

PROGNOZOWANIE NA PODSTAWIE MODELI ARIMA

redaktor naukowy Joanicjusz Nazarko

Redaktor naukowy
Joanicjusz Nazarko

PROGNOZOWANIE W ZARZĄDZANIU PRZEDSIĘBIORSTWEM

Cz. V
Prognozowanie
na podstawie modeli ARIMA

Autorzy:
Joanicjusz Nazarko
Ewa Chodakowska



OFICyna WYDAWNICZA POLITECHNIKI BIAŁOSTOCKIEJ
BIAŁYSTOK 2022

Recenzenci:
prof. dr hab. Dorota Witkowska
dr hab. Stanisław Łobejko, prof. SGH

Redakcja i korekta językowa:
Agnieszka Polecka

Skład, grafika:
Oficyna Wydawnicza Politechniki Białostockiej

© Copyright by Politechnika Białostocka, Białystok 2022

ISBN 978-83-67185-13-4
ISBN 978-83-67185-14-1 (eBook)
DOI: 10.24427/978-83-67185-14-1



Publikacja jest udostępniona na licencji
Creative Commons Uznanie autorstwa-Użycie niekomercyjne-
-Bez utworów zależnych 4.0
(CC BY-NC-ND 4.0).

Pełną treść licencji udostępniono na stronie
creativecommons.org/licenses/by-nc-nd/4.0/legalcode.pl.
Publikacja jest dostępna w Internecie na stronie Oficyny Wydawniczej PB.

Druk: PPH Remigraf sp. z o.o.

Oficyna Wydawnicza Politechniki Białostockiej
ul. Wiejska 45C, 15-351 Białystok
e-mail: oficyna.wydawnicza@pb.edu.pl
www.pb.edu.pl

SPIS TREŚCI

WPROWADZENIE.....	4
1. KORELACJE W SZEREGACH CZASOWYCH.....	7
2. STACJONARNE I NIESTACJONARNE SZEREGI CZASOWE.....	27
3. MODELE ARIMA SZEREGÓW CZASOWYCH	56
3.1. Rodzaje modeli klasy ARIMA.....	56
3.2. Budowa modeli ARIMA	68
3.3. Estymacja parametrów modeli ARIMA	83
4. Prognozowanie oparte na modelach ARIMA.....	96
4.1. Wprowadzenie do modelu prognozy.....	96
4.2. Przykłady budowy prognoz	101
4.2.1. Niesezonowy model ARIMA(0,2,1)	101
4.2.2. Sezonowy model ARIMA(0,1,0)(0,1,1) ₄	126
4.2.3. Sezonowy model ARIMA(0,1,1)(2,1,0) ₁₂	138
5. Przykłady zastosowań	152
5.1. Prognozowanie przepustowości portu	152
5.2. Prognozowanie liczby zachorowań na COVID-19.....	154
5.3. Prognozowanie parytetu dwóch walut	157
5.4. Prognozowanie cen owoców i warzyw	160
5.5. Prognozowanie nasłonecznienia	163
6. Problemy do samodzielnego rozwiązania.....	168
Literatura.....	171
Portale informacyjne, strony internetowe, dokumenty elektroniczne.....	177
Spis tabel.....	178
Spis rysunków	179

WPROWADZENIE

Tematem części piątej podręcznika *Prognosowanie w zarządzaniu przedsiębiorstwem* jest prognozowanie szeregów czasowych z zastosowaniem modeli klasy ARIMA.

Modele klasy ARIMA (ang. *Autoregressive Integrated Moving Average*), czyli zintegrowane (scałkowane) modele autoregresji i średniej ruchomej są jedną z najbardziej znanych i powszechnie stosowanych metod analizy i prognozowania szeregów czasowych o różnej strukturze. Ze względu na przełomowy wkład, który w ich rozwój wnieśli G.E.P. Box i G.M. Jenkins, w literaturze nazywane są także modelami Boxa-Jenkinsa.

Modele ARIMA stanowią uogólnienie i rozwinięcie, opisanych w części trzeciej podręcznika, modeli adaptacyjnych. Znajdują one zastosowanie do prognozowania szeregów czasowych zawierających oprócz wahań przypadkowych zarówno tendencję rozwojową, jak i wahania periodyczne.

Istota modeli ARIMA opiera się na zjawisku autokorelacji, czyli zależności bieżącej wartości zmiennej od jej wartości przeszłych. Wśród klasy modeli ARIMA wyróżnić można: modele autoregresji (AR), modele średniej ruchomej (MA), modele mieszane (ARMA) oraz modele zintegrowane (scałkowane) ARIMA, w których pierwotny szereg czasowy poddawany jest różnicowaniu pierwszego lub wyższego rzędu (wyznaczenie pierwszych lub dalszych różnic między wartościami szeregu) w celu uzyskania jego stacjonarności ze względu na średnią. Kolejną grupę w klasie modeli ARIMA stanowią modele sezonowe, często zapisywane jako SARIMA (*Seasonal* ARIMA). Odnoszą się one do zjawisk opisywanych szeregami czasowymi zawierającymi składowe sezonowe i dodatkowe człony odpowiadające tym składowym.

Modele klasy ARIMA stosowane są szeroko w modelowaniu i prognozowaniu szeregów czasowych opisujących różnorodne zjawiska w wielu obszarach, zarówno gospodarczych, społecznych, jak i technicznych.

W podręczniku wyróżniono 6 rozdziałów. Rozdział pierwszy poświęcono omówieniu zjawiska korelacji w szeregach czasowych. Wprowadzono pojęcia autokorelacji i autokorelacji częściowej oraz wskazano sposoby wyznaczania wartości tych funkcji. W rozdziale drugim podjęto problem stacjonarności szeregów czasowych. Podano niezbędne definicje oraz przedyskutowano sposoby eliminowania niestacjonarności w szeregach czasowych. Wyjaśniono zachowanie się przebiegów funkcji autokorelacji i autokorelacji częściowej dla szeregów stacjonarnych i niestacjonarnych. Omówiono podstawowe testy stacjonarności szeregów czasowych. W rozdziale trzecim przybliżono podstawowe rodzaje modeli ARIMA. Wyjaśniono ich elementy i konstrukcję. Przedstawiono proces budowy modeli i szczegółowo omówiono jego poszczególne etapy. Wskazano na istotną rolę interpretacji przebiegów funkcji autokorelacji i autokorelacji częściowej w identyfikacji rodzaju modelu. Przybliżono algorytmy estymacji parametrów modelu oraz kryteria wyboru najlepszych modeli spośród alternatywnych wariantów. Rozdział czwarty dedykowano wykorzystaniu modeli ARIMA w zadaniach prognozowania szeregów czasowych. Nakreślono procedurę postępowania przy prognozowaniu opartym na modelach klasy ARIMA i dokładnie przeanalizowano jej kolejne kroki. W rozdziale piątym przedstawiono szeroki przegląd zastosowań modeli ARIMA w różnych obszarach, opracowany na podstawie publikacji w międzynarodowych czasopismach naukowych. Rozdział szósty zawiera problemy do samodzielnego rozwiązania. Podręcznik zaopatrzono w wykaz literatury, która uzupełnia i uszczegóławia omawiane w nim zagadnienia.

Rozważania teoretyczne i metodyczne w poszczególnych rozdziałach są bogato ilustrowane szczegółowymi przykładami opracowanymi przy wykorzystaniu pakietu statystycznego Statistica. Zamierzeniem autorów było, aby podręcznik był przystępny i zrozumiały zarówno dla studentów, jak i dla szerszego grona czytelników, dlatego też, starając się zachować niezbędną ścisłość wywodów, unikali dyskusji niuansów teoretycznych.

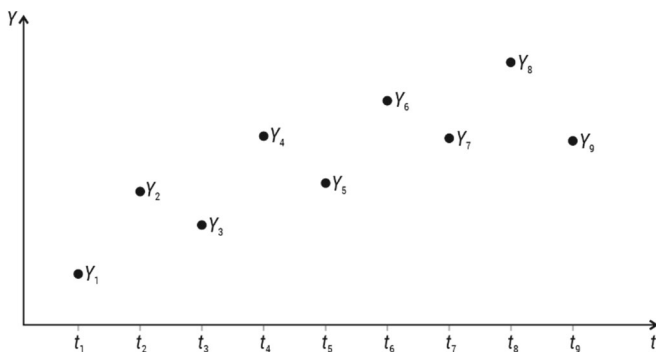
Oczywiście zakres materiału przedstawionego w podręczniku nie wyczerpuje obszernej problematyki prognozowania z zastosowaniem modeli ARIMA. Autorzy, dokonując wyboru materiału

i jego zakresu merytorycznego, kierowali się swoją wiedzą oraz doświadczeniem badawczym i dydaktycznym z zakresu modelowania i prognozowania szeregów czasowych za pomocą modeli klasy ARIMA. Sugerowana przez autorów na końcu każdego rozdziału literatura podstawowa i pomocnicza może pomóc zainteresowanym czytelnikom w samodzielnym pogłębieniu i poszerzeniu swojej wiedzy na temat omawianych zagadnień.

Autorzy podręcznika wyrażają nadzieję, że Czytelnicy znajdą w nim wartościowe źródło wiedzy i inspiracji do własnych działań prognostycznych z wykorzystaniem modeli klasy ARIMA. Będą jednocześnie wdzięczni za wszelkie sugestie i opinie dotyczące możliwego udoskonalenia następnych wydań i kolejnych części podręcznika.

1. KORELACJE W SZEREGACH CZASOWYCH

Szereg czasowy jest to uporządkowana seria obserwacji pokazująca kształtowanie się poziomu badanego zjawiska w kolejnych momentach lub okresach, np. sekundach, godzinach, dniach, tygodniach, miesiącach, latach, itp. (rys. 1.1)¹. Szereg czasowy momentów przedstawia zjawisko w punktach pomiarowych (momentach), np. w określonej godzinie danego dnia. Szereg czasowy okresów to dane z pewnych odcinków czasu, np. uśrednione obserwacje z danego dnia². Zbiór stanów jednej zmiennej Y , uporządkowanych według wartości zmiennej czasowej t ($t = 1, \dots, n$), opisujących dane zjawisko zachodzące w obiekcie objętym prognozą, określa się jako jednowymiarowy szereg czasowy³.



Rys. 1.1. Przykład jednowymiarowego szeregu czasowego

Źródło: opracowanie własne.

¹ A. Frenkel, G. Kildiszew (1976), *Analiza szeregów czasowych i prognozowanie*, Państwowe Wydawnictwo Ekonomiczne, Warszawa, s. 13.

² S. Łobjko (red.), K. Masłowska, R. Wojdan (2015), *Analiza i prognozowanie szeregów czasowych z programem SAS*, Szkoła Główna Handlowa w Warszawie – Oficyna Wydawnicza, Warszawa, s. 14, https://ssl-administracja.sgh.waw.pl/pl/OW/publikacje/Documents/Analiza_lobejko_maslowska_wojdan.pdf [27.07.2020].

³ J. Nazarko (red.) (2011), *Prognozowanie w zarządzaniu przedsiębiorstwem. Cz. II. Prognozowanie na podstawie szeregów czasowych*, Wydawnictwo Politechniki Białostockiej, Białystok.

Jednowymiarowy szereg czasowy $\{Y_t: t = 0, \pm 1, \pm 2, \pm 3, \dots\}$ ⁴, jest realizacją procesu stochastycznego – zjawiska statystycznego zmieniającego się w czasie zgodnie z rozkładami prawdopodobieństwa⁵. Oznacza to, że jest zbiorem zmiennych losowych Y_t (ciągłych lub dyskretnych) generowanych przez proces stochastyczny, zależnych od parametru czasu (indeksowanych czasem) $t \in R$ (przyjmującego wartości ze zbioru liczb rzeczywistych)⁶. W tej części podręcznika zawężono rozważania do szeregów czasowych w tradycyjnym ujęciu, tzn. zawierających obserwacje zmieniające się w pewien określony sposób i obarczone zakłóceniami, ale dokonane w kolejnych stałych odstępach: momentach lub okresach⁷.

Proces stochastyczny można scharakteryzować za pomocą kilku podstawowych miar. **Średnia stacjonarnego procesu stochastycznego*** μ_Y określa poziom, wokół którego proces oscyluje, zaś **wariancja** σ_Y^2 mierzy rozrzut wokół tego poziomu⁸:

$$\mu_Y = E[Y_t], \quad (1.1)$$

$$\sigma_Y^2 = E[(Y_t - \mu_Y)^2], \quad (1.2)$$

gdzie: $Y_t, t = 1, 2, \dots$ – obserwacje szeregu czasowego.

Powyższe wielkości estymować można odpowiednio przez średnią szeregu czasowego i wariancję szeregu czasowego⁹:

$$\bar{Y} = \frac{1}{n} \sum_{t=1}^n Y_t, \quad (1.3)$$

$$\hat{\sigma}_Y^2 = \frac{1}{n} \sum_{t=1}^n (Y_t - \bar{Y})^2, \quad (1.4)$$

gdzie: $Y_t, t = 1, 2, \dots, n$ – obserwacje szeregu czasowego.

⁴ J.D. Cryer, K.-S. Chan (2008), *Time Series Analysis. With Applications in R*, Second Edition, Springer, s. 11.

⁵ G.E.P. Box, G.M. Jenkins, G.C. Reinsel, G.M. Ljung (2015), *Time Series Analysis: Forecasting and Control. Fifth Edition*, John Wiley & Sons, s. 6.

⁶ W. Tarczyński, D. Witkowska, K. Kompa (2013), *Współczynnik beta. Teoria i praktyka*, Pielaszek Research, Warszawa, s. 11.

⁷ Ibidem.

* Stacjonarność jest tematem rozdziału drugiego podręcznika.

⁸ G.E.P. Box, G.M. Jenkins (1983), *Analiza szeregów czasowych. Prognozowanie i sterowanie*, PWN, Warszawa, s. 35.

⁹ Ibidem.

W wypadku obliczania wariancji niezbyt licznych zbiorów obserwacji polecane jest wykorzystanie tzw. nieobciążonego estymatora wariancji¹⁰:

$$\hat{\sigma}_Y^2 = \frac{1}{n-1} \sum_{t=1}^n (Y_t - \bar{Y})^2. \quad (1.5)$$

Odchylenie standardowe to pierwiastek kwadratowy z wariancji¹¹:

$$\sigma_Y = \sqrt{E[(Y_t - \mu_Y)^2]} \quad (1.6)$$

oraz:
$$\hat{\sigma}_Y = \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - \bar{Y})^2}, \quad (1.7)$$

lub:
$$\hat{\sigma}_Y = \sqrt{\frac{1}{n-1} \sum_{t=1}^n (Y_t - \bar{Y})^2}. \quad (1.8)$$

Funkcją autokorelacji (ang. *Autocorrelation Function* – ACF) nazywa się wartości współczynnika autokorelacji wyznaczone dla kolejnych odstępów. Funkcja autokorelacji pokazuje jak zmienia się korelacja pomiędzy dwiema dowolnymi wartościami szeregu wraz ze zmianą ich odstęp¹². Można ją traktować jako funkcję korelacji liniowej Pearsona^{*}, ale liczoną nie dla dwóch odrębnych szeregów czasowych, a pomiędzy badanym szeregiem czasowym a szeregiem czasowym utworzonym przez jego przesunięcie o wartość opóźnienia $k = |t - s|$ dla $t, s = 0, \pm 1, \pm 2, \pm 3, \dots$ (rys. 1.2). Autokorelacja występuje w szeregu czasowym wówczas, gdy bieżąca wartość zmiennej losowej zależy od przeszłych wartości tej zmiennej. Występowanie autokorelacji w szeregu czasowym pozwala na wnioskowanie o przyszłych wartościach zmiennej losowej na podstawie jej przeszłych wartości.

¹⁰ R.A. Johnson, D.A. Wichern (2007), *Applied Multivariate Statistical Analysis. Sixth Edition*, Pearson Education, Inc., s. 7.

¹¹ D.J. Denis (2015), *Applied Univariate, Bivariate, and Multivariate Statistics*, John Wiley & Sons, Inc., s. 109.

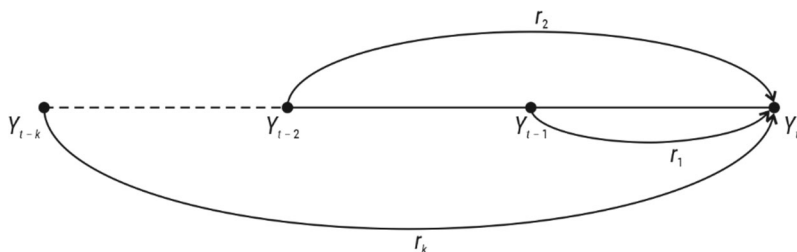
¹² G.E.P. Box, G.M. Jenkins (1983), *Analiza szeregów...*, op. cit., s. 39.

^{*} Analiza współzależności zjawisk i obserwacji została omówiona w części II podręcznika: J. Nazarko (red.) (2011), *Prognozowanie w zarządzaniu przedsiębiorstwem. Cz. II. Prognozowanie na podstawie szeregów czasowych*, Wydawnictwo Politechniki Białostockiej, Białystok.

Autokorelacja przy odstępzie k jest równa¹³:

$$\rho_{t,s} = \rho_k = \frac{E[(Y_t - \mu_Y)(Y_{t+k} - \mu_Y)]}{\sqrt{E[(Y_t - \mu_Y)^2]E[(Y_{t+k} - \mu_Y)^2]}} = \frac{E[(Y_t - \mu_Y)(Y_{t+k} - \mu_Y)]}{\sigma_Y^2}. \quad (1.9)$$

Jest to konsekwencja faktu, że w wypadku stacjonarnego procesu czasowego* (czyli takiego, który może być podstawą formułowania prognoz modelami ARIMA) wariancja jest taka sama w momencie k oraz $t + k$.



Rys. 1.2. Ilustracja zjawiska autokorelacji w szeregu czasowym

Źródło: opracowanie własne.

Ponieważ w praktyce dysponuje się skończonym szeregiem czasowym złożonym z n obserwacji, można otrzymać jedynie estymatory funkcji autokorelacji. Zadowalającym estymatorem autokorelacji jest¹⁴:

$$r_{t,s} = r_k = \frac{\sum_{t=1}^{n-k} (Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2}. \quad (1.10)$$

Współczynnik autokorelacji jest wartością unormowaną autokowariancji – γ_k i przyjmuje wartości z zakresu: $-1 \leq \rho_k \leq 1$ ($-1 \leq r_k \leq 1$).

Autokowariancja przy odstępzie k jest kowariancją pomiędzy Y_t i Y_{t+k} :

$$\gamma_{t,s} = \gamma_k = E[(Y_t - \mu_Y)(Y_{t+k} - \mu_Y)]. \quad (1.11)$$

¹³ G.E.P. Box, G.M. Jenkins (1983), *Analiza szeregów...*, op. cit., s. 36.

* Stacjonarność jest tematem rozdziału drugiego podręcznika.

¹⁴ S. Makridakis, S.C. Wheelwright, R.J. Hyndman (1998), *Forecasting. Methods and Applications*, John Wiley & Sons, Inc. s. 313.

Jej estymator można zapisać jako:

$$c_{t,s} = c_k = \frac{1}{n} \sum_{t=1}^{n-k} (Y_t - \bar{Y})(Y_{t+k} - \bar{Y}). \quad (1.12)$$

Zależności pomiędzy autokorelacją a autokowariancją wyrażają się wzorami:

$$\rho_k = \frac{\gamma_k}{\gamma_0}, \quad (1.13)$$

$$r_k = \frac{c_k}{c_0}. \quad (1.14)$$

Autokorelacja i autokowariancja mają następujące właściwości¹⁵:

$$\begin{aligned} c_{t,t} &= \hat{\sigma}_Y^2, \quad r_{t,t} = 1, \\ c_{t,s} &= c_{s,t}, \quad r_{t,s} = r_{s,t}, \\ |c_{t,s}| &\leq \sqrt{c_{t,t}c_{s,s}}, \quad |r_{t,s}| \leq 1. \end{aligned} \quad (1.15)$$

Wartości $r_{t,s}$ bliskie ± 1 wskazują na silną liniową zależność, natomiast wartości bliskie 0 na słabą lub brak liniowej zależności. Uzyskanie użytecznej oceny funkcji autokorelacji wymaga co najmniej $n = 50$ obserwacji, a estymowane autokorelacje powinny być obliczane dla $k = 0, 1, \dots, K$, gdzie K jest w przybliżeniu nie większe niż $n/4$ ¹⁶.

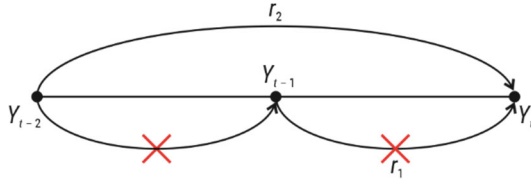
Funkcja autokorelacji cząstkowej (ang. *Partial Autocorrelation Function* – PACF) mierzy zależność pomiędzy Y_t a Y_{t-k} , gdy wpływ pozostałych 1, 2, 3, ..., $k-1$ obserwacji zostanie usunięty. Inaczej mówiąc, autokorelacja cząstkowa to korelacja szeregu z nim samym, przesuniętym o określone opóźnienie k obserwacji przy wyeliminowaniu wpływu korelacji dla wszystkich przesunięć od 1 do $k-1$ ¹⁷. Dla rzędu opóźnienia równego 1 funkcje ACF i PACF mają równe wartości¹⁸.

¹⁵ J.D. Cryer, K.-S. Chan (2008), *Time Series Analysis...*, op. cit., s. 12.

¹⁶ G.E.P. Box, G.M. Jenkins (1983), *Analiza szeregów...*, op. cit., s. 41.

¹⁷ S. Makridakis, S.C. Wheelwright, R.J. Hyndman (1998), *Forecasting. Methods and...*, op. cit., s. 320.

¹⁸ N. Iwaszczuk, P. Drygás, P. Pusz, R. Pusz (2013), *Prognozowanie gospodarcze*, Uniwersytet Rzeszowski, Rzeszów, s. 133.



Rys. 1.3. Ilustracja zjawiska autokorelacji cząstkowej w szeregu czasowym

Źródło: opracowanie własne.

Wyjaśnienie funkcji autokorelacji cząstkowej polega na tym, że w wypadku istnienia znaczącej korelacji pomiędzy Y_t a Y_{t-1} , występuje również znacząca korelacja pomiędzy Y_{t-1} a Y_{t-2} , gdyż odstęp pomiędzy Y_{t-1} i Y_{t-2} wynosi także 1. Wobec tego, pomiędzy Y_t a Y_{t-2} istnieje zależność wynikająca z faktu, iż obie są powiązane z Y_{t-1} . W celu pomiaru rzeczywistej wartości korelacji pomiędzy Y_t i Y_{t-2} należy wyeliminować efekt wpływu Y_{t-1} (rys. 1.3)¹⁹.

Funkcję autokorelacji cząstkowej można estymować, dopasowując kolejno procesy autoregresji (procesy, w których wartość obserwacji w danym okresie jest sumą liniowej kombinacji wartości poprzednich i nieskorelowanego szumu, tj. $Y_t = \sum_{j=1}^p \phi_j Y_{t-j} + e_t$) rzędu 1, 2, 3, ... metodą najmniejszych kwadratów i obliczając na każdym kroku estymatory próbkowe $\hat{\phi}_{11}$, $\hat{\phi}_{22}$, $\hat{\phi}_{33}$, ... ostatniego dopasowanego współczynnika²⁰.

Jeżeli wartości parametrów nie są zbyt bliskie granicom niestacjonarności, można skorzystać z przybliżonych estymatorów Yule'a-Walkera, co sprowadza się do rozwiązywania równania (1.16) dla $k = 1, 2, \dots$ ²¹:

$$r_j = \hat{\phi}_{k1}r_{j-1} + \hat{\phi}_{k2}r_{j-2} + \dots + \hat{\phi}_{k(k-1)}r_{j-k+1} + \hat{\phi}_{kk}r_{j-k},$$

$$j = 1, 2, \dots, k. \quad (1.16)$$

¹⁹ S. Makridakis, S.C. Wheelwright, R.J. Hyndman (1998), *Forecasting. Methods and...*, op. cit., s. 320.

²⁰ G.E.P. Box, G.M. Jenkins (1983), *Analiza szeregów...*, op. cit., s. 72.

²¹ Ibidem.

Metoda rekurencyjna takiego rozwiązania, pochodząca od Durбина, wykorzystuje równania²²:

$$\hat{\phi}_{kk} = \begin{cases} r_1, & k = 1 \\ \frac{r_k - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j} r_j}, & k = 2, 3, \dots, \end{cases} \quad (1.17)$$

$$\hat{\phi}_{kj} = \hat{\phi}_{k-1,j} - \hat{\phi}_{kk} \hat{\phi}_{k-1,k-j}, \quad k = 2, \dots, \quad j = 1, 2, \dots, k-1.$$

Oszacowania takie są jednak bardzo czułe na błędy zaokrąglenia i nie powinny być stosowane w przypadkach, gdy wartości parametrów są bliskie granicom niestacjonarności²³. Szczególną uwagę należy zwrócić na istnienie obserwacji odstających (nietypowych) w szeregu. Pojedyncza wartość odstająca może przybliżyć wartość autokorelacji przy opóźnieniu k do zera, podczas gdy dla $k+1$ kilka wartości odstających może sprawić, że będzie ona bliska jedności²⁴.

Mimo możliwości samodzielnego oszacowania wartości funkcji PACF, w praktyce, ze względu na skomplikowanie formuł, do obliczania wartości autokorelacji cząstkowej wykorzystuje się dostępne oprogramowanie komputerowe. Wyniki prezentowane w tej części podręcznika zostały uzyskane przy wykorzystaniu pakietu Statistica.

Szczególnymi szeregami czasowymi są szeregi bazujące na **procesie białego szumu**. Biały szum (ang. *White Noise*) to ciąg niezależnych zmiennych losowych $\{e_t\}$ o jednakowym rozkładzie ze skończonymi wartościami przeciętnymi i wariancjami²⁵. Jeżeli jest generowany zgodnie z rozkładem normalnym z wartością przeciętną równą 0 i daną wariancją σ_e^2 , tj. $e_t \sim N(0, \sigma_e^2)$ nazywany jest gaussowskim białym szumem (ang. *Gaussian White*

²² Ibidem, s. 492.

²³ Ibidem, s. 72.

²⁴ A. Dürre, R. Fried, T. Liboschik (2015), *Robust estimation of (partial) autocorrelation*, „WIREs Computational Statistics” 7(3), s. 205–222, doi: 10.1002/wics.1351.

²⁵ A. Sokołowski (2003), *Prognozowanie finansowych szeregów czasowych*, Statsoft Polska, https://media.statsoft.pl/_old_dnn/downloads/sokolowski.pdf [27.07.2020].

Noise). Proces białego szumu ma wartości autokorelacji równe 0 dla każdego opóźnienia różnego od 0²⁶:

$$r_k = \begin{cases} 1, & \text{dla } k = 0 \\ 0, & \text{dla } k \neq 0 \end{cases} \quad (1.18)$$

oraz autokowariancję wynoszącą²⁷:

$$\gamma_k = \begin{cases} \sigma_e^2, & \text{dla } k = 0 \\ 0, & \text{dla } k \neq 0. \end{cases} \quad (1.19)$$

Bazujący na procesie białego szumu dwupunktowy szereg średniej ruchomej ma następującą konstrukcję²⁸:

$$Y_t = (e_t + e_{t-1})/2. \quad (1.20)$$

Średnia takiego procesu jest równa 0:

$$\mu_Y = E[Y_t] = E\left[\frac{e_t + e_{t-1}}{2}\right] = \frac{E[e_t] + E[e_{t-1}]}{2} = 0. \quad (1.21)$$

Natomiast wariancja, przy założeniu, że zmienne losowe e_t i e_{t-1} są niezależne:

$$\begin{aligned} \sigma_Y^2 &= E[(Y_t - \mu_Y)^2] = E\left[\left(\frac{e_t + e_{t-1}}{2} - 0\right)^2\right] = \\ &= \frac{E[(e_t)^2] + 2E[e_t]E[e_{t-1}] + E[(e_{t-1})^2]}{4} = \frac{\sigma_e^2 + \sigma_e^2}{4} = 0,5\sigma_e^2. \end{aligned} \quad (1.22)$$

Autokowariancja $\gamma_{t,t-1} = 0,25\sigma_e^2$ oraz $\gamma_{t,t-2} = 0$, co można przedstawić wzorem²⁹:

$$\gamma_k = \begin{cases} 0,5\sigma_e^2, & \text{dla } k = 0 \\ 0,25\sigma_e^2, & \text{dla } |k| = 1 \\ 0, & \text{dla } |k| \geq 2. \end{cases} \quad (1.23)$$

²⁶ J.D. Cryer, K.-S. Chan (2008), *Time Series Analysis...*, op. cit., s. 17.

²⁷ R.H. Shumway, D.S. Stoffer (2017), *Time Series and Its Applications with R Examples*, Springer International Publishing AG, s. 21.

²⁸ J.D. Cryer, K.-S. Chan (2008), *Time Series Analysis...*, op. cit., s. 14.

²⁹ Ibidem, s. 15.

Autokorelacja zaś przyjmuje wartości:

$$r_k = \begin{cases} 1, & \text{dla } k = 0 \\ 0,5, & \text{dla } |k| = 1 \\ 0, & \text{dla } |k| \geq 2. \end{cases} \quad (1.24)$$

Podobnie można rozważyć trzypunktową średnią ruchomą³⁰:

$$Y_t = (e_{t-1} + e_t + e_{t+1})/3. \quad (1.25)$$

W tym wypadku wartość przeciętna także wynosi 0:

$$\mu_Y = E[Y_t] = 0. \quad (1.26)$$

Autokowariancja natomiast³¹:

$$\gamma_{t,s} = \frac{1}{9} E[(e_{s-1} + e_s + e_{s+1})(e_{t-1} + e_t + e_{t+1})], \quad (1.27)$$

$$\gamma_{t,s} = \begin{cases} 3/9, & |k| = 0 \\ 2/9, & |k| = 1 \\ 1/9, & |k| = 2 \\ 0, & |k| \geq 3. \end{cases} \quad (1.28)$$

Innym przykładem szczególnego szeregu czasowego jest proces **błądzenia przypadkowego** (ang. *Random Walk*)³²:

$$\begin{aligned} Y_1 &= e_1, \\ Y_2 &= e_1 + e_2, \end{aligned} \quad (1.29)$$

...

$$Y_t = e_1 + e_2 + \dots + e_t,$$

czyli:

$$Y_t = Y_{t-1} + e_t, \text{ oraz } Y_1 = e_1. \quad (1.30)$$

³⁰ R.H. Shumway, D.S. Stoffer (2017), *Time Series Analysis...*, op. cit., s. 17.

³¹ Ibidem, s. 18.

³² J.D. Cryer, K.-S. Chan (2008), *Time Series Analysis...*, op. cit., s. 12.

Średnia procesu błędzenia przypadkowego jest równa 0:

$$\begin{aligned}\mu_Y &= E[Y_t] = E[e_1 + e_2 + \dots + e_t] = \\ &= E[e_1] + E[e_2] + \dots + E[e_t] = 0.\end{aligned}\quad (1.31)$$

Natomiast wariancja rośnie liniowo wraz z czasem³³:

$$\begin{aligned}\sigma_Y^2 &= E[(e_1 + e_2 + \dots + e_t)^2] = \sigma_{e_1}^2 + \sigma_{e_2}^2 + \dots + \sigma_{e_t}^2 = \\ &= \sigma_e^2 + \sigma_e^2 + \dots + \sigma_e^2 = t\sigma_e^2.\end{aligned}\quad (1.32)$$

Kowariancja dla wszystkich wartości t i s takich, że $1 \leq t \leq s$ przyjmuje wartości:

$$\gamma_{t,s} = t\sigma_e^2. \quad (1.33)$$

Funkcję autokorelacji zgodnie z definicją (1.9) można zapisać³⁴:

$$r_{t,s} = \frac{\gamma_{t,s}}{\sqrt{\gamma_{t,t}\gamma_{s,s}}} = \sqrt{\frac{t}{s}}, \text{ dla } 1 \leq t \leq s. \quad (1.34)$$

Oznacza to, że np.:

$$r_{1,2} = \sqrt{\frac{1}{2}} = 0,707, r_{1,3} = \sqrt{\frac{1}{3}} = 0,577, r_{2,3} = \sqrt{\frac{2}{3}} = 0,816.$$

Oczywiście wartości uzyskane z próby mogą różnić się nieznacznie od przedstawionych teoretycznych wyliczeń. Warto podkreślić, że proces błędzenia przypadkowego jest procesem autoregresyjnym pierwszego rzędu, w którym wartość w czasie t zależy wyłącznie od wartości w chwili poprzedniej i szumu.

W wypadku rzeczywistych szeregów czasowych nie są znane faktyczne wartości autokorelacji, a estymowane wartości obarczone są błędem, dlatego istotna jest możliwość oceny, kiedy wartości autokorelacji i autokorelacji cząstkowych są praktycznie równe zero dla odstępów większych od konkretnych wartości p czy q (p i q są rzędami wybranego modelu ARIMA)³⁵. Dla dużych

³³ P.J. Brockwell, R.A. Davis (2002), *Introduction to Time Series and Forecasting*, Second Edition, Springer, s. 17.

³⁴ J.D. Cryer, K.-S. Chan (2008), *Time Series Analysis...*, op. cit., s. 13.

³⁵ G.E.P. Box, G.M. Jenkins (1983), *Analiza szeregów...*, op. cit., s. 178.

odstępów można obliczyć błędy standardowe estymowanych autokorelacji³⁶:

$$\begin{aligned}\hat{\sigma}[r_k] &\cong \frac{1}{n^{1/2}} \{1 + 2(r_1^2 + r_2^2 + \dots + r_q^2)\}^{1/2} = \\ &= \frac{1}{n^{1/2}} (1 + 2 \sum_{i=1}^{k-1} r_i^2)^{1/2}, \quad k > q.\end{aligned}\quad (1.35)$$

Przy założeniu, że proces jest procesem białego szumu wynoszącą one³⁷:

$$\hat{\sigma}[r_k] \cong \left[\frac{1}{n} \left(\frac{n-k}{n+2} \right) \right]^{1/2}. \quad (1.36)$$

Dla autokorelacji cząstkowych, zakładając że proces jest procesem autoregresji rzędu p , błąd standardowy rzędu większego lub równego $p+1$ wynosi³⁸:

$$\sigma[\hat{\phi}_{kk}] \cong \frac{1}{n^{1/2}}, \quad k > p. \quad (1.37)$$

Nawet dla dość dużych n rozkład estymowanego współczynnika autokorelacji i autokorelacji cząstkowej o wartości teoretycznej 0 jest w przybliżeniu rozkładem Gaussa. Zaś dla rozkładu normalnego, odchylenia przekraczające jeden błąd standardowy w dowolnym kierunku mają prawdopodobieństwo około 1/3, a odchylenia przekraczające dwa błędy standardowe około 1/20. Powyższe fakty można wykorzystać do nieformalnej oceny czy autokorelacje i autokorelacje cząstkowe są istotnie różne od zera przy większych wartościach odstępów³⁹.

³⁶ Ibidem, s. 179.

³⁷ *IBM SPSS Statistics 24 Algorithms* (1989, 2016), IBM Corporation, s. 5, <https://www.ibm.com/support/pages/node/616931#en> [11.08.2020].

³⁸ G.E.P. Box, G.M. Jenkins (1983), *Analiza szeregów...*, op. cit., s. 179.

³⁹ Ibidem.

Alternatywnym podejściem do sprawdzenia istotności funkcji autokorelacji może być wykorzystanie statystyki Ljunga-Boxa, która dla opóźnienia k jest definiowana następująco⁴⁰:

$$Q_k = n(n+2) \sum_{i=1}^k \frac{r_i^2}{n-i}. \quad (1.38)$$

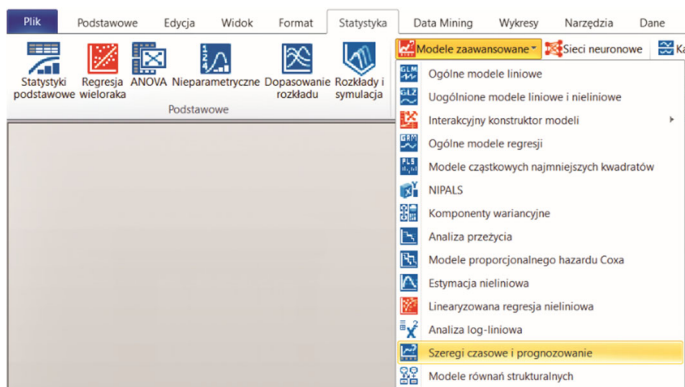
W przypadku, gdy n jest duże, Q_k ma rozkład χ^2 z $k-p-q$ stopniami swobody. Test ten pozwala sprawdzić hipotezę zerową H_0 : czy funkcja autokorelacji danego szeregu nie różni się znacząco od funkcji autokorelacji procesu białego szumu. W przypadku $Q_k \leq \chi_{\alpha}^2$, nie ma podstaw do odrzucenia hipotezy zerowej przy zadanym prawdopodobieństwie α . Inaczej mówiąc, ACF badanego procesu nie jest statystycznie istotna. Jeżeli zaś, $Q_k > \chi_{\alpha}^2$, odrzuca się hipotezę zerową na rzecz hipotezy alternatywnej H_1 : ACF jest znacząco różna od ACF procesu białego szumu⁴¹.

Funkcję autokorelacji i autokorelacji częściowej zilustrowano na przykładzie rzeczywistego szeregu czasowego średniej dobowej temperatury w Białymstoku zanotowanej w kwietniu i maju 2019 r.

Wszystkie obliczenia i wykresy wykonano w programie Statistica 13.3. Wykresy wartości szeregów można otrzymać w programie poprzez wybór na zakładce Wykresy → 2W → Wykresy liniowe (Zmienne).... Wygodną opcją jest skorzystanie już na etapie preliminaryjnej analizy danych od razu z modułu Statystyka → Modele zaawansowane → Szeregi czasowe i prognozowanie → OK (przekształcenia, autokorelacje, korelacje wzajemne, wykresy) (rys. 1.4 i 1.5).

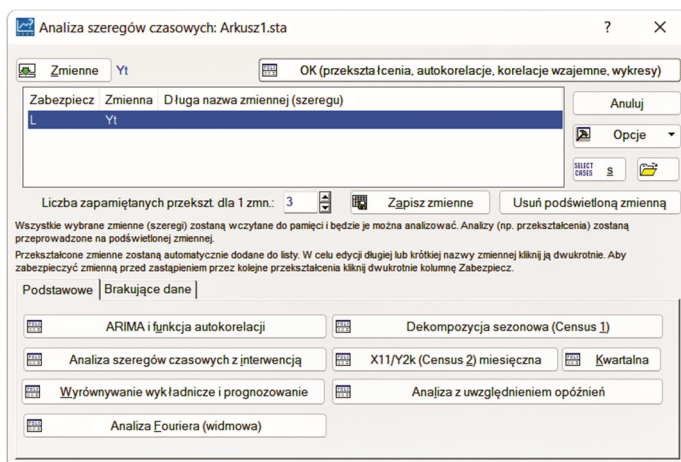
⁴⁰ J. Fan, Q. Yao (2005), *Nonlinear Time Series Nonparametric and Parametric Methods*, Springer, s. 300; D.A. Belsley (red.), E.J. Kontoghiorghes (red.) (2009), *Handbook of Computational Econometrics*, John Wiley & Sons, Ltd, s. 387, <https://thenigerianprofessionalaccountant.files.wordpress.com/2013/04/handbook-of-computational-econometrics-belsley.pdf> [15.12.2020].

⁴¹ S.A. DeLurgio (1998), *Forecasting Principles and Applications*, McGraw-Hill, s. 280.



Rys. 1.4. Pakiet Statistica – menu

Źródło: opracowanie własne.



Rys. 1.5. Pakiet Statistica – okno Szeregi czasowe i prognozowanie

Źródło: opracowanie własne.

Wykres zależności liniowej pomiędzy wartościami szeregu i szeregu opóźnionego to Wykresy rozrzutu dla wybranego opóźnienia. Wartości funkcji autokorelacji i autokorelacji cząstkowej uzyskuje się poprzez wybór na zakładce Statystyka → Modele zaawansowane → Szeregi czasowe i prognozowanie → OK (przekształcenia, autokorelacje, korelacje wzajemne, wykresy) → opcji Autokorelacje

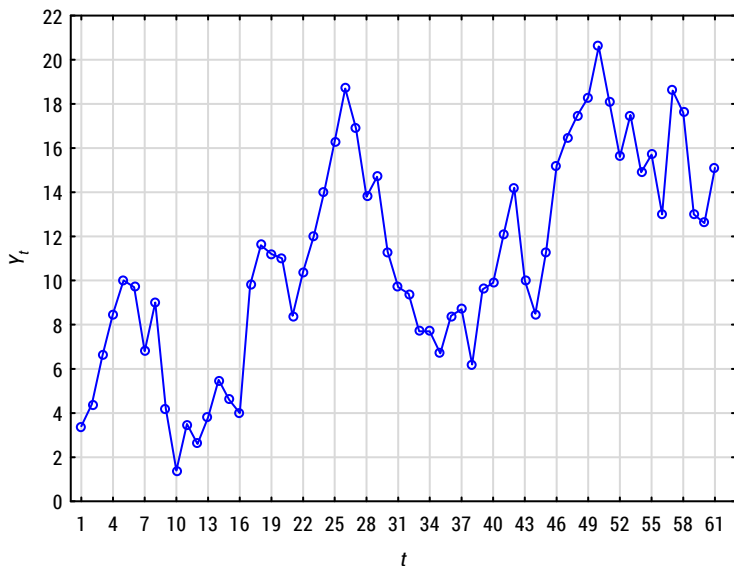
oraz kolejno Autokorelacje cząstkowe lub Modele zaawansowane → Szeregi czasowe i prognozowanie → ARIMA i funkcja autokorelacji → Autokorelacje → oraz Autokorelacje cząstkowe.

W tabeli 1.1 przedstawiono 61 wartości średniej temperatury zanotowanej w Białymstoku w okresie od 1 kwietnia do 31 maja 2019 r. Na rys. 1.6 zilustrowano przebieg szeregu.

Tabela 1.1. Średnia temperatura w Białymstoku w dniach 1.04–31.05.2019 r. w °C

t	data	Y_t	t	data	Y_t	t	data	Y_t	t	data	Y_t
1	01.04.2019	3,4	16	16.04.2019	4,0	31	01.05.2019	9,7	46	16.05.2019	15,2
2	02.04.2019	4,4	17	17.04.2019	9,8	32	02.05.2019	9,4	47	17.05.2019	16,5
3	03.04.2019	6,6	18	18.04.2019	11,6	33	03.05.2019	7,7	48	18.05.2019	17,5
4	04.04.2019	8,5	19	19.04.2019	11,2	34	04.05.2019	7,7	49	19.05.2019	18,3
5	05.04.2019	10,0	20	20.04.2019	11,0	35	05.05.2019	6,7	50	20.05.2019	20,6
6	06.04.2019	9,7	21	21.04.2019	8,4	36	06.05.2019	8,4	51	21.05.2019	18,1
7	07.04.2019	6,8	22	22.04.2019	10,4	37	07.05.2019	8,7	52	22.05.2019	15,6
8	08.04.2019	9,0	23	23.04.2019	12,0	38	08.05.2019	6,2	53	23.05.2019	17,5
9	09.04.2019	4,2	24	24.04.2019	14,0	39	09.05.2019	9,6	54	24.05.2019	14,9
10	10.04.2019	1,4	25	25.04.2019	16,3	40	10.05.2019	9,9	55	25.05.2019	15,7
11	11.04.2019	3,5	26	26.04.2019	18,7	41	11.05.2019	12,1	56	26.05.2019	13,0
12	12.04.2019	2,6	27	27.04.2019	16,9	42	12.05.2019	14,2	57	27.05.2019	18,6
13	13.04.2019	3,8	28	28.04.2019	13,8	43	13.05.2019	10,0	58	28.05.2019	17,6
14	14.04.2019	5,5	29	29.04.2019	14,7	44	14.05.2019	8,5	59	29.05.2019	13,0
15	15.04.2019	4,6	30	30.04.2019	11,3	45	15.05.2019	11,3	60	30.05.2019	12,6
									61	31.05.2019	15,1

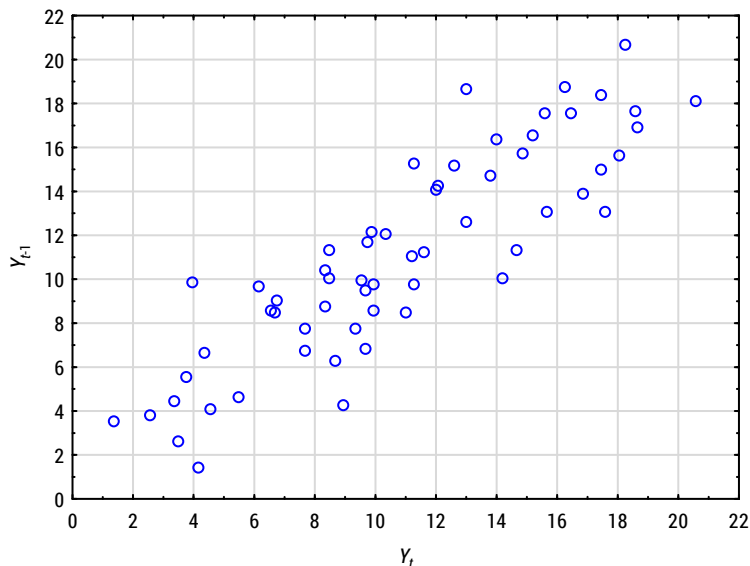
Źródło: Serwis MeteoModel.pl, <https://meteoModel.pl/dane/histeryczne-dane-pomiarowe/> na podstawie danych udostępnionych przez Instytut Meteorologii i Gospodarki Wodnej – Państwowy Instytut Badawczy [8.12.2019].



Rys. 1.6. Szereg średniej temperatury zanotowanej w Białymstoku w dniach 1.04–31.05.2019 r.

Źródło: opracowanie własne danych z serwisu [Meteomodel.pl](https://meteomodel.pl/dane/historyczne-dane-pomiarowe/), <https://meteomodel.pl/dane/historyczne-dane-pomiarowe/> na podstawie danych udostępnionych przez Instytut Meteorologii i Gospodarki Wodnej – Państwowy Instytut Badawczy [8.12.2019].

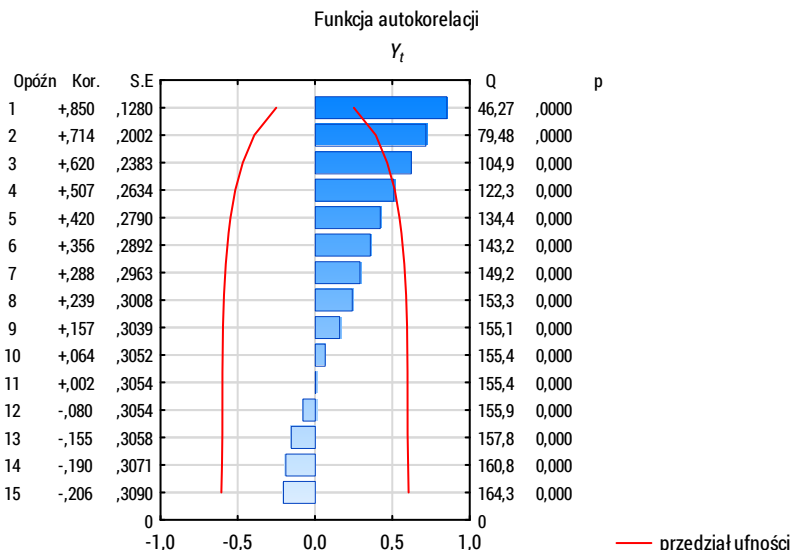
Z wykresu rozrzutu na rys. 1.7 można wnioskować o silnej zależności liniowej pomiędzy wartościami szeregu i szeregu opóźnionego dla $k = 1$.



Rys. 1.7. Zależność autokorelacyjna w szeregu średniej temperatury zanotowanej w Białymstoku w dniach 1.04–31.05.2019 r. dla $k = 1$

Źródło: opracowanie własne danych z serwisu [Meteomodel.pl](https://meteomodel.pl/dane/historyczne-dane-pomiarowe/), <https://meteomodel.pl/dane/historyczne-dane-pomiarowe/> na podstawie danych udostępnionych przez Instytut Meteorologii i Gospodarki Wodnej – Państwowy Instytut Badawczy [8.12.2019].

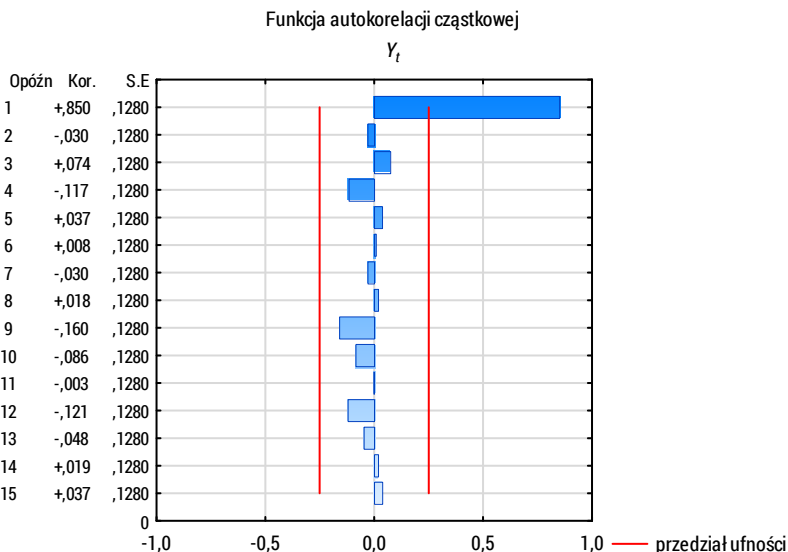
Na rys. 1.8 i 1.9 przedstawiono odpowiednio wartości funkcji autokorelacji i autokorelacji cząstkowej dla opóźnień od $k = 1$ do $k = 15$ wraz z zaznaczonymi przedziałami ufności dla $p = 0,05$.



Rys. 1.8. Wartości autokorelacji szeregu średniej temperatury zanotowanej w Białymstoku w dniach 1.04–31.05.2019 r.

Źródło: opracowanie własne danych z serwisu meteomodel.pl/dane/historyczne-dane-pomiarowe/ na podstawie danych udostępnionych przez Instytut Meteorologii i Gospodarki Wodnej – Państwowy Instytut Badawczy [8.12.2019].

Na podstawie analizy przebiegu funkcji ACF i PACF możliwa jest ocena, na ile dane (obserwacje) z przeszłości wpływają na przyszłe obserwacje. Możliwa jest też ocena wstecznego horyzontu czasowego (liczby przeszłych obserwacji) tego wpływu.



Rys. 1.9. Wartości autokorelacji cząstkowej szeregu średniej temperatury zanotowanej w Białymstoku w dniach 1.04–31.05.2019 r.

Źródło: opracowanie własne danych z serwisu meteomodel.pl/dane/histeryczne-dane-pomiarowe/ na podstawie danych udostępnionych przez Instytut Meteorologii i Gospodarki Wodnej – Państwowy Instytut Badawczy [8.12.2019].

Analizując przedstawione wykresy na rys. 1.8 i 1.9 należy zwrócić uwagę na łagodne zmniejszanie się wartości autokorelacji (trzy znaczące wartości dla opóźnień $k = 1, 2$ i 3) oraz tylko jedną znaczącą wartość funkcji autokorelacji cząstkowej dla opóźnienia $k = 1$. Analiza przebiegów funkcji ACF i PACF jest podstawą doboru do szeregu czasowego właściwego modelu opisującego zachowanie się szeregu. Będzie to tematem rozdziału trzeciego podręcznika.

