RECURSIVE ESTIMATION OF NON-LINEAR TIME SERIES MODELS

by

A. Thavaneswaran and B. Abraham

Department of Statistics
and Actuarial Science
University of Waterloo
Waterloo, Ontario

Institute of Statistics Mimeo Series No. 1835

August 1987

Recursive Estimation of Non-linear Time Series Models

A. Thavaneswaran and B. Abraham

Department of Statistics
and Actuarial Science
University of Waterloo
Waterloo, Ontario

ABSTRACT

Recursive estimation procedure based on optimal estimating function is derived. Applications to the case of missing observations and time-varying parameters are also given for nonlinear time series models. A heteroscedastic model with lagged dependent variables is treated as a special case. A recursive scheme for simultaneous optimal estimation of conditional mean and variance in a nonlinear ARCH (autoregressive conditional heteroscedastic) model is also proposed.

Keywords: Heteroscedasticity, Kalman filter, Missing observations, 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

Parametric analysis and modelling of signals using linear stationary time series models have found applications in a variety of contexts including speech and seismic signal processing, spectral estimation, medical imaging, process control and others. The statistical inference of such linear models is well established and efficient computational algorithms exist in a number of cases. Recently there has been a growing interest in inference for nonlinear time series models. Computational procedures for determining parameter estimates for such models together with their theoretical properties have been investigated in Tjøstheim (1986), Thavanesawaran and Abraham (1987).

In this paper we will develop a more systematic approach and discuss a general framework for optimal finite sample nonlinear time series estimation. It can be shown that the most recent estimation results for nonlinear time series (see for example Charbonnier et al. (1987)) are special cases of the method presented here. This approach also allows us to deal with Nicholls and Pagan's (1983) heteroscedasticity model with lagged dependent variables. We obtain a recursive version of the optimal estimate and show that it can be applied to handle time-varying parameters as well as missing observations. Recursive approach to time-varying parameters allows computational advantages in obtaining the estimates and also allows us to identify any structural change in the time-varying parameter.

One of the unfortunate facts facing most data analysts is missing data. Data that are known to have been observed erroneously can farily safely be categorized as missing. Erroneous data can also wreak havoc with the estimation and forecasting of time series models. Abraham (1981) proposed a procedure to interpolate the adjacent missing values on the basis of the unknown segments of an ARIMA (p,d,r) (autoregressive integrated moving average) model. Recently Jones (1985) proposed a state space Kalman filter approach to handle unequally spaced data in the linear case. In Section 3 we indicate how our recursive scheme can be adapted to handle missing observations for a nonlinear multiparameter model.

One of the main objectives in fitting nonlinear time series models is to obtain some idea of the nonlinear structure (trend) of the process without first committing oneself to a particular form of a nonlinear model. This may be achieved by introducing time-varying parameters. In addition time-varying parameters can be used to characerise nonstationary processes; see for example Priestly (1980), Charbonnier et al. (1987). In Section 4 we discuss the problem of estimating time-varying parameters in nonlinear time series models.

2. Godambe's Theorem and Applications

In this section we recall Godambe's (1985) theorem on stochastic processes and apply its extended version (c.f. Bhapkar (1972) for extended version in the i.i.d. case) to obtain optimal estimates for a multivariate multiparameter nonlinear time series model. Let $\{y_t: t \epsilon I\}$ be a discrete time stochastic process taking values in R^q and defined on a probability space (Ω, A, F). The index set I is the set of all positive integers. We assume that observations $(y_1, \cdots, y_n)$ are available and that the parameter $\theta \epsilon \Theta$, a compact subset of R^k. Let $F \epsilon F$ be a class of distributions and F_t^y be the σ- field generated by y up to time t. Following Godambe (1985) we say that any R^k valued function g of the variates $y_1, \cdots, y_n$ and the parameter θ, satisfying certain regularity conditions, is a regular unbiased estimating function if,

$$E_F[g(y_1, \ldots, y_n \; ; \theta(F))] = 0 \; , F \in F \; .$$

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

$$g = \sum_{t=2}^{n} a_{t-1} h_t \tag{2.1}$$

where the R^q valued functions h_t are such that $E[h_t \mid F_{t-1}^y] = 0$, $(t=1, \ldots, n)$ and a_{t-1} is a function of $y_1, \cdots, y_{t-1}$ and θ, for $t=1, \ldots, n$. Let $b = [E(\frac{\partial g}{\partial \theta})]^+ g$, $b^o = [E(\frac{\partial g^o}{\partial \theta})]^+ g^o$, where Z^+ denotes the pseudoinverse of Z and $A_b = E[bb^T]$ and $A_{b^o} = E[b^o b^{oT}]$ be the corresponding covariance matrices.

