

Maria Szmuksta-Zawadzka, Jan Zawadzki

Zachodniopomorski Uniwersytet Technologiczny w Szczecinie

ZASTOSOWANIE MODELOWANIA EKONOMETRYCZNEGO W PROGNOZOWANIU BRAKUJĄCYCH DANYCH W SZEREGACH O WYSOKIEJ CZĘSTOTLIWOŚCI

Streszczenie: W pracy przedstawione zostały wyniki modelowania i prognozowania w warunkach braku pełnej informacji na przykładzie zapotrzebowania na energię elektryczną w regionie dla danych półgodzinnych. Prognozy zostaną zbudowane na podstawie modeli opisujących za pomocą zmiennych zero-jedynkowych trzy rodzaje wahań periodycznych o cyklach: rocznym, tygodniowym i dobowym. Ponadto będą one zawierać tego rodzaju zmienne dla oznaczenia dni świątecznych oraz opóźnioną zmienną prognozowaną. W toku empirycznej weryfikacji prognoz przeprowadzona zostanie analiza dokładności prognoz ekstrapolacyjnych i interpolacyjnych.

Słowa kluczowe: dane o wysokiej częstotliwości, prognozowanie, brakujące dane.

1. Wstęp

Celem pracy jest wykazanie przydatności modeli szeregu czasowego z trzema zespółami zmiennych zero-jedynkowych w prognozowaniu zmiennej ekonomicznej o bardzo wysokiej częstotliwości obserwowania w warunkach występowania niesystematycznych luk w danych. Dane statystyczne o kształtowaniu się zapotrzebowania w okresach półgodzinnych wykorzystane do modelowania, a następnie prognozowania dotyczą wybranej jednostki terytorialnej firmy zajmującej się dystrybucją energii elektrycznej. Otrzymanie równania o dość dobrych własnościach predyktywnych pozwoli na wyznaczenie prognoz interpolacyjnych i ekstrapolacyjnych oraz przeprowadzenie analizy ich dokładności. Uzyskanie prognoz dopuszczalnych świadczyć będzie o praktycznej przydatności proponowanej procedury.

2. Metodologia badawcza

Przykład wykorzystania ekonometrycznych modeli zgodnych do budowy prognoz zapotrzebowania na energię elektryczną dla okresów godzinnych można znaleźć w pracy [Kufel 2010].

Natomiast w pracy [Szmuksta-Zawadzka, Zawadzki 2011] przedstawione zostały wyniki modelowania i prognozowania na energię w okresach półgodzinnych w mikroskali dla pełnych danych. Do modelowania, a następnie prognozowania zapotrzebowania na moc elektryczną w okresach półgodzinnych zostały w niej wykorzystane modele z liniowym trendem oraz trzema zespołami zmiennych zero-jedynkowych opisujących wahania okresowe o cyklach wynoszących odpowiednio: 12 miesięcy, 7 dni oraz 24 godziny (48 okresów półgodzinnych). Kolejne zmienne zero-jedynkowe opisywały występowanie świąt oraz jednego z dni przedświątecznych. Modele zawierały także opóźnioną zmienną endogeniczną z opóźnieniem 24-godzinnym lub jednogodzinnym.

Szacowane były modele, w których parametry spełniały warunek sumowalności do zera:

$$Y_t = \alpha_1 t + \alpha_0 + \sum_{i=1}^{12} b_{0i} M_{it} + \sum_{j=1}^7 c_{0j} D_{jt} + \sum_{k=1}^{48} d_{0k} P_{kt} + \sum_{l=1}^r a_{l0} S_{lt} + \delta_{1s} Y_{t-s} + U_t, \quad (1)$$

$$\sum_{j=1}^7 c_{0j} = \sum_{i=1}^{12} b_{0i} = \sum_{k=1}^{48} d_{0k} = 0; (i = 1, 2, \dots, 12; j = 1, \dots, 7; k = 1, 2, \dots, n). \quad (2)$$

Oznacza to, że w macierzy X w wierszach dla zmiennych zero-jedynkowych: M_{it} , D_{jt} oraz P_{kt} , odpowiadających ostatnim okresom cykli, występowały elementy równe minus jeden. Otrzymane wyniki prognozowania dla pełnych danych będą punktem wyjścia i jednocześnie punktem odniesienia do przeprowadzonych w niniejszej pracy badań dotyczących prognozowania w warunkach braku pełnej informacji.

