Ann. Inst. Statist. Math.
Vol. 43, No. 3, 493–504 (1991)

A NONLINEAR TIME SERIES MODEL AND ESTIMATION OF MISSING OBSERVATIONS

BOVAS ABRAHAM[1]* AND A. THAVANESWARAN[2]*

[1] *Department of Statistics and Actuarial Sciences, University of Waterloo, Waterloo, Ontario, Canada, N2L 3G1*
[2] *Department of Statistics, University of Manitoba, Winnipeg, Manitoba, Canada, R3T 2N2*

(Received August 10, 1989; revised January 8, 1990)

Abstract. This paper formulates a nonlinear time series model which encompasses several standard nonlinear models for time series as special cases. It also offers two methods for estimating missing observations, one using prediction and fixed point smoothing algorithms and the other using optimal estimating equation theory. Recursive estimation of missing observations in an autoregressive conditionally heteroscedastic (ARCH) model and the estimation of missing observations in a linear time series model are shown to be special cases. Construction of optimal estimates of missing observations using estimating equation theory is discussed and applied to some nonlinear models.

Key words and phrases: Kalman filter, missing observations, nonlinear time series, optimal estimation, robustness.

1. Introduction

Quite often data analysts are faced with the problem of missing data. Data that are known to have been observed erroneously can fairly safely be categorized as missing. Erroneous data can also wreak havoc with the estimation and forecasting of linear or nonlinear time series models. Abraham (1981) proposed a procedure to interpolate the adjacent missing values on the basis of the known segments of an autoregressive integrated moving average (ARIMA (p, d, q)) process. Recently Jones (1985) proposed a state space Kalman filter approach to handle unequally spaced data in linear time series models.

In Section 2 we are concerned with the so-called conditionally Gaussian system which is treated in a filtering theoretic context in Shiryayev (1984) and in Ruskeepaa (1985). Exploiting this theory, we develop a state space approach and discuss a general framework for estimating missing observations in a nonlinear

* Authors were supported in part by a grant from the Natural Sciences and Engineering Research Council of Canada.

time series model. In Section 3 we discuss a number of special cases of the model in Section 2.

In Section 4 we describe the method of optimal estimation of Godambe (1985) to estimate missing observations in a nonlinear time series model. This approach extends the work of Ferreiro (1987) to nonlinear heteroscedastic time series models.

2. Nonlinear state space models

The linear state space system is given by

$$\theta_{t+1} = \alpha_t\theta_t + \beta_t u_{t+1}, \qquad y_t = A_{t-1}\theta_t + B_{t-1}v_t \tag{2.1}$$

where θ_t and u_t are $p \times 1$ vectors, y_t and v_t are $q \times 1$ vectors, α_t and β_t are $p \times p$ matrices, and A_t and B_t are matrices of dimensions $q \times p$ and $q \times q$, respectively. $\{y_t\}$ represents the observed time series, whereas α_t, A_t, β_t, B_t are known matrices of nonrandom functions while the vectors $\{u_t\}$, $\{v_t\}$ are independent, each being a sequence of independent normal random vectors having components with zero mean and unit variances. In order to handle various deviations which may occur in practice, several generalizations of (2.1) have been suggested. Among these are conditionally Gaussian sequences given in Shiryayev (1984), nonlinear state space models treated in Broemeling (1985) and Priestley's state-dependent models (1980).

In this paper we consider the model in (2.1) with random coefficients. We allow the coefficients in (2.1) to depend on past observations: $\alpha_t = \alpha(t, \mathcal{F}_t^y)$, $\beta_t = \beta(t, \mathcal{F}_t^y)$, $A_{t-1} = A(t-1, \mathcal{F}_{t-1}^y)$ and $B_{t-1} = B(t-1, \mathcal{F}_{t-1}^y)$, where $\mathcal{F}_t^y$ denotes the σ-field generated by the observations up to time t. We refer to (2.1) under these settings as the generalized model (2.1). This generalized model encompasses some of the nonlinear time series models that have been proposed in the literature.

(i) ARCH models: Suppose that $\alpha_t = \alpha$, $A_{t-1} = A$ and $B_t \equiv 0$ so that

$$\theta_{t+1} = \alpha\theta_t + \beta_t u_{t+1}, \qquad y_t = A\theta_t.$$

This is the ARCH model described in Engle (1982).

(ii) Dynamic linear state space models: When $\{\alpha_t\}$, $\{\beta_t\}$ and $\{B_t\}$ are constant matrices and A_{t-1} is a matrix of "known functions" at $t-1$, (i.e. A_{t-1} is $\mathcal{F}_{t-1}^y$ measurable) the generalized model (2.1) becomes

$$\theta_t = \alpha\theta_{t-1} + u_t, \qquad y_t = A_{t-1}\theta_t + v_t$$

which is the state space model described in Harrison and Stevens (1976).

