Time Series lecture 1

Introduction

Sofia Olhede

EPFL

February 19, 2025

Lecture outline

1. Practical
2. Motivation
3. Basic Notation
4. Stationarity
5. Dependence

Practical

Practical information

- Problem sheets, solutions, slides and lecture notes will be uploaded on Moodle.
- There is an exam in the summer for the course, and a midterm assessed coursework.

Course topics

- ARIMA models
- Frequency domain analysis
- Forecasting
- ARCH models
- Multivariate Time Series

Useful resources

- Brockwell, P. J. and Davis, R. A. (2002). *Introduction to time series and forecasting.*
  Springer
- Shumway, R. H., Stoffer, D. S., and Stoffer, D. S. (2000). *Time series analysis and its applications*, volume 3.
  Springer
- Tsay, R. S. (2005). *Analysis of financial time series.*
  John Wiley & Sons
- Percival, D. B. and Walden, A. T. (1993). *Spectral analysis for physical applications.*
  Cambridge University Press

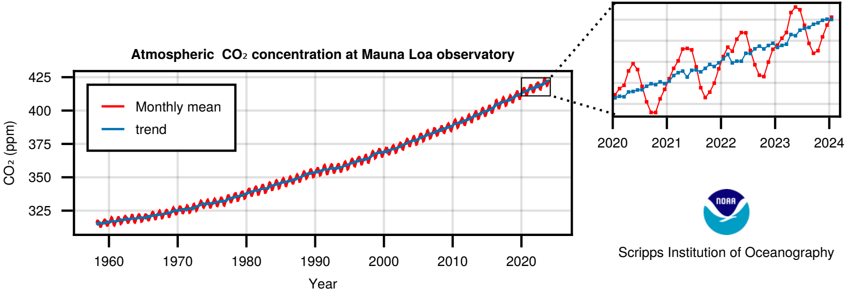
Motivation

Dependence

In basic data analysis we assume that observations are independent or even independent and identically distributed,

$$X_1, \ldots, X_n \overset{\text{ind}}{\sim} F_1, \ldots, F_n, \quad X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{N}\left(\mu, \sigma^2\right).$$

- Time series is the study of observations that arise in some order (usually time) and so are dependent.
- There are many more ways to be dependent than to be independent, and almost all data are collected in time order, so time series arise in a vast range of disciplines.
- Many of these disciplines have developed special techniques to deal with their specific problems, and we will only scratch the surface of them in this course, by surveying some main ideas.



Scripps Institution of Oceanography

Figure: Atmospheric CO2 concentration at Mauna Loa Observatory.

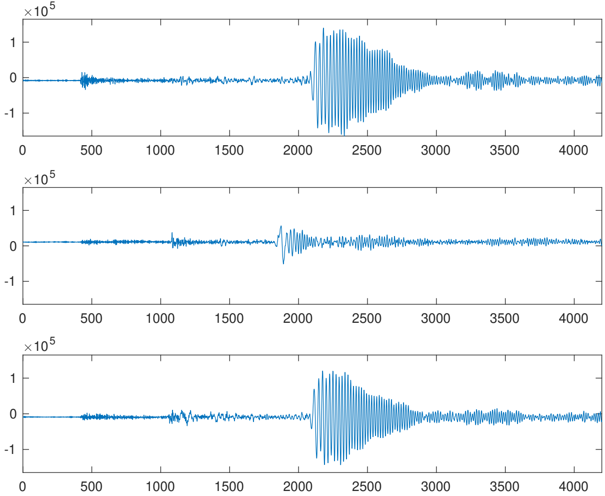
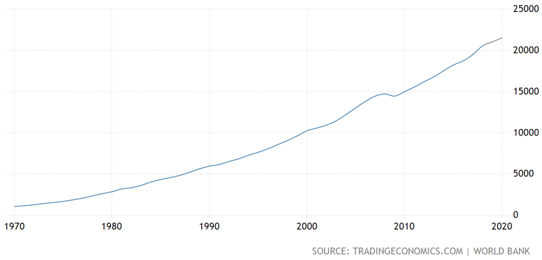


Figure: Seismic traces from the Feb. 9th, 1991 Solomon Islands earthquake as measured from the Pasadena recording station in California.



