

Authors' contribution/
Wkład autorów:
A. Study design/
Zaplanowanie badań
B. Data collection/
Zebranie danych
C. Statistical analysis/
Analiza statystyczna
D. Data interpretation/
Interpretacja danych/
E. Manuscript preparation/
Przygotowanie tekstu
F. Literature search/
Opracowanie
piśmiennictwa
G. Funds collection/
Pozyskanie funduszy

REVIEW ARTICLE

JEL code: C22

Submitted:
November 2022

Accepted:
January 2023

Tables: 5
Figures: 1
References: 19

ARTYKUŁ PRZEGLĄDOWY

Klasyfikacja JEL: C22

Zgłoszony:
listopad 2022

Zaakceptowany:
styczeń 2023

Tabele: 5
Rysunki: 1
Literatura: 19

COMPARATIVE ANALYSIS OF METHODS FOR HOURLY
ELECTRICITY DEMAND FORECASTING IN THE ABSENCE OF
DATA – A CASE STUDY

ANALIZA PORÓWNAWCZA METOD PROGNOZOWANIA GODZINNEGO
ZAPOTRZEBOWANIA NA ENERGIĘ ELEKTRYCZNĄ PRZY BRAKACH
W DANYCH – STUDIUM PRZYPADKU

Jan Zawadzki^(A,B,C,D,E,F,G)

Faculty of Economics, Department of Applications of Mathematics in Economics
Professor emeritus, West Pomeranian University of Technology in Szczecin,
Poland

Wydział Ekonomiczny, Katedra Zastosowań Matematyki w Ekonomii
Profesor emerytowany, Zachodniopomorskiego Uniwersytetu Technologicznego
w Szczecinie, Polska

Citation:

Zawadzki, J. (2023). Comparative analysis of methods for hourly electricity demand forecasting in the absence of data – a case study/ Analiza porównawcza metod prognozowania godzinnego zapotrzebowania na energię elektryczną przy brakach w danych – studium przypadku. *Economic and Regional Studies*, 16(1), strony 34-50. <https://doi.org/10.2478/ers-2023-0003>

Guest Editor Prof. Karol Kukuła, Ph.D., University of Applied Sciences in Tarnow

Abstract

Scope and purpose of work: This paper examines the impact of the number of gaps in data, the analytical form, and the model type selection criterion on the accuracy of interpolation and extrapolation forecasts for hourly data.

Materials and methods: Forecasts were developed on the basis of predictors that are based on: classical time series forecasting models and regression time series forecasting models, hybrid time series forecasting models and hybrid regression forecasting models for uncleared series, and exponential smoothing models for cleared series of two or three types of seasonal fluctuations, with minimum estimates of errors in interpolation or extrapolation forecasts.

Results: Adaptive and hybrid regression models have proved to have the most favorable predictive properties. Most hybrid time series models for systematic and non-systematic gaps and for both analytical forms are single models that generally describe fluctuations within a 24-hour cycle.

Conclusions: The lowest estimators of prediction errors involving interpolation were obtained for exponential smoothing models, followed by hybrid regression models. A reverse sequence was obtained for extrapolative forecasting.

Keywords: forecasting, missing data, time series, high frequency

Streszczenie

Przedmiot i cel pracy: Celem artykułu było zbadanie wpływu: liczby luk w danych, postaci analitycznej, kryterium wyboru rodzaju modelu na dokładność prognoz interpolacyjnych i ekstrapolacyjnych dla danych godzinnych.

Materiały i metody: Do budowy prognoz wykorzystano predyktory oparte na: klasycznych i regresyjnych modelach szeregu czasowego, modelach hybrydowych szeregu czasowego i hybrydowych modelach regresyjnych dla szeregów nieoczyszczonych oraz modelach wyrównywania

Address for correspondence / Adres korespondencyjny: prof. Jan Zawadzki (ORCID: 0000-0002-2996-3180 (jzawadzki@zut.edu.pl), Faculty of Economics, Department of Applications of Mathematics in Economics, West Pomeranian University of Technology in Szczecin; address: 71-270 Szczecin, ul. Janickiego 31; phone: 91 449 68 68.

Journal included in: ERIH PLUS; AgEcon Search; AGRO; Arianta; Baidu Scholar; BazEkon; Cabell's Whitelist; CNKI Scholar; CNPIEC – cnpLINKer; EBSCO Discovery Service; EBSCO-CEEAS; EuroPub; Google Scholar; Index Copernicus ICV 2017-2020: 100,00; J-Gate; KESLI-NDSL; MyScienceWork; Naver Academic; Naviga (Softweco); Polish Ministry of Science and Higher Education 2021: 20 points; Primo Central; QOAM; ReadCube; Semantic Scholar; Summon (ProQuest); TDNet; WanFang Data; WorldCat.
Copyright: © The Authors, 2022. **Publisher:** John Paul II University of Applied Sciences in Biala Podlaska, Poland.

wykładniczego dla szeregów oczyszczonych z dwóch albo trzech rodzajów wahań sezonowych, charakteryzujących się minimalnymi ocenami błędów prognoz interpolacyjnych lub ekstrapolacyjnych.

Wyniki: Najkorzystniejsze własności predyktywne posiadają modele adaptacyjne i hybrydowe modele regresyjne. Większość modeli hybrydowych szeregu czasowego dla luk systematycznych i niesystematycznych oraz obu postaci analitycznych, są to modele pojedyncze na ogół opisujące wahania o cyklu 24-godzinnym.

Wnioski: Najniższe oceny błędów prognoz interpolacyjnych otrzymano dla modeli wyrównywania wykładniczego a następnie hybrydowych modeli regresyjnych. Natomiast dla prognoz ekstrapolacyjnych kolejność była odwrotna.

Słowa kluczowe: prognozowanie, brakujące dane, szeregi czasowe, wysoka częstotliwość

Introduction

Companies that deal with the production or distribution of electricity compile three types of electricity forecasts: long-term, medium-term, and short-term. Short-term forecasts covering the time period of no more than an hour play the key role in this respect. Electricity production and consumption is a product of three types of seasonality: a daily pattern, a weekly pattern, and an annual pattern, and also involves other weather-related factors (temperature, precipitation, wind power), holidays etc. Short-term forecasts covering the time period of no more than an hour play the key role in this respect. There is an extensive body of research on short-term modeling and forecasting of electricity demand based on complete time series using various statistical methods and artificial neural networks. The following authors have researched this field: Dordonnat et al., 2008; Engle et al., 1986; Harvey, Koopman, 1993; Lichota, 2006; Misiorek and Weron, 2004; Nowicka-Zagrajek, Weron, 2002; Rammanth et al., 1997; Szmuksta-Zawadzka, Zawadzki, 2011, Taylor et al., 2006; Taylor 2010; Tomaszewski, 2005; Weron, 2006; Witkowska et al., 2000.

In contrast, there is scarcity of literature that explores energy demand for hourly series with data gaps.

The work of Weron (2006) is a notable exception. The author focuses on forecasting using time series models with gaps in data where the energy demand is measured at 5-minute intervals. This method involves making up for the missing data using average values from long-time observations, adjacent observations or, for example, from the previous week, followed by the elimination of outliers.

Szmuksta-Zawadzka and Zawadzki (2011, 2014, 2015, 2016), and then Zawadzki (2018, 2020) also conducted long-standing research on the methods used for forecasting missing data in hourly time series with triple-complex seasonal fluctuations, for both systematic and non-systematic gaps. This research also explores the missing variants, and summarizes the obtained findings.

Wstęp

Firmy zajmujące się produkcją lub dystrybucją wyznaczają trzy rodzaje prognoz energii elektrycznej: długo-okresowe, średniookresowe oraz krótkookresowe. Podstawowe znaczenie w odgrywają w tym względzie prognozy krótkookresowe o okresie jednostkowym nie dłuższym niż godzina. Produkcja i zużycie energii elektrycznej jest wypadkową trzech rodzajów sezonowości o cyklach: dobowych, tygodniowym i rocznym oraz innych czynników związanych m.in. z pogodą (temperatura, opady, siła wiatru), występowanie świąt itp. Podstawowe znaczenie odgrywają w tym względzie prognozy krótkookresowe o okresie jednostkowym nie dłuższym niż godzina. Literatura poświęcona krótkookresowemu modelowaniu i prognozowaniu zapotrzebowania na energię elektryczną na podstawie kompletnych szeregów czasowych z zastosowaniem różnych metod statystycznych i sztucznych sieci neuronowych jest dość obszerna. Przykładowo można tu wymienić prace następujących autorów: Dordonnat i inni, 2008; Engle i inni, 1986; Harvey, Koopman, 1993; Lichota, 2006; Misiorek i Weron, 2004; Nowicka-Zagrajek, Weron, 2002; Rammanth i inni, 1997; Szmuksta-Zawadzka, Zawadzki, 2011, Taylor i inni, 2006; Taylor 2010; Tomaszewski, 2005; Weron, 2006; Witkowska i inni, 2000.

Natomiast zupełnie do rzadkości należą prace dotyczące zapotrzebowania na energię dla szeregów godzinnych z lukami w danych.

Na uwagę w tym względzie zasługuje książka Weron (2006). Autor przywołanej pracy rozpatruje prognozowanie w szeregach z lukami, w których zapotrzebowanie na energię jest mierzone w odstępach 5-minutowych. Metoda polega na zastępowaniu brakujących danych: średnimi z obserwacji z dłuższych odcinków czasu, z sąsiadujących obserwacji lub np. z poprzedniego tygodnia, po wcześniejszym wyeliminowaniu obserwacji odstających.

Wieloletnie badania nad zastosowaniami metod prognozowania brakujących danych w godzinnych szeregach czasowych z potrójnie złożonymi wahaniami sezonowymi, dla luk systematycznych i niesystematycznych, były prowadzone do roku 2016 przez Szmuksta-Zawadzka i Zawadzkiego (2011, 2014, 2015, 2016), a następnie przez Zawadzkiego (2018, 2020). Niniejsza praca uzupełniona o brakujące warianty, stanowi podsumowanie otrzymanych wyników.

Theoretical introduction

Time series models with triple complex seasonal fluctuations can be used for modeling and forecasting variables for hourly data: classical models, hybrid models, and exponential smoothing models, as well as classical and hybrid regression models. This paper attempts to reflect on time series models since the classical and hybrid regression models are extended time series models with endogenous variable delayed by 24 hours and dummy variables that account for holidays and one day before holidays (Easter Saturday).

