

Preface

Statistical forecasting procedures are used to solve many applied problems in engineering, economics, finance, medicine, environmental studies, etc. For the majority of the developed statistical forecasting procedures, optimality (or asymptotic optimality as observation time increases) is proved w.r.t. the mean square forecast risk under the assumptions of an underlying hypothetical model. In practice, however, the observed data usually deviates from hypothetical models: random observation errors may be non-Gaussian, correlated, or inhomogeneous; the data may be contaminated by outliers, level shifts, or missing values; trend, regression, and autoregression functions do not necessarily belong to the declared parametric families, etc. Unfortunately, the forecasting procedures which have been proved to be optimal under the hypothetical model often become unstable under even small model distortions, resulting in forecast risks or mean square errors which are significantly higher than the theoretical values obtained in the absence of distortion. This necessitates the development of *robust* statistical algorithms, which are designed to retain most of their properties under small deviations from model assumptions.

The available textbooks on the subjects of statistical forecasting and robust statistical methods can be split into two distinct clusters. *The first cluster* includes books on theoretical and applied aspects of statistical forecasting where little or no attention is paid to robustness. The focus of these books is on various hypothetical models, methods, and computer algorithms used in forecasting, as well as their performance in the absence of model distortions.

The second cluster includes books on robust statistics which are dedicated to such diverse subjects as robust statistical parameter estimation, robust hypothesis testing in parametric (e.g., shift-scale) families of probability distributions, regression analysis, discriminant analysis, cluster analysis, time series analysis, etc. However, the topic of robustness in statistical forecasting remains barely touched upon, and little or no information is provided on such important aspects of forecasting as analysis of risk increments due to different types and levels of distortion, estimation of critical distortion levels for the traditional forecasting procedures,

and development of robust forecasting procedures tailored to the distortion types that are commonly encountered in applications.

This monograph is an attempt to fill the described gap in the literature by going beyond the fundamental subjects of robust statistical estimation and robust statistical hypothesis testing and presenting a systematic collection of known and new results related to the following topical problems:

- Construction of mathematical models and descriptions of typical distortions in applied forecasting problems;
- Quantitative evaluation of the robustness of traditional forecasting procedures;
- Evaluation of critical distortion levels;
- Construction of new robust forecasting procedures satisfying the minimax-risk criterion.

Solving these problems answers the following questions, which are highly relevant to both theoretical and applied aspects of statistical forecasting:

- Which distortion types can be accommodated by forecasting procedures?
- What are the maximal distortion levels allowing for “safe” use of the traditional forecasting algorithms?
- How can we estimate the effect of distortions on the mean square risk of traditional forecasting algorithms?
- Which robust forecasting statistics are the most suitable under different types of distortions?

The monograph is organized into ten chapters. Chapter 1 serves as a general introduction to the subject of statistical forecasting, presenting its history and some of the possible applications. Chapter 2 describes the decision-theoretic approach to forecasting, which is different from the general statistical approach used in Chaps. 3–10. Chapter 3 presents mathematical models of the time series commonly used in statistical forecasting. Chapter 4 classifies types of model distortions and defines metrics for optimality and robustness in statistical forecasting. Chapter 5 presents methods for optimal parametric and nonparametric time series regression forecasting. In Chap. 6, robustness of these methods is evaluated, and robust forecasting statistics are constructed. A similar treatment of the $ARIMA(p, d, q)$ autoregressive integrated moving average time series model is presented in Chap. 7. Chapter 8 presents an analysis of optimality and robustness of forecasting based on autoregressive time series models under missing values. Robustness of multivariate time series forecasting based on systems of simultaneous equations is investigated in Chaps. 9, and 10 discusses forecasting of discrete time series.

The interdependence of the chapters is illustrated in Fig. 1 below; solid lines represent prerequisites, and dashed lines indicate weaker relations.

Presentation of the material within the chapters follows the pattern “model → method → algorithm → computation results based on simulated or real-world data.” The theoretical results are illustrated by computer experiments using real-world statistical data (eight instances) and simulated data (ten instances).

