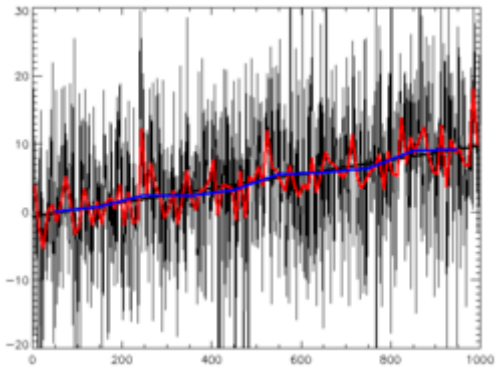
Time series

From Wikipedia, the free encyclopedia

A **time series** is a series of data points indexed (or listed or graphed) in time order. Most commonly, a time series is a sequence taken at successive equally spaced points in time. Thus it is a sequence of discrete-time data. Examples of time series are heights of ocean tides, counts of sunspots, and the daily closing value of the Dow Jones Industrial Average.



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

Time series are very frequently plotted via line charts. Time series are used in statistics, signal processing, pattern recognition, econometrics, mathematical finance, weather forecasting, intelligent transport and trajectory forecasting [1], earthquake prediction, electroencephalography, control engineering, astronomy, communications engineering, and largely in any domain of applied science and engineering which involves temporal measurements.

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. While regression analysis is often employed in such a way as to test theories that the current values of one or more independent time series affect the current value of another time series, this type of analysis of time series is not called "time series analysis", which focuses on comparing values of a single time series or multiple dependent time series at different points in time.[2] Interrupted time series analysis is the analysis of interventions on a single time series

Time series data have a natural temporal ordering. This makes time series analysis distinct from cross-sectional studies, 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.)

Time series analysis can be applied to real-valued, continuous data, discrete numeric data, or discrete symbolic data (i.e. sequences of characters, such as letters and words in the English language[3]).

Contents

Methods for analysis

Methods for time series analysis may be divided into two classes: frequency-domain methods and time-domain methods. The former include spectral analysis and wavelet analysis; the latter include auto-correlation and cross-correlation analysis. In the time domain, correlation and analysis can be made in a filter-like manner using scaled correlation, thereby mitigating the need to operate in the frequency domain.

Additionally, time series analysis techniques may be divided into parametric and non-parametric methods. The parametric approaches assume that the underlying stationary stochastic process has a certain structure which can be described using a small number of parameters (for example, using an autoregressive or moving average model). In these approaches, the task is to estimate the parameters of the model that describes the stochastic process. By contrast, non-parametric approaches explicitly estimate the covariance or the spectrum of the process without assuming that the process has any particular structure.

Methods of time series analysis may also be divided into linear and non-linear, and univariate and multivariate.

Time series and panel data

A time series is one type of Panel data. Panel data is the general class, a multidimensional data set, whereas a time series data set is a one-dimensional panel (as is a cross-sectional dataset). A data set may exhibit characteristics of both panel data and time series data. One way to tell is to ask what makes one data record unique from the other records. If the answer is the time data field, then this is a time series data set candidate. If determining a unique record requires a time data field and an additional identifier which is unrelated to time (student ID, stock symbol, country code), then it is panel data candidate. If the differentiation lies on the non-time identifier, then the data set is a cross-sectional data set candidate.

Analysis

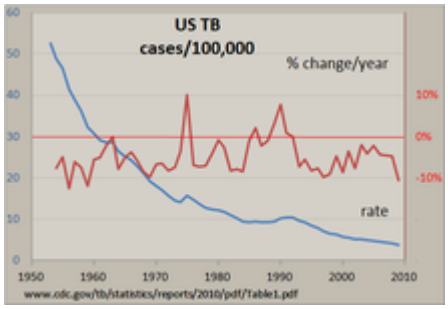
There are several types of motivation and data analysis available for time series which are appropriate for different purposes and etc.

Motivation

In the context of statistics, econometrics, quantitative finance, seismology, meteorology, and geophysics the primary goal of time series analysis is forecasting. In the context of signal processing, control engineering and communication engineering it is used for signal detection and estimation, while in the context of data mining, pattern recognition and machine learning time series analysis can be used for clustering, classification, query by content, anomaly detection as well as forecasting.

Exploratory analysis

The clearest way to examine a regular time series manually 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.



Tuberculosis incidence US 1953-2009

Other techniques include:

- Autocorrelation analysis to examine serial dependence
- Spectral analysis to examine cyclic behavior which need not be related to seasonality. For example, sun spot activity varies over 11 year cycles.[4][5] Other common examples include celestial phenomena, weather patterns, neural activity, commodity prices, and economic activity.
- Separation into components representing trend, seasonality, slow and fast variation, and cyclical irregularity: see trend estimation and decomposition of time series

Curve fitting

Curve fitting[6][7] is the process of constructing a curve, or mathematical function, that has the best fit to a series of data points,[8] possibly subject to constraints.[9][10] Curve fitting can involve either interpolation,[11][12] where an exact fit to the data is required, or smoothing,[13][14] in which a "smooth" function is constructed that approximately fits the data. A related topic is regression analysis,[15][16] which focuses more on questions of statistical inference such as how much uncertainty is present in a curve that is fit to data observed with random errors. Fitted curves can be used as an aid for data visualization,[17][18] to infer values of a function where no data are available,[19] and to summarize the relationships among two or more variables.[20] Extrapolation refers to the use of a fitted curve beyond the range of the observed data,[21] and is subject to a degree of uncertainty[22] since it may reflect the method used to construct the curve as much as it reflects the observed data.

The construction of economic time series involves the estimation of some components for some dates by interpolation between values ("benchmarks") for earlier and later dates. Interpolation is estimation of an unknown quantity between two known quantities (historical data), or drawing conclusions about missing information from the available information ("reading between the lines").[23] Interpolation is useful where the data surrounding the missing data is available and its trend, seasonality, and longer-term cycles are known. This is often done by using a related series known for all relevant dates.[24] Alternatively polynomial interpolation or spline interpolation is used where piecewise polynomial functions are fit into time intervals such that they fit smoothly together. A different problem which is closely related to interpolation is the approximation of a complicated function by a simple function (also called regression).The main difference between regression and interpolation is that polynomial regression gives a single polynomial that models the entire data set. Spline interpolation, however, yield a piecewise continuous function composed of many polynomials to model the data set.

Extrapolation is the process of estimating, beyond the original observation range, the value of a variable on the basis of its relationship with another variable. It is similar to interpolation, which produces estimates between known observations, but extrapolation is subject to greater uncertainty and a higher risk of producing meaningless results.

Function approximation

In general, a function approximation problem asks us to select a function among a well-defined class that closely matches ("approximates") a target function in a task-specific way. One can distinguish two major classes of function approximation problems: First, for known target functions approximation theory is the branch of numerical analysis that investigates how certain known functions (for example, special functions) can be approximated by a specific class of functions (for example, polynomials or rational functions) that often have desirable properties (inexpensive computation, continuity, integral and limit values, etc.).

Second, the target function, call it *g*, may be unknown; instead of an explicit formula, only a set of points (a time series) of the form (*x*, *g*(*x*)) is provided. Depending on the structure of the domain and codomain of *g*, several techniques for approximating *g* may be applicable. For example, if *g* is an operation on the real numbers, techniques of interpolation, extrapolation, regression analysis, and curve fitting can be used. If the codomain (range or target set) of *g* is a finite set, one is dealing with a classification problem instead. A related problem of *online* time series approximation[25] is to summarize the data in one-pass and construct an approximate representation that can support a variety of time series queries with bounds on worst-case error.

To some extent the different problems (regression, classification, fitness approximation) have received a unified treatment in statistical learning theory, where they are viewed as supervised learning problems.

Prediction and forecasting

In statistics, prediction is a part of statistical inference. One particular approach to such inference is known as predictive inference, but the prediction can be undertaken within any of the several approaches to statistical inference. Indeed, one description of statistics is that it provides a means of transferring knowledge about a sample of a population to the whole population, and to other related populations, which is not necessarily the same as prediction over time. When information is transferred across time, often to specific points in time, the process is known as 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).
- Forecasting on time series is usually done using automated statistical software packages and programming languages, such as R, S, SAS, SPSS, Minitab, pandas (Python) and many others.

Classification

Assigning time series pattern to a specific category, for example identify a word based on series of hand movements in sign language

Signal estimation

This approach is based on harmonic analysis and filtering of signals in the frequency domain using the Fourier transform, and spectral density estimation, the development of which was significantly accelerated during World War II by mathematician Norbert Wiener, electrical engineers Rudolf E. Kálmán, Dennis Gabor and others for filtering signals from noise and predicting signal values at a certain point in time. See Kalman filter, Estimation theory, and Digital signal processing

Segmentation

Splitting a time-series into a sequence of segments. It is often the case that a time-series can be represented as a sequence of individual segments, each with its own characteristic properties. For example, the audio signal from a conference call can be partitioned into pieces corresponding to the times during which each person was speaking. In time-series segmentation, the goal is to identify the segment boundary points in the time-series,

and to characterize the dynamical properties associated with each segment. One can approach this problem using change-point detection, or by modeling the time-series as a more sophisticated system, such as a Markov jump linear system.

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 on previous data points.[26] 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", as in VAR for vector autoregression. 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. Further references on nonlinear time series analysis: (Kantz and Schreiber),[27] and (Abarbanel) [28]

Among other types of non-linear time series models, there are models to represent the changes of variance over 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.

A Hidden Markov model (HMM) is a statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved (hidden) states. An HMM can be considered as the simplest dynamic Bayesian network. HMM models are widely used in speech recognition, for translating a time series of spoken words into text.

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.[29]

Tools

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 to remove contributions of slow components[30]
- Performing a Fourier transform to investigate the series in the frequency domain
- Use of a filter to remove unwanted noise
- Principal component 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
  - Gaussian Processes
- Hidden Markov model
- Queueing Theory Analysis
- Control chart
  - Shewhart individuals control chart
  - CUSUM chart
  - EWMA chart
- Detrended fluctuation analysis
- Dynamic time warping[31]
- Cross-correlation[32]
- 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

Measures

Time series metrics or features that can be used for time series classification or regression analysis:[33]

- **Univariate linear measures**
  - Moment (mathematics)
  - Spectral band power
  - Spectral edge frequency
  - Accumulated Energy (signal processing)
  - Characteristics of the autocorrelation function
  - Hjorth parameters
  - FFT parameters
  - Autoregressive model parameters
  - Mann–Kendall test
- **Univariate non-linear measures**
  - Measures based on the correlation sum
  - Correlation dimension
  - Correlation integral
  - Correlation density
  - Correlation entropy
  - Approximate entropy[34]
  - Sample entropy
  - Fourier entropy[uk]
  - Wavelet entropy
  - Rényi entropy
  - Higher-order methods
  - Marginal predictability
  - Dynamical similarity index
  - State space dissimilarity measures
  - Lyapunov exponent
  - Permutation methods
  - Local flow
- **Other univariate measures**
  - Algorithmic complexity
  - Kolmogorov complexity estimates
  - Hidden Markov Model states
  - Rough path signature[35]
  - Surrogate time series and surrogate correction
  - Loss of recurrence (degree of non-stationarity)
- **Bivariate linear measures**
  - Maximum linear cross-correlation
  - Linear Coherence (signal processing)
- **Bivariate non-linear measures**
  - Non-linear interdependence
  - Dynamical Entrainment (physics)
  - Measures for Phase synchronization
  - Measures for Phase locking
- **Similarity measures**:[36]
  - Cross-correlation
  - Dynamic Time Warping[31]
  - Hidden Markov Models
  - Edit distance
  - Total correlation
  - Newey–West estimator
  - Prais–Winsten transformation
  - Data as Vectors in a Metrizable Space
    - Minkowski distance

        - Mahalanobis distance
    - Data as Time Series with Envelopes
        - Global Standard Deviation
        - Local Standard Deviation
        - Windowed Standard Deviation
    - Data Interpreted as Stochastic Series
        - Pearson product-moment correlation coefficient
        - Spearman's rank correlation coefficient
    - Data Interpreted as a Probability Distribution Function
        - Kolmogorov–Smirnov test
        - Cramér–von Mises criterion

Visualization

Time series can be visualized with two categories of chart: Overlapping Charts and Separated Charts. Overlapping Charts display all time series on the same layout while Separated Charts presents them on different layouts (but aligned for comparison purpose)[37]

Overlapping charts

- Braided Graphs
- Line Charts
- Slope Graphs
- GapChart

Separated charts

- Horizon Graphs
- Reduced Line Charts (small multiples)
- Silhouette Graph
- Circular Silhouette Graph

Software

Working with Time Series data is a relatively common use for statistical analysis software. As a result of this, there are many offerings both commercial and open source. Some examples include:

- CRAN supplementary statistics package for R[38]
- Analysis and Forecasting with Weka[39]
- Predictive modeling with GMDH Shell[40]
- Functions and Modeling in the Wolfram Language[41]
- Time Series Objects in MATLAB[42]
- SAS/ETS in SAS software[43]
- Expert Modeler in IBM SPSS Statistics and IBM SPSS Modeler
- Automatic Time series Forecasting with LDT[44]
- EViews is a statistical package for Windows, used mainly for time-series oriented econometric analysis.
- bayesloop: Probabilistic programming framework that facilitates objective model selection for time-varying parameter models[45]

See also

- Anomaly time series
- Chirp
- Decomposition of time series
- Detrended fluctuation analysis
- Digital signal processing
- Distributed lag

- Estimation theory
- Forecasting
- Hurst exponent
- Monte Carlo method
- Panel analysis
- Random walk
- Scaled correlation
- Seasonal adjustment
- Sequence analysis
- Signal processing
- Trend estimation
- Unevenly spaced time series
- Time series database

References

1. Zissis, Dimitrios; Xidias, Elias; Lekkas, Dimitrios (2015). "Real-time vessel behavior prediction" (http://link.springer.com/article/10.1007/s12530-015-9133-5#). *Evolving Systems*. **7**: 1–12. doi:10.1007/s12530-015-9133-5 (https://doi.org/10.1007%2Fs12530-015-9133-5).
2. Imdadullah. "Time Series Analysis" (http://itfeature.com/time-series-analysis-and-forecasting/time-series-analysis-forecasting). *Basic Statistics and Data Analysis*. itfeature.com. Retrieved 2 January 2014.
3. Lin, Jessica; Keogh, Eamonn; Lonardi, Stefano; Chiu, Bill (2003). "A symbolic representation of time series, with implications for streaming algorithms". *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery*. New York: ACM Press. doi:10.1145/882082.882086 (https://doi.org/10.1145%2F882082.882086).
4. Bloomfield, P. (1976). *Fourier analysis of time series: An introduction*. New York: Wiley. ISBN 0471082562.
5. Shumway, R. H. (1988). *Applied statistical time series analysis*. Englewood Cliffs, NJ: Prentice Hall. ISBN 0130415006.
6. Sandra Lach Arlinghaus, PHB Practical Handbook of Curve Fitting. CRC Press, 1994.
7. William M. Kolb. Curve Fitting for Programmable Calculators. Syntec, Incorporated, 1984.
8. S.S. Halli, K.V. Rao. 1992. Advanced Techniques of Population Analysis. ISBN 0306439972 Page 165 (*cf*. ... functions are fulfilled if we have a good to moderate fit for the observed data.)
9. (https://books.google.com/books?id=SI-VqAT4_hYC)The Signal and the Noise: Why So Many Predictions Fail-but Some Don't. By Nate Silver
10. Data Preparation for Data Mining (https://books.google.com/books?id=hhdVr9F-JfAC): Text. By Dorian Pyle.
11. Numerical Methods in Engineering with MATLAB®. By Jaan Kiusalaas. Page 24.
12. Numerical Methods in Engineering with Python 3. By Jaan Kiusalaas. Page 21.
13. Numerical Methods of Curve Fitting. By P. G. Guest, Philip George Guest. Page 349.
14. See also: Mollifier
15. Fitting Models to Biological Data Using Linear and Nonlinear Regression. By Harvey Motulsky, Arthur Christopoulos.
16. Regression Analysis By Rudolf J. Freund, William J. Wilson, Ping Sa. Page 269.
17. Visual Informatics. Edited by Halimah Badioze Zaman, Peter Robinson, Maria Petrou, Patrick Olivier, Heiko Schröder. Page 689.
18. Numerical Methods for Nonlinear Engineering Models. By John R. Hauser. Page 227.
19. Methods of Experimental Physics: Spectroscopy, Volume 13, Part 1. By Claire Marton. Page 150.
20. Encyclopedia of Research Design, Volume 1. Edited by Neil J. Salkind. Page 266.
21. Community Analysis and Planning Techniques. By Richard E. Klosterman. Page 1.
22. An Introduction to Risk and Uncertainty in the Evaluation of Environmental Investments. DIANE Publishing. Pg 69 (https://books.google.com/books?id=rJ23LWaZAqsC&pg=PA69)
23. Hamming, Richard. Numerical methods for scientists and engineers. Courier Corporation, 2012.
24. Friedman, Milton. "The interpolation of time series by related series." Journal of the American Statistical Association 57.300 (1962): 729-757.
25. Gandhi, Sorabh, Luca Foschini, and Subhash Suri. "Space-efficient online approximation of time series data: Streams, amnesia, and out-of-order." Data Engineering (ICDE), 2010 IEEE 26th International Conference on. IEEE, 2010.
26. Gershenfeld, N. (1999). *The Nature of Mathematical Modeling*. New York: Cambridge University Press. pp. 205–208. ISBN 0521570956.
27. Kantz, Holger; Thomas, Schreiber (2004). *Nonlinear Time Series Analysis*. London: Cambridge University Press. ISBN 978-0521529020.

