Introduction to Time Series Analysis. Lecture 1.

Peter Bartlett

1. Organizational issues.
2. Objectives of time series analysis. Examples.
3. Overview of the course.
4. Time series models.
5. Time series modelling: Chasing stationarity.

Organizational Issues

- Peter Bartlett. bartlett@stat. Office hours: Tue 9:30-10:30 (Evans 399). Thu 3-4 (Soda 723).
- Nate Coehlo. nate@stat. Office hours: Mon/Wed 12-1pm (Evans 342).
- http://www.stat.berkeley.edu/∼bartlett/courses/fall2007/
  Check it for announcements, assignments, slides, ...
- Text: *Time Series Analysis and its Applications*, Shumway and Stoffer. 2006.

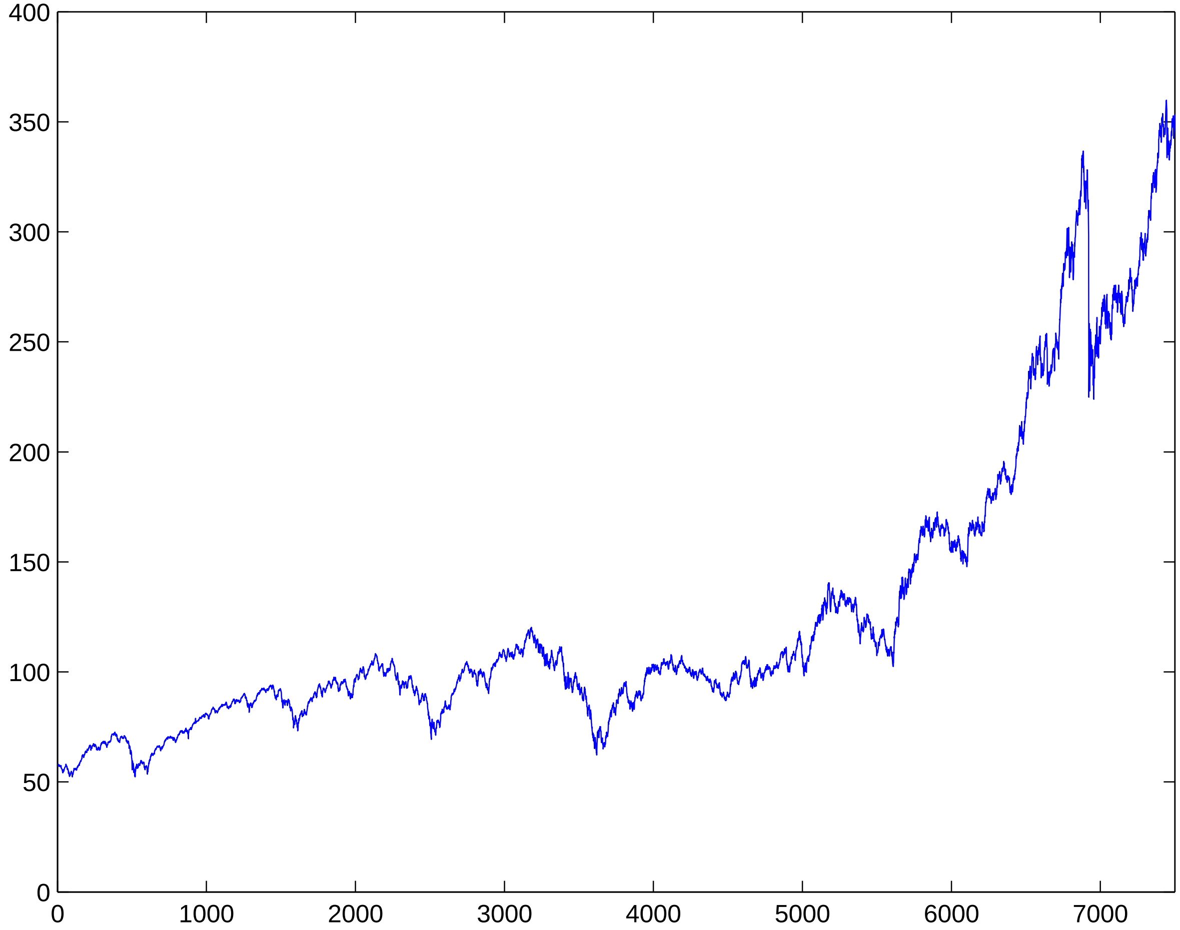
Organizational Issues

Classroom Section: Mon 11–12, in 332 Evans.
Starting tomorrow, August 29.

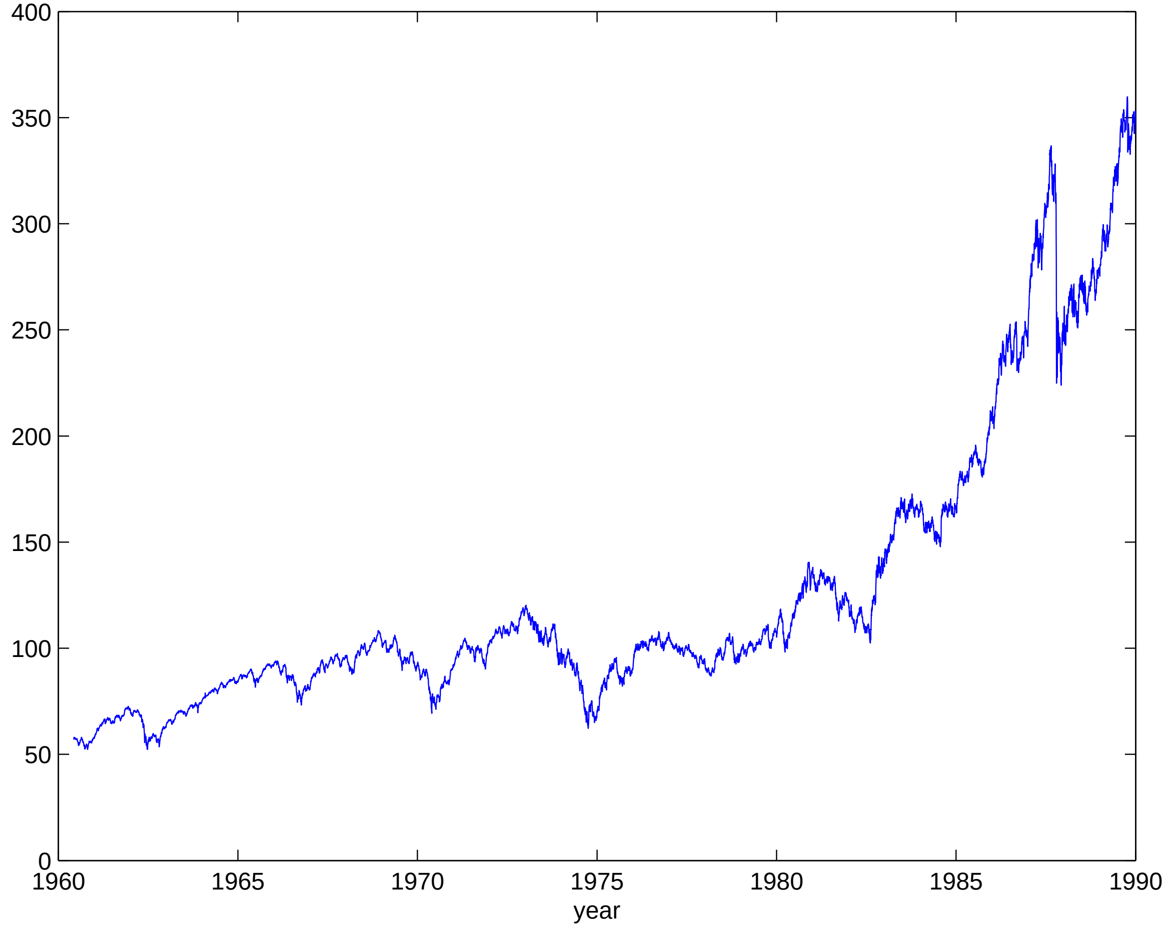
Computer Labs: Wed 11–12, in 342 Evans.
Starting next week, September 3.

Assessment:
Lab/Homework Assignments (35%): posted on the website.
These involve a mix of pen-and-paper and computer exercises. You may use any programming language you choose (R, Splus, Matlab). The last assignment will involve analysis of a data set that you choose.
Midterm Exam (25%): scheduled for October 16, at the lecture.
Final Exam (40%): scheduled for Saturday, December 15.

