

Maria Szmuksta-Zawadzka, Jan Zawadzki

Zachodniopomorski Uniwersytet Technologiczny w Szczecinie

e-mails: maria.szmuksta-zawadzka@zut.edu.pl; jan.zawadzki@zut.edu.pl

MODELE HIERARCHICZNE W PROGNOZOWANIU ZMIENNYCH O WYSOKIEJ CZĘSTOTLIWOŚCI OBSERWOWANIA W WARUNKACH BRAKU PEŁNEJ INFORMACJI

Streszczenie: W pracy zostanie przedstawiona procedura wykorzystania regularnych modeli hierarchicznych w prognozowaniu brakujących danych w szeregach czasowych z wahaniami cyklicznymi o wysokiej częstotliwości obserwowania. Wahania złożone o cyklach: rocznym, tygodniowym i dobowym będą nakładać się na trend w sposób addytywny. Dla wahań o parzystej długości cykli zostały zbudowane osobne regularne modele hierarchiczne. Rozważania teoretyczne zostały zilustrowane przykładem empirycznym dotyczącym zapotrzebowania na energię elektryczną w okresach godzinnych.

Słowa kluczowe: dane o wysokiej częstotliwości obserwowania, modele hierarchiczne, niekompletne szeregi czasowe.

DOI: 10.15611/ekt.2014.4.07

1. Wstęp

W dotychczasowych badaniach dotyczących modelowania i prognozowania zmiennych ze złożoną sezonowością zarówno dla pełnych, jak i niepełnych danych o bardzo wysokiej częstotliwości obserwowania były wykorzystywane modele ze zmiennymi zero-jedynkowymi, uwzględniające trzy rodzaje wahań złożonych o cyklach: 12-miesięcznym (rocznym), tygodniowym (7-dniowym) i dziennym (24-godzinny). Każdy z rodzajów wahań może być opisany także za pomocą składowych harmonicznych. Autorzy w pracy [Szmuksta-Zawadzka, Zawadzki 2013] zaproponowali wykorzystanie do tego celu regularnych modeli hierarchicznych dla pełnych danych. Ze względu na to, że modele tego rodzaju mogą być budowane do opisu wahań o parzystych długościach cykli, można je wykorzystać do opisu wahań o cyklu rocznym oraz dobowym. Zapisy analityczne wymienionych wyżej modeli dla wahań nakładających się addytywnie na trend zostaną podane w następnym podrozdziale.

Celem artykułu jest przedstawienie procedury wykorzystania regularnych modeli hierarchicznych dla danych o bardzo wysokich częstotliwościach do prognozowania brakujących danych w szeregach z lukami niesystematycznymi. Z tego rodzaju lukami mamy do czynienia wtedy, gdy dostępne są przynajmniej pojedyncze dane o każdym z podokresów składowych (miesiącu, dniu tygodnia i godzinie). Rozważania teoretyczne zostały zilustrowane przykładem empirycznym dotyczącym zapotrzebowania na energię elektryczną w okresach godzinnych w sytuacji, gdy luki niesystematyczne występują w każdym z podokresów składowych.

2. Wprowadzenie teoretyczne

Rozważania prowadzone w tym podrozdziale rozpoczniemy, zgodnie z zapowiedzią, od przedstawienia zapisów analitycznych modeli ze zmiennymi zero-jedynkowymi oraz z wielomianami trygonometrycznymi. Zasadnicza część podrozdziału zostanie poświęcona prezentacji modeli hierarchicznych dla pełnych danych. W końcowej jego części przedstawione zostaną uwagi dotyczące modelowania i prognozowania brakujących danych z lukami niesystematycznymi.

Zapis modelu ze zmiennymi zero-jedynkowymi, uwzględniającego trzy rodzaje wahań złożonych o cyklach: 12-miesięcznym (rocznym), tygodniowym (7-dniowym) i dziennym (24 godzinnym), jest następujący (por. np. [Kufel 2010; Szmuksta-Zawadzka, Zawadzki 2011]):

$$Y_t = \alpha_1 t + \alpha_0 + \sum_{i=1}^{12} b_{0i} M_{it} + \sum_{j=1}^7 c_{0j} D_{jt} + \sum_{k=1}^{24} d_{0k} G_{kt} + U_t \quad (1)$$

przy warunkach:

$$\sum_{i=1}^{12} b_{0i} = \sum_{j=1}^7 c_{0j} = \sum_{k=1}^{24} d_{0k} = 0,$$

gdzie:

- M_{it} – miesiąc,
- D_{jt} – dzień tygodnia,
- G_{kt} – okres godzinny w cyklu dobowym.

