Unit 1: Introduction to Time Series

Taylor R. Brown PhD

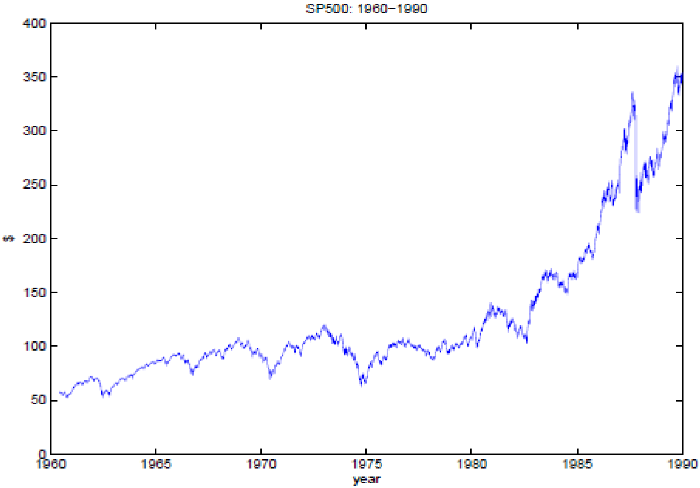
Department of Statistics, University of Virginia

Fall 2020

Textbook

This unit (tentatively) follows section 1.1 of *Time Series Analysis and Its Applications.*

Example: S&P 500 stock



Example: S&P 500 stock

Let X_t denote the closing price of S&P 500 stocks at the end of each month. We want to forecast the closing price of the stock for future observations.

Question: Should we regress X_t against the time t for our forecasts?

Terminology

- A **stochastic process** is a collection of random variables X_t indexed by a set T, i.e. $t \in T$.
- If T consists of real numbers (or a subset), the process is called a **continuous time** stochastic process.
- If T is restricted to integers (or a subset), the process is called a **discrete time** stochastic process.
- These processes may take on values which are real or restricted to integers and are called **continuous state space** or **discrete state space** respectively.

Question: In the S&P stock example, X_t is a **discrete** time, **continuous** state space stochastic process.

Terminology

- Time series analysis is generally restricted to discrete time, continuous state space stochastic processes.
- Continuous time, continuous state space stochastic processes are generally covered in a class for stochastic processes.

Introduction to Time Series

Time Series are data collected in a sequence. They are usually evenly spaced and because of the sequential nature are statistically **dependent** observations.

Introduction to Time Series

In a regression setting, an assumption for data is that the observations are **independent and identically distributed (iid),** i.e. the outcome at one point in the sequence does not effect the outcome at another point in the sequence. The nature of our observations then will be a sequence of (typically real) numbers $x_1, \cdots, x_n$ where the index represents some type of ordering. Unlike iid sequences, order matters in time series.

Introduction to Time Series

Question: What are some consequences of using regression analysis on time series data?

Features

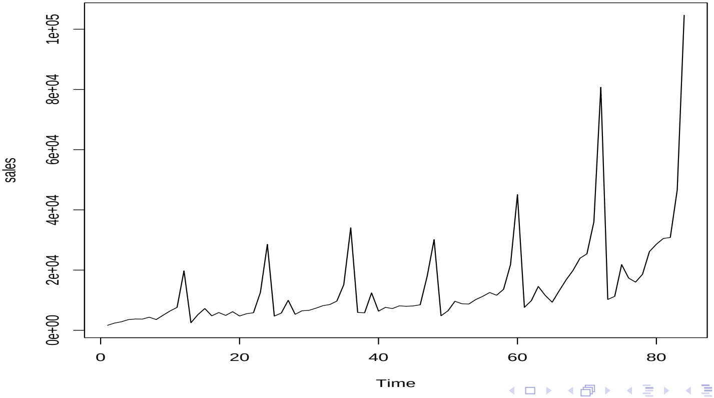
Some of the features of time series data we look out for are:

- Trend.
- Periodicity / Seasonality.
- Is the mean changing over time?
- Is the variation changing over time?
- Are there abrupt changes?
- Are there outliers?

Time series plots will be an important tool.