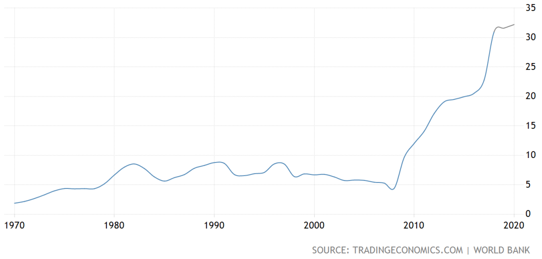
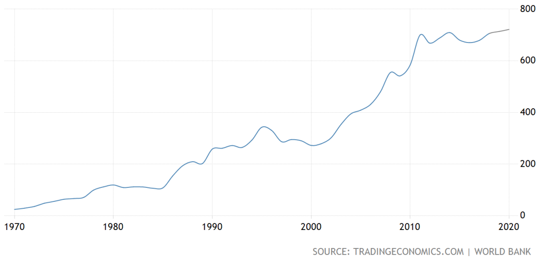


Figure: Gross domestic product (USA), (Zimbabwe).



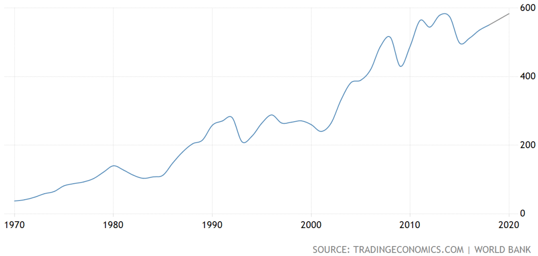


Figure: Gross domestic product (Switzerland) and (Sweden).

Example of dependent data: drifting buoys

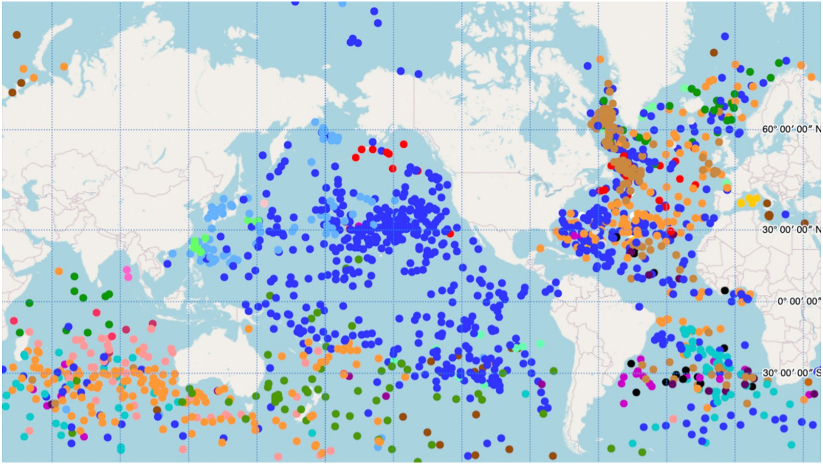


Image from the Global Drifter Programme, NOAA Atlantic Oceanographic and Meteorological Laboratory.

Drifting buoys, time evolution

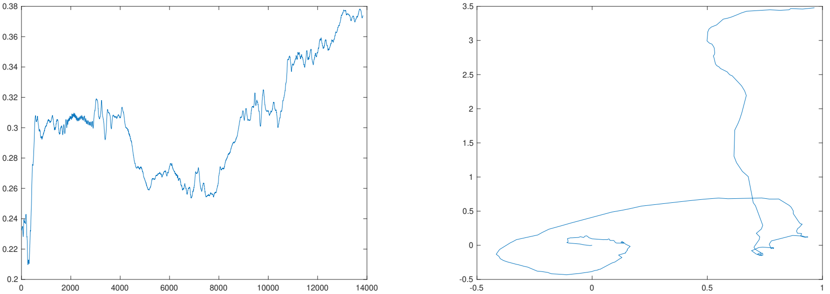
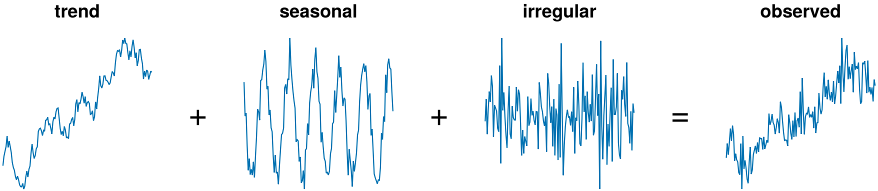


Figure: Measurements of velocity (cm/s) and position from a buoy.