Model, którym każdy z trzech wymienionych wyżej rodzajów wahań może być opisany za pomocą wielomianu trygonometrycznego, przyjmuje postać [Szmuksta-Zawadzka, Zawadzki 2011]:

$$Y_t = \alpha_1 t + \alpha_0 + \sum_{i=1}^6 (a_{0i} \sin \omega_i t_m + b_{0i} \cos \omega_i t_m) + \sum_{j=1}^3 (a_{0j} \sin \omega_j t_d + b_{0j} \cos \omega_j t_d) + \sum_{k=1}^{24} (a_{0k} \sin \omega_k t + b_{0k} \cos \omega_k t) + U_t \quad (2)$$

gdzie:

- $t_m = 1, 2, \dots, 12, \dots, 12 \times r$,
 - $t_d = 1, 2, \dots, 7, \dots, 365 \times r$,
 - $t = 1, 2, \dots, 48, \dots, 365 \times 48 \times r$,
- (r oznacza nr roku).

W modelu (2) występują trzy zmienne oznaczające czas, przy czym zmienne t_m i t_d mają charakter pomocniczy i odnoszą się one odpowiednio do kolejnych miesięcy lub dni tygodnia.

W pracy [Szmuksta-Zawadzka, Zawadzki 2013] przedstawiano wykorzystanie do tego celu regularnych modeli hierarchicznych dla pełnych danych. Za ich pomocą można opisać wahania o parzystych długościach cykli (12-miesięczne i 24-godzinne).

Regularne modele hierarchiczne to takie, dla których podzielniki p_i długości cyklu wahań m spełniają jednocześnie dwa warunki:

$$2 \leq p_i \leq \frac{m}{2} \text{ oraz } \prod_i p_i = m. \quad (3)$$

Parametry modeli hierarchicznych są średnimi z parametrów modelu ze zmiennymi zero-jedynkowymi (1).

Liczba regularnych modeli hierarchicznych jest równa liczbie permutacji oraz permutacji z powtórzeniami podzielników p_i . Dla danych miesięcznych ($i = 12$) wynosi ona 7, w tym:

- 4 dwustopniowe (HM_p1_p2),
- 3 trójstopniowe (HM_p1_p2_p3).

Natomiast dla danych godzinnych ($k = 24$) ich liczba jest równa 19, w tym:

- 6 dwustopniowych (HG_p1_p2),
- 9 trójstopniowych (HG_p1_p2_p3),
- 4 czterostopniowe (HG_p1_p2_p3_p4).

W modelach klasycznych danych równaniem (1) do opisu każdego rodzaju wahań periodycznych, ze względu na warunki sumowalności do zera, szacuje się o jeden parametr mniej, niż wynoszą długości cykli. Podobnie wygląda sprawa w przypadku modelu (2). Natomiast w modelach hierarchicznych dla każdego rodzaju wahań szacuje się tyle parametrów, ile wynosi suma podzielników pomniejszona o ich liczbę. Jest ich nie więcej, niż wynosi połowa długości cyklu. Najmniej parametrów szacuje się w modelach 4-stopniowych. Przykładowo dla cyklu dobowego zamiast 23 w modelu H2_2_2_3 szacuje się pięć parametrów ($(2+2+2+3) - 4 = 5$).

Ze względu na to, że zapisy analityczne modeli hierarchicznych wraz z ich specyfikacją dla danych miesięcznych i godzinnych można znaleźć m.in. w pracach ([Szmuksta-Zawadzka, Zawadzki 2013; Zawadzki (red.) 2003]), poniżej podany zostanie zapis analityczny jedynie modelu czterostopniowego (HG_p1_p2_p3_p4). Zapis tego modelu jest następujący:

$$Y_{srltk} = \alpha_1 t + \alpha_0 + \sum_{s=1}^{p_1} b_{0s} G_{st} + \sum_{r=1}^{p_2} b_{0sr} G_{str} + \sum_{l=1}^{p_3} b_{0srl} G_{srlt} + \sum_{l=1}^{p_4} b_{0srlk} G_{srltk} + U_{srltk} \quad (4)$$

przy warunkach:

$$\sum_{s=1}^{p_1} b_{0s} = \sum_{r=1}^{p_2} b_{0sr} = \sum_{l=1}^{p_3} b_{0srl} = \sum_{k=1}^{p_4} b_{0srkl} = 0.$$

Zapisy modeli o mniejszej liczbie stopni hierarchii otrzymuje się w sposób natychmiastowy poprzez uwzględnienie tylu początkowych składników, ile wynosi ta liczba. W przypadku danych miesięcznych wystarczy podstawić zmienną M w miejsce zmiennej G .

