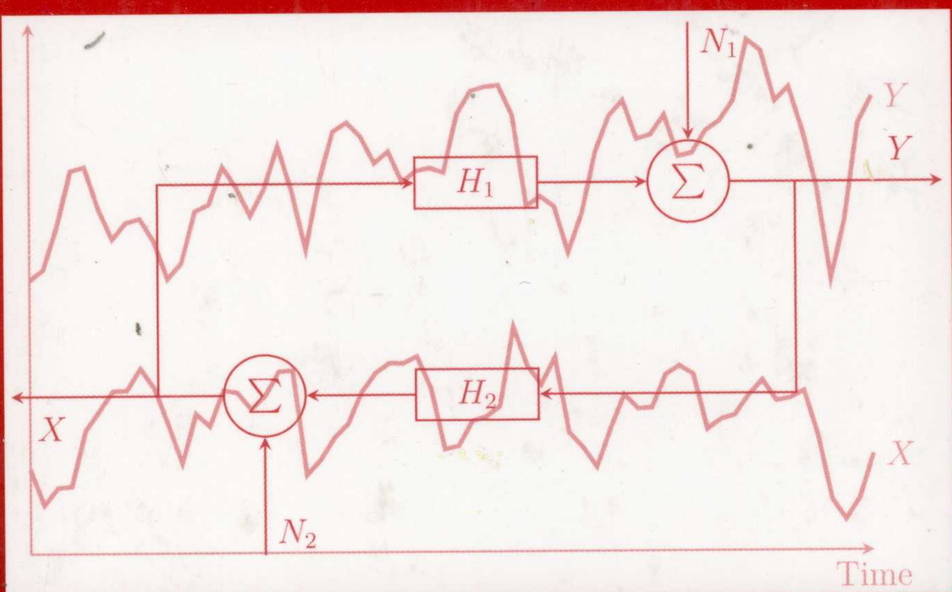
Texts in Statistical Science

Time Series Analysis



Henrik Madsen

Chapman & Hall/CRC
Taylor & Francis Group

Texts in Statistical Science

Time Series Analysis

Henrik Madsen
Technical University of Denmark

Chapman & Hall/CRC
Taylor & Francis Group
Boca Raton London New York

Chapman & Hall/CRC is an imprint of the
Taylor & Francis Group, an informa business

Chapman & Hall/CRC
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

International Standard Book Number-13: 978-1-4200-5967-0 (Hardcover)

Library of Congress Cataloging-in-Publication Data

Madsen, Henrik, 1955-
Time series analysis / Henrik Madsen.
p. cm. -- (Chapman & Hall/CRC texts in statistical science series ; v. 72)
Includes bibliographical references and index.
ISBN 978-1-4200-5967-0 (hardback : alk. paper)
1. Time-series analysis. I. Title. II. Series.

QA280.M32 2007
519.5'5--dc22 2007036211

Preface

The aim of this book is to give an introduction to time series analysis. The emphasis is on methods for modeling of linear stochastic systems. Both time domain and frequency domain descriptions will be given; however, emphasis is on the time domain description. Due to the highly different mathematical approaches needed for linear and non-linear systems, it is instructive to deal with them in seperate textbooks, which is why non-linear time series analysis is not a topic in this book—instead the reader is referred to Madsen, Holst, and Lindström (2007).

Theorems are used to emphasize the most important results. Proofs are given only when they clarify the results. Small problems are included at the end of most chapters, and a separate chapter with real-life problems is included as the final chapter of the book. This also serves as a demonstration of the many possible applications of time series analysis in areas such as physics, engineering, and econometrics.

During the sequence of chapters, more advanced stochastic models are gradually introduced; with this approach, the family of linear time series models and methods is put into a clear relationship. Following an initial chapter covering static models and methods such as the use of the general linear model for time series data, the rest of the book is devoted to stochastic dynamic models which are mostly formulated as difference equations, as in the famous ARMA or vector ARMA processes. It will be obvious to the reader of this book that even knowing how to solve difference equations becomes important for understanding the behavior of important aspects such as the autocovariance functions and the nature of the optimal predictions.

