Introduction to time series analysis

- Introduction to time series
- De-trend: regression analysis
- Discrete Fourier analysis

What is a time series?

- A series of data points organized in time order
- Theoretically, time series can be observations of a contnuous function of variable time t.
  In practice, time series data are discrete series taken at equally spacing time intervals.
  If collected data points are not equally spacing in time, some interpolation may be performed at the preprocessing stage.
- Contrary to spatial data, time series data close in time are more closely related than data further apart.

Objectives of time series analysis

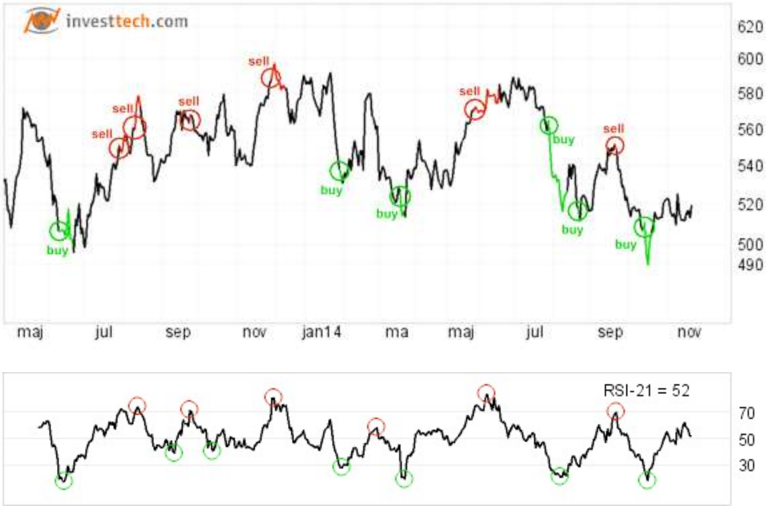
- Extract meaningful information, statistics and other characteristics, e.g. trends and patterns
- Understand potential mechanism which affects the series
- Use previous observed data to develop a model in order to forecast (or predict) future values in longer time frame (or in short-term, near future)
- Investigate correlations between two or more different time series, instead of studying relationships between different data points in time within a single series

Applications of time series analysis

- Natural science, seismology, geophysics, geodesy, signa processing, real-time navigation, climate change, etc etc
- Social sciences, statistics of various kinds
  opinion poll, economical data, inflation, etc etc
- Financial analysis (financial mathematics): macro- / micro-economics
- Prediction of stock prices: foundamental analysis vs *technical analysis*

Lars E.O. Svensson
MSc in Mathematics, KTH 1971
PhD in Economics, MIT 1975
Deputy Governor, Riksbanken
(2007-2013)

Technical analysis of stock prices using RSI (Relative Strength Index)



RSI (Relative Strength Index)

Closing prices of a stock: $x_i,\ i = 1,\ 2,\ 3,\ \ldots,\quad N$

$$if \quad x_i > x_{i-1} \qquad u_i = x_i - x_{i-1} \qquad d_i = 0$$

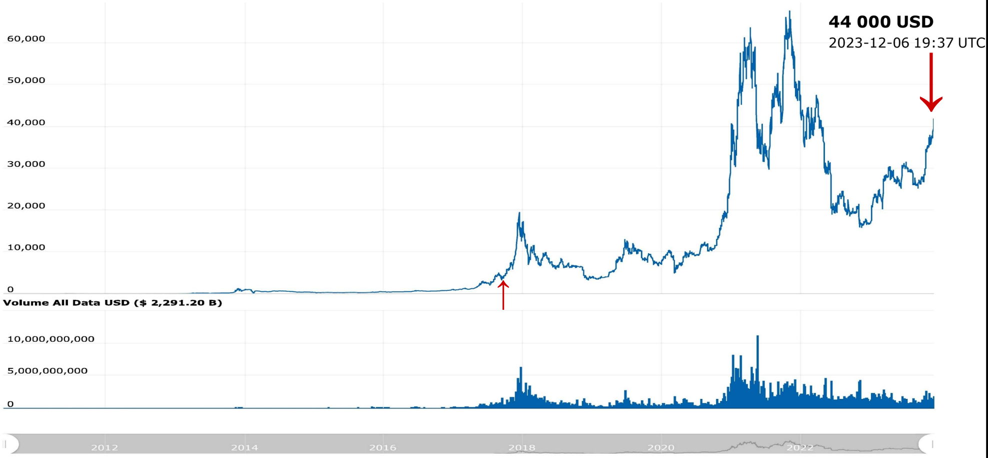
$$if \quad x_i < x_{i-1} \qquad u_i = 0 \qquad d_i = x_{i-1} - x_i$$