Zaletą modeli hierarchicznych jest to, że liczba parametrów opisujących wahań o cyklu m jest nie większa niż $m/2$. Własność ta jest szczególnie istotna w przypadku modelowania i prognozowania dla danych o długości jednostkowej nie krótszej niż dekada.

W sytuacji, gdy korzystamy z szeregów o wysokiej częstotliwości, fakt ten nie ma większego znaczenia. Natomiast w wielu przypadkach korzyści z ich stosowania wynikają z tego, że parametry modeli hierarchicznych, o czym była mowa wcześniej, są uśrednionymi parametrami modelu (1).

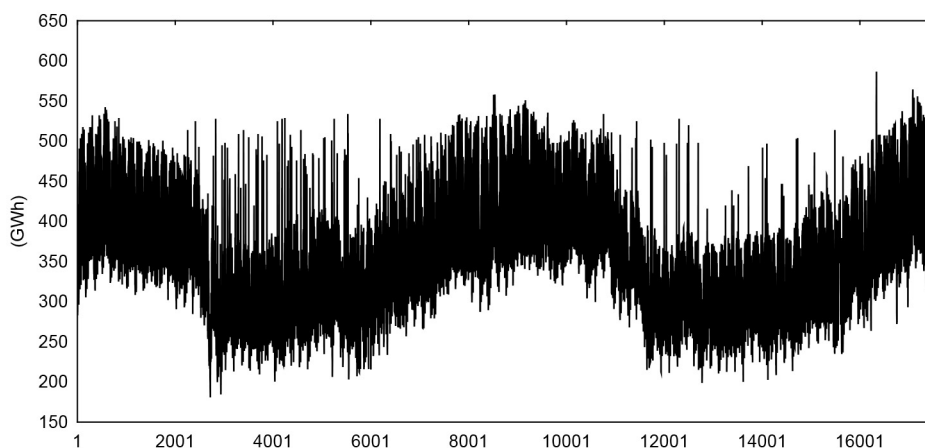
W modelach (1) i (2) dla każdego rodzaju wahań szacuje się o jeden parametr mniej, niż wynoszą długości cykli. Natomiast w modelach hierarchicznych szacowanych jest nie więcej parametrów niż połowa długości cyklu. Oznacza to, że zazwyczaj otrzymuje się modele o nieznacznie gorszych własnościach predyktywnych. Jednak w następstwie zastosowania procedury uśredniania ma miejsce swego rodzaju „wyrównanie” szeregu, polegające na osłabieniu wpływu wartości skrajnych, zakłócających przebieg badanego procesu. Własność ta jest szczególnie istotna w przypadku wykorzystania modeli dla celów prognozowania, ponieważ to dość często prognozy otrzymywane na podstawie najlepszych modeli hierarchicznych charakteryzują się niższymi ocenami błędów od prognoz otrzymanych dla modeli niehierarchicznych.

W przypadku wykorzystania modeli hierarchicznych do prognozowania brakujących danych szacuje się dwu-, trzy- i czterostopniowe na podstawie dostępnych danych, a więc w taki sam sposób jak dla pełnych danych. Różnica polega na tym, że dla podokresów, w których wystąpiły luki, wyznacza się prognozy interpolacyjne. Dla obu rodzajów danych buduje się prognozy wybiegające poza przedział czasowy próby (ekstrapolacyjne).

3. Modelowanie i prognozowanie zapotrzebowania na energię elektryczną dla niepełnych danych

Modelowaniu, a następnie prognozowaniu inter- i ekstrapolacyjnemu poddano zapotrzebowanie na energię elektryczną w okresach godzinnych w aglomeracji A. Dane statystyczne o kształtowaniu się zapotrzebowania na energię elektryczną (w GWh)

pochodzą z Banku Danych Katedry. Obejmowały one okres dwóch lat, tj. 17 520 obserwacji. Trzeci rok był okresem empirycznej weryfikacji prognoz.



Rys. 1. Kształtowanie zapotrzebowania na energię elektryczną (GWh)

Źródło: Banku Danych Katedry.

Rozpatrywany był jeden wariant niesystematycznych luk danych. Luki występowały od 1 lipca pierwszego roku do 30 czerwca drugiego roku i dotyczyły:

- dwóch miesięcy (kwiecień, październik),
- dwóch dni tygodnia (wtorek, czwartek),
- ośmiu okresów godzinnych (2, 6, 8, 10, 14, 16, 18, 22).

Udział luk wynosił ponad 30% długości przedziału czasowego próby (5288).

Na podstawie szeregu z lukami zostało oszacowanych 159 modeli hierarchicznych:

- 7 modeli hierarchicznych dla miesięcy (HMp1_p2_(p3)_D01_G01),
- 19 modeli hierarchicznych dla godzin (M01_D01_HG_p1_p2_(p3_p4),
- 133 modele (7*19 modeli) ze złożoną sezonowością (HMp1_p2_(p3)_D01_HG_p1_p2_(p3_p4).

