Institute of Mathematical Statistics

LECTURE NOTES — MONOGRAPH SERIES

ON THE PREDICTION FOR SOME NONLINEAR TIME SERIES MODELS USING ESTIMATING FUNCTIONS

by

Bovas Abraham
University of Waterloo, Canada
A. Thavaneswaran
University of Manitoba, Canada
and
Shelton Peiris
University of Sydney, Australia

Abstract

Godambe's (1960, 1985) theorems on optimal estimating equations are applied to some non-linear, non-Gaussian time series prediction problems. (Examples are considered from the usual class of time series models.) Recently many researchers in applied time series analysis attracted the information and valid analysis provided by the estimating equation approach. Therefore this article places an interest of estimating equation (EE) prediction theory and building a link between it and the well-known minimum mean square error (MMSE) prediction methodology. Superiority of this EE prediction method over the MMSE is investigated. In particular a random coefficient autoregressive model is discussed in some detail using these EE and MMSE theories.

Keywords: Non-Gaussian models, non-linear time series; optimal estimation; optimal prediction, random coefficient autoregressive, minimum mean square.

1 Introduction

There are many examples of random vibrations in the real world. For example a ship rolling at a sea, car vibration on the road, brain-wave records in neurophysiology, and so on. Recently, there has been a growing interest in modelling these events as non-linear time series. See for instance Tjøstheim (1986), Abraham and Thavaneswaran (1991). In order to use non-linear time series models in practice, one must be able to fit models to data, estimate the respective parameters and obtain valid predictors. Computational procedures for determining parameters for various model classes, together with the theoretical properties of the resulting estimates are outlined in Tjøstheim (1986), Thavaneswaran and Abraham (1988).

The theory of generalized estimation equation (GEE) was originally proposed by Godambe (1960) for identically distributed independent observations and recently extended to discrete time stochastic processes (see Godambe (1985)). The particular statistical relevance and lucidity of the GEE prediction method for statistical models under present study should be appreciated against the fundamental difficulties encountered in

(a) likelihood prediction when the variance of the observation error depends on the parameter of interest (Godambe, 1985, §3.2) and

(b) when the variance of the observation error becomes infinity and MMSE method does not apply.

A special class of nonlinear models called Random Coefficient Autoregressive (RCA) models play an important role in the modern era of time series analysis.

The class of RCA models are defined by allowing random additive perturbations of the autoregressive (AR) coefficients of ordinary AR models. That is we assume that the process $\{Z_t\}$ is given by

$$Z_t - \sum_{i=1}^{p} (\phi_i + b_i(t)) Z_{t-i} = e_t, \tag{1.1}$$

where $\phi_i, i = 2, \cdots, p,$ are the parameters assumed to be known, $\{e_t\}$ and $\{b_i(t)\}$ are zero mean square integrable independent processes and the variances are denoted by σ_e^2 and σ_b^2; $b_i(t)(i = 1, 2, \cdots, p)$ are independent of $\{e_t\}$ and $\{Z_{t-i}\}$.

$\{b_i(t)\}$ may be thought of as incorporating environmental stochasticity. For example, weather conditions might make $\{b_i(t)\}$ random variables having binomial distribution.

For RCA models, the superiority of optimal estimate had been demonstrated in Thavaneswaran and Abraham (1988) and the superiority of the interpolation had been given in Abraham and Thavaneswaran (1991). Godamble (1994) had briefly looked at the prediction problem in the Bayesian context assuming the future values as random. Naik-Nimbalkar and Rajarshi (1995) have used the estimating function method to study the smoothing and filtering problem in the Bayesian point of view.

In this paper, we shall attempt to develop a more systematic approach and discuss a general framework for finite sample non-linear time series prediction. Our approach yields the most recent forecasting results as special cases and, in fact, we are able to improve the efficiency of the predicting equations.

This approach of using estimating function ideas to study the prediction problem is very similar to the one used to study the smoothing problem as in Thavaneswaran and Peiris (1996).

