

Wiesława MALSKA¹
Henryk WACHTA²

WYKORZYSTANIE MODELU ARIMA DO ANALIZY SZEREGU CZASOWEGO

W artykule zaprezentowano zastosowanie metody ARIMA służącej do analizy szeregu czasowego z trendem i sezonowością. Szereg czasowy jest jednym z rodzajów szeregów statystycznych, który można zdefiniować jako ciąg obserwacji pewnego zjawiska w kolejnych jednostkach czasu (latach, kwartałach, miesiącach, itp.). Analiza szeregów czasowych opiera się na głównym założeniu, że kolejne wartości rozważanej cechy (zmiennej) reprezentują kolejne pomiary wykonane w takiej samej jednostce czasu (w równych odstępach czasu). Zmienną niezależną jest czas (jednostka czasu). Obserwując różne zjawiska (w tym także związane z gospodarką elektroenergetyczną) często chcemy wiedzieć czy i jak zmieniają się w czasie, czyli jaka jest ich dynamika. Analiza szeregów czasowych stosowana jest głównie do podejmowania decyzji związanych z przyszłością. Rozważane zjawisko może podlegać pewnym prawidłowościom, których wykrycie i opis jest głównym celem analizy szeregów czasowych. W wielu przypadkach modele szeregów czasowych wykorzystywane są w celu wnioskowania o przyszłości badanego zjawiska (do prognozowania). Prognozowanie w ujęciu statystycznym to wnioskowanie o przyszłych wartościach szeregu czasowego, które oparte jest na danych czasowych lub analizie wartości, jakie przyjmują rozważane cechy statystyczne (zmienne). Przy analizie w dziedzinie czasu w szeregu czasowym można wyodrębnić pewne składowe (stały przeciętny poziom zjawiska, trend, cykle długookresowe, wahania sezonowe, wahania krótkookresowe, interwencje, składnik losowy (zakłócenie losowe)), przy czym nie wszystkie one muszą występować w konkretnym analizowanym szeregu. Metoda prognozowania zależy od składowych szeregu czasowego. Wyniki obliczeń z wykorzystaniem modelu ARIMA zaprezentowano korzystając z pakietu STATISTICA v. 10.0.

Słowa kluczowe: szereg czasowy, metody analizy szeregów czasowych, prognozowanie, model ARIMA.

¹ Autor do korespondencji: Wiesława Malska, Politechnika Rzeszowska, Katedra Energoelektroniki i Elektroenergetyki, ul. W. Pola 2, 35-959 Rzeszów, 17 865 1974, wmalska@prz.edu.pl

² Henryk Wachta, Politechnika Rzeszowska, Katedra Energoelektroniki i Elektroenergetyki, ul. W. Pola 2, 35-959 Rzeszów, 17 865 1977, hwachta@prz.edu.pl

1. Wstęp

Różnorodne aspekty zjawisk gospodarczych, które związane są z zapotrzebowaniem, produkcją i zużyciem energii elektrycznej wymagają różnorodnych informacji o charakterze tego procesu. Informacji dotyczących przeszłości tego procesu dostarczają nam instytucje (Główny Urząd Statystyczny, Eurostat, itp.), a informacji dotyczących przyszłości, czyli prognoz elektroenergetycznych wymaga się w procesach planowania rozwoju i eksploatacji całego systemu elektroenergetycznego. Znaczenie i ważność prognozowania elektroenergetycznego znalazło formalne uzasadnienie w ustawie Prawo energetyczne i w rozporządzeniach Ministra Gospodarki. Analizę różnego typu zjawisk masowych, wykorzystując metody statystyczne przeprowadza się na podstawie analizy szeregów czasowych. W szeregach czasowych zmienną niezależną jest czas, a zmienną zależną są wartości liczbowe badanego zjawiska. Analiza szeregów czasowych należy do działu statystyki, która wykorzystywana jest w praktyce gospodarczej. Podstawowym celem analizy jest uzyskanie możliwości prognozowania przebiegu zjawisk, a także uzyskanie odpowiedzi na pytanie, jakie mechanizmy powodowały taki, a nie inny przebieg rozważanego zjawiska. Celem jest także możliwość symulacji w oparciu o model zawierający zmienne, których zmiana daje możliwość obserwacji efektu. Szeregi czasowe są realizacjami pewnych procesów stochastycznych. Szereg czasowy jest próbą z procesu stochastycznego. Na podstawie analizy szeregu czasowego chcemy poznać właściwości mechanizmu, który go wygenerował, czyli procesu stochastycznego. Wybór odpowiedniej metody analizy szeregu czasowego często jest zdeterminowany jednostką czasu, wg której „mierzone” określone zjawisko.