The important concept of time-varying systems is dealt with using a state space approach and the Kalman filter. However, the strength of also using adaptive estimation methods for on-line forecasting and control is often not adequately recognized. For instance, in finance the classical methods for forecasting are often not very useful, but, by using adaptive techniques, interesting results are often obtained.

The last chapter of this book is devoted to problems inspired by real life. Solutions to the problems are found at `http://www.imm.dtu.dk/~hm/time.series.analysis`. This home page also contains additional exercises, called assignments, intended for being solved using a computer with dedicated software for time series analysis.

I am grateful to all who have contributed with useful comments and suggestions for improvement. Especially, I would like to thank my colleagues Jan Holst, Henrik Spliid, Leif Mejlbro, Niels Kjølstad Poulsen, and Henrik Aalborg Nielsen for their valuable comments and suggestions. Furthermore, I would like to thank former students Morten Høier Olsen, Rasmus Tamstorf, and Jan Nygaard Nielsen for their great effort in proofreading and improving the first manuscript in Danish. For this 2007 edition in English, I would like to thank Devon Yates, Stig Mortensen, and Fannar Örn Thordarson for proofreading and their very useful suggestions. In particular, I am grateful to Anna Helga Jónsdóttir for her assistance with figures and examples. Finally, I would like to thank Morten Høgholm for both proofreading and for proposing and creating a new layout in LaTeX.

Lyngby, Denmark *Henrik Madsen*

Contents

CHAPTER 1

Introduction

Time series analysis deals with statistical methods for analyzing and modeling an ordered sequence of observations. This modeling results in a stochastic process model for the system which generated the data. The ordering of observations is most often, but not always, through time, particularly in terms of equally spaced time intervals. In some applied literature, time series are often called signals. In more theoretical literature a time series is just an observed or measured realization of a stochastic process.

This book on time series analysis focuses on modeling using linear models. During the sequence of chapters more and more advanced models for dynamic systems are introduced; by this approach the family of linear time series models and methods are placed in a structured relationship. In a subsequent book, non-linear time series models will be considered.

At the same time the book intends to provide the reader with an understanding of the mathematical and statistical background for time series analysis and modeling. In general the theory in this book is kept in a second order theory framework, focussing on the second order characteristics of the persistence in time as measured by the autocovariance and autocorrelation functions.

The separation of linear and non-linear time series analysis into two books facilitates a clear demonstration of the highly different mathematical approaches that are needed in each of these two cases. In linear time series analysis some of the most important approaches are linked to the fact that superposition is valid, and that classical frequency domain approaches are directly usable. For non-linear time series superposition is not valid and frequency domain approaches are in general not very useful.

The book can be seen as a text for graduates in engineering or science departments, but also for statisticians who want to understand the link between models and methods for linear dynamical systems and linear stochastic processes. The intention of the approach taken in this book is to bridge the gap between scientists or engineers, who often have a good understanding of methods for describing dynamical systems, and statisticians, who have a good understanding of statistical theory such as likelihood-based approaches.

In classical statistical analysis the correlation of data in time is often disregarded. For instance in regression analysis the assumption about serial uncorrelated residuals is often violated in practice. In this book it will be demonstrated that it is crucial to take this autocorrelation into account in the modeling procedure. Also for applications such as simulations and forecasting, we will most often be able to provide much more reasonable and realistic results by taking the autocorrelation into account.

On the other hand adequate methods and models for time series analysis can often be seen as a simple extension of linear regression analysis where previous observations of the dependent variable are included as explanatory variables in a simple linear regression type of model. This facilitates a rather easy approach for understanding many methods for time series analysis, as demonstrated in various chapters of this book.

There are a number of reasons for studying time series. These include a characterization of time series (or signals), understanding and modeling the data generating system, forecasting of future values, and optimal control of a system.

In the rest of this chapter we will first consider some typical time series and briefly mention the reasons for studying them and the methods to use in each case. Then some of the important methodologies and models are introduced with the help of an example where we wish to predict the monthly wheat prices. Finally the contents of the book is outlined while focusing on the model structures and their basic relations.

1.1 Examples of time series

In this section we will show examples of time series, and at the same time indicate possible applications of time series analysis. The examples contain both typical examples from economic studies and more technical applications.

