Journal of Statistical Planning and Inference 139 (2009) 2791 – 2799

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Forecasting time series with missing data using Holt's model

José D. Bermúdez*, Ana Corberán-Vallet, Enriqueta Vercher

Department of Statistics and O.R., University of Valencia, Doctor Moliner 50, E-46100 Burjassot, Spain

ARTICLE INFO

Article history:
Received 4 May 2007
Received in revised form
29 December 2008
Accepted 1 January 2009
Available online 14 January 2009

Keywords:
Forecasting
Exponential smoothing
Linear model
EM algorithm
Data transformation

ABSTRACT

This paper deals with the prediction of time series with missing data using an alternative formulation for Holt's model with additive errors. This formulation simplifies both the calculus of maximum likelihood estimators of all the unknowns in the model and the calculus of point forecasts. In the presence of missing data, the EM algorithm is used to obtain maximum likelihood estimates and point forecasts. Based on this application we propose a leave-one-out algorithm for the data transformation selection problem which allows us to analyse Holt's model with multiplicative errors. Some numerical results show the performance of these procedures for obtaining robust forecasts.

1. Introduction

Most time series models assume that the observations are sampled with the same frequency, but it is common to find time series with missing data. In order to carry out a precise analysis of these time series and obtain reliable forecasts, it is necessary to deal effectively with the missing data.

The missing data problem has been dealt with successfully using the state-space methodology. Jones (1980) obtained the maximum likelihood estimates of the parameters of an ARMA model in the presence of missing data using the Kalman filter. Kohn and Ansley (1986) proposed a modified Kalman filter to generalise those previous results to the case of ARIMA models. Gómez and Maravall (1994) showed a new definition of the likelihood of an ARIMA model with missing observations that permits the use of the ordinary Kalman filter. Recently, Gómez et al. (1999) proposed filling in the holes in the missing data with arbitrary values and carrying out the maximum likelihood estimation with additive outliers.

Analogously, Wright (1986) suggested an extension for simple exponential smoothing and Holt's method in the case of missing data. Cipra et al. (1995) extended the previous approach to the Holt–Winters method: the level, trend and seasonal terms are updated each time a new observation becomes available, using modified transition equations which take into account the possible presence of missing data between two consecutive observations. Cipra and Romera (1997) proposed new transition equations for the Holt–Winters method in the presence of missing data using a robust version of the Kalman filter based on the M-estimation methodology: if the observation at one time point is missing, the estimated level, trend and seasonality remain unchanged. The problem of missing data in time series prediction has also been dealt with using neural networks (Hofmann and Tresp, 1998) and Monte Carlo methods (Chen and Liu, 1998).

Since exponential smoothing methods are widely used for short-term prediction in business and industry (Gardner, 2006), in this paper, we present a new approach to the prediction of time series with missing data based on an alternative formulation for Holt's model with additive errors. In this new formulation the stochastic component of the model is introduced by means of additive, independent, homoscedastic and normal errors. Then the data vector is multivariate normal and its mean and

* Corresponding author.
E-mail addresses: bermudez@uv.es (J.D. Bermúdez), ana.corberan@uv.es (A. Corberán-Vallet), vercher@uv.es (E. Vercher).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2009.01.004

covariance matrix are functions of the model parameters. Hence Holt's model can be formulated as a heteroscedastic linear model with coefficients given by the initial conditions and the covariance matrix relying on the smoothing parameters. This formulation allows us to obtain the maximum likelihood estimates of all the unknowns, the smoothing parameters and the initial values of level and trend jointly, as in some other proposals (Harvey, 1989; Ord et al., 1997; Segura and Vercher, 2001; Bermúdez et al., 2006a,b; Bermúdez et al., 2007, 2008). In the presence of missing data, the EM algorithm (Dempster et al., 1977) is used in the estimation of the model's parameters and the calculus of point forecasts.

The paper is organised as follows. In Section 2 we define the multivariate linear model for Holt's model and the formulae for the calculation of the maximum likelihood estimators and point forecasts. In Section 3 we apply the EM algorithm to the estimation of the model parameters and the calculus of point forecasts when missing data are presented, and show the performance of the procedure with some numerical results. In Section 4 we develop a new data transformation selection method based on the leave-one-out technique and present the results corresponding to the prediction of the yearly time series of the M3 Competition when the proposed data transformation selection mechanism is used. The last section is concerned with some concluding remarks.