Kluczowe zagadnienia:

autokorelacje, autokorelacje cząstkowe,
proces błędzenia przypadkowego, proces białego szumu

Literatura podstawowa

- [1] Box G.E.P., Jenkins G.M. (1983), *Analiza szeregów czasowych. Prognozowanie i sterowanie*, PWN, Warszawa
- [2] Box G.E.P., Jenkins G.M., Reinsel G.C., Ljung G.M. (2015), *Time Series Analysis: Forecasting and Control. Fifth Edition*, John Wiley & Sons
- [3] DeLurgio S.A. (1998), *Forecasting Principles and Applications*, McGraw-Hill Educations
- [4] Makridakis S., Wheelwright S.C., Hyndman R.J. (1998), *Forecasting. Methods and Applications*, John Wiley & Sons, Inc.
- [5] Nazarko J. (red.) (2011), *Prognozowanie w zarządzaniu przedsiębiorstwem. Cz. II. Prognozowanie na podstawie szeregów czasowych*, Wydawnictwo Politechniki Białostockiej, Białystok

Literatura uzupełniająca

- [1] Belsley D.A. (red.), Kontoghiorghes E.J. (red.) (2009), *Handbook of Computational Econometrics*, John Wiley & Sons, Ltd, <https://thenigerianprofessionalaccountant.files.wordpress.com/2013/04/handbook-of-computational-econometrics-belsley.pdf>
- [2] Brockwell P.J., Davis R.A. (2002), *Introduction to Time Series and Forecasting*, Second Edition, Springer
- [3] Cryer J.D., Chan K.-S. (2008), *Time Series Analysis. With Applications in R*, Second Edition, Springer
- [4] Denis D.J. (2015), *Applied Univariate, Bivariate, and Multivariate Statistics*, John Wiley & Sons, Inc.
- [5] Dürre A., Fried R., Liboschik T. (2015), *Robust estimation of (partial) autocorrelation*, „WIREs Computational Statistics” 7(3), s. 205–222, doi:10.1002/wics.1351
- [6] Fan J., Yao Q. (2005), *Nonlinear Time Series Nonparametric and Parametric Methods*, Springer
- [7] Frenkel A., Kildiszew G. (1976), *Analiza szeregów czasowych i prognozowanie*, Państwowe Wydawnictwo Ekonomiczne, Warszawa
- [8] *IBM SPSS Statistics 24 Algorithms* (1989,2016), IBM Corporation, <https://www.ibm.com/support/pages/node/616931#en> [11.08.2020]
- [9] Iwaszczuk N., Drygaś P., Pusz P., Pusz R. (2013), *Prognozowanie gospodarcze*, Uniwersytet Rzeszowski, Rzeszów

- [10] Johnson R.A., Wichern D.A. (2007), *Applied Multivariate Statistical Analysis. Sixth Edition*, Pearson Education, Inc.
- [11] Łobejko S. (red.), Masłowska K., Wojdan R. (2015), *Analiza i prognozowanie szeregów czasowych z programem SAS*, Szkoła Główna Handlowa w Warszawie – Oficyna Wydawnicza, Warszawa, https://ssl-administracja.sgh.waw.pl/pl/OW/publikacje/Documents/Analiza_lobejko_maslowska_wojdan.pdf [27.07.2020]
- [12] Shumway R.H., Stoffer D.S. (2017), *Time Series and Its Applications with R Examples*, Springer International Publishing AG
- [13] Sokołowski A. (2003), *Prognozowanie finansowych szeregów czasowych*, Statsoft Polska, https://media.statsoft.pl/_old_dnn/downloads/sokolowski.pdf [27.07.2020]
- [14] Tarczyński W., Witkowska D., Kompa K. (2013), *Współczynnik beta. Teoria i praktyka*, Pielaszek Research, Warszawa

2. STACJONARNE I NIESTACJONARNE SZEREGI CZASOWE

Szereg czasowy, czyli uporządkowana sekwencja obserwacji⁴²: $\{Y_t: t = 0, \pm 1, \dots\}$, jak zdefiniowano w rozdziale pierwszym, odzwierciedla proces stochastyczny zmieniający się w czasie zgodnie z rozkładami prawdopodobieństwa⁴³. Proces stochastyczny może być **ściśle stacjonarny**, jeżeli łączny rozkład prawdopodobieństwa związany z m obserwacjami: $Y_1, Y_2, \dots, Y_m$ poczynionymi w dowolnych momentach czasu $1, 2, 3, \dots, m$ jest taki sam, jak związany z m obserwacjami $Y_{1+k}, Y_{1+k}, \dots, Y_{m+k}$ poczynionymi w momentach czasu $1+k, 2+k, \dots, m+k$ ⁴⁴. Wyrażając definicję równaniem dystrybuanty, dla wszystkich $t = 1, 2, 3, \dots, m$ i dowolnych liczb $c_1, c_2, \dots, c_m$ oraz dowolnych przesunięć $k = 1, 2, 3, \dots$, dla stacjonarnych szeregów czasowych prawdziwa jest zależność⁴⁵:

$$P\{Y_1 \leq c_1, \dots, Y_m \leq c_m\} = P\{Y_{1+k} \leq c_1, \dots, Y_{m+k} \leq c_m\}, \quad (2.1)$$

gdzie: P – prawdopodobieństwo.

Mniej restrykcyjne jest założenie słabej stacjonarności. Proces stochastyczny jest **słabo stacjonarny**, jeżeli jego wartość oczekiwana – μ_Y i wariancja – σ_Y^2 są skończone i stałe, a wartość kowariancji między obserwacjami z dwóch okresów zależy jedynie od odległości (odstępu) między nimi⁴⁶. Jeżeli proces nie jest stacjonarny w podanym sensie nazywany jest niestacjonarnym⁴⁷.

⁴² C.M. Mastrangelo, J.R. Simpson, D.C. Montgomery (2013), *Time Series Analysis*, in: Gass S.I., Fu M.C. (red.) *Encyclopedia of Operations Research and Management Science*, Springer, Boston, MA, doi: 10.1007/978-1-4419-1153-7_1045.

⁴³ G.E.P. Box, G.M. Jenkins (1983), *Analiza szeregów...*, op. cit., s. 32.

⁴⁴ Ibidem, s. 35.

⁴⁵ R.H. Shumway, D.S. Stoffer (2017), *Time Series Analysis...*, op. cit., s. 19.

⁴⁶ S. Łobejko (red.), K. Masłowska, R. Wojdan (2015), *Analiza i prognozowanie...*, op. cit., s. 102.

⁴⁷ M. Gruszczyński, M. Podgórska (red.) (1996), *Ekonometria*, Oficyna Wydawnicza Szkoła Główna Handlowa, Warszawa, s. 182.

W rzeczywistych zastosowaniach proces spełniający warunki słabej stacjonarności traktowany jest jako stacjonarny⁴⁸. Takie założenie przyjęto też w tym podręczniku.

Klasyczne statystyczne metody prognozowania bazują na założeniu, że analizowane szeregi czasowe są stacjonarne lub można je przekształcić do przybliżenia stacjonarnych za pomocą transformacji matematycznych. Tym samym można przyjąć założenie o stałości w czasie właściwości statystycznych, to znaczy średniej, wariancji i autokorelacji, i traktować je jako deskryptory kształtowania się przyszłych wartości szeregu czasowego, bądź przynajmniej przyjąć założenie o stałości tych statystyk w serii między okresami lub między sezonami⁴⁹.

Większość ekonomicznych procesów, szczególnie wyrażających wartości w naturalnych jednostkach miary, nie jest stacjonarna: wykazują trendy, cykle, nie mają stałego przeciętnego poziomu, charakteryzują się różną amplitudą wahań*. Na rys. 2.1–2.6 przedstawiono przykłady różnych szeregów z zaznaczonym ich przeciętnym poziomem – linia ciągła oraz odchyleniem standardowym – linia przerywana.

Na rysunku 2.1 przedstawiono przykład stacjonarnego (o średniej równej 0 i wariancji niezależnej od czasu) szeregu autoregresyjnego:

$$Y_t = \phi Y_{t-1} + e_t, \quad (2.2)$$

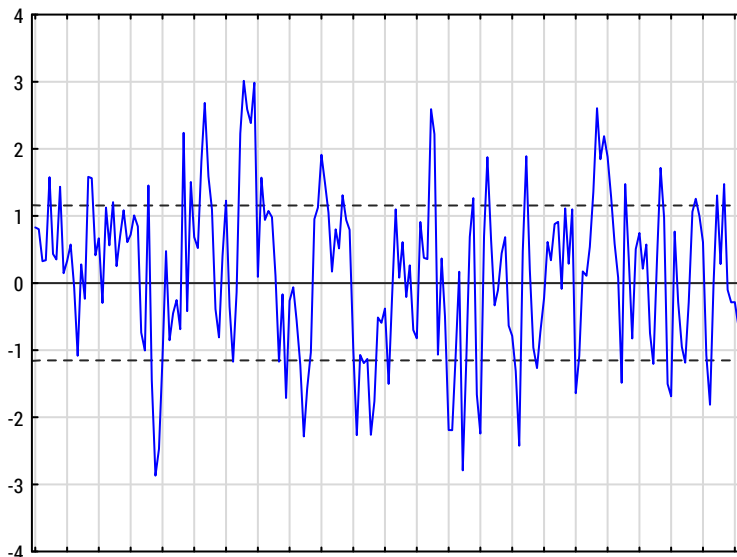
przy czym: $\phi = 0,5$,

$$e_t \sim N(0, \sigma_e^2), \quad \sigma_e^2 = 1.$$

⁴⁸ W. Yang, I. Zurbenko (2010), *Nonstationarity*, „WIREs Computational Statistics” 2(1), s. 107–115, doi: 10.1002/wics.64.

⁴⁹ R. Nau, *Statistical forecasting: notes on regression and time series analysis*, Fuqua School of Business, Duke University, <https://people.duke.edu/~rnau/411home.htm> [7.08.2020].

* Składowe szeregów czasowych zostały omówione w części II podręcznika: J. Nazarko (red.) (2011), *Prognozowanie w zarządzaniu przedsiębiorstwem. Cz. II. Prognozowanie na podstawie szeregów czasowych*, Wydawnictwo Politechniki Białostockiej, Białystok.



Rys. 2.1. Przykład szeregu stacjonarnego ze stałą średnią i wariancją

Źródło: opracowanie własne na podstawie: R. Salles, K. Belloze, F. Porto, P.H. Gonzalez, E. Ogasawara (2019), *Nonstationary time series transformation methods: An experimental review*, „Knowledge-Based Systems” 164, s. 274–291, doi: 10.1016/j.knosys.2018.10.041.

Wartość wariancji można oszacować, korzystając z równania (1.2), tj.:

$$\begin{aligned}\sigma_Y^2 &= E[(Y_t)^2] = E[(\phi Y_{t-1} + e_t)^2] = \\ &= \phi^2 E[(Y_{t-1})^2] + 2\phi E[(Y_{t-1}e_t)] + E[(e_t)^2].\end{aligned}\quad (2.3)$$

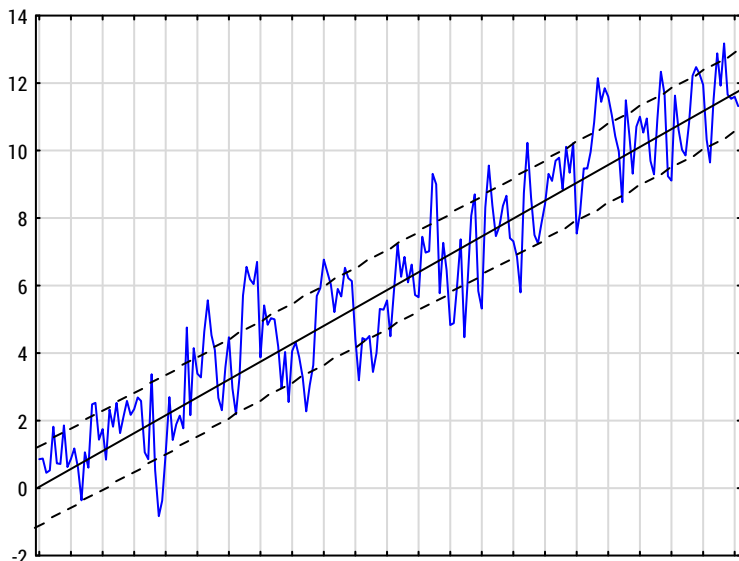
Prowadzi ono do formuły:

$$\sigma_Y^2 = \phi^2 \sigma_Y^2 + \sigma_e^2, \text{ gdy } E[e_t] = 0, E[(e_t)^2] = \sigma_e^2. \quad (2.4)$$

Z tego wynika, że:

$$\sigma_Y^2 = \frac{\sigma_e^2}{1-\phi^2}, \text{ dla } |\phi| < 1. \quad (2.5)$$

W wypadku $\phi = 0,5$ oraz $\sigma_e^2 = 1$, $\sigma_Y^2 \sim 1,33$. Na wykresie szeregu (rys. 2.1) zaobserwować można stały poziom średniej oraz wariancji szeregu w rozpatrywanym czasie.



Rys. 2.2. Przykład szeregu niestacjonarnego ze względu na średnią (z trendem)

Źródło: opracowanie własne na podstawie: R. Salles, K. Belloze, F. Porto, P.H. Gonzalez, E. Ogasawara (2019), *Nonstationary time series transformation methods: An experimental review*, „Knowledge-Based Systems” 164, s. 274–291, doi:10.1016/j.knosys.2018.10.041.

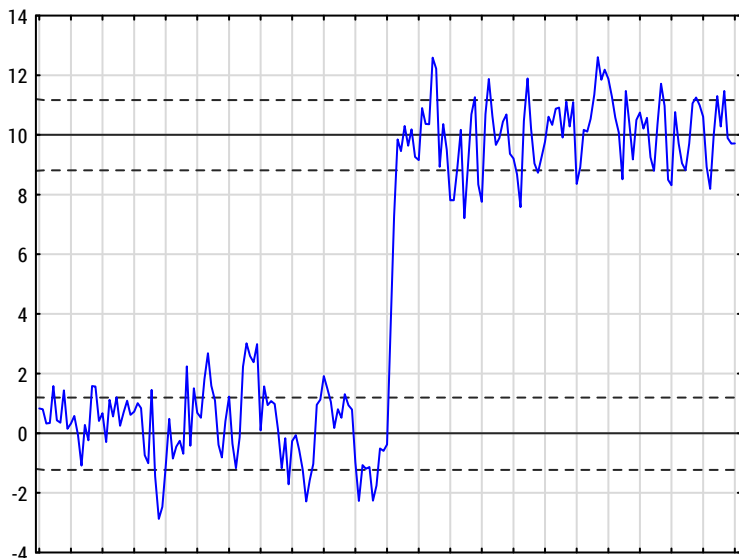
Na rysunku 2.2 przedstawiono przykład szeregu autoregresyjnego z deterministycznym trendem:

$$Y_t = \phi Y_{t-1} + \beta t + e_t, \quad (2.6)$$

przy czym: $\phi = 0,5$,

$$\beta = 0,03, e_t \sim N(0, \sigma_e^2), \sigma_e^2 = 1.$$

Na wykresie szeregu zaobserwować można rosnący poziom średniej oraz stałą wariancję szeregu w rozpatrywanym czasie.



Rys. 2.3. Przykład szeregu niestacjonarnego ze względu na średnią (z różnym poziomem)

Źródło: opracowanie własne na podstawie: R. Salles, K. Belloze, F. Porto, P.H. Gonzalez, E. Ogasawara (2019), *Nonstationary time series transformation methods: An experimental review*, „Knowledge-Based Systems” 164, s. 274–291, doi: 10.1016/j.knosys.2018.10.041.

Szereg na rysunku 2.3 to przykład szeregu autoregresyjnego z różnym poziomem, który można opisać wzorem:

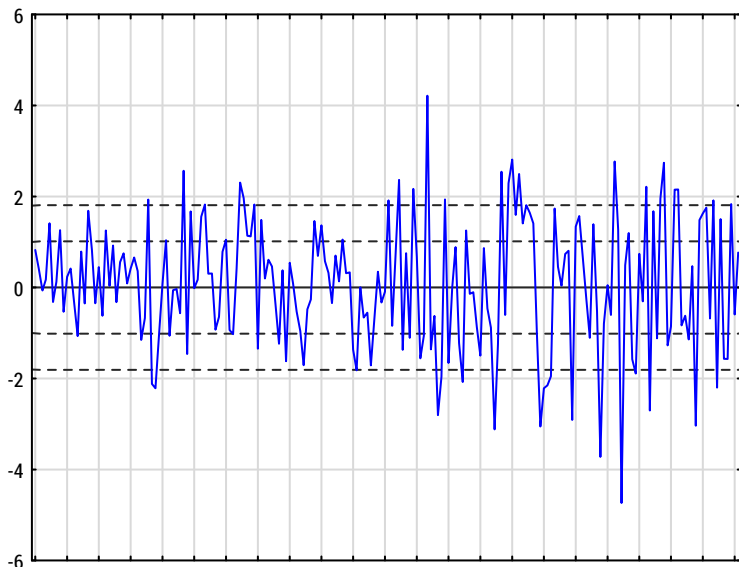
$$Y_t = \begin{cases} \phi Y_{t-1} + e_t, & t = 1, \dots, 100 \\ \alpha + \phi Y_{t-1} + e_t, & t = 101, \dots, 200, \end{cases} \quad (2.7)$$

przy czym: $\alpha = 5, \phi = 0,5, e_t \sim N(0, \sigma_e^2), \sigma_e^2 = 1$.

Średnia tego szeregu wynosi:

$$\mu_Y = \begin{cases} \sim 0, & t = 1, \dots, 100 \\ \alpha/(1 - \phi) \sim 10, & t = 101, \dots, 200. \end{cases} \quad (2.8)$$

Na wykresie szeregu zaobserwować można dwa poziomy średniej oraz stałą wariancję szeregu w rozpatrywanym czasie.



Rys. 2.4. Przykład szeregu niestacjonarnego ze względu na wariancję

Źródło: opracowanie własne na podstawie: R. Salles, K. Belloze, F. Porto, P.H. Gonzalez, E. Ogasawara (2019), *Nonstationary time series transformation methods: An experimental review*, „Knowledge-Based Systems” 164, s. 274–291, doi: 10.1016/j.knosys.2018.10.041.

Szereg przedstawiony na rysunku 2.4 jest realizacją procesu autoregresyjnego z różną wartością wariancji składnika losowego σ_e^2 $e_t \sim N(0, \sigma_e^2)$ wynoszącą odpowiednio 1 lub 4, tj.:

$$Y_t = \begin{cases} \phi Y_{t-1} + e_t, e_t \sim N(0,1), & t = 1, \dots, 100 \\ \phi Y_{t-1} + e_t, e_t \sim N(0,4), & t = 101, \dots, 200, \end{cases} \quad (2.9)$$

przy czym: $\phi = 0,5$.

Na wykresie szeregu zaobserwować można stały poziom średniej oraz dwa poziomy wariacji szeregu w rozpatrywanym czasie.

Przedstawione na rys. 2.1–2.4 wykresy otrzymano za pomocą programu Statistica wybierając dla uprzednio przygotowanych danych na zakładce Wykresy: 2W → Wykresy liniowe (Zmienne) dopasowując regresję liniową oraz rysując odpowiednie linie odniesienia. Wartości szeregu białego szumu e_t wygenerowano za pomocą narzędzia portalu RANDOM.ORG: <https://www.random.org/>.

Niestacjonarność może manifestować się w różny sposób. Trend powodujący zmianę średniej w czasie to chyba najprostszy i najczęściej spotykany rodzaj niestacjonarności. Zmiana poziomu średniej może być konsekwencją zajścia zmiany strukturalnej procesu. Innym rodzajem niestacjonarności jest zmiana wariancji w czasie – zjawisko, które nosi nazwę heteroskedastyczności (ang. *heteroscedasticity*). Należy też pamiętać o szczególnej formie niestacjonarności, która jest obecna w szeregach z wahaniami cyklicznymi lub sezonowymi. W tym wypadku wariancja i średnia mogą zmieniać się okresowo w zależności od t . Każda niestacjonarność może mieć wpływ na budowę modeli prognostycznych. Rozwiązanie problemu niestacjonarności jest koniecznym warunkiem otrzymania wiarygodnych prognoz na podstawie modeli ARIMA⁵⁰.

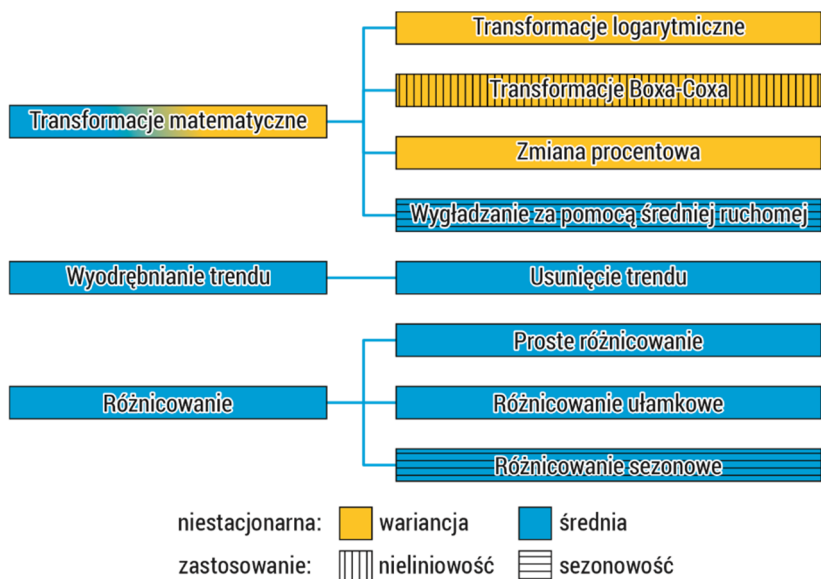
Zostało opracowanych wiele podejść do radzenia sobie z niestacjonarnością danych. Przegląd najczęściej stosowanych parametrycznych transformacji bazujących na odwzorowaniu wartości szeregu przedstawiono na rys. 2.5.

Stacjonarność szeregu ze względu na średnią ocenić można poprzez analizę wartości szeregu, w tym jego wizualizację lub poprzez przeprowadzenie statystycznych testów. Czy wartości szeregu oscylują wokół stałej średniej można sprawdzić za pomocą testów pierwiastków jednostkowych (ang. *unit root tests*), np. testem Dickey’*a*-Fullera⁵¹ lub testów stacjonarności np. testem Kwiatkowskiego, Phillipsa, Schmidta i Shina⁵².

⁵⁰ R. Salles, K. Belloze, F. Porto, P.H. Gonzalez, E. Ogasawara (2019), *Nonstationary time series transformation methods: An experimental review*, „Knowledge-Based Systems” 164, s. 274–291, doi: 10.1016/j.knosys.2018.10.041.

⁵¹ D.A. Dickey, W.A. Fuller (1979), *Distribution of the Estimators for Autoregressive Time Series with a Unit Root*, „Journal of the American Statistical Association” 74(366), s. 427–431, doi: 10.2307/2286348.

⁵² D. Lcc, P. Schmidt (1996), *On the power of the KPSS test of stationarity against fractionally-integrated alternatives*, „Journal of Econometrics” 73, s. 285–302, doi: 10.1016/0304-4076(95)01741-0; D. Kwiatkowski, P.C.B. Phillips, P. Schmidt, Y. Shin (1992), *Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?*, „Journal of Econometrics” 54(1–3), s. 159–178, doi: 10.1016/0304-4076(92)90104-Y.



Rys. 2.5. Metody eliminacji niestacjonarności w szeregach czasowych

Źródło: opracowanie własne na podstawie: R. Salles, K. Belloze, F. Porto, P.H. Gonzalez, E. Ogasawara (2019), *Nonstationary time series transformation methods: An experimental review*, „Knowledge-Based Systems” 164, s. 274–291, doi: 10.1016/j.knosys.2018.10.041.

Przedstawienie statystyk testu Dickey’a-Fullera wymaga wprowadzenia definicji operatora przesunięcia wstecz oraz operatora różnicowania.

Operator przesunięcia wstecz B określa się jako⁵³:

$$BY_t = Y_{t-1}. \quad (2.10)$$

Operator ten ma następujące właściwości⁵⁴:

$$B(BY_t) = BY_{t-1} = Y_{t-2} \quad (2.11)$$

$$B^k Y_t = Y_{t-k}. \quad (2.12)$$

⁵³ B. Pfaff (2008), *Analysis of Integrated and Cointegrated Time Series with R*, Springer-Verlag New York, s. 55.

⁵⁴ P. Strawiński, *Materiały do zajęć: statystyczna analiza danych finansowych*, http://coin.wne.uw.edu.pl/pstrawinski/dane_fin/adf03.pdf [5.09.2021].

Z kolei operator różnicowy zdefiniować można jako:

$$\nabla Y_t = Y_t - Y_{t-1}. \quad (2.13)$$

Zależność między operatorami wyraża się wzorem:

$$\nabla Y_t = (1 - B)Y_t. \quad (2.14)$$

Uogólniając można zapisać:

$$\nabla^d = (1 - B)^d. \quad (2.15)$$

Z powyższego wynika, że np.:

$$(1 - B)Y_t = Y_t - Y_{t-1}, \quad (2.16)$$

$$(1 - B)^2 Y_t = (1 - B)(1 - B)Y_t = Y_t - 2Y_{t-1} + Y_{t-2}. \quad (2.17)$$

Testy pierwiastków jednostkowych zostały nazwane od wartości pierwiastka wielomianu charakterystycznego (tzn. rozwiązania równania) procesu AR, tj. $Y_t = \phi Y_{t-1} + e_t$, który gdy wynosi 1 oznacza proces niestacjonarny⁵⁵. Przypominając informacje z rozdziału pierwszego, model $Y_t = Y_{t-1} + e_t$ to model czystego błędzenia przypadkowego⁵⁶. Przyjmuje się, że jeżeli proces stochastyczny zawiera pierwiastek, który leży wewnątrz bądź na obrzeżu koła jednostkowego, to jest procesem niestacjonarnym. Analogicznie, proces jest stacjonarny, jeżeli wszystkie pierwiastki równania charakterystycznego są większe co do wartości bezwzględnej od 1.

Przedstawioną zasadę można zilustrować na przykładzie procesu autoregresyjnego:

$$Y_t = 0,5Y_{t-1} + e_t, \text{ czyli } Y_t - 0,5Y_{t-1} = e_t. \quad (2.18)$$

Proces ten za pomocą operatora przesunięcia wstecz można zapisać jako:

$$(1 - 0,5B)Y_t = e_t. \quad (2.19)$$

⁵⁵ S. Łobejko (red.), K. Masłowska, R. Wojdan (2015), *Analiza i prognozowanie...*, op. cit., s. 108.

⁵⁶ B. Pfaff (2008), *Analysis of Integrated and...*, op. cit., s. 55.

Jego pierwiastek wynosi 2, co oznacza, że jest to proces stacjonarny. Z kolei proces:

$$Y_t = -1,4Y_{t-1} - 0,4Y_{t-2} + e_t, \quad (2.20)$$

może być zapisany jako:

$$(1 + 1,4B + 0,4B^2)Y_t = e_t. \quad (2.21)$$

Jego równanie charakterystyczne to:

$$(1 + 0,4B)(1 + B)Y_t = e_t. \quad (2.22)$$

Wartości jego pierwiastków, równe odpowiednio -2,5 i -1, biorąc pod uwagę wartości bezwzględne nie są większe od 1, zatem ten proces jest niestacjonarny.

Test Dickey'a-Fullera weryfikuje hipotezę, że w równaniu jest pierwiastek jednostkowy. Hipoteza H_0 ma postać: zmienna Y_t jest niestacjonarna wskutek występowania pierwiastka jednostkowego, zaś H_1 : zmienna Y_t jest stacjonarna. Odpowiadają one sytuacjom, gdy moduł parametru δ (tj. $|\delta|$) jest odpowiednio: równy lub mniejszy od 1 w równaniu:

$$Y_t = \delta Y_{t-1} + e_t. \quad (2.23)$$

Stosując test należy metodą najmniejszych kwadratów oszacować wartość parametru równania, w którym odjęto od obu stron Y_{t-1} , tj.:

$$Y_t - Y_{t-1} = \delta Y_{t-1} - Y_{t-1} + e_t. \quad (2.24)$$

Czyli zapisując równanie z wykorzystaniem operatora różnicowego, następującego równania:

$$\nabla Y_t = (\delta - 1)Y_{t-1} + e_t. \quad (2.25)$$

Warto zwrócić uwagę, że jest to równoznaczne z testowaniem $H_0: \beta = 0$ oraz $H_1: \beta < 0$ dla $\nabla Y_t = \beta Y_{t-1} + e_t$, gdzie $\beta = \delta - 1$ ⁵⁷. Statystyka jest obliczana podobnie jak statystyka t-Studenta: $\hat{\beta}/s_{\hat{\beta}}$, gdzie: $\hat{\beta}$ – ocena parametru $s_{\hat{\beta}}$ – średni błąd szacunku parametru,

⁵⁷ S. Łobejko (red.), K. Masłowska, R. Wojdan (2015), *Analiza i prognozowanie...*, op. cit., s. 109.

lecz stosuje się tablicę wartości krytycznych Dickey'a-Fullera⁵⁸. W wypadku, gdy wartość statystyki testowej jest mniejsza od wartości krytycznej odrzuca się H_0 i przyjmuje H_1 o stacjonarności Y_t . Natomiast, gdy wartość statystyki testowej jest większa od wartości krytycznej, brak jest podstaw do odrzucenia H_0 o niestacjonarności.

Stosunkowo łatwo o istnieniu trendu w szeregu można także wnioskować z kształtu funkcji autokorelacji. Dla szeregu niestacjonarnego funkcja autokorelacji nie będzie szybko zanikać dla średnich i dużych opóźnień k , a przeciwnie, będzie opadać wolno i prawie liniowo. Przy czym nie jest konieczne, aby korelacja przy małych odstępach była bardzo duża⁵⁹.

Jednym ze sposobów usunięcia niestacjonarności ze względu na średnią jest różnicowanie szeregu z opóźnieniem 1 lub wyższym. Sezonowe szeregi czasowe z liczbą obserwacji w jednym cyklu sezonowym wynoszącą s można spróbować sprowadzić do szeregów stacjonarnych, korzystając z różnicowania sezonowego.

Różnicowanie z opóźnieniem 1 prowadzi do szeregu⁶⁰:

$$\hat{Y}_t = \nabla Y_t = Y_t - Y_{t-1}, \quad (2.26)$$

gdzie: ∇ – operator różnicowy.

Wykorzystując operator różnicowy dla opóźnienia rzędu d , tj. ∇^d , otrzymuje się:

$$\hat{Y}_t = \nabla^d(Y_t) = \nabla(\nabla^{d-1}(Y_t)). \quad (2.27)$$

gdzie: d – rząd różnicowania.

⁵⁸ E.M. Syczewska (2010), *Empirical power of the Kwiatkowski-Phillips-Schmidt-Shin test*, Department of Applied Econometrics Working Papers 3, <http://yadda.icm.edu.pl/yadda/element/bwmeta1.element.ekon-element-000171231155>.

⁵⁹ G.E.P. Box, G.M. Jenkins (1983), *Analiza szeregów...*, op. cit., s. 176.

⁶⁰ R.M. Heiberger, P. Teles (2002), *Displays for Direct Comparison of ARIMA Models*, „The American Statistician” 56(2), s. 134, doi: 10.1198/000313002317572808.

W wypadku różnicowania szeregów z wykrytą sezonowością stosuje się operator różnicowania sezonowego ∇_s^D :

$$\nabla_s^D = (1 - B^s)^D, \quad (2.28)$$

gdzie:

- D – rząd różnicowania sezonowego,
- s – liczba sezonów w cyklu sezonowym.

Oznacza to, że np.: $(1 - B^4)Y_t = Y_t - Y_{t-4}$.

Przyjmuje się, że szereg, w którym nie występuje składowa okresowa, osiągnie stacjonarność, jeżeli tylko jego funkcja autokorelacji dostatecznie szybko zanika. Zazwyczaj rząd różnicowania wynosi od 0 do 2 i wystarczy przeanalizować tylko 20 pierwszych autokorelacji oryginalnego szeregu oraz jego pierwszej lub drugiej różnicy⁶¹. Niestacjonarny proces stochastyczny, który można sprowadzić do stacjonarnego w wyniku różnicowania rzędu d nazywany jest procesem zintegrowanym w stopniu d . Podobnie proces sezonowo zintegrowany w stopniu (d, D) to proces, który można sprowadzić do stacjonarnego, różnicując D razy sezonowo, a następnie z rzędem d ⁶².

Odmianą prostego różnicowania jest różnicowanie ułamkowe, które dopuszcza $-0,5 < d < 0,5$ i jest stosowane do szeregów z długą pamięcią⁶³.

Poza najczęściej stosowanym różnicowaniem, w wypadku istnienia w szeregu deterministycznego trendu, którego postać funkcjonalna oraz parametry dają się określić, można zastosować operację usunięcia trendu⁶⁴:

$$\hat{Y}_t = Y_t - \eta_t, \quad (2.29)$$

gdzie: η_t – trend.

⁶¹ G.E.P. Box, G.M. Jenkins (1983), *Analiza szeregów...*, op. cit., s. 176.

⁶² M. Cieślak (red.) (2011), *Prognozowanie gospodarcze. Metody i zastosowania*. Wydanie czwarte zmienione. Wydawnictwo Naukowe PWN, Warszawa, s. 116.

⁶³ A. Maynard, A. Smallwood, M.E. Wohar (2013), *Long Memory Regressors and Predictive Testing: A Two-stage Rebalancing Approach*, „Econometric Reviews” 32(3), s. 318–360, doi: 10.1080/07474938.2012.690663.

⁶⁴ R. Salles, K. Belloze, F. Porto, P.H. Gonzalez, E. Ogasawara (2019), *Nonstationary time series...*, op. cit.

Należy także wspomnieć o metodzie eliminacji niestacjonarności, jaką jest wygładzanie za pomocą modelu średniej ruchomej*, wartej rozważenia w wypadku istnienia sezonowości i długookresowych trendów⁶⁵. Podstawowy model można zapisać jako⁶⁶:

$$\hat{Y}_t = \frac{1}{k} \sum_{i=t}^{t+k-1} Y_i, 1 \leq t \leq n - k + 1, \quad (2.30)$$

gdzie: k – stała wygładzania.

Podobnie można ocenić stacjonarność szeregu ze względu na wariancję na podstawie wizualnej oceny szeregu bądź statystycznych testów istotności sprawdzających homogeniczność wariancji, np. testu jednorodności wariancji Levene'a lub Bartletta (gdy rozkłady zmiennej są normalne). Test Levene'a testuje hipotezę: H_0 : wariancje w grupach są równe w stosunku do hipotezy alternatywnej H_1 : wariancje w grupach istotnie się od siebie różnią. Wykorzystuje on wartości bezwzględnych odchyłeń od średnich w próbkach. Przyjmując n jako liczbę wszystkich elementów (obserwacji), m jako liczbę grup, n_i jako rozmiar i -tej grupy, Y_{ij} jako j -ty element i -tej grupy, statystykę można zdefiniować jako⁶⁷:

$$F = \frac{(n-m) \sum_i^m n_i (\bar{Z}_i - \bar{Z})^2}{(m-1) \sum_{i=1}^m \sum_{j=1}^{n_i} (Z_{ij} - \bar{Z}_i)^2}, \quad (2.31)$$

gdzie: $Z_{ij} = |Y_{ij} - \bar{Y}_i|$, $\bar{Z}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Z_{ij}$ oraz $\bar{Z} = \sum_{i=1}^m \sum_{j=1}^{n_i} Z_{ij}$.

Statystyka ma rozkład F o $m - 1$ i $n - 1$ stopniach swobody. Wartość istotności testu Levene'a $p < 0,05$ (lub $p < 0,01$) oznacza,

* Model średniej ruchomej prostej został omówiony w części III podręcznika: J. Nazarko (red.) (2005), *Prognosowanie w zarządzaniu przedsiębiorstwem. Cz. III. Prognosowanie na podstawie modeli adaptacyjnych*, Wydawnictwo Politechniki Białostockiej, Białystok.

⁶⁵ R. Salles, K. Belloze, F. Porto, P.H. Gonzalez, E. Ogasawara (2019), *Nonstationary time series...*, op. cit.

⁶⁶ Ibidem.

⁶⁷ M.E. O'Neill, K. Mathews (2000), *A weighted least squares approach to Levene's test of homogeneity of variance*, „Australian and New Zealand Journal of Statistics” 42(1), s. 81-100, doi: 10.1111/1467-842X.00109.

że hipotezę o homogeniczności wariancji można odrzucić⁶⁸. Zmodyfikowany test Levene'a, w którym obliczane są odchylenia od median $\tilde{Y}_i$, a nie od średnich arytmetycznych, tzn. $Z_{ij} = |Y_{ij} - \tilde{Y}_i|$, nosi nazwę testu Browna-Forsythe'a⁶⁹.

Test Bartletta zakłada istnienie m populacji normalnych, z których losuje się niezależnie n_i elementów. Sprawdzana jest także hipoteza H_0 : wariancje są równe wobec hipotezy alternatywnej H_1 : nie wszystkie wariancje są równe⁷⁰. Test oparty jest na badaniu proporcji między wariancją łączną a iloczynem wariancji⁷¹:

$$\chi^2 = \frac{(n-m) \ln(S_p^2) - \sum_{i=1}^m (n_i - 1) \ln S_i^2}{1 + \frac{1}{3(m-1)} \left(\sum_{i=1}^m \left(\frac{1}{n_i - 1} \right) - \frac{1}{n-m} \right)}, \quad (2.32)$$

gdzie S_i^2 jest nieobciążoną wariancją i -tej grupy, $S_p^2 = \frac{1}{n-m} \sum_{i=1}^m \left(\frac{1}{n_i - 1} \right) S_i^2$ jest wariancją łączną.

Statystyka testowa ma rozkład χ^2 o $k - 1$ stopniach swobody.

W wypadku, gdy szereg ma zmienną w czasie wariancję, można spróbować przekształcić go stosując transformacje matematyczne⁷². Najprostszą z nich jest logarytmowanie, a najczęściej wykorzystywany jest logarytm naturalny, tzn. dla $b = e$ ⁷³ (liczba Eulera):

$$\hat{Y}_t = \log_b Y_t, \quad (2.33)$$

gdzie: $\hat{Y}_t$ – obserwacje szeregu czasowego po transformacji.

⁶⁸ J.L. Gastwirth, Y.R. Gel, W. Miao (2009), *The Impact of Levene's Test of Equality of Variances on Statistical Theory and Practice*, „Statistical Science” 24(3), 343–360, doi: 10.1214/09-STS301.

⁶⁹ Ibidem.

⁷⁰ E. Chodakowska (2004), *Analiza wartości sprzedaży w przedsiębiorstwie handlowym z zastosowaniem modeli ARIMA*, „Zeszyty Naukowe Politechniki Białostockiej. Ekonomia i Zarządzanie” 9, s. 37–52.

⁷¹ D. Allingham, J.C.W. Rayner (2012), *Testing Equality of Variances for Multiple Univariate Normal Populations*, „Journal of Statistical Theory and Practice” 6, s. 524–535, doi: 10.1080/15598608.2012.695703.

⁷² S. Makridakis, S.C. Wheelwright, R.J. Hyndman (1998), *Forecasting. Methods and...*, op. cit., s. 63–67.

⁷³ R. Salles, K. Belloze, F. Porto, P.H. Gonzalez, E. Ogasawara (2019), *Nonstationary time series...*, op. cit.

Stosowane są także transformacje Boxa-Coxa, które jednak wymagają obliczenia argumentu λ związanego z rozkładem danych⁷⁴:

$$\hat{Y}_t = \begin{cases} (Y_t^\lambda - 1), \lambda \neq 0 \\ \log Y_t, \lambda = 0. \end{cases} \quad (2.34)$$

Inną możliwością, w wypadku stabilności względnych zmian pomiędzy kolejnymi obserwacjami w szeregu, jest wykorzystanie transformacji procentowej⁷⁵:

$$\hat{Y}_t \approx \log \left(\frac{Y_t}{Y_{t-1}} \right). \quad (2.35)$$

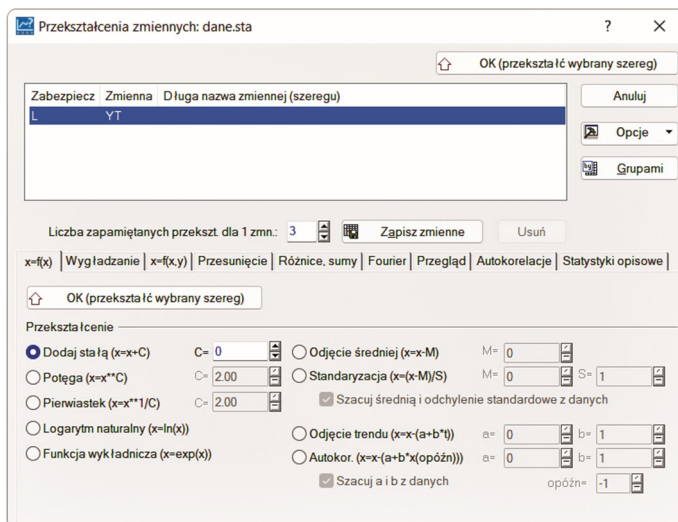
Znalezienie sekwencji transformacji potrzebnych do ustabilizowania szeregu czasowego dostarcza ważnych wskazówek w poszukiwaniu odpowiedniego modelu prognozowania. Szczególnie różnicowanie szeregów czasowych jest ważną częścią procesu dopasowywania modelu ARIMA⁷⁶.

W programie Statistica przekształcenia zmiennych dostępne są np. poprzez wybór na zakładce Statystyka → Modele zaawansowane → Szeregi czasowe i prognozowanie → OK (przekształcenia, autokorelacje, korelacje wzajemne, wykresy). Okno programu przedstawiono na rys. 2.6.

⁷⁴ Ibidem.

⁷⁵ Ibidem.

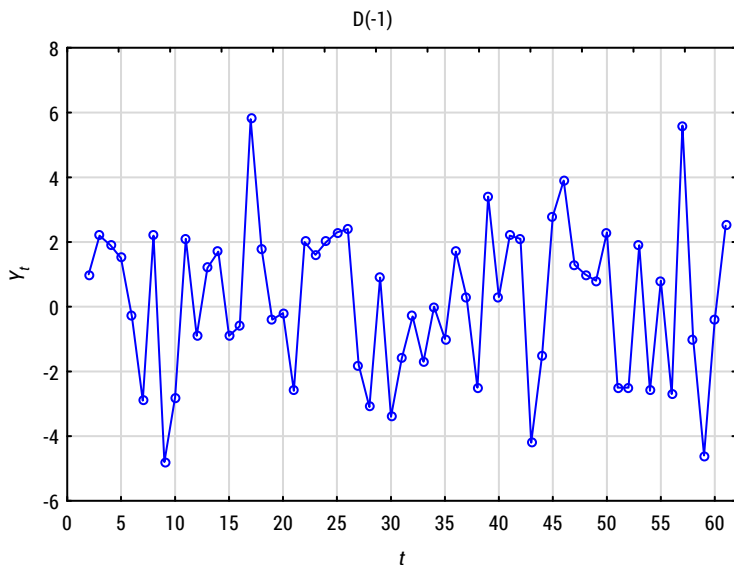
⁷⁶ R. Nau, *Statistical forecasting...*, op. cit.



Rys. 2.6. Pakiet Statistica – okno przekształcania zmiennych

Źródło: opracowanie własne.

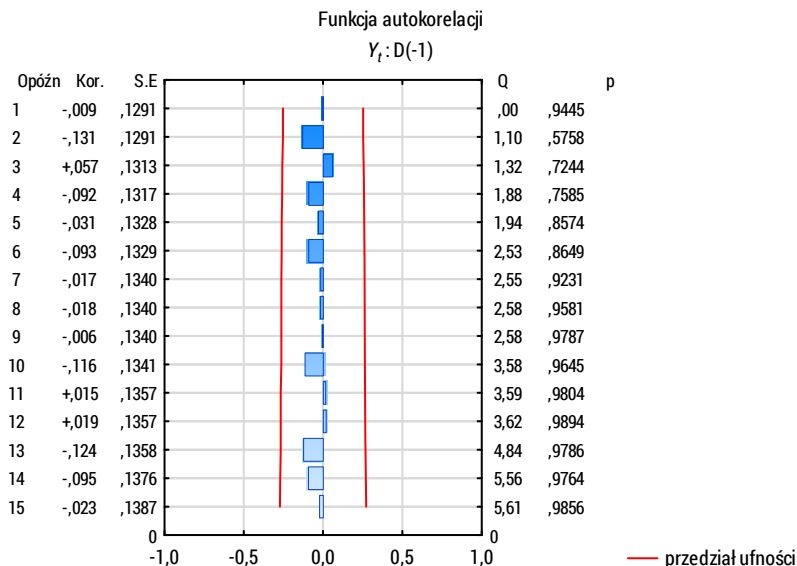
Operację różnicowania z opóźnieniem 1 zastosowano dla danych przedstawiających średnie temperatury w Białymstoku wykazujące wyraźny trend rosnący, zaprezentowany na rys. 1.6 w rozdziale pierwszym. Na rys. 2.7 zilustrowano szereg zróżnicowany z opóźnieniem 1. Na wybór takiego różnicowania wskazywały: (i) trend widoczny na rys. 1.6 oraz (ii) kształty funkcji autokorelacji cząstkowej (znacząca wartość przy opóźnieniu 1) i autokorelacji (bardzo powolne zanikanie) (rys. 1.8 i 1.9).



Rys. 2.7. Szereg średniej temperatury zanotowanej w Białymstoku w dniach 1.04–31.05.2019 r. po operacji różnicowania $d = 1$

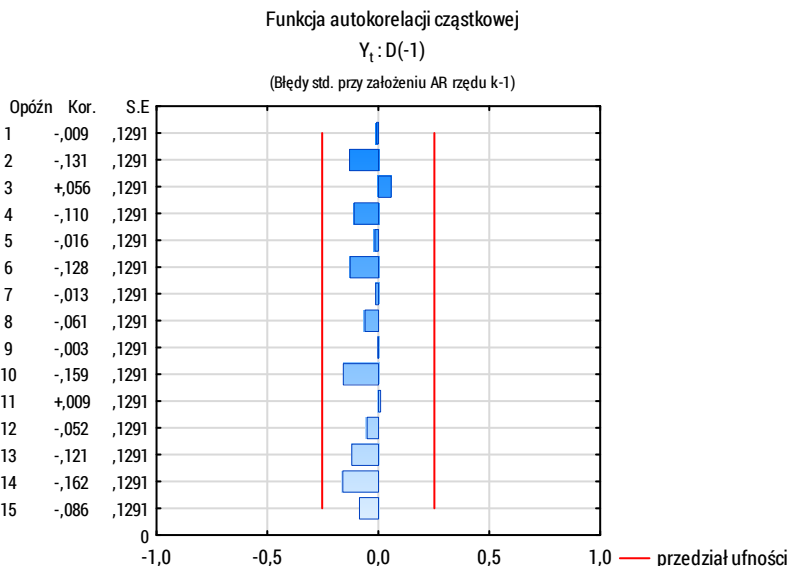
Źródło: opracowanie własne danych z serwisu [Meteomodel.pl](https://meteomodel.pl/dane/historyczne-dane-pomiarowe/), <https://meteomodel.pl/dane/historyczne-dane-pomiarowe/> na podstawie danych udostępnionych przez Instytut Meteorologii i Gospodarki Wodnej – Państwowy Instytut Badawczy [8.12.2019].

Ponowne obliczenie wartości funkcji autokorelacji i autokorelacji cząstkowej zróżnicowanego szeregu wskazują, że szereg osiągnął stacjonarność ze względu na średnią (nie ma znaczących wartości funkcji autokorelacji i autokorelacji cząstkowej). Co ilustrują wykresy na rys. 2.8 i 2.9.



Rys. 2.8. Wartości autokorelacji zróżnicowanego $d = 1$ szeregu średniej temperatury zanotowanej w Białymstoku w dniach 1.04–31.05.2019 r.

Źródło: opracowanie własne danych z serwisu [Meteomodel.pl](https://meteomodel.pl/dane/histeryczne-dane-pomiarowe/), <https://meteomodel.pl/dane/histeryczne-dane-pomiarowe/> na podstawie danych udostępnionych przez Instytut Meteorologii i Gospodarki Wodnej – Państwowy Instytut Badawczy [8.12.2019].



Rys. 2.9. Wartości autokorelacji cząstkowej zróżnicowanego szeregu średniej temperatury zanotowanej w Białymstoku w dniach 1.04–31.05.2019 r.

Źródło: opracowanie własne danych z serwisu [Metemodel.pl](https://meteomodel.pl/dane/histeryczne-dane-pomiarowe/), <https://meteomodel.pl/dane/histeryczne-dane-pomiarowe/> na podstawie danych udostępnionych przez Instytut Meteorologii i Gospodarki Wodnej – Państwowy Instytut Badawczy [8.12.2019].

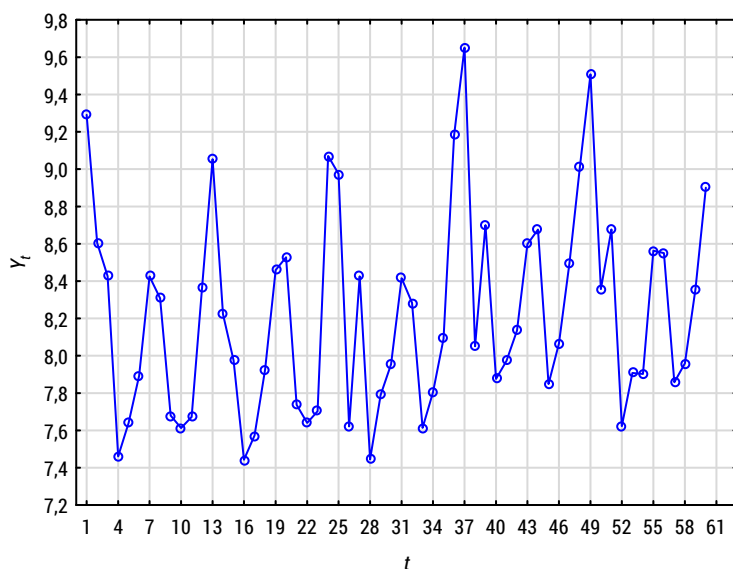
Kolejny przykład szeregu niestacjonarnego to szereg czasowy całkowitego zużycia energii pierwotnej w Stanach Zjednoczonych w bilingach BTU (ang. *British Thermal Unit*), przedstawiony w tabeli 2.1 i zilustrowany na rys. 2.10.

