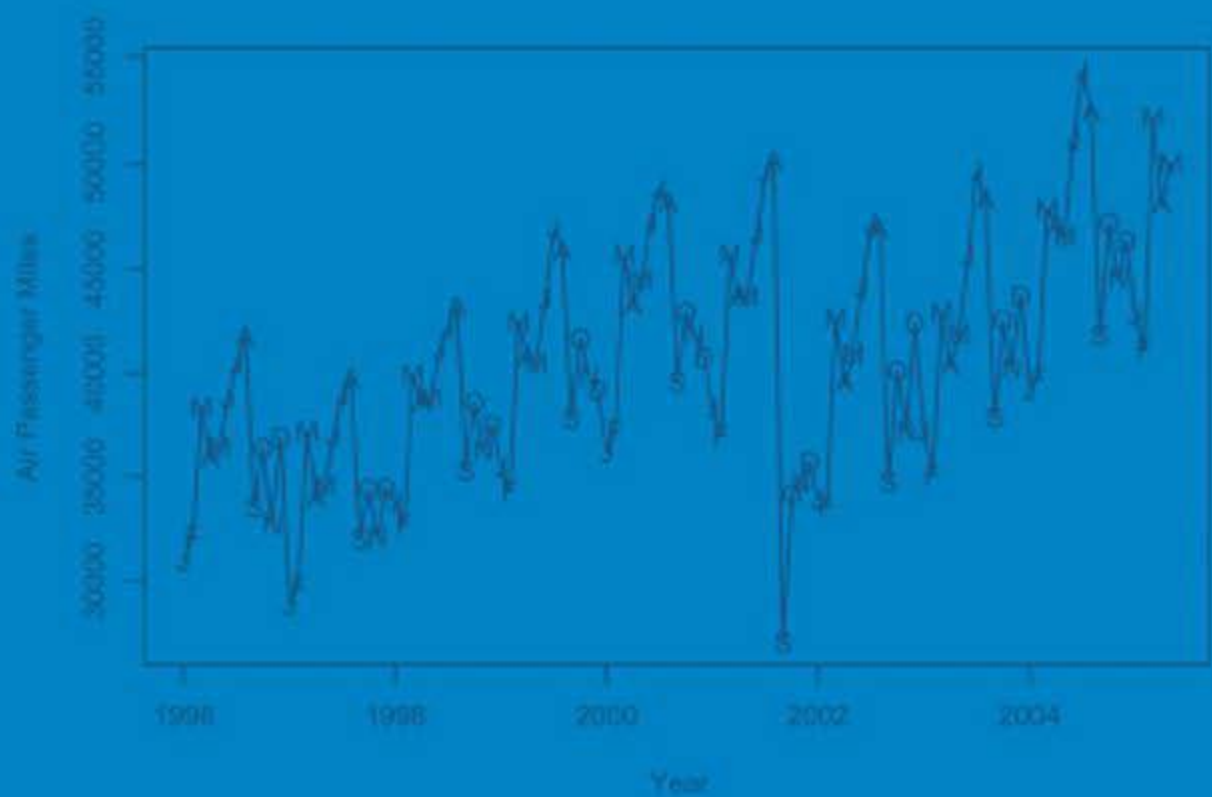


Springer Texts in Statistics

Jonathan D. Cryer  
Kung-Sik Chan

Time Series Analysis  
With Applications in R



Second Edition



Full file at

<https://buklibry.com/download/solutions-manual-time-series-analysis-with-applications-in-r-2bd-edition-by-cryer-chan/>