Do typowych elementów składowych procesu generującego szereg należy trend, wahania okresowe, wahania losowe oraz interwencje w proces (nagłe zmiany spowodowane czynnikami zewnętrznymi), itp. W większości analiz statystycznych przyjmuje się, że wahania losowe są generowane przez rozkład normalny o wartości oczekiwanej równej zero i stałej wariancji. Prognozowanie to wnioskowanie o przyszłym przebiegu zdarzeń oparte na danych czasowych lub też na analizie wartości przyjmowanych przez zmienne losowe. Wybór metody prognozowania zależy od elementów składowych szeregu czasowego [1, 3, 4]. Jeżeli szereg czasowy zawiera tylko składową systematyczną w postaci stałego poziomu i wahania losowe (zakłócenia przypadkowe) to wówczas oblicza się współczynnik zmienności. Jeżeli wartość tego współczynnika nie przekracza 0,05 korzysta się w takim przypadku z wygładzania średnimi ruchomymi lub wykorzystuje się tzw. metodę naiwną. W przypadku występowania jedynie trendu do identyfikacji struktury szeregu czasowego w statystyce stosuje się różne metody wygładzania i dopasowania odpowiedniej krzywej. Do opracowania prognozy w takim przypadku wykorzystuje się model Holta i model Wintersa [1]. Szeregi czasowe z trendem można także aproksymować za pomocą funkcji liniowej. Do analizy szeregów czasowych z wahaniami sezonowymi stosuje się

metodę wskaźników, analizę harmoniczną Fouriera i funkcje autokorelacji. Jeżeli w szeregu czasowym występują wszystkie składniki systematyczne analiza polega na wyodrębnianiu ich z szeregu czasowego, czyli tzw. dekompozycji szeregu czasowego [1, 3, 5, 11]. W artykule wykorzystano model ARIMA dostępny w pakiecie STATISTICA 10.0.

2. Model ARIMA

Model ARIMA należy do najbardziej efektywnych metod prognozowania szeregów czasowych. Zapoczątkowany został przez Boxa i Jenkinsa [11,12].

W modelu tym wyróżnia się trzy parametry: parametr autoregresyjny (p), rząd różnicowania (d), parametr średniej ruchomej (q) [3, 12].

Model ARIMA (p, q, d), opisywany jest również za pomocą cyfr np. (0, 1, 1), gdzie 0 oznacza zerową ilość parametrów autoregresyjnych, 1-jednokrotne różnicowanie dla 1-jednego parametru średniej ruchomej [11].

Metoda ARIMA zawiera poszczególne etapy wykonania analizy: identyfikację, estymację i diagnozę [3]. Pierwszą fazą analizy jest identyfikacja. Na tym etapie należy zidentyfikować liczbę oraz typ parametrów modelu ARIMA, które następnie zostają oszacowane. W celu dokonania identyfikacji, należy posłużyć się wykresami szeregów oraz autokorelacją. Głównym warunkiem procesu identyfikacji dla metody ARIMA jest stacjonarność wejściowego szeregu. Powinien mieć stałą w czasie średnią, wariancję oraz autokorelację. Za stacjonarność szeregu odpowiada parametr (d). Określając liczbę parametrów modelu zwracamy szczególną uwagę na korelogramy autokorelacji (ACF) oraz autokorelacji cząstkowej (PACF) [1, 3, 12]. Model ARIMA może być stosowany wówczas gdy szereg jest stacjonarny i wejściowy zbiór danych zawiera minimum 50 obserwacji [1, 13]. Aby osiągnąć stacjonarność szeregu, należy poddać go różnicowaniu, a następnie przeanalizować wykres danych oraz korelogram.