W niniejszej pracy szacowana będzie postać równoważna modelowi (1):

$$Y_t = \alpha_1 t + \alpha_{jik} + \sum_{i=1}^{11} b_{0i} M_{it} + \sum_{j=1}^6 c_{0j} D_{jt} + \sum_{k=1}^{47} d_{0k} P_{kt} + \sum_{l=1}^r a_{l0} S_{lt} + \delta_s Y_{t-s} + U_t, \quad (3)$$

($s = 48; j = 7; i = 12; k = 48$).

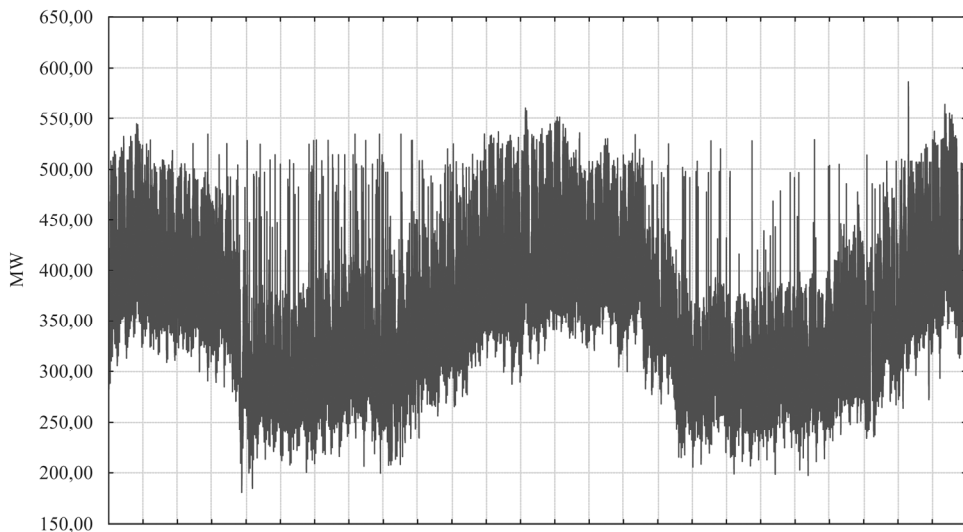
Z zapisu powyższego równania wynika, że wyraz wolny α_{jik} ($j = 7, i = 12, k = 48$) jest wielkością odnoszącą się do 7 dnia tygodnia, 12 miesięcy oraz 48 okresu półgodzinnych. Oznacza to, że parametry: b_{0i} , c_{0j} , d_{0k} interpretowane będą jako odchylenia od odpowiedniego elementu składowego wyrazu wolnego.

Szacowanie parametrów modelu danego równaniem (3) na podstawie danych, w których występują luki, jest wskazane ze względów zarówno numerycznych, jak i interpretacyjnych.

3. Wyniki estymacji modelowania zapotrzebowania na energię elektryczną dla okresów półgodzinnych dla pełnych i niepełnych danych

Dane statystyczne o kształtowaniu się zapotrzebowania na moc w okresach półgodzinnych (w MWh) pochodzą z Banku Danych Katedry Zastosowań Matematyki w Ekonomii ZUT w Szczecinie. Dane obejmowały dwa lata, tj. 35 040 obserwacji. Trzeci rok był okresem empirycznej weryfikacji prognoz. Dane te zostały wykorzystane w szacowaniu parametrów modeli postaci (1) i (3). W modelowaniu i prognozowaniu zostanie wykorzystany model ze zmienną prognozowaną opóźnioną o 48 okresów półgodzinnych. Rozpatrywany będzie wariant, w którym luki występują w miesiącach nieparzystych w pierwszym roku oraz w miesiącach parzystych w drugim roku. Oznacza to, że parametry będą szacowane na podstawie szeregu zawierającego połowę spośród 35 040 obserwacji. Ze względu na to, że nie występują podokresy, dla których nie dysponujemy przynajmniej 48 obserwacjami, parametry modelu (3) mogą być szacowane MNK. Na podstawie oszacowanego modelu zostaną obliczone prognozy interpolacyjne (dla okresów, w których występowały luki) oraz prognozy ekstrapolacyjne na kolejny rok. Zostanie przeprowadzona analiza dokładności prognoz dla okresu interpolacyjnego i ekstrapolacyjnego oraz w dezagregacji na: dni tygodnia, miesiące, poszczególne okresy półgodzinne oraz dni świąteczne.

Kształtowanie się zmiennej prognozowanej w okresie estymacyjnym (bez luk) przedstawiono na rys. 1.



Rys. 1. Kształtowanie się zapotrzebowania na energię elektryczną w okresach półgodzinnych (w MW)

Źródło: opracowanie własne.