W trakcie omawiania wyników modelowania i prognozowania stosowane będą skrócone zapisy modeli hierarchicznych ze złożoną sezonowością roczną i dobową. Będą one składać się z dwóch członów odnoszących się do wymienionych rodzajów wahań oraz modeli mieszanych złożonych z części hierarchicznej i klasycznej (zero-jedynkowej). W zapisach tych zostaną pominięte podkreślniki oddzielające podzbiorniki długości cyklu wahań p_i oraz składniki D01. Przykładowo modele trójstopniowe: HMp1_p2_p3_HG_p1_p2_p3 zostaną zapisane jako: HMp1p2p3_HGp1p2p3.

W celach porównawczych, zarówno dla danych z lukami, jak i pełnych danych, zostaną oszacowane modele dane równaniem (1). Ich skrócone zapisy z pominięciem składowej D01 będą wyglądać następująco: M01_G01 oraz M01_G01(p).

Szacowane modele zawierały także zmienne zero-jedynkowe opisujące występowanie świąt oraz jednego dnia przedświątecznego (Wielkiej Soboty). Będą one zawierać ponadto zmienną endogeniczną opóźnioną o 24 godziny. Wprowadzenie jej wynika z faktu 24-godzinnego wyprzedzenia dotyczącego składania zamówień na moc elektryczną.

Na podstawie predyktorów opartych na oszacowanych równaniach zostały wyznaczone prognozy inter- i ekstrapolacyjne oraz obliczone ich średnie błędy względne.

W tabeli 1 zestawione zostały syntetyczne charakterystyki opisujące własności predyktywne równań (R^2 , SE, V_s) oraz błędy prognoz interpolacyjnych (MAPE_I) dla najlepszych modeli hierarchicznych najpierw opisujących jeden rodzaj wahań sezonowych, a następnie ich dwa rodzaje. Dla celów porównawczych podano także wartości tych charakterystyk otrzymanych dla modelu klasycznego z dwiema zmiennymi zero-jedynkowymi.

Tabela 1. Charakterystyki najlepszych modeli w prognozowaniu interpolacyjnym

Model	R^2	SE	V_s (w %)	MAPE_I	Ranga_ R^2	Ranga_I
M01_G01	0,851	28,33	7,78	5,529	1	1
M01_HG122	0,848	28,68	7,87	5,573	2	2
HM43_G01	0,834	29,93	8,22	5,757	18	18
HM43_HG122	0,832	30,12	8,27	5,773	26	21

Źródło: opracowanie własne.

Ponadto dwie ostatnie kolumny zawierają rangi modeli ze względu na R^2 oraz MAPE_I. Z informacji zawartych w kolumnie trzeciej wynika, że najwyższą oceną współczynnika determinacji R^2 oraz najniższą oceną współczynnika zmienności losowej charakteryzuje się model klasyczny ze zmiennymi zero-jedynkowymi (M01_G01). Oceny tych parametrów wynoszą odpowiednio: 0,851% oraz 7,78%. Nie jest to niespodzianką, ponieważ w przypadku modeli hierarchicznych (nawet dla jednego rodzaju wahań) szacuje się mniej parametrów. Parametry te są, jak powiedziano wcześniej, przeciętnymi z parametrów modelu klasycznego i stąd ma miejsce utrata pewnego kwantum informacji. Z bardzo znikomą jej utratą, wynoszącą tylko 0,03 punktu procentowego (p.p.), mamy do czynienia w przypadku współczynnika determinacji dla modelu mieszanego M01_H122. Dla modelu z dwoma rodzajami wahań (HM43_H122) otrzymano nieznacznie niższą, wynoszącą o 0,19 p.p., ocenę współczynnika determinacji oraz wyższą – o 0,49 p.p. – ocenę współczynnika zmienności losowej. W przedostatniej kolumnie podane zostały rangi otrzymane dla najlepszych modeli, uszeregowanych według malejących wartości współczynników determinacji wyznaczonych dla 160 oszacowanych modeli. Na pierwszych dwóch miejscach uplasowały się modele: ze zmiennymi zero-jedynkowymi (M01_G01) oraz model mieszany (M01_HG122). Modele HM43_G01 oraz HM43_HG122 sklasyfikowane zostały odpowiednio na 18 i 26 miejscu. W przypadku błędów prognoz

interpolacyjnych (MAPE_I) najniższe ich oceny wynoszące odpowiednio: 5,529% oraz 5,573% otrzymano dla predyktorów opartych na modelach o najlepszych własnościach predyktywnych. Zajęły one te same dwie czołowe lokaty co poprzednio. Dla dwóch pozostałych modeli otrzymane oceny były wyższe odpowiednio: o 0,23 p.p. i 0,25 p.p. (tj. o 4,12% i 4,41%). W przypadku trzech pierwszych modeli występuje zgodność rang błędów prognoz i współczynników determinacji. Natomiast model HM43_HG122 „awansował” z 26 na 21 miejsce. Ocena współczynnika korelacji między R^2 a MAPE_I obliczona dla wszystkich modeli wynosząca $-0,132$ świadczy o istnieniu słabej ujemnej zależności tych miar.

