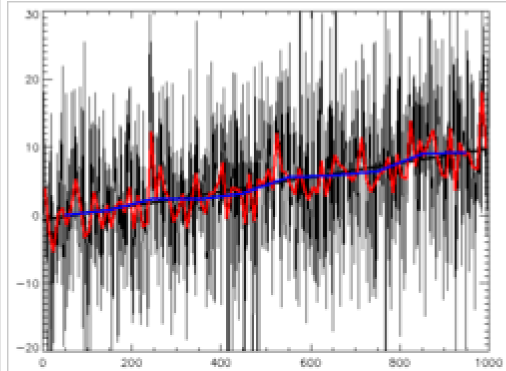
Date visited 10/4/12

Time series

From Wikipedia, the free encyclopedia

In statistics, signal processing, econometrics and mathematical finance, a **time series** is a sequence of data points, measured typically at successive time instants spaced at uniform time intervals. Examples of time series are the daily closing value of the Dow Jones index or the annual flow volume of the Nile River at Aswan. **Time series *analysis*** comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data. **Time series *forecasting*** is the use of a model to predict future values based on previously observed values. Time series are very frequently plotted via line charts.



Time series: random data plus trend, with best-fit line and different smoothings

Time series data have a natural temporal ordering. This makes time series analysis distinct from other common data analysis problems, in which there is no natural ordering of the observations (e.g. explaining people's wages by reference to their respective education levels, where the individuals' data could be entered in any order). Time series analysis is also distinct from spatial data analysis where the observations typically relate to geographical locations (e.g. accounting for house prices by the location as well as the intrinsic characteristics of the houses). A stochastic model for a time series will generally reflect the fact that observations close together in time will be more closely related than observations further apart. In addition, time series models will often make use of the natural one-way ordering of time so that values for a given period will be expressed as deriving in some way from past values, rather than from future values (see time reversibility.)

Methods for time series analyses may be divided into two classes: frequency-domain methods and time-domain methods. The former include spectral analysis and recently wavelet analysis; the latter include auto-correlation and cross-correlation analysis.

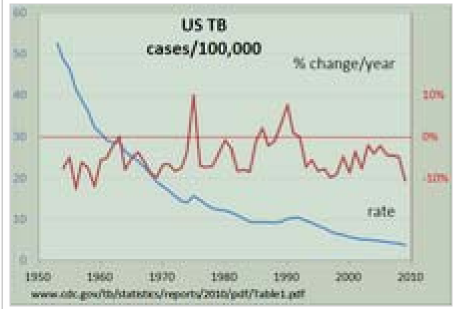
Contents

Analysis

There are several types of data analysis available for time series which are appropriate for different purposes.

General exploration

The clearest way to examine a regular time series is with a line chart such as the one shown for tuberculosis in the United States, made with a spreadsheet program. The number of cases was standardized to a rate per 100,000 and the percent change per year in this rate was calculated. The nearly steadily dropping line shows that the TB incidence was decreasing in most years, but the percent change in this rate varied by as much as +/- 10%, with 'surges' in 1975 and around the early 1990s. The use of both vertical axes allows the comparison of two time series in one graphic. Other techniques include:



Tuberculosis incidence US 1953-2009

- Autocorrelation analysis to examine serial dependence
- Spectral analysis to examine cyclic behaviour which need not be related to seasonality. For example, sun spot activity varies over 11 year cycles.[1][2] Other common examples include celestial phenomena, weather patterns, neural activity, commodity prices, and economic activity.

Description

- Separation into components representing trend, seasonality, slow and fast variation, cyclical irregular: see decomposition of time series
- Simple properties of marginal distributions

Prediction and forecasting

- Fully formed statistical models for stochastic simulation purposes, so as to generate alternative versions of the time series, representing what might happen over non-specific time-periods in the future
- Simple or fully formed statistical models to describe the likely outcome of the time series in the immediate future, given knowledge of the most recent outcomes (forecasting).

