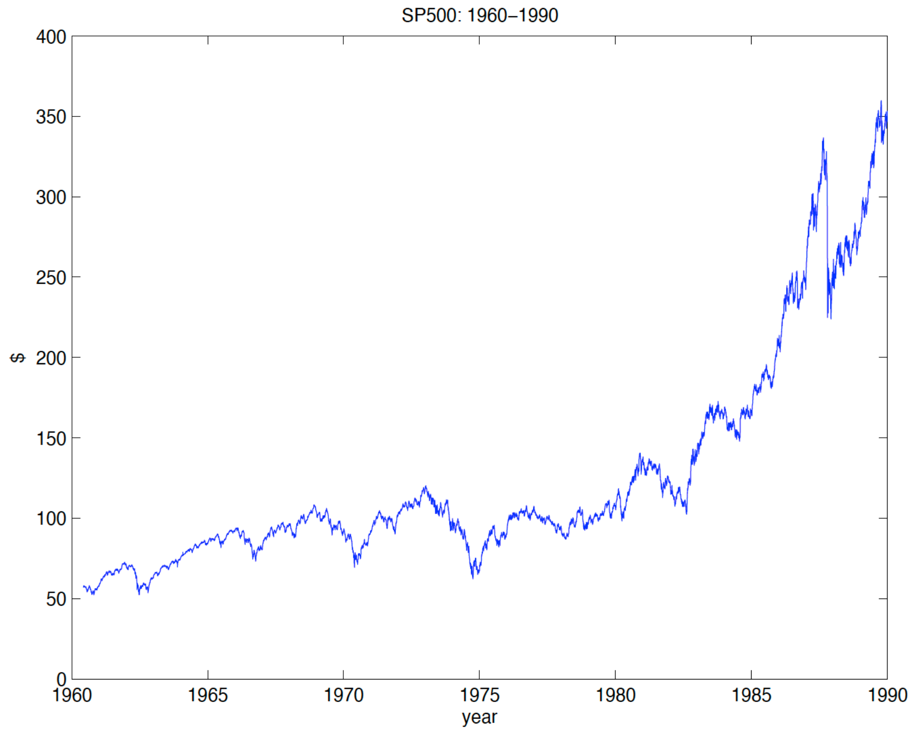
MATH 7339 - Machine Learning and Statistical Learning Theory 2

Section. Times Series Introduction

1. Overview
2. Times Series Models
3. Examples



Overview of Times Series

1. Time series models
    (a) Stationarity.
    (b) Autocorrelation function.
    (c) Transforming to stationarity.
2. Time domain methods
    (a) AR/MA/ARMA models.
    (b) ACF and partial autocorrelation function.
    (c) Forecasting
    (d) Parameter estimation
    (e) ARIMA models/seasonal ARIMA models
3. Spectral analysis
    (a) Spectral density
    (b) Periodogram
    (c) Spectral estimation

- "Time Series Analysis and Its Applications", 4th ed. 2017, by Shumway and Stoffer.

- Introduction to time series and exploratory techniques. Time plots, calculation of the sample autocorrelation. (Class Notes. Shumway and Stoffer Ch. 1 and 2)
- Time Series Regression (Class Notes. Shumway and Stoffer Ch. 2)
- ARMA modeling. Estimation of autoregressive moving averages processes via frequentist and Bayesian approaches. Model diagnostics, forecasting and applications. (Class Notes. Shumway and Stoffer Ch. 3)
- Spectral estimation using Fourier analysis and Filtering. Bayesian Approach (Class Notes. Shumway and Stoffer Ch. 4)

Notation and Terminology :

- $X_1, X_2, \ldots, X_t$ is a stochastic process, i.e., a collection of random variables indexed by a set T.
- $x_1, x_2, \ldots, x_t$ is a single realization.
- 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.

- **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 **stochastic processes**.

➢ Time Series Models

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

A **time series model** specifies the *joint distribution* of the sequence of random variables: $\{X_t\}$.

For example:

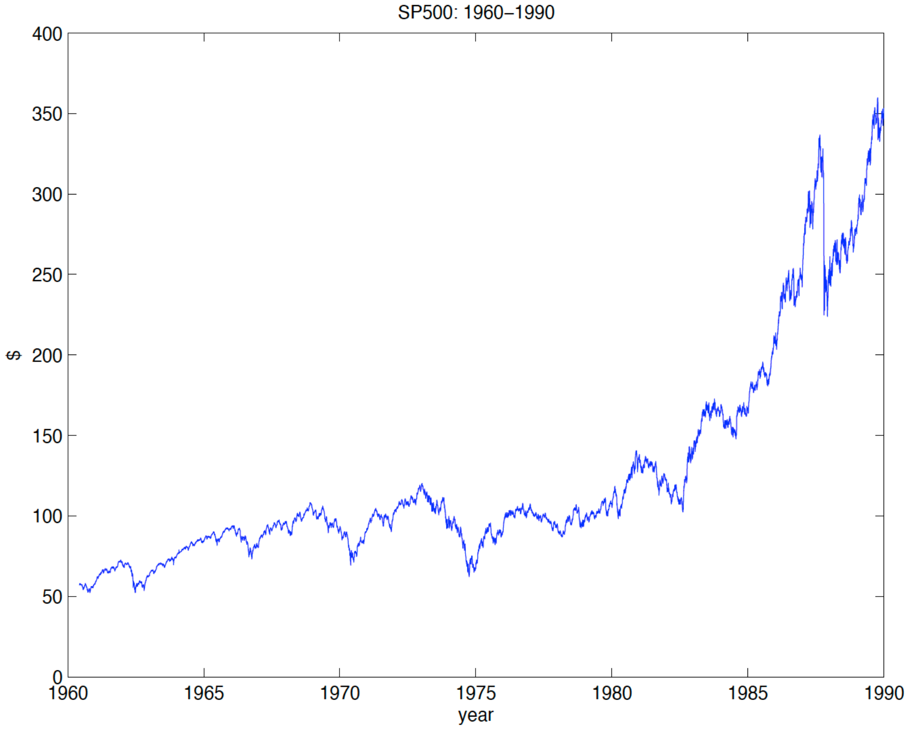
$$P(X_1 \leq x_1, \ldots, X_t \leq x_t) \text{ for all } t \text{ and } x_1, x_2, \ldots, x_t$$

We'll mostly restrict our attention to **second-order moments** properties only:

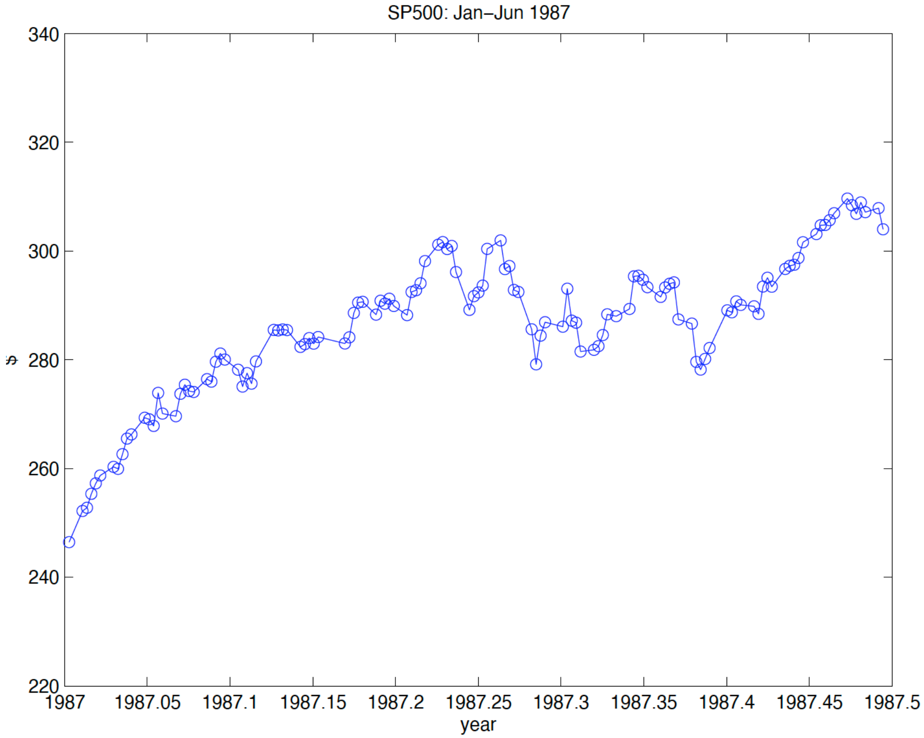
$$E[X_t], \qquad Cov(X_s, X_t).$$

Closing price of S&P 500 stocks

At the end of each month



At the end of each week



X_t is a **discrete** time, **continuous** state space stochastic process.

Example: White noise

Typically, we are thinking of a sequence of random variables that may be dependent on one another, $x_1, \ldots, x_n$. There may be times when we want to think of this as an infinite list $\ldots, x_{-1}, x_0, x_1, \ldots$

One model $\{X_t\}$ with which we are already familiar consists of a sequence of **uncorrelated** random variables. When the **mean** is **zero** and the sequence is indexed by time (t), this is usually called **white noise.**

$$E[X_t] = 0,\ Var(X_t) = \sigma^2\ and\ Cov(X_s, X_t) = 0 \text{ for all } s \neq t$$

Denote $X_t \sim WN(0, \sigma^2)$

Note: Uncorrelated RVs does not imply they are independent. Independent RVs implies they are uncorrelated.