W tabeli 2 zestawione zostały oceny błędów prognoz ekstrapolacyjnych. Zawiera ona o dwie kolumny i dwa wiersze więcej niż tab. 1. Kolumny dotyczą odpowiednio: ocen błędów prognoz ekstrapolacyjnych (MAPE_E) i rang modeli ze względu na kształtowanie się ocen prognoz tego rodzaju. Modele zostały uszeregowane ze względu na wielkości stałych błędów. Natomiast w dodatkowych wierszach, wyróżnionych kolorem szarym, zamieszczono w celach porównawczych charakterystyki modeli dla pełnych danych: ze zmiennymi zero-jedynkowymi (M01_G01(p)) oraz modelu hierarchicznego z wahaniami złożonymi (HM43_HG122(p)).

Tabela 2. Charakterystyki najlepszych modeli w prognozowaniu ekstrapolacyjnym

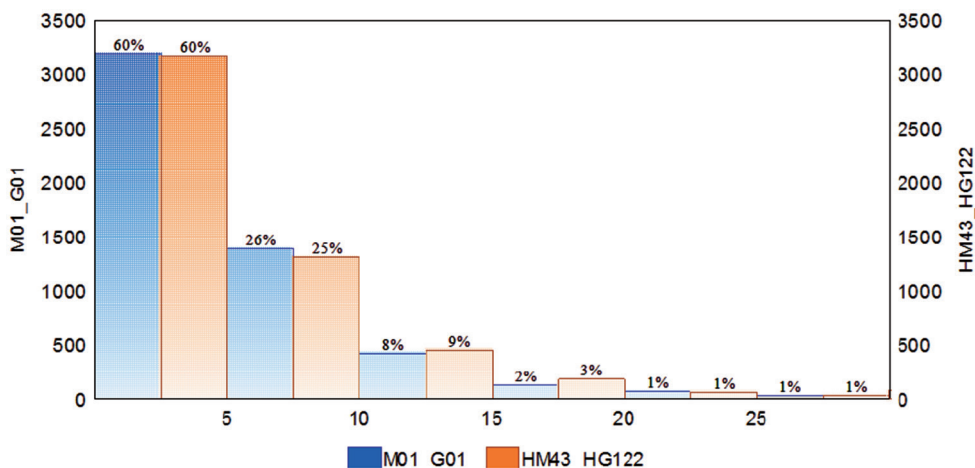
Modele	R^2	SE	V_s (w %)	MAPE I	MAPE E	Ranga_E	Ranga_I
HM43_G01	0,834	29,93	8,22	5,757	4,838	1	18
HM43_HG122	0,832	30,12	8,27	5,773	4,845	2	21
M01_HG122	0,848	28,68	7,87	5,573	5,47	146	2
M01_G/1	0,851	28,33	7,78	5,529	5,557	148	1
M01_G01(p)	0,847	28,48	7,77	–	5,274		
HM43_HG122(p)	0,824	30,32	8,33	–	5,271		

Źródło: opracowanie własne.

Z informacji zawartych w tab. 2 wynika, że najniższe i niemal identyczne oceny błędów prognoz ekstrapolacyjnych otrzymano dla modelu mieszanego HM43_G01 oraz modelu hierarchicznego ze złożoną sezonowością (HM43_HG122). W przypadku modeli M01_HG122 oraz M01_G01 oceny błędów były wyższe o ok. 0,63 i 0,72 p.p. Predyktory te znalazły się na miejscach 18 oraz 26 i 21 w rankingu ze względu na R^2 oraz MAPE_I. Na szczególne podkreślenie zasługuje fakt, że predyktory charakteryzujące się najlepszymi własnościami predyktywnymi i najniższymi ocenami błędów prognoz interpolacyjnych (M01_G01 oraz M01_HG122) w rankingu błędów prognoz ekstrapolacyjnych znalazły się dopiero na 146 i 148 miejscu. Oznacza to, że w przypadku występowania niesystematycznych luk w danych kryterium wyboru modelu dla celów prognozowania *ex ante* nie mogą być zarówno własności predyktywne, jak i oceny błędów prognoz interpolacyjnych.