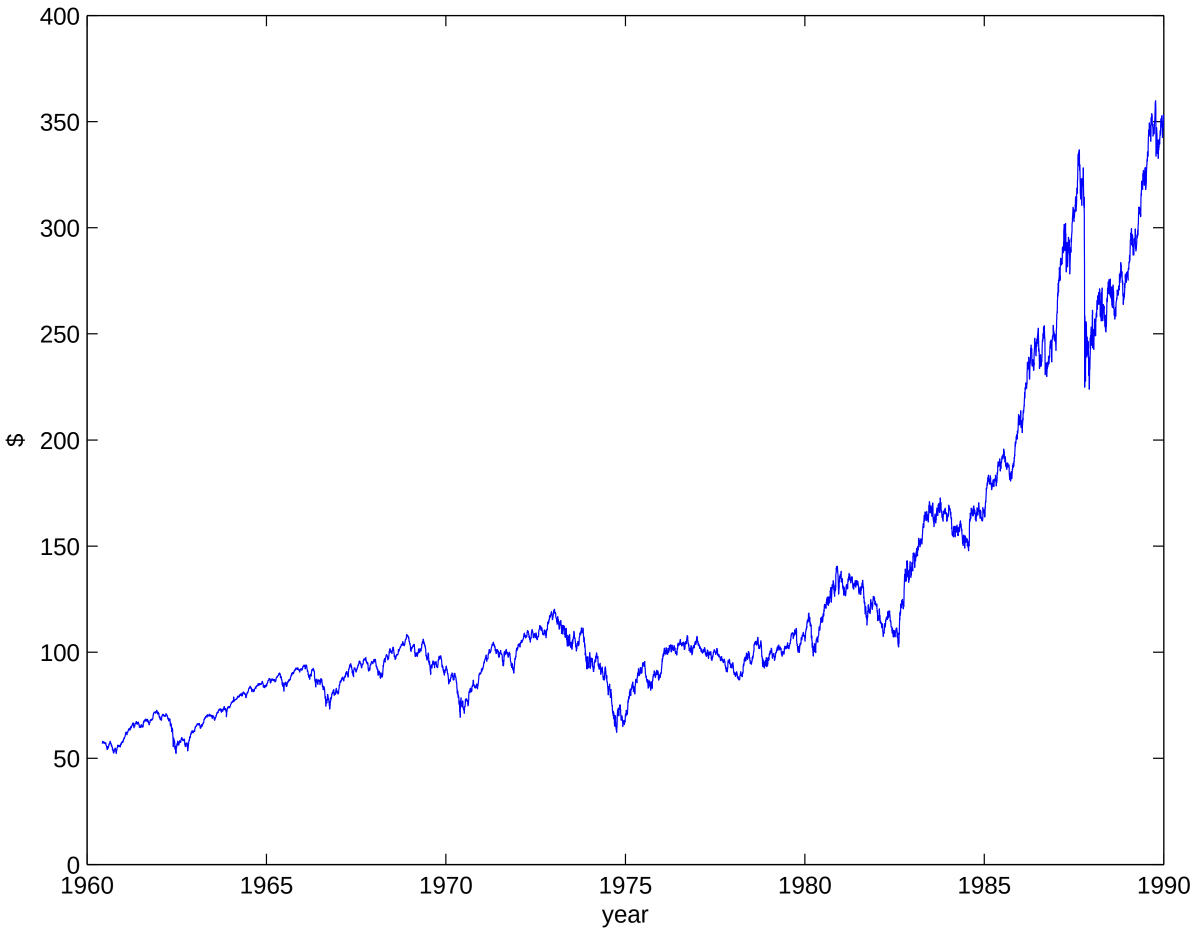
A Time Series



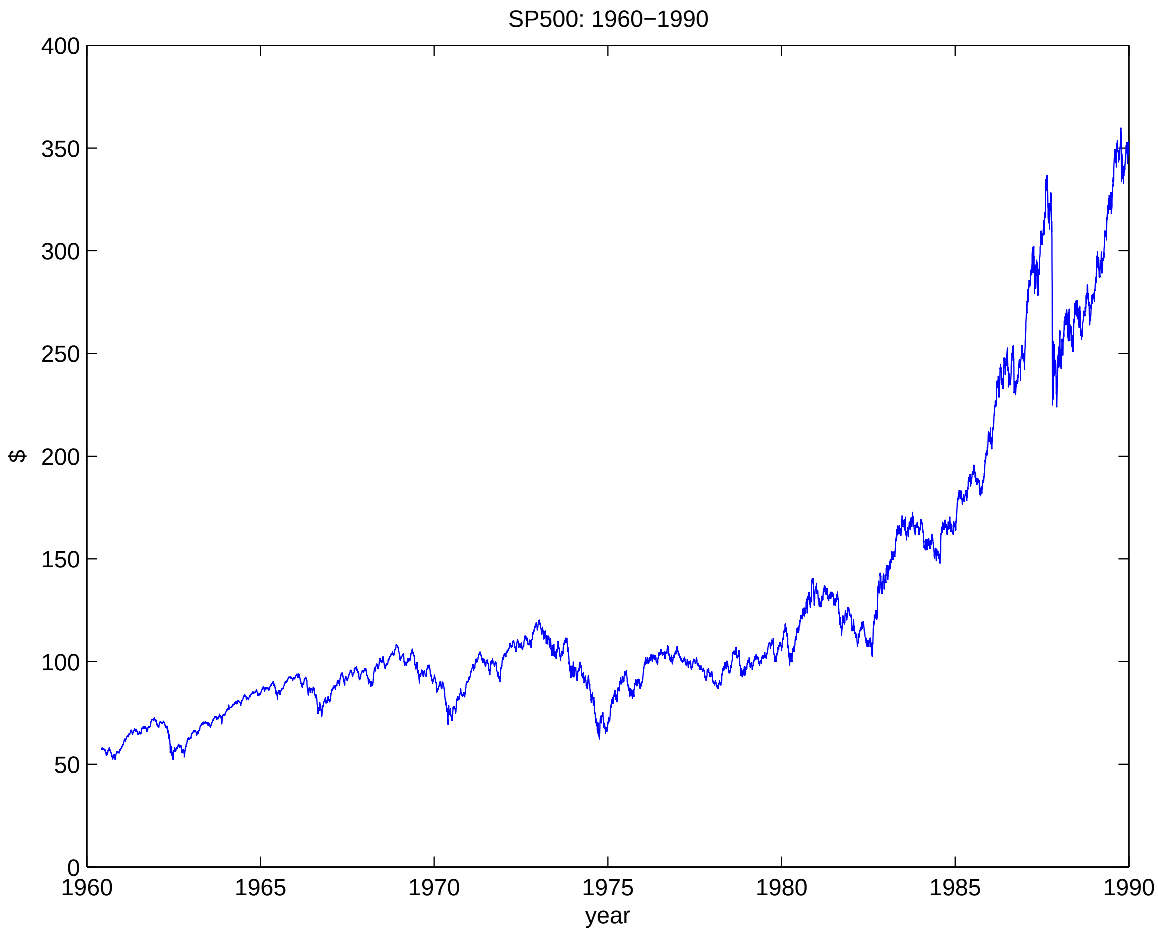
A Time Series



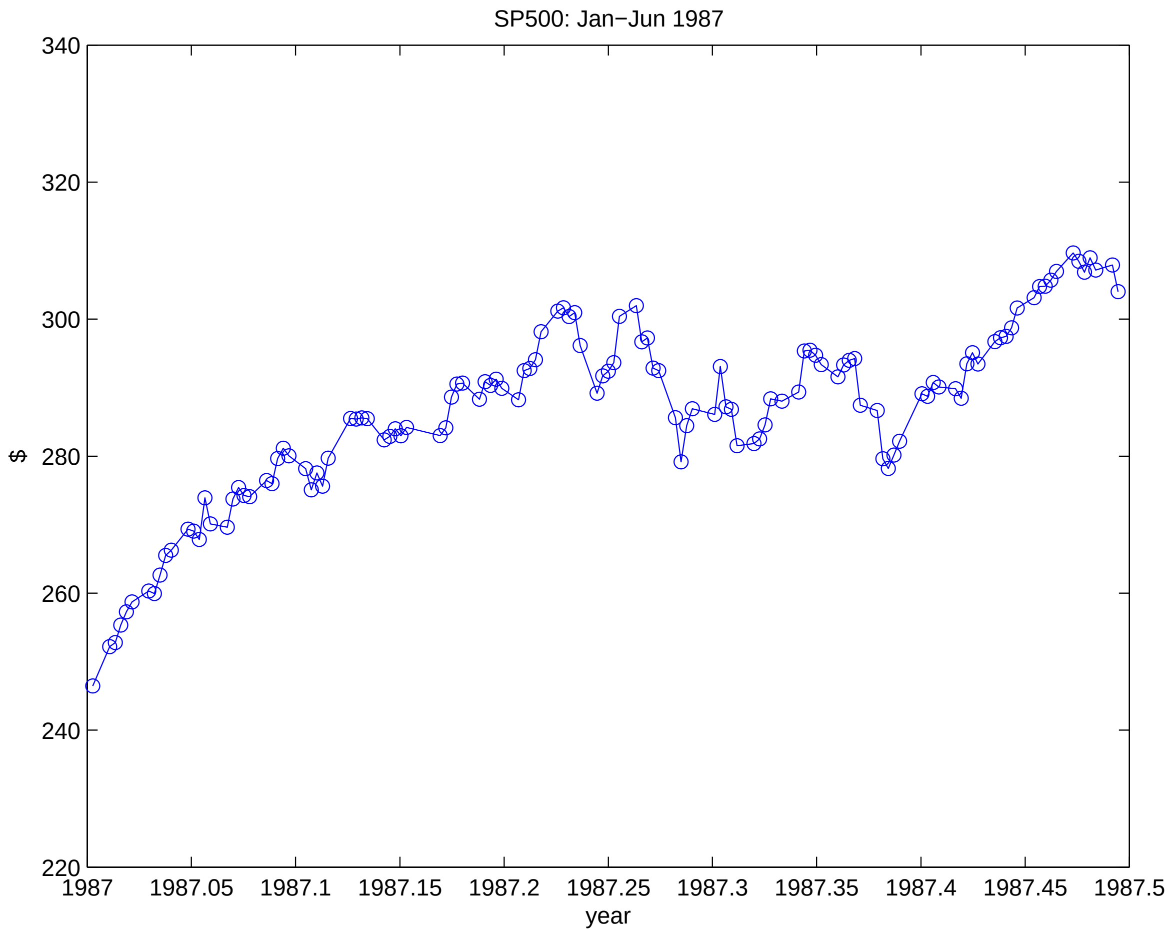
A Time Series



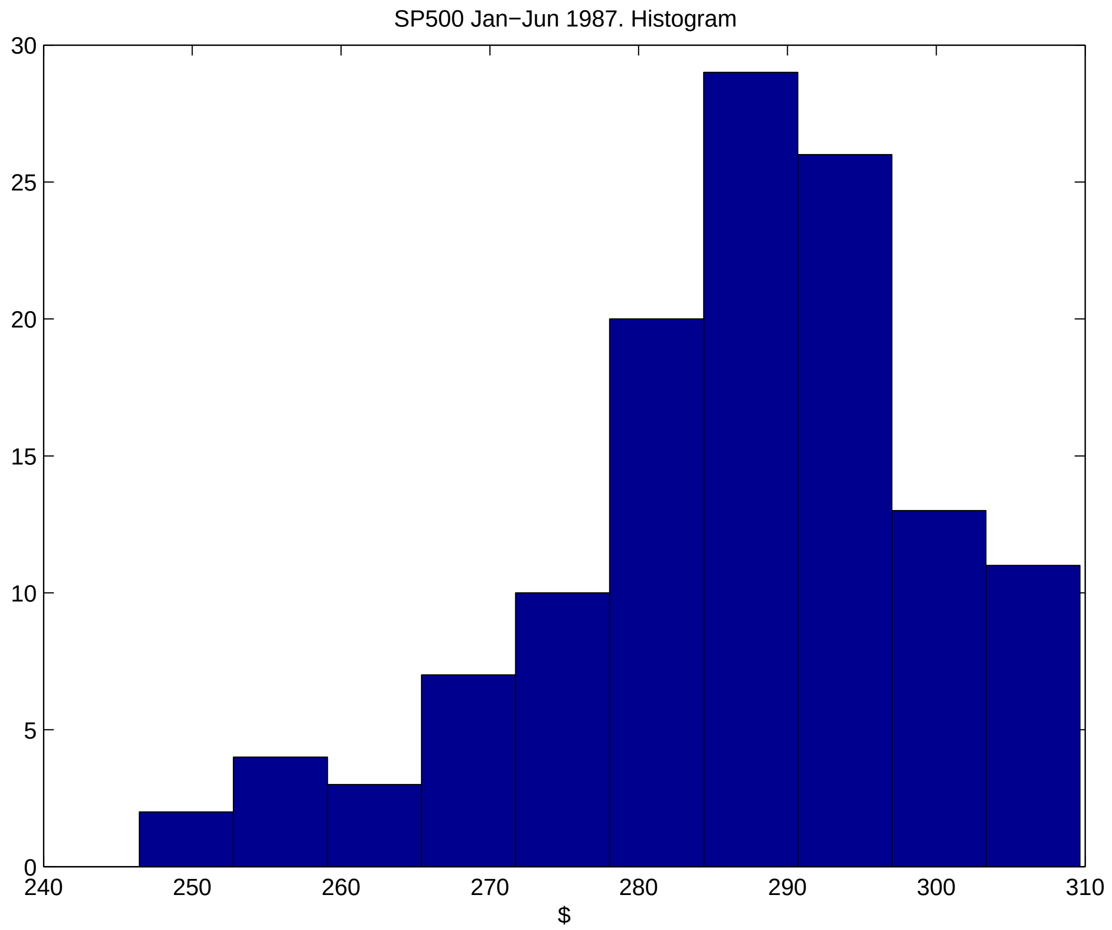
A Time Series



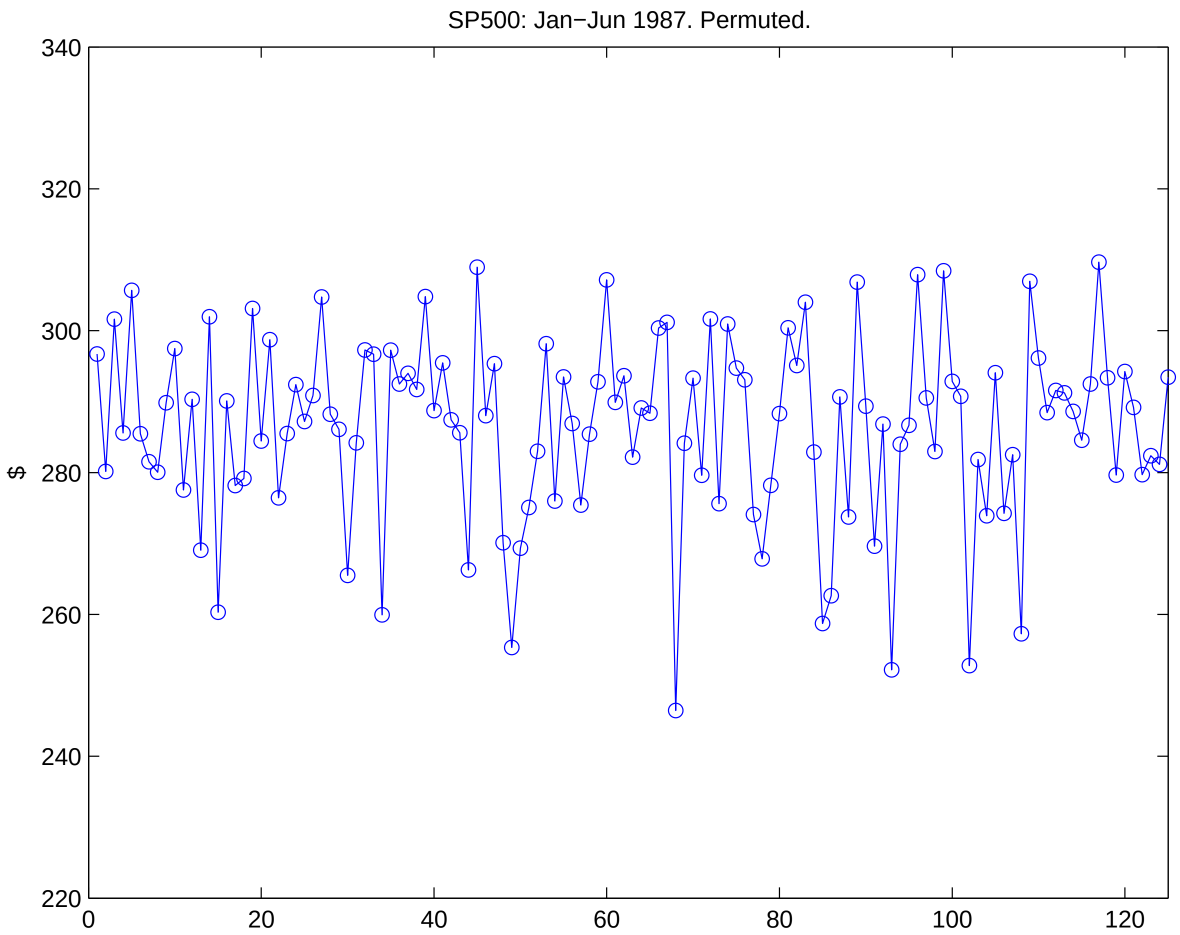
A Time Series



A Time Series



A Time Series