In section 2, we present a theorem on optimal forecasting for discrete time stochastic processes based on estimating functions with applications in non-linear, non-Gaussian time series models. Section 3 deals with quasi-likelihood non-linear estimating functions and compares the efficiency for MMSE and optimal predictors.

2 A theorem on optimal prediction

Let $\{Z_t : t \in I\}$ be a discrete-time stochastic process taking values in $\mathcal{R}$ and defined on a probability space (Ω, A, F). The index set I is the set of all positive integers. We assume that the observations $(Z_1, Z_2, \cdots, Z_n)$ are available and that $\hat{Z}_t(1) = E[Z_{t+1}|F_t^z]$ is a function of F_t^z, the σ-field generated by $Z_1, \cdots, Z_t$. Then the following theorem gives the form of the optimal one step ahead forecast of Z_{n+1} based on observed values $Z_1, \cdots, Z_n$. Let $Z_{n+1} - \hat{Z}_n(1) = a_{n+1}$, where $\{a_1, \cdots a_t, \cdots a_{n+1}\}$ is an iid sequence with probability density function $f(\ )$. Assume that $f(\ )$ is known and that $E\left[-\frac{d^2}{da^2} f(a)\right] < \infty$ and $\int_{-\infty}^{\infty} f(a)\, da$ is twice differentiable under the integral sign.

Let G be the class of unbiased estimating functions $g(a_{n+1})$ such that $E\ g(a_{n+1}) = 0$.

Consider the following theorem.

Theorem 2.1 : In the class G, the optimal predictor of a_{n+1}, which minimizes $\dfrac{\text{Var } g(a_{n+1})}{(E\frac{dg}{da})^2}$ is given by

(i) $m_a(f)$, where $m_a(f)$ is the mode of f ,

(ii) the optimal predictor of Z_{n+1} is given by $Z_n^{\text{opt.}}(1) = \hat{Z}_n(1) + m_a(f)$,

(iii) the efficiency of the optimal estimating function, g°, $Eff(g^\circ)$ for a_{n+1} is

$$\frac{\{E(\frac{dg^\circ}{da})\}^2}{E(g^{\circ 2})} = E(g^{\circ 2}). \tag{2.1}$$

Proof:

Parts (i) and (ii) of the theorem follow by observing that $g^\circ = \frac{d}{da} log\ f(a)$ is an unbiased estimating function in the class G and using the Cauchy-Schwarz inequality for unbiased estimating functions as in Godambe (1960).

It is easy to show that $E[g^{\circ 2}] = E(-\frac{dg}{da})$ and hence part (iii) follows.

Note: If $g =$ identify, $\frac{Var\ g(a_{t+1})}{(E\frac{dg}{da})^2} = \sigma_a^2$. Thus the minimum mean square error forecast is a special case of the theorem.

If $g = \frac{d \log f(a)}{da}$, then $\frac{\text{Var } g(a_{t+1})}{(E\frac{dg}{da})^2} = \frac{1}{\text{Var}(\frac{d \log f}{da})} = \frac{-1}{E\frac{d^2 \log f}{da^2}}$.

Hence the maximum likelihood predictor is also a special case of the theorem.

It is of interest to note that when the distribution of $\{a_t\}$ is stable such as Cauchy, the MMSE predictor cannot be defined but the MLE could be defined and it has finite information.

Example 2.1: Consider a stationary time series having moving average (linear filter) representation

$$Z_t = \Psi(B)a_t = a_t + \Psi_1 a_{t-1} + \Psi_2 a_{t-2} + \cdots \tag{2.2}$$

where $\{a_t\}$'s are independent mean zero with probability density function f. Existence in mean square requires that $\{a_t\}$ have finite variance σ_a^2 and $\sum_{j=1}^{\infty} \Psi_j^2 < \infty$.