Tabela 1. Oceny parametrów modeli: dla pełnych danych (PD) oraz danych z lukami (LD)

ZMIENNA	PD	LD	ZMIENNA	PD	LD
Wyraz wolny	227,67	219,06	P22	26,50	25,97
PONIEDZIAŁEK	18,17	19,57	P23	28,07	28,05
WTOREK	1,17	0,87	P24	28,37	28,60
ŚRODA	1,64	2,40	P25	27,98	27,50
CZWARTEK	0,53	0,77	P26	27,79	27,59
PIĄTEK	0,00	0,00	P27	27,60	27,32
SOBOTA	-24,83	-26,56	P28	26,82	27,07
NIEDZIELA	-42,59	-42,54	P29	23,73	24,29
STYCZEŃ	0,53	7,11	P30	22,42	22,52
LUTY	-6,96	-5,44	P31	19,40	19,63
MARZEC	-13,25	-6,93	P32	19,26	20,25
KWIECIEŃ	-39,63	-33,95	P33	20,46	19,92
MAJ	-66,23	-63,98	P34	22,48	22,57
CZERWIEC	-64,70	-59,88	P35	25,76	25,65
LIPIEC	-60,03	-61,03	P36	26,22	25,73
SIERPIEŃ	-62,77	-56,87	P37	24,13	24,24
WRZESIEŃ	-53,68	-51,05	P38	28,15	27,91
PAŹDZIERNIK	-40,58	-33,11	P39	29,29	28,00
LISTOPAD	-13,18	-15,43	P40	30,17	29,42
GRUDZIEŃ	0,00	0,00	P41	30,67	30,65
P1	-7,96	-7,90	P42	30,42	30,21
P2	-12,69	-12,95	P43	29,75	29,57
P3	-17,93	-17,67	P44	26,59	25,88
P4	-21,08	-19,41	P45	21,99	21,93
P5	-24,91	-24,38	P46	17,34	17,27
P6	-25,70	-25,18	P47	8,71	8,48
P7	-26,58	-25,89	P48	0,0	0,0
P8	-25,77	-24,49	t	$1,4833 \cdot 10^{-5}$	$1,1333 \cdot 10^{-4}$
P9	-27,38	-26,77	OP48	0,4681	0,4767
P10	-26,11	-25,83	N_ROK	-37,55	-70,43
P11	-23,49	-22,04	WIELKA SOBOTA	-49,31	-34,69
P12	-21,28	-21,37	WIELKANOC	-49,31	-74,43
P13	-16,30	-16,06	1-3 MAJ	-37,04	-30,04
P14	-5,43	-4,59	BOŻE CIAŁO	-29,47	4,22
P15	4,91	5,09	WNMP	-44,99	-43,91
P16	12,32	11,56	WSZYSTKICH ŚWIĘTYCH	-89,07	-87,53
P17	16,90	17,29	ŚWIĘTO NIEPODLEGŁOŚCI	-14,49	-6,39
P18	22,02	22,69	BOŻE NARODZENIE	-60,97	-57,53
P19	25,29	25,40	R^2	0,8565	0,8612
P20	26,11	26,04	SE	27,53	27,40
P21	24,68	25,01	Vs(%)	7,52	7,90

Źródło: opracowanie własne.

W tabeli 1 zestawione zostały oceny parametrów równań oszacowanych odpowiednio dla pełnych danych (PD) i danych z lukami (LD).

W celu bardziej przejrzystej prezentacji wyników modelowania, a następnie prognozowania, zmienne D_{jt} , M_{it} oraz S_{it} zastąpiono nazwami: dni tygodnia, miesiący i świąt.

Z informacji zawartych w tab. 1 wynika, że dodatnie oceny parametrów otrzymano dla dni tygodnia od poniedziałku do czwartku, przy czym ocenami maksymalnymi dla modelu bez luk oraz z lukami wynoszącymi odpowiednio: 19,57 i 18,17 MW charakteryzuje się poniedziałek. Natomiast dla soboty i niedzieli otrzymano oceny ujemne, świadczące o mniejszym zapotrzebowaniu na moc energetyczną.

Dla miesięcy od lutego do listopada otrzymano ujemne oceny parametrów. Oznaczają one, że zapotrzebowanie na energię w tych miesiącach jest niższe niż w grudniu. Oceny te od lutego do maja wykazują, co do wartości bezwzględnej, tendencję wzrostową. Natomiast w miesiącach od czerwca do sierpnia mamy do czynienia ze stabilizacją na poziomie ok. 60 MW. W kolejnych miesiącach oceny parametrów szybko maleją, co oznacza rosnące zapotrzebowanie na energię. Jedynym miesiącem o dodatniej ocenie wynoszącej 0,53 MW jest styczeń. O tę wielkość przeciętne zapotrzebowanie było wyższe niż w grudniu.

