

O METODZIE PROGNOZOWANIA BRAKUJĄCYCH DANYCH W DZIENNYCH SZEREGACH CZASOWYCH Z LUKAMI SYSTEMATYCZNYMI

Maria Szmuksta-Zawadzka

Zachodniopomorski Uniwersytet Technologiczny w Szczecinie

Jan Zawadzki

Katedra Zastosowań Matematyki w Ekonomii

Zachodniopomorski Uniwersytet Technologiczny w Szczecinie

e-mail: jan.zawadzki@zut.edu.pl

Streszczenie: W pracy przedstawiono metodę modelowania a następnie prognozowania w sytuacji, gdy w szeregu czasowym dla danych dziennych występują luki systematyczne. Podstawą budowy prognoz były regularne hierarchiczne modele szeregu czasowego opisujące wahania o rocznym. Wahania o cyklu tygodniowym były opisywane za pomocą zmiennej grupującej, w skład której wchodziły dni podobne oraz tego rodzaju zmiennych dla pozostałych dni. W modelach wystąpiły także zmienne o charakterze migawkowym oznaczające występowanie świąt oraz dni około świątecznych. Rozważania o charakterze teoretycznym zostały zilustrowane przykładem empirycznym dla założonego wariantu luk w danych. Przeprowadzona została analiza dokładności błędów prognoz inter- i ekstrapolacyjnych ogółem oraz w dezagregacji na dni tygodnia, miesiące i święta oraz dni około świąteczne.

Słowa kluczowe: szeregi czasowe, dane dzienne, prognozowanie brakujących danych, luki systematyczne

WPROWADZENIE TEORETYCZNE

W modelowaniu i prognozowaniu na podstawie kompletnych szeregów czasowych dla danych dziennych można wykorzystywać m.in. modele szeregu czasowego z wahaniami okresowymi zawierające dwa zespoły zmiennych zerjedynkowych opisujących wahania o cyklu 7-dniowym oraz 12-miesięcznym.

Zapis analityczny takiego modelu z liniowym trendem jest następujący [por. Kufel 2010; Szmuksta-Zawadzka, Zawadzki 2011]:

$$Y_t = \alpha_1 t + \alpha_0 + \sum_{k=1}^7 d_{0k} Q_{kt} + \sum_{s=1}^{12} d_{0s} Q_{sMt} + \sum_{j=1}^r a_{0j} D_{jt} + U_t, \quad (1)$$

gdzie: $\sum_{k=1}^7 d_{0k} = \sum_{s=1}^{12} d_{0s} = 0$.

W modelu tym uwzględniono także zmienne D_{jt} oznaczające występowanie dni świątecznych oraz około świątecznych. Na parametry d_{0k} oraz d_{0i} nałożony został warunek sumowalności do zera. Następstwem użycia do opisu wahań dwóch zespołów zmiennych zerojedynkowych będzie to, że dla każdego rodzaju wahań szacowanych będzie o jeden parametr mniej niż wynosi długość każdego z cykli. Oznacza to, że w podmacierzach zawierających zmienne Q_{kt} oraz Q_{sMt} w jednym z wierszy (zazwyczaj ostatnim w ramach cyklu) wystąpią elementy równe minus jeden. Parametry d_{0k} oraz d_{0i} interpretuje się wówczas jako odchylenia od wyrazu wolnego, będącego średnią z wyrazów wolnych dla poszczególnych dni i miesięcy.

W innym, równoważnym zapisie, nie nakłada się na parametry warunku sumowalności do zera, zastępując elementy równe minus jeden zerami. Wyraz wolny odnosić się będzie wtedy na ogół do ostatnich okresów cykli na przykład niedzieli i grudnia. Parametry dla poszczególnych dni i miesięcy będą odchyleniami od tak zdefiniowanego wyrazu wolnego.

Z literatury przedmiotu wiadomo, że w procesie modelowania i prognozowania dla danych w postaci szeregów czasowych z wahaniami sezonowymi dla okresów jednostkowych nie krótszych niż dekada użyteczne są także regularne modele hierarchiczne. W modelach tych podzielniki p_i cyklu wahań m spełniają jednocześnie dwa warunki [Szmuksta-Zawadzka, Zawadzki 2002]:

$$2 \leq p_i \leq 0 \frac{m}{2} \quad (2)$$

$$\prod_{i=1}^{\frac{m}{2}} p_i = m \quad (3)$$

Dla danych miesięcznych wyróżnia się siedem modeli hierarchicznych (cztery dwustopniowe i trzy trójstopniowe). Natomiast dla danych o cyklu 36 dekadowym modeli tych 25, w tym 5 dwu-, 14 trój-, 6 czterostopniowych.

