Available online at www.sciencedirect.com
ScienceDirect
International Journal of Forecasting 27 (2011) 252–265
international journal of forecasting
www.elsevier.com/locate/ijforecast

Forecasting correlated time series with exponential smoothing models

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

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

Abstract

This paper presents the Bayesian analysis of a general multivariate exponential smoothing model that allows us to forecast time series jointly, subject to correlated random disturbances. The general multivariate model, which can be formulated as a seemingly unrelated regression model, includes the previously studied homogeneous multivariate Holt-Winters' model as a special case when all of the univariate series share a common structure. MCMC simulation techniques are required in order to approach the non-analytically tractable posterior distribution of the model parameters. The predictive distribution is then estimated using Monte Carlo integration. A Bayesian model selection criterion is introduced into the forecasting scheme for selecting the most adequate multivariate model for describing the behaviour of the time series under study. The forecasting performance of this procedure is tested using some real examples.

Keywords: Bayesian forecasting; Exponential smoothing; Innovations state space models; Model selection; Monte Carlo methods; Multivariate time series

1. Introduction

Exponential smoothing methods are forecasting techniques which are used widely for the analysis of univariate time series, due to their simplicity and robustness as automatic forecasting procedures (Bermúdez, Segura, & Vercher, 2008; Gardner, 2006; Hyndman, Koehler, Snyder, & Grose, 2002). They originated in the work of Brown and Holt (Brown, 1959; Holt, 1957), but became well known through the paper by Winters (1960). The general form of the exponential smoothing forecast function, involving a set of adaptive coefficients, was given, possibly for the first time, by Box and Jenkins (1976, Appendix A5.3). Snyder (1985) introduced the linear single source of error state space models and showed how they were related to exponential smoothing, while their generalisation to nonlinear state space models was given by Ord, Koehler, and Snyder (1997); see also Koehler, Snyder, and Ord (2001) for a general multiplicative Holt-Winters' model. Without drawing out links with exponential smoothing, single source of error state space models, also known as innovations

* Corresponding author.
E-mail address: bermudez@uv.es (J.D. Bermúdez).

0169-2070/$ - see front matter © 2010 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.ijforecast.2010.06.003

models, had previously been used by Anderson and Moore (1979, pp. 230–238). Innovations state space models provide a general framework where the level, trend and seasonality components of exponential smoothing are stated explicitly in the models (Hyndman, Koehler, Ord, & Snyder, 2008).

For the analysis of dependent time series, that is, series which are subject to correlated random disturbances, or where the observations of a time series are related to the past and present values of other series, the use of multivariate time series models allows information to be borrowed from one series in order to improve the predictions of another series. On such occasions, some multivariate generalisations of the exponential smoothing methods (de Silva, Hyndman, & Snyder, 2007; Enns, Machak, Spivey, & Wrobleski, 1982; Fernández, 1990; Harvey, 1986; Pfeffermann & Allon, 1989) have been shown to provide more satisfactory results than those derived from the univariate analysis of each series. Recently, Bermúdez, Corberán-Vallet, and Vercher (2009a) introduced a new formulation for a multivariate Holt-Winters' model. This model formulation is based on the assumptions that each of the individual time series comes from the univariate Holt-Winters' model, that all of them share a common structure, that is, common smoothing parameters, and that corresponding errors in the univariate models are contemporaneously correlated. Expressing this multivariate Holt-Winters' model as an innovations state space model, the assumption of common smoothing parameters for the univariate models is equivalent to the homogeneity condition assumed in previous studies. This condition implies that all of the individual series have identical time series properties and that the sum of the series also has the same properties (Fernández & Harvey, 1990).

In the case of a common structure for the univariate time series, the estimation of the homogeneous multivariate model is straightforward. Its use both allows the fitting and forecast accuracies to be improved with respect to the univariate model and reduces the computing time required for the analysis of the series considerably (Bermúdez et al., 2009a). However, although univariate series may follow similar processes in some practical situations, this does not necessarily hold in general. In such cases, the use of the homogeneous multivariate model may lead to misleading results. Therefore, it is advisable to use a general multivariate exponential smoothing model where the univariate series are subject to correlated random disturbances but do not necessarily share a common structure.

In this paper, we describe the Bayesian analysis of a general multivariate linear innovations state space model. This general model cannot be formulated as a traditional multivariate regression model, as the homogeneous one can be; instead, it is formulated as a seemingly unrelated regression model (Zellner, 1962), which complicates its analysis. The posterior distribution of all of the unknowns is then obtained from conventional non-informative prior distributions. This posterior distribution is not analytically tractable, but can be approached using MCMC simulation techniques. In particular, we propose a Metropolis-within-Gibbs algorithm that allows us to simulate from the full conditional posterior distributions of the model parameters. The predictive distribution, which encapsulates all of the information concerning the future values of the time series, is finally estimated using Monte Carlo integration.

The paper is organised as follows. In the next section we briefly review linear univariate exponential smoothing, based on innovations state space models, and its multivariate extensions. The Bayesian analysis of the general multivariate model proposed in this paper is developed in Section 3. In Section 4, the Bayesian forecasting procedure, which allows us to obtain point forecasts and prediction intervals, is developed. Section 5 shows the results obtained from the prediction of some correlated time series data sets using our Bayesian forecasting procedure. The last section gives some concluding remarks.

2. Linear exponential smoothing models

2.1. Linear univariate exponential smoothing models

The general linear innovations state space model provides a general framework for linear exponential smoothing. It is defined through the equations (Hyndman, Koehler et al., 2008, p. 34)

$$y_t = w'x_{t-1} + \varepsilon_t \quad \text{(measurement equation)}$$
$$x_t = Fx_{t-1} + g\varepsilon_t, \quad \text{(transition equation)}$$

where y_t denotes the observation at time t, and x_t is the state vector which, in exponential smoothing, is the vector of the level, trend and seasonality components at time t. The transition matrix F and the vectors w and g usually contain some unknown parameters. In the usual exponential smoothing models, the smoothing parameters appear in the vector g, while the damped parameters, if included in the model, appear in F or w. In the following, θ denotes the vector of all of these parameters.

All of the usual linear exponential smoothing models are special cases of the above state space model, including multiple seasonality models (Gould et al., 2008; Taylor, 2003). Some interesting extensions could also be considered by allowing w and g to vary with time, such as the inclusion of covariates in the model (Hyndman, Koehler et al., 2008, Chapter 9) or non-regular seasonal cycles (Gould et al., 2008).

For example, the additive Holt-Winters' model is given by the measurement equation $y_t = a_{t-1} + b_{t-1} + c_{t-s} + \varepsilon_t$. The local level, trend and seasonal components are updated using the equations

$$\begin{aligned} a_t &= \alpha(y_t - c_{t-s}) + (1-\alpha)(a_{t-1} + b_{t-1}) \\ &\quad \text{(level equation)} \\ b_t &= \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1} \\ &\quad \text{(trend equation)} \\ c_t &= \gamma(y_t - a_{t-1} - b_{t-1}) + (1-\gamma)c_{t-s}, \\ &\quad \text{(seasonal equation)} \end{aligned}$$

where s is the length of the seasonal cycle and $\theta = (\alpha, \beta, \gamma)'$ is the vector of smoothing parameters. Note that the above equation for updating the seasonal index is similar to the one originally proposed by Winters (1960) but not the same; instead, it is the one used by Hyndman et al. (2002). Plugging the measurement equation into the above updating equations, the following transition equations are obtained: $a_t = a_{t-1} + b_{t-1} + \alpha\varepsilon_t$, $b_t = b_{t-1} + \alpha\beta\varepsilon_t$ and $c_t = c_{t-s} + \gamma\varepsilon_t$.