$$S_u = \sum_{i=2}^{N} u_i$$

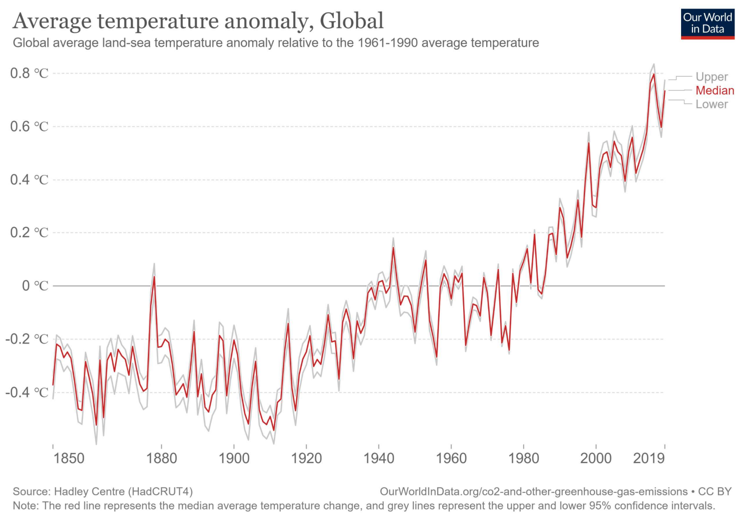
$$S_d = \sum_{i=2}^{N} d_i$$

$$\boxed{RSI = \frac{S_u}{S_u + S_d} \times 100 \quad \text{(in percent)}}$$

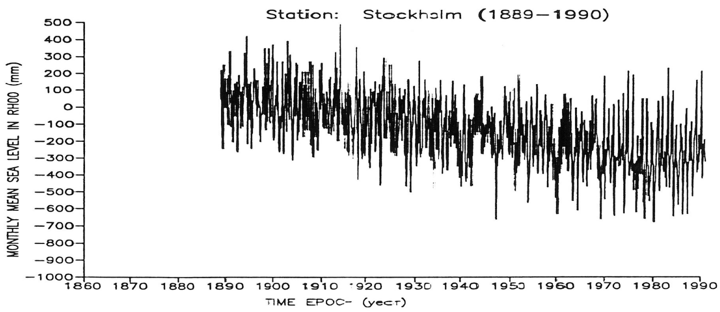
Bitcoin price in USD (2011-2023)



