15	0,01	0,14	4,453		83,436							0,08	0,01	0,07	8,047	4,820	
Klas. szer.	a_KI				6,905						6,539						8,208		
Czasowego	a_KI _p				6,666												8,192		
Modele multiplikatywne																			
Browna	m_B	0,94	-	-	4,308	4,335	7,564	0,95	-	-	4,334	4,308	7,590	0,66	-	-	7,254	4,465	4,584
	m_B _p	0,97	-	-	3,983		7,646							0,66	-	-	7,254	4,191	
	m_H	0,20	0,01	-	5,302	5,429	8,850	0,20	0,01	-	5,429	5,302	8,850	0,15	0,01	-	8,071	5,454	5,516
	m_H _p	0,15	0,02	-	5,303		87,289							0,15	0,01	-	8,071	5,313	
Holta-Wintersa	m_HW	0,15	0,01	0,14	4,454	6,069	94,868	0,15	0,01	0,11	6,061	4,470	83,173	0,11	0,01	0,01	7,081	6,181	6,307
	m_HW _p	0,15	0,01	0,14	4,288		94,870							0,11	0,01	0,01	7,068	5,927	
Klas. szer.	a_KI				6,663						6,544						7,653		
Czasowego	a_KI _p				6,372												7,857		

Dla pełnych danych bardzo wysokie błędy prognoz ekstrapolacyjnych (K_{WW_EX}), oprócz trzech wyżej wymienionych modeli, otrzymano także dla multiplikatywnego modelu Browna (m_{B_p}).

Minimalną ocenę błędu prognoz interpolacyjnych (K_{IN}), wynoszącą 4,323%, otrzymano na podstawie predyktora opartego na addytywnym modelu Browna (a_B) dla $\alpha = 0,96$. Dla modelu multiplikatywnego m_B i $\alpha = 0,95$ była ona o 0,11 p.p. wyższa. Oceny błędów tego rodzaju prognoz otrzymanych dla pozostałych modeli są od 0,98 p.p. (m_H) do 1,738 p.p. (m_{HW}) wyższe.

Oceny błędów K_{IN_EX} dla addytywnej postaci modelu Holta i obu postaci modelu Holta-Wintersa przyjmują, podobnie jak w przypadku kryterium K_{WW} , bardzo wysokie wartości. Zawarte są one w przedziale od 50,267% dla modelu a_{HW} do 91,758% dla a_H .

Najniższą ocenę błędu prognoz ekstrapolacyjnych (K_{EX}), wynoszącą 7,081%, otrzymano dla modelu Holta-Wintersa w postaci multiplikatywnej (m_{HW}) o stałych wygładzania: $\alpha = 0,96$, $\beta = 0,01$, $\gamma = 0,01$. Drugim w kolejności był model Browna o tej samej postaci (m_B) i stałej wygładzania $\alpha = 0,66$ (7,254%). Jeszcze dwa modele w postaci addytywnej (a_B oraz a_H) charakteryzowały się ocenami w granicach 7,5%. Dla dwóch pozostałych (a_{HW} oraz m_H) przekroczyły one 8%.

Z porównania ocen błędów dla kryterium (K_{EX}) dla danych z lukami i pełnych danych wynika, że z wyjątkiem modeli Holta w postaci addytywnej, otrzymano nie tylko identyczne oceny stałych wygładzania, ale także identyczne oceny błędów prognoz ekstrapolacyjnych. Oznacza to, iż występowanie luk w danych w ośmiu okresach godzinnych od połowy roku pierwszego do połowy roku drugiego nie miało wpływu na wartości i dokładność prognoz. W przypadku miernika K_{EX_WW} , różnice te nie przekraczają 0,3 p.p.

Należy zwrócić uwagę, że oceny błędów: wartości wyrównanych (K_{EX_WW}) prognoz interpolacyjnych (K_{EX_IN}) przyjęły wartości tylko niewiele wyższe od minimalnych ocen błędów K_{WW} oraz K_{IN} . Oznacza to, że kryterium wyboru modelu dla celów prognozowania ekstrapolacyjnego z całą pewnością nie mogą być predyktory, oparte na modelach Holta oraz Holta-Wintersa. Ewentualnie mogą być brane pod uwagę niezależnie od postaci jedynie modele Browna.