Under the assumption of a finite start-up with fixed initial state variables x_0, any linear innovations state space model has a linear model representation. Applying the transition equations recursively, the measurement equation of the innovations model is given by

$$y_t = w'F^{t-1}x_0 + \sum_{j=1}^{t-1} w'F^{t-j-1}g\varepsilon_j + \varepsilon_t. \tag{1}$$

Staking the above equations for all observed data, the linear model $y = Dx_0 + L\varepsilon$ is obtained, with y being the observed data vector and ε the error vector. L is a unit lower triangular matrix, which is a function of the vector θ of parameters which appear in matrix F and vectors w and g. The design matrix D is usually not of complete rank, so some linear restrictions on the vector x_0 are needed for the model to be identifiable (seasonal components adding up to zero, for example). Considering all of these restrictions, a reduced initial conditions vector ψ could be defined such that $x_0 = T\psi$ and $M = DT$ is a complete rank matrix which could also depend on the unknown parameter vector θ. The linear heteroscedastic model with a complete rank design matrix is

$$y = M\psi + L\varepsilon. \tag{2}$$

For the additive Holt-Winters' model, M is a completely known $n \times (s+1)$ matrix. In the first column, all of the components are equal to one and in the second column they are given by the vector $(1, 2, \ldots, n)'$, while the last $s-1$ columns are built with blocks of identity $(s-1) \times (s-1)$ matrices, with a row with all of its elements equal to -1 between each pair of identity matrices. $\psi = (a_0, b_0, c_{1-s}, \ldots, c_{-1})'$ is the reduced vector of initial conditions (note that c_0 is not considered, as this is not necessary due to the constraint $c_0 + c_{-1} + \cdots + c_{1-s} = 0$). L is a $n \times n$ lower triangular matrix with all of the elements in its main diagonal equal to one and $L_{ij} = \alpha + \alpha\,\beta(i - j) + \gamma(i = j \bmod s)$ for $j < i$; thus, L is a function of the unknown vector of smoothing parameters $\theta = (\alpha, \beta, \gamma)'$, and consequently, is unknown.

The linear model in Eq. (2) was used by Snyder (1986) to estimate the parameters of a linear innovations model. He obtained maximum likelihood estimators of the unknown parameters assuming uncorrelated, homoscedastic and normally distributed errors; i.e., assuming that ε follows a $\mathrm{N}(0, \sigma^2 I_n)$ distribution. Prediction intervals for the model are given by Bermúdez, Segura, and Vercher (2007). The Bayesian analysis of the linear model in Eq. (2) is given by Bermúdez, Segura, and Vercher (2010). The Bayesian analysis of state space models with multiple sources of error goes back to Harrison and Stevens (1976), see also West and Harrison (1989). The first Bayesian analysis of the innovations state space model is given by Forbes, Snyder, and Sharmi (2000), but instead of using model (2), the authors built the likelihood function using the conditional distributions of each data observation given the preceding ones. Their predictive distribution is based on a discretisation of the marginal posterior distribution of θ.

The general linear innovations state space model and its related linear model (Eq. (2)) are equivalent formulations of exponential smoothing models. They share the same likelihood function, and therefore they will produce the same forecasting results with either maximum likelihood or Bayesian techniques. The innovations model formulation is closely related to the usual formulation of exponential smoothing, and is intuitive and appealing: if the parameters are known or have already been estimated, when a new observation becomes available, the state vector is updated and the new one-step-ahead forecast is available immediately. Moreover, nonparametric estimation techniques based on the minimisation of measures of the one-step-ahead errors are easily to implement (Bermúdez, Segura, & Vercher, 2006).

The linear model in Eq. (2) is more general, but it could have too many free parameters. It is necessary to give some structure to the matrices M and L, such that they depend on a reduced set of parameters θ. The parametric structure of the matrices can easily be obtained if an innovations state space model has previously been given, so we propose model (2) as an intermediate step between model formulation and statistical analysis. The main advantage of the linear formulation is that it shows the joint distribution of the data vector, Eq. (2), explicitly, giving a better probabilistic understanding of the model. In fact, the correlation matrix of the time series, LL', is given explicitly in the model formulation. Moreover, given that knowing the joint distribution implies that any conditional distribution is also known, it is easy to implement the EM algorithm in the presence of missing data (Bermúdez, Corberán-Vallet, & Vercher, 2009b). From a Bayesian point of view, the joint distribution of the data set is an essential piece of the analysis, and it is provided by Eq. (2).

2.2. Linear multivariate exponential smoothing models

The multivariate extension proposed in this paper is based on the assumption that the observations come from m time series, $y_i = (y_{1i}, y_{2i}, \ldots, y_{ni})'$ for $i = 1, 2, \ldots, m$, each one of them following univariate linear model

$$y_i = M_i\psi_i + L_i\varepsilon_i, \tag{3}$$

sharing a similar structure for matrices M_i and L_i; i.e., all of the L_i matrices are equal except for the value of the unknown parameter vector θ_i, which could differ from series to series, and similarly for the M_i matrices. The condition of a similar structure is given, for example, if the m series follow an additive Holt-Winters' model with possibly different smoothing parameters. We also assume that there is a contemporaneous correlation between the corresponding errors in different equations, and specifically assume that the distribution of the error matrix $[\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m]$ is the matrix normal distribution $\mathrm{N}(0, \Sigma, I_n)$, or equivalently, $(\varepsilon_1', \varepsilon_2', \ldots, \varepsilon_m')' \sim \mathrm{N}(0, \Sigma \otimes I_n)$, where $\otimes$ denotes the Kronecker product; i.e., all of the rows of the error matrix are independent and share the same multinormal distribution, with a zero mean and covariance matrix Σ. Under these assumptions, the series are only related through the error term.

The homogeneous case is given when $\theta_1 = \cdots = \theta_m$. Then all of the L_i matrices are equal, as are the M_i matrices, and the m univariate models can be combined as a multivariate heteroscedastic linear model (Bermúdez et al., 2009a)

$$\begin{aligned}[y_1, y_2, \ldots, y_m] = {} & M[\psi_1, \psi_2, \ldots, \psi_m] \\ & + L[\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m].\end{aligned} \tag{4}$$

However, if the individual series under study do not share a common θ parameter, the above multivariate model becomes inadequate and its use may lead to misleading results.

The general case, a different θ parameter vector in each series, needs an alternative analysis. Multiplying Eq. (3) by L_i^{-1}, the homoscedastic model $z_i = X_i\psi_i + \varepsilon_i$ is obtained, where $z_i = L_i^{-1}y_i$ and $X_i = L_i^{-1}M_i$. A general multivariate model which takes into account the dependence between the errors in different equations while allowing the univariate series to follow different processes, can be obtained by stacking the m equations as follows:

$$z = X_B\psi + \varepsilon,$$

where $z = (z_1', z_2', \ldots, z_m')'$, $\psi = (\psi_1', \psi_2', \ldots, \psi_m')'$, and X_B is the block-diagonal matrix whose ith diagonal block is matrix X_i. The distribution of the error vector $\varepsilon = (\varepsilon_1', \varepsilon_2', \ldots, \varepsilon_m')'$ is the multivariate normal $N(0, \Sigma \otimes I_n)$.

