Statistics & Probability Letters 6 (1988) 237–241
North-Holland March 1988

A BOUND FOR ESTIMATION IN NONLINEAR TIME SERIES MODELS BY INDEPENDENCE TESTING METHODS

Andrey FEUERVERGER
University of Toronto, Toronto, Canada M5S 1A1

Received April 1986
Revised July 1987

Abstract: The bound (19) is derived for the asymptotic efficiency of estimates of the parameter β in the nonlinear time series model (1) when the estimation is based on testing for independence of the estimated residuals from the series past. For intrinsically linear time series models efficiency is attainable regardless of the distribution of the error terms.

Keywords: time series analysis, nonlinear time series models, estimation, asymptotic efficiency, testing independence.

The object of this note is to point out that estimation in time series models through procedures based on independence testing is subject to certain efficiency limitations. Let $X_1, X_2, \ldots, X_T$ be a realization from a stationary time series model of the general type

$$X_t = h_\beta(X_{t-1}, X_{t-2}, \ldots) + \xi_t \tag{1}$$

where the ξ_t are iid and independent of the process past. The nature of h_β is required to be such that its values can be suitably approximated using only a finite segment of the process past, but otherwise h_β is not assumed to be linear and the error terms ξ_t are not assumed to be Gaussian. We do assume however that sufficient regularity holds so that the maximum likelihood procedure results in an asymptotically normal and (in the appropriate sense) asymptotically efficient estimator which may be taken as being essentially equivalent to a consistent root of the equation

$$\sum_{t=1}^{T} \frac{\partial \log f\left(X_t - h_\beta(X_{t-1}\ldots)\right)}{\partial\beta} = 0 \tag{2}$$

where $f(\cdot)$ is the density of the ξ's. For simplicity here, the parameter β is taken as univariate; the multiparameter extension is straightforward and given below. Evaluating the derivative in (2) results in

$$\sum_{t=1}^{T} g(\xi_t(\beta)) \cdot h'_\beta(X_{t-1}, \ldots) = 0 \tag{3}$$

where primes on h_β denote differentiation with respect to β and where

$$g(y) = \frac{\mathrm{d} \log f(y)}{\mathrm{d}\, y} \tag{4}$$

and

$$\xi_t(\beta) = X_t - h_\beta(X_{t-1}, \ldots). \tag{5}$$

0167-7152/88/$3.50

Observe that $\xi_t(\beta_0) \equiv \xi_t$. It should be noted that the estimating equation (3) is an *orthogonality*-type assertion involving the estimated residuals and the series past; indeed (3) is the (asymptotically) *optimal* estimation equation of the form $\Sigma_{t=1}^T U(\xi_t(\beta)) \cdot V_\beta(X_{t-1}, \ldots) = 0$ essentially because of the optimality of maximum likelihood that has been assumed.

Now the idea of estimation through testing for *independence* is a highly natural one. Thus, in the context of Box and Jenkins (1970) for example, one might ask whether the minimization of 'portmanteau' statistics (ibid., Section 2.2 of Chapter 8) could lead to meaningful procedures. In nonlinear times series modelling, estimation based on independence testing is proposed, for example, in Feuerverger and McDunnough (1981, Section 7). Some background and general discussion of nonlinear time series is given, for example, in Priestley (1981, Chapter 11), and references therein. An example of the most general models of the type (1) occurs when the h_β are infinite Volterra series as on the right side of equations 11.5.7 in Priesley (ibid., page 869). Unfortunately — with the specific exception of the linear case — when the Volterra expansion is truncated, a stationary solution generally does not exist, and for this reason nonlinear model families, while abundant, are generally less 'natural' in appearance than linear families. As a simple example of (1) consider $X_t = \beta \tan^{-1} X_{t-1} + \xi_t$. Note that although the parameter β enters here in a linear way, we nevertheless shall refer to this model as being *intrinsically nonlinear* since the dependence on the series past is of a nonlinear nature. By the same token, a model such as $X_t = e^\beta X_{t-1} + \xi_t$ is considered to be *intrinsically linear*. The exponential autoregressive model (ibid., page 889)

$$X_t = \left\{ \phi + \pi \exp\left(-\gamma X_{t-1}^2 \right) \right\} X_{t-1} + \xi_t$$

provides an example of a practical, multiparameter, intrinsically nonlinear model of the form (1).

Now independence among two sets of variates is equivalent to the absence of correlation, or covariance, among arbitrary functions of the variate sets. In the model (1) at the true value $\beta = \beta_0$ we would have $\xi_t(\beta_0) = \xi_t$ uncorrelated with every function of the series past $(X_{t-1}, X_{t-2}, \ldots)$. This motivates the examination of estimation procedures based on measuring the correlation between some function of $\xi_t(\beta)$ and some function of $(X_{t-1}, X_{t-2}, \ldots)$. We are thus led to consider a modification of (3), namely the covariance based equation

$$\operatorname{cov}_{t=1}^T \left[U\left(X_t - h_\beta(X_{t-1}, \ldots) \right), V(X_{t-1}, \ldots) \right] = 0 \tag{6}$$