Unobserved Components Models

- In econometrics for example, the notion that a time series is an aggregation of different phenomenological behaviours is common.



Basic Notation

What is a time series?

- Informally, a time series X_t is just data recorded over time.
- We shall use the word ‘time series’ to mean both the data, and the process from which the data is a realisation.
- More formally, we think of X_t as a stochastic process, i.e. as a family of random variables $\{X_t : \; t \in \mathcal{T}\}$ defined on a probability space $(\Omega, \mathscr{F}, P)$.
- In time series analysis the index (or parameter) set $\mathcal{T}$ is a set of time points, very often $\mathbb{R}$ or $\Delta\mathbb{Z}$ (or a subset of them).
- Here $\Delta \in \mathbb{R}$ is the time step between observations.

What is a time series practically?

- Whilst it is mathematically useful to think of processes with infinite index sets, in practice we can only make finitely many observations.
- Therefore, the set of observations X_t we actually record are in some set of time points $\mathcal{S} \subset \mathcal{T}$.
- Normally $\mathcal{S}$ is a discrete set (often with a regular sampling interval) $\{\Delta, \ldots, n\Delta\}$.
- The time series may also be recorded over an interval $\mathcal{S} = [0, T_0]$ (though it obviously cannot be stored digitally in this way directly).

Kolmogorov's theorem

Let $\mathscr{F} := \{\mathbf{t} = (t_1, \ldots, t_n)^T \in \mathcal{T}^n : \; t_1 < \cdots < t_n, \; n = 1, 2, \ldots\}$. Then the (finite-dimensional) distribution functions of $\{X_t\}_{t \in \mathcal{T}}$ are the functions $\{F_{\mathbf{t}}(\cdot), \; \mathbf{t} \in \mathcal{T}^n\}$ defined by

$$F_{\mathbf{t}}(\mathbf{x}) = \Pr\left(X_{t_1} \leq x_1, \ldots, X_{t_n} \leq x_n\right), \quad \mathbf{x} = (x_1, \ldots, x_n)^T \in \mathbb{R}^n.$$

Theorem 1.1 (Kolmogorov's theorem)

The probability distribution functions $\{F_{\mathbf{t}}(\cdot), \; \mathbf{t} \in \mathscr{F}\}$ are the distribution functions of a given stochastic process if and only if for any natural number n and $\mathbf{t} \in \mathscr{F}$ and $1 \leq i \leq n$ we have

$$\lim_{x_i \to \infty} F_{\mathbf{t}}(\mathbf{x}) = F_{\mathbf{t}(i)}(\mathbf{x}(i)),$$

where we have defined $\mathbf{t}(i)$ and $\mathbf{x}(i)$ as the $(n-1)$-component vectors obtained by deleting the ith component of $\mathbf{t}$ and $\mathbf{x}$ respectively.

Stationarity

Weak stationarity

Definition 1.2 ((Weak) Stationarity)

The time series $\{X_t\}$ is said to be second-order/weak or covariance stationary if for all $n \geq 1$ for any $t_1, \ldots, t_n \in \mathcal{T}$ and for all τ such that $t_1 + \tau, \ldots, t_n + \tau \in \mathcal{T}$ all the joint moments of order 1 and 2 of $X_{t_1}, \ldots, X_{t_n}$ exist, are all finite and equal to the corresponding joint moments of $X_{t_1+\tau}, \ldots, X_{t_n+\tau}$.

In fact this corresponds to that, for all $t, s, \tau \in \mathcal{T}$

1. $\mathbb{E}[X_t] = \mu$,
2. $\text{Var}(X_t) = \sigma^2 < \infty$,
3. $\mathbb{E}[X_t X_{t+\tau}] = \mathbb{E}[X_s X_{s+\tau}]$.

One may deduce from this that $\mathbb{E}[X_t X_{t+\tau}]$ can be written as a function of τ only.

Strong stationarity

We can go beyond the first two moments and define strong stationarity.

