APPLIED TIME SERIES ANALYSIS WITH R
Second Edition

Wayne A. Woodward
Southern Methodist University Dallas, Texas, USA

Henry L. Gray
Southern Methodist University Dallas, Texas, USA

Alan C. Elliott
Southern Methodist University Dallas, Texas, USA
Contents

Preface for Second Edition
Acknowledgments

1 Stationary Time Series
   1.1 Time Series
   1.2 Stationary Time Series
   1.3 Autocovariance and Autocorrelation Functions for Stationary Time Series
   1.4 Estimation of the Mean, Autocovariance, and Autocorrelation for Stationary Time Series
      1.4.1 Estimation of \( \mu \)
         1.4.1.1 Ergodicity of \( \bar{X} \)
         1.4.1.2 Variance of \( \bar{X} \)
      1.4.2 Estimation of \( y_k \)
      1.4.3 Estimation of \( p_k \)
   1.5 Power Spectrum
   1.6 Estimating the Power Spectrum and Spectral Density for Discrete Time Series
   1.7 Time Series Examples
      1.7.1 Simulated Data
      1.7.2 Real Data
   Appendix 1A: Fourier Series
   Appendix 1B: R Commands
   Exercises

2 Linear Filters
   2.1 Introduction to Linear Filters
      2.1.1 Relationship between the Spectra of the Input and Output of a Linear Filter
   2.2 Stationary General Linear Processes
      2.2.1 Spectrum and Spectral Density for a General Linear Process
   2.3 Wold Decomposition Theorem
   2.4 Filtering Applications
      2.4.1 Butterworth Filters
   Appendix 2A: Theorem Proofs
   Appendix 2B: R Commands
   Exercises

3 ARMA Time Series Models
   3.1 MA Processes
      3.1.1 MA(1) Model
      3.1.2 MA(2) Model
3.2 AR Processes
   3.2.1 Inverting the Operator
   3.2.2 AR(1) Model
   3.2.3 AR(p) Model for $p \geq 1$
   3.2.4 Autocorrelations of an AR(p) Model
   3.2.5 Linear Difference Equations
   3.2.6 Spectral Density of an AR(p) Model
   3.2.7 AR(2) Model
      3.2.7.1 Autocorrelations of an AR(2) Model
      3.2.7.2 Spectral Density of an AR(2)
      3.2.7.3 Stationary/Causal Region of an AR(2)
      3.2.7.4 $\psi$-Weights of an AR(2) Model
   3.2.8 Summary of AR(1) and AR(2) Behavior
   3.2.9 AR(p) Model
   3.2.10 AR(1) and AR(2) Building Blocks of an AR(p) Model
   3.2.11 Factor Tables
   3.2.12 Invertibility/Infinite-Order AR Processes
   3.2.13 Two Reasons for Imposing Invertibility

3.3 ARMA Processes
   3.3.1 Stationarity and Invertibility Conditions for an ARMA(p,q) Model
   3.3.2 Spectral Density of an ARMA(p,q) Model
   3.3.3 Factor Tables and ARMA(p,q) Models
   3.3.4 Autocorrelations of an ARMA(p,q) Model
   3.3.5 $\psi$-Weights of an ARMA(p,q)
   3.3.6 Approximating ARMA(p,q) Processes Using High-Order AR(p) Models

3.4 Visualizing AR Components

3.5 Seasonal ARMA(p,q) × ($P_S, Q_S$)S Models

3.6 Generating Realizations from ARMA(p,q) Processes
   3.6.1 MA(q) Model
   3.6.2 AR(2) Model
   3.6.3 General Procedure

3.7 Transformations
   3.7.1 Memoryless Transformations
   3.7.2 AR Transformations

Appendix 3A: Proofs of Theorems
Appendix 3B: R Commands
Exercises

4 Other Stationary Time Series Models
   4.1 Stationary Harmonic Models
      4.1.1 Pure Harmonic Models
      4.1.2 Harmonic Signal-Plus-Noise Models
      4.1.3 ARMA Approximation to the Harmonic Signal-Plus-Noise Model
   4.2 ARCH and GARCH Processes
      4.2.1 ARCH Processes
4.2.1.1 The ARCH(1) Model
4.2.1.2 The ARCH(q₀) Model
4.2.2 The GARCH(p₀, q₀) Process
4.2.3 AR Processes with ARCH or GARCH Noise