Example: iid white noise.

$\{X_t\}$ independent and identically distributed, i.e.,

$$P(X_1 \leq x_1, \dots, X_t \leq x_t) = P(X_1 \leq x_1) \cdots P(X_t \leq x_t)$$

It is NOT interesting for forecasting:

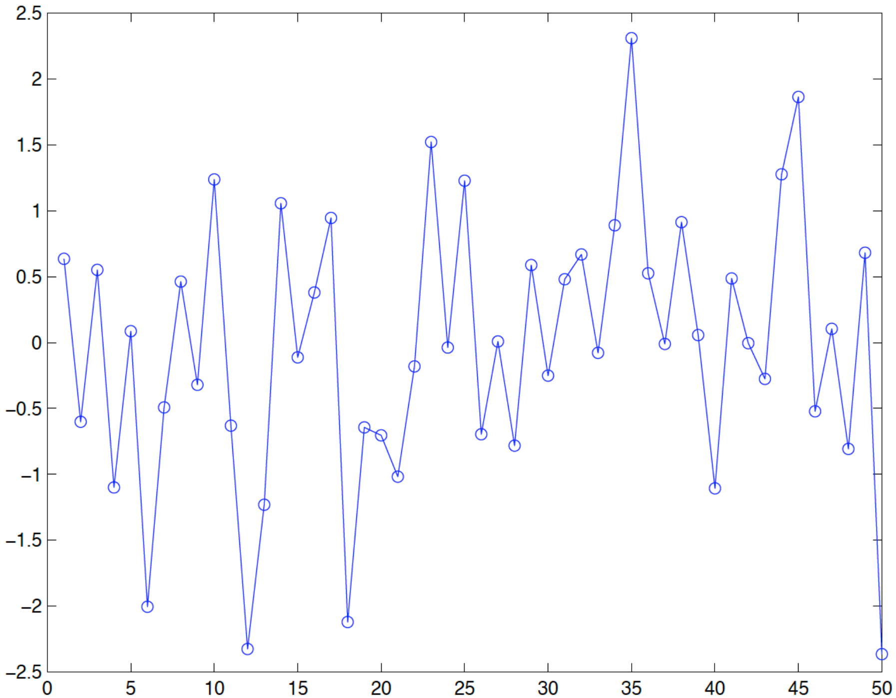
$$P(X_t \leq x_t | X_1, \dots, X_{t-1}) = P(X_t \leq x_t)$$

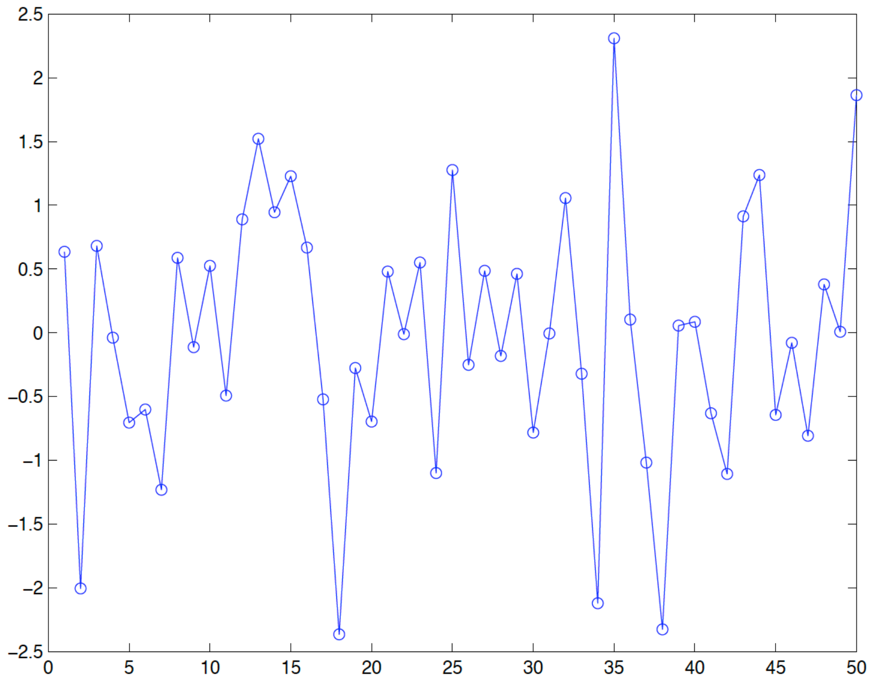
Example: Gaussian white noise

$$X_t \sim Normal(0, \sigma^2)$$

For example, when $\sigma = 1$,

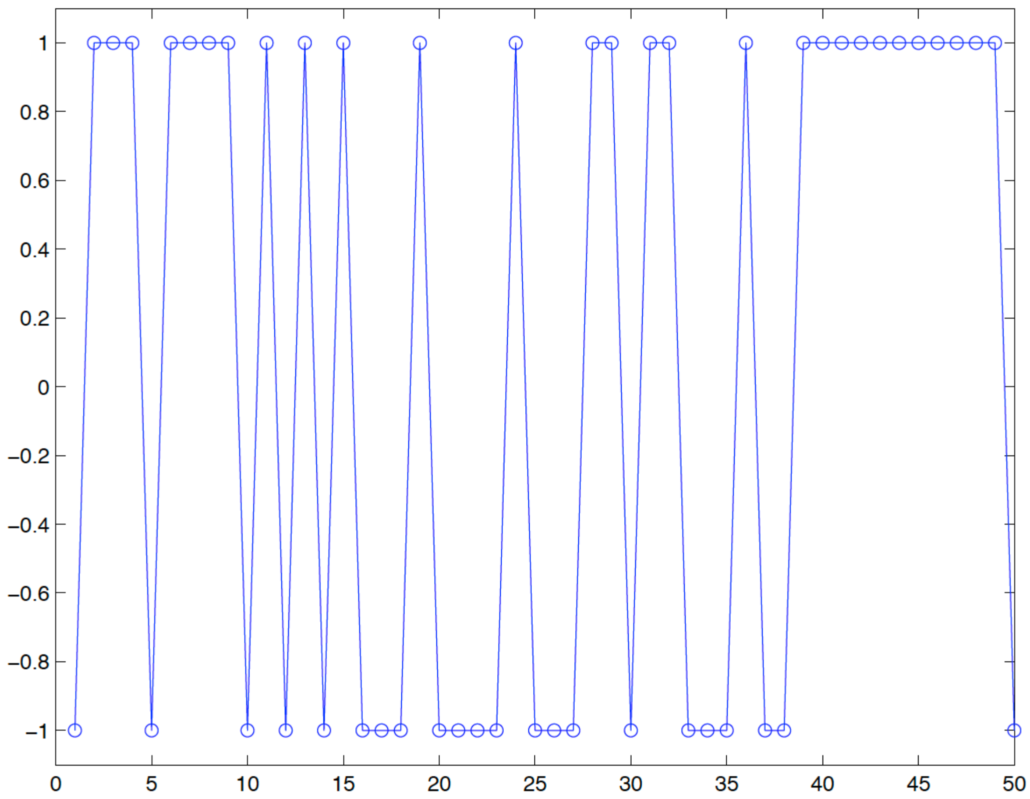
$$P(X_t \leq x_t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x_t} e^{-x^2/2}\, dx$$





Example: Binary i.i.d. Process

$$P(X_t = 1) = p; \quad P(X_t = -1) = 1 - p$$



Example $P(X_t = 1) = P(X_t = -1) = 1/2$

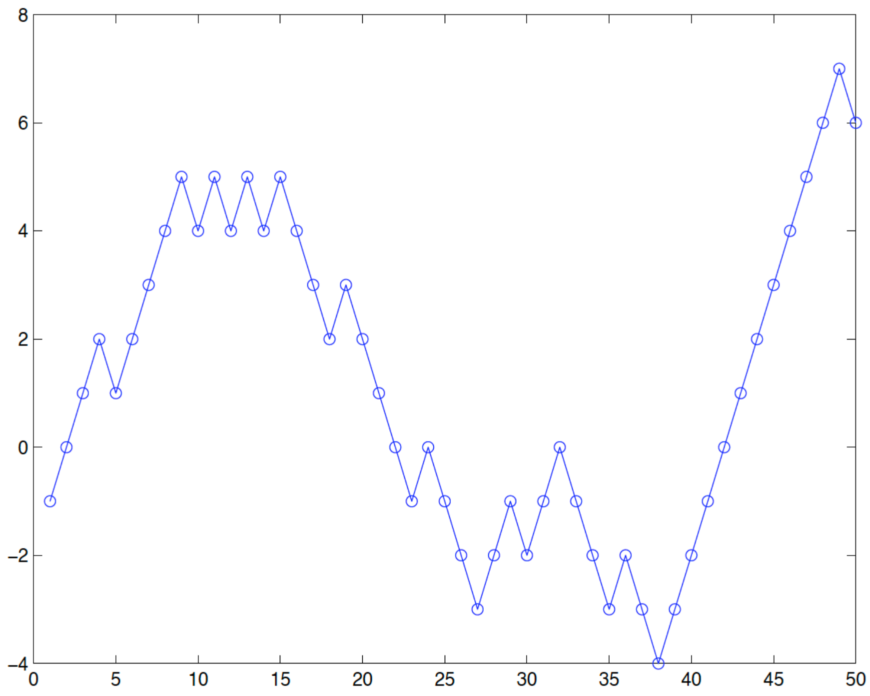
Example: Random Walk

A model for analyzing trend is the random walk model. Your current position is determined by where you were at the last step plus the random step that you just took. So, the equation would be

$$S_t = \sum_{i=1}^{t} X_i$$

where X_i is the **iid** noise.