(iii) Doubly stochastic time series model (cf. Tjøstheim (1986)): When $\alpha_t = 1$, $\beta_t = 1$, $u_{t+1} = \epsilon_{t+1} - \epsilon_{t-1}$ and $B_t = 1$, (2.1) becomes

$$\theta_{t+1} = \theta_t + u_{t+1}, \qquad y_t = A_{t-1}\theta_t + v_t.$$

This corresponds to the doubly stochastic time series model

$$\theta_t = \theta + \epsilon_t + \epsilon_{t-1}, \qquad y_t = \theta_t f(t, \mathcal{F}_{t-1}^y) + e_t$$

considered in Thavaneswaran and Abraham (1988). When $f(t, \mathcal{F}_{t-1}^y) = y_{t-1}$, this turns out to be a special case of the RCA model of Nicholls and Quinn (1982). Moreover, if we take $\theta_t = \alpha_{t-1}\theta_{t-1} + u_t$, $y_t = \theta_t$ with $\alpha_{t-1} = \phi + \pi \exp(-\gamma y_{t-1}^2)$ then the generalized model (2.1) describes the exponential autoregressive model of Ozaki (1985).

The following theorems give prediction and fixed point smoothing algorithms for the generalized model (2.1).

THEOREM 2.1. *Let* $\hat{\theta}_t = E[\theta_t \mid \mathcal{F}_{t-1}^y]$, $\Sigma_t = E((\theta_t - \hat{\theta}_t)(\theta_t - \hat{\theta}_t)^T \mid \mathcal{F}_{t-1}^y)$. *Then*

$$\hat{\theta}_{t+1} = \alpha_t \hat{\theta}_t + k_t[y_t - \hat{y}_t],$$
$$\Sigma_{t+1} = \beta_t \beta_t^T + (\alpha_t - k_t A_{t-1})\Sigma_t(\alpha_t - k_t A_t)^T + k_t B_{t-1} B_{t-1}^T k_t^T,$$

where $k_t = \alpha_t \Sigma_t A_{t-1}^T [A_{t-1}\Sigma_t A_{t-1}^T + B_{t-1}B_{t-1}^T]^+$ *and* $\hat{y}_t = E(y_t \mid \mathcal{F}_{t-1}^y)$; M^T *and* M^+ *denote the transpose and the pseudo inverse, respectively, of a matrix* M.

PROOF. A straightforward extension of results in Brockwell and Davis (1987).

We now introduce another theorem on fixed-point smoother to obtain recursive estimates of m missing observations say, $y_m = (y_{t_1}, \dots, y_{t_j}, \dots, y_{t_m})$. The basic idea here is the same as that in the derivation of the recursive estimate of a parameter θ_{t_j} $(j = 1, \dots, m)$, based on the observations up to time t $(t > t_j)$, as a function of the estimate based on $t-1$ $(t > t_j + 1)$ and the observation at time t. This will also enable us to get an idea of how the estimate of the parameter (missing value) changes when a new observation becomes available.

THEOREM 2.2. *For* $t > t_j$, *let* $\tilde{\theta}_{t_j|t} = E[\theta_{t_j} \mid \mathcal{F}_t^y]$ *be the estimate of* θ_{t_j} *based on the observations up to time* t, $\tilde{\Sigma}_t$ *be the covariance matrix*

$$\tilde{\Sigma}_t = E[(\theta_{t_j} - \hat{\theta}_{t_j})(\theta_t - \hat{\theta}_t)^T \mid \mathcal{F}_{t-1}^y]$$

and

$$\Sigma_t^* = E[(\theta_{t_j} - \tilde{\theta}_{t_j|t})(\theta_{t_j} - \tilde{\theta}_{t_j|t})^T \mid \mathcal{F}_{t-1}^y].$$

Then

$$\tilde{\theta}_{t_j|t} = \tilde{\theta}_{t_j|t-1} + \tilde{k}_t(y_t - A_{t-1}\hat{\theta}_t), \qquad t > t_j$$

where

$$\tilde{k}_t = \tilde{\Sigma}_t A_{t-1}^T [A_{t-1}\Sigma_t A_{t-1}^T + B_{t-1}B_{t-1}^T]^+,$$
$$\tilde{\Sigma}_{t+1} = \tilde{\Sigma}_t[\alpha_t - k_t A_{t-1}]^T, \qquad \Sigma_{t-1}^* = \Sigma_t \qquad \textit{for} \qquad t < t_j \qquad \textit{and}$$
$$\Sigma_t^* = \Sigma_{t-1}^* - \tilde{\Sigma}_t A_{t-1}^T \tilde{k}_t^T, \qquad t \geq t_j.$$

PROOF. A straightforward generalization of results in Brockwell and Davis (1987) or Shiryayev (1984).

