Estimation for Nonlinear Time Series Models Using Estimating Equations

by

A. Thavaneswaran
Department of Biostatistics
University of North Carolina at Chapel Hill

and

B. Abraham
Department of Statistics and Actuarial Science
University of Waterloo

Institute of Statistics Mimeo Series No. 1821

March 1987

ESTIMATION FOR NONLINEAR TIME SERIES MODELS USING ESTIMATING EQUATIONS

A. Thavaneswaran and B. Abraham

Department of Statistics and Actuarial Science

University of Waterloo

Abstract. Godambe's (1985) theorem on optimal estimating equation for stochastic processes is applied to nonlinear time series estimation problems. Examples are considered from the usual classes of nonlinear time series models. Recursive estimation procedure based on optimal estimating equation is provided. It is also shown that prefiltered estimates can be used to obtain the optimal estimate from a nonlinear state-space model.

Keywords: Kalman filter, Nonlinear time series, Optimal estimation, Recursive estimates.

Currently with the Department of Biostatistics at the University of North Carolina at Chapel Hill.

Research partly supported by the Office of Naval Research under Contract Number N00014-83-K-0387.

1. INTRODUCTION

There are many examples of random vibrations in the real world, eg. a ship rolling at sea, car vibration on the road, brain-wave records in neurophysiology etc. Recently there has been a growing interest in modeling these events as nonlinear time series models (Ozaki (1985)). In order to use nonlinear time series models in practice one must be able to fit the models to data and estimate the parameters. Computational procedures for determining parameters for various model classes together with the theoretical properties of the resulting estimates are outlined in Tjøstheim (1986) and the references therein.

The theory of estimating equations was originally proposed by Godambe (1960) for i.i.d. observations and recently extended to discrete time stochastic processes by Godambe (1985). The basic ideas of Godambe have been further adapted and generalized for time continuous real valued semimartingales by Thavaneswaran and Thompson (1986). The particular statistical relevance and lucidity of the estimating equation method for statistical models under present study should be appreciated against the background of the fundamental difficulties encountered in likelihood estimation when the variance of the observation error depends on the parameter of interest Godambe, (1985, 3.2).

In this paper we will try to develop a more systematic approach and discuss a general framework for finite sample nonlinear time series estimation. Our approach yields the most recent estimation results for nonlinear time series as special cases, and, infact, we are able to weaken the conditions in the maximum likelihood case. We derive the recursive version of the optimal estimate and apply it to obtain a recursive estimate for a parameter in a state space model.

In section 2 we present Godambe's (1985) theorem with applications in nonlinear time series. Section 3 deals with a recursive version (on line procedure) of the optimal estimate and shows through an example that this procedure can be used to obtain a recursive estimate without making any distributional assumptions on the errors. In section 4 the theory of estimating equations together with the Kalman filter algorithm are applied to obtain an optimal estimate for a parameter which is usually assumed known in the state-space set up.

2. GODAMBE'S THEOREM AND SOME APPLICATIONS

In this section we recall Godambe's (1985) theorem on stochastic processes and apply it to obtain optimal estimates for recently proposed nonlinear time series models.

Let $\{y_t, t \epsilon I\}$ be a discrete time stochastic process taking values in R and defined on a probability space (Ω, A, F). The index set I is the set of all positive integers. We assume that observations $(y_1, y_2, \ldots, y_n)$ are available and that the parameter $\theta \epsilon \Theta$, a compact subset of R. Let $\mathcal{F}$ be a class of distributions and F_t^y be the σ field generated by y upto time t. Following Godambe (1985) we say that any real function g of the variates $y_1, \ldots, y_n$ and the parameter θ, satisfying certain regularity conditions, is called a regular unbiased estimating function if,

$$E_F[\, g(y_1, \ldots, y_n; \theta(F))\,] = 0, \qquad F \epsilon \mathcal{F} .$$

Let L be the class of estimating functions g of the form

$$g = \sum_{t=1}^{n} h_t \, a_{t-1}$$

where the function h_t is such that $E[\, h_t \mid F_{t-1}^y \,] = 0$ ($t = 1, \ldots, n$) and a_{t-1} is a function of $y_1, \ldots, y_{t-1}$ and θ, for $t = 1, \ldots, n$.

