TIME SERIES MODELS

L. Billard
Department of Statistics, University of Georgia, Athens, GA USA

Keywords: Autoregressive, moving average, seasonal, linear, bilinear spatial, autocorrelation, estimation, Gaussian, exponential.

Contents

Summary

Time series data measure phenomena when there are dependencies over time between prior and current values of the observations. This chapter provides a brief introduction to the standard (linear autoregressive moving-average) time series models including how such models are identified and fitted to the data. A review of more generalized models follows. This includes bilinear models which are useful to model processes which might contain sudden and short spurts, i.e., higher/lower values than those of the basic series as might occur in an earthquake, cyclone, etc. Where dependence also exists in space as well as time, spatial models, be these standard linear or bilinear models, are required; these are described briefly also. All these models have an underlying assumption that the errors and hence the observations are normally distributed. A brief review of models where the marginal distribution of the observations follows an exponential distribution is included.

1. Introduction

Time series data measure many phenomena in the life sciences, such as diseases (including epidemic trends), environmental data (such as pollution counts), meteorological data (cloud cover, temperatures, wind speeds, rainfall, barometric pressures, etc.), oceanographic data (ocean currents and temperatures, etc) as well as fisheries data (e.g., catching trends), agriculture (such as farm/crop outputs, land usage, water quality, and crop diseases), economic trends (time, productivity levels, prices, etc.), and geological and geographical data (such as earthquakes); the list is endless. All produce data that are best modeled by time series.

A key feature of time series data is that the current observations are dependent on values of previous observations; that is, the order in which the observations appear is important. For example, consider a disease outbreak where the data consist of the number of occurrences (number of infectives, or other measures of infectivity) for each unit of time (day, week, month, etc.). Then, it is quite reasonable to assume that the number of occurrences at the present time t depends (in some way to be determined) on the number of occurrences at the previous time (t - 1). This may be represented as

$$y(t) = \phi y(t-1) + e(t)\,, \tag{1}$$

where $y(t)$ is the number of occurrences at time t, ϕ is a parameter [in effect representing a measure of the degree of dependence on the previous observation $y(t-1)$] and $e(t)$ is the error term associated with the observation $y(t)$. Indeed, the dependence may go back as far as the p previous times $\{t-1, \ldots, t-p\}$. In ways that shall be seen shortly, this dependence can also (and/or alternatively) be expressed through the (q) preceding error terms. Notice that, when $p = 0$ and $q = 0$, the data are independent. This basic so-called standard linear time series model is discussed in Section 2.

There are some extensions to, and variations of, this standard model. Let us consider data recording average monthly temperatures at a given location. It is easy to see that the temperature this month (June, say) is related to the temperature for the previous month (May) as in Eq. (1). It is also likely that the temperature is related to the previous June figure. That is, there is seasonal dependence as well as current dependence. Such data give rise to seasonal time series models; these are also discussed briefly in Section 2.

The standard linear model (including the seasonal model) fits data that are stationary in the sense that the underlying patterns in the data are reasonably consistent. However, these models are typically unable to detect changes in historical patterns for processes which may exhibit sudden outbursts of activity such as when there is an epidemic outbreak in a disease pattern, or an earthquake, or the like. In these cases, bilinear models are better suited to model such data. This is particularly important when the data/models are being utilized as a surveillance mechanism for monitoring and control purposes to detect any such sudden changes. These will be described in Section 3.

In a different direction, consider data representing incidence of disease (e.g., mumps incidence at different locations, or fungi on wheat, etc.). Many (if not all) diseases occur in locations that are part of a broader region, rather than occurring in isolation at any particular site. That is, in addition to time dependence, the numbers of occurrences of a disease at a specific site are typically spatially dependent on the numbers at adjacent sites. Clearly, the standard models (linear, bilinear, and/or seasonal) which model dependence over time but all at one location cannot deal with this. Therefore, it is necessary to consider spatial time series models; see Section 4. It is easy to see that many environmental, oceanographic, etc., processes have both the time and spatial dependences as inherent components. Many will be spatial but linear time series models. When both spatial and sudden bursts of changes (as in disease outbreaks) occur, then spatial bilinear models should be considered; see Section 5.

An underlying assumption of all the above models is that the marginal distribution of the observations follows a normal distribution. There are however many applications in which these distributions are non-normal. Processes which give non-negative values, and/or have heavy tails, (such as water levels, or flood level measures, etc.), may be, and often are, nonnormal. Wind velocity amplitudes or acoustical data are typically uniformly or exponentially or more generally Laplace distributed and so are nonnormal. This leads to an exponential (broadly defined) class of time series models discussed in Section 6.

2. Standard Linear ARMA Models

The standard linear autoregressive moving average model, developed in the 1920s, is given by

$$y(t) = \sum_{i=1}^{p} \phi_i y(t-i) + \sum_{j=1}^{q} \theta_j e(t-j) + e(t), \qquad (2)$$

where $\{y(t)\}$, t = 1, 2, …, is a sequence of observations and $\{e(t)\}$, t = 1, 2, …, is a white noise process with $E\{e(t)\} = 0$ and $Var\{e(t)\} = \sigma^2$, and where ϕ_i, i = 1, …, p, are the autoregressive parameters and θ_j , j = 1, …, q, are the moving average parameters. Let us assume further that these $\{e(t)\}$ are normally distributed. The model is denoted by ARMA (p,q) with p and q representing the autoregressive order and moving average order of the model, respectively.