Definition 1.3 (Strong Stationarity)

The time series $\{X_t\}$ is said to be completely/strong or strictly stationary if for all $n \geq 1$ for any $t_1, \ldots, t_n \in \mathcal{T}$ and for all τ such that $t_1 + \tau, \ldots, t_n + \tau \in \mathcal{T}$ the joint distribution of $X_{t_1}, \ldots, X_{t_n}$ is the same as $X_{t_1+\tau}, \ldots, X_{t_n+\tau}$.

In general:

- second order stationary $\not\Rightarrow$ strictly stationary,
- strict stationarity $\not\Rightarrow$ 2nd order stationarity.

For example iid student t with non-finite variance.

Gaussian processes

Definition 1.4 (Gaussian processes)

A stochastic process is called a Gaussian process if, for all $n \geq 1$ and for any $t_1, \ldots t_n \in \mathcal{T}$, the joint distribution of $X_{t_1}, \ldots, X_{t_n}$ is multivariate normal.

- In other words, the distribution of the process at any finite collection of time points is multivariate normal.
- Gaussian processes are then completely characterised by their first two moments (which are finite).
- Thus for Gaussian processes, strong stationarity and weak stationarity are the same!

Dependence

Measures of dependence

- As we discussed earlier, often in statistics one works with collections of independent observations.
- Here, we do not have independent data.
- To understand dependence better for finite collections of random variables, we often compute their covariance matrix.
- For a time series $\{X_t\}$ the extension of the covariance matrix corresponds to the autocovariance function.
- We usually only have one time series, but if the process is stationary we might hope to still be able to do some kind of averaging.
- For notational simplicity, assume that $\Delta = 1$ for the remainder of this lecture.

Autocovariance Sequence (ACVS)

Definition 1.5 (ACVS)

For a discrete time second-order stationary process $\{X_t\}$ we define the autocovariance sequence (ACVS) by

$$\gamma_\tau = \operatorname{Cov}(X_0, X_\tau), \tag{1.1}$$

where $\tau \in \mathbb{Z}$ is the lag.

- By stationarity, we have $\gamma_\tau = \operatorname{Cov}(X_t, X_{t+\tau})$ for any $t \in \mathbb{Z}$.
- Clearly $\gamma_0 = \operatorname{Var}(X_t)$ and $\gamma_{-\tau} = \gamma_\tau$.
- But how do we estimate this from a single realisation?

Estimating the ACVS: moment matching

- A natural estimator for the acvs is based on matching moments, i.e.

$$\tilde{\gamma}_\tau = \begin{cases} \frac{1}{n-|\tau|} \sum_{t=1}^{n-|\tau|} (X_t - \bar{X})(X_{t+|\tau|} - \bar{X}) & \text{if } |\tau| \leq n-1, \\ 0 & \text{otherwise.} \end{cases} \tag{1.2}$$

- If we knew the true mean μ, and replace the sample mean by this we obtain for $|\tau| \leq n-1$

$$\mathbb{E}\left[\tilde{\gamma}_\tau\right] = \frac{1}{n-|\tau|} \sum_{t=1}^{n-|\tau|} \mathbb{E}\left[(X_t - \mu)(X_{t+|\tau|} - \mu)\right] = \gamma_\tau. \tag{1.3}$$

- However, $\tilde{\gamma}_\tau$ is not always non-negative definite, motivating a slightly different estimator.

Estimating the ACVS: a non-negative definite estimate

Definition 1.6 (The sample autocovariance)

Given observations of a time series $X_1, \ldots, X_n$. We define the sample autocovariance to be

$$\hat{\gamma}_\tau = \begin{cases} \frac{1}{n}\sum_{t=1}^{n-|\tau|}(X_t - \bar{X})(X_{t+|\tau|} - \bar{X}) & \text{if } |\tau| \leq n-1, \\ 0 & \text{otherwise,} \end{cases} \tag{1.4}$$

where $\bar{X}$ is the sample mean.

- Again in the known mean case

$$\mathbb{E}\left[\hat{\gamma}_\tau\right] = \frac{n-|\tau|}{n}\gamma_\tau, \tag{1.5}$$

for $|\tau| \leq n-1$, which is biased.

- But, $\hat{\gamma}_\tau$ is non-negative definite.

Autocorrelation sequence (ACF)

Definition 1.7 (ACF)

The autocorrelation sequence, usually called autocorrelation function, (ACF) is defined as

$$\rho_\tau = \text{Corr}\left(X_t, X_{t+\tau}\right), \tag{1.6}$$

for $\tau \in \mathbb{Z}$.