These fluctuations can overlap the trend either additively or multiplicatively. In classical models, each type of fluctuations is described using separate subsets of dummy variables. An example of an analytical notation of the time series model in the additive form with a linear trend is as follows (Kufel, 2010; Szmuksta-Zawadzka, Zawadzki, 2011):

$$Y_t = \kappa_1 t + \kappa_0 + \sum_{i=1}^{12} b_{0i} M_{it} + \sum_{j=1}^7 c_{0j} D_{jt} + \sum_{k=1}^{24} d_{0k} G_{kt} + U_t \quad (1)$$

subject to the following conditions:

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

where:

M_{it} – months,

D_{jt} – days of the week,

G_{kt} – hours in a daily cycle.

The exponential model differs in that there is a natural logarithm instead of the variable .

Hybrid models consisting of classical and hierarchical parts are also used for modeling and forecasting triple complex seasonal fluctuations. The classical part refers to fluctuations with an odd cycle length (here: a weekly cycle), and is described using dummy variables as in model 1.

Regular hierarchical models can be used to describe fluctuations for hourly data in even cycle lengths (here: annual and daily cycles). Models in which one type of fluctuations is described using a hierarchical model will be referred to as single hybrid models. If both types of fluctuations are described using a hierarchical model, these models will be described as dual hybrid models. As defined by Szmuksta-Zawadzka and Zawadzki (2002), a regular hierarchical model is a model with divisors p_i of a periodic (seasonal) fluctuation cycle m , which is an even number, that meets two conditions at the same time:

$$2 \leq p_i \leq \frac{m}{2} \text{ and } \prod_i p_i = m \quad (2)$$

Wprowadzenie teoretyczne

W modelowaniu i prognozowaniu zmiennych dla danych godzinnych mogą być wykorzystywane modele szeregu czasowego z potrójnie złożonymi wahaniami sezonowymi: klasyczne, modele hybrydowe i modele wyrównywania wykładniczego oraz klasyczne i hybrydowe modele regresyjne. Ograniczymy się do syntetycznego omówienia modeli szeregu czasowego, ponieważ klasyczne i hybrydowe modele regresyjne są rozszerzonymi modelami szeregu czasowego o: opóźnioną o 24 godziny zmienną endogeniczną oraz zmienne zerojedynkowe opisujące występowanie świąt i jednego dnia przedświątecznego (Wielkiej Soboty)

Wahania te mogą nakładać się na trend w sposób addytywny lub multiplikatywny. W modelach klasycznych każdy rodzaj wahań opisywany jest za pomocą osobnych podzbiorów zmiennych zerojedynkowych. Przykładowy zapis analityczny modelu szeregu czasowego w postaci addytywnej z trendem liniowym jest następujący (Kufel, 2010; Szmuksta-Zawadzka, Zawadzki, 2011):

$$Y_t = \kappa_1 t + \kappa_0 + \sum_{i=1}^{12} b_{0i} M_{it} + \sum_{j=1}^7 c_{0j} D_{jt} + \sum_{k=1}^{24} d_{0k} G_{kt} + U_t \quad (1)$$

przy warunkach:

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

gdzie:

M_{it} – miesiące,

D_{jt} – dni tygodnia,

G_{kt} – godziny w cyklu dobowym.

Model wykładniczy różni się tym, że zamiast zmiennej występuje jej logarytm naturalny.

W modelowaniu i prognozowaniu dla potrójnie złożonych wahań sezonowych wykorzystywane są także modele hybrydowe, składające się z dwóch części: klasycznej i hierarchicznej. Część klasyczna odnosi się do wahań o nieparzystej długości cyklu, w naszym przypadku tygodniowym, i jest opisywana za pomocą zmiennych zerojedynkowych jak w modelu 1.

Do opisu wahań o parzystej długości cykli dla danych godzinnych (w naszym przypadku o cyklu: rocznym i dobowym) mogą być wykorzystywane regularne modele hierarchiczne. Modele, w których jeden rodzaj wahań opisywany jest za pomocą modelu hierarchicznego nazywać będziemy pojedynczymi modelami hybrydowymi. Jeżeli za ich pomocą opisywane będą obydwa rodzaje wahań to będą to podwójne modele hybrydowe. W myśl definicji zamieszczonej w pracy (Szmuksta-Zawadzka i Zawadzki, 2002) regularnym modelem hierarchicznym jest model o podzielnikach p_i cyklu wahań okresowych

Hierarchical models for fluctuations within a 12-month and 24-hour cycle will be referred to respectively as: HM and HG. The numbers noted after these symbols shall mean successive divisors of the length of the fluctuation cycle.

The number of regular hierarchical models for a given fluctuation cycle m is equal to the number of permutations and permutations with repetitions of divisors p_i .

Szmuksta-Zawadzka, Zawadzki (2014) list hierarchical models for individual cycle lengths and provide their analytical notations.

There are 7 hierarchical models for monthly data within an annual cycle ($m=12$), including 4 two-stage models (HM26, HM34, HM43, HM62), and 3 three-stage models (HM223, HM232, HM322).

There are 19 models for data within a daily cycle ($m=24$), including 6 two-stage models (HG2_12, HG38, HG46, HG83, HG12_2), 9 three-stage models (HG226, HG234, HG243, HG324, HG342, HG423, HG432, HG226, HG262), and 4 four-stage models (HG2223, HG2232, HG2322, HG3222).

The number of estimated dual hybrid models is equal to the product of the numbers of single models. Hence, a total of hybrid models are estimated for a single analytical form. One classical model (with three sets of dummy variables) as per equation (1) should be added as the benchmark for hybrid models.

An analytical notation of a four-stage linear hierarchical model with a linear trend is as follows (Szmuksta-Zawadzka and Zawadzki, 2014):

(3)

$$Y_t = \kappa_1 t + \kappa_0 + \sum_{s=1}^{p_1} b_{0s} Q_{st} + \sum_{r=1}^{p_2} b_{0sr} Q_{srt} + \sum_{l=1}^{p_3} b_{0srlt} Q_{srlt} + \sum_{k=1}^{p_4} b_{0srklt} Q_{srklt} + U_{srklt}, \quad t = 1, 2, \dots, n$$

subject to the following conditions:

$$\sum_{s=1}^{p_1} b_{0s} = \sum_{r=1}^{p_2} b_{0sr} = \sum_{l=1}^{p_3} b_{0srlt} = \sum_{k=1}^{p_4} b_{0srklt} = 0$$

The two-stage model consists of the first two components and the three-stage model will feature the first three components. The form of variables Q depends on the number of hierarchy levels, the number and sequence of divisors in the model, and the fluctuation cycle length. The dummy variables Q in the formula provided above for a 24-hour cycle refer to 19 two-, three- and four-stage models. The number of variables at each level of the hierarchy is 1 lower than the value of the divisors p_1 , p_2 , p_3 or p_4 as the parameters must add up to 0. For example, there are 11 and 1 variables, respectively, in the

(sezonowych) m , będącego liczbą parzystą, spełniającej jednocześnie dwa warunki:

$$2 \leq p_i \leq \frac{m}{2} \text{ and } \prod_i p_i = m \quad (2)$$

Modele hierarchiczne dla wahań o cyklu 12 miesięcznym i 24 godzinnym oznaczane będą odpowiednio jako: HM i HG. Liczby występujące po tych symbolach oznaczać będą kolejne podzielniki długości cyklu wahań.

Liczba regularnych modeli hierarchicznych dla danego cyklu wahań m jest równa liczbie permutacji oraz permutacji z powtórzeniami podzielników p_i .

W pracy (Szmuksta-Zawadzka, Zawadzki, 2014) wymienione zostały modele hierarchiczne dla poszczególnych długości cykli i podane zostały ich zapisy analityczne.

Dla danych miesięcznych o cyklu rocznym ($m=12$) liczba modeli hierarchicznych wynosi 7, w tym:

- 4 dwustopniowe (HM26, HM34, HM43, HM62) oraz
- 3 trójstopniowe (HM223, HM232, HM322).

Dla danych o cyklu dobowym ($m=24$) ich liczba wynosi 19, w tym:

- 6 dwustopniowych (HG2_12, HG38, HG46, HG83, HG12_2),
- 9 trzystopniowych (HG226, HG234, HG243, HG324, HG342, HG423, HG432, HG226, HG262),
- 4 czterostopniowe (HG2223, HG2232, HG2322, HG3222).

Natomiast liczba szacowanych podwójnych modeli hybrydowych równa jest iloczynowi liczb modeli pojedynczych i wynosi . Zatem dla jednej postaci analitycznej szacowanych jest łącznie modeli hybrydowych. Do tego należy doliczyć jeden model klasyczny (z trzema zespołami zmiennych zerojedynkowych) dany równaniem (1). Będzie on punktem odniesienia dla modeli hybrydowych.

Poniżej podano zapis analityczny czterostopniowego liniowego modelu hierarchicznego z trendem liniowym (Szmuksta-Zawadzka i Zawadzki, 2014):

(3)

$$Y_t = \kappa_1 t + \kappa_0 + \sum_{s=1}^{p_1} b_{0s} Q_{st} + \sum_{r=1}^{p_2} b_{0sr} Q_{srt} + \sum_{l=1}^{p_3} b_{0srlt} Q_{srlt} + \sum_{k=1}^{p_4} b_{0srklt} Q_{srklt} + U_{srklt}, \quad t = 1, 2, \dots, n$$

przy warunkach:

$$\sum_{s=1}^{p_1} b_{0s} = \sum_{r=1}^{p_2} b_{0sr} = \sum_{l=1}^{p_3} b_{0srlt} = \sum_{k=1}^{p_4} b_{0srklt} = 0$$

HG12_2 two-stage model. The dummy variables Q_{st} ($s = 1, 2, \dots, 11$) take values equal to 1 for each two-hour period: $Q_{1t} = 1$ for hour 1 and 2, and 0 for the remaining hours of the daily cycle, Q_{2t} for 3 and 4, ..., Q_{11t} for 21 and 22. There is only one variable Q_{srt} ($r = 1, 2$) since $p_2 - 1 = 1$. The variable Q_{s1t} takes the value of 1 for the first (odd) hour of each two-hour period. In seven two- and three-stage hierarchical models describing fluctuations within an annual cycle, the variables Q take different values. For example, the divisors in the HM43 model are $p_1 = 4$ and $p_2 = 3$. The number of Q variables is lower by 1. The dummy variables $Q_{st(M)}$ ($s = 1, 2, 3$) take the value of 1 in the s -th quarter and 0 in the remaining quarters, and the variables $Q_{srt(M)}$ ($r = 1, 2$) take the value of 1 in the r -th months of each quarter. All of the discussed models have complex seasonality: the classical time series models, regression models, and hierarchical (hybrid) models are single-equation models.