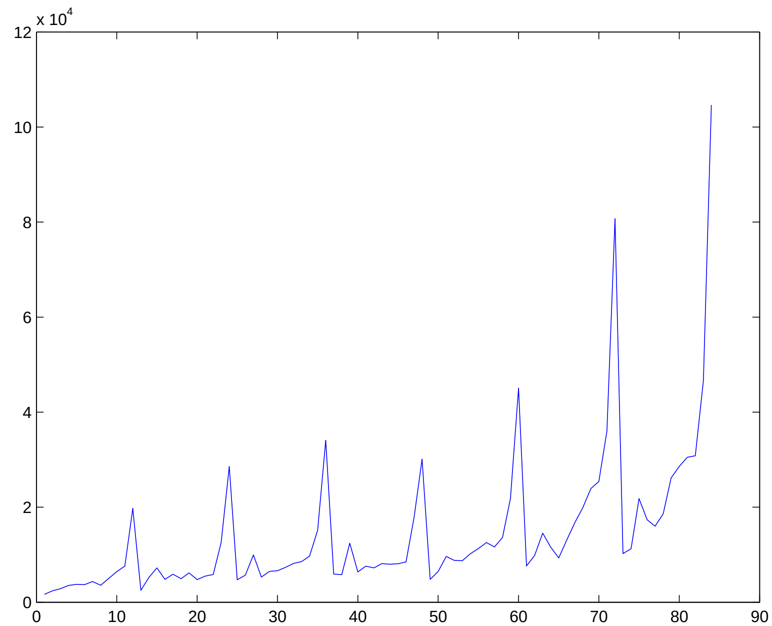
Objectives of Time Series Analysis

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

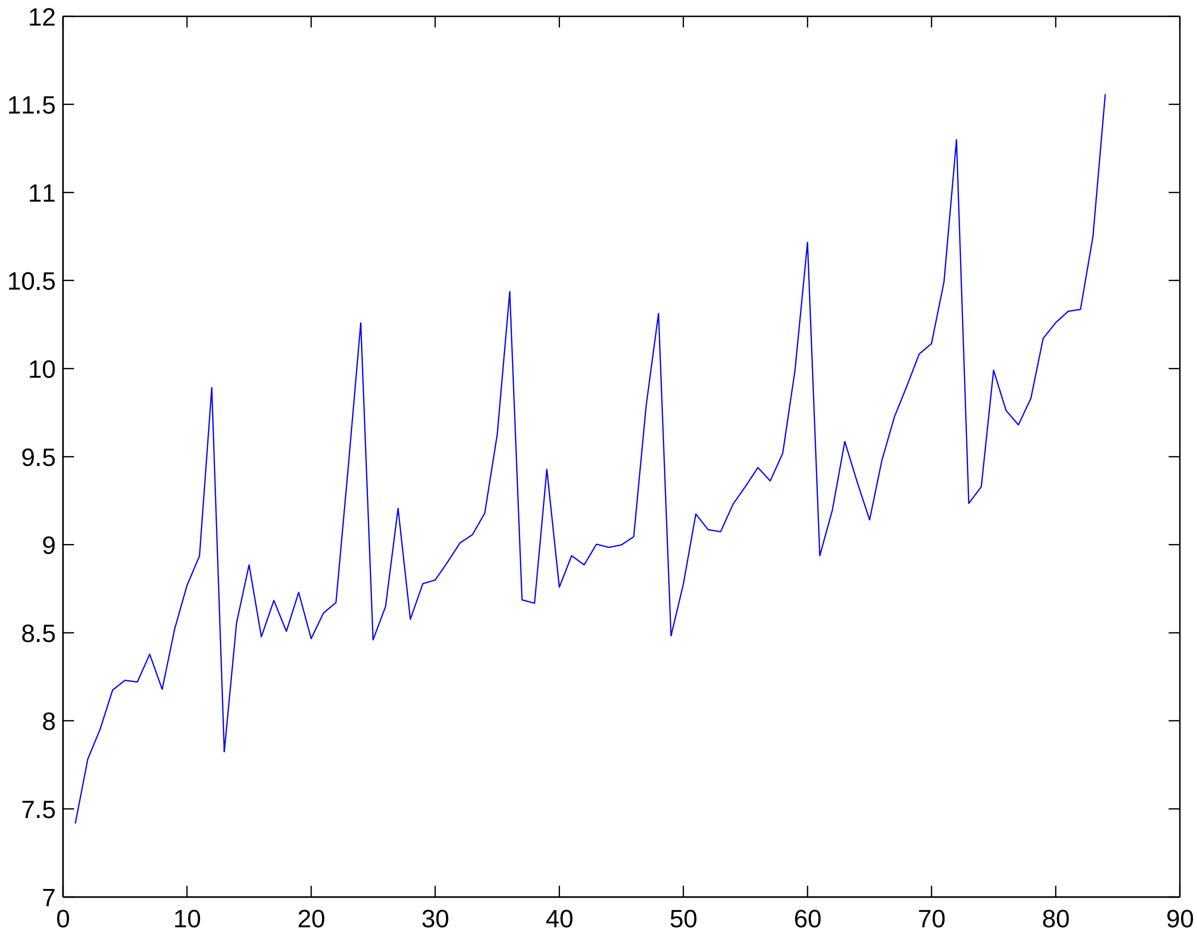
Classical decomposition: An example

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

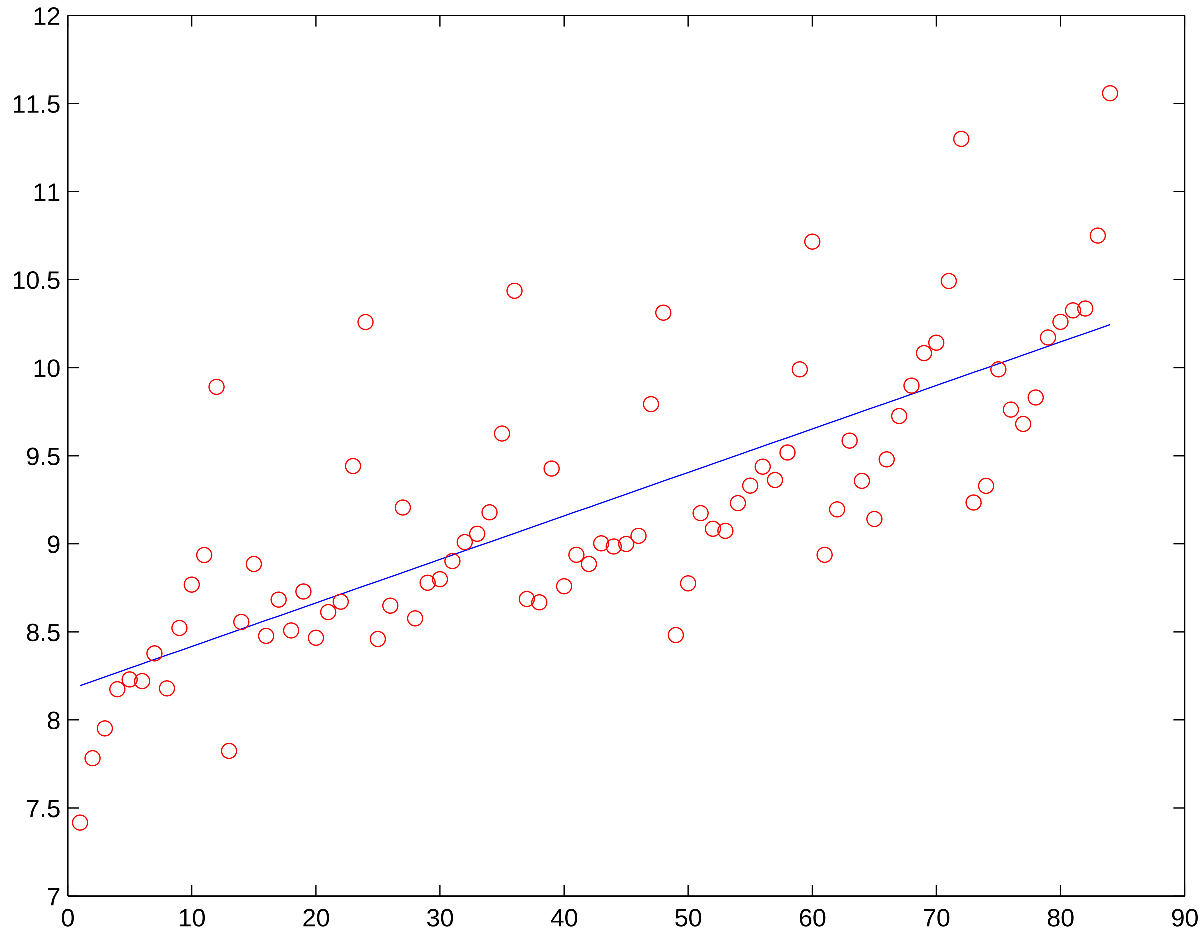
(Makridakis, Wheelwright and Hyndman, 1998)