Appendix 4A: R Commands
Exercises

5 Nonstationary Time Series Models
5.1 Deterministic Signal-Plus-Noise Models
5.1.1 Trend-Component Models
5.1.2 Harmonic Component Models
5.2 ARIMA(p,d,q) and ARUMA(p,d,q) Processes
5.2.1 Extended Autocorrelations of an ARUMA(p,d,q) Process
5.2.2 Cyclical Models
5.3 Multiplicative Seasonal ARUMA (p,d,q) × (Pₛ, Dₛ, Qₛ)ₛ Process
5.3.1 Factor Tables for Seasonal Models of the Form of Equation 5.17 with s = 4 and s = 12
5.4 Random Walk Models
5.4.1 Random Walk
5.4.2 Random Walk with Drift
5.5 G-Stationary Models for Data with Time-Varying Frequencies
Appendix 5A: R Commands
Exercises

6 Forecasting
6.1 Mean-Square Prediction Background
6.2 Box–Jenkins Forecasting for ARMA(p,q) Models
6.2.1 General Linear Process Form of the Best Forecast Equation
6.3 Properties of the Best Forecast \( \hat{X}_{t₀}(ℓ) \)
6.4 π-Weight Form of the Forecast Function
6.5 Forecasting Based on the Difference Equation
6.5.1 Difference Equation Form of the Best Forecast Equation
6.5.2 Basic Difference Equation Form for Calculating Forecasts from an ARMA(p,q) Model
6.6 Eventual Forecast Function
6.7 Assessing Forecast Performance
6.7.1 Probability Limits for Forecasts
6.7.2 Forecasting the Last k Values
6.8 Forecasts Using ARUMA(p,d,q) Models
6.9 Forecasts Using Multiplicative Seasonal ARUMA Models
6.10 Forecasts Based on Signal-Plus-Noise Models
Appendix 6A: Proof of Projection Theorem
Appendix 6B: Basic Forecasting Routines
Exercises
7 Parameter Estimation
7.1 Introduction
7.2 Preliminary Estimates
   7.2.1 Preliminary Estimates for AR(p) Models
      7.2.1.1 Yule–Walker Estimates
      7.2.1.2 Least Squares Estimation
      7.2.1.3 Burg Estimates
   7.2.2 Preliminary Estimates for MA(q) Models
      7.2.2.1 MM Estimation for an MA(q)
      7.2.2.2 MA(q) Estimation Using the Innovations Algorithm
   7.2.3 Preliminary Estimates for ARMA(p,q) Models
      7.2.3.1 Extended Yule–Walker Estimates of the AR Parameters
      7.2.3.2 Tsay–Tiao Estimates of the AR Parameters
      7.2.3.3 Estimating the MA Parameters
   7.3 ML Estimation of ARMA(p,q) Parameters
   7.3.1 Conditional and Unconditional ML Estimation
   7.3.2 ML Estimation Using the Innovations Algorithm
   7.4 Backcasting and Estimating \( \sigma_a^2 \)
   7.5 Asymptotic Properties of Estimators
      7.5.1 AR Case
         7.5.1.1 Confidence Intervals: AR Case
      7.5.2 ARMA(p,q) Case
         7.5.2.1 Confidence Intervals for ARMA(p,q) Parameters
      7.5.3 Asymptotic Comparisons of Estimators for an MA(1)
   7.6 Estimation Examples Using Data
   7.7 ARMA Spectral Estimation
   7.8 ARUMA Spectral Estimation
   Appendix
   Exercises