Tabela 2.1. Szereg zużycia energii pierwotnej w Stanach Zjednoczonych w latach 2015–2019 w bilingach BTU

t	data	Y_t	t	data	Y_t	t	data	Y_t	t	data	Y_t	t	data	Y_t
1	2015	I 9,2910	13	2016	I 9,0531	25	2017	I 8,9744	37	2018	I 9,6511	49	2019	I 9,5078
2		II 8,6043	14		II 8,2223	26		II 7,6221	38		II 8,0509	50		II 8,3568
3		III 8,4266	15		III 7,9754	27		III 8,4260	39		III 8,6998	51		III 8,6770
4		IV 7,4616	16		IV 7,4425	28		IV 7,4451	40		IV 7,8760	52		IV 7,6234
5		V 7,6389	17		V 7,5720	29		V 7,7903	41		V 7,9718	53		V 7,9083

t	data	Y_t	t	data	Y_t	t	data	Y_t	t	data	Y_t	t	data	Y_t
6	VI	7,8940	18	VI	7,9257	30	VI	7,9542	42	VI	8,1350	54	VI	7,9027
7	VII	8,4249	19	VII	8,4590	31	VII	8,4160	43	VII	8,5987	55	VII	8,5633
8	VIII	8,3079	20	VIII	8,5278	32	VIII	8,2839	44	VIII	8,6765	56	VIII	8,5504
9	IX	7,6803	21	IX	7,7366	33	IX	7,6139	45	IX	7,8494	57	IX	7,8585
10	X	7,6125	22	X	7,6408	34	X	7,8044	46	X	8,0652	58	X	7,9589
11	XI	7,6721	23	XI	7,7032	35	XI	8,0912	47	XI	8,4985	59	XI	8,3579
12	XII	8,3637	24	XII	9,0702	36	XII	9,1814	48	XII	9,0117	60	XII	8,9008

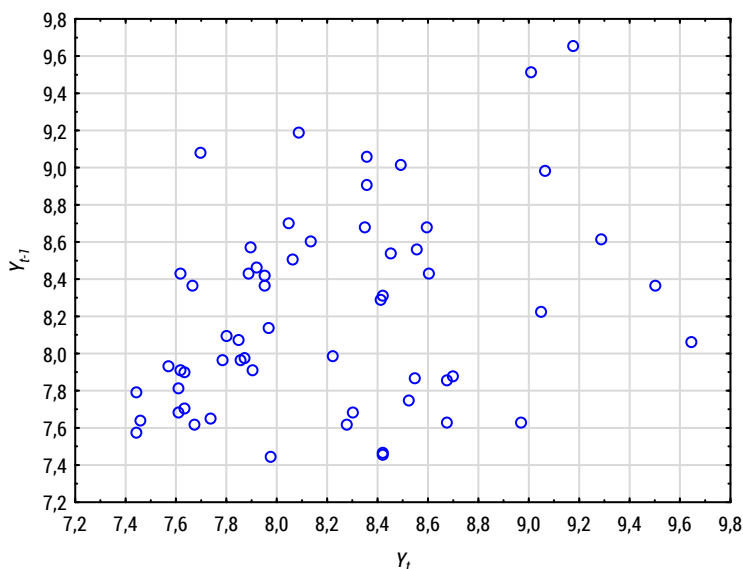
Źródło: serwis U.S. Energy Information Administration, <https://www.eia.gov/totalenergy/data/monthly/> [28.07.2020].



Rys. 2.10. Szereg zużycia energii pierwotnej w Stanach Zjednoczonych w latach 2015–2019

Źródło: opracowanie własne danych z serwisu U.S. Energy Information Administration, <https://www.eia.gov/totalenergy/data/monthly/> [28.07.2020].

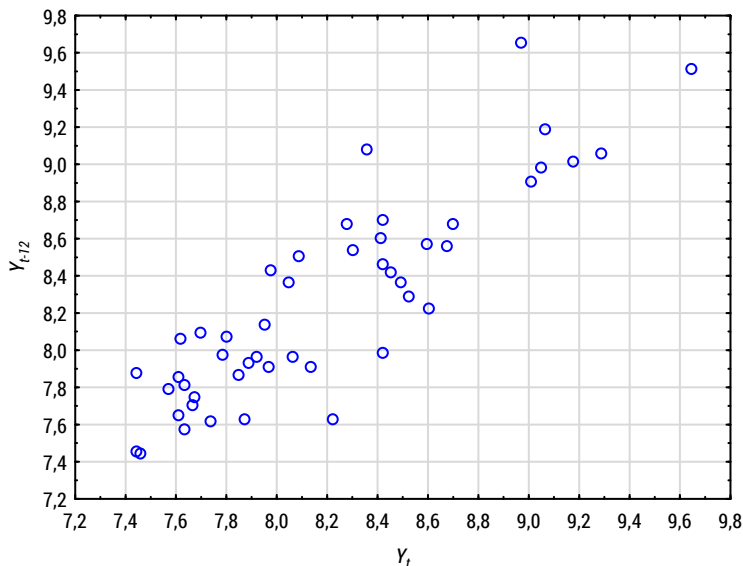
Na rys. 2.11 pokazano wykres rozrzutu obrazujący zależność pomiędzy wartościami zużycia energii w danym miesiącu (Y_t) i w miesiącu poprzedzającym (Y_{t-1}). Rozrzut punktów wykresu nie wskazuje na istnienie znaczącej zależności autokorelacyjnej dla opóźnienia równego 1.



Rys. 2.11. Wykres rozrzutu obrazujący zależność pomiędzy wartościami zużycia energii w Stanach Zjednoczonych w latach 2015–2019 w danym miesiącu (Y_t) i w miesiącu poprzedzającym (Y_{t-1})

Źródło: opracowanie własne danych z serwisu U.S. Energy Information Administration, <https://www.eia.gov/totalenergy/data/monthly/> [28.07.2020].

Inaczej sytuacja przedstawia się na rys. 2.12, na którym pokazano zależność pomiędzy wartościami zużycia energii w danym miesiącu (Y_t) i w miesiącu poprzedzającym o rok (Y_{t-12}).

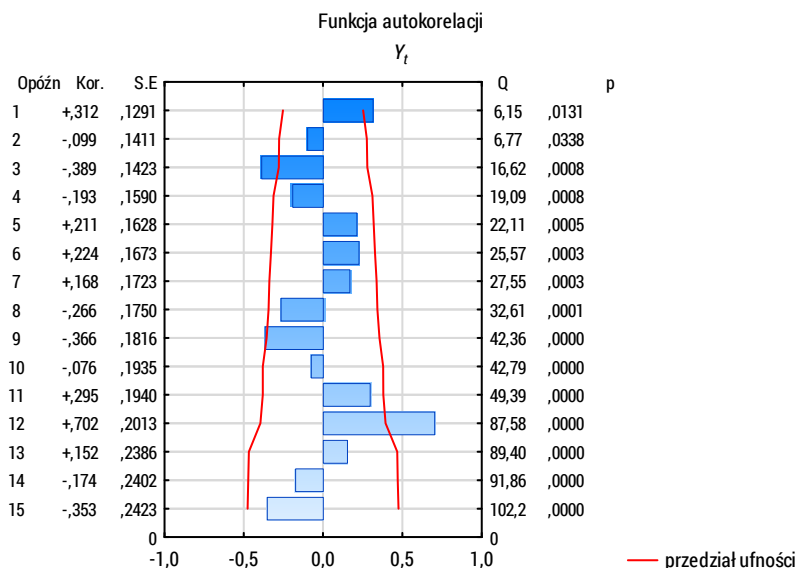


Rys. 2.12. Wykres rozrzutu obrazujący zależność pomiędzy wartościami zużycia energii w Stanach Zjednoczonych w latach 2015–2019 w danym miesiącu (Y_t) i w miesiącu poprzedzającym o rok (Y_{t-12})

Źródło: opracowanie własne danych z serwisu U.S. Energy Information Administration, <https://www.eia.gov/totalenergy/data/monthly/> [28.07.2020].

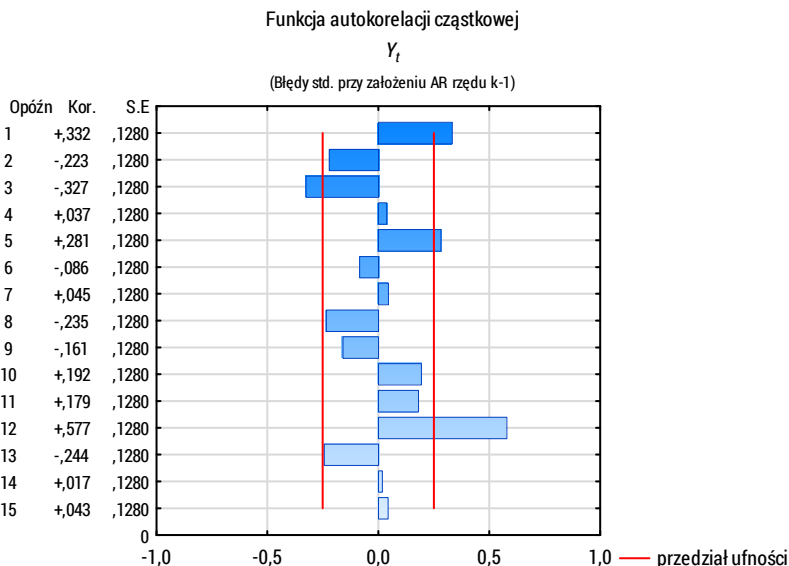
Na wykresie zaobserwować można wyraźną zależność liniową między zmiennymi, co świadczy o istnieniu autokorelacji w szeregu dla opóźnienia $k = 12$.

Wnioski z obserwacji wykresów rozrzutu (rys. 2.11 i 2.12) potwierdzają wykresy funkcji autokorelacji (rys. 2.13) i autokorelacji cząstkowej (rys. 2.14).



Rys. 2.13. Wartości autokorelacji szeregu zużycia energii pierwotnej w Stanach Zjednoczonych w latach 2015–2019

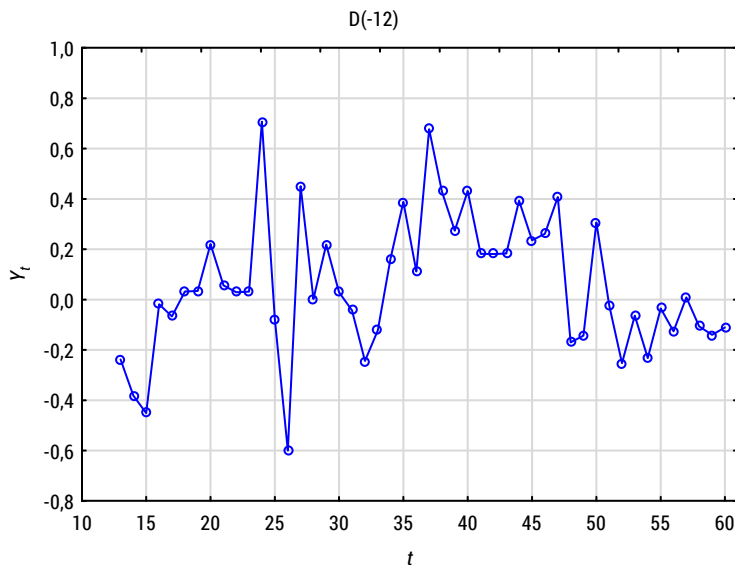
Źródło: opracowanie własne danych z serwisu U.S. Energy Information Administration, <https://www.eia.gov/totalenergy/data/monthly/> [28.07.2020].



Rys. 2.14. Wartości autokorelacji cząstkowej zużycia energii pierwotnej w Stanach Zjednoczonych w latach 2015–2019

Źródło: opracowanie własne danych z serwisu U.S. Energy Information Administration, <https://www.eia.gov/totalenergy/data/monthly/> [28.07.2020].

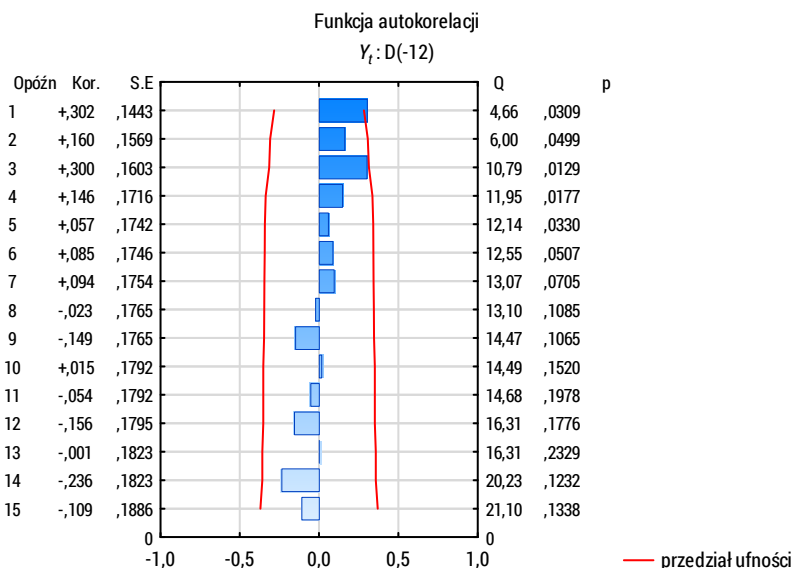
Na obu wykresach wyraźnie zaznacza się skokowy wzrost wartości funkcji dla opóźnienia równego 12. Ponieważ obserwacje dokonywane były w okresach miesięcznych, występowanie autokorelacji dla opóźnienia $k = 12$ świadczy o obecności w szeregu sezonowości rocznej ($s = 12$). W celu jej eliminacji szereg różnicowano z opóźnieniem równym 12. Szereg po operacji różnicowania sezonowego przedstawiono na rys. 2.15.



Rys. 2.15. Szereg zużycia energii pierwotnej w Stanach Zjednoczonych w latach 2015–2019 po operacji różnicowania sezonowego $s = 12$

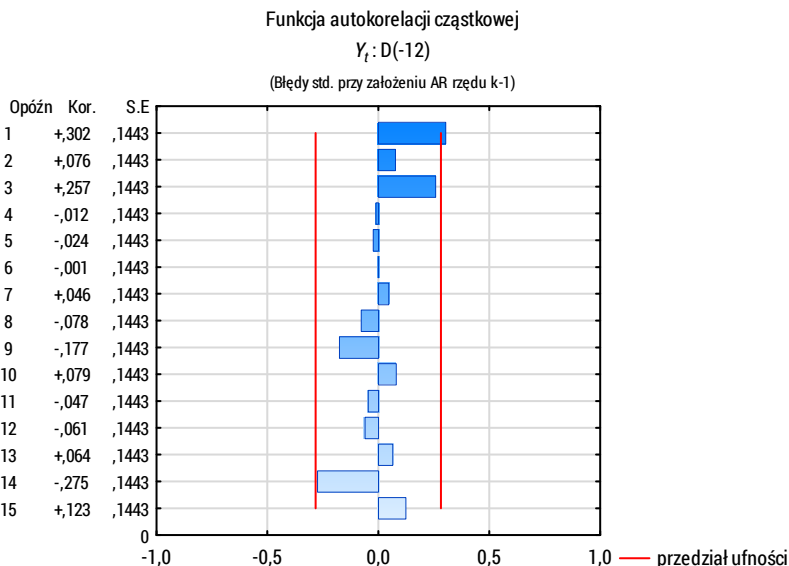
Źródło: opracowanie własne danych z serwisu U.S. Energy Information Administration, <https://www.eia.gov/totalenergy/data/monthly/> [28.07.2020].

Dla zróżnicowanego sezonowo szeregu obciążeń sporządzono następnie wykresy funkcji autokorelacji (rys. 2.16) i autokorelacji cząstkowej (rys. 2.17).



Rys. 2.16. Wartości autokorelacji zróżnicowanego z $s = 12$ szeregu zużycia energii pierwotnej w Stanach Zjednoczonych w latach 2015–2019

Źródło: opracowanie własne danych z serwisu U.S. Energy Information Administration, <https://www.eia.gov/totalenergy/data/monthly/> [28.07.2020].



Rys. 2.17. Wartości autokorelacji częściowej zróżnicowanego $s = 12$ szeregu zużycia energii pierwotnej w Stanach Zjednoczonych w latach 2015–2019

Źródło: opracowanie własne danych z serwisu U.S. Energy Information Administration, <https://www.eia.gov/totalenergy/data/monthly/> [28.07.2020].

Zarówno wykres funkcji ACF, jak i PACF nie wykazują statystycznie istotnych wartości co pozwala przyjąć, że szereg osiągnął stacjonarność i można opisać szereg modelem ARIMA. Problem doboru adekwatnego modelu jest tematem rozdziału trzeciego.

Kluczowe zagadnienia:

szereg stacjonarny, szereg niestacjonarny,
 różnicowanie, logarytmowanie

Literatura podstawowa

- [1] Box G.E.P., Jenkins G.M. (1983), *Analiza szeregów czasowych. Prognozowanie i sterowanie*, PWN, Warszawa
- [2] Cieślak M. (red.) (2011), *Prognozowanie gospodarcze. Metody i zastosowania*, Wydanie czwarte zmienione, Wydawnictwo Naukowe PWN, Warszawa
- [3] Makridakis S., Wheelwright S.C., Hyndman R.J. (1998), *Forecasting. Methods and Applications*, John Wiley & Sons, Inc.
- [4] Nazarko J. (red.) (2011), *Prognozowanie w zarządzaniu przedsiębiorstwem. Cz. II. Prognozowanie na podstawie szeregów czasowych*, Wydawnictwo Politechniki Białostockiej, Białystok
- [5] Nazarko J. (red.) (2005), *Prognozowanie w zarządzaniu przedsiębiorstwem. Cz. III. Prognozowanie na podstawie modeli adaptacyjnych*, Wydawnictwo Politechniki Białostockiej, Białystok
- [6] Shumway R.H., Stoffer D.S. (2017), *Time Series and Its Applications with R Examples*, Springer International Publishing AG

Literatura uzupełniająca

- [1] Allingham D., Rayner J.C.W. (2012), *Testing Equality of Variances for Multiple Univariate Normal Populations*, „Journal of Statistical Theory and Practice” 6, s. 524–535, doi: 10.1080/15598608.2012.695703
- [2] Gastwirth J.L., Gel Y.R., Miao W. (2009), *The Impact of Levene's Test of Equality of Variances on Statistical Theory and Practice*, „Statistical Science” 24(3), s. 343–360, doi: 10.1214/09-STS301
- [3] Gruszczyński M., Podgórska M. (red.) (1996), *Ekonometria*, Oficyna Wydawnicza Szkoła Główna Handlowa, Warszawa
- [4] Heiberger R.M., Teles P. (2002), *Displays for Direct Comparison of ARIMA Models*, „The American Statistician” 56(2), s. 131–138, doi: 10.1198/000313002317572808
- [5] Lcc D., Schmidt P. (1996), *On the power of the KPSS test of stationarity against fractionally-integrated alternatives*, „Journal of Econometrics” 73, s. 285–302, doi: 10.1016/0304-4076(95)01741-0
- [6] Kwiatkowski D., Phillips P.C.B., Schmidt P., Shin Y. (1992), *Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?*, „Journal

- of Econometrics” 54(1–3), s. 159–178, doi: 10.1016/0304-4076(92)90104-Y
- [7] Łobejko S. (red.), Maślowska K., Wojdan R. (2015), *Analiza i prognozowanie szeregów czasowych z programem SAS*, Szkoła Główna Handlowa w Warszawie – Oficyna Wydawnicza, Warszawa
 - [8] Mastrangelo C.M., Simpson J.R., Montgomery D.C. (2013), *Time Series Analysis*, in: Gass S.I., Fu M.C. (red.) *Encyclopedia of Operations Research and Management Science*. Springer, Boston, MA, doi: 10.1007/978-1-4419-1153-7_1045
 - [9] Maynard A., Smallwood A., Wohar M.E. (2013), *Long Memory Regressors and Predictive Testing: A Two-stage Rebalancing Approach*, „Econometric Reviews” 32(3), s. 318–360, doi: 10.1080/07474938.2012.690663
 - [10] Nau R., *Statistical forecasting: notes on regression and time series analysis*, Fuqua School of Business, Duke University, <https://people.duke.edu/~rnau/411home.htm> [7.08.2020]
 - [11] O'Neill M.E., Mathews K. (2000), *A weighted least squares approach to Levene's test of homogeneity of variance*, „Australian and New Zealand Journal of Statistics” 42(1), s. 81–100, doi: 10.1111/1467-842X.00109
 - [12] Pfaff B. (2008), *Analysis of Integrated and Cointegrated Time Series with R*, Springer-Verlag New York
 - [13] Salles R., Belloze K., Porto F., Gonzalez P.H., Ogasawara E. (2019), *Nonstationary time series transformation methods: An experimental review*, „Knowledge-Based Systems” 164, s. 274–291, doi: 10.1016/j.knosys.2018.10.041
 - [14] Strawiński P., *Materiały do zajęć: statystyczna analiza danych finansowych*, http://coin.wne.uw.edu.pl/pstrawinski/dane_fin/adf03.pdf [5.09.2021]
 - [15] Syczewska E.M. (2010), *Empirical power of the Kwiatkowski-Phillips-Schmidt-Shin test*, Department of Applied Econometrics Working Papers 3, <http://yadda.icm.edu.pl/yadda/element/bwmeta1.element.ekon-element-000171231155>
 - [16] Yang W., Zurbenko I. (2010), Nonstationarity, „WIREs Computational Statistics” 2(1), s. 107–115, doi:10.1002/wics.64

3. MODELE ARIMA SZEREGÓW CZASOWYCH

3.1. Rodzaje modeli klasy ARIMA

Modele klasy ARIMA, czyli autoregresyjne zintegrowane modele średniej ruchomej lub scałkowane modele autoregresji i średniej ruchomej (ang. *Autoregressive Integrated Moving Average*), są jedną z najbardziej znanych i popularnych metod analizy i prognozowania szeregów czasowych o różnej strukturze, których różnice rzędu d są procesami stacjonarnymi ze względu na wartość średnią (zintegrowanymi rzędu d)⁷⁷. Z uwagi na przełomowy wkład, który w ich rozwój wnieśli G.E.P. Box i G.M. Jenkins (1970)⁷⁸ w literaturze nazywane są także modelami Boxa-Jenkinsa. Modele charakteryzują się dużą elastycznością i uniwersalnością, lecz pewnym ograniczeniem ich stosowania jest konieczność dysponowania stosunkowo długim, co najmniej pięćdziesięcioelementowym, a w praktyce często znacznie dłuższym szeregiem obserwacji.

W modelach ARIMA bieżąca wartość zmiennej jest przedstawiana jako liniowa funkcja jej przeszłych wartości i wahań przypadkowych⁷⁹:

$$Y_t = f[Y_{t-k}, e_{t-k}] + e_t, \quad k > 0, \quad (3.1)$$

gdzie:

Y_t – wartość zmiennej w momencie lub okresie t ,

e_t – wahania przypadkowe w momencie lub okresie t .

Konstrukcja modeli ARIMA opiera się na wyjaśnionym w rozdziale pierwszym zjawisku autokorelacji w szeregach czasowych, czyli zależności liniowej między wartościami danej zmiennej a wartościami tej samej zmiennej opóźnionymi w czasie.

⁷⁷ C.M. Mastrangelo, J.R. Simpson, D.C. Montgomery (2013), *Time Series Analysis...*, op. cit.

⁷⁸ G.E.P. Box, G.M. Jenkins (1976), *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day. 1970 (Revised edition published 1976).

⁷⁹ S.A. DeLurgio (1998), *Forecasting Principles and...*, op. cit., s. 268.

Wyróżnia się dwa podstawowe typy zależności, na podstawie których budowane są pozostałe bardziej skomplikowane modele⁸⁰:

- model autoregresji rzędu 1: AR(1), w którym każda obserwacja Y_t jest skorelowana z poprzedzającą obserwacją Y_{t-1} , a poprzez to także – choć w mniejszym stopniu – z wcześniejszymi obserwacjami:

$$Y_t = \phi_1 Y_{t-1} + e_t; \quad (3.2)$$

- model średniej ruchomej rzędu 1: MA(1), w którym każda wartość obserwacji Y_t jest skorelowana z odchyleniem losowym (resztą) e_{t-1} poprzedzającej obserwacji Y_{t-1} i jest nieskorelowana z wcześniejszymi obserwacjami (Y_t zależy od wahań przypadkowych występujących w okresie t oraz $t - 1$):

$$Y_t = e_t - \theta_1 e_{t-1}. \quad (3.3)$$

Te dwa podstawowe modele: AR(1) i MA(1), zazwyczaj są niewystarczające do opisu struktury autokorelacyjnej rzeczywistych szeregów czasowych. Są one jedynie bazą, na podstawie której buduje się bardziej złożone modele ARIMA. Ogólne modele Boxa-Jenkinsa zawierają możliwość wprowadzenia również długo-terminowych zależności (tzn. takich, w których opóźnienie jest większe od 1)⁸¹.

Model autoregresji rzędu p – AR(p) można zapisać jako⁸²:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + e_t, \quad (3.4)$$

przy czym:

ϕ_j – parametr modelu autoregresji stojący przy wyrazie Y_{t-j} ,

e_t – reszta dla momentu lub okresu t .

Przy wykorzystaniu wprowadzonego w rozdziale drugim operatora B , tj. $B^j Y_t = Y_{t-j}$, równanie modelu AR(p) (3.4) przyjmuje postać:

⁸⁰ R.M. Heiberger, P. Teles (2002), *Displays for...*, op. cit.

⁸¹ Ibidem, s. 131.

⁸² S. Makridakis, S.C. Wheelwright, R.J. Hyndman (1998), *Forecasting. Methods...*, op. cit., s. 339.

$$(1 - \phi_1 B - \dots - \phi_p B^p) Y_t = e_t. \quad (3.5)$$

Można je zapisać w skrócie:

$$\phi(B) Y_t = e_t, \quad (3.6)$$

przy czym:

$\phi(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$ – operator autoregresyjny rzędu p .

Model średniej ruchomej rzędu q – $MA(q)$ można zapisać jako⁸³:

$$Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q}, \quad (3.7)$$

przy czym:

θ_j – parametr modelu średniej ruchomej stojący przy wyrazie e_{t-j} ,

e_{t-k} – reszta dla momentu lub okresu $t - k$.

Także model MA można przedstawić z wykorzystaniem operatora B jako:

$$Y_t = (1 - \theta_1 B - \dots - \theta_q B^q) e_t, \quad (3.8)$$

lub:

$$Y_t = \theta(B) e_t, \quad (3.9)$$

przy czym:

$\theta(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$ – operator średniej ruchomej rzędu q .

Model AR bazuje na założeniu, że istnieje autokorelacja pomiędzy wartością zmiennej prognozowanej a jej wartościami opóźnionymi w czasie. Wartość zmiennej wyrażona jest przez funkcję przeszłych obserwacji. W modelu MA wartość zmiennej prognozowanej

⁸³ Ibidem, s. 342.

jest zaś liniową funkcją reszt modelu. Model ARMA(p, q) jest kombinacją modelu AR(p) i MA(q) i zapisuje się go jako⁸⁴:

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q}. \quad (3.10)$$

Przedstawiając równanie (3.10) z wykorzystaniem operatora B otrzymuje się:

$$(1 - \phi_1 B - \dots - \phi_p B^p) Y_t = (1 - \theta_1 B - \dots - \theta_q B^q) e_t. \quad (3.11)$$

Formuła (3.11) jest równoznaczna z zapisem:

$$\phi(B) Y_t = \theta(B) e_t, \quad (3.12)$$

przy czym:

$\phi(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$ – operator autoregresyjny rzędu p ,
 $\theta(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$ – operator średniej ruchomej rzędu q .

Ogólny model ARIMA jest kombinacją modelu AR(p) i MA(q), a także pozwala ująć różnicowanie z opóźnieniem 1 lub wyższym, w celu likwidacji niestacjonarności danych ze względu na średnią (tzn. poprzez operowanie nie bezpośrednio na obserwacji Y_t , ale na – wprowadzonym w rozdziale drugim – operatorze różnicowym $\nabla^d(Y_t)$)⁸⁵. Innymi słowy, model ARIMA jest modelem z przekształceniem różnicowym sprowadzającym szereg niestacjonarny ze względu na średnią do postaci stacjonarnej. Oznacza to, że proces niestacjonarny można poddać d -krotnemu różnicowaniu w celu uzyskania stacjonarności i zastosować do tego nowego stacjonarnego szeregu czasowego model ARMA lub bezpośrednio budować model ARIMA szeregu pierwotnego niestacjonarnego⁸⁶.

⁸⁴ E. Chodakowska, J. Nazarko, Ł. Nazarko (2021), *ARIMA Models in Electrical Load Forecasting and Their Robustness to Noise*, „Energies” 14(23), 7952, s. 1–22, doi: 10.3390/en14237952.

⁸⁵ R.M. Heiberger, P. Teles (2002), *Displays for...*, op. cit., s. 131.

⁸⁶ D. Witkowska (2012), *Podstawy ekonometrii i teorii prognozowania*, Wolters Kluwer Polska SA, s. 169–176.

Wszystkie niesezonowe modele klasy ARIMA zapisuje się posługując się uniwersalną notacją: $ARIMA(p,d,q)$, gdzie:

p – rząd autoregresji AR,

d – stopień różnicowania (integracji),

q – liczba parametrów średniej ruchomej MA.

Jeżeli któryś z parametrów p , d lub q jest równy 0, nazwę modelu można zapisać, opuszczając odpowiedni człon w nazwie, np.: $ARIMA(2,0,0) \equiv AR(2)$, $ARIMA(1,0,1) \equiv ARMA(1,1)$. Wskazywanie 0 może być konieczne podczas pracy z oprogramowaniem, podczas gdy w opisie czy interpretacji przyjęte jest opuszczanie zerowych parametrów w nazwie modelu, tj. raczej proces $AR(p)$ niż $ARIMA(p,0,0)$.

Przy okazji można zwrócić uwagę, że model $ARIMA(1,1,0) \equiv ARMA(1 + 1,0)$. Uogólniając, każdy scałkowany model autoregresji tj. $ARI(p,d)$, można zapisać jako model autoregresji wyższego rzędu, czyli $AR(p + d)$, gdyż:

$$\begin{aligned} (1 - \phi_1 B - \dots - \phi_p B^p)(1 - B)^d Y_t &= e_t \\ \equiv (1 - \phi_1^* B - \dots - \phi_{p+d}^* B^{p+d}) Y_t &= e_t. \end{aligned} \quad (3.13)$$

Zazwyczaj parametry modelu $AR(p + d)$ nie spełniają jednak warunku stacjonarności⁸⁷.

Przedstawiany w rozdziale pierwszym i drugim przykład szeregu temperatury zanotowanej w Białymstoku, który osiągnął stacjonarność po jednokrotnym różnicowaniu można próbować opisać modelem $ARIMA(p,1,q)$, gdzie określenie właściwych wartości p i q będzie tematem podrozdziału 3.2 i 3.3.

Przykładem zintegrowanych modeli ARIMA jest zintegrowany model autoregresji pierwszego rzędu $ARIMA(1,1,0)$:

$$\phi(B)(1 - B)Y_t = e_t. \quad (3.14)$$

Interpretacja modelu jest następująca: szereg w celu osiągnięcia stacjonarności został poddany niesezonowemu różnicowaniu

⁸⁷ M. Cieślak (red.) (2011), *Prognozowanie gospodarcze...*, op. cit., s. 115–116.

pierwszego rzędu, a otrzymany w ten sposób szereg opisywany jest modelem autoregresji pierwszego rzędu. Model (3.14) można zapisać, wykorzystując równoznaczne formuły:

$$(1 - \phi_1 B)(1 - B)Y_t = e_t, \quad (3.15)$$

$$(1 - \phi_1 B - B + \phi_1 B^2)Y_t = e_t, \quad (3.16)$$

$$Y_t - \phi_1 Y_{t-1} - Y_{t-1} + \phi_1 Y_{t-2} = e_t, \quad (3.17)$$

$$Y_t = (1 + \phi_1)Y_{t-1} - \phi_1 Y_{t-2} + e_t. \quad (3.18)$$

Podobnie można zapisać zintegrowany w stopniu pierwszym model autoregresji rzędu 3, tzn. ARIMA(3,1,0):

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3)(1 - B)Y_t = e_t, \quad (3.19)$$

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - B + \phi_1 B^2 + \phi_2 B^3 + \phi_3 B^4)Y_t = e_t, \quad (3.20)$$

$$Y_t - (1 + \phi_1)Y_{t-1} - (\phi_2 - \phi_1)Y_{t-2} - (\phi_3 - \phi_2)Y_{t-3} + \phi_3 Y_{t-4} = e_t, \quad (3.21)$$

$$Y_t = (1 + \phi_1)Y_{t-1} + (\phi_2 - \phi_1)Y_{t-2} + (\phi_3 - \phi_2)Y_{t-3} - \phi_3 Y_{t-4} + e_t. \quad (3.22)$$

Prezentowany model ARIMA(3,1,0) może być interpretowany jako model AR(4):

$$Y_t = \phi_1^* Y_{t-1} + \phi_2^* Y_{t-2} + \phi_3^* Y_{t-3} + \phi_4^* Y_{t-4} + e_t, \quad (3.23)$$

przy czym:

$$\phi_1^* = 1 + \phi_1,$$

$$\phi_2^* = \phi_2 - \phi_1,$$

$$\phi_3^* = \phi_3 - \phi_2,$$

$$\phi_4^* = -\phi_3.$$

Zintegrowany w stopniu pierwszym model autoregresji i średniej ruchomej pierwszego rzędu ARIMA(1,1,1) oznacza, że szereg został jednokrotnie zróżnicowany i można go zamodelować kombinacją modelu AR(1) i MA(1), tzn.:

$$\phi(B)(1 - B)Y_t = \theta(B)e_t, \quad (3.24)$$

$$(1 - \phi_1 B)(1 - B)Y_t = (1 - \theta_1 B)e_t, \quad (3.25)$$

$$(1 - \phi_1 B - B + \phi_1 B^2)Y_t = e_t - \theta_1 e_{t-1}, \quad (3.26)$$

$$Y_t - \phi_1 Y_{t-1} - Y_{t-1} + \phi_1 Y_{t-2} = e_t - \theta_1 e_{t-1}, \quad (3.27)$$

$$Y_t = (1 + \phi_1)Y_{t-1} - \phi_1 Y_{t-2} + e_t - \theta_1 e_{t-1}. \quad (3.28)$$

Ogólny autoregresyjny zintegrowany model średniej ruchomej ARIMA(p, d, q) można zapisać w postaci⁸⁸:

$$\phi(B)(1 - B)^d Y_t = \theta(B)e_t. \quad (3.29)$$

Równanie (3.29) jest równoznaczne z:

$$\phi(B)\nabla^d Y_t = \theta(B)e_t, \quad (3.30)$$

gdzie:

$\phi(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$ – operator autoregresyjny rzędu p ,

$\theta(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$ – operator średniej ruchomej rzędu q ,

$\nabla^d = (1 - B)^d$ – operator różnicowy.

W przypadku szeregów, w których występuje składowa sezonowa można je modelować za pomocą sezonowego modelu ARIMA (ang. *Seasonal ARIMA*). Ogólna notacja sezonowego modelu ARIMA to: ARIMA(p, d, q)(P, D, Q) $_s^s$ ⁸⁹, gdzie s oznacza liczbę sezonów w cyklu sezonowym⁹⁰:

$$\phi(B^s)\phi(B)\nabla_s^D \nabla^d Y_t = \theta(B^s)\theta(B)e_t, \quad (3.31)$$

gdzie:

s – liczba sezonów w cyklu sezonowym,

$\nabla_s^D = (1 - B^s)^D$ – operator różnicowania sezonowego,

D – rząd różnicowania sezonowego,

$\phi(B^s) = (1 - \phi_1 B^s - \dots - \phi_P B^{Ps})$ – sezonowy operator autoregresyjny rzędu P ,

⁸⁸ R.M. Heiberger, P. Teles (2002), *Displays for...*, op. cit., s. 131.

⁸⁹ W. Zalewski, J. Nazarko, A. Jurczuk (2005), *ARIMA models in load modelling with clustering approach*, „PowerTech IEEE'2005”, St. Petersburg, June 27–30.

⁹⁰ R.M. Heiberger, P. Teles (2002), *Displays for...*, op. cit., s. 131.

$\theta(B^s) = (1 - \theta_1 B^s - \dots - \theta_Q B^{Qs})$ – sezonowy operator średniej ruchomej rzędu Q .

Przykładem szeregu ze składową sezonową jest zaprezentowany w rozdziale drugim szereg zużycia energii pierwotnej w Stanach Zjednoczonych, który wymagał jednokrotnego ($D = 1$) różnicowania sezonowego $s = 12$. Można próbować opisać go modelem: $ARIMA(p, 0, q)(P, 1, Q)_{12}$, gdzie dobór wartości p , q , P i Q będzie tematem podrozdziału 3.2.

Model opisujący szereg, który został poddany jednokrotnemu różnicowaniu niesezonowemu i różnicowaniu sezonowemu z opóźnieniem 12 i który jest kombinacją niesezonowego i sezonowego modelu średniej ruchomej, tzn. $ARIMA(0, 1, 1)(0, 1, 1)_{12}$, można wyrazić wzorem:

$$(1 - B)(1 - B^{12})Y_t = \theta_1(B^{12})\theta_1(B)e_t. \quad (3.32)$$

Rozpisując równanie (3.32) otrzymuje się kolejno:

$$(1 - B - B^{12} + B^{13})Y_t = (1 - \theta_1 B^{12})(1 - \theta_1 B)e_t, \quad (3.33)$$

$$Y_t - Y_{t-1} - Y_{t-12} + Y_{t-13} = (1 - \theta_1 B^{12} - \theta_1 B + \theta_1 \theta_1 B^{13})e_t, \quad (3.34)$$

$$Y_t = Y_{t-1} + Y_{t-12} - Y_{t-13} + e_t - \theta_1 e_{t-12} - \theta_1 e_{t-1} + \theta_1 \theta_1 e_{t-13}. \quad (3.35)$$

Inny przykład to złożony model autoregresji pierwszego rzędu i średniej ruchomej pierwszego rzędu niesezonowej oraz autoregresji i średniej ruchomej sezonowej szeregu zróżnicowanego z opóźnieniem 1 i 4, tzn. $ARIMA(1, 1, 1)(1, 1, 1)_4$, który można wyrazić równaniami:

$$\Phi(B^4)\phi(B)(1 - B)(1 - B^4)Y_t = \theta(B^4)\theta(B)e_t, \quad (3.36)$$

$$\begin{aligned} (1 - \phi_1 B^4)(1 - \phi_1 B)(1 - B)(1 - B^4)Y_t = \\ = (1 - \theta_1 B)(1 - \theta_1 B^4)e_t, \end{aligned} \quad (3.37)$$

$$\begin{aligned}
Y_t = & (1 + \phi_1)Y_{t-1} - \phi_1 Y_{t-2} + (1 + \Phi_1)Y_{t-4} - \\
& -(1 + \phi_1 + \Phi_1 + \phi_1 \Phi_1)Y_{t-5} + (\phi_1 + \phi_1 \Phi_1)Y_{t-6} + \\
& + \Phi_1 Y_{t-8} + (\Phi_1 + \phi_1 \Phi_1)Y_{t-9} - \phi_1 \Phi_1 Y_{t-10} + e_t - \\
& - \theta_1 e_{t-1} - \theta_1 e_{t-4} + \theta_1 \theta_1 e_{t-5}.
\end{aligned} \tag{3.38}$$

Modele ARIMA, zarówno niesezonowe jak i sezonowe, oprócz parametrów autoregresyjnych i średniej ruchomej mogą zawierać także stałą c . Wówczas opisujące je równania (3.30) i (3.31) zapisuje się odpowiednio⁹¹:

$$\phi(B)\nabla^d Y_t = \theta(B)e_t + c \tag{3.39}$$

oraz

$$\Phi(B^s)\phi(B)\nabla_s^D \nabla^d Y_t = \Theta(B^s)\theta(B)e_t + c, \tag{3.40}$$

gdzie: c – stała.

Warto podkreślić, że modele ARIMA umożliwiają także uwzględnienie wielokrotnej sezonowości – tygodniowej i rocznej np. ARIMA(1,0,0)(0,1,0)₇(0,1,1)₃₆₅ czy dobowej i tygodniowej np. ARIMA(0,1,1)(0,1,1)₂₄(0,1,1)₁₆₈⁹². Proces ich budowy dopuszcza stosowanie różnych kombinacji, np.⁹³:

$$\begin{aligned}
& (1 - \phi_1 B - \phi_2 B^2)(1 - \phi_{24} B^{24} - \phi_{48} B^{48})(1 - \phi_{168} B^{168}) \\
& (1 - B)(1 - B^{24})Y_t = (1 - \theta_{24} B^{24})(1 - \theta_{168} B^{168})e_t.
\end{aligned} \tag{3.41}$$

Model przedstawiony równaniem (3.41) zawiera różnicowanie z opóźnieniem 1 i 24, proces autoregresji rzędu 2, ale także procesy sezonowej autoregresji rzędu 2 dla $s = 24$ oraz rzędu 1 dla $s = 168$. Dodatkowo są dwa procesy średniej ruchomej rzędu 1, odpowiednio $s = 24$ oraz $s = 168$. Ponadto w literaturze

⁹¹ S. Makridakis, S.C. Wheelwright, R.J. Hyndman (1998), *Forecasting. Methods and...*, op. cit., s. 346.

⁹² J. Nazarko, J. Chraślowska, M. Rybaczuk (2004), *Zastosowanie wielosezonowego modelu ARIMA w prognozowaniu obciążeń mocą elektryczną*, „Taksonomia” 11, s. 173–182.

⁹³ J. Contreras, R. Espínola, F.J. Nogales, A.J. Conejo (2003), *ARIMA Models to Predict Next-Day Electricity Prices*, „IEEE Transactions on Power Systems” 18(3), s. 1014–1020, doi: 10.1109/TPWRS.2002.804943.

występują modele nazwane w języku angielskim *subset models*⁹⁴ (modelami podzbiorów) lub *models with the nonzero coefficients at nonconsecutive lags*⁹⁵ (modelami z niezerowymi współczynnikami przy nienastępujących po sobie opóźnieniach). Są to modele, które zawierają parametry tylko dla niektórych opóźnień. Oprogramowanie Statistica, wykorzystane w tym podręczniku, nie pozwala na estymację parametrów takich modeli. Możliwe jest to na przykład w środowisku Matlab czy SAS.

Omawiając różne modele ARIMA można zauważyć, że niektóre omówione we wcześniejszych częściach podręcznika *Prognozowanie w zarządzaniu przedsiębiorstwem* modele prognozowania, są szczególnymi przypadkami modelu ARIMA, np.⁹⁶:

- ARIMA(0,1,1) jest równoznaczny z prostym wygładzaniem wykładniczym, gdzie $\theta = 1 - \alpha$;
- ARIMA(0,2,2) to metoda liniowa Holta, w której $\theta_1 = 2 - \alpha - \alpha\beta$ a $\theta_2 = \alpha - 1$;
- ARIMA(0,1, $s + 1$)(0,1,0)_s jest addytywną metodą Wintersa, z pewnymi ograniczeniami co do parametrów.

Wszystkie modele ARIMA muszą być modelami stacjonarnymi i odwracalnymi. Wyjaśnienie problemu stacjonarności i odwracalności modeli ARIMA wymaga rozbudowania wprowadzonych w rozdziałach pierwszym i drugim definicji i przedstawienia pojęcia ogólnego procesu liniowego definiowanego jako sumę ważoną bieżącej i przeszłych wartości procesu białego szumu e_t ⁹⁷:

$$\begin{aligned}\tilde{Y}_t &= e_t + \psi_1 e_{t-1} + \psi_2 e_{t-2} + \dots = e_t + \sum_{j=1}^{\infty} \psi_j e_{t-j} = \\ &= \left(1 + \sum_{j=1}^{\infty} \psi_j B^j\right) e_t,\end{aligned}\tag{3.42}$$

⁹⁴ SAS/ETS 13.2 User's Guide. The ARIMA Procedure (2014), SAS Institute Inc., Cary, NC, USA, <https://support.sas.com/documentation/onlinedoc/ets/132/arima.pdf> [29.11.2021].

⁹⁵ MathWorks. Help Center, <https://www.mathworks.com/help/econ/how-to-specify-arima-models.html> [29.11.2021].

⁹⁶ S. Makridakis, S.C. Wheelwright, R.J. Hyndman (1998), *Forecasting. Methods and...*, op. cit., s. 373.

⁹⁷ G.E.P. Box, G.M. Jenkins (1983), *Analiza szeregów...*, op. cit., s. 54–56.

gdzie: $\tilde{Y}_t = Y_t - \mu$ jest odchyleniem procesu od jakiegoś punktu lub od swojej średniej.

Przyjmując, że $\psi(B) = 1 + \sum_{j=1}^{\infty} \psi_j B^j$ otrzymuje się:

$$\tilde{Y}_t = \psi(B)e_t. \quad (3.43)$$

Proces liniowy (3.42–3.43) można także zapisać jako sumę ważoną przeszłych wartości procesu i składnika losowego e_t ⁹⁸:

$$\begin{aligned} \tilde{Y}_t &= \pi_1 \tilde{Y}_{t-1} + \pi_2 \tilde{Y}_{t-2} + \dots + e_t = \sum_{j=1}^{\infty} \pi_j \tilde{Y}_{t-j} + e_t = \\ &= \left(\sum_{j=1}^{\infty} \pi_j B^j \right) \tilde{Y}_t + e_t. \end{aligned} \quad (3.44)$$

Przekształcając równanie (3.44) otrzymuje się:

$$\left(1 - \sum_{j=1}^{\infty} \pi_j B^j \right) \tilde{Y}_t = e_t. \quad (3.45)$$

Jeżeli założyć się, że $1 - \sum_{j=1}^{\infty} \pi_j B^j = \pi(B)$, to w konsekwencji powstanie równanie:

$$\pi(B) \tilde{Y}_t = e_t. \quad (3.46)$$

Z równań (3.43) oraz (3.46) wynika, że:

$$\pi(B) = \psi^{-1}(B). \quad (3.47)$$

Proces autoregresyjny skończonego rzędu można przedstawić jako nieskończony proces średniej ruchomej, zaś proces średniej ruchomej skończonego rzędu jako nieskończony proces autoregresyjny. Oznacza to, że oba podejścia są w stosunku do siebie dualne.

Proces liniowy jest stacjonarny, jeżeli szereg $\psi(B)$ – funkcja tworząca dla wag ψ , zwana także funkcją przenoszenia filtru liniowego – jest zbieżny na okręgu jednostkowym lub w jego wnętrzu. Proces jest odwracalny, gdy $\pi(B)$ jest zbieżny na okręgu jednostkowym lub w jego wnętrzu⁹⁹.

W przypadku modelu AR, $\psi(B) = \phi^{-1}(B)$, warunek zbieżności sprowadza się do wymogu, aby pierwiastki jego równania $\phi(B) = 0$ leżały na zewnątrz okręgu jednostkowego. Na przykład proces AR(1) jest stacjonarny dla $|\phi_1| < 1$, gdyż wtedy szereg jest

⁹⁸ Ibidem, s. 55.

⁹⁹ Ibidem, s. 59.

zbieżny ($1 - \phi_1 B = 0$ ma pierwiastek $B = 1/\phi_1$, który powinien leżeć na zewnątrz koła jednostkowego, tzn. $|B| > 1$). Stacjonarność procesu MA jest zagwarantowana bez żadnych dodatkowych ograniczeń dla parametrów, ponieważ szereg $\psi(B) = \theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ jest skończony¹⁰⁰. Odwracalność procesu AR jest zapewniona zawsze, gdyż szereg $\pi(B) = \phi(B)$ jest skończony. Dla procesów MA szereg $\pi(B) = \theta^{-1}(B)$ jest zbieżny, gdy pierwiastki równania charakterystycznego: $\theta(B) = 0$ leżą na zewnątrz okręgu jednostkowego¹⁰¹.

Obszary dopuszczalne dla parametrów gwarantujące stacjonarność i odwracalność pięciu podstawowych modeli ARIMA zamieszczono w tabeli 3.1.

Tabela 3.1. Obszary dopuszczalne dla parametrów procesu średniej ruchomej i procesu autoregresji rzędu pierwszego i drugiego oraz procesu mieszanego

Model	ARIMA(1,d,0)	ARIMA(0,d,1)
Obszar dopuszczalny	$ \phi_1 < 1$	$ \theta_1 < 1$
Model	ARIMA(2,d,0)	ARIMA(0,d,2)
Obszar dopuszczalny	$\begin{cases} \phi_2 < 1 \\ \phi_2 + \phi_1 < 1 \\ \phi_2 - \phi_1 < 1 \end{cases}$	$\begin{cases} \theta_2 < 1 \\ \theta_2 + \theta_1 < 1 \\ \theta_2 - \theta_1 < 1 \end{cases}$
Model	ARIMA(1,d,1)	
Obszar dopuszczalny	$\begin{cases} \phi_1 < 1 \\ \theta_1 < 1 \end{cases}$	

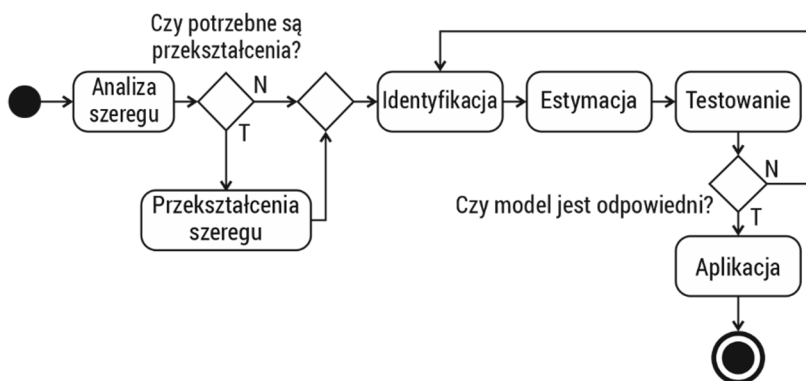
Źródło: G.E.P. Box, G.M. Jenkins (1983), *Analiza szeregów czasowych: prognozowanie i sterowanie*, Państwowe Wydawnictwo Naukowe, Warszawa, s. 178.

¹⁰⁰ Ibidem, s. 74.
¹⁰¹ Ibidem, s. 62, 74.

3.2. Budowa modeli ARIMA

Procedura budowania modeli ARIMA zgodnie z metodyką Boxa-Jenkinsa składa się z pięciu podstawowych etapów¹⁰²:

1. Określenia składowych szeregu i zbadania stacjonarności.
2. Przekształcenia szeregu do postaci stacjonarnej.
3. Identyfikacji, czyli określenia typu modelu i wyboru liczby parametrów procesów AR i MA na podstawie przekształconego szeregu.
4. Estymacji parametrów, czyli oszacowania wartości parametrów.
5. Sprawdzenia diagnostycznego modelu, tzn. sprawdzenia adekwatności i dopasowania wyestymowanego modelu do danego szeregu czasowego.



Rys. 3.1. Proces budowy modelu ARIMA

Źródło: opracowanie własne na podstawie: E. Chodakowska, J. Nazarko, Ł. Nazarko (2021), *ARIMA Models in Electrical Load Forecasting and Their Robustness to Noise*, „Energies” 14(23), 7952, s.1–22, doi:10.3390/en14237952.

¹⁰² S. Makridakis, S.C. Wheelwright, R.J. Hyndman (1998), *Forecasting. Methods and...*, op. cit., s. 314.

Proces budowy rozpoczyna się od analizy stacjonarności szeregu i podjęcia decyzji co do ewentualnej transformacji i różnicowania danych. Następnie wykorzystuje się funkcję autokorelacji i autokorelacji cząstkowej do określenia rzędów p i q operatorów autoregresji i średniej ruchomej¹⁰³. Funkcja autokorelacji procesu autoregresji rzędu p maleje w sposób łagodny (wykładniczy), natomiast funkcja autokorelacji cząstkowej urywa się po p -tym odstępie. W wypadku procesu średniej ruchomej rzędu q , funkcja autokorelacji urywa się po odstępie q , podczas gdy funkcja autokorelacji cząstkowej łagodnie maleje. Jeżeli zarówno ACF i PACF zanikają wykładniczo, sugeruje to proces mieszany. Funkcja autokorelacji procesu mieszanego zawierającego składnik autoregresji rzędu p i składnik średniej ruchomej rzędu q , po pierwszych $q-p$ odstępach jest sumą funkcji wykładniczych i sinusoid tłumionych. Analogicznie funkcja autokorelacji cząstkowej procesu mieszanego zdominowana jest po pierwszych $p-q$ odstępach przez kombinację funkcji wykładniczych i sinusoid tłumionych¹⁰⁴. Powyższe informacje zebrano w tabeli 3.2.