28. Abarbanel, Henry (Nov 25, 1997). *Analysis of Observed Chaotic Data*. New York: Springer. ISBN 978-0387983721.
29. Boashash, B. (ed.), (2003) *Time-Frequency Signal Analysis and Processing: A Comprehensive Reference*, Elsevier Science, Oxford, 2003 ISBN 0-08-044335-4
30. Nikolić, D.; Muresan, R. C.; Feng, W.; Singer, W. (2012). "Scaled correlation analysis: a better way to compute a cross-correlogram". *European Journal of Neuroscience*. **35** (5): 742–762. doi:10.1111/j.1460-9568.2011.07987.x (https://doi.org/10.1111%2Fj.1460-9568.2011.07987.x).
31. Sakoe, Hiroaki; Chiba, Seibi (1978). "Dynamic programming algorithm optimization for spoken word recognition". *IEEE Transactions on Acoustics, Speech and Signal Processing*. doi:10.1109/TASSP.1978.1163055 (https://doi.org/10.1109%2FTASSP.1978.1163055).
32. Goutte, Cyril; Toft, Peter; Rostrup, Egill; Nielsen, Finn Å.; Hansen, Lars Kai (1999). "On Clustering fMRI Time Series". *NeuroImage*. doi:10.1006/nimg.1998.0391 (https://doi.org/10.1006%2Fnimg.1998.0391).
33. Mormann, Florian; Andrzejak, Ralph G.; Elger, Christian E.; Lehnertz, Klaus (2007). "Seizure prediction: the long and winding road". *Brain*. **130** (2): 314–333. PMID 17008335 (https://www.ncbi.nlm.nih.gov/pubmed/17008335). doi:10.1093/brain/awl241 (https://doi.org/10.1093%2Fbrain%2Fawl241).
34. Land, Bruce; Elias, Damian. "Measuring the 'Complexity' of a time series" (http://www.nbb.cornell.edu/neurobio/land/PROJECTS/Complexity/).
35. [1] Chevyrev, I., Kormilitzin, A. (2016) "A Primer on the Signature Method in Machine Learning, arXiv:1603.03788v1"
36. Ropella, G. E. P.; Nag, D. A.; Hunt, C. A. (2003). "Similarity measures for automated comparison of in silico and in vitro experimental results". *Engineering in Medicine and Biology Society*. **3**: 2933–2936. doi:10.1109/IEMBS.2003.1280532 (https://doi.org/10.1109%2FIEMBS.2003.1280532).
37. Tominski, Christian; Aigner, Wolfgang. "The TimeViz Browser:A Visual Survey of Visualization Techniques for Time-Oriented Data" (http://survey.timeviz.net/). Retrieved 1 June 2014.
38. Hyndman, Rob J (2016-01-22). "CRAN Task View: Time Series Analysis" (https://cran.r-project.org/web/views/TimeSeries.html).
39. "Time Series Analysis and Forecasting with Weka - Pentaho Data Mining - Pentaho Wiki" (http://wiki.pentaho.com/display/DATAMINING/Time+Series+Analysis+and+Forecasting+with+Weka). *wiki.pentaho.com*. Retrieved 2016-07-07.
40. "Time Series Analysis & Forecasting Software 2016 [Free Download]" (https://www.gmdhshell.com/time-series-analysis-software). Retrieved 2016-07-07.
41. "Time Series—Wolfram Language Documentation" (http://reference.wolfram.com/language/guide/TimeSeries.html). *reference.wolfram.com*. Retrieved 2016-07-07.
42. "Time Series Objects - MATLAB & Simulink" (http://www.mathworks.com/help/matlab/data_analysis/time-series-objects.html). *www.mathworks.com*. Retrieved 2016-07-07.
43. "Econometrics and Time Series Analysis, SAS/ETS Software" (http://www.sas.com/en_us/software/analytics/ets.html). Retrieved 2016-07-07.
44. "LDT" (https://sourceforge.net/projects/letdatalk/). *SourceForge*. Retrieved 2016-09-04.
45. "bayesloop: Probabilistic programming framework that facilitates objective model selection for time-varying parameter models" (http://bayesloop.com). Retrieved 2016-12-06.

Further reading

- Box, George; Jenkins, Gwilym (1976), *Time Series Analysis: forecasting and control, rev. ed.*, Oakland, California: Holden-Day
- Cowpertwait P.S.P., Metcalfe A.V. (2009), *Introductory Time Series with R*, Springer.
- Durbin J., Koopman S.J. (2001), *Time Series Analysis by State Space Methods*, Oxford University Press.
- Gershenfeld, Neil (2000), *The Nature of Mathematical Modeling*, Cambridge University Press, ISBN 978-0-521-57095-4, OCLC 174825352
- Hamilton, James (1994), *Time Series Analysis*, Princeton University Press, ISBN 0-691-04289-6
- Priestley, M. B. (1981), *Spectral Analysis and Time Series*, Academic Press. ISBN 978-0-12-564901-8
- Shasha, D. (2004), *High Performance Discovery in Time Series*, Springer, ISBN 0-387-00857-8
- Shumway R. H., Stoffer (2011), *Time Series Analysis and its Applications*, Springer.
- Weigend A. S., Gershenfeld N. A. (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), Addison-Wesley.

- Wiener, N. (1949), *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*, MIT Press.
- Woodward, W. A., Gray, H. L. & Elliott, A. C. (2012), *Applied Time Series Analysis*, CRC Press.

External links

- Time series at Encyclopaedia of Mathematics.
- A First Course on Time Series Analysis — An open source book on time series analysis with SAS.
- Introduction to Time series Analysis (Engineering Statistics Handbook) — A practical guide to Time series analysis.
- MATLAB Toolkit for Computation of Multiple Measures on Time Series Data Bases.
- A Matlab tutorial on power spectra, wavelet analysis, and coherence on website with many other tutorials.
- TimeViz survey
- Gaussian Processes for Machine Learning: Book webpage
- CRAN Time Series Task View - Time Series in R
- TimeSeries Analysis with Pandas

Forecasting

From Wikipedia, the free encyclopedia

Forecasting is the process of making predictions of the future based on past and present data and most commonly by analysis of trends. A commonplace example might be estimation of some variable of interest at some specified future date. Prediction is a similar, but more general term. Both might refer to formal statistical methods employing time series, cross-sectional or longitudinal data, or alternatively to less formal judgmental methods. Usage can differ between areas of application: for example, in hydrology the terms "forecast" and "forecasting" are sometimes reserved for estimates of values at certain specific future times, while the term "prediction" is used for more general estimates, such as the number of times floods will occur over a long period.

Risk and uncertainty are central to forecasting and prediction; it is generally considered good practice to indicate the degree of uncertainty attaching to forecasts. In any case, the data must be up to date in order for the forecast to be as accurate as possible.[1]

Contents

Categories of forecasting methods

Qualitative vs. quantitative methods

Qualitative forecasting techniques are subjective, based on the opinion and judgment of consumers, experts; they are appropriate when past data are not available. They are usually applied to intermediate- or long-range decisions. Examples of qualitative forecasting methods are informed opinion and judgment, the Delphi method, market research, and historical life-cycle analogy.

Quantitative forecasting models are used to forecast future data as a function of past data. They are appropriate to use when past numerical data is available and when it is reasonable to assume that some of the patterns in the data are expected to continue into the future. These methods are usually applied to short- or intermediate-range

decisions. Examples of quantitative forecasting methods are last period demand, simple and weighted N-Period moving averages, simple exponential smoothing, poisson process model based forecasting [2] and multiplicative seasonal indexes. Previous research shows that different methods may lead to different level of forecasting accuarcy. For example, GMDH neural network was found to have better forecasting performance than the classical forecasting algorithms such as Single Exponential Smooth, Double Exponential Smooth, ARIMA and back-propagation neural network. [3]

Average approach

In this approach, the predictions of all future values are equal to the mean of the past data. This approach can be used with any sort of data where past data is available. In time series notation:

$\hat{y}_{T+h|T} = \bar{y} = (y_1 + \ldots + y_T)/T$ [4]

where $y_1, \ldots, y_T$ is the past data.

Although the time series notation has been used here, the average approach can also be used for cross-sectional data (when we are predicting unobserved values; values that are not included in the data set). Then, the prediction for unobserved values is the average of the observed values.

Naïve approach

Naïve forecasts are the most cost-effective forecasting model, and provide a benchmark against which more sophisticated models can be compared. This forecasting method is only suitable for time series data.[4] Using the naïve approach, forecasts are produced that are equal to the last observed value. This method works quite well for economic and financial time series, which often have patterns that are difficult to reliably and accurately predict.[4] If the time series is believed to have seasonality, seasonal naïve approach may be more appropriate where the forecasts are equal to the value from last season. The naïve method may also use a drift, which will take the last observation plus the average change from the first observation to the last observation.[4] In time series notation:

$\hat{y}_{T+h|T} = y_T$

Drift method

A variation on the naïve method is to allow the forecasts to increase or decrease over time, where the amount of change over time (called the drift) is set to be the average change seen in the historical data. So the forecast for time $T + h$ is given by

$\hat{y}_{T+h|T} = y_T + \frac{h}{T-1}\sum_{t=2}^{T}(y_t - y_{t-1}) = y_T + h\left(\frac{y_T - y_1}{T-1}\right)$. [4]

This is equivalent to drawing a line between the first and last observation, and extrapolating it into the future.

Seasonal naïve approach

The seasonal naïve method accounts for seasonality by setting each prediction to be equal to the last observed value of the same season. For example, the prediction value for all subsequent months of April will be equal to the previous value observed for April. The forecast for time $T + h$ is:[4]

$\hat{y}_{T+h|T} = y_{T+h-km}$ where m=seasonal period and k is the smallest integer greater than $(h-1)/m$.

The seasonal naïve method is particularly useful for data that has a very high level of seasonality.

Time series methods

Time series methods use historical data as the basis of estimating future outcomes.

- Moving average
- Weighted moving average
- Kalman filtering
- Exponential smoothing
- Autoregressive moving average (ARMA)
- Autoregressive integrated moving average (ARIMA)

  e.g. Box–Jenkins
  Seasonal ARIMA or SARIMA

- Extrapolation
- Linear prediction
- Trend estimation
- Growth curve (statistics)

Causal / econometric for ecasting methods

Some forecasting methods try to identify the underlying factors that might influence the variable that is being forecast. For example, including information about climate patterns might improve the ability of a model to predict umbrella sales. Forecasting models often take account of regular seasonal variations. In addition to climate, such variations can also be due to holidays and customs: for example, one might predict that sales of college football apparel will be higher during the football season than during the off season.[5]

Several informal methods used in causal forecasting do not employ strict algorithms , but instead use the judgment of the forecaster. Some forecasts take account of past relationships between variables: if one variable has, for example, been approximately linearly related to another for a long period of time, it may be appropriate to extrapolate such a relationship into the future, without necessarily understanding the reasons for the relationship.

Causal methods include:

- Regression analysis includes a large group of methods for predicting future values of a variable using information about other variables. These methods include both parametric (linear or non-linear) and non-parametric techniques.
- Autoregressive moving average with exogenous inputs (ARMAX)[6]

Quantitative forecasting models are often judged against each other by comparing their in-sample or out-of-sample mean square error, although some researchers have advised against this.[7] Different forecasting approach has different level of accuracy. For example, it was found that GMDH has higher forecasting accuracy than traditional ARIMA [8]

Judgmental methods

Judgmental forecasting methods incorporate intuitive judgement, opinions and subjective probability estimates. Judgmental forecasting is used in cases where there is lack of historical data or during completely new and unique market conditions.[9]

Judgmental methods include:

- Composite forecasts
- Cooke's method
- Delphi method

- Forecast by analogy
- Scenario building
- Statistical surveys
- Technology forecasting

Artificial intelligence methods

- Artificial neural networks
- Group method of data handling
- Support vector machines

Often these are done today by specialized programs loosely labeled

- Data mining
- Machine Learning
- Pattern Recognition

Other methods

- Simulation
- Prediction market
- Probabilistic forecasting and Ensemble forecasting
- Some socioeconomic forecasters often try to include a humanist factor. They claim that humans, through deliberate action, can have a profound influence on the future. They argue that it should be regarded a real possibility within our current socioeconomic system that its future may be influenced by, to a varying degree, individuals and small groups of individuals. Recent popular publications like Capital in the Twenty-First Century are regarded as major contributors to the increasingly apparent possibility of such reality. It is argued that the influence private and public investment have on our future can never be discomposed of the individual Machiavelian human character. All methods that disregard this factor can not only never accurately predict our socioeconomic future, but can even be used as strong coercion tools. Such theories are generally regarded conspiracy theories, but the increasingly worrying socioeconomic development in the world grants some of these theories a persistent credibility.

Forecasting accuracy

The forecast error (also known as a residual) is the difference between the actual value and the forecast value for the corresponding period.

$$E_t = Y_t - F_t$$

where E is the forecast error at period t, Y is the actual value at period t, and F is the forecast for period t.

A good forecasting method will yield residuals that are **uncorrelated** and have **zero mean**. If there are correlations between residual values, then there is information left in the residuals which should be used in computing forecasts. If the residuals have a mean other than zero, then the forecasts are biased.

Measures of aggregate error:

Scaled Errors: The forecast error, E, is on the same scale as the data, as such, these accuracy measures are scale-dependent and cannot be used to make comparisons between series on different scales.	
Mean absolute error (MAE) or mean absolute deviation (MAD)	$MAE = \frac{\sum_{t=1}^{N} \lvert E_t \rvert}{N}$ $MAD = \frac{\sum_{t=1}^{N} \lvert E_t \rvert}{N}$
Mean squared error (MSE) or mean squared prediction error (MSPE)	$MSE = \frac{\sum_{t=1}^{N} E_t^2}{N}$
Root mean squared error (RMSE)	$RMSE = \sqrt{\frac{\sum_{t=1}^{N} E_t^2}{N}}$
Average of Errors (E)	$\bar{E} = \frac{\sum_{i=1}^{N} E_i}{N}$
Percentage Errors: These are more frequently used to compare forecast performance between different data sets because they are scale-independent. However, they have the disadvantage of being infinite or undefined if Y is close to or equal to zero.	
Mean absolute percentage error (MAPE) or mean absolute percentage deviation (MAPD)	$MAPE = 100 * \frac{\sum_{t=1}^{N} \lvert \frac{E_t}{Y_t} \rvert}{N}$ $MAPD = \frac{\sum_{t=1}^{N} \lvert E_t \rvert}{\sum_{t=1}^{N} \lvert Y_t \rvert}$
Scaled Errors: Hyndman and Koehler (2006) proposed using scaled errors as an alternative to percentage errors.	
Mean absolute scaled error (MASE)	$MASE = \frac{\sum_{t=1}^{N} \lvert \frac{E_t}{\frac{1}{N-m} \sum_{t=m+1}^{N} \lvert Y_t - Y_{t-m} \rvert} \rvert}{N}$ *$m = seasonal period$ or 1 if non-seasonal
Other Measures:	
Forecast skill (SS)	$SS = 1 - \frac{MSE_{forecast}}{MSE_{ref}}$

Business forecasters and practitioners sometimes use different terminology in the industry. They refer to the PMAD as the MAPE, although they compute this as a volume weighted MAPE.[10] For more information see Calculating demand forecast accuracy.

When comparing the accuracy of different forecasting methods on a specific data set, the measures of aggregate error are compared with each other and the method that yields the lowest error is preferred.

Training and test sets

It is important to evaluate forecast accuracy using genuine forecasts. That is, it is invalid to look at how well a model fits the historical data; the accuracy of forecasts can only be determined by considering how well a model performs on new data that were not used when fitting the model. When choosing models, it is common to use a portion of the available data for fitting, and use the rest of the data for testing the model, as was done in the above examples.[11]

Cross Validation

A more sophisticated version of training/test set.

for cross sectional data, cross-validation works as follows:

1. Select observation *i* for the test set, and use the remaining observations in the training set. Compute the error on the test observation.
2. Repeat the above step for *i = 1,2,..., N* where *N* is the total number of observations.
3. Compute the forecast accuracy measures based on the errors obtained.

This is a much more efficient use of the available data, as you only omit one observation at each step

for time series data, the training set can only include observations prior to the test set. therefore no future observations can be used in constructing the forecast. Suppose *k* observations are needed to produce a reliable forecast then the process works as:

1. Select the observation *k + i* for test set, and use the observations at times *1, 2, ..., k+i-1* to estimate the forecasting model. Compute the error on the forecast for k+i.
2. Repeat the above step for *i = 1,2,...,T-k* where *T* is the total number of observations.
3. Compute the forecast accuracy over all errors

This procedure is sometimes known as a "rolling forecasting origin" because the "origin" (*k+i -1)* at which the forecast is based rolls forward in time[12]

Limitations of Errors

The two most popular measures of accuracy that incorporate the forecast error are the Mean Absolute Error (MAE) and the Root Mean Squared Error (RMSE). Thus these measures are considered to be scale-dependent, that is, they are on the same scale as the original data. Consequently, these cannot be used to compare models of differing scales.

Percentage errors are simply forecast errors converted into percentages and are given by $P_t = 100E_t/Y_t$. A common accuracy measure that utilizes this is the Mean Absolute Percentage Error (MAPE). This allows for comparison between data on different scales. However, percentage errors are not quite meaningful when Y_t is close to or equal to zero, which results in extreme values or simply being undefined.[13] Scaled errors are a helpful alternative to percentage errors when comparing between different scales. They do not have the shortfall of giving unhelpful values if Y_t is close to or equal to zero.

See also

- Calculating demand forecast accuracy
- Consensus forecasts
- Forecast error
- Predictability
- Prediction intervals, similar to confidence intervals
- Reference class forecasting

Seasonality and cyclic behaviour

Seasonality

Seasonality is a characteristic of a time series in which the data experiences regular and predictable changes which recur every calendar year. Any predictable change or pattern in a time series that recurs or repeats over a one-year period can be said to be seasonal. It is common in many situations – such as grocery store[14] or even in a Medical Examiner's office[15]—that the demand depends on the day of the week. In such situations, the forecasting procedure calculates the seasonal index of the "season" – seven seasons, one for each day – which is the ratio of the average demand of that season (which is calculated by Moving Average or Exponential

Smoothing using historical data corresponding only to that season) to the average demand across all seasons. An index higher than 1 indicates that demand is higher than average; an index less than 1 indicates that the demand is less than the average.

Cyclic behaviour

The cyclic behaviour of data takes place when there are regular fluctuations in the data which usually last for an interval of at least two years, and when the length of the current cycle cannot be predetermined. Cyclic behavior is not to be confused with seasonal behavior. Seasonal fluctuations follow a consistent pattern each year so the period is always known. As an example, during the Christmas period, inventories of stores tend to increase in order to prepare for Christmas shoppers. As an example of cyclic behaviour, the population of a particular natural ecosystem will exhibit cyclic behaviour when the population increases as its natural food source decreases, and once the population is low, the food source will recover and the population will start to increase again. Cyclic data cannot be accounted for using ordinary seasonal adjustment since it is not of fixed period.

Applications

Forecasting has applications in a wide range of fields where estimates of future conditions are useful. Not everything can be forecasted reliably, if the factors that relate to what is being forecast are known and well understood and there is a significant amount of data that can be used very reliable forecasts can often be obtained. If this is not the case or if the actual outcome is effected by the forecasts, the reliability of the forecasts can be significantly lower.[16]

Climate change and increasing energy prices have led to the use of Egain Forecasting for buildings. This attempts to reduce the energy needed to heat the building, thus reducing the emission of greenhouse gases. Forecasting is used in Customer Demand Planning in everyday business for manufacturing and distribution companies.

While the veracity of predictions for actual stock returns are disputed through reference to the Efficient-market hypothesis, forecasting of broad economic trends is common. Such analysis is provided by both non-profit groups as well as by for-profit private institutions (including brokerage houses[17] and consulting companies[18]).