8 Model Identification
8.1 Preliminary Check for White Noise
8.2 Model Identification for Stationary ARMA Models
   8.2.1 Model Identification Based on AIC and Related Measures
8.3 Model Identification for Nonstationary ARUMA(p,d,q) Models
   8.3.1 Including a Nonstationary Factor in the Model
   8.3.2 Identifying Nonstationary Component(s) in a Model
   8.3.3 Decision Between a Stationary or a Nonstationary Model
   8.3.4 Deriving a Final ARUMA Model
   8.3.5 More on the Identification of Nonstationary Components
      8.3.5.1 Including a Factor \((1 - B)^d\) in the Model
      8.3.5.2 Testing for a Unit Root
      8.3.5.3 Including a Seasonal Factor \((1 - B^s)\) in the Model
   Appendix 8A: Model Identification Based on Pattern Recognition
   Appendix 8B: Model Identification Functions in tswge
Exercises

9 Model Building
9.1 Residual Analysis
  9.1.1 Check Sample Autocorrelations of Residuals versus 95% Limit Lines
  9.1.2 Ljung–Box Test
  9.1.3 Other Tests for Randomness
  9.1.4 Testing Residuals for Normality
9.2 Stationarity versus Nonstationarity
9.3 Signal-Plus-Noise versus Purely Autocorrelation-Driven Models
  9.3.1 Cochrane–Orcutt and Other Methods
  9.3.2 A Bootstrapping Approach
  9.3.3 Other Methods for Trend Testing
9.4 Checking Realization Characteristics
9.5 Comprehensive Analysis of Time Series Data: A Summary
Appendix 9A: R Commands
Exercises

10 Vector-Valued (Multivariate) Time Series
10.1 Multivariate Time Series Basics
10.2 Stationary Multivariate Time Series
  10.2.1 Estimating the Mean and Covariance for Stationary Multivariate Processes
    10.2.1.1 Estimating $\mu$
    10.2.1.2 Estimating $T(k)$
10.3 Multivariate (Vector) ARMA Processes
  10.3.1 Forecasting Using VAR($p$) Models
  10.3.2 Spectrum of a VAR($p$) Model
  10.3.3 Estimating the Coefficients of a VAR($p$) Model
    10.3.3.1 Yule–Walker Estimation
    10.3.3.2 Least Squares and Conditional ML Estimation
    10.3.3.3 Burg-Type Estimation
  10.3.4 Calculating the Residuals and Estimating $\Gamma_a$
  10.3.5 VAR($p$) Spectral Density Estimation
  10.3.6 Fitting a VAR($p$) Model to Data
    10.3.6.1 Model Selection
    10.3.6.2 Estimating the Parameters
    10.3.6.3 Testing the Residuals for White Noise
10.4 Nonstationary VARMA Processes
10.5 Testing for Association between Time Series
  10.5.1 Testing for Independence of Two Stationary Time Series
  10.5.2 Testing for Cointegration between Nonstationary Time Series
10.6 State-Space Models
  10.6.1 State Equation
  10.6.2 Observation Equation
  10.6.3 Goals of State-Space Modeling
10.6.4 Kalman Filter
   10.6.4.1 Prediction (Forecasting)
   10.6.4.2 Filtering
   10.6.4.3 Smoothing Using the Kalman Filter
   10.6.4.4 h-Step Ahead Predictions

10.6.5 Kalman Filter and Missing Data

10.6.6 Parameter Estimation

10.6.7 Using State-Space Methods to Find Additive Components of a
      Univariate AR Realization
   10.6.7.1 Revised State-Space Model
   10.6.7.2 $\Psi_j$ Real
   10.6.7.3 $\Psi_j$ Complex

Appendix 10A: Derivation of State-Space Results
Appendix 10B: Basic Kalman Filtering Routines
Exercises

11 Long-Memory Processes
   11.1 Long Memory
   11.2 Fractional Difference and FARMA Processes
   11.3 Gegenbauer and GARMA Processes
      11.3.1 Gegenbauer Polynomials
      11.3.2 Gegenbauer Process
      11.3.3 GARMA Process
   11.4 k-Factor Gegenbauer and GARMA Processes
      11.4.1 Calculating Autocovariances
      11.4.2 Generating Realizations
   11.5 Parameter Estimation and Model Identification
   11.6 Forecasting Based on the k-Factor GARMA Model
   11.7 Testing for Long Memory
      11.7.1 Testing for Long Memory in the Fractional and FARMA Setting
      11.7.2 Testing for Long Memory in the Gegenbauer Setting
   11.8 Modeling Atmospheric CO$_2$ Data Using Long-Memory Models
Appendix 11A: R Commands
Exercises