Tabela 3.2. Zestawienie własności funkcji autokorelacji i autokorelacji cząstkowej procesów autoregresji, średniej ruchomej i mieszanych

	Procesy autoregresji AR(p)	Procesy średniej ruchomej MA(q)	Procesy mieszane ARMA(p,q), $p>0,q>0$
Funkcja autokorelacji ACF	maleje w sposób łagodny (zanikające funkcje wykładnicze oraz/lub sinusoidy tłumione)	urywa się po opóźnieniu q	maleje w sposób łagodny (po pierwszych $q-p$ odstępach zanikające funkcje wykładnicze oraz/lub sinusoidy tłumione)

¹⁰³ Z. Yang, L. Ce, L. Lian (2017), *Electricity price forecasting by a hybrid model, combining wavelet transform, ARMA and kernel-based extreme learning machine methods*, „Applied Energy” 190, s. 291-305, doi: 10.1016/j.apenergy.2016.12.130.

¹⁰⁴ G.E.P. Box, G.M. Jenkins (1998), *Analiza szeregów...*, op. cit., s. 177.

	Procesy autoregresji $AR(p)$	Procesy średniej ruchomej $MA(q)$	Procesy mieszane $ARMA(p,q), p>0, q>0$
Funkcja autokorelacji częstkowej PACF	urywa się po opóźnieniu p	maleje w sposób łagodny (zanikające funkcje wykładnicze oraz/lub sinusoidy tłumione)	maleje w sposób łagodny (po pierwszych $p-q$ odstępach zanikające funkcje wykładnicze oraz/lub sinusoidy tłumione)

Źródło: G.E.P. Box, G.M. Jenkins (1983), *Analiza szeregów czasowych: prognozowanie i sterowanie*, Państwowe Wydawnictwo Naukowe, Warszawa, s. 86, 177; J.D. Cryer, K.-S. Chan (2008), *Time Series Analysis. With Applications in R*, Second Edition, Springer, s. 116.

W praktycznych przykładach szeregów czasowych najczęściej spotykane są procesy autoregresji oraz średniej ruchomej rzędu pierwszego i drugiego. Dlatego własności teoretycznych funkcji autokorelacji i autokorelacji częstkowych dla tych procesów zebrano w tabeli 3.3.

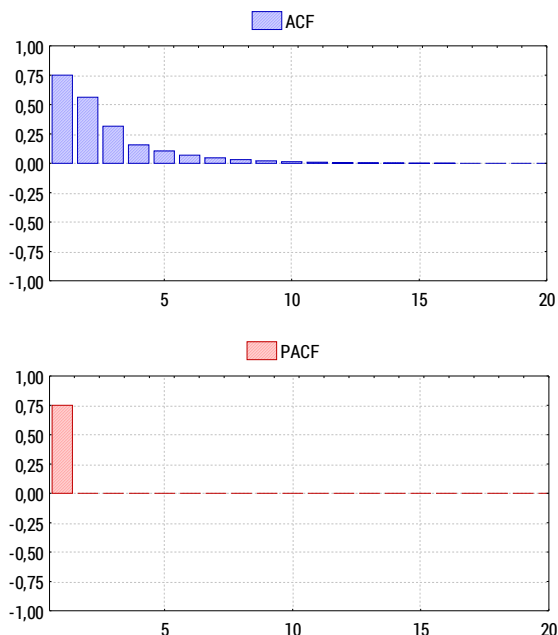
Tabela 3.3. Zachowanie się funkcji autokorelacji dla wybranych procesów $ARIMA(p,d,q)$

Proces	$ARIMA(1,d,0)$	$ARIMA(0,d,1)$
Funkcja autokorelacji	wykładniczo zanika	tylko $r_1 \neq 0$
Funkcja autokorelacji częstkowej	tylko $\hat{\phi}_{11} \neq 0$	wykładniczo zanika
Proces	$ARIMA(2,d,0)$	$ARIMA(0,d,2)$
Funkcja autokorelacji	kombinacja funkcji wykładniczych i sinusoid tłumionych	tylko $r_1 \neq 0$ i $r_2 \neq 0$
Funkcja autokorelacji częstkowej	tylko $\hat{\phi}_{11} \neq 0$ i $\hat{\phi}_{22} \neq 0$	kombinacja funkcji wykładniczych i sinusoid tłumionych
Proces	$ARIMA(1,d,1)$	
Funkcja autokorelacji	wykładniczo zanika, począwszy od pierwszego odstępu	
Funkcja autokorelacji częstkowej	wykładniczo zanika, począwszy od pierwszego odstępu	

Źródło: G.E.P. Box, G.M. Jenkins (1983), *Analiza szeregów czasowych: prognozowanie i sterowanie*, Państwowe Wydawnictwo Naukowe, Warszawa, s. 178.

Wyprowadzenie wniosków o przebiegu funkcji ACF i PACF można przeprowadzić na podstawie korelogramów – graficznie przedstawionej funkcji korelacji. Porównanie wykresów autokorelacji rzeczywistego szeregu czasowego i teoretycznych kształtów funkcji ACF i PACF różnych procesów ARIMA jest prostym sposobem odrzucenia danego modelu lub pozostawienia go do dalszej diagnostyki.

Na rysunkach 3.2 i 3.3 pokazano przebieg funkcji ACF i PACF dla procesu AR(1).

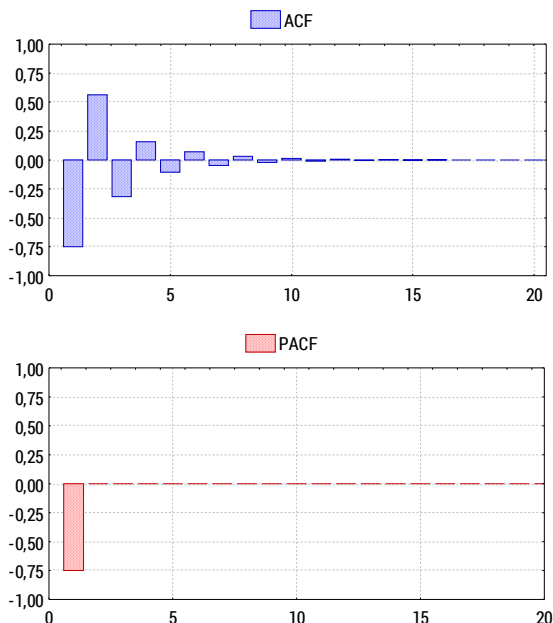


Rys. 3.2. Przebieg funkcji autokorelacji i autokorelacji cząstkowej procesu AR(1) $\phi_1 > 0$

Źródło: opracowanie własne na podstawie: S.A. DeLurgio (1998), *Forecasting Principles and Applications*, Irwin/McGraw-Hill, s. 277.

Wartości funkcji autokorelacji opadają wykładniczo, zaś funkcja autokorelacji cząstkowej jest największa przy opóźnieniu 1 i nie wykazuje korelacji dla innych opóźnień. Istnieje jednak zasadnicza różnica w przebiegu funkcji autokorelacji dla modeli

z $\phi_1 < 0$ oraz $\phi_1 > 0$. W przypadku, gdy $\phi_1 > 0$ wartości funkcji autokorelacji dla wszystkich opóźnień są dodatnie, jeżeli $\phi_1 < 0$ wartości zmieniają się naprzemiennie¹⁰⁵.



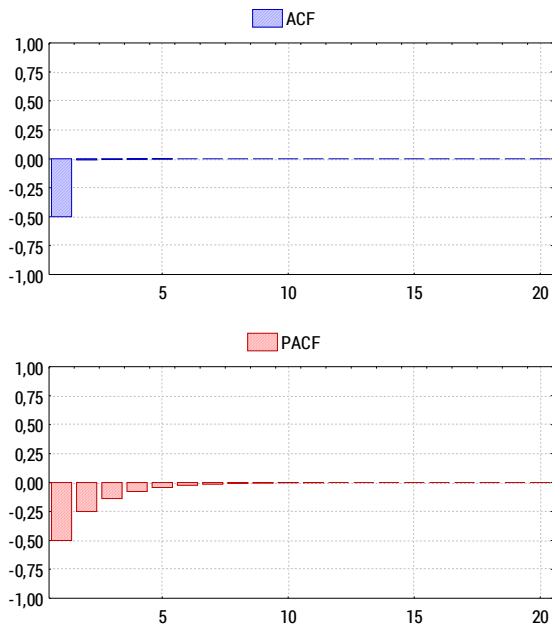
Rys. 3.3. Przebieg funkcji autokorelacji i autokorelacji cząstkowej procesu AR(1) $\phi_1 < 0$

Źródło: opracowanie własne na podstawie: S.A. DeLurgio (1998), *Forecasting Principles and Applications*, Irwin/McGraw-Hill, s. 277.

Dla procesu średniej ruchomej MA(1) teoretyczna autokorelacja, z wyjątkiem opóźnienia równego 1, jest zerowa. Jeżeli $\theta_1 > 0$, ta pojedyncza znacząca wartość jest ujemna, zaś dla $\theta_1 < 0$ dodatnia. PACF dla procesu MA(1) zachowuje się podobnie do ACF modeli AR(1). PACF opada wykładniczo i ma wartości ujemne, jeżeli $\theta_1 > 0$, lub jeżeli $\theta_1 < 0$ znaki wartości zmieniają się na przemian (rys. 3.4 i 3.5)¹⁰⁶.

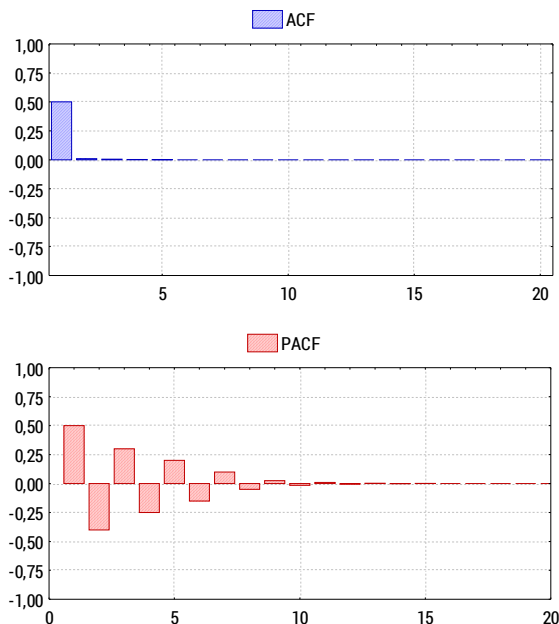
¹⁰⁵ S.A. DeLurgio (1998), *Forecasting Principles and...*, op. cit., s. 276.

¹⁰⁶ Ibidem, s. 278.



Rys. 3.4. Przebieg funkcji autokorelacji i autokorelacji cząstkowej procesu MA(1) $\theta_1 < 0$

Źródło: opracowanie własne na podstawie: S.A. DeLurgio (1998), *Forecasting Principles and Applications*, Irwin/McGraw-Hill, s. 277.

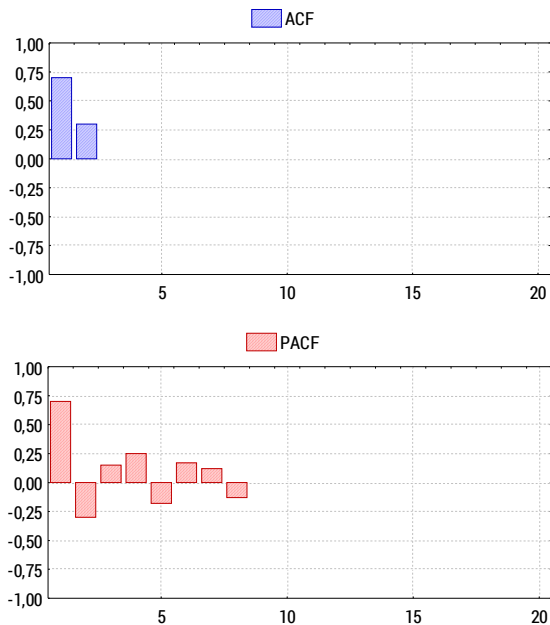


Rys. 3.5. Przebieg funkcji autokorelacji i autokorelacji cząstkowej procesu MA(1) $\theta_1 < 0$

Źródło: opracowanie własne na podstawie: S.A. DeLurgio (1998), *Forecasting Principles and Applications*, Irwin/McGraw-Hill, s. 277.

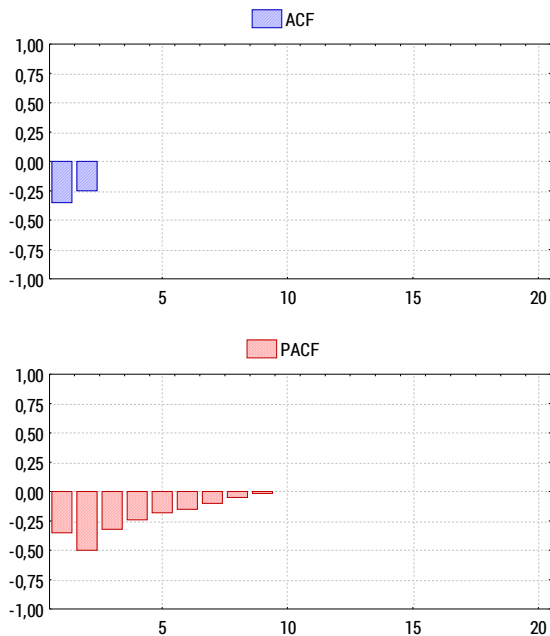
Charakterystyki przebiegów funkcji ACF i PACF dla procesów drugiego rzędu przedstawiono poniżej.

W wypadku modelu MA(2) funkcja autokorelacji ma duże wartości dla opóźnień 1 i 2, a funkcja autokorelacji cząstkowej wykładniczo lub sinusoidalnie wykładniczo zanika. Na rys. 3.6–3.9 pokazano cztery, z wielu możliwych kształtów funkcji AC i PACF procesów średniej ruchomej.



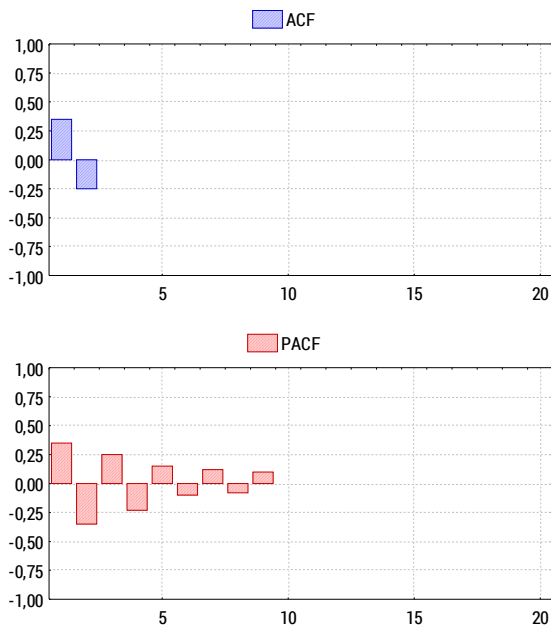
Rys. 3.6. Przebieg funkcji autokorelacji i autokorelacji cząstkowej procesu MA(2) $\theta_1 < 0$, $\theta_2 < 0$

Źródło: opracowanie własne na podstawie: S.A. DeLurgio (1998), *Forecasting Principles and Applications*, Irwin/McGraw-Hill, s. 307.



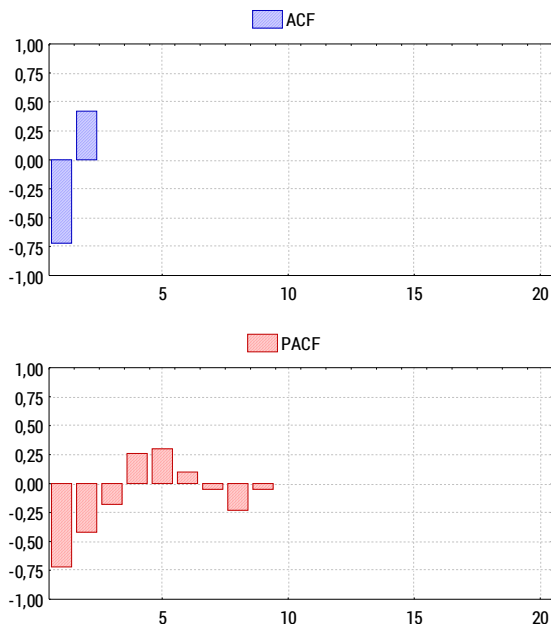
Rys. 3.7. Przebieg funkcji autokorelacji i autokorelacji cząstkowej procesu MA(2) $\theta_1 < 0$, $\theta_2 > 0$

Źródło: opracowanie własne na podstawie: S.A. DeLurgio (1998), *Forecasting Principles and Applications*, Irwin/McGraw-Hill, s. 307.



Rys. 3.8. Przebieg funkcji autokorelacji i autokorelacji cząstkowej procesu MA(2) $\theta_1 < 0$, $\theta_2 > 0$

Źródło: opracowanie własne na podstawie: S.A. DeLurgio (1998), *Forecasting Principles and Applications*, Irwin/McGraw-Hill, 1998, s. 307.

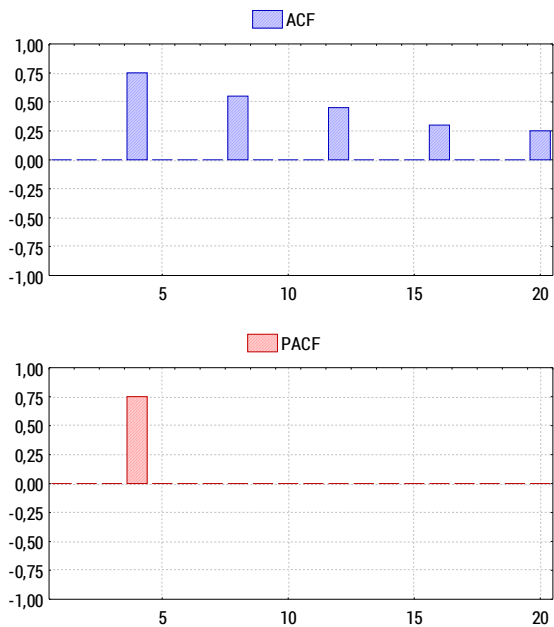


Rys. 3.9. Przebieg funkcji autokorelacji i autokorelacji cząstkowej procesu MA(2) $\theta_1 > 0$, $\theta_2 < 0$

Źródło: opracowanie własne na podstawie: S.A. DeLurgio (1998), *Forecasting Principles and Applications*, Irwin/McGraw-Hill, s. 307.

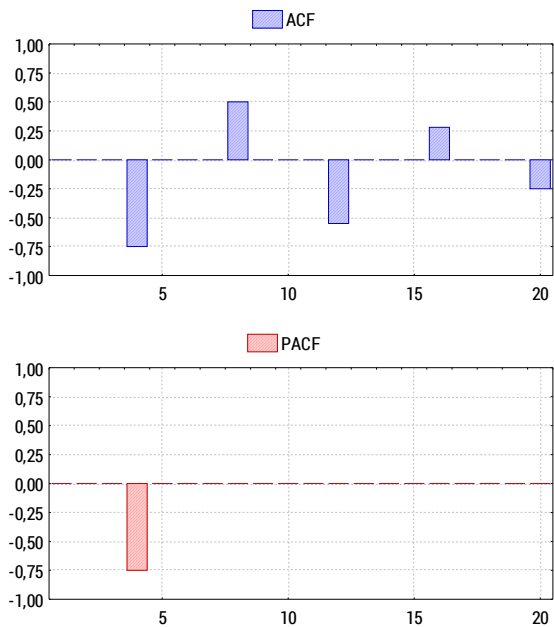
Przebieg funkcji autokorelacji dla procesów AR(2) jest lustrzanym odbiciem kształtu funkcji autokorelacji cząstkowej procesów MA(2). Z kolei modele ARMA(1,1) często posiadają taki kształt funkcji autokorelacji jak procesy AR(1), zaś autokorelacji cząstkowej jak MA(1).

W przypadku sezonowych modeli ARIMA charakterystyczne kształty ACF i PACF z niesezonowych modeli AR i MA pojawiają się tylko dla opóźnień będących wielokrotnością s . Na rys. 3.10–3.13 zilustrowano cztery, z wielu możliwych teoretycznych przebiegów ACF i PACF dla danych wykazujących sezonowość kwartalną, tzn. $s = 4$.



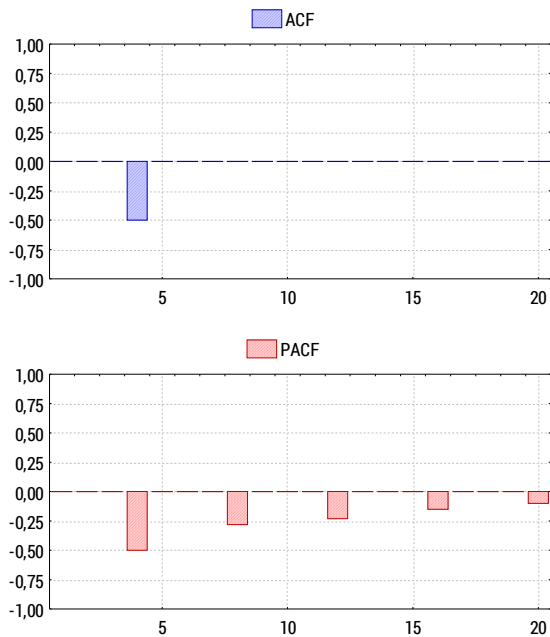
Rys. 3.10. Przebieg funkcji autokorelacji i autokorelacji cząstkowej procesu $ARIMA(1,0,0)_4$ $\phi_1 = 0,75$

Źródło: opracowanie własne na podstawie: S.A. DeLurgio (1998), *Forecasting Principles and Applications*, Irwin/McGraw-Hill, s. 309.



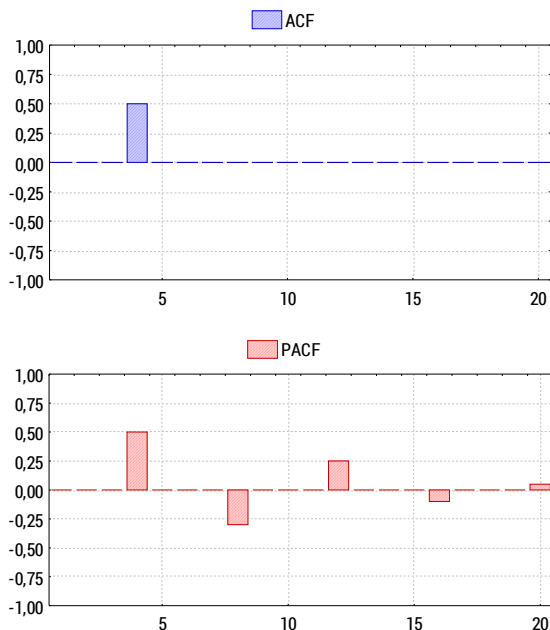
Rys. 3.11. Przebieg funkcji autokorelacji i autokorelacji cząstkowej procesu $ARIMA(1,0,0)_4$ $\phi_1 = -0,75$

Źródło: opracowanie własne na podstawie: S.A. DeLurgio (1998), *Forecasting Principles and Applications*, Irwin/McGraw-Hill, s. 309.



Rys. 3.12. Przebieg funkcji autokorelacji i autokorelacji cząstkowej procesu $ARIMA(0,0,1)_4$ $\theta_1 = 0,5$

Źródło: opracowanie własne na podstawie: S.A. DeLurgio (1998), *Forecasting Principles and Applications*, Irwin/McGraw-Hill, s. 309.



Rys. 3.13. Przebieg funkcji autokorelacji i autokorelacji cząstkowej procesu $ARIMA(0,0,1)_4$ $\theta_1 = -0,5$

Źródło: opracowanie własne na podstawie: S.A. DeLurgio (1998), *Forecasting Principles and Applications*, Irwin/McGraw-Hill, s. 309.

Jeżeli w procesie istnieje zarówno składnik sezonowy, jak i niesezonowy, dość trudno odseparować je od siebie i jednoznacznie określić model. Ogólnie, wartości autokorelacji rzeczywistych szeregów czasowych bardzo rzadko pasują idealnie do opisu teoretycznych przebiegów funkcji autokorelacji procesów AR czy MA¹⁰⁷. W praktyce bardzo często wypróbowuje się kilka różnych modeli i dopiero po etapie estymacji i sprawdzania

¹⁰⁷ J. Nazarko, M. Rybaczuk, A. Jurczuk (2005), *Wpływ poziomu zakłóceń losowych na możliwość identyfikacji modeli ARIMA*, „Taksonomia” 12, s. 238–247.

diagnostycznego podejmuje się ostateczną decyzję¹⁰⁸. Umożliwiają to pakiety programów, takich jak: R¹⁰⁹, Matlab¹¹⁰ czy SAS¹¹¹.

3.3. Estymacja parametrów modeli ARIMA

Otrzymanie efektywnych oszacowań parametrów jest kolejnym etapem budowy modelu ARIMA. Estymacja parametrów modelu ARIMA jest zadaniem dość trudnym. Parametry modeli AR mogą być szacowane metodą najmniejszych kwadratów. W pozostałych modelach (tzn. gdy $q > 0$) trzeba stosować procedury iteracyjne minimalizujące sumę kwadratów reszt. Często wykorzystuje się też metody największej wiarygodności¹¹². Należy pamiętać o ograniczeniach wartości parametrów procesów, które wynikają z konieczności zapewniania stacjonarności i odwracalności modeli (o czym była mowa w podrozdziale 3.1).

W celu określenia wartości początkowych parametrów dla stosowanych metod iteracyjnych można wykorzystać funkcję autokorelacji i autokorelacji cząstkowej¹¹³.

¹⁰⁸ S.A. DeLurgio (1998), *Forecasting Principles and...*, op. cit., s. 305.

¹⁰⁹ R. Hyndman, Y. Khandakar (2008), *Automatic time series forecasting: the forecast package for R*, „Journal of Statistical Software” 26(3), s. 1–22, doi: 10.18637/jss.v027.i03; T.M. Awan, F. Aslam (2020), *Prediction of daily COVID-19 cases in European countries using automatic ARIMA model*, „Journal of Public Health Research” 9(3), 1765, s. 227–233, doi: 10.4081/jphr.2020.1765.

¹¹⁰ M.A. Villegas, D.J. Pedregal (2019), *Automatic selection of unobserved components models for supply chain forecasting*, „International Journal of Forecasting” 35(1), s. 157–169, doi: 10.1016/j.ijforecast.2017.11.001.

¹¹¹ M.A. Villegas, D.J. Pedregal (2018), *SSpace: A Toolbox for State Space Modeling*, „Journal of Statistical Software” 87(1), s. 1–26, doi: 10.18637/jss.v087.i05.

¹¹² S. Makridakis, S.C. Wheelwright, R.J. Hyndman (1998), *Forecasting. Methods and...*, op. cit., s. 358–359.

¹¹³ G.E.P. Box, G.M. Jenkins (1983), *Analiza szeregów...*, op. cit., s. 175.

Dla procesu średniej ruchomej MA(q) funkcję autokorelacji można wyrazić jako¹¹⁴:

$$\rho_k = \frac{-\theta_k + \theta_1\theta_{k+1} + \dots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2}, \quad k = 1, 2, \dots, q. \quad (3.48)$$

Podstawiając w miejsce ρ_k ich estymatory r_k i rozwiązując q nieliniowych równań otrzymuje się wstępne oszacowania parametrów θ_k . W przypadku ogólnym równania (3.48) mają wiele rozwiązań, jednak tylko jedno spełnia warunek odwracalności¹¹⁵.

Przy założeniu, że badany szereg jest procesem AR oszacowania początkowe parametrów można otrzymać, wykorzystując równania autokorelacji Yule'a-Walkera¹¹⁶:

$$\boldsymbol{\phi} = \mathbf{P}_p^{-1} \boldsymbol{\rho}_p, \quad (3.49)$$

gdzie:

$$\boldsymbol{\phi} = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix}, \mathbf{P}_p = \begin{bmatrix} 1 & \rho_1 & \dots & \rho_{p-1} \\ \rho_1 & 1 & \dots & \rho_{p-2} \\ & \dots & \dots & \dots \\ \rho_{p-1} & \rho_{p-2} & \dots & 1 \end{bmatrix}, \boldsymbol{\rho}_p = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix}.$$

Zamieniając teoretyczne wartości autokorelacji ρ_k na autokorelacje próbkowe r_k otrzymuje się oszacowania¹¹⁷:

$$\hat{\boldsymbol{\phi}} = \mathbf{R}_p^{-1} \mathbf{r}_p, \quad (3.50)$$

gdzie:

$$\mathbf{R}_p = \begin{bmatrix} 1 & r_1 & \dots & r_{p-1} \\ r_1 & 1 & \dots & r_{p-2} \\ & \dots & \dots & \dots \\ r_{p-1} & r_{p-2} & \dots & 1 \end{bmatrix}, \mathbf{r}_p = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_p \end{bmatrix}.$$

Ogólny algorytm otrzymywania oszacowań początkowych parametrów dowolnego procesu ARMA(p, q) (tzn. od (0,1) i (1,0)

¹¹⁴ Ibidem, s. 187.

¹¹⁵ Ibidem, s. 188.

¹¹⁶ Ibidem, s. 63.

¹¹⁷ Ibidem, s. 189.

do (p, q) włącznie) składa się z oszacowania parametrów autoregresji na podstawie autokowariancji c_k ¹¹⁸. Gdy $p > 0$, rozwiązuje się układ p równań liniowych¹¹⁹:

$$A\hat{\phi} = x, \quad (3.51)$$

gdzie:

$$A_{ij} = c_{|q+i-j|}, x_i = c_{q+i}, \quad i, j = 1, 2, \dots, p,$$

$$c_k = \frac{1}{n} \sum_{t=k+1}^n (Y_t - \bar{Y})(Y_{t-k} - \bar{Y}), \quad k = 0, 1, \dots, K, \quad K \geq p + q.$$

Następnie, korzystając z otrzymanych oszacowań $\hat{\phi}$, oblicza się ciąg zmodyfikowanych kowariancji c'_j ¹²⁰:

$$c'_j = \begin{cases} \sum_{i=0}^p \sum_{k=0}^p \hat{\phi}_i \hat{\phi}_k c_{|j+i-k|}, & p > 0 \\ c_j, & p = 0, \end{cases} \quad (3.52)$$

dla $\hat{\phi}_0 = -1$, $j = 0, 1, \dots, q$.

Oszacowania początkowe parametrów procesu średniej ruchomej można otrzymać, wykorzystując procedurę iteracyjną Newtona-Raphsona¹²¹. Obliczając wektor τ^{i+1} w $i + 1$ iteracji na podstawie jego wartości τ^i w i -tej iteracji¹²²:

$$\tau^{i+1} = \tau^i - (T^i)^{-1} f^i, \quad (3.53)$$

gdzie:

$$\tau' = [\tau_0 \quad \tau_1 \quad \dots \quad \tau_q], f_j = \sum_{i=0}^{q-j} \tau_i \tau_{i+j} - c'_j, f' = [f_0 \quad f_1 \quad \dots \quad f_q],$$

$$T = \begin{bmatrix} \tau_0 & \tau_1 & \dots & \tau_{q-1} & \tau_q \\ \tau_1 & \tau_2 & \dots & \tau_q & 0 \\ & & \dots & & \\ \tau_q & 0 & \dots & 0 & 0 \end{bmatrix} + \begin{bmatrix} \tau_0 & \tau_1 & \tau_2 & \dots & \tau_q \\ 0 & \tau_0 & \tau_1 & \dots & \tau_{q-1} \\ & & \dots & & \\ 0 & 0 & 0 & \dots & \tau_0 \end{bmatrix}$$

z wartościami początkowymi: $\tau_0 = \sqrt{c'_0}$, $\tau_1 = \tau_2 = \dots = \tau_q = 0$.

¹¹⁸ Ibidem, s. 202.

¹¹⁹ Ibidem, s. 493.

¹²⁰ Ibidem, s. 494.

¹²¹ Ibidem, s. 203.

¹²² Ibidem, s. 494.

W przypadku, gdy $|f_j| < \varepsilon$, $j = 0, 1, \dots, q$ dla zadanej wartości ε procedurę iteracyjną uważa się za zakończoną, a oszacowania parametrów otrzymuje się z końcowej wartości τ ¹²³:

$$\hat{\theta}_j = -\frac{\tau_j}{\tau_0}, \quad j = 1, 2, \dots, q. \quad (3.54)$$

W modelach ze stałą, jej początkową wartość można wyliczyć, korzystając ze wzoru¹²⁴:

$$\hat{\theta}_0 = \begin{cases} \bar{Y}(1 - \sum_{i=1}^p \hat{\phi}_i), & p > 0 \\ \bar{Y}, & p = 0, \end{cases} \quad (3.55)$$

gdzie: $\bar{Y}$ – średnia szeregu stacjonarnego.

W tabeli 3.4 zamieszczono wstępne oszacowania parametrów pięciu podstawowych modeli ARIMA.

Tabela 3.4. Przybliżone oszacowania parametrów procesów średniej ruchomej i procesów autoregresji rzędu pierwszego i drugiego

Model	ARIMA(1,d,0)	ARIMA(0,d,1)
Wstępne oszacowania	$\hat{\phi}_1 = r_1$	$r_1 = \frac{-\hat{\theta}_1}{1 + \hat{\theta}_1^2}$
Model	ARIMA(2,d,0)	ARIMA(0,d,2)
Wstępne oszacowania	$\hat{\phi}_1 = \frac{r_1(1 - r_2)}{1 - r_1^2}$ $\hat{\phi}_2 = \frac{r_2 - r_1^2}{1 - r_1^2}$	$r_1 = \frac{-\hat{\theta}_1(1 - \hat{\theta}_2)}{1 + \hat{\theta}_1^2 + \hat{\theta}_2^2}$ $r_2 = \frac{-\hat{\theta}_2}{1 + \hat{\theta}_1^2 + \hat{\theta}_2^2}$
Model	ARIMA(1,d,1)	
Wstępne oszacowania	$r_1 = \frac{(1 - \hat{\theta}_1\phi_1)(\hat{\phi}_1 - \hat{\theta}_1)}{1 + \hat{\theta}_1^2 - 2\hat{\phi}_1\theta_1}$ $r_2 = r_1\hat{\phi}_1$	

Źródło: G.E.P. Box, G.M. Jenkins (1983), *Analiza szeregów czasowych: prognozowanie i sterowanie*, Państwowe Wydawnictwo Naukowe, Warszawa, s. 178.

¹²³ Ibidem, s. 494.

¹²⁴ Ibidem, s. 494.

Oszacowane wstępne wartości parametrów są podstawą ustalenia wartości optymalnych w procedurach iteracyjnych¹²⁵. Problem estymacji parametrów jest tematem wielu publikacji naukowych (poruszają go między innymi: Caines i Rissan¹²⁶, Gardner Harvey i Phillips¹²⁷, Lindquist¹²⁸). Różne podejścia, jako wstęp do prezentacji własnego rozwiązania, cytują Mélard, Roch i Saidi¹²⁹. Główną trudnością jest określenie początkowych wartości (szczególnie reszt w modelach MA, gdyż w wypadku AR można ewentualnie skrócić szereg). Iteracyjny algorytm można odnaleźć w pracy Cieślak¹³⁰, gdzie przedstawiony jest algorytm Marquardta¹³¹. W programie Statistica, który wykorzystano w kolejnym rozdziale do zaprezentowania prognozowania opartego na modelach ARIMA zaimplementowane są metody: przybliżona Mcleoda i Salesa¹³² oraz dokładna Mélarda¹³³.

¹²⁵ M. Cieślak (red.) (2011), *Prognozowanie gospodarcze...*, op. cit., s. 122.

¹²⁶ P.E. Caines, J. Rissanen (1974), *Maximum likelihood estimation of parameters in multivariate Gaussian stochastic processes*, „IEEE Transactions on Information Theory” 20, s. 102–104, doi: 10.1109/TIT.1974.1055155.

¹²⁷ G. Gardner, A.C. Harvey, G.D.A. Phillips (1980), *An algorithm for exact maximum likelihood estimation of autoregressive-moving average models by means of Kalman filtering*, „Journal of the Royal Statistical Society” C, 29, s. 311–322, doi: 10.2307/2346910.

¹²⁸ A. Lindquist (1974), *A new algorithm for optimal filtering of discrete time stationary processes*, „SIAM Journal on Control and Optimization” 12, s. 736–746, doi: 10.1137/0312057.

¹²⁹ G. Mélard, R. Roch, A. Saidi (2006), *Exact maximum likelihood estimation of structured or unit root multivariate time series models*, „Computational Statistics & Data Analysis” 50(11), s. 2958–2986, doi: 10.1016/j.csda.2005.06.009.

¹³⁰ M. Cieślak (red.) (2011), *Prognozowanie gospodarcze...*, op. cit., s. 129.

¹³¹ D.W. Marquardt (1963), *An Algorithm for Least-Squares Estimation of Nonlinear Parameters*, „Journal of the Society for Industrial and Applied Mathematics” 11(2), s. 431–441, doi: 10.1137/0111030.

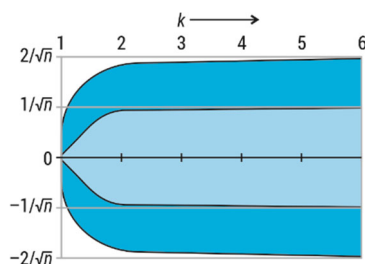
¹³² A.I. Mcleod, P.R.H. Sales (1983), *An Algorithm for Approximate Likelihood Calculation of Arma and Seasonal Arma Models*, „Journal of the Royal Statistical Society: Series C (Applied Statistics)” 32(2), s. 211–223, doi: 10.2307/2347301.

¹³³ G. Mélard (1984), *A fast algorithm for the exact likelihood of autoregressive-moving average models*, „Journal of the Royal Statistical Society: Series C (Applied Statistics)” 33(1), s. 104–114, doi: 10.2307/2347672.

Diagnostyka zbudowanych modeli ARIMA może opierać się na analizie reszt. Jeżeli model został dobrze dobrany, to wartości reszt modelu $\hat{e}_t$ wraz ze wzrostem długości szeregu powinny zbliżać się do białego szumu e_t . Badanie reszt modelu może więc wykryć istnienie i wskazać charakter nieadekwatności modelu. Szczególnie, dostrzeżony kształt estymowanej funkcji autokorelacji reszt $\hat{e}_t$ może wskazywać na konieczność przeprowadzenia odpowiedniej modyfikacji modelu¹³⁴.

Jeżeli znane są prawdziwe wartości parametrów ϕ i θ , estymowane autokorelacje szeregu e_t będą nieskorelowane i będą posiadać w przybliżeniu rozkład normalny o średniej równej 0 i błędzie standardowym $n^{-1/2}$ (n – liczba obserwacji w szeregu). W praktyce nie są znane prawdziwe wartości parametrów, a jedynie estymatory $(\hat{\phi}, \hat{\theta})$, na podstawie których można obliczyć $\hat{e}_t$. Dlatego też wykorzystanie $n^{-1/2}$ jako błędu standardowego dla $\hat{e}_t$ może prowadzić do niedocenienia istotności odchyłeń autokorelacji od zera przy małych odstępach, ale można stosować je dla średnich i dużych odstępów¹³⁵.

Na rys. 3.14 pokazano przedziały jednego i dwóch błędów standardowych dla dużej próbki w przypadku procesu autoregresji oraz średniej ruchomej rzędu pierwszego.



Rys. 3.14. Granice błędów standardowych dla autokorelacji reszt procesu AR(1) $\phi_1 = 0,3$ oraz MA(1) $\theta_1 = 0,3$

Źródło: G.E.P. Box, G.M. Jenkins (1983), *Analiza szeregów czasowych: prognozowanie i sterowanie*, Państwowe Wydawnictwo Naukowe, Warszawa, s. 292.

¹³⁴ G.E.P. Box, G.M. Jenkins (1983), *Analiza szeregów...*, op. cit., s. 290.

¹³⁵ Ibidem, s. 291.

W praktyce, do oceny autokorelacji określa się poziom prawdopodobieństwa, z którym badane autokorelacje reszt powinny się znaleźć pomiędzy wartościami granicznymi $r_{krytyczne} = \pm z/\sqrt{n}$ (z – wartość krytyczna standaryzowanego rozkładu normalnego). Przedziały te narysowane na korelogramach pozwalają szybko rozpoznać sytuację, w której szereg reszt różni się od białego szumu.

W wyborze najlepszego modelu spośród kilku alternatywnych z różną liczbą parametrów może pomóc też graficzna prezentacja wartości AIC (ang. *Akaike's Information Criterion*). Wielkość ta definiowana jest jako¹³⁶:

$$AIC = -2 \ln \hat{L} + 2m, \quad (3.56)$$

gdzie:

m – liczba estymowanych parametrów ($m = p + q + P + Q$),

$\hat{L}$ – funkcja największej wiarygodności.

Wskaźnik AIC dostarcza informacji o jakości dopasowania różnych modeli dla tych samych danych, biorąc pod uwagę liczbę parametrów modelu, tzw. „karę” (ang. *penalty*). W przypadku, gdy nie jest znana wartość funkcji $\hat{L}$, wygodnym sposobem na obliczenie AIC jest aproksymowanie jej wartości zgodnie ze wzorem¹³⁷:

$$AIC \approx n(1 + \ln(2\pi)) + n \ln \hat{\sigma}^2 + 2m, \quad (3.57)$$

gdzie:

$\hat{\sigma}^2$ – wariancja reszt modelu,

n – liczba obserwacji w szeregu.

Ponieważ dla tych samych danych: $n(1 + \ln(2\pi))$ ma identyczną wartość dla wszystkich modeli, ten człon równania często jest pomijany¹³⁸. Niekiedy wariancję reszt we wzorze (3.57) zastępuje się sumą kwadratów reszt¹³⁹. Na rys. 3.15 pokazano

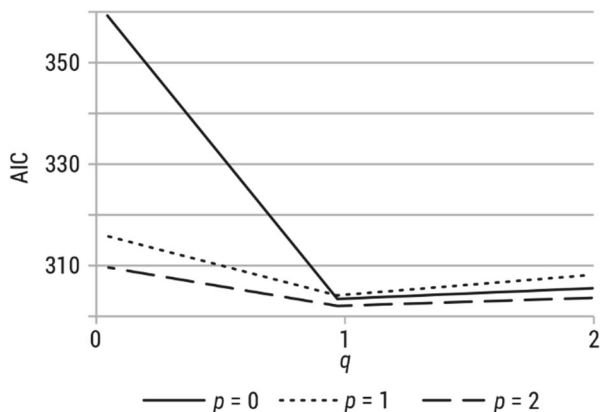
¹³⁶ S. Makridakis, S.C. Wheelwright, R.J. Hyndman (1998), *Forecasting. Methods and...*, op. cit., s. 361.

¹³⁷ Ibidem.

¹³⁸ Ibidem.

¹³⁹ S.A. DeLurgio (1998), *Forecasting Principles and...*, op. cit., s. 735.

hipotetyczny przebieg takiej funkcji zbudowanej z wartości AIC dla modelu, w którym występuje zarówno proces średniej ruchomej $MA(q)$ oraz autokoregresji $AR(p)$, a $0 \leq p, q \leq 2$.



Rys. 3.15. Wykres wartości AIC dla hipotetycznych modeli, w których $0 \leq p, q \leq 2$

Źródło: R.M. Heiberger, P. Teles (2002), *Displays for Direct Comparison of ARIMA Models*, „The American Statistician” 56(2), s. 134.

Wartość AIC interpretuje się tylko poprzez zestawienie jej z wartościami tego współczynnika dla innych modeli dla tych samych danych. Najlepszy model to taki, który ma najmniejszą wartość AIC . Z tym, że różnica w wielkości 2 lub mniejsza nie ma znaczenia – można wybrać model bądź prostszy, bądź lepiej dopasowany do danych¹⁴⁰. Przedstawione na rys. 3.15 wartości AIC modelu z $q = 1$ są mniejsze niż w przypadku, gdy $q = 2$. Dodatkowo, gdy $q = 1$, AIC jest niemal identyczna, niezależnie od złożoności procesu autokoregresji. Wnioskować więc można, że w tym hipotetycznym modelu tylko jeden parametr średniej ruchomej jest niezbędny.

¹⁴⁰ S. Makridakis, S.C. Wheelwright, R.J. Hyndman (1998), *Forecasting. Methods and...*, op. cit., s. 362.

Wykorzystywane jest też Bayesowskie kryterium informacyjne Schwartza (BIC, ang. *Bayesian Information Criterion*, SBC ang. *Schwartz Bayesian Criterion*)¹⁴¹:

$$BIC = m \ln(n) - 2\ln(\hat{L}). \quad (3.58)$$

Kryterium porównania różnych modeli ARIMA mogą być też błędy: średnia wartość modułu błędu (błąd przeciętny, ang. *mean absolute error*), średniokwadratowy błąd (ang. *root mean squared error*) czy przeciętny procentowy błąd (ang. *mean absolute percentage error*)¹⁴², które zostały omówione w poprzednich częściach podręcznika *Prognozowanie w zarządzaniu przedsiębiorstwem*. W ocenie modelu ARIMA szczególnie często wykorzystuje się standardowy błąd oceny modelu:

$$\hat{\sigma}_k = \sqrt{\frac{1}{n-p-q} \sum_{t=1}^n \hat{e}_t^2}, \quad (3.59)$$

gdzie: $\hat{e}_t$ – reszty modelu.

Błąd ten jest też wykorzystywany do wyznaczenia wariancji oceny parametrów i oceny ich istotności. Ogólnie, im mniejsza wartość błędów, tym lepsze dopasowanie modelu do danych.

W celu sprawdzenia losowości reszt modelu można zastosować test statystyczny zaproponowany przez Ljunga-Boxa, będący rozszerzeniem statystyki Boxa Pierce'a, omówiony w rozdziale pierwszym. Weryfikowana jest hipoteza H_0 : reszty są niezależne, w stosunku do H_1 : reszty nie są niezależne. W wypadku reszt statystyka ma wartość¹⁴³:

¹⁴¹ K.P. Burnham, D.R. Anderson (2004), *Multimodel Inference: Understanding AIC and BIC in Model Selection*. „Sociological Methods Research” 33(2), s. 261-304, doi: 10.1177/0049124104268644.

¹⁴² R.J. Hyndman, G. Athanasopoulos (2018), *Forecasting: principles and practice*, 2nd ed., OTexts, Melbourne, Australia. Section 3.4 Evaluating forecast accuracy, <https://otexts.org/fpp2/accuracy.html>.

¹⁴³ D. Witkowska, D. Żebrowska-Suchodolska (2008), *Badanie efektywności GPW na przykładzie wybranych indeksów: test autokorelacji*, „Oeconomia” 7(4), s. 155-162.

$$Q_k = n(n+2) \sum_{k=1}^K \frac{\hat{\varrho}_k^2}{n-k}, \quad (3.60)$$

gdzie: $\hat{\varrho}_k^2 = \frac{\sum_{t=k+1}^n \hat{e}_t^2 \hat{e}_{t-k}^2}{\sum_{t=k+1}^n \hat{e}_t^2}$, K - liczba uwzględnianych autokorelacji.

Przypominając, wartość krytyczna pochodzi z asymptotycznego rozkładu χ^2 z $K-p-q$ stopniami swobody. W przypadku $Q_k \leq \chi_\alpha^2$ nie ma podstaw do odrzucenia hipotezy zerowej przy zadanym prawdopodobieństwie α .

Jeżeli model jest poprawny można zastosować go do prognozowania. Prognozą realizacji procesu stochastycznego ARMA(p, q) w okresie $t+i$ jest¹⁴⁴:

$$Y_{t+i}^* = E \left[\hat{\phi}_1 Y_{t-1+i} + \dots + \hat{\phi}_p Y_{t-p+i} + e_t - \hat{\theta}_1 e_{t-1+i} - \dots - \hat{\theta}_q e_{t-q+i} \right]. \quad (3.61)$$

Prognozowanie na podstawie modelu ARIMA polega na podstawieniu do zbudowanego modelu odpowiednich wartości za Y_{t-1} , Y_{t-2} , ..., Y_{t-p} oraz e_{t-1} , e_{t-2} , ..., e_{t-q} . Prognozy wyznacza się rekurencyjnie.

Kluczowe zagadnienia:

model autoregresji, model średniej ruchomej,
model autoregresji i średniej ruchomej,
zintegrowany model autoregresji i średniej ruchomej

Literatura podstawowa

- [1] Box G.E.P., Jenkins G.M. (1998), *Analiza szeregów czasowych. Prognozowanie i sterowanie*, PWN, Warszawa
- [2] Cieślak M. (red.) (2011), *Prognozowanie gospodarcze. Metody i zastosowania*, Wydanie czwarte zmienione, Wydawnictwo Naukowe PWN, Warszawa

¹⁴⁴ M. Cieślak (red.) (2011), *Prognozowanie gospodarcze...*, op. cit., s. 124.