1.1.1 Dollar to Euro exchange rate

The first example is the daily US dollar to Euro interbank exchange rate shown in Figure 1.1. This is a typical economic time series where time series analysis could be used to formulate a model for forecasting future values of the exchange rate. The analysis of such a problem relates to the models and methods described in Chapters 3, 5, and 6.

1.1.2 Number of monthly airline passengers

Next we consider the number of monthly airline passengers in the US shown in Figure 1.2. For this series a clear annual variation is seen. Again it might be useful to construct a model for making forecasts of the future number of airline passengers. Models and methods for analyzing time series with seasonal variation are described in Chapters 3, 5, and 6.

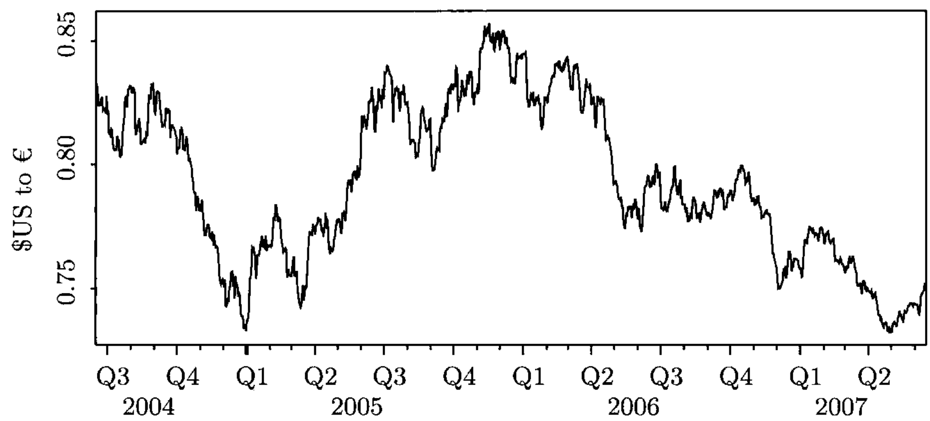


Figure 1.1: *Daily US dollar to Euro interbank exchange rate.*

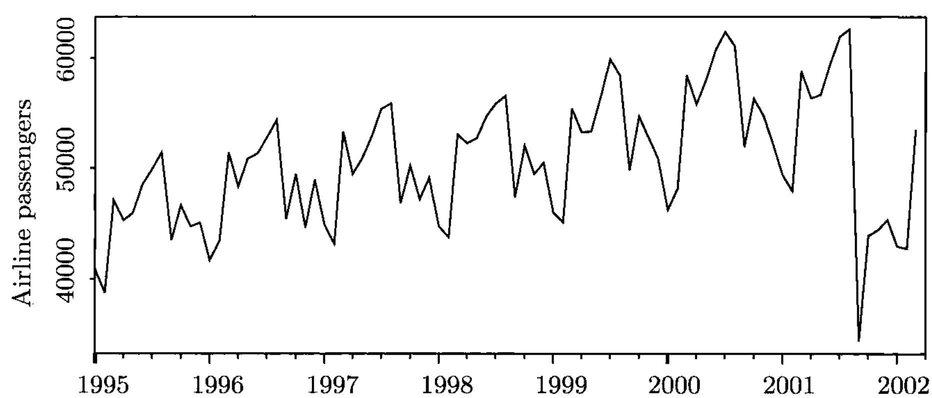


Figure 1.2: *Number of monthly airline passengers in the US. A clear annual variation can be seen in the series.*

1.1.3 Heat dynamics of a building

Now let us consider a more technical example. Figure 1.3 on the following page shows measurements from an unoccupied test building. The data on the lower plot show the indoor air temperature, while on the upper plot the ambient air temperature, the heat supply, and the solar radiation are shown.

For this example it might be interesting to characterize the thermal behavior of the building. As a part of that the so-called resistance against heat flux from inside to outside can be estimated. The resistance characterizes the insulation of the building. It might also be useful to establish a dynamic model for the building and to estimate the time constants. Knowledge of the time constants can be used for designing optimal controllers for the heat supply.

