

Recenzja książki pt.
Analiza i prognozowanie szeregów czasowych.
Praktyczne wprowadzenie na podstawie
środowiska R

autorstwa: A. Zagdański i A. Suchwałko



Niniejsza książka [6]¹ stanowi praktyczne wprowadzenie do modelowania w środowisku **R** (zob. [3], [4], [5], [1], [2]) różnorodnych danych zbieranych w regularnych odstępach czasu. Książka adresowana jest do wszystkich zainteresowanych modelami szeregów czasowych a szczególnie do studentów i absolwentów kierunków ścisłych, ekonomicznych oraz technicznych.

Książka² składa się z ośmiu rozdziałów, w tym pierwszego, który jest wstępem. Zawiera ponadto jeden dodatek, bibliografię oraz skorowidz³. Wydanie liczy 341 stron. Autorzy udostępniają również bibliotekę (pakiet) **TSAFbook**, opracowana dla środowiska **R** a dostępna w repozytorium **CRAN**⁴.

Rozdział drugi stanowi wprowadzenie do podstawowych pojęć używanych w książce. Przedstawiono w nim etapy analizowania danych modelami szeregów czasowych. Co bardzo cenne, od razu zilustrowano wymienione etapy na przykładzie danych o liczbie turystów korzystających z noclegów na terenie Dolnego Śląska. Przykład, z konieczności, analizowany jest pobieżnie, ale daje czytelnikowi możliwość popróbowania od razu własnych sił z pakietem **R**.

W rozdziale trzecim przedstawiono techniczne aspekty związane z analizą szeregów czasowych w środowisku **R**. Umówiono w nim formaty danych używane w pakiecie, sposoby wczytywania i importowania danych oraz podstawowe struktury wykorzystywane w **R** do reprezentacji szeregów czasowych.

¹A. Zagdański, A. Suchwałko. *Analiza i prognozowanie szeregów czasowych. Praktyczne wprowadzenie na podstawie środowiska R*. Wydawnictwo Naukowe PWN, Warszawa 2016
2010 Mathematics Subject Classification[2010]: 62-01;62M10;62M20;62-07

Słowa kluczowe: modele szeregów czasowych; prognozowanie; analiza danych

²2010 Mathematics Subject Classification: 62-01;62M10;62M20;62-07

³Słowa kluczowe: modele szeregów czasowych; prognozowanie; analiza danych

⁴Bibliotekę instalujemy z konsoli **R** poleceniem `install.packages("TSAFbook")`.

Opisano także ideę podziału danych na część uczącą i testową. Przedstawiono sposoby symulowania podstawowych modeli szeregów czasowych takich jak biały szum czy błędzenie losowe.

Rozdział czwarty zawiera przegląd najważniejszych metod wykorzystywanych do graficznej prezentacji szeregów graficznych. Zostały opisane podstawowe narzędzia do wizualizacji danych takie jak funkcja „plot” a także narzędzia do tworzenia wykresów sezonowych oraz wykresów autokorelacji. Do celów ilustracji wybrano dwa przykładowe zbiory danych, jeden rozważany już wcześniej oraz dane o dochodzie krajowym brutto w USA.

W rozdziale piątym przedstawiono najbardziej popularne transformacje szeregów czasowych. Zaczęto od omawiania korekt związanych z układem kalendarza a następnie opisano transformację Boxa-Coxa. Następnym rozważanym rodzajem transformacji jest różnicowanie i transformacje prowadzące do eliminacji trendów i sezonowości. Rozdział kończy, bardzo przydatny z punktu widzenia niedoświadczonego czytelnika, opis prawidłowej kolejności wykonywania transformacji.

Rozdział szósty poświęcony jest problemowi dekompozycji szeregów czasowych. Umówiona została koncepcja dekompozycji szeregów oraz zostały przedstawione metody do wyodrębnienia składowych takich jak trend, cykliczność i sezonowość w analizowanych szeregach. W rozdziale przedstawiono także algorytm dekompozycji w wersji zaimplementowanej w funkcji „decompose” pakietu **R**. Wszystkie metody zostały szczegółowo zilustrowane na przykładowych danych. Na końcu rozdziału zostały krótko wymienione zaawansowane metody dekompozycji szeregów czasowych takie jak regresja harmoniczna, wraz z odpowiednimi odwołaniami do literatury.