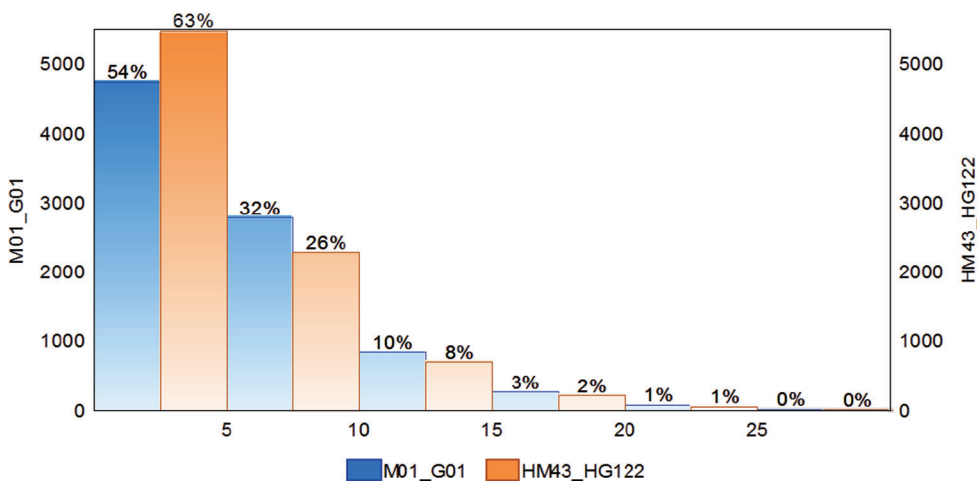
O słabej lub bardzo słabej zależności świadczą oceny współczynników korelacji między: R^2 a $MAPE_E$ oraz $MAPE_I$ i $MAPE_E$. Otrzymano następujące ich oceny: $r(R^2, MAPE_E) = 0,109$ oraz $r(R^2, MAPE_I, MAPE_E) = 0,222$.

Kształtowanie się empirycznych rozkładów błędów prognoz inter- i ekstrapolacyjnych przedstawione zostało w postaci histogramów na rys. 2 i 3 poniżej (dla przedziału 0-30% z odstępem wynoszącym 5 punktów procentowych).



Rys. 2. Rozkłady empiryczne błędów prognoz interpolacyjnych

Źródło: opracowanie własne.



Rys. 3. Rozkłady empiryczne błędów prognoz ekstrapolacyjnych

Źródło: opracowanie własne.

Z rysunku 2 wynika, że rozkłady te są niemal identyczne, różnice między nimi nie przekraczają 1 p.p. Nie jest to niespodzianką, ponieważ różnica ocen błędów prognoz interpolacyjnych wynosiła 0,05 p.p. W przypadku ocen błędów prognoz ekstrapolacyjnych różnica rozkładów jest znacznie bardziej widoczna, odnosi się to zwłaszcza do przedziału 0-5%. Dla trzech kolejnych przedziałów wyższe odsetki otrzymano dla modelu klasycznego M01_G01.

Z porównania ocen współczynników determinacji modelu dla niepełnych danych M01_G01 z modelem dla pełnych danych M01_G01(p) wynika, że pierwszy z nich jest ok. 0,4 p.p. wyższy. Sytuacja wygląda podobnie w przypadku modeli HM43_HG122 i HM43_HG122(p) – różnica wynosi ok. 0,8 p.p. Oceny błędów prognoz ekstrapolacyjnych otrzymane na podstawie predyktorów dla pełnych danych były ok. 10% wyższe od otrzymanych dla predyktora HM43_HG122.

Obecnie przechodzimy do analizy błędów obu rodzajów prognoz zdezagregowanych na miesiące, dni tygodnia oraz godziny. W przypadku prognoz interpolacyjnych dotyczyć one będą łącznie 5288 podokresów (miesiące, dni tygodnia oraz godzin), w których wystąpiły niesystematyczne luki w danych. Porównanie błędów obu rodzajów prognoz otrzymanych na podstawie modelu hierarchicznego i klasycznego możliwe będzie jedynie dla okresów, w których wystąpiły luki w danych.

W tabeli 3 zestawione zostały zdezagregowane błędy prognoz interpolacyjnych otrzymanych na podstawie modelu hierarchicznego (HM43_HG122) oraz modelu ze zmiennymi zero-jedynkowymi (M01_G01). W tabeli podane zostały także liczby prognoz dla poszczególnych podokresów. Spośród 12 miesięcy niższe od 1,51 do 2,96% oceny błędów prognoz dla predyktora hierarchicznego otrzymano dla 5 miesięcy: marca, czerwca, sierpnia, listopada i grudnia. Dla pozostałych miesięcy lepszy okazał się predyktor klasyczny. Maksymalną różnicę wynoszącą 1,70 p.p. (23,18%) otrzymano dla października. W przypadku dni tygodnia niższą ocenę błędu prognoz interpolacyjnych otrzymano jedynie dla niedzieli (0,41 p.p. – 3,95%). Dla wtorku i czwartku, a więc dni, w których wystąpiły luki, otrzymano oceny nieznaczące wyższe (odpowiednio o 0,39 p.p. i 1,83%).