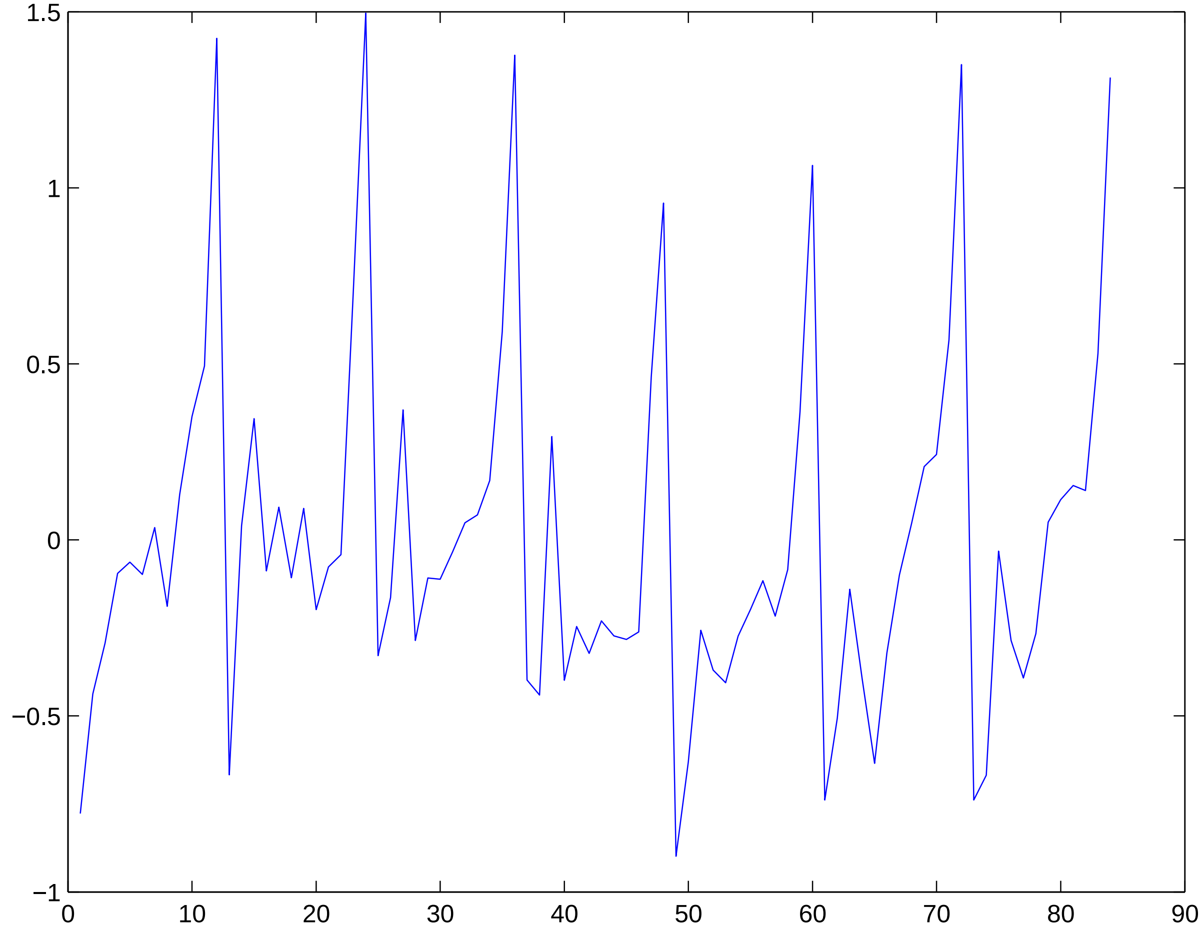
Transformed data



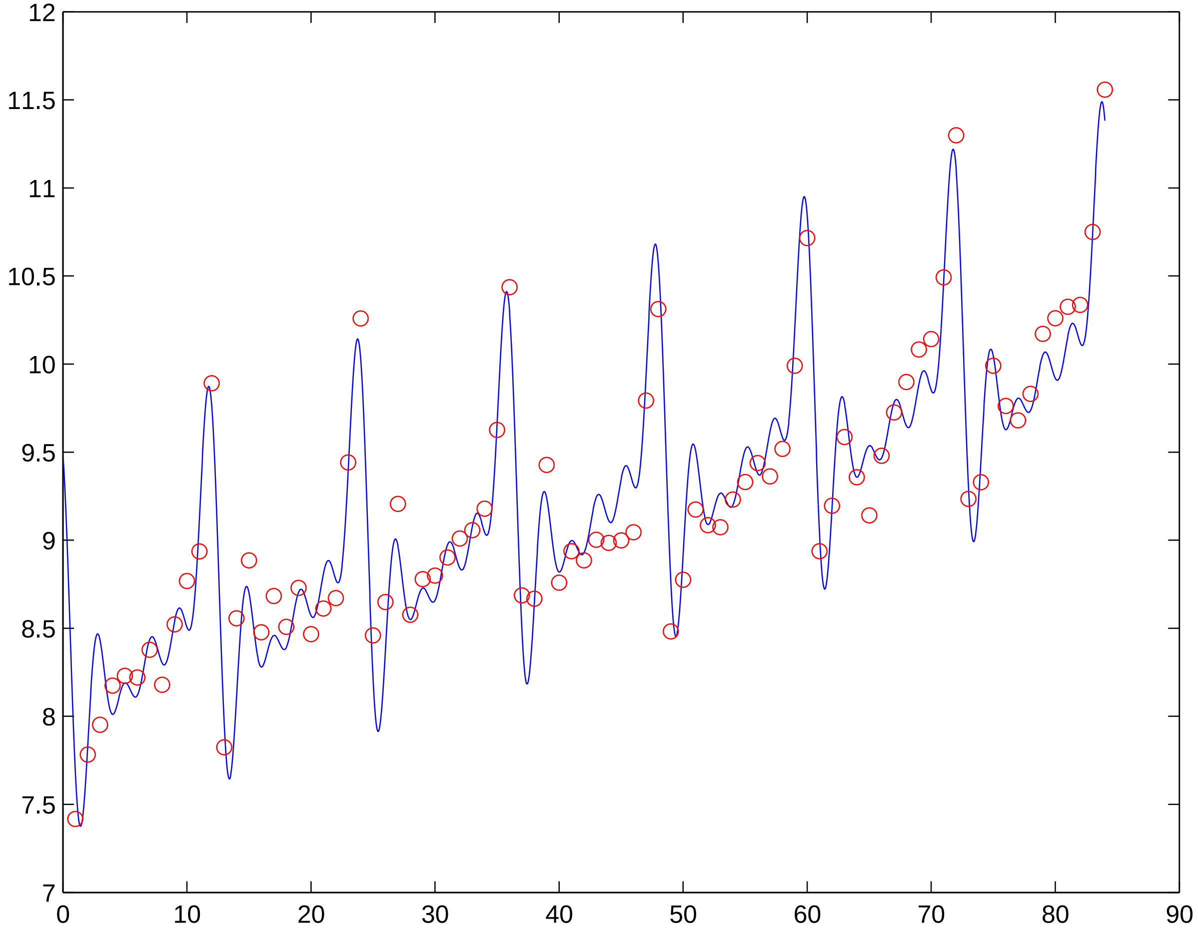
Trend



Residuals



Trend and seasonal variation



Objectives of Time Series Analysis

1. Compact description of data.
   Example: Classical decomposition: $X_t = T_t + S_t + Y_t$.

2. Interpretation. Example: Seasonal adjustment.

3. Forecasting. Example: Predict sales.

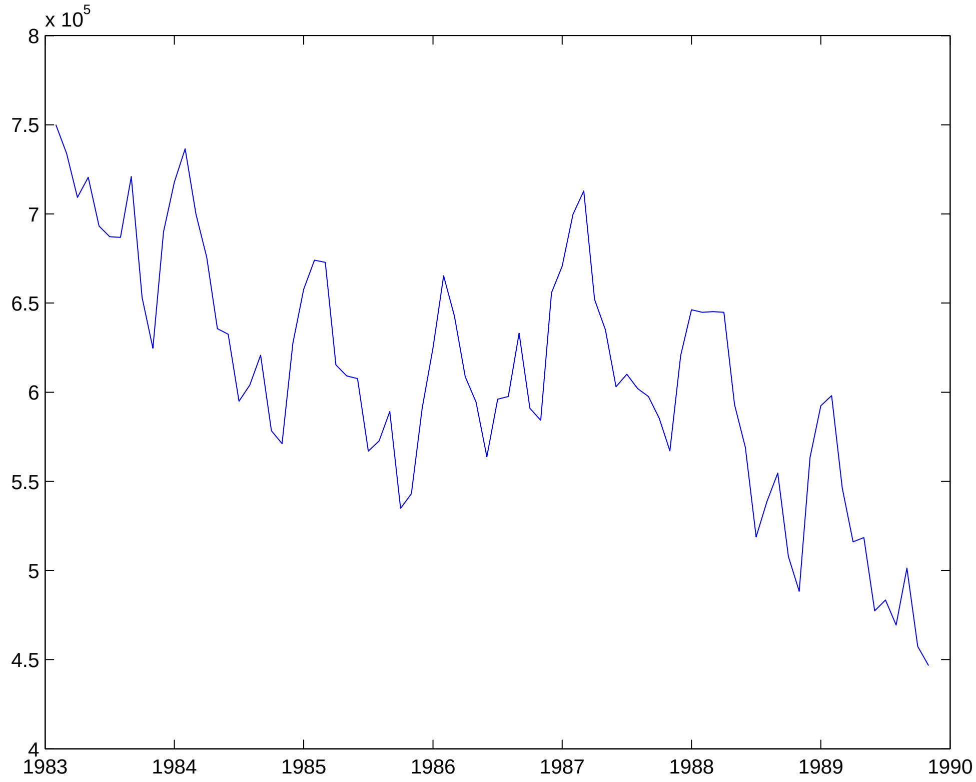
4. Control.

5. Hypothesis testing.

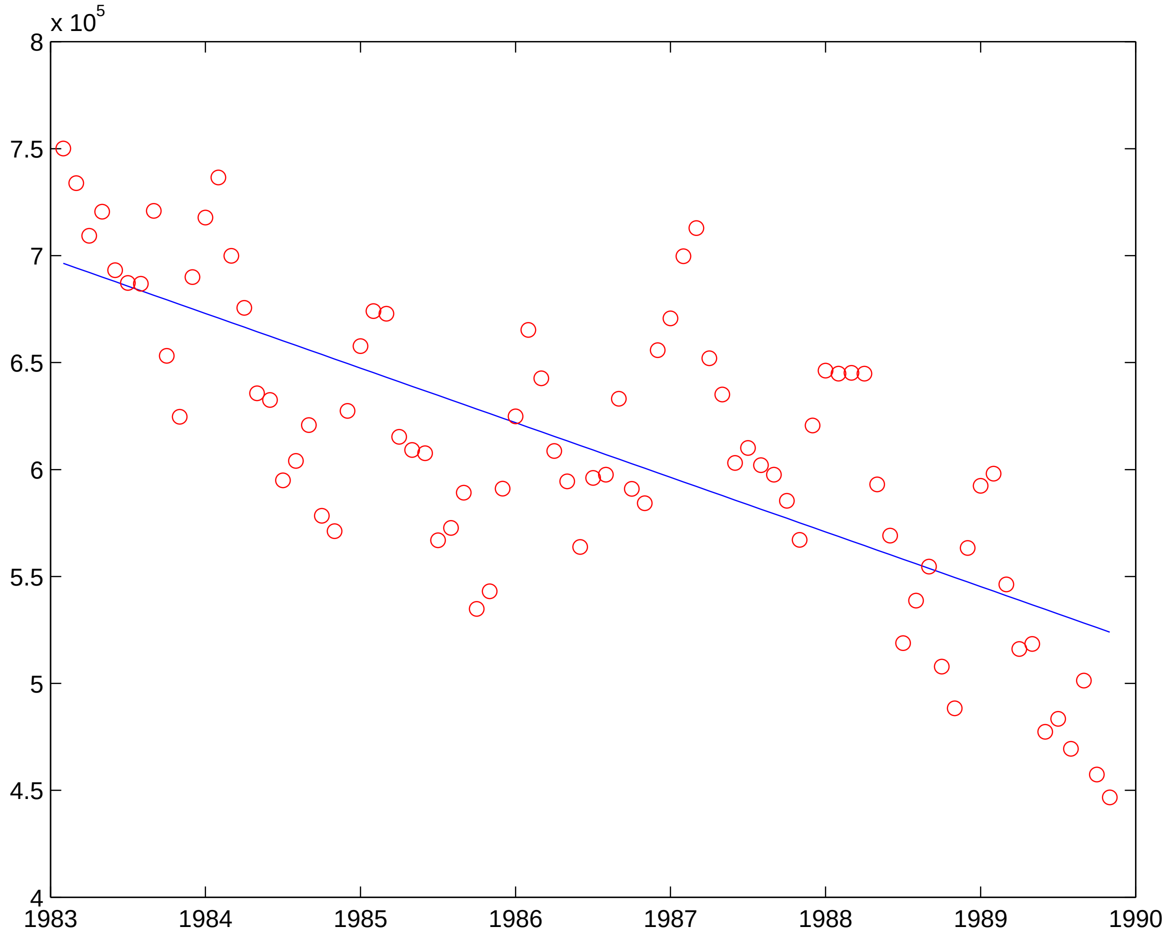
6. Simulation.