## **Statistics Texts in Statistics**

*Series Editors:*

G. Casella

S. Fienberg

I. Olkin

# Springer Texts in Statistics

*Athreya/Lahiri*: Measure Theory and Probability Theory  
*Bilodeau/Brenner*: Theory of Multivariate Statistics  
*Brockwell/Davis*: An Introduction to Time Series and Forecasting  
*Carmona*: Statistical Analysis of Financial Data in S-PLUS  
*Chow/Teicher*: Probability Theory: Independence, Interchangeability, Martingales, 3<sup>rd</sup> ed.  
*Christensen*: Advanced Linear Modeling: Multivariate, Time Series, and Spatial Data;  
Nonparametric Regression and Response Surface Maximization, 2<sup>nd</sup> ed.  
*Christensen*: Log-Linear Models and Logistic Regression, 2<sup>nd</sup> ed.  
*Christensen*: Plane Answers to Complex Questions: The Theory of Linear Models, 2<sup>nd</sup> ed.  
*Cryer/Chan*: Time Series Analysis, Second Edition  
*Davis*: Statistical Methods for the Analysis of Repeated Measurements  
*Dean/Voss*: Design and Analysis of Experiments  
*Dekking/Kraaikamp/Lopuhaä/Meester*: A Modern Introduction to Probability and Statistics  
*Durrett*: Essential of Stochastic Processes  
*Edwards*: Introduction to Graphical Modeling, 2<sup>nd</sup> ed.  
*Everitt*: An R and S-PLUS Companion to Multivariate Analysis  
*Gentle*: Matrix Algebra: Theory, Computations, and Applications in Statistics  
*Ghosh/Delampady/Samanta*: An Introduction to Bayesian Analysis  
*Gut*: Probability: A Graduate Course  
*Heiberger/Holland*: Statistical Analysis and Data Display; An Intermediate Course with Examples  
in S-PLUS, R, and SAS  
*Jobson*: Applied Multivariate Data Analysis, Volume I: Regression and Experimental Design  
*Jobson*: Applied Multivariate Data Analysis, Volume II: Categorical and Multivariate Methods  
*Karr*: Probability  
*Kulkarni*: Modeling, Analysis, Design, and Control of Stochastic Systems  
*Lange*: Applied Probability  
*Lange*: Optimization  
*Lehmann*: Elements of Large Sample Theory  
*Lehmann/Romano*: Testing Statistical Hypotheses, 3<sup>rd</sup> ed.  
*Lehmann/Casella*: Theory of Point Estimation, 2<sup>nd</sup> ed.  
*Longford*: Studying Human Populations: An Advanced Course in Statistics  
*Marin/Robert*: Bayesian Core: A Practical Approach to Computational Bayesian Statistics  
*Nolan/Speed*: Stat Labs: Mathematical Statistics Through Applications  
*Pitman*: Probability  
*Rawlings/Pantula/Dickey*: Applied Regression Analysis  
*Robert*: The Bayesian Choice: From Decision-Theoretic Foundations to Computational  
Implementation, 2<sup>nd</sup> ed.  
*Robert/Casella*: Monte Carlo Statistical Methods, 2<sup>nd</sup> ed.  
*Rose/Smith*: Mathematical Statistics with *Mathematica*  
*Ruppert*: Statistics and Finance: An Introduction  
*Sen/Srivastava*: Regression Analysis: Theory, Methods, and Applications.  
*Shao*: Mathematical Statistics, 2<sup>nd</sup> ed.  
*Shorack*: Probability for Statisticians  
*Shumway/Stoffer*: Time Series Analysis and Its Applications, 2<sup>nd</sup> ed.  
*Simonoff*: Analyzing Categorical Data  
*Terrell*: Mathematical Statistics: A Unified Introduction  
*Timm*: Applied Multivariate Analysis  
*Toutenberg*: Statistical Analysis of Designed Experiments, 2<sup>nd</sup> ed.  
*Wasserman*: All of Nonparametric Statistics  
*Wasserman*: All of Statistics: A Concise Course in Statistical Inference  
*Weiss*: Modeling Longitudinal Data  
*Whittle*: Probability via Expectation, 4<sup>th</sup> ed.

Jonathan D. Cryer • Kung-Sik Chan

# Time Series Analysis

With Applications in R

Second Edition



Jonathan D. Cryer  
Department of Statistics & Actuarial Science  
University of Iowa  
Iowa City, Iowa 52242  
USA  
jon-cryer@uiowa.edu

Kung-Sik Chan  
Department of Statistics & Actuarial Science  
University of Iowa  
Iowa City, Iowa 52242  
USA  
kung-sik-chan@uiowa.edu

*Series Editors:*

George Casella  
Department of Statistics  
University of Florida  
Gainesville, FL 32611-8545  
USA

Stephen Fienberg  
Department of Statistics  
Carnegie Mellon University  
Pittsburgh, PA 15213-3890  
USA

Ingram Okin  
Department of Statistics  
Stanford University  
Stanford, CA 94305  
USA

ISBN: 978-0-387-75958-6

e-ISBN: 978-0-387-75959-3

Library of Congress Control Number: 2008923058

©2008 Springer Science+Business Media, LLC

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.

The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed on acid-free paper.

9 8 7 6 5 4 3 2 1

springer.com

*To our families*

# PREFACE

The theory and practice of time series analysis have developed rapidly since the appearance in 1970 of the seminal work of George E. P. Box and Gwilym M. Jenkins, *Time Series Analysis: Forecasting and Control*, now available in its third edition (1994) with co-author Gregory C. Reinsel. Many books on time series have appeared since then, but some of them give too little practical application, while others give too little theoretical background. This book attempts to present both application, and theory at a level accessible to a wide variety of students and practitioners. Our approach is to mix application and theory throughout the book as they are naturally needed.

The book was developed for a one-semester course usually attended by students in statistics, economics, business, engineering, and quantitative social sciences. Basic applied statistics through multiple linear regression is assumed. Calculus is assumed only to the extent of minimizing sums of squares, but a calculus-based introduction to statistics is necessary for a thorough understanding of some of the theory. However, required facts concerning expectation, variance, covariance, and correlation are reviewed in appendices. Also, conditional expectation properties and minimum mean square error prediction are developed in appendices. Actual time series data drawn from various disciplines are used throughout the book to illustrate the methodology. The book contains additional topics of a more advanced nature that can be selected for inclusion in a course if the instructor so chooses.

All of the plots and numerical output displayed in the book have been produced with the R software, which is available from the R Project for Statistical Computing at [www.r-project.org](http://www.r-project.org). Some of the numerical output has been edited for additional clarity or for simplicity. R is available as free software under the terms of the Free Software Foundation's GNU General Public License in source code form. It runs on a wide variety of UNIX platforms and similar systems, Windows, and MacOS.