W przypadku okresów półgodzinnych parametry przyjmują wartości ujemne dla zmiennych od P1 do P14, tj. w godzinach 0^{30} - 7^{00} . Oceny minimalne w granicach -24 to -26 MW przypadają na godziny od 2^{30} do 5^{00} . Oceny maksymalne przypadają na pierwszy i ostatni okres półgodzinny. Widoczne jest „falowanie” tych ocen. Od godziny 7^{30} do 10^{00} wykazują tendencję wzrostową. Po godzinym spadku ma miejsce ich wzrost, a następnie stabilizacja na poziomie 27-28 MW w szczycie południowo-popołudniowym (od 11^{30} do 14^{00}). Następnie w godzinach 14^{30} - 16^{30} ma miejsce niewielki spadek zapotrzebowania, a następnie jego wzrost. Szczyt wieczorny przypada na godziny od 19^{00} do 21^{30} . Potem oceny parametrów wykazują szybką tendencję spadkową. Dla dni świątecznych oceny prawie wszystkich parametrów są ujemne. Charakteryzują się one dużym zróżnicowaniem. Dlatego omówimy je jedynie dla modelu z lukami. Oceny maksymalne, co do wartości bezwzględnej, przypadają na dzień Wszystkich Świętych oraz Święta Bożego Narodzenia. Natomiast minimalne przypadają w Święto Niepodległości oraz Wielką Sobotę.

Z porównania ocen parametrów obu równań dla tych samych zmiennych wynika, że różnice między nimi nie są zazwyczaj zbyt duże. Spostrzeżenie to odnosi się zwłaszcza do ocen parametrów dla okresów półgodzinnych. Maksymalna różnica dodatnia wynosi 1,94 MW i przypada o godz. 2^{00} (zmienna P4), a maksymalna różnica ujemna, co do wartości bezwzględnej, o godz. 19^{30} (zmienna P39). Większe różnice w ocenach parametrów mają miejsce w przypadku miesięcy. Wahają się one od -7,47 i -6,58 MW w październiku i styczniu do +0,99 MW w lipcu. Spośród dni świątecznych największe różnice dodatnie występują w dniu Nowego Roku (32,87 MW) oraz Świąt Wielkanocnych (25,13 MW), a największa różnica ujemna przypada w dniu Bożego Ciała (-33,70 MW). Stałe „przeniesienia”, tzn. oceny parametrów

Tabela 2. Oceny błędów prognoz inter- i ekstrapolacyjnych

ZMIENNA	Błędy prognoz (w %)			ZMIENNA	Błędy prognoz (w %)		
	Int_LD	Ex_PD	Ex_LD		Int_LD	Ex_PD	Ex_LD
OGÓLEM	5,70	5,26	5,66	P20	5,00	3,97	4,41
PONIEDZIA-	5,49	9,88	10,83	P21	4,84	3,97	4,52
WTOREK	6,32	4,99	5,36	P22	4,97	3,92	4,35
ŚRODA	5,59	4,76	5,48	P23	5,08	3,92	4,38
CZWARTEK	5,54	4,71	5,13	P24	5,27	3,80	4,35
PIĄTEK	4,46	4,63	4,85	P25	4,76	3,84	4,23
SOBOTA	5,54	3,52	3,41	P26	4,66	3,69	4,11
NIEDZIELA	5,38	5,16	5,25	P27	4,74	3,68	4,12
STYCZEŃ	4,16	4,24	4,61	P28	4,30	3,40	3,85
LUTY	5,09	5,04	5,08	P29	4,57	3,68	4,23
MARZEC	6,42	5,56	6,90	P30	4,58	3,40	3,76
KWIECIEŃ	5,13	5,36	6,18	P31	4,71	3,62	4,03
MAJ	5,79	6,05	6,48	P32	5,46	3,95	4,59
CZERWIEC	5,90	6,23	6,99	P33	5,85	4,44	4,85
LIPIEC	5,83	5,54	5,17	P34	5,69	4,97	5,51
SIERPIEŃ	5,34	4,75	5,64	P35	5,96	5,74	6,22
WRZESIEŃ	6,84	5,48	5,83	P36	5,88	5,54	5,94
PAŹDZIERNIK	7,07	4,16	4,90	P37	5,98	5,65	6,08
LISTOPAD	4,94	4,73	4,62	P38	5,62	5,91	6,18
GRUDZIEŃ	5,07	4,00	4,27	P39	5,15	5,84	5,95
P1	6,97	6,31	6,71	P40	5,32	5,68	5,85
P2	7,38	6,78	7,20	P41	4,97	5,12	5,40
P3	7,63	6,92	7,42	P42	4,04	4,88	5,11
P4	7,08	7,08	7,80	P43	3,94	4,50	4,69
P5	7,08	7,07	7,64	P44	4,23	4,25	4,47
P6	7,22	7,23	7,80	P45	4,48	4,29	4,69
P7	7,38	7,16	7,78	P46	5,27	4,37	4,78
P8	7,28	7,51	8,22	P47	5,60	4,94	5,33
P9	7,09	7,10	7,68	P48	5,57	4,54	4,63
P10	6,25	7,18	7,74	NOWY ROK	3,70	10,55	9,03
P11	5,53	6,94	7,68	WIELKA SOBOTA	3,65	13,03	25,02
P12	5,52	6,46	6,83	WIELKANOC	4,53	10,83	16,00
P13	5,54	6,07	6,50	1_3 MAJA	5,43	6,94	6,76
P14	5,58	5,51	6,04	BOŻE CIAŁO	5,19	2,57	11,92
P15	6,36	5,60	6,01	WNMP	4,30	5,76	6,13
P16	6,58	5,76	5,96	WSZYSTKICH ŚWIĘTYCH	3,13	9,77	10,30
P17	6,26	5,48	6,01	ŚWIĘTO NIEPODLEGŁOŚCI	5,15	11,22	13,13
P18	6,02	4,90	5,51	BOŻE NARODZENIE	3,96	9,38	8,99
P19	5,45	4,36	4,84				