As for exponential smoothing models, data cleared of two (Y_t^{**}) or three types of seasonality (Y_t^{***}) are the basis for modelling and then for creating the output forecasts. The forecasts will be built on predictors based on three-parameter Holt-Winters models and single-parameter Brown's models or two-equation Holt's models, respectively.

The analytical notations of these models and the resulting predictors were described by Szmuksta-Zawadzka, Zawadzki (2015). The smoothing constants take values on a scale of 0 to 1. Final forecasts are obtained in additive models by adding components, and in multiplicative models – by multiplying by seasonality indices.

In the first case, the forecasts are built on predictors based on Holt-Winters models with three equations and three smoothing constants (α, β, δ). In this model, the seasonal effects within a 24-hour cycle are described by the third equation. Equations and predictors of the additive Holt-Winters model:

$$m_t^{**} = \alpha(Y_t^{**} - c_{t-L}) + (1 - \kappa)m_{t-1}^{**} \quad (4)$$

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

$$C_t^{**} = \delta(Y_t^{**} - m_t^{**}) + (1 - \delta)C_{t-m}^{**} \quad (6)$$

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

$$\Pi_{HW} = \Pi^{**} + M^{(a)}(t) \quad (8)$$

If the three-parameter Holt-Winters models are used, there is a numerical problem, specifically if the parameters α, β, δ change from 0.01 to 0.99, every 0.01, then $99^3 = 970299$ versions should be estimated.

Choosing the "optimal" values of the smoothing constants without having to estimate close to a million versions is one of the biggest challenges. Earlier research revealed that, if the variable is highly inert,

Model dwustopniowy zawierać będzie dwie pierwsze składowe a model trójstopniowy trzy. Postać zmiennych Q zależy od: liczby stopni hierarchii, liczby i kolejności podzielników modelu oraz długości cyklu wahań. Występujące w powyższym wzorze zmienne zero-jedynkowe Q dla cyklu 24-godzinnego odnoszą się do 19 modeli dwu-, trzy- i czterostopniowych. Liczba zmiennych na każdym stopniu hierarchii, z uwagi na warunek sumowalności parametrów do 0, jest o 1 mniejsza od wartości podzielników p_1, p_2, p_3 lub p_4 . Przykładowo w modelu dwustopniowym HG12_2 ich liczba wynosi odpowiednio 11 i 1. Zmienne zero-jedynkowe Q_{st} ($s = 1, 2, \dots, 11$) przyjmują wartości równe 1, dla każdego okresu dwugodzinnego: $Q_{1t} = 1$ dla godziny 1 i 2 oraz 0 dla pozostałych godzin cyklu dobowego, Q_{2t} dla 3 i 4, ..., Q_{11t} dla 21 i 22. Natomiast zmienna Q_{srt} ($r = 1, 2$) jest tylko jedna, ponieważ $p_2 - 1 = 1$. Zmienna Q_{s1t} przyjmuje wartość 1 dla pierwszej (nieparzystej) godziny każdego okresu dwugodzinnego. W siedmiu dwu- i trzystopniowych modelach hierarchicznych opisujących wahania o cyklu rocznym zmienne Q przyjmują inne wartości. Przykładowo dla modelu HM43 podzielniki wynoszą: $p_1 = 4, p_2 = 3$. Liczby zmiennych Q są o 1 mniejsze. Zmienne zero-jedynkowe $Q_{st(M)}$ ($s = 1, 2, 3$) przyjmują wartości 1 w s -tym kwartale i 0 w pozostałych, a zmienne $Q_{srt(M)}$ ($r = 1, 2$) – wartości 1 w r -tych miesiącach każdego kwartału. Wszystkie omawiane w pracy modele wykazujące złożoną sezonowość: klasyczne szeregu czasowego, regresyjne oraz hierarchiczne (hybrydowe) są modelami jednorównaniowymi.

W przypadku modeli wyrównywania wykładniczego podstawa modelowania a następnie budowy prognoz wyjściowych są dane oczyszczone z dwóch (Y_t^{**}) lub trzech rodzajów sezonowości (Y_t^{***}). Podstawą budowy prognoz będą predyktory oparte odpowiednio na: trzy parametrowych modelach Holta-Wintersa oraz jednoparametrowych modelach Browna lub dwu-równaniowych modelach Holta.

Zapisy analityczne tych modeli i opartych na nich predyktorów można znaleźć w pracy (Szmuksta-Zawadzka, Zawadzki, 2015). Stałe wygładzania przyjmują wartości z przedziału (0,1). Prognozy końcowe otrzymuje się w modelach addytywnych przez dodanie składników, a modelach multiplikatywnych przez przemnożenie przez wskaźniki sezonowości. W pierwszym przypadku podstawą budowy prognoz są predyktory oparte na modelach Holta-Wintersa o trzech równaniach i z trzema stałymi wyrównywania (α, β, δ). W modelu tym równanie trzecie opisuje efekty sezonowe o cyklu 24 godzinnym. Równania i predyktory addytywnego modelu Holta-Wintersa przyjmują postać:

$$m_t^{**} = \alpha(Y_t^{**} - c_{t-L}) + (1 - \kappa)m_{t-1}^{**} \quad (4)$$

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

i.e. the latest observations have a relatively weaker impact, it is enough to assume that the smoothing constants in the Holt-Winters models vary from 0.01 to 0.25. This translates into a significant reduction in the number of versions to be estimated.

The equation and predictors based on the Brown's model are as follows:

$$m_t^{***} = \alpha Y_t^{**} + (1 - \kappa) m_{t-1}^{***} \quad (9)$$

$$\Pi_B^{***} = m_t^{***} \quad (10)$$

$$\Pi_B = \Pi_B^{***} + M^{(a)}(t) + D^{(a)}(t) \quad (11)$$

Equations and predictors of the linear Holt's model:

$$m_t^{***} = \alpha Y_t^{***} + (1 - \kappa)(m_{t-1}^{***} + \delta_{1t-1}^{***}) \quad (12)$$

$$\delta_{1t}^{***} = \beta(m_t^{***} - m_{t-1}^{***}) + (1 - \beta)\delta_{1t-1}^{***} \quad (13)$$

$$\Pi_H^{***} = m_{t_0}^{***} + \delta_{1t_0}^{***} h \quad (14)$$

In the Holt's model, changes in parameters varying from 0.01 to 0.25 are considered sufficient.

$$C_t^{**} = \delta(Y_t^{**} - m_t^{**}) + (1 - \delta)C_{t-m}^{**} \quad (6)$$

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

$$\Pi_{HW} = \Pi^{**} + M^{(a)}(t) \quad (8)$$

Wykorzystanie trójparametrycznych modeli Holta-Wintersa. związane jest z problemem natury numerycznej, zwłaszcza w przypadku, jeżeli parametry α , β , i δ zmieniają się od 0,01 do 0,99, co 0,01, należy szacować wówczas $99^3 = 970299$ wersji.

Jednym z trudniejszych zagadnień, jest wybór „optymalnych” wartości stałych wygładzania bez konieczności szacowania blisko miliona wersji. Z przeprowadzonych wcześniej badań wynika, że jeżeli zmienna charakteryzuje znacznym stopniem inercji tzn. stosunkowo słabszym wpływem najnowszych obserwacji, to wystarczającym jest przyjęcie założenia o tym, że w modelach Holta-Wintersa stałe wygładzania zmieniają się od 0,01 do 0,25. Oznacza to znaczne zmniejszenie liczby szacowanych wersji.

Równanie i predyktory oparte na modelu Browna są następujące:

$$m_t^{***} = \alpha Y_t^{**} + (1 - \kappa) m_{t-1}^{***} \quad (9)$$

$$\Pi_B^{***} = m_t^{***} \quad (10)$$

$$\Pi_B = \Pi_B^{***} + M^{(a)}(t) + D^{(a)}(t) \quad (11)$$

Równania i predyktory liniowego modelu Holta przyjmują postać:

$$m_t^{***} = \alpha Y_t^{***} + (1 - \kappa)(m_{t-1}^{***} + \delta_{1t-1}^{***}) \quad (12)$$

$$\delta_{1t}^{***} = \beta(m_t^{***} - m_{t-1}^{***}) + (1 - \beta)\delta_{1t-1}^{***} \quad (13)$$

$$\Pi_H^{***} = m_{t_0}^{***} + \delta_{1t_0}^{***} h \quad (14)$$

W modelu Holta, także wystarczającym jest przyjęcie założenia o zmianach parametrów z przedziału (0,01-0,25).

Subject and scope of empirical research

The demand for electricity in hourly cycles in agglomeration A was modelled and forecasted. Statistical data on trends in power demand (in MW) originated from the Data Bank of the Department of Applications of Mathematics in Economics at the West Pomeranian University of Technology in Szczecin. The baseline series (without gaps in data) covered the period of two years and 17,520 observations. The forecasts were empirically verified during the third year of analysis. The evolution of the variable in the period of estimation is presented graphically in Figure 1.

Przedmiot i zakres badań empirycznych

Modelowaniu a następnie prognozowaniu poddano zapotrzebowanie na energię elektryczną w okresach godzinnych w aglomeracji A. Dane statystyczne o kształtowaniu się zapotrzebowania na moc (w MW) pochodzą z Banku Danych Katedry Zastosowań Matematyki w Ekonomii ZUT w Szczecinie. Szereg wyjściowy (bez luk) obejmował okres dwóch lat, tj. 17520 obserwacji. Trzeci rok był okresem empirycznej weryfikacji prognoz. Kształtowanie się zmiennej w okresie estymacyjnym zostało przedstawione w sposób graficzny na Rysunku 1.