The general multivariate model can then be formulated as a seemingly unrelated regression (SUR) model (Zellner, 1962), which is an extension of the linear regression model that allows correlated errors between equations that appear to be independent of each other. Note that the homogeneous model given by Eq. (4) is a special case of this general multivariate model if the θ parameters for each of the univariate models are the same.

Taking into account the fact that the L_i matrices are unit lower triangular, their determinants are equal to one and the likelihood function is given by

$$f(y|\psi, \Sigma, \theta) \propto |\Sigma|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}(z - X_B\psi)' \times (\Sigma^{-1} \otimes I_n)(z - X_B\psi)\right\}, \tag{5}$$

where $\theta = (\theta_1', \theta_2', \ldots, \theta_m')'$ is the vector of parameters of the m time series, which is present in all matrices L_i, $i = 1, 2, \ldots, m$, and therefore in both z and X_B. In practice, since these parameters are always unknown, the design matrix X_B is unknown, which complicates the statistical analysis of this multivariate model.

The matrices in the quadratic form in the above likelihood function are sparse. Using the definition of the Kronecker product and taking into account the fact that X_B is block diagonal, the quadratic form becomes

$$(z - X_B\psi)'(\Sigma^{-1} \otimes I_n)(z - X_B\psi) = \sum_{i=1}^{m}\sum_{j=1}^{m} \sigma^{ij}(z_i - X_i\psi_i)'(z_j - X_j\psi_j), \tag{6}$$

where σ^{ij} is the (i, j) element of matrix Σ^{-1}. Note that the matrices in this last expression are the univariate model matrices, so the matrix calculus for the multivariate case is no more complex than for the univariate case.

The vector exponential smoothing model (Hyndman, Koehler et al., 2008, Section 17.1) is the general multivariate extension of the innovations state space models. It is defined through the vector measurement and transition equations

$$y_t = W x_{t-1} + \varepsilon_t$$

$$x_t = F x_{t-1} + G\varepsilon_t,$$

where y_t is the vector of observations at time t in the m series, the state vector x_t is the vector of all state variables from the m time series, and the innovations vector ε_t follows a multivariate normal distribution with a zero mean and covariance matrix Σ. W, F and G are matrices which could be functions of the unknown parameter vector θ. Recursively applying the transition equation of the above vector model, the measurement equation allows a vector representation similar to Eq. (1). Then, a linear model $y = Dx_0 + L\varepsilon$ is obtained by stacking the n multivariate observations y_t. However, this general multivariate innovations state space model could have an excessive number of parameters.

As a particular example, if the original matrices W, F and G are block diagonal, a seemingly unrelated model is obtained (Hyndman, Koehler et al., 2008, Section 17.1.1) which is mathematically equivalent to the one proposed in this paper. That linear model is the same as the one obtained by stacking the m vectors given in Eq. (3), but for a reordering of its elements: the observations are ordered by time in one case and by series in the other. The likelihood function given by Eq. (5) will be the same in both cases, so either of those alternatives could be used when implementing the numerical routines.

3. Bayesian estimation of the multivariate model

Zellner (1971) popularised Bayesian inference in econometrics and described the SUR model within the context of Bayesian inference, although at that time there were no convenient methods for its estimation. Unlike the estimation of the traditional multivariate regression model, the Bayesian estimation of the SUR model was not accessible until the application of MCMC methods to Bayesian inference. In the case where the parameter vector θ was known, which would imply that the L_i matrices ($i = 1, 2, \ldots, m$) and the design matrix X_B were also known, the Bayesian analysis of the SUR model is well known (Griffiths, 2001). However, we are chiefly interested in the general case where θ is unknown, since this is the only real case in practice.

In this paper we are going to use a non-informative prior distribution, but similar expressions could be obtained if a prior on ψ and Σ from the conjugate family, the normal inverted Wishart distribution, is used. If some restrictions on the θ vector are needed, they should be included in its prior distribution. We propose a non-informative constant prior on θ on the parameter space Θ where those restrictions are fulfilled; this prior will be proper if Θ is bounded. In the formulation introduced in the previous section for the additive Holt-Winters' model, the three smoothing parameters may be restricted to the interval (0, 1) (see for instance Hyndman et al., 2002), and therefore $\Theta = (0,1)^3$. Hyndman, Akram, and Archibald (2008) propose an admissible parameter space for exponential smoothing models based on conditions on forecastability and stationarity, and Θ could be defined to satisfy those conditions.

From the conventional non-informative prior distribution given by

$$f(\psi,\Sigma,\theta)=f(\psi)f(\Sigma)f(\theta)\propto|\Sigma|^{-(m+1)/2}, \tag{7}$$

the posterior distribution — proportional to the product of the likelihood function, Eqs. (5) and (6), and the prior distribution, Eq. (7) — is

$$\begin{aligned} f(\psi,\Sigma,\theta|y) &\propto |\Sigma|^{-\frac{n+m+1}{2}}\exp\left\{-\frac{1}{2}(z-X_B\psi)'\right.\\ &\quad\left.\times\ (\Sigma^{-1}\otimes I_n)(z-X_B\psi)\right\} \qquad (8)\\ &\propto |\Sigma|^{-\frac{n+m+1}{2}}\exp\left\{-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\sigma^{ij}(z_i-X_i\psi_i)'\right.\\ &\quad\left.\times\ (z_j-X_j\psi_j)\right\}. \qquad (9)\end{aligned}$$

Since the posterior distribution for the model parameters is not analytically tractable, we propose to obtain a sample $\{(\psi^{(j)},\Sigma^{(j)},\theta^{(j)})\}_{j=1}^{N}$ from this posterior distribution which will be used to obtain Monte Carlo approaches to any characteristics in which we are interested. In order to obtain such a sample, simulation is required.

For the Bayesian analysis of the homogeneous multivariate model, using the prior distribution given by Eq. (7), it is known that only a sample from the non-analytically tractable posterior distribution of the common parameter vector, $f(\theta|y)$, is required (Bermúdez et al., 2009a); the marginal posterior distributions for the initial conditions and the covariance matrix can then be estimated using Monte Carlo integration. In particular, for dealing with the additive Holt-Winters' model the authors propose an acceptance sampling algorithm, which is easy to implement and fast enough to draw observations from the three dimensional distribution with bounded support $f(\theta|y)$. Unfortunately, this approach is not feasible for the estimation of the general multivariate model, due to the intractability of the conditional posterior distributions for the initial conditions and the covariance matrix given θ (Griffiths, 2001). Therefore, in order to obtain a sample from the joint posterior distribution of the model parameters, Eq. (9), we propose to use the Gibbs sampling procedure, which is easy to implement but requires the full conditional posterior distribution to be known.