which we may write as

$$\sum_{t=1}^T U\left(X_t - h_\beta(X_{t-1}, \ldots) \right) \cdot \left[V(X_{t-1}, \ldots) - \bar{V} \right] = 0 \tag{7}$$

where $\bar{V} = (1/T)\Sigma_{t=1}^T V(X_{t-1}, \ldots)$. [The notation employed in (6) is defined by $\operatorname{cov}_{t=1}^T(Y_t, Z_t) \equiv \Sigma_{t=1}^T (Y_t - \bar{Y})(Z_t - \bar{Z})$.]. To investigate the *maximum* efficiency attainable by means of estimation procedures of the general type (6, 7) we shall need to optimize over (U, V). To determine the optimal selections for $U(\cdot)$ and $V(\cdot)$ Taylor expand (7) about β_0 to obtain

$$\sum_{t=1}^T \left[U(\xi_t) - (\beta - \beta_0) U'(\xi_t) h'_{\beta_0}(X_{t-1}, \ldots) - \tfrac{1}{2}(\beta - \beta_0)^2 R_t \right] \left[V(X_{t-1}, \ldots) - \bar{V} \right] = 0 \tag{8}$$

where $R_t = U''(\xi_t)(h'_{\beta_*})^2 - U'(\xi_t) h''_{\beta_*}$ for some β_* between β and β_0 and then drop the second order term to obtain the approximate solution $\hat{\beta}$ given by

$$\sqrt{T}\left(\hat{\beta} - \beta_0 \right) = \frac{\frac{1}{\sqrt{T}} \sum U(\xi_t) V(X_{t-1}, \ldots) - \left(\frac{1}{\sqrt{T}} \sum U(\xi_t) \right)\left(\frac{1}{T} \sum V(X_{t-1}, \ldots) \right)}{\frac{1}{T} \sum U'(\xi_t) h'_{\beta_0}(X_{t-1}, \ldots)\left[V(X_{t-1}, \ldots) - \bar{V} \right]}. \tag{9}$$

Now since (6) is unchanged if a constant term is subtracted from either argument we may introduce constraints

$$E_f U(\xi)=0 \tag{10}$$

and

$$E_{\beta_0} V(X_{t-1},\ \cdots)=0 \tag{11}$$

under which (9) is seen to take on the asymptotic behaviour

$$\sqrt{T}(\hat{\beta}-\beta_0) \sim \frac{\frac{1}{\sqrt{T}} \sum_1^T U(\xi_t) V(X_{t-1},\ \ldots)}{E_{\beta_0} U'(\xi_t) \cdot h'_{\beta_0}(X_{t-1},\ \ldots) \cdot V(X_{t-1},\ \ldots)}. \tag{12}$$

This is asymptotically normal with mean zero and variance

$$\frac{EU^2(\xi_t) \cdot E_{\beta_0} V^2(X_{t-1},\ \ldots)}{\left(E_{\beta_0}\left[U'(\xi_t) \cdot h'_{\beta_0}(X_{t-1},\ \ldots) \cdot V(X_{t-1},\ \ldots)\right]\right)^2} \tag{13}$$

$$=\left[\frac{EU^2(\xi)}{(EU'(\xi))^2}\right] \cdot \left[\frac{E_{\beta_0} V^2(X_{t-1},\ \ldots)}{\left(E_{\beta_0} h'_{\beta_0}(X_{t-1},\ \ldots) \cdot V(X_{t-1},\ \ldots)\right)^2}\right]. \tag{14}$$

Thus it turns out that the optimization problem for U and V separates and we find — except for arbitrary constants — the optimal choices to be $U=g$ and $V=h'_{\beta_0}-Eh'_{\beta_0}$. (The result for V follows by virtue of the Cauchy–Schwartz inequality upon noting that because V is constrained by (11) then h' in (14) can be replaced there by $h'-Eh'$. To obtain that for U, observe that $EU^2/(EU')^2$ does not depend on the scale of U, and minimize the numerator $\int U^2(\xi) f(\xi)\ \mathrm{d}\xi$ subject to the denominator being unity; applying integration by parts, the denominator constraint may be written as $\int U(\xi) f'(\xi)\ \mathrm{d}\xi=1$ and the result of U now follows using straightforward variational arguments.) Substituting these values in (14) yields

$$\frac{Eg^2(\xi)}{(Eg'(\xi))^2} \cdot \frac{1}{\mathrm{VAR}_{\beta_0} h'_{\beta_0}(X_{t-1},\ \ldots)} \tag{15}$$

$$=I^{-1}(f) \cdot \left(\mathrm{VAR}_{\beta_0} h'_{\beta_0}(X_{t-1},\ \ldots)\right)^{-1} \tag{16}$$

where $I(f)$ is the Fisher information per observation from the location family based on $f(\cdot)$. (We remark here that V in (6, 7) is considered constant; the fact that the optimal V depends on the unknown true β_0 does not affect the argument towards the efficiency bound given below.)