Forecasting foreign exchange movements is typically achieved through a combination of chart and fundamental analysis. An essential difference between chart analysis and fundamental economic analysis is that chartists study only the price action of a market, whereas fundamentalists attempt to look to the reasons behind the action.[19] Financial institutions assimilate the evidence provided by their fundamental and chartist researchers into one note to provide a final projection on the currency in question.[20]

Forecasting has also been used to predict the development of conflict situations.[21] Forecasters perform research that uses empirical results to gauge the effectiveness of certain forecasting models.[22] However research has shown that there is little difference between the accuracy of the forecasts of experts knowledgeable in the conflict situation and those by individuals who knew much less.[23]

Similarly, experts in some studies argue that role thinking does not contribute to the accuracy of the forecast.[24] The discipline of demand planning, also sometimes referred to as supply chain forecasting, embraces both statistical forecasting and a consensus process. An important, albeit often ignored aspect of forecasting, is the relationship it holds with planning. Forecasting can be described as predicting what the future *will* look like, whereas planning predicts what the future *should* look like.[25][26] There is no single right forecasting method to use. Selection of a method should be based on your objectives and your conditions (data etc.).[27] A good place to find a method, is by visiting a selection tree. An example of a selection tree can be found here.[28] Forecasting has application in many situations:

- Supply chain management - Forecasting can be used in supply chain management to ensure that the right product is at the right place at the right time. Accurate forecasting will help retailers reduce excess inventory and thus increase profit margin. Studies have shown that extrapolations are the least accurate, while company earnings forecasts are the most reliable.[29] Accurate forecasting will also help them meet consumer demand.
- Economic forecasting
- Earthquake prediction
- Egain forecasting
- Finance against risk of default via credit ratings and credit scores
- Land use forecasting
- Player and team performance in sports
- Political forecasting
- Product forecasting
- Sales forecasting
- Technology forecasting
- Telecommunications forecasting
- Transport planning and Transportation forecasting
- Weather forecasting, Flood forecasting and Meteorology

Limitations

Limitations pose barriers beyond which forecasting methods cannot reliably predict. There are many events and values that cannot be forecast reliably. Events such as the roll of a die or the results of the lottery cannot be forecast because they are random events and there is no significant relationship in the data. When the factors that lead to what is being forecast are not known or well understood such as in stock and foreign exchange markets forecasts are often inaccurate or wrong as there is not enough data about everything that affects these markets for the forecasts to be reliable, in addition the outcomes of the forecasts of these markets change the behavior of those involved in the market further reducing forecast accuracy.[16]

Performance limits of fluid dynamics equations

As proposed by Edward Lorenz in 1963, long range weather forecasts, those made at a range of two weeks or more, are impossible to definitively predict the state of the atmosphere, owing to the chaotic nature of the fluid dynamics equations involved. Extremely small errors in the initial input, such as temperatures and winds, within numerical models double every five days.[30]

See also

- Accelerating change
- Collaborative planning, forecasting, and replenishment
- Earthquake prediction
- Energy forecasting
- Forecasting bias
- Foresight (future studies)
- Futures studies
- Futurology
- Kondratiev wave
- Optimism bias
- Planning
- Risk management
- Scenario planning
- Spending wave
- Strategic foresight
- Technology forecasting
- Time Series

- Weather forecasting
- Wind power forecasting

References

1. Scott Armstrong; Fred Collopy; Andreas Graefe; Kesten C. Green. "Answers to Frequently Asked Questions" (http://www.forecastingprinciples.com/index.php/faq). Retrieved May 15, 2013.
2. Mahmud, Tahmida; Hasan, Mahmudul; Chakraborty, Anirban; Roy-Chowdhury, Amit (19 August 2016). *A poisson process model for activity forecasting* (http://dx.doi.org/10.1109/ICIP.2016.7532978). 2016 IEEE International Conference on Image Processing (ICIP). IEEE.
3. Li, Rita Yi Man, Fong, S., Chong, W.S. (2017) Forecasting the REITs and stock indices: Group Method of Data Handling Neural Network approach (http://www.tandfonline.com/doi/full/10.1080/14445921.2016.1225149), Pacific Rim Property Research Journal, 23(2), 1-38
4. https://www.otexts.org/fpp/2/3
5. Nahmias, Steven (2009). *Production and Operations Analysis*.
6. Ellis, Kimberly (2008). *Production Planning and Inventory Control Virginia Tech*. McGraw Hill. ISBN 978-0-390-87106-0.
7. J. Scott Armstrong and Fred Collopy (1992). "Error Measures For Generalizing About Forecasting Methods: Empirical Comparisons" (http://marketing.wharton.upenn.edu/ideas/pdf/armstrong2/armstrong-errormeasures-empirical.pdf) (PDF). *International Journal of Forecasting*. **8**: 69–80. doi:10.1016/0169-2070(92)90008-w (https://doi.org/10.1016%2F0169-2070%2892%2990008-w).
8. 16. Li, Rita Yi Man, Fong, S., Chong, W.S. (2017) Forecasting the REITs and stock indices: Group Method of Data Handling Neural Network approach (http://www.tandfonline.com/eprint/pgjIcAMrJBP4WcHIZH7F/full), Pacific Rim Property Research Journal, 23(2), 1-38
9. https://www.otexts.org/fpp/3/1
10. http://www.forecastingblog.com/?p=134
11. "2.5 Evaluating forecast accuracy | OTexts" (https://www.otexts.org/fpp/2/5). *www.otexts.org*. Retrieved 2016-05-14.
12. "2.5 Evaluating forecast accuracy | OTexts" (https://www.otexts.org/fpp/2/5). *www.otexts.org*. Retrieved 2016-05-17.
13. https://www.otexts.org/fpp/2/5
14. Erhun, F.; Tayur, S. (2003). "Enterprise-Wide Optimization of Total Landed Cost at a Grocery Retailer". *Operations Research*. **51** (3): 343. doi:10.1287/opre.51.3.343.14953 (https://doi.org/10.1287%2Fopre.51.3.343.14953).
15. Omalu, B. I.; Shakir, A. M.; Lindner, J. L.; Tayur, S. R. (2007). "Forecasting as an Operations Management Tool in a Medical Examiner's Office". *Journal of Health Management*. **9**: 75. doi:10.1177/097206340700900105 (https://doi.org/10.1177%2F097206340700900105).
16. https://www.otexts.org/fpp/1/1
17. Fidelity. "2015 Stock Market Outlook" (https://www.fidelity.com/viewpoints/investing-ideas/2015-outlook-for-stocks), a sample outlook report by a brokerage house.
18. McKinsey Insights & Publications. "Insights & Publications" (http://www.mckinsey.com/insights).
19. Helen Allen; Mark P. Taylor (1990). "Charts, Noise and Fundamentals in the London Foreign Exchange Market". JSTOR 2234183 (https://www.jstor.org/stable/2234183).
20. Pound Sterling Live. "Euro Forecast from Institutional Researchers" (https://www.poundsterlinglive.com/euroforecast), A list of collated exchange rate forecasts encompassing technical and fundamental analysis in the foreign exchange market.
21. T. Chadefaux (2014). "Early warning signals for war in the news". Journal of Peace Research, 51(1), 5-18
22. J. Scott Armstrong; Kesten C. Green; Andreas Graefe (2010). "Answers to Frequently Asked Questions" (http://qbox.wharton.upenn.edu/documents/mktg/research/FAQ.pdf) (PDF).
23. Kesten C. Greene; J. Scott Armstrong (2007). "The Ombudsman: Value of Expertise for Forecasting Decisions in Conflicts" (http://marketing.wharton.upenn.edu/documents/research/Value%20of%20expertise.pdf) (PDF). *Interfaces*. INFORMS. **0**: 1–12.
24. Kesten C. Green; J. Scott Armstrong (1975). "Role thinking: Standing in other people's shoes to forecast decisions in conflicts" (http://www.forecastingprinciples.com/paperpdf/Escalation%20Bias.pdf) (PDF). *Role thinking: Standing in other people's shoes to forecast decisions in conflicts*. **39**: 111–116.

25. "FAQ" (http://www.forecastingprinciples.com/index.php?option=com_content&task=view&id=3&Itemid=3). Forecastingprinciples.com. 1998-02-14. Retrieved 2012-08-28.
26. Greene, Kesten C.; Armstrong, J. Scott. "Structured analogies for forecasting" (http://www.qbox.wharton.upenn.edu/documents/mktg/research/INTFOR3581%20-%20Publication%2015.pdf) (PDF). University of Pennsylvania.
27. "FAQ" (http://www.forecastingprinciples.com/index.php?option=com_content&task=view&id=3&Itemid=3#D._Choosing_the_best_method). Forecastingprinciples.com. 1998-02-14. Retrieved 2012-08-28.
28. "Selection Tree" (http://www.forecastingprinciples.com/index.php?option=com_content&task=view&id=17&Itemid=17). Forecastingprinciples.com. 1998-02-14. Retrieved 2012-08-28.
29. J. Scott Armstrong (1983). "Relative Accuracy of Judgmental and Extrapolative Methods in Forecasting Annual Earnings" (http://www.forecastingprinciples.com/paperpdf/Monetary%20Incentives.pdf) (PDF). *Journal of Forecasting*. **2**: 437–447. doi:10.1002/for.3980020411 (https://doi.org/10.1002%2Ffor.3980020411).
30. Cox, John D. (2002). *Storm Watchers*. John Wiley & Sons, Inc. pp. 222–224. ISBN 0-471-38108-X.

- Armstrong, J. Scott, ed. (2001). *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Norwell, Massachusetts: Kluwer Academic Publishers. ISBN 0-7923-7930-6.
- Ellis, Kimberly (2010). *Production Planning and Inventory Control*. McGraw-Hill. ISBN 0-412-03471-9.
- Geisser, Seymour (June 1993). *Predictive Inference: An Introduction*. Chapman & Hall, CRC Press. ISBN 0-390-87106-0.
- Gilchrist, Warren (1976). *Statistical Forecasting*. London: John Wiley & Sons. ISBN 0-471-99403-0.
- Hyndman, Rob J.; Koehler, Anne B. (October–December 2006). "Another look at measures of forecast accuracy" (PDF). *International Journal of Forecasting*. **22** (4): 679–688. doi:10.1016/j.ijforecast.2006.03.001.
- Makridakis, Spyros; Wheelwrigt, Steven; Hyndman, Rob J. (1998). *Forecasting: Methods and Applications*. John Wiley & Sons. ISBN 0-471-53233-9.
- Malakooti, Behnam (February 2014). *Operations and Production Systems with Multiple Objectives*. John Wiley & Sons. ISBN 978-0-470-03732-4.
- Kaligasidis, Angela Sasic; Taesler, Roger; Andersson, Cari; Nord, Margitta (August 2006). "Upgraded weather forecast control of building heating systems". In Fazio, Paul. *Research in Building Physics and Building Engineering*. Taylor & Francis, CRC Press. pp. 951–958. ISBN 0-415-41675-2.
- Kress, George J.; Snyder, John (May 1994). *Forecasting and Market Analysis Techniques: A Practical Approach*. Quorum Books. ISBN 0-89930-835-X.
- Rescher, Nicholas (1998). *Predicting the Future: An Introduction to the Theory of Forecasting*. State University of New York Press. ISBN 0-7914-3553-9.
- Taesler, Roger (1991). "Climate and Building Energy Management". *Energy and Buildings*. **15** (1-2): 599–608. doi:10.1016/0378-7788(91)90028-2.
- Turchin, Peter (2007). "Scientific Prediction in Historical Sociology: Ibn Khaldun meets Al Saud". *History & Mathematics: Historical Dynamics and Development of Complex Societies*. Moscow: KomKniga. pp. 9–38. ISBN 978-5-484-01002-8.
- US patent 6098893, Berglund, Ulf Stefan & Lundberg, Bjorn Henry, "Comfort control system incorporating weather forecast data and a method for operating such a system", issued August 8, 2000.

External links

- Forecasting Principles: *"Evidence-based forecasting"*
- International Institute of Forecasters
- Introduction to Time series Analysis (Engineering Statistics Handbook) - A practical guide to Time series analysis and forecasting
- Time Series Analysis
- Global Forecasting with IFs
- Earthquake Electromagnetic Precursor Research
- Forecasting Science and Theory of Forecasting

Seasonality

From Wikipedia, the free encyclopedia

In time series data, **seasonality** is the presence of variations that occur at specific regular intervals less than a year, such as weekly, monthly, or quarterly. Seasonality may be caused by various factors, such as weather, vacation, and holidays[1] and consists of periodic, repetitive, and generally regular and predictable patterns in the levels[2] of a time series.

Seasonal fluctuations in a time series can be contrasted with cyclical patterns. The latter occur when the data exhibits rises and falls that are not of a fixed period. These fluctuations are usually due to economic conditions and are often related to the "business cycle." The period of time usually extends beyond a single year and the fluctuations are usually of at least two years.[3]

Organisations facing seasonal variations, such as ice-cream vendors, are often interested in knowing their performance relative to the normal seasonal variation. Seasonal variations in the labour market can be attributed to the entrance of school leavers into the job market as they aim to contribute to the workforce upon the completion of their schooling. These regular changes are of less interest to those who study employment data than the variations that occur due to the underlying state of the economy; their focus is on how unemployment in the workforce has changed, despite the impact of the regular seasonal variations.[3]

It is necessary for organisations to identify and measure seasonal variations within their market to help them plan for the future. This can prepare them for the temporary increases or decreases in labour requirements and inventory as demand for their product or service fluctuates over certain periods. This may require training, periodic maintenance, and so forth that can be organized in advance. Apart from these considerations, the organisations need to know if variation they have experienced has been more or less than the expected amount, beyond what the usual seasonal variations account for.

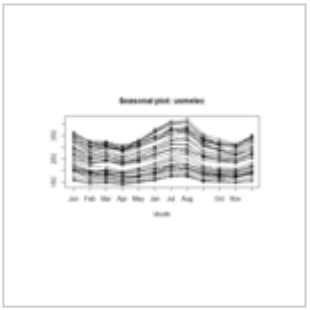
Contents

Reasons for studying seasonal variation

There are several main reasons for studying seasonal variation:

- The description of the seasonal effect provides a better understanding of the impact this component has upon a particular series.
- After establishing the seasonal pattern, methods can be implemented to eliminate it from the time-series to study the effect of other components such as cyclical and irregular variations. This elimination of the seasonal effect is referred to as de-seasonalizing or seasonal adjustment of data.

- To use the past patterns of the seasonal variations to contribute to forecasting and the prediction of the future trends.

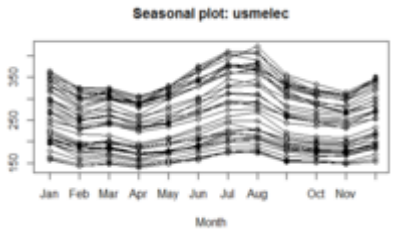


A seasonality plot of
US Electricity Usage

Detecting seasonality

The following graphical techniques can be used to detect seasonality:

- A run sequence plot will often show seasonality
- A seasonal plot will show the data from each season overlapped[4]
- A seasonal subseries plot is a specialized technique for showing seasonality
- Multiple box plots can be used as an alternative to the seasonal subseries plot to detect seasonality
- An autocorrelation plot (ACF) and a spectral plot can help identify seasonality.
- Seasonal Index measures how much the average for a particular period tends to be above (or below) the expected value



A seasonality plot of US electricity usage

A really good way to find periodicity, including seasonality, in any regular series of data is to remove any overall trend first and then to inspect time periodicity.[5]

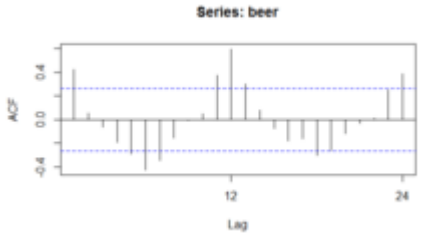
The run sequence plot is a recommended first step for analyzing any time series. Although seasonality can sometimes be indicated by this plot, seasonality is shown more clearly by the seasonal subseries plot or the box plot. The seasonal subseries plot does an excellent job of showing both the seasonal differences (between group patterns) and also the within-group patterns. The box plot shows the seasonal difference (between group patterns) quite well, but it does not show within group patterns. However, for large data sets, the box plot is usually easier to read than the seasonal subseries plot.

The seasonal plot, seasonal subseries plot, and the box plot all assume that the seasonal periods are known. In most cases, the analyst will in fact, know this. For example, for monthly data, the period is 12 since there are 12 months in a year. However, if the period is not known, the autocorrelation plot can help. If there is significant seasonality, the autocorrelation plot should show spikes at lags equal to the period. For example, for monthly data, if there is a seasonality effect, we would expect to see significant peaks at lag 12, 24, 36, and so on (although the intensity may decrease the further out we go).

An autocorrelation plot (ACF) can be used to identify seasonality, as it calculates the difference (residual amount) between a Y value and a lagged value of Y. The result gives some points where the two values are close together (no seasonality), but other points where there is a large discrepancy. These points indicate a level of seasonality in the data.

Semiregular cyclic variations might be dealt with by spectral density estimation.

Measuring seasonality



An ACF (autocorrelation) plot, of Australia beer consumption data.

Seasonal variation is measured in terms of an index, called a seasonal index. It is an average that can be used to compare an actual observation relative to what it would be if there were no seasonal variation. An index value is attached to each period of the time series within a year. This implies that if monthly data are considered there are 12 separate seasonal indices, one for each month. The following methods use seasonal indices to measure seasonal variations of a time-series data.

- Method of simple averages
- Ratio to trend method
- Ratio-to-moving-average method
- Link relatives method

The measurement of seasonal variation by using the ratio-to-moving-average method provides an index to measure the degree of the seasonal variation in a time series. The index is based on a mean of 100, with the degree of seasonality measured by variations away from the base. For example, if we observe the hotel rentals in a winter resort, we find that the winter quarter index is 124. The value 124 indicates that 124 percent of the average quarterly rental occur in winter. If the hotel management records 1436 rentals for the whole of last year, then the average quarterly rental would be 359= (1436/4). As the winter-quarter index is 124, we estimate the number of winter rentals as follows:

359*(124/100)=445;

Here, 359 is the average quarterly rental. 124 is the winter-quarter index. 445 the seasonalized winter-quarter rental.

This method is also called the percentage moving average method. In this method, the original data values in the time-series are expressed as percentages of moving averages. The steps and the tabulations are given below.

Steps

1. Find the centered 12 monthly (or 4 quarterly) moving averages of the original data values in the time-series.

2. Express each original data value of the time-series as a percentage of the corresponding centered moving average values obtained in step(1).In other words, in a multiplicative time-series model, we get(Original data values)/(Trend values) *100 = (T*C*S*I)/(T*C)*100 = (S*I) *100. This implies that the ratio–to-moving average represents the seasonal and irregular components.

3. Arrange these percentages according to months or quarter of given years. Find the averages over all months or quarters of the given years.

4. If the sum of these indices is not 1200(or 400 for quarterly figures), multiply then by a correction factor = 1200/ (sum of monthly indices). Otherwise, the 12 monthly averages will be considered as seasonal indices.

Let us calculate the seasonal index by the ratio-to-moving-average method from the following data:

Table (1):				
Year/Quarters	I	II	III	IV
1996	75	60	54	59
1997	86	65	63	80
1998	90	72	66	85
1999	100	78	72	93

Now calculations for 4 quarterly moving averages and ratio-to-moving-averages are shown in the below table.