W przypadku danych dziennych, ze względu na nieparzystą liczbę dni w tygodniu, nie jest możliwa budowa regularnych modeli hierarchicznych. Dlatego, podobnie jak w modelach klasycznych, osobno należy modelować wahania o cyklu tygodniowym oraz o cyklu rocznym.

Wahania o cyklu rocznym można opisać za pomocą regularnych modeli hierarchicznych, w których parametry dla poszczególnych stopni hierarchii są średnimi z ocen parametrów dla odpowiednich miesięcy.

Sprawa opisu wahań o cyklu tygodniowym jest bardziej skomplikowana ze względu na wspomnianą wyżej, nieparzystą liczbę dni w tygodniu. Dlatego też zaproponować można procedurę polegającą na grupowaniu dni podobnych. Dokonując wyboru dni podobnych należy brać pod uwagę przede wszystkim kryteria oparte na przesłankach merytorycznych. Przykładem takiego kryterium może być podział na dni robocze i dni świąteczne. Z kolei dni robocze można podzielić na dni z początku tygodnia oraz jego końca itp. Wyodrębnienie dni podobnych można przeprowadzić także kierując się zbliżonymi wielkościami średnich wartości zmiennej prognozowanej lub też zbliżonymi ocenami parametrów w modelach dla pełnych danych. Ocena parametru przy zmiennej grupującej będzie równa średniej z ocen dla dni podobnych. Natomiast pozostałe dni będą uwzględnione w modelu bez zmian. Ponadto w modelach mogą występować także zerojedynkowe zmienne migawkowe, charakteryzujące występowanie dni świątecznych i około świątecznych.

Sposób budowy oraz liczba dni wchodzących w skład zmiennej grupującej odgrywa istotną rolę w przypadku modelowania i prognozowania w warunkach występowania systematycznych luk w danych. Występowanie luk tego rodzaju oznacza, że przynajmniej o jednym dniu tygodnia lub (i) przynajmniej o jednym miesiącu nie posiadamy żadnej informacji.

Do modelowania i prognozowania, w omawianej sytuacji, nie mogą być wykorzystane klasyczne modele szeregu czasowego, ze względu na zerowe wariancje zmiennych odnoszących się do tych podokresów (miesiące i dni), w których występują luki. Natomiast do tego celu, mogą być wykorzystane regularne modele hierarchiczne opisujące wahania o cyklu rocznym. Wahania o cyklu tygodniowym będą opisywane za pomocą zmiennej grupującej oraz zmiennych zerojedynkowych nie wchodzących w jej skład. Zmienna grupująca powinna zawierać przynajmniej jeden dzień, w którym nie występują luki. Prognozy zarówno interpolacyjne jak i ekstrapolacyjne dla okresów, w których występują z luki, będą sumami: iloczynów parametru kierunkowego trendu, wyrazu wolnego, parametrów modelu hierarchicznego dla poszczególnych stopni hierarchii oraz parametru dla dni podobnych, wyznaczonego dla dnia, o których dostępne były informacje. Dla pozostałych dni dodawany będzie, zamiast parametru przy zmiennej grupującej, parametr dla określonego dnia.

Występowanie luk systematycznych może pociągnąć za sobą daleko idące następstwa związane z wyborem istotnych składowych w modelach hierarchicznych. W zależności od liczby i rozmieszczenia luk systematycznych następstwa te mogą być następujące [Zawadzki(red.) 2003]:

1. Przynajmniej część składowych należących do różnych stopni hierarchii może być ze sobą silnie skorelowana, a niektóre z nich mogą być nawet dokładnie współliniowe.
2. Część składowych może przyjmować stałe wartości, a więc charakteryzować się zerowymi wariancjami - składowe te należy wyeliminować jeszcze przed rozpoczęciem procesu modelowania.
3. Ponadto niektóre składowe mogą być liniowymi kombinacjami innych składowych. Oznaczać to będzie konieczność szacowania wielu wersji modeli różniących się zestawami składowych nie tworzących takich kombinacji. Wszystkie te wersje charakteryzować się będą identycznymi ocenami parametrów struktury stochastycznej. Tak więc będą one nierozróżnialne z punktu widzenia własności prognostycznych. Różnić się będą natomiast dokładnością prognoz *ex post*. Dlatego też dla celów prognozowania *ex ante* wybierane będą modele o minimalnych ocenach błędów prognoz *ex post*.