R is a language and environment for statistical computing and graphics, provides a wide variety of statistical (e.g., time-series analysis, linear and nonlinear modeling, classical statistical tests) and graphical techniques, and is highly extensible. The extensive appendix An Introduction to R, provides an introduction to the R software specially designed to go with this book. One of the authors (KSC) has produced a large number of new or enhanced R functions specifically tailored to the methods described in this book. They are listed on page 468 and are available in the package named TSA on the R Project's Website at [www.r-project.org](http://www.r-project.org). We have also constructed R command script files for each chapter. These are available for download at [www.stat.uiowa.edu/~kchan/TSA.htm](http://www.stat.uiowa.edu/~kchan/TSA.htm). We also show the required R code beneath nearly every table and graphical display in the book. The datasets required for the exercises are named in each exercise by an appropriate filename; for example, `larain` for the Los Angeles rainfall data. However, if you are using the TSA package, the datasets are part of the package and may be accessed through the R command `data(larain)`, for example.

All of the datasets are also available at the textbook website as ACSCII files with variable names in the first row. We believe that many of the plots and calculations

described in the book could also be obtained with other software, such as SAS<sup>®</sup>, Splus<sup>®</sup>, Statgraphics<sup>®</sup>, SCA<sup>®</sup>, EViews<sup>®</sup>, RATS<sup>®</sup>, Ox<sup>®</sup>, and others.

This book is a second edition of the book *Time Series Analysis* by Jonathan Cryer, published in 1986 by PWS-Kent Publishing (Duxbury Press). This new edition contains nearly all of the well-received original in addition to considerable new material, numerous new datasets, and new exercises. Some of the new topics that are integrated with the original include unit root tests, extended autocorrelation functions, subset ARIMA models, and bootstrapping. Completely new chapters cover the topics of time series regression models, time series models of heteroscedasticity, spectral analysis, and threshold models. Although the level of difficulty in these new chapters is somewhat higher than in the more basic material, we believe that the discussion is presented in a way that will make the material accessible and quite useful to a broad audience of users. Chapter 15, Threshold Models, is placed last since it is the only chapter that deals with nonlinear time series models. It could be covered earlier, say after Chapter 12. Also, Chapters 13 and 14 on spectral analysis could be covered after Chapter 10.

We would like to thank John Kimmel, Executive Editor, Statistics, at Springer, for his continuing interest and guidance during the long preparation of the manuscript. Professor Howell Tong of the London School of Economics, Professor Henghsiu Tsai of Academia Sinica, Taipei, Professor Noelle Samia of Northwestern University, Professor W. K. Li and Professor Kai W. Ng, both of the University of Hong Kong, and Professor Nils Christian Stenseth of the University of Oslo kindly read parts of the manuscript, and Professor Jun Yan used a preliminary version of the text for a class at the University of Iowa. Their constructive comments are greatly appreciated. We would like to thank Samuel Hao who helped with the exercise solutions and read the appendix: An Introduction to R. We would also like to thank several anonymous reviewers who read the manuscript at various stages. Their reviews led to a much improved book. Finally, one of the authors (JDC) would like to thank Dan, Marian, and Gene for providing such a great place, *Casa de Artes*, Club Santiago, Mexico, for working on the first draft of much of this new edition.

Iowa City, Iowa  
January 2008

Jonathan D. Cryer  
Kung-Sik Chan



# CONTENTS

CHAPTER 1 INTRODUCTION .....	1
1.1 Examples of Time Series .....	1
1.2 A Model-Building Strategy .....	8
1.3 Time Series Plots in History .....	8
1.4 An Overview of the Book .....	9
Exercises .....	10
CHAPTER 2 FUNDAMENTAL CONCEPTS .....	11
2.1 Time Series and Stochastic Processes .....	11
2.2 Means, Variances, and Covariances .....	11
2.3 Stationarity .....	16
2.4 Summary .....	19
Exercises .....	19
Appendix A: Expectation, Variance, Covariance, and Correlation ..	24
CHAPTER 3 TRENDS .....	27
3.1 Deterministic Versus Stochastic Trends .....	27
3.2 Estimation of a Constant Mean .....	28
3.3 Regression Methods .....	30
3.4 Reliability and Efficiency of Regression Estimates .....	36
3.5 Interpreting Regression Output .....	40
3.6 Residual Analysis .....	42
3.7 Summary .....	50
Exercises .....	50
CHAPTER 4 MODELS FOR STATIONARY TIME SERIES .....	55
4.1 General Linear Processes .....	55
4.2 Moving Average Processes .....	57
4.3 Autoregressive Processes .....	66
4.4 The Mixed Autoregressive Moving Average Model .....	77
4.5 Invertibility .....	79
4.6 Summary .....	80
Exercises .....	81
Appendix B: The Stationarity Region for an AR(2) Process .....	84
Appendix C: The Autocorrelation Function for ARMA( $p, q$ ) .....	85

## CHAPTER 5 MODELS FOR NONSTATIONARY TIME SERIES .87

5.1	Stationarity Through Differencing . . . . .	88
5.2	ARIMA Models. . . . .	92
5.3	Constant Terms in ARIMA Models. . . . .	97
5.4	Other Transformations . . . . .	98
5.5	Summary . . . . .	102
	Exercises. . . . .	103
	Appendix D: The Backshift Operator. . . . .	106