From Eq. (9), the full conditional posterior distribution for parameter θ is

$$\begin{aligned} f(\theta|y,\psi,\Sigma) &\propto \exp\left\{-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\sigma^{ij}(z_i-X_i\psi_i)'\right.\\ &\quad\left.\times\ (z_j-X_j\psi_j)\right\}. \end{aligned} \tag{10}$$

In order to obtain the full conditional posterior distribution of parameter ψ, note that the quadratic form $(z-X_B\psi)'(\Sigma^{-1}\otimes I_n)(z-X_B\psi)$ in Eq. (8) can be decomposed as $(z-X_B\widehat{\psi})'(\Sigma^{-1}\otimes I_n)(z-X_B\widehat{\psi})+(\psi-\widehat{\psi})'X_B'(\Sigma^{-1}\otimes I_n)X_B(\psi-\widehat{\psi})$, where $\widehat{\psi}=(X_B'(\Sigma^{-1}\otimes I_n)X_B)^{-1}X_B'(\Sigma^{-1}\otimes I_n)z$ is the generalized least squares estimator of ψ when the value of θ is known (Zellner, 1962). Therefore, the posterior distribution can be expressed as

$$\begin{aligned} f(\psi,\Sigma,\theta|y) &\propto |\Sigma|^{-\frac{n+m+1}{2}}\\ &\times\exp\left\{-\frac{1}{2}(z-X_B\widehat{\psi})'(\Sigma^{-1}\otimes I_n)(z-X_B\widehat{\psi})\right\}\\ &\times\exp\left\{-\frac{1}{2}(\psi-\widehat{\psi})'X_B'(\Sigma^{-1}\otimes I_n)X_B(\psi-\widehat{\psi})\right\}, \end{aligned}$$

from which the full conditional posterior of parameter ψ is the multivariate normal distribution

$$f(\psi|y,\Sigma,\theta)=\mathrm{N}(\widehat{\psi},(X_B'(\Sigma^{-1}\otimes I_n)X_B)^{-1}). \tag{11}$$

The size of the covariance matrix of the above distribution could be computationally difficult to handle due to its great size. In such a case, for any fixed i, rearranging the terms in ψ_i in Eq. (9) and completing its quadratic form, the complete conditional posterior of ψ_i is obtained:

$$f(\psi_i|y,\psi_{(i)},\Sigma,\theta) = \mathrm{N}\Bigg((X_i'Xi)^{-1}X_i' \times \left[z_i + \sum_{j\neq i}\frac{\sigma^{ij}}{\sigma^{ii}}(z_j - X_j\psi_j)\right], (\sigma^{ii}X_i'Xi)^{-1}\Bigg), \tag{12}$$

where $\psi_{(i)} = (\psi_1,\ldots,\psi_{i-1},\psi_{i+1},\ldots,\psi_m)'$ and σ^{ij} is the (i,j) element of matrix Σ^{-1}.

Finally, from Eq. (8) and using the properties of the Kronecker product, the posterior distribution can also be expressed in the following form:

$$f(\psi,\Sigma,\theta|y) \propto |\Sigma|^{-\frac{n+m+1}{2}}\exp\left\{-\frac{1}{2}\mathrm{tr}[\Sigma^{-1}H'H]\right\},$$

where H is the $n\times m$ matrix whose ith column is the vector $z_i - X_i\psi_i$. The full conditional posterior for the covariance matrix Σ is then

$$f(\Sigma|y,\psi,\theta) = \mathrm{IW}(n, H'H), \tag{13}$$

which is an inverted Wishart distribution with n degrees of freedom and an expectation given by $H'H/(n-m-1)$.

Since only two of the three full conditional posterior distributions are known, we propose a Metropolis-within-Gibbs algorithm which allows us to use the Metropolis algorithm to draw observations from the unknown posterior distribution $f(\theta|y,\psi,\Sigma)$ within the Gibbs sampling procedure.

3.1. Metropolis-within-Gibbs sampling

To obtain a sample from the posterior distribution $f(\psi,\Sigma,\theta|y)$, we propose to use the following Metropolis-within-Gibbs algorithm, whose jth step can be implemented as follows:

1. Simulate $\Sigma^{(j)}$ from $f(\Sigma|y,\psi^{(j-1)},\theta^{(j-1)})$, given by Eq. (13).
2. Simulate $\psi^{(j)}$ from $f(\psi|y,\Sigma^{(j)},\theta^{(j-1)})$, given by Eq. (11), if computationally feasible, or compute $\psi_i^{(j)}$ $(i=1,\ldots,m)$ from $f(\psi_i|y,\psi_1^{(j)},\ldots,\psi_{i-1}^{(j)},\psi_{i+1}^{(j-1)},\ldots,\psi_m^{(j-1)},\Sigma^{(j)},\theta^{(j-1)})$, given by Eq. (12).
3. Given $\theta^{(j-1)} = (\theta_1^{(j-1)},\theta_2^{(j-1)},\ldots,\theta_m^{(j-1)})$, where $\theta_i^{(j-1)}$ is the last simulated parameter vector for the ith time series, simulate $\theta^{(j)}$ using the following Metropolis algorithm:
   3.1 For $i = 1,2,\ldots,m$, simulate θ_i^* from a uniform distribution on the hypercube centered at $\theta_i^{(j-1)}$ and with side length $2l$, where l is a tuning parameter. Let $\theta^* = (\theta_1^*,\theta_2^*,\ldots,\theta_m^*)$ be the resulting parameter vector.
   3.2 Calculate $\alpha(\theta^{(j-1)},\theta^*) = \min\left\{1, \frac{f(\theta^*|y,\psi^{(j)},\Sigma^{(j)})}{f(\theta^{(j-1)}|y,\psi^{(j)},\Sigma^{(j)})}\right\}$ using Eq. (10).
   3.3 Simulate $u \sim Un(0,1)$.
   3.4 If $u \le \alpha(\theta^{(j-1)},\theta^*)$, set $\theta^{(j)} = \theta^*$. Otherwise, $\theta^{(j)} = \theta^{(j-1)}$.

After repeating the algorithm $N_0 + N$ times and discarding the first N_0 burn-in iterations, we obtain a sample of size N from the posterior distribution of the model parameters that allows us to estimate any of its characteristics.

Note that for each simulated value $\Sigma^{(j)}$ it is possible to compute the corresponding correlation coefficients $\rho_{kl}^{(j)}$, for $k = 1,2,\ldots,m-1$ and $l = k+1,\ldots,m$. Therefore, the proposed simulation procedure allows us to also obtain samples from the posterior distributions of the correlation coefficients, $f(\rho_{kl}|y)$, which can be used to verify the assumed correlation between the time series.

Since the draws produced by MCMC simulation methods are correlated, larger samples are required to achieve the level of estimation accuracy that would be achieved by independent draws. In addition, these draws need to be tested to assess whether or not convergence has taken place. Therefore, in the case of common parameters for the univariate models, the estimation of the homogeneous multivariate model is easier than the estimation of the general model, and therefore it should be used on those occasions where it is justified.

3.2. Model selection

Since model choice is a fundamental part of any statistical analysis, it has been the subject of considerable effort over past years, and as a consequence, many approaches for selecting the best model have been suggested in the literature. For a review of some of the most-used Bayesian techniques, see for example Robert (2007). In this paper, we propose to use the DIC (Deviance Information Criterion, see Spiegelhalter, Best, Carlin, & van der Linde, 2002) to decide between the two competing models: M_G, the general multivariate model proposed in this paper, and M_H, the homogeneous multivariate model. For any model, the DIC is defined as

$$DIC = E(D(\varphi)|y) + P_D,$$

where φ is the vector of parameters in the model and $D(\varphi) = -2 \log f(y|\varphi)$ is the Bayesian deviance. The first term in the definition, the posterior mean deviance, is a Bayesian measure of goodness of fit, while $P_D = E(D(\varphi)|y) - D(E(\varphi|y))$, known as the effective number of parameters in the model, is a measure of complexity. Models with smaller DIC values, with negative values being possible, are better supported by the data. Note that the DIC can easily be calculated from samples generated by MCMC simulation techniques. Besides, it can prove more satisfactory than either the AIC (Akaike's Information Criterion, see Akaike, 1974) or the BIC (Bayesian Information Criterion, see Schwarz, 1978), because it takes prior information into account and provides a natural penalization factor for the log-likelihood.