Egzemplifikacją przeprowadzonych wyżej rozważań o charakterze teoretycznym będzie przykład empiryczny dotyczący modelowania i prognozowania dziennej sprzedaży paliw płynnych w jednej ze stacji benzynowych.

WYNIKI BADAŃ EMPIRYCZNYCH.

Rozpatrywany będzie jeden wariant występowania systematycznych luk w danych. Luki będą występować w dwóch spośród siedmiu dni tygodnia (w środy i w piątki) oraz w dwóch spośród 12 miesięcy (w sierpniu i w październiku). Oznacza to, że dostępnych będzie 430 spośród 724 obserwacji dla całego szeregu, a więc obejmować będą około 40 procent jego długości. Luki otrzymano przez „wymazanie” 294 obserwacji z pełnego szeregu.

Szacowane będą hierarchiczne modele szeregu czasowego z liniowym trendem oraz składowymi opisującymi wahania o cyklu rocznym i tygodniowym. Do opisu wahań o cyklu 12-miesięcznym zostaną wykorzystane klasyczne regularne dwu- i trójstopniowe modele hierarchiczne. Natomiast wahania o cyklu tygodniowym opisywane będą za pomocą zmiennej grupującej obejmującej cztery spośród pięciu dni roboczych (oprócz czwartku). Oznacza to, że prognozy dla tych dni, w tym dla środy i piątku, będą średnią z ocen dla poniedziałku i wtorku. Dokonując wyboru zmiennej grupującej (GR) kierowano się zbliżonymi ocenami parametrów modelu dla tych dni dla pełnych danych [por. Szmuksta-Zawadzka, Zawadzki 2011]. W modelach wystąpią także zmienne zerojedynkowe dla czwartku i soboty oraz tego rodzaju zmienne dla dni świątecznych i około świątecznych. W celach porównawczych, odnoszących się do: stopnia opisu oraz dokładności *ex post* prognoz ekstrapolacyjnych, został oszacowany model (1) a następnie na jego zostały wyznaczone prognozy.

W tabeli 1 zestawione zostały oceny parametrów struktury stochastycznej: współczynników determinacji (R^2), odchyłeń standardowych składników losowych (S_e), współczynników zmienności losowej (V_{SE}) oraz statystyk empirycznych testu Durbin-Watsona (DW).

Tabela 1 Oceny parametrów struktury stochastycznej modeli hierarchicznych

Model	R^2	SE(l)	VS(%)	DW
PEL	0,527	682,8	13,50	1,94
H26	0,402	764,8	15,50	1,86
H34	0,468	720,6	14,61	2,10
H43	0,481	711,4	14,42	2,14
H62	0,470	719,9	14,60	2,11
H223	0,401	763,3	15,47	1,86
H232	0,389	771,4	15,64	1,83
H322	0,444	735,8	14,91	2,02

Źródło: opracowanie własne

Z informacji zawartych w tabeli wynika, że najwyższą ocenę współczynnika determinacji wynoszącą 52,65% oraz najniższą oceną współczynnika zmienności losowej (13,50%) charakteryzuje się model klasyczny (PEL). Spośród modeli hierarchicznych zbliżone do niego własności predyktywne posiadają trzy modele dwustopniowe; H43, H34 i H62 oraz jeden model trójstopniowy - H322. Oceny współczynników determinacji dla tych modeli mieszczą się w przedziale od 44,4 do 48,1% a współczynniki zmienności losowej od 14,42 do 14,91%.

W przypadku pozostałych trzech modeli oceny współczynników zmienności losowej kształtują się w granicach 15,47-15,64%.

Dokonując oceny kształtowania się własności predyktywnych oszacowanych równań możemy uznać je za zadowalające. Na zjawisko malejącej dokładności opisu wraz ze skracaniem długości okresu jednostkowego z jednocześnie dopuszczalnymi ocenami współczynników zmienności losowej, zwracają uwagę Hozer, Zawadzki [1990]. Statystyki empiryczne testu Durbin-Watsona zawarte w przedziale od 1,830 do 2,289 informują o bardzo słabym dodatnim lub ujemnym skorelowaniu reszt rzędu pierwszego.