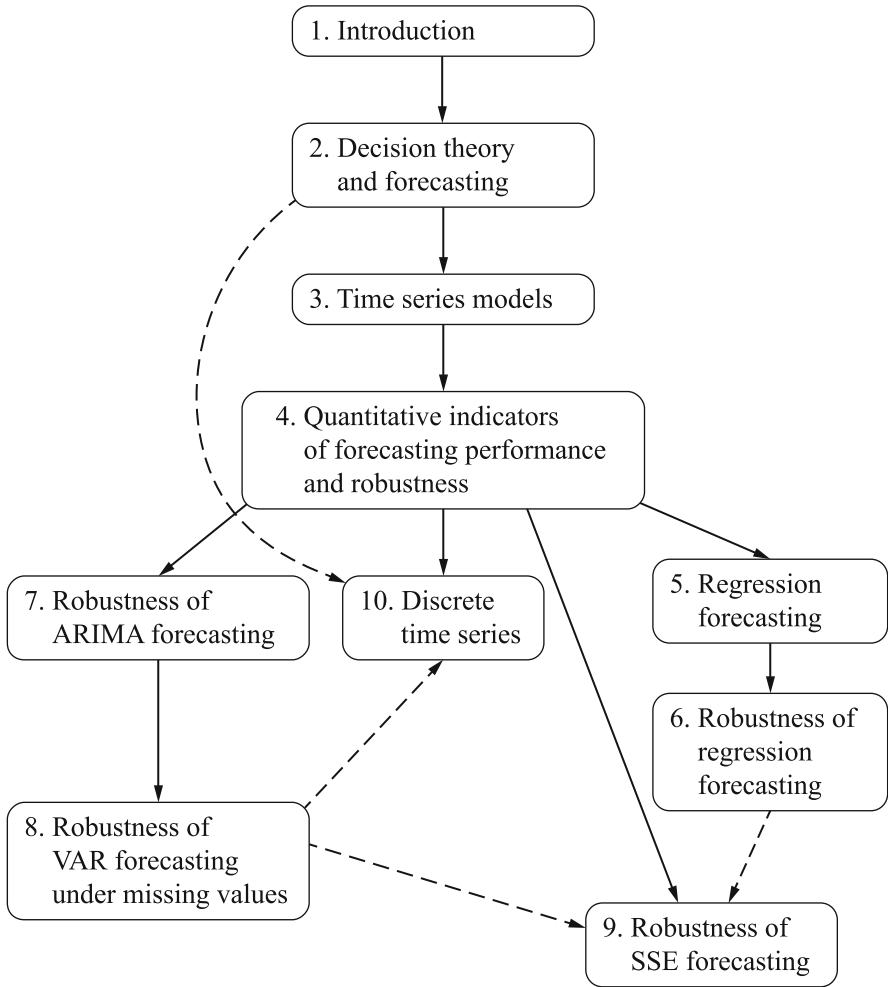


Fig. 1 The interdependence of the chapters

The robust forecasting algorithms described in the monograph have been implemented as computer programs; however, the developed software package cannot be made available to the reader (in particular, only the Russian-language user interface has been designed). The author has intentionally deferred from connecting the material presented in this book to any specific software package or framework. Instead, the methods and algorithms are presented in the most general form, allowing the reader to implement them within a statistical package of their choice (e.g., S-PLUS or R). This also ensures that software developers can easily incorporate the developed methods into their products.

The book is primarily intended for mathematicians, statisticians, and software developers in applied mathematics, computer science, data analysis, econometrics, financial engineering, and biometrics. It can also be recommended as a textbook for a one-semester course for advanced undergraduate and postgraduate students of the mentioned disciplines.

The author would like to thank Serguei Ayvazyan, Christophe Croux, Peter Filzmoser, Ursula Gather, Peter Huber, William Mallios, Helmut Rieder, and Peter Rousseeuw for the discussions that aroused his interest in robust statistics. The author is grateful to the anonymous reviewers for their attention in analyzing the manuscript, and the resulting comments and proposals on improving the book. The author also thanks his colleagues from the Research Institute for Applied Problems of Mathematics and Informatics, as well as the Department of Mathematical Modeling and Data Analysis of the Belarusian State University. The author thanks his Ph.D. students Alexander Huryn, Andrey Kostevich, Vladimir Malugin, Maksim Pashkevich, Andrey Piatlistkiy, Valery Voloshko, and Andrey Yarmola for their contributions to the joint research in the field of robust statistical methods and algorithms. The author would like to give special thanks to Nikolai Kalosha for his assistance in editing the English version of the manuscript. To conclude the list of people who made a significant impact on the book, the author must mention the wise counsel and helpful advice provided by Dr. Niels Peter Thomas and Veronika Rosteck of Springer.

Minsk, Belarus
April 2013

Yuriy Kharin



<http://www.springer.com/978-3-319-00839-4>

Robustness in Statistical Forecasting

Kharin, Y.

2013, XVI, 356 p. 47 illus., Hardcover

ISBN: 978-3-319-00839-4