Źródło: opracowanie własne.

występujące przy zmiennej OP48, będącej opóźnioną o 48 okresów półgodzinnych zmienną Y_t , są bardzo zbliżone i wynoszą odpowiednio: 0,4681 oraz 0,4767.

Także zbliżone są oceny parametrów struktury stochastycznej: R^2 , Se i V_s . Oznacza to, że wykorzystanie do szacowania parametrów jedynie połowy obserwacji nie wpłynęło na obniżenie własności predyktywnych równania z lukami w danych (LD).

W tabeli 2 zamieszczone zostały błędy względne prognoz inter- i ekstrapolacyjnych. Błędy prognoz interpolacyjnych (IN_LD) zostały obliczone jako względne różnice pomiędzy tymi prognozami a realizacjami w miesiącach, w których wystąpiły luki. W przypadku prognoz ekstrapolacyjnych dla modelu LD (EX_LD) w obliczeniach względnych błędów prognoz *ex post* wykorzystano realizację zmiennej prognozowanej w trzecim roku. Ponadto punktem odniesienia w ocenie ich dokładności będą błędy prognoz wyznaczonych na podstawie modelu bez luk (EX_PD).

Z informacji zawartych w drugim wierszu tab. 2 wynika, że błąd ogólny prognoz interpolacyjnych wynosi 5,70%. Spośród dni tygodnia oceny błędów przyjmują wartości od 4,46 dla piątku do 6,32% dla wtorku. Nieco wyższym zróżnicowaniem charakteryzują się błędy prognoz dla miesięcy. Oceny skrajne zawarte są w przedziale od 4,16% dla stycznia do 7,07% dla października. Oceny błędów prognoz interpolacyjnych dla okresów półgodzinnych zawarte są w przedziale od 3,94 o godzinie 21³⁰ do 7,63 o godzinie 1³⁰. W kształtowaniu się ocen błędów prognoz widoczna jest prawidłowość polegająca na tym, że oceny błędów powyżej 7% przypadają na godziny od 1⁰⁰ do 4³⁰, a więc w okresie o najniższym zapotrzebowaniu na energię. Błędy powyżej 6% otrzymano dla godzin 7³⁰-9⁰⁰. Można wyróżnić także dwa okresy z ocenami błędów 4-5%, tj. godz. 10³⁰-15³⁰ oraz od 20³⁰ do 24⁰⁰.

Dla dni świątecznych oceny tych błędów są dość wyrównane i mieszczą się w przedziale od 3,13 dla Wszystkich Świętych do 5,43% dla dni majowych (1-3 maja). W kolejnych kolumnach podane są błędy *ex post* prognoz ekstrapolacyjnych. Błędy te będą analizowane łącznie, przy zwróceniu szczególnej uwagi na prognozy otrzymane na podstawie modelu z lukami (LD). Niższą oceną ogólnego błędu prognoz wynoszącą 5,28% charakteryzują się prognozy otrzymane na podstawie modelu bez luk. Dla modelu z lukami jest ona nieznacznie wyższa (o 0,44 punktu procentowego). Oceny błędów dla dni tygodnia są bardzo zróżnicowane. Zdecydowanie *in minus* odbiega dokładność prognoz dla poniedziałku – ocena tego błędu wyniosła 10,83%. Dla modelu pełnego błąd ten przyjął wartość 9,88%. Dla pozostałych dni tygodnia błędy prognoz dla modelu LD kształtowały się w przedziale od 3,41% dla sobót do 5,48% dla śród. Dla prawie wszystkich dni tygodnia (poza sobotą) oceny błędów prognoz otrzymanych na podstawie modelu LD są wyższe od błędów prognoz dla pełnych danych. Różnice te są niewielkie (od 0,09 punktu procentowego dla niedziel do 0,95 dla poniedziałków). Jedynie dla sobót ocena ta okazała się o 0,1 punktu procentowego niższa.