Rozdziały siódmy i ósmy są zdecydowanie najobszerniejsze i stanowią kluczową część książki. Rozdział siódmy poświęcony jest procesom ARIMA i ich uogólnieniom umożliwiającym modelowanie danych posiadających efekty trendu i sezonowości. Na początku podane są definicje modeli ruchomej średniej (MA) i autoregresji (AR) wraz z warunkami na ich stacjonarność. Następnie zdefiniowany jest model autoregresji ruchomej średniej (ARMA) i wymienione są warunki na jego stacjonarność, przyczynowość i odwracalność. Modele stacjonarne kończy pojęcie modelu sezonowego ARMA. W drugiej części rozdziału znajduje się przegląd modeli niestacjonarnych takich jak ARIMA i SARIMA, które są uogólnieniem procesów ARMA. Przedstawiono w jaki sposób symulować te procesy w pakiecie **R** i przeanalizowano metody identyfikacji modeli w tym problem wyboru ich rzędów. Metody te zostały zilustrowane danymi rzeczywistymi rozważanymi w poprzednich rozdziałach. Rozdział kończą sposoby walidacji modelu oparte na analizie reszt oraz kryteria wyboru optymalnego modelu. Wymienione zostały także kryteria oceniające dokładność prognoz polegające na obliczaniu różnych miar błędu prognozy. Ostatni punkt zawiera szczegółowy opis procedury automatycznego wyboru optymalnego modelu, która została zaimplementowana w

środoowisku **R**.

Rozdział ósmy, najdluższy w książce, poświęcony jest prognozowaniu, które jest jednym z głównych zadań analizy szeregów czasowych. Przedstawiono w nim proste metody prognozowania, metody oparte na modelach ARIMA, algorytmy wygładzania wykładniczego oraz prognozy oparte na dekompozycji szeregu. Rozdział ten wykorzystuje wiedzę z poprzednich rozdziałów, ponieważ wiele z używanych tu pojęć pojawiło się już wcześniej. Rozważane proste metody to metody oparte na średniej ruchomej i jej szczególnych przypadkach. Dokładność prognoz została przeanalizowana za pomocą liczbowych kryteriów takich jak średni błąd kwadratowy i za pomocą przedziałów predykcyjnych. Rozważano w tym celu dwa podejścia: (i) prognozy uzyskane na podstawie tych samych danych, które były wykorzystane do konstrukcji prognoz oraz (ii) przy użyciu części uczącej do dopasowania modelu i konstrukcji prognoz oraz części testowej wykorzystanej do oceny dokładności prognoz. Rozdział ten kończą bardzo praktyczne rozważania jak wybrać optymalną metodę prognozowania.

W dodatku A przedstawiono najważniejsze informacje w jaki sposób pogłębić wiedzę z pakietu **R**.

Każdy z rozdziałów kończy się ćwiczeniami, które dobrze ilustrują zagadnienia w nim badane. W książce znajdują się fragmenty kodów pozwalające na samodzielne wykonanie opisywanych analiz. W książce oprócz definicji są też 64 uwagi, które zawierają trafne komentarze do dopiero co wprowadzonych pojęć. Są one interesującym źródłem praktycznych informacji dla osób zajmujących się analizą danych. Spis literatury liczy 26 pozycji w większości napisanych w języku angielskim. Zawiera on książki o tematyce szeregów czasowych a także artykuły, w których wprowadzone niestandardowe metody rozważane w książce.

Jedyna moja krytyczna uwagi dotyczy nie najlepszej jakości rysunków w książce, czasami trudno jest rozróżnić rodzaje poszczególnych linii i rodzaje kolorowanych powierzchni. Przy okazji następnego wydania można by także uwzględnić zdefiniowanie podstawowych pojęć statystycznych, które często pojawiają się w tekście a mianowicie pojęcie p-wartości czy poziomu istotności testu.

Podsumowując, omawiana książka jest wartościową pozycją adresowaną do czytelnika zainteresowanego analizą danych przy pomocy modeli szeregów czasowych. W książce przedstawiono w czytelny sposób metody służące do identyfikacji i walidacji modeli szeregów czasowych typu ARIMA. Zawiera ona wiele ciekawych przykładów połączonych z ćwiczeniami do samodzielnego rozwiązania, a ponadto ilustruje omawiane metody danymi rzeczywistymi z różnych dziedzin. Wydaje mi się, że ta pozycja wypełnia lukę w polskiej literaturze dotyczącej modelowania szeregów czasowych. Często słyszę od swoich studentów pytanie o książki napisane w języku polskim zawierające wprowadzenie do tematyki modeli szeregów czasowych. Ta książka bardzo

dobrze odpowiada na to zapotrzebowanie.