Theorem 1. In the class L of unbiased estimating functions g , the optimum estimating function g^* is the one which minimizes

$$E(\, g^2) \,/\, E(\, \frac{\partial g}{\partial \theta}\,)^2$$

and this is given by

$$g^* = \sum_{t=1}^{n} h_t \, a_{t-1}^* \quad \text{where} \quad a_{t-1}^* = [\, E(\, \partial h_t / \partial \theta \mid F_{t-1}^y \,)\,] \,/\, E(\, h_t^2 \mid F_{t-1}^y \,) \qquad (2.1)$$

(see Godambe (1985)). For further motivation of this optimality based on efficiency considerations see Lindsay (1985).

2.1 RANDOM COEFFICIENT AUTOREGRESSIVE (RCA) MODEL

Random coefficient autoregressive models are defined by allowing random additive perturbations of the Autoregressive (AR) coefficients of ordinary AR models. We assume that the random process $\{y_t\}$ is given by

$$y_t - \sum_{i=1}^{p} (\, \theta_i + b_i(t)\,)\, y_{t-i} = e_t \qquad (2.2)$$

where θ_i ; $i = 1,2, \ldots ,p$ are the parameters to be estimated, $\{e_t\}$ and $\{b_i(t)\}$ are zero mean square integrable independent processes and the variances are denoted by σ_e^2 and σ_b^2; $b_i(t)$ ($i = 1,2, \ldots ,p$) are independent of $\{e_t\}$ and $\{y_{t-i}\}$. $b(t)$ may be thought of as incorporating environmental stochasticity. For example weather conditions might make $b(t)$ a random variable having a binomial distribution.

Let $g_i = \sum_{t=1}^{n} h_t \, a_{i,t-1}$ where $h_t = y_t - E[\, y_t \mid F_{t-1}^{y} \,] = y_t - \sum_{i=1}^{p} \theta_i \, y_{t-i}$ be the estimating function for θ_i. Then it follows from Theorem 1 that the optimal estimating function for θ_i is given by

$$g_i^* = \sum_{t=1}^{n} h_t \, a_{i,t-1}^*$$

where $a_{i,t-1}^* = [\, E(\, \partial h_t / \partial \theta_i \mid F_{t-1}^{y} \,)\,] \,/\, E(\, h_t^2 \mid F_{t-1}^{y} \,)$. Now it can be shown that

$$a_{i,t-1}^* = -y_{t-i} \,/\, \{\, \sigma_e^2 + \sum_{i=1}^{p} y_{t-i}^2 \, \sigma_b^2 \,\}$$

and the optimal estimate for $\theta' = (\, \theta_1, \ldots, \theta_p \,)$ can be obtained by solving the equations

$$\sum_{t=1}^{n} h_t \, a_{i,t-1} = 0\,, \quad i = 1, 2, \ldots, p.$$

This leads to

$$\hat{\theta}_n = (\, \sum_{t=p+1}^{n} Y_{t-1} Y_{t-1}' \,/\, W_t \,)^{-1} (\, \sum_{t=p+1}^{n} Y_{t-1} y_t \,/\, W_t \,)$$

where $Y_{t-1}' = (\, y_{t-1}, \ldots, y_{t-p} \,)$ and $W_t = \sigma_e^2 + Y_{t-1}' \, Y_{t-1} \, \sigma_b^2$. In the special case of a model with one parameter θ given by

$$y_t - (\, \theta + b_t \,)\, y_{t-1} = e_t \tag{2.3}$$

we have $a_{t-1}^* = -y_{t-1} \,/\, (\sigma_e^2 + y_{t-1}^2 \, \sigma_b^2)$ and the optimal estimate is given by

$$\hat{\theta}_n = \sum_{t=2}^{n} a_{t-1}^{*} y_t \Big/ \sum_{t=2}^{n} a_{t-1}^{*} y_{t-1} \quad . \tag{2.4}$$

Nicholls and Quinn (1980) obtained least squares estimate of θ_i $i = 1, \ldots, p$ and their estimate is somewhat different from what is given here. For comparison, we consider only the special case with one parameter. In this case the optimal estimate given in (2.4) simplifies down to

$$\hat{\theta}_n = \sum_{t=2}^{n} \frac{y_t \, y_{t-1}}{\sigma_e^2 + \sigma_b^2 \, y_{t-1}^2} \Big/ \sum_{t=2}^{n} \frac{y_{t-1}^2}{\sigma_e^2 + \sigma_b^2 \, y_{t-1}^2} \tag{2.5}$$

while the one given by Nicholls and Quinn (1980), Tjøstheim (1986) is

$$\tilde{\theta}_n = \sum_{t=2}^{n} y_{t-1} \, y_t \Big/ \sum_{t=2}^{n} y_{t-1}^2 \quad .$$