Table (2)							
Year	Quarter	Original Values(Y)	4 Figures Moving Total	4 Figures Moving Average	2 Figures Moving Total	2 Figures Moving Average(T)	Ratio-to-Moving-Average(%)(Y)/ (T)*100
1996	1	75					
	2	60					
			248	62.00			
	3	54			126.75	63.375	85.21
			259	64.75			
	4	59			130.75	65.375	90.25
			264	66.00			
1997	1	86			134.25	67.125	128.12
			273	68.25			
	2	65			141.75	70.875	91.71
			294	73.50			
	3	63			148	74	85.13
			298	74.50			
	4	80			150.75	75.375	106.14
			305	76.25			
1998	1	90			153.25	76.625	117.45
			308	77.00			
	2	72			155.25	77.625	92.75
			313	78.25			
	3	66			159.00	79.50	83.02
			323	80.75			
	4	85			163	81.50	104.29
			329	82.25			
1999	1	100			166	83	120.48
			335	83.75			
	2	78			169.50	84.75	92.03
			343	85.75			
	3	72					
	4	93					

Calculation of seasonal index

Table (3)				
Years/Quarters	1	2	3	4
1996	-	-	85.21	90.25
1997	128.12	91.71	85.13	106.14
1998	117.45	92.75	83.02	104.29
1999	120.48	92.04	-	-
Total	366.05	276.49	253.36	300.68
Seasonal Average	122.01	92.16	84.45	100.23
Adjusted Seasonal Average	122.36	92.43	84.69	100.52

Now the total of seasonal averages is 398.85. Therefore, the corresponding correction factor would be 400/398.85 = 1.00288. Each seasonal average is multiplied by the correction factor 1.00288 to get the adjusted seasonal indices as shown in the above table.

Remarks

1. In an additive time-series model, the seasonal component is estimated as: S = Y − (T + C + I)

Where S is for Seasonal values

Y is for actual data values of the time-series

T is for trend values

C is for cyclical values

I is for irregular values.

2. In a multiplicative time-series model, the seasonal component is expressed in terms of ratio and percentage as

$$Seasonal\ effect = \frac{T \cdot S \cdot C \cdot I}{T \cdot C \cdot I} \times 100 \ = \frac{Y}{T \cdot C \cdot I} \times 100;$$

However, in practice the detrending of time-series is done to arrive at $S \cdot C \cdot I$.

This is done by dividing both sides of $Y = T \cdot S \cdot C \cdot I$ by trend values T so that $\frac{Y}{T} = S \cdot C \cdot I$.

3. The deseasonalized time-series data will have only trend (T) cyclical(C) and irregular (I) components and is expressed as:

- Multiplicative model : $\frac{Y}{S} \times 100 = \frac{T \cdot S \cdot C \cdot I}{S} \times 100 = (T \cdot C \cdot I) \times 100$
- Additive model: Y – S = (T + S + C + I) – S = T + C + I

Modeling seasonality

A completely regular cyclic variation in a time series might be dealt with in time series analysis by using a sinusoidal model with one or more sinusoids whose period-lengths may be known or unknown depending on the context. A less completely regular cyclic variation might be dealt with by using a special form of an

ARIMA model which can be structured so as to treat cyclic variations semi-explicitly. Such models represent cyclostationary processes.

Another method of modelling periodic seasonality is the use of pairs of Fourier terms. Similar to using the sinusoidal model, Fourier terms added into regression models utilize sine and cosine terms in order to simulate seasonality. However, the seasonality of such a regression would be represented as the sum of sine or cosine terms, instead of a single sine or cosine term in a sinusoidal model. Every periodic function can be approximated with the inclusion of Fourier terms.

The difference between a sinusoidal model and a regression with Fourier terms can be simplified as below:

Sinusoidal Model:

$$Y_i = a + bt + \alpha \sin(2\pi\omega T_i + \phi) + E_i$$

Regression With Fourier Terms:

$$Y_i = a + bt + (\sum_{k=1}^{K} \alpha_k \cdot \sin(\tfrac{2\pi kt}{m}) + \beta_k \cdot \cos(\tfrac{2\pi kt}{n})) + E_i$$

Seasonal adjustment

Seasonal adjustment is any method for removing the seasonal component of a time series. The resulting seasonally adjusted data are used, for example, when analyzing or reporting non-seasonal trends over durations rather longer than the seasonal period. An appropriate method for seasonal adjustment is chosen on the basis of a particular view taken of the decomposition of time series into components designated with names such as "trend", "cyclic", "seasonal" and "irregular", including how these interact with each other. For example, such components might act additively or multiplicatively. Thus, if a seasonal component acts additively, the adjustment method has two stages:

- estimate the seasonal component of variation in the time series, usually in a form that has a zero mean across series;
- subtract the estimated seasonal component from the original time series, leaving the seasonally adjusted series: $Y_t - S_t = T_t + E_t$.[3]

If it is a multiplicative model,the magnitude of the seasonal fluctuations will vary with the level, which is more likely to occur with economic series.[3] When taking seasonality into account, the seasonally adjusted multiplicative decomposition can be written as $Y_t / S_t = T_t * E_t$; whereby the original time series is divided by the estimated seasonal component.

The multiplicative model can be transformed into an additive model by taking the log of the time series;

SA Multiplicative decomposition: $Y_t = S_t * T_t * E_t$

Taking log of the time series of the multiplicative model: $logY_t = logS_t + logT_t + logE_t$ [3]

One particular implementation of seasonal adjustment is provided by X-12-ARIMA.

Regression analysis

In regression analysis such as ordinary least squares, with a seasonally varying dependent variable being influenced by one or more independent variables, the seasonality can be accounted for and measured by including *n*-1 dummy variables, one for each of the seasons except for an arbitrarily chosen reference season, where *n* is the number of seasons (e.g., 4 in the case of meteorological seasons, 12 in the case of months, etc.).

Each dummy variable is set to 1 if the data point is drawn from the dummy's specified season and 0 otherwise. Then the predicted value of the dependent variable for the reference season is computed from the rest of the regression, while for any other season it is computed using the rest of the regression and by inserting the value 1 for the dummy variable for that season.

Difference between seasonal patterns and cyclic patterns

A **seasonal pattern** occurs when a time series is affected by seasonal factors such as the time of the year or the day of the week.
A **cycle** occurs when the data exhibit rises and falls that are not of a fixed period. These fluctuations are usually due to economic conditions and are often related to the "business cycle".
It is important to distinguish cyclic patterns and seasonal patterns. Seasonal patterns have a fixed and known length, while cyclic patterns have variable and unknown length. The average length of a cycle is usually longer than that of seasonality, and the magnitude of cyclic variation is usually more variable than that of seasonal variation.[6]

See also

- Oscillation
- Periodic function
- Periodicity (disambiguation)
- Photoperiodism

References

1. http://www.allbusiness.com/barrons_dictionary/dictionary-seasonality-4946957-1.html. Missing or empty `|title=` (help) |title=Influencing Factors|
2. http://www.businessdictionary.com/definition/seasonality.html. Missing or empty `|title=` (help)
3. "6.1 Time series components - OTexts" (https://www.otexts.org/fpp/6/1).
4. "2.1 Graphics - OTexts" (https://www.otexts.org/fpp/2/1).
5. "What method can be used to detect seasonality in data?" (http://stats.stackexchange.com/questions/16117/what-method-can-be-used-to-detect-seasonality-in-data).
6. "2.1 Graphics - OTexts" (https://www.otexts.org/fpp/2/1).

- Barnett, A.G.; Dobson, A.J. (2010). *Analysing Seasonal Health Data*. Springer. ISBN 978-3-642-10747-4.
- *Complete Business Statistics* (Chapter 12) by Amir D. Aczel.
- *Business Statistics: Why and When* (Chapter 15) by Larry E. Richards and Jerry J. Lacava.
- *Business Statistics* (Chapter 16) by J.K. Sharma.
- *Business Statistics, a decision making approach* (Chapter 18) by David F. Groebner and Patric W. Shannon.
- *Statistics for Management* (Chapter 15) by Richard I. Levin and David S. Rubin.
- *Forecasting: practice and principles* by Rob J. Hyndman and George Athansopoulos

External links

- Seasonality at NIST/SEMATECH e-Handbook of Statistical Methods

Additive models

The models that we have considered in earlier sections have been **additive models**, and there has been an implicit assumption that the different components affected the time series additively.

$$\text{Data} = \text{Seasonal effect} + \text{Trend} + \text{Cyclical} + \text{Residual}$$

For monthly data, an additive model assumes that the difference between the January and July values is approximately the same each year. In other words, the **amplitude** of the seasonal effect is the same each year.

The model similarly assumes that the residuals are roughly the same size throughout the series -- they are a random component that adds on to the other components in the same way at all parts of the series.

Multiplicative models

In many time series involving **quantities** (e.g. money, wheat production, ...), the absolute differences in the values are of less interest and importance than the percentage changes.

For example, in seasonal data, it might be more useful to model that the July value is the same **proportion** higher than the January value in each year, rather than assuming that their difference is constant. Assuming that the seasonal and other effects act proportionally on the series is equivalent to a **multiplicative model**,

$$\text{Data} = (\text{Seasonal effect}) \times \text{Trend} \times \text{Cyclical} \times \text{Residual}$$

Fortunately, multiplicative models are equally easy to fit to data as additive models! The trick to fitting a multiplicative model is to take logarithms of both sides of the model,

$$\begin{aligned}\log(\text{Data}) &= \log(\text{Seasonal effect} \times \text{Trend} \times \text{Cyclical} \times \text{Residual}) \\ &= \log(\text{Seasonal effect}) + \log(\text{Trend}) \\ &\quad + \log(\text{Cyclical}) + \log(\text{Residual})\end{aligned}$$

After taking logarithms (either natural logarithms or to base 10), the four components of the time series again act additively.

To fit a multiplicative model, take logarithms of the data, then analyse the log data as before.

It is important to recognise when multiplicative models are appropriate. However fitting the models is no harder than fitting additive models.

Decomposition of time series

From Wikipedia, the free encyclopedia

The **decomposition of time series** is a statistical method that deconstructs a time series into several components, each representing one of the underlying categories of patterns.[1] There are two principal types of decomposition which are outlined below.

Contents

Decomposition based on rates of change

This is an important technique for all types of time series analysis, especially for seasonal adjustment.[2] It seeks to construct, from an observed time series, a number of component series (that could be used to reconstruct the original by additions or multiplications) where each of these has a certain characteristic or type of behaviour. For example, time series are usually decomposed into:

- T_t, the trend component at time *t*, that reflects the long-term progression of the series (secular variation). A trend exists when there is an increasing or decreasing direction in the data. The trend component does not have to be linear.[1]
- C_t, the cyclical component at time t, that describes repeated but non-periodic fluctuations. The duration of these fluctuations is usually of at least two years.
- S_t, the seasonal component at time *t*, reflecting seasonality (seasonal variation). A seasonal pattern exists when a time series is influenced by seasonal factors. Seasonality is always of a fixed and known period (e.g., the quarter of the year, the month, or day of the week).[1]
- I_t, the irregular component (or "noise") at time t, that describes random, irregular influences. It represents the residuals or remainder of the time series after the other components have been removed.

Hence a time series using an additive model can be thought of as

$$y_t = T_t + C_t + S_t + I_t,$$

whereas a multiplicative model would be

$$y_t = T_t \times C_t \times S_t \times I_t.$$

An additive model would be used when the variations around the trend does not vary with the level of the time series where as a multiplicative model would be appropriate if the trend is proportional to the level of the time series.[3]

Sometimes the trend and cyclic components are grouped into one, called the trend-cycle component. The trend-component can just be referred to as the "trend" component, even though it may contain cyclic behaviour.[3] For example, a seasonal decomposition of time series by Loess (STL)[4] plot decomposes a time series into seasonal, trend and irregular components using loess and plots the components separately, whereby the cyclic component (if present in the data) is included in the "trend" component plot.

Decomposition based on predictability

The theory of time series analysis makes use of the idea of decomposing a times series into deterministic and non-deterministic components (or predictable and unpredictable components).[2] See Wold's theorem and Wold decomposition.

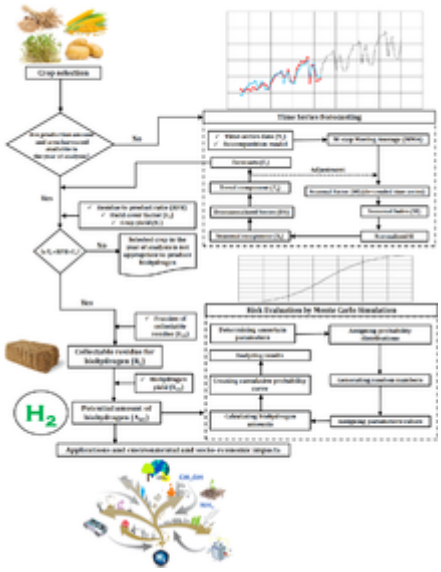
Examples

Kendall shows an example of a decomposition into smooth, seasonal and irregular factors for a set of data containing values of the monthly aircraft miles flown by UK airlines.[5]

Furthermore, in policy analysis, forecasting future production of biofuels is key data for making better decisions. Thereby, statistical time series models have recently been developed to forecast renewable energy sources. In other research, a multiplicative decomposition method was designed to forecast future production of biohydrogen. The optimum length of the moving average (seasonal length) and start point, where the averages are placed, were indicated based on the best coincidence between the present forecast and actual values.[6]

Software

An example of statistical software for this type of decomposition is the program BV4.1 that is based on the so-called Berlin procedure.



An example of using multiplicative decomposition in biohydrogen production forecast.[6]

See also

- Hilbert–Huang transform
- Stochastic drift

References

1. "6.1 Time series components | OTexts" (https://www.otexts.org/fpp/6/1). *www.otexts.org*. Retrieved 2016-05-14.
2. Dodge, Y. (2003). *The Oxford Dictionary of Statistical Terms*. New York: Oxford University Press. ISBN 0-19-920613-9.
3. "6.1 Time series components | OTexts" (https://www.otexts.org/fpp/6/1). *www.otexts.org*. Retrieved 2016-05-18.
4. "6.5 STL decomposition | OTexts" (https://www.otexts.org/fpp/6/5). *www.otexts.org*. Retrieved 2016-05-18.
5. Kendall, M. G. (1976). *Time-Series* (Second ed.). Charles Griffin. (Fig. 5.1). ISBN 0-85264-241-5.
6. Asadi, Nooshin; Karimi Alavijeh, Masih; Zilouei, Hamid (2016). "Development of a mathematical methodology to investigate biohydrogen production from regional and national agricultural crop residues: A case study of Iran" (https://www.researchgate.net/publication/309382052). *International Journal of Hydrogen Energy*. doi:10.1016/j.ijhydene.2016.10.021 (https://doi.org/10.1016%2Fj.ijhydene.2016.10.021).

Further reading

- Enders, Walter (2004). "Models with Trend". *Applied Econometric Time Series* (Second ed.). New York: Wiley. pp. 156–238. ISBN 0-471-23065-0.

Autoregressive model

From Wikipedia, the free encyclopedia

In statistics and signal processing, an **autoregressive** (**AR**) **model** is a representation of a type of random process; as such, it is used to describe certain time-varying processes in nature, economics, etc. The autoregressive model specifies that the output variable depends linearly on its own previous values and on a stochastic term (an imperfectly predictable term); thus the model is in the form of a stochastic difference equation.

Together with the moving-average (MA) model, it is a special case and key component of the more general ARMA and ARIMA models of time series, which have a more complicated stochastic structure; it is also a special case of the vector autoregressive model (VAR), which consists of a system of more than one interlocking stochastic difference equation in more than one evolving random variable.

Contrary to the moving-average model, the autoregressive model is not always stationary as it may contain a unit root.

Contents

Definition

The notation $AR(p)$ indicates an autoregressive model of order *p*. The AR(*p*) model is defined as

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t$$

where $\varphi_1, \ldots, \varphi_p$ are the *parameters* of the model, c is a constant, and ε_t is white noise. This can be equivalently written using the backshift operator *B* as

$$X_t = c + \sum_{i=1}^{p} \varphi_i B^i X_t + \varepsilon_t$$

so that, moving the summation term to the left side and using polynomial notation, we have

$$\phi(B) X_t = c + \varepsilon_t \,.$$

An autoregressive model can thus be viewed as the output of an all-pole infinite impulse response filter whose input is white noise.

Some parameter constraints are necessary for the model to remain wide-sense stationary. For example, processes in the AR(1) model with $|\varphi_1| \geq 1$ are not stationary. More generally, for an AR(p) model to be wide-sense stationary, the roots of the polynomial $z^p - \sum_{i=1}^{p} \varphi_i z^{p-i}$ must lie within the unit circle, i.e., each root z_i must satisfy $|z_i| < 1$.

Intertemporal effect of shocks

In an AR process, a one-time shock affects values of the evolving variable infinitely far into the future. For example, consider the AR(1) model $X_t = c + \varphi_1 X_{t-1} + \varepsilon_t$. A non-zero value for ε_t at say time t=1 affects X_1 by the amount ε_1. Then by the AR equation for X_2 in terms of X_1, this affects X_2 by the amount $\varphi_1 \varepsilon_1$. Then by the AR equation for X_3 in terms of X_2, this affects X_3 by the amount $\varphi_1^2 \varepsilon_1$. Continuing this process shows that the effect of ε_1 never ends, although if the process is stationary then the effect diminishes toward zero in the limit.

Because each shock affects X values infinitely far into the future from when they occur, any given value X_t is affected by shocks occurring infinitely far into the past. This can also be seen by rewriting the autoregression

$$\phi(B) X_t = \varepsilon_t$$

(where the constant term has been suppressed by assuming that the variable has been measured as deviations from its mean) as

$$X_t = \frac{1}{\phi(B)} \varepsilon_t \,.$$

When the polynomial division on the right side is carried out, the polynomial in the backshift operator applied to ε_t has an infinite order—that is, an infinite number of lagged values of ε_t appear on the right side of the equation.

Characteristic polynomial

The autocorrelation function of an AR(p) process can be expressed as

$$\rho(\tau) = \sum_{k=1}^{p} a_k y_k^{-|\tau|},$$

where y_k are the roots of the polynomial

$$\phi(B) = 1 - \sum_{k=1}^{p} \varphi_k B^k$$

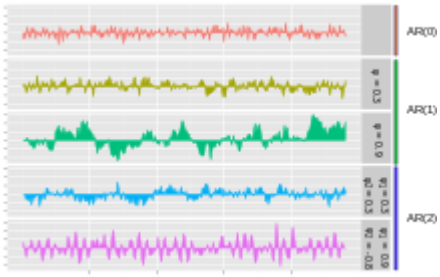
where B is the backshift operator, where $\phi(\cdot)$ is the function defining the autoregression, and where φ_k are the coefficients in the autoregression.

The autocorrelation function of an AR(p) process is a sum of decaying exponentials.

- Each real root contributes a component to the autocorrelation function that decays exponentially.
- Similarly, each pair of complex conjugate roots contributes an exponentially damped oscillation.

Graphs of AR(p) processes

The simplest AR process is AR(0), which has no dependence between the terms. Only the error/innovation/noise term contributes to the output of the process, so in the figure, AR(0) corresponds to white noise.



AR(0); AR(1) with AR parameter 0.3; AR(1) with AR parameter 0.9; AR(2) with AR parameters 0.3 and 0.3; and AR(2) with AR parameters 0.9 and −0.8

For an AR(1) process with a positive φ, only the previous term in the process and the noise term contribute to the output. If φ is close to 0, then the process still looks like white noise, but as φ approaches 1, the output gets a larger contribution from the previous term relative to the noise. This results in a "smoothing" or integration of the output, similar to a low pass filter.