3. Applications to missing data

Missing values in time series have been usually estimated using two different approaches. The first one, a Bayesian approach, uses the Kalman filtering technique, while in the second one, a non Bayesian approach, the missing values are treated as parameters (fixed). In this section we follow the Kalman type recursive approach to estimate the missing values by replacing them with normal random variables. This approach may be viewed as one which uses a prior for the parameter which replaces the missing value.

3.1 *ARCH type models with one missing observation*

Now we indicate an appropriate way to modify a given nonlinear time series to reflect the fact that the observation at time m is missing. Let $\{X_t\}$ be a time series in which X_m is missing and $\boldsymbol{X}_n' = (X_1, \ldots, X_{m-1}, X_{m+1}, \ldots, X_n)$. If we know the first two conditional moments $E[X_{t+1} \mid \mathcal{F}_t^x]$ and $\mathrm{Var}[X_{t+1} \mid \mathcal{F}_t^x]$, then X_{t+1} can be written as

$$X_{t+1} = E[X_{t+1} \mid \mathcal{F}_t^x] + X_{t+1} - E[X_{t+1} \mid \mathcal{F}_t^x]. \tag{3.1}$$

Suppose that the time series X_t satisfies

$$E[X_{t+1} \mid \mathcal{F}_t^x] = \alpha_{t-1} X_t \quad \text{and} \quad X_{t+1} - E[X_{t+1} \mid \mathcal{F}_t^x] = \beta_{t-1} u_{t+1} \tag{3.2}$$

where α_{t-1} and β_{t-1} are $\mathcal{F}_{t-1}^x$ measurable and $\{u_t\}$ is an i.i.d. $N(0, 1)$ sequence. Then X_{t+1} has the ARCH representation

$$X_{t+1} = \alpha_{t-1} X_t + \beta_{t-1} u_{t+1}. \tag{3.3}$$

Note. The restriction in (3.2) is introduced to apply the recursive approach. However, the method of Section 4 can be applied in the more general set up in which the coefficients of X_t and u_{t+1} are $\mathcal{F}_t^x$ measurable.

Now we consider the estimation of a missing observation as a parameter estimation problem in a particular formulation of the generalized model (2.1):

$$\begin{aligned} \theta_{t+1} &= \alpha_{t-1}\theta_t + \beta_{t-1} u_{t+1}, \\ X_t &= A_{t-1}\theta_t, \\ y_t &= A_{t-1}\theta_t + B_{t-1} v_t \end{aligned} \tag{3.4}$$

with $A_{m-1} = 0$, $B_{m-1} = 1$, $A_t = 1$, $t \neq m-1$; $B_t = 0$, $t \neq m-1$. Then $\boldsymbol{Y} = (X_1, \ldots, X_{m-1}, v_m, X_{m+1}, \ldots, X_n)$ is the extended observed series. Here v_m is a normal random variable replacing the missing observation. Such a formulation was also considered in Brockwell and Davis (1987).

Using Theorems 2.1 and 2.2 we have

$$\begin{aligned} k_t &= \alpha_{t-1}\Sigma_t A_{t-1}[A_{t-1}^2\Sigma_t + B_{t-1}^2]^+, \\ \Sigma_{t+1} &= \beta_{t-1}^2 + (\alpha_{t-1} - k_t A_{t-1})^2\Sigma_t + k_t^2 B_{t-1}^2. \end{aligned}$$

This implies that $k_t = \alpha_{t-1}$, $\Sigma_{t+1} = \beta_{t-1}^2$, $t \neq m$ and

$$k_m = 0 \quad \text{and} \quad \Sigma_{m+1} = \beta_{m-1}^2 + \alpha_{m-1}^2 \Sigma_m.$$

Moreover,

$$\begin{aligned} \tilde{\Sigma}_{t+1} &= \tilde{\Sigma}_t[\alpha_{t-1} - k_t A_{t-1}] = 0 \quad \text{for} \quad t \neq m, \\ \tilde{\Sigma}_{m+1} &= \tilde{\Sigma}_m \alpha_{m-1} = \Sigma_m \alpha_{m-1} = \alpha_{m-1}\beta_{m-2}^2. \end{aligned}$$