Na podstawie oszacowanych modeli zostały wyznaczone dwa rodzaje prognoz: interpolacyjne i ekstrapolacyjne. Prognozy interpolacyjne dotyczą tych okresów (dni, miesięcy i świąt oraz dni przed- i poświętecznych), w których

wystąpiły luki w danych. Natomiast prognozy ekstrapolacyjne zostały wyznaczone dla kolejnego roku, bezpośrednio następującego po okresie estymacyjnym.

W tabeli 2 zestawione zostały oceny błędów prognoz interpolacyjnych. Odnosić się one będą do tych dni i miesięcy, w których wystąpiły luki w danych, a także do jedyne go święta (tj. Święta Wniebowzięcia Najświętszej Marii Panny (WNMP)) przypadające w dniu 15 sierpnia oraz dnia je poprzedzającego i po nim następującego. Zostaną także przeanalizowane błędy prognoz dla zmiennej grupującej (GR).

Tabela 2 Oceny średnich względnych błędów prognoz interpolacyjnych(w %)

	Liczba prognoz	H26	H34	H43	H62	H223	H232	H322
OG	294	10,14	9,35	9,01	10,97	10,20	9,62	8,86
Q3	104	12,16	11,79	11,46	14,74	12,42	12,29	11,57
Q5	103	12,41	11,40	11,04	13,86	12,56	12,33	11,26
GR	240	12,10	11,28	10,80	13,53	12,22	11,87	10,99
Q8M	62	13,07	10,15	8,79	10,70	12,96	9,93	8,79
Q10M	62	10,33	10,95	11,25	10,44	10,06	9,90	9,74
WNMP	2	5,21	12,60	4,04	2,13	5,05	2,81	7,48
P_WNMP	4	22,17	26,25	23,56	22,68	22,20	23,01	24,14

Źródło: opracowanie własne

Z informacji zawartych w tabeli wynika, że błędy względne prognoz interpolacyjnych ogółem, otrzymanych na podstawie modeli hierarchicznych, zawierają się w przedziale od 8,86 procent dla modelu H322 do 10,97 procent dla modelu H62.

Z analizy kształtowania się ocen błędów prognoz dla środy wynika, że dla sześciu spośród siedmiu modeli hierarchicznych (z wyłączeniem modelu H62), oceny te różnią się mniej niż 1,5 punktu procentowego od oceny maksymalnej prognoz ogółem. Zdecydowanie wyższą ocenę błędu, wynoszącą 14,74%, otrzymano dla modelu H62. Oceny błędów dla piątku kształtują się na poziomie niższym o około 0,5 punktu procentowego w porównaniu ze środą. Minimalna jego ocena jest niższa o 0,42 punktu procentowego. Niższe jest także zróżnicowanie ocen – ich rozstępy przyjęły następujące wartości: 3,28 punktu dla środy i 2,82 punktu dla piątku. Minimalną ocenę łączną dla dni wchodzących w skład zmiennej grupującej (obejmującej cztery dni robocze (oprócz czwartku)), wynoszącą 10,80% otrzymano dla modelu H43. Jest ona nieznacznie niższa od minimalnych ocen dla środy i piątku (odpowiednio o 0,66 i 0,24 punktu). Oznacza to, że błędy prognoz dla poniedziałku i wtorku, nie dotyczące dni, w których wystąpiły luki, powinny być niższe od otrzymanych dla środy i piątku. Tak jest w istocie - przyjęły one wartości wynoszące one odpowiednio: 7,33 i 8,49%.

Z porównania minimalnych ocen błędów prognoz dla miesięcy wynika, że dla sierpnia jest ona bardzo nieznacznie niższa (o 0,07 punktu), a dla października mniej niż o 1 punkt procentowy wyższa od oceny minimalnej prognoz ogółem. Wprawdzie dla października otrzymano ocenę minimalną nieco wyższą, ale błędy prognoz dla tego miesiąca charakteryzują się niższym zróżnicowaniem. Ich rozstęp dla października wynosi 1,51 punktu procentowego a dla sierpnia 4,28 punktu.

Dla święta WNMP oceny błędów prognoz otrzymanych na podstawie modeli hierarchicznych są znacznie niższe od ocen otrzymanych na podstawie tych modeli dla dnia je poprzedzającego i po nim następującego. Różnica ocen minimalnych błędów prognoz otrzymanych odpowiednio dla predyktorów H62 i H26 przekracza 20 punktów procentowych.