Let $(Z_1, \cdots, Z_n)$ be n observations from a series $\{Z_t\}$. Then the optimal one-step ahead forecast is given by MMSE forecast, $\hat{Z}_n(1)$, plus the mode of f

i.e. $Z_n^{\text{opt.}}(1) = \hat{Z}_n(1) + m_a(f) = E[Z_{n+1}|F_n^z] + m_a(f)$.

For an AR(1) model of the form $Z_t = \phi Z_{t-1} + a_t$ with $|\phi| < 1$, the optimal forecast of Z_{n+1} based on observed values $Z_1, \cdots, Z_n$ is $Z_n^{\text{opt.}}(1) = \hat{Z}_n(1) + m_a(f)$, where $m_a(f)$ is the mode of the probability density of $\{a_t\}$.

If a_t's are i.i.d. $N(0, \sigma_a^2)$ then $m_a(f) = 0$ and $Z_n^{\text{opt.}}(1) = \hat{Z}_n(1) =$ MMSE forecast.

Now suppose that f corresponds to a double exponential distribution, with the density $f(x) = \frac{1}{2}e^{-|x|}$, $-\infty < x < \infty$.
In this case, with

$$g = \frac{d \log f(a)}{da}, \text{ one has } \text{Eff.}(g^\circ(a)) = \text{Var}\left[\frac{d\log f(a)}{da}\right] = 1. \tag{2.3}$$

On the other hand,

$$\sigma_a^2 = 1/2 \int_{-\infty}^{+\infty} x^2 \, e^{-|x|} \, dx = \int_0^\infty x^2 e^{-x} dx = 2.$$

That is

$$\text{Eff.}(g_{\text{MMSE}}(a)) = 2. \tag{2.4}$$

Thus, Eff.(g_{MMSE}) is twice as large as Eff.(g°) and, therefore, the MMSE forecast $\hat{Z}_n(1)$ of Z_{n+1} in (2.1) entails about 50% loss of efficiency in this case.

Example 2.2: Consider an ARCH (autoregressive conditionally heterocedastic) model of the form

$$Z_t = \phi \; Z_{t-1} + Z_{t-1}^2 \; a_t,$$

where $\{a_t\}$ is an iid sequence having pdf $f(a)$ and variance σ_a^2. It can be easily shown that the optimal predictor of Z_{n+1} based on observed values $Z_1, \cdots, Z_n$ is given by

$$\hat{Z}_n(1) = \phi Z_n + Z_n^2 \; m_a(f).$$

Similarly the two steps ahead forecast is

$$\hat{Z}_n(2) = \phi^2 Z_n + \hat{Z}_n^2(1) \; m_a(f)$$

and the ℓ-steps ahead forecast is given by

$$\hat{Z}_n(\ell) = \phi^\ell Z_n + \hat{Z}_n^2(\ell - 1) m_a(f).$$

Now we consider a more general situation in the next section.

3 Non-linear non-Gaussian models

Theorem 2.1 of this paper gives the optimal predictor when $\{a_t\}$ is an i.i.d. sequence with known p.d.f. $f(a)$. In the case when $\{a_t\}$ is not an identically distributed independent sequence and when the first two conditional moments are specified the following theorem gives the form of the optimal predictor.

Let $Z_1, \cdots, Z_n$ be n observations from a series having first two conditional moments

$$E[Z_t|F_{t-1}^z] = f(\theta,\ F_{t-1}^z)$$

and

$$\text{Var}[Z_t|F_{t-1}^z] = \sigma^2(\theta,\ F_{t-1}^z).$$

Let $h_{1t} = Z_{t+1} - E[Z_{t+1}|F_t^z]$
and
let $h_{2t} = Z_{t+1}^2 - E[Z_{t+1}^2|F_t^z]$.

The following theorem reports the MMSE and otpimal predictors of Z_{n+1} .

Note: The choice of elementary estimating functions is subjective.