Models

Models for time series data can have many forms and represent different stochastic processes. When modeling variations in the level of a process, three broad classes of practical importance are the *autoregressive* (AR) models, the *integrated* (I) models, and the *moving average* (MA) models. These three classes depend linearly[3] on previous data points. Combinations of these ideas produce autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA) models. The autoregressive fractionally integrated moving average (ARFIMA) model generalizes the former three. Extensions of these classes to deal with vector-valued data are available under the heading of multivariate time-series models and sometimes the preceding acronyms are extended by including an initial "V" for "vector". An additional set of extensions of these models is available for use where the observed time-series is driven by some "forcing" time-series (which may not have a causal effect on the observed series): the distinction from the multivariate case is that the forcing series may be deterministic or under the experimenter's control. For these models, the acronyms are extended with a final "X" for "exogenous".

Non-linear dependence of the level of a series on previous data points is of interest, partly because of the possibility of producing a chaotic time series. However, more importantly, empirical investigations can indicate

the advantage of using predictions derived from non-linear models, over those from linear models, as for example in nonlinear autoregressive exogenous models.

Among other types of non-linear time series models, there are models to represent the changes of variance along time (heteroskedasticity). These models represent autoregressive conditional heteroskedasticity (ARCH) and the collection comprises a wide variety of representation (GARCH, TARCH, EGARCH, FIGARCH, CGARCH, etc.). Here changes in variability are related to, or predicted by, recent past values of the observed series. This is in contrast to other possible representations of locally varying variability, where the variability might be modelled as being driven by a separate time-varying process, as in a doubly stochastic model.

In recent work on model-free analyses, wavelet transform based methods (for example locally stationary wavelets and wavelet decomposed neural networks) have gained favor. Multiscale (often referred to as multiresolution) techniques decompose a given time series, attempting to illustrate time dependence at multiple scales. See also Markov switching multifractal (MSMF) techniques for modeling volatility evolution.

Notation

A number of different notations are in use for time-series analysis. A common notation specifying a time series X that is indexed by the natural numbers is written

$$X = \{X_1, X_2, ...\}.$$

Another common notation is

$$Y = \{Y_t: t \in T\},$$

where T is the index set.

Conditions

There are two sets of conditions under which much of the theory is built:

- Stationary process
- Ergodic process

However, ideas of stationarity must be expanded to consider two important ideas: strict stationarity and second-order stationarity. Both models and applications can be developed under each of these conditions, although the models in the latter case might be considered as only partly specified.

In addition, time-series analysis can be applied where the series are seasonally stationary or non-stationary. Situations where the amplitudes of frequency components change with time can be dealt with in time-frequency analysis which makes use of a time–frequency representation of a time-series or signal.[4]

Models

Main article: Autoregressive model

The general representation of an autoregressive model, well known as AR(p), is

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \cdots + \alpha_p Y_{t-p} + \varepsilon_t$$

where the term ε_t is the source of randomness and is called white noise. It is assumed to have the following characteristics:

- $E[\varepsilon_t] = 0\,,$
- $E[\varepsilon_t^2] = \sigma^2\,,$
- $E[\varepsilon_t \varepsilon_s] = 0 \quad \text{for all } t \neq s\,.$

With these assumptions, the process is specified up to second-order moments and, subject to conditions on the coefficients, may be second-order stationary.

If the noise also has a normal distribution, it is called normal or Gaussian white noise. In this case, the AR process may be strictly stationary, again subject to conditions on the coefficients.

Tools for investigating time-series data include:

- Consideration of the autocorrelation function and the spectral density function (also cross-correlation functions and cross-spectral density functions)
- Scaled cross- and auto-correlation functions [5]
- Performing a Fourier transform to investigate the series in the frequency domain
- Use of a filter to remove unwanted noise
- Principal components analysis (or empirical orthogonal function analysis)
- Singular spectrum analysis
- "Structural" models:
  - General State Space Models
  - Unobserved Components Models
- Machine Learning
  - Artificial neural networks
  - Support Vector Machine
  - Fuzzy Logic
- Hidden Markov model
- Control chart
  - Shewhart individuals control chart
  - CUSUM chart
  - EWMA chart
  - Real-time contrasts chart
- Detrended fluctuation analysis
- Dynamic time warping
- Dynamic Bayesian network
- Time-frequency analysis techniques:
  - Fast Fourier Transform
  - Continuous wavelet transform