For an AR(2) process, the previous two terms and the noise term contribute to the output. If both φ_1 and φ_2 are positive, the output will resemble a low pass filter, with the high frequency part of the noise decreased. If φ_1 is positive while φ_2 is negative, then the process favors changes in sign between terms of the process. The output oscillates. This can be likened to edge detection or detection of change in direction.

Example: An AR(1) process

An AR(1) process is given by:

$$X_t = c + \varphi X_{t-1} + \varepsilon_t$$

where ε_t is a white noise process with zero mean and constant variance σ_ε^2. (Note: The subscript on φ_1 has been dropped.) The process is wide-sense stationary if $|\varphi| < 1$ since it is obtained as the output of a stable filter whose input is white noise. (If $\varphi = 1$ then X_t has infinite variance, and is therefore not wide sense stationary.) Assuming $|\varphi| < 1$, the mean $\mathrm{E}(X_t)$ is identical for all values of t by the very definition of wide sense stationarity. If the mean is denoted by μ, it follows from

$$\mathrm{E}(X_t) = \mathrm{E}(c) + \varphi\,\mathrm{E}(X_{t-1}) + \mathrm{E}(\varepsilon_t),$$

that

$$\mu = c + \varphi\mu + 0,$$

and hence

$$\mu = \frac{c}{1-\varphi}.$$

In particular, if $c = 0$, then the mean is 0.

The variance is

$$\operatorname{var}(X_t) = \operatorname{E}(X_t^2) - \mu^2 = \frac{\sigma_\varepsilon^2}{1-\varphi^2},$$

where σ_ε is the standard deviation of ε_t. This can be shown by noting that

$$\operatorname{var}(X_t) = \varphi^2 \operatorname{var}(X_{t-1}) + \sigma_\varepsilon^2,$$

and then by noticing that the quantity above is a stable fixed point of this relation.

The autocovariance is given by

$$B_n = \operatorname{E}(X_{t+n} X_t) - \mu^2 = \frac{\sigma_\varepsilon^2}{1-\varphi^2}\ \varphi^{|n|}.$$

It can be seen that the autocovariance function decays with a decay time (also called time constant) of $\tau = -1/\ln(\varphi)$ [to see this, write $B_n = K\varphi^{|n|}$ where K is independent of n. Then note that $\varphi^{|n|} = e^{|n| \ln \varphi}$ and match this to the exponential decay law $e^{-n/\tau}$].

The spectral density function is the Fourier transform of the autocovariance function. In discrete terms this will be the discrete-time Fourier transform:

$$\Phi(\omega) = \frac{1}{\sqrt{2\pi}} \sum_{n=-\infty}^{\infty} B_n e^{-i\omega n} = \frac{1}{\sqrt{2\pi}} \left(\frac{\sigma_\varepsilon^2}{1 + \varphi^2 - 2\varphi\cos(\omega)} \right).$$

This expression is periodic due to the discrete nature of the X_j, which is manifested as the cosine term in the denominator. If we assume that the sampling time ($\Delta t = 1$) is much smaller than the decay time (τ), then we can use a continuum approximation to B_n:

$$B(t) \approx \frac{\sigma_\varepsilon^2}{1-\varphi^2}\ \varphi^{|t|}$$

which yields a Lorentzian profile for the spectral density:

$$\Phi(\omega) = \frac{1}{\sqrt{2\pi}}\ \frac{\sigma_\varepsilon^2}{1-\varphi^2}\ \frac{\gamma}{\pi(\gamma^2 + \omega^2)}$$

where $\gamma = 1/\tau$ is the angular frequency associated with the decay time τ.

An alternative expression for X_t can be derived by first substituting $c + \varphi X_{t-2} + \varepsilon_{t-1}$ for X_{t-1} in the defining equation. Continuing this process *N* times yields

$$X_t = c\sum_{k=0}^{N-1} \varphi^k + \varphi^N X_{t-N} + \sum_{k=0}^{N-1} \varphi^k \varepsilon_{t-k}.$$

For *N* approaching infinity, φ^N will approach zero and:

$$X_t = \frac{c}{1-\varphi} + \sum_{k=0}^{\infty} \varphi^k \varepsilon_{t-k}.$$

It is seen that X_t is white noise convolved with the φ^k kernel plus the constant mean. If the white noise ε_t is a Gaussian process then X_t is also a Gaussian process. In other cases, the central limit theorem indicates that X_t will be approximately normally distributed when φ is close to one.

Explicit mean/difference form of AR(1) process

The AR(1) model is the discrete time analogy of the continuous Ornstein-Uhlenbeck process. It is therefore sometimes useful to understand the properties of the AR(1) model cast in an equivalent form. In this form, the AR(1) model is given by:

$X_{t+1} = X_t + \theta(\mu - X_t) + \epsilon_{t+1}$, where $|\theta| < 1$ and μ is the model mean.

By putting this in the form $X_{t+1} = c + \phi X_t$, and then expanding the series for X_{t+n}, one can show that:

$\mathrm{E}(X_{t+n}|X_t) = \mu\left[1 - (1-\theta)^n\right] + X_t(1-\theta)^n$, and
$\mathrm{Var}(X_{t+n}|X_t) = \sigma^2 \dfrac{\left[1 - (1-\theta)^{2n}\right]}{1-(1-\theta)^2}$.

Choosing the maximum lag

The partial autocorrelation of an AR(p) process is zero at lag p + 1 and greater, so the appropriate maximum lag is the one beyond which the partial autocorrelations are all zero.

Calculation of the AR parameters

There are many ways to estimate the coefficients, such as the ordinary least squares procedure or method of moments (through Yule–Walker equations).

The AR(p) model is given by the equation

$$X_t = \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t.$$

It is based on parameters φ_i where i = 1, ..., p. There is a direct correspondence between these parameters and the covariance function of the process, and this correspondence can be inverted to determine the parameters from the autocorrelation function (which is itself obtained from the covariances). This is done using the Yule–Walker equations.

Yule–Walker equations

The Yule–Walker equations, named for Udny Yule and Gilbert Walker,[1][2] are the following set of equations.[3]

$$\gamma_m = \sum_{k=1}^{p} \varphi_k \gamma_{m-k} + \sigma_\varepsilon^2 \delta_{m,0},$$

where m = 0, ..., p, yielding p + 1 equations. Here γ_m is the autocovariance function of X_t, σ_ε is the standard deviation of the input noise process, and $\delta_{m,0}$ is the Kronecker delta function.

Because the last part of an individual equation is non-zero only if m = 0, the set of equations can be solved by representing the equations for $m > 0$ in matrix form, thus getting the equation

$$\begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \vdots \\ \gamma_p \end{bmatrix} = \begin{bmatrix} \gamma_0 & \gamma_{-1} & \gamma_{-2} & \cdots \\ \gamma_1 & \gamma_0 & \gamma_{-1} & \cdots \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \\ \gamma_{p-1} & \gamma_{p-2} & \gamma_{p-3} & \cdots \end{bmatrix} \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \varphi_3 \\ \vdots \\ \varphi_p \end{bmatrix}$$

which can be solved for all $\{\varphi_m; m = 1, 2, \cdots, p\}$. The remaining equation for *m* = 0 is

$$\gamma_0 = \sum_{k=1}^{p} \varphi_k \gamma_{-k} + \sigma_\varepsilon^2,$$

which, once $\{\varphi_m; m = 1, 2, \cdots, p\}$ are known, can be solved for σ_ε^2.

An alternative formulation is in terms of the autocorrelation function. The AR parameters are determined by the first p+1 elements $\rho(\tau)$ of the autocorrelation function. The full autocorrelation function can then be derived by recursively calculating [4]

$$\rho(\tau) = \sum_{k=1}^{p} \varphi_k \rho(k - \tau)$$

Examples for some Low-order AR(*p*) processes

- p=1
  - $\gamma_1 = \varphi_1 \gamma_0$
  - Hence $\rho_1 = \gamma_1 / \gamma_0 = \varphi_1$
- p=2
  - The Yule–Walker equations for an AR(2) process are

$$\begin{aligned} \gamma_1 &= \varphi_1 \gamma_0 + \varphi_2 \gamma_{-1} \\ \gamma_2 &= \varphi_1 \gamma_1 + \varphi_2 \gamma_0 \end{aligned}$$

    - Remember that $\gamma_{-k} = \gamma_k$
    - Using the first equation yields $\rho_1 = \gamma_1 / \gamma_0 = \dfrac{\varphi_1}{1 - \varphi_2}$
    - Using the recursion formula yields $\rho_2 = \gamma_2 / \gamma_0 = \dfrac{\varphi_1^2 - \varphi_2^2 + \varphi_2}{1 - \varphi_2}$

Estimation of AR parameters

The above equations (the Yule–Walker equations) provide several routes to estimating the parameters of an AR(*p*) model, by replacing the theoretical covariances with estimated values. Some of these variants can be described as follows:

- Estimation of autocovariances or autocorrelations. Here each of these terms is estimated separately, using conventional estimates. There are different ways of doing this and the choice between these affects the properties of the estimation scheme. For example, negative estimates of the variance can be produced by some choices.
- Formulation as a least squares regression problem in which an ordinary least squares prediction problem is constructed, basing prediction of values of X_t on the *p* previous values of the same series. This can be thought of as a forward-prediction scheme. The normal equations for this problem can be seen to correspond to an approximation of the matrix form of the Yule–Walker equations in which each appearance of an autocovariance of the same lag is replaced by a slightly different estimate.

- Formulation as an extended form of ordinary least squares prediction problem. Here two sets of prediction equations are combined into a single estimation scheme and a single set of normal equations. One set is the set of forward-prediction equations and the other is a corresponding set of backward prediction equations, relating to the backward representation of the AR model:

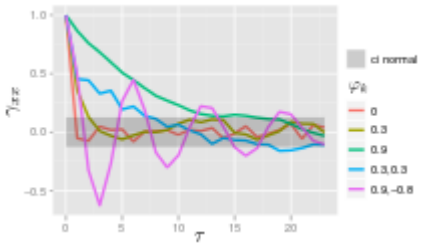
$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t^* \,.$$

  Here predicted of values of X_t would be based on the p future values of the same series. This way of estimating the AR parameters is due to Burg,[5] and is called the Burg method:[6] Burg and later authors called these particular estimates "maximum entropy estimates",[7] but the reasoning behind this applies to the use of any set of estimated AR parameters. Compared to the estimation scheme using only the forward prediction equations, different estimates of the autocovariances are produced, and the estimates have different stability properties. Burg estimates are particularly associated with maximum entropy spectral estimation.[8]

Other possible approaches to estimation include maximum likelihood estimation. Two distinct variants of maximum likelihood are available: in one (broadly equivalent to the forward prediction least squares scheme) the likelihood function considered is that corresponding to the conditional distribution of later values in the series given the initial p values in the series; in the second, the likelihood function considered is that corresponding to the unconditional joint distribution of all the values in the observed series. Substantial differences in the results of these approaches can occur if the observed series is short, or if the process is close to non-stationarity.

Spectrum

The power spectral density of an AR(p) process with noise variance $\operatorname{Var}(Z_t) = \sigma_Z^2$ is[4]

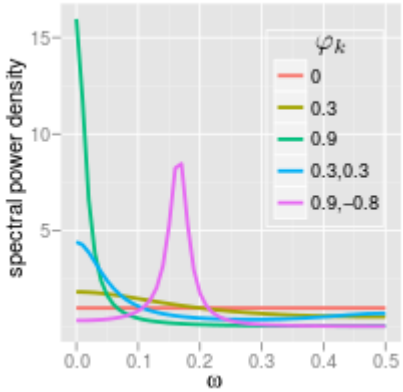


$$S(f) = \frac{\sigma_Z^2}{\left|1 - \sum_{k=1}^{p} \varphi_k e^{-i2\pi fk}\right|^2}.$$

AR(0)

For white noise (AR(0))

$$S(f) = \sigma_Z^2.$$



AR(1)

For AR(1)

$$S(f) = \frac{\sigma_Z^2}{\left|1 - \varphi_1 e^{-2\pi i f}\right|^2} = \frac{\sigma_Z^2}{1 + \varphi_1^2 - 2\varphi_1 \cos 2\pi f}$$

- If $\varphi_1 > 0$ there is a single spectral peak at f=0, often referred to as red noise. As φ_1 becomes nearer 1, there is stronger power at low frequencies, i.e. larger time lags. This is then a low-pass filter, when applied to full spectrum light, everything except for the red light will be filtered.
- If $\varphi_1 < 0$ there is a minimum at f=0, often referred to as blue noise. This similarly acts as a high-pass filter, everything except for blue light will be filtered.

AR(2)

AR(2) processes can be split into three groups depending on the characteristics of their roots:

$$z_1, z_2 = \frac{1}{2}\left(\varphi_1 \pm \sqrt{\varphi_1^2 + 4\varphi_2}\right)$$

- When $\varphi_1^2 + 4\varphi_2 < 0$, the process has a pair of complex-conjugate roots, creating a mid-frequency peak at:

$$f^* = \frac{1}{2\pi}\cos^{-1}\left(\frac{\varphi_1(\varphi_2 - 1)}{4\varphi_2}\right)$$

Otherwise the process has real roots, and:

- When $\varphi_1 > 0$ it acts as a low-pass filter on the white noise with a spectral peak at $f = 0$
- When $\varphi_1 < 0$ it acts as a high-pass filter on the white noise with a spectral peak at $f = 1/2$.

The process is non-stationary when the roots are outside the unit circle. The process is stable when the roots are within the unit circle, or equivalently when the coefficients are in the triangle $-1 \leq \varphi_2 \leq 1 - |\varphi_1|$.

The full PSD function can be expressed in real form as:

$$S(f) = \frac{\sigma_Z^2}{1 + \varphi_1^2 + \varphi_2^2 - 2\varphi_1(1 - \varphi_2)\cos(2\pi f) - 2\varphi_2\cos(4\pi f)}$$

Implementations in statistics packages

- R, the *stats* package includes an *ar* function.[9]
- MATLAB's Econometrics Toolbox [10] and System Identification Toolbox [11] includes autoregressive models [12]
- Matlab and Octave: the *TSA toolbox* contains several estimation functions for uni-variate, multivariate and adaptive autoregressive models.[13]

Impulse response

The impulse response of a system is the change in an evolving variable in response to a change in the value of a shock term *k* periods earlier, as a function of *k*. Since the AR model is a special case of the vector autoregressive model, the computation of the impulse response in Vector autoregression#Impulse response applies here.

n-step-ahead forecasting

Once the parameters of the autoregression

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t$$

have been estimated, the autoregression can be used to forecast an arbitrary number of periods into the future. First use *t* to refer to the first period for which data is not yet available; substitute the known prior values X_{t-i} for *i*=1, ..., *p* into the autoregressive equation while setting the error term ε_t equal to zero (because we forecast X_t to equal its expected value, and the expected value of the unobserved error term is zero). The output of the

autoregressive equation is the forecast for the first unobserved period. Next, use *t* to refer to the *next* period for which data is not yet available; again the autoregressive equation is used to make the forecast, with one difference: the value of *X* one period prior to the one now being forecast is not known, so its expected value—the predicted value arising from the previous forecasting step—is used instead. Then for future periods the same procedure is used, each time using one more forecast value on the right side of the predictive equation until, after *p* predictions, all *p* right-side values are predicted values from prior steps.

There are four sources of uncertainty regarding predictions obtained in this manner: (1) uncertainty as to whether the autoregressive model is the correct model; (2) uncertainty about the accuracy of the forecasted values that are used as lagged values in the right side of the autoregressive equation; (3) uncertainty about the true values of the autoregressive coefficients; and (4) uncertainty about the value of the error term ε_t for the period being predicted. Each of the last three can be quantified and combined to give a confidence interval for the *n*-step-ahead predictions; the confidence interval will become wider as *n* increases because of the use of an increasing number of estimated values for the right-side variables.

Evaluating the quality of forecasts

The predictive performance of the autoregressive model can be assessed as soon as estimation has been done if cross-validation is used. In this approach, some of the initially available data was used for parameter estimation purposes, and some (from available observations later in the data set) was held back for out-of-sample testing. Alternatively, after some time has passed after the parameter estimation was conducted, more data will have become available and predictive performance can be evaluated then using the new data.

In either case, there are two aspects of predictive performance that can be evaluated: one-step-ahead and *n*-step-ahead performance. For one-step-ahead performance, the estimated parameters are used in the autoregressive equation along with observed values of *X* for all periods prior to the one being predicted, and the output of the equation is the one-step-ahead forecast; this procedure is used to obtain forecasts for each of the out-of-sample observations. To evaluate the quality of *n*-step-ahead forecasts, the forecasting procedure in the previous section is employed to obtain the predictions.

Given a set of predicted values and a corresponding set of actual values for *X* for various time periods, a common evaluation technique is to use the mean squared prediction error; other measures are also available (see Forecasting#Forecasting accuracy).

The question of how to interpret the measured forecasting accuracy arises—for example, what is a "high" (bad) or a "low" (good) value for the mean squared prediction error? There are two possible points of comparison. First, the forecasting accuracy of an alternative model, estimated under different modeling assumptions or different estimation techniques, can be used for comparison purposes. Second, the out-of-sample accuracy measure can be compared to the same measure computed for the in-sample data points (that were used for parameter estimation) for which enough prior data values are available (that is, dropping the first *p* data points, for which *p* prior data points are not available). Since the model was estimated specifically to fit the in-sample points as well as possible, it will usually be the case that the out-of-sample predictive performance will be poorer than the in-sample predictive performance. But if the predictive quality deteriorates out-of-sample by "not very much" (which is not precisely definable), then the forecaster may be satisfied with the performance.