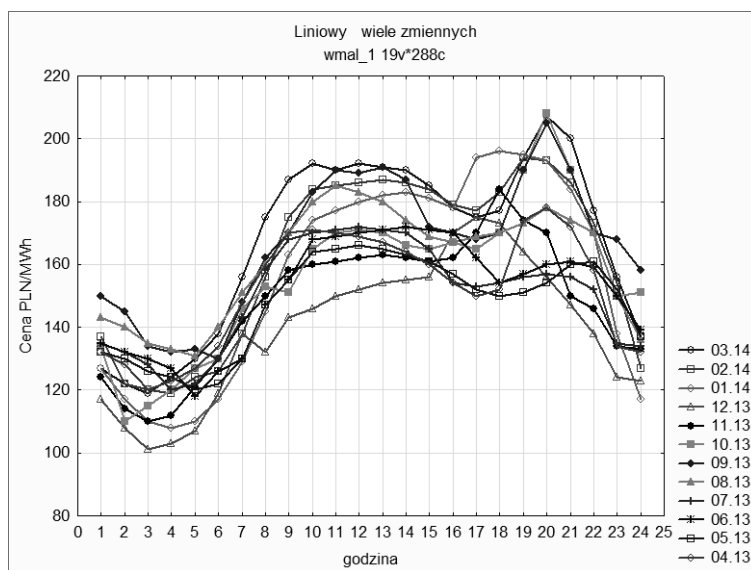
Przy doborze parametrów modelu sezonowego ARIMA, posługujemy się analizą ACF oraz PACF jak w modelu prostym ARIMA [12]. W szeregu sezonowym autokorelogram (ACF) i autokorelogram cząstkowy (PACF) wykazuje duże wartości dla wielokrotności opóźnienia sezonowego.

Kolejną fazą modelu ARIMA jest estymacja parametrów. Jest to zadanie trudne. Podczas estymacji parametrów stosuje się metodę najmniejszych kwadratów. Polega ona na znalezieniu najmniejszej sumy kwadratów reszt wartości parametrów modelu. Niejednokrotnie trzeba tutaj posłużyć się metodą iteracyjną. W przypadku dobrania nieodpowiedniego modelu, podczas procesu estymacji może dojść do niezgodności parametrów. Wówczas program przypisze sumie kwadratów bardzo dużą wartość. Wartość taką nazywamy wartością kary [3, 12].

Etap diagnozy dokonuje oceny wybranego modelu. Na podstawie znanych części danych ocenia się trafność wygenerowanej prognozy. Podczas diagnozy

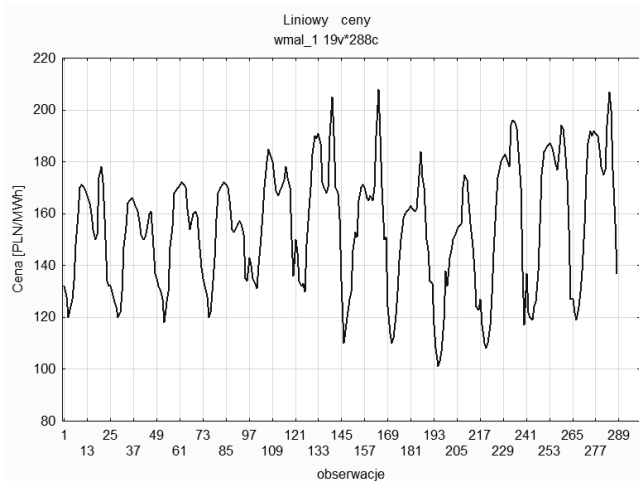
możemy sprawdzić również reszty przy pomocy (ACF) i (PACF). Przy dobrze dobranym modelu, funkcje nie powinny przyjmować wartości różnych od zera. O niepoprawności modelu ARIMA mówi się, gdy reszty wykazują nieprawidłowość lub autokorelację [3]. Metoda ARIMA jest wykorzystywana do analizy szeregów czasowych z trendem i sezonowością. Do analizy wybrano dane dostępne na stronie Towarowej Giełdy Energii [6, 13] odnoszące się do miesięcznych średnioważonych cen dla 24 godzin doby w okresie od 1 kwietnia 2013 roku do 31 marca 2014 roku (jednostka PLN/MWh). Szereg czasowy przedstawiono na rys. 1. dla miesięcznych średnioważonych cen odniesionych do każdej godziny doby. Szereg czasowy jest podzielony na miesiące ($k = 12$) i w obrębie każdego miesiąca na 24 godziny ($d = 24$). Wykorzystując dostępne w pakiecie STATISTICA wykresy można przedstawić rozważany szereg czasowy dla poszczególnych miesięcy analizowanego roku, a także można przedstawić szereg czasowy dla analizowanego okresu. Na rys. 1 przedstawiono szereg czasowy dla miesięcznych średnioważonych cen energii elektrycznej dla 24 godzin doby w okresie od kwietnia 2013 do marca 2014 roku, na którym widać zależność wartości cen w odniesieniu do godziny doby (czyli tzw. sezonowość).