There are many properties of the models which need to be satisfied in some aspect. Typically, the most important are stationarity and invertibility. For example, take the ARMA (1,0) model, or simply the pure autoregressive model of order one, AR(1) model, given by Eq. (1). When $|\phi| < 1$, the model is stationary. Otherwise, it is nonstationary. When $|\phi| > 1$, we see from (1) that the underlying process is explosive. When $|\phi| = 1$, we have

$$y(t) = y(t-1) + e(t),$$

or

$$w(t) \equiv y(t) - y(t-1) = e(t);$$

that is,

$$w(t) = (1-B)y(t), \quad By(t) = y(t-1),$$

where B is the backward shift operator.

That is, by differencing, the nonstationary $\{y(t)\}$ process is transformed into a stationary $\{w(t)\}$ process. In general, we difference d times to produce stationarity. The general model is denoted as ARIMA (p,d,q). A pure autoregressive model will always be invertible. In contrast, an ARMA(0,q) pure moving average process is always stationary but requires conditions on its parameters to achieve invertibility; so, e.g., the ARMA (0,1) or equivalently the MA(1) model,

$$y(t) = \theta e(t-1) + e(t) \tag{3}$$

is invertible only if $|\theta| < 1$.

Determination of the (p,d,q) which identifies the model is done through the autocorrelation functions defined, at lag k, by

$$\rho_k = \frac{\mathrm{Cov}\{y(t), y(t+k)\}}{\sqrt{\mathrm{Var}\{y(t)\}\mathrm{Var}\{y(t+k)\}}}, \tag{4}$$

where

$$\mathrm{Cov}\{y(t), y(t+k)\} = \mathrm{E}\{y(t) - \bar{y}\}\{y(t+k) - \bar{y}\}$$

is the autovariance function at lag k and $\bar{y}$ is the usual average of the observed observations. For each ARMA (p,q) model, we know the theoretical patterns for ρ_k, k = 1, 2, … .

For example, for an AR(1) model,

$$\rho_k = \phi^k, \quad k = 1, 2, \ldots, \tag{5}$$

that is, the autocorrelation function ρ_k decays exponentially; while for an MA(q) model,

$$\begin{aligned} \rho_k &\neq 0, \quad k \leq q, \\ &= 0, \quad k > q, \end{aligned} \tag{6}$$

that is, the autocorrelation function ρ_k cuts off at $k = q$.

The partial autocorrelation functions, ϕ_{pp}, have the reverse pattern. By this we mean that while for pure AR(p) models, ρ_k decays exponentially in some manner, the partial autocorrelation function cuts off at p. Likewise, for a pure MA(q) model, the autocorrelation function cuts off at q, but the partial autocorrelation function decays. For the mixed ARMA (p,q) model, both the autocorrelation functions and the partial autocorrelation functions decay (and do not cut off).

These theoretical autocorrelation and partial autocorrelation functions are therefore compared with the sample autocorrelations $\hat{\rho}_k$, k = 1, 2, …, and the sample partial autocorrelation functions, calculated from the data, to identify tentative values for p, q and d. If these plots of $\hat{\rho}_k$ as k increases do not decay towards zero, then the data need further differencing to produce stationarity.

Standard seasonal models are so-called multiplicative model extensions of the nonseasonal models of (2). Let us consider the pure autoregressive AR(1) model. Eq. (1) can be written as $(1 - B\phi)y(t) = e(t)$.

If there is also seasonal dependence of order *s* of the autoregressive component, then the model becomes

$$(1-B\phi)(1-B^s\Phi)y(t)=e(t), \tag{7}$$

where Φ is the seasonal autoregressive parameter. Typically s = 12 (monthly data), s = 52 (weekly), s = 4 (quarterly), etc. This equation becomes

$$y(t)=\phi y(t-1)+\Phi y(t-s)-\phi\Phi y(t-s-1)+e(t),$$

and represents the ARMA(1,0) × $(1,0)_s$ model. The general ARMA (p, d, q) × $(P, D, Q)_s$ model is

$$(1-\phi_1 B-\ldots-\phi_p B^p)(1-\Phi_1 B^s-\ldots-\Phi_p B^{sP})(1-B)^d(1-B^s)^D y(t)$$

$$=(1-\theta_1 B-\ldots-\theta_q B^q)(1-\Theta_1 B^s-\ldots-\Theta_Q B^{sQ})e(t). \tag{8}$$

Properties for the seasonal components of this model are analogous to those for the nonseasonal model. In particular, the seasonal model orders (P,D,Q) are identified from the patterns in the autocorrelation and partial autocorrelation functions but at s-unit apart lags. When P = 0, D = 0 and Q = 0, the model reduces to the standard (nonseasonal) linear ARMA model of (2).