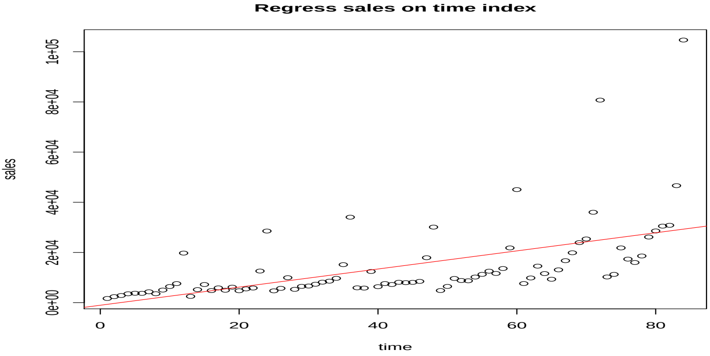
Example: Monthly Sales at Souvenir Shop

Figure: Monthly sales for a souvenir shop at a beach resort town in Queensland, Australia, Jan 1987-Dec 1993.



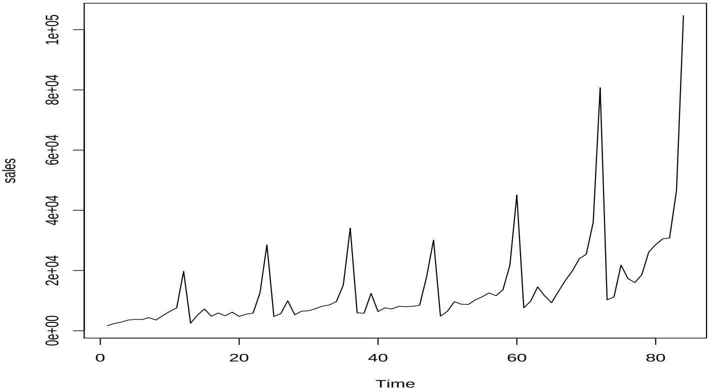
Example: Monthly Sales at Souvenir Shop

Regress sales on time index.



Appears to be increasing trend.

Example: Monthly Sales at Souvenir Shop



Notice peaks at every 12 time period interval. Suggests presence of seasonality.

Two Approaches to Time Series

There are two primary approaches to time series. One is the **time domain** approach. This approach focuses on the rules for a time series to move forward. For example, how do yesterday's and today's observations affect tomorrow's observation?

Two Approaches to Time Series

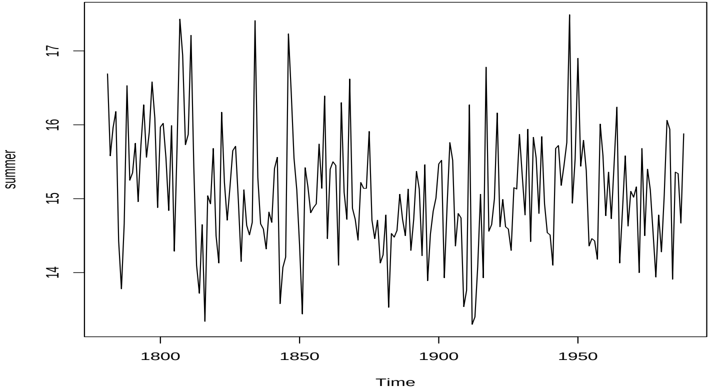
The other approach is the **frequency domain** approach. This approach tries to understand how differing oscillations can contribute to current observations. For example, taking hourly temperatures in Charlottesville, VA. There will be a very clear 24 hour oscillation. There will be another clear 8,760 hour oscillation. The current temperature is a sum of these two sinusoids (plus a lot of noise and fluctuation).

Summer Temperature in Munich

The next example is a series that involves the average summer temperature for each year in Munich from 1781 to 1988.

Summer Temperature in Munich

Figure: Average Summer Temperature, Munich 1781-1988.