12 Wavelets
   12.1 Shortcomings of Traditional Spectral Analysis for TVF Data
   12.2 Window-Based Methods that Localize the “Spectrum” in Time
      12.2.1 Gabor Spectrogram
      12.2.2 Wigner–Ville Spectrum
   12.3 Wavelet Analysis
      12.3.1 Fourier Series Background
      12.3.2 Wavelet Analysis Introduction
      12.3.3 Fundamental Wavelet Approximation Result
      12.3.4 Discrete Wavelet Transform for Data Sets of Finite Length
      12.3.5 Pyramid Algorithm
12.3.6 Multiresolution Analysis
12.3.7 Wavelet Shrinkage
12.3.8 Scalogram: Time-Scale Plot
12.3.9 Wavelet Packets
12.3.10 Two-Dimensional Wavelets

12.4 Concluding Remarks on Wavelets

Appendix 12A: Mathematical Preliminaries for This Chapter
Appendix 12B: Mathematical Preliminaries
Exercises

13 G-Stationary Processes

13.1 Generalized-Stationary Processes
   13.1.1 General Strategy for Analyzing G-Stationary Processes

13.2 M-Stationary Processes
   13.2.1 Continuous M-Stationary Process
   13.2.2 Discrete M-Stationary Process
   13.2.3 Discrete Euler(p) Model
   13.2.4 Time Transformation and Sampling

13.3 G(λ)-Stationary Processes
   13.3.1 Continuous G(p; λ) Model
   13.3.2 Sampling the Continuous G(λ)-Stationary Processes
      13.3.2.1 Equally Spaced Sampling from G(p; λ) Processes
   13.3.3 Analyzing TVF Data Using the G(p; λ) Model
      13.3.3.1 G(p; λ) Spectral Density

13.4 Linear Chirp Processes
   13.4.1 Models for Generalized Linear Chirps

13.5 G-Filtering

13.6 Concluding Remarks

Appendix 13A: G-Stationary Basics
Appendix 13B: R Commands
Exercises

References

Index
Preface for Second Edition

We continue to believe that this book is a one-of-a-kind book for teaching introductory time series. We make every effort to not only present a compendium of models and methods supplemented by a few examples along the way. Instead, we dedicate extensive coverage designed to provide insight into the models, we discuss features of realizations from various models, and we give caveats regarding the use and interpretation of results based on the models. We have used the book with good success teaching PhD students as well as professional masters’ students in our program.

Suggestions concerning the first edition were as follows: (1) to base the computing on R and (2) to include more real data examples. To address item (1) we have created an R package, tsige, which is available in CRAN to accompany this book. Extensive discussion of the use of tsige functions is given within the chapters and in appendices following each chapter. The tsige package currently has about 40 functions and that number may continue to grow. Check the book’s website, http://www.texasoft.com/ATSA/index.html, for updates. We have added guidance concerning R usage throughout the entire book. Of special note is the fact that R support is now provided for Chapters 10 through 13. In the first edition, the accompanying software package GW-WINKS contained only limited computational support related to these chapters.

Concerning item (2), the CRAN package tsige contains about 100 data files, many of them real data sets along with a large collection of data sets associated with figures and examples in the book. We have also included about 20 new examples, many of these related to the analysis of real data sets.