Oceny mierników dokładności prognoz dla miesięcy mieszczą się w przedziale 4-7%. Oceny minimalne otrzymano dla grudnia (4,20 i 4,27%), a maksymalne dla

czerwca (6,23 i 6,99%). Nieznacznie wyższą dokładność prognoz na podstawie modelu z lukami otrzymano jedynie dla lipca i listopada.

Dla dziesięciu spośród dwunastu miesięcy bardziej efektywne okazały się prognozy otrzymane na podstawie modelu dla pełnych danych. Różnice ocen prognoz otrzymanych na podstawie modeli dla pełnych i niepełnych danych wahały się od 0,04 punktu procentowego dla lutego do 1,34 punktu procentowego dla marca. Różnice dla pozostałych miesięcy nie przekraczają 0,9 punktu procentowego. Natomiast dla lipca i listopada otrzymano niższe błędy prognoz na podstawie modelu LD wynoszące odpowiednio 0,38 i 0,1 punktu procentowego.

W kształtowaniu się ocen błędów prognoz dla okresów półgodzinnych widoczna jest prawidłowość polegająca na tym, że najwyższe oceny błędów (w granicach 5-8%) otrzymano dla godzin 0³⁰-9⁰⁰. W godzinach od 9³⁰ do 16³⁰-17⁰⁰ przyjmują one wartości z przedziału od 3 do 5%. W godzinach 17³⁰ do 21⁰⁰ kształtują się one na poziomie 5,12-6,22%. Od 21³⁰ ich oceny są ok. 1-2 punkty procentowe niższe. Dla wszystkich okresów półgodzinnych prognozy otrzymane na podstawie modelu szacowanego dla pełnych danych charakteryzują się ocenami niższymi od ocen uzyskanych dla modelu z lukami w danych (LD) od 0,11 punktu dla zmiennej P39 (godz. 19³⁰) do 0,71, 0,72 i 0,74 punktu w godzinach 4⁰⁰, 2⁰⁰ oraz 5³⁰.

Dla wszystkich okresów półgodzinnych nieznacznie wyższą efektywnością charakteryzowały się prognozy otrzymane na podstawie modelu dla pełnych danych (PD). Różnice te zawarte były w przedziale od 0,09 punktu procentowego dla godziny 24⁰⁰ (zmienna P48) do 0,74 dla godziny 5³⁰ (zmienna P11). Oceny błędów prognoz otrzymanych na podstawie modelu LD zawarte są w przedziale od 3,76% dla godziny 15⁰⁰ do 8,22% dla godziny 4⁰⁰.

W przypadku świąt widoczne jest bardzo duże zróżnicowanie błędów prognoz. Są one o kilka, a niekiedy nawet o kilkanaście punktów procentowych wyższe od błędów prognoz interpolacyjnych. Błędy prognoz EX_LD niższe od 10% otrzymano dla: Nowego Roku, 1-3 maja oraz Bożego Narodzenia. Z porównania ich z ocenami błędów prognoz EX_PD wynika, że większość różnic ocen przyjmuje wartości dodatnie, tzn. prognozy otrzymane na podstawie modelu dla pełnych danych charakteryzują się niższymi błędami. Najwyższą różnicę dodatnią ocen błędów wynoszącą blisko 12 punktów procentowych otrzymano dla Wielkiej Soboty, a następnie dla Bożego Ciała (9,35 punktu procentowego). Ponad 5-procentową różnicę otrzymano dla Wielkanocy. Różnice błędów prognoz dla pozostałych świąt nie przekraczają 2 punktów procentowych. Otrzymanie tych różnic nie jest w pełni niespodziewane w sytuacji, gdy parametry modelu LD szacowane były na podstawie danych z jednego roku. Zaskakujące jest natomiast to, że dla trzech świąt (Nowego Roku, 1-3 maja oraz Bożego Narodzenia) dla modelu z lukami otrzymano niższe błędy (z maksymalną różnicą 1,52 punktu procentowego dla Nowego Roku).