Najwyższe różnice ocen błędów, w granicach 8-10%, otrzymano dla poniedziałku, środy i piątku. Model hierarchiczny okazał się bardziej efektywny dla godzin 7-9, 18-21. Najwyższą różnicę otrzymano dla godziny 19. Dla pozostałych 19 okresów godzinnych niższymi godzinami charakteryzował się model klasyczny. Maksymalną różnicę, wynoszącą 1,44 p.p. (21,34%), otrzymano dla godziny 24.

W tabeli 4 zestawione zostały zdezagregowane oceny błędów prognoz ekstrapolacyjnych. Ogólna ocena błędu tego rodzaju prognoz otrzymanych na podstawie modelu hierarchicznego to 0,71 p.p., tj. była ona o 14,76% niższa od błędu otrzymanego dla modelu klasycznego. Oceny niższe o 2,34 p.p. dla modelu klasycznego w układzie miesięcznym otrzymano dla października oraz lipca (0,16 p.p.). Dla pozostałych miesięcy bardziej efektywny okazał się predyktor hierarchiczny.

Tabela 3. Oceny zdezagregowanych błędów prognoz interpolacyjnych

Dni/ miesiące	L.prog.	H01_G01	HM43_ HG122	Godziny	L.prog.	H01_G01	HM43_ HG122
OG.	5288	5,54	5,76	G1	148	6,54	7,31
ST	392	4,40	4,41	G2	365	7,28	7,46
LU	352	3,45	3,53	G3	148	5,30	5,80
MARZ	392	3,81	3,70	G4	148	6,15	6,67
KW	720	6,68	6,80	G5	148	6,03	6,65
MAJ	408	5,25	5,40	G6	365	5,25	5,50
CZW	368	6,20	6,12	G7	148	5,68	5,10
LI	376	5,09	5,11	G8	365	6,49	6,86
SIE	408	6,97	6,77	G9	148	5,77	5,75
WRZ	368	6,62	6,84	G10	365	5,37	5,58
PAZ	744	5,64	7,34	G11	148	4,87	5,27
LIS	384	5,12	4,99	G12	148	4,95	5,44
GRUDZ	376	5,89	5,80	G13	148	4,86	5,42
PONIEDZA	552	7,61	8,43	G14	365	4,62	5,08
WTOREK	576	4,85	5,02	G15	148	4,81	5,65
ŚRODA	576	4,35	4,78	G16	365	4,88	5,46
CZWARTEK	1248	5,02	5,11	G17	148	5,55	5,73
PIĄTEK	544	4,38	4,76	G18	365	5,92	5,76
SOBOTA	1248	5,90	6,30	G19	148	7,61	6,17
NIEDZIELA	544	7,87	7,57	G20	148	5,87	4,79
				G21	148	5,31	5,19
				G22	365	4,21	4,36
				G23	148	4,37	4,96
				G24	148	5,32	6,77

Źródło: opracowanie własne.

Tabela 4. Oceny zdezagregowanych błędów prognoz ekstrapolacyjnych

Miesiące/dni	H01_G01	HM43_ HG122	Godziny	H01_G01	HM43_ HG122
1	2	3	4	5	6
Ogółem	5,56	4,84	G1	7,08	6,19
ST	4,32	3,93	G2	7,16	7,05
LU	5,40	3,53	G3	7,66	6,90
MARZ	6,27	5,08	G4	7,83	6,97
KW	6,05	3,99	G5	7,43	6,92

Tabela 4, cd.

1	2	3	4	5	6
MAJ	6,97	5,97	G6	6,55	6,06
CZW	6,90	5,74	G7	5,80	5,53
LI	5,47	5,63	G8	6,21	3,59
SIE	4,82	4,31	G9	5,30	4,04
WRZ	5,81	4,58	G10	4,37	3,22
PAZ	4,15	6,50	G11	4,28	3,84
LIS	4,64	3,92	G12	4,24	3,40
GRUDZ	5,86	4,81	G13	4,06	3,47
			G14	3,71	3,39
			G15	3,62	3,35
			G16	3,95	3,82
			G17	5,60	4,74
			G18	6,12	4,79
			G19	6,45	5,21
			G20	6,37	4,78
			G21	5,33	4,71
			G22	4,47	4,38
			G23	4,57	4,24
			G24	5,22	5,68

Źródło: opracowanie własne.