Definition 2.1: In the class L of unbiased estimating functions g, g^o is said to be the optimum estimating function if $Q = A_b - A_{b^o}$, is a non-negative definite matrix.

The following theorem gives the extended version of Godambe's Theorem (1985) to the multiparameter multi-valued stochastic process.

Theorem 2.1: The optimal estimating function in L is given by

$$g^o = \sum_{t=2}^{n} a_{t-1}^0 h_t$$

where

$$a_{t-1}^o = E[\frac{\partial h_t}{\partial \theta} \mid F_{t-1}^y](E[h_t h_t^T \mid F_{t-1}^y])^+ \tag{2.2}$$

provided the inverse $E[h_t h_t^T \mid F_{t-1}^y]^+$ exists.

Proof: (See Appendix.)

2.1 Heteroscedasticity model with lagged dependent variables.

White (1980) considered a general specification of a regression model by allowing the exogenous regressors to be stochastic. The important case when some of the exogenous variables are lagged values of the endogeneous variables have been treated in Nicholls and Pagan (1983).

They considered the model

$$y_t = \sum_{j=1}^{p} \theta_j y_{t-j} + e_t \tag{2.3}$$

where the error sequence $\{e_t\}$ satisfies

$$E(e_t \,|F_{t-1}^y) = 0 \quad , \quad E(e_t^2 \,|F_{t-1}^y) = \sigma_t^2 \quad .$$

Let $h_t = y_t - E[y_t \mid F_{t-1}^y] = y_t - \theta^T X_{t-1}$, and $E(h_t^2 \,|F_{t-1}^y) = \sigma_t^2$

where $\theta^T = (\theta_1, \theta_2, \cdots, \theta_p)$, $X_{t-1}^T = (y_{t-1}, \ldots, y_{t-p})$ and θ^T denotes the transpose of θ. Then the optimal estimating function $g_{n,\theta}^0$ is given by

$$g_{n,\theta}^0 = \sum_{t=p+1}^{n} \sigma_t^{-2} X_{t-1}(y_t - \theta^T X_{t-1}) \quad ,$$

and the optimal estimate is given by

$$\hat{\theta}_n = (\sum_{t=p+1}^{n} X_{t-1}X_{t-1}^T/\sigma_t^2)^{-1}(\sum_{t=p+1}^{n} X_{t-1}y_t/\sigma_t^2) \quad .$$

The least-squares estimate of θ is given by

$$\tilde{\theta}_n = (\sum_{t=p+1}^{n} X_{t-1}X_{t-1}^T)^{-1}(\sum_{t=p+1}^{n} X_{t-1}y_t) \quad ,$$

and is independent of the conditional variance σ_t^2. Moreoever, if e_t is assumed to be normal then $\hat{\theta}_n$ becomes the maximum likelihood estimate (m.l.e.) and is more efficient than $\tilde{\theta}_n$.

3. Recursive Estimation

So far we have considered the parameter estimation based on all the available data. When data come successively in time or the data are unequally spaced due to missing values, errors, etc. it is natural to look for a recursive estimate for the parameters involved. For an ARIMA (autoregressive integrated moving average process), Jones (1985) used a state space representation and Kalman filter to handle missing observations. Our concern here is to obtain a recursive scheme for nonlinear time series to handle missing observations.

Let $h_t = y_t - \theta^T f(t-1,y)$ where $f(t-1,y) = f(t-1,F_{t-1}^y)$ is a $p \times 1$ column vector. This choice of h_t covers most of the nonlinear time series models considered in the literature. For example, taking $f(t-1,y) = X_{t-1}$ and $\theta^T = (\theta_1, \ldots, \theta_p)$ leads to model 2.3.

Then the optimal estimating function can be written as

$$g^0 = \sum_{t=2}^{n} a_{t-1}^o (y_t - \theta^T f(t-1,y))$$

and the optimal estimate based on the first n observations is given by

$$\hat{\theta}_n = \left(\sum_{t=p+1}^{n} a_{t-1}^o f^T(t-1,y) \right)^{-1} \left(\sum_{t=p+1}^{n} a_{t-1}^o y_t \right) \quad .$$

When the $(n+1)^{th}$ observation becomes available, the estimate based on all the observiations is given by

$$\hat{\theta}_{n+1} = \left(\sum_{t=p+1}^{n+1} a_{t-1}^{o} f^{T}(t-1,y)\right)^{-1}\left(\sum_{t=p+1}^{n+1} a_{t-1}^{o} y_t\right) \quad .$$