See also

- Moving average model
- Linear difference equation
- Predictive analytics
- Linear predictive coding
- Resonance

Notes

1. Yule, G. Udny (1927) "On a Method of Investigating Periodicities in Disturbed Series, with Special Reference to Wolfer's Sunspot Numbers" (http://visualiseur.bnf.fr/Visualiseur?Destination=Gallica&O=NUMM-56031), *Philosophical Transactions of the Royal Society of London*, Ser. A, Vol. 226, 267–298.]
2. Walker, Gilbert (1931) "On Periodicity in Series of Related Terms" (http://visualiseur.bnf.fr/Visualiseur?Destination=Gallica&O=NUMM-56224), *Proceedings of the Royal Society of London*, Ser. A, Vol. 131, 518–532.
3. Theodoridis, Sergios. "Chapter 1. Probability and Stochastic Processes". *Machine Learning: A Bayesian and Optimization Perspective*. Academic Press, 2015,. pp. 9–51. ISBN 978-0-12-801522-3.
4. Von Storch, H.; F. W Zwiers (2001). *Statistical analysis in climate research*. Cambridge Univ Pr. ISBN 0-521-01230-9.
5. Burg, J. P. (1968). "A new analysis technique for time series data". In *Modern Spectrum Analysis* (Edited by D. G. Childers), NATO Advanced Study Institute of Signal Processing with emphasis on Underwater Acoustics. IEEE Press, New York.
6. Brockwell, Peter J.; Dahlhaus, Rainer; Trindade, A. Alexandre (2005). "Modified Burg Algorithms for Multivariate Subset Autoregression" (https://web.archive.org/web/20121021015413/http://www3.stat.sinica.edu.tw/statistica/oldpdf/A15n112.pdf) (PDF). *Statistica Sinica*. **15**: 197–213. Archived from the original (http://www3.stat.sinica.edu.tw/statistica/oldpdf/A15n112.pdf) (PDF) on 2012-10-21.
7. Burg, J.P. (1967) "Maximum Entropy Spectral Analysis", *Proceedings of the 37th Meeting of the Society of* Exploration Geophysicists, *Oklahoma City, Oklahoma.*
8. Bos, R.; De Waele, S.; Broersen, P. M. T. (2002). "Autoregressive spectral estimation by application of the burg algorithm to irregularly sampled data". *IEEE Transactions on Instrumentation and Measurement*. **51** (6): 1289. doi:10.1109/TIM.2002.808031 (https://doi.org/10.1109%2FTIM.2002.808031).
9. "Fit Autoregressive Models to Time Series" (http://finzi.psych.upenn.edu/R/library/stats/html/ar.html) (in R)
10. Econometrics Toolbox Overview (http://www.mathworks.com/products/econometrics/)
11. System Identification Toolbox overview (http://www.mathworks.com/products/sysid/)
12. "Autoregressive modeling in MATLAB" (http://www.mathworks.com/help/econ/autoregressive-model.html)
13. "Time Series Analysis toolbox for Matlab and Octave" (http://pub.ist.ac.at/~schloegl/matlab/tsa/)

References

- Mills, Terence C. (1990). *Time Series Techniques for Economists*. Cambridge University Press.
- Percival, Donald B.; Walden, Andrew T. (1993). *Spectral Analysis for Physical Applications*. Cambridge University Press.
- Pandit, Sudhakar M.; Wu, Shien-Ming (1983). *Time Series and System Analysis with Applications*. John Wiley & Sons.

External links

- AutoRegression Analysis (AR) by Paul Bourke
- Econometrics lecture (topic: Autoregressive models) on YouTube by Mark Thoma

Moving-average model

From Wikipedia, the free encyclopedia

In time series analysis, the **moving-average** (**MA**) **model** is a common approach for modeling univariate time series. The moving-average model specifies that the output variable depends linearly on the current and various past values of a stochastic (imperfectly predictable) term.

Together with the autoregressive (AR) model, the moving-average model is a special case and key component of the more general ARMA and ARIMA models of time series, which have a more complicated stochastic structure.

The moving-average model should not be confused with the moving average, a distinct concept despite some similarities.

Contrary to the AR model, the finite MA model is always stationary.

Contents

Definition

The notation MA(q) refers to the moving average model of order q:

$$X_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

where μ is the mean of the series, the θ_1, ..., θ_q are the parameters of the model and the ε_t, ε_{t-1},..., $\varepsilon_{t-\mathrm{q}}$ are white noise error terms. The value of q is called the order of the MA model. This can be equivalently written in terms of the backshift operator B as

$$X_t = \mu + (1 + \theta_1 B + \cdots + \theta_q B^q)\varepsilon_t.$$

Thus, a moving-average model is conceptually a linear regression of the current value of the series against current and previous (unobserved) white noise error terms or random shocks. The random shocks at each point are assumed to be mutually independent and to come from the same distribution, typically a normal distribution, with location at zero and constant scale.

Interpretation

The moving-average model is essentially a finite impulse response filter applied to white noise, with some additional interpretation placed on it. The role of the random shocks in the MA model differs from their role in the autoregressive (AR) model in two ways. First, they are propagated to future values of the time series directly: for example, ε_{t-1} appears directly on the right side of the equation for X_t. In contrast, in an AR model ε_{t-1} does not appear on the right side of the X_t equation, but it does appear on the right side of the

X_{t-1} equation, and X_{t-1} appears on the right side of the X_t equation, giving only an indirect effect of ε_{t-1} on X_t. Second, in the MA model a shock affects X values only for the current period and *q* periods into the future; in contrast, in the AR model a shock affects X values infinitely far into the future, because ε_t affects X_t, which affects X_{t+1}, which affects X_{t+2}, and so on forever (see Vector autoregression#Impulse response).

Deciding appropriateness of the MA model

Sometimes the autocorrelation function (ACF) and partial autocorrelation function (PACF) will suggest that an MA model would be a better model choice and sometimes both AR and MA terms should be used in the same model (see Box–Jenkins method#Identify p and q).

Fitting the model

Fitting the MA estimates is more complicated than with autoregressive models (AR models) because the lagged error terms are not observable. This means that iterative non-linear fitting procedures need to be used in place of linear least squares.

Choosing the order q

The autocorrelation function of an MA(*q*) process becomes zero at lag *q* + 1 and greater, so we determine the appropriate maximum lag for the estimation by examining the sample autocorrelation function to see where it becomes insignificantly different from zero for all lags beyond a certain lag, which is designated as the maximum lag *q*.

See also

- Autoregressive–moving-average model
- Autoregressive model

Further reading

- Enders, Walter (2004). "Stationary Time-Series Models". *Applied Econometric Time Series* (Second ed.). New York: Wiley. pp. 48–107. ISBN 0-471-45173-8.

External links

- Common approaches to univariate time series

Autoregressive integrated moving average

From Wikipedia, the free encyclopedia

In statistics and econometrics, and in particular in time series analysis, an **autoregressive integrated moving average (ARIMA)** model is a generalization of an autoregressive moving average (ARMA) model. Both of these models are fitted to time series data either to better understand the data or to predict future points in the series (forecasting). ARIMA models are applied in some cases where data show evidence of non-stationarity, where an initial differencing step (corresponding to the "integrated" part of the model) can be applied one or more times to eliminate the non-stationarity.[1]

The AR part of ARIMA indicates that the evolving variable of interest is regressed on its own lagged (i.e., prior) values. The MA part indicates that the regression error is actually a linear combination of error terms whose values occurred contemporaneously and at various times in the past. The I (for "integrated") indicates that the data values have been replaced with the difference between their values and the previous values (and this differencing process may have been performed more than once). The purpose of each of these features is to make the model fit the data as well as possible.

Non-seasonal ARIMA models are generally denoted ARIMA(p,d,q) where parameters p, d, and q are non-negative integers, p is the order (number of time lags) of the autoregressive model, d is the degree of differencing (the number of times the data have had past values subtracted), and q is the order of the moving-average model. Seasonal ARIMA models are usually denoted ARIMA$(p,d,q)(P,D,Q)_m$, where m refers to the number of periods in each season, and the uppercase P,D,Q refer to the autoregressive, differencing, and moving average terms for the seasonal part of the ARIMA model.[2][3]

When two out of the three terms are zeros, the model may be referred to based on the non-zero parameter, dropping "AR", "I" or "MA" from the acronym describing the model. For example, ARIMA (1,0,0) is AR(1), ARIMA(0,1,0) is I(1), and ARIMA(0,0,1) is MA(1).

ARIMA models can be estimated following the Box–Jenkins approach.

Contents

Definition

Given a time series of data X_t where t is an integer index and the X_t are real numbers, an ARMA(p,q) model is given by

$$X_t - \alpha_1 X_{t-1} - \cdots - \alpha_{p'} X_{t-p'} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q},$$

or equivalently by

$$\left(1 - \sum_{i=1}^{p'} \alpha_i L^i\right) X_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \varepsilon_t$$

where L is the lag operator, the α_i are the parameters of the autoregressive part of the model, the θ_i are the parameters of the moving average part and the ε_t are error terms. The error terms ε_t are generally assumed to be independent, identically distributed variables sampled from a normal distribution with zero mean.

Assume now that the polynomial $\left(1 - \sum_{i=1}^{p'} \alpha_i L^i\right)$ has a unit root (a factor $(1 - L)$) of multiplicity *d*. Then it can be rewritten as:

$$\left(1 - \sum_{i=1}^{p'} \alpha_i L^i\right) = \left(1 - \sum_{i=1}^{p'-d} \phi_i L^i\right) (1 - L)^d.$$

An ARIMA(*p*,*d*,*q*) process expresses this polynomial factorisation property with $p = p' - d$, and is given by:

$$\left(1 - \sum_{i=1}^{p} \phi_i L^i\right) (1 - L)^d X_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \varepsilon_t$$

and thus can be thought as a particular case of an ARMA(*p*+*d*,*q*) process having the autoregressive polynomial with *d* unit roots. (For this reason, no ARIMA model with $d > 0$ is wide sense stationary.)

The above can be generalized as follows.

$$\left(1 - \sum_{i=1}^{p} \phi_i L^i\right) (1 - L)^d X_t = \delta + \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \varepsilon_t.$$

This defines an ARIMA(*p*,*d*,*q*) process with **drift** $\delta/(1 - \Sigma\varphi_i)$.

Other special forms

The explicit identification of the factorisation of the autoregression polynomial into factors as above, can be extended to other cases, firstly to apply to the moving average polynomial and secondly to include other special factors. For example, having a factor $(1 - L^s)$ in a model is one way of including a non-stationary seasonality of period *s* into the model; this factor has the effect of re-expressing the data as changes from *s* periods ago. Another example is the factor $\left(1 - \sqrt{3}L + L^2\right)$, which includes a (non-stationary) seasonality of period 2. The effect of the first type of factor is to allow each season's value to drift separately over time, whereas with the second type values for adjacent seasons move together.

Identification and specification of appropriate factors in an ARIMA model can be an important step in modelling as it can allow a reduction in the overall number of parameters to be estimated, while allowing the imposition on the model of types of behaviour that logic and experience suggest should be there.

Differencing

Differencing in statistics is a transformation applied to time-series data in order to make it stationary. A stationary time series' properties do not depend on the time at which the series is observed.

In order to difference the data, the difference between consecutive observations is computed. Mathematically, this is shown as

$$y'_t = y_t - y_{t-1}$$

Differencing removes the changes in the level of a time series, eliminating trend and seasonality and consequently stabilizing the mean of the time series.

Sometimes it may be necessary to difference the data a second time to obtain a stationary time series, which is referred to as **second order differencing**:

$$\begin{aligned} y^*_t &= y'_t - y'_{t-1} \\ &= (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) \\ &= y_t - 2y_{t-1} + y_{t-2} \end{aligned}$$

Another method of differencing data is **seasonal differencing**, which involves computing the difference between an observation and the corresponding observation in the previous year. This is shown as:

$$y'_t = y_t - y_{t-m} \quad \text{where } m = \text{number of seasons.}$$

The differenced data is then used for the estimation of an ARMA model.

Examples

Some well-known special cases arise naturally or are mathematically equivalent to other popular forecasting models. For example:

- An ARIMA(0,1,0) model (or I(1) model) is given by $X_t = X_{t-1} + \varepsilon_t$ — which is simply a random walk.
- An ARIMA(0,1,0) with a constant, given by $X_t = c + X_{t-1} + \varepsilon_t$ — which is a random walk with drift.
- An ARIMA(0,0,0) model is a white noise model.
- An ARIMA(0,1,2) model is a Damped Holt's model.
- An ARIMA(0,1,1) model without constant is a basic exponential smoothing model.[4]
- An ARIMA(0,2,2) model is given by $X_t = 2X_{t-1} - X_{t-2} + (\alpha + \beta - 2)\varepsilon_{t-1} + (1 - \alpha)\varepsilon_{t-2} + \varepsilon_t$ — which is equivalent to Holt's linear method with additive errors, or double exponential smoothing.[4]

Choosing the order

To determine the order of a non-seasonal ARIMA model, a useful criterion is the Akaike information criterion (AIC) . It is written as

$$\text{AIC} = -2\log(L) + 2(p + q + k + 1),$$

where L is the likelihood of the data, p is the order of the autoregressive part and q is the order of the moving average part. The parameter k in this criterion is defined as the number of parameters in the model being fitted to the data. For AIC, if $k = 1$ then $c \neq 0$ and if $k = 0$ then $c = 0$.

The corrected AIC for ARIMA models can be written as

$$AICc = AIC + (2(p + q + k + 1)(p + q + k + 2))/(T - p - q - k - 2).$$

The Bayesian Information Criterion can be written as

$$BIC = AIC + (\log(T) - 2)(p + q + k + 1).$$

The objective is to minimize the AIC, AICc or BIC values for a good model. The lower the value of one of these criteria for a range of models being investigated, the better the model will suit the data. It should be noted however that the AIC and the BIC are used for two completely different purposes. Whilst the AIC tries to approximate models towards the reality of the situation, the BIC attempts to find the perfect fit. The BIC approach is often criticized as there never is a perfect fit to real-life complex data; however, it is still a useful method for selection as it penalizes models more heavily for having more parameters than the AIC would.

AICc can only be used to compare ARIMA models with the same orders of differencing. For ARIMAs with different orders of differencing, RMSE can be used for model comparison.

Estimation of coefficients

Forecasts using ARIMA models

The ARIMA model can be viewed as a "cascade" of two models. The first is non-stationary:

$$Y_t = (1-L)^d X_t$$

while the second is wide-sense stationary:

$$\left(1 - \sum_{i=1}^{p} \phi_i L^i\right) Y_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right) \varepsilon_t \,.$$

Now forecasts can be made for the process Y_t, using a generalization of the method of autoregressive forecasting.

Forecast intervals

The forecast intervals (confidence intervals for forecasts) for ARIMA models are based on assumptions that the residuals are uncorrelated and normally distributed. If either of these assumptions does not hold, then the forecast intervals may be incorrect. For this reason, researchers plot the ACF and histogram of the residuals to check the assumptions before producing forecast intervals.

95% forecast interval: $\hat{y}_{T+h|T} \pm 1.96\sqrt{v_{T+h|T}}$, where $v_{T+h|T}$ is the variance of $y_{T+h}|y_1, \ldots, y_T$.

For $h = 1$, $v_{T+h|T} = \hat{\sigma}^2$ for all ARIMA models regardless of parameters and orders.

For ARIMA(0,0,q), $y_t = e_t + \sum_{i=1}^{q} \theta_i e_{t-i}$.

$$v_{T+h|T} = \hat{\sigma}^2 \left[1 + \sum_{i=1}^{h-1} \theta_i e_{t-i}\right], \text{ for } h = 2, 3, \ldots$$

In general, forecast intervals from ARIMA models will increase as the forecast horizon increases.

Variations and extensions

A number of variations on the ARIMA model are commonly employed. If multiple time series are used then the X_t can be thought of as vectors and a VARIMA model may be appropriate. Sometimes a seasonal effect is suspected in the model; in that case, it is generally better to use a SARIMA (seasonal ARIMA) model than to

increase the order of the AR or MA parts of the model. If the time-series is suspected to exhibit long-range dependence, then the *d* parameter may be allowed to have non-integer values in an autoregressive fractionally integrated moving average model, which is also called a Fractional ARIMA (FARIMA or ARFIMA) model.

Software implementations

Various packages that apply methodology like Box–Jenkins parameter optimization are available to find the right parameters for the ARIMA model.

- EViews: has extensive ARIMA and SARIMA capabilities.
- Julia: contains an ARIMA implementation in the TimeModels package[5]
- Mathematica: includes ARIMAProcess function.
- MATLAB: the Econometrics Toolbox includes ARIMA models and regression with ARIMA errors
- NCSS: includes several procedures for `ARIMA` fitting and forecasting.[6][7][8]
- Python: the "statsmodels" package includes models for time series analysis – univariate time series analysis: AR, ARIMA – vector autoregressive models, VAR and structural VAR – descriptive statistics and process models for time series analysis.
- R: the standard R *stats* package includes an *arima* function, which is documented in "ARIMA Modelling of Time Series". Besides the *ARIMA(p,d,q)* part, the function also includes seasonal factors, an intercept term, and exogenous variables (*xreg*, called "external regressors"). The CRAN task view on Time Series is the reference with many more links. The "forecast" package in R can automatically select an ARIMA model for a given time series with the auto.arima() function. The package can also simulate seasonal and non-seasonal ARIMA models with its simulate.Arima() function. It also has a function Arima(), which is a wrapper for the arima from the "stats" package.[9]
- Ruby: the "statsample-timeseries" gem is used for time series analysis, including ARIMA models and Kalman Filtering.
- SAFE TOOLBOXES: includes ARIMA modelling and regression with ARIMA errors.
- SAS: includes extensive ARIMA processing in its Econometric and Time Series Analysis system: SAS/ETS.
- IBM SPSS: includes ARIMA modeling in its Statistics and Modeler statistical packages. The default Expert Modeler feature evaluates a range of seasonal and non-seasonal autoregressive (*p*), integrated (*d*), and moving average (*q*) settings and seven exponential smoothing models. The Expert Modeler can also transform the target time-series data into its square root or natural log. The user also has the option to restrict the Expert Modeler to ARIMA models, or to manually enter ARIMA nonseasonal and seasonal *p*, *d*, and *q* settings without Expert Modeler. Automatic outlier detection is available for seven types of outliers, and the detected outliers will be accommodated in the time-series model if this feature is selected.
- SAP: the APO-FCS package[10] in SAP ERP from SAP allows creation and fitting of ARIMA models using the Box–Jenkins methodology.
- SQL Server Analysis Services: from Microsoft includes ARIMA as a Data Mining algorithm.
- Stata includes ARIMA modelling (using its arima command) as of Stata 9.
- TOL (Time Oriented Language) is designed to model ARIMA models (including SARIMA, ARIMAX and DSARIMAX variants) [1].
- Scala: spark-timeseries library contains ARIMA implementation for Scala, Java and Python. Implementation is designed to run on Apache Spark.

See also

- Autocorrelation
- ARMA
- X-12-ARIMA
- Partial autocorrelation

References

1. For further information on Stationarity and Differencing see https://www.otexts.org/fpp/8/1
2. "Notation for ARIMA Models" (https://support.sas.com/documentation/cdl/en/etsug/63939/HTML/default/viewer.htm#etsug_tffordet_sect016.htm). *Time Series Forecasting System*. SAS Institute. Retrieved 19 May 2015.
3. Hyndman, Rob J; Athanasopoulos, George. "8.9 Seasonal ARIMA models" (https://www.otexts.org/fpp/8/9). *Forecasting: principles and practice*. oTexts. Retrieved 19 May 2015.
4. "Introduction to ARIMA models" (http://people.duke.edu/~rnau/411arim.htm#arima010). *people.duke.edu*. Retrieved 2016-06-05.
5. TimeModels.jl (https://github.com/JuliaStats/TimeModels.jl) www.github.com
6. ARIMA in NCSS (http://ncss.wpengine.netdna-cdn.com/wp-content/themes/ncss/pdf/Procedures/NCSS/ARIMA-Box-Jenkins.pdf),
7. Automatic ARMA in NCSS (http://ncss.wpengine.netdna-cdn.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Automatic_ARMA.pdf),
8. Autocorrelations and Partial Autocorrelations in NCSS (http://ncss.wpengine.netdna-cdn.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Autocorrelations.pdf)
9. "8.7 ARIMA modelling in R | OTexts" (https://www.otexts.org/fpp/8/7). *www.otexts.org*. Retrieved 2016-05-12.
10. "Box Jenkins model" (http://help.sap.com/saphelp_45b/helpdata/en/35/8a524b52060634e10000009b38f9b9/content.htm). SAP. Retrieved 8 March 2013.