W tabeli 3 zestawione zostały błędy prognoz ekstrapolacyjnych. Zmienne zerojedynkowe Q_{kt} oraz Q_{smt} oznaczają w tej tabeli odpowiednio: dni tygodnia oraz miesiące w roku, dni świąteczne zastąpione zostały skrótami nazw tych świąt a nazwy dni przed- i po świątecznych są skrótami nazw świąt poprzedzonych symbolem P_:

- NROK - zmienna przyjmująca wartość jeden w dniu Nowego Roku,
- P_NROK - zmienna przyjmująca wartość jeden w dniu poprzedzającym Nowy Rok i dzień po nim następujący,
- WIELK - zmienna przyjmująca wartość jeden w pierwszym i drugim dniu Wielkanocy,
- P_WIELK - zmienna przyjmująca wartość jeden w dniu poprzedzającym Wielkanoc i dzień następujący po jej zakończeniu,
- M1_3 - zmienna przyjmująca wartość jeden w dniach 1-3 maja,
- P_M1_3 - zmienna przyjmująca wartość jeden w dniu 30 kwietnia i 4 maja,
- BC - zmienna przyjmująca wartość jeden w dniu święta Bożego Ciała,
- P_BC - zmienna przyjmująca wartość jeden w dniu poprzedzającym święto Bożego Ciała i dzień po nim następujący,
- WNMP - zmienna przyjmująca wartość jeden w dniu Święta Wniebowzięcia Najświętszej Panny Marii (15 sierpnia),
- P_WNMP - zmienna przyjmująca wartość jeden w dniach 14 i 16 sierpnia,
- WSW - zmienna przyjmująca wartość jeden w dniu święta Wszystkich Świętych (1 listopada),
- P_WSW - zmienna przyjmująca wartość jeden w dniu 31 października oraz 2 listopada,
- Li11 - zmienna przyjmująca wartość jeden w dniu Święta Niepodległości (11 listopada),
- P_Li11 - zmienna przyjmująca wartość jeden w dniach: 10 i 12 listopada,
- BN - zmienna przyjmująca wartość jeden w pierwszy i drugi dzień Świąt Bożego Narodzenia,

- P_BN - zmienna przyjmująca wartość jeden w dniach 24 i 27 grudnia.

Tabela 3. Oceny błędów prognoz ekstrapolacyjnych

	PEL	H34	H43	H26	H62	H223	H232	H322
Og.	13,46	16,01	15,19	14,30	16,28	11,34	11,39	13,18
Q1	16,83	15,08	14,36	14,52	15,02	12,83	12,98	13,63
Q2	10,95	14,80	14,17	13,59	16,38	11,26	12,27	13,07
Q3	11,92	12,63	12,18	10,78	14,27	9,01	9,07	11,25
Q4	12,46	19,96	19,86	20,01	21,63	16,72	17,11	18,46
Q5	13,32	12,78	11,72	10,23	10,90	7,04	7,30	8,50
Q6	17,85	26,53	24,27	22,02	25,46	15,23	14,80	19,63
Q7	12,23	14,28	14,19	15,26	16,91	15,39	14,99	19,10
GR	13,32	18,53	17,66	16,65	19,48	16,59	13,34	15,64
Q1M	14,77	17,31	17,09	18,88	17,33	19,09	17,94	17,85
Q2M	11,46	10,13	10,38	9,93	15,04	9,92	10,10	10,35
Q3M	6,58	11,46	11,22	13,23	14,18	8,57	8,95	7,24
Q4M	10,29	13,53	15,73	14,88	16,05	14,40	13,88	13,54
Q5M	18,11	25,49	22,01	15,02	25,76	14,95	16,06	26,56
Q6M	16,35	18,97	18,99	13,77	15,63	14,35	12,36	22,22
Q7M	10,19	13,66	10,87	8,73	17,90	8,82	8,63	12,12
Q8M	11,67	20,21	11,86	10,86	10,62	10,80	11,40	15,61
Q9M	14,57	13,96	19,79	15,78	17,35	15,57	16,19	14,83
Q10M	16,52	14,87	14,90	17,30	14,74	16,63	16,49	14,64
Q11M	17,08	16,46	15,04	18,01	15,32	17,88	19,21	17,17
Q12M	15,57	16,46	14,02	17,26	15,31	17,44	16,93	15,78
NROK	14,62							
P NROK								
WIELK	16,77	21,87	22,32	21,82	22,65	21,78	21,59	22,31
P WIELK								
M1 3	17,35							
P M1 3	3,84	17,62	18,95	16,52	23,38	16,82	17,85	17,29
BC	7,84	8,45	10,26	11,34	9,12	11,79	9,69	10,30
P BC								
WNMP	14,74	6,91	3,20	14,57	6,80	14,41	8,14	1,69
P WNMP	18,48	10,92	1,41	11,52	3,55	11,34	4,77	5,56
WSW	1,29	15,45	15,52	15,63	12,82	15,65	15,53	14,46
P WSW	19,74	6,31	6,70	6,67	7,17	6,67	6,63	5,75
Li11								
P Li11	1,67	5,49	5,40	5,51	5,33	5,52	5,58	5,60