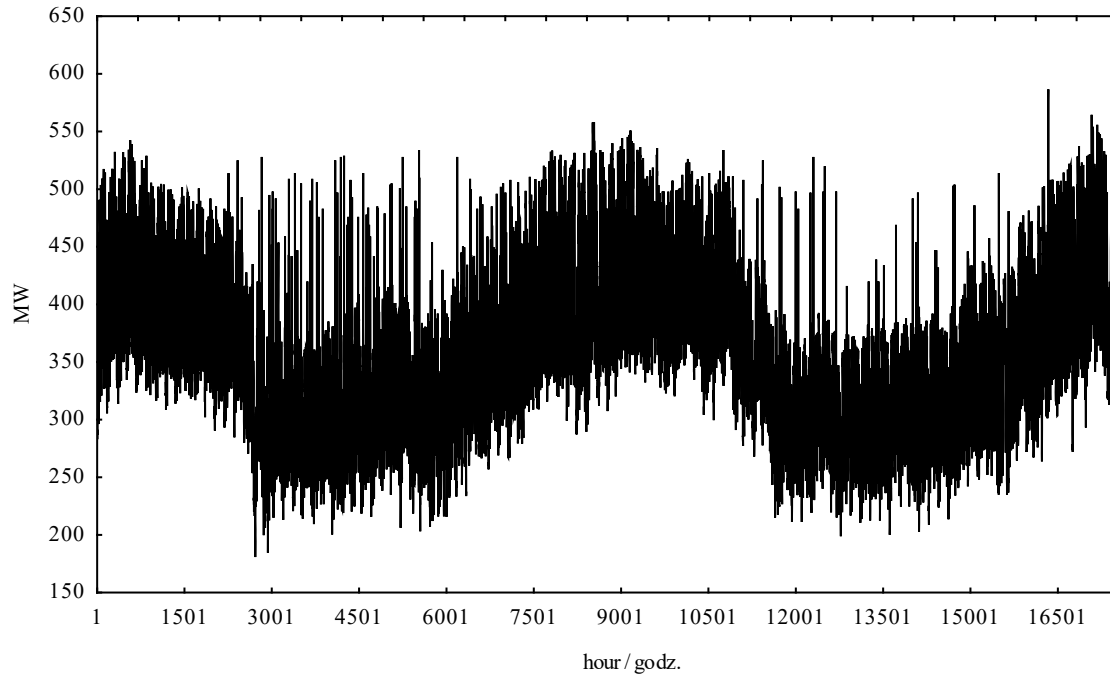


Figure 1. Power demand (MW)

Rysunek 1. Zapotrzebowanie na moc energetyczną (MW)

Source: Data Bank of the Department of Applications of Mathematics in Economics at the West Pomeranian University of Technology in Szczecin
 Źródło: Bank Danych Katedry Zastosowań Matematyki ZUT w Szczecinie.

The best predictors based on classical and hybrid time series models as well as classical and hybrid regression models were used to create forecasts for incomplete data.

Forecasting based on predictors from exponential smoothing models was done for series cleared of two or three types of seasonal fluctuations of the variable.

Two types of data gaps were considered: non-systematic (NS) and systematic (S). Non-systematic gaps occurred in selected hours, days of the week and months, in the period from July 1 of the first year to June 30 of the second year of analysis, while systematic gaps were reported in the same time units throughout the sampled period. Two options were considered for each type of gaps: G and GDM. In the first variant, the gaps occurred within 8 hourly periods (2, 6, 8, 10, 14, 16, 18 and 22): NS_G and S_G, respectively. In the second variant, gaps also occurred on two days of the week (Tuesday and Thursday) and in two months (August and October), NS_GDM and S_GDM, respectively, apart from the specified hours.

Two analytical forms were analyzed for all classes of models: linear and exponential for the classical and regressive time series models, as well as additive and multiplicative for exponential smoothing models.

Two types of forecasts were determined on the basis of the estimated equations: interpolation (I) and extrapolation (E). Interpolation forecasts were calculated for periods within the sample time range,

Do budowy prognoz dla niepełnych danych wykorzystano najlepsze predyktory oparte na: modelach klasycznych i hybrydowych szeregu czasowego oraz klasycznych i hybrydowych modelach regresyjnych.

Prognozowanie na podstawie predyktorów opartych na modelach wyrównywania wykładniczego prowadzono dla szeregów oczyszczonych z dwóch albo trzech rodzajów wahań sezonowych zmiennej.

Rozpatrywane były dwa rodzaje luk w danych: niesystematyczne (NS) i systematyczne (S). Luki niesystematyczne występowały w wybranych: godzinach, dniach tygodnia i miesiącach w okresie od 1 lipca pierwszego roku do 30 czerwca drugiego roku, a luki systematyczne w tych samych jednostkach czasowych w całym przedziale czasowym próby. W ramach każdego rodzaju luk rozpatrywano dwa warianty: G i GDM. W wariantcie pierwszym luki występowały w 8 okresach godzinnych (2, 6, 8, 10, 14, 16, 18 i 22) - oznaczono je odpowiednio: NS_G i S_G. Natomiast w wariantcie drugim, obok wymienionych godzin, także w dwóch dniach tygodnia (wtorek i czwartek) oraz dwóch miesiącach (sierpień i październik) – zostały one oznaczone przez: NS_GDM oraz S_GDM.

Dla wszystkich klas modeli rozpatrywane były dwie postaci analityczne: liniowa i wykładnicza dla klasycznych i regresyjnych modeli szeregu czasowego oraz addytywna i multiplikatywna dla modeli wyrównywania wykładniczego.

Na podstawie oszacowanych równań wyznaczano dwa rodzaje prognoz: interpolacyjne (I)

i.e. those in which gaps occurred, while extrapolation forecasts were calculated for periods beyond this range.

16,349 versions were estimated for each variant, type of gaps, model classes and analytical forms, which makes a total of versions.

Table 1 presents interpolation forecasts in absolute numbers and percentages for each type of gap and variant.

i ekstrapolacyjne (E). Prognozy interpolacyjne oblicza się dla okresów, należących do przedziału czasowego próby, tj. tych w których wystąpiły luki, a prognozy ekstrapolacyjne dla okresów wybiegających poza ten przedział.

Dla każdego wariantu, rodzaju luk, klas modelu i postaci analitycznej szacowano po 16349 wersji, co czyni łącznie wersji.

W Tabeli 1 zestawione zostały prognozy interpolacyjne w liczbach bezwzględnych i odsetkach dla każdego rodzaju luk i wariantu.

Table 1. Number of gaps (interpolation forecasts) by gap types and variants

Tabela 1. Liczby luk (prognoz interpolacyjnych) według rodzajów luk i wariantów

Type of gaps / Variant / Rodzaj luk / Wariant	G		GDM	
	in absolute numbers / w liczbach bezwzględnych	in percentage values / w odsetkach	in absolute numbers / w liczbach bezwzględnych	in percentage values / w odsetkach
Non-systematic (NS) / Niesystematyczne (NS)	2920	16.67%	5272	30.09%
Systematic (S) / Systematyczne (S)	5840	33.33%	10544	60.18%

Source: Own study.

Źródło: Opracowanie własne.

The gaps were obtained by clearing the corresponding numbers of observations from the overall series for the sample time interval of 17520. The parameters of the models were estimated on the basis of the number of observations added to the 17,520 numbers provided in Table 1.

The number of ex post extrapolation forecasts for the third year (empirical verification) was always 8,760 (365 days x 24 hours).

This paper draws on and supplements the previous research on systematic gaps in exponential smoothing models (GDM variant) and in regression models (both variants). It covers all classes of models, analytical forms, variants and types of gaps.

The findings will be analyzed to examine the impact of the number of gaps, the analytical form, the selection criterion, and the type and class of the model on the accuracy of inter- and extrapolation forecasts in forecasting a microeconomic variable.

The research conducted by Szmuksta-Zawadzka and Zawadzki (2014, 2016) and Zawadzki (2018) revealed that forecasting models should be selected according to predictors with minimal estimates of relative errors of interpolation forecasts (MAPE_I) or extrapolation forecasts (MAPE_E). Accordingly, for each of these criteria, the authors listed above calculated type II error of forecasts as auxiliary values: MAPE_{I,E} and MAPE_{E,I}, respectively.

The modeling and forecasting results are presented in Tables 1-4. Their analysis was carried out in two cross-sections. The first cross-section involved comparing the accuracy of inter- and extrapolation

Luki otrzymano przez wymazanie odpowiednich liczb obserwacji z szeregu pełnego liczącego dla przedziału czasowego próby 17520. Parametry modeli szacowane były na podstawie liczb obserwacji będących dopełnieniem do 17520 liczb podanych w Tabeli 1.

Liczba prognoz ekstrapolacyjnych *ex post*, odnosząca się do trzeciego roku (okresu ich empirycznej weryfikacji) każdorazowo wynosiła 8760 (365 dni x 24 godziny).

Praca, jest rozszerzeniem dotychczasowych badań na przypadek występowania luk systematycznych w modelach: wyrównywania wykładniczego (wariant GDM) oraz w modelach regresyjnych (obydwa warianty) i stanowi swoiste domknięcie, obejmujące wszystkie klasy modeli, ich postaci analityczne, warianty i rodzaje luk.

Analiza wyników pozwoli na zbadanie wpływu: liczby luk, postaci analitycznej, kryterium wyboru, rodzaju i klasy modelu na dokładność prognoz inter- i ekstrapolacyjnych w prognozowaniu zmiennej mikroekonomicznej.

Z badań przeprowadzonych w pracach: Szmuksta-Zawadzka i Zawadzki (2014, 2016 oraz Zawadzki (2018) wynika, że kryterium wyboru modeli dla celów prognozowania powinny być predyktory charakteryzujące się minimalnymi ocenami względnych błędów prognoz interpolacyjnych (MAPE_I) bądź prognoz ekstrapolacyjnych (MAPE_E). W przywołanych pracach dla każdego z tych kryteriów obliczono, jako wielkości pomocnicze, błędy drugiego rodzaju prognoz, oznaczonych odpowiednio jako: MAPE_{I,E} oraz MAPE_{E,I}.

forecasts, separately for non-systematic (Table 2) and systematic (Table 3) gaps, for variants G and GDM, obtained on the basis of time series models: hybrid, exponential smoothing and classical ones, as well as hybrid and classical regression models.

In the second cross-section (Tables 4 and 5), the structure, random variation coefficients and forecast errors for were compared for the same classes and variants of models for non-systematic (NS) and systematic (S) gaps, separately for interpolation forecasts (Table 4) and extrapolation forecasts (Table 5). The impact of doubling the number of gaps on the accuracy of the forecasts was examined.

Tables 2 and 3 for variants G and GDM list the following: the model structure and components, the characteristics of model estimates, and errors in inter- and extrapolation forecasts for non-systematic gaps or systematic gaps, respectively.

The first column in each table lists alphanumeric acronyms denoting: type of gaps (NS – non-systematic, S – systematic), analytical form (L – linear, W – exponential in classical and hybrid models, and a – additive and m – multiplicative in exponential smoothing models), types of forecasts (I – interpolation, E – extrapolation), gap occurrence variants (G, GDM), and the model class and components. Columns 2 to 4 for time series models and regression models present: assessments of the following coefficients: determination coefficient (R^2), standard deviations of random components (S_e), and coefficients of random variation (V_{se}). As for exponential smoothing models, estimates of the average errors of the adjusted values are provided ($MAPE_{ww}$) instead of the coefficients of random variation (V_{se}). The last two columns contain the estimates of interpolation ($MAPE_i$) or extrapolation errors ($MAPE_e$). For regression predictors, the letter denoting the analytic form is preceded by the letter P.