- [3] DeLurgio S.A. (1998), *Forecasting Principles and Applications*, Irwin/McGraw-Hill, Boston

Literatura uzupełniająca

- [1] Awan T.M., Aslam F. (2020), *Prediction of daily COVID-19 cases in European countries using automatic ARIMA model*, „Journal of Public Health Research” 9(3), 1765, s. 227–233, doi: 10.4081/jphr.2020.1765
- [2] Box G.E.P., Jenkins G.M. (1976), *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day
- [3] Burnham K.P., Anderson D.R. (2004), *Multimodel Inference: Understanding AIC and BIC in Model Selection*, „Sociological Methods Research” 33(2), s. 261–304, doi: 10.1177/0049124104268644
- [4] Caines P.E., Rissanen J. (1974), *Maximum likelihood estimation of parameters in multivariate Gaussian stochastic processes*, „IEEE Transactions on Information Theory” 20, s. 102–104, doi: 10.1109/TIT.1974.1055155
- [5] Chodakowska E., Nazarko J., Nazarko Ł. (2021), *ARIMA Models in Electrical Load Forecasting and Their Robustness to Noise*, „Energies” 14(23), 7952, s. 1–22, doi: 10.3390/en14237952
- [6] Contreras J., Espínola R., Nogales F.J., Conejo A.J. (2003) *ARIMA Models to Predict Next-Day Electricity Prices*, „IEEE Transactions on Power Systems” 18(3), s. 1014–1020, doi: 10.1109/TPWRS.2002.804943
- [7] Gardner G., Harvey A.C., Phillips G.D.A. (1980), *An algorithm for exact maximum likelihood estimation of autoregressive-moving average models by means of Kalman filtering*, „Journal of the Royal Statistical Society” C, 29, s. 311–322, doi: 10.2307/2346910
- [8] Heiberger R.M., Teles P. (2002), *Displays for Direct Comparison of ARIMA Models*, „The American Statistician” 56(2)
- [9] Hyndman R., Khandakar Y. (2008), *Automatic time series forecasting: the forecast package for R*, „Journal of Statistical Software” 26(3), s. 1–22, doi: 10.18637/jss.v027.i03
- [10] Hyndman R.J., Athanasopoulos G. (2018), *Forecasting: principles and practice*, 2nd ed., OTexts, Melbourne, Australia. Section 3.4 Evaluating forecast accuracy, <https://otexts.org/fpp2/accuracy.html>

- [11] Lindquist A. (1974), *A new algorithm for optimal filtering of discrete time stationary processes*, „SIAM Journal on Control and Optimization” 12, s. 736–746, doi: 10.1137/0312057
- [12] Makridakis S., Wheelwright S.C., Hyndman R.J. (1998), *Forecasting. Methods and Applications*, John Wiley & Sons, Inc.
- [13] Marquardt D.W. (1963), *An Algorithm for Least-Squares Estimation of Nonlinear Parameters*, „Journal of the Society for Industrial and Applied Mathematics” 11(2), s. 431–441, doi: 10.1137/0111030
- [14] Mastrangelo C.M., Simpson J.R., Montgomery D.C. (2013), *Time Series Analysis*, in: Gass S.I., Fu M.C. (red.) *Encyclopedia of Operations Research and Management Science*. Springer, Boston, MA, doi: 10.1007/978-1-4419-1153-7_1045
- [15] MathWorks. *Help Center*, <https://www.mathworks.com/help/econ/how-to-specify-arima-models.html> [29.11.2021]
- [16] Mcleod A.I., Sales P.R.H. (1983), *An Algorithm for Approximate Likelihood Calculation of Arma and Seasonal Arma Models*, „Journal of the Royal Statistical Society: Series C (Applied Statistics)” 32(2), s. 211–223, doi: 10.2307/2347301
- [17] Mélard G. (1984), *A fast algorithm for the exact likelihood of autoregressive-moving average models*, „Journal of the Royal Statistical Society: Series C (Applied Statistics)” 33(1), s. 104–114, doi: 10.2307/2347672
- [18] Mélard G., Roch R., Saidi A. (2006), *Exact maximum likelihood estimation of structured or unit root multivariate time series models*, „Computational Statistics & Data Analysis” 50(11), s. 2958–2986, doi: 10.1016/j.csda.2005.06.009
- [19] Nazarko J., Chraołowska J., Rybaczuk M. (2004), *Zastosowanie wielosezonowego modelu ARIMA w prognozowaniu obciążeń mocą elektryczną*, „Taksonomia” 11, s. 173–182
- [20] Nazarko J., Rybaczuk M., Jurczuk A. (2005), *Wpływ poziomu zakłóceń losowych na możliwość identyfikacji modeli ARIMA*, „Taksonomia” 12, s. 238–247
- [21] *SAS/ETS 13.2 User's Guide. The ARIMA Procedure* (2014), SAS Institute Inc., Cary, NC, USA, <https://support.sas.com/documentation/onlinedoc/ets/132/arima.pdf> [29.11.2021]
- [22] Villegas M.A., Pedregal D.J. (2019), *Automatic selection of unobserved components models for supply chain forecasting*, „International

- Journal of Forecasting”, 35(1), s. 157–169, doi: 10.1016/j.ijforecast.2017.11.001
- [23] Villegas M.A., Pedregal D.J., (2018), *SSpace: A Toolbox for State Space Modeling*, „Journal of Statistical Software” 87(1), s. 1–26, doi: 10.18637/jss.v087.i05
 - [24] Witkowska D. (2012), *Podstawy ekonometrii i teorii prognozowania*, Wolters Kluwer Polska SA
 - [25] Witkowska D., Żebrowska-Suchodolska D. (2008), *Badanie efektywności GPW na przykładzie wybranych indeksów: test autokorelacji*, „Oeconomia” 7(4), s. 155–162
 - [26] Yang Z., Ce L., Lian L. (2017), *Electricity price forecasting by a hybrid model, combining wavelet transform, ARMA and kernel-based extreme learning machine methods*, „Applied Energy” 190, s. 291–305, doi: 10.1016/j.apenergy.2016.12.130
 - [27] Zalewski W., Nazarko J., Jurczuk A. (2005), *ARIMA models in load modelling with clustering approach*, PowerTech IEEE'2005, St. Petersburg, June 27–30

4. PROGNOZOWANIE OPARTE NA MODELACH ARIMA

4.1. Wprowadzenie do modelu prognozy

Przedstawione w poprzednich rozdziałach modele klasy ARIMA mogą być w efektywny sposób zastosowane do prognozowania stacjonarnych szeregów czasowych, w których występują kombinacje różnych składowych (wahania przypadkowe, stały średni poziom, trend, wahania sezonowe i cykliczne). Podstawą sporządzania prognozy jednokrokowej jest model (3.1) przeindeksowany dla czasu $t+1$:

$$Y_{t+1}^* = f[Y_{t+1-k}, e_{t+1-k}] + e_{t+1}, k > 0, \quad (4.1)$$

gdzie:

- Y_{t+1}^* – wartość prognozowana zmiennej w momencie lub okresie $t+1$,
- Y_t – wartość zmiennej w momencie lub okresie t ,
- e_t – wahania przypadkowe w momencie lub okresie t .

Zależność (4.1) dotyczy prognozy jednokrokowej sporządzanej w czasie t (po uzyskaniu obserwacji Y_t) dla czasu $t+1$. Wyznaczona w ten sposób wartość prognozowana Y_{t+1}^* może być wykorzystana do wyznaczenia kolejnej prognozy Y_{t+2}^* (dwukrokowej), przy przyjęciu założenia, że $Y_{t+1} = Y_{t+1}^*$. Wykonując powyższe postępowanie iteracyjnie można wyznaczać prognozy dla kolejnych przyszłych wartości zmiennej na i kroków w przód.

Ogólny model autoregresji i średniej ruchomej ARMA(p, q) opisany zależnością (3.10) można po przeindeksowaniu dla czasu $t+i$ zapisać do celów prognostycznych następująco:

$$Y_{t+i}^* = \hat{\phi}_1 \hat{Y}_{t-1+i} + \dots + \hat{\phi}_p \hat{Y}_{t-p+i} + e_t - \hat{\theta}_1 \hat{e}_{t-1+i} - \dots - \hat{\theta}_q \hat{e}_{t-q+i}, \quad (4.2)$$

przy czym:

$\hat{Y}_{t-k+i}$ – wartość rzeczywista szeregu lub oszacowana modelem w krokach poprzednich,

$\hat{e}_{t-k+i}$ – wartość rzeczywista błędu lub oszacowana modelem w krokach poprzednich.

Podobnie można przekształcić scałkowany model autoregresji i średniej przedstawiony równaniem (3.29). Na przykład, w wypadku modelu ARIMA(1,1,1), tj. $Y_t = (1 + \phi_1)Y_{t-1} - \phi_1Y_{t-2} + e_t - \theta_1e_{t-1}$, prognozy na kolejne okresy będą wyznaczone iteracyjnie, korzystając z wartości obserwowanych lub oszacowanych tym modelem w krokach poprzednich:

$$Y_{t+1}^* = (1 + \hat{\phi}_1)Y_t - \hat{\phi}_1Y_{t-1} - \hat{\theta}_1e_t^*, \quad (4.3)$$

$$Y_{t+2}^* = (1 + \hat{\phi}_1)Y_{t+1}^* - \hat{\phi}_1Y_t - \hat{\theta}_1e_{t+1}^*, \quad (4.4)$$

$$Y_{t+3}^* = (1 + \hat{\phi}_1)Y_{t+2}^* - \hat{\phi}_1Y_{t+1}^* - \hat{\theta}_1e_{t+2}^*. \quad (4.5)$$

W podanych zależnościach problematyczne jest wyznaczenie pierwszych reszt modelu. Jednym z rozwiązań jest skorzystanie z tak zwanego prognozowania wstecznego, które wykorzystuje fakt, że charakterystyki stacjonarnego procesu pozostają stałe bez względu na kierunek osi czasu¹⁴⁵. Tym samym model ARMA: $\phi(B)\hat{Y}_t = \theta(B)e_t$ można zapisać wykorzystując operator przesunięcia w przód F , tj.: $F^jY_t = Y_{t+j}$, otrzymując¹⁴⁶:

$$\phi(F)\hat{Y}_t = \theta(F)e_t', \quad (4.6)$$

gdzie:

$\phi(F) = (1 - \phi_1F - \phi_2F^2 - \dots - \phi_pF^p)$ – operator autoregresyjny rzędu p ,

$\theta(F) = (1 - \theta_1F - \theta_2F^2 - \dots - \theta_qF^q)$ – operator średniej ruchomej rzędu q ,

e_t' – proces białego szumu o wariancji $\sigma_{e'}^2 = \sigma_e^2$.

¹⁴⁵ M. Cieślak (red.) (2011), *Prognozowanie gospodarcze...*, op. cit., s. 128.

¹⁴⁶ Ibidem.

Wartości $\hat{Y}_t, \hat{Y}_{t-1}, \hat{Y}_{t-2}$ i wcześniejsze mogą zostać obliczone na podstawie znanych wartości procesu za pomocą przekształconego równania (4.6) do postaci¹⁴⁷:

$$\hat{Y}_t = \phi_1 \hat{Y}_t + \phi_2 \hat{Y}_{t+1} + \dots + \phi_p \hat{Y}_{t+p} + e'_t - \theta_1 e'_{t+1} - \dots - \theta_q e'_{t+q}. \quad (4.7)$$

Natomiast reszty jako:

$$e'_t = \hat{Y}_t - \phi_1 \hat{Y}_t - \phi_2 \hat{Y}_{t+1} - \dots - \phi_p \hat{Y}_{t+p} + \theta_1 e'_{t+1} + \dots + \theta_q e'_{t+q}. \quad (4.8)$$

Na przykład, rozważając wspomniany model ARIMA(1,1,1) za pomocą operatorów przesunięcia wstecz i w przód można zapisać odpowiednio:

$$(1 - \hat{\phi}_1 B)(1 - B)Y_t = (1 - \hat{\theta}_1 B)\hat{e}_t \quad (4.9)$$

oraz

$$(1 - \hat{\phi}_1 F)(1 - F)Y_t = (1 - \hat{\theta}_1 F)e'_t. \quad (4.10)$$

Wzory (4.9) i (4.10) są równoznaczne z zapisem:

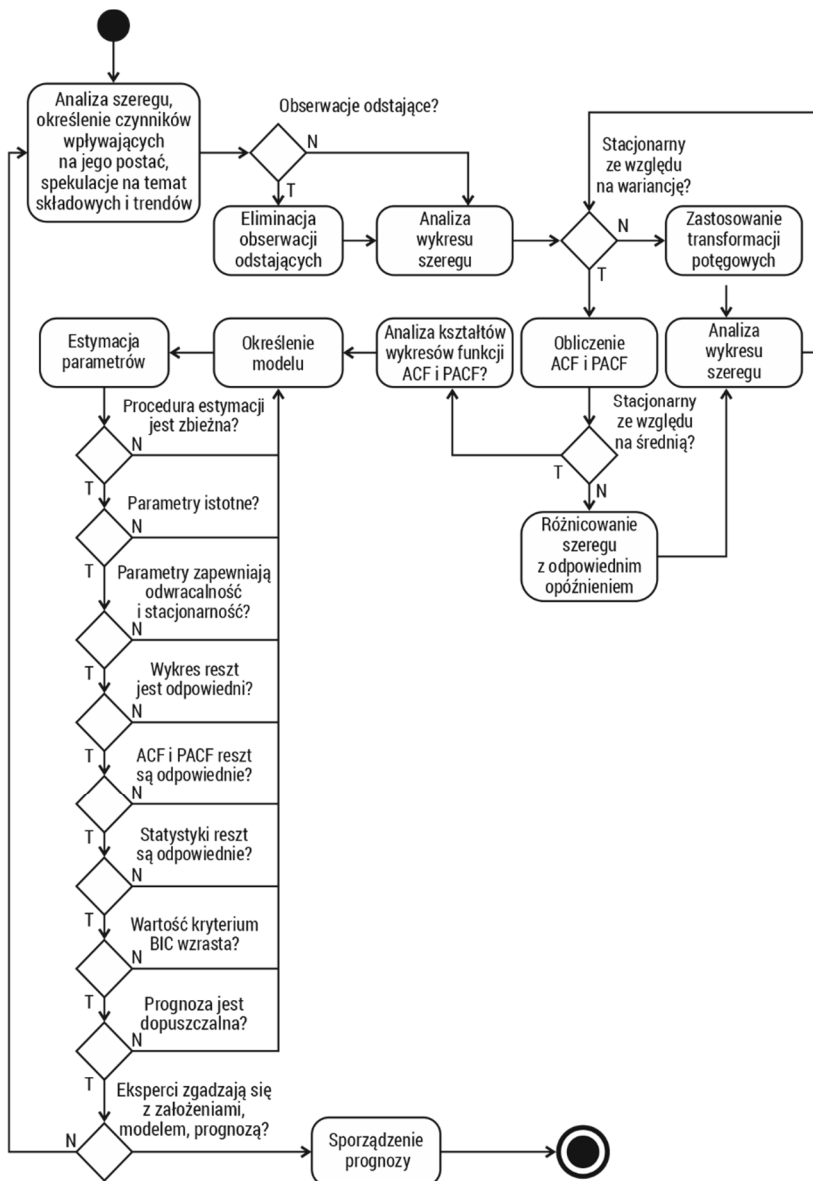
$$\hat{e}_t = \hat{Y}_t - (1 + \hat{\phi}_1)\hat{Y}_{t-1} + \hat{\phi}_1 \hat{Y}_{t-2} + \hat{\theta}_1 \hat{e}_{t-1} \quad (4.11)$$

oraz

$$e'_t = \hat{Y}_t - (1 + \hat{\phi}_1)\hat{Y}_{t+1} + \hat{\phi}_1 \hat{Y}_{t+2} + \hat{\theta}_1 e'_{t+1}. \quad (4.12)$$

Sposób postępowania przy prognozowaniu szeregów czasowych oparty na modelach klasy ARIMA zaprezentowano w postaci diagramu na rys. 4.1. Uszczegóławia on przedstawione w rozdziale trzecim na rys. 3.1 etapy budowy modelu oraz podsumowuje opisane w poprzednich rozdziałach informacje.

¹⁴⁷ Ibidem.



Rys. 4.1. Prognozowanie oparte na modelach ARIMA

Źródło: opracowanie własne na podstawie: S.A. DeLurgio (1998), *Forecasting Principles and Applications*, Irwin/McGraw-Hill, s. 332.

Budowa modelu prognostycznego ARIMA jest procesem często wymagającym powtórzeń, cofania się do wykonanych już etapów, porównywania otrzymywanych wyników¹⁴⁸. Rozpoczyna się od analizy szeregu czasowego, co obejmuje określenie czynników kształtujących proces stochastyczny, wyodrębnienie składowych szeregu, eliminację obserwacji odstających nietypowych, które mogą zaburzyć ocenę zmienności szeregu i jako incydentalne, z małym prawdopodobieństwem, w jakimś wariancie powtórzą się w przyszłości. Kolejnym bardzo istotnym etapem jest ewentualna transformacja szeregu do postaci stacjonarnej. Obejmuje ona zarówno operacje stabilizujące wariancję, jak i średnią. Niekiedy potrzebne jest zastosowanie różnych transformacji lub wielokrotne różnicowanie szeregu. Na diagramie ilustrują to pętle: *Stacjonarny ze względu na wariancję?* oraz *Stacjonarny ze względu na średnią?* Wybór modelu tylko na podstawie funkcji autokorelacji i autokorelacji cząstkowych nieraz jest problematyczny, gdyż przebiegi korelogramów rzeczywistych szeregów czasowych niejednokrotnie dosyć znacznie różnią się od charakterystyk teoretycznych. Proces wyboru adekwatnego modelu wymaga często porównania kilku wstępnie określonych modeli. Może też dokonywać się drogą eliminacji, co odzwierciedla zestaw pytań testowych na diagramie. Parametry modelu, oszacowane wybranym algorytmem optymalizacji, muszą spełniać warunki stacjonarności i odwracalności modelu oraz być statystycznie istotne. Ponadto model powinien ujmować wszystkie systematyczne składowe szeregu. W celu sprawdzenia tego warunku przeprowadza się test losowości szeregu reszt, który może obejmować zarówno wizualizację szeregu reszt, autokorelacji i autokorelacji cząstkowej, ale też testy statystyczne. W wyborze właściwego modelu, z kilku rozważanych, pomocne mogą być wartości kryterium informacyjnego. O dopasowaniu modelu do danych świadczą błędy prognoz wygasłych (wariancja reszt). Model powinien przejść też weryfikację merytoryczną, tłumaczyć badane zjawisko, co na diagramie

¹⁴⁸ J. Fattah, L. Ezzine, Z. Aman, H. El Moussami, A. Lachhab (2018), *Forecasting of demand using ARIMA model*, „International Journal of Engineering Business Management” 10, s. 1–9, doi: 10.1177/1847979018808673.

wyróżnione jest jako ostatnie pytanie sprawdzające: *Eksperci zgadzają się z założeniami, modelem, prognozą?*

Prognozowanie za pomocą modelu ARIMA można wykonać za pomocą wielu programów komputerowych. Na budowę modelu i prognozowanie pozwalają pakiety statystyczne, takie jak: R¹⁴⁹ (R jest projektem opartym o licencje GNU GPL), SAS¹⁵⁰ (dostępny bezpłatnie dla studentów w ramach licencji SAS University Edition lub OnDemand for Academics [stan na 2.12.2021]), Matlab¹⁵¹ (wdrożony i dostępny dla studentów na bardzo wielu uczelniach akademickich), SPSS¹⁵², Mathematica¹⁵³, ale też dodatki do MS Excela, np. XLSTAT¹⁵⁴. Przykłady zastosowania algorytmu przedstawiono z wykorzystaniem pakietu Statistica 13 (dostępny dla studentów w ramach licencji akademickiej). Za pomocą programu można przejść wszystkie etapy budowy prognozy, rozpoczynając od wstępnej eksploracji danych, poprzez budowę modeli i ich weryfikację, a kończąc na obliczeniu wartości prognozy¹⁵⁵.

4.2. Przykłady budowy prognoz

4.2.1. Niesezonowy model ARIMA(0,2,1)

Pierwszy przykład dotyczy opracowania prognozy zużycia energii pochodzącej ze źródeł odnawialnych na świecie – OZE (wiatru, promieniowania słonecznego, wody, biomasy, energii geotermalnej oraz innych) na podstawie danych historycznych z lat 1965–2020. Dane pochodzą z serwisu internetowego Our World

¹⁴⁹ CRAN: <https://cran.r-project.org/> [2021-12-12].

¹⁵⁰ SAS: https://www.sas.com/pl_pl/home.html [2021-12-12].

¹⁵¹ MATHWorks: <https://uk.mathworks.com/> [2021-12-12].

¹⁵² SPSS Polska: <https://spss.pl/> [2021-12-12].

¹⁵³ WOLFRAM: <https://www.wolfram.com/mathematica/> [2021-12-12].

¹⁵⁴ XLSTAT: <https://www.xlstat.com/en/> [2021-12-12].

¹⁵⁵ E. Chodakowska (2002), *Zastosowanie pakietu STATISTICA w prognozowaniu na podstawie szeregów czasowych*, „Zeszyty Naukowe Politechniki Białostockiej. Ekonomia i Zarządzanie” 7, s. 77–86.

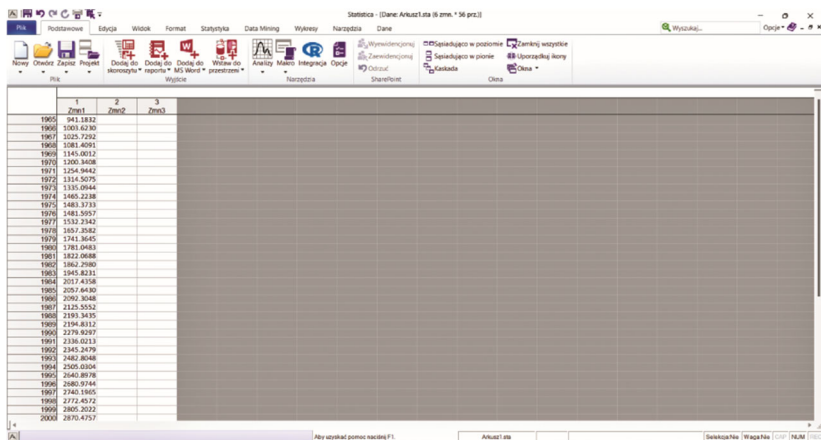
in Data i są dostępne pod adresem: https://ourworldindata.org/grapher/modern-renewable-energy-consumption?country=~OWID_WRL [2.12.2021]. Szereg czasowy przedstawiono w tabeli 4.1. W celu przeprowadzenia analiz dane pobrane z serwisu zapisano w arkuszu programu Statistica (rys. 4.2).

Tabela 4.1. Zużycie energii pochodzącej ze źródeł odnawialnych na świecie w terawatogodzinach (TWh)

t	Rok	Konsumpcja energii [TWh]	t	Rok	Konsumpcja energii [TWh]	t	Rok	Konsumpcja energii [TWh]	t	Rok	Konsumpcja energii [TWh]
1	1965	941,183	15	1979	1741,365	29	1993	2482,805	43	2007	3552,429
2	1966	1003,623	16	1980	1781,048	30	1994	2505,030	44	2008	3805,907
3	1967	1025,729	17	1981	1822,069	31	1995	2640,898	45	2009	3888,276
4	1968	1081,409	18	1982	1862,298	32	1996	2680,974	46	2010	4196,936
5	1969	1145,001	19	1983	1945,823	33	1997	2740,197	47	2011	4409,302
6	1970	1200,341	20	1984	2017,436	34	1998	2772,457	48	2012	4717,686
7	1971	1254,944	21	1985	2057,643	35	1999	2805,202	49	2013	5041,431
8	1972	1314,507	22	1986	2092,305	36	2000	2870,476	50	2014	5308,287
9	1973	1335,094	23	1987	2125,555	37	2001	2815,481	51	2015	5519,439
10	1974	1465,224	24	1988	2193,343	38	2002	2892,232	52	2016	5869,965
11	1975	1483,373	25	1989	2194,831	39	2003	2910,627	53	2017	6246,874
12	1976	1481,596	26	1990	2279,930	40	2004	3144,044	54	2018	6655,333
13	1977	1532,234	27	1991	2336,021	41	2005	3279,194	55	2019	7017,106
14	1978	1657,358	28	1992	2345,248	42	2006	3437,884	56	2020	7443,809

Źródło: opracowanie własne danych z serwisu Our World in Data, dostępnych pod adresem: https://ourworldindata.org/grapher/modern-renewable-energy-consumption?country=~OWID_WRL [2.12.2021].

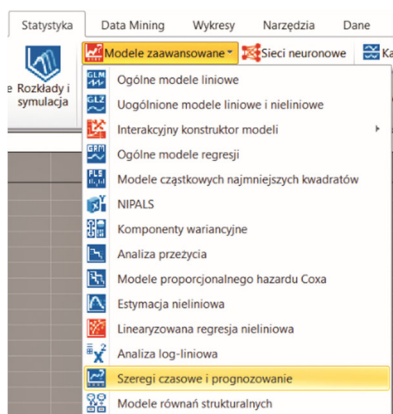
Efektem wzrostu znaczenia zrównoważonej energii są zmiany w branży energetycznej i ciągły wzrost produkcji energii ze źródeł odnawialnych. W latach 1965–2003 obserwowany był wzrost zużycia „zielonej energii” o ok. 3% rocznie. Znaczne przyspieszenie produkcji z wykorzystaniem OZE nastąpiło od roku 2004, wynosząc przeciętnie 5,5% rocznie.



Rys. 4.2. Pakiet Statistica – arkusz danych

Źródło: opracowanie własne.

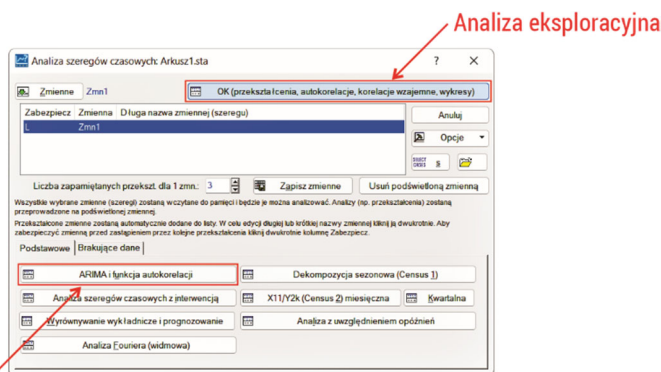
Do budowy modelu ARIMA, tj. weryfikacji stacjonarności, transformacji zmiennych, a także estymacji parametrów służy w programie Statistica moduł dostępny poprzez: Statystyka → Modele zaawansowane → Szeregi czasowe i prognozowanie (rys. 4.3).



Rys. 4.3. Pakiet Statistica – menu Szeregi czasowe i prognozowanie

Źródło: opracowanie własne.

Prognozowanie na podstawie modeli ARIMA rozpoczyna się od wstępnej eksploracji danych, które w wypadku modeli ARIMA oznacza ocenę kształtowania się szeregu czasowego. Obejmuje ona przede wszystkim wizualizację zarówno samego szeregu, jak i jego funkcji autokorelacji i autokorelacji cząstkowej. Wywołane poleceniem *Szeregi czasowe i prognozowanie* okno *Analiza szeregów czasowych* (rys. 4.4), umożliwia dostęp zarówno do wstępnej analizy szeregu i jego transformacji, tj. *OK (przekształcenia, autokorelacje, korelacje wzajemne, wykresy)*, jak i estymacji modelu ARIMA: *ARIMA i funkcja autokorelacji*.

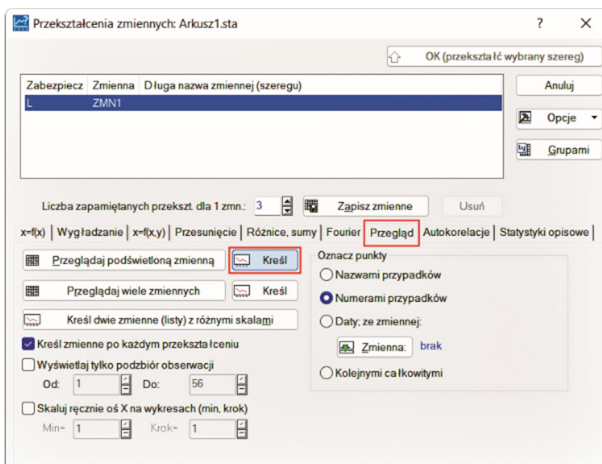


Estymacja i weryfikacja modelu

Rys. 4.4. Pakiet Statistica – okno Analiza szeregów czasowych

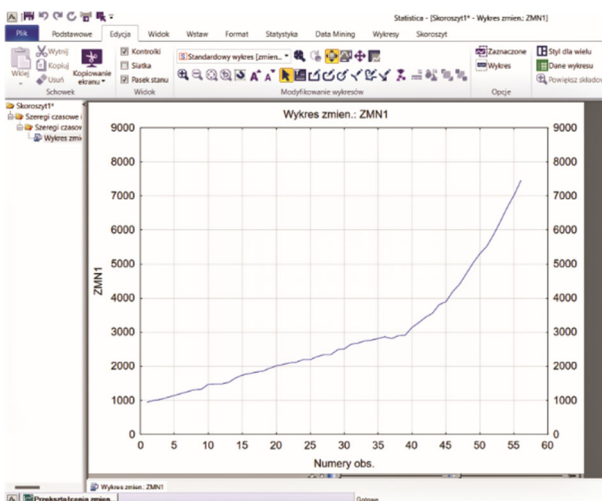
Źródło: opracowanie własne.

Widok okna *Przekształcenia zmiennych* pokazano na rys. 4.5. Wykreślenie przebiegu zmiennej dostępne jest z poziomu zakładki *Przegląd* poprzez wybór opcji *Kreśl*. Otrzymany szereg przedstawiono na wykresie na rys. 4.6.



Rys. 4.5. Pakiet Statistica – okno Przekształcenia zmiennych (Przegląd)

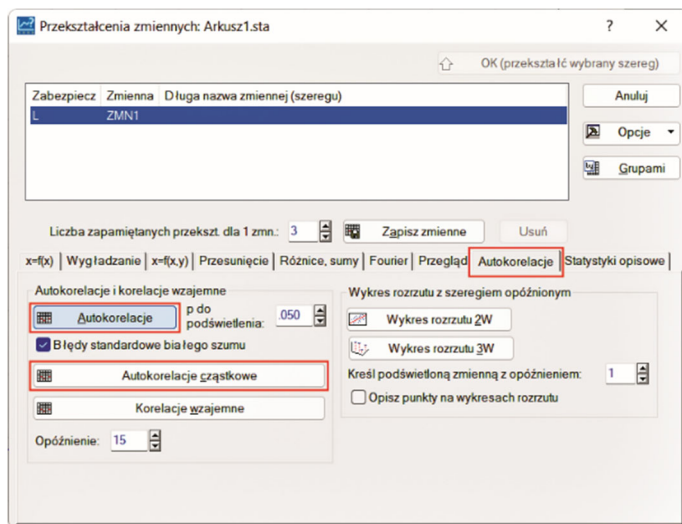
Źródło: opracowanie własne.



Rys. 4.6. Szereg zużycia energii pochodzącej ze źródeł odnawialnych na świecie w programie Statistica

Źródło: opracowanie własne danych z serwisu Our World in Data, dostępnych pod adresem: https://ourworldindata.org/grapher/modern-renewable-energy-consumption?country=~OWID_WRL [2.12.2021].

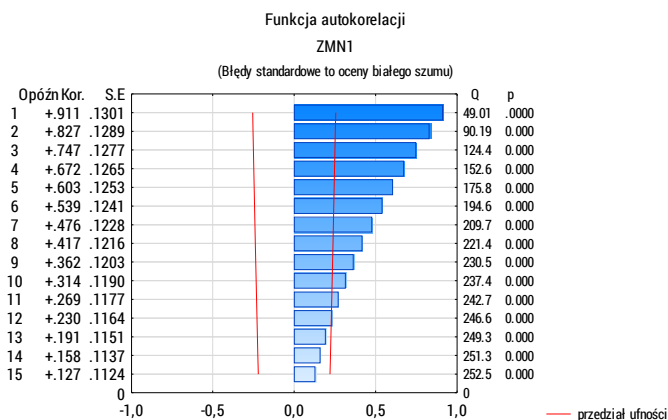
Analiza kształtu krzywej wskazuje, że główną cechą charakteryzującą szereg jest trend. Niestacjonarność ze względu na średnią potwierdzają kształty funkcji autokorelacji i autokorelacji cząstkowej przedstawione na rys. 4.8 (wolne, liniowo zmniejszające się wartości ACF i pojedyncza duża wartość PACF). Otrzymano je, wybierając zakładkę **Autokorelacje** w oknie dialogowym **Przekształcenia zmiennych** i szacując kolejno **Autokorelacje** i **Autokorelacje cząstkowe** (rys. 4.7).



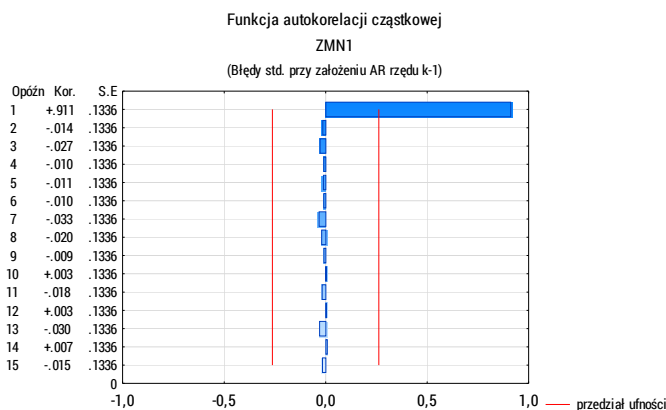
Rys. 4.7. Pakiet Statistica – okno Przekształcenia zmiennych (Autokorelacje)

Źródło: opracowanie własne.

a)



b)



Rys. 4.8. ACF (a) i PACF (b) szeregu zużycia energii pochodzącej ze źródeł odnawialnych na świecie

Źródło: opracowanie własne danych z serwisu Our World in Data, dostępnych pod adresem: https://ourworldindata.org/grapher/modern-renewable-energy-consumption?country=~OWID_WRL [2.12.2021].

Wnioskowanie o stacjonarności wariancji w wypadku tych danych, na podstawie wizualizacji danych bez wcześniejszego usunięcia trendu, jest utrudnione. Na niestacjonarność wskazuje różne tempo wzrostu wartości szeregu. W celu formalnego sprawdzenia, czy szereg jest niestacjonarny ze względu na wariancję

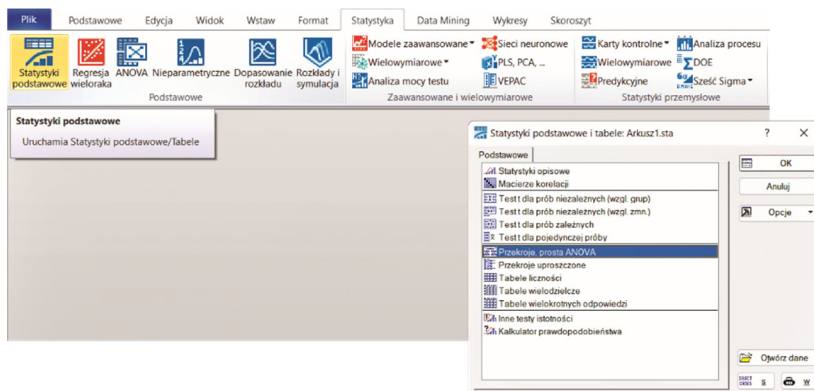
wykonano testy Levene'a oraz Browna-Forsythe'a. Testy, przywołując informacje z rozdziału drugiego, weryfikują hipotezę H_0 : wariancje w grupach są równe, wobec hipotezy alternatywnej H_1 : nie są równe. Zaobserwowane przyśpieszenie wzrostu wartości szeregu na początku lat 2000 może być podstawą do podziału obserwacji na dwie równoliczne grupy po 28 obserwacji. Do arkusza dodano zmienną grupującą, przyjmującą wartości 1 dla obserwacji od 1 do 28 i 2 dla obserwacji od 29 do 56 (rys. 4.9). Należy podkreślić, że wartości zmiennej grupującej nie mają znaczenia, mogą być także etykietami, gdyż służą tylko wskazaniu obserwacji należących do poszczególnych grup.

	1 Zmn1	2 Zmn2	3 Zmn3
1965	941.1832	1	
1966	1003.6230	1	
1967	1025.7292	1	
1968	1081.4091	1	
1969	1145.0012	1	
1970	1200.3408	1	
1971	1254.9442	1	
1972	1314.5075	1	
1973	1335.0944	1	
1974	1465.2238	1	
1975	1483.3733	1	
1976	1481.5957	1	
1977	1532.2342	1	
1978	1657.3582	1	
1979	1741.3645	1	
1980	1781.0483	1	
1981	1822.0688	1	
1982	1862.2980	1	
1983	1945.8211	1	
1984	2017.4358	1	
1985	2057.6430	1	
1986	2092.3048	1	
1987	2125.5552	1	
1988	2193.3435	1	
1989	2194.8312	1	
1990	2279.9297	1	
1991	2336.0213	1	
1992	2345.2479	1	
1993	2482.8048	2	
1994	2505.0304	2	
1995	2640.8978	2	
1996	2680.9744	2	
1997	2740.1965	2	

Rys. 4.9. Pakiet Statistica – arkusz danych wraz ze zmienną grupującą

Źródło: opracowanie własne.

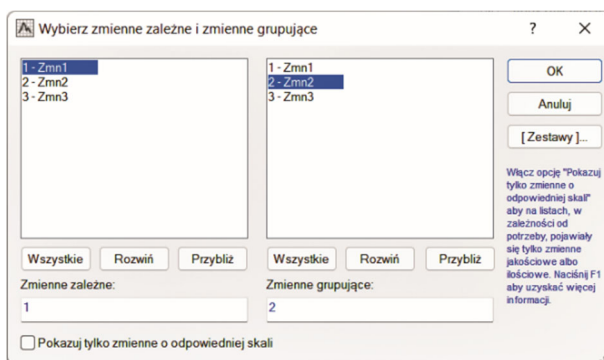
Test Levene'a oraz Browna-Forsythe'a jest udostępniony w module Statystyki podstawowe poprzez Przekroje, prosta ANOVA (rys. 4.10).



Rys. 4.10. Pakiet Statistica – Statystyki podstawowe

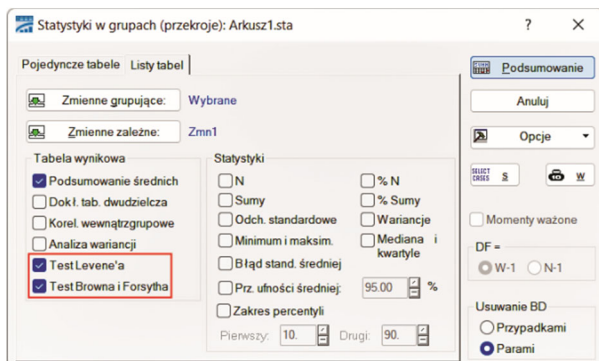
Źródło: opracowanie własne.

Właściwy wybór zmiennej zależnej (szeregu zużycia energii) i grupującej (zmiennej sztucznej) jest kluczowy do otrzymania poprawnych wyników (rys. 4.11). W celu otrzymania wyników testów jednorodności wariancji w oknie dialogowym Statystyki w grupach (przekroje) należy zaznaczyć dwa pola wyboru: Test Levene'a oraz Browna-Forsythe'a (rys. 4.12).



Rys. 4.11. Pakiet Statistica – wybór zmiennych zależnych i grupujących

Źródło: opracowanie własne.



Rys. 4.12. Pakiet Statistica – okno Statystyki w grupach

Źródło: opracowanie własne.

Obliczone, wysokie wartości statystyki F i prawdopodobieństwa wynoszące odpowiednio 0,00000 (test Levene'a) oraz 0,000195 (test Browna-Forsythe'a) mniejsze od poziomu istotności 0,05 (rys. 4.13) wskazują, że należy odrzucić hipotezę zerową na rzecz hipotezy alternatywnej mówiącej: wariancja w grupach nie jest równa.

a)

we	Edycja	Widok	Format	Statystyka	Data Mining	Wykresy	Narzędzia	Dane	Skoroszyt
									
ANOVA	Nieparametryczne	Dopasowanie	Rozkłady i symulacja	Modely zaawansowane * Wielowymiarowe *	Sieci neuronowe PLS, PCA, ... VEPAC	Karty kontrolne *	Wielowymiarowe	Predykcje	
Podstawowe				Zaawansowane i wielowymiarowe			Statystyki p		
Test Levene'a jednorodności wariancji (Arkusz1.sta)									
Zaznaczone efekty są istotne z p < .05000									
Zmienna	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	F	p	
Zmn1	10878935	1	10878935	17878294	54	331079.5	32.85898	0.000000	

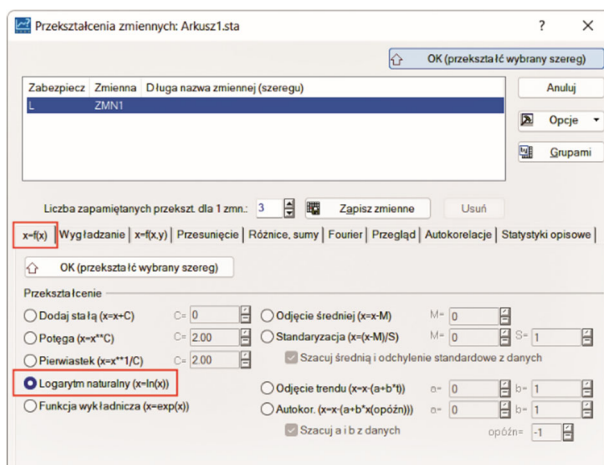
b)

dok	Format	Statystyka	Data Mining	Wykresy	Narzędzia	Dane	Skoroszyt	
								
yczne	Dopasowanie i rozkładu	Rozkłady i symulacja	Modele zaawansowane i wielowymiarowe	Sieci neuronowe i PLS PCA...	VEPAC	Karty kontrolne i Wielowymiarowe	Analiza procesu i DOE	
			Analiza mocy testu			Predykcyjne	Sześć Sigma	
Zaawansowane i wielowymiarowe					Statystyki przemysłowe			
Test jednorod. wariancji Browna-Forsythe'a (Arkusz1.sta)								
Zaznaczone efekty są istotne z p < .05000								
Zmienna	SS Efekt	df Efekt	SS Efekt	MS Błąd	SS Błąd	F	p	
Zmn1	9300766	1	9300766	31406117	54	581594.8	15.99183	0.000195

Rys. 4.13. Wyniki test Levene'a (a) oraz testu Browna-Forsythe'a (b)

Źródło: opracowanie własne.

Ponieważ szereg jest niestacjonarny ze względu na wariancję, poddano go operacji logarytmowania. W pakiecie Statistica możliwe jest to w oknie dialogowym **Przekształcenia zmiennych**, na zakładce $x = f(x)$ poprzez wybór **Logarytm naturalny ($x = \ln(x)$)** (rys. 4.14).



Rys. 4.14. Pakiet Statistica – okno Przekształcenia zmiennych (logarytmowanie)

Źródło: opracowanie własne.

Ponownie obliczone wartości testów wykonane dla szeregu zlogarytmowanego wskazują, że szereg osiągnął stacjonarność ze względu na wariancję (rys. 4.15).

a)

dok

Format

Statystyka

Data Mining

Wykresy

Narzędzia

Dane

Skoroszyt

yczne

Dopasowanie

Rozkłady i rozkładu

rozkładu

symulacja

Modele zaawansowane *

Sieci neuronowe

Karty kontrolne *

Analiza procesu

Wielowymiarowe *

PLS, PCA, ...

Wielowymiarowe

DOE

Analiza mocy testu

VEPAC

Predykcyjne

Sześć Sigma *

Zaawansowane i wielowymiarowe

Statystyki przemysłowe

Test Levene'a jednorodności wariancji (Arkusz1.sta)

Zaznaczone efekty są istotne z $p < .05000$

Zmienna	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	F	p
Zmn3	0.043506	1	0.043506	1.271356	54	0.023544	1.847883	0.179680

b)

dok

Format

Statystyka

Data Mining

Wykresy

Narzędzia

Dane

Skoroszyt

Modelce zaawansowane *

Sieci neuronowe

Karty kontrolne *

Analiza procesu

yczne

Dopasowanie

Rozkłady i rozkładu

symulacja

Wielowymiarowe *

PLS, PCA, ...

Wielowymiarowe

DOE

Analiza mocy testu

VEPAC

Predykcyjne

Sześć Sigma *

Zaawansowane i wielowymiarowe

Statystyki przemysłowe

Test jednorod. wariancji Browna-Forsythe'a (Arkusz1 sta)

Zaznaczone efekty są istotne z $p < .05000$

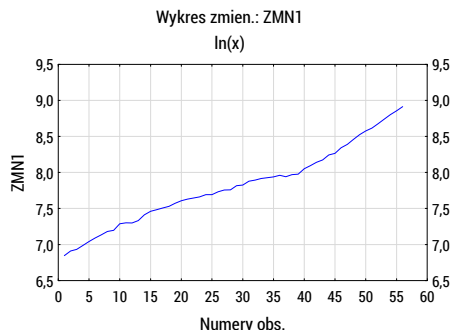
Zmienna	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	F	p
Zmn3	0.038679	1	0.038679	1.695777	54	0.031403	1.231694	0.271996

Rys. 4.15. Wyniki test Levene'a (a) oraz testu Browna-Forsythe'a (b) dla szeregu zlogarytmowanego

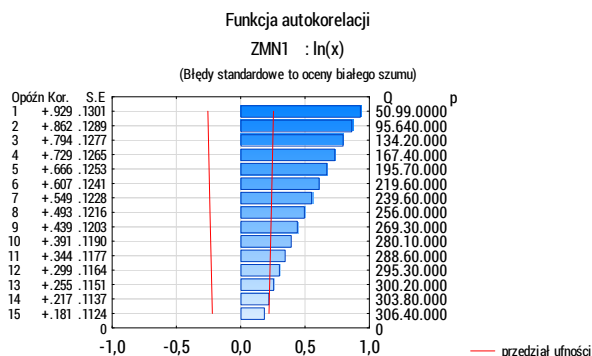
Źródło: opracowanie własne.

Kolejnym etapem jest weryfikacja stacjonarności ze względu na średnią. Ponownie zilustrowano kształtowanie się szeregu zlogarytmowanego oraz jego funkcji autokorelacji i autokorelacji cząstkowej (rys. 4.16).

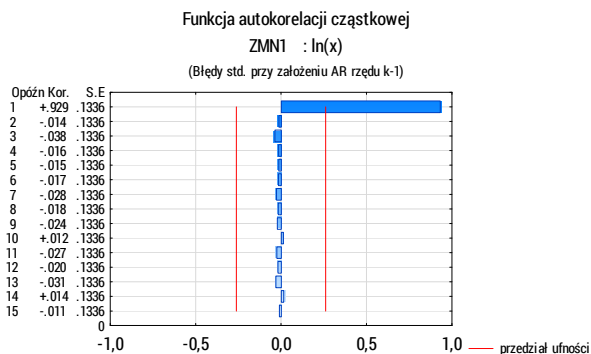
a)



b)



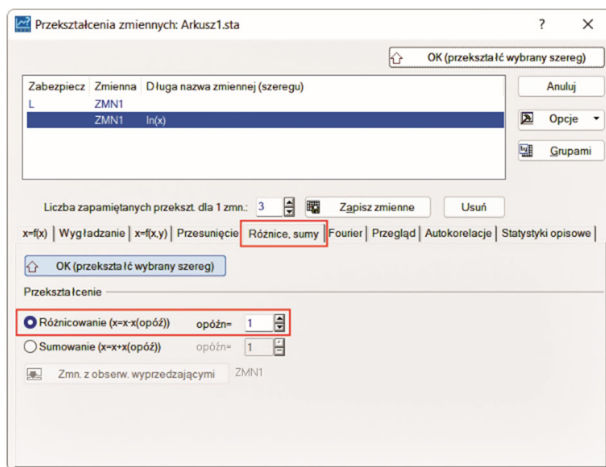
c)



Rys. 4.16. Zlogarytmowany szereg zużycia energii pochodzącej ze źródeł odnawialnych (a), ACF (b) i PACF (c)

Źródło: opracowanie własne danych z serwisu Our World in Data, dostępnych pod adresem: https://ourworldindata.org/grapher/modern-renewable-energy-consumption?country=~OWID_WRL [2.12.2021].

Komentując kształtowanie się zlogarytmowanego szeregu można powtórzyć, dokonane na początku analizy, wnioski: szereg zdecydowanie nie jest stacjonarny ze względu na średnią. Dlatego też zlogarytmowany szereg poddano operacji różnicowania z opóźnieniem 1. Sposób wykonania tej operacji przedstawiono na rys. 4.17.



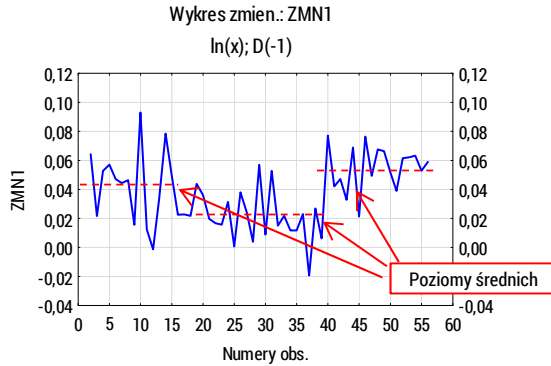
Rys. 4.17. Pakiet Statistica – okno Przekształcenia zmiennych (różnicowanie)

Źródło: opracowanie własne.

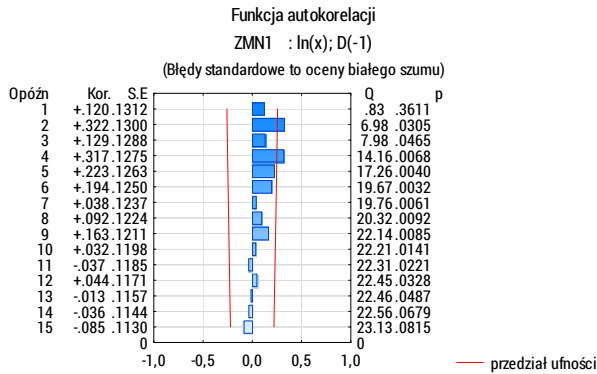
Po jednokrotnym zróżnicowaniu zlogarytmowanego szeregu ponownie wykreślono wartości szeregu oraz jego funkcji autokorelacji i autokorelacji cząstkowych w celu oceny osiągnięcia stacjonarności (rys. 4.18).

Na podstawie oceny wizualnej można stwierdzić, że szereg w dalszym ciągu nie osiągnął stacjonarności. Na rys. 4.18 średnia szeregu przez pierwsze okresy oscyluje wokół 0,04, następnie spada do ok. 0,02, by po 40. obserwacji wzrosnąć do ok. 0,05. Dlatego też zdecydowano się ponownie zróżnicować szereg. Wyniki przedstawiono na rys. 4.19.

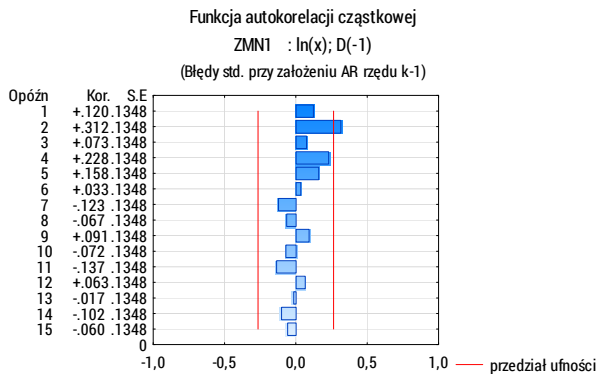
a)



b)



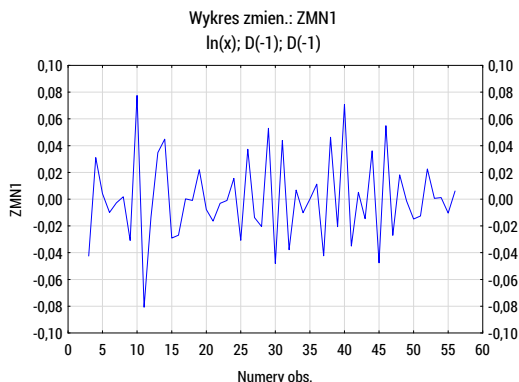
c)



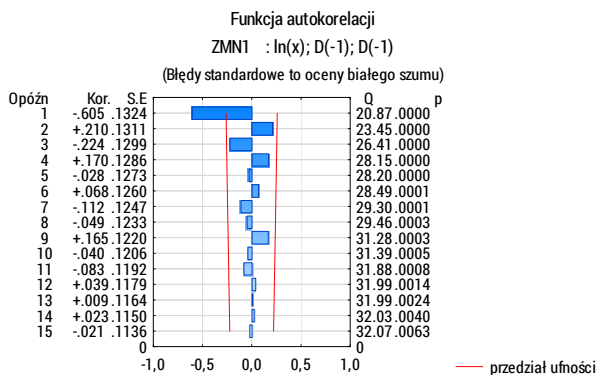
Rys. 4.18. Zlogarytmowany i jednokrotnie zróżnicowany szereg zużycia energii pochodzącej ze źródeł odnawialnych (a), ACF (b) i PACF (c)

Źródło: opracowanie własne.

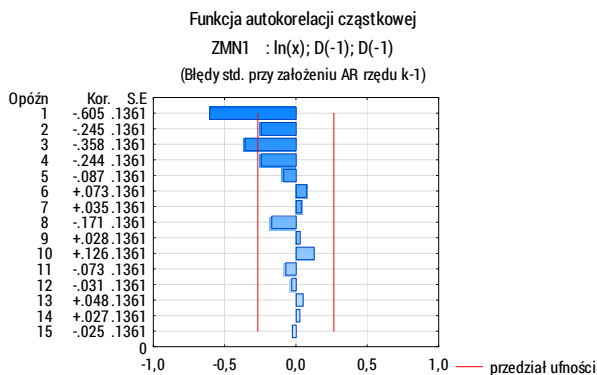
a)



b)



c)



Rys. 4.19. Zlogarytmowany i dwukrotnie zróżnicowany szereg zużycia energii pochodzącej ze źródeł odnawialnych (a), ACF (b) i PACF (c)

Źródło: opracowanie własne.