Źródło: opracowanie własne

W tabeli zamieszczone zostały oceny błędów prognoz dla tych dni świątecznych i około świątecznych, które nie przekraczają przyjętego kryterium dopuszczalności prognoz na poziomie 25%. Dla dni i miesięcy przyjęto kryterium o 5 punktów procentowych niższe. Z informacji zawartych w tabeli wynika,

że oceny błędów prognoz ekstrapolacyjnych ogółem otrzymane dla modeli hierarchicznych mieszczą się w przedziale od 11,34 dla modelu H223 do 16,28 % dla modelu H62. W ich kształtowaniu się widoczna jest prawidłowość polegająca na tym, że oceny błędów prognoz otrzymanych na podstawie predyktorów dwustopniowych (H26, H34, H46 i H62) są wyższe od ocen błędów prognoz dla predyktorów trójstopniowych (H223, H232, H322). Maksymalna ocena błędu prognoz otrzymana dla predyktora H322, wynosząca 13,18% jest o 1,12 punktu procentowego niższa od minimalnej oceny błędu dla predyktora dwustopniowego. Oceny błędów prognoz dla predyktorów trójstopniowych są też niższe od oceny błędu prognoz otrzymanych na podstawie predyktora, którego parametry były szacowane z wykorzystaniem pełnych danych.

W przypadku prognoz ekstrapolacyjnych dla dni tygodnia otrzymanych na podstawie modeli hierarchicznych, także widoczna jest różnica w ich efektywności między modelami dwu- i trójstopniowymi. Prognozy otrzymane na podstawie predyktorów trójstopniowych charakteryzują się ocenami błędów niższymi od 1 do 3 punktów procentowych od ocen otrzymanych dla predyktorów dwustopniowych. Dla dni roboczych najbardziej efektywny, tzn. charakteryzujący się najniższą oceną względnego błędu prognoz, okazał się predyktor H223, a dla soboty był to predyktor H232. Jedynie dla niedzieli nieco wyższą dokładnością charakteryzowały się dwa predyktory dwustopniowe (H34 i H43). Z informacji zawartych w tabeli wynika, że minimalne oceny błędów prognoz otrzymanych na podstawie modeli hierarchicznych dla trzech dni tygodnia (poniedziałku, środy i piątku) oraz błęd prognoz ogółem, są niższe od odpowiadających im błędów prognoz otrzymanych na podstawie predyktora szacowanego na podstawie pełnych danych.

Z informacji dotyczących kształtowania się ocen błędów prognoz ekstrapolacyjnych dla miesięcy otrzymanych na podstawie predyktora bez luk wynika, że wahają się one w przedziale od 6,58 procent dla marca do 20,39 procent dla maja. Ocenę poniżej 10 procent otrzymano jeszcze dla kwietnia. Dokładność prognoz otrzymanych na podstawie predyktorów hierarchicznych była bardziej zróżnicowana niż w przypadku prognoz ogółem i niż dla dni tygodnia