The information in column one reveals that all hybrid models for both types of forecasts are single models, predominantly with a hierarchical structure describing fluctuations within a 24-hour cycle. This structure was identified for the fluctuations within a 12-month cycle only in the extrapolation forecasts for the GDM variant. Modeling and forecasting results for non-systematic gaps, for variants G and GDM, are presented in Table 2. According to the data contained in columns 2-4, the estimated values of the coefficients of determination (R^2) for regression models were higher than for time series models. At the same time, standard deviations of random components and coefficients of variation were lower. This proves that they have a higher predictive value.

Wyniki modelowania i prognozowania prezentowane będą w postaci tabelarycznej (Tabele 1-4). Ich analiza prowadzona będzie w dwóch przekrojach. Istota pierwszego przekroju polegać będzie na porównaniu, osobno dla luk niesystematycznych (Tabela 2) i systematycznych (Tabela 3), dokładności prognoz inter- i ekstrapolacyjnych, dla wariantów G i GDM, otrzymanych na podstawie modeli szeregu czasowego: hybrydowych, wyrównywania wykładniczego i modeli klasycznych oraz hybrydowych i klasycznych modeli regresyjnych.

Natomiast w ramach przekroju drugiego, obejmującego Tabele 4 i 5, przeprowadzone zostanie równoczesne porównanie struktury, ocen współczynników zmienności losowej oraz błędów prognoz dla otrzymanych dla tych samych klas i wariantów modeli dla luk niesystematycznych (NS) i systematycznych (S) osobno dla prognoz interpolacyjnych (tabela 4) i osobno dla prognoz ekstrapolacyjnych (Tabela 5). Pozwoli to na dokonanie oceny wpływu podwojenia liczby luk na dokładność prognoz.

W Tabelach 2 i 3 dla wariantów G i GDM zestawione zostały informacje dotyczące: struktury i składowych modeli, charakterystyk oszacowań modeli, błędów prognoz inter i ekstrapolacyjnych odpowiednio dla luk niesystematycznych lub luk systematycznych.

W kolumnie pierwszej każdej z tabel podane zostały kolejno akronimy literowo-liczbowe odnoszące się do: rodzaju luk (NS – niesystematyczne, S – systematyczne) postaci analitycznej (L – liniowa, W – wykładnicza w modelach klasycznych i hybrydowych oraz a – addytywna i m – multiplikatywnej w modelach wyrównywania wykładniczego), rodzajów prognoz (I – interpolacyjne, E – ekstrapolacyjne), wariantów występowania luk (G, GDM), klasy modelu i jego składowych. Następnie w kolumnach od 2 do 4 dla modeli szeregu czasowego i modeli regresyjnych podano kolejno: oceny współczynników: determinacji (R^2), odchyłeń standardowych składników losowych (S_e) oraz współczynników zmienności losowej (V_{se}). W przypadku modeli wyrównywania wykładniczego zamiast współczynników zmienności losowej (V_{se}) podane zostały oceny przeciętnych błędów wartości wyrównanych ($MAPE_{ww}$). Dwie ostatnie kolumny zawierają oceny błędów prognoz interpolacyjnych ($MAPE_i$) lub ekstrapolacyjnych ($MAPE_e$). W przypadku predyktorów regresyjnych przed literą oznaczającą postać analityczną występuje litera P.

Z informacji zawartych w kolumnie pierwszej wynika, że wszystkie modele hybrydowe dla obu rodzajów prognoz są modelami pojedynczymi, w większości o strukturze hierarchicznej opisującej wahania o cyklu, 24 godzinnym. Jedynie w przypadku prognoz ekstrapolacyjnych dla wariantu GDM taką strukturę posiadają wahania o cyklu 12-miesięcznym. Prezentacja wyników modelowania i prognozowania dla luk niesystematycznych, dla wariantów G i GDM

przedstawiona została w Tabeli 2. Z analizy danych zawartych w kolumnach 2-4 wynika, że dla modeli regresyjnych oceny współczynników determinacji (R^2) przyjęły wyższe wartości niż dla modeli szeregu czasowego. Jednocześnie oceny odchyłeń standardowych składników losowych i współczynników zmienności były niższe. Świadczy to o ich lepszych własnościach predykcyjnych.

Table 2. Evaluation of the parameters of stochastic structure and errors of inter- and extrapolation forecasts for selected models for non-systematic gaps

Tabela 2. Oceny parametrów struktury stochastycznej oraz błędów prognoz inter- i ekstrapolacyjnych dla wybranych modeli dla luk niesystematycznych

Model / Model	R^2	S_e [MW]	V_{se} [%]	MAPE _i [%]	MAPE _E [%]
NS_WI_G_HG432_0/1	0.7661	36.04	9.785	6.391	-
NS_WI_GDM_HG122_0/1	0.8106	32.49	8.823	6.584	
NS_WE_G_HG122_0/1	0.8106	32.44	8.808	-	8.775
NS_WE_GDM_0/1_HM43	0.753	37.11	10.075	-	9.257
NS_aI_G_B(0.95)	-	-	4,335*	4.323	-
NS_mI_GDM_H(0.20;0.01)	-	-	5,429*	5.302	-
NS_aE_G_HW(0.11;0.01;0.01)	-	-	6,181*	-	7.081
NS_mE_GDM_B(0.66)	-	-	4,371*	-	7.254
NS_W_G_0/1_0/1	0.8207	31.57	8.572	6.194	8.554
NS_W_GDM_0/1_0/1	0.8227	31.45	8.540	6.308	9.129
NS_PLI_G_HG432_0/1	0.8344	29.63	8.107	5.406	-
NS_PLI_GDM_HG122_0/1	0.8454	28.66	7.847	5.452	-
NS_PWE_G_HG432_0/1	0.8379	29.75	8.145	-	6.622
NS_PWE_GDM_0/1_HM43	0.8364	29.96	8.202	-	6.535
NS_PL_G_0/1_0/1	0.8485	28.36	7.760	5.374	6.906
NS_PW_GDM_0/1_0/1	0.8553	28.18	7.716	5.406	6.983

*MAPE_{ww}

Source: Own study based on: Szmuksta-Zawadzka and Zawadzki (2014, 2015, 2016); Zawadzki, 2020, and own calculations.

Źródło: Opracowanie własne na podstawie: Szmuksta-Zawadzka i Zawadzki (2014, 2015, 2016); Zawadzki, 2020 oraz obliczenia własne.

For exponential smoothing models, the estimates of MAPE_{ww} coefficients are marked with an asterisk (*). Their values ranged from 4.335* to 6.181*% and were lower than the coefficients of random variation of other models by at least 1.5 percentage points (p.p.). This may be indicative of a higher predictive value of adaptive models.

The data in the penultimate column indicate that the estimates of interpolation forecast errors were the lowest for adaptive predictors. They equaled 4.323% for the Brown's exponential smoothing model for variant G, and 5.302% for the Holt's model for variant GDM. They differed only by about one percentage point. Estimates for the hybrid regression model for variant G were higher by approx. 1.1 p.p., and by approx. 0.4 p.p. for variant GDM. Similar values were obtained for hybrid regression models, with a difference of up to 0.046 p.p., and were slightly lower for the classical models.

Higher values of more than 6 p.p. were obtained for the classical and hybrid time series models. The error relationship was similar to that in the regression models.

Dla modeli wyrównywania wykładniczego oceny współczynników MAPE_{ww} zostały oznaczone (*). Z ich kształtowania się wynika, że przyjęły one wartości z przedziału od 4,335* do 6,181*% i były niższe od współczynników zmienności losowej pozostałych modeli o co najmniej 1,5 punktu procentowego (p.p.). Może to świadczyć o wyższych własnościach predykcyjnych modeli adaptacyjnych.

Z informacji zawartych w przedostatniej kolumnie wynika, że najniższe oceny błędów prognoz interpolacyjnych otrzymano dla predyktorów adaptacyjnych. Wyniosły one odpowiednio: 4,323% dla modelu wyrównywania wykładniczego Browna dla wariantu G oraz 5,302% dla modelu Holta – wariant GDM – różniły się jedynie o około jeden p.p.. Oceny wyższe ok. 1,1 p.p. otrzymano dla hybrydowego modelu regresyjnego dla wariantu G i ok. 0,4 p.p dla wariantu GDM. Dla hybrydowych modeli regresyjnych, otrzymano oceny bardzo zbliżone, różniące się o nie więcej niż o 0,046 p.p., przy czym były one nieznacznie niższe dla modeli klasycznych.

Wyższe oceny, przekraczające 6 p.p., otrzymano dla klasycznych i hybrydowych modelu szeregu

The variant GDM values for extrapolation forecast errors presented in the last column were generally higher than the values obtained for variant G. Hybrid regression models with the lowest error estimates of 6.622% and 6.535%, respectively, were the only exception. Error estimates for predictors based on exponential smoothing models exceeded 7% by a narrow margin, and were higher by approx. 0.4-0.5 p.p. The highest values ranging from 8.775 to 9.257% were obtained for hybrid and classical time series models.

Table 3 lists model characteristics, parameters of the stochastic structure, and estimates of errors in inter- and extrapolation forecasts for systematic gaps, in the same arrangement. According to the first column, single models with hierarchical components within a 24-hour cycle were obtained for variant G, for two hybrid time series models and one regression model (S_PLI_G_HG432_0/1). Unlike the non-systematic gaps, 4 dual hybrid models were obtained for the variant GDM: two for time series models and two for regression models. A dual model was also obtained for variant G for the regression predictor: S_PLE_G_HG2232_HM43.

czasowego. Relacja błędów była podobna jak dla modeli regresyjnych.

W przypadku zamieszczonych w ostatniej kolumnie błędów prognoz ekstrapolacyjnych ich oceny dla wariantu GDM były generalnie wyższe od ocen otrzymanych dla wariantu G. Jedyny wyjątek dotyczy hybrydowych modeli regresyjnych – otrzymano dla nich najniższe oceny błędów wynoszące odpowiednio: 6,622% i 6,535%. Oceny błędów dla predyktorów opartych na modelach wyrównywania wykładniczego nieznacznie przekraczały 7%, a więc były wyższe o ok. 0,4-0,5 p.p. Najwyższe oceny, kształtujące się w granicach 8,775 – 9,257% otrzymano dla hybrydowych i klasycznych modeli szeregu czasowego.