- Short-time Fourier transform
- Chirplet transform
- Fractional Fourier transform

- Chaotic analysis
  - Correlation dimension
  - Recurrence plots
  - Recurrence quantification analysis
  - Lyapunov exponents
  - Entropy encoding

See also

- Anomaly time series
- Decomposition of time series
- Seasonal adjustment
- Signal processing
- Trend estimation
- Unevenly spaced time series
- Scaled correlation

References

1. ^ Bloomfield, P. (1976). *Fourier analysis of time series: An introduction*. New York: Wiley.
2. ^ Shumway, R. H. (1988). *Applied statistical time series analysis*. Englewood Cliffs, NJ: Prentice Hall.
3. ^ Gershenfeld, N. (1999). *The nature of mathematical modeling*. p.205-08
4. ^ Boashash, B. (ed.), (2003) *Time-Frequency Signal Analysis and Processing: A Comprehensive Reference*, Elsevier Science, Oxford, 2003 ISBN ISBN 0-08-044335-4
5. ^ Nikolić D, Muresan RC, Feng W, Singer W (2012) Scaled correlation analysis: a better way to compute a cross-correlogram. *European Journal of Neuroscience*, pp. 1–21, doi:10.1111/j.1460-9568.2011.07987.x http://www.danko-nikolic.com/wp-content/uploads/2012/03/Scaled-correlation-analysis.pdf

Further reading

- Bloomfield, P. (1976). *Fourier analysis of time series: An introduction*. New York: Wiley.
- Box, George; Jenkins, Gwilym (1976), *Time series analysis: forecasting and control, rev. ed.*, Oakland, California: Holden-Day
- Brillinger, D. R. (1975). *Time series: Data analysis and theory*. New York: Holt, Rinehart. & Winston.
- Brigham, E. O. (1974). *The fast Fourier transform*. Englewood Cliffs, NJ: Prentice-Hall.
- Elliott, D. F., & Rao, K. R. (1982). *Fast transforms: Algorithms, analyses, applications*. New York: Academic Press.
- Gershenfeld, Neil (2000), *The nature of mathematical modeling*, Cambridge: Cambridge Univ. Press, ISBN 978-0-521-57095-4, OCLC 174825352 (//www.worldcat.org/oclc/174825352)
- Hamilton, James (1994), *Time Series Analysis*, Princeton: Princeton Univ. Press, ISBN 0-691-04289-6
- Jenkins, G. M., & Watts, D. G. (1968). *Spectral analysis and its applications*. San Francisco: Holden-Day.
- Priestley, M. B. (1981). *Spectral Analysis and Time Series*. London: Academic Press. ISBN 978-0-12-564901-8
- Shasha, D. (2004), *High Performance Discovery in Time Series*, Berlin: Springer, ISBN 0-387-00857-8
- Shumway, R. H. (1988). *Applied statistical time series analysis*. Englewood Cliffs, NJ: Prentice Hall.
- Wiener, N.(1964). *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*.The MIT Press.
- Wei, W. W. (1989). *Time series analysis: Univariate and multivariate methods*. New York: Addison-Wesley.

- Weigend, A. S., and N. A. Gershenfeld (Eds.) (1994) *Time Series Prediction: Forecasting the Future and Understanding the Past*. Proceedings of the NATO Advanced Research Workshop on Comparative Time Series Analysis (Santa Fe, May 1992) MA: Addison-Wesley.
- Durbin J., and Koopman S.J. (2001) *Time Series Analysis by State Space Methods*. Oxford University Press.

External links

- A First Course on Time Series Analysis (http://statistik.mathematik.uni-wuerzburg.de/timeseries/) - an open source book on time series analysis with SAS
- Introduction to Time series Analysis (Engineering Statistics Handbook) (http://www.itl.nist.gov/div898/handbook/pmc/section4/pmc4.htm) - a practical guide to Time series analysis

Categories: Time series analysis | Formal sciences