W tab. 1 podane zostały także zarówno dla luk, jak i dla pełnych danych oceny błędów wartości wyrównanych, prognoz interpolacyjnych (dla luk) oraz ekstrapolacyjnych, otrzymanych na podstawie modeli klasycznych w postaci addytywnej (a_{Kl} , a_{Kl_p}) i multiplikatywnej (m_{Kl} , m_{Kl_p}). Średnie względne błędy wartości teoretycznych (wyrównanych) dla modelu addytywnego i multiplikatywnego wynoszą odpowiednio: 6,905% oraz 6,663% i są one o ok. 2,3 p.p.

oraz 2,6 p.p. wyższe od otrzymanych dla najlepszego modelu adaptacyjnego (m_B) i jednocześnie o ok. 0,3 p.p. wyższe od otrzymanych dla modeli bez luk (a_{Kl_p} , m_{Kl_p}). W przypadku prognoz interpolacyjnych względne ich błędy kształtują się na poziomie: 6,542% (a_{Kl}) oraz 6,545% (m_{Kl}) i są one wyższe od minimalnego średniego błędu prognoz otrzymanego na podstawie addytywnych modeli Browna (a_B) o ok. 2,2 p.p.-2,3 p.p.

W przypadku prognoz ekstrapolacyjnych otrzymanych na podstawie modelu w postaci multiplikatywnej, błąd ten jest o 0,421 p.p. (6,6%) wyższy od błędu minimalnego otrzymanego na podstawie modelu m_{HW} . Dla modelu addytywnego różnica ta wynosi ok. 0,9 p.p.

W tab. 2 zamieszczone zostały oceny błędów średnich prognoz interpolacyjnych, otrzymanych na podstawie najlepszych predyktorów adaptacyjnych (a_B oraz m_{HW}), w dezagregacji na godziny, dni tygodnia i miesiące.

Tabela 2. Przeciętne względne błędy prognoz zdezagregowanych

Zmienne	Prognozy interpolacyjne			Prognozy ekstrapolacyjne		
	Liczba	a_B	m_{Kl}	m_{HW}	m_{Kl}	m_{Kl_p}
		0,96	-	0.11	0.01	0.01
	2920	Błędy względne prognoz (MAPE) w %				
		4,323	6,544	7,068	7,653	7,857
1	2	3	4	5	6	7
G1	0	-	-	5,979	7,175	7,312
G2	365	4,087	6,574	6,190	6,973	7,672
G3	0	-	-	6,500	7,843	7,977
G4	0	-	-	6,548	8,519	8,657
G5	0	-	-	6,964	8,446	8,559
G6	365	5,335	7,335	8,394	8,106	9,000
G7	0	-	-	9,181	9,832	9,887
G8	365	4,365	8,195	9,431	10,001	10,661
G9	0	-	-	7,693	9,145	9,235
G10	365	3,581	6,023	6,421	7,071	7,400
G11	0	-	-	5,416	6,704	6,791
G12	0	-	-	5,707	6,952	7,039
G13	0	-	-	5,734	6,479	6,534
G14	365	2,684	5,401	5,655	5,940	6,005
G15	0	-	-	5,815	5,874	5,932
G16	365	3,644	6,072	6,162	6,212	6,697
G17	0	-	-	8,837	8,513	8,580
G18	365	4,857	7,415	9,638	8,979	9,237
G19	0	-	-	9,459	9,393	9,399
G20	0	-	-	9,398	9,235	9,238
G21	0	-	-	8,083	7,706	7,776
G22	365	6,034	5,332	6,090	6,129	6,263