Na rys. 2 zaprezentowano szereg empiryczny dla okresu od kwietnia 2013 do marca 2014 roku dla danych zaczerpniętych z raportów miesięcznych Towarowej Giełdy Energii S.A.



Rys. 1. Szereg czasowy dla miesięcznych średnioważonych cen energii elektrycznej dla 24 godzin doby w okresie od kwietnia 2013 do marca 2014 roku

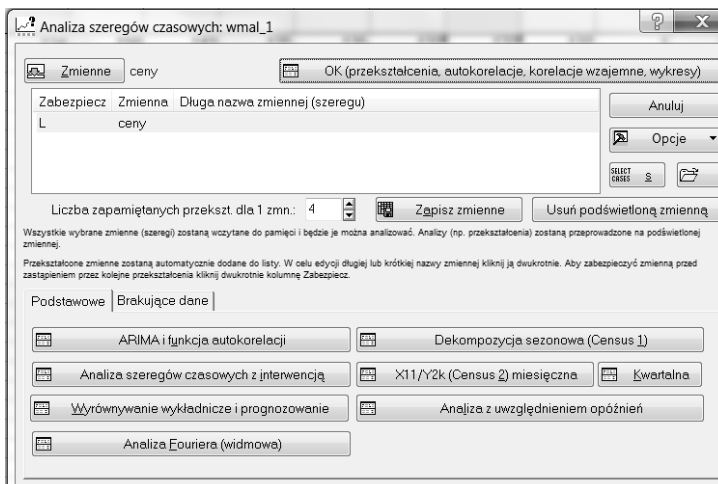
Fig. 1. The time series representing weight-averaged monthly the prices of electricity for 24 hours of the day in the period from April 2013 to March 2014



Rys. 2. Szereg empiryczny dla okresu od kwietnia 2013 do marca 2014 roku, na podstawie [6]

Fig. 2. The empirical series for the period from April 2013 to March 2014, based on [6]

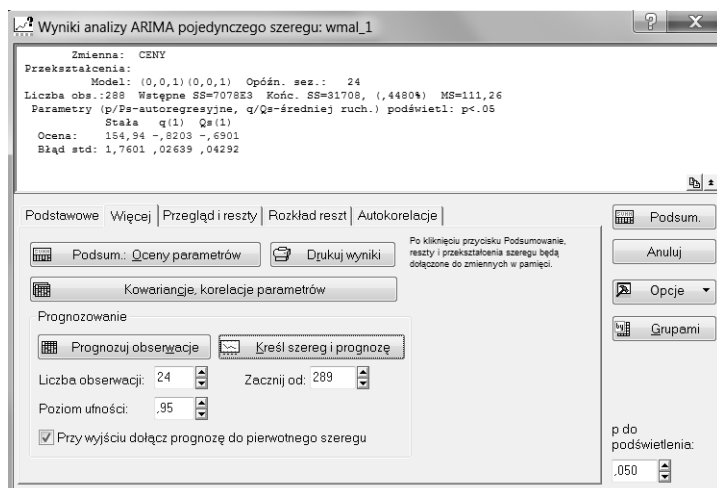
W programie STATISTICA są dostępne narzędzia do dekompozycji szeregów czasowych z wahaniami sezonowymi (rys 3.).