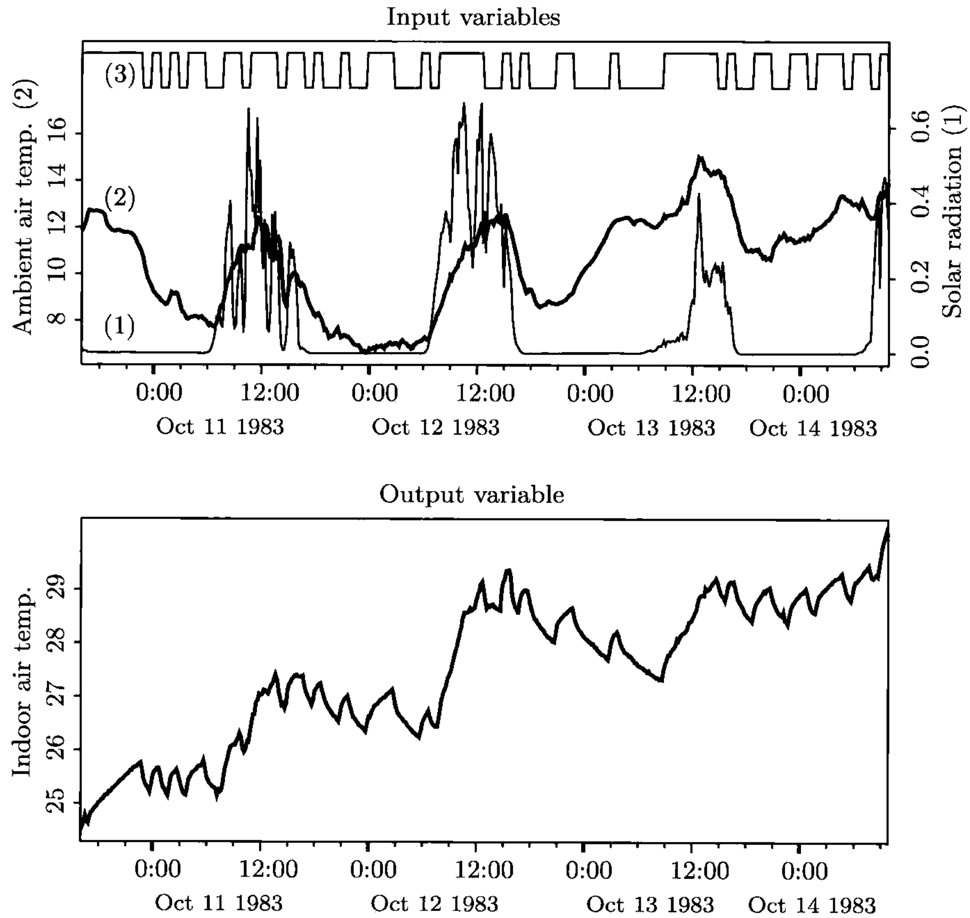


Figure 1.3: *Measurements from an unoccupied test building. The input variables are (1) solar radiation, (2) ambient air temperature, and (3) heat input. The output variable is the indoor air temperature.*

For this case methods for transfer function modeling as described in Chapter 8 can be used, where the input (explanatory) variables are the solar radiation, heat input, and outdoor air temperature, while the output (dependent) variable is the indoor air temperature. For the methods in Chapter 8 it is crucial that all the signals can be classified as either input or output series related to the system considered.

1.1.4 Predator-prey relationship

This example illustrates a typical multivariate time series, since it is not possible to classify one of the series as input and the other series as output. Figure 1.4 shows a widely studied predator-prey case, namely the series of annually traded skins of muskrat and mink by the Hudson's Bay Company