For example, if $\{X_i\}$ is the binary process as last example,

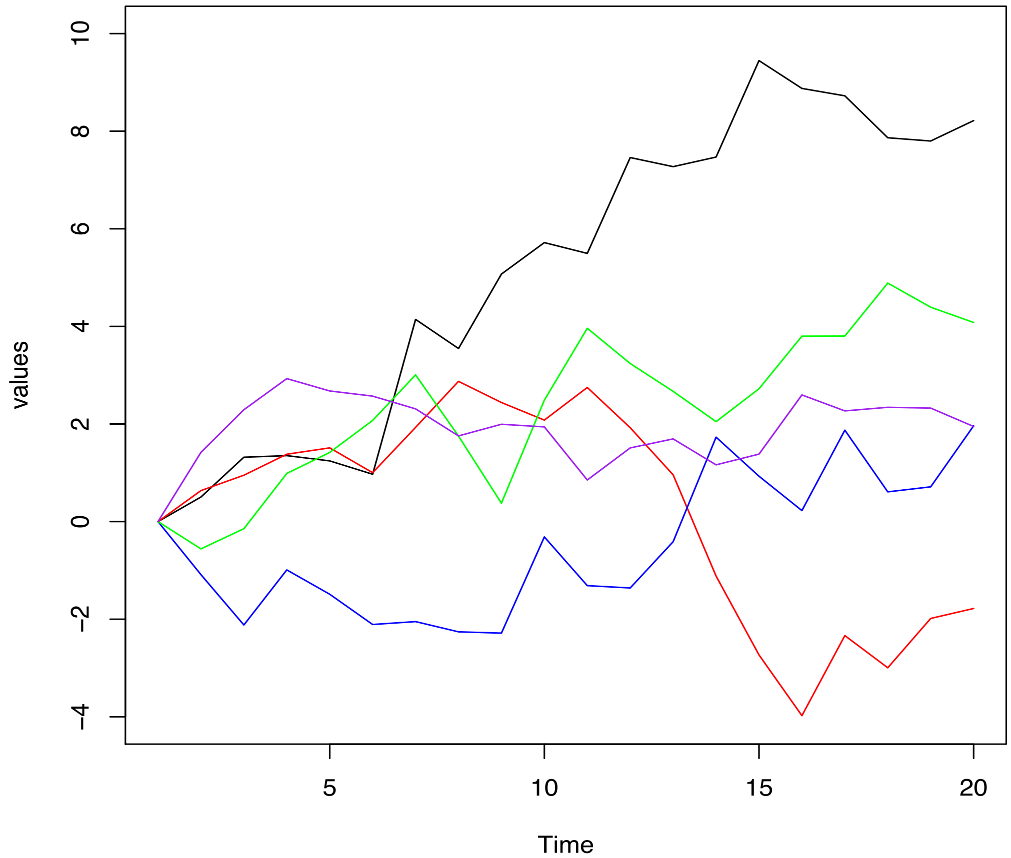


Random walks:

Differences: $\nabla S_t = S_t - S_{t-1} = X_t$

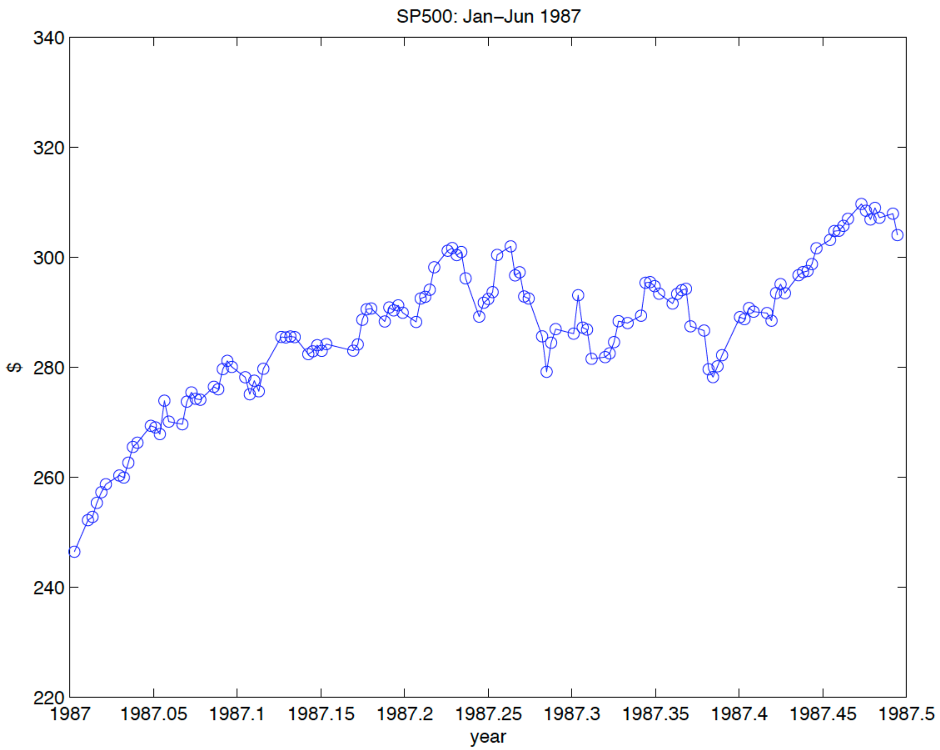
Mean: $E(S_t) =$

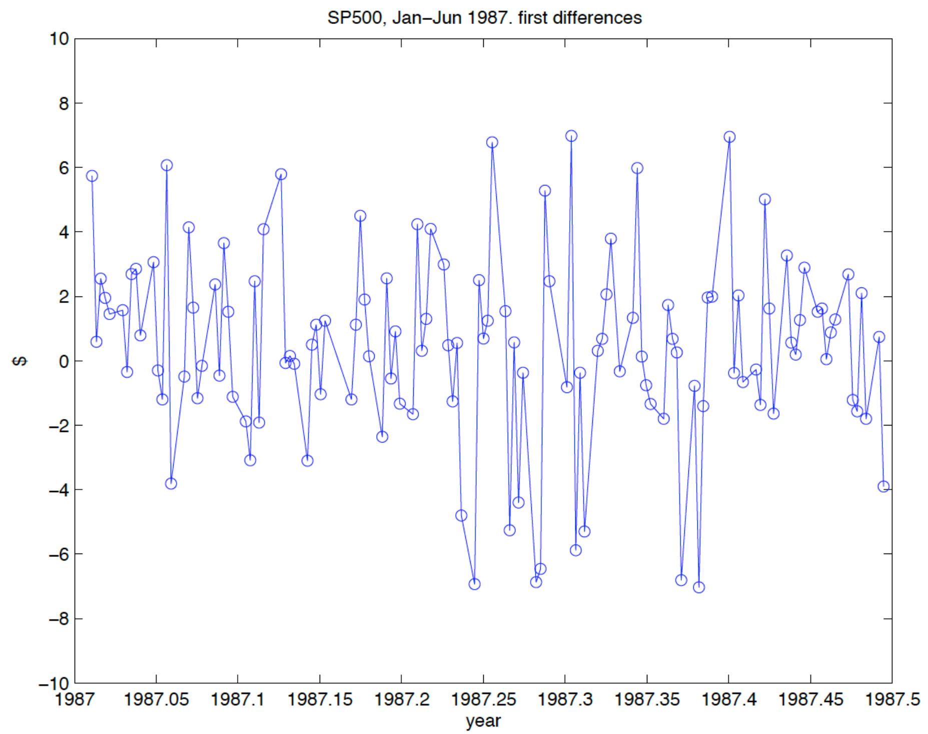
Variance: $\mathbf{Var}(S_t) =$



S&P500 data

Differences: $\nabla S_t = S_t - S_{t-1} = X_t$

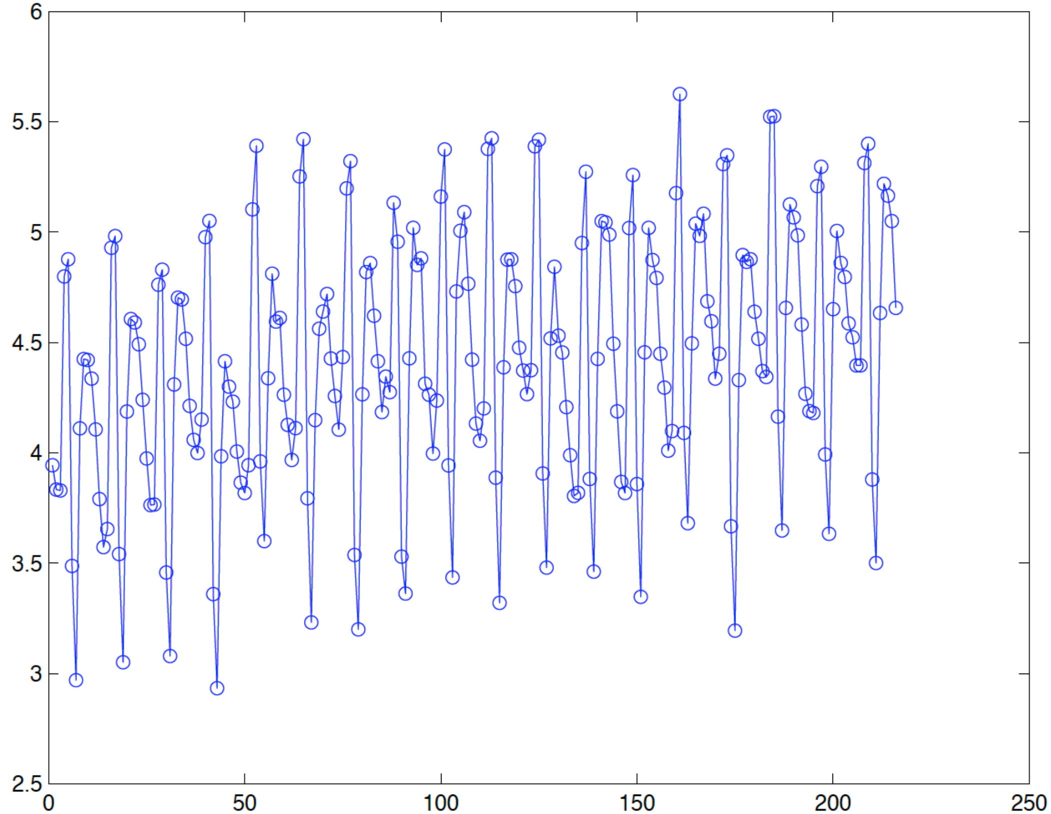




Trend and Seasonal Models

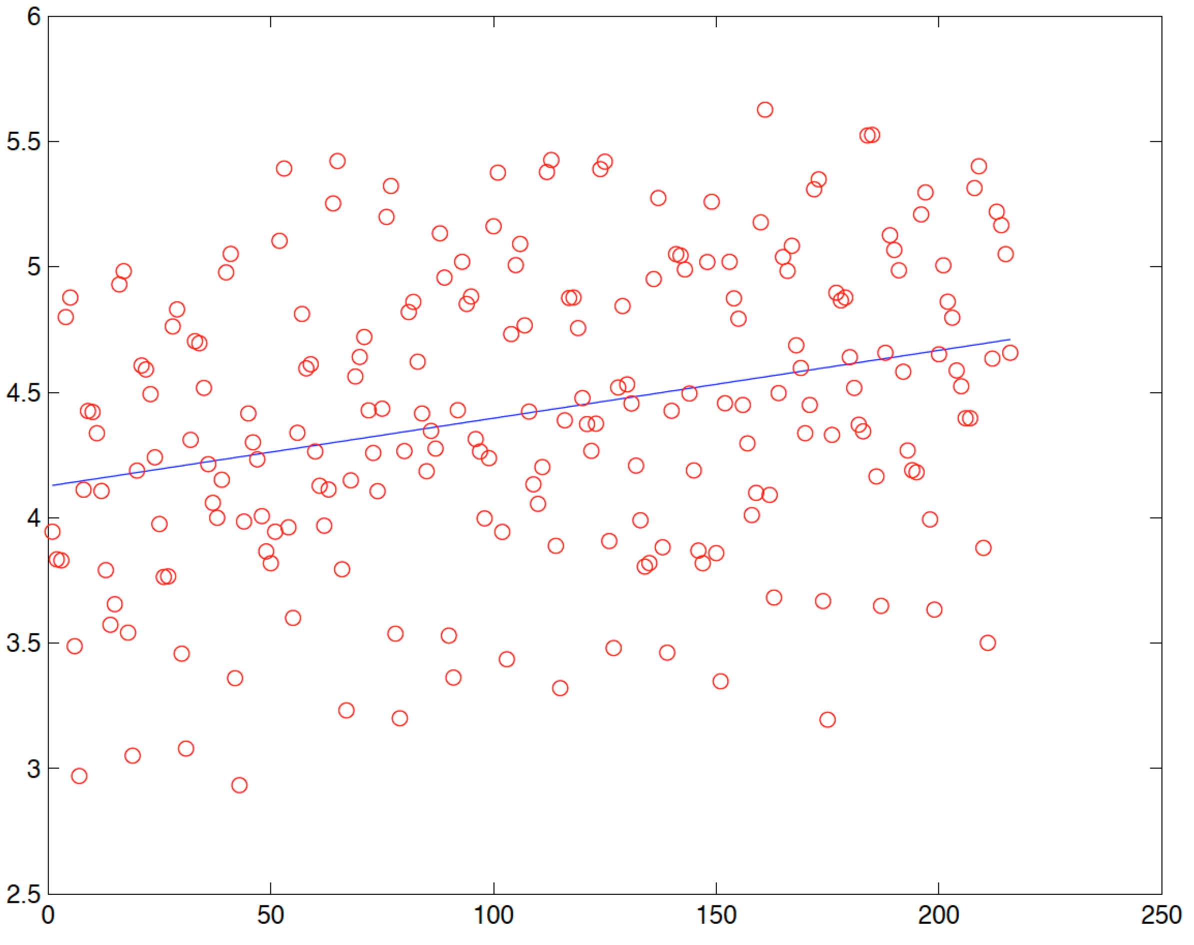
$$X_t = T_t + S_t + E_t$$

$$= \beta_0 + \beta_1 t + \sum_i \alpha_i \cos(\lambda_i t) + \gamma_i \sin(\lambda_i t) + E_t$$