- We have $\rho_\tau = \gamma_\tau / \gamma_0$.
- Even though it is a sequence, it is still usually referred to as the autocorrelation function.
- Estimation uses the plug in estimator

$$\hat{\rho}_\tau = \hat{\gamma}_\tau / \hat{\gamma}_0. \tag{1.7}$$

Example time series vs sample ACF

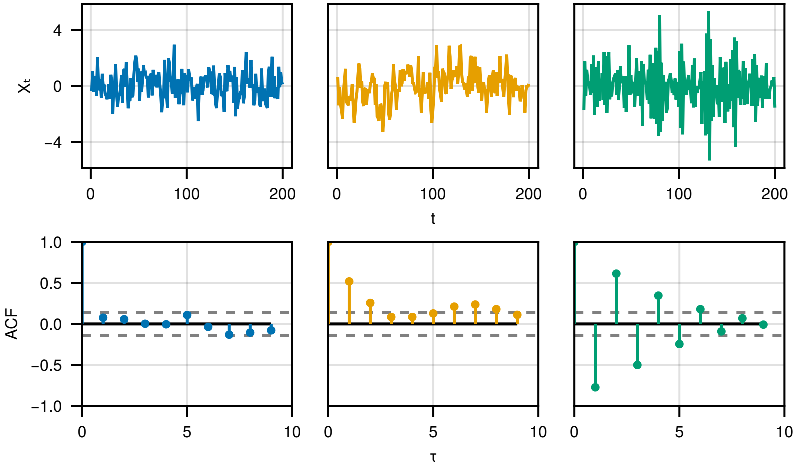


Figure: Simulated time series and their sample ACF.

ACVS is positive semi-definite

Theorem 1.8

The sequence $\{\gamma_\tau\}$ *is positive semi-definite, that is for all* $n \geq 1$ *for any* $t_1, \ldots, t_n \in \mathbb{Z}$ *and for any* $a_1, \ldots, a_n \in \mathbb{R}$ *we have*

$$\sum_{j=1}^{n} \sum_{k=1}^{n} a_j a_k \gamma_{j-k} \geq 0. \tag{1.8}$$

Proof.

Consider the random variable $W = \sum_{j=1}^{n} a_j X_j$. Now we have

$$0 \leq \operatorname{Var}(W) = \sum_{j=1}^{n} \sum_{k=1}^{n} a_j a_k \operatorname{Cov}(X_j, X_k) = \sum_{j=1}^{n} \sum_{k=1}^{n} a_j a_k \gamma_{j-k}$$

as required. □

The impact of dependence on estimation

- In estimation we will only have a single realisation available.
- We will use a time-average to give time replication, as we saw in the estimation of the autocovariance.
- However, we have yet to examine the effect of the dependence on our estimates.
- Assume that the autocovariance satisfies $\sum_{\tau=-\infty}^{\infty} |\gamma_\tau| < \infty$.
- Define

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

- What are the properties of this estimator?

$$\mathbb{E}\left[\bar{X}\right] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[X_i\right] = \mu.$$

- So $\bar{X}$ is <u>unbiased</u> as an estimator of μ.

- What about the variance? We say that $\bar{X}$ will converge to μ in mean square if

$$\lim_{n\to\infty} \mathrm{Var}\left(\bar{X}\right) = 0.$$

- How do we figure this out?
- We calculate

$$\begin{aligned}
\mathrm{Var}\left(\bar{X}\right) &= \mathbb{E}\left[(\bar{X}-\mu)^2\right] \\
&= \mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right)^2\right] \\
&= \frac{1}{n^2}\sum_{i,j=1}^{n} \mathbb{E}\left[(X_i-\mu)(X_j-\mu)\right].
\end{aligned}$$

How can we simply this?

- We need to acknowledge the correlation in the process. If the covariance was σ^2 everywhere then we could not have mean square convergence.
- We find that

$$\begin{aligned}\mathrm{Var}\left(\bar{X}\right) &= \frac{1}{n^2}\sum_{i,j=1}^{n}\mathbb{E}\left[(X_i-\mu)(X_j-\mu)\right] \\ &= \frac{1}{n^2}\sum_{i,j=1}^{n}\gamma_{j-i} \\ &= \frac{1}{n^2}\sum_{\tau=-(n-1)}^{n-1}(n-|\tau|)\gamma_\tau\end{aligned}$$

- We now need the Césaro summability theorem which says that if $\sum_{\tau=-(n-1)}^{n-1}\gamma_\tau$ converges to a limit then $\sum_{\tau=-(n-1)}^{n-1}\frac{(n-|\tau|)}{n}\gamma_\tau$ converges to the same limit.