## CHAPTER 6 MODEL SPECIFICATION . . . . .109

6.1	Properties of the Sample Autocorrelation Function . . . . .	109
6.2	The Partial and Extended Autocorrelation Functions . . . . .	112
6.3	Specification of Some Simulated Time Series. . . . .	117
6.4	Nonstationarity. . . . .	125
6.5	Other Specification Methods . . . . .	130
6.6	Specification of Some Actual Time Series. . . . .	133
6.7	Summary . . . . .	141
	Exercises. . . . .	141

## CHAPTER 7 PARAMETER ESTIMATION . . . . .149

7.1	The Method of Moments . . . . .	149
7.2	Least Squares Estimation . . . . .	154
7.3	Maximum Likelihood and Unconditional Least Squares . . . . .	158
7.4	Properties of the Estimates . . . . .	160
7.5	Illustrations of Parameter Estimation . . . . .	163
7.6	Bootstrapping ARIMA Models . . . . .	167
7.7	Summary . . . . .	170
	Exercises. . . . .	170

## CHAPTER 8 MODEL DIAGNOSTICS . . . . .175

8.1	Residual Analysis . . . . .	175
8.2	Overfitting and Parameter Redundancy. . . . .	185
8.3	Summary . . . . .	188
	Exercises. . . . .	188

<b>CHAPTER 9 FORECASTING</b>	<b>191</b>
9.1 Minimum Mean Square Error Forecasting	191
9.2 Deterministic Trends	191
9.3 ARIMA Forecasting	193
9.4 Prediction Limits	203
9.5 Forecasting Illustrations	204
9.6 Updating ARIMA Forecasts	207
9.7 Forecast Weights and Exponentially Weighted Moving Averages	207
9.8 Forecasting Transformed Series	209
9.9 Summary of Forecasting with Certain ARIMA Models	211
9.10 Summary	213
Exercises	213
Appendix E: Conditional Expectation	218
Appendix F: Minimum Mean Square Error Prediction	218
Appendix G: The Truncated Linear Process	221
Appendix H: State Space Models	222
<b>CHAPTER 10 SEASONAL MODELS</b>	<b>227</b>
10.1 Seasonal ARIMA Models	228
10.2 Multiplicative Seasonal ARMA Models	230
10.3 Nonstationary Seasonal ARIMA Models	233
10.4 Model Specification, Fitting, and Checking	234
10.5 Forecasting Seasonal Models	241
10.6 Summary	246
Exercises	246
<b>CHAPTER 11 TIME SERIES REGRESSION MODELS</b>	<b>249</b>
11.1 Intervention Analysis	249
11.2 Outliers	257
11.3 Spurious Correlation	260
11.4 Prewhitening and Stochastic Regression	265
11.5 Summary	273
Exercises	274

CHAPTER 12 TIME SERIES MODELS OF HETEROSCEDASTICITY . . . . .	277
12.1 Some Common Features of Financial Time Series . . . . .	278
12.2 The ARCH(1) Model . . . . .	285
12.3 GARCH Models . . . . .	289
12.4 Maximum Likelihood Estimation . . . . .	298
12.5 Model Diagnostics . . . . .	301
12.6 Conditions for the Nonnegativity of the Conditional Variances . . . . .	307
12.7 Some Extensions of the GARCH Model . . . . .	310
12.8 Another Example: The Daily USD/HKD Exchange Rates . .	311
12.9 Summary . . . . .	315
Exercises . . . . .	316
Appendix I: Formulas for the Generalized Portmanteau Tests . .	318
CHAPTER 13 INTRODUCTION TO SPECTRAL ANALYSIS . . . . .	319
13.1 Introduction . . . . .	319
13.2 The Periodogram . . . . .	322
13.3 The Spectral Representation and Spectral Distribution . . .	327
13.4 The Spectral Density . . . . .	330
13.5 Spectral Densities for ARMA Processes . . . . .	332
13.6 Sampling Properties of the Sample Spectral Density . . . .	340
13.7 Summary . . . . .	346
Exercises . . . . .	346
Appendix J: Orthogonality of Cosine and Sine Sequences . . . .	349
CHAPTER 14 ESTIMATING THE SPECTRUM . . . . .	351
14.1 Smoothing the Spectral Density . . . . .	351
14.2 Bias and Variance . . . . .	354
14.3 Bandwidth . . . . .	355
14.4 Confidence Intervals for the Spectrum . . . . .	356
14.5 Leakage and Tapering . . . . .	358
14.6 Autoregressive Spectrum Estimation . . . . .	363
14.7 Examples with Simulated Data . . . . .	364
14.8 Examples with Actual Data . . . . .	370
14.9 Other Methods of Spectral Estimation . . . . .	376
14.10 Summary . . . . .	378
Exercises . . . . .	378
Appendix K: Tapering and the Dirichlet Kernel . . . . .	381