As indicated by Nicholls and Quinn (1980) $\tilde{\theta}_n$ will not be efficient but strongly consistent and asymptotically normally distributed. $\hat{\theta}_n$ also has the consistency and asymptotic normality properties. In addition the optimal estimator $\hat{\theta}_n$ uses a weighting factor, depending upon the variance of b_t , for the numerator and denominator. It should be noted that σ_e^2 and σ_b^2 are not known in practice. However, $\tilde{\theta}_n$ could be used initially to estimate σ_e^2 and σ_b^2 as in Nicholls and Quinn (1980). Then $\hat{\theta}_n$ may be calculated with the estimated values of σ_e^2 and σ_b^2 .

2.2 DOUBLY STOCHASTIC TIME SERIES

Random coefficient autoregressive sequences given in (2.3) are special cases of what Tjøstheim (1986) refers to as doubly stochastic time series models. In the nonlinear case these models are given by

$$y_t - \theta_t \, f(\, t, F^y_{t-1}\,) = e_t \tag{2.6}$$

where $\{\theta + b_t\}$ of (2.3) is now replaced by a more general stochastic sequence $\{\theta_t\}$ and y_{t-1} is replaced by a function of the past, $f(\, t, F^y_{t-1}\,)$. When θ_t is a Moving Average (MA) sequence of the form

$$\theta_t = \theta + \epsilon_t + \epsilon_{t-1} \tag{2.7}$$

where θ_t, e_t are square integrable independent random variables and $\{\epsilon_t\}$ consists of zero mean square integrable random variables independent of $\{e_t\}$. In this case $E(\, y_t \mid F^y_{t-1}\,)$ depends on the posterior mean, $m_t = E(\, \epsilon_t \mid F^y_t\,)$ and variance $\gamma_t = E[(\, \epsilon_t - m_t\,)^2 \mid F^y_t\,]$ of ϵ_t . Thus for the evaluation of m_t and γ_t we further assume that $\{e_t\}$ and $\{\epsilon_t\}$ are Gaussian and that $y_0 = 0$. Then m_t and γ_t satisfy the following Kalman-like recursive algorithms (see Shiryayev (1984) p.439)

$$m_t = \frac{\sigma_\epsilon^2 \, f(t, F^y_{t-1}) \, [y_t - (\theta + m_{t-1}) \, f(t, F^y_{t-1})]}{\sigma_e^2 + f^2(t, F^y_{t-1}) \, (\sigma_\epsilon^2 + \gamma_{t-1})}$$

and

$$\gamma_t = \sigma_\epsilon^2 - \frac{f^2(t, F^y_{t-1}) \, \sigma_\epsilon^4}{\sigma_e^2 + f^2(t, F^y_{t-1}) \, (\sigma_\epsilon^2 + \gamma_{t-1})}$$

where $\gamma_0 = \sigma_\epsilon^2$ and $m_0 = 0$. Hence

$$E(y_t \mid F^y_{t-1}) = (\theta + m_{t-1}) \, f(t, F^y_{t-1})$$

and

$$E(h_t^2 \mid F_t^y) = E\{[y_t - E(y_t \mid F_{t-1}^y)]^2 \mid F_{t-1}^y\}$$

$$= \sigma_\epsilon^2 + f^2(t, F_{t-1}^y)\,(\sigma_e^2 + \gamma_{t-1})$$

can be calculated recursively. Then the optimal estimating function turns out to be

$$g_n^* = \sum_{t=1}^{n} h_t a_{t-1}^*$$

where $a_{t-1}^* = E[(\partial h_t/\partial \theta) \mid F_{t-1}^y] \,/\, E[h_t^2 \mid F_{t-1}^y]$.