Unemployment data

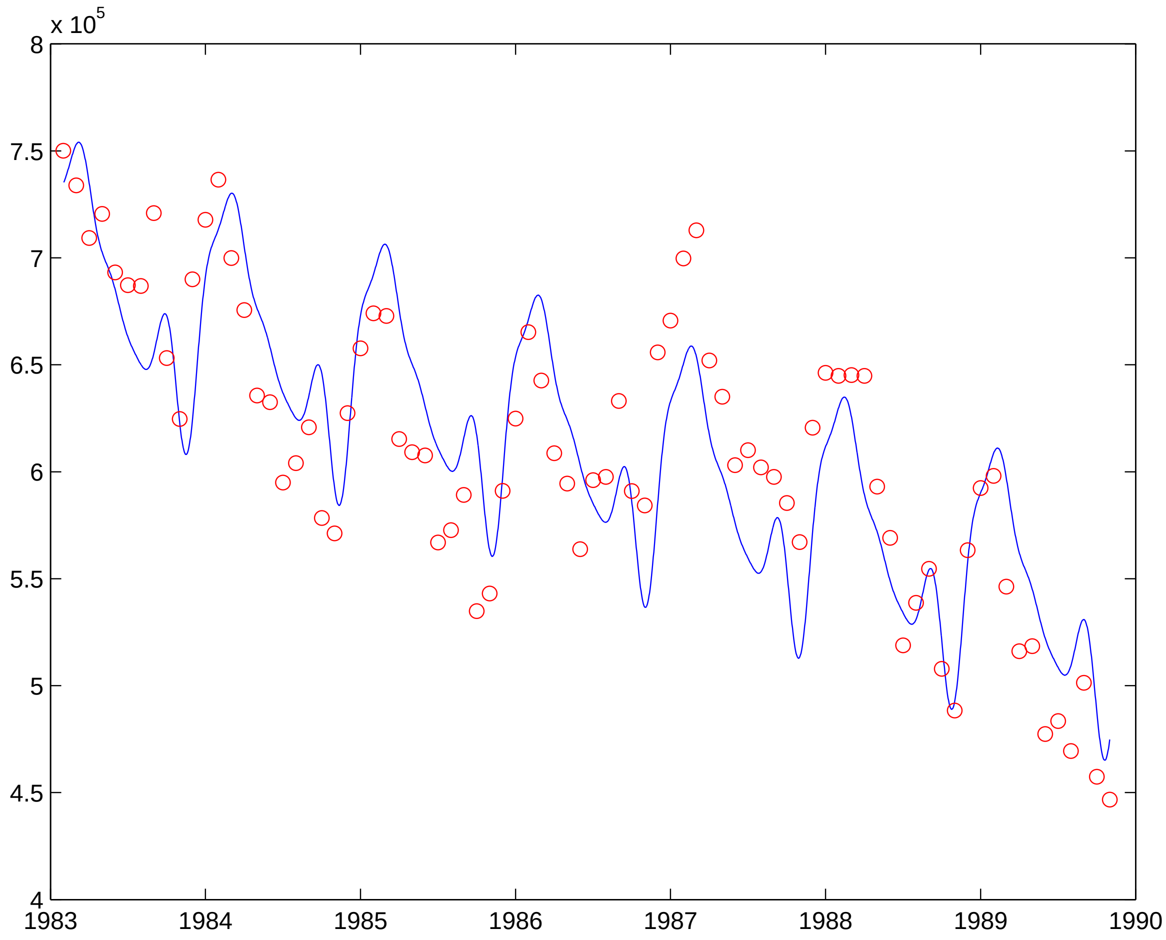
Monthly number of unemployed people in Australia. (Hipel and McLeod, 1994)



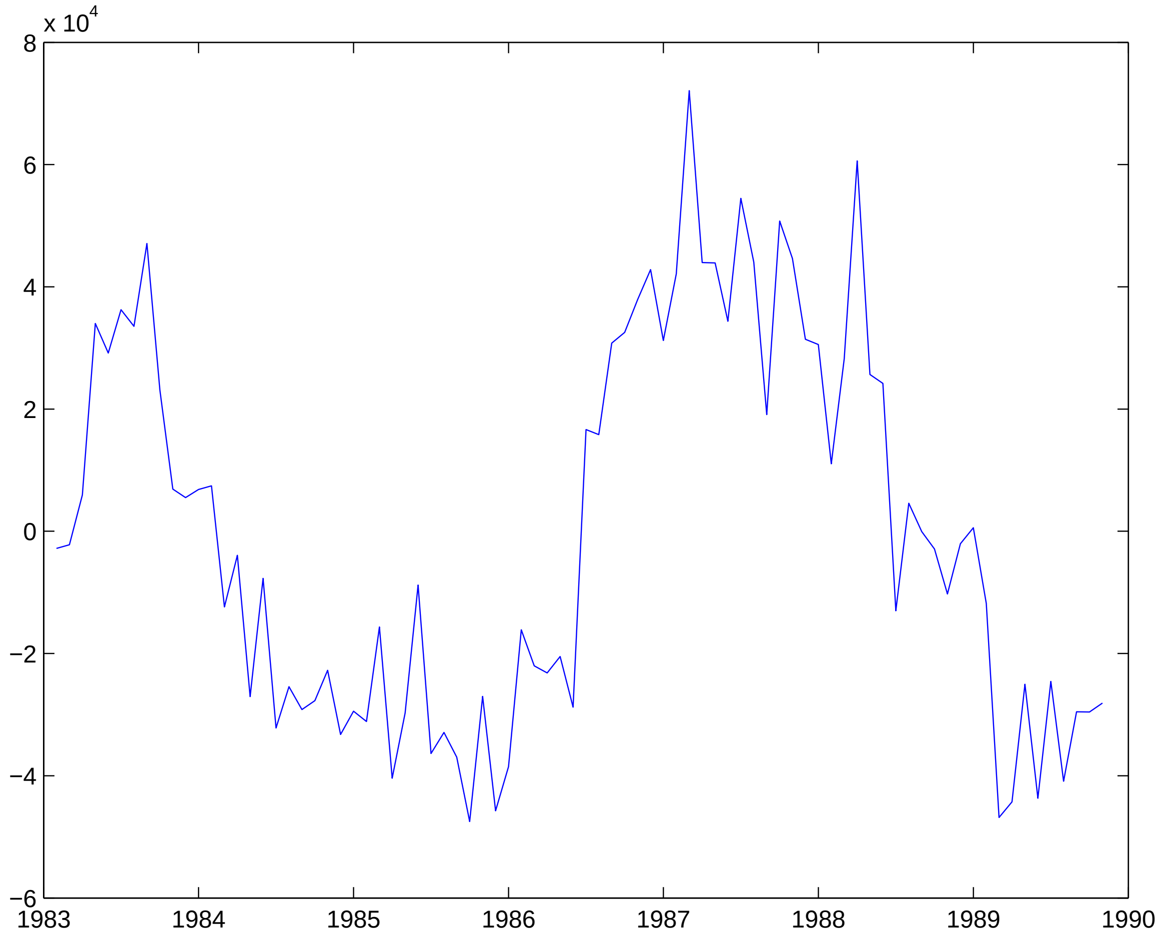
Trend



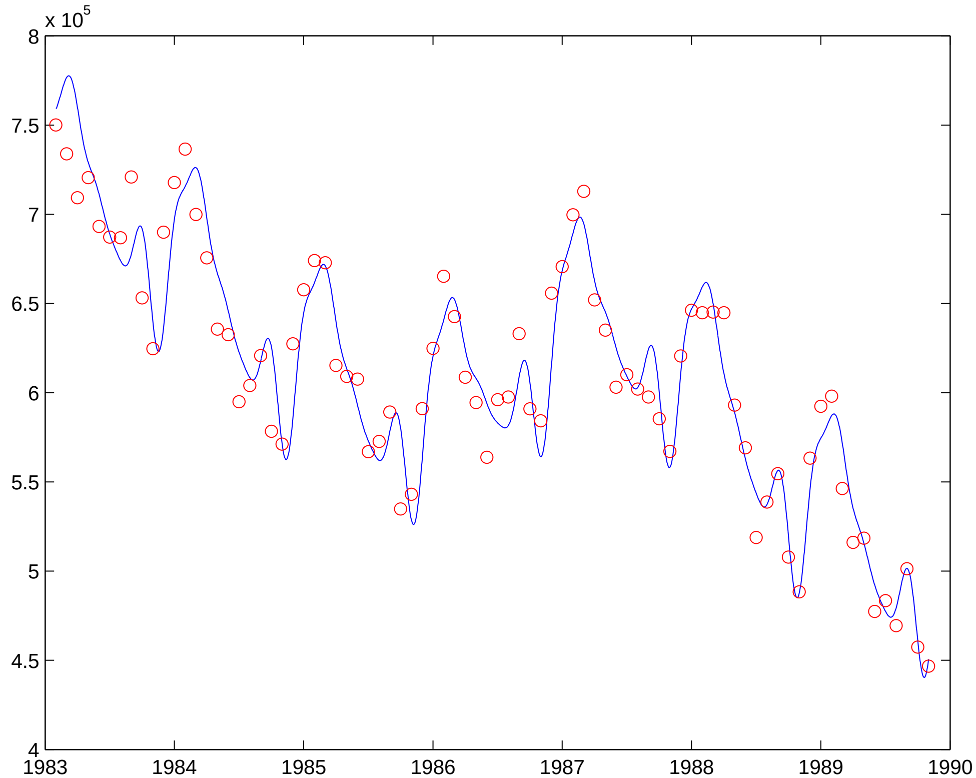
Trend plus seasonal variation



Residuals



Predictions based on a (simulated) variable



Objectives of Time Series Analysis

1. Compact description of data:

$$X_t = T_t + S_t + f(Y_t) + W_t.$$

2. Interpretation. Example: Seasonal adjustment.
3. Forecasting. Example: Predict unemployment.
4. Control. Example: Impact of monetary policy on unemployment.
5. Hypothesis testing. Example: Global warming.
6. Simulation. Example: Estimate probability of catastrophic events.

Overview of the Course

1. Time series models
2. Time domain methods
3. Spectral analysis
4. State space models(?)

Overview of the Course

1. Time series models
   (a) Stationarity.
   (b) Autocorrelation function.
   (c) Transforming to stationarity.
2. Time domain methods
3. Spectral analysis
4. State space models(?)