cd. tabeli 2

1	2	3	4	5	6	7
G23	0	-	-	5,435	5,750	5,869
G24	0	-	-	4,907	6,699	6,842
PN	416	4,110	7,323	7,267	7,382	7,604
WT	416	4,308	6,613	6,372	7,343	7,578
ŚR	416	3,880	6,033	6,786	7,314	7,501
CZ	416	3,852	6,476	6,969	9,349	9,480
PT	416	4,130	5,451	6,790	8,149	8,375
SO	424	4,468	6,150	6,741	6,916	7,079
ND	416	5,514	7,766	8,566	7,133	7,394
STYCZEŃ	248	3,123	6,047	6,080	6,220	6,277
LUTY	224	3,126	4,764	5,250	8,101	8,507
MARZEC	248	2,788	5,299	10,968	9,128	9,712
KWIECIEŃ	240	3,597	6,907	6,168	7,163	7,467
MAJ	248	5,297	6,225	8,291	8,696	8,766
CZERWIEC	240	4,252	6,786	6,654	9,336	9,375
LIPIEC	248	4,994	5,700	7,782	7,439	7,912
SIERPIEŃ	248	5,544	8,191	6,306	6,536	6,717
WRZESIEŃ	240	5,044	7,006	5,673	7,225	7,291
PAŹDZIERNIK	248	5,356	6,275	6,050	5,930	5,974
LISTOPAD	240	4,354	7,538	6,453	6,972	7,199
GRUDZIEŃ	248	4,289	7,679	8,860	9,137	9,143

W celach porównawczych podane zostały, w takim samym układzie, błędy obu rodzajów prognoz, otrzymanych na podstawie modelu klasycznego w postaci multiplikatywnej (m_{KI}). W kolumnie drugiej podano liczby prognoz interpolacyjnych, obejmujących drugą połowę pierwszego roku oraz pierwszą połowę drugiego. Dotyczyły one tylko tych godzin, w których występowały luki, tj.: 2, 4, 6, 8, 10, 14, 18, 22. W przypadku godzin liczba prognoz odpowiada ilości dni w roku (365). Dla dni tygodnia jest ona iloczynem liczby tygodni w roku (52) oraz liczby luk w każdym dniu (8). W odniesieniu do miesięcy są one iloczynem liczby dni w poszczególnych miesiącach oraz liczby luk.

Z porównania ocen błędów prognoz interpolacyjnych dla godzin wynika, że dla predyktora adaptacyjnego (a_B) są one w siedmiu na osiem przypadków niższe od uzyskanych dla predyktora klasycznego (m_{KI}). Różnice te zawarte są w przedziale od 2,0 p.p. (godz. 6) do 3,83 p.p. (godz. 8). Jedynie dla godz. 22 lepszym o 0,702 p.p. okazał się predyktor klasyczny. Dla wszystkich dni tygodnia bardziej efektywny okazał się predyktor adaptacyjny. Różnice ocen kształtowały się od 1,32 p.p. dla piątku do 3,213 p.p. dla poniedziałku. Podobnie sytuacja wygląda w przypadku miesięcy. Wyższe dokładności prognoz adaptacyjnych zawarte były w przedziale od 0,706 p.p. dla lipca do 3,39 dla grudnia. Bardzo zbliżone oceny do różnicy maksymalnej otrzymano także dla kwietnia (3,31 p.p.) oraz listopada (3,184 p.p.).