Średnia szeregu zilustrowanego na rys. 4.19 oscyluje wokół zera. Nie widać też znaczących różnic w amplitudach wahań. Dlatego można przyjąć, że szereg po transformacjach osiągnął stacjonarność i na podstawie kształtowania się wartości funkcji autokorelacji i autokorelacji dla kolejnych opóźnień można wnioskować o postaci modelu ARIMA. Malejące wartości autokorelacji częściowej i jedna znacząca wartość autokorelacji przy opóźnieniu 1 wskazuje, że zachowanie szeregu można spróbować opisać dwukrotnie scałkowanym modelem średniej ruchomej dla szeregu zlogarytmowanego – ARIMA(0,2,1):

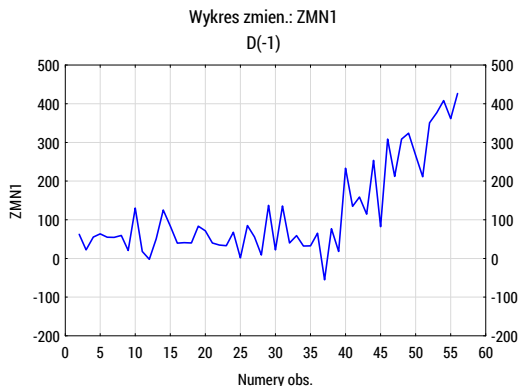
$$(1 - B)^2 Y_t = (1 - \theta_1 B) e_t, \quad (4.13)$$

Równoznaczny zapis równania (4.13) to:

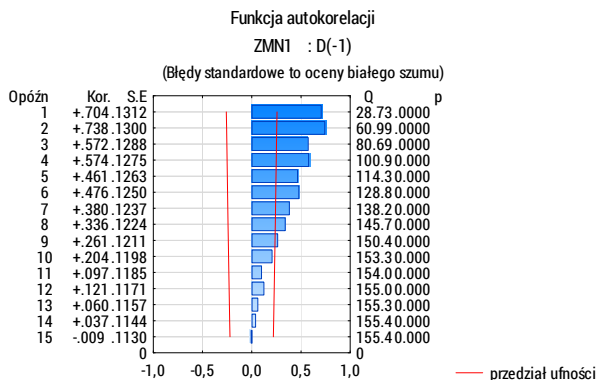
$$Y_t = 2Y_{t-1} - Y_{t-2} - \theta_1 e_{t-1} + e_t. \quad (4.14)$$

Wracając do wykonanych na początku testów jednorodności wariancji w celu określenia zasadności logarytmowania można dodać, że alternatywnym podejściem jest ocena wizualna szeregu zróżnicowanego. W pojedynczo zróżnicowanym szeregu, który w dalszym ciągu jest niestacjonarny ze względu na średnią, można zaobserwować większe zmiany amplitudy w drugiej części szeregu (rys. 4.20). Niestacjonarność ze względu na wariancję jest bardzo wyraźnie widoczna na wykresach dwukrotnie zróżnicowanego szeregu (rys. 4.21). Kolejność wybranych operacji transformacji wynika z faktu, że logarytm naturalny z liczby ujemnej jest nieokreślony w zbiorze liczb rzeczywistych (wartości ujemne są bardzo prawdopodobne do uzyskania w wyniku operacji różnicowania).

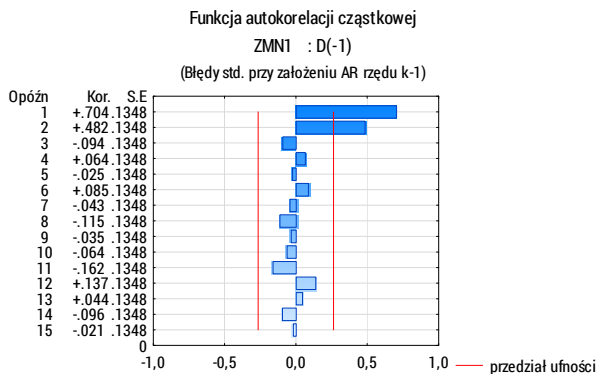
a)



b)



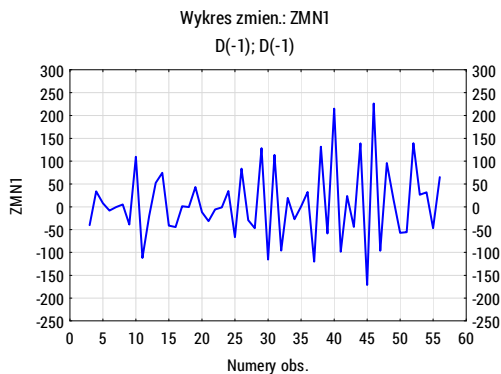
c)



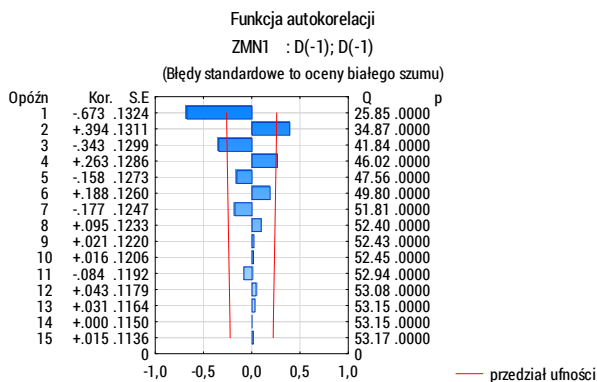
Rys. 4.20. Zróżnicowany szereg zużycia energii pochodzącej ze źródeł odnawialnych (a), ACF (b) i PACF (c)

Źródło: opracowanie własne.

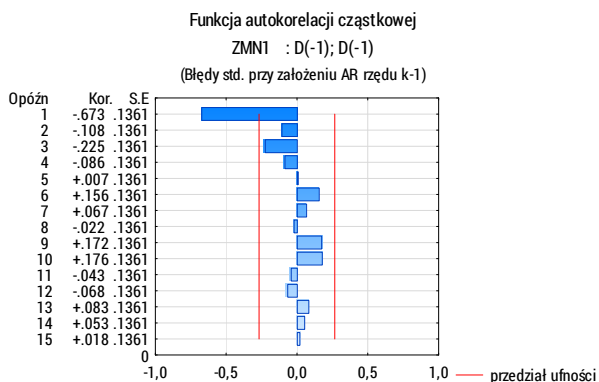
a)



b)



c)



Rys. 4.21. Dwukrotnie zróżnicowany szereg zużycia energii pochodzący ze źródeł odnawialnych (a), ACF (b) i PACF (c)

Źródło: opracowanie własne.

Kolejnym etapem w budowie modelu ARIMA jest estymacja jego parametrów. W celu oszacowania parametrów modelu, w module *Analiza szeregów czasowych* → *ARIMA i funkcja autokorelacji* w oknie dialogowym *Modele ARIMA dla pojedynczego szeregu* należy wskazać wybrany model wraz z transformacjami (rys. 4.22). Warto zwrócić uwagę, że podstawą estymacji jest zmienna oryginalna, zaś logarytmowanie, jak i różnicowanie jest zintegrowane w modelu i wybierane w oknie dialogowym.

Rys. 4.22. Pakiet Statistica – okno Model ARIMA dla pojedynczego szeregu: określenie parametrów

Źródło: opracowanie własne.

Otrzymane wyniki, uzyskane domyślną metodą estymacji McLeoda i Salesa, przedstawiono na rys. 4.23.

ok	Format	Statystyka	Data Mining	Wykresy	Narzędzia	Dane	Skoroszyt
znie	Dopasowanie	Rozkłady i symulacja	Modele zaawansowane	Wielowymiarowe	Analiza procesy	Statystika VB	Statystyki bloku
			Model: ZMN1 (Arkusz1.sla)	Przekształcenia: ln(x), 2 ^o D(1)	Model: (0,2,1) Resztowy MS= .00054		
Parametr	Asympt. Błąd std	Asympt. t(53)	p	Dolna gr 95%p. ufn	Górna gr 95%p. ufn		
q(1)	0.779339	0.069717	11.17865	0.000000	0.639505	0.919173	

Rys. 4.23. Pakiet Statistica – okno Wyniki analizy ARIMA pojedynczego szeregu

Źródło: opracowanie własne.

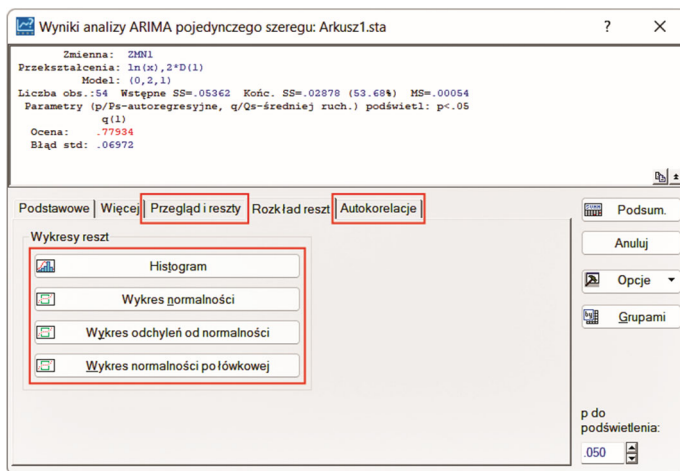
Parametr średniej ruchomej ma wartość 0,77934 i jest statystycznie istotny na podstawie statystyki t , przyjmując $p < 0,05$. Tym samym model ARIMA(0,2,1), inaczej IMA(2,1), można zapisać jako:

$$Y_t = 2Y_{t-1} - Y_{t-2} - 0,77934e_{t-1} + e_t. \quad (4.15)$$

Należy zwrócić uwagę, że model ten mógłby być traktowany jako model ARMA(2,1) z parametrem $\phi_1 = 2, \phi_2 = -1$ oraz $\theta_1 = 0,77934$ przy założeniu, że proces jest stacjonarny. Jednak w takiej interpretacji wartości parametrów autoregresji przekraczają obszar dopuszczalny.

Błąd standardowy oceny parametru wynosi 0,06972. MS w oknie wyników wskazuje średni błąd kwadratowy (średni kwadrat reszt), który wynosi 0,00054. Dopuszczalność rozwiązania, tj. stacjonarność i odwracalność parametrów kontrolowana jest przez algorytmy estymacji programu Statistica.

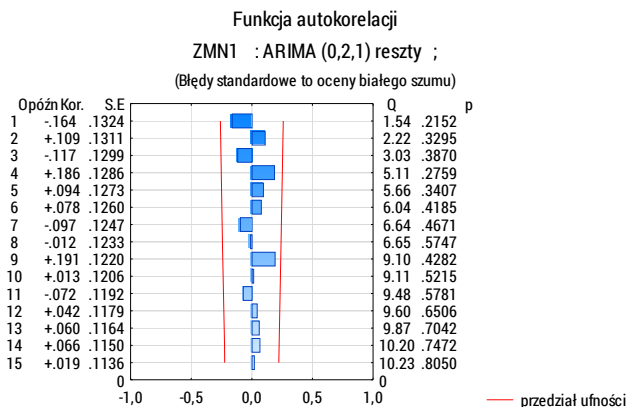
Obowiązkowym etapem, przed wykorzystaniem modelu do prognozowania, jest diagnostyka poprawności i dopasowania modelu poprzez analizę jego reszt. Przypominając, reszty powinny być procesem białego szumu. W module ARIMA i funkcja autokorelacji jest możliwość sprawdzenia kształtowania się reszt (opcja Przeglądaj reszty), autokorelacji reszt i autokorelacji częstotkowej reszt, wykreślenia histogramu reszt oraz wykresów normalności i odchylenia od normalności (rys. 4.24). Wyniki zaprezentowano na rys. 4.25.



Rys. 4.24. Pakiet Statistica – okno Wyniki analizy ARIMA pojedynczego szeregu

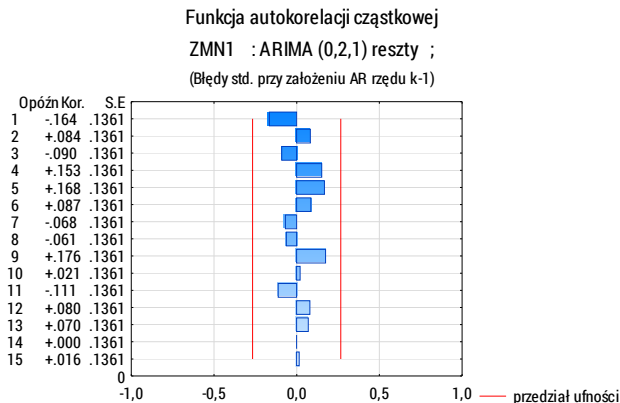
Źródło: opracowanie własne.

a)

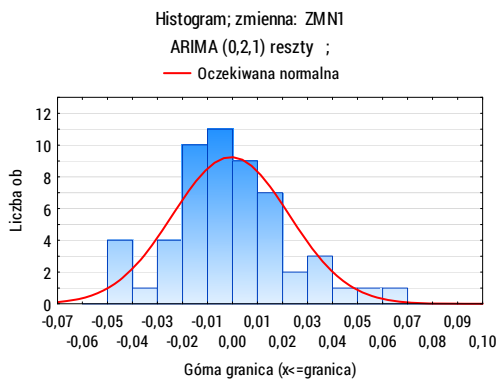


Rys. 4.25. Weryfikacja reszt: ACF (a)

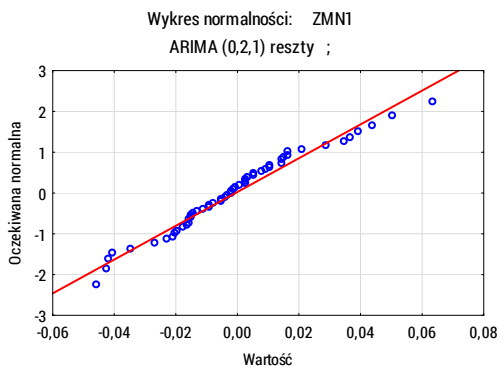
b)



c)



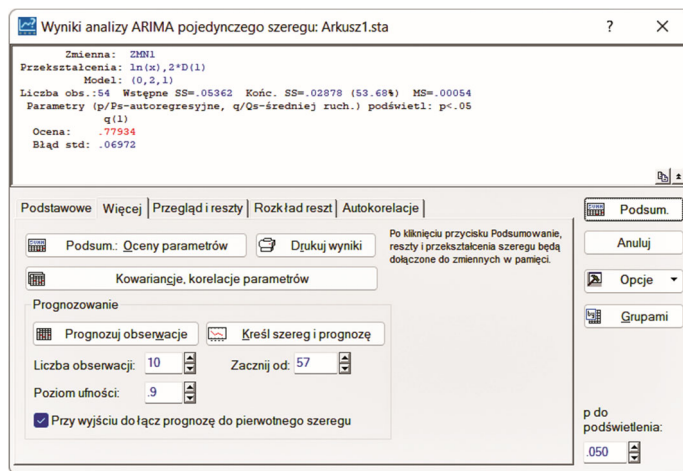
d)



Rys. 4.25. (cd.) Weryfikacja reszt: PACF (b), histogram (c), wykres normalności (d)

Źródło: opracowanie własne.

Ocena graficznego przedstawienia reszt wybranego modelu pozwala przyjąć, że nie odbiegają znacząco od rozkładu normalnego (wykorzystanie formalnych testów z rozkładem normalnym wymaga uruchomienia modułu programu Statystyka → Statystyki nieparametryczne). Autokorelacje także nie wskazują na obecność w szeregu reszt składowej systematycznej. Dla żadnego opóźnienia wartość funkcji autokorelacji nie przekracza wartości krytycznej. Oznacza to, że reszty nie są skorelowane. Tym samym wyznaczony model można wykorzystać do opracowania prognozy. Pakiet Statistica umożliwia wybór okresu, od którego prognoza ma zostać obliczona i liczby prognozowanych wartości (rys. 4.26). Należy pamiętać, że wraz ze wzrostem liczby prognozowanych obserwacji poszerza się przedział ufności. Wartości prognoz oraz ich ilustrację przedstawiono na rys. 4.27.



Rys. 4.26. Pakiet Statistica – okno dialogowe Wyniki analizy ARIMA dla pojedynczego szeregu – prognozowanie

Źródło: opracowanie własne.

a)

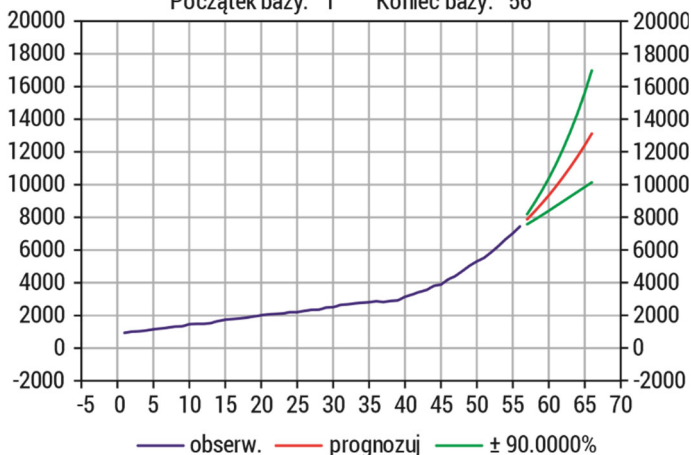
Prognoza; Model: (0,2,1) Opóź. sezon.: 12 (Arkusz1.sta)			
Dane: ZMN1			
Początek bazy: 1 Koniec bazy: 56			
Nr obs.	Prognozu	Dolne 90.0000%	Górne 90.0000%
57	7877.971	7578.53	8191.39
58	8337.45	7839.64	8866.86
59	8823.72	8117.84	9590.98
60	9338.37	8402.31	10378.70
61	9883.02	8689.94	11239.91
62	10459.45	8979.20	12183.71
63	11069.49	9269.17	13219.47
64	11715.11	9559.20	14357.25
65	12396.39	9848.78	15608.04
66	13121.53	10137.47	16983.96

b)

Prognoza; Model: (0,2,1) Opóź. sezon.: 12

Dane: ZMN1

Początek bazy: 1 Koniec bazy: 56



Rys. 4.27. Prognoza zużycia energii pochodzącej ze źródeł odnawialnych na świecie na lata 2021–2030: wartość (a) oraz wykres (b)

Źródło: opracowanie własne.

Szybkość transformacji ku zrównoważonej energetyce jest kluczową kwestią zajmującą naukowców i praktyków. Zgodnie z wyliczonymi wartościami prognozy w 2021 r. można oczekiwać już prawie 8 tys. TWh, a w 2030 ponad 13 tys. TWh. Podstawą szacowania jest dotychczasowe tempo wzrostu i założenie, że utrzyma się ono w kolejnym dziesięcioleciu. Sprawdzenie trafności prognozy wymaga zrealizowania się wartości szeregu. Alternatywnym rozwiązaniem jest podział na zbiór uczący i testowy, co zostało wykorzystane w następnym przykładzie.

4.2.2. Sezonowy model ARIMA(0,1,0)(0,1,1)₄

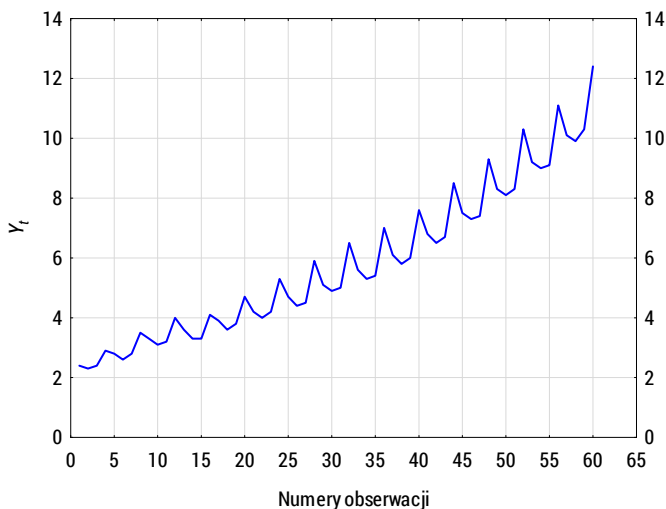
Kolejny przykład dotyczy zastosowania algorytmu dla danych ze składową okresową. Szereg dokumentuje kwartalną sprzedaż detaliczną w handlu elektronicznym, wyrażoną jako procent całkowitej sprzedaży, nieskorygowaną sezonowo. Dane pochodzą z Federal Reserve Economic Data: <https://fred.stlouisfed.org> i obejmują lata 2005–2019 (60 obserwacji). Przedstawiono je w tabeli 4.2.

Tabela 4.2. Kwartalna sprzedaż e-commerce jako procent całkowitej sprzedaży, nieskorygowana sezonowo

Rok	I kwartał	II kwartał	III kwartał	IV kwartał
2005	2,4	2,3	2,4	2,9
2006	2,8	2,6	2,8	3,5
2007	3,3	3,1	3,2	4,0
2008	3,6	3,3	3,3	4,1
2009	3,9	3,6	3,8	4,7
2010	4,2	4,0	4,2	5,3
2011	4,7	4,4	4,5	5,9
2012	5,1	4,9	5,0	6,5
2013	5,6	5,3	5,4	7,0
2014	6,1	5,8	6,0	7,6
2015	6,8	6,5	6,7	8,5
2016	7,5	7,3	7,4	9,3
2017	8,3	8,1	8,3	10,3
2018	9,2	9,0	9,1	11,1
2019	10,1	9,9	10,3	12,4

Źródło: Serwis Federal Reserve Economic, <https://fred.stlouisfed.org/series/ECOMPCTNSA> [5.08.2021].

Podobnie jak w poprzednim przykładzie budowę modelu rozpoczyna analiza graficznej prezentacji danych. Wykres szeregu przedstawiono na rys. 4.28.



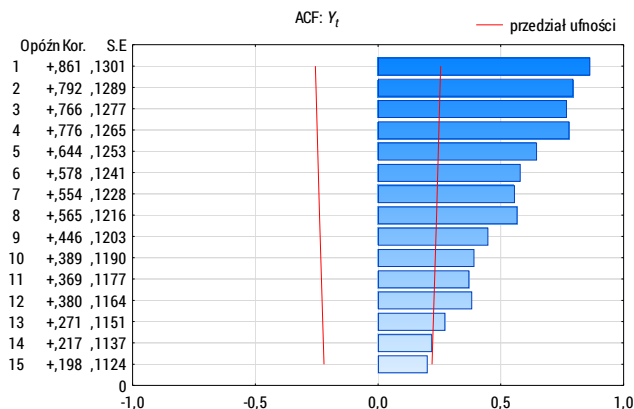
Rys. 4.28. Szereg kwartalnej sprzedaży detalicznej w handlu elektronicznym wyrażonej jako procent całkowitej sprzedaży, nieskorygowany sezonowo 2005–2019

Źródło: opracowanie własne na podstawie danych z Federal Reserve Economic, <https://fred.stlouisfed.org/series/ECOMPCTNSA> [6.08.2021].

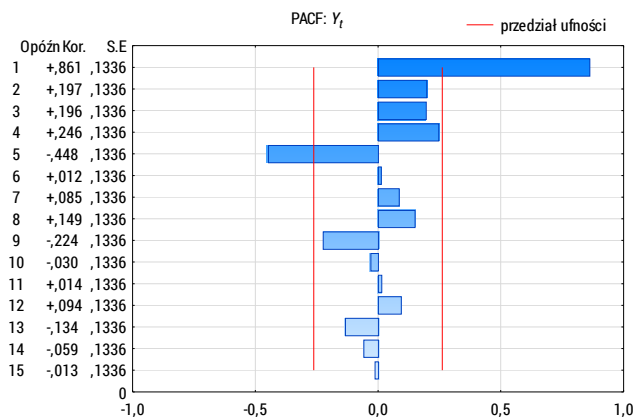
Rynek e-commerce rozwija się bardzo dynamicznie. Analizowany szereg potwierdza systematyczną zmianę preferencji i zachowań konsumentów oraz powiększanie się udziału sprzedaży internetowej. Na wykresie zwracają uwagę też silne wahania okresowe. Analiza danych liczbowych wskazuje, że obserwowany w latach 2005–2019 kwartalny wzrost w stosunku do analogicznego kwartału roku poprzedniego wahał się od ok. 2,5% do aż ok. 20%. Dodatkowo wartości procentowe są realizacją trendu i oznaczają niestałą w czasie średnią. Natomiast bardzo różne tempo zmian wskazuje na niestałą w czasie wariancję. Ponadto udział sprzedaży elektronicznej jest zdecydowanie najwyższy w kwartale czwartym każdego roku i stanowi od ok. 120% do ok. 130% sprzedaży w kwartale trzecim tego samego roku, co z kolei wskazuje na silną sezonowość kwartalną. Potwierdzeniem tych obserwacji są kształty korelogramów przedstawione na rys. 4.29. W przykładzie zdecydowano się zbudować model

na podstawie danych z 2005–2018 (56 obserwacji). Dane z roku 2019 (4 obserwacje) pozostawiono do oceny jakości prognoz modelu.

a)



b)



Rys. 4.29. ACF (a) i PACF (b) szeregu kwartalnej sprzedaży detalicznej w handlu elektronicznym wyrażonej jako procent całkowitej sprzedaży, nieskorygowanego sezonowo 2005–2018

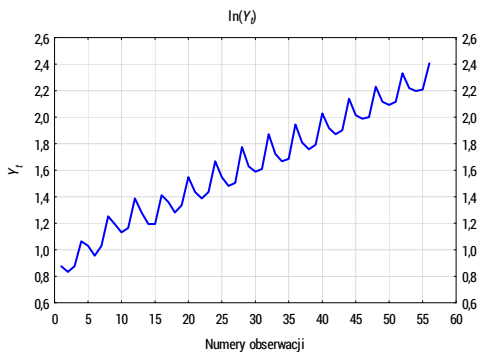
Źródło: opracowanie własne na podstawie danych z Federal Reserve Economic, <https://fred.stlouisfed.org/series/ECOMPCTNSA> [6.08.2021].

W celu uzyskania stacjonarności szeregu ze względu na wariancję, zlogarytmowano szereg (wynik tej operacji pokazano na rys. 4.30a). Otrzymano szereg stacjonarny ze względu na wariancję. Oszacowane ponownie wartości ACF i PACF nie zmieniły się znacząco (rys. 4.30b i 4.30c). Szereg następnie poddano różnicowaniu z opóźnieniem 1 i 4 za każdym razem kreśląc przebieg szeregu oraz korelogramy (rys. 4.31 i 4.32). Przyjęto założenie, że różnicowanie z opóźnieniem 1 pozwoli usunąć trend, zaś z opóźnieniem 4 sezonowość. Warto zaznaczyć, że autokorelacje dla kolejnych opóźnień są wzajemnie zależne i w związku z tym usunięcie niektórych autokorelacji zmienia inne autokorelacje, usuwając lub uwypuklając inne składniki okresowe¹⁵⁶.

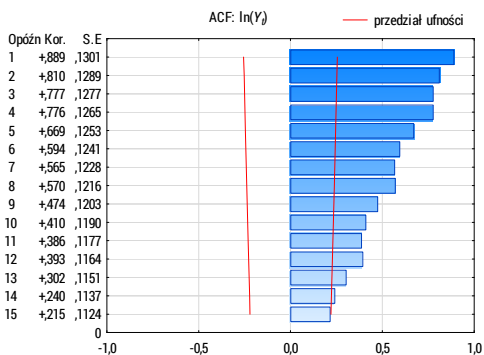
Analizując kolejno wykresy 4.30 - 4.32 można zauważyć, że zlogarytmowany szereg przed różnicowaniem (niestacjonarny ze względu na średnią) ma wyróżniające się wartości PACF dla opóźnień: 5, 9 i 13 (rys. 4.30c). Szereg zróżnicowany z opóźnieniem 1 wykazuje przebieg ACF (rys. 4.31b) charakterystyczny dla procesu sezonowego autoregresji – powtarzające się istotne wartości autokorelacji dla opóźnień będących wielokrotnością 4, tj. 4, 8, 12. Kształt PACF może zaś sugerować niesezonowy proces średniej ruchomej (rys. 4.31c). Szereg udziału sprzedaży zróżnicowany dodatkowo sezonowo z opóźnieniem 4 przedstawiono na rys. 4.32.

¹⁵⁶ S. Łobejko (red.), K. Maślowska, R. Wojdan (2015), *Analiza i prognozowanie...*, op. cit., s. 85.

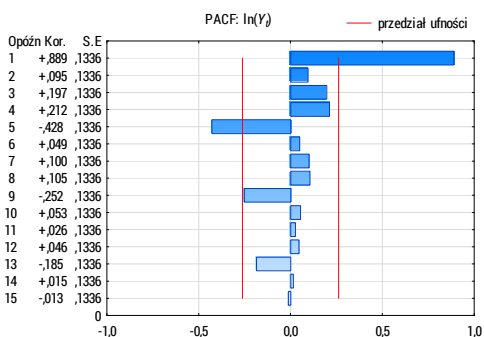
a)



b)



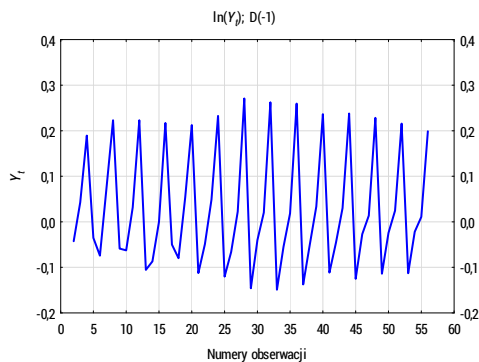
c)



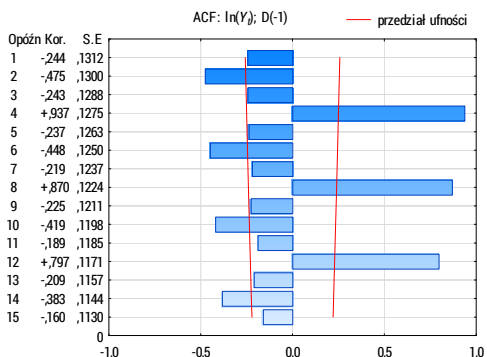
Rys. 4.30. Szereg kwartalnej sprzedaży detalicznej w handlu elektronicznym wyrażonej jako procent całkowitej sprzedaży, nieskorygowanej sezonowo po operacji logarytmowania (a), ACF (b) i PACF (c) zlogarytmowanego szeregu 2005–2018

Źródło: opracowanie własne na podstawie danych z Federal Reserve Economic, <https://fred.stlouisfed.org/series/ECOMPCTNSA> [6.08.2021].

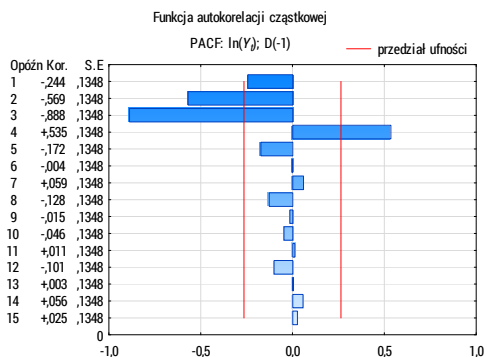
a)



b)



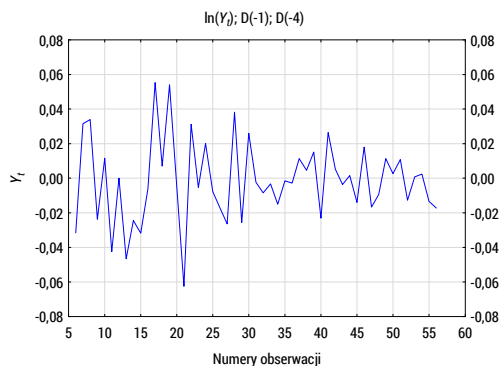
c)



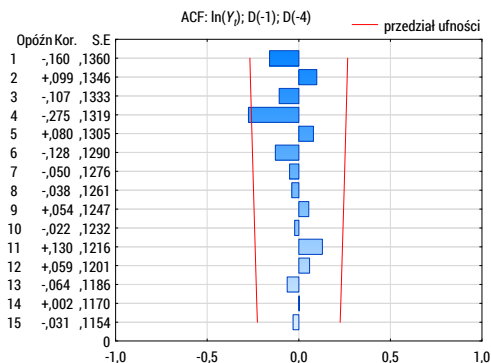
Rys. 4.31. Zlogarytmowany i zróżnicowany z opóźnieniem 1 szereg kwartalnej sprzedaży detalicznej w handlu elektronicznym wyrażonej jako procent całkowitej sprzedaży, nieskorygowanej sezonowo 2005–2018 (a) oraz jego ACF (b) i PACF (c)

Źródło: opracowanie własne na podstawie danych z Federal Reserve Economic, <https://fred.stlouisfed.org/series/ECOMPCTNSA> [6.08.2021].

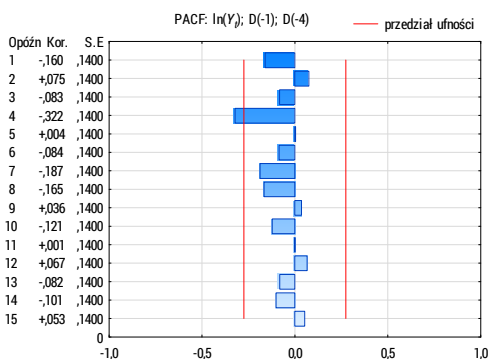
a)



b)



c)



Rys. 4.32. Zlogarytmowany i zróżnicowany z opóźnieniem 1 i 4 szereg kwartalnej sprzedaży detalicznej w handlu elektronicznym wyrażonej jako procent całkowitej sprzedaży, nieskorygowanej sezonowo 2005–2018 (a) oraz jego ACF (b) i PACF (c)

Źródło: opracowanie własne na podstawie danych z Federal Reserve Economic, <https://fred.stlouisfed.org/series/ECOMPCTNSA> [6.08.2021].

Pomiary rzeczywistych szeregów są zawsze obarczone błędami stochastycznymi¹⁵⁷, a możliwość poprawnej identyfikacji zależy od poziomu zakłóceń¹⁵⁸. Bardzo często szereg czasowy można opisać kilkoma modelami ARIMA. Wybór modelu jest często kompromisem między dopasowaniem modelu a prostotą. W tym przykładzie ze względu na wykraczające poza przedział ufności wartości ACF i PACF dla opóźnienia 4 dopasowano początkowo dwa modele: ARIMA(0,1,0)(0,1,1)₄ oraz ARIMA(0,1,0)(1,1,0)₄. Zestawienie wyników estymacji zawarto w tabeli 4.3.

Tabela 4.3. Wyniki estymacji wybranych modeli

Model	Parametr	Prawdopodobieństwo	Błąd standardowy oceny parametru	Błąd kwadratowy
ARIMA(0,1,0)(0,1,1) ₄	$\theta_1 = 0,30542$	0,026	0,13327	0,00052
ARIMA(0,1,0)(1,1,0) ₄	$\phi_1 = -0,2746$	0,053	0,13864	0,00053

Źródło: opracowanie własne.

Uzyskane charakterystyki modelu ARIMA(0,1,0)(0,1,1)₄ są następujące: $\theta_1 = 0,30542$, błąd standardowy oceny parametru 0,13327, najmniejsze prawdopodobieństwo, przy którym odrzucono by hipotezę o istotności $p = 0,026$, resztowy MS = 0,00052. Parametr sezonowego procesu autoregresyjnego ARIMA(0,1,0)(1,1,0)₄ to $\phi_1 = -0,2746$. Model ten ma większy błąd standardowy oceny parametru (0,13864) i wyższą wartość p (0,053), a wartość resztkowego MS to 0,00053. Ze względu na wyższą od zwyczajowo przyjętego poziomu istotności 0,05 wartość statystyki p i nieznacznie, ale jednak mniejszą wartość MS (tym samym większą wartość AIC przy tej samej liczbie szacowanych parametrów), wybrano model ARIMA(0,1,0)(0,1,1)₄:

$$(1 - B)(1 - B^4)Y_t = (1 - \theta_1 B^4)e_t, \quad (4.16)$$

¹⁵⁷ J. Nazarko, M. Rybaczuk, A. Jurczuk (2005), *Wpływ poziomu zakłóceń losowych...*, op. cit.

¹⁵⁸ E. Chodakowska, J. Nazarko, Ł. Nazarko (2021), *ARIMA Models...*, op. cit.

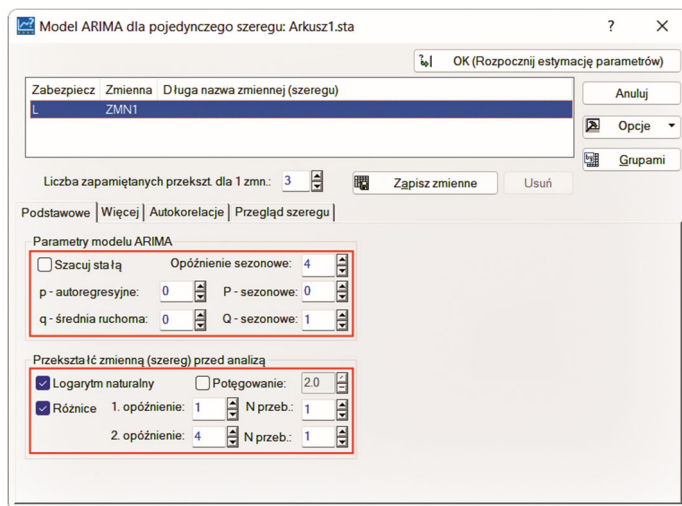
czyli:

$$Y_t = Y_{t-1} + Y_{t-4} - Y_{t-5} - \theta_1 e_{t-4} + e_t. \quad (4.17)$$

Podstawiając obliczoną wartość parametru θ_1 otrzymuje się:

$$Y_t = Y_{t-1} + Y_{t-4} - Y_{t-5} - 0,30542 e_{t-4} + e_t. \quad (4.18)$$

Na rys. 4.33 przedstawiono wybór parametrów w oknie dialogowym programu Statistica Model ARIMA dla pojedynczego szeregu. Podobnie jak w przykładzie pierwszym podstawą estymacji jest zmienna oryginalna, zaś logarytmowanie, jak i różnicowanie jest zintegrowane w modelu i wybierane w oknie dialogowym.

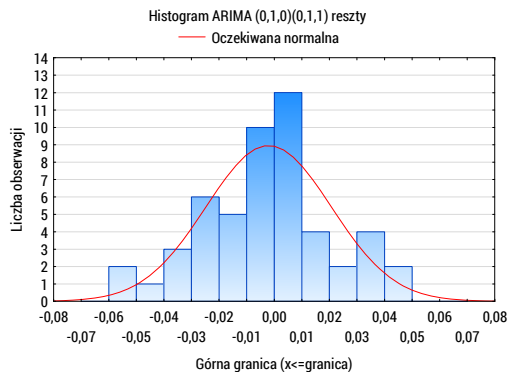


Rys. 4.33. Statistica – okno Model ARIMA dla pojedynczego szeregu

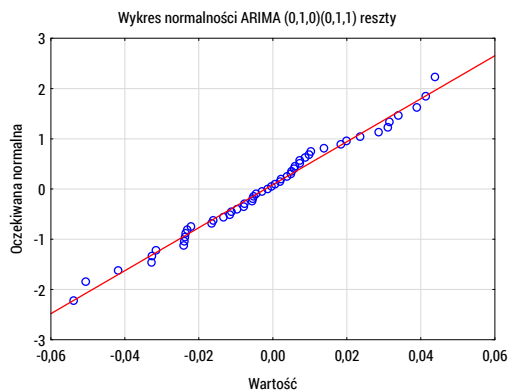
Źródło: opracowanie własne

Przeprowadzona analiza reszt wskazuje, że model jest zadowalający. Rozkład reszt pasuje do rozkładu normalnego (rys. 4.34a i 4.34b). Funkcja autokorelacji nie wykazuje składników systematycznych (rys. 4.34c i 4.35d). Tym samym można przyjąć, że proces reszt nie różni się znacząco od procesu białego szumu.

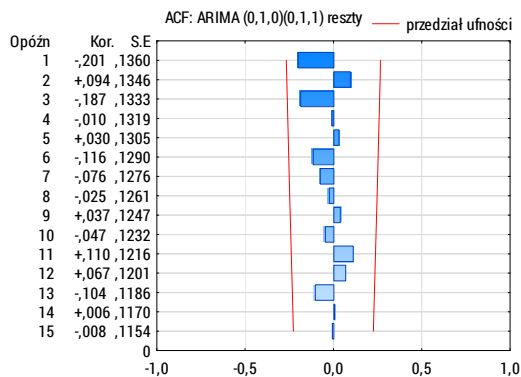
a)



b)

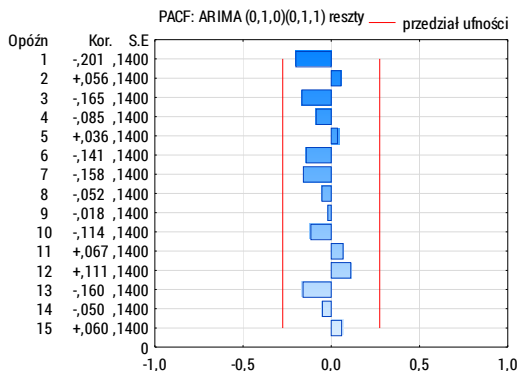


c)



Rys. 4.34. Statistica – ocena reszt modelu: histogram (a), wykres normalności (b), ACF (c)

d)



Rys. 4.34. (cd.) Statistica – ocena reszt modelu: PACF (d)

Źródło: opracowanie własne na podstawie danych z Federal Reserve Economic, <https://fred.stlouisfed.org/series/ECOMPCTNSA> [6.08.2021].

Akceptacja reszt modelu pozwala przejść do etapu budowy prognozy na rok 2019. Wartości prognozy obliczone z modelu oraz rzeczywistą realizację zmiennej zamieszczono w tabeli 4.4 oraz zilustrowano na rys. 4.35.

Tabela 4.4. Prognoza wartości zmiennej i jej realizacja

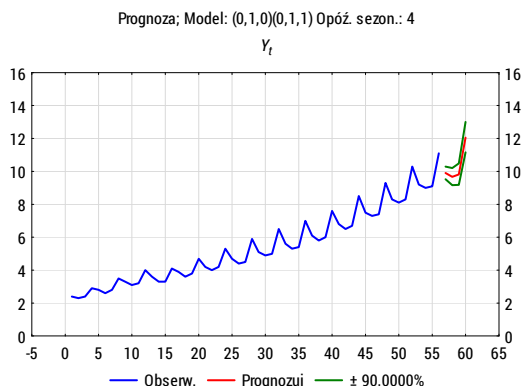
	Realizacja zmiennej Y_t	Prognoza Y^*_{t+1}	Dolne 90%	Górne 90%	Reszty (błąd prognozy) $e_{(t+1)}$
I kwartał 2019	10,1	9,9	9,532	10,289	0,197
II kwartał 2019	9,9	9,7	9,164	10,210	0,227
III kwartał 2019	10,3	9,8	9,186	10,487	0,485
IV kwartał 2019	12,4	12,1	11,166	13,011	0,347

Źródło: opracowanie własne otrzymane za pomocą programu Statistica.

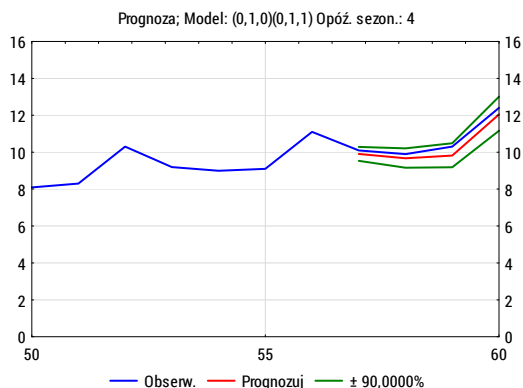
Analiza wartości przedstawionych w tabeli wskazuje, że model nie doszacowuje wartości prognozowanych względem wartości rzeczywistych (zrealizowanych) zmiennej. Można z tego wyciągnąć wniosek, że dynamika zmian wzrostu udziału sprzedaży internetowej w ogólnej sprzedaży zwiększa się i trudno

ją szacować na podstawie lat ubiegłych. Niemniej wartości rzeczywiste mieszczą się w 90% przedziale ufności, czyli model można byłoby wykorzystać w 2018 roku do szacowania udziału w 2019 roku.

a)



b)



Rys. 4.35. Szereg i jego prognoza

Źródło: opracowanie własne na podstawie danych z Federal Reserve Economic, <https://fred.stlouisfed.org/series/ECOMPCTNSA> [6.08.2021].

Pod koniec 2019 roku rozpoczęła się epidemia COVID-19, uznana w 2020 roku przez WHO za pandemię. Podjęte działania w celu ograniczenia zakażenia oraz ich konsekwencje przyspieszyły

jeszcze rozwój e-handlu. Udział wzrósł dynamicznie¹⁵⁹ i tym samym model z pewnością będzie musiał zostać zaktualizowany w celu oszacowania prognoz na lata kolejne. Spekulując o długoterminowych skutkach i kolejnych realizacjach procesu jednym z rozwiązań może być zastosowanie rozwinięcia modeli ARIMA w postaci modeli z interwencją¹⁶⁰. W pakiecie Statistica dostępne są trzy główne typy możliwych zmian: (1) nagły, trwały, (2) stopniowy, trwały oraz (3) nagły, zanikający. Temat modeli z interwencją wykracza jednak poza zakres niniejszego podręcznika.

4.2.3. Sezonowy model ARIMA(0,1,1)(2,1,0)₁₂

Ostatni przykład jest inspirowany szeregiem obrazującym pasażerów międzynarodowych linii lotniczych w latach 1949–1960 przytoczonym jako przykład niestacjonarnego sezonowego szeregu czasowego pod nazwą szereg G w G.E.P. Box, G.M. Jenkins (1976), *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day¹⁶¹. Jest on najczęściej omawianym zbiorem obserwacji w kontekście modeli ARIMA (tab. 4.5, rys. 4.36). Szereg ten posiada bardzo wyraźną sezonowość, a dobrany model ARIMA(0,1,1)(0,1,1)₁₂ jest znany jako model pasażerów linii lotniczej (ang. *the airline model*).

¹⁵⁹ Federal Reserve Economic, <https://fred.stlouisfed.org/series/ECOMPCTNSA> [6.8.2021]

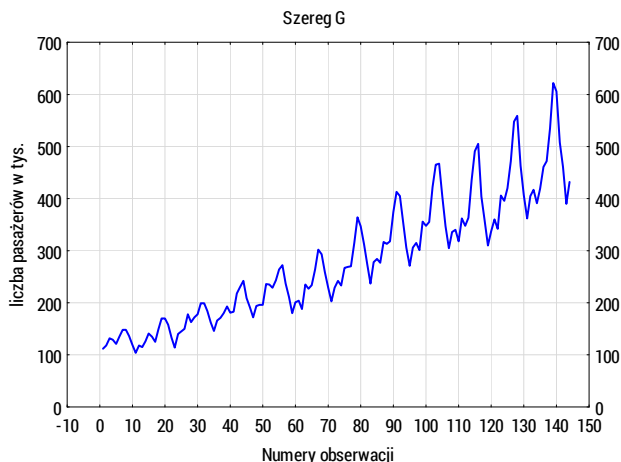
¹⁶⁰ R.K. Raman, T.V. Sathianandan, A.P. Sharma, B.P. Mohanty (2017), *Modelling and Forecasting Marine Fish Production in Odisha Using Seasonal ARIMA Model*, „National Academy Science Letters” 40, s. 393–397 doi: 10.1007/s40009-017-0581-2; L. Wu, H. Hayashi (2014), *The Impact of Disasters on Japan's Inbound Tourism Demand*, „Journal of Disaster Research” 9, s. 699–708, doi: 10.20965/jdr.2014.p0699.

¹⁶¹ G.E.P. Box, G.M. Jenkins, G.C. Reinsel, G.M. Ljung, (2015), *Time Series Analysis...*, op. cit.

Tabela 4.5. Liczba pasażerów międzynarodowych linii lotniczych w tysiącach (szereg G)

	1949	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960
I	112	115	145	171	196	204	242	284	315	340	360	417
II	118	126	150	180	196	188	233	277	301	318	342	391
III	132	141	178	193	236	235	267	317	356	362	406	419
IV	129	135	163	181	235	227	269	313	348	348	396	461
V	121	125	172	183	229	234	270	318	355	363	420	472
VI	135	149	178	218	243	264	315	374	422	435	472	535
VII	148	170	199	230	264	302	364	413	465	491	548	622
VIII	148	170	199	242	272	293	347	405	467	505	559	606
IX	136	158	184	209	237	259	312	355	404	404	463	508
X	119	133	162	191	211	229	274	306	347	359	407	461
XI	104	114	146	172	180	203	237	271	305	310	362	390
XII	118	140	166	194	201	229	278	306	336	337	405	432

Źródło: SAS Institute Inc. https://documentation.sas.com/doc/en/etscdc/14.2/etsug/etsug_code_ariex02.htm [9.08.2021] za: G.E.P. Box, G.M. Jenkins (1976), *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day.

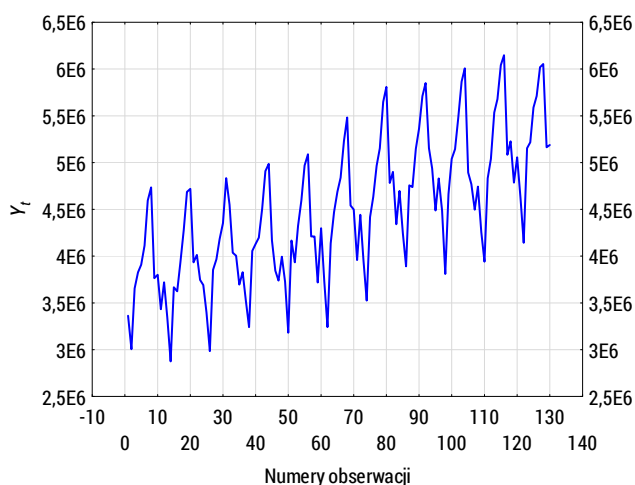


Rys. 4.36. Liczba pasażerów międzynarodowych linii lotniczych (szereg G)

Źródło: opracowanie własne na podstawie: SAS Institute Inc. https://documentation.sas.com/doc/en/etscdc/14.2/etsug/etsug_code_ariex02.htm [9.08.2021] za: G.E.P. Box, G.M. Jenkins (1976), *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day.

Estymując model zlogarytmowanego szeregu, to znaczy $(1 - B)(1 - B^{12})Y_t = \theta(B^{12})\theta(B)e_t$, w pakiecie Statistica za pomocą metody dokładnej Mélardarda otrzymuje się: $\theta_1 = 0,55694$ oraz $\theta_1 = 0,40182$, przy $MS = 0,00142$.

W tabeli 4.6 przedstawiono pasażerów komercyjnych lotniska John F. Kennedy International Airport w latach 2009–2019. Można zauważyć, że zwyczaje pasażerów i preferencje co do podróży w miesiącach letnich nie zmieniły się istotnie w XXI w. Liczba pasażerów systematycznie rośnie, ale nie ma już znaczących różnic w wariancji szeregu (rys. 4.37). Tym samym można przyjąć, że szereg jest stacjonarny ze względu na wariancję.