Overview of the Course

1. Time series models
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
4. State space models(?)

Overview of the Course

1. Time series models
2. Time domain methods
3. Spectral analysis
   (a) Spectral density
   (b) Periodogram
   (c) Spectral estimation
4. State space models(?)

Overview of the Course

1. Time series models
2. Time domain methods
3. Spectral analysis
4. State space models(?)
   (a) ARMAX models.
   (b) Forecasting, Kalman filter.
   (c) Parameter estimation.

Time Series Models

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

For example:

$$P[X_1 \leq x_1, \ldots, X_t \leq x_t] \text{ for all } t \text{ and } x_1, \ldots, x_t.$$

Notation:

$X_1, X_2, \ldots$ is a stochastic process.

$x_1, x_2, \ldots$ is a single realization.

We'll mostly restrict our attention to **second-order properties** only:
$\mathrm{E}X_t, \mathrm{E}(X_{t_1}, X_{t_2})$.

Time Series Models

Example: White noise: $X_t \sim WN(0, \sigma^2)$.
i.e., $\{X_t\}$ uncorrelated, $\mathrm{E}X_t = 0$, $\mathrm{Var}X_t = \sigma^2$.

Example: i.i.d. noise: $\{X_t\}$ independent and identically distributed.

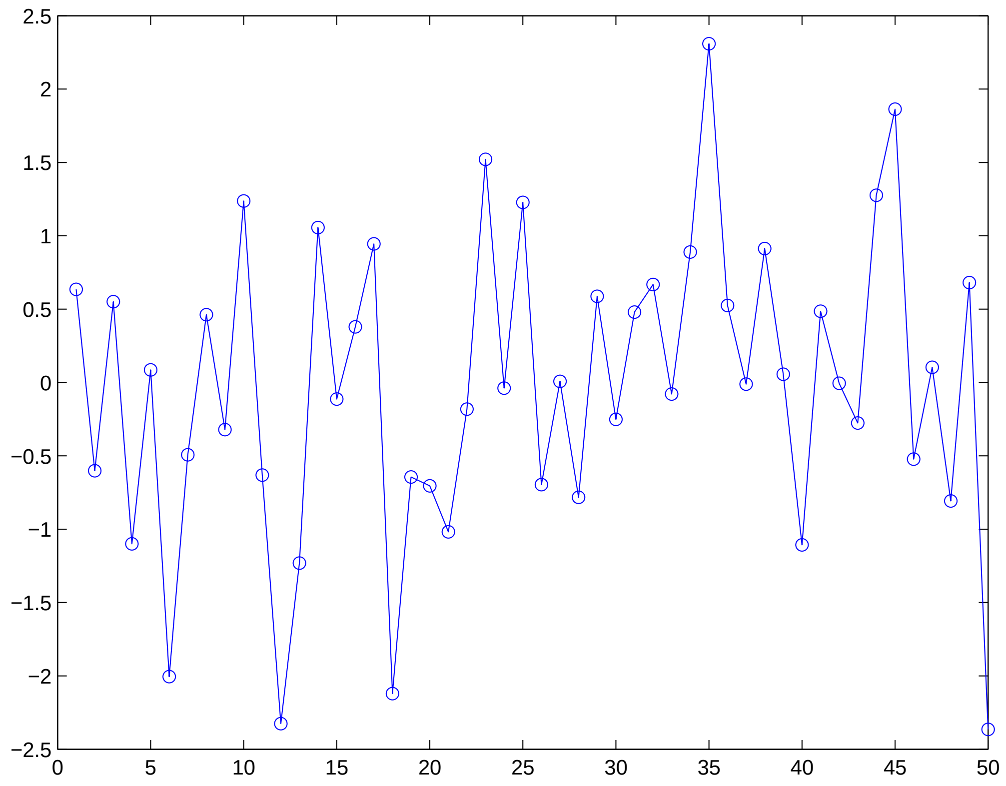
$$P[X_1 \le x_1, \ldots, X_t \le x_t] = P[X_1 \le x_1] \cdots P[X_t \le x_t].$$

Not interesting for forecasting:

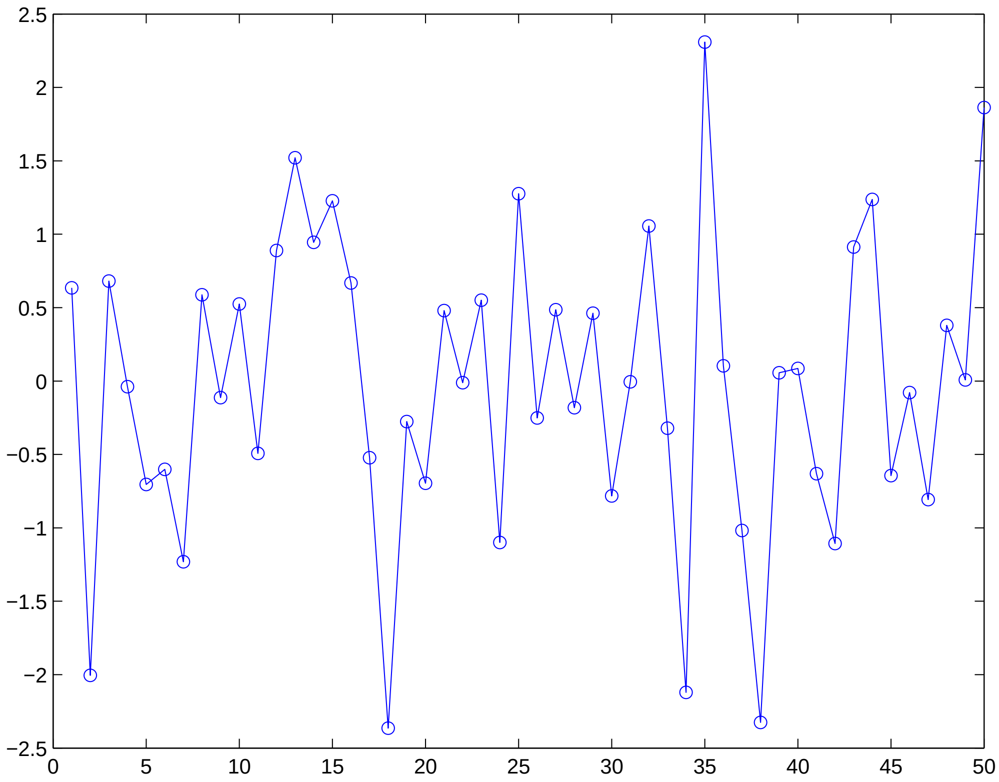
$$P[X_t \le x_t | X_1, \ldots, X_{t-1}] = P[X_t \le x_t].$$

Gaussian white noise

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

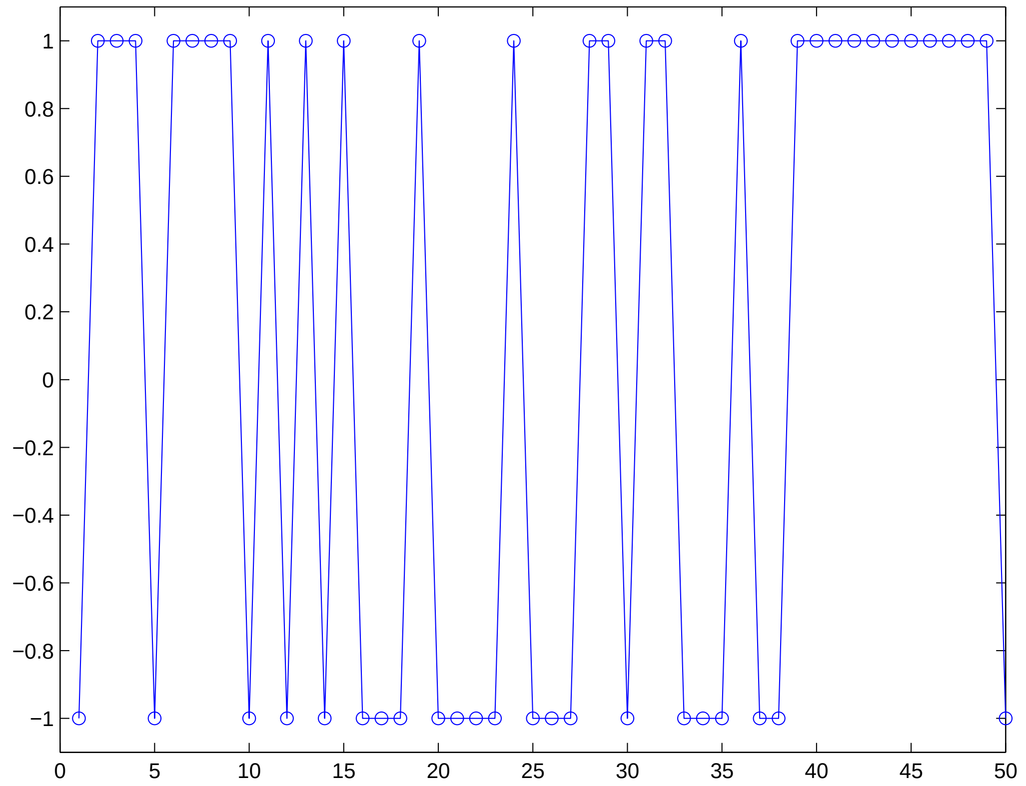


Gaussian white noise



Time Series Models

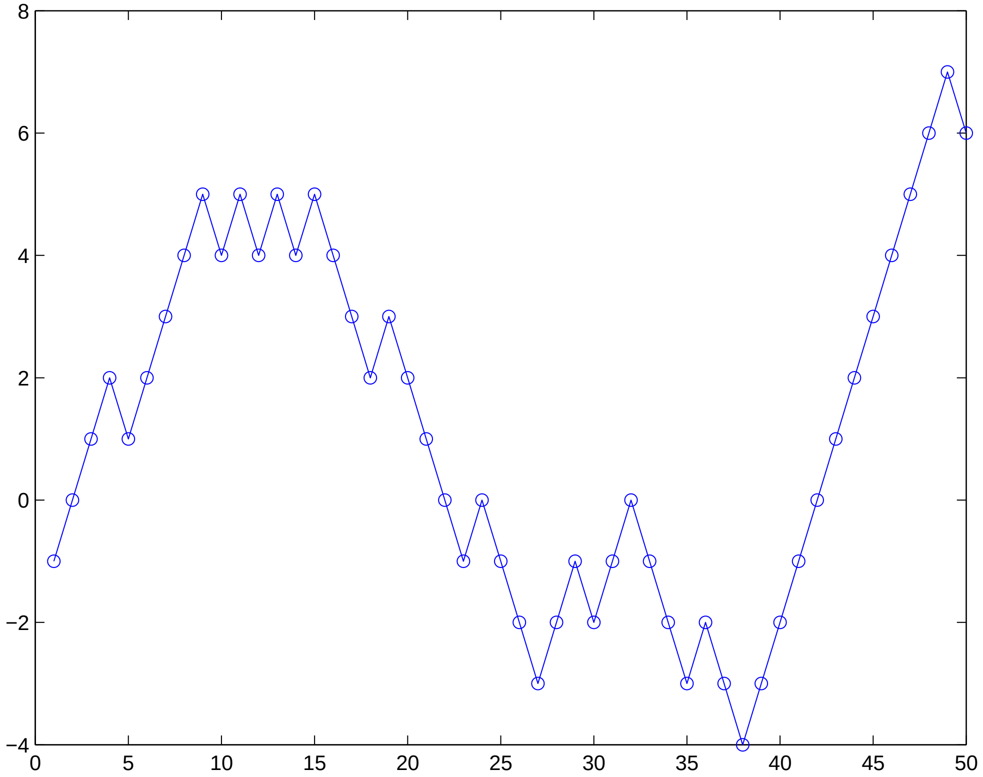
Example: Binary i.i.d. $P[X_t = 1] = P[X_t = -1] = 1/2.$