Once the model orders have been identified, the model parameters $\{\phi_i, i=1,\ldots,p\}$, $\{\theta_j, j=1,\ldots,q\}$, $\{\Phi_i, i=1,\ldots,P\}$, $\{\Theta_j, j=1,\ldots,Q\}$ and σ^2 are estimated using the usual techniques. Since the $\{y(t)\}$ and $\{e(t)\}$ are all normally distributed, maximum likelihood estimators can easily be calculated, as well as the properties of these estimators (such as their asymptotic normality and the like). If further the model is to be used to predict observations at other (often future) values, then these predictors and their associated properties can also be calculated easily. Many of the well-known statistical packages now perform these tasks routinely.

-
-
-

Bibliography

Billard L. (1997). Exponential autoregressive models. In: *Encyclopedia of Biostatistics* (eds. S. Kotz, C. B. Read and D. C. Banks) Oxford University Press, p. 227-231 [This article provides a review of exponential time series models and includes an extensive bibliography.]

Box G. E. P., Jenkins G. M. and Reinsel G. (2000). *Time Series Analysis Forecasting and Control* Holden-Day, CA, USA. [This provides an exhaustive coverage of the identification, estimation and prediction methods for standard seasonal and nonseasonal ARMA models, including theoretical aspects.]

Cliff A. D., Haggett P., Ord J. K., Bassett K. A. and Davies R. B. (1975). *Elements of Statistical Structure: An Quantitative Approach*, Cambridge University Press, New York. [Describes the STARMA model and gives applications in geographical settings.]

Dai Y. and Billard L. (1998). A space time bilinear model and its identification. *Journal of Time Series Analysis* 19, 657-679. [Describes the STBL model and its properties and the procedure for identifying model orders.]

Dai Y. and Billard L. (2003). Maximum likelihood estimation in space time bilinear models. *Journal of Time Series Analysis* 24, 25-44. [Provides a recursive algorithm for calculating parameter estimates, as well as a bibliography of other estimation procedures for this and the pure bilinear model.]

Granger C. W. and Andersen A. P. (1978). *An Introduction to Bilinear Time Series Models*}, Vandenhoeck and Ruprecht, Göttingen, Germany. [Describes the BL model especially the need to use the squared observation in the identification stage, which need carries over to the STBL model.]

Lawrance A. J. and Lewis P. A. W. (1985). Modelling and residual analysis of nonlinear autoregressive time series in exponential variables. *Journal of the Royal Statistical Society B* 47, 165-202. [This discusses several models from the exponential class and fits them to meteorological data.]

Pfeifer P. E. and Deutsch S. J. (1980). Identification and interpretation of first order space-time ARMA models. *Technometrics* 22, 397-408. [Provides coverage of patterns for the theoretical autocorrelation functions used to identify STARMA models and used in the first stage of identifying STBL models.]

Subba Rao T. and Gabr M. M. (1984). *An Introduction to Bispectral Analysis and Bilinear Time Series Models*, Lecture Notes in Statistics, Springer-Verlag, Berlin, Germany. [Provides an extensive coverage of results for bilinear time series models up to1984.]

Biographical Sketch

Lynne Billard was born in Toowoomba, Australia; she obtained her BS First Class Honors degree (1966) and PhD degree (1969) in Statistics from the University of New South Wales Australia. She has spent 13 years in administration including 9 years as Head of Statistics and 2 years as Associate to the Dean. Has held academic positions in Australia, Britain, Canada and the USA. Currently, she is University Professor and Professor of Statistics at the University of Georgia where she has taught design of experiments to graduate students since 1980; she has also taught a wide range of courses (including time series, introductory statistics, theoretical statistics) to both undergraduate and graduate students. She has over 150 publications mostly in the major journals including 6 books edited or co-edited, in sequential analysis, AIDS and epidemics, time series, and inference, with applications in agriculture, biology, epidemiology, education and social sciences. She has been accorded many honors and awards including the 1990 American Statistical Association's (ASA) Award for Outstanding Statistical Application paper (shared), the ASA 1999 Wilks Award and the ASA's 2003 Founders Award, and the University of Georgia Creative Research Award. She has held numerous professional offices including International President 1994 and 1995 of the International Biometric Society and was the 1996 President of the American Statistical Association. She has served on the International Council of both the International Biometric Society and of the International Statistical Institute, and was the 1985 President of the Eastern North American Region of the International Biometric Society, and served on the executive committee as Program Secretary of the Institute of Mathematical Statistics. She has served on many US national committees including the Advisory Committee for DMS National Science Foundation, Panel on AIDS and Panel on Microsimulation Modeling of Social Welfare Policy both for the National Research Council, the National Academy of Sciences' Board of Mathematical Sciences, was Chair of the Conference Board of Mathematical Sciences, and numerous review panels for the National Institute of Health and the National Science Foundation as well as the UK Research Council. She currently serves on the US Secretary of Commerce Census Advisory Committee. She is a Fellow of the American Statistical Association and the Institute of Mathematical Statistics and an elected Member of the International Statistical Institute.