Hence,

$$\tilde{\theta}_{m|t} = \tilde{\theta}_{m|t-1} + \tilde{k}_t[y_t - A_{t-1}\hat{\theta}_t] \tag{3.5}$$

so that at $t = m+1$

$$\tilde{\theta}_{m|m+1} = \tilde{\theta}_{m|m} + \tilde{k}_{m+1}[y_{m+1} - A_m\hat{\theta}_{m+1}]$$

where

$$\tilde{k}_{m+1} = \frac{\tilde{\Sigma}_{m+1}}{\Sigma_{m+1}} = \frac{\beta_{m-2}^2 \alpha_{m-1}}{\beta_{m-1}^2 + \alpha_{m-1}^2 \beta_{m-2}^2}.$$

Thus the estimate of the m-th observation based on $\boldsymbol{X}_{m+1}$ is

$$\begin{aligned} \tilde{X}_{m|m+1} &= \alpha_{m-2}X_{m-1} + \frac{\beta_{m-2}^2 \alpha_{m-1}}{\beta_{m-1}^2 + \alpha_{m-1}^2 \beta_{m-2}^2}[X_{m+1} - \alpha_{m-1}\alpha_{m-2}X_{m-1}] \\ &= \frac{\beta_{m-2}^2 \alpha_{m-1} X_{m+1} + \beta_{m-1}^2 \alpha_{m-2} X_{m-1}}{\beta_{m-1}^2 + \alpha_{m-1}^2 \beta_{m-2}^2}. \end{aligned} \tag{3.6}$$

It should be noted that $\tilde{\Sigma}_{t+1} = 0$ for $t \neq m$ and hence $\tilde{k}_{t+1} = 0$ for $t \neq m$. Therefore, (3.5) yields

$$\tilde{X}_{m|t} = \tilde{X}_{m|m+1} \quad \text{for} \quad t > m+1.$$

In the special case of a model with a constant conditional variance, $\beta_m^2 = \beta_{m-1}^2 = \sigma^2$, the estimate of the missing value is given by

$$\tilde{X}_{m|m+1} = \frac{\alpha_{m-1}X_{m+1} + \alpha_{m-2}X_{m-1}}{1 + \alpha_{m-1}^2}. \tag{3.7}$$

When $\beta_m = \text{const}$ and $\alpha_m = \phi$, the estimate of the missing value (X_m) for an AR(1) model, $X_{t+1} = \phi X_t + u_{t+1}$, becomes

$$\tilde{X}_{m|m+1} = \frac{\phi}{1+\phi^2}[X_{m+1} + X_{m-1}].$$

Moreover, for a nonlinear model of the form

$$X_{t+1} = \phi X_{t-1} X_t + u_{t+1}$$

in which the m-th observation X_m is missing, the estimate of X_m based on $\mathcal{F}^x_{m+1}$ or $\mathcal{F}^x_n$ is given by

$$\tilde{X}_{m|m+1} = \frac{\phi X_{m-1}}{1+\phi^2 X^2_{m-1}}[X_{m-2} + X_{m+1}].$$

Autoregressive models with deterministic time varying coefficients:

Models of the form

$$X_t - \alpha(t, \phi)X_{t-1} = u_t \tag{3.8}$$

have been found to be quite useful, in particular in signal processing (c.f. Charbonnier *et al.* (1987)). As in (3.7), it can be shown that the estimate $\tilde{X}_{m|m+1}$ of the missing observation based on $\mathcal{F}^x_{m+1}$ is given by

$$\tilde{X}_{m|m+1} = \frac{\alpha(m+1, \phi)X_{m+1} + \alpha(m, \phi)X_{m-1}}{1+\alpha^2(m+1, \phi)}.$$

Bilinear models:

Consider the model

$$X_t - \phi X_{t-1} = cu_t + \beta X_{t-2}u_t$$

The estimation of a missing observation, X_m, can be obtained by writing the model as

$$X_t = \alpha_{t-2}X_{t-1} + \beta_{t-2}u_t$$

where

$$\alpha_{t-2} = \phi \quad \text{and} \quad \beta_{t-2} = c + \beta X_{t-2}.$$

Hence the estimate of X_m, $\tilde{X}_{m|m+1} = E[X_m \mid \mathcal{F}^x_{m+1}]$ can be obtained as in the case of model (3.8) and is given by

$$\tilde{X}_{m|m+1} = \frac{\phi\beta^2_{m-2}X_{m+1} + \phi\beta^2_{m-1}X_{m-1}}{\beta^2_{m-1} + \phi^2\beta^2_{m-2}}.$$