Random walk

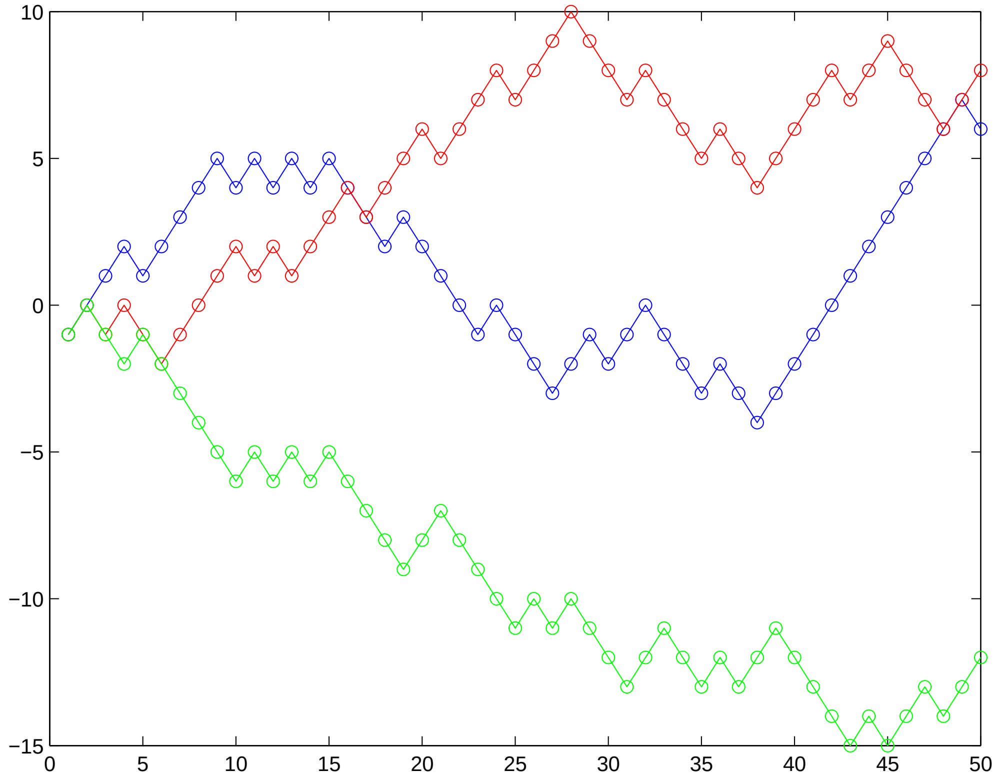
$S_t = \sum_{i=1}^{t} X_i.$

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



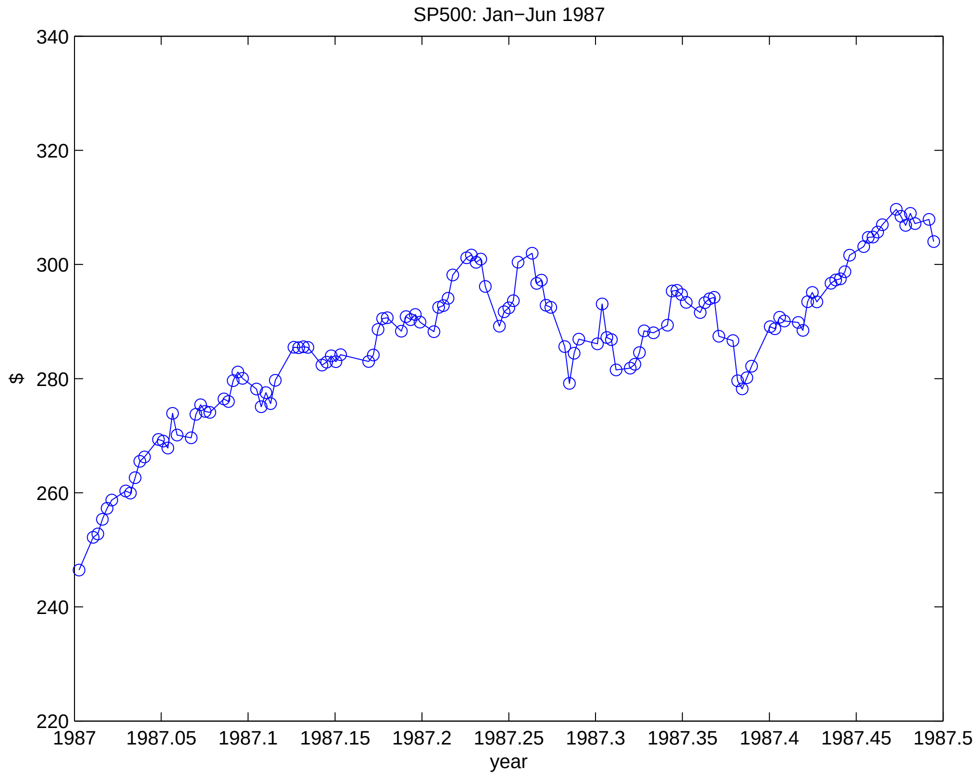
Random walk

ES_t? $\mathrm{Var}S_t$?



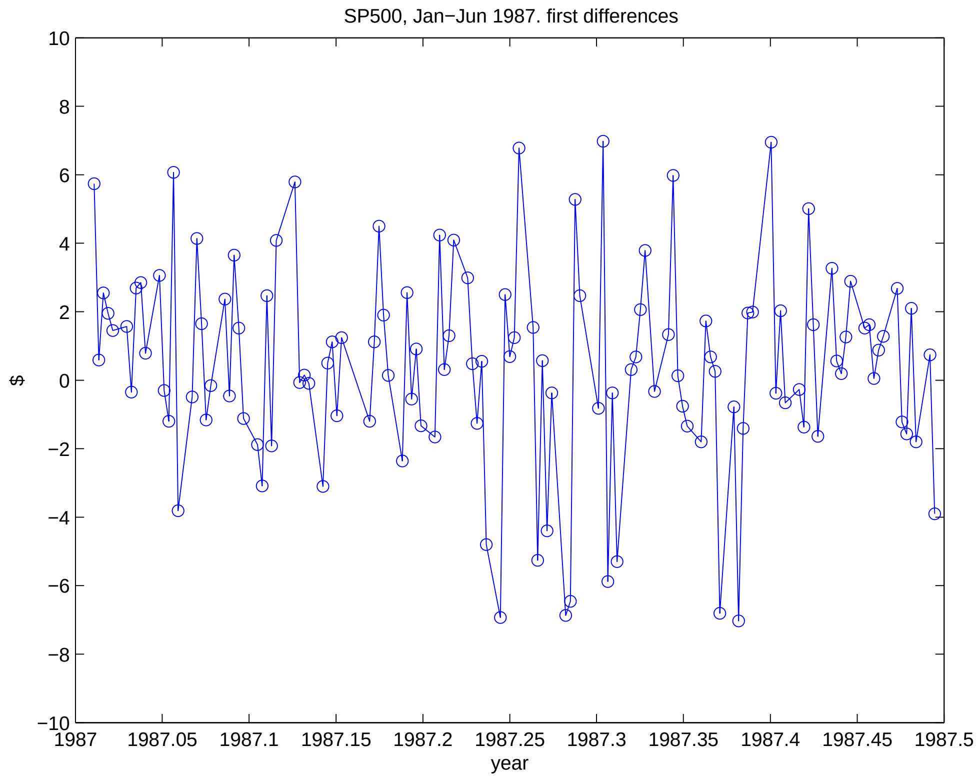
Random Walk

Recall S&P500 data. (Notice that it's smooth)



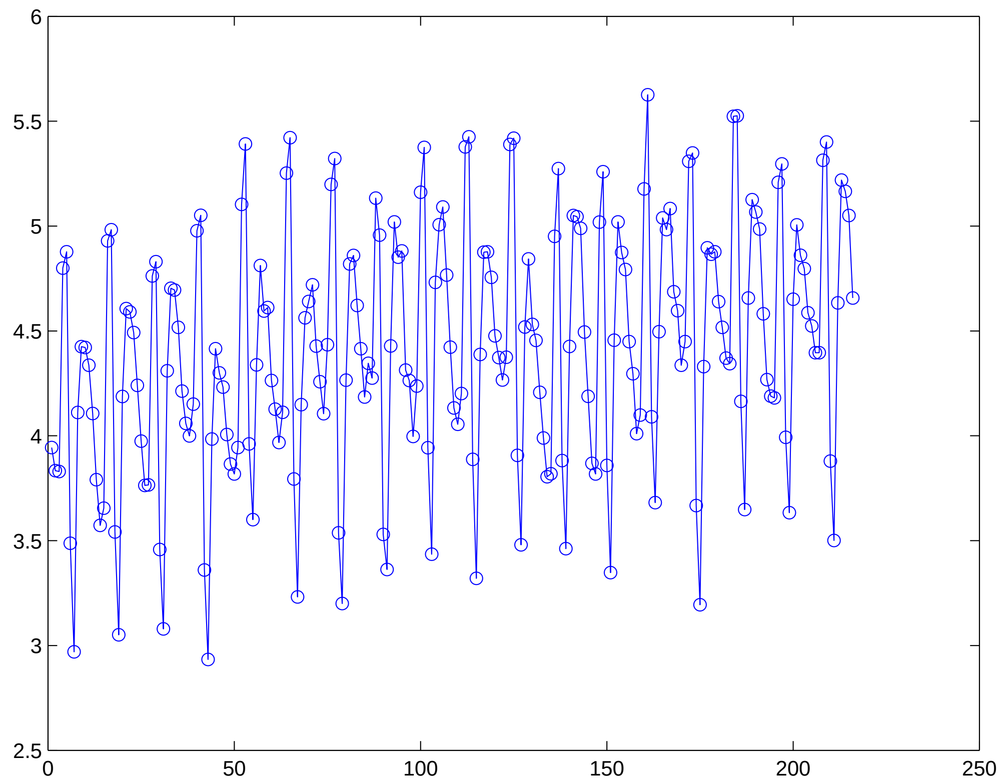
Random Walk

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



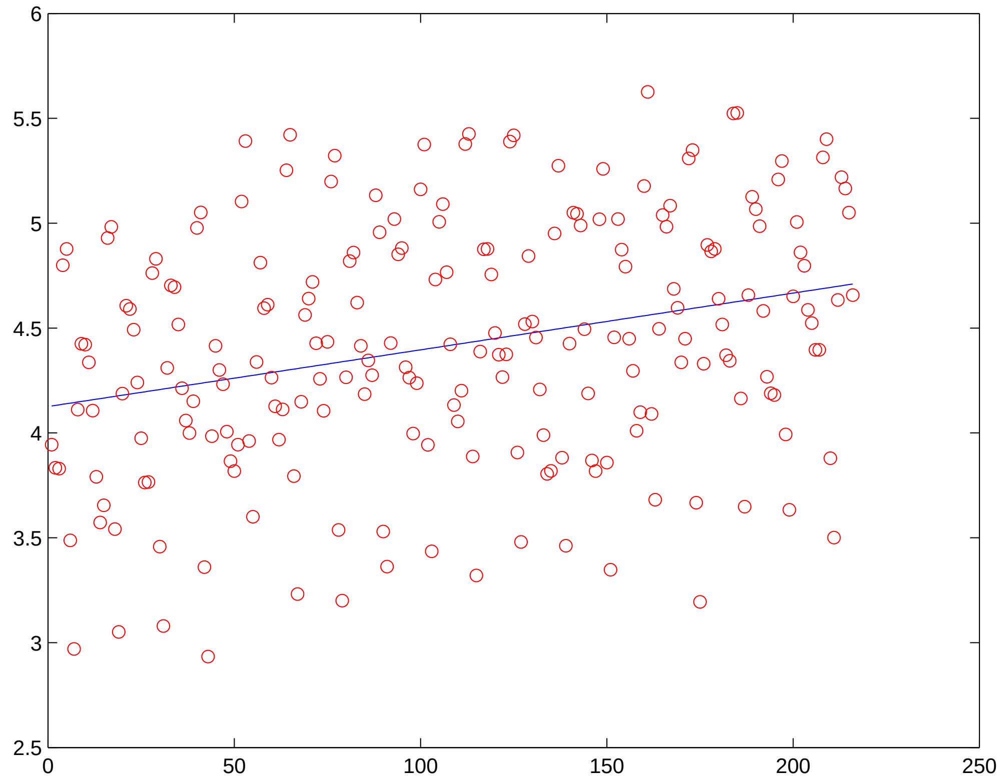
Trend and Seasonal Models

$$X_t = T_t + S_t + E_t = \beta_0 + \beta_1 t + \sum_i \left(\beta_i \cos(\lambda_i t) + \gamma_i \sin(\lambda_i t)\right) + E_t$$