W tabeli 3 przedstawiono, w postaci tablicy rozdzielczej, rozkłady błędów prognoz interpolacyjnych (IN_LD) oraz prognoz ekstrapolacyjnych otrzymanych na

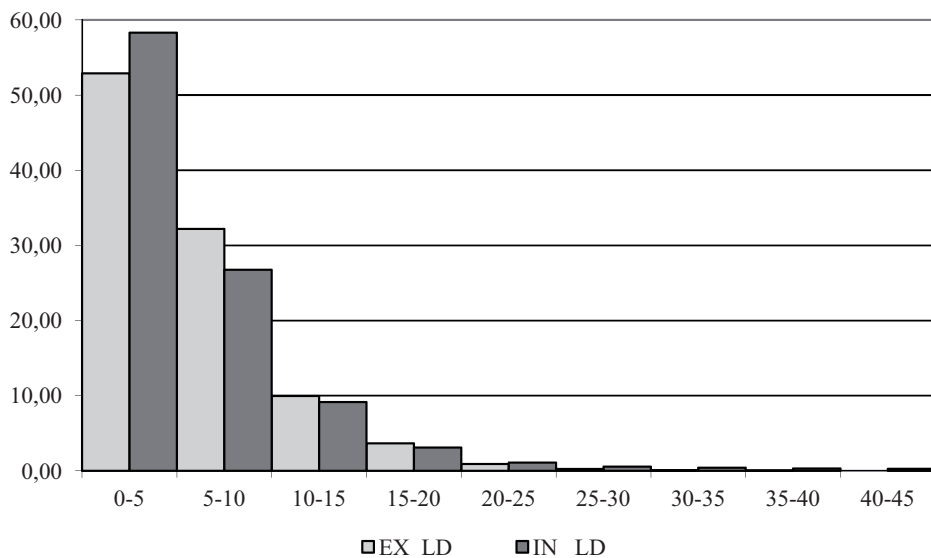
podstawie modelu dla pełnych danych (EX_PD) oraz modelu dla danych z lukami (EX_LD).

Tabela 3. Rozkłady empiryczne błędów prognoz interpolacyjnych i ekstrapolacyjnych (w odsetkach)

Granice przedziałów błędów prognoz (w %)	IN_LD	EX_PD	EX_LD
0-5	58,30	57,15	52,88
5-10	26,75	29,78	32,19
10-15	9,16	9,25	9,96
15-20	3,11	2,91	3,67
20-25	1,10	0,69	0,91
25-30	0,56	0,19	0,25
30-35	0,42	0,03	0,07
35-40	0,31	0,00	0,05
40-45	0,27	0,00	0,00

Źródło: opracowanie własne.

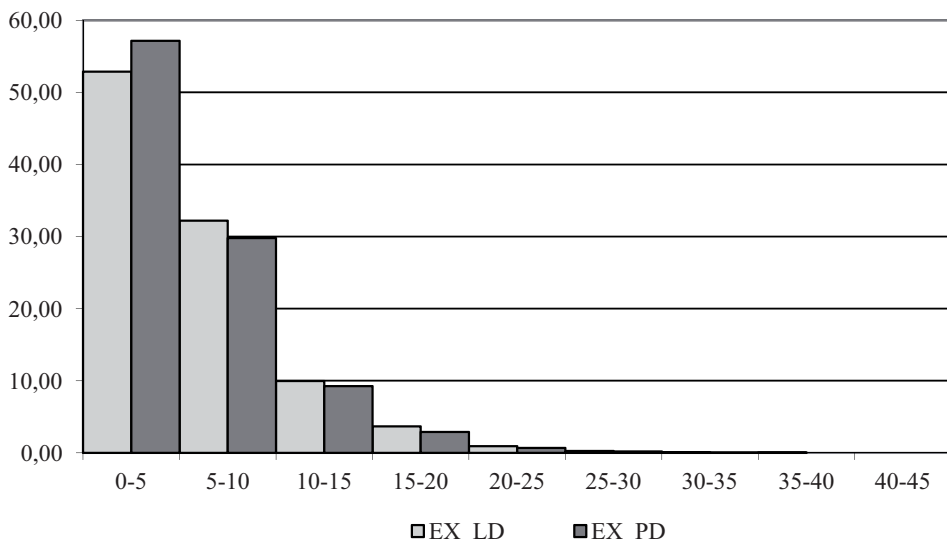
Z tabeli 3 wynika, że ponad 85% błędów prognoz zawartych jest w przedziale 0-10%. Największym odsetkiem wynoszącym 86,93% charakteryzują się prognozy ekstrapolacyjne otrzymane na podstawie modelu bez luk (EX_PD). Zwraca uwagę