Now

$$\hat{\theta}_{n+1} - \hat{\theta}_n = K_{n+1}\left[\sum_{t=p+1}^{n+1} a_{t-1}^{o} y_t - K_{n+1}^{-1}\hat{\theta}_n\right]$$

where

$$K_{n+1}^{-1} = \sum_{t=p+1}^{n+1} a_{t-1}^{o} f^{T}(t-1,y).$$

Using the relation

$$K_{n+1}^{-1} = K_n^{-1} + a_n^{o} f^{T}(n,y)$$

we have

$$\hat{\theta}_{n+1} - \hat{\theta}_n = K_{n+1}\left[\sum_{t=p+1}^{n+1} a_{t-1}^{o} y_t - (K_n^{-1} + a_{n-1}^{o} f^{T}(n,y))\hat{\theta}_n\right]$$

$$= K_{n+1}[a_n^{o} y_{n+1} - a_n^{o} f^{T}(n,y)\hat{\theta}_n]$$

$$= K_{n+1} a_n^{o}[y_{n+1} - f^{T}(n,y)\hat{\theta}_n] \quad .$$

Hence we have

$$K_{n+1}^{-1} = K_n^{-1} + a_n^{o} f(n,y) \tag{3.1}$$

$$\hat{\theta}_{n+1} = \hat{\theta}_n + K_{n+1} a_n^{o}[y_{n+1} - f^{T}(n,y)\hat{\theta}_n] \tag{3.2}$$

For the model of Section 2, $y_t = \theta^T X_{t-1} + e_t$

$$K_{n+1}^{-1} = K_n^{-1} + X_n X_n^T / \sigma^2(n,y)$$

$$\hat{\theta}_{n+1} = \hat{\theta}_n + K_{n+1}X_n[y_{n+1} - X_n^T\hat{\theta}_n]/\sigma^2(n,y) \quad .$$

The algorithm in (3.2) gives the new estimate at time $n+1$ as the old estimate at time n plus an adjustment. This adjustment is based on the prediction error $y_{n+1} - E[y_{n+1}|F_n^y]$, since the term $f^T\hat{\theta}_n = E[y_{n+1}|F_n^y]$ can be considered as an estimated forecast of y_{n+1} given F_n^y. Given starting values θ_0 and K_0 one can compute the estimate recursively using (3.1) and (3.2). The recursive estimate in (3.2) 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 any distributional assumption of the error. This algorithm also extends the model reference adaptive system algorithm proposed by Aase (1983) to derive an algorithm for a multiparameter case. The algorithm may be interpreted in the Bayesian setup by considering the following state space form

$$y_t = \theta_t^T f(t-1,y) + h_t$$

$$\theta_t^T = \theta^T$$

and assuming that h_t and θ^T are independently normally distributed. Then the algorithm obtained here is the same as the nonlinear version of the Kalman filter. Hence any missing observations can be handled using (3.1) and (3.2) as in Jones (1985) for the ARIMA model. When an observation is missing in a univariate time series (with equal spacing) steps given by equations (3.1)-(3.2) become simply

$$\hat{\theta}_{n+1} = \hat{\theta}_n \text{ and } K_{n+1}^{-1} = K_n^{-1} \text{ respectively} \quad .$$

These two equations require no calculation since the values that are in memory are not changed. For multivariate time series with missing observations within the observation vector, $f(t-1,F_{t-1}^{y})$, it is only necessary to reduce the number of rows in the θ^T vector to allow for these missing observations. It should be noted that we have not made any distributional assumptions on e_t or on θ^T to obtain the recursive algorithm (3.1) and (3.2). In the case of missing observations it is difficult and in some cases impossible to obtain the likelihood even if we make distributional assumptions on e_t. Since the estimate here is calculated successively, it is also easy to see the effect of a particular observation on the estimate and the influence of it may be studied by estimating the parameters with and without it.

If we solve the recursive relations (3.1) and (3.2) using initial values θ_0 and K_0 we obtain an expression for $\hat{\theta}_n$, the "off-line" version.

$$\hat{\theta}_n = (\theta_0 K_0^{-1} + \sum_{t=p+1}^{n} a_{t-1}^{o} f^T(t-1,y))^{-1}(K_0^{-1} + \sum_{t=p+1}^{n} a_{t-1}^{o} y_t) \quad .$$

As $K_0 \to \infty$,

$$\hat{\theta}_n \to (\sum_{t=p+1}^{n} a_{t-1}^{o} f^T(t-1,y))^{-1}(\sum_{t=p}^{n} a_{t-1}^{o} y_t). \tag{3.3}$$

This is the same as the m.l.e. if e_t has normal distribution. Hence the recursive estimate constitute m.l.e. as a subclass.