Summer Temperature in Munich

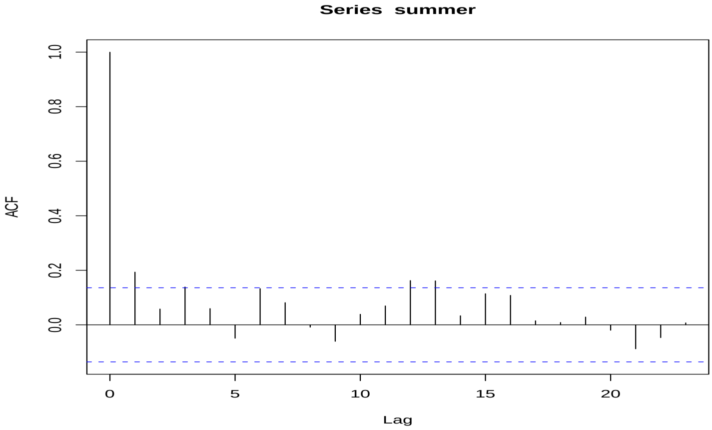
The level of variance seems to be changing. There seems to be a very mild sinusoidal trend. It is usually not trivial to detect long range weather/climate trends at one location.

Summer Temperature in Munich

One way to assess independence is through **correlation**. Independent random variables (RVs) are uncorrelated, but uncorrelated RVs may or may not be independent. Mostly we'll be dealing with normal RVs in which case we won't have to worry about this. An autocorrelation plot is a tool which can aid us in assessing correlation

Summer Temperature in Munich

Figure: ACF for summer data.

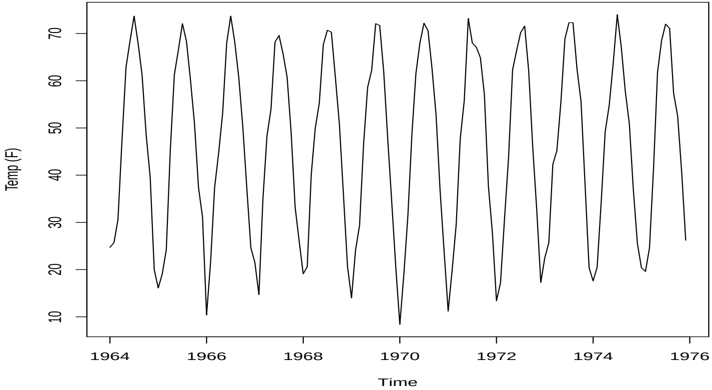


Average Monthly Temperature in Dubuque, IA

Another temperature time series is the average monthly temperature in Dubuque, IA. This is highly **periodic**. Later we'll talk about taking seasonal effects into account. We'll also discuss decomposing variation into Fourier components.

Average Monthly Temperature in Dubuque, IA

Figure: Average monthly temperature: Dubuque, IA.



Marriages in the Church of England

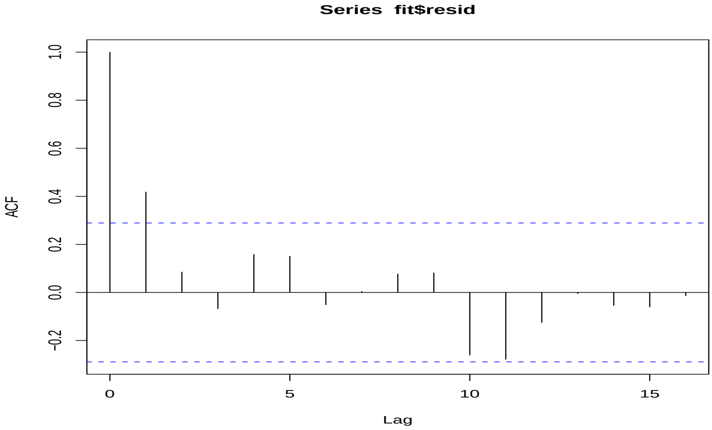
There was a data set made famous by Yule concerning the number of marriages in the Church of England as a percentage of all marriages. It appears as a somewhat negative linear trend. If we perform a linear regression, then the residuals are correlated. We may want to tackle trends in a slightly different way to deal with the dependency.

Marriages in the Church of England

Figure: Percent of Marriages in Church of England.

Marriages in the Church of England

Figure: ACF of residuals for marriage data.



IBM Stock Prices

In another data set with a clear trend, we look at the closing stock price for IBM stock from Jan 1980 to Oct 1992. A common transformation for stock price is to look at returns $diff(log(x_n))$. The resulting series looks fairly **stationary** except for a huge shock. Different transformations get us data that's easier to analyze.

IBM Stock Prices