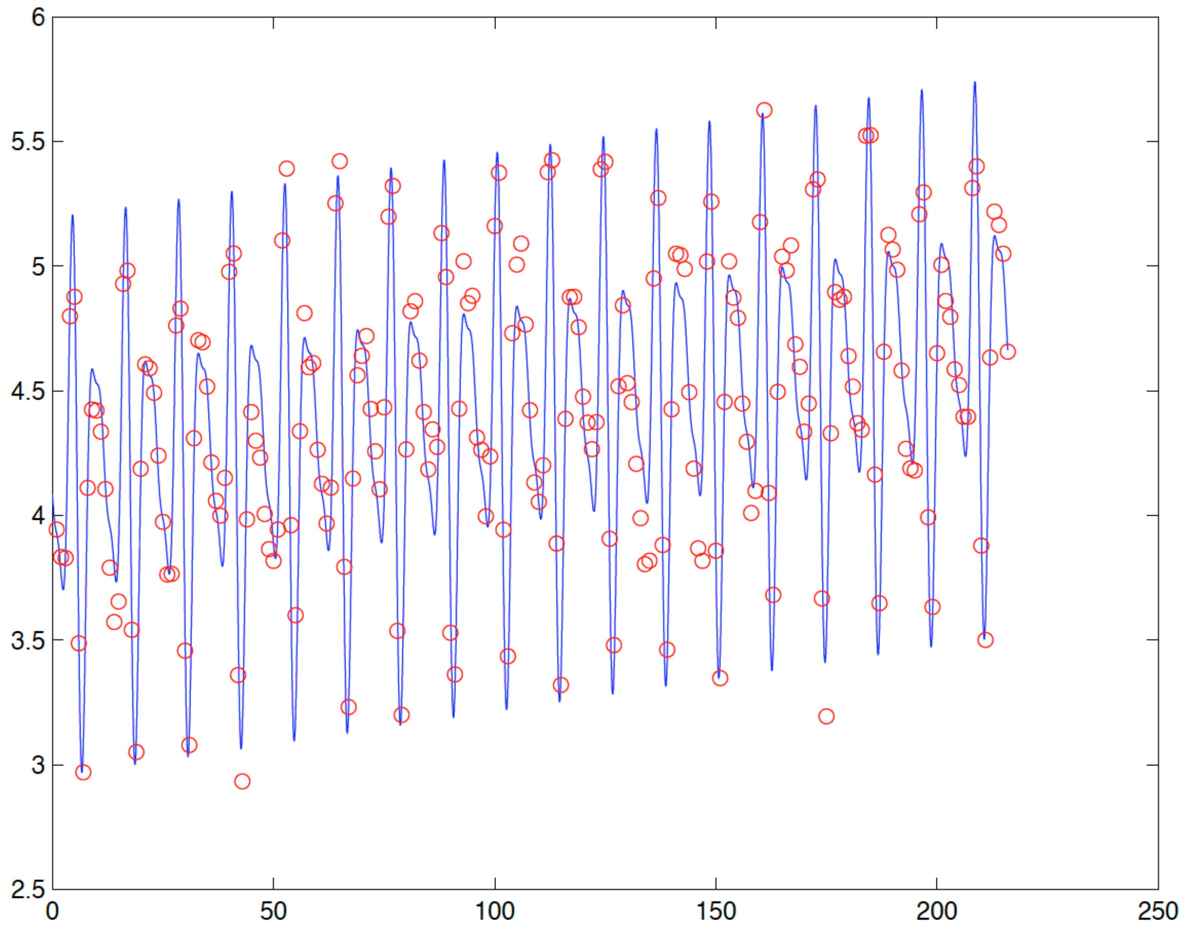
$$X_t = T_t + E_t$$

$$= \beta_0 + \beta_1 t + E_t$$

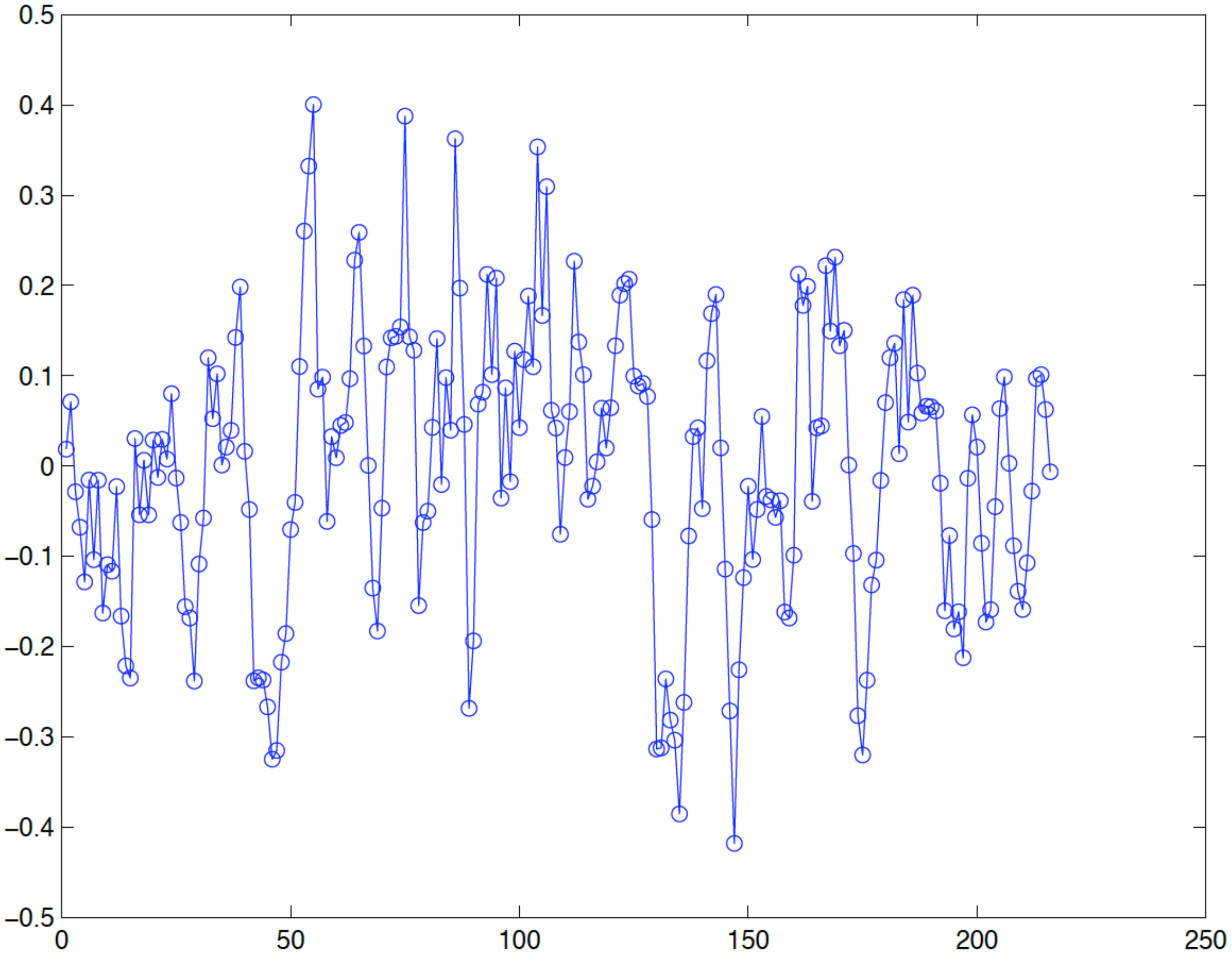


$$X_t = T_t + S_t + E_t$$

$$= \beta_0 + \beta_1 t + \sum_i \alpha_i \cos(\lambda_i t) + \gamma_i \sin(\lambda_i t) + E_t$$



Trend and Seasonal Models: Residuals



Time Series Modelling (Chasing stationarity**)**

Step 1. 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/step** changes?
- Are there **outliers**?

Time series plots will be an important tool.

Step 2. Transform data so that **residuals** are **stationary**

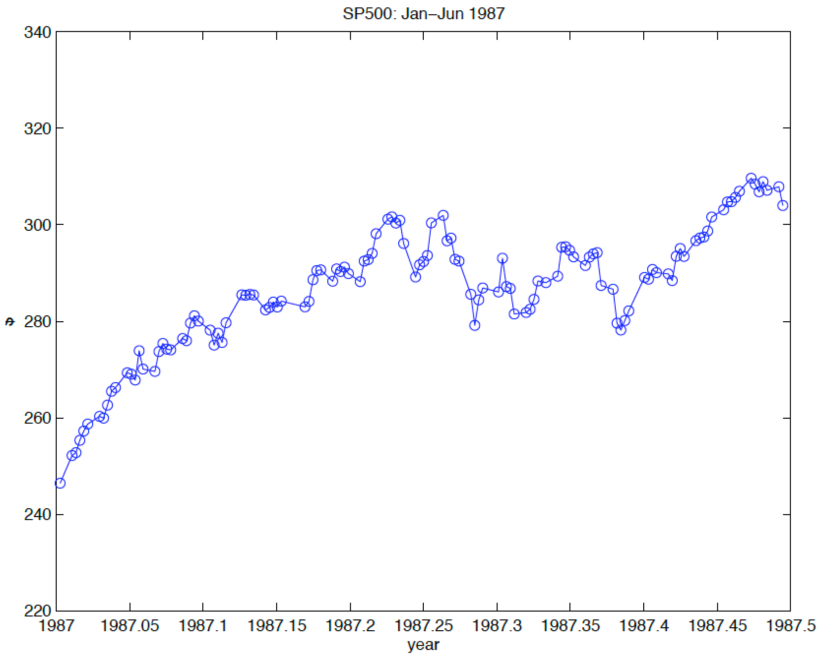
(1) Estimate and subtract T_t and S_t.

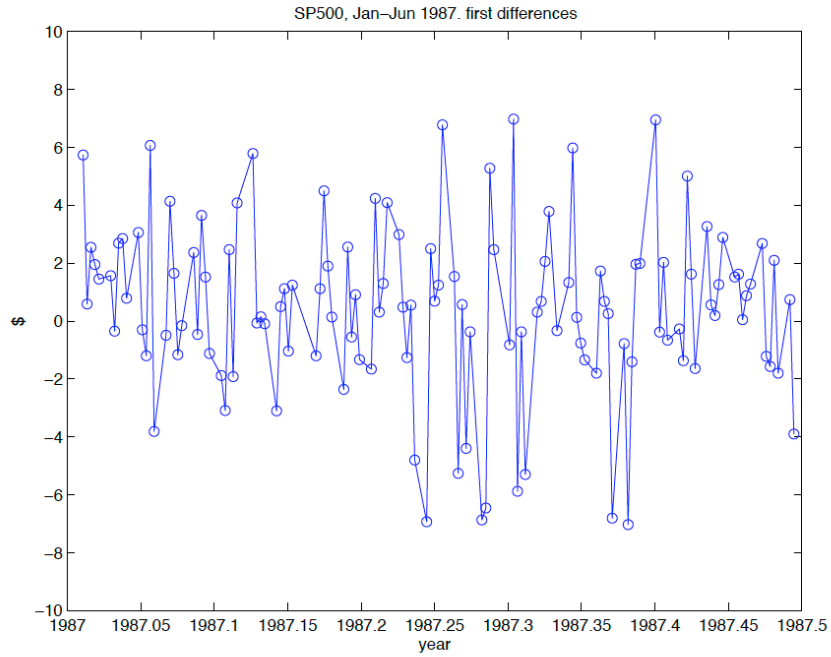
(2) Differencing.

(3) Nonlinear transformations (log, root function).

Step 3. Fit model to **residuals**.

S&P 500 data. Differencing and Trend





Define the **lag-1** difference operator, (think 'first derivative)

$$\nabla X_t = X_t - X_{t-1} = (1 - B)X_t$$

where B is the **backshift** operator, $BX_t := X_{t-1}$

- If $X_t = \beta_0 + \beta_1 t + Y_t$, then

$$\nabla X_t = \beta_1 + \nabla Y_t$$

- If $X_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \cdots + \beta_k t^k + Y_t$, then

$$\nabla X_t = \beta_1 + \beta_2(t^2 - (t-1)^2) + \cdots + \beta_k \left(t^k - (t-1)^k\right) + \nabla Y_t$$

and

$$\nabla^k X_t = k!\, \beta_k + \nabla^k Y_t$$

Define the lag-s difference operator,

$$\nabla_s X_t := X_t - X_{t-s} = (1 - B^s) X_t$$

where B^s is the backshift operator applied s times.