Różnice w dokładności wahały się od 10,03% w styczniu do 52,9% w lutym. Spośród dni tygodnia jedynie dla poniedziałku otrzymano nieznacznie niższą ocenę błędu (0,95%) dla predyktora klasycznego. Dla dni od wtorku do piątku różnica ocen błędów wahała się od 26,72% dla czwartku do 33,10% dla wtorku. Mimo braku ponad 30% obserwacji, predyktor hierarchiczny charakteryzował się wyższą dokładnością. Różnica w dokładności prognoz jest szczególnie widoczna w dezagregacji prognoz w układzie godzinnym. Jedynie dla godziny 24 oceną niższą o 8,04% charakteryzował się predyktor klasyczny. Dla pozostałych godzin niższe oceny otrzymano dla predyktora hierarchicznego. Błędy prognoz wahały się od 2,00% dla godziny 22 do 72,74% dla godziny 8.

4. Podsumowanie

1. Regularne modele hierarchiczne dla danych o wysokiej częstotliwości obserwowania mogą być budowane zarówno dla pojedynczych, jak i złożonych wahań sezonowych.

2. Dokładność prognoz interpolacyjnych otrzymanych na podstawie modelu klasycznego (M01_G01) była o ok. 4% wyższa niż dla najlepszego modelu hierarchicznego ze złożoną sezonowością (HM43_HG122). Model ten został sklasyfikowany na 21 miejscu wśród 160 oszacowanych równań uszeregowanych według rosnących wartości błędów prognoz interpolacyjnych.

3. Prognozy *ex post* zbudowane na podstawie modelu HM43_HG122 charakteryzowały się zdecydowanie wyższą dokładnością niż otrzymane na podstawie modelu klasycznego – różnica ocen wynosiła ponad 23%. Model klasyczny został sklasyfikowany dopiero na 146 miejscu.

4. Rozbieżność klasyfikacji ocen własności predyktywnych i błędów prognoz interpolacyjnych z ocenami błędów prognoz ekstrapolacyjnych świadczy o tym, że dwa pierwsze mierniki nie mogą być kryterium wyboru równań dla celów prognozowania *ex ante*.

5. Ocena błędu prognoz otrzymanych na podstawie modelu hierarchicznego dla luk niesystematycznych (ponad 30% obserwacji) była o ponad 10% niższa niż dla modeli dla pełnych danych. Potwierdza to istnienie „wyrównujących” własności modeli hierarchicznych.

6. W toku badań empirycznych potwierdzona została użyteczność modeli hierarchicznych w modelowaniu i prognozowaniu zmiennych ze złożoną sezonowością dla szeregów o wysokiej częstotliwości obserwowania z lukami w danych. Odnosi się to zwłaszcza do budowy prognoz ekstrapolacyjnych.

Literatura

- Kufel T., 2010, *Ekonometryczna analiza cykliczności procesów gospodarczych o wysokiej częstotliwości obserwacji*, Wydawnictwo Naukowe UMK w Toruniu, Toruń.
- Szmuksta-Zawadzka M., Zawadzki J., 2002, *Hierarchiczne modele szeregów czasowych z wahaniami sezonowymi. Budowa. Estymacja. Prognozowanie. Przestrzenno-czasowe modelowanie i prognozowanie zjawisk gospodarczych*, Wydawnictwo AE, Kraków, s. 193-204.
- Szmuksta-Zawadzka M., Zawadzki J., 2011, *Modelowanie i prognozowanie zapotrzebowania na energię elektryczną w mikroskali*, Prace Naukowe UE we Wrocławiu nr 165, s. 152-161.
- Szmuksta-Zawadzka M., Zawadzki J., 2013, *Regularne modele hierarchiczne dla szeregów czasowych o wysokiej częstotliwości* (w druku).
- Zawadzki J. (red.), 2003, *Zastosowanie hierarchicznych modeli szeregów czasowych w prognozowaniu zmiennych ekonomicznych z wahaniami sezonowymi*, Akademia Rolnicza, Szczecin.

HIERARCHICAL MODELS IN FORECASTING OF THE HIGH-FREQUENCY VARIABLES IN THE CONDITIONS OF LACK OF FULL INFORMATION

Summary: The paper presents a procedure of application of regular hierarchical models in forecasting missing data in high-frequency time series with cyclical fluctuations. Annual, weekly and daily cycles of seasonal fluctuation have additive character. Separately regular hierarchical models have been built for even length cycles. Theoretical considerations are illustrated with an empirical example.

Keywords: high-frequency data, hierarchical models, incomplete time series.