NOTE: Although it is no longer discussed within the text, the Windows-based software package GW-WINKS that accompanied the first edition, is still available on our website http://www.texasoft.com/ATSA/index.html along with instructions for downloading and analysis. Although we have moved to R because of user input, we continue to believe that GW-WINKS is easy to learn and use, and it provides a “learning environment” that enhances the understanding of the material. After the first edition of this book became available, a part of the first homework assignment in our time series course has been to load GW-WINKS and perform some rudimentary procedures. We are yet to have a student come to us for help getting started. It’s very easy to use.
Acknowledgments

As we have used the first edition of this book and began developing the second edition, many students in our time series courses have provided invaluable help in copy editing and making suggestions concerning the functions in the new tswge package. Of special note is the fact that Ranil Samaranatunga and Yi Zheng provided much appreciated software development support on tswge. Peter Vanev provided proofreading support of the entire first edition although we only covered Chapters 1 through 9 in the course. Other students who helped find typos in the first edition are Chelsea Allen, Priyangi Bulathsinhala, Shiran Chen, Xusheng Chen, Wejdan Deebani, Mahesh Fernando, Sha He, Shuang He, Lie Li, Sha Li, Bingchen Liu, Shuling Liu, Yuhang Liu, Wentao Lu, Jin Luo, Guo Ma, Qida Ma, Ying Meng, Amy Nussbaum, Yancheng Qian, Xiangwen Shang, Charles South, Jian Tu, Nicole Wallace, Yixun Xing, Yibin Xu, Ren Zhang, and Qi Zhou. Students in the Fall 2015 section used a beta version of the revised book and R software and were very helpful. These students include Gunes Alkan, Gong Bai, Heng Cui, Tian Hang, Tianshi He, Chuqiao Hu, Tingting Hu, Ailin Huang, Lingyu Kong, Dateng Li, Ryan McShane, Shaoling Qi, Lu Wang, Qian Wang, Benjamin Williams, Kangyi Xu, Ziyuan Xu, Yuzhi Yan, Rui Yang, Shen Yin, Yifan Zhong, and Xiaojie Zhu.
1

Stationary Time Series

In basic statistical analysis, attention is usually focused on data samples, $X_1, X_2, \ldots, X_n$, where the $X_i$s are independent and identically distributed random variables. In a typical introductory course in univariate mathematical statistics, the case in which samples are not independent but are in fact correlated is not generally covered. However, when data are sampled at neighboring points in time, it is very likely that such observations will be correlated. Such time-dependent sampling schemes are very common. Examples include the following:

- Daily Dow Jones stock market closes over a given period
- Monthly unemployment data for the United States
- Annual global temperature data for the past 100 years
- Monthly incidence rate of influenza
- Average number of sunspots observed each year since 1749
- West Texas monthly intermediate crude oil prices
- Average monthly temperatures for Pennsylvania

Note that in each of these cases, an observed data value is (probably) not independent of nearby observations. That is, the data are correlated and are therefore not appropriately analyzed using univariate statistical methods based on independence. Nevertheless, these types of data are abundant in fields such as economics, biology, medicine, and the physical and engineering sciences, where there is interest in understanding the mechanisms underlying these data, producing forecasts of future behavior, and drawing conclusions from the data. Time series analysis is the study of these types of data, and in this book we will introduce you to the extensive collection of tools and models for using the inherent correlation structure in such data sets to assist in their analysis and interpretation.

As examples, in Figure 1.1a we show monthly West Texas intermediate crude oil prices from January 2000 to October 2009, and in Figure 1.1b we show the average monthly temperatures in degrees Fahrenheit for Pennsylvania from January 1990 to December 2004. In both cases, the monthly data are certainly correlated. In the case of the oil process, it seems that prices for a given month are positively correlated with the prices for nearby (past and future) months. In the case of Pennsylvania temperatures, there is a clear 12 month (annual) pattern as would be expected because of the natural seasonal weather cycles.
FIGURE 1.1
Two time series data sets. (a) West Texas intermediate crude. (b) Pennsylvania average monthly temperatures.

Using R: The CRAN package tswge associated with this book contains over 100 data sets containing data related to the material in this book. Throughout the book, whenever a data set being discussed is included in the tswge package, the data set name will be noted. In this case the data sets associated with Figure 1a and b are wtcrude and patemp, respectively.

Time series analysis techniques are often classified into two major categories: time domain and frequency domain techniques. Time domain techniques include the analysis of the correlation structure, development of models that describe the manner in which such data evolve in time, and forecasting future behavior. Frequency domain approaches are designed to develop an understanding of time series data by examining the data from the perspective of their underlying cyclic (or frequency) content. The observation that the Pennsylvania temperature data tend to contain 12 month cycles is an example of examination of the frequency domain content of that data set. The basic frequency domain analysis tool is the power spectrum.

