Statistics and Computing

Series Editor
Wolfgang Karl Härdle, Humboldt-Universität zu Berlin, Berlin, Germany

Statistics and Computing (SC) includes monographs and advanced texts on statistical computing and statistical packages.

Changquan Huang • Alla Petukhina

Applied Time Series Analysis and Forecasting with Python

Springer

Changquan Huang
School of Economics
Xiamen University
Xiamen, China

Alla Petukhina
School of Computing, Communication and Business
HTW Berlin
Berlin, Germany

ISSN 1431-8784 ISSN 2197-1706 (electronic)
Statistics and Computing
ISBN 978-3-031-13583-5 ISBN 978-3-031-13584-2 (eBook)
https://doi.org/10.1007/978-3-031-13584-2

Mathematics Subject Classification: 62-01, 62M10, 62-04, 62-07, 62M20, 62M45, 62F03, 62F10

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

It is a data-rich era, and data exist almost everywhere. One of the greatest challenges nowadays is how to deal with various kinds of data. It is well known that time series data are the most common data type. And thus methods and techniques for analyzing and forecasting time series have become one of the indispensable tools to handle real-world data problems. Part of this book is rightly concerned with these methods and techniques. It will introduce a wide range of time series models and approaches to building adequate models. Another part is about the general-purpose programming language Python. Python's history is relatively short, but its popularity has been rising steadily. In recent years, Python has been continually taking leading position in solving data-scientific problems and artificial intelligence challenges. The book will show you how to use Python and its extension packages to implement time series analysis and forecasting. Therefore, it is an organic combination of the principle of time series analysis and Python programming.

This book has grown out of a course in time series analysis that Changquan Huang has been teaching at Xiamen University since 2003. More than 18 years of experience in teaching the time series analysis course has made him realize and understand the difficulties of students taking this course. For this reason, during the course of writing the book, he has always been doing his best to let the book be reader friendly and interesting in the hope that the reader can grasp the essence of time series analysis thoroughly and quickly.

The book is intended for an undergraduate and graduate audience as well as for everyone interested in time series analysis and forecasting with Python. To understand the book, only a prerequisite knowledge in probability theory and statistics is needed, which is equivalent to an undergraduate's probability and statistics course for two semesters. Besides, a knowledge of linear (matrix) algebra is helpful in better understanding Chaps. 7–9 of the book.

Changquan Huang wrote every chapter of this book as well as the Python code, and is responsible for the whole book. Alla Petukhina validated the Python code and created Quantlets, and the code of numerical examples has been indicated with a small sign  . We believe that these publicly available Quantlets on www.quantlet.

com and https://github.com/QuantLet/pyTSA/ create a valuable contribution to the distribution of knowledge in statistical science. We welcome all readers of this book to propose changes to our existing codes or add codes in other programming languages. A free online companion course to the book developed together with Professor Wolfgang Karl Härdle is available through  https://quantinar.com.

Our thanks go to Guido van Rossum, the Python Software Foundation, and all the open-source Python package developers for making Python applications in various fields possible. In particular, our thanks go to Kevin Sheppard from University of Oxford for answering our consultation about his excellent Python package `arch` and to the anonymous referee for suggestions. We also thank Veronika Rosteck, Daniel Ignatius Jagadisan, and the Springer team for their support and patience.

Alla Petukhina would like to thank Prof. Wolfgang Karl Härdle and all colleagues from IRTG 1792 at Humboldt University of Berlin for their help and valuable advice, as well as her parents, sister, and son for unconditional support and inspiration.

Changquan Huang would like to thank Prof. Yang Xiangqun, Prof. Wang Zikun, and Prof. GU Ming-gao for mentoring him during the course of his postgraduate study. He particularly thanks Prof. Wolfgang Karl Härdle for inviting him to visit Humboldt University of Berlin. He is grateful to Jiancheng Jiang from University of North Carolina at Charlotte, Hongtu Zhu from University of North Carolina at Chapel Hill, and Xin Yuan Song from The Chinese University of Hong Kong for their encouragement. Finally, he is indebted to his family for their love and understanding.

Xiamen, China — Changquan Huang
Berlin, Germany — Alla Petukhina
May 2022

Contents

1 Time Series Concepts and Python 1
1.1 The Concept of Time Series 1
1.1.1 What Is Time Series 1
1.1.2 Brief History of Time Series Analysis 4
1.1.3 Objectives of Time Series Analysis 6
1.2 The Programming Language Python 6
1.2.1 Introduction and Installing 7
1.2.2 Demonstrations 7
1.2.3 Python Extension Packages and Some Usages 11
1.3 Time Series Moment Functions and Stationarity 15
1.3.1 Moment Functions 15
1.3.2 Stationarity and Ergodicity 16
1.3.3 Sample Autocorrelation Function 18
1.3.4 White Noise and Random Walk 21
1.4 Time Series Data Visualization 29
Problems 34

2 Exploratory Time Series Data Analysis 37
2.1 Partial Autocorrelation Functions 37
2.1.1 Definition of PACF 37
2.1.2 Sample PACF and PACF Plot 39
2.2 White Noise Test 42
2.3 Simple Time Series Compositions 47
2.4 Time Series Decomposition and Smoothing 53
2.4.1 Deterministic Components and Decomposition Models 53
2.4.2 Decomposition and Smoothing Methods 58
2.4.3 Example 61
Problems 68