Z porównania minimalnych ocen błędów otrzymanych na podstawie modeli hierarchicznych i modelu dla pełnych danych wynika, że dla pierwszych siedmiu miesięcy oraz września i października bardziej dokładne okazały się predyktory oparte na równaniach hierarchicznych. Różnice te mieściły się w przedziale od 0,18 punktu procentowego dla lutego do 7,95 punktu dla lipca. Dla pozostałych trzech miesięcy nieznacznie lepszy okazał się model dla pełnych danych z różnicą w dokładności od 0,11 punktu dla grudnia do 0,41 dla listopada. Jednocześnie zwraca uwagę fakt, że w odróżnieniu od dni tygodnia, dokładność dwu- i trójstopniowych predyktorów hierarchicznych dla miesięcy była zbliżona. Minimalne oceny błędów prognoz w poszczególnych miesiącach zawarte są w przedziale od 7,24% dla lutego do 17,09% procent dla stycznia, otrzymano dla sześciu z siedmiu predyktorów (poza H62) - zostały one zaznaczone „tłustym

drukiem”. Dla każdego z tych modeli oceny minimalne ocen błędów otrzymano dla dwóch miesięcy. Oznacza to, że należy wykorzystywać różne modele hierarchiczne dla celów prognozowania. Oceny niższe od 10 procent otrzymano jeszcze dla lipca i lutego. Ocenami powyżej 15 procent charakteryzują się ostatnie miesiące roku a bliską tej wielkości jest także ocena dla maja.

Z analizy błędów prognoz dla dni świątecznych oraz przed- i poświątecznych wynika, że prognozy dopuszczalne otrzymano jedynie dla Święta Wniebowzięcia Najświętszej Marii Panny oraz Wszystkich Świętych.

W pozostałych przypadkach dopuszczalne były prognozy bądź tylko dla samych Świąt (Nowego Roku, Wielkanocy, Bożego Ciała) bądź dni poprzedzających i następujących po świątach: 1_3 Maja i Niepodległości oraz dni po nich następujących. Niemożność otrzymania prognoz dopuszczalnych dla pozostałych świąt i dni około świątecznych wynikała z dużej skali wahań o charakterze losowym zakłócającym wyrazistość prawidłowości.

PODSUMOWANIE

Z analizy błędów prognoz inter- i ekstrapolacyjnych ogółem oraz w dezagregacji na dni tygodnia, miesiące oraz część dni świątecznych oraz około świątecznych wynika praktyczna użyteczność predyktorów hierarchicznych w prognozowaniu brakujących danych, w którym luki systematyczne obejmowały około 40 procent długości szeregu. W pełni zasadna okazała się także procedura wykorzystania zmiennej grupującej obejmującej cztery dni robocze (bez czwartku), z których w dwóch wystąpiły braki w danych.

BIBLIOGRAFIA

- Hozer J., Zawadzki J. (1990) *Zmienna czasowa i jej rola w badaniach ekonometrycznych*, PWN, Warszawa.
- Kufel T. (2010) *Ekonometryczna analiza cykliczności procesów gospodarczych o wysokiej częstotliwości prognozowania*, Wydawnictwo Uniwersytetu Mikołaja Kopernika, Toruń.
- Szmuksta – Zawadzka M., Zawadzki J. (2002) *Hierarchiczne modele szeregów czasowych z wahaniami sezonowymi. Budowa. Estymacja. Prognozowanie.*, [w:] *Przestrzenno-czasowe modelowanie i prognozowanie zjawisk gospodarczych*, Wyd. AE, Kraków, str. 193-204.
- Szmuksta-Zawadzka, M., Zawadzki, J. (2011) *The ex post analysis of the forecast accuracy of liquid fuel daily sales*, *Research Papers of Wrocław University of Economics, Econometrics* nr 32, s. 194-204.
- Zawadzki J. (red) (2003) *Zastosowanie hierarchicznych modeli szeregów czasowych w prognozowaniu zmiennych ekonomicznych z wahaniami sezonowymi*, Akademia Rolnicza, Szczecin.

**ABOUT METHOD OF FORECASTING THE MISSING DATA
IN DAILY TIME SERIES WITH SYSTEMATIC GAPS**

Abstract: This paper presents a method for modeling and then forecasting in situation, when in time series for daily data contain systematic gaps. Base of construction were regular hierarchical time series models describing annual fluctuations. Weekly fluctuations were described as a grouping variable, which contains similar days and this type variables for other days. In models were used also dummy variables describing holidays and days pre- and post- holidays. Theoretical considerations were illustrated by empirical example for selected variant of gaps. Based on the same estimated equations, inter- and extrapolation predictions were built. For both types of prediction – in general and in disaggregation to weekdays and months and holidays and days pre- and post holidays.

Keywords: data gaps, daily data, forecasting