Rys. 3. Widok karty Statystyka/Zaawansowane modele liniowe i nieliniowe/Szeregi czasowe i prognozowanie z dostępnymi narzędziami do dekompozycji szeregów czasowych

Fig. 3. View the card Statistics/ Advanced Linear Model / Time series and forecasting tools available to the decomposition of the time series

Rysunek 4 przedstawia okno wynikowe analizy ARIMA wraz z dostępnymi kartami do dalszej analizy i obliczeń.



Rys. 4. Okno wynikowe analizy ARIMA

Fig. 4. Result of analysis in ARIMA

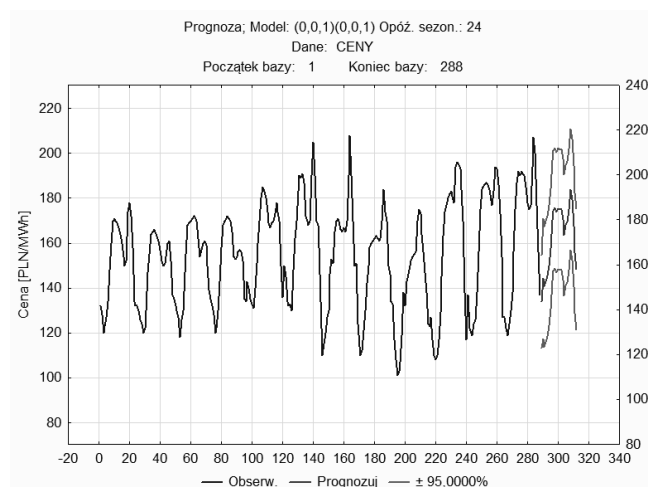
Prognoza; Model: (0,0,1)(0,0,1) Opóź. sezon.: Dane: CENY Początek bazy: 1 Koniec bazy: 288				
Nr obs.	Prognozu	Dolne 95,0000%	Górne 95,0000%	Błąd std
289	133,9747	113,2132	154,7363	10,54784
290	144,1890	117,3361	171,0419	13,64253
291	140,5626	113,7097	167,4155	13,64253
292	142,8261	115,9732	169,6790	13,64253
293	145,3426	118,4897	172,1955	13,64253
294	149,9281	123,0752	176,7810	13,64253
295	157,5308	130,6779	184,3837	13,64253
296	170,0284	143,1755	196,8813	13,64253
297	174,4093	147,5564	201,2622	13,64253
298	175,6571	148,8042	202,5100	13,64253
299	173,8001	146,9473	200,6530	13,64253
300	175,3528	148,4999	202,2056	13,64253
301	174,9044	148,0515	201,7573	13,64253
302	175,0195	148,1666	201,8724	13,64253
303	170,5884	143,7355	197,4413	13,64253
304	163,5094	136,6565	190,3623	13,64253
305	167,8539	141,0010	194,7068	13,64253
306	169,4102	142,5573	196,2631	13,64253
307	174,9850	148,1321	201,8379	13,64253
308	183,8460	156,9931	210,6989	13,64253
309	179,4457	152,5928	206,2986	13,64253
310	167,5414	140,6885	194,3943	13,64253
311	155,6631	128,8102	182,5160	13,64253
312	148,4491	121,5962	175,3020	13,64253

Rys. 5. Prognoza modelu ARIMA dla średnioważonych cen energii dla 24 godzin doby dla miesiąca kwietnia 2014 roku

Fig. 5. Forecast of ARIMA model for the average of the prices of energy for 24 hours a day for the month of April 2014

Na rysunku 5 przedstawiony jest arkusz wynikowy prognozy w miesiącu kwietniu 2014r. przy miesięcznych średnioważonych cenach energii elektrycznej dla 24 godzin doby. Natomiast na rysunku 6 przedstawiono przebieg szeregu

empirycznego wraz prognozą i 95% przedziałem ufności dla miesięcznych średnioważonych cen dla 24 godzin doby.



Rys. 6. Przebieg szeregu empirycznego wraz prognozą (dla miesiąca kwietnia 2014r.) i 95% przedziałem ufności dla miesięcznych średnioważonych cen dla 24 godzin doby

Fig. 6. The course of a number of empirical and forecast (for the month of April 2014) and 95% confidence intervals for the monthly average of the prices for 24 hours a day

Pakiet STATISTICA umożliwia analizę szeregów czasowych, a także wykonanie prognozy na kolejny okres czasu [1, 3, 5].