The estimate studied in Nicholls and Pagan (1983) is the ordinary least squares estimate and it does not take into account the conditional variance of e_t into consideration. In other words the OLS estimate of θ in the following models

(i) AR model

$$y_t = \theta y_{t-1} + e_t \tag{3.4}$$

with

$$E[y_t | F^y_{t-1}] = \theta y_{t-1} \quad , \quad Var[y_t | F^y_{t-1}] = \sigma^2$$

(ii) and the ARCH (autoregressive conditional heteroscedasticity) model (see Engle (1982))

$$y_t = \theta y_{t-1} + \epsilon_t$$

with

$$E[y_t | f^y_{t-1}] = \theta y_{t-1} \quad , \quad Var(y_t | F^y_{t-1}) = \alpha_0 + \alpha_1 y^2_{t-1}$$

would be the same. But the optimal estimate will be different and it depends on the conditional variance. In practice initial estimate for θ may be obtained by the least squares method and $\sigma^2_{t-1} = \alpha_0 + \alpha_1 y^2_{t-1}$ may be estimated initially, using least square residuals. Then the optimal estimate as well as the recursive estimate can be calculated as if the conditional variances are known. The following example indicates how one can jointly estimate the conditional mean and variance parameters optimally.

Example 3.1. Consider an ARCH regression model

$$y_t = x_t \theta + e_t \quad \text{with} \quad E[y_t | F^{y,x}_{t-1}] = x_t \theta$$

where x_t is an observable exogenous variable or a lagged value of y,

$$\sigma^2_{t-1} = Var[y_t | F^{y,x}_{t-1}] = \alpha_0 + \alpha_1 e^2_{t-1} + \cdots + \alpha_p e^2_{t-p} \quad .$$

Engle (1982) studied the maximum likelihood estimation by assuming that the conditional distribution of e_t is normal. However we can obtain the same estimates as optimal in Godambe's sense without making any distributional assumptions. Godambe (1985) has treated the variance estimation (i.e. estimation of $\alpha = (\alpha_0, \cdots, \alpha_p)$) by assuming that θ is known. Here we propose the following scheme for simultaneous estimation of θ and α.

Step 1. Obtain least squares estimate $\hat{\theta}_{LS}$ of θ.

Step 2. Using residuals $y_t - x_t \hat{\theta}_{LS}$ obtain the estimate of α (see, for example, Nicholls and Quinn (1982)).

Step 3. Using the estimate of α, construct the optimal estimate $\hat{\theta}_{0p}$ for θ.

Step 4. Using the optimal estimate, calculate the residuals $y_t - x_t \hat{\theta}_{0p}$ and re-estimate α optimally.

This four steps procedure will enable us to obtain the estimates of θ and α.

4. Time-Varying Parameters

Until recently, estimation procedures have only been developed for the identification of stationary signals. Unfortunately, the hypothesis of stationarity is unverified by most of the physical signals. In order to extend stationary cases to nonstationary ones we consider, processes with time-varying coefficients. For example in an $AR(p)$ (Autoregressive) model with p time-varying parameters each new observation adds p more parameters to the model. This is analogous to the density estimation problem in the i.i.d. setup in which one tries to estimate the density $p(y)$, $y \in \boldsymbol{R}$. This may be viewed as an estimation problem with infinite dimensional parameter '$p(y)$' based on a finite number of observations $y_1, \ldots, y_n$. In the density estimation problem, the 'series' method has been used to approximate the density with finite dimensional parameters (c.f. Tapia and Thompson (1978)).

A nonlinear nonstationary *AR*-process y_t may be represented by the following relation

$$y_t = f(t, F_{t-1}^y, \Theta_{t-1}) + e_t$$

where

$$E[y_t | F_{t-1}^y] = f(t, F_{t-1}^y, \Theta_{t-1}),$$

$$Var[y_t | F_{t-1}^y] = \sigma^2(y_1, \ldots, y_{t-1}),$$

and $\Theta_{t-1} = (\theta_1(t-1), \ldots, \theta_p(t-p))$ the corresponding time-varing parameter. It is of interest to note that when

$$f(t, F_{t-1}^y, \Theta_{t-1}) = -[\theta_1(t-1)y_{t-1} + \cdots + \theta_p(t-p)y_{t-p}]$$