CHAPTER 15 THRESHOLD MODELS . . . . .	383
15.1 Graphically Exploring Nonlinearity . . . . .	384
15.2 Tests for Nonlinearity . . . . .	390
15.3 Polynomial Models Are Generally Explosive . . . . .	393
15.4 First-Order Threshold Autoregressive Models . . . . .	395
15.5 Threshold Models . . . . .	399
15.6 Testing for Threshold Nonlinearity . . . . .	400
15.7 Estimation of a TAR Model . . . . .	402
15.8 Model Diagnostics . . . . .	411
15.9 Prediction . . . . .	415
15.10 Summary . . . . .	420
Exercises . . . . .	420
Appendix L: The Generalized Portmanteau Test for TAR . . . . .	421
APPENDIX: AN INTRODUCTION TO R . . . . .	423
Introduction . . . . .	423
Chapter 1 R Commands . . . . .	429
Chapter 2 R Commands . . . . .	433
Chapter 3 R Commands . . . . .	433
Chapter 4 R Commands . . . . .	438
Chapter 5 R Commands . . . . .	439
Chapter 6 R Commands . . . . .	441
Chapter 7 R Commands . . . . .	442
Chapter 8 R Commands . . . . .	446
Chapter 9 R Commands . . . . .	447
Chapter 10 R Commands . . . . .	450
Chapter 11 R Commands . . . . .	451
Chapter 12 R Commands . . . . .	457
Chapter 13 R Commands . . . . .	460
Chapter 14 R Commands . . . . .	461
Chapter 15 R Commands . . . . .	462
New or Enhanced Functions in the TSA Library . . . . .	468
DATASET INFORMATION . . . . .	471
BIBLIOGRAPHY . . . . .	477
INDEX . . . . .	487

# CHAPTER 1

## INTRODUCTION

Data obtained from observations collected sequentially over time are extremely common. In business, we observe weekly interest rates, daily closing stock prices, monthly price indices, yearly sales figures, and so forth. In meteorology, we observe daily high and low temperatures, annual precipitation and drought indices, and hourly wind speeds. In agriculture, we record annual figures for crop and livestock production, soil erosion, and export sales. In the biological sciences, we observe the electrical activity of the heart at millisecond intervals. In ecology, we record the abundance of an animal species. The list of areas in which time series are studied is virtually endless. The purpose of time series analysis is generally twofold: to understand or model the stochastic mechanism that gives rise to an observed series and to predict or forecast the future values of a series based on the history of that series and, possibly, other related series or factors.

This chapter will introduce a variety of examples of time series from diverse areas of application. A somewhat unique feature of time series and their models is that we usually cannot assume that the observations arise independently from a common population (or from populations with different means, for example). Studying models that incorporate dependence is the key concept in time series analysis.

### 1.1 Examples of Time Series

---

In this section, we introduce a number of examples that will be pursued in later chapters.

#### **Annual Rainfall in Los Angeles**

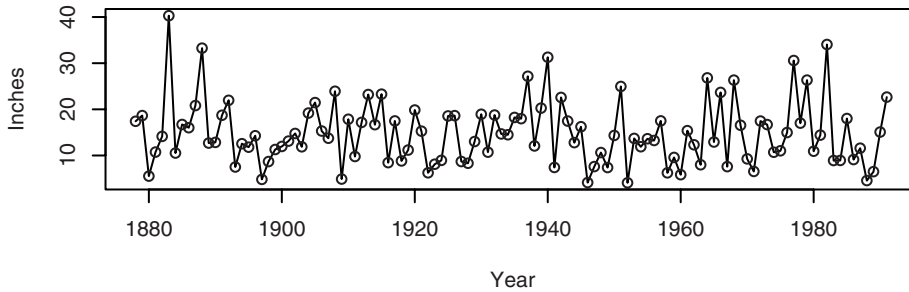
Exhibit 1.1 displays a time series plot of the annual rainfall amounts recorded in Los Angeles, California, over more than 100 years. The plot shows considerable variation in rainfall amount over the years—some years are low, some high, and many are in-between in value. The year 1883 was an exceptionally wet year for Los Angeles, while 1983 was quite dry. For analysis and modeling purposes we are interested in whether or not consecutive years are related in some way. If so, we might be able to use one year's rainfall value to help forecast next year's rainfall amount. One graphical way to investigate that question is to pair up consecutive rainfall values and plot the resulting scatterplot of pairs.

Exhibit 1.2 shows such a scatterplot for rainfall. For example, the point plotted near the lower right-hand corner shows that the year of extremely high rainfall, 40 inches in 1883, was followed by a middle of the road amount (about 12 inches) in 1884. The point

near the top of the display shows that the 40 inch year was preceded by a much more typical year of about 15 inches.