If $X_t = T_t + S_t + Y_t$ and S_t has period s, (i.e., $S_t = S_{t-s}$), then

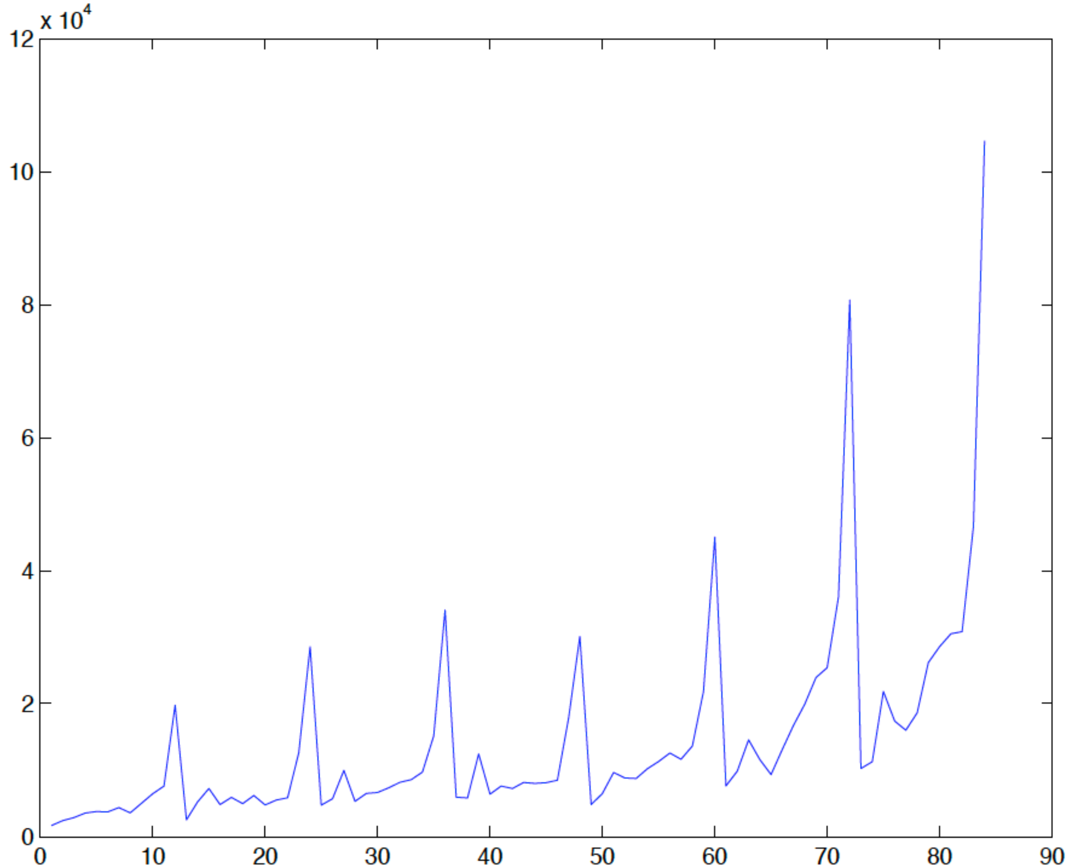
$$\nabla_s X_t = T_t - T_{t-s} + \nabla_s Y_t$$

Objectives of Time Series Analysis

1. Compact description of data.
2. Interpretation.
3. Forecasting.
4. Control.
5. Hypothesis testing.
6. Simulation.

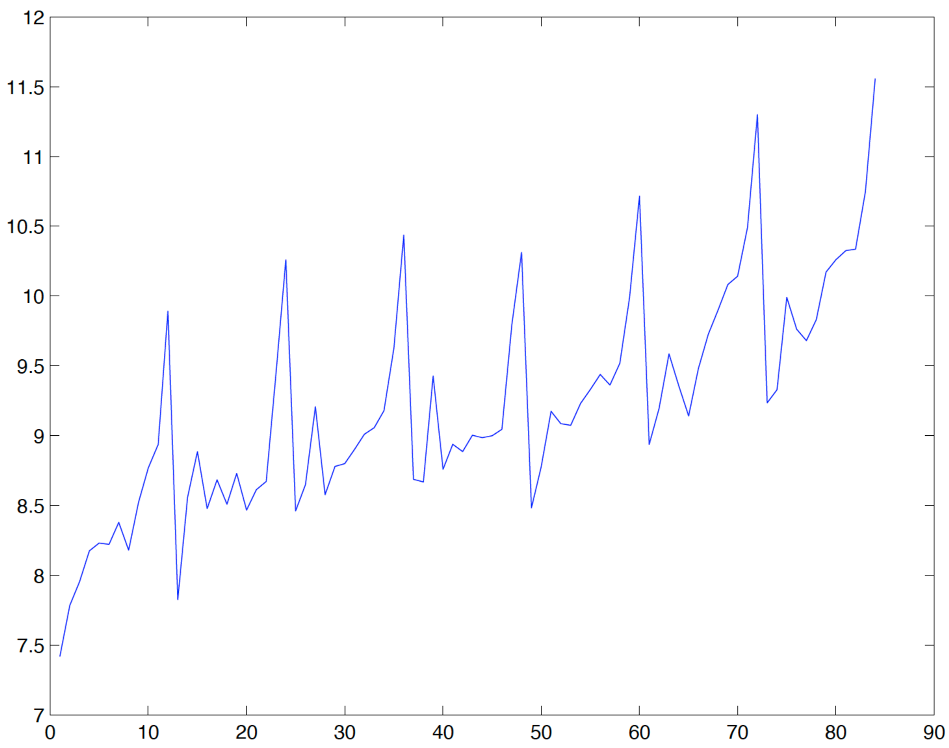
Classical decomposition: An example

Monthly sales for a souvenir shop at a beach resort town in Queensland.

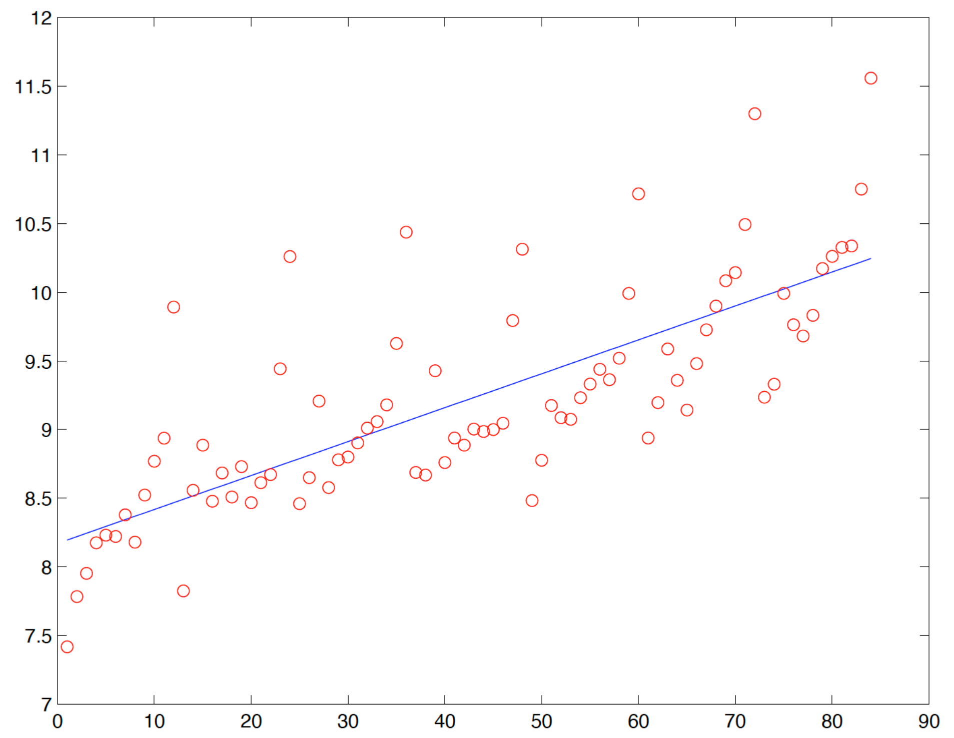


(Makridakis, Wheelwright and Hyndman, 1998)