Theorem 3.1: (a) The MMSE predictor of Z_{n+1} is given by

$$\hat{Z}_n(1) = E[Z_{n+1}|F_n^z]$$

and (b) The optimal predictor of Z_{n+1} is given by

$$Z_n^{*^2}(1) = E[Z_{n+1}^2|F_n^z].$$

The proof of this theorem follows by taking the elementary estimating function $h_{2n} = Z_{n+1}^2 - E[Z_{n+1}^2|F_n^z]$ (cf. Godambe (1985)). The following theorem reports the MMSE and optimal predictors of Z_{n+1} .

Example 3.1: Consider the Random Coefficient Autoregressive (RCA) model given in (1.1).

By considering a class of estimating functions of the form $g_n = \sum_{t=2}^n a_{t-1}\, h_t$, where

$$h_t = Z_t - E[Z_t|F_{t-1}^z] = Z_t - \sum_{i=1}^p \phi_i Z_{t-i}\ , \tag{3.1}$$

optimal estimates for the model parameters were obtained in Thavaneswaran and Abraham (1988) and the superiority of the optimal estimate over the least squares had been discussed.

Here if we restrict ourselves to a class of estimating functions of the above form then we will get the forecast of the future value of Z_{n+1} based on the observed values $Z_1, Z_2, \cdots, Z_n$ as $\hat{Z}_n(1) = E\left[Z_{n+1}|Z_n, Z_{n-1}, \cdots,\right]$.

That is whether we have an AR(p) model or RCA(p) model we will get the same linear predictor of Z_{n+1}. However, for the RCA model under consideration we have

$$E\left[Z_t|F_{t-1}^z\right] = \sum_{i=1}^p \phi_i Z_{t-i}$$

and

$$\text{Var}\,[Z_t|F_{t-1}^z] = \sigma_e^2 + \sum_{i=1}^{p} Z_{t-1}^2 \sigma_b^2 \; .$$

Thus the conditional variance is a nonlinear function and hence the RCA model (1.1) may be viewed as a nonlinear time series model.

Nicholls and Quinn (1980) studied linear as well as some nonlinear (proposed) predictors by fitting a nonlinear (RCA) model for the lynx data.

By giving heuristic reasoning they proposed a nonlinear predictor $\tilde{Z}_{n+1} = \text{sgn}(\phi_1 Z_n)[\phi_1^2 Z_n^2 + \sigma_e^2]^{\frac{1}{2}}$ and have shown empirically that the predictor $\tilde{Z}_{n+1}$ is a better predictor (having smaller prediction errors when compared with the actual observations) than the linear predictor $\hat{Z}_{n+1} = \phi Z_n$, for the lynx data.

It is of interest to note that by defining $h_t = Z_t^2 - E[Z_t^2|F_{t-1}^z]$, the optimal predictor for Z_{n+1} can be obtained as $Z_n^*(1) = \text{sqrt}\,[E[Z_t^2|F_{t-1}^z]] = \text{Sgn}(\phi_1 Z_n)\,[(\phi_1^2 + \sigma_b^2)Z_n^2 + \sigma_e^2]^{\frac{1}{2}}$. i.e. estimating function method could be used to obtain a nonlinear predictor for a nonlinear model by considering a class of elementary martingale estimating functions generated by nonlinear functions of the observations. Using a similar argument we could also propose a nonlinear forecast for the ARCH process.