W Tabeli 3 zamieszczono, w takim samym układzie jak poprzednio: charakterystyki modeli, oceny parametrów struktury stochastycznej oraz oceny błędów prognoz inter- i ekstrapolacyjnych dla luk systematycznych. Z kolumny pierwszej wynika, że dla wariantu G dla dwóch modeli hybrydowych szeregu czasowego i jednego modelu regresyjnego (S_PLI_G_HG432_0/1) otrzymano modele pojedyncze z hierarchicznymi składowymi dotyczącymi cyklu 24 godzinnego. Dla wariantu GDM, w odróżnieniu od luk niesystematycznych otrzymano 4 podwójne modele hybrydowe, po dwa dla modeli szeregu czasowego i modeli regresyjnych. Model podwójny otrzymano także dla wariantu G dla predyktora regresyjnego: S_PLE_G_HG2232_HM43.

Table 3. Evaluation of the parameters of stochastic structure and errors of inter- and extrapolation forecasts for selected models and systematic gaps

Tabela 3. Oceny parametrów struktury stochastycznej oraz błędów prognoz inter- i ekstrapolacyjnych dla wybranych modeli i luk systematycznych

Models / Modele	R ²	S _e [MW]	V _{se} [%]	MAPE _i [%]	MAPE _e [%]
S_WI_G_HG432_0/1	0.753	36.40	9.965	6.581	-
S_LI_GDM_HG122_HM62	0.772	35.15	9.579	7.798	-
S_WE_G_HG122_0/1	0.864	31.72	8.683	-	8.633
S_WE_GDM_HG122_HM62	0.789	34.35	9.361	-	9.786
S_mI_G_HW(0.15;0.01;0.15)	-	-	4,835*	4.051	-
S_mI_GDM_B(0.45)	-	-	5,364*	5.540	-
S_mE_G_B(0.57)	-	-	4,935*	-	7.009
S_mE_GDM_B(0.57)	-	-	5,214*	-	7.009
S_W_G_0/1_0/1	0.877	31.23	8.551	12.343	10.095
S_W_GDM_0/1_0/1	0.818	31.91	8.695	14.881	14.539
S_PLI_G_HG432_0/1	0.833	29.69	8.128	5.472	-
S_PLI_GDM_HG432_HM62	0.822	31.04	8.458	5.6557	-
S_PLE_G_HG2232_HM43	0.816	31.20	8.541	-	6.573
S_PLE_GDM_HG2232_HM34	0.804	32.55	8.870	-	6.678
S_PL_G_0/1_0/1	0.849	28.28	7.742	7.989	7.497
S_PL_GDM_0/1_0/1	0.842	29.24	7.969	9.613	9.886

*MAPE_{ww}

Source: Own study based on: Szmuksta-Zawadzka and Zawadzki (2014, 2015, 2016); Zawadzki, 2020, and own calculations.

Źródło: Opracowanie własne na podstawie: Szmuksta-Zawadzka i Zawadzki (2014, 2015, 2016); Zawadzki, 2020 oraz obliczenia własne.

Likewise, models with a different structure were obtained for pairs of predictors belonging to the same class, for both variants. The analytical form was at least one discriminating factor. This is most evident for the GDM variant for dual time series hybrid predictors – S_LI_GDM_HG122_HM62 and S_WE_GDM_HG122_HM62. These are either linear or exponential, respectively.

For variants G and GDM, the minimum estimates of interpolation forecast errors of 4.051* and 5.540*% were obtained for exponential smoothing models. The first one was the Holt-Winters model with smoothing constants: 0.15; 0.01 and 0.15; the Brown's model with the equalization constant 0.45 was the second one. Hybrid regression models with error estimates of 5.472 and 5.656% were ranked second. By far the highest error estimates – 12.343% for variant G and 14.881% for variant GDM – were obtained for the classical time series models. For the classical regression models, they equaled 7.798% and 9.613%, respectively. This means that predictors based on both classical time series models and classical regression models, having the highest estimates of errors in interpolation forecasts, are not eligible to be used in the forecasting process.

In terms of extrapolation forecasts, the lowest error estimates for variants G and GDM of 6.573% and 6.678%, respectively, were obtained in regression hybrid equations. Very small differences in the error estimates of 0.105 p.p. for regression models provide evidence that the accuracy of extrapolation forecasts was not affected by doubling the share of gaps from 30.08% to 60.16%. Only slightly higher error estimates of 7.009% for both variants were obtained for predictors based on the multiplicative Brown's model with the same smoothing constant of 0.57.

The highest error estimates of 10.095% and 14.539% were also obtained for the classical time series models. This unequivocally proves their poor forecasting value. These error estimates were higher by approx. 3.1 and 5.5 p.p., respectively, for the classical regression models.

With the data provided in Tables 2 and 3, the accuracy of inter- and extrapolation forecasts was compared, separately for non-systematic (Table 2) and systematic (Table 3) gaps, for variants G and GDM, obtained on the basis of time series models: hybrid, exponential smoothing and classical ones, as well as hybrid and classical regression models.

The accuracy for non-systematic and systematic gaps, estimated separately, for interpolation (Table 4) and extrapolation (Table 5) forecasts will now be compared, obtained from predictors for the same classes of time series models, exponential smoothing models, and regression models for variants G and GDM. The gap percentage points for variants G and GDM are as follows: 16.67% and 33.33%, and 30.09%

Ponadto, podobnie jak poprzednio, dla par predyktorów należących do tej samej klasy, dla obu wariantów, otrzymano modele o różnej strukturze, różniące się przynajmniej postacią analityczną. Najlepiej widoczne jest to w odniesieniu do wariantu GDM dla podwójnych predyktorów hybrydowych szeregu czasowego – S_LI_GDM_HG122_HM62 oraz S_WE_GDM_HG122_HM62. W pierwszym przypadku jest to postać liniowa a w drugim wykładnicza.

Dla wariantów G i GDM minimalne oceny błędów prognoz interpolacyjnych, wynoszące 4,051* i 5,540*% otrzymano dla modeli wyrównywania wykładniczego. W pierwszym przypadku był to model Holta-Wintersa o stałych wygładzania: 0,15; 0,01 i 0,15 a w drugim model Browna o stałej wyrównywania 0,45. Drugimi w kolejności były hybrydowe modele regresyjne z ocenami błędów wynoszącymi: 5,472 i 5,656%. Zdecydowanie najwyższe oceny błędów otrzymano dla klasycznych modeli szeregu czasowego. Kształtowały się one na poziomie 12,343% dla wariantu G i 14,881% dla wariantu GDM. Dla klasycznych modeli regresyjnych przyjęły one wartości wynoszące odpowiednio: 7,798% i 9,613%. Oznacza to, że predyktory oparte, zarówno o klasyczne modele szeregu czasowego jak i klasyczne modele regresyjne, charakteryzujące się najwyższymi ocenami błędów prognoz interpolacyjnych, nie będą mogły być wykorzystane w procesie prognozowania.

W przypadku prognoz ekstrapolacyjnych najniższe oceny błędów dla wariantów G i GDM wynoszące odpowiednio: 6,573% i 6,678% otrzymano na podstawie regresyjnych równań hybrydowych. Bardzo niewielkie różnice ocen błędów dla modeli regresyjnych, wynosząca 0,105 p.p., świadczy o tym, że podwojenie udziału luk z 30,08% do 60,16% nie miało wpływu na dokładność prognoz ekstrapolacyjnych. Tylko nieznacznie wyższe oceny błędów, wynoszące 7,009% dla obu wariantów, otrzymano dla predyktorów opartych na multiplikatywnym modelu Browna o tej samej stałej wygładzania, wynoszącej 0,57.

Najwyższe oceny błędów, wynoszące 10,095% i 14,539% otrzymano także dla klasycznych modeli szeregu czasowego. Jednoznacznie przesądza to o braku ich przydatności do budowy prognoz. Dla klasycznych modeli regresyjnych oceny te były wyższe odpowiednio o ok. 3,1 i 5,5 p.p.

Informacje zawarte w Tabelach 2 i 3 pozwoliły na dokonane porównanie dokładności prognoz inter- i ekstrapolacyjnych, osobno dla luk niesystematycznych i systematycznych, dla wariantów G i GDM, otrzymanych na podstawie modeli szeregu czasowego: hybrydowych, wyrównywania wykładniczego i modeli klasycznych oraz hybrydowych i klasycznych modeli regresyjnych.

Obecnie przechodzimy do porównania dokładności dla luk niesystematycznych i systematycznych osobno prognoz interpolacyjnych (Tabela 4)

and 60.18%, respectively. The first and fifth columns present alpha-numeric acronyms referring to the same classes of models and variants of gaps per row. The lowest estimates for variability coefficients and $MAPE_{ww}$ were obtained for exponential smoothing models (4.323*% – 5.421*%). The remaining estimates ranges from 7.716% for predictor NS_PL_GDM_0/1_0/1 to 9.785% for predictor NS_WI_G_HG432_0/1.

The penultimate column lists differences in the estimates of forecasting errors for systematic (S) and non-systematic (NS) gaps (p.p.=MAPE(S)-MAPE(NS)). However, to facilitate interpretation, the last column contains quotients $[(MAPE(S)-MAPE(NS))/(MAPE(S))*100]$ expressed as percentage points. The data in the penultimate column demonstrate that the error estimates in interpolation forecasts for systematic gaps are generally higher, as evidenced by the positive values of differences. The only exception is the higher error estimate of 4.335*% for the exponential smoothing model for variant G for non-systematic gaps NS_aI_G_B (0.95) compared to that for systematic gaps (S_mI_G_HW (0.15; 0.01; 0.15) of - 4.051*%. It is an additive (a) Brown's model for NS gaps, and a multiplicative (m) Holt-Winters model for gaps S.

i ekstrapolacyjnych (Tabela 5) otrzymanych na podstawie predyktorów dla tych samych klas modelach szeregu czasowego, wyrównywania wykładniczego i modelach regresyjnych dla wariantów G i GDM. Dla wariantu G odsetki luk wynoszą: 16,67 i 33,33% a dla wariantu GDM: 30,09% i 60,18%. W kolumnach pierwszej i piątej podane zostały akronimy literowo-liczbowe odnoszące się w poszczególnych wierszach do tych samych klas modeli i wariantów luk. Najniższe oceny spośród współczynników: zmienności i $MAPE_{ww}$ otrzymano dla modeli wyrównywania wykładniczego – zawarte one były w przedziale (4,323*% – 5,421*%). Pozostałe oceny mieszczą się przedziale od 7,716% dla predyktora NS_PL_GDM_0/1_0/1 do 9,785% dla predyktora NS_WI_G_HG432_0/1.