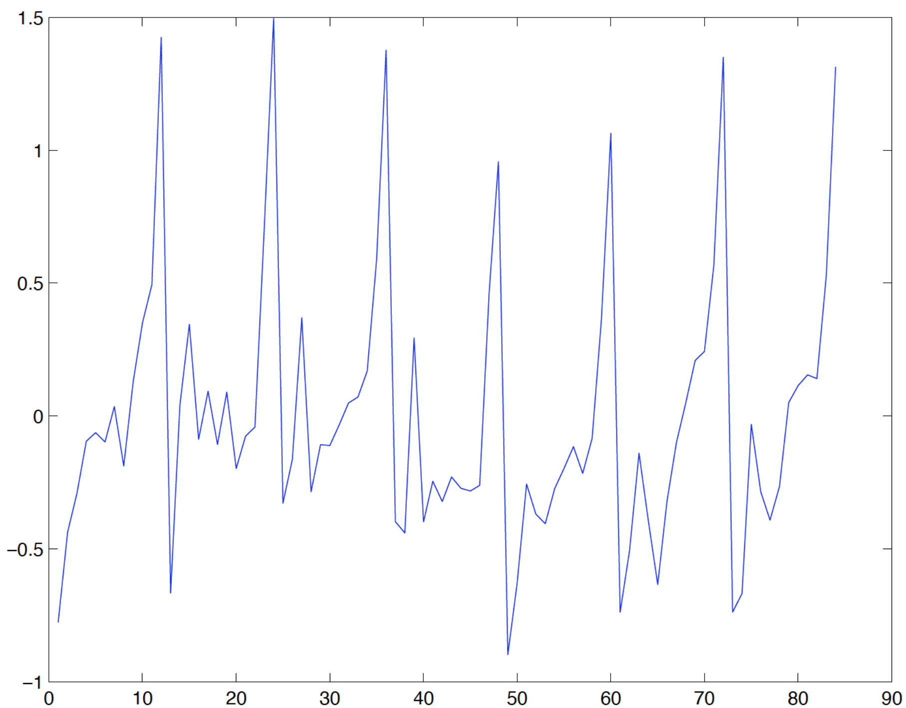
Transformed data



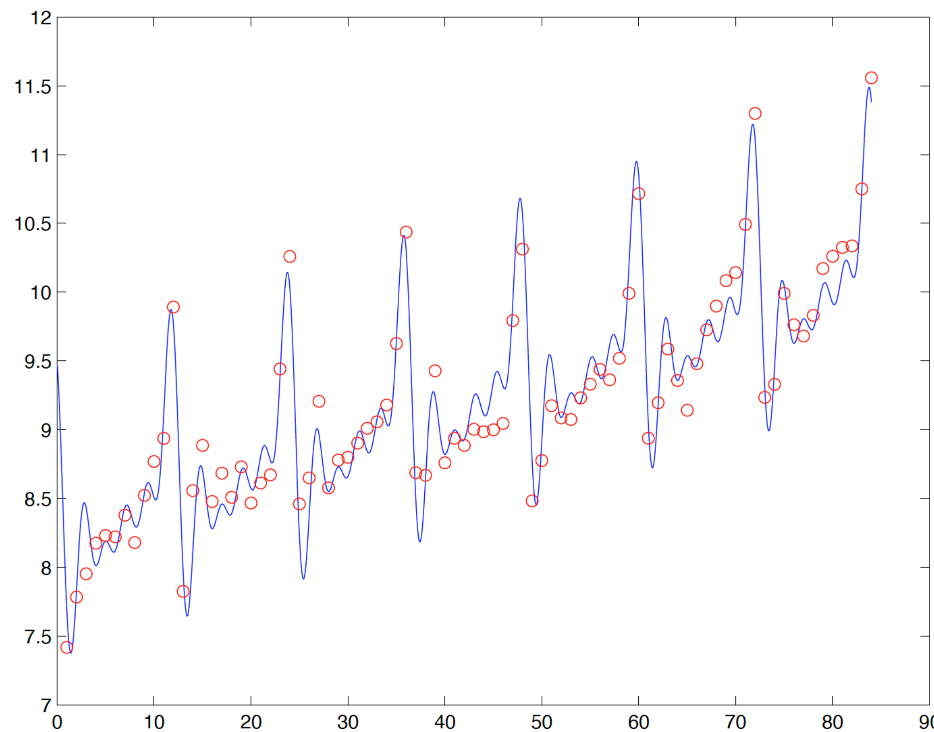
Trend



Residuals



Trend and seasonal variation



Objectives of Time Series Analysis

1. Compact description of data.

Example: Classical decomposition: $X_t = T_t + S_t + f(Y_t) + W_t$.

2. Interpretation. (e.g., Seasonal adjustment.)

3. Forecasting. (e.g., Predict sales, unemployment.)

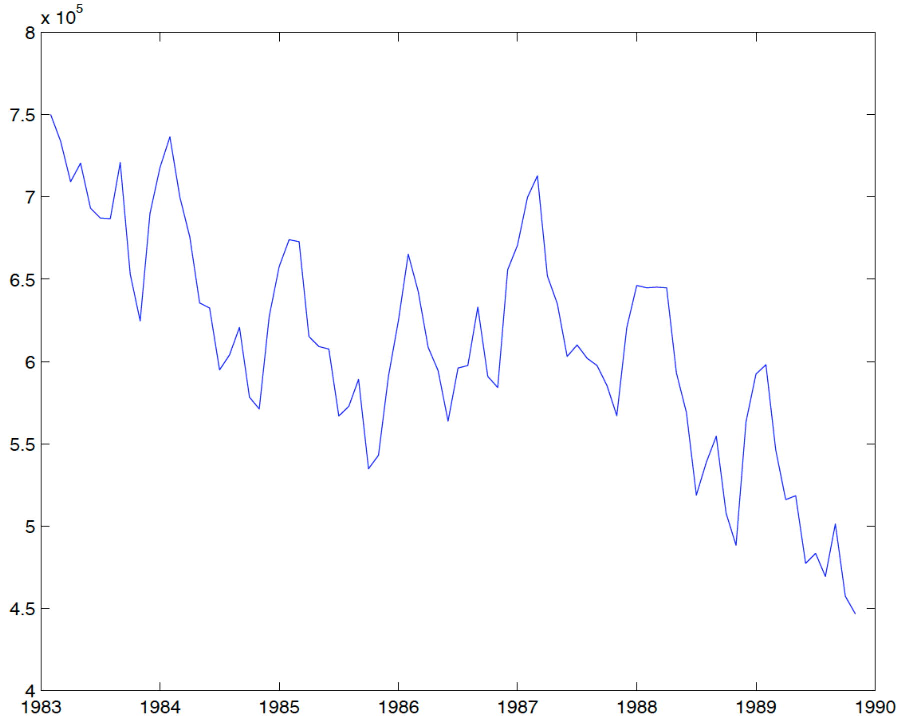
4. Control. (e.g., Example: Impact of monetary policy on unemployment)

5. Hypothesis testing. (e.g., Global warming.)

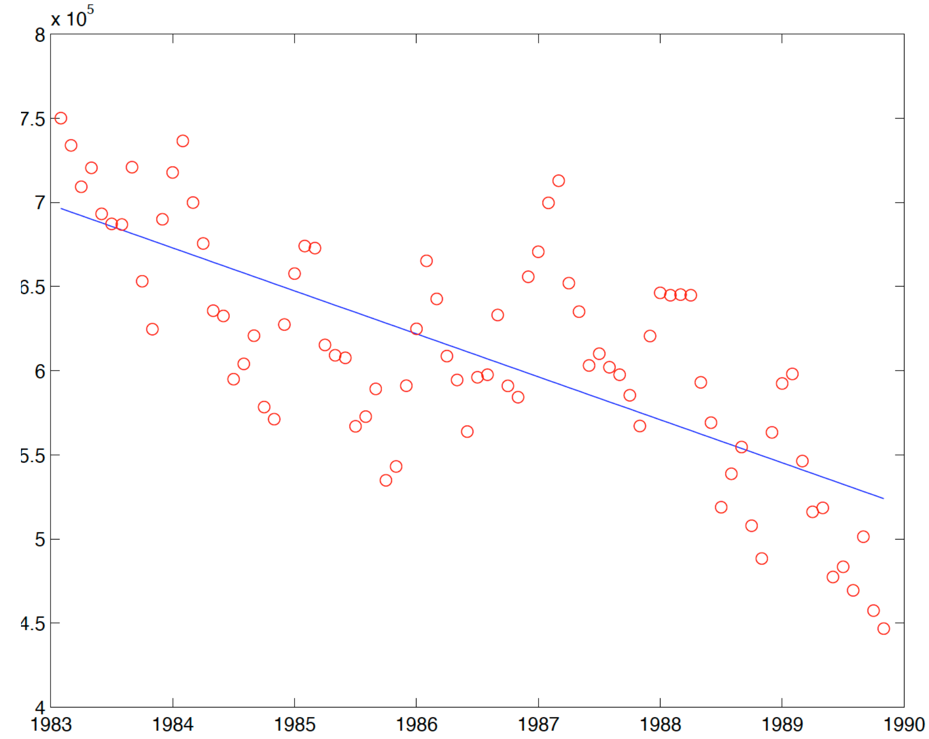
6. Simulation. (e.g.,Estimate probability of catastrophic events)

Unemployment data

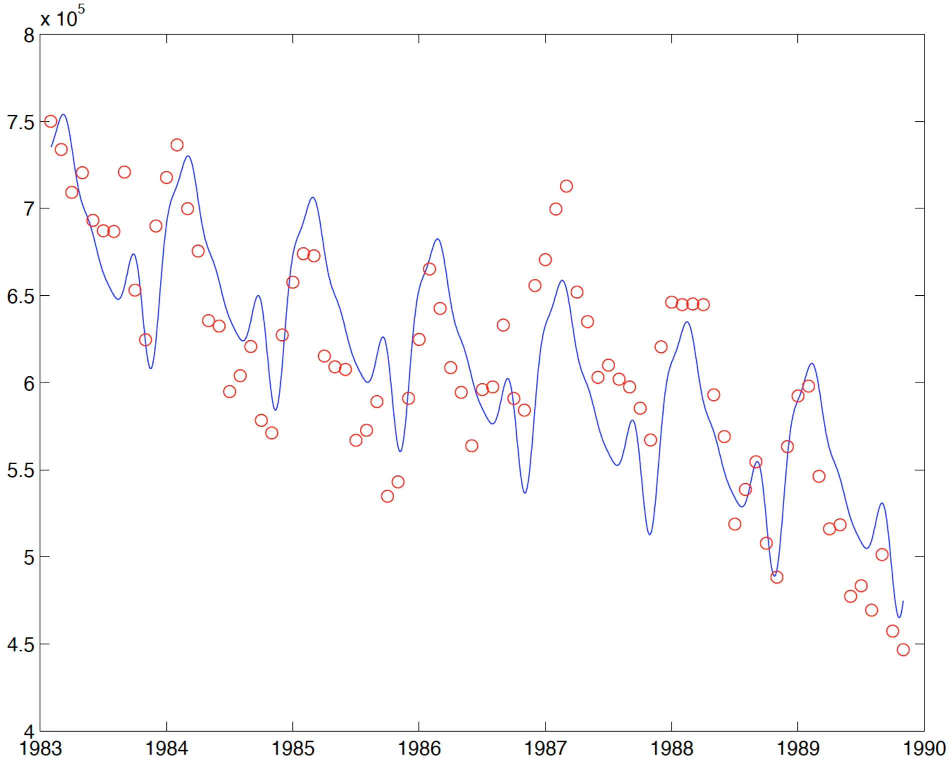
Monthly number of unemployed people in Australia. (Hipel and McLeod, 1994)