---

### Exhibit 1.1 Time Series Plot of Los Angeles Annual Rainfall



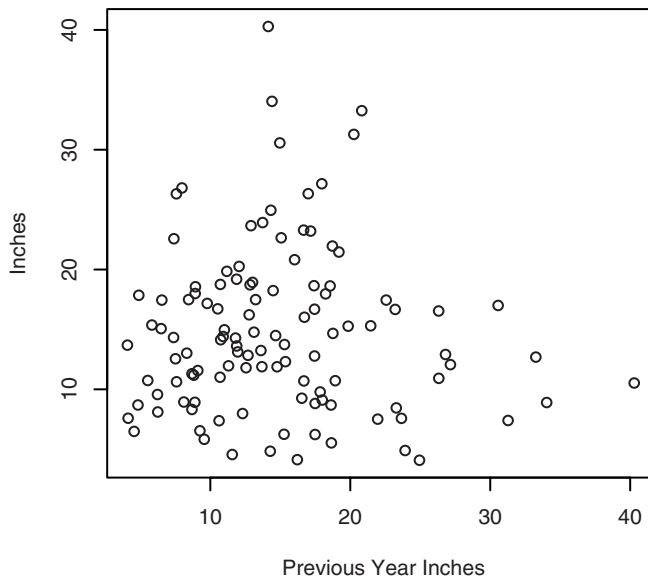
---

```
> library(TSA)
> win.graph(width=4.875, height=2.5, pointsize=8)
> data(larain); plot(larain, ylab='Inches', xlab='Year', type='o')
```

---

---

### Exhibit 1.2 Scatterplot of LA Rainfall versus Last Year's LA Rainfall

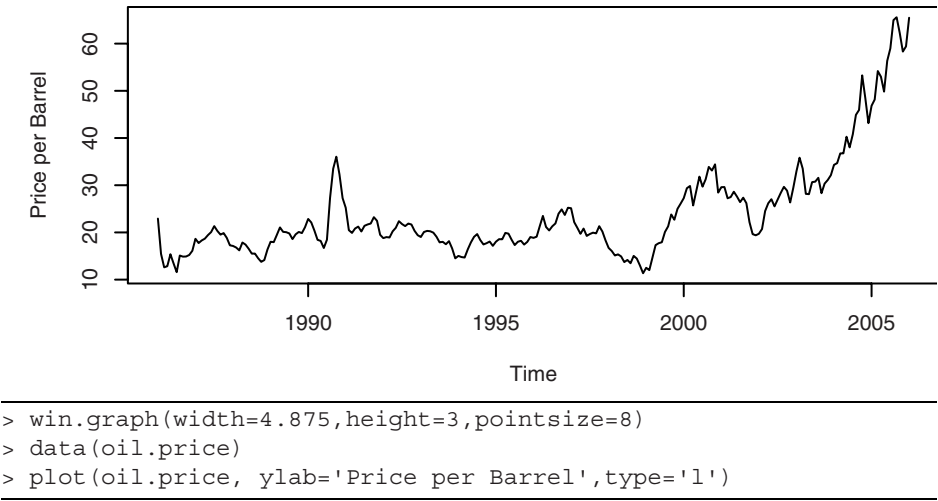


---

```
> win.graph(width=3, height=3, pointsize=8)
> plot(y=larain, x=zlag(larain), ylab='Inches',
       xlab='Previous Year Inches')
```

---

**Exhibit 5.1    Monthly Price of Oil: January 1986–January 2006**



**5.1    Stationarity Through Differencing**

Consider again the AR(1) model

$$Y_t = \phi Y_{t-1} + e_t \tag{5.1.1}$$

We have seen that assuming  $e_t$  is a true “innovation” (that is,  $e_t$  is uncorrelated with  $Y_{t-1}, Y_{t-2}, \dots$ ), we must have  $|\phi| < 1$ . What can we say about solutions to Equation (5.1.1) if  $|\phi| \geq 1$ ? Consider in particular the equation

$$Y_t = 3Y_{t-1} + e_t \tag{5.1.2}$$

Iterating into the past as we have done before yields

$$Y_t = e_t + 3e_{t-1} + 3^2e_{t-2} + \cdots + 3^{t-1}e_1 + 3^tY_0 \tag{5.1.3}$$

We see that the influence of distant past values of  $Y_t$  and  $e_t$  does not die out—indeed, the weights applied to  $Y_0$  and  $e_1$  grow exponentially large. In Exhibit 5.2, we show the values for a very short simulation of such a series. Here the white noise sequence was generated as standard normal variables and we used  $Y_0 = 0$  as an initial condition.

<b>Exhibit 5.2    Simulation of the Explosive “AR(1) Model” <math>Y_t = 3Y_{t-1} + e_t</math></b>								
<b><math>t</math></b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>
$e_t$	0.63	−1.25	1.80	1.51	1.56	0.62	0.64	−0.98
$Y_t$	0.63	0.64	3.72	12.67	39.57	119.33	358.63	1074.91



- 8.9** The data file named `robot` contains a time series obtained from an industrial robot. The robot was put through a sequence of maneuvers, and the distance from a desired ending point was recorded in inches. This was repeated 324 times to form the time series. Compare the fits of an  $AR(1)$  model and an  $IMA(1,1)$  model for these data in terms of the diagnostic tests discussed in this chapter.
- 8.10** The data file named `deere3` contains 57 consecutive values from a complex machine tool at Deere & Co. The values given are deviations from a target value in units of ten millionths of an inch. The process employs a control mechanism that resets some of the parameters of the machine tool depending on the magnitude of deviation from target of the last item produced. Diagnose the fit of an  $AR(1)$  model for these data in terms of the tests discussed in this chapter.
- 8.11** Exhibit 6.31 on page 139, suggested specifying either an  $AR(1)$  or possibly an  $AR(4)$  model for the difference of the logarithms of the oil price series. (The file-name is `oil.price`).
- (a) Estimate both of these models using maximum likelihood and compare the results using the diagnostic tests considered in this chapter.
  - (b) Exhibit 6.32 on page 140, suggested specifying an  $MA(1)$  model for the difference of the logs. Estimate this model by maximum likelihood and perform the diagnostic tests considered in this chapter.
  - (c) Which of the three models  $AR(1)$ ,  $AR(4)$ , or  $MA(1)$  would you prefer given the results of parts (a) and (b)?