In our particular setting, see Eqs. (5) and (6), the deviance corresponding to the general multivariate model M_G is of the form

$$\begin{aligned} D(\psi, \Sigma, \theta) &= n \log |\Sigma| \\ &+ \sum_{i=1}^{m} \sum_{j=1}^{m} \sigma^{ij} (z_i - X_i \psi_i)'(z_j - X_j \psi_j). \end{aligned} \tag{14}$$

The expectation of this deviance, and the corresponding DIC value, can be approximated by Monte Carlo estimation once a sample from the posterior distribution of the model parameters has been obtained.

The homogeneous multivariate model is a special case of the general one, so its expected deviance could also be approached by Monte Carlo estimation using Eq. (14), constrained to a common smoothing parameter vector for all series.

4. Forecasting procedure

Once the posterior distribution for the parameters in the model has been obtained, the predictive distribution is calculated. This distribution is the final outcome of a forecasting problem, since it encapsulates all of the information concerning the future values of the given time series. In particular, it allows us to calculate point forecasts and prediction intervals (Geweke & Whiteman, 2006).

Let y_i be the $n \times 1$ vector of observed data for the ith series and let P_i be the $h \times 1$ vector of future data. Consider the joint $(n + h) \times 1$ vector $(y_i, P_i)'$ and suppose that it still follows model (3), where the vector ε_i and the matrices M_i and L_i are partitioned in such a way that:

$$\begin{pmatrix} y_i \\ P_i \end{pmatrix} = \begin{pmatrix} M_i^1 \\ M_i^2 \end{pmatrix} \psi_i + \begin{pmatrix} L_i^1 & 0 \\ L_i^{21} & L_i^2 \end{pmatrix} \begin{pmatrix} \varepsilon_i \\ v_i \end{pmatrix}.$$

We also assume that the correlation structure of the m series is the same during the forecasting period as during the training period. Under these assumptions, the predictive distribution of the joint vector of future values $P = (P_1', P_2', \ldots, P_m')'$ is computed in the Appendix using Monte Carlo integration. Then

$$f(P_i|y) \approx \frac{1}{N} \sum_{j=1}^{N} \text{MSt}(\mu_{P_i}^{(j)}, V_{P_i}^{(j)}, n - m + 1),$$

where MSt is the multivariate Student-t distribution with $n - m + 1$ degrees of freedom, mean vector $\mu_{P_i}^{(j)}$ and covariance matrix $V_{P_i}^{(j)}$, given by the following expressions computed at $(\psi, \theta) = (\psi^{(j)}, \theta^{(j)})$:

$$\begin{aligned} \mu_{P_i} &= L_i^{21} z_i + (M_i^2 - L_i^{21} X_i) \psi_i \\ V_{P_i} &= \frac{(z_i - X_i \psi_i)'(z_i - X_i \psi_i)}{\nu - 2} L_i^2 (L_i^2)'. \end{aligned}$$

The mean vector and variance matrix of the predictive distribution can be approached by

$$E(P_i|y) \approx \frac{1}{N} \sum_{j=1}^{N} \mu_{P_i}^{(j)}$$

$$\begin{aligned} V(P_i|y) \approx\ & \frac{1}{N} \sum_{j=1}^{N} V_{P_i}^{(j)} + \frac{1}{N} \sum_{j=1}^{N} \mu_{P_i}^{(j)} (\mu_{P_i}^{(j)})' \\ & - E(P_i|y) E'(P_i|y). \end{aligned}$$

Both prediction intervals for any step ahead forecasts, from 1 to h, and cumulative prediction intervals can also be approached by means of the corresponding quantiles of the mixture of univariate Student-t distributions.

When the selected model for the joint analysis of the time series is the homogeneous multivariate model, the integrations with respect to both Σ and ψ are analytical and the conditional predictive distributions $f(P_i|y, \theta)$ are obtained instead, which correspond to multivariate Student-t distributions (Bermúdez et al., 2009a). Given a sample $\{\theta^{(j)}\}_{j=1}^{N}$ from the posterior distribution of the common smoothing parameters, the marginal predictive distributions, together with their main moments and prediction intervals, can be estimated in a similar way using Monte Carlo integration.

5. Numerical results

In this section we illustrate the performance of our Bayesian forecasting procedure using two real correlated time series data sets. The first example uses the Industrial Production Index in three autonomous communities in Spain: Catalonia, the Valencian Community and Madrid. In the second, we analyse the total number of rural tourists in Asturias, Cantabria and Galicia, three autonomous communities in northern Spain. Both time series data sets are available from the web page of the Spanish National Institute of Statistics (http://www.ine.es, accessed on 30 November, 2009).

5.1. Industrial Production Index

The Industrial Production Index (IPI) measures the monthly evolution of the productive activity of industrial branches; that is, of the extractive, manufacturing and production industries, and the distribution of electrical energy, water and gas. It reflects the complete evolution of the quantity and quality, eliminating the influence of prices. In this section, we analyse the time series corresponding to the Industrial Production Index in Catalonia, the Valencian Community and Madrid from January 2002 to December 2008. Plots of the three time series are shown in Fig. 1. A seasonal cycle of 12 periods, which justifies the use of the Holt-Winters' model, is observed. Moreover, it is reasonable to assume correlated errors at each time point, and thus the use of the multivariate model may allow us to obtain more accurate forecasts than those derived from the univariate analysis. In order to test the performance of our method, we consider as historical data the observations for the first 6 years, 2002–2007, and forecast the last year, 2008, to measure the post-sample accuracy.

The first step in the analysis of the time series data is to select the most adequate multivariate Holt-Winters' model for describing their behaviour. This model selection is based on the DIC (see Section 3.2), and therefore simulation is needed. Specifically, we simulate $N = 20{,}000$ values from the posterior distribution of the parameters in the general multivariate model M_G using the Metropolis-within-Gibbs algorithm described in Section 3.1. Similarly, using the acceptance sampling algorithm proposed by Bermúdez et al. (2009a), we obtain a random sample of size $N = 20{,}000$ from the posterior distribution of the smoothing parameter vector in the homogeneous multivariate model M_H. The values of the DIC corresponding to the simulated samples are 812.11 and 817.46 respectively for M_G and M_H. Thus, the advisable multivariate model for the joint analysis of those series is the general multivariate Holt-Winters' model. The sample posterior means for the correlation coefficients are $\bar{\rho}_{12} = 0.86$, $\bar{\rho}_{13} = 0.72$ and $\bar{\rho}_{23} = 0.73$, where the subindexes 1, 2 and 3 correspond to the time series of Catalonia, Valencia and Madrid respectively, with the three 95% confidence intervals fluctuating between 0.57 and 0.92. Hence, as we suggested, the three time series are actually correlated, and therefore the use of the multivariate model is justified.