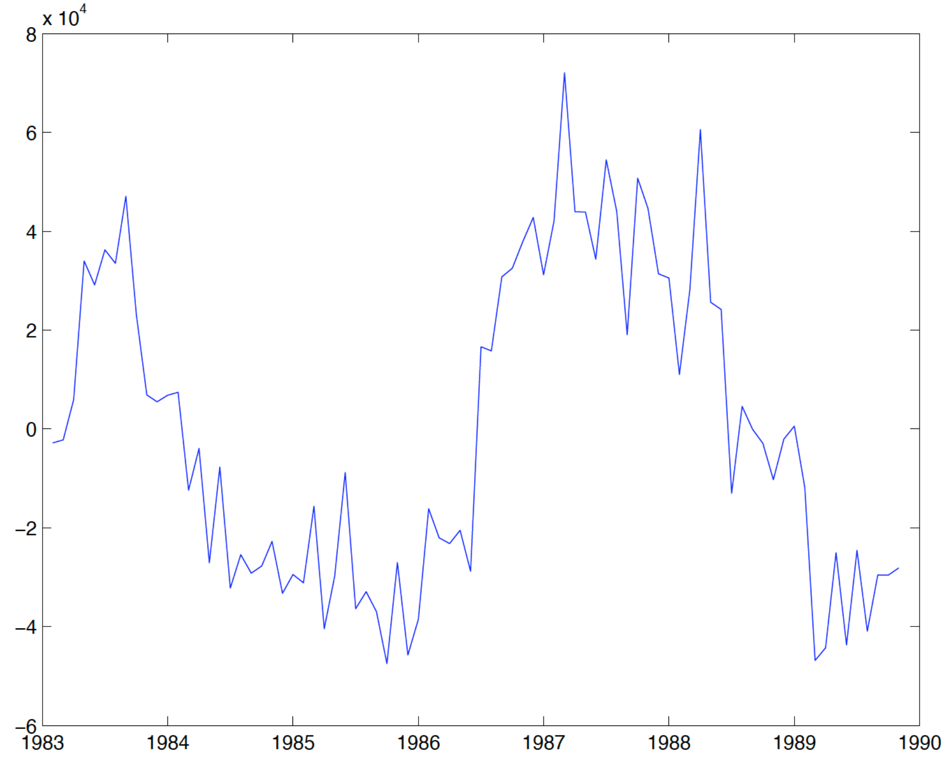
Trend



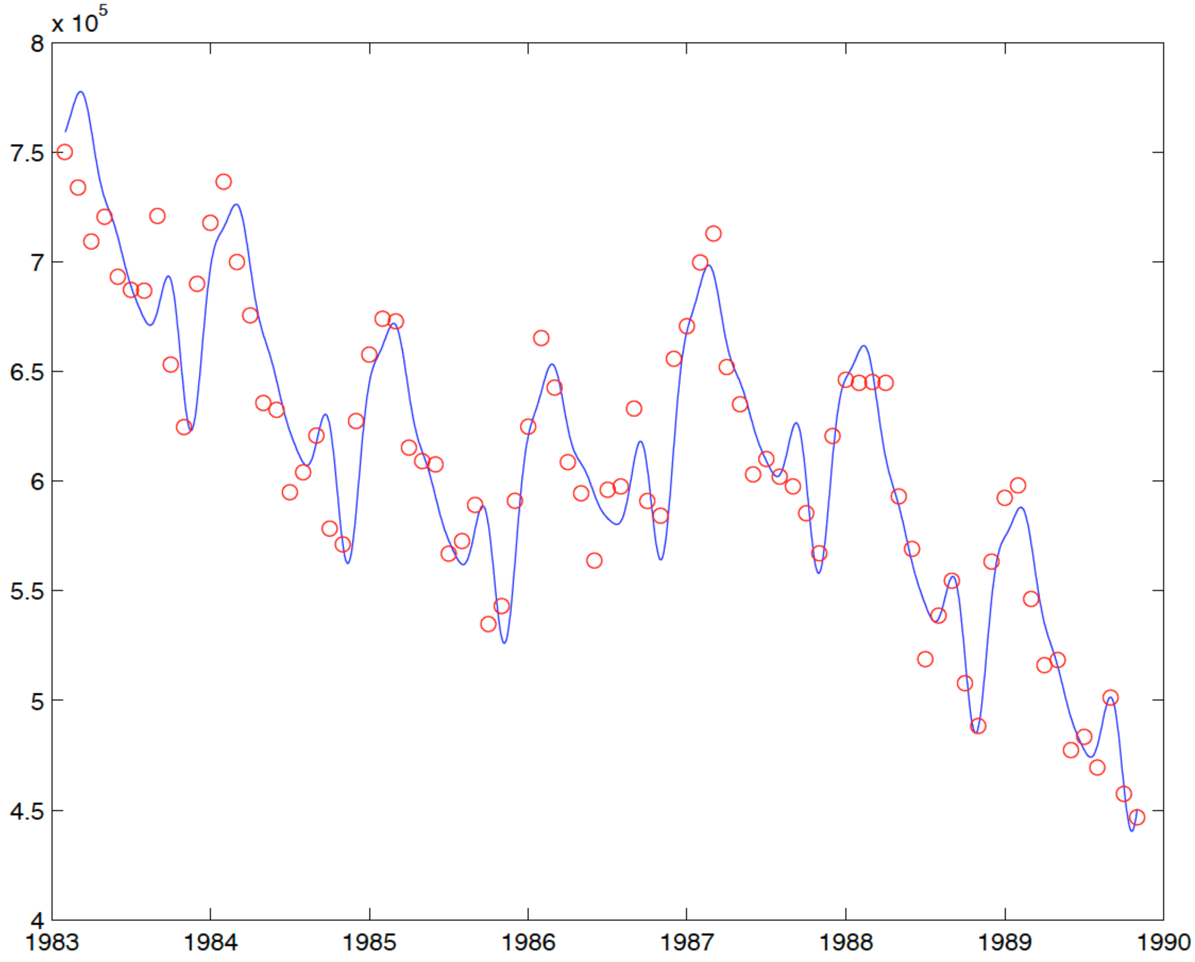
Trend plus seasonal variation



Residuals



Predictions based on a (simulated) variable



Two Approaches to Time Series

There are two primary approaches to time series.

1. **Time domain** approach. This approach focuses on the rules for a time series to move forward.

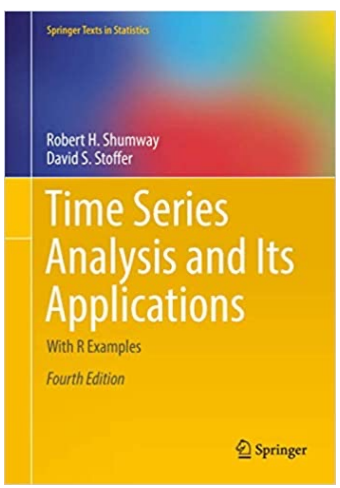
   For example, how do yesterday's and today's observations affect tomorrow's observation?

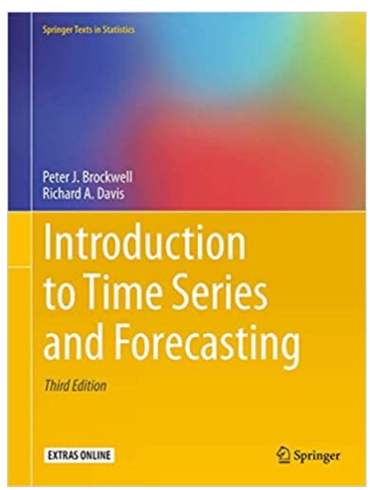
2. **Frequency domain** approach. This approach tries to understand how differing oscillations can contribute to current observations.

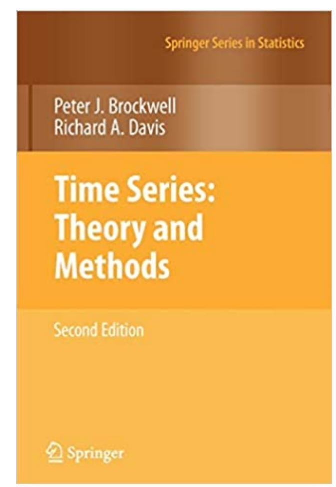
   For example, taking hourly temperatures in Boston. There will be a very clear 24 hour oscillation. There will be another clear 8,760 hour oscillation. The current temperature is a sum of these two sinusoids (plus a lot of noise and fluctuation).

References:

- "Time Series Analysis and Its Applications", 4th ed. 2017, by Shumway and Stoffer.
- "Introduction to Time Series and Forecasting", 3rd ed. 2016, by Brockwell and Davis.
- Time Series Analysis 1st Edition, by James Douglas Hamilton
- Time Series: Theory and Methods (Springer Series in Statistics) by Peter Brockwell, Richard Davis
- **Facebook: Forecasting at scale: https://facebook.github.io/prophet/**







Interesting and useful sources of economic and finance data:

- General data assembled by the St. Louis Fed: research.stlouisfed.org/fred2/.
- Interest rate data from the Fed Board of Governors: www.federalreserve.gov/econresdata/statisticsdata.htm.
- Other data from the Fed Board of Governors: www.federalreserve.gov/releases/h15/update/.
- Data from the World Bank: data.worldbank.org/.
- Stock price data: finance.yahoo.com/.

Reading and manipulating stock prices from **Yahoo Finance**.

- There is a useful R program written by John Nolan.
- **Python**: https://pypi.org/project/yfinance/
- **MATLAB**: https://www.mathworks.com/matlabcentral/fileexchange/68361-yahoo-finance-and-quandl-data-downloader