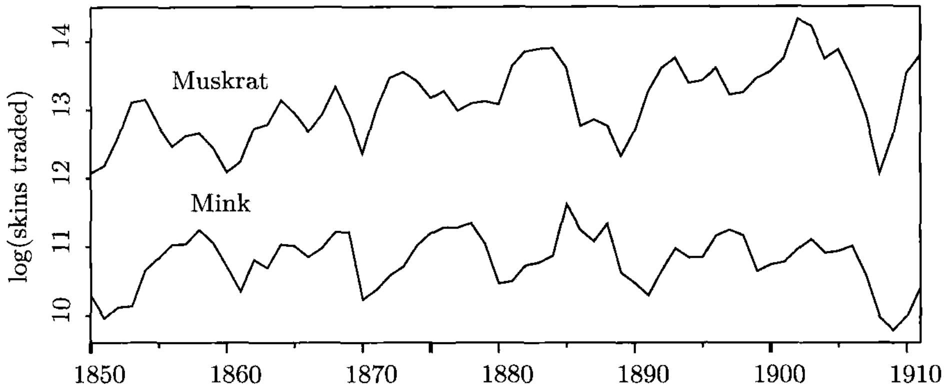


Figure 1.4: *Annually traded skins of muskrat and mink by the Hudson's Bay Company after logarithmic transformation. It is not possible to classify one of the series as input and the other series as output.*

during the 62 year period 1850–1911. In fact the population of muskrats depends on the population of mink, and the population of mink depends on the number of muskrats. In such cases both series must be included in a multivariate time series. This series has been considered in many texts on time series analysis, and the purpose is to describe in general the relation between populations of muskrat and mink. Methods for analyzing such multivariate series are considered in Chapter 9.

1.2 A first crash course

Let us introduce some of the most important concepts of time series analysis by considering an example where we look for simple models for predicting the monthly prices of wheat.

In the following, let P_t denote the price of wheat at time (month) t. The first naive guess would be to say that the price next month is the same as in this month. Hence, the *predictor* is

$$\widehat{P}_{t+1|t} = P_t \,. \tag{1.1}$$

This predictor is called the *naive predictor* or the *persistent predictor*. The syntax used is short for a prediction (or estimate) of P_{t+1} given the observations $P_t, P_{t-1}, \ldots$.

Next month, i.e., at time $t+1$, the actual price is P_{t+1}. This means that the *prediction error* or *innovation* may be computed as

$$\varepsilon_{t+1} = P_{t+1} - \widehat{P}_{t+1|t} \,. \tag{1.2}$$

By combining Equations (1.1) and (1.2) we obtain the *stochastic model* for the wheat price

$$P_t = P_{t-1} + \varepsilon_t \tag{1.3}$$

If $\{\varepsilon_t\}$ is a sequence of uncorrelated zero mean random variables (*white noise*), the process (1.3) is called a *random walk*. The random walk model is very often seen in finance and econometrics. For this model the optimal predictor is the naive predictor (1.1).

The random walk can be rewritten as

$$P_t = \varepsilon_t + \varepsilon_{t-1} + \cdots \tag{1.4}$$

which shows that the random walk is an integration of the noise, and that the variance of P_t is unbounded; therefore, no stationary distribution exists. This is an example of a *non-stationary process*.

However, it is obvious to try to consider the more general model

$$P_t = \varphi P_{t-1} + \varepsilon_t \tag{1.5}$$

called the $AR(1)$ *model* (the autoregressive first order model). For this process a stationary distribution exists for $|\varphi| < 1$. Notice that the random walk is obtained for $\varphi = 1$.

Another candidate for a model for wheat prices is

$$P_t = \psi P_{t-12} + \varepsilon_t \tag{1.6}$$

which assumes that the price this month is explained by the price in the same month last year. This seems to be a reasonable guess for a simple model, since it is well known that wheat price exhibits a *seasonal variation*. (The noise processes in (1.5) and (1.6) are, despite the notation used, of course, not the same).

For wheat prices it is obvious that both the actual price and the price in the same month in the previous year might be used in a description of the expected price next month. Such a model is obtained if we assume that the innovation ε_t in model (1.5) shows an annual variation, i.e., the combined model is

$$(P_t - \varphi P_{t-1}) - \psi(P_{t-12} - \varphi P_{t-13}) = \varepsilon_t\,. \tag{1.7}$$

Models such as (1.6) and (1.7) are called *seasonal models*, and they are used very often in econometrics.

Notice, that for $\psi = 0$ we obtain the AR(1) model (1.5), while for $\varphi = 0$ the most simple seasonal model in (1.6) is obtained.

By introducing the *backward shift operator* B by

$$\mathrm{B}^k P_t = P_{t-k} \tag{1.8}$$