Fig. 2 shows, in solid and dashed lines respectively, the point forecasts and 95% prediction intervals obtained from the general multivariate model M_G for each time series. Solid points represent the real values. It is worth pointing out that the forecasts for March and April are affected by the movable Easter holidays. Thus, in the case of Valencia, the effect of Easter, which took place in March 2008, was aggravated by the local Fallas holidays. Similarly, we can appreciate the effect of the global crisis towards the end of year 2008, which was difficult to predict using data from previous years.

The fitting and forecast errors are given in Table 1, where, with the aim of comparing the forecast

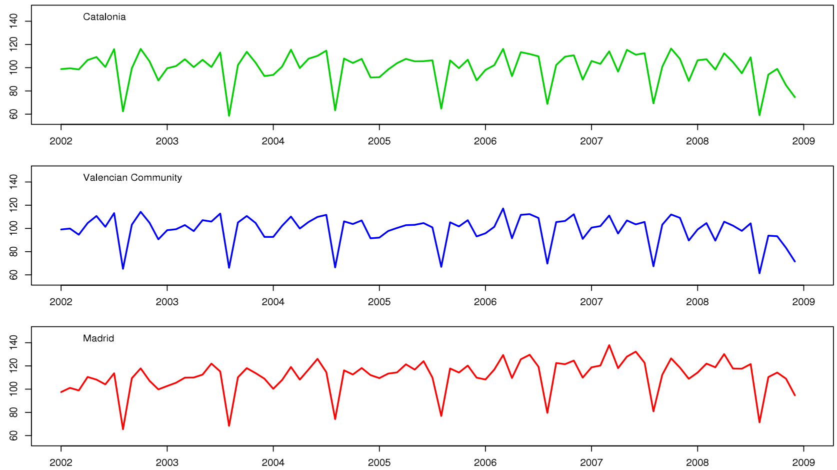

Fig. 1. Time series of the Industrial Production Index in three autonomous communities in Spain: Catalonia, Valencia and Madrid.

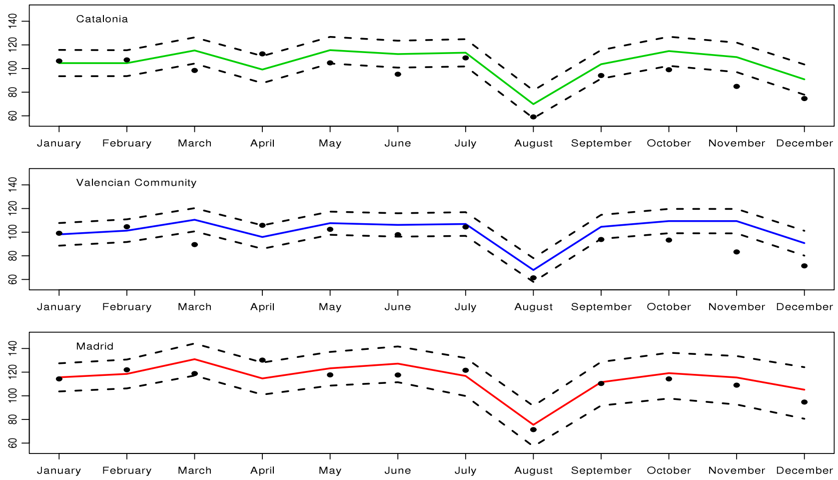

Fig. 2. Monthly point forecasts (solid line) and 95% prediction intervals (dashed lines) for the year 2008 obtained from the general multivariate Holt-Winters' model. Real data are represented by solid points.

Table 1
Fitting and forecast RMSE values obtained from both the general and homogeneous multivariate models and the univariate models, for each time series and their means.

	Catalonia		Valencia		Madrid		Means	
	Fitting	Forecast	Fitting	Forecast	Fitting	Forecast	Fitting	Forecast
General mult. model M_G	4.87	13.64	4.33	13.30	4.99	7.86	4.73	11.60
Homogeneous mult. model M_H	4.73	13.15	4.25	13.18	4.90	11.25	4.63	12.53
Univariate models	4.74	13.79	4.61	13.33	4.86	12.74	4.74	13.29

performance of the general multivariate model M_G with the performances of other models, we also include the corresponding errors obtained from the homogeneous model M_H and the univariate analysis of the series. The accuracy measure used in this study is the root mean squared error (RMSE).

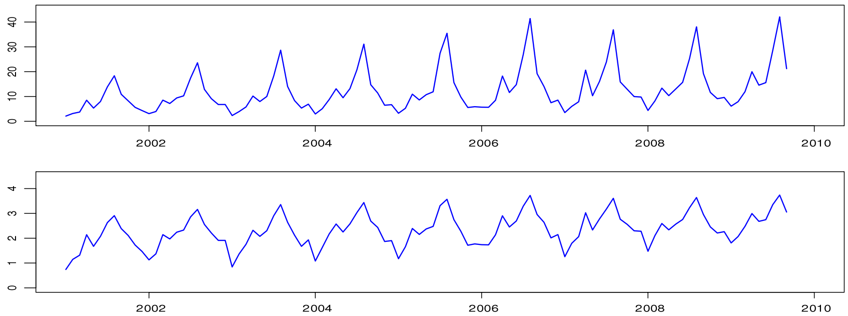

Fig. 3. Time plot of the total number of rural tourists in Cantabria, in thousands. Top: raw data. Bottom: log-transformed data.

The results given in Table 1 show that while the fitting errors for the three time series are similar in the two multivariate models, the general multivariate Holt-Winters' model, M_G, provides more accurate forecasts. Using this model it is possible to reduce the forecast error associated with Madrid considerably, with the mean of the forecasting errors for the three time series being smaller than that of the homogeneous multivariate model. Therefore, in this example the multivariate Holt-Winters' model proposed in this paper outperforms the previously studied homogeneous multivariate model. Moreover, the use of a multivariate model which takes the dependence between the series into account allows us to improve the forecast accuracy with respect to the univariate analysis.

5.2. Rural tourism in Northern Spain

Since tourism is a significant source of jobs and income, it is one of the mainstays of Spain's economy. Due to the high seasonality of the traditional 'sun and beach' tourism and the increasing competition from destinations abroad, the alternative of rural tourism is becoming increasingly important to the Spanish economy. Northern Spain offers some of the best natural surroundings for rural tourism, with Asturias, Cantabria and Galicia being some of the most popular destinations. Therefore, a good estimate of the demand is necessary to ensure that investment is at a level which is appropriate for the future. Specifically, in this example we study the data set of the monthly total number of rural tourists, in thousands, in Asturias, Cantabria and Galicia from January 2001 to September 2009.

The Cantabria series is plotted at the top of Fig. 3, and shows a regular growth and a seasonal cycle of 12 periods; however, the variability seems to increase with the level of the series, and thus a multiplicative Holt-Winters' model could be advisable. Instead, however, we proposed a logarithmic transformation, plotted at the bottom of Fig. 3, to stabilise the variance. The other two series behave similarly and seem to be highly correlated. We therefore forecast the three series on a logarithmic scale using a multivariate additive Holt-Winters' model, and recover the original units at the end of the study. In order to evaluate the performance of our forecasting procedure, we consider the observations for the first 8 years as historical data and use the remaining 9 values, corresponding to the year 2009, to measure the post-sample accuracy.