3 Stationary Time Series Models 71
3.1 Backshift Operator, Differencing, and Stationarity Test 71
3.1.1 Backshift Operator 71
3.1.2 Differencing and Stationarity 72
3.1.3 KPSS Stationarity Test 73
3.2 Moving Average Models 80
3.2.1 Definition of Moving Average Models 80
3.2.2 Properties of MA Models 84
3.2.3 Invertibility 85
3.3 Autoregressive Models 88
3.3.1 Definition of Autoregressive Models 88
3.3.2 Durbin-Levinson Recursion Algorithm 90
3.3.3 Properties of Autoregressive Models 92
3.3.4 Stationarity and Causality of AR Models 94
3.4 Autoregressive Moving Average Models 98
3.4.1 Definitions 98
3.4.2 Properties of ARMA Models 100
Problems 105

4 ARMA and ARIMA Modeling and Forecasting 107
4.1 Model Building Problems 107
4.2 Estimation Methods 108
4.2.1 The Innovations Algorithm 109
4.2.2 Method of Moments 110
4.2.3 Method of Conditional Least Squares 111
4.2.4 Method of Maximum Likelihood 113
4.3 Order Determination 115
4.4 Diagnosis of Models 116
4.5 Forecasting 118
4.6 Examples 119
Problems 142

5 Nonstationary Time Series Models 143
5.1 The Box-Jenkins Method 143
5.1.1 Seasonal Differencing 143
5.1.2 SARIMA Models 147
5.2 SARIMA Model Building 155
5.2.1 General Idea 155
5.2.2 Case Studies 156
5.3 REGARMA Models 165
Problems 174

6 Financial Time Series and Related Models 177
6.1 Stylized Facts of Financial Time Series 177
6.1.1 Examples of Return Series 177
6.1.2 Stylized Facts of Financial Time Series 182

6.2 GARCH Models 183
6.2.1 ARCH Models 183
6.2.2 GARCH Models 185
6.2.3 Estimation and Testing 188
6.2.4 Examples 190
6.3 Other Extensions 204
6.3.1 EGARCH Models 204
6.3.2 TGARCH Models 205
6.3.3 An Example 205
Problems 212

7 Multivariate Time Series Analysis 215
7.1 Basic Concepts 215
7.1.1 Covariance and Correlation Matrix Functions 215
7.1.2 Stationarity and Vector White Noise 217
7.1.3 Sample Covariance and Correlation Matrices 219
7.1.4 Multivariate Portmanteau Test 220
7.2 VARMA Models 226
7.2.1 Definitions 227
7.2.2 Properties 229
7.3 VAR Model Building and Analysis 233
7.3.1 VAR(1) Representation of VARMA Processes 233
7.3.2 VAR Model Building Steps 233
7.3.3 Granger Causality 235
7.3.4 Impulse Response Analysis 236
7.4 Examples 237
Problems 255

8 State Space Models and Markov Switching Models 257
8.1 State Space Models and Representations 257
8.1.1 State Space Models 258
8.1.2 State Space Representations of Time Series 259
8.2 Kalman Recursions 261
8.3 Local-Level Model and SARIMAX Models 263
8.3.1 Local-Level Model 263
8.3.2 SARIMAX Models 265
8.4 Markov Switching Models 271
8.4.1 Definitions 271
8.4.2 Examples 273
Problems 285

9 Nonstationarity and Cointegrations 287
9.1 Stochastic Trend and Stochastic Seasonality 287
9.1.1 Deterministic Trend and Stochastic Trend 287
9.1.2 Deterministic Seasonality and Stochastic Seasonality 293

9.2 Brownian Motions and Simulation 302
9.2.1 Probability Space 302
9.2.2 Brownian Motions 304
9.3 Stationarity, Nonstationarity, and Unit Root Tests 306
9.3.1 Trend Stationarity and Difference Stationarity 306
9.3.2 Unit Root Tests 308
9.3.3 Stationarity Tests 316
9.4 Cointegrations and Granger's Representation Theorem 318
9.4.1 Spurious Regressions and $I(d)$ Processes 318
9.4.2 Cointegrations 322
9.4.3 Granger's Representation Theorem 324
9.4.4 Estimation of Vector Error Correction Models 328
9.4.5 Real Case of Spurious Regression and Noncointegration 334
Problems 339

10 Modern Machine Learning Methods for Time Series Analysis 341
10.1 Introduction 341
10.1.1 Brief History of Artificial Intelligence 341
10.1.2 AI in Time Series Analysis 343
10.2 Artificial Neural Networks 344
10.2.1 Artificial Neural Network Developments 344
10.2.2 Neural Network Models 347
10.3 Deep Learning and Backpropagation Algorithms 350
10.3.1 What Is Deep Learning? 350
10.3.2 Gradient Descent and Backpropagation Algorithms 350
10.4 Time Series Forecasting and TensorFlow 351
10.4.1 Time Series Forecasting 351
10.4.2 TensorFlow and Keras 351
10.5 Implementation and Example 352
10.5.1 Implementation Steps 352
10.5.2 An Example 353
10.6 Concluding Remarks 360
Problems 360

References 363

Index 369