It is important to remember that the local polynomial approach assumes that the true lag 1 regression function is a smooth function. If the true lag 1 regression function is discontinuous, then the local polynomial approach may yield misleading estimates. However, a sharp turn in the estimated regression function may serve as a warning that the smoothness condition may not hold for the true lag 1 regression function.

## 15.2 Tests for Nonlinearity

Several tests have been proposed for assessing the need for nonlinear modeling in time series analysis. Some of these tests, such as those studied by Keenan (1985), Tsay (1986), and Luukkonen et al. (1988), can be interpreted as Lagrange multiplier tests for specific nonlinear alternatives.

Keenan (1985) derived a test for nonlinearity analogous to Tukey's one degree of freedom for nonadditivity test (see Tukey, 1949). Keenan's test is motivated by approximating a nonlinear stationary time series by a second-order Volterra expansion (Wiener, 1958)

$$Y_t = \mu + \sum_{\mu=-\infty}^{\infty} \theta_{\mu} \varepsilon_{t-\mu} + \sum_{\nu=-\infty}^{\infty} \sum_{\mu=-\infty}^{\infty} \theta_{\mu\nu} \varepsilon_{t-\mu} \varepsilon_{t-\nu} \quad (15.2.1)$$

where  $\{\varepsilon_t, -\infty < t < \infty\}$  is a sequence of independent and identically distributed zero-mean random variables. The process  $\{Y_t\}$  is linear if the double sum on the right-hand side of (15.2.1) vanishes. Thus, we can test the linearity of the time series by testing whether or not the double sum vanishes. In practice, the infinite series expansion has to be truncated to a finite sum. Let  $Y_1, \dots, Y_n$  denote the observations. Keenan's test can be implemented as follows:

- (i) Regress  $Y_t$  on  $Y_{t-1}, \dots, Y_{t-m}$ , including an intercept term, where  $m$  is some pre-specified positive integer; calculate the fitted values  $\{\hat{Y}_t\}$  and the residuals  $\{\hat{\varepsilon}_t\}$ , for  $t = m+1, \dots, n$ ; and set  $RSS = \sum \hat{\varepsilon}_t^2$ , the residual sum of squares.
- (ii) Regress  $\hat{Y}_t^2$  on  $Y_{t-1}, \dots, Y_{t-m}$ , including an intercept term, and calculate the residuals  $\{\hat{\xi}_t\}$  for  $t = m+1, \dots, n$ .
- (iii) Regress  $\hat{\varepsilon}_t$  on the residuals  $\hat{\xi}_t$  without an intercept for  $t = m+1, \dots, n$ , and Keenan's test statistic, denoted by  $\hat{F}$ , is obtained by multiplying  $(n-2m-2)/(n-m-1)$  to the  $F$ -statistic for testing that the last regression function is identically zero. Specifically, let

$$\eta = \eta_0 \sqrt{\sum_{t=m+1}^n \hat{\xi}_t^2} \quad (15.2.2)$$

where  $\eta_0$  is the regression coefficient. Form the test statistic

$$\hat{F} = \frac{\eta^2 (n-2m-2)}{RSS - \eta^2} \quad (15.2.3)$$

time-invariant linear filter 331  
    causal 331  
trading days 278  
transfer-function model 267  
trend  
    cosine 34  
    cyclical 32  
    deterministic 27  
    linear 30  
    quadratic 30  
    seasonal 32  
    seasonal means 32  
    stochastic 27  
triangular lag window 377  
trombone 374, 381  
truncated  
    lag window 377  
    linear process 200, 221  
TSA Library 468  
tuba 381

**U**

unbiased forecast 192  
unconditional  
    least squares 160  
    sum-of-squares function 159  
unit-root test 128  
unperturbed process 250

updating equation 207

**V**

Value at Risk (VaR) 277  
variance 25  
volatility 285  
    clustering 279

**W**

weakly stationary 17  
weight function 351  
white light 332  
white noise 17, 332  
whitening 265  
window closing 355  
Wold decomposition 383  
working directory 423  
workspace 423

**Y**

$\Psi$ -coefficients  
    AR(2) 75  
Yule-Walker equations  
    AR(2) 72  
    AR( $p$ ) 76  
Yule-Walker estimates 150

```

      ma1      ma2      sma1  intercept      A0      IO-63      IO-106
0.2432  0.1729  0.0899    27.7658  37.7247  28.1777  23.2698
s.e.  0.0767  0.0698  0.0713    0.7544   5.4619   5.5982   5.5740

sigma^2 estimated as 30.67:  log likelihood = -407.06,  aic = 828.13