Z uwagi na to, iż prognozy ekstrapolacyjne dotyczyły całego roku trzeciego ich liczba wynosiła 8760 (365 dni x 24 godziny). Jak wspomniano wcześniej przeciętny błąd względny prognoz ekstrapolacyjnych, otrzymanych na podstawie predyktora adaptacyjnego (m_{HW}) wynosił 7,068% i był niższy o 0,421 p.p. (5,95%) od błędu prognoz dla predyktora klasycznego (m_{KI}). Ta ostatnia ocena była niższa o ok. 0,2 p.p. od oceny uzyskanej dla modelu klasycznego bez luk w danych (m_{KI_p}). Spośród 24 godzin, dla 18 niższe oceny błędów otrzymano dla predyktora adaptacyjnego. Kształtowały się one w przedziale od 0,049 p.p. (dla godziny 16) do 1,971 p.p. (dla godz. 4) i 1,792 p.p. (dla godz. 24). Niższe oceny błędów prognoz dla modelu m_{KI} otrzymano dla godz. 6 oraz godzin od 17 do 21. Różnice ocen zawarte były w granicach do 0,065 (godz. 19) do 0,659 (godz. 18). Dla sześciu dni tygodnia bardziej efektywny okazał się model m_{HW} . Różnice ocen błędów kształtowały się w przedziale od 0,115 p.p. dla poniedziałku do 2,380 p.p. dla czwartku. Jedynie dla niedziel otrzymano ocenę o 1,434 p.p. niższą dla predyktora klasycznego. Dla 9 z 12 miesięcy bardziej efektywny okazał się predyktor adaptacyjny. Różnice ocen błędów zawarte były w przedziale 0,14 p.p.-2,85 p.p. i dotyczyły odpowiednio: stycznia oraz czerwca. Niższe oceny błędów dla predyktora klasycznego otrzymano dla: marca (1,841 p.p.), lipca (0,343 p.p.) oraz października (0,12 p.p.).

Podsumowanie

Z przeprowadzonych w pracy rozważań można wyprowadzić następujące wnioski syntetyczne:

1. Podstawą wyboru modelu dla celów prognozowania *ex ante* nie mogą być błędy wartości wyrównanych (KWW), a także błędy prognoz interpolacyjnych (KIN), ponieważ minimalnym ich ocenom zazwyczaj odpowiadają bardzo duże błędy prognoz ekstrapolacyjnych. Uwaga ta odnosi się do modeli Holta i Holta-Wintersa.
2. Kryterium wyboru powinny być oceny błędów prognoz ekstrapolacyjnych *ex post*. Minimalnym ocenom błędów tego rodzaju prognoz, odpowiadają nieznacznie wyższe od minimalnych oceny błędów wartości wyrównanych oraz prognoz interpolacyjnych.
3. Oceny błędów prognoz ekstrapolacyjnych dla modeli adaptacyjnych dla danych z lukami były zbliżone do błędów dla pełnych danych.
4. Prognozy ekstrapolacyjne dla modeli adaptacyjnych dla niesystematycznych luk w danych oczyszczonych z sezonowości, pozwoliło na uzyskanie prze-

ciętnego błędu prognoz o ok. 6,6 % niższego niż dla klasycznego modelu szeregu czasowego. W przypadku prognoz interpolacyjnych różnica ta wyniosła ok. 50%.

5. Tym samym potwierdzona została użyteczność modeli adaptacyjnych dla danych oczyszczonych z sezonowości w przypadku występowania niesystematycznych luk w danych.

Literatura

- Dittmann P. (2006), *Prognozowanie w przedsiębiorstwie. Metody i ich zastosowanie*, Wolters Kluwer, Kraków.
- Pawłowski Z. (1973), *Prognozowanie ekonometryczne*, PWN, Warszawa.
- Szmuksta-Zawadzka M., Zawadzki J. (2015), *Wykorzystanie danych oczyszczonych o wysokiej częstotliwości w prognozowaniu zmiennych ze złożoną sezonowością*, *Metody Ilościowe w Badaniach Ekonomicznych* Vol. 16, No. 4, Warszawa, s. 147-159.
- Zeliaś A., Pawełek B., Wanat S. (2003), *Prognozowanie ekonomiczne. Teoria, przykłady, zadania*, Wydawnictwo Naukowe PWN, Warszawa.

FORECASTING MISSING DATA FOR SEASONAL ADJUSTED HIGH FREQUENCY TIME SERIES

Summary: In this paper was presented application of selected exponential smoothing models in forecasting very high frequency variables on the basis of time series with un-systematic gaps, from which two or three types of seasonal fluctuations were eliminated. Exemplification of theoretical considerations will be an empirical example, concerning the power demand in agglomeration A in hourly periods.

Keywords: forecasting, high frequency time series, complex seasonality, exponential smoothing, un-systematic gaps.