Thus the optimal estimate is given by

$$\hat{\theta}_n = \sum_{t=2}^{n} a_{t-1}^* y_t \,/ \sum_{t=2}^{n} a_{t-1}^* f(t, F_{t-1}^y) \tag{2.8}$$

where

$$a_{t-1}^* = f(t, F_{t-1}^y)(1 + (\partial m_{t-1}/\partial \theta)) \,/\, \{\sigma_\epsilon^2 + f^2(t, F_{t-1}^y)\,(\sigma_e^2 + \gamma_{t-1})\} \tag{2.9}$$

Since γ_t is independent of θ, the relation

$$\partial m_t/\partial \theta = -\{\sigma_\epsilon^2 f^2(t, F_{t-1}^y)(1 + \partial m_{t-1}/\partial \theta)\} \,/\, \{\sigma_\epsilon^2 + f^2(t, F_{t-1}^y)\,(\sigma_e^2 + \gamma_{t-1})\}$$

can be used to calculate this derivative recursively.

Conditional least squares approach of Tjøstheim (1986) leads to an estimator

$$\tilde{\theta}_n = \sum_{t=2}^{n} f(t, F_{t-1}^y)\,(1 + (\partial m_{t-1}/\partial \theta))\, y_t \,/ \sum_{t=2}^{n} [\, f(t, F_{t-1}^y)\,(1 + (\partial m_{t-1}/\partial \theta))\,]$$

which does not take into account the variances σ_ϵ^2 and σ_e^2 . However as can be seen from (2.8) and (2.9), the optimal estimate $\hat{\theta}_n$ adopts a weighting scheme based on σ_ϵ^2 and σ_e^2 . In practice these quantities may be obtained using a nonlinear optimization algorithm as indicated later in section 4.

2.3 THRESHOLD AUTOREGRESSIVE PROCESS

Now we consider an application of the theory of estimating equations in the context of the threshold autoregressive model with only one residual process given in Tjøstheim (1986):

$$y_t - \sum_{j=1}^{p} \theta_j \, y_{t-1} H_j(y_{t-1}) = e_t \tag{2.10}$$

where $H_j(y_{t-1}) = I(y_{t-1} \epsilon D_j)$, $I(.)$ being the indicator function and $D_1, D_2, \ldots, D_m$ are disjoint regions of R such that $UD_j = R$. Then we have

$$h_t = y_t - E(y_t \mid F^y_{t-1}) = y_t - \sum_{j=1}^{p} \theta_j \, y_t \, H_j(y_{t-1})$$

and $E(h_t^2 \mid F^y_{t-1}) = E(e_t^2) = \sigma_e^2$. Hence the optimal estimate for θ_j based on the n observations is given by

$$\hat{\theta}_j = \sum_{t=1}^{n} y_t \, y_{t-1} \, H_j(y_{t-1}) \, / \sum_{t=2}^{n} y_{t-1}^2 \, H_j(y_{t-1}) \tag{2.11}$$

which turns out to be the same estimate obtained in Tjostheim (1986). This is because $E(h_t^2 \mid F^y_{t-1})$ is a constant which need not be the case in general.

The nonlinear time series models of this section are related to the Kalman filtering set up, except for the important difference that the parameter θ is replaced by a random process $\{\theta_t\}$ with $E(\theta_t \mid F^y_{t-1})$ satisfying a recursive relation.

3. RECURSIVE ESTIMATION

So far we have considered the parameter estimation based on all the available data. When data come successively in time it is natural to look for a recursive estimate for the parameter involved. Aase (1983) proposed a recursive parameter estimation procedure for a nonlinear time series model based on Kalman filtering. In this section we develop a recursive scheme based on the optimal estimating equation.

We now consider estimating functions of the form

$$g^* = \sum_{t=1}^{n} h_t \, a_{t-1}^*$$

based on n observations. Suppose $h_t = y_t - \theta f(t-1, y)$ where $f(t-1, y) = f(t-1, F_{t-1}^y)$. This choice of h_t covers most of the nonlinear time series models of section 2 except the doubly stochastic case. (In the doubly stochastic case $h_t = y_t - (\theta + m_{t-1}) f(t-1, y)$ where m_{t-1} is a function of θ and this is more difficult to handle).