While frequency domain analysis is commonly used in the physical and engineering sciences, students with a statistics, mathematics, economics, or finance background may not be familiar with these methods. We do not assume a prior familiarity with frequency domain methods, and throughout the book we will introduce and discuss both time domain and frequency domain procedures for analyzing time series. In Sections 1.1 through 1.4, we discuss time domain analysis of time series data while in Sections 1.5 and 1.6 we present a basic introduction to frequency domain analysis and tools. In Section 1.7, we discuss several simulated and real-time series data sets from both time and frequency domain perspectives.

1.1 Time Series
Loosely speaking, a time series can be thought of as a collection of observations made sequentially in time. Our interest will not be in such series that are deterministic but rather in those whose values behave according to the laws of probability. In this
chapter, we will discuss the fundamentals involved in the statistical analysis of time series. To begin, we must be more careful in our definition of a time series. Actually, a time series is a special type of stochastic process.

**Definition 1.1**

A stochastic process \( \{ X(t); \ t \in T \} \) is a collection of random variables, where \( T \) is an index set for which all of the random variables, \( X(t), t \in T \), are defined on the same sample space. When \( T \) represents time, we refer to the stochastic process as a time series.

If \( T \) takes on a continuous range of values (e.g., \( T = (-\infty, \infty) \) or \( T = (0, \infty) \)), the process is said to be a continuous parameter process. If, on the other hand, \( T \) takes on a discrete set of values (e.g., \( T = \{0, 1, 2, \ldots\} \) or \( T = \{0, \pm 1, \pm 2, \ldots\} \)), the process is said to be a discrete parameter process. Actually, it is typical to refer to these as continuous and discrete processes, respectively.

We will use the subscript notation, \( X_t \), when we are dealing specifically with a discrete parameter process. However, when the process involved is either continuous parameter or of unspecified type, we will use the function notation, \( X(t) \). Also, when no confusion will arise, we often use the notation \( \{X(t)\} \) or simply \( X(t) \) to denote a time series. Similarly, we will usually shorten \( \{X_t; t = 0, \pm 1, \ldots\} \) to \( X_t, t = 0, \pm 1, \ldots \) or simply to \( X_t \).

Recall that a random variable, \( \gamma \), is a function defined on a sample space \( \Omega \) whose range is the real numbers. An observed value of the random variable \( \gamma \) is a real number \( \gamma = \gamma'(\omega) \) for some \( \omega \in \Omega \). For a time series \( \{X(t)\} \), its “value,” \( \{X(t, \omega); t \in T\} \) for some fixed \( \omega \in \Omega \), is a collection of real numbers. This leads to the following definition.

**Definition 1.2**

A realization of the time series \( \{X(t); \ t \in T\} \) is the set of real-valued outcomes, \( \{X(t, \omega); t \in T\} \) for a fixed value of \( \omega \in \Omega \).

That is, a realization of a time series is simply a set of values of \( \{X(t)\} \), that result from the occurrence of some observed event. A realization of the time series \( \{X(t); t \in T\} \) will be denoted \( \{x(t); t \in T\} \). As before, we will sometimes use the notation \( \{x(t)\} \) or simply \( x(t) \) in the continuous parameter case and \( \{x_i\} \) or \( x_i \) in the discrete parameter case when these are clear. The collection of all possible realizations is called an ensemble, and, for a given \( t \), the expectation of the random variable \( X(t) \), is called the ensemble mean and will be denoted \( E[X(t)] = \mu(t) \). The variance of \( X(t) \) is given by \( \text{Var} \{X(t)\} = E\{[X(t) - \mu(t)]^2\} \) and is often denoted by \( \sigma^2(t) \) since it also can depend on \( t \).