Further reading

- Asteriou, Dimitros; Hall, Stephen G. (2011). "ARIMA Models and the Box–Jenkins Methodology". *Applied Econometrics* (Second ed.). Palgrave MacMillan. pp. 265–286. ISBN 978-0-230-27182-1.
- Mills, Terence C. (1990). *Time Series Techniques for Economists*. Cambridge University Press. ISBN 0-521-34339-9.
- Percival, Donald B.; Walden, Andrew T. (1993). *Spectral Analysis for Physical Applications*. Cambridge University Press. ISBN 0-521-35532-X.

External links

- The US Census Bureau uses ARIMA for "seasonally adjusted" data (programs, docs, and papers here)
- Lecture notes on ARIMA models (Robert Nau, Duke University)

Exponential smoothing

From Wikipedia, the free encyclopedia

Exponential smoothing is a rule of thumb technique for smoothing time series data. Whereas in the simple moving average the past observations are weighted equally, exponential functions are used to assign exponentially decreasing weights over time. It is an easily learned and easily applied procedure for making some determination based on prior assumptions by the user, such as seasonality. Exponential smoothing is used for analysis of financial time-series data as well as the field of signal processing.

Exponential smoothing is commonly applied to smooth data, as many window functions are in signal processing, acting as low-pass filters to remove high frequency noise. This method is proceeded by Poisson's use of recursive exponential window functions in convolutions from the 19th century, as well as Kolmogorov and Zurbenko's use of recursive moving averages from their studies of turbulence in the 1940s.

The raw data sequence is often represented by $\{x_t\}$ beginning at time $t = 0$, and the output of the exponential smoothing algorithm is commonly written as $\{s_t\}$, which may be regarded as a best estimate of what the next value of x will be. When the sequence of observations begins at time $t = 0$, the simplest form of exponential smoothing is given by the formulas:[1]

$$s_0 = x_0$$
$$s_t = \alpha x_t + (1-\alpha)s_{t-1},\ t > 0$$

where α is the *smoothing factor*, and $0 < \alpha < 1$.

Contents

Background

Window functions

The simple moving average (SMA)

Intuitively, the simplest way to smooth a time series is to calculate a simple, or unweighted, moving average. This is known as using a rectangular or "boxcar" window function. The smoothed statistic s_t is then just the mean of the last k observations:

$$s_t = \frac{1}{k}\sum_{n=0}^{k-1} x_{t-n} = \frac{x_t + x_{t-1} + x_{t-2} + \cdots + x_{t-k+1}}{k} = s_{t-1} + \frac{x_t - x_{t-k}}{k},$$

where the choice of an integer $k > 1$ is arbitrary. A small value of k will have less of a smoothing effect and be more responsive to recent changes in the data, while a larger k will have a greater smoothing effect, and produce a more pronounced lag in the smoothed sequence. One disadvantage of this technique is that it cannot be used on the first $k - 1$ terms of the time series without the addition of values created by some other means. This means effectively extrapolating outside the existing data, and the validity of this section would therefore be questionable and not a direct representation of the data. However, as long as the time series contains at least k values, this has no effect on forecasts of future values.

It also introduces a phase shift into the data of half the window length. For example, if the data were all the same except for one high data point, the peak in the "smoothed" data would appear half a window length later than when it actually occurred. Where the phase of the result is important, this can be simply corrected by shifting the resulting series back by half the window length.

A major drawback of the SMA is that it lets through a significant amount of the signal shorter than the window length. Worse, it actually inverts it. This can lead to unexpected artifacts, such as peaks in the "smoothed" result appearing where there were troughs in the data. It also leads to the result being less "smooth" than expected since some of the higher frequencies are not properly removed.

The weighted moving average

A slightly more intricate method for smoothing a raw time series $\{x_t\}$ is to calculate a weighted moving average by first choosing a set of weighting factors

$$\{w_1, w_2, \ldots, w_k\} \text{ such that } \sum_{n=1}^{k} w_n = 1$$

and then using these weights to calculate the smoothed statistics $\{s_t\}$:

$$s_t = \sum_{n=1}^{k} w_n x_{t+1-n} = w_1 x_t + w_2 x_{t-1} + \cdots + w_k x_{t-k+1}.$$

In practice the weighting factors are often chosen to give more weight to the most recent terms in the time series and less weight to older data. Notice that this technique has the same disadvantage as the simple moving average technique (i.e., it cannot be used until at least k observations have been made), and that it entails a more complicated calculation at each step of the smoothing procedure. In addition to this disadvantage, if the data from each stage of the averaging is not available for analysis, it may be difficult if not impossible to reconstruct a changing signal accurately (because older samples may be given less weight). If the number of stages missed is known however, the weighting of values in the average can be adjusted to give equal weight to all missed samples to avoid this issue.

Basic exponential smoothing

The use of the exponential window function is first attributed to Poisson[2] as an extension of a numerical analysis technique from the 17th century, and later adopted by the signal processing community in the 1940s. Here, exponential smoothing is the application of the exponential, or Poisson, window function. Exponential

smoothing was first suggested in the statistical literature without citation to previous work by Robert Goodell Brown in 1956,[3] and then expanded by Charles C. Holt in 1957.[4] The formulation below, which is the one commonly used, is attributed to Brown and is known as "Brown's simple exponential smoothing".[5] All the methods of Holt, Winters and Brown may be seen as a simple application of recursive filtering, first found in the 1940s[2] to convert FIR filters to IIR filters.

The simplest form of exponential smoothing is given by the formula:

$$s_t = \alpha \cdot x_t + (1 - \alpha) \cdot s_{t-1}.$$

where α is the *smoothing factor*, and 0 < α < 1. In other words, the smoothed statistic s_t is a simple weighted average of the current observation x_t and the previous smoothed statistic s_{t-1}. The term *smoothing factor* applied to α here is something of a misnomer, as larger values of α actually reduce the level of smoothing, and in the limiting case with α = 1 the output series is just the current observation. Simple exponential smoothing is easily applied, and it produces a smoothed statistic as soon as two observations are available.

Values of α close to one have less of a smoothing effect and give greater weight to recent changes in the data, while values of α closer to zero have a greater smoothing effect and are less responsive to recent changes. There is no formally correct procedure for choosing α. Sometimes the statistician's judgment is used to choose an appropriate factor. Alternatively, a statistical technique may be used to *optimize* the value of α. For example, the method of least squares might be used to determine the value of α for which the sum of the quantities $(s_t - x_{t+1})^2$ is minimized.[6]

Unlike some other smoothing methods, such as the simple moving average, this technique does not require any minimum number of observations to be made before it begins to produce results. In practice, however, a "good average" will not be achieved until several samples have been averaged together; for example, a constant signal will take approximately *3/α* stages to reach 95% of the actual value. To accurately reconstruct the original signal without information loss all stages of the exponential moving average must also be available, because older samples decay in weight exponentially. This is in contrast to a simple moving average, in which some samples can be skipped without as much loss of information due to the constant weighting of samples within the average. If a known number of samples will be missed, one can adjust a weighted average for this as well, by giving equal weight to the new sample and all those to be skipped.

This simple form of exponential smoothing is also known as an exponentially weighted moving average (EWMA). Technically it can also be classified as an autoregressive integrated moving average (ARIMA) (0,1,1) model with no constant term.[7]

Time Constant

The time constant of an exponential moving average is the amount of time for the smoothed response of a unit set function to reach $1 - 1/e \approx 63.2\,\%$ of the original signal. The relationship between this time constant, τ, and the smoothing factor, α, is given by the formula:

$$\alpha = 1 - e^{\frac{-\Delta T}{\tau}}$$

Where ΔT is the sampling time interval of the discrete time implementation. If the sampling time is fast compared to the time constant then

$$\alpha \approx \frac{\Delta T}{\tau}$$

Choosing the initial smoothed value

Note that in the above definition s_1 is being initialized to x_0. Because exponential smoothing requires that at each stage we have the previous forecast, it is not obvious how to get the method started. We could assume that the initial forecast is equal to the initial value of demand; however, this approach has a serious drawback. Exponential smoothing puts substantial weight on past observations, so the initial value of demand will have an unreasonably large effect on early forecasts. This problem can be overcome by allowing the process to evolve for a reasonable number of periods (10 or more) and using the average of the demand during those periods as the initial forecast. There are many other ways of setting this initial value, but it is important to note that the smaller the value of α, the more sensitive your forecast will be on the selection of this initial smoother value s_1.[8]

Optimization

For every exponential smoothing method we also need to choose the value for the smoothing parameters. For simple exponential smoothing, there is only one smoothing parameter (α), but for the methods that follow there is usually more than one smoothing parameter.

There are cases where the smoothing parameters may be chosen in a subjective manner — the forecaster specifies the value of the smoothing parameters based on previous experience. However, a more robust and objective way to obtain values for the unknown parameters included in any exponential smoothing method is to estimate them from the observed data.

The unknown parameters and the initial values for any exponential smoothing method can be estimated by minimizing the SSE. The errors are specified as $e_t = y_t - \hat{y}_{t|t-1}$ for t=1,…,T (the one-step-ahead within-sample forecast errors). Hence we find the values of the unknown parameters and the initial values that minimize

$$SSE = \sum_{t=1}^{T}(y_t - \hat{y}_{t|t-1})^2 = \sum_{t=1}^{T} e_t^2$$ [9]

Unlike the regression case (where we have formulas that return the values of the regression coefficients which minimize the SSE) this involves a non-linear minimization problem and we need to use an optimization tool to perform this.

Why is it “exponential”?

The name 'exponential smoothing' is attributed to the use of the exponential window function during convolution. It is no longer attributed to Holt, Winters & Brown.

By direct substitution of the defining equation for simple exponential smoothing back into itself we find that

$$\begin{aligned} s_t &= \alpha x_t + (1-\alpha)s_{t-1} \\ &= \alpha x_t + \alpha(1-\alpha)x_{t-1} + (1-\alpha)^2 s_{t-2} \\ &= \alpha\left[x_t + (1-\alpha)x_{t-1} + (1-\alpha)^2 x_{t-2} + (1-\alpha)^3 x_{t-3} + \cdots + (1-\alpha)^{t-1} x_1\right] + (1-\alpha)^t x_0. \end{aligned}$$

In other words, as time passes the smoothed statistic s_t becomes the weighted average of a greater and greater number of the past observations x_{t-n}, and the weights assigned to previous observations are in general proportional to the terms of the geometric progression {1, (1 − α), $(1 - \alpha)^2$, $(1 - \alpha)^3$, ...}. A geometric progression is the discrete version of an exponential function, so this is where the name for this smoothing method originated according to Statistics lore.

Comparison with moving average

Exponential smoothing and moving average have similar defects of introducing a lag relative to the input data. While this can be corrected by shifting the result by half the window length for a symmetrical kernel, such as a moving average or gaussian, it is unclear how appropriate this would be for exponential smoothing. They also both have roughly the same distribution of forecast error when $\alpha = 2/(k+1)$. They differ in that exponential smoothing takes into account all past data, whereas moving average only takes into account k past data points. Computationally speaking, they also differ in that moving average requires that the past k data points, or the data point at lag $k+1$ plus the most recent forecast value, to be kept, whereas exponential smoothing only needs the most recent forecast value to be kept.[10]

In the signal processing literature, the use of non-causal (symmetric) filters is commonplace, and the exponential window function is broadly used in this fashion, but a different terminology is used: exponential smoothing is equivalent to a first-order Infinite Impulse Response or IIR filter and moving average is equivalent to a Finite Impulse Response or FIR filter with equal weighting factors.

Double exponential smoothing

Simple exponential smoothing does not do well when there is a trend in the data, which is inconvenient.[1] In such situations, several methods were devised under the name "double exponential smoothing" or "second-order exponential smoothing.", which is the recursive application of an exponential filter twice, thus being termed "double exponential smoothing". This nomenclature is similar to quadruple exponential smoothing, which also references its recursion depth.[11] The basic idea behind double exponential smoothing is to introduce a term to take into account the possibility of a series exhibiting some form of trend. This slope component is itself updated via exponential smoothing.

One method, sometimes referred to as "Holt-Winters double exponential smoothing"[12] works as follows:[13]

Again, the raw data sequence of observations is represented by $\{x_t\}$, beginning at time $t = 0$. We use $\{s_t\}$ to represent the smoothed value for time t, and $\{b_t\}$ is our best estimate of the trend at time t. The output of the algorithm is now written as F_{t+m}, an estimate of the value of x at time $t+m$, $m>0$ based on the raw data up to time t. Double exponential smoothing is given by the formulas

$$s_1 = x_1$$
$$b_1 = x_1 - x_0$$

And for t > 2 by

$$s_t = \alpha x_t + (1-\alpha)(s_{t-1} + b_{t-1})$$
$$b_t = \beta(s_t - s_{t-1}) + (1-\beta)b_{t-1}$$

where α is the *data smoothing factor*, 0 < α < 1, and β is the *trend smoothing factor*, 0 < β < 1.

To forecast beyond x_t

$$F_{t+m} = s_t + mb_t$$

Setting the initial value b_0 is a matter of preference. An option other than the one listed above is $(x_n - x_0)/n$ for some $n > 1$.

Note that F_0 is undefined (there is no estimation for time 0), and according to the definition $F_1=s_0+b_0$, which is well defined, thus further values can be evaluated.

A second method, referred to as either Brown's linear exponential smoothing (LES) or Brown's double exponential smoothing works as follows.[14]

$$
\begin{aligned}
s'_0 &= x_0 \\
s''_0 &= x_0 \\
s'_t &= \alpha x_t + (1-\alpha)s'_{t-1} \\
s''_t &= \alpha s'_t + (1-\alpha)s''_{t-1} \\
F_{t+m} &= a_t + mb_t,
\end{aligned}
$$

where a_t, the estimated level at time t and b_t, the estimated trend at time t are:

$$
\begin{aligned}
a_t &= 2s'_t - s''_t \\
b_t &= \frac{\alpha}{1-\alpha}(s'_t - s''_t).
\end{aligned}
$$

Triple exponential smoothing

Triple exponential smoothing takes into account seasonal changes as well as trends (all of which are trends). Seasonality is defined to be the tendency of time-series data to exhibit behavior that repeats itself every L periods, much like any periodic function. The term season is used to represent the period of time before behavior begins to repeat itself. There are different types of seasonality: 'multiplicative' and 'additive' in nature, much like addition and multiplication are basic operations in mathematics.

If every month of December we sell 10,000 more apartments than we do in November the seasonality is *additive* in nature. Can be represented by an 'absolute' increase. However, if we sell 10% more apartments in the summer months than we do in the winter months the seasonality is *multiplicative* in nature. Multiplicative seasonality can be represented as a constant factor, not an absolute amount. [15]

Triple exponential smoothing was first suggested by Holt's student, Peter Winters, in 1960 after reading a signal processing book from the 1940s on exponential smoothing.[16] Holt's novel idea was to repeat filtering an odd number of times (ignoring 1) . While recursive filtering had been used previously, it was applied twice and four times to coincide with the Hadamard conjecture, while triple application required more than double the operations of singular convolution.

Suppose we have a sequence of observations $\{x_t\}$, beginning at time $t = 0$ with a cycle of seasonal change of length L.

The method calculates a trend line for the data as well as seasonal indices that weight the values in the trend line based on where that time point falls in the cycle of length L.

$\{s_t\}$ represents the smoothed value of the constant part for time t. $\{b_t\}$ represents the sequence of best estimates of the linear trend that are superimposed on the seasonal changes. $\{c_t\}$ is the sequence of seasonal correction factors. c_t is the expected proportion of the predicted trend at any time $t \bmod L$ in the cycle that the observations take on. As a rule of thumb, a minimum of two full seasons (or 2L periods) of historical data is needed to initialize a set of seasonal factors.

The output of the algorithm is again written as F_{t+m}, an estimate of the value of x at time $t+m$, $m>0$ based on the raw data up to time t. Triple exponential smoothing with multiplicative seasonality is given by the formulas[1]

$$
\begin{aligned}
s_0 &= x_0 \\
s_t &= \alpha \frac{x_t}{c_{t-L}} + (1-\alpha)(s_{t-1} + b_{t-1}) \\
b_t &= \beta(s_t - s_{t-1}) + (1-\beta)b_{t-1} \\
c_t &= \gamma \frac{x_t}{s_t} + (1-\gamma)c_{t-L} \\
F_{t+m} &= (s_t + mb_t)c_{t-L+1+(m-1) \mod L},
\end{aligned}
$$

where α is the *data smoothing factor*, 0 < α < 1, β is the *trend smoothing factor*, 0 < β < 1, and γ is the *seasonal change smoothing factor*, 0 < γ < 1.

The general formula for the initial trend estimate b_0 is:

$$
b_0 = \frac{1}{L}\left(\frac{x_{L+1} - x_1}{L} + \frac{x_{L+2} - x_2}{L} + \ldots + \frac{x_{L+L} - x_L}{L}\right)
$$

Setting the initial estimates for the seasonal indices c_i for i = 1,2,...,L is a bit more involved. If N is the number of complete cycles present in your data, then:

$$
c_i = \frac{1}{N}\sum_{j=1}^{N}\frac{x_{L(j-1)+i}}{A_j} \quad \forall i \qquad = 1, 2, \ldots, L
$$

where

$$
A_j = \frac{\sum_{i=1}^{L} x_{L(j-1)+i}}{L} \quad \forall j \qquad = 1, 2, \ldots, N
$$

Note that A_j is the average value of x in the jth cycle of your data.