Then the optimal estimating function can be written as

$$g^* = \sum_{t=1}^{n} a_{t-1}^* \left(y_t - \theta \, f(t-1, y) \right)$$

and the optimal estimate based on the first $t-1$ observations is given by

$$\hat{\theta}_{t-1} = \sum_{s=2}^{t-1} a_{s-1}^* \, y_s \Big/ \sum_{s=2}^{t-1} a_{s-1}^* \, f(s-1, y). \tag{3.1}$$

When the t^{th} observation becomes available, the estimate based on the first t observations is given by

$$\hat{\theta}_t = \sum_{s=2}^{t} a_{s-1}^{*} y_s \Big/ \sum_{s=2}^{t} a_{s-1}^{*} f(s-1, y) \ . \tag{3.2}$$

Then $\hat{\theta}_t - \hat{\theta}_{t-1} = K_t [\sum_{s=2}^{t} a_{s-1}^{*} y_s - \hat{\theta}_{t-1} K_t^{-1}]$

where $\quad K_t^{-1} = \sum_{s=2}^{t} a_{s-1}^{*} f(s-1, y)$.

Thus $\quad \hat{\theta}_t - \hat{\theta}_{t-1} = K_t [\sum_{s=2}^{t} a_{s-1}^{*} y_s - K_{t-1} (\sum_{s=2}^{t-1} a_{s-1}^{*} y_s)(K_{t-1}^{-1} + a_{t-1}^{*} f(t-1, y))]$

$$= K_t [a_{t-1}^{*} y_t - \hat{\theta}_{t-1} a_{t-1}^{*} f(t-1, y)] \ .$$

Now it is easy to show that

$$K_t = \frac{K_{t-1}}{1 + f(t-1, y)\, a_{t-1}^{*} K_{t-1}} \tag{3.3}$$

and

$$\hat{\theta}_t = \hat{\theta}_{t-1} + \frac{K_{t-1} a_{t-1}^{*}}{1 + f(t-1, y)\, a_{t-1}^{*} K_{t-1}} [y_t - \hat{\theta}_{t-1} f(t-1, y)] \tag{3.4}$$

The algorithm in (3.4) gives the new estimate at time t as the old estimate at time $t-1$ plus an adjustment. This adjustment is based on the prediction error $y_t - E(y_t \mid F_{t-1}^{y})$, since the term $\hat{\theta}_{t-1} f(t-1, y) = E(y_t \mid F_{t-1}^{y})$ can be considered as an estimated forecast of y_t given F_{t-1}^{y}. Given starting values θ_0 and K_0 we can compute the estimator recursively using (3.3) and (3.4). The recursive estimate $\hat{\theta}_t$ in (3.4) is usually referred to as an "on-line" estimate and it is very appealing computationally, especially when data are gathered sequentially. θ_0 and K_0 can usually be obtained from an initial stretch of data. It is of interest to note that the recursive estimate obtained here is derived from the optimal estimating equation and it does not depend on Aase's (1983) proposal. This algorithm may be interpreted in the Bayesian framework by considering the following state space form

$$y_t = \theta_t \, f(\, t, F_{t-1}^y \,) + h_t \quad ,$$

$$\theta_t = \theta$$

and assuming that h_t and θ are independently normally distributed. Then the algorithm obtained here is the same as the nonlinear version of the Kalman filter. In particular if $f(t, F_{t-1}^y) = y_{t-1}$ then this algorithm is the same as the usual Kalman filter. It should be noted that we have not made any distributional assumptions for h_t or θ to obtain the recursive algorithms (3.3) and (3.4).

If we solve the recursive relations (3.3) and (3.4), using initial values θ_0 and K_0 we obtain an expression for $\hat{\theta}_t$, the "off-line" version:

$$\hat{\theta}_t = \frac{\theta_0 \, K_0^{-1} + \sum_{s=2}^{t-1} a_{s-1}^* \, y_s}{K_0^{-1} + \sum_{s=2}^{t-1} a_{s-1}^* f(\, s-1, \, y \,)} \, . \tag{3.5}$$

This version will sometimes be better studied for certain theoretical investigations. It should be mentioned that until recently very little was known about the theoretical properties of these procedures and the corresponding estimates except the results for the random coefficient model (see Nicholls and Quinn (1982)). Tjøstheim (1984a,b) considered a wide class of nonlinear time series models and developed a systematic asymptotic theory. The asymptotic properties such as consistency and normality of $\hat{\theta}_t$ can be proved as in Tjøstheim (1986) or Thavaneswaran and Thompson (1986) subject to certain regularity conditions.