**EXAMPLE 1.1:** A TIME SERIES WITH TWO POSSIBLE REALIZATIONS
Consider the stochastic process $Y(t)$ for $t \in (-\infty, \infty)$ defined by $Y(t) = \sin(t + \varphi)$ where $P[\varphi = 0] = 0.5$ and $P[\varphi = \pi / 2] = 0.5$ and $P$ denotes probability. This process has only two possible realizations or sample functions, and these are shown in Figure 1.2 for $t \in [0,25]$.

The individual curves are the realizations while the collection of the two possible curves is the ensemble. For this process,

$$E[Y(t)] = 0.5\sin(t + 0) + 0.5\sin(t + \frac{\pi}{2}).$$

So, for example,

$$E[Y(0)] = 0.5\sin(0) + 0.5\sin(\frac{\pi}{2}) = 0.5,$$

and

$$E[Y(\frac{\pi}{4})] = \frac{\sqrt{2}}{2}.$$

Thus, $E[Y(t)] = \mu(t)$, that is, the expectation depends on $t$. Note that this expectation is an average “vertically” across the ensemble and not “horizontally” down the time axis. In Section 1.4, we will see how these “different ways of averaging” can be related. Of particular interest in the analysis of a time series is the covariance between $X(t_1)$ and $X(t_2)$, $t_1, t_2 \in T$. Since this is covariance within the same time series, we refer to it as the autocovariance.

**FIGURE 1.2**
The two distinct realizations for $Y(t)$ in Example 1.1.

**Definition 1.3**
If $\{X(t); t \in T\}$ is a time series, then for any $t_1, t_2 \in T$, we define

1. The autocovariance function, $\gamma(\cdot, \cdot)$, by

$$\gamma(t_1, t_2) = E\{[X(t_1) - \mu(t_1)][X(t_2) - \mu(t_2)]\}$$
2. The autocorrelation function, $\rho(\cdot)$, by

$$\rho(t_1, t_2) = \frac{\gamma(t_1, t_2)}{\sigma(t_1)\sigma(t_2)}$$

### 1.2 Stationary Time Series

In the study of a time series, it is common that only a single realization from the series is available. Analysis of a time series on the basis of only one realization is analogous to analyzing the properties of a random variable on the basis of a single observation. The concepts of stationarity and ergodicity will play an important role in enhancing our ability to analyze a time series on the basis of a single realization in an effective manner. A process is said to be stationary if it is in a state of “statistical equilibrium.” The basic behavior of such a time series does not change in time. As an example, for such a process, $\mu(t)$ would not depend on time and thus could be denoted $\mu$ for all $t$. It would seem that, since $x(t)$ for each $t \in T$ provides information about the ensemble mean, $\mu$, it may be possible to estimate $\mu$ on the basis of a single realization. An ergodic process is one for which ensemble averages such as $\mu$ can be consistently estimated from a single realization. In this section, we will present more formal definitions of stationarity, but we will delay further discussion of ergodicity until Section 1.4.

The most restrictive notion of stationarity is that of strict stationarity, which we define as follows.

**Definition 1.4**

A process $\{X(t); t \in T\}$ is said to be strictly stationary if for any $t_1, t_2, \ldots, t_k \in T$ and any $h \in T$, the joint distribution of $\{X(t_1), X(t_2), \ldots, X(t_k)\}$ is identical to that of $\{X(t_1 + h), X(t_2 + h), \ldots, X(t_k + h)\}$.

**NOTE:** We have tacitly assumed that $T$ is closed under addition, and we will continue to do so.

Strict stationarity requires, among other things, that for any $t_1, t_2 \in T$, the distributions of $X(t_1)$ and $X(t_2)$ must be the same, and further that all bivariate distributions of pairs $\{X(t), X(t + h)\}$ are the same for all $h$, etc. The requirement of strict stationarity is a severe one and is usually difficult to establish mathematically. In fact, for most applications, the distributions involved are not known. For this reason, less restrictive notions of stationarity have been developed. The most common of these is covariance stationarity.
**Definition 1.5 (Covariance Stationarity)**

The time series \( \{X(t); t \in T\} \) is said to be covariance stationary if

1. \( E[X(t)] = \mu \) (constant for all \( t \))
2. \( \text{Var} [X(t)] = \sigma^2 < \infty \) (i.e., a finite constant for all \( t \))
3. \( \gamma(t_1, t_2) \) depends only on \( t_2 - t_1 \)

Covariance stationarity is also called weak stationarity, stationarity in the wide sense, and second-order stationarity. In the remainder of this book, unless specified otherwise, the term stationarity will refer to covariance stationarity.