LITERATURA

- [1] R.A. Becker, J.M. Chambers, A.R. Wilks (1988), *The New S Language*. Chapman & Hall, London, 1988.
- [2] P. Biecek (2013), *Analiza danych z programem R. Modele liniowe z efektami stałymi, losowymi i mieszanymi*. Wydawnictwo Naukowe PWN, Warszawa 2013.
- [3] *Notatka o R w Oberwolfach References on Mathematical Software*, dostępne pod adresem <http://orms.mfo.de/project?id=266>, data dostępu: 23.12.2016.
- [4] *Notatka o R w swMath*, dostępne pod adresem <http://www.swmath.org/software/771>, data dostępu: 23.12.2016.
- [5] *The R Project for Statistical Computing*, dostępne pod adresem <https://www.r-project.org/>, data dostępu: 23.12.2016.
- [6] A. Zagdański, A. Suchwałko (2016), *Analiza i prognozowanie szeregów czasowych. Praktyczne wprowadzenie na podstawie środowiska R*. Wydawnictwo Naukowe PWN, Warszawa 2016, ISBN 978-83-01-18356-1.

On the book „The analysis and forecasting of time series. Practical introduction on the basis of the R environment” by: A. Zagdanski and A. Suchwałko

Abstract The book under review provides a practical introduction to modelling in the R environment various data collected at regular intervals. It is addressed to the general audience interested in time series modelling, and especially to students and graduates in mathematics, economics and engineering. The book is a valuable source of information for anyone interested in data analysis by means of the autoregressive integrated moving average (ARIMA) processes.

2010 Mathematics Subject Classification: 62-01;62M10;62M20;62-07.

Key words and phrases: time series models; prediction; data analysis.

This book [6] provides a practical introduction to modelling with **R** (see [3], [4], [5], [1], [2]) various data collected at regular intervals. The book is addressed to the general audience interested in time series modelling, and especially to students and graduates in mathematics, economics and engineering.

The book consists of eight chapters. It includes an appendix, bibliography and index. This edition has 341 pages. For the book purposes the authors have developed a library (package) **TSAFbook** which is available in the repository **CRAN**.

Chapters 7 and 8 are by far the most extensive, and represent a key part of the book. Chapter 7 is devoted to autoregressive moving average (ARMA) processes and their generalizations that allow for modeling data with observed trend and seasonality. First, ARMA models are defined and conditions for their stationarity, causality and reversibility are provided. In the second

part of the chapter non-stationary models such as ARIMA and SARIMA, which are a generalization of the ARMA, are studied. The authors clearly explain how to simulate them in the **R** environment and present methods for their identification, including the problem of selecting the order. Next, methods for the model validation, which are based on the analysis of the residuals and the criteria for choosing the optimal model, are discussed. The last section provides a detailed description of the automatic optimal model selection procedure, which was implemented in the **R** environment. All methods are illustrated on real empirical data which were introduced in the previous chapters. Chapter 8, the longest in the book, is devoted to forecasting, which is one of the main tasks of time series analysis. Simple forecasting methods, methods based on ARIMA models, exponential smoothing algorithms and forecasts based on the series decomposition, are presented. This chapter summarizes the ideas from the previous chapters, because many of the concepts used here were introduced earlier. The accuracy of the forecasts was analyzed by means of numerical criteria such as the mean square error and with the help of prediction intervals. This chapter ends with practical considerations how to choose the optimal forecasting method. Each chapter ends with exercises that well illustrate the discussed problems. Moreover, the reader may find computer codes that allow to practice the acquired skills. In addition, the book contains 64 numbered remarks which contain relevant comments on the newly introduced concepts. They are an interesting source of practical information for data analysts.

My only critical remark concerns the quality of pictures. It is sometimes difficult to distinguish between the different types of lines and types of shaded areas. It would be also beneficial to consider, in the next edition, defining the crucial statistical concepts, often used in the book, namely, p-value and significance level of a statistical test.

To summarize, I strongly believe that the book is a valuable source of information for the reader interested in data analysis by means of time series models. In the book methods for identification and validation of ARIMA time series models along with prediction techniques are presented. The problems are adequately illustrated the real-world data from different fields.

KRZYSZTOF BURNECKI
FACULTY OF PURE AND APPLIED MATHEMATICS
CHAIR OF APPLIED MATHEMATICS
WROCLAW UNIVERSITY OF SCIENCE AND TECHNOLOGY
UL. WYBRZEŻE WYSPIAŃSKIEGO 27, PL-50-370 WROCLAW, POLAND
E-mail: Krzysztof.Burnecki@pwr.edu.pl

(Received: 23rd December 2017; revised: 31st December 2017)