Repeating the simulation process described in the previous example, we obtain samples from the posterior distributions of the model parameters in M_G, the general multivariate Holt-Winters' model, and M_H, the homogeneous multivariate model. The values of the DIC corresponding to these simulated samples are -667.08 and -688.12 for M_G and M_H respectively. Therefore, the best multivariate model for the analysis of the time series is the homogeneous multivariate Holt-Winters' model. With the homogeneous model, the RMSE values for Asturias, Cantabria and Galicia are 0.24, 0.14 and 0.10, which are slightly smaller than those derived from the joint analysis of the series using the general multivariate Holt-Winters' model, namely 0.37, 0.15 and 0.10, respectively.

The point forecasts and 95% prediction intervals obtained from the homogeneous multivariate Holt-Winters' model are displayed in Fig. 4. The posterior

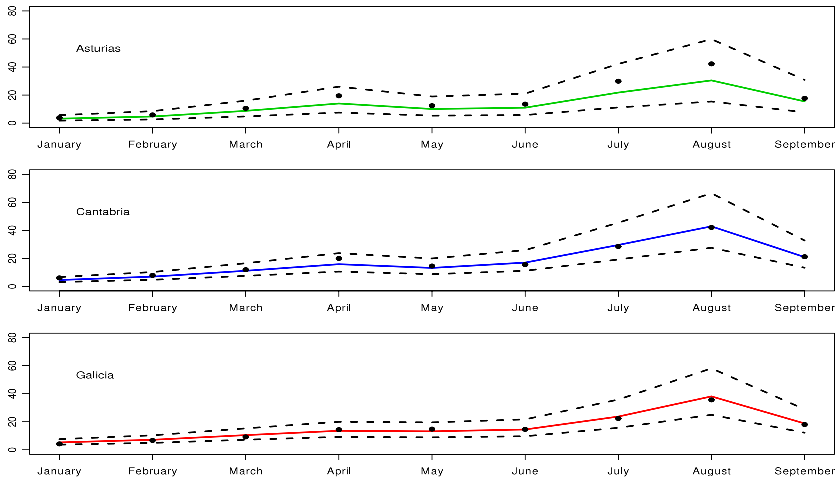

Fig. 4. Monthly point forecasts (solid line) and 95% prediction intervals (dashed lines) for the year 2009 derived from the analysis of the series using the homogeneous multivariate Holt-Winters' model. Real data are represented by solid points.

sample means for the correlation coefficients are 0.63, 0.61 and 0.66, while their 95% confidence intervals lie in the interval (0.46, 0.77). The existing correlation between the three time series justifies a multivariate analysis.

6. Concluding remarks

A multivariate extension of additive exponential smoothing, which allows us to jointly analyse time series with correlated errors at each time point, has been presented in this paper. The model is a generalisation of the homogeneous multivariate Holt-Winters' model proposed by Bermúdez et al. (2009a), which assumes a common structure for the univariate models describing the individual time series. This assumption, although sometimes reasonable, does not necessarily satisfy every time series data set. The Bayesian analysis of the general model, formulated as a seemingly unrelated regression model, is straightforward. The non-analytically tractable posterior distribution of the unknowns can be approached using MCMC simulation techniques. The predictive distribution of the future data of the time series is then estimated using Monte Carlo integration. A Bayesian model selection criterion is incorporated into the forecasting scheme to test for homogeneity.

The numerical calculus involved in the forecasting procedure is no more complex than that required in the univariate analysis of each series. The matrices that appear in the joint distribution of all observed data are very large, but they can be handled using only the matrices appearing in the univariate analysis, using the properties of the Kronecker product.

The results obtained from the prediction of real correlated time series are encouraging. With the use of the general multivariate exponential smoothing model proposed in this paper, it is possible to improve the forecast accuracy with respect to the homogeneous multivariate model when the assumption of a common structure for the univariate models is inappropriate.

This paper deals with linear innovations state space models only, as exponential smoothing models with multiplicative seasonal indexes need an alternative Bayesian analysis. If the innovations are still assumed to be homoscedastic and normally distributed, which is often a reasonable approximation, the likelihood function could be built as the product of the conditional distribution of each observation given the previous ones. The posterior distribution is then easy to obtain, except for the proportionality constant, whatever the prior used. However, the simulation technique that will allow the analysis of the posterior distribution and the obtaining of predictive distributions will depend on the specific model considered, as no general simulation procedure seems feasible for a general nonlinear innovations state space model. On those occasions where linear models are inadequate, they might still be useful after an adequate data transformation, as the rural tourism data series shows.

Acknowledgements

The authors want to acknowledge the editor and two anonymous referees; their comments and suggestions helped us to greatly improve a previous version of this paper. This research has been partially supported by the Ministerio de Ciencia e Innovación of Spain, Grant MTM2008-03993.

Appendix. Predictive distributions

Let $P = (P_1', P_2', \ldots, P_m')'$ be the $hm \times 1$ vector of future values of the m time series. The predictive distribution, $f(P|y)$, is obtained by solving the following expectation

$$f(P|y) = \int\int\int f(P|y, \psi, \Sigma, \theta) \times f(\psi, \Sigma, \theta|y)\mathrm{d}\psi\, \mathrm{d}\Sigma \mathrm{d}\theta,$$

which requires the calculation of the conditional predictive distribution $f(P|y, \psi, \Sigma, \theta)$. We assume that the joint $(n + h) \times 1$ vector $(y_i', P_i')'$ still follows model (3). Therefore, the vector of future values of the ith time series is given by $P_i = M_i^2 \psi_i + L_i^{21} \varepsilon_i + L_i^2 v_i$, for $i = 1, 2, \ldots, m$, where the matrices M_i and L_i have been partitioned in a similar way to the vector $(y_i', P_i')'$ and v_i is the error vector associated with the future data. Similarly, the matrix form of the observed data vector is $y_i = M_i^1 \psi_i + L_i^1 \varepsilon_i$, and consequently the error vector ε_i can be stated as $\varepsilon_i = z_i - X_i \psi_i$, where $z_i = (L_i^1)^{-1} y_i$ and $X_i = (L_i^1)^{-1} M_i^1$. The vector of future values of each time series, given the observed data, can then be expressed as $P_i = L_i^{21} z_i + (M_i^2 - L_i^{21} X_i)\psi_i + L_i^2 v_i = \widetilde{z}_i + D_i \psi_i + L_i^2 v_i$, and hence the joint distribution of P is multivariate normal

$$f(P|y, \psi, \Sigma, \theta) = \mathrm{N}(\widetilde{z} + D_B \psi, L_B^2(\Sigma \otimes I_h)(L_B^2)'),$$

where $\widetilde{z} = (\widetilde{z}_1', \widetilde{z}_2', \ldots, \widetilde{z}_m')'$; D_B and L_B^2 are the block-diagonal matrices whose ith diagonal blocks are D_i and L_i^2, respectively, for $i = 1, 2, \ldots, m$; and $v = (v_1', v_2', \ldots, v_m')'$ is the normal distributed error vector associated with the future data, $v \sim \mathrm{N}(0, \Sigma \otimes I_h)$.