3.2 *Two consecutive missing observations*

We now consider a slightly modified form of the model (3.3):

$$X_{t+1} = \alpha_{t-2}X_t + \beta_{t-2}u_{t+1}$$

where $\beta^2_t = \sigma^2$, X_m and X_{m+1} are missing and α_t is $\mathcal{F}^x_t$ measurable. The problem is to estimate X_m based on the available data $(X_1, X_2, \ldots, X_{m-1}, X_{m+2}, \ldots, X_n)$. The corresponding state-space model may be written as

$$\theta_{t+1} = \alpha_{t-2}\theta_t + \beta_{t-2}u_{t+1}, \qquad X_t = A_{t-1}\theta_t, \qquad y_t = A_{t-1}\theta_t + B_{t-1}v_t$$

where $A_{m-1} = 0$, $A_m = 0$, $B_m = 1$, $B_{m-1} = 1$ and $B_t = 0$, $A_t = 1$ for $t \neq m,\ m-1$. Then using Theorems 2.1 and 2.2 it is easy to show that for $t \neq m,\ m-1$, $k_t = \alpha_{t-1}$

and $k_m = 0 = k_{m+1}$. Also it can be shown that for $t \neq m,\ m+1$, $\Sigma_{t+1} = \sigma^2$ and $\tilde{\Sigma}_{t+1} = \tilde{\Sigma}_t[\alpha_{t-1} - k_t A_{t-1}] = 0$.

$$\Sigma_{m+1} = \sigma^2(1+\alpha_{m-2}^2), \qquad \Sigma_{m+2} = \sigma^2[1+\alpha_{m-1}^2+\alpha_{m-2}^2\alpha_{m-1}^2],$$
$$\tilde{\Sigma}_{m+1} = \sigma^2\alpha_{m-2}, \qquad \tilde{\Sigma}_{m+2} = \sigma^2\alpha_{m-1}\alpha_{m-2}.$$

Also $\tilde{k}_t = 0$, $t \neq m+2$ and

$$\tilde{k}_{m+2} = \frac{\alpha_{m-1}\alpha_{m-2}}{[1+\alpha_{m-1}^2+\alpha_{m-1}^2\alpha_{m-2}^2]}.$$

Then the estimate of X_m based on $\mathcal{F}_{m+2}^x$, is given by

$$\frac{(\alpha_{m-2}+\alpha_{m-3}\alpha_{m-1}^2)X_{m-1}+\alpha_{m-1}\alpha_{m-2}X_{m+2}}{1+\alpha_{m-1}^2(1+\alpha_{m-2}^2)}.$$

It should be noted that when $\alpha_t = \phi$ and the model becomes AR(1), the estimate of the m-th observation becomes

$$\tilde{X}_{m|m+2} = \frac{(\phi+\phi^3)X_{m-1}+\phi^2X_{m+2}}{1+\phi^2+\phi^4}$$
$$= \frac{(1-\phi^4)\phi X_{m-1}+\phi^2(1-\phi^2)X_{m+2}}{(1-\phi^6)}.$$

Similarly, we can also obtain $\hat{X}_{m+1|m+2}$. These estimates are the same as those obtained by Abraham (1981) and Miller and Ferreiro (1984).

As noted before (see the note after equation (3.3)), the approach presented in this section can only handle some of the nonlinear models mentioned before. Hence we consider a more general approach in the next section.

4. Optimal estimation of missing observations

Following Godambe (1985), the optimal estimation of parameters in adaptive as well as nonadaptive nonlinear time series has been discussed in Thavaneswaran and Abraham (1988). In this section, we briefly describe the estimation of missing observations considering them as parameters.

Let $y_1, \ldots, y_n$ be an observed time series with y_m $(1 < m < n)$ missing and the parameters θ_t be those known from the generalized model (2.1). Then, when considering y_m as a parameter we can obtain its optimal estimate as in Thavaneswaran and Abraham (1988).

Example 4.1. (ARCH model) Consider the model $y_{t+1} = \alpha_t y_t + \beta_t u_{t+1}$, where $\alpha_t = \alpha(t, \mathcal{F}_t^y, \theta)$, $\beta_t = \beta(t, \mathcal{F}_t^y, \theta)$ and $\{u_t\}$ are a sequence of i.i.d. random variables having mean zero and finite variance σ^2. Here it should be noted that we are not making any distributional assumption on the errors. It can be shown that the optimal estimate of y_m satisfies

$$\sum_{t=1}^{n} a_t^*(y_{t+1} - \alpha_t y_t) = 0 \tag{4.1}$$

where

$$a_t^* = \frac{E\left[\frac{\partial}{\partial y_m}(y_{t+1} - \alpha_t y_t) \mid \mathcal{F}_t^y\right]}{\beta_t^2 \sigma^2}. \tag{4.2}$$

In the special case of an AR(1) model, $y_{t+1} = \phi y_t + u_{t+1}$, the estimate of y_m turns out to be the solution of

$$(\hat{y}_m - \phi y_{m-1}) - \phi(y_{m+1} - \phi \hat{y}_m) = 0$$

and the optimal estimate is

$$\hat{y}_m = \frac{\phi(y_{m+1} + y_{m-1})}{1 + \phi^2}.$$

This is the same as what we obtained in Subsection 3.1.