In time series, as in most other areas of statistics, uncorrelated data play an important role. There is no difficulty in defining such a process in the case of a discrete parameter time series. That is, the time series \( \{X_t; t = 0, \pm 1, \pm 2, \ldots\} \) is called a “purely random process” if the \( X_t \)'s are uncorrelated random variables. When considering purely random processes, we will only be interested in the case in which the \( X_t \)'s are also identically distributed. In this situation, it is more common to refer to the time series as white noise. The following definition summarizes these remarks.

**Definition 1.6 (Discrete White Noise)**

The time series is called discrete white noise if

1. The \( X_t \)'s is identically distributed
2. \( \gamma(t_1, t_2) = 0 \) when \( t_2 \neq t_1 \)
3. \( \gamma(t, t) = \sigma^2 \) where \( 0 < \sigma^2 < \infty \)

We will also find the two following definitions to be useful.

**Definition 1.7 (Gaussian Process)**

A time series is said to be Gaussian (normal) if for any positive integer \( k \) and any \( t_1, t_2, \ldots, t_k \in T \), the joint distribution of \( \{X(t_1), X(t_2), \ldots, X(t_k)\} \) is multivariate normal.

Note that for Gaussian processes, the concepts of strict stationarity and covariance stationarity are equivalent. This can be seen by noting that if the Gaussian process \( X(t) \) is covariance stationary, then for any \( t_1, t_2, \ldots, t_k \in T \) and any \( h \in T \), the multivariate normal distributions of \( \{X(t_1 + h), X(t_2 + h), \ldots, X(t_k + h)\} \) and \( \{X(t_1), X(t_2), \ldots, X(t_k)\} \) have the same means and covariance matrices and thus the same distributions.

**Definition 1.8 (Complex Time Series)**
A complex time series is a sequence of complex random variables $Z(t)$, such that

$$Z(t) = X(t) + iY(t),$$

where $X(t)$ and $Y(t)$ are real-valued random variables for each $t$.

It is easy to see that for a complex time series, $Z(t)$, the mean function, $\mu_Z(t)$, is given by

$$\mu_Z(t) = E[Z(t)]$$

$$= E[X(t)] + iE[Y(t)]$$

$$= \mu_X(t) + i\mu_Y(t).$$

A time series will be assumed to be real-valued in this book unless it is specifically indicated to be complex.

### 1.3 Autocovariance and Autocorrelation Functions for Stationary Time Series

In this section, we will examine the autocovariance and autocorrelation functions for stationary time series. If a time series is covariance stationary, then the autocovariance function $\gamma(t, t + h)$ only depends on $h$. Thus, for stationary processes, we denote this autocovariance function by $\gamma(h)$. Similarly, the autocorrelation function for a stationary process is given by $\rho(h) = \gamma(h)/\sigma^2$. Consistent with our previous notation, when dealing with a discrete parameter time series, we will use the subscript notation $\gamma_h$ and $\rho_h$. The autocovariance function of a stationary time series satisfies the following properties:

1. $\gamma(0) = \sigma^2$
2. $|\gamma(h)| \leq \gamma(0)$ for all $h$  
   The inequality in (2) can be shown by noting that for any random variables $X$ and $Y$ it follows that
   $$E[ |(X - \mu_X)(Y - \mu_Y)| ] \leq \{E[(X - \mu_X)^2]E[(Y - \mu_Y)^2]\}^{1/2},$$
   by the Cauchy–Schwarz inequality. Now letting $X = X(t)$ and $Y = X(t + h)$, we see that
   $$|\gamma(h)| \leq E |(X(t) - \mu)(X(i + h) - \mu)| \leq \sigma^2 = \gamma(0).$$
3. $\gamma(h) = \gamma(-h)$
   This result follows by noting that