W przedostatniej kolumnie zamieszczono różnice ocen błędów prognoz dla luk systematycznych (S) i niesystematycznych (NS)(p.p.=MAPE(S)-MAPE(NS)) Natomiast ze względów interpretacyjnych w ostatniej kolumnie zamieszczono wyrażone w % ilorazy $[(MAPE(S)-MAPE(NS))/(MAPE(S))*100]$. Z informacji zawartych w kolumnie przedostatniej wynika, że oceny błędów prognoz interpolacyjnych dla luk systematycznych są generalnie wyższe – świadczą o tym dodatnie wartości różnic. Jedynym wyjątkiem jest wyższa ocen błędu dla modelu wyrównywania wykładniczego dla wariantu G dla luk niesystematycznych NS_aI_G_B (0,95) wynosząca 4,335*% w porównaniu z modelem dla tego wariantu dla luk systematycznych (S_mI_G_HW (0,15; 0,01; 0,15) – 4.051*%. Dla luk NS jest to addytywny (a) model Browna a dla luk S jest to model multiplikatywny (m) Holta-Wintersa.

Table 4. Comparison of the accuracy of extrapolation forecasts for non-systematic and systematic gaps for variants G and GDM
Tabela 4. Porównanie dokładności prognoz ekstrapolacyjnych dla luk niesystematycznych i systematycznych dla wariantów G i GDM

Non-systematic gaps (NS) / Luki niesystematyczne (NS)			Systematic gaps (NS) / Luki systematyczne (S)			p.p.	%
Models / Modele	V_{se} [%]	MAPE (NS)	Models / Modele	V_{se} [%]	MAPE(S)		
NS_WI_G_HG432_0/1	9.785	6.391	S_WI_G_HG432_0/1	9.965	6.581	0.190	2.89
NS_WI_GDM_HG122_0/1	8.823	6.584	S_LI_GDM_HG122_HM62	9.579	7.798	1.214	15.57
NS_aI_G_B(0.95)	4,335*	4.323	S_mI_G_HW(0.15;0.01;0.15)	4,835*	4.051	-0.272	-6.71
NS_mI_GDM_H(0.20;0.01)	5,421*	5.302	S_mI_GDM_B(0.45)	5,364*	5.540	0.238	4.30
NS_W_G_0/1_0/1	8.572	6.194	S_W_G_0/1_0/1	8.551	12.343	6.149	49.82
NS_W_GDM_0/1_0/1	8.54	6.308	S_W_GDM_0/1_0/1	8.695	14.881	8.573	57.61
NS_PLI_G_HG432_0/1	8.107	5.406	S_PLI_G_HG432_0/1	8.128	5.472	0.066	1.21
NS_PLI_GDM_HG122_0/1	7.847	5.452	S_PLI_GDM_HG432_HM62	8.458	5.6557	0.204	3.60
NS_PL_G_0/1_0/1	7.76	5.374	S_PL_G_0/1_0/1	7.742	7.989	2.615	32.73
NS_PL_GDM_0/1_0/1	7.716	5.406	S_PL_GDM_0/1_0/1	7.969	9.613	4.207	43.76

* $MAPE_{ww}$

Source: Own calculations based on data in Tables 2 and 3.

Źródło: Opracowanie własne na podstawie Tabel 2 i 3 i obliczenia własne.

The highest estimates of differences in interpolation forecast errors of 6.149 and 8.572 p.p. (49.8 and 87.6%), respectively, were obtained for the classical time series models, and followed by the classical regression models: 2.615 and 4.207 p.p. (32.7 and 43.8%), respectively. This difference exceeded 1.214 p.p. (15.6%) only for variant GDM of the hybrid time series model (S_LI_GDM_HG122_HM62 and NS_WI_GDM_HG122_0/1). If the differences exceed the commonly recognized criterion of 8%, the classical time series models and classical regression models as well as the S_LI_GDM_HG122_HM62 model are unequivocally found ineligible for developing forecasts for systematic gaps for variants G and GDM. The lowest error estimates and the lowest difference of 0.066 p.p. were obtained for predictors: S_PLI_G_HG432_0/1 and NS_PLI_G_HG432_0/1 among the hybrid models. The absolute differences for remaining models did not exceed 0.272 p.p. (6.7%). Accordingly, doubling the number of gaps has no significant impact on the accuracy of interpolation predictions.

The layout of information provided in Table 5 and Table 4 is the same and serves as a basis for comparing the accuracy of extrapolation forecasts for non-systematic and systematic gaps.

The maximum values of errors in extrapolation forecasts are higher than the estimates of errors in interpolation forecasts. The estimates were the lowest for hybrid regression models.

Najwyższe oceny wartości różnic błędów prognoz interpolacyjnych, wynoszące odpowiednio: 6,149 i 8,572 p.p. (49,8 i 87,6%) otrzymano dla klasycznych modeli szeregu czasowego a następnie dla klasycznych modeli regresyjnych: 2,615 i 4,207 p.p. (32,7 i 43,8%). Ponadto jedynie dla modeli hybrydowych szeregu czasowego (S_LI_GDM_HG122_HM62 i NS_WI_GDM_HG122_0/1), dla wariantu GDM różnica ta przekroczyła 1,214 p.p (15,6%). Otrzymanie wartości różnic, przekraczających przyjęte umowne kryterium 8%, przesądza jednoznacznie o tym, że dla wariantów G i GDM klasyczne modele szeregu czasowego i klasyczne modele regresyjne oraz model S_LI_GDM_HG122_HM62 nie mogą być wykorzystywane do budowy prognoz dla luk systematycznych. Spośród modeli hybrydowych najniższe oceny błędów i najniższą różnicę, wynoszącą 0,066 p.p., otrzymano dla predyktorów: S_PLI_G_HG432_0/1 i NS_PLI_G_HG432_0/1. Dla pozostałych modeli wartości ich bezwzględnych różnic nie przekraczały 0,272 p.p. (6,7%). Oznacza to, że podwojenie liczby luk nie wpływa istotnie na dokładność prognoz interpolacyjnych.

W Tabeli 5 zamieszczone zostały w takim samym układzie jak w Tabeli 4 informacje będące podstawą do porównania dokładności prognoz ekstrapolacyjnych dla luk niesystematycznych i systematycznych.

Wartości maksymalne błędów prognoz ekstrapolacyjnych kształtują się na poziomie wyższym niż oceny błędów prognoz interpolacyjnych. Najniższe oceny otrzymano dla hybrydowych modeli regresyjnych.

Table 5. Comparison of the accuracy of extrapolation forecasts for non-systematic and systematic gaps for variants G and GDM
Tabela 5. Porównanie dokładności prognoz ekstrapolacyjnych dla luk niesystematycznych i systematycznych dla wariantów G i GDM

Non-systematic gaps (NS) / Luki niesystematyczne (NS)			Systematic gaps (NS) / Luki systematyczne (S)			p.p.	%
Models / Modele	V _{se} [%]	MAPE (NS)	Models / Modele	V _{se} [%]	MAPE (S)		
NS_WE_G_HG122_0/1	8.808	8.775	S_WE_G_HG122_0/1	8.683	8.633	-0.142	-1.64
NS_WE_GDM_0/1_HM43	10.075	9.257	S_WE_GDM_HG122_HM62	9.361	9.786	0.529	5.41
NS_aE_G_HW(0.11;0.01;0.01)	6,181*	7.081	S_mE_G_B(0.57)	4,935*	7.009	-0.072	-1.03
NS_mE_GDM_B(0.66)	4,371*	7.254	S_mE_GDM_B(0.57)	5,214*	7.009	-0.245	-3.50
NS_W_G_0/1_0/1	8.572	8.554	S_W_G_0/1_0/1	8.551	10.095	1.541	15.26
NS_W_GDM_0/1_0/1	8.54	9.129	S_W_GDM_0/1_0/1	8.695	14.539	5.410	37.21
NS_PWE_G_HG432_0/1	8.145	6.622	S_PLE_G_HG2232_HM43	8.541	6.573	-0.049	-0.75
NS_PWE_GDM_0/1_HM43	8.202	6.535	S_PLE_GDM_HG2232_HM34	8.870	6.678	0.143	2.14
NS_PL_G_0/1_0/1	7.76	6.906	S_PL_G_0/1_0/1	7.742	7.497	0.591	7.88
NS_PW_GDM_0/1_0/1	8.54	9.129	S_W_GDM_0/1_0/1	8.695	9.886	0.757	7.66

*MAPE_{ww}

Source: Own calculations based on data in Tables 2 and 3.

Źródło: Opracowanie własne na podstawie Tabel 2 i 3 i obliczenia własne.

The highest estimates of 1.541 and 5.410 p.p. (15.3 and 37.2%) of the differences in extrapolation forecast errors for variants G and GDM, respectively, were obtained for the classical time series models. This means that these models are not eligible to be used for extrapolative forecasting with systematic gaps. As for classical regression models, these differences equaled 0.591 and 0.757 p.p. (7.88 and 7.65%) and only narrowly missed the eight percent threshold of admissibility.

It is notable that five negative differences were identified between the forecasting errors for systematic and non-systematic gaps: four for variant G and one for variant GDM. This means that the errors of extrapolation forecasts obtained for models with systematic gaps were lower than the errors of models with non-systematic gaps. The maximum value of the error difference in question was 3.50%, and these differences should qualify as negligible.

Conclusions

The classical, hybrid and exponential smoothing time series models, as well as classical and hybrid regression models were used for modeling and forecasting of hourly series with complex seasonality for incomplete data. The following conclusions can be drawn from these analyzes:

1. Models should be selected based on criteria designed to minimize errors in interpolation ($MAPE_I$) or extrapolation ($MAPE_E$) forecasts.
2. The best models selected for inter- and extrapolation forecasting for non-systematic and systematic gaps in variants G and GDM, for individual classes, usually differ in structure, analytical form and predictive properties.
3. Adaptive and hybrid regression models have proved to have the most favorable predictive properties.
4. Most hybrid time series models for both types of gaps and both variants are single models that mainly describe fluctuations within a 24-hour cycle. As for regression models, these are dual models that also describe fluctuations within a 12-month cycle.
5. The estimates of extrapolation forecast errors for both types of gaps were higher than for interpolation forecasts, and lower for variant G than for variant GDM.
6. The lowest estimators of prediction errors involving interpolation were obtained for exponential smoothing models, followed by hybrid regression models. A reverse sequence was obtained for extrapolative forecasting.
7. The highest error estimates for both types of forecasts were obtained for the classical time

Najwyższe oceny wartości różnic błędów prognoz ekstrapolacyjnych, dla wariantów G i GDM wynoszące odpowiednio: 1,541 i 5,410 p.p. (15,3 i 37,2%) otrzymano dla klasycznych modeli szeregu czasowego. Oznacza to, że modele te nie mogą być wykorzystywane do budowy prognoz ekstrapolacyjnych dla luk systematycznych. Dla klasycznych modeli regresyjnych różnice te wyniosły: 0,591 i 0,757 p.p. (7,88 i 7,65%) i mieściły się na granicy ośmioprocentowego progu dopuszczalności.