2. Holt's model with additive errors

We assume that $\{y_t\}_{t=1}^n$ are the observed data. The Holt model with additive errors assumes that the observation at time t comes from the random variable

$$Y_t = \mu_t + \varepsilon_t \tag{1}$$

where $\mu_t = a_{t-1} + b_{t-1}$, a_t and b_t are the level and trend at time t, respectively, and $\{\varepsilon_t\}_{t=1}^n$ are independent homoscedastic normal random variables, $N(0, \sigma^2)$. When a new observation becomes available the level and trend terms are updated through the transition equations

$$a_t = \mu_t + \alpha\varepsilon_t \tag{2}$$

$$b_t = b_{t-1} + \alpha\beta\varepsilon_t \tag{3}$$

Let $\mathbf{Y} = (y_1, y_2, \ldots, y_n)'$ be the data vector, $\theta = (\alpha, \beta)'$ the vector of smoothing parameters and $\omega = (a_0, b_0)'$ the vector of initial values. Given the initial values and applying the transition recursively the h steps ahead prediction is usually obtained as $\hat{y}_{n+h} = a_n + hb_n$ for $h \geqslant 1$. In practice, both θ and ω vectors are unknown and have to be estimated from the data. Applying (1), (2) and (3) recursively the data can be stated (Bermúdez et al., 2007) as

$$Y_1 = a_0 + b_0 + \varepsilon_1$$

$$Y_2 = a_0 + 2b_0 + \alpha(1+\beta)\varepsilon_1 + \varepsilon_2$$

$$Y_3 = a_0 + 3b_0 + \alpha(1+2\beta)\varepsilon_1 + \alpha(1+\beta)\varepsilon_2 + \varepsilon_3$$

$$\vdots$$

$$Y_t = a_0 + tb_0 + \alpha\sum_{r=1}^{t-1}(1+(t-r)\beta)\varepsilon_r + \varepsilon_t$$

The data vector can then be described through the equation

$$\mathbf{Y} = A\omega + L\varepsilon \tag{4}$$

where A is the $n \times 2$ matrix whose first column is the vector $(1, 1, \ldots, 1)'$ and the second one the vector $(1, 2, \ldots, n)'$; L is the $n \times n$ lower triangular matrix whose elements in the main diagonal are equal to 1 and $l_{i,j} = \alpha(1+(i-j)\beta)$ if $i > j$; $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)'$ is the error vector. Therefore, the data vector follows a multivariate normal distribution with mean $\mathrm{E}(\mathbf{Y}) = A\omega$ and covariance matrix $\mathrm{V}(\mathbf{Y}) = \sigma^2 LL'$. This covariance matrix depends on the smoothing parameters but is always positive-definite because it is symmetric and its determinant is always positive: $|\mathrm{V}(\mathbf{Y})| = \sigma^{2n}|L|^2 = \sigma^{2n} > 0$, due to $|L| = 1$. We assume that the value of each smoothing parameter lies in the interval [0,1], although this restriction is not necessary.

From Eq. (4), the log-likelihood function of the data vector $\mathbf{Y}$ is given by the logarithm of the multivariate normal density function

$$-\frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}(\mathbf{Y} - A\omega)'(LL')^{-1}(\mathbf{Y} - A\omega) \tag{5}$$

The quadratic form in (5), $(\mathbf{Y} - A\omega)'(LL')^{-1}(\mathbf{Y} - A\omega)$, can be decomposed as $(\tilde{\omega} - \omega)'X'X(\tilde{\omega} - \omega) + (L^{-1}\mathbf{Y})'(I - P_X)(L^{-1}\mathbf{Y})$, where X is the matrix $L^{-1}A$, $P_X = X(X'X)^{-1}X'$ is the orthogonal projection matrix on the vector space generated by the columns of the matrix X and $\tilde{\omega} = (X'X)^{-1}X'L^{-1}\mathbf{Y}$ is the mean square estimator of ω when θ is known. So, the log-likelihood function can be