Example 4.2. (RCA model) Let

$$y_t = (\phi + \beta_t) y_{t-1} + u_t$$

where $\{u_t\}$ and $\{\beta_t\}$ are zero mean square integrable independent sequences and $V(u_t) = \sigma_u^2$ and $V(\beta_t) = \sigma_\beta^2$; β_t is independent of $\{u_t\}$ and $\{y_{t-1}\}$. Then the optimal estimate of y_m (treated as a parameter) can be given as a solution of the nonlinear equation

$$[(y_m - \phi y_{m-1})/(\sigma_u^2 y_{m-1}^2 + \sigma_\beta^2)] - \phi(y_{m+1} - \phi y_m)/(\sigma_u^2 + y_m^2 \sigma_\beta^2) = 0. \tag{4.3}$$

It can also be seen that the least square estimate of y_m is the solution of

$$(y_m - \phi y_{m-1}) - \phi(y_{m+1} - \phi y_m) = 0 \tag{4.4}$$

and is given by

$$\hat{y}_m(\mathrm{LS}) = \phi(y_{m-1} + y_{m+1})/(1 + \phi^2).$$

This is the same as that previously obtained for an AR(1) process. However, the optimal estimate will not be the same in both cases.

This estimate depends on the conditional variance of the observed series which in turn depends on the missing value, y_m. Hence we first find the least square estimate $\hat{y}_m(\mathrm{LS})$ of y_m and then use it to obtain the weights $w_1 = \sigma_u^2 + \sigma_\beta^2 y_{m-1}^2$ and $w_2 = \sigma_u^2 + \sigma_\beta^2 \hat{y}_m^2(\mathrm{LS})$ to calculate the optimal estimate.

We propose to find the optimal estimate using the following steps.

Step 1. Obtain $\hat{y}_m(\mathrm{LS})$

Step 2.

$$\hat{y}_m(op) = \frac{\phi(w_2 y_{m-1} + w_1 y_{m+1})}{w_1 + w_2}$$

where $w_1 = \sigma_u^2 + \sigma_\beta^2 y_{m-1}^2$, $w_2 = \sigma_u^2 + \sigma_\beta^2 \hat{y}_m^2(\mathrm{LS})$.

In this algorithm we assume that the model parameters ϕ, σ_u^2 and σ_β^2 are known. Such an assumption about the model parameters is not uncommon in the context of estimating missing observations (for example, see Abraham (1981)). However, in practice, model parameters may be estimated using part of the data (Abraham (1981)). Such an approach for optimal estimation in nonlinear models will be pursued in a subsequent paper.

The extension of the results to the p-th order RCA model

$$y_t = \sum_{i=1}^{p} (\phi_i + \beta_{it}) y_{t-i} + u_t$$

is immediate.

Example 4.3. (Doubly Stochastic Time Series) Consider the model

$$y_t = \phi_t y_{t-1} + u_t \tag{4.5}$$

where ϕ_t is a moving average sequence of the form

$$\phi_t = \phi + \epsilon_t + \epsilon_{t-1} \tag{4.6}$$

such that $\{\phi_t\}$, $\{\epsilon_t\}$ are square integrable independent sequences; $\{\epsilon_t\}$ and $\{u_t\}$ are zero mean independent Gaussian sequences with variances σ_ϵ^2 and σ_u^2. Then $\mu_t = E(\epsilon_t \mid \mathcal{F}_t^y)$ and $\gamma_t = E[(\epsilon_t - \mu_t)^2 \mid \mathcal{F}_t^y]$ satisfy the recursive algorithms (see Thavaneswaran and Abraham (1988)),

$$\mu_t = \sigma_\epsilon^2 y_{t-1}[(y_t - (\phi + \mu_{t-1}))y_{t-1}]/[\sigma_u^2 + y_{t-1}^2(\sigma_u^2 + \gamma_{t-1})], \tag{4.7}$$

$$\gamma_t = \sigma_\epsilon^2 - (y_{t-1}^2)/[\sigma_u^2 + y_{t-1}^2(\sigma_u^2 + \gamma_{t-1})] \tag{4.8}$$

with the initial values $\gamma_0 = \sigma_\epsilon^2$ and $\mu_0 = 0$. Suppose that y_m is missing, then μ_t and γ_t can be computed up to $t = m - 1$ using (4.7) and (4.8). It can be shown that the optimal estimate for y_m is the solution of

$$[y_m - (\phi + \mu_{m-1})\gamma_{m-1}]/w_{m-1} - (\phi + \mu_m)[y_{m+1} - (\phi + \mu_m)y_m]/w_m = 0 \tag{4.9}$$

where $w_m = \sigma_u^2 + y_m^2(\sigma_u^2 + \gamma_m)$. Since the estimator depends on the unknown μ_m and γ_m we propose to use the following algorithm to estimate y_m.