Rys. 4.37. Liczba pasażerów komercyjnych lotniska John F. Kennedy International Airport w latach 2009–2019

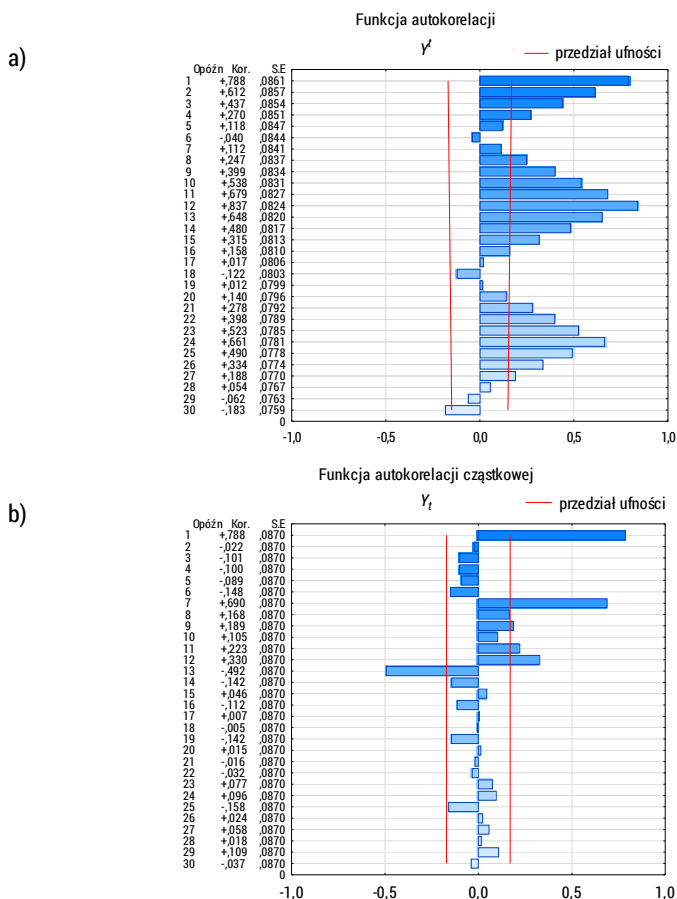
Źródło: opracowanie własne na podstawie: Annual Airport Traffic Report z lat 2009–2019, <https://www.panynj.gov/airports/en/statistics-general-info.html> [9.08.2021].

Tabela 4.6. Liczba pasażerów komercyjnych lotniska John F. Kennedy International Airport

	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
I	3003610	2874742	2984436	3239977	3182186	3240578	3525008	3890737	3808065	3940279	4142980
II	3654594	3666012	3854583	4054413	4167173	4134490	4417086	4755865	4663061	4832541	5153203
III	3825614	3624713	3965636	4131110	3934978	4469904	4634132	4737920	5038142	5038239	5217506
IV	3908798	3932920	4182902	4196869	4323261	4688566	4965684	5147681	5142597	5533739	5587477
V	4113916	4282071	4353031	4508524	4596311	4838571	5153820	5362291	5477601	5679229	5711772
VI	4594578	4687045	4832825	4909743	4968173	5231355	5647863	5708707	5865760	6041910	6018658
VII	4734772	4718586	4549014	4986731	5088447	5482066	5809724	5849941	6007903	6148330	6053873
VIII	3763901	3933255	4036426	4167416	4210409	4540212	4782097	5152464	4893624	5084305	5164589
IX	3800620	4013193	4003395	3844052	4209188	4498493	4900066	4936656	4768329	5228051	5186908
X	3431917	3745653	3696617	3738653	3717622	3957769	4342058	4488489	4496948	4785508	4591203
XI	3719804	3690195	3828426	3992540	4299071	4442421	4697493	4830679	4743659	5055084	5101501
XII	3345769	3396238	3522705	3726946	3730108	3952123	4244083	4486811	4256541	4621402	4657768

Źródło: Annual Airport Traffic Report z lat 2009–2019, <https://www.panynj.gov/airports/en/statistics-general-info.html> [9.08.2021].

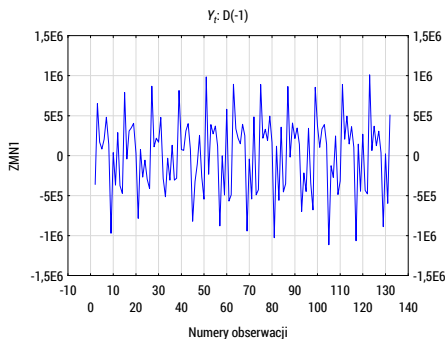
Na rysunku 4.38 przedstawiono wykresy autokorelacji i autokorelacji cząstkowej szeregu. Oprócz trendu główną cechą charakterystyczną szeregu jest składowa okresowa. W analizowanym przykładzie zastosowano kolejno różnicowanie z opóźnieniem 1 oraz sezonowe z opóźnieniem 12. Wyniki operacji zilustrowano na rys. 4.39 oraz 4.40.



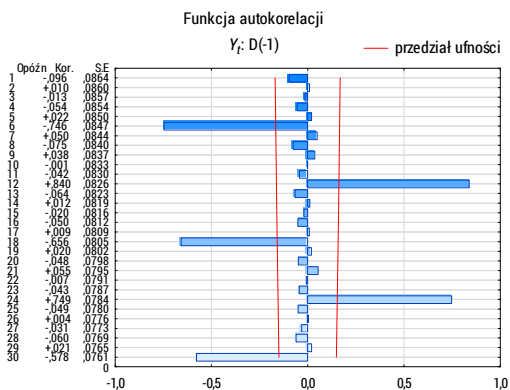
Rys. 4.38. ACF (a) i PACF (b) szeregu liczby pasażerów komercyjnych lotniska John F. Kennedy International Airport w latach 2009–2019

Źródło: opracowanie własne na podstawie: Annual Airport Traffic Report z lat 2009–2019, <https://www.panynj.gov/airports/en/statistics-general-info.html> [9.08.2021].

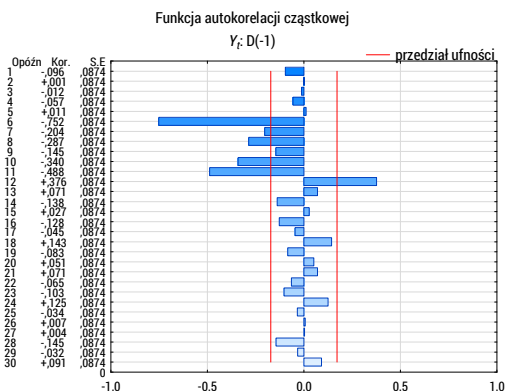
a)



b)



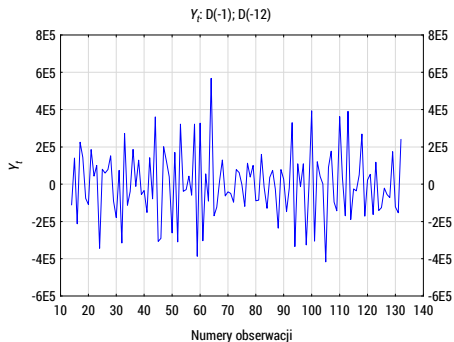
c)



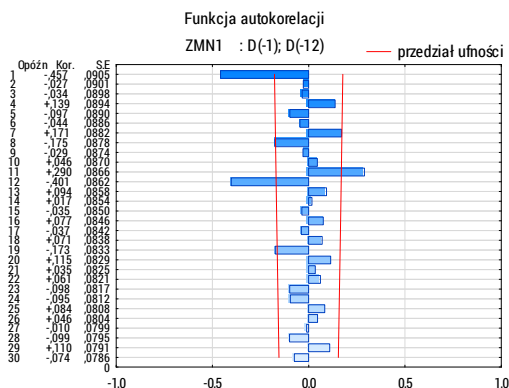
Rys. 4.39. Zróżnicowany z opóźnieniem 1 szereg liczby pasażerów komercyjnych lotniska John F. Kennedy International Airport (a), ACF (b), PACF (c)

Źródło: opracowanie własne na podstawie: Annual Airport Traffic Report z lat 2009–2019, <https://www.panynj.gov/airports/en/statistics-general-info.html> [9.08.2021].

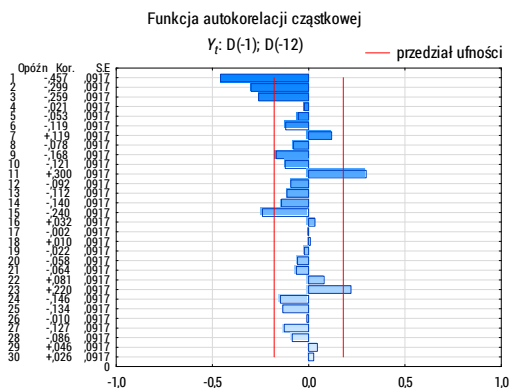
a)



b)



c)



Rys. 4.40. Zróżnicowany z opóźnieniem 1 i 12 szereg liczby pasażerów komercyjnych lotniska John F. Kennedy International Airport (a), ACF (b), PACF (c)

Źródło: opracowanie własne na podstawie: Annual Airport Traffic Report z lat 2009–2019, <https://www.panynj.gov/airports/en/statistics-general-info.html> [9.08.2021].

Przebieg ACF wskazuje na model niesezonowy MA (zanikające wartości funkcji PACF i pojedyncza wartość ACF dla opóźnienia 1) oraz sezonowy AR (powtarzające się, znaczące dla opóźnienia będącego wielokrotnością 12), tj. $ARIMA(0,1,1)(1,1,0)_{12}$. Wyniki estymacji metodą przybliżoną McLeoda i Salesa przedstawiono na rys. 4.41.

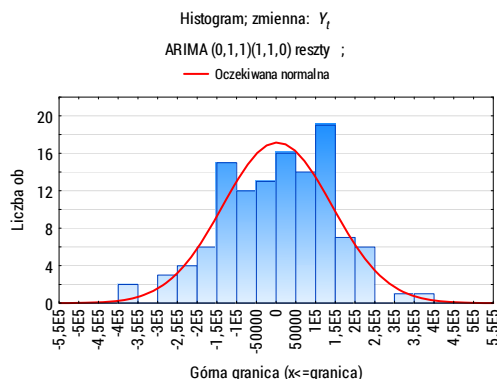
```
Zmienna: ZMN1
Przekształcenia: D(1),D(12)
Model: (0,1,1)(1,1,0) Opóźn. sez.: 12
Liczba obs.:119 Wstępne SS=397E10 Końc. SS=226E10 (56,87%) MS=1930E7
Parametry (p/Ps-autoregresyjne, q/Qs-średniej ruch.) podświetl: p<.05
q(1) Ps(1)
Ocena: ,69265 -,4118
Błąd std: ,07080 ,08921
```

Rys. 4.41. Wyniki estymacji modelu $ARIMA(0,1,1)(1,1,0)_{12}$ szeregu liczby pasażerów komercyjnych lotniska John F. Kennedy International Airport

Źródło: opracowanie własne na podstawie: Annual Airport Traffic Report z lat 2009–2019, <https://www.panynj.gov/airports/en/statistics-general-info.html> [9.08.2021].

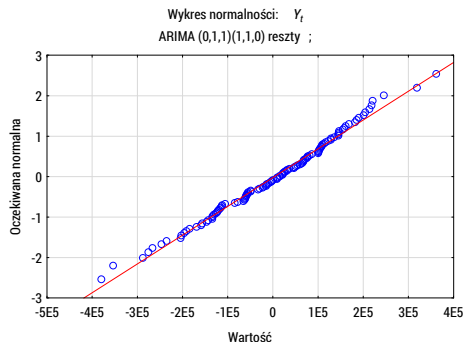
Analizując reszty można zaobserwować, że nie wszystkie systematyczne składowe szeregu zostały ujęte w modelu. W szeregu reszt pozostaje autokorelacja przy opóźnieniu 24 (rys. 4.42).

a)

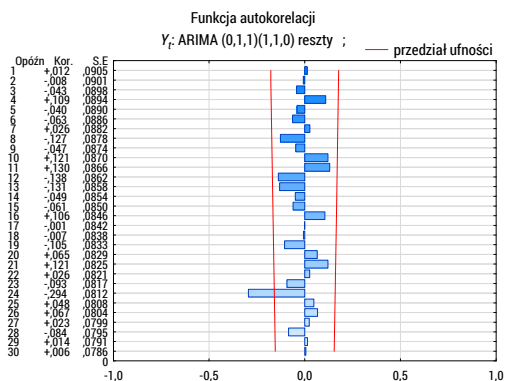


Rys. 4.42. Analiza reszt modelu $ARIMA(0,1,1)(1,1,0)_{12}$ szeregu liczby pasażerów komercyjnych lotniska John F. Kennedy International Airport: histogram (a)

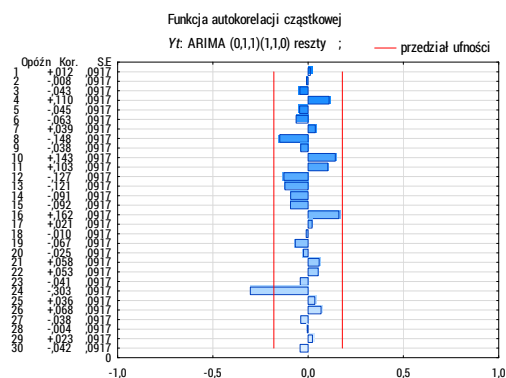
b)



c)



d)



Rys. 4.42. (cd.) Analiza reszt modelu ARIMA(0,1,1)(1,1,0)₁₂ szeregu liczby pasażerów komercyjnych lotniska John F. Kennedy International Airport: wykres normalności (b), ACF (c), PACF (d)

Źródło: opracowanie własne na podstawie: Annual Airport Traffic Report z lat 2009–2019, <https://www.panynj.gov/airports/en/statistics-general-info.html> [9.08.2021].

Dlatego zdecydowano się rozszerzyć model, zmieniając rząd procesu autoregresji i tym samym dodając dodatkowy sezonowy parametr autoregresyjny: $ARIMA(0,1,1)(2,1,0)_{12}$. Parametry tego modelu estymowane metodą przybliżoną przedstawiono na rys. 4.43.

```

Zmienna: ZMN1
Przekształcenia: D(1),D(12)
Model: (0,1,1)(2,1,0) Opóźn. sez.: 12
Liczba obs.:119 Wstępne SS=397E10 Końc. SS=194E10 (48,84%) MS=1671E7
Parametry (p/Ps-autoregresyjne, q/Qs-średniej ruch.) podświetl: p<.05
      q(1) Ps(1) Ps(2)
Ocena: ,68520 -,5625 -,4236
Błąd std: ,06922 ,08995 ,09737

```

Rys. 4.43. Wyniki estymacji modelu $ARIMA(0,1,1)(2,1,0)_{12}$ szeregu liczby pasażerów komercyjnych lotniska John F. Kennedy International Airport

Źródło: opracowanie własne na podstawie: Annual Airport Traffic Report z lat 2009–2019, <https://www.panynj.gov/airports/en/statistics-general-info.html> [9.08.2021].

Opracowany model ma postać:

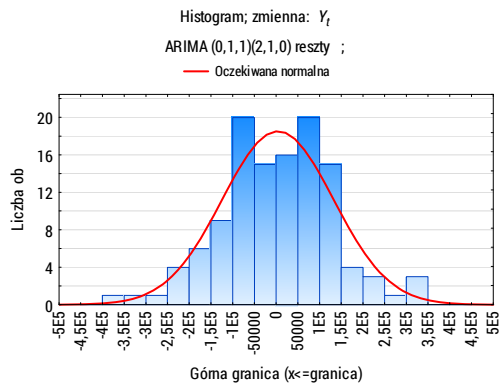
$$(1 - \phi_1 B^{12} - \phi_2 B^{24})(1 - B)(1 - B^{12})Y_t = (1 - \theta_1 B)e_t. \quad (4.19)$$

Podstawiając wartości oszacowanych parametrów otrzymuje się:

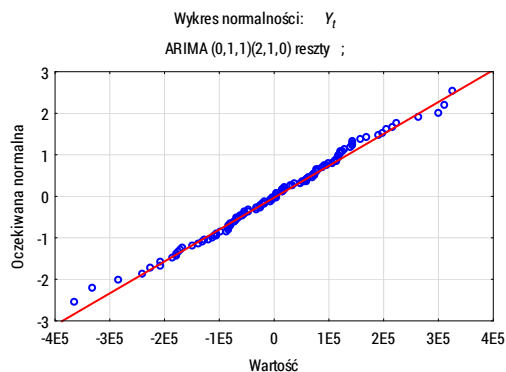
$$\begin{aligned} (1 + 0,5625B^{12} + 0,4236B^{24})(1 - B)(1 - B^{12})Y_t = \\ = (1 - 0,06852B)e_t. \end{aligned} \quad (4.20)$$

Diagnostyka wizualna reszt modelu (rys. 4.44) pozwala stwierdzić, że w wypadku modelu $ARIMA(0,1,1)(2,1,0)_{12}$ reszty spełniają założenie losowości reszt modelu.

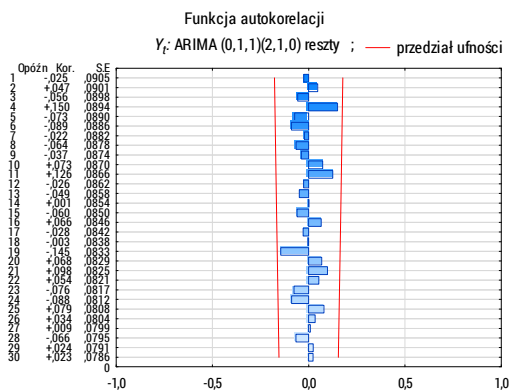
a)



b)

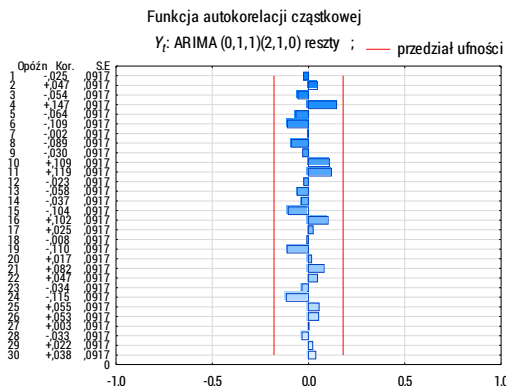


c)



Rys. 4.44. Analiza reszt modelu $ARIMA(0,1,1)(2,1,0)_{12}$ szeregu liczby pasażerów komercyjnych lotniska John F. Kennedy International Airport: histogram (a), wykres normalności (b), ACF (c)

d)



Rys. 4.44. (cd.) Analiza reszt modelu ARIMA(0,1,1)(2,1,0)₁₂ szeregu liczby pasażerów komercyjnych lotniska John F. Kennedy International Airport: PACF (d)

Źródło: opracowanie własne na podstawie: Annual Airport Traffic Report z lat 2009–2019, <https://www.panynj.gov/airports/en/statistics-general-info.html> [9.08.2021].

Oznacza to, że wybrany model ARIMA(0,1,1)(2,1,0)₁₂ można wykorzystać do prognozowania. Podobnie jednak jak w wypadku szeregu z przykładu 4.2.2 epidemia COVID-19 zmieniła całkowicie dotychczasowe wzorce zachowań. W roku 2020 nastąpiło załamanie rynku lotniczego. Można wyrazić nadzieję, że model będzie trafnie opisywać ruch pasażerski dopiero w kolejnych latach.

Przytoczone przykłady budowy modeli rzeczywistego szeregu czasowego pokazują, że modele ARIMA są uniwersalnym narzędziem modelowania i prognozowania. Niemniej identyfikacja modeli ARIMA nie jest ściśle skodyfikowana i w dużej mierze zależy od wiedzy, doświadczenia i intuicji badacza¹⁶².

Kluczowe zagadnienia:

usuwanie niestacjonarności, wybór modelu,
losowość reszt modelu, prognoza

¹⁶² E. Chodakowska, J. Nazarko, Ł. Nazarko (2021), *ARIMA Models...*, op. cit.

Literatura podstawowa

- [1] Cieślak M. (red.) (2011), *Prognozowanie gospodarcze. Metody i zastosowania*, Wydanie czwarte zmienione, Wydawnictwo Naukowe PWN, Warszawa
- [2] Box G.E.P., Jenkins G.M., Reinsel G.C., Ljung G.M., (2015), *Time Series Analysis: Forecasting and Control. Fifth Edition*, John Wiley & Sons

Literatura uzupełniająca

- [1] Chodakowska E. (2002), *Zastosowanie pakietu STATISTICA w prognozowaniu na podstawie szeregów czasowych*, „Zeszyty Naukowe Politechniki Białostockiej. Ekonomia i Zarządzanie” 7, s. 77–86
- [2] Chodakowska E., Nazarko J., Nazarko Ł. (2021), *ARIMA Models in Electrical Load Forecasting and Their Robustness to Noise*, „Energies” 14(23), 7952, s.1–22, doi: 10.3390/en14237952
- [3] Fattah J., Ezzine L., Aman Z., El Moussami H., Lachhab A. (2018), *Forecasting of demand using ARIMA model*, „International Journal of Engineering Business Management” 10, s. 1–9, doi: 10.1177/1847979018808673
- [4] Łobejko S. (red.), Maśłowska K., Wojdan R. (2015), *Analiza i prognozowanie szeregów czasowych z programem SAS*, Szkoła Główna Handlowa w Warszawie – Oficyna Wydawnicza, Warszawa, https://ssl-administracja.sgh.waw.pl/pl/OW/publikacje/Documents/Analiza_lobejko_maslowska_wojdan.pdf [27.07.2020]
- [5] Nazarko J., Rybaczuk M., Jurczuk A. (2005), *Wpływ poziomu zakłóceń losowych na możliwość identyfikacji modeli ARIMA*, „Taksonomia” 12, s. 238–247
- [6] Raman R.K., Sathianandan T.V., Sharma A.P., Mohanty B.P. (2017), *Modelling and Forecasting Marine Fish Production in Odisha Using Seasonal ARIMA Model*, „National Academy Science Letters” 40, s. 393–397 doi:10.1007/s40009-017-0581-2
- [7] Wu L., Hayashi H. (2014), *The Impact of Disasters on Japan's Inbound Tourism Demand*, „Journal of Disaster Research” 9, s. 699–708, doi: 10.20965/jdr.2014.p0699

Strony internetowe

- [1] Federal Reserve Economic, <https://fred.stlouisfed.org/series/ECOMPCTNSA> [6.08.2021]
- [2] CRAN: <https://cran.r-project.org/> [12.12.2021]
- [3] MATHWorks: <https://uk.mathworks.com/> [12.12.2021]
- [4] XLSTAT: <https://www.xlstat.com/en/> [12.12.2021]
- [5] SAS: https://www.sas.com/pl_pl/home.html [12.12.2021]
- [6] SPSS Polska: <https://spss.pl/> [12.12.2021]
- [7] WOLFRAM: <https://www.wolfram.com/mathematica/> [12.12.2021]

5. PRZYKŁADY ZASTOSOWAŃ

Modele klasy ARIMA znajdują szerokie zastosowanie progностyczne w wielu obszarach gospodarczych, społecznych i technicznych. W kolejnych podrozdziałach tego rozdziału przedstawiono zaczerpnięte z literatury praktyczne przykłady wykorzystania modeli ARIMA w prognozowaniu różnorodnych zjawisk. Poniższy przegląd ma na celu uświadomienie czytelnikowi szerokiego spektrum możliwych zastosowań i implementowanych typów modeli klasy ARIMA.

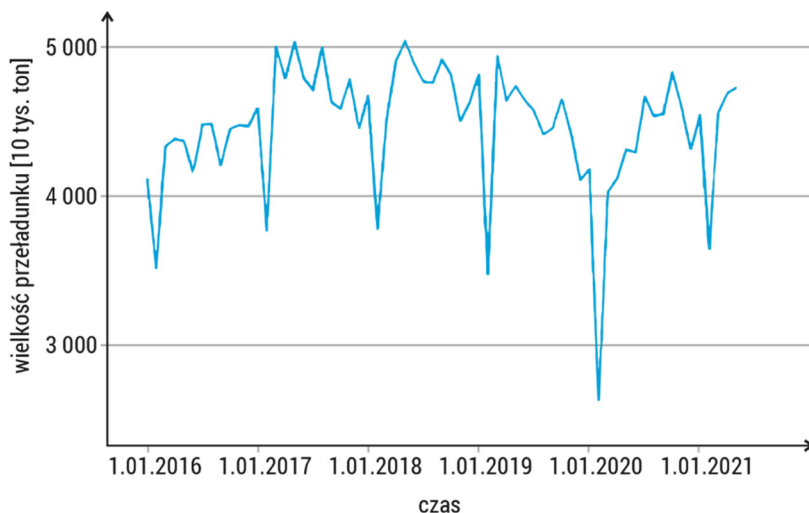
5.1. Prognozowanie przepustowości portu

Przykład zaczerpnięto z artykułu J. Zhang, *Forecast of Shanghai Port Throughput Based on ARIMA*¹⁶³.

Port w Szanghaju jest największym na świecie pod względem wielkości przeładunku kontenerowego¹⁶⁴. W artykule przedstawiono model oraz prognozę całkowitej przepustowości portu. Analizą objęto dane miesięczne za okres od stycznia 2016 do maja 2021 roku (rys. 5.1).

¹⁶³ J. Zhang (2021), *Forecast of Shanghai Port Throughput Based on ARIMA*, IOP Conf. Ser.: Earth Environ. Sci. 831 012042, doi: 10.1088/1755-1315/831/1/012042.

¹⁶⁴ Międzynarodowe Centrum Wysyłkowe w Szanghaju w pierwszej trójce na świecie, China Radio International: <http://polish.cri.cn/wideo/3868/20200713/505021.html>, 2020-07-13 [29.11.2021].

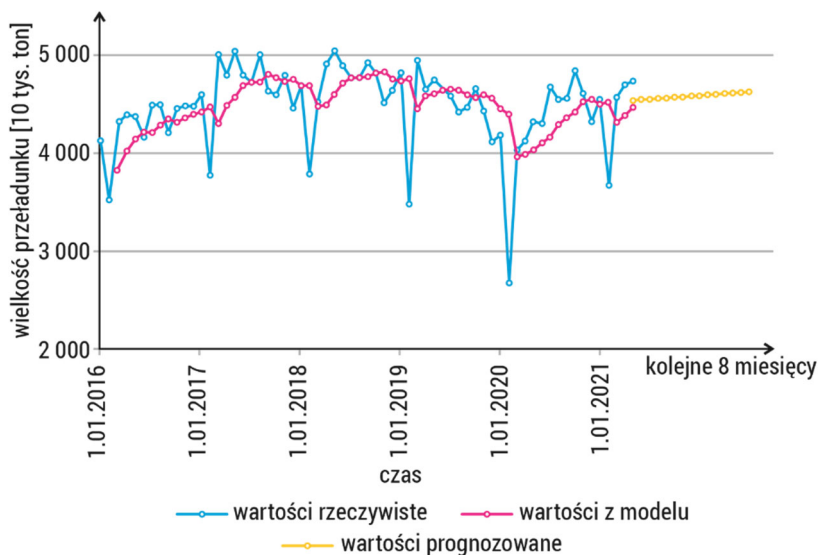


Rys. 5.1. Miesięczne wielkości przeładunków w porcie Szanghaj w okresie od stycznia 2016 do maja 2021 r.

Źródło: J. Zhang (2021), *Forecast of Shanghai Port Throughput Based on ARIMA*, IOP Conf. Ser.: Earth Environ. Sci. 831 012042, doi: 10.1088/1755-1315/831/1/012042.

Na podstawie analizy przebiegu funkcji ACF i PACF szeregu czasowego wielkości przeładunków, do jego modelowania zaproponowano model ARIMA(0,1,1). W artykule nie podano wartości parametrów modelu. Testy dopasowania modelu (m.in. R^2 , RMSE, MAE, BIC) spełniły oczekiwania autora. Pozytywnie zweryfikowano też hipotezę o losowości reszt modelu.

Następnie model posłużył do wykonania prognozy na piętnaście kolejnych miesięcy. Wyniki modelowania szeregu oraz wyznaczone prognozy przedstawiono na rys. 5.2.



Rys. 5.2. Szereg czasowy miesięcznych wielkości przeładunków w porcie Szanghaj, wartości uzyskane z modelu oraz prognozowane wielkości przeładunków

Źródło: J. Zhang (2021), *Forecast of Shanghai Port Throughput Based on ARIMA*, IOP Conf. Ser.: Earth Environ. Sci. 831, 012042, doi:10.1088/1755-1315/831/1/012042.

Prognoza wskazuje na dalszy oczekiwany wzrost przeładunków w szanghajskim porcie. Powinno to być podstawą do podjęcia działań w kierunku większej automatyzacji terminali kontenerowych w celu zwiększenia efektywności operacyjnej portu.

5.2. Prognozowanie liczby zachorowań na COVID-19

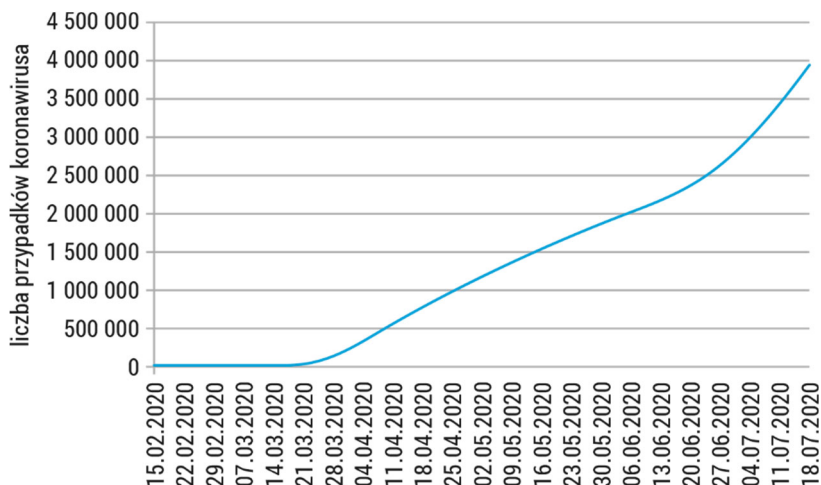
Przykład zaczerpnięto z artykułu A. K Sahai i in., *ARIMA modelling & forecasting of COVID-19 in top five affected countries*¹⁶⁵.

Pandemia COVID-19, która rozpoczęła się w grudniu 2019 roku w Wuhan w Chinach, rozprzestrzeniła się na cały świat

¹⁶⁵ A.K. Sahai, N. Rath, V. Sood, M.P. Singh (2020), *ARIMA modelling & forecasting of COVID-19 in top five affected countries*, „Diabetes & Metabolic Syndrome: Clinical Research & Reviews” 14(5), s. 1419–1427, doi: 10.1016/j.dsx.2020.07.042.

i w ciągu kolejnych miesięcy spowodowała bezprecedensowe spustoszenie. W artykule przedstawiono model oraz prognozę całkowitej liczby osób chorych na COVID-19 w pięciu krajach: USA, Brazylii, Indiach, Rosji i Hiszpanii. Analizą objęto dane dzienne za okres od 15 lutego do 18 lipca 2020 roku¹⁶⁶.

Na rysunku 5.3 przedstawiono szereg dziennej całkowitej liczby chorych na COVID-19 w USA w badanym okresie.



Rys. 5.3. Szereg czasowy dziennej całkowitej liczby chorych na COVID-19 w USA w okresie od 15 lutego do 18 lipca 2020 roku

Źródło: opracowanie własne na podstawie Worldmeter: <https://www.worldometers.info/coronavirus/>.

Na podstawie wstępnej analizy przebiegu szeregów czasowych oraz ich funkcji ACF i PACF zdecydowano o wyborze modeli typu $ARIMA(p,d,q)$. Estymacji parametrów i wyboru najlepszych modeli dla poszczególnych szeregów wykonano za pomocą oprogramowania JMulti¹⁶⁷, wykorzystującego algorytm Hannana-Rissanena¹⁶⁸. Do estymacji parametrów modelu wykorzystano

¹⁶⁶ Worldmeter: <https://www.worldometers.info/coronavirus/> [29.11.2021].

¹⁶⁷ JMulti: <http://www.jmulti.de/> [29.11.2021].

¹⁶⁸ E.J. Hannan, J. Rissanen (1982), *Recursive estimation of mixed autoregressive-moving average order*, „Biometrika” 69, s. 81–94, doi: 10.1093/biomet/69.1.81.

zmienne z zakresu od 15 lutego do 30 czerwca 2020 roku. Jako kryterium wyboru modelu posłużono się wartością AIC.

Modele wybrane dla poszczególnych krajów przedstawiono w tabeli 5.1.

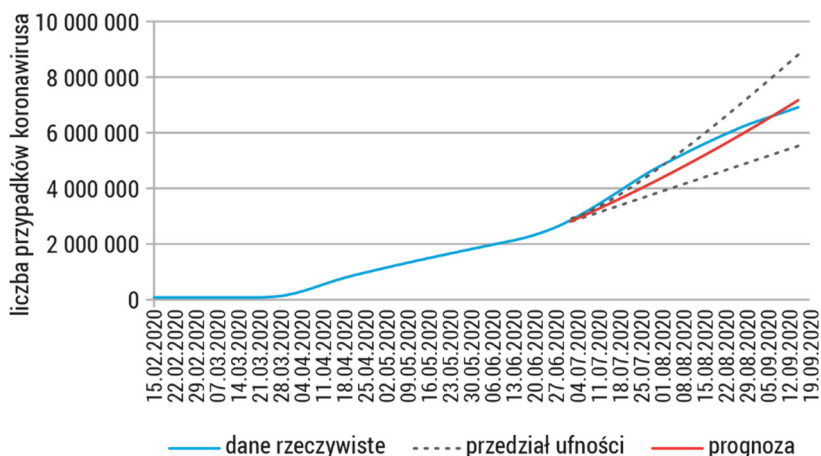
Tabela 5.1. Specyfikacja modeli ARIMA dla badanych krajów

	Indie	Brazylia	Rosja	Hiszpania	USA
Model	(4,2,4)	(3,1,2)	(3,0,0)	(4,2,4)	(1,2,1)
AR1	1,3759	1,3822	1,8218	-0,1032	0,2243
AR2	-1,0232	-0,8340	-0,6470	0,2658	-
AR3	1,3695	0,4483	-0,1757	-0,4359	-
AR4	-0,7242	-	-	-0,6305	-
MA1	1,7944	0,8246	-	0,3806	0,5943
MA2	-1,5084	-0,1108	-	0,1500	-
MA3	1,5919	-	-	-0,3751	-
MA4	-0,8983	-	-	-0,4913	-

Źródło: opracowanie własne na podstawie: A.K. Sahai, N. Rath, V. Sood, M.P. Singh (2020), *ARIMA modelling & forecasting of COVID-19 in top five affected countries*, „Diabetes & Metabolic Syndrome: Clinical Research & Reviews” 14(5), s. 1419–1427, doi: 10.1016/j.dsx.2020.07.042.

Testy dopasowania modeli oraz prognozy wygaśle wykonane dla okresu 1–18 lipca 2020 roku potwierdziły dobrą jakość modeli.

Na podstawie opracowanych modeli wykonano dla poszczególnych krajów prognozy na 77 dni w przód do 15 września 2020 roku. Na rysunku 5.4 przedstawiono prognozę całkowitej liczby chorych na COVID-19 w USA wraz z danymi rzeczywistymi.



Rys. 5.4. Prognoza dziennej całkowitej liczby chorych na COVID-19 w USA na okres od 19 lipca do 15 września 2020 roku

Źródło: opracowanie własne na podstawie: A.K. Sahai, N. Rath, V. Sood, M.P. Singh (2020), *ARIMA modelling & forecasting of COVID-19 in top five affected countries*, „Diabetes & Metabolic Syndrome: Clinical Research & Reviews” 14(5), s. 1419–1427, doi: 10.1016/j.dsx.2020.07.042.

Prognoza wskazuje na szybkie zwiększanie się liczby chorych i dalszy rozwój pandemii COVID-19 w kolejnych dwóch miesiącach.

5.3. Prognozowanie parytetu dwóch walut

Przykład zaczerpnięto z artykułu C. U. Yildiran, A. Fettahoğlu, *Forecasting USDTRY rate by ARIMA method*¹⁶⁹.

Parytet dwóch walut określa stosunek wymiany między tymi walutami. W systemie płynnych kursów walutowych stosunek wartości dwóch walut ustalony jest na giełdzie walutowej na podstawie relacji podaży i popytu tych walut. Ponieważ międzynarodowe transakcje gospodarcze rozliczane są często w różnych walutach, przewidywanie przyszłych wartości kursów walutowych

¹⁶⁹ C.U. Yildiran, A. Fettahoğlu (2017), *Forecasting USDTRY rate by ARIMA method*, „Cogent Economics & Finance” 5(1), 1335968, doi: 10.1080/23322039.2017.1335968.

jest bardzo ważne dla niemal każdego podmiotu gospodarczego, a w szczególności dla banków i rządów.

W artykule zastosowano modelowanie ARIMA do prognozowania parytetu dolara amerykańskiego (USD) i liry tureckiej (TRY). Jako dane do budowy i testowania modelu posłużyło 3069 dziennych obserwacji z okresu od 3 stycznia 2005 roku do 8 marca 2017 roku¹⁷⁰. Szereg czasowy parytetu walut w badanym okresie przedstawiono na rysunku 5.5.



Rys. 5.5. Dienne wartości parytetu USD/TRY w okresie od 3 stycznia 2005 roku do 8 marca 2017 roku

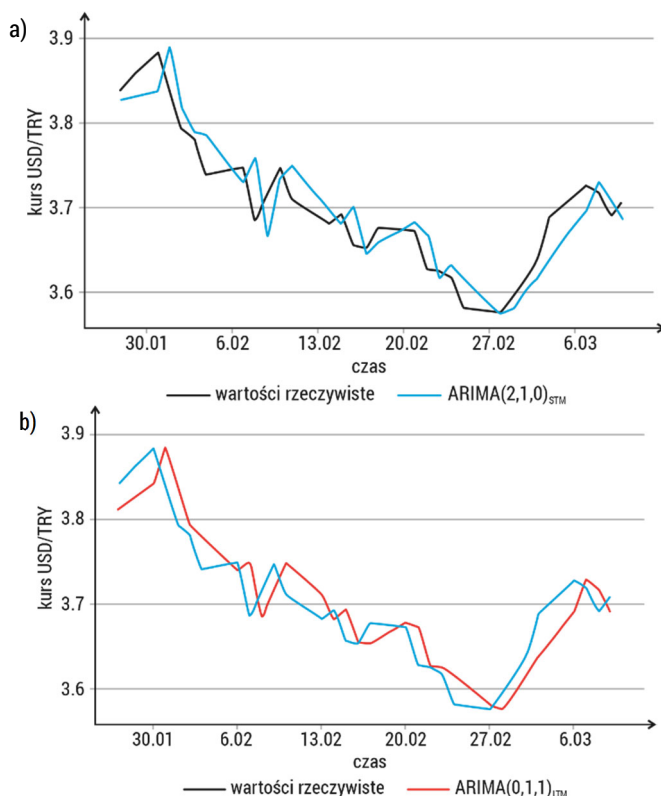
Źródło: opracowanie własne na podstawie Central Bank of the Republic of Turkey: https://evds2.tcmb.gov.tr/index.php?/evds/serieMarket/#collapse_2 [29.11.2021].

W celu osiągnięcia stacjonarności, oryginalny szereg został zlogarytmowany a następnie zróżnicowany. Autorzy założyli opracowanie dwóch różnych modeli, opartych na długim (LTM) i krótkim (STM) szeregu danych. Do budowy modelu LTM wykorzystano wszystkie obserwacje (3069), a do budowy modelu STM ostatnie 150 obserwacji. W obu wypadkach zbudowano i przetestowano po 30 modeli typu $ARIMA(p,d,q)$. Do wyboru najlepszych modeli zastosowano kryterium najmniejszego średniego błędu

¹⁷⁰ Central Bank of the Republic of Turkey: https://evds2.tcmb.gov.tr/index.php?/evds/serieMarket/#collapse_2 [29.11.2021].

bezwzględnej prognozy (MAE). Prognozę wyznaczono dla kolejnych 30 wartości od 27 stycznia do 9 marca 2017 roku. Finalnie, do sporządzania prognoz parytetu USD/TRY opartych na krótkim szeregu czasowym wybrano model $ARIMA(2,1,0)_{STM}$, a dla długiego szeregu danych model $ARIMA(0,1,1)_{LTM}$. W artykule nie podano wartości parametrów modelu.

Na rysunkach 5.6 a) i b) przedstawiono wyniki prognozowania parytetu USD i TRY za pomocą obu modeli.



Rys. 5.6. Wartości rzeczywiste i prognozowane dziennych wartości parytetu USD/TRY w okresie od 27 stycznia do 9 marca 2017 roku: model STM (a), model LTM (b)

Źródło: opracowanie własne na podstawie: C.U. Yildiran, A. Fettahoğlu (2017), *Forecasting USDTRY rate by ARIMA method*, „Cogent Economics & Finance” 5(1), 1335968, doi:10.1080/23322039.2017.1335968.

W konkluzji autorzy stwierdzają, że lepsze wyniki prognozowania w badanym horyzoncie czasowym wykazuje model STM oparty na krótszym szeregu czasowym danych.

5.4. Prognozowanie cen owoców i warzyw

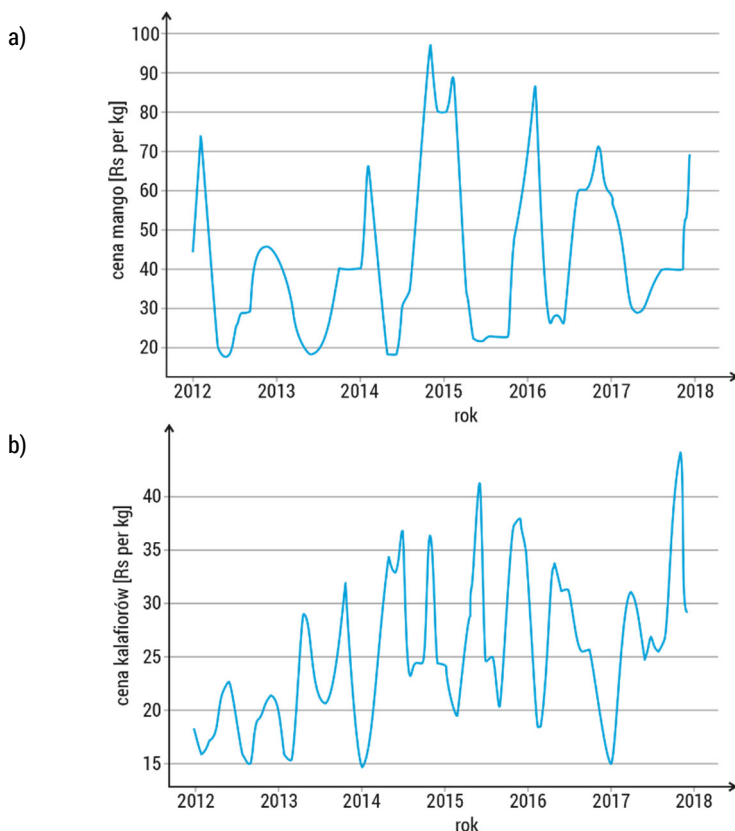
Przykład zaczerpnięto z artykułu R. Dharavath, E. Khosla, *Seasonal ARIMA to Forecast Fruits and Vegetable Agricultural Prices*¹⁷¹.

Rolnictwo ze względu na swoje znaczenie dla bezpieczeństwa żywnościowego ludzkości jest strategicznym sektorem gospodarki światowej. Ceny produktów rolnych wykazują silną sezonowość zarówno ze względu na warunki ich uprawy, przechowywania, dostaw, jak i sprzedaży. Przedmiotem badań przedstawionych w artykule są zmiany cen owoców i warzyw w aglomeracji Bengaluru (ang. Bangalore) w Indiach.

Analizą objęto miesięczne ceny wybranych owoców i warzyw z lat 2010–2018¹⁷². Nie dla wszystkich owoców i warzyw dostępne są ceny z pełnego zakresu dat. Na rysunku 5.7 przedstawiono szereg czasowy cen: owoców mango (a) i kalafiorów (b) za lata 2012–2017.

¹⁷¹ R. Dharavath, E. Khosla (2019), *Seasonal ARIMA to Forecast Fruits and Vegetable Agricultural Prices*, IEEE International Symposium on Smart Electronic Systems (iSES), s. 47–52, doi: 10.1109/iSES47678.2019.00023.

¹⁷² Data.World: <https://data.world/raghu543/fruits-and-vegetables-prices-in-india> [02.11.2021].

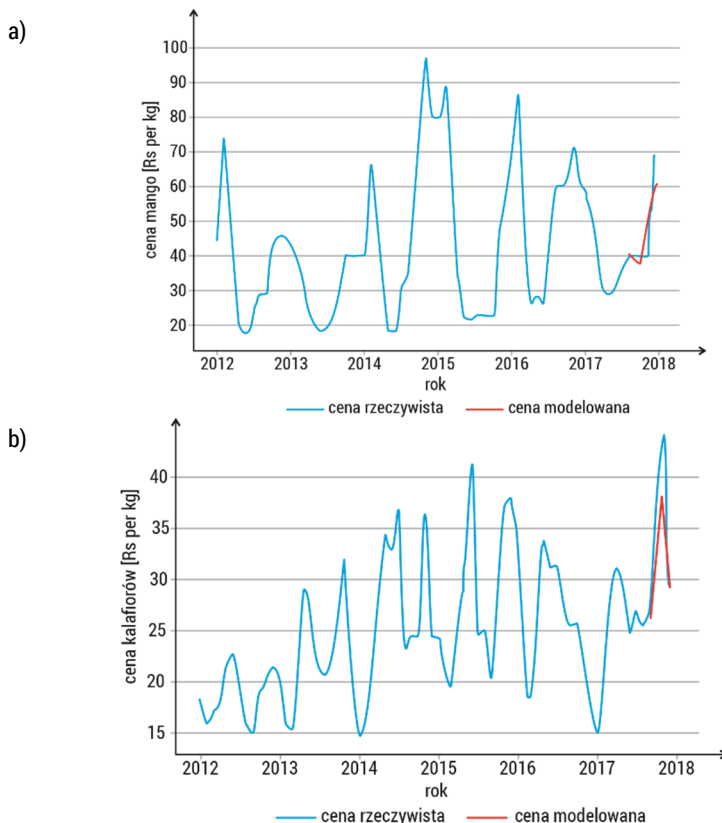


Rys. 5.7. Ceny w Bengaluru w latach 2012–2017: owoców mango (a), kalafiorów (b)

Źródło: opracowanie własne na podstawie: R. Dharavath, E. Khosla (2019), *Seasonal ARIMA to Forecast Fruits and Vegetable Agricultural Prices*, 2019 IEEE International Symposium on Smart Electronic Systems (iSES), doi: 10.1109/iSES47678.2019.00023.

Wykres cen owoców mango (rys. 5.7a) wskazuje na wyraźną sezonowość szeregu. Ceny mango najniższe są w czerwcu, a najwyższe w lutym. Do budowy modelu wykorzystano dane z okresu od stycznia 2012 roku do sierpnia 2017 roku. Jako model prognostyczny zaproponowano sezonowy model $ARIMA(2,2,1)(1,1,1)_{12}$. W artykule nie podano wartości parametrów modelu. Prognozę wygasłą miesięcznych cen owoców mango wykonano na okres wrzesień–grudzień 2017 roku. Wyniki prognozy przedstawiono

na rysunku 5.8a. Autorzy stwierdzają, że zaproponowany model wystarczająco dobrze prognozuje ceny owoców mango w rozpatrywanym horyzoncie czasowym.



Rys. 5.8. Prognoza ceny w Bengaluru w latach 2012–2017: owoców mango (a), kalafiorów (b)

Źródło: opracowanie własne na podstawie: R. Dharavath, E. Khosla, (2019) *Seasonal ARIMA to Forecast Fruits and Vegetable Agricultural Prices*, 2019 IEEE International Symposium on Smart Electronic Systems (iSES), doi:10.1109/iSES47678.2019.00023.

Również wykres cen kalafiorów (rys. 5.7b) wskazuje na wyraźną sezonowość. Ceny kalafiorów najniższe są w lutym, a najwyższe w listopadzie. Do budowy modelu wykorzystano dane z okresu od stycznia 2012 roku do sierpnia 2017 roku. Jako model

prognostyczny zaproponowano sezonowy model ARIMA(2,1,0)(1,1,1)₁₂. W artykule nie podano wartości parametrów modelu. Prognozę wygasłą miesięcznych cen kalafiorów wykonano na okres wrzesień–grudzień 2017 roku. Wyniki prognozy przedstawiono na rysunku 5.8b. Autorzy stwierdzają, że zaproponowany model wystarczająco dobrze prognozuje ceny kalafiorów w rozpatrywanym horyzoncie czasowym.

Autorzy konkludują, że problem prognozowania cen warzyw i owoców jest w warunkach Indii bardzo ważny społecznie ze względu na skalę ubóstwa i związaną z tym potrzebę ciągłej kontroli kosztów utrzymania ludności.

5.5. Prognozowanie nasłonecznienia

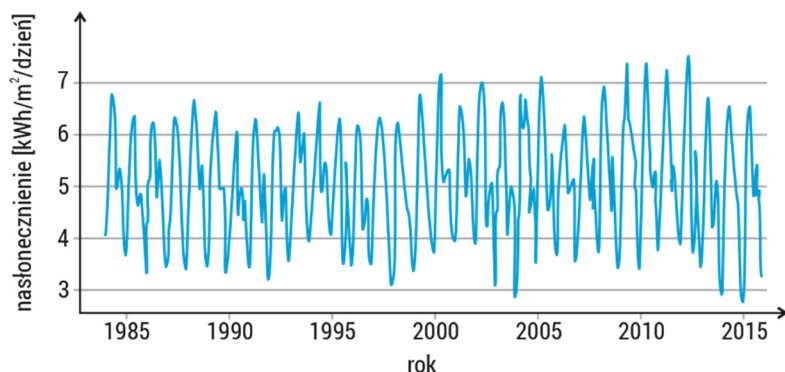
Przykład zaczerpnięto z artykułu A. Shadab, S. Ahmad, S. Said, *Spatial forecasting of solar radiation using ARIMA model*¹⁷³.

Rosnące gwałtownie zapotrzebowanie na energię z jednej strony, a wymagania klimatyczne i środowiskowe z drugiej strony powodują konieczność zwiększania produkcji energii ze źródeł odnawialnych. Energia słarna jest najważniejszym odnawialnym źródłem energii. W artykule podjęto problem wykorzystania modeli ARIMA do obszarowego prognozowania miesięcznych wartości nasłonecznienia dla celów projektowania lokalizacji elektrowni słonecznych. Analizą objęto obszar wokół New Delhi.

Dane źródłowe wykorzystane do stworzenia szeregów czasowych będących podstawą analiz zaprezentowanych w artykule, pozyskane zostały z bazy danych NASA POWER o warunkach słarnych i atmosferycznych tworzonej na podstawie obserwacji satelitarnych¹⁷⁴. Na rysunku 5.9 przedstawiono szereg czasowy średniego miesięcznego nasłonecznienia w New Delhi w latach 1984–2015.

¹⁷³ A. Shadab, S. Ahmad, S. Said (2020), *Spatial forecasting of solar radiation using ARIMA model*, „Remote Sensing Applications: Society and Environment” 20, 100427, doi: 10.1016/j.rsase.2020.100427.

¹⁷⁴ NASA POWER: <https://power.larc.nasa.gov/> [3.12.2021].



Rys. 5.9. Średnie miesięczne wartości nasłonecznienia ($\text{kWh/m}^2/\text{dzień}$) w New Delhi w latach 1984–2015

Źródło: opracowanie własne na podstawie: A. Shadab, S. Ahmad, S. Said (2020), *Spatial forecasting of solar radiation using ARIMA model*, „Remote Sensing Applications: Society and Environment” 20, 100427, doi:10.1016/j.rsase.2020.100427.