3.1 AASE'S MODEL (1983)

For recursive estimation in the case of a nonlinear time series model Aase (1983) considered a model of the form

$$y_t = g(\, t-1,\, F^y_{t-1}\,) + \theta\ f(\, t-1,\, F^y_{t-1}\,) + \sigma\,(\, t-1,\, F^y_{t-1}\,)\, e_t$$

where g, f and σ are general (nonlinear) real valued functions of the process y, whose conditional distribution is determined by the values taken during the period $[0, t-1]$. The parameter θ is to be estimated and it is assumed that the error sequence e_t satisfies

$$E(\, e_t \mid F^y_{t-1}\,) = 0$$

$$E(\, e_t^2 \mid F^y_{t-1}\,) = 1 \quad \text{for} \quad t = 0, 1, 2, \ldots\ .$$

Under this set up Aase (1983) proposed a recursive scheme based on the model reference adaptive system (MRAS) approach (see Landau (1976)). Here based on the theory of estimating equations we arrive at the same scheme. In fact it can be shown that by taking

$$h_t = y_t - E(y_t \mid F^y_{t-1}) = y_t - \theta f(t-1, y) - g(t-1, y)$$

$$a^*_{t-1} = f(\, t-1, y\,) \,/\, \sigma^2(\, t-1, y\,)\ ,$$

$$K_t = K_{t-1} - \frac{(\, K_{t-1} f(\, t-1, y\,)\,)^2}{\sigma^2(\, t-1, y\,) + f^2(\, t-1, y\,)\, K_{t-1}}$$

and

$$\hat{\theta}_t = \hat{\theta}_{t-1} + \frac{K_{t-1}\, f(\, t-1,\, y\,)}{\sigma^2(\, t-1,\, y\,) + f^2(\, t-1,\, y)\, K_{t-1}} [\, y_t - g(\, t-1,\, y\,) - \hat{\theta}_{t-1}\, f(\, t-1,\, y\,)\,]$$

This is the same recursive estimate for θ motivated some what differently in Aase (1983). If we assume that the e_t's are normally distributed then the above estimate becomes the maximum likelihood estimate and the algorithm may be interpreted as the Newton-Raphson algorithm.

4. PREFILTERED OPTIMAL ESTIMATION

Sections 2 and 3 dealt with optimal estimation and recursive estimation for adaptive systems. Now we consider optimal parameter estimation for a nonadaptive system. This system is governed by a state space model in which the states (random) as well as an additional parameter (deterministic) are unknown. To illustrate the problem we focus on a state-space representation of a particular autoregressive moving average (ARMA) process of order (1,1) considered by Jones (1985). This may be given as follows.

$$x_t = \theta\, x_{t-1} + \sigma u_t$$

$$y_t = x_t + v_t$$

where $\{u_t\}$, $\{v_t\}$ are zero mean square integrable independent normal sequences having unit variance. Often a major problem in a state-space model is to estimate x_t. This problem can be set up in two stages. In the first stage we have to solve a filtering problem in which $m_t = E[\, x_t \mid F_t^y\,]$ is to be determined. Then m_t is the optimal (in the mean squared error sense) estimate given the observations up to time t and given the value of the deterministic parameter θ. In the second stage we look for an optimal estimate for θ and a corresponding estimate for x_t.

For fixed θ, Kalman filter algorithms can be used as follows.

(1) Calculate a one step prediction:

$E[\, x_t \mid F^y_{t-1} \,] = \theta\, m_{t-1}$ where $m_t = E[\, x_t \mid F^y_t \,]$.

(2) Calculate its variance: $P(\, t \mid F^y_{t-1} \,) = P(\, t-1 \mid F^y_{t-1} \,)\, \theta^2 + \sigma^2$.

(3) The prediction of the next observation is $E[\, y_t \mid F^y_{t-1} \,] = E[\, x_t \mid F^y_{t-1} \,]$

(4) Calculate the innovation: $I_t = y_t - E[\, y_t \mid F^y_{t-1} \,]$

(5) The innovation variance is $V_t = P(\, t \mid F^y_{t-1} \,) + 1$

(6) The Kalman gain is $K_t = \dfrac{P(\, t \mid F^y_{t-1} \,)}{V_t}$

(7) Update the estimate of the state: $m_t = \theta\, m_{t-1} + K_t\, I_t$

(8) Update its variance: $p_t = P(\, t \mid t \,) = P(\, t \mid t-1 \,) - K_t\, P(\, t \mid t-1 \,)$