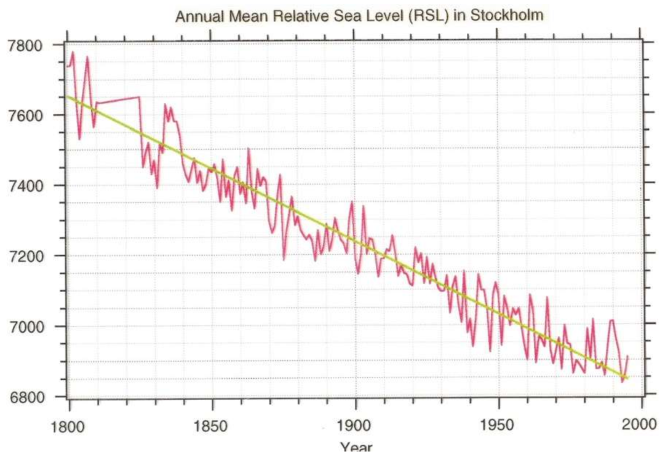
Global warming 1885-2019



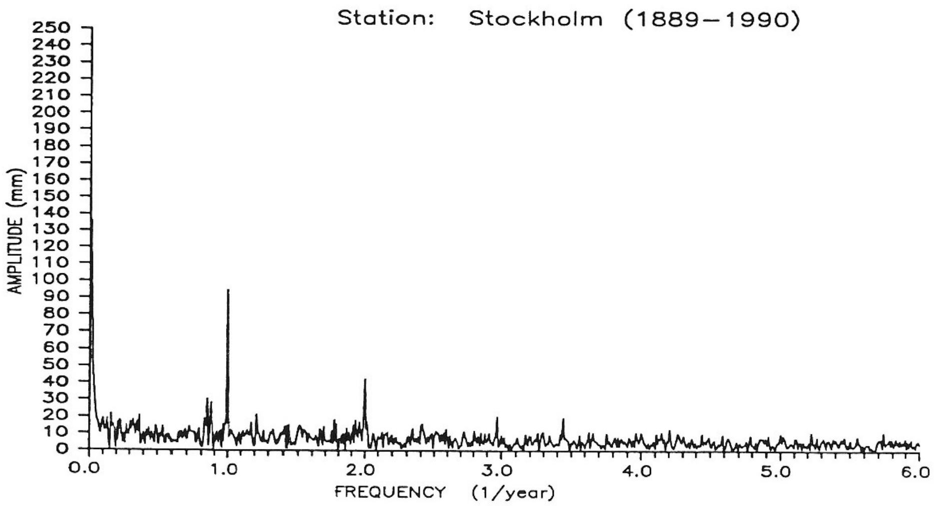
Relative sea level and land uplift



Relative sea level and land uplift



Spectral analysis of relative sea level



Coordinate time series at SWEPOS stations

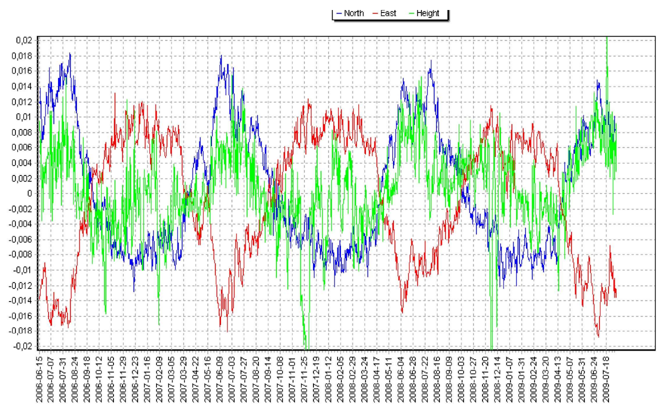


Figure 3: Example time serie analysis of Class B SWEPOS station Älvsbyn

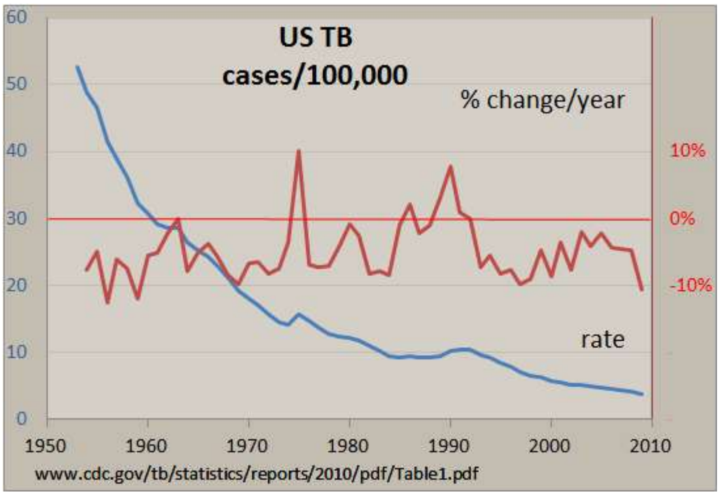
Visualization of a SWEPOS Coordinate Analysis

Axel Viking Bronder

Degree Project in Built Environment, First Level

Stockholm 2011
KTH, Department of Urban Planning and Environment
Division of Geodesy and Geoinformatics
Kungliga Tekniska högskolan

TB cases and change per year in US



Types of time series

- Characteristics of time series:
  - Trends or other systematic variations
  - Periodic oscilations: with fixed or varing patterns
  - Irregularity: unexpected situations or events or scenarios, spikes in short time span
- Stationary time series (*stationary stochastic process*):
  - Constant mean
  - Constant variance
  - Constant covariance in case of two series
- White noise:

  uniform uncorrelated normally distributed variations

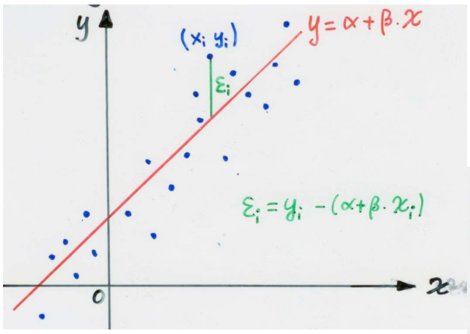
General procedure of time series analysis

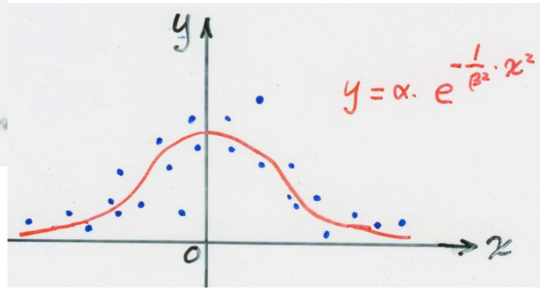
- Collect and pre-process data
  quality control, data filling by interpolation, etc
- Plott the data for possible features
- Identify and remove possible trends (de-trend)
- Modelling
- Extract insights for prediction