```

## Exercise 11.19

Let us see the plot of the log-transformed weekly unit sales of lite potato chips and the weekly average price over a period of 104 weeks.

```

> library(TSA)
> data(bluebirdlite)
> ts.bluebirdlite=ts(bluebirdlite)
> plot(ts.bluebirdlite,yax.flip=T)

```

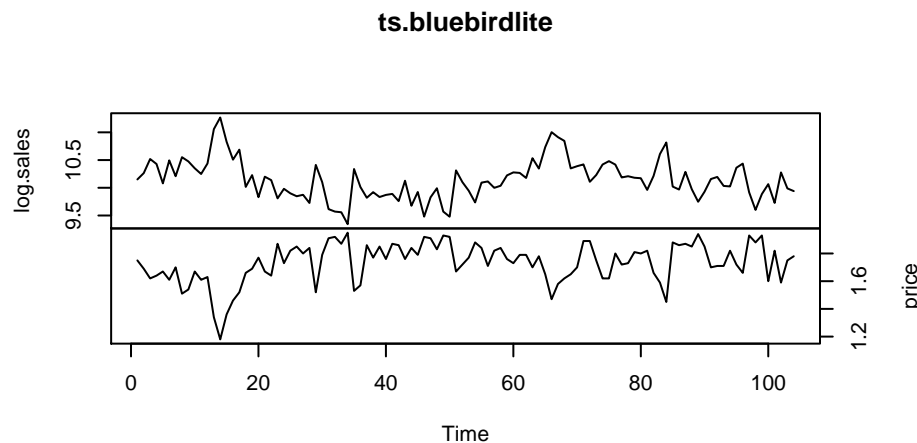


Figure 17: Weekly Log(Sales) and Price Series for Bluebirdlite Potato Chips

Next, after differencing and using prewhitened data, we draw the plot for CCF, which is significant only at lag 0, suggesting a strong contemporaneous negative relationship between lag 1 of price and sales. Higher prices are associated with lower sales.

```

> prewhiten(y=diff(ts.bluebirdlite)[,1],x=diff(ts.bluebirdlite)
+ [,2],ylab='CCF')

```

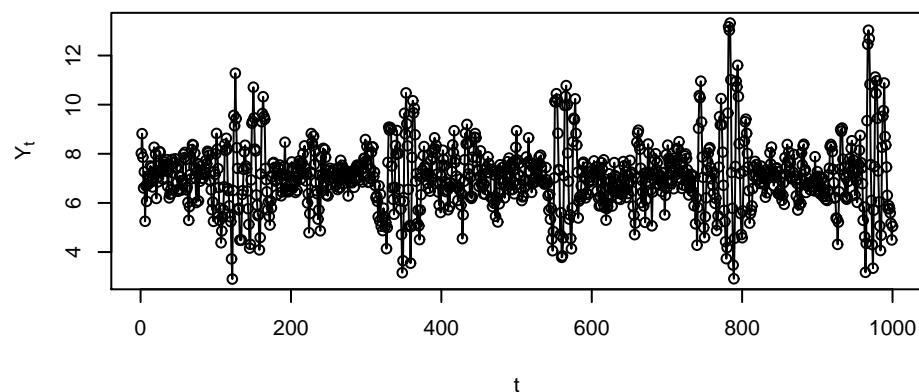


Figure 6: Simulated Series of the Fitted Model ( $n = 1000$ )

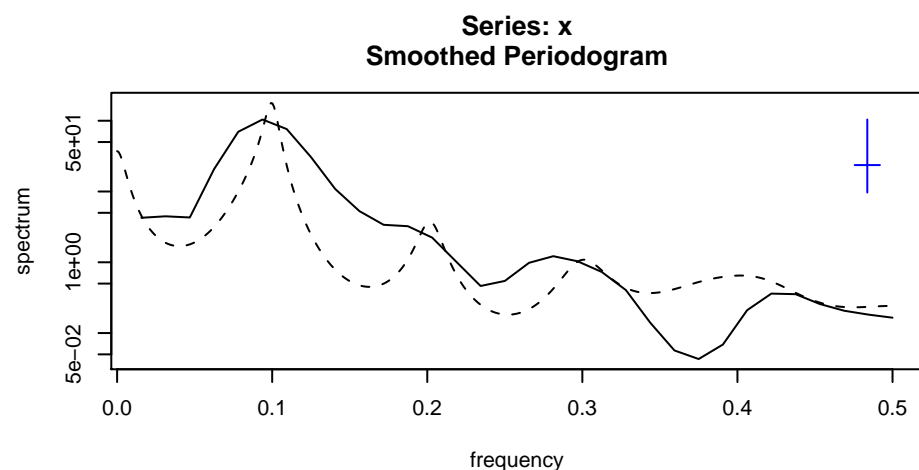


Figure 7: Spectra of Simulated Series and Sqrt Transformed Data

## Exercise 15.6

The lagged regression plots for the square-root transformed `hare` series is given below. We can see from Figure 8 that the regression function estimates appear to be strongly nonlinear for lags 2, 3 and 6, suggesting a nonlinear data mechanism.

```
> data(hare)
> win.graph(width=4.875, height=6.5, pointsize=8)
> set.seed(2534567)
> par(mfrow=c(3,2))
> lagplot(sqrt(hare))
```