Example 3.2: *Doubly stochastic times series*

Random coefficient autoregressive sequences given in (1.1) are special cases of what Tjøstheim (1986) refers to as doubly stochastic time series models. In the nonlinear case these models are given by

$$z_t - \theta_t\, f(t,\, F_{t-1}^z) = e_t, \tag{3.2}$$

where $\{\theta_t + b_t\}$ of (3.2) is now replaced by a more general stochastic sequence $\{\theta_t\}$ and z_{t-1} is replaced by a function of the past, $f(t,\, F_{t-1}^z)$. When $\{\theta_t\}$ is a moving average (MA) sequence of the form

$$\theta_t = \theta + \epsilon_t + \epsilon_{t-1}, \tag{3.3}$$

where $\{\theta_t\}, \{e_t\}$ are square integrable-independent random variables and $\{\epsilon_t\}$ consists of zero mean square integrable random variables independent of $\{e_t\}$. In this case $E(z_t|F_{t-1}^z)$ depends on the posterior mean, $m_t = E(\epsilon_t|F_t^z)$, and variance $\nu_t = E[(\epsilon_t - m_t)^2|F_t^z]$ of ϵ_t. Thus, for the evaluation of m_t and ν_t we further assume that $\{e_t\}$ and $\{\epsilon_t\}$ are Gaussian and that $z_0 = 0$. Then m_t and ν_t satisfy the following Kalman-like recursive algorithms (see Shiryayev, 1984, p. 439) :

$$m_t = \frac{\sigma_\epsilon^2\, f(t, F_{t-1}^z)[z_t - (\theta + m_{t-1})\, f(t, F_{t-1}^z)]}{\sigma_e^2 + f^2(t, F_{t-1}^z)(\sigma_e^2 + \nu_{t-1})}$$

and

$$\nu_t = \sigma_\epsilon^2 - \frac{f^2(t, F_{t-1}^z)\sigma_\epsilon^4}{\sigma_e^2 + f^2(t, F_{t-1}^z)(\sigma_e^2 + \nu_{t-1})}$$

where $\nu_0 = \sigma_\epsilon^2$ and $m_t = 0$. Hence

$$E(z_t|F_{t-1}^2) = (\theta + m_{t-1})f(t, F_{t-1}^z)$$

and

$$\begin{aligned} E(h_t^2|F_t^z) &= E\{[z_t - E(z_t|F_{t-1}^z)]^2|F_{t-1}^z\} \\ &= \sigma_e^2 + f^2(t, F_{t-1}^z)(\sigma_e^2 + \nu_{t-1}). \end{aligned}$$

Now the optimal predictor based on h_t is given by

$$\begin{aligned} z_n^*(1) &= \text{sqrt}[E(z_t^2|F_{t-1}^z)] \\ &= \text{sgn}(\theta + m_{t-1})f(t, F_{t-1}^z)[\sigma_e^2 + f^2(t, F_{t-1}^z)(\sigma_e^2 + \nu_{t-1}^2)]^{1/2}. \end{aligned}$$

ACKNOWLEDGEMENT

This paper was finalised while A. Thavaneswaran was at the University of Sydney. He thanks the School of Mathematics and Statistics for the support during his visit. The authors thank the referee and Professors I.V. Basawa and M. E. Thompson for very useful comments on an earlier version of this paper.

References

Abraham, B. and Thavaneswaran, A. (1991). A nonliner time series model and estimation of missing observations. *Ann. Inst. Statist. Math.* 43, 493-504.

Godambe, V.P. (1960). An optimal property of regular maximum likelihood equations. *Ann. Math. Statist.*, 31, 1208-1211.

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

Godambe, V.P. (1994). Linear Bayes and Optimal Estimation. Technical Report No. 11, University of Waterloo.

Naik-Nimbalkar, U.V. and Rajarshi, M.B. (1995). Filtering and Smoothing via Estimating Functions. *J. Amer. Statist. Assoc.* 90, 301-306.

Nicholls, D.F. and Quinn, B.G. (1982) *Random Coefficient Autoregressive Models. An Introduction.* Springer, New York.

Shiryayev, A.N. (1984). *Probability.* 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.

Thavaneswaran, A. and Peiris, S. (1996). Nonparametric estimation for some nonlinear models. *Stats. and Prob. Letters* 28, 227-233.

Tjøstheim, D. (1986) Estimation in nonlinear time series models. *Stochastic Processes and Appl.*, 21, 251-273.