Na wyraźne podkreślenie zasługuje fakt otrzymania pięciu ujemnych różnic między błędami prognoz dla luk systematycznych i niesystematycznych, w tym czterech dla wariantu G i jednej dla wariantu GDM. Świadczy to o tym, że błędy prognoz ekstrapolacyjnych otrzymane dla modeli z lukami systematycznymi były niższe od błędów modeli z lukami niesystematycznymi. Z uwagi na to, że maksymalna wartość omawianej różnicy błędów wyniosła 3,50%, różnice te należy uznać za nieistotne.

Wnioski

W modelowaniu i prognozowaniu dla szeregów godzinnych ze złożoną sezonowością dla niekompletnych danych wykorzystano modele szeregu czasowego: klasyczne, hybrydowe i wyrównywania wykładniczego oraz klasyczne i hybrydowe modele regresyjne). Z analiz przeprowadzonych w pracy można wyprowadzić następujące wnioski:

1. Podstawą ich wyboru modeli być powinny być kryteria minimalizujące błędy prognoz interpolacyjnych ($MAPE_I$) lub ekstrapolacyjnych ($MAPE_E$).
2. Najlepsze modele wybrane dla celów prognozowania inter- i ekstrapolacyjnego dla luk niesystematycznych i systematycznych dla wariantów G i GDM dla poszczególnych klas zazwyczaj różnią się: strukturą, postacią analityczną i własnościami predyktywnymi.
3. Najkorzystniejsze własności predyktywne posiadają modele adaptacyjne i hybrydowe modele regresyjne.
4. Większość modeli hybrydowych szeregu czasowego dla obu rodzajów luk i obu wariantów są to modele pojedyncze na ogół opisujące wahania o cyklu 24-godzinnym. Natomiast dla modeli regresyjnych modele podwójne, opisujące także wahania o cyklu 12 miesięcznym.
5. Dla obu rodzajów luk oceny błędów prognoz ekstrapolacyjnych przyjmowały oceny wyższe niż dla prognoz interpolacyjnych, przy czym dla wariantu G były one niższe niż dla wariantu GDM.
6. Najniższe oceny błędów prognoz interpolacyjnych otrzymano dla modeli wyrównywania wykładniczego a następnie hybrydowych modeli regresyjnych. Natomiast dla prognoz ekstrapolacyjnych kolejność była odwrotna.

- series models, and were higher for variant GDM. This makes this class of models ineligible for forecasting.
8. The classical time series models and the classical regression models cannot be used in forecasting, as revealed by a comparative analysis of errors in interpolation forecasts for both variants of non-systematic gaps and systematic gaps, for both analytical forms.
 9. This is true only for the classical time series models of extrapolation forecasts. As for classical regression models, the relative percentage differences do not exceed the assumed 8% criterion of acceptability, which appears reasonable given the 100% increase in the number of gaps.
 10. There were five cases where the relative errors were lower for systematic gaps, by no more than 3.5%, than for models of the same class for non-systematic gaps. This proves that the accuracy of extrapolation forecasts is not affected if the number of gaps increases.
7. Najwyższe oceny błędów obu rodzajów prognoz otrzymano dla klasycznych modeli szeregu czasowego, przy czym przyjmowały one wartości wyższe dla wariantu GDM. Przesądza to o niemożności wykorzystania tej klasy modeli.
 8. Z analizy porównawczej błędów prognoz interpolacyjnych dla obu wariantów luk niesystematycznych i luk systematycznych dla obu postaci analitycznych, że modele klasyczne szeregu czasowego i klasyczne modele regresyjne nie mogą być wykorzystane w prognozowaniu.
 9. W przypadku prognoz ekstrapolacyjnych dotyczy to jedynie modeli klasycznych szeregu czasowego. Natomiast dla klasycznych modeli regresyjnych względne różnice procentowe nie przekraczają założonego 8 procentowego kryterium dopuszczalności, co przy 100 procentowym wzroście liczby luk nie jest wielkością wygórowaną.
 10. W pięciu przypadkach dla luk systematycznych otrzymano błędy względne niższe, nie więcej niż o 3,5%, niż dla modeli tej samej klasy dla luk niesystematycznych. Świadczy to o braku wpływu zwiększenia liczby luk na dokładność prognoz ekstrapolacyjnych.

Bibliografia

1. Dordonnat, V., Koopman, S.J., Ooms, M., Dessertaine, A. (2008). An Hourly Periodic State Space Model for Modeling French National Electricity Load. *International Journal of Forecasting*, 24: 588-587. DOI: <https://doi.org/10.1016/j.ijforecast.2008.08.010>.
2. Engle, R. F., Granger, C. W. J., Rice, J., Weiss, A. (1986). Semiparametric Estimates of the Relation between Weather and Electricity. *Journal of the American Statistical Association*, 81: 310-320.
3. Harvey, A.C., Koopman, S.J. (1993). Forecasting Hourly Electricity Demand Using Time-varying Splines. *Journal of the American Statistical Association*, 88: 1228-1237.
4. Kufel, T. (2010). *Ekonometryczna analiza cykliczności procesów gospodarczych o wysokiej częstotliwości obserwacji*. Toruń: Wydawnictwo Naukowe.
5. Lichota, A. (2006). *Prognozowanie krótkoterminowe na lokalnym rynku energii elektrycznej* [Rozprawa doktorska, Akademia Górniczo-Hutnicza w Krakowie]. Pobrane z: <https://winntbg.bg.agh.edu.pl/rozprawy/9746/full.pdf> (21.01.2023).
6. Misiorek, A., Weron, R. (2004). Modelowanie sezonowości a prognozowanie zapotrzebowania na energię elektryczną. *Energetyka*, 12: 794-799. Pobrane z: http://prac.im.pwr.wroc.pl/~hugo/publ/MisiorekWeron04_Energetyka_szcotka.pdf (21.01.2023).
7. Nowicka-Zagrajek, J., Weron, R. (2002). Modeling Electricity Loads in California: ARMA Models with Hyperbolic Noise. *Signal Processing*, 82: 1903-1915. DOI: [https://doi.org/10.1016/S0165-1684\(02\)00318-3](https://doi.org/10.1016/S0165-1684(02)00318-3).
8. Ramanathan, R., Engle, R., Granger, C. J. V., Vahid-Araghi, F., Brace, C. (1997). Short-run Forecast of Electricity Loads and Peaks. *International Journal of Forecasting*, 13: 161-174. DOI: [https://doi.org/10.1016/S0169-2070\(97\)00015-0](https://doi.org/10.1016/S0169-2070(97)00015-0).
9. Szmuksta-Zawadzka, M., Zawadzki, J. (2002). Hierarchiczne modele szeregów czasowych z wahaniami sezonowymi. Budowa. Estymacja. Prognozowanie. W: A. Zeliaś (red.), *Przestrzenno-czasowe modelowanie i prognozowanie zjawisk gospodarczych* (s. 193-204). Kraków: Wydawnictwo Akademii Ekonomicznej.
10. Szmuksta-Zawadzka, M., Zawadzki, J. (2011). Modelowanie i prognozowanie zapotrzebowania na energię elektryczną w mikroskali. *Prace Naukowe UE we Wrocławiu*, 165: 152-161.
11. Szmuksta-Zawadzka, M., Zawadzki, J. (2014). Modele hierarchiczne w prognozowaniu zmiennych o wysokiej częstotliwości obserwowania w warunkach braku pełnej informacji. *Ekonometria*, 46(4): 72-84. DOI: 10.15611/ekt.2014.4.07.
12. Szmuksta-Zawadzka, M., Zawadzki, J. (2015). Wykorzystanie danych oczyszczonych o wysokiej częstotliwości w prognozowaniu zmiennych ze złożoną sezonowością. *Metody Ilościowe w Badaniach Ekonomicznych*, 16(4): 147-159. Pobrane z: <https://qme.sggw.edu.pl/article/view/3817/3444> (17.09.2022).

13. Szmuksta-Zawadzka, M., Zawadzki, J. (2016). Prognozowanie na podstawie szeregów czasowych o wysokiej częstotliwości oczyszczonych z sezonowości dla luk niesystematycznych. *Metody Ilościowe w Badaniach Ekonomicznych*, 17(1): 121-136. Pobrane z: <https://qme.sggw.edu.pl/article/view/3851/3478> (17.09.2022)
14. Taylor, J.W, Menezes, L.M., McSharry, P. E. (2006). A comparison of univariate methods for forecasting electricity demand up to a day ahead. *International Journal of Forecasting*, 22: 1-16. DOI: <https://doi.org/10.1016/j.ijforecast.2005.06.006>
15. Taylor, J.W. (2010). Triple Seasonal Methods for Short-term Load Forecasting. *European Journal of Operational Research*, 204: 139-152. DOI: <https://doi.org/10.1016/j.ejor.2009.10.003>.
16. Tomaszewski, M. (2005). *Model przedsiębiorstwa dystrybucyjnego działającego na otwartym rynku energii elektrycznej* [Rozprawa doktorska, Politechnika Opolska]. s. 122, Pobrane z: <https://silo.tips/download/politechnika-opolska-wydzia-elektrotechniki-i-automatyki> (21.01.2023).
17. Witkowska, D., Górecka, A., Szadkowska, D., Szymczak, Z. (2000). The forecasting of the demand for electric energy: comparative analysis. *Dynamic Econometric Models*, 4: 45–59.
18. Zawadzki, J. (2018). Modele hybrydowe w prognozowaniu brakujących danych w szeregach o bardzo wysokiej częstotliwości obserwowania. *Folia Pomeranae Universitatis Technologiae Stetinensis seria OECONOMICA*, 346(92): 81–96.
19. Zawadzki, J. (2020). Prognozowanie brakujących danych w szeregach czasowych przy zastosowaniu modeli hybrydowych – podejście teoretyczne i empiryczne. *Wiadomości Statystyczne. The Polish Statistician*, 65(10): 24-48. Pobrane z: http://cejsh.icm.edu.pl/cejsh/element/bwmeta1.element.ojs-doi-10_5604_01_3001_0014_4315/c/1-a7f95192-0b38-4cb9-b923-953a0d1507ef.pdf.pdf (17.09.2022).



This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International (CC BY 4.0) License (<https://creativecommons.org/licenses/by/4.0/deed.pl>) allowing third parties to copy and redistribute the material in any medium or format and remix, transform, and build upon the material for any purpose, even commercially.