For given θ the steps (1) to (8) can be used to obtain the estimate m_{t+1} and p_{t+1} from m_t and p_t starting from the initial values m_0 and p_0. Now we consider optimal estimation of θ focussing on the equation given in step (7). If we take $g = \sum_{t=2}^{n} a_{t-1} h_t$ with $h_t = K_t I_t$, then $E(h_t \mid F^y_{t-1}) = 0$ and $E(h_t^2 \mid F^y_{t-1}) = K_t^2 V_t = W_t$. Note that V_t depends on θ. Thus this corresponds to Godambe's (1985) set up in which $E(\, h_t^2 \mid F^y_{t-1} \,)$ depends on θ and hence the superiority of this approach can be argued as in Godambe (1985) . Using Theorem 1 it is easy to show that the optimal estimating function is given by

$$g^{*} = \sum_{t=2}^{n} a_{t-1}^{*} h_t$$

where $a_{t-1}^{*} = m_{t-1} / W_t$. Then the optimal estimate of θ satisfies the equation

$$\sum_{t=2}^{n} m_t \, m_{t-1} / W_t = \hat{\theta} \sum_{t=2}^{n} m_{t-1}^{2} / W_t \; . \tag{4.1}$$

This estimate may be obtained numerically. For example, we can calculate the difference between the L.H.S. and R.H.S. of (4.1) for a given value of θ and search for a θ which will make this difference as small as we please. Same steps can be followed to obtain optimal estimate of a parameter in a nonlinear state-space model having normal errors. The same procedure can be used to obtain the optimal estimate in a more general nonlinear state space model in which the state and observation vectors are nonlinear time series models as in Shiryayev (1984). Such extensions will be treated in a subsequent paper.

5. SUMMARY AND CONCLUSIONS

In this paper we have applied the theory of estimating equations to obtain optimal estimates for a number of nonlinear time series models of adaptive as well as non adaptive nature. In particular we considered estimation for RCA and Threshold AR models. We also discussed the recursive version of the optimal estimate which led to an algorithm similar to the Kalman filter algorithm. Also we looked at parameter estimation for a state-space system which cannot be observed directly.

ACKNOWLEDGEMENTS

We would like to thank an anonymous referee for making some valuable comments on an earlier version of this paper. B. Abraham was supported in part by a grant from the National Sciences and Engineering Research Council of Canada.

REFERENCES

AASE, K.K. (1983). Recursive estimation in nonlinear time series models of autoregressive type. *J. Roy. Stat. Soc. Ser. B* 45, 228-237.

ABRAHAM, B. & LEDOLTER, J. (1983). *Statistical Methods for Forecasting.* New York, Wiley.

GODAMBE, V.P. (1960). An optimum property of regular maximum likelihood equation. *Ann. Math. Stat.*, 31, 1208-11.

GODAMBE, V.P. (1985). The foundations of finite sample estimation in stochastic processes. *Biometrika*, 72, 419-428.

JONES, R.H. (1985). Time series analysis with unequally spaced data. *Handbook of Statistics*, Vol. 5 , 157-177, North Holland.

LANDAU, I.D. (1976). Unbiased recursive identification using model reference adaptive techniques. *IEEE Trans. Automatic Control* AC-21, 194-202.

LINDSAY, B.G. (1985). Using empirical partially Bayes inference for increased efficiency. *Annals of Statistics*, 13(3), 914-931.

NICHOLLS, D.F. & QUINN, B.G. (1980). The estimation of autoregressive models with random coefficients I. *J. Time Series Anal.*, 1, 37-46.

NICHOLLS, D.F. & 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, North Holland.

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

THAVANESWARAN, A. and THOMPSON, M.E. (1986). Optimal estimation for semimartingales. *J. Appl. Prob.*,23, 409-417.

TJOSTHEIM, D. (1984a). Estimation in nonlinear time series I: Stationary series, TR#70, *Centre for Stochastic Processes, Dept. of Statistics, University of North Carolina, Chapel Hill, NC27514.*

TJOSTHEIM, D. (1984b). Estimation in nonlinear time series II: Some nonstationary series, TR#71, *Centre for Stochastic Processes, Dept. of Statistics, University of North Carolina, Chapel Hill, NC27514.*

TJOSTHEIM, D. (1986). Estimation in nonlinear time series models. *Stoch. Proc. Appl.*, 21, 251-273.

TONG, H. (1983). Threshold models in nonlinear time series analysis, *Lecture Notes in Statistics*, 21, Springer, New York.