Step 1. Use $\hat{y}_m = (\phi + \mu_{m-1})y_{m-1}$ as an initial value for y_m and obtain μ_m and γ_m.

Step 2. Obtain the least square estimate of y_m:

$$\hat{y}_m(\mathrm{LS}) = [(\phi + \mu_{m-1}y_{m-1}) + y_{m+1}]/[1 + \phi + \mu_m]\hat{y}_m.$$

Step 3. Calculate the weight

$$\hat{w}_m = \sigma_u^2 + \hat{y}_m^2(\mathrm{LS})(\sigma_u^2 + \gamma_m)$$

and obtain the optimal estimate from (4.9) by replacing w_m with $\hat{w}_m$.

Limited experience with this algorithm indicates that few iterations are necessary before the final estimate is obtained. As in Example 4.2, we assume that the model parameters are known. These parameters may be estimated using part of the data (see the last part of Example 4.2).

5. Concluding remarks

The occurrences of missing observations is quite common in time series and the generalised model (2.1) may be used to characterise such situations. This paper offers two alternatives for estimating missing observations. The methodology in Section 3 can be applied for a restricted class of models whenever the normality assumption is made on the errors while the optimal estimation method in Section 4 is more general and is useful when a practitioner has doubts about specifying a particular distribution for the errors. As can be seen from Sections 3 and 4, these procedures yield some of the known results as special cases. For example, the well known results of Abraham (1981) and Miller and Ferreiro (1985) for linear time series models become special cases of the results obtained here. It should be noted, however, that the procedures may not cover all the non-linear time series situations and the methods should be adapted to meet particular needs.

Appendix

A.1 *Proof of Theorem* 2.1

Let $\nu_t = y_t - \hat{y}_t$. Then

$$\begin{aligned}\nu_t &= A_{t-1}\theta_t + B_{t-1}v_t - A_{t-1}\hat{\theta}_t \\ &= A_{t-1}(\theta_t - \hat{\theta}_t) + B_{t-1}v_t.\end{aligned}$$

$E[\nu_t\nu_t^T \mid \mathcal{F}_{t-1}^y] = A_{t-1}\Sigma_t A_{t-1}^T + B_{t-1}B_{t-1}^T$, $E[\theta_t\nu_t^T \mid \mathcal{F}_{t-1}^y] = E[(\theta_t - \hat{\theta}_t)\nu_t^T \mid \mathcal{F}_{t-1}^y]$ since $E[\hat{\theta}_t\nu_t^T \mid \mathcal{F}_{t-1}^y] = 0$. Moreover it follows from the definition of Σ_t that

$$\begin{aligned}E[(\theta_t - \hat{\theta}_t)\nu_t^T \mid \mathcal{F}_{t-1}^y] &= E[(\theta_t - \hat{\theta}_t)\{A_{t-1}(\theta_t - \hat{\theta}_t) + B_{t-1}v_t\}^T \mid \mathcal{F}_{t-1}^y] \\ &= \Sigma_t A_{t-1}^T.\end{aligned}$$

Using the fact that the σ-field generated by the observations up to time t, namely $\mathcal{F}_t^y$, is the same as the σ-field generated by ν_t, $\mathcal{F}_t^\nu$, we have

$$\hat{\theta}_{t+1} = E[\theta_{t+1} \mid \mathcal{F}_t^y] = a_t E[\theta_t \mid \mathcal{F}_t^y] = a_t E[\theta_t \mid \mathcal{F}_{t-1}^y, \nu_t].$$

Using the properties of normal random vectors we have

$$a_t E[\theta_t \mid \mathcal{F}_{t-1}^y, \nu_t] = a_t E[\theta_t \mid \mathcal{F}_{t-1}^y] + k_t[y_t - \hat{y}_t]$$

where $k_t = a_t E[\theta_t\nu_t^T \mid \mathcal{F}_{t-1}^y][E[\nu_t\nu_t^T \mid \mathcal{F}_{t-1}^y]]^+$. Hence

$$\hat{\theta}_{t+1} = a_t\hat{\theta}_t + k_t(y_t - \hat{y}_t) \quad \text{and} \quad k_t = a_t\Sigma_t A_{t-1}^T[A_{t-1}\Sigma_t A_{t-1}^T + B_{t-1}B_{t-1}^T]^+.$$