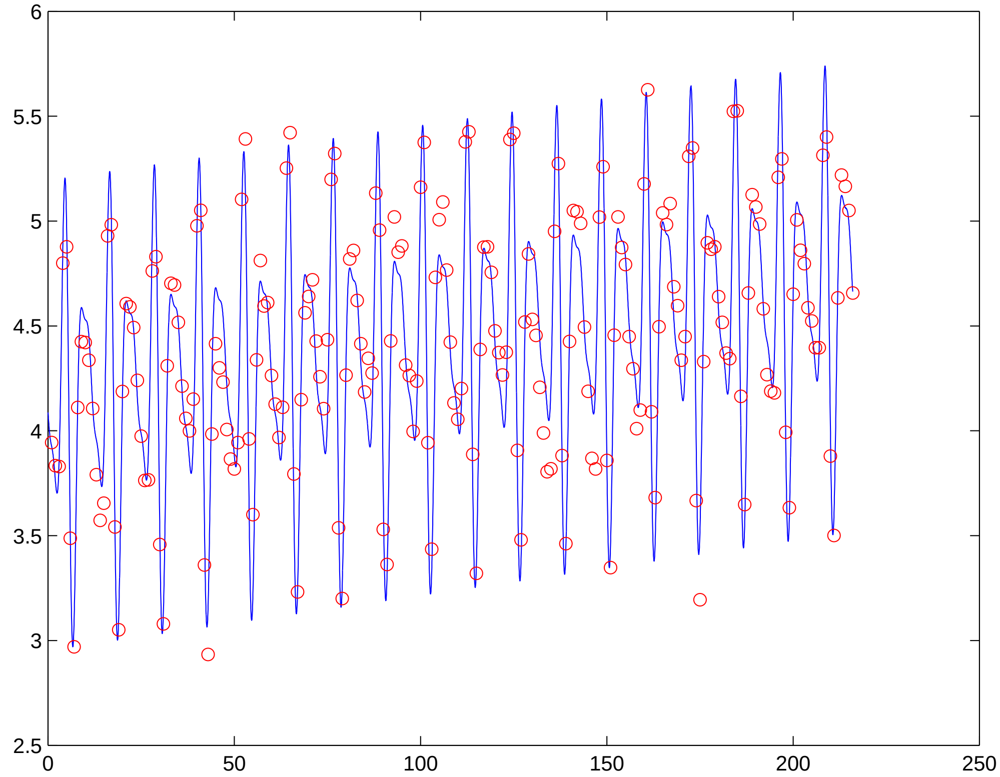
Trend and Seasonal Models

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

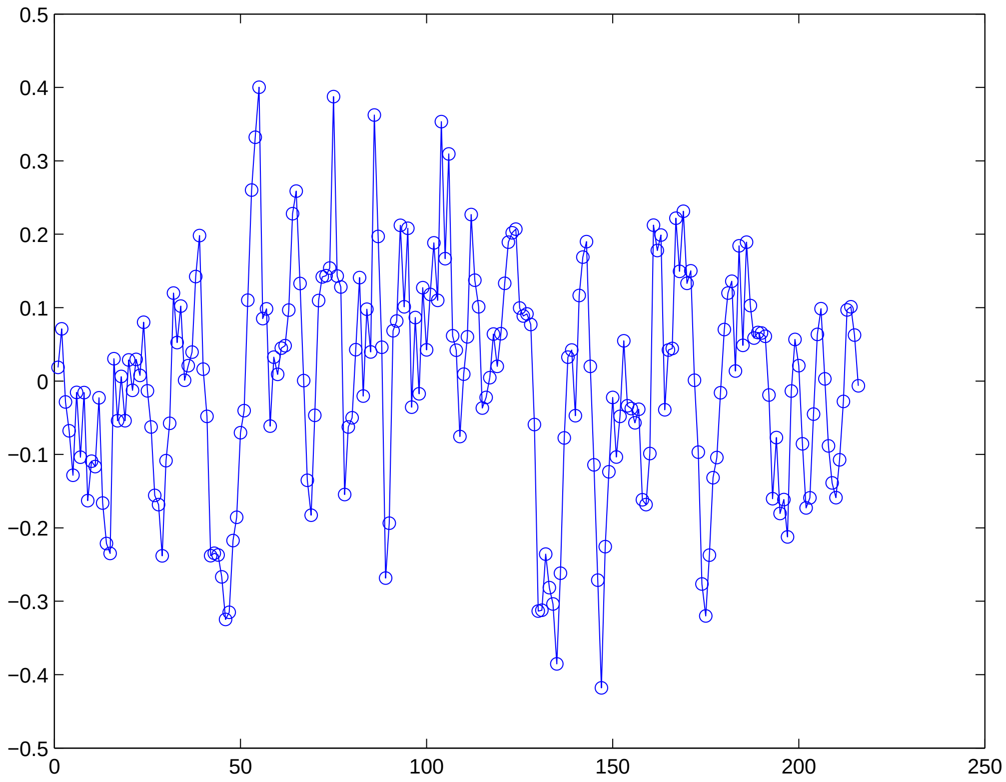


Trend and Seasonal Models

$$X_t = T_t + S_t + E_t = \beta_0 + \beta_1 t + \sum_i \left(\beta_i \cos(\lambda_i t) + \gamma_i \sin(\lambda_i t)\right) + E_t$$



Trend and Seasonal Models: Residuals



Time Series Modelling

1. Plot the time series.

   Look for trends, seasonal components, step changes, outliers.

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

   (a) Estimate and subtract T_t, S_t.

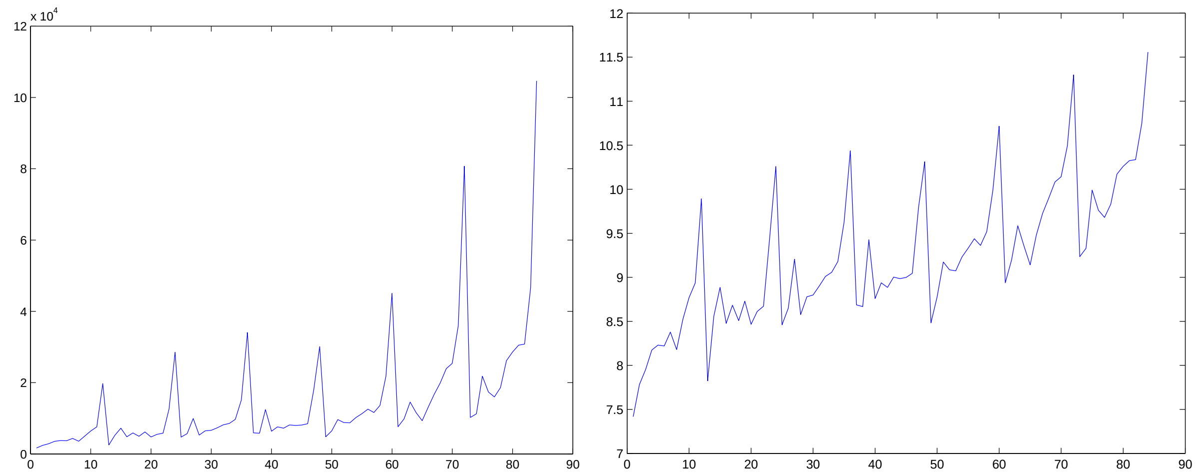
   (b) Differencing.

   (c) Nonlinear transformations (log, $\sqrt{\cdot}$).

3. Fit model to residuals.

Nonlinear transformations

Recall: Monthly sales. (Makridakis, Wheelwright and Hyndman, 1998)



Time Series Modelling

1. Plot the time series.

   Look for trends, seasonal components, step changes, outliers.

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

   (a) Estimate and subtract T_t, S_t.

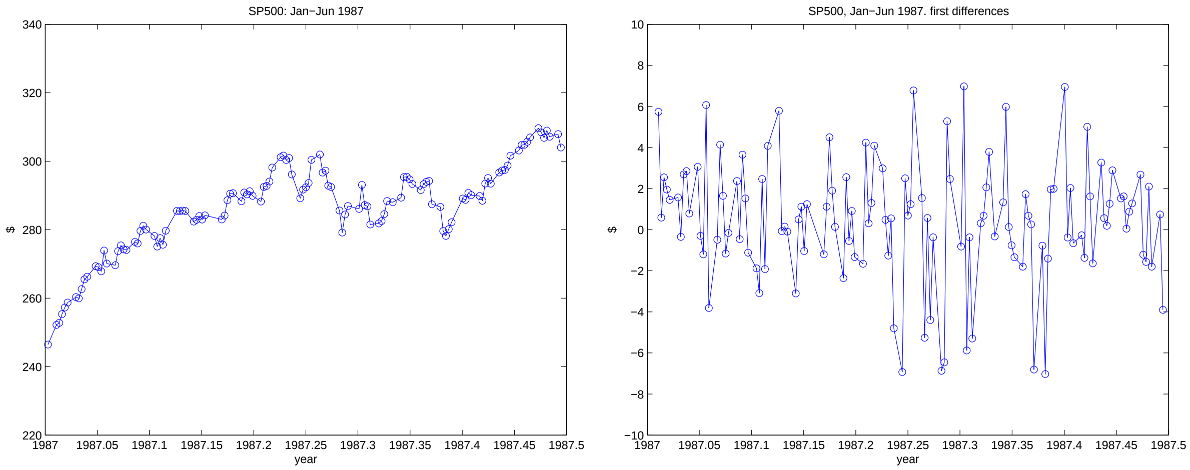
   (b) Differencing.

   (c) Nonlinear transformations (log, $\sqrt{\cdot}$).

3. Fit model to residuals.

Differencing

Recall: 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 $X_t = \sum_{i=0}^{k} \beta_i t^i + Y_t$, then

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

  where $\nabla^k X_t = \nabla(\nabla^{k-1} X_t)$ and $\nabla^1 X_t = \nabla X_t$.

Differencing and Seasonal Variation

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, $B^s X_t = B(B^{s-1} X_t)$ and $B^1 X_t = B X_t$.

If $X_t = T_t + S_t + Y_t$, and S_t has period s (that is, $S_t = S_{t-s}$ for all t), then

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

Time Series Modelling

1. Plot the time series.

   Look for trends, seasonal components, step changes, outliers.

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

   (a) Estimate and subtract T_t, S_t.

   (b) Differencing.

   (c) Nonlinear transformations (log, $\sqrt{\cdot}$).

3. Fit model to residuals.

Outline

1. Objectives of time series analysis. Examples.
2. Overview of the course.
3. Time series models.
4. Time series modelling: Chasing stationarity.