Triple exponential smoothing with additive seasonality is given by:

$$
\begin{aligned}
s_0 &= x_0 \\
s_t &= \alpha(x_t - c_{t-L}) + (1-\alpha)(s_{t-1} + b_{t-1}) \\
b_t &= \beta(s_t - s_{t-1}) + (1-\beta)b_{t-1} \\
c_t &= \gamma(x_t - s_{t-1} - b_{t-1}) + (1-\gamma)c_{t-L} \\
F_{t+m} &= s_t + mb_t + c_{t-L+1+(m-1) \mod L},
\end{aligned}
$$

Implementations in statistics packages

- R: the HoltWinters function in the stats package[17] and ets function in the forecast package[18] (a more complete implementation, generally resulting in a better performance[19]).
- IBM SPSS includes Simple, Simple Seasonal, Holt's Linear Trend, Brown's Linear Trend, Damped Trend, Winters' Additive, and Winters' Multiplicative in the Time-Series modeling procedure within its Statistics and Modeler statistical packages. The default Expert Modeler feature evaluates all seven exponential smoothing models and ARIMA models with a range of nonseasonal and seasonal p, d, and q values, and selects the model with the lowest Bayesian Information Criterion statistic.
- LibreOffice 5.2[20]
- Microsoft Excel 2016[21]

See also

- Autoregressive moving average model (ARMA)
- Errors and residuals in statistics
- Moving average
- Continued fraction

Notes

1. "NIST/SEMATECH e-Handbook of Statistical Methods" (http://www.itl.nist.gov/div898/handbook/). NIST. Retrieved 2010-05-23.
2. Oppenheim, Alan V.; Schafer, Ronald W. (1975). *Digital Signal Processing*. Prentice Hall. p. 5. ISBN 0-13-214635-5.
3. Brown, Robert G. (1956). *Exponential Smoothing for Predicting Demand* (http://legacy.library.ucsf.edu/tid/dae94e00;jsessionid=104A0CEFFA31ADC2FA5E0558F69B3E1D.tobacco03). Cambridge, Massachusetts: Arthur D. Little Inc. p. 15.
4. Holt, Charles C. (1957). "Forecasting Trends and Seasonal by Exponentially Weighted Averages". *Office of Naval Research Memorandum*. **52**. reprinted in Holt, Charles C. (January–March 2004). "Forecasting Trends and Seasonal by Exponentially Weighted Averages" (http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6V92-4BJVV07-3&_user=1535420&_coverDate=03%2F31%2F2004&_rdoc=1&_fmt=high&_orig=gateway&_origin=gateway&_sort=d&_docanchor=&view=c&_searchStrId=1755927165&_rerunOrigin=google&_acct=C000053610&_version=1&_urlVersion=0&_userid=1535420&md5=0e881c06b0512cbf976a92268d48edc4&searchtype=a). *International Journal of Forecasting*. **20** (1): 5–10. doi:10.1016/j.ijforecast.2003.09.015 (https://doi.org/10.1016%2Fj.ijforecast.2003.09.015).
5. Brown, Robert Goodell (1963). *Smoothing Forecasting and Prediction of Discrete Time Series* (http://babel.hathitrust.org/cgi/pt?id=mdp.39015004514728;view=1up;seq=1). Englewood Cliffs, NJ: Prentice-Hall.
6. "NIST/SEMATECH e-Handbook of Statistical Methods, 6.4.3.1. Single Exponential Smoothing" (http://www.itl.nist.gov/div898/handbook/pmc/section4/pmc431.htm). NIST. Retrieved 2017-07-05.
7. "Averaging and Exponential Smoothing Models" (http://www.duke.edu/~rnau/411avg.htm). Retrieved 26 July 2010.
8. "Production and Operations Analysis" Nahmias. 2009.
9. https://www.otexts.org/fpp/7/1
10. Nahmias, Steven. *Production and Operations Analysis* (6th ed.). ISBN 0-07-337785-6.
11. "Model: Second-Order Exponential Smoothing" (http://help.sap.com/saphelp_45b/helpdata/en/7d/c27a14454011d182b40000e829fbfe/content.htm). SAP AG. Retrieved 23 January 2013.
12. Prajakta S. Kalekar. "Time series Forecasting using Holt-Winters Exponential Smoothing" (https://labs.omniti.com/people/jesus/papers/holtwinters.pdf) (PDF).
13. "6.4.3.3. Double Exponential Smoothing" (http://www.itl.nist.gov/div898/handbook/pmc/section4/pmc433.htm). *itl.nist.gov*. Retrieved 25 September 2011.
14. "Averaging and Exponential Smoothing Models" (http://www.duke.edu/~rnau/411avg.htm). *duke.edu*. Retrieved 25 September 2011.
15. Kalehar, Prajakta S. "Time series Forecasting using Holt-Winters Exponential Smoothing" (http://www.it.iitb.ac.in/~praj/acads/seminar/04329008_ExponentialSmoothing.pdf) (PDF). Retrieved 23 June 2014.
16. Winters, P. R. (April 1960). "Forecasting Sales by Exponentially Weighted Moving Averages" (http://mansci.journal.informs.org/cgi/content/abstract/6/3/324). *Management Science*. **6** (3): 324–342. doi:10.1287/mnsc.6.3.324 (https://doi.org/10.1287%2Fmnsc.6.3.324).
17. "R: Holt-Winters Filtering" (https://stat.ethz.ch/R-manual/R-patched/library/stats/html/HoltWinters.html). *stat.ethz.ch*. Retrieved 2016-06-05.
18. "ets {forecast} | inside-R | A Community Site for R" (http://www.inside-r.org/packages/cran/forecast/docs/ets). *www.inside-r.org*. Retrieved 2016-06-05.
19. "Comparing HoltWinters() and ets()" (http://robjhyndman.com/hyndsight/estimation2/). *Hyndsight*. 2011-05-29. Retrieved 2016-06-05.
20. https://wiki.documentfoundation.org/ReleaseNotes/5.2#New_spreadsheet_functions
21. http://www.real-statistics.com/time-series-analysis/basic-time-series-forecasting/excel-2016-forecasting-functions/

External links

- Lecture notes on exponential smoothing (Robert Nau, Duke University)
- Data Smoothing by Jon McLoone, The Wolfram Demonstrations Project
- The Holt-Winters Approach to Exponential Smoothing: 50 Years Old and Going Strong by Paul Goodwin (2010) Foresight: The International Journal of Applied Forecasting
- Algorithms for Unevenly Spaced Time Series: Moving Averages and Other Rolling Operators by Andreas Eckner

---

8 Holt-Winters' Exponential Smoothing with Seasonality

8.1 So far on Exponential Smoothing and Holt's Linear model

In exponential smoothing, we've so far met:

Simple exponential smoothing : This model is good for non-seasonal data that is fairly level (no trend). It follows the equation

$$F_{t+1} = F_t + \alpha\ (y_t - F_t) = \alpha\ y_t + (1-\alpha)\ F_t,$$

for $0 \leq \alpha \leq 1$.

For a longer range forecasts, it is assumed that the forecast function is flat:

$$\boxed{F_{t+k} = F_{t+1}}$$

Holt's linear method/Double exponential smoothing : This model is good for non-seasonal data with a trend. It follows the equations

$$\begin{aligned} \text{level} \quad L_t &= \alpha y_t + (1-\alpha)(L_{t-1} + b_{t-1}); \\ \text{trend} \quad b_t &= \beta(L_t - L_{t-1}) + (1-\beta) b_{t-1}, \\ \text{forecast } F_{t+1} &= L_t + b_t \end{aligned}$$

for $0 \leq \alpha \leq 1$ and $0 \leq \beta \leq 1$.

For a longer range forecasts, it is assumed that the forecast function follows the trend:

$$\boxed{F_{t+k} = F_{t+k} = L_t + k\ b_t}$$

8.2 Holt Winters' Additive Model

What about an exponential smoothing for data with a trend and seasonal behaviour? Winters generalised Holt's linear method to come up with such a technique, now called Holt Winters. A seasonal equation is added to Holt's linear method equations:

$$\begin{aligned} \text{level} \quad L_t &= \alpha(y_t - S_{t-s}) + (1-\alpha)(L_{t-1} + b_{t-1}); \\ \text{trend} \quad b_t &= \beta(L_t - L_{t-1}) + (1-\beta) b_{t-1}, \\ \text{seasonal} \quad S_t &= \gamma(y_t - L_t) + (1-\gamma) S_{t-s} \\ \text{forecast } F_{t+k} &= L_t + k b_t + S_{t+k-s}, \end{aligned}$$

where s is the length of the seasonal cycle, for $0 \leq \alpha \leq 1$, $0 \leq \beta \leq 1$ and $0 \leq \gamma \leq 1$. The seasonal equation picks up differences between the current level and the data at that time in the seasonal cycle. This is then added to a forecast at the same point in the cycle.

We have to pick the values of α, β and γ. As with the other methods, we can use RMSE or MAPE to choose the best values.

8.1 Definition This particular model is known as the *additive model* (seasonality factor is added to forecast).

EXERCISE: what do these updating equations say?

8.2.1 Getting Initial Values

To get started, we need initial values of the level, trend and seasonality. To initialise the seasonality, *at least one complete seasonal cycle $y_1, y_2, \ldots, y_s$ is needed.* We cannot make forecasts until one complete cycle has been observed. We therefore initialise L_s, b_s and $S_1, \ldots, S_s$, then use the formulae above to make forecasts from time s onwards. It is in fact best to initialise the trend with 2 complete cycles. Sensible starting values are

$$\begin{aligned} L_s &= \frac{1}{s}\sum_{i=1}^{s} y_i; \\ b_s &= \frac{1}{s}\Big[\frac{y_{s+1}-y_1}{s} + \frac{y_{s+2}-y_2}{s} + \cdots + \frac{y_{2s}-y_s}{s}\Big]; \\ S_i &= y_i - L_s,\ i = 1, \ldots, s. \end{aligned}$$

8.2.2 Holt Winter's Additive Model Algorithm

Init:

$L_s = \frac{1}{s}\sum_{i=1}^{s} y_i$

$b_s = \frac{1}{s}\big[\frac{y_{s+1}-y_1}{s} + \frac{y_{s+2}-y_2}{s} + \cdots + \frac{y_{2s}-y_s}{s}\big]$

$S_i = y_i - L_s,\ i = 1, \ldots, s$

and choose $0 < \alpha < 1$ and $0 < \beta < 1$

Compute:

level	$L_t = \alpha\ (y_t - S_{t-s}) + (1-\alpha)\ (L_{t-1} + b_{t-1})$
trend	$b_t = \beta\ (L_t - L_{t-1}) + (1-\beta)\ b_{t-1},$
seasonal	$S_t = \gamma\ (y_t - L_t) + (1-\gamma)\ S_{t-s}$
forecast	$F_{t+1} = L_t + b_t + S_{t+1-s}$

Until no more observation y_t are available

and subsequent forecasts: $F_{t+k} = L_t + k\ b_t + S_{t+k-s}$

Table 3: Holt Winter's Additive Model Algorithm.

8.3 Holt Winters' Multiplicative Model

An alternative Holt Winters' model multiplies the forecast by a seasonal factor. Its equations are:

$$\begin{aligned} \text{level} \quad L_t &= \alpha \frac{y_t}{S_{t-s}} + (1-\alpha)(L_{t-1} + b_{t-1}); \\ \text{trend} \quad b_t &= \beta(L_t - L_{t-1}) + (1-\beta) b_{t-1}, \\ \text{seasonal} \quad S_t &= \gamma \frac{y_t}{L_t} + (1-\gamma) S_{t-s} \\ \text{forecast} \quad F_{t+k} &= (L_t + k b_t) S_{t+k-s}, \end{aligned}$$

where, as before, s is the length of the seasonal cycle, for $0 \le \alpha \le 1$, $0 \le \beta \le 1$ and $0 \le \gamma \le 1$. Initial values for L_s and b_s are the same as in the additive case, but the initial seasonal estimates are: $S_1 = y_1/L_s$, $S_2 = y_2/L_s, \ldots, S_s = y_s/L_s$.

8.4 Example

In the table on the next page are the first 14 months beer production data. Since the data have a 12 month seasonal cycle, we initialise L_{12}, b_{12} and $S_1, \ldots, S_{12}$. Use the additive model formulae to calculate month 13 and 14's level, trend and seasonality, and make a 1-step ahead forecast for months 13, 14 and 15. Use $\alpha = 0.5, \beta = 0.3$ and $\gamma = 0.9$.

Month No.	Production	Level L_t	Trend b_t	Seasonal S_t	Forecast F_t
1	164	–	–	5.75	–
2	148	–	–	−10.25	–
3	152	–	–	−6.25	–
4	144	–	–	−14.25	–
5	155	–	–	−3.25	–
6	125	–	–	−33.25	–
7	153	–	–	−5.25	–
8	148	–	–	−12.25	–
9	138	–	–	−20.25	–
10	190	–	–	31.75	–
11	192	–	–	33.75	–
12	192	158.25	−0.65	33.75	–
13	147				
14	133				
15	163				

Figures 18.9 show the one-step ahead forecast for the complete data set. The RMSE is 13.6.

8.5 Conclusion

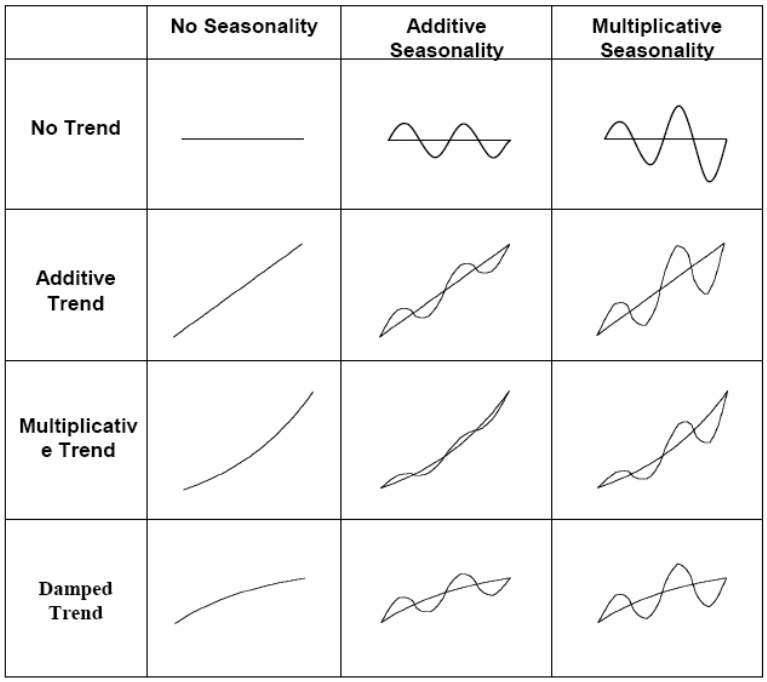


Figure 1: Trend and seasonality.

18.8 Comparing Forecasting Methods

18.9 Holt-Winters' Exponential Smoothing with Seasonality

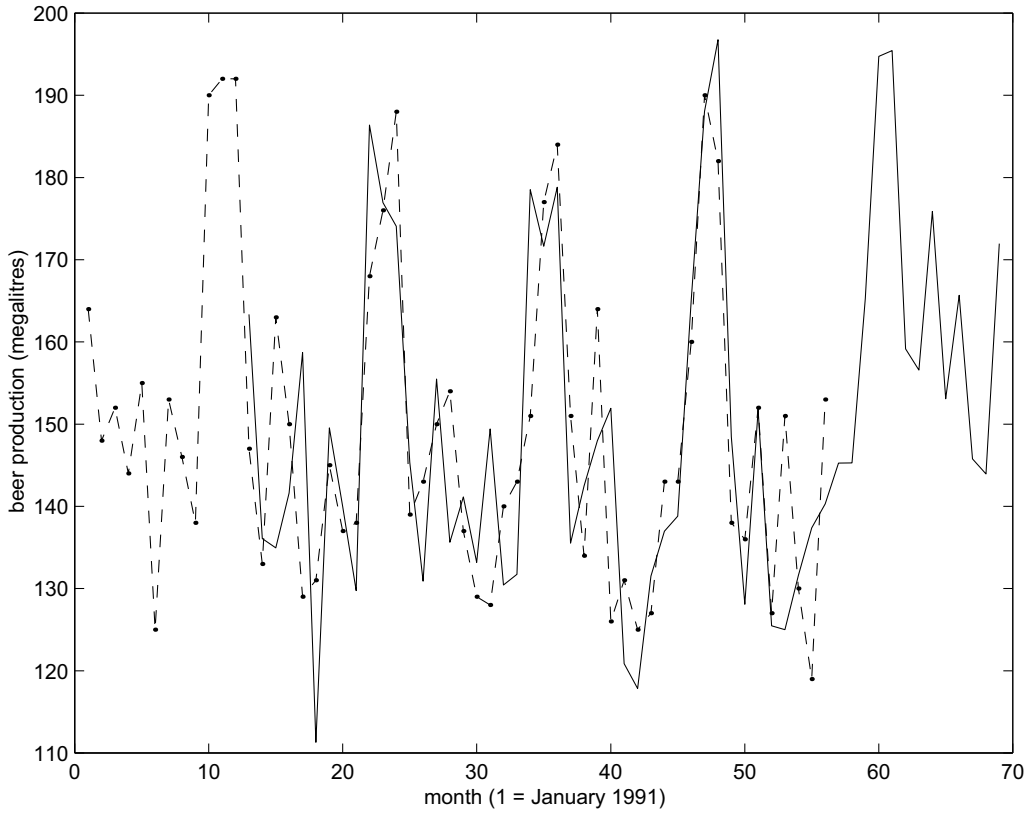


Figure 28: Monthly Australian beer production (dashed line) and forecasts by Holt Winters' additive model (one step ahead for observed values), with $\alpha = 0.5$, $\beta = 0.3$ and $\gamma = 0.9$.

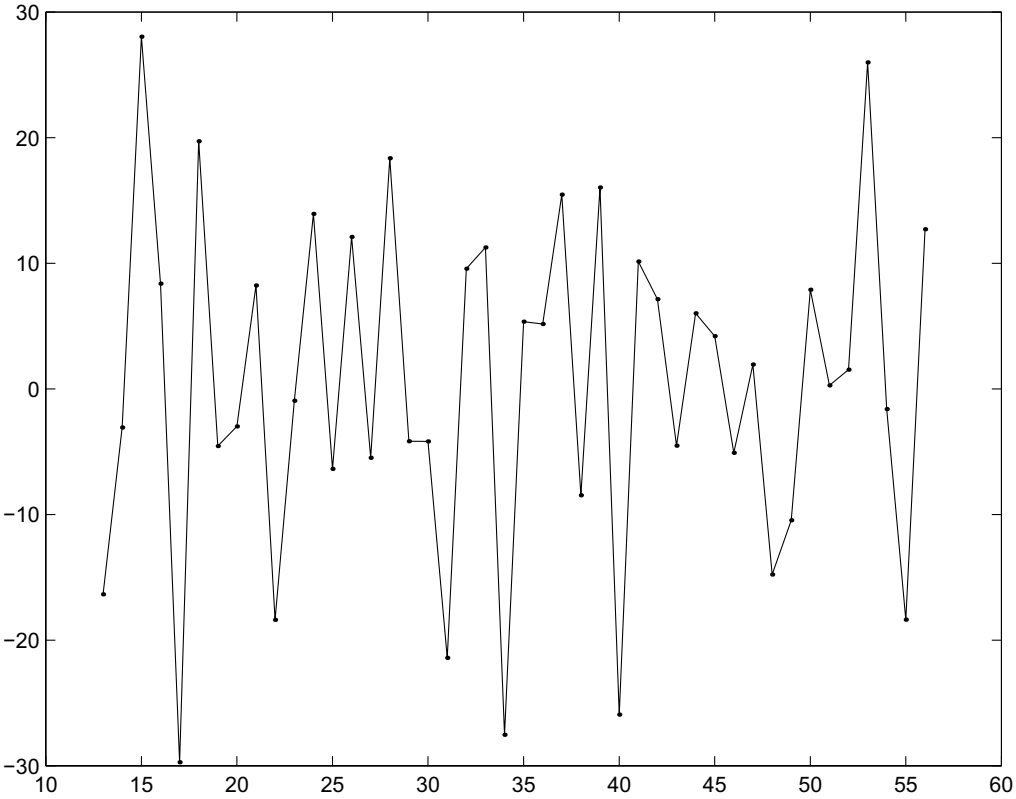


Figure 29: Timeplot of residuals in fitting by Holt Winters'.