and e_t has normal distribution with constant variance σ^2, this corresponds to the nonstationary model studied in Charbonnier et al. (1987). In order to estimate Θ_{t-1}, we may use the series expansion method and formulate the problem as a parametric one. If we assume that each of the $\theta_i(t-i)$ are smooth functions of t then we may approximate them by a projection on a subspace Θ_{m+1} of Θ with dimension $m+1$. i.e. each $\theta_i(t-i)$ can be written as

$$\theta_i(t-i) = \sum_{j=0}^{m} \theta_{ij} q_j(t-i), \quad i = 1, \ldots, p$$

where θ_{ij}'s are constant parameters and $\{q_j(t)\}$, $j = 0, \ldots, m$ is a known sequence of functions in t, usually a complete orthogonal system of Θ_{m+1}. Now the model may be written as

$$y_t = f(t, F_{t-1}^y, \theta) + e_t$$

with $\theta = (\theta_{ij};\ i = 1, \ldots, p,\quad j = 0, \ldots, m) \in R^{mp}$, a finite dimensional parameter. Hence we can directly apply the results of Section 2 to obtain the optimal estimates of θ and then the estimates of Θ_{t-1}. Moreover, when $f(t, F_{t-1}^y, \Theta_{t-1})$ is a linear function of the parameters, the recursive scheme of Section 3 can be applied directly to estimate the parameters successively. The recursive approach allows us to detect abrupt changes (structural changes) in the nonlinear model, helps detection of convergence of the estimates and avoids unnecessary computations by stopping the recursion as soon as the convergence is attained.

It should be noted that for the proposed procedure one doesn't have to assume normality, constant variance or linearity of f as was required by Charbonnier et al. (1987).

5. Concluding Remarks

Theory of estimating functions is used to obtain optimal estimates and recursive estimates for nonlinear time series. A recursive scheme is proposed for joint estimation of conditional mean and variance parameters in an ARCH model. Application of the recursive algorithm to handle missing observations and time-varying parameters are also discussed in some detail.

Appendix

Proof of Theorem 2.1

We now present the proof of Theorem 2.1 by verifying a sufficient condition for optimality is to hold, due to M.E. Thompson (c.f. Godambe (1985)).

Let $b = Vg$ and $b^o = V^o g^o$ where $V = [E(\frac{\partial g}{\partial \theta})]^+$ and $V^o = [E(\frac{\partial g^o}{\partial \theta})]^+$ and Z^+ denotes the pseudo inverse of Z. Now let $Q = A_b - A_{b^o} = E(bb^T) - E(b^o b^{oT})$. Then

$$A_{b-b^o} = Cov(b - b^o) = Cov(Vg - V^o g^o)$$

$$= VA_g V^T + V^o A_{go} V^{oT} - 2Cov(Vg, V^o g^o)$$

where $A_g = Cov(g)$. Hence

$$A_{b-b^o} = A_b + A_{b^o} - 2VA_{gg^o}V^{oT}$$

$A_{b-b^o} = Q$ if $A_{b^o} = VA_{gg^o}V^{oT}$ where $A_{gg^o} = E(gg^o)$.

This is possible only if $E(\frac{\partial g}{\partial \theta}) = A_{gg^o} \cdot K \ \forall \ g \in L$ where K is a constant matrix.

Thus Q is non negative definite if

$$E(\frac{\partial g}{\partial \theta}) = A_{gg^o} K \qquad \forall g \in L$$

Acknowledgements

B. Abraham was supported in part by a grant from the Natural and Engineering Sciences 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. (1981). Missing observations in time series. *Commun. Statist. Theor. Math.* A10(16), 1645-1653.

Abraham, B. and Ledolter, J. (1983). Statistical Methods for Forecasting. New York, Wiley.

Bhapkar, V.P. (1972). On a measure of efficiency in an estimating equation. *Sankhya* A34, 467-472.

Charbonnier, R., Barlaud, M., Alengrin, G. and Menez, J. (1987). Results on *AR*-modelling of nonstationary signals. *Signal Processing* 12, 143-151.

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

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, 414-428.

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

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

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

Nicholls, D.F. and Pagan, A.R. (1983). Heteroscedasticity in models with lagged dependent variables. *Econometrica,* 51, 1233-1242.

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

Tapia, R.A. and Thompson, J.R. (1978). Nonparametric Probability Density Estimation. The John Hopkins University Press, Baltimore.

Thavaneswaran, A. and Abraham, B. (1987). Estimation for nonlinear time series models using estimating equations. To appear in J. of Time Series Analysis.

Thavaneswaran, A. and Thompson, M.E. (1986). Optimal estimation for semimartingales. *Jour. Appl. Prob.* 23, 409-417.

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

White, H. (1980). A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity", *Econometrica,* 48, 817-838.