- Thus

$$\begin{aligned}
\lim_{n\to\infty} n \operatorname{Var}\left(\bar{X}\right) &= \lim_{n\to\infty} \sum_{\tau=-(n-1)}^{n-1} \frac{(n-|\tau|)}{n}\gamma_\tau \\
&= \lim_{n\to\infty} \sum_{\tau=-(n-1)}^{n-1} \gamma_\tau \\
&= \sum_{\tau=-\infty}^{\infty} \gamma_\tau \\
&= C^{(\gamma)}
\end{aligned}$$

say.

- Now we know that $C^{(\gamma)} < \infty$ by assumption, therefore

$$\lim_{n\to\infty} \operatorname{Var}\left(\bar{X}\right) = 0.$$

- We just showed that the sample mean was consistent, e.g. $\bar{X} \xrightarrow{P} \mu$, if the autocovariance satisfies $\sum_\tau |\gamma_\tau| < \infty$.
- Seems like a general idea: "when can we replace a sample average by a population average"? But what about the correlation? Does it matter? Does it change things?
- For example, consider the AR(1) process:

$$X_t = \phi X_{t-1} + \varepsilon_t,$$

for $X_0 \sim \mathcal{N}\left(0, \frac{1}{1-\phi^2}\right)$. Can we always average this? Do other statistical operations? What happens as ϕ changes?
- (Note we will see more on the AR process next week!)

Effect of correlation on mean estimates

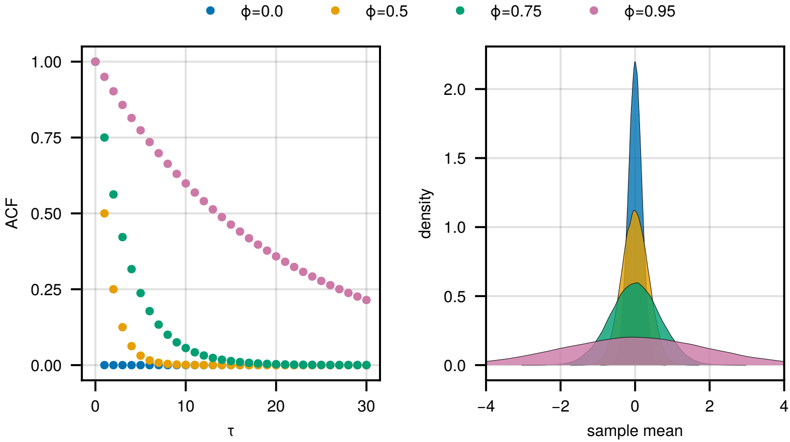


Figure: Density estimates of the distribution of the sample mean for different AR(1) processes. Time series of length 30.

Definition 1.9 ((Mean) Ergodic)

The time series $\{X_t\}$ is said to be mean ergodic if its first and second moments are finite and

$$\lim_{n\to\infty} \bar{X} \xrightarrow{P} \mathbb{E}\left[X_t\right].$$

- The funny squiggle $\xrightarrow{P}$ means "converges in probability" and informally it implies that the mean stabilises as the expectation tends to a constant and the variance goes to zero.
- The concept can be generalised to the d^{th} moment for $d \geq 2$, not just for the mean.
- The informal understanding of this is "sample means converge to population means", or "temporal averages are equivalent to population averages".
- Ergodicity and stationarity are not equivalent. The former concept is popular in econometrics.

Bibliography

Brockwell, P. J. and Davis, R. A. (2002). *Introduction to time series and forecasting*. Springer.

Percival, D. B. and Walden, A. T. (1993). *Spectral analysis for physical applications*. Cambridge University Press.

Shumway, R. H., Stoffer, D. S., and Stoffer, D. S. (2000). *Time series analysis and its applications*, volume 3. Springer.

Tsay, R. S. (2005). *Analysis of financial time series*. John Wiley & Sons.