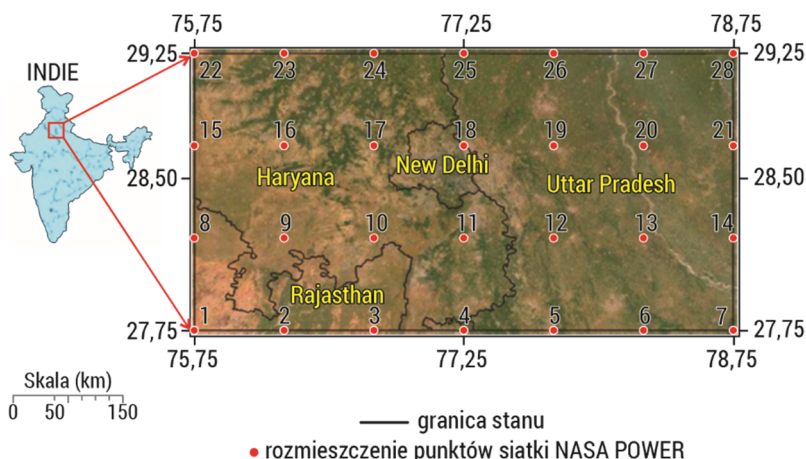
Na obszar geograficzny będący przedmiotem badań (rys. 5.10) naniesiono siatkę zdefiniowaną szerokością i długością geograficzną o kroku $0,5^\circ$, otrzymując w ten sposób 28 węzłów siatki. Dla każdego z węzłów siatki opracowano osobny model ARIMA modelujący szereg czasowy średniego miesięcznego nasłonecznienia w tym węźle. Do budowy modeli posłużyły szeregi czasowe średnich miesięcznych wartości nasłonecznienia w poszczególnych węzłach za okres od stycznia 1984 roku do grudnia 2017 roku (34 lata).

Dla wszystkich węzłów Autorzy zaproponowali modele klasy $\text{ARIMA}(p,d,q)(P,D,Q)_{12}$ o różnych wartościach parametrów. Wyboru najlepszego spośród alternatywnych modeli dla każdego z węzłów dokonano na podstawie kryterium AIC. Do obliczeń wykorzystano pakiet statystyczny R¹⁷⁵. Modele charakteryzują się dobrą jakością określoną miarami RMSE, MAE, MAPE, MASE, AIC i BIC¹⁷⁶. Modele zweryfikowano także, wyznaczając prognozy wygaśnię nasłonecznienia dla okresu od stycznia 2018 roku

¹⁷⁵ The R Foundation: <https://www.r-project.org/> [3.12.2021].

¹⁷⁶ J. Nazarko (red.) (2011), *Prognozowanie w zarządzaniu przedsiębiorstwem. Cz. II. ...*, op. cit.

do grudnia 2019 roku. Prognozy wygaśle charakteryzowały się bardzo dobrą jakością określoną miarami R^2 , RMSE, MAE i MAPE.

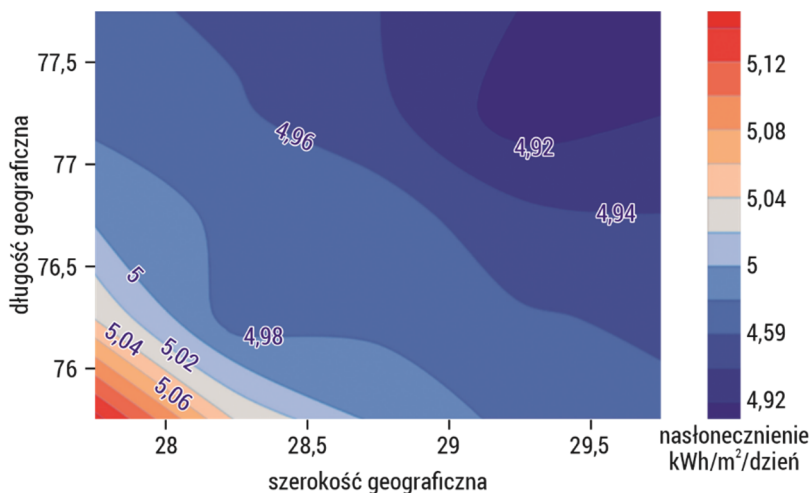


Rys. 5.10. Obszar objęty obszarowym prognozowaniem nasłonecznienia

Źródło: A. Shadab, S. Ahmad, S. Said (2020), *Spatial forecasting of solar radiation using ARIMA model*, „Remote Sensing Applications: Society and Environment” 20, 100427, doi:10.1016/j.rsase.2020.100427.

Wypracowane modele ARIMA posłużyły do sporządzenia miesięcznych prognoz nasłonecznienia w poszczególnych węzłach siatki geograficznej na dwa kolejne lata 2020–2021. Na podstawie wyznaczonych prognoz opracowano dla badanego obszaru wykresy konturowe nasłonecznienia dla poszczególnych miesięcy oraz średnioroczne. Wykresy konturowe wykonano, korzystając z pakietu statystycznego R.

Na rys. 5.11 pokazano wykres konturowy prognozowanego na rok 2021 średniego rocznego nasłonecznienia badanego obszaru.



Rys. 5.11. Prognoza średniorocznego nasłonecznienia badanego obszaru w roku 2021

Źródło: A. Shadab, S. Ahmad, S. Said (2020), *Spatial forecasting of solar radiation using ARIMA model*, „Remote Sensing Applications: Society and Environment” 20, 100427, doi:10.1016/j.rsase.2020.100427.

Na podstawie sporządzonych obszarowych prognoz nasłonecznienia wytypowano najlepszy region (Rajasthan-Haryana) do lokalizacji elektrowni słonecznych. Nasłonecznienie w tym regionie ma największą wartość średnioroczną oraz charakteryzuje się najmniejszą zmiennością miesięczną.

Literatura uzupełniająca

- [1] Dharavath R., Khosla E. (2019), *Seasonal ARIMA to Forecast Fruits and Vegetable Agricultural Prices*, IEEE International Symposium on Smart Electronic Systems (iSES), s. 47–52, doi:10.1109/iSES47678.2019.00023
- [2] Hannan E.J., Rissanen J. (1982), *Recursive estimation of mixed autoregressive-moving average order*, „Biometrika” 69, s. 81–94, doi: 10.1093/biomet/69.1.81
- [3] Nazarko J. (red.) (2011), *Prognozowanie w zarządzaniu przedsiębiorstwem. Cz. II. Prognozowanie na podstawie szeregów czasowych*, Wydawnictwo Politechniki Białostockiej, Białystok

- [4] Sahai A.K., Rath N., Sood V., Singh M.P. (2020), *ARIMA modelling & forecasting of COVID-19 in top five affected countries*, „Diabetes & Metabolic Syndrome: Clinical Research & Reviews” 14(5), s. 1419–1427, doi: 10.1016/j.dsx.2020.07.042
- [5] Shadab A., Ahmad S., Said S. (2020), *Spatial forecasting of solar radiation using ARIMA model*, „Remote Sensing Applications: Society and Environment” 20, 100427, doi: 10.1016/j.rsase.2020.100427
- [6] Yildiran C.U., Fettahoğlu A. (2017), *Forecasting USDTRY rate by ARIMA method*, „Cogent Economics & Finance” 5(1), 1335968, doi: 10.1080/23322039.2017.1335968
- [7] Zhang J. (2021), *Forecast of Shanghai Port Throughput Based on ARIMA*, IOP Conf. Ser.: Earth Environ. Sci. 831 012042, doi: 10.1088/1755-1315/831/1/012042

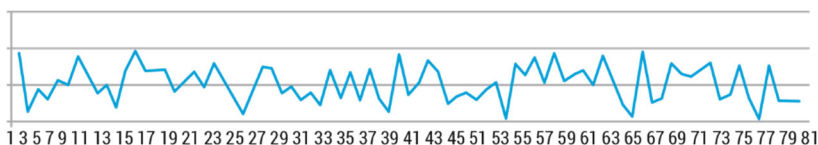
Strony internetowe

- [1] Central Bank of the Republic of Turkey: https://evds2.tcmb.gov.tr/index.php?/evds/serieMarket/#collapse_2 [29.11.2021]
- [2] Data.World: <https://data.world/raghu543/fruits-and-vegetables-prices-in-india> [2.11.2021]
- [3] JMulTi: <http://www.jmulti.de/> [29.11.2021]
- [4] Międzynarodowe Centrum Wysyłkowe w Szanghaju w pierwszej trójce na świecie, China Radio International: <http://polish.cri.cn/wideo/3868/20200713/505021.html>, 2020-07-13 [29.11.2021]
- [5] NASA POWER: <https://power.larc.nasa.gov/>, [3.12.2021]
- [6] The R Foundation: <https://www.r-project.org/>, [3.12.2021]
- [7] Worldmeter: <https://www.worldometers.info/coronavirus/> [29.11.2021]

6. PROBLEMY

DO SAMODZIELNEGO ROZWIĄZANIA

- Na podstawie oceny wizualnej szeregu przedstawionego na rysunku 6.1 można powiedzieć, że szereg jest:
 - niestacjonarny względem średniej i wariancji,
 - stacjonarny względem średniej i wariancji,
 - stacjonarny względem średniej i niestacjonarny względem wariancji,
 - niestacjonarny względem średniej i stacjonarny względem wariancji.

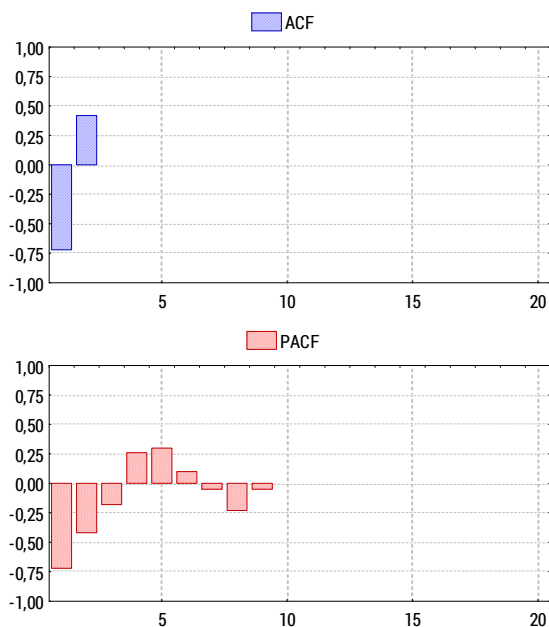


Rys. 6.1. Przykład szeregu

Źródło: opracowanie własne.

- Rozwinięciem zapisu operatorowego modelu $ARIMA(1 - B)(1 - B^4)Y_t = (1 - \theta_1 B)(1 - \Theta_1 B^4)e_t$ jest:
 - $Y_t = Y_{t-4} - Y_{t-5} - \theta_1 e_{t-4} - \Theta_1 e_{t-4} + \Theta_1 \theta_1 e_{t-5} + e_t$
 - $Y_t = Y_{t-1} + Y_{t-4} - \theta_1 e_{t-1} - \Theta_1 e_{t-4} + \Theta_1 \theta_1 e_{t-5} + e_t$
 - $Y_t = Y_{t-1} + Y_{t-4} - Y_{t-5} - \theta_1 e_{t-1} - \Theta_1 e_{t-4} + \Theta_1 \theta_1 e_{t-5} + e_t$
 - $Y_t = Y_{t-1} + Y_{t-4} - \theta_1 e_{t-1} - \Theta_1 e_{t-4} + e_t$
- Model $ARIMA(1,1,0)(1,0,0)_4$ w postaci operatorowej to:
 - $(1 - \phi_1 B)(1 - \phi_1 B^4)(1 - B)Y_t = e_t$
 - $(1 - \phi_1 B)(1 - \phi_1 B^4)Y_t = (1 - B)e_t$
 - $(1 - \phi_1 B)Y_t = (1 - \phi_1 B)^4(1 - B)e_t$
 - $Y_t = (1 - \phi_1 B)(1 - \phi_1 B^4)(1 - B)e_t$
- Przedstawione na rysunku 6.2 przebiegi funkcji ACF i PACF wskazują na model:
 - AR(2)
 - ARMA(1,1)

- c) ARMA(2,2)
- d) MA(2)



Rys. 6.2. Przykład ACF i PACF

Źródło: opracowanie własne.

5. Kryterium informacyjne AIC to kryterium wyboru pomiędzy modelami statystycznymi o różnej liczbie parametrów. Czy korzystając z niego wybiera się model, dla którego wartość jest:
 - a) najmniejsza,
 - b) największa.
6. Które zdanie(a) jest/są prawdziwe:
 - a) Procesy AR są stacjonarne zawsze, ale wymagają nałożenia ograniczeń na parametry ze względu na konieczność zapewnienia odwracalności.

- b) Procesy MA są odwracalne zawsze, ale wymagają ograniczenia ze względu na konieczność zapewnienia stacjonarności.
7. W rozdziale czwartym przytoczony został szereg G. Dokonaj samodzielnej estymacji jego parametrów.
8. W tabeli 6.1 zamieszczono dane dotyczące liczby pasażerów Lotniska Chopina w Warszawie w latach 2008–2019. Zbuduj model ARIMA tego szeregu czasowego.

Tabela 6.1. Liczba pasażerów lotniska Chopina w Warszawie w latach 2008–2019

rok	I kwartał	II kwartał	III kwartał	IV kwartał
2019	3 696 371	4 891 498	5 744 461	4 512 261
2018	3 418 730	4 577 301	5 672 623	4 068 577
2017	2 978 217	3 989 410	5 065 287	3 697 416
2016	2 314 051	3 264 233	4 069 807	3 147 265
2015	2 165 311	2 999 272	3 572 135	2 449 970
2014	1 990 629	2 820 154	3 391 928	2 371 828
2013	2 106 811	2 923 440	3 454 526	2 185 102
2012	1 888 414	2 868 215	2 884 649	1 925 785
2011	1 782 918	2 485 521	2 971 790	2 082 256
2010	1 712 986	2 191 768	2 759 235	2 002 563
2009	1 588 265	2 203 481	2 618 804	1 868 197
2008	1 984 486	2 550 700	2 960 291	1 941 481

Źródło: Urząd Lotnictwa Cywilnego, Statystyki wg portów lotniczych, <https://www.ulc.gov.pl/pl/statystyki-analizy/statystyki-i-analizy-rynku-transportu-lotniczego/3724-statystyki-wg-portow-lotniczych> [10.08.2021].

9. Z ogólnie dostępnych baz danych wybierz miesięczny, kwartalny lub roczny wskaźnik makroekonomiczny opisujący poziom rozwoju społeczno-gospodarczego dowolnego kraju lub regionu. Zbuduj dla szeregu czasowego wskaźnika model ARIMA.

Literatura

- [1] Allingham D., Rayner J.C.W. (2012), *Testing Equality of Variances for Multiple Univariate Normal Populations*, „Journal of Statistical Theory and Practice” 6, s. 524–535, doi: 10.1080/15598608.2012.695703
- [2] Awan T.M., Aslam F. (2020), *Prediction of daily COVID-19 cases in European countries using automatic ARIMA model*, „Journal of Public Health Research” 9(3), 1765, s. 227–233, doi: 10.4081/jphr.2020.1765
- [3] Belsley D.A. (red.), Kontoghiorghes E.J. (red.) (2009), *Handbook of Computational Econometrics*, John Wiley & Sons, Ltd, <https://thenigerianprofessionalaccountant.files.wordpress.com/2013/04/handbook-of-computational-econometrics-belsley.pdf> [15.12.2020]
- [4] Box G.E.P., Jenkins G.M., Reinsel G.C., Ljung G.M. (2015), *Time Series Analysis: Forecasting and Control*. Fifth Edition, John Wiley & Sons
- [5] Box G.E.P., Jenkins G.M. (1976), *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day 1970 (Revised edition published 1976)
- [6] Brockwell P.J., Davis R.A. (2002), *Introduction to Time Series and Forecasting*, Second Edition, Springer
- [7] Burnham K.P., Anderson D.R. (2004), *Multimodel Inference: Understanding AIC and BIC in Model Selection*. „Sociological Methods Research” 33(2), s. 261–304, doi:10.1177/0049124104268644
- [8] Caines P.E., Rissanen J. (1974), *Maximum likelihood estimation of parameters in multivariate Gaussian stochastic processes*, „IEEE Transactions on Information Theory” 20, s. 102–104, doi: 10.1109/TIT.1974.1055155
- [9] Chodakowska E. (2002), *Zastosowanie pakietu STATISTICA w prognozowaniu na podstawie szeregów czasowych*, „Zeszyty Naukowe Politechniki Białostockiej. Ekonomia i Zarządzanie” 7, s. 77–86
- [10] Chodakowska E. (2004), *Analiza wartości sprzedaży w przedsiębiorstwie handlowym z zastosowaniem modeli ARIMA*, „Zeszyty Naukowe Politechniki Białostockiej. Ekonomia i Zarządzanie” 9, s. 37–52

- [11] Chodakowska E., Nazarko J., Nazarko Ł. (2021), *ARIMA Models in Electrical Load Forecasting and Their Robustness to Noise*, „Energies” 14(23), 7952, s. 1–22, doi:10.3390/en14237952
- [12] Cieślak M. (red.) (2011), *Prognozowanie gospodarcze. Metody i zastosowania*. Wydanie czwarte zmienione. Wydawnictwo Naukowe PWN, Warszawa
- [13] Contreras J., Espínola R., Nogales F.J., Conejo A.J. (2003), *ARIMA Models to Predict Next-Day Electricity Prices*, „IEEE Transactions on Power Systems” 18(3), s. 1014–1020, doi: 10.1109/TPWRS.2002.804943
- [14] Cryer J.D., Chan K.-S. (2008), *Time Series Analysis. With Applications in R*, Second Edition, Springer
- [15] DeLurgio S.A. (1998), *Forecasting Principles and Applications*, Irwin/McGraw-Hill, Boston
- [16] Denis D.J. (2015), *Applied Univariate, Bivariate, and Multivariate Statistics*, John Wiley & Sons, Inc.
- [17] Dharavath R., Khosla E. (2019), *Seasonal ARIMA to Forecast Fruits and Vegetable Agricultural Prices*, IEEE International Symposium on Smart Electronic Systems (iSES), s. 47–52, doi: 10.1109/iSES47678.2019.00023
- [18] Dickey D.A., Fuller W.A. (1979), *Distribution of the Estimators for Autoregressive Time Series with a Unit Root*, „Journal of the American Statistical Association” 74(366), s. 427–431, doi: 10.2307/2286348
- [19] Dürre A., Fried R., Liboschik T. (2015), *Robust estimation of (partial) autocorrelation*, „WIREs Computational Statistics” 7(3), s. 205–222, doi: 10.1002/wics.1351
- [20] Fan J., Yao Q. (2005), *Nonlinear Time Series Nonparametric and Parametric Methods*, Springer
- [21] Fattah J., Ezzine L., Aman Z., El Moussami H., Lachhab A. (2018), *Forecasting of demand using ARIMA model*, „International Journal of Engineering Business Management” 10, s. 1–9, doi: 10.1177/1847979 018808673
- [22] Frenkel A., Kildiszew G. (1976), *Analiza szeregów czasowych i prognozowanie*, Państwowe Wydawnictwo Ekonomiczne, Warszawa
- [23] Gardner G., Harvey A.C., Phillips G.D.A. (1980), *An algorithm for exact maximum likelihood estimation of autoregressive-moving average*

- models by means of Kalman filtering*, „Journal of the Royal Statistical Society” C, 29, s. 311–322, doi: 10.2307/2346910
- [24] Gastwirth J.L., Gel Y.R., Miao W. (2009), *The Impact of Levene's Test of Equality of Variances on Statistical Theory and Practice*, „Statistical Science” 24(3), 343–360, doi: 10.1214/09-STS301
 - [25] Gruszczyński M., Podgórska M. (red.) (1996), *Ekonometria*, Oficyna Wydawnicza Szkoła Główna Handlowa, Warszawa
 - [26] Hannan E.J., Rissanen J. (1982), *Recursive estimation of mixed autoregressive-moving average order*, „Biometrika” 69, s. 81–94, doi: 10.1093/biomet/69.1.81
 - [27] Heiberger R.M., Teles P. (2002), *Displays for Direct Comparison of ARIMA Models*, „The American Statistician” 56(2), s. 134, doi: 10.1198/000313002317572808
 - [28] Hyndman R., Khandakar Y. (2008), *Automatic time series forecasting: the forecast package for R*, „Journal of Statistical Software” 26(3), s. 1–22, doi: 10.18637/jss.v027.i03
 - [29] Hyndman R.J., Athanasopoulos G. (2018), *Forecasting: principles and practice*, 2nd ed., OTexts, Melbourne, Australia. Section 3.4 Evaluating forecast accuracy, <https://otexts.org/fpp2/accuracy.html>
 - [30] Iwaszczuk N., Drygaś P., Pusz P., Pusz R. (2013), *Prognozowanie gospodarcze*, Uniwersytet Rzeszowski, Rzeszów
 - [31] Johnson R.A., Wichern D.A. (2007), *Applied Multivariate Statistical Analysis*. Sixth Edition, Pearson Education, Inc.
 - [32] Kwiatkowski D., Phillips P.C.B., Schmidt P., Shin Y. (1992), *Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?*, „Journal of Econometrics”: 54(1–3), s. 159–178, doi:10.1016/0304-4076(92)90104-Y
 - [33] Lcc D., Schmidt P. (1996), *On the power of the KPSS test of stationarity against fractionally-integrated alternatives*, „Journal of Econometrics” 73, s. 285–302, doi:10.1016/0304-4076(95)01741-0
 - [34] Lindquist A. (1974), *A new algorithm for optimal filtering of discrete time stationary processes*, „SIAM Journal on Control and Optimization” 12, s. 736–746, doi:10.1137/0312057
 - [35] Łobejko S. (red.), Maślowska K., Wojdan R. (2015), *Analiza i prognozowanie szeregów czasowych z programem SAS*, Szkoła Główna Handlowa w Warszawie – Oficyna Wydawnicza, Warszawa,

- https://ssl-administracja.sgh.waw.pl/pl/OW/publikacje/Documents/Analiza_lobejko_maslowska_wojdan.pdf [27.07.2020]
- [36] Makridakis S., Wheelwright S.C., Hyndman R.J. (1998), *Forecasting. Methods and Applications*, John Wiley & Sons, Inc. s. 313
 - [37] Marquardt D.W. (1963), *An Algorithm for Least-Squares Estimation of Nonlinear Parameters*, „Journal of the Society for Industrial and Applied Mathematics” 11(2), s. 431–441, doi: 10.1137/0111030
 - [38] Mastrangelo C.M., Simpson J.R., Montgomery D.C. (2013), *Time Series Analysis*, in: Gass S.I., Fu M.C. (red.) *Encyclopedia of Operations Research and Management Science*, Springer, Boston, MA, doi: 10.1007/978-1-4419-1153-7_1045
 - [39] MathWorks. Help Center, <https://www.mathworks.com/help/econ/how-to-specify-arima-models.html> [29.11.2021]
 - [40] Maynard A., Smallwood A., Wohar M.E. (2013), *Long Memory Regressors and Predictive Testing: A Two-stage Rebalancing Approach*, „Econometric Reviews” 32(3), s. 318–360, doi: 10.1080/07474938.2012.690663
 - [41] Mcleod A.I., Sales P.R.H. (1983), *An Algorithm for Approximate Likelihood Calculation of Arma and Seasonal Arma Models*, „Journal of the Royal Statistical Society: Series C (Applied Statistics)”, 32(2), s. 211–223, doi: 10.2307/2347301
 - [42] Mélard G. (1984), *A fast algorithm for the exact likelihood of autoregressive-moving average models*, „Journal of the Royal Statistical Society: Series C (Applied Statistics)”, 33(1), s. 104–114, <https://doi.org/10.2307/2347672>
 - [43] Mélard G., Roch R., Saidi A. (2006), *Exact maximum likelihood estimation of structured or unit root multivariate time series models*, „Computational Statistics & Data Analysis” 50(11), s. 2958–2986, doi: 10.1016/j.csda.2005.06.009
 - [44] Nau R., *Statistical forecasting: notes on regression and time series analysis*, Fuqua School of Business, Duke University, <https://people.duke.edu/~rnau/411home.htm> [7.08.2020]
 - [45] Nazarko J. (red.) (2005), *Prognozowanie w zarządzaniu przedsiębiorstwem. Cz. III. Prognozowanie na podstawie modeli adaptacyjnych*, Wydawnictwo Politechniki Białostockiej, Białystok

- [46] Nazarko J. (red.) (2011), *Prognozowanie w zarządzaniu przedsiębiorstwem. Cz. II. Prognozowanie na podstawie szeregów czasowych*, Wydawnictwo Politechniki Białostockiej, Białystok
- [47] Nazarko J., Chraobowska J., Rybaczuk M. (2004), *Zastosowanie wielosezonowego modelu ARIMA w prognozowaniu obciążeń mocą elektryczną*, „Taksonomia” 11, s. 173–182
- [48] Nazarko J., Rybaczuk M., Jurczuk A. (2005), *Wpływ poziomu zakłóceń losowych na możliwość identyfikacji modeli ARIMA*, „Taksonomia” 12, s. 238–247
- [49] O'Neill M.E., Mathews K. (2000), *A weighted least squares approach to Levene's test of homogeneity of variance*, „Australian and New Zealand Journal of Statistics” 42(1), s. 81–100, doi: 10.1111/1467-842X.00109
- [50] Pfaff B. (2008), *Analysis of Integrated and Cointegrated Time Series with R*, Springer-Verlag New York
- [51] Raman R.K., Sathianandan T.V., Sharma A.P., Mohanty B.P. (2017), *Modelling and Forecasting Marine Fish Production in Odisha Using Seasonal ARIMA Model*, „National Academy Science Letters” 40, s. 393–397, doi: 10.1007/s40009-017-0581-2
- [52] Sahai A.K., Rath N., Sood V., Singh M.P. (2020), *ARIMA modelling & forecasting of COVID-19 in top five affected countries*, „Diabetes & Metabolic Syndrome: Clinical Research & Reviews” 14(5), s. 1419–1427, doi: 10.1016/j.dsx.2020.07.042
- [53] Salles R., Belloze K., Porto F., Gonzalez P.H., Ogasawara E. (2019), *Nonstationary time series transformation methods: An experimental review*, „Knowledge-Based Systems” 164, s. 274–291, doi: 10.1016/j.knosys.2018.10.041
- [54] Shadab A., Ahmad S., Said S. (2020), *Spatial forecasting of solar radiation using ARIMA model*, „Remote Sensing Applications: Society and Environment” 20, 100427, doi:10.1016/j.rsase.2020.100427
- [55] Shumway R.H., Stoffer D.S. (2017), *Time Series and Its Applications with R Examples*, Springer International Publishing AG
- [56] Sokołowski A. (2003), *Prognozowanie finansowych szeregów czasowych*, Statsoft Polska, https://media.statsoft.pl/_old_dnn/downloads/sokolowski.pdf [27.07.2020]

- [57] Strawiński P., *Materiały do zajęć: statystyczna analiza danych finansowych*, http://coin.wne.uw.edu.pl/pstrawinski/dane_fin/adf03.pdf [5.09.2021]
- [58] Syczewska E.M. (2010), *Empirical power of the Kwiatkowski-Phillips-Schmidt-Shin test*, Department of Applied Econometrics Working Papers 3, <http://yadda.icm.edu.pl/yadda/element/bwmeta1.element.ekon-element-000171231155>
- [59] Tarczyński W., Witkowska D., Kompa K. (2013), *Współczynnik beta. Teoria i praktyka*, Pielaszek Research, Warszawa
- [60] Villegas M.A., Pedregal D.J. (2018), *SSpace: A Toolbox for State Space Modeling*, „Journal of Statistical Software” 87(1), s. 1–26, doi: 10.18637/jss.v087.i05
- [61] Villegas M.A., Pedregal D.J. (2019), *Automatic selection of unobserved components models for supply chain forecasting*, „International Journal of Forecasting” 35(1), s. 157–169, doi: 10.1016/j.ijforecast.2017.11.001
- [62] Witkowska D. (2012), *Podstawy ekonometrii i teorii prognozowania*, Wolters Kluwer Polska SA
- [63] Witkowska D., Żebrowska-Suchodolska D. (2008), *Badanie efektywności GPW na przykładzie wybranych indeksów: test autokorelacji*, „Oeconomia” 7(4), s. 155–162
- [64] Wu L., Hayashi H. (2014), *The Impact of Disasters on Japan's Inbound Tourism Demand*, „Journal of Disaster Research” 9, s. 699–708, doi: 10.20965/jdr.2014.p0699
- [65] Yang W., Zurbenko I. (2010), *Nonstationarity*, „WIREs Computational Statistics” 2(1), s. 107–115, doi: 10.1002/wics.64
- [66] Yang Z., Ce L., Lian L. (2017), *Electricity price forecasting by a hybrid model, combining wavelet transform, ARMA and kernel-based extreme learning machine methods*, „Applied Energy” 190, s. 291–305, doi: 10.1016/j.apenergy.2016.12.130
- [67] Yildiran C.U., Fettahoğlu A. (2017), *Forecasting USDTRY rate by ARIMA method*, „Cogent Economics & Finance” 5(1), 1335968, doi: 10.1080/23322039.2017.1335968
- [68] Zalewski W., Nazarko J., Jurczuk A. (2005), *ARIMA models in load modelling with clustering approach*, „PowerTech IEEE'2005”, St. Petersburg, June 27–30

- [69] Zhang J. (2021), *Forecast of Shanghai Port Throughput Based on ARIMA*, IOP Conf. Ser.: Earth Environ. Sci. 831 012042, doi: 10.1088/1755-1315/831/1/012042

Portale informacyjne, strony internetowe, dokumenty elektroniczne

- [1] Data.World: <https://data.world/raghu543/fruits-and-vegetables-prices-in-india> [2.11.2021]
- [2] Federal Reserve Economic, <https://fred.stlouisfed.org/series/ECOMPCTNSA> [6.08.2021]
- [3] Międzynarodowe Centrum Wysyłkowe w Szanghaju w pierwszej trójce na świecie, China Radio International: <http://polish.cri.cn/wideo/3868/20200713/505021.html>, 2020-07-13 [29.11.2021]
- [4] Central Bank of the Republic of Turkey: https://evds2.tcmb.gov.tr/index.php?/evds/serieMarket/#collapse_2 [29.11.2021]
- [5] Worldmeter: <https://www.worldometers.info/coronavirus/> [29.11.2021]
- [6] *IBM SPSS Statistics 24 Algorithms* (1989,2016), IBM Corporation, <https://www.ibm.com/support/pages/node/616931#en> [11.08.2020]
- [7] *SAS/ETS 13.2 User's Guide. The ARIMA Procedure (2014)*, SAS Institute Inc., Cary, NC, USA, <https://support.sas.com/documentation/onlinedoc/ets/132/arima.pdf> [29.11.2021]
- [8] JMulTi: <http://www.jmulti.de/> [29.11.2021]
- [9] The R Foundation: <https://www.r-project.org/> [3.12.2021]
- [10] NASA POWER: <https://power.larc.nasa.gov/> [3.12.2021]
- [11] CRAN: <https://cran.r-project.org/> [12.12.2021]
- [12] SAS: https://www.sas.com/pl_pl/home.html [12.12.2021]
- [13] MATHWorks: <https://uk.mathworks.com/> [12.12.2021]
- [14] SPSS Polska: <https://spss.pl/> [12.12.2021]
- [15] WOLFRAM: <https://www.wolfram.com/mathematica/> [12.12.2021]
- [16] XLSTAT: <https://www.xlstat.com/en/> [12.12.2021]

Spis tabel

Tabela 1.1. Średnia temperatura w Białymstoku w dniach 1.04–31.05.2019 r. w °C.....	20
Tabela 2.1. Szereg zużycia energii pierwotnej w Stanach Zjednoczonych w latach 2015–2019 w bilingach BTU	45
Tabela 3.1. Obszary dopuszczalne dla parametrów procesu średniej ruchomej i procesu autoregresji rzędu pierwszego i drugiego oraz procesu mieszanego	67
Tabela 3.2. Zestawienie własności funkcji autokorelacji i autokorelacji częściowej procesów autoregresji, średniej ruchomej i mieszanych.....	69
Tabela 3.3. Zachowanie się funkcji autokorelacji dla wybranych procesów ARIMA(p,d,q)	70
Tabela 3.4. Przybliżone oszacowania parametrów procesów średniej ruchomej i procesów autoregresji rzędu pierwszego i drugiego	86
Tabela 4.1. Zużycie energii pochodzącej ze źródeł odnawialnych na świecie w terawatogodzinach (TWh).....	102
Tabela 4.2. Kwartalna sprzedaż e-commerce jako procent całkowitej sprzedaży, nieskorygowana sezonowo	126
Tabela 4.3. Wyniki estymacji wybranych modeli.....	133
Tabela 4.4. Prognoza wartości zmiennej i jej realizacja	136
Tabela 4.5. Liczba pasażerów międzynarodowych linii lotniczych w tysiącach (szereg G).....	139
Tabela 4.6. Liczba pasażerów komercyjnych lotniska John F. Kennedy International Airport.....	141
Tabela 5.1. Specyfikacja modeli ARIMA dla badanych krajów	156
Tabela 6.1. Liczba pasażerów lotniska Chopina w Warszawie w latach 2008–2019.....	170

Spis rysunków

Rys. 1.1.	Przykład jednowymiarowego szeregu czasowego.....	7
Rys. 1.2.	Ilustracja zjawiska autokorelacji w szeregu czasowym	10
Rys. 1.3.	Ilustracja zjawiska autokorelacji cząstkowej w szeregu czasowym	12
Rys. 1.4.	Pakiet Statistica – menu.....	19
Rys. 1.5.	Pakiet Statistica – okno Szeregi czasowe i prognozowanie	19
Rys. 1.6.	Szereg średniej temperatury zanotowanej w Białymstoku w dniach 1.04–31.05.2019 r.	21
Rys. 1.7.	Zależność autokorelacyjna w szeregu średniej temperatury zanotowanej w Białymstoku w dniach 1.04–31.05.2019 r. dla $k = 1$	22
Rys. 1.8.	Wartości autokorelacji szeregu średniej temperatury zanotowanej w Białymstoku w dniach 1.04–31.05.2019 r.	23
Rys. 1.9.	Wartości autokorelacji cząstkowej szeregu średniej temperatury zanotowanej w Białymstoku w dniach 1.04–31.05.2019 r.	24
Rys. 2.1.	Przykład szeregów stacjonarnego ze stałą średnią i wariancją	29
Rys. 2.2.	Przykład szeregu niestacjonarnego ze względu na średnią (z trendem)	30
Rys. 2.3.	Przykład szeregu niestacjonarnego ze względu na średnią (z różnym poziomem)	31
Rys. 2.4.	Przykład szeregu niestacjonarnego ze względu na wariancję	32
Rys. 2.5.	Metody eliminacji niestacjonarności w szeregach czasowych	34
Rys. 2.6.	Pakiet Statistica – okno przekształcania zmiennych	42
Rys. 2.7.	Szereg średniej temperatury zanotowanej w Białymstoku w dniach 1.04–31.05.2019 r. po operacji różnicowania $d = 1$	43

Rys. 2.8.	Wartości autokorelacji zróżnicowanego $d = 1$ szeregu średniej temperatury zanotowanej w Białymstoku w dniach 1.04–31.05.2019 r.	44
Rys. 2.9.	Wartości autokorelacji cząstkowej zróżnicowanego szeregu średniej temperatury zanotowanej w Białymstoku w dniach 1.04–31.05.2019 r.	45
Rys. 2.10.	Szereg zużycia energii pierwotnej w Stanach Zjednoczonych w latach 2015–2019	46
Rys. 2.11.	Wykres rozrzutu obrazujący zależność pomiędzy wartościami zużycia energii w Stanach Zjednoczonych w latach 2015–2019 w danym miesiącu (Y_t) i w miesiącu poprzedzającym (Y_{t-1})	47
Rys. 2.12.	Wykres rozrzutu obrazujący zależność pomiędzy wartościami zużycia energii w Stanach Zjednoczonych w latach 2015–2019 w danym miesiącu (Y_t) i w miesiącu poprzedzającym o rok (Y_{t-12}).....	48
Rys. 2.13.	Wartości autokorelacji szeregu zużycia energii pierwotnej w Stanach Zjednoczonych w latach 2015–2019.....	49
Rys. 2.14.	Wartości autokorelacji cząstkowej zużycia energii pierwotnej w Stanach Zjednoczonych w latach 2015–2019.....	50
Rys. 2.15.	Szereg zużycia energii pierwotnej w Stanach Zjednoczonych w latach 2015–2019 po operacji różnicowania sezonowego $s = 12$	51
Rys. 2.16.	Wartości autokorelacji zróżnicowanego $z s = 12$ szeregu zużycia energii pierwotnej w Stanach Zjednoczonych w latach 2015–2019.....	52
Rys. 2.17.	Wartości autokorelacji cząstkowej zróżnicowanego $s = 12$ szeregu zużycia energii pierwotnej w Stanach Zjednoczonych w latach 2015–2019.....	53
Rys. 3.1.	Proces budowy modelu ARIMA.....	68
Rys. 3.2.	Przebieg funkcji autokorelacji i autokorelacji cząstkowej procesu AR(1) $\phi_1 > 0$	71

Rys. 3.3.	Przebieg funkcji autokorelacji i autokorelacji częstkowej procesu AR(1) $\phi_1 < 0$	72
Rys. 3.4.	Przebieg funkcji autokorelacji i autokorelacji częstkowej procesu MA(1) $\theta_1 < 0$	73
Rys. 3.5.	Przebieg funkcji autokorelacji i autokorelacji częstkowej procesu MA(1) $\theta_1 < 0$	74
Rys. 3.6.	Przebieg funkcji autokorelacji i autokorelacji częstkowej procesu MA(2) $\theta_1 < 0, \theta_2 < 0$	75
Rys. 3.7.	Przebieg funkcji autokorelacji i autokorelacji częstkowej procesu MA(2) $\theta_1 < 0, \theta_2 > 0$	76
Rys. 3.8.	Przebieg funkcji autokorelacji i autokorelacji częstkowej procesu MA(2) $\theta_1 < 0, \theta_2 > 0$	77
Rys. 3.9.	Przebieg funkcji autokorelacji i autokorelacji częstkowej procesu MA(2) $\theta_1 > 0, \theta_2 < 0$	78
Rys. 3.10.	Przebieg funkcji autokorelacji i autokorelacji częstkowej procesu ARIMA(1,0,0) ₄ $\phi_1 = 0,75$	79
Rys. 3.11.	Przebieg funkcji autokorelacji i autokorelacji częstkowej procesu ARIMA(1,0,0) ₄ $\phi_1 = -0,75$	80
Rys. 3.12.	Przebieg funkcji autokorelacji i autokorelacji częstkowej procesu ARIMA(0,0,1) ₄ $\theta_1 = 0,5$	81
Rys. 3.13.	Przebieg funkcji autokorelacji i autokorelacji częstkowej procesu ARIMA(0,0,1) ₄ $\theta_1 = -0,5$	82
Rys. 3.14.	Granice błędów standardowych dla autokorelacji reszt procesu AR(1) $\phi_1 = 0,3$ oraz MA(1) $\theta_1 = 0,3$	88
Rys. 3.15.	Wykres wartość AIC dla hipotetycznych modeli, w których $0 \leq p, q \leq 2$	90
Rys. 4.1.	Prognozowanie oparte na modelach ARIMA	99
Rys. 4.2.	Pakiet Statistica – arkusz danych	103
Rys. 4.3.	Pakiet Statistica – menu Szeregi czasowe i prognozowanie	103
Rys. 4.4.	Pakiet Statistica – okno Analiza szeregów czasowych	104
Rys. 4.5.	Pakiet Statistica – okno Przekształcenia zmiennych (Przegląd)	105
Rys. 4.6.	Szereg zużycia energii pochodzącej ze źródeł odnawialnych na świecie w programie Statistica	105

Rys. 4.7. Pakiet Statistica – okno Przekształcenia zmiennych (Autokorelacje)	106
Rys. 4.8. ACF (a) i PACF (b) szeregu zużycia energii pochodzącej ze źródeł odnawialnych na świecie.....	107
Rys. 4.9. Pakiet Statistica – arkusz danych wraz ze zmienną grupującą.....	108
Rys. 4.10. Pakiet Statistica – Statystyki podstawowe	109
Rys. 4.11. Pakiet Statistica – wybór zmiennych zależnych i grupujących	109
Rys. 4.12. Pakiet Statistica – okno Statystyki w grupach.....	110
Rys. 4.13. Wyniki test Levene’a (a) oraz testu Browna-Forsythe’a (b)	110
Rys. 4.14. Pakiet Statistica – okno Przekształcenia zmiennych (logarytmowanie)	111
Rys. 4.15. Wyniki test Levene’a (a) oraz testu Browna-Forsythe’a (b) dla szeregu zlogarytmowanego.....	112
Rys. 4.16. Zlogarytmowany szereg zużycia energii pochodzącej ze źródeł odnawialnych (a), ACF (b) i PACF (c).....	113
Rys. 4.17. Pakiet Statistica – okno Przekształcenia zmiennych (różnicowanie)	114
Rys. 4.18. Zlogarytmowany i jednokrotnie zróżnicowany szereg zużycia energii pochodzącej ze źródeł odnawialnych (a), ACF (b) i PACF (c).....	115
Rys. 4.19. Zlogarytmowany i dwukrotnie zróżnicowany szereg zużycia energii pochodzącej ze źródeł odnawialnych (a), ACF (b) i PACF (c).....	116
Rys. 4.20. Zróżnicowany szereg zużycia energii pochodzącej ze źródeł odnawialnych (a), ACF (b) i PACF (c).....	118
Rys. 4.21. Dwukrotnie zróżnicowany szereg zużycia energii pochodzącej ze źródeł odnawialnych (a), ACF (b) i PACF (c)	119
Rys. 4.22. Pakiet Statistica – okno Model ARIMA dla pojedynczego szeregu: określenie parametrów.....	120
Rys. 4.23. Pakiet Statistica – okno Wyniki analizy ARIMA pojedynczego szeregu	121

Rys. 4.24. Pakiet Statistica – okno Wyniki analizy ARIMA pojedynczego szeregu	122
Rys. 4.25. Weryfikacja reszt: ACF (a)	122
Rys. 4.25. Weryfikacja reszt: PACF (b), histogram (c), wykres normalności (d)	123
Rys. 4.26. Pakiet Statistica – okno dialogowe Wyniki analizy ARIMA dla pojedynczego szeregu – prognozowanie	124
Rys. 4.27. Prognoza zużycia energii pochodzącej ze źródeł odnawialnych na świecie na lata 2021–2030: wartość (a) oraz wykres (b)	125
Rys. 4.28. Szereg kwartalnej sprzedaży detalicznej w handlu elektronicznym wyrażonej jako procent całkowitej sprzedaży, nieskorygowany sezonowo 2005–2019	127
Rys. 4.29. ACF (a) i PACF (b) szeregu kwartalnej sprzedaży detalicznej w handlu elektronicznym wyrażonej jako procent całkowitej sprzedaży, nieskorygowanego sezonowo 2005–2018	128
Rys. 4.30. Szereg kwartalnej sprzedaży detalicznej w handlu elektronicznym wyrażonej jako procent całkowitej sprzedaży, nieskorygowanej sezonowo po operacji logarytmowania (a), ACF (b) i PACF (c) zlogarytmowanego szeregu 2005–2018	130
Rys. 4.31. Zlogarytmowany i zróżnicowany z opóźnieniem 1 szereg kwartalnej sprzedaży detalicznej w handlu elektronicznym wyrażonej jako procent całkowitej sprzedaży, nieskorygowanej sezonowo 2005–2018 (a) oraz jego ACF (b) i PACF (c)	131
Rys. 4.32. Zlogarytmowany i zróżnicowany z opóźnieniem 1 i 4 szereg kwartalnej sprzedaży detalicznej w handlu elektronicznym wyrażonej jako procent całkowitej sprzedaży, nieskorygowanej sezonowo 2005–2018 (a) oraz jego ACF (b) i PACF (c)	132
Rys. 4.33. Statistica – okno Model ARIMA dla pojedynczego szeregu	134
Rys. 4.34. Statistica – ocena reszt modelu	135

Rys. 4.34. Statistica – ocena reszt modelu	136
Rys. 4.35. Szereg i jego prognoza.....	137
Rys. 4.36. Liczba pasażerów międzynarodowych linii lotniczych (szereg G)	139
Rys. 4.37. Liczba pasażerów komercyjnych lotniska John F. Kennedy International Airport w latach 2009–2019.....	140
Rys. 4.38. ACF (a) i PACF (b) szeregu liczby pasażerów komercyjnych lotniska John F. Kennedy International Airport w latach 2009–2019	142
Rys. 4.39. Zróznicowany z opóźnieniem 1 szereg liczby pasażerów komercyjnych lotniska John F. Kennedy International Airport (a), ACF (b), PACF (c)	143
Rys. 4.40. Zróznicowany z opóźnieniem 1 i 12 szereg liczby pasażerów komercyjnych lotniska John F. Kennedy International Airport (a), ACF (b), PACF (c)	144
Rys. 4.41. Wyniki estymacji modelu $ARIMA(0,1,1)(1,1,0)_{12}$ szeregu liczby pasażerów komercyjnych lotniska John F. Kennedy International Airport.....	145
Rys. 4.42. Analiza reszt modelu $ARIMA(0,1,1)(1,1,0)_{12}$ szeregu liczby pasażerów komercyjnych lotniska John F. Kennedy International Airport.....	145
Rys. 4.42. Analiza reszt modelu $ARIMA(0,1,1)(1,1,0)_{12}$ szeregu liczby pasażerów komercyjnych lotniska John F. Kennedy International Airport.....	146
Rys. 4.43. Wyniki estymacji modelu $ARIMA(0,1,1)(2,1,0)_{12}$ szeregu liczby pasażerów komercyjnych lotniska John F. Kennedy International Airport.....	147
Rys. 4.44. Analiza reszt modelu $ARIMA(0,1,1)(2,1,0)_{12}$ szeregu liczby pasażerów komercyjnych lotniska John F. Kennedy International Airport: histogram (a), wykres normalności (b), ACF (c)	148
Rys. 4.44. Analiza reszt modelu $ARIMA(0,1,1)(2,1,0)_{12}$ szeregu liczby pasażerów komercyjnych lotniska John F. Kennedy International Airport: PACF (d)	149

Rys. 5.1.	Miesięczne wielkości przeładunków w porcie Szanghaj w okresie od stycznia 2016 do maja 2021 r.....	153
Rys. 5.2.	Szereg czasowy miesięcznych wielkości przeładunków w porcie Szanghaj, wartości uzyskane z modelu oraz prognozowane wielkości przeładunków	154
Rys. 5.3.	Szereg czasowy dziennej całkowitej liczby chorych na COVID-19 w USA w okresie od 15 lutego do 18 lipca 2020 roku	155
Rys. 5.4.	Prognoza dziennej całkowitej liczby chorych na COVID-19 w USA na okres od 19 lipca do 15 września 2020 roku	157
Rys. 5.5.	Dzienne wartości parytetu USD/TRY w okresie od 3 stycznia 2005 roku do 8 marca 2017 roku	158
Rys. 5.6.	Wartości rzeczywiste i prognozowane dziennych wartości parytetu USD/TRY w okresie od 27 stycznia 2017 do 9 marca 2017 roku model STM (a), model LTM (b).....	159
Rys. 5.7.	Ceny w Bengaluru w latach 2012–2017 owoców mango (a), kalafiorów (b).....	161
Rys. 5.8.	Prognoza ceny w Bengaluru w latach 2012–2017 owoców mango (a), kalafiorów (b)	162
Rys. 5.9.	Średnie miesięczne wartości nasłonecznienia (kWh/m ² /dzień) w New Delhi w latach 1984–2015.....	164
Rys. 5.10.	Obszar objęty obszarowym prognozowaniem nasłonecznienia	165
Rys. 5.11.	Prognoza średniorocznego nasłonecznienia badanego obszaru w roku 2021	166
Rys. 6.1.	Przykład szeregu	168
Rys. 6.2.	Przykład ACF i PACF	169

Oddawany do rąk Czytelników podręcznik „Prognozowanie w zarządzaniu przedsiębiorstwem” przekazuje wiedzę, zarówno podstawową, jak i zaawansowaną, z zakresu metod i narzędzi prognozowania gospodarczego. Podręcznik zaprojektowany został jako składający się z dziewięciu odrębnych części, które obejmują:

- wprowadzenie do metodyki prognozowania,
- podstawy prognozowania szeregów czasowych,
- zastosowanie modeli adaptacyjnych,
- zastosowanie modeli trendu,
- wykorzystanie modeli ARIMA do prognozowania gospodarczego,
- prognozowanie na podstawie modeli ekonometrycznych,
- zastosowanie w prognozowaniu metod sztucznej inteligencji,
- prognozowanie z zastosowaniem metod heurystycznych,
- wykorzystanie różnych metod prognozowania w zarządzaniu przedsiębiorstwem.

Każda część kończy się pytaniami i zadaniami, których samodzielne rozwiązanie pomoże Czytelnikowi ugruntować zdobytą wiedzę.

Autorami V części podręcznika są pracownicy Wydziału Inżynierii Zarządzania Politechniki Białostockiej: prof. n. ekon. i n. techn. dr hab. inż. Joanicjusz Nazarko oraz dr hab. inż. Ewa Chodakowska, prof. PB.

W podręczniku skoncentrowano się na metodach i technikach prognozowania, które są szczególnie przydatne w zarządzaniu przedsiębiorstwem. Autorzy, dokonując wyboru materiału i jego zakresu merytorycznego, oparli się na swoim wieloletnim doświadczeniu w nauczaniu prognozowania na studiach magisterskich, podyplomowych i MBA oraz na doświadczeniach zdobytych w uczelniach zagranicznych. Dzięki stałej współpracy z gospodarką zawarli w swojej pracy wiedzę nie tylko teoretyczną, lecz także praktyczną.

Recenzowana publikacja to wartościowy podręcznik akademicki, który może być polecony zarówno studentom, jak i szerszemu gronu czytelników zainteresowanych problematyką modelowania i prognozowania szeregów czasowych w procesach planowania i zarządzania przedsiębiorstwem. Autorzy zdołali przedstawić skomplikowane nieraz zagadnienia związane ze stosowaniem modeli ARIMA w sposób przystępny, zachowując jednocześnie niezbędną ścisłość wywodów. Materiał teoretyczny w poszczególnych rozdziałach jest bogato ilustrowany szczegółowymi przykładami opracowanymi przy wykorzystaniu pakietu statystycznego STATISTICA, co czyni podręcznik cennym poradnikiem metodycznym stosowania modeli ARIMA.

prof. dr hab. Dorota Witkowska

Katedra Zarządzania Finansami Przedsiębiorstwa
Uniwersytet Łódzki

Prognozowanie na podstawie modeli ARIMA to podręcznik o wysokim poziomie merytorycznym oraz praktycznym. Zaprezentowano w nim w sposób przyjazny dla czytelnika zarówno zagadnienia teoretyczne, w tym określone twierdzenia oraz wzory matematyczne, jak i możliwości aplikacyjne w wielu dziedzinach życia gospodarczego i społecznego. Po każdej części zawierającej zagadnienia teoretyczne przedstawiono w jaki sposób można dane zagadnienie modelować w programie komputerowym Statistica. Sposób postępowania opisano za pomocą kolejnych kroków, które należy wykonać oraz zilustrowano je za pomocą zrzutów ekranów dialogowych programu Statistica. Dzięki temu czytelnik posługując się podręcznikiem może zarówno opanować zagadnienia teoretyczne jak i samodzielnie nauczyć się prowadzenia analiz i modelowania szeregów czasowych za pomocą modeli ARIMA w programie komputerowym Statistica. Zwłaszcza że dla każdego z omawianych zagadnień zaprezentowano proces modelowania na przykładowych danych wraz z uzyskanymi wynikami oraz ich dogłębną analizą, interpretacją i komentarzem Autorów.

dr hab. Stanisław Łobjko, prof. SGH

Kolegium Nauk o Przedsiębiorstwie
Szkoła Główna Handlowa