Introduction to Time Series

Dr. Bo Li

January 8, 2013
Introduction

Examples of time series
A time series problem
Terminology
Objectives of Time Series Analysis

Get start with R
A **time series** is a collection of observations $x_t$ made sequentially through time. Examples occur in a variety of fields, ranging from economics to engineering.

Examples of time series:

- Monthly sales of U.S. houses (thousands) 1965 - 1975
- The Beveridge wheat price annual index series from 1500 to 1869
  - Average wheat price in nearly 50 places in various countries
  - Particular interest to economic historians.
- Time series related to temperature
  - Northern hemisphere mean temperature
  - 14 Tree ring series
  - Important to past temperature reconstruction
Average wheat price (1500 - 1869)
Past temperature reconstruction
Instrumental temperature series
14 tree ring series
Long Climate Proxies and NH temperatures

Exploit where there is overlap in two sets of data
Reconstructed temperatures

![Graph showing reconstructed temperatures over time.](image-url)
Continuous: A time series is continuous when observations are made continuously through time, even when the measured variable can only take a discrete set of values. E.g., a binary process at continuous time is a continuous time series.

Discrete: A time series is discrete when observations are taken only at specific times, usually equally spaced, even the measured variable is a continuous variable.

We will more focus on discrete time series.
Terminology

- Discrete time series can arise in several ways:
  - Sampled: Given a continuous time series, we could read off the values at equal intervals of time to give a discrete time series, sometimes called a sampled series. The sampling interval between successive readings must be carefully chosen so as to lose little information.
  - Aggregated: Aggregate the values over equal intervals of a continuous time series. E.g., monthly exports and daily rainfalls.

- What are the main difference between a time series and random samples of independent observations?
Answer: The special feature of time-series analysis is the fact that successive observations are usually not independent and that the analysis must taken into account the *time order* of the observations. When successive observations are dependent, future values may be predicted from past observations.

- **Deterministic:** If a time series can be predicted exactly, it is said to be deterministic, e.g., $x_t = 3x_{t-1}$

- **Stochastic:** Most time series are stochastic in that the future is only partly determined by past values, so that exact predictions are impossible and must be replaced by the idea that future values have a probability distribution, which is conditioned by a knowledge of past values, e.g.,
  \[ X_t = 3X_{t-1} + \epsilon, \quad \epsilon \sim N(0, \sigma^2). \]
Objectives

▶ Description
  ▶ Time plot: plot the time series against time, and then to obtain simple descriptive measures of the main properties of the series. Cyclic pattern? Seasonal effect? Upward trend? Outlier (“wild observations”)? Anyone who tries to analyze a time series without plotting it first is asking for trouble!

▶ Explanation: when observations are taken on two or more variables, it may be possible to use the variation in one time series to explain the variation in another series. e.g., it is interesting to see how sea level is affected by temperature and pressure, and to see how sales are affected by price and economic conditions.
Objectives

- **Prediction**: Given an observed time series, one may want to predict the future values of the series. This is an important task in sales forecasting, and in the analysis of economic and industrial time series. Here we also call it “forecasting”.

- **Control**: Time series are sometimes collected or analyzed so as to improve control over some physical or economic system. For example, when a time series is generated that measures the “quality” of a manufacturing process, the aim of the analysis may be to keep the process operating at a “high” level. Control problems are closely related to predictions in many situations. For example, if one can predict that a manufacturing process is going to move off target, then appropriate corrective action can be taken.

To some people R is just the 18th letter of the alphabet. To others, its the rating on racy movies, a measure of an attic's insulation or what pirates in movies say.

R is also the name of a popular programming language used by a growing number of data analysts inside corporations and academia. It is becoming their lingua franca partly because data mining has entered a golden age, whether being used to set ad prices, find new drugs more quickly or fine-tune financial models. Companies as diverse as Google, Pfizer, Merck, Bank of America, the InterContinental Hotels Group and Shell use it.
continued:

But R has also quickly found a following because statisticians, engineers and scientists without computer programming skills find it easy to use.

“R is really important to the point that its hard to overvalue it, said Daryl Pregibon, a research scientist at Google, which uses the software widely. It allows statisticians to do very intricate and complicated analyses without knowing the blood and guts of computing systems.”
R code to make the wheat time plot:

```r
file <- read.table('wheat.txt', header=F, sep='')
file <- as.matrix(file)
wheat <- as.vector(t(file))
plot(1500:1869, wheat, xlab='year', ylab='price index', type='l', lwd=2.5, col='blue')
```
More R code examples:

```r
x <- c(17, 19, 20, 15, 13, 14, 14, 14, 14)
plot(x)
plot(1500:1509, x)
points(1500:1509, x)
mean(x)
var(x)
```