Methods for time series analysis

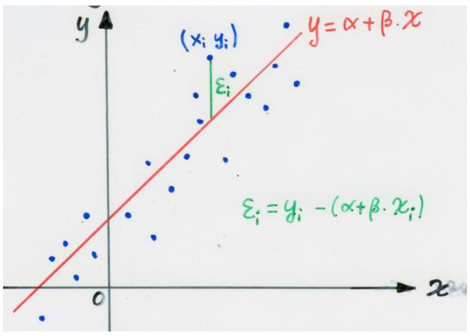
- Visualization: simple line plot
- Frequency-domain: spectral analysis / Fourier transform
- Time-domain: covariance,
  auto-correlation, cross-correlation
- Parametric method: moving average,
  auto-regression
- Non-parametric method: covariance, spectrum

Visualization of time series





De-trend by linear regression



$$y_i = \alpha + \beta \cdot x_i + \varepsilon_i$$

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \alpha - \beta \cdot x_i)^2 = \text{minimum}$$

$$
\begin{aligned}
q_{11} &= n \\
q_{12} &= \textstyle\sum_{i=1}^{n} x_i \\
q_{22} &= \textstyle\sum_{i=1}^{n} x_i^2 \\
w_1 &= \textstyle\sum_{i=1}^{n} y_i \\
w_2 &= \textstyle\sum_{i=1}^{n} x_i y_i
\end{aligned}
$$

$$\begin{bmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{bmatrix} = \frac{1}{q_{11} q_{22} - q_{12}^2} \begin{bmatrix} q_{22} & -q_{12} \\ -q_{12} & q_{11} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \frac{1}{q_{11} q_{22} - q_{12}^2} \begin{bmatrix} q_{22} w_1 - q_{12} w_2 \\ -q_{12} w_1 + q_{11} w_2 \end{bmatrix}$$

Removal of the linear trend

$$\boxed{\widehat{\varepsilon}_i = y_i - \widehat{\alpha} - \widehat{\beta} \cdot x_i}$$

$$\widehat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}\widehat{\varepsilon}_i^2 = \frac{1}{n-2}\sum_{i=1}^{n}\left(y_i - \widehat{\alpha} - \widehat{\beta}\cdot x_i\right)^2$$

$$\sigma_{\widehat{\alpha}} = \widehat{\sigma}\cdot\sqrt{\frac{q_{22}}{q_{11}q_{22}-q_{12}^2}}, \quad \sigma_{\widehat{\beta}} = \widehat{\sigma}\cdot\sqrt{\frac{q_{11}}{q_{11}q_{22}-q_{12}^2}}$$

$\longrightarrow$ After removing the linear trend from the time series, the residuals can be further analyzed to identify additional characteristics of the time series and posisbly any mechanism behind it.

Fourier Series

Let $x = f(t)$ be a function of t $(0 \leq t \leq 2\pi)$.

$$x = f(t) = a_0 + \sum_{n=1}^{\infty}\left[a_n \cos(nt) + b_n \sin(nt)\right]$$

Fourier coefficients:

$$a_0 = \frac{1}{2\pi}\int_{-\pi}^{+\pi} f(t)\ dt$$

$$a_n = \frac{1}{\pi}\int_{-\pi}^{+\pi} f(t)\ \cos(nt)\ dt$$

$$b_n = \frac{1}{\pi}\int_{-\pi}^{+\pi} f(t)\ \sin(nt)\ dt$$

$$n = 1,\ 2,\ 3,\ \ldots\ldots$$

Discrete Fourier analysis of time series

Discrete time series

$$x_t = f(t) \qquad t = 1,\ 2,\ 3,\ ...,\ N$$

$$x_t = a_0 + \sum_{n=1}^{N/2} \left[a_n \ \cos\left(\frac{2n\pi}{N}\ t\right) + b_n \ \sin\left(\frac{2n\pi}{N}\ t\right)\right]$$

$$a_n = \frac{2}{N} \sum_{t=1}^{N} \left[x_t \ \cos\left(\frac{2n\pi}{N}\ t\right)\right] \qquad n = 1,\ 2,\ 3,\ ...,\ \boxed{\frac{N}{2}}$$

$$b_n = \frac{2}{N} \sum_{t=1}^{N} \left[x_t \ \sin\left(\frac{2n\pi}{N}\ t\right)\right] \qquad a_0 = \frac{1}{N} \sum_{t=1}^{N} x_t$$

Amplitude, frequency and period

- Amplitude:

$$A_n = \sqrt{a_n^2 + b_n^2}$$

- Frequency:

$$f_n = \frac{n}{N} \qquad \text{(in cycles per time unit of the data interval)}$$

- Period:

$$T_n = \frac{N}{n} \qquad \text{(in time unit of the data interval)}$$

Comments on DFT

- Nyquist frequency (highest frequency posisble to resolve): 1/2

- We cannot resolve slowl variations with a period longer than the data period

- For large data sets, DFT computations can be time-consuming. Fast Fourier Transform (FFT) algorithm may be needed.

- IF FFT is not employed, DFT can be speeded up if we can drastically reduce the number of repeated computations of sine and cosin functions