Rys. 2. Rozkłady błędów prognoz ekstrapolacyjnych dla modelu z lukami oraz błędów prognoz interpolacyjnych (w odsetkach)

Źródło: opracowanie własne.

ok. 1-procentowy udział błędów prognoz interpolacyjnych (IN_LD) wynoszących ponad 30%. W przypadku prognoz ekstrapolacyjnych odsetki te są znacznie niższe i wynoszą 0,03% dla modelu z pełnymi danymi i 0,12 dla modelu z lukami danych. Porównanie graficzne rozkładów błędów prognoz dla prognoz inter- i ekstrapolacyjnych otrzymanych na podstawie modelu z lukami przedstawiono na rys. 2. Rysunek 3 zaś zawiera porównanie rozkładów błędów prognoz ekstrapolacyjnych.



Rys. 3. Rozkłady błędów prognoz ekstrapolacyjnych dla modelu z lukami oraz dla modelu bez luk (w odsetkach)

Źródło: opracowanie własne.

Z rysunku 2 wynika, że dla modelu z lukami w danych w przedziale 0-5% wyższy jest odsetek błędów prognoz ekstrapolacyjnych, a w przedziale 5-10% – prognoz interpolacyjnych. Z analizy rysunku 3 wynika, że podobna relacja zachodzi między odsetkami błędów prognoz ekstrapolacyjnych otrzymanych na podstawie modeli dla pełnych i niepełnych danych.

4. Wnioski

1. Z oszacowań parametrów struktury stochastycznej modelu dla pełnych danych i danych z lukami wynika, że własności predyktywne równań są zbliżone.
2. Otrzymano także zbliżone oceny parametrów dla dni tygodnia, miesięcy oraz okresów półgodzinnych.
3. Odnotowano natomiast znaczne różnice w ich ocenach dla dni świątecznych. Wynikały one z różnicy liczby obserwacji, na podstawie których były szacowane –

dla pełnych danych oceny parametrów są średnimi z dwóch lat, a w modelu z lukami – tylko z jednego roku.

4. Z przedstawionych w pracy obliczeń wynika, że ogólne oceny błędów prognoz interpolacyjnych i ekstrapolacyjnych otrzymanych na podstawie modelu są bardzo zbliżone. Kształtują się one w granicach 5,70%. Ocena ta dla prognoz ekstrapolacyjnych wyznaczonych na podstawie modelu bez luk jest jedynie o ok. 0,4 punktu procentowego niższa.

5. Błędy prognoz we wszystkich przekrojach, zarówno dla dni tygodnia, miesięcy, jak i świąt, były dość stabilne, nie przekraczając na ogół 7%. Dla okresów półgodzinnych zawarte są one w przedziale od 3,94 do 7,63%. Oceny błędów przekraczające 7% przypadały w godzinach 0^{30} - 4^{30} .

6. Prognozy ekstrapolacyjne otrzymane na podstawie modelu z lukami wykazywały tę samą prawidłowość dla dni tygodnia (poza poniedziałkiem), miesięcy i niektórych świąt. Dla 38 z 48 okresów półgodzinnych ich oceny były niższe od 6,22%. Przekraczały one 7% w godzinach o najmniejszym zapotrzebowaniu na energię.

Reasumując, można stwierdzić, że metody ekonometryczne mogą być przydatne w prognozowaniu brakujących informacji w szeregach czasowych dla danych o bardzo wysokiej częstotliwości obserwowania.

Literatura

- Kufel T., *Ekonometryczna analiza cykliczności procesów gospodarczych o wysokiej częstotliwości obserwacji*, Wydawnictwo Naukowe UMK w Toruniu, Toruń 2010.
- Szmuksta-Zawadzka M., Zawadzki J., *Modelowanie i prognozowanie zapotrzebowania na energię elektryczną w mikroskali*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu nr 165, UE, Wrocław 2011.

APPLICATION OF ECONOMETRICAL MODELING FOR PREDICTING MISSING DATA IN HIGH-FREQUENCY TIME SERIES

Summary: The paper presents the results of modeling and forecasting the lack of full information, with the example of electric energy demand in a certain area for half-hourly data. The forecasts were built based on models depicting three types of periodic fluctuations with the annual, weekly and 24-hourly cycles, by the means of dummy variables. Moreover, they included such variables for determining holidays and a delayed predicted variable. In the course of empiric verification of the predictions, the analysis of precision of interpolation and extrapolation forecasts was performed.

Keywords: high-frequency data, forecasting, lack of data.