The expectation with respect to Σ is analytical, and the result is

$$f(P|y, \psi, \theta) \propto |H'H + (\widetilde{P} - \widetilde{\mu}_P)'(\widetilde{P} - \widetilde{\mu}_P)|^{-\frac{n+h}{2}},$$

where $\widetilde{P}$ and $\widetilde{\mu}_P$ are the matrices defined as $\widetilde{P} = [(L_1^2)^{-1} P_1, \ldots, (L_m^2)^{-1} P_m]$ and $\widetilde{\mu}_P = [(L_1^2)^{-1}(\widetilde{z}_1 + (M_1^2 - L_1^{21} X_1)\psi_1), \ldots, (L_m^2)^{-1}(\widetilde{z}_m + (M_m^2 - L_m^{21} X_m)\psi_m)]$. Hence, the conditional predictive distribution of matrix $\widetilde{P}$ is the generalised Student-t distribution $f(\widetilde{P}|y, \psi, \theta) = \mathrm{GSt}(I_h, H'H, \widetilde{\mu}_P, n + h)$. As a property of that distribution (Zellner, 1971, p. 397), each one of the columns of $\widetilde{P}$ (and therefore the vectors P_i of future values of the time series) has a multivariate Student-t distribution $\mathrm{MSt}(\mu_{P_i}, V_{P_i}, n - m + 1)$, with $n - m + 1$ degrees of freedom and a mean vector and covariance matrix given by

$$E(P_i|y, \psi, \theta) = \mu_{P_i} = L_i^{21} z_i + (M_i^2 - L_i^{21} X_i)\psi_i$$

$$V(P_i|y, \psi, \theta) = V_{P_i} = \frac{(z_i - X_i \psi_i)'(z_i - X_i \psi_i)}{\nu - 2} L_i^2 (L_i^2)'.$$

The other two integrals in the calculus of $f(P|y)$ are not analytical, so we propose to use Monte Carlo integration to estimate the predictive distributions $f(P_i|y)$ and their main moments. Let $\{(\psi^{(j)}, \theta^{(j)})\}_{j=1}^N$ be a sample from the posterior distribution, then

$$f(P_i|y) \approx \frac{1}{N} \sum_{j=1}^N \mathrm{MSt}(\mu_{P_i}^{(j)}, V_{P_i}^{(j)}, n - m + 1).$$

References

Anderson, B. D. O., & Moore, J. B. (1979). *Optimal filtering* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723.

Bermúdez, J. D., Corberán-Vallet, A., & Vercher, E. (2009a). Multivariate exponential smoothing: a Bayesian forecast approach based on simulation. *Mathematics and Computers in Simulation*, *79*, 1761–1769.

Bermúdez, J. D., Corberán-Vallet, A., & Vercher, E. (2009b). Forecasting time series with missing data using Holt's model. *Journal of Statistical Planning and Inference*, *139*, 2791–2799.

Bermúdez, J. D., Segura, J. V., & Vercher, E. (2006). A decision support system methodology for forecasting of time series based on soft computing. *Computational Statistics and Data Analysis*, *51*, 177–191.

Bermúdez, J. D., Segura, J. V., & Vercher, E. (2007). Holt-Winters forecasting: an alternative formulation applied to UK air passenger data. *Journal of Applied Statistics*, *34*, 1075–1090.

Bermúdez, J. D., Segura, J. V., & Vercher, E. (2008). SIOPRED: a prediction and optimisation integrated system for demand. *TOP*, *16*, 258–271.

Bermúdez, J. D., Segura, J. V., & Vercher, E. (2010). Bayesian forecasting with the Holt-Winters model. *Journal of the Operational Research Society*, *61*, 164–171.

Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis, forecasting and control* (2nd ed.). San Francisco: Holden-Day.

Brown, R. G. (1959). *Statistical forecasting for inventory control*. New York: McGraw-Hill.

de Silva, A., Hyndman, R. J., & Snyder, R. D. (2007). *The vector innovation structural time series framework: a simple approach to multivariate forecasting*. Working paper 03/07, Department of Econometrics and Business Statistics, Monash University.

Enns, P. G., Machak, J. A., Spivey, W. A., & Wrobleski, W. J. (1982). Forecasting applications of an adaptive multiple exponential smoothing model. *Management Science*, *28*, 1035–1044.

Fernández, F. J. (1990). Estimation and testing of a multivariate exponential smoothing model. *Journal of Time Series Analysis*, *11*, 89–105.

Fernández, F. J., & Harvey, A. C. (1990). Seemingly unrelated time series equations and a test for homogeneity. *Journal of Business and Economic Statistics*, *8*, 71–81.

Forbes, C. S., Snyder, R. D., & Sharmi, R. G. (2000). *Bayesian exponential smoothing*. Working paper 07/00, Department of Econometrics and Business Statistics, Monash University.

Gardner, E. S., Jr. (2006). Exponential smoothing: The state of the art—Part II. *International Journal of Forecasting*, *22*, 637–666.

Geweke, J. F., & Whiteman, C. H. (2006). Bayesian forecasting. In G. Elliot, C. W. J. Granger, & A. Timmermann (Eds.), *The handbook of economic corecasting*. Amsterdam: North-Holland.

Gould, P. G., Koehler, A. B., Ord, J. K., Snyder, R. D., Hyndman, R. J., & Vahid-Araghi, F. (2008). Forecasting time series with multiple seasonal patterns. *European Journal of Operational Research*, *191*, 207–222.

Griffiths, W. E. (2001). *Bayesian inference in the seemingly unrelated regression model*. Technical Report, Department of Economics, The University of Melbourne.

Harrison, P. J., & Stevens, C. F. (1976). Bayesian forecasting. *Journal of the Royal Statistical Society, Series B*, *38*, 205–247.

Harvey, A. C. (1986). Analysis and generalisation of a multivariate exponential smoothing model. *Management Science*, *32*, 374–380.

Holt, C. C. (1957). *Forecasting trends and seasonals by exponentially weighted averages.* O.N.R. Memorandum 52/1957. Carnegie Institute of Technology.

Hyndman, R. J., Akram, M., & Archibald, B. C. (2008). The admissible parameter space for exponential smoothing models. *Annals of the Institute of Statistical Mathematics*, *60*, 407–426.

Hyndman, R. J., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with exponential smoothing: the state space approach*. Berlin: Springer.

Hyndman, R. J., Koehler, A. B., Snyder, R. D., & Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, *18*, 439–454.

Koehler, A. B., Snyder, R. D., & Ord, J. K. (2001). Forecasting models and prediction intervals for the multiplicative Holt-Winters' method. *International Journal of Forecasting*, *17*, 269–286.

Ord, J. K., Koehler, A. B., & Snyder, R. D. (1997). Estimation and prediction for a class of dynamic nonlinear statistical models. *Journal of the American Statistical Association*, *92*, 1621–1629.

Pfeffermann, D., & Allon, J. (1989). Multivariate exponential smoothing: method and practice. *International Journal of Forecasting*, *5*, 83–98.

Robert, C. P. (2007). *The Bayesian choice: from decision-theoretic foundations to computational implementation*. New York: Springer.

Snyder, R. D. (1985). Recursive estimation of dynamic linear models. *Journal of the Royal Statistical Society, Series B*, *47*, 272–276.

Snyder, R. D. (1986). *Estimation of a dynamic linear model with unknown starting values*. Working paper 12/86, Department of Econometrics and Business Statistics, Monash University.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*, 461–464.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B*, *64*, 583–639.

Taylor, J. W. (2003). Short-term electricity demand forecasting using double seasonal exponential smoothing. *Journal of the Operational Research Society*, *54*, 799–805.

West, M., & Harrison, P. J. (1989). *Bayesian forecasting and dynamic models*. New York: Springer-Verlag.

Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*, *6*, 324–342.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and test for aggregation bias. *Journal of the American Statistical Association*, *57*, 348–368.

Zellner, A. (1971). *An introduction to Bayesian inference in econometrics*. New York: Wiley.