Moreover

$$\begin{aligned}\theta_{t+1} - \hat{\theta}_{t+1} &= a_t\theta_t + b_t u_{t+1} - a_t\hat{\theta}_t - k_t(y_t - \hat{y}_t) \\ &= a_t(\theta_t - \hat{\theta}_t) + b_t u_{t+1} - k_t[A_{t-1}(\theta_t - \hat{\theta}_t) + B_{t-1}\nu_t] \\ &= b_t u_{t+1} + (a_t - k_t A_{t-1})(\theta_t - \hat{\theta}_t) - k_t B_{t-1}\nu_t.\end{aligned}$$

Thus

$$\begin{aligned}\Sigma_{t+1} &= E[(\theta_{t+1} - \hat{\theta}_{t+1})(\theta_{t+1} - \hat{\theta}_{t+1})^T \mid \mathcal{F}_t^y] \\ &= b_t b_t^T + (a_t - k_t A_{t-1})\Sigma_t (a_t - k_t A_{t-1})^T + k_t B_{t-1} B_{t-1}^T k_t^T.\end{aligned}$$

Hence the theorem follows.

A.2 *Proof of Theorem* 2.2

For a fixed t_j, we observe the following

$$\tilde{\theta}_{t_j|t} = E[\theta_{t_j} \mid \mathcal{F}_t^y] = E[\theta_{t_j} \mid \mathcal{F}_{t-1}^y, \nu_t]$$

where $\nu_t = y_t - \hat{y}_t$. Applying again the results from normal theory we have

$$\tilde{\theta}_{t_j|t} = \tilde{\theta}_{t_j|t-1} + \tilde{k}_t \nu_t$$

where

$$\tilde{k}_t = E[\theta_{t_j} \nu_t^T \mid \mathcal{F}_{t-1}^y]\{E[\nu_t \nu_t^T \mid \mathcal{F}_{t-1}]\}^+.$$

Note that the second factor remains the same as the "innovation" variance as in k_t. Hence it can be obtained as in Theorem 2.1. The rest of the proof is similar to that of Theorem 2.1 and hence is omitted.

References

Abraham, B. (1981). Missing observations in time series, *Comm. Statist. A—Theory Methods*, **10**, 1645–1653.

Brockwell, P. J. and Davis, R. A. (1987). *Time Series: Theory and Methods*, Springer, New York.

Broemeling, L. D. (1985). *Bayesian Analysis of Linear Models*, Dekker, New York.

Charbonnier, R., Barlaud, M., Alengrin, G. and Menez, J. (1987). Results on AR-modelling of nonstationary signals, *Signal Process.*, **12**, 143–157.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica*, **50**, 987–1007.

Ferreiro, O. (1987). Methodologies for the estimation of missing observations in time series, *Statist. Probab. Lett.*, **5**, 65–69.

Godambe, V. P. (1985). The foundations of finite sample estimation in stochastic processes, *Biometrika*, **12**, 419–428.

Harrison, P. J. and Stevens, C. F. (1976). Bayesian forecasting (with discussion), *J. Roy. Statist. Soc. Ser. B*, **38**, 205–248.

Jones, R. H. (1985). Time series analysis with unequally spaced data, *Handbook of Statistics*, Vol. 5 (eds. E. J. Hannan, P. R. Krishnaiah and M. M. Rao), 157–177, North Holland, Amsterdam.

Miller, R. B. and Ferreiro, O. (1984). A strategy to complete a time series with missing observations, *Lecture Notes in Statistics*, **25**, 251–275, Springer, New York.

Nicholls, D. F. and Quinn, B. G. (1982). Random coefficient autoregressive models: An introduction, *Lecture Notes in Statistics*, **11**, Springer, New York.

Ozaki, T. (1985). Nonlinear time series models and dynamical systems, *Handbook of Statistics*, Vol. 5 (eds. E. J. Hannan, P. R. Krishnaiah and M. M. Rao), 25–83, North Holland, Amsterdam.

Priestley, M. B. (1980). State-dependent models: A general approach to time series analysis, *J. Time Ser. Anal.*, **1**, 47–71.

Ruskeepaa, H. (1985). Conditionally Gaussian distributions and an application to Kalman filtering with stochastic regressors, *Comm. Statist. A—Theory Methods*, **14**, 2919–2942.

Shiryayev, A. N. (1984). Probability, *Graduate Texts in Math.*, **95**, Springer, New York.

Thavaneswaran, A. and Abraham, B. (1988). Estimation for nonlinear time series models using estimating equations, *J. Time Ser. Anal.*, **9**, 99–108.

Tjøstheim, D. (1986). Estimation in nonlinear time series models, *Stochastic. Process. Appl.*, **21**, 251–273.