3. Podsumowanie

Zaprezentowany w artykule przykład zastosowania metody ARIMA jest przykładem praktycznego wykorzystania analizy szeregu czasowego do prognozowania przyszłych wartości zmiennej objaśnianej (zależnej od czasu). Konkretna wartość liczbową, będąca wynikiem procesu prognozowania (predykcji) jest prognozą. Natomiast okres, na który prognozujemy tzw. horyzont prognozy zależy jest od typu i charakteru zmiennej objaśnianej. Prognozowanie jest praktycznym wykorzystaniem metod statystycznych i modeli, które dostępne są w narzędziach informatycznych wspomagających często procesy gospodarcze, w tym także w dziedzinie gospodarki elektrotechnicznej.

Literatura

- [1] Rabiej M., Statystyka z programem Statistica, *Helion* 2012
- [2] Nowak E., Prognozowanie gospodarcze, *Agencja Wyd. PLACET*, Warszawa 1998
- [3] Kot S., Jakubowski J., Sokołowski A., Statystyka, *Wyd. Difin*, Warszawa 2011

- [4] Sokołowski A., Analiza szeregów czasowych i prognozowanie, *Statistica w badaniach naukowych i nauczaniu statystyki*, Statsoft, Kraków 2010
- [5] <http://www.statsoft.pl>
- [6] <http://tge.pl/pl/155/raporty-miesieczne>, Urząd Regulacji Energetyki, Raporty Roczne Prezesa URE, dostęp kwiecień 2015r.
- [7] <http://www.ure.gov.pl/>
- [8] Nowak E., Prognozowanie gospodarcze. Metody, modele, zastosowania, przykłady. Agencja wydawnicza Placet, Warszawa 1998
- [9] Cieślak M., *Prognozowanie gospodarcze, metody i zastosowania*, Wydawnictwo Naukowe PWN, Warszawa 2005
- [10] Snarska A., *Statystyka, ekonometria, prognozowanie*, Wydawnictwo Placet, Warszawa 2005
- [11] Sobczyk M., *Statystyka*, Wydawnictwo Naukowe PWN, Warszawa 2002
- [12] Box E. P., Jenkins G. M., *Szeregi czasowe – analiza i prognozowanie*, PWN, Warszawa 1983
- [13] Malska W.: Zastosowanie metody wskaźników do analizy szeregów czasowych, „POSTĘPY w ELEKTROTECHNICIE STOSOWANEJ”, Kościelisko, 16 -20 czerwca 2014r.

ARIMA MODEL USING THE TIME SERIES ANALYSIS

Summary

The paper presents the application of the method used for the analysis of ARIMA time series with trend and seasonality. Time series is one of the types of statistical series, which can be defined as a series of observations of a phenomenon in the following units of time (years, quarters, months, etc.). Time series analysis based on the main idea that a further consideration of the characteristics (variable) represent the more measurements made in the same unit of time (at regular intervals). The independent variable is the time (unit of time). Observing different phenomena (including related to the economy electricity) often want to know whether and how they are changing over time, that is what is their dynamics. Time series analysis is mainly used to make decisions about the future. Considered phenomenon may be subject to certain regularities, which detect and description is the main objective of the analysis of time series. In many cases, time series models are used to apply for the future of the studied phenomenon (to predict). Forecasting is statistically inference about future values of the time series, which is based on the analysis of data or time values which take under consideration the statistical characteristics (variables). At the time domain analysis in time series can extract some components (constant average level of the phenomenon, a trend long-term cycles, seasonal fluctuations, fluctuations in short-term, interventions, random component (random disturbance)), and not all of them must be analyzed in a specific number of. Forecasting method depends on the components of the time series. The results of calculations using the ARIMA model is presented using STATISTICA v. 10.0.

Keywords: time series, time series analysis methods, forecasting, ARIMA model.

DOI: 10.7862/re.2015.27

Tekst złożono w redakcji: maj 2015

Przyjęto do druku: wrzesień 2015