To determine the asymptotic efficiency of this optimal covariance procedure we need to compare (16) with the result from the likelihood equation. Thus, expanding (3) now about β_0 as before gives, approximately,

$$\sqrt{T}(\hat{\beta}_{\mathrm{MLE}}-\beta_0)=\frac{\frac{1}{\sqrt{T}} \sum g(\xi_t) h'_{\beta_0}(X_{t-1},\ \ldots)}{\frac{1}{T} \sum\left[g'(\xi_t) \cdot h'_{\beta_0}(X_{t-1},\ \ldots)+g(\xi_t) h''_{\beta_0}(X_{t-1},\ \ldots)\right]} \tag{17}$$

which is asymptotically normal with mean zero and variance

$$I^{-1}(f) \cdot\left[E_{\beta_0}\left(h'_{\beta_0}(X_{t-1},\ \ldots)\right)^2\right]^{-1} \tag{18}$$

since the second denominator term in (17) is negligible by the Central Limit Theorem. Consequently the asymptotic efficiency of the best covariance-based procedure, given as the ratio of (18) to (16) is

$$e = \frac{\mathrm{VAR}_{\beta_0}\left(\dfrac{\partial h(X_{t-1}, \ldots)}{\partial \beta_0}\right)}{E_{\beta_0}\left(\dfrac{\partial h_\beta(X_{t-1}, \ldots)}{\partial \beta_0}\right)^2}. \tag{19}$$

The expression (19) apparently provides a measure of the *intrinsic* nonlinearity of a time series model of the form (1). In the essentially linear cases

$$X_t = \beta X_{t-1} + \xi_t \quad \text{or} \quad X_t = \rho(\beta) X_{t-1} + \xi_t \tag{20}$$

(where $\rho(\beta)$ need not be linear in β) for example, we obtain $e = 1$ regardless of the unknown distribution $f(\cdot)$ of the errors. Here, as before, the term linearity relates to the nature of the dependence of X_t on the previous X's and not on the nature of the parameterization

We remark here that alternative to procedures based on the independence of $\xi_t(\beta)$ from $(X_{t-1}, X_{t-2}, \ldots)$ are procedures based on measuring the extent to which the $\{\xi_t(\beta)\}$ sequence is iid. Under stationarity and mild regularity, independence of the $\{\xi_t(\beta)\}$ sequence is equivalent to the independence of $\xi_t(\beta)$ from $(\xi_{t-1}(\beta), \xi_{t-2}(\beta), \ldots)$ with this latter being a function of $(X_{t-1}, X_{t-2}, \ldots)$. Now under suitable invertibility conditions the use of $(\xi_{t-1}(\beta), \xi_{t-2}(\beta), \ldots)$ will be equivalent to $(X_{t-1}, X_{t-2}, \ldots)$; failing this, the former leads to a subset of the procedures based on the latter. Hence the bound (19) will remain valid but not necessarily attainable.

Straightforward multiparameter extension of the calculations above leads to expressions (16) and (18) except with the second terms in these expressions replaced by the obvious matrices. Specifically, the expressions become $I^{-1}(f) \cdot \boldsymbol{A}^{-1}$ and $I^{-1}(f) \cdot \boldsymbol{B}^{-1}$ respectively where $\boldsymbol{A} = \operatorname{cov}(\boldsymbol{Y})$, and $\boldsymbol{B} = E\boldsymbol{Y}\boldsymbol{Y}'$ are the covariance and second moment matrices for the $p \times 1$ vector $\boldsymbol{Y}$ whose entries are $\partial h_{\beta_0}(X_{t-1}, \ldots)/\beta_i$, $i = 1, 2, \ldots, p$ where $\beta = (\beta_1, \beta_2, \ldots, \beta_p)$. Note that $\boldsymbol{B} - \boldsymbol{A} = (E\boldsymbol{Y})(E\boldsymbol{Y})'$ is in general a nonnegative definite matrix of unit rank and will be zero if $E\boldsymbol{Y} = \boldsymbol{0}$. We see from these multiparameter expressions that for linear models such as

$$(X_t - \mu) = \sum_{l=1}^{p} \beta_l (X_{t-1} - \mu) + \xi_t \tag{21}$$

the optimal independence based estimation procedure will be asymptotically efficient for the autoregressive parameters, i.e. for estimating the β_l's and, for example, that the same will continue to hold for the $\boldsymbol{\theta}$'s if the β_l's in (21) should be linked in some narrower and not necessarily linear parameterization $\beta_l = \beta_l(\boldsymbol{\theta})$. This holds regardless of the distribution of the error terms. The parameter μ of course cannot be estimated in this manner.

We remark that for models having essentially nonlinear character (19) may yield values $e < 1$, but not necessarily poor efficiencies.

Acknowledgement

This research was supported by a grant from the National Sciences and Engineering Research Council of Canada. Portions of this work were carried out at Tel Aviv University whose generous hospitality and facilities provided are gratefully acknowledged.

References

Box, G.E.P. and G.M. Jenkins (1970), *Time Series Analysis: Forecasting and Control* (Holden Day, San Fransisco).

Feuerverger, A. and P. McDunnough (1981), On some Fourier methods for inference, *J. Amer. Statist. Assoc.* **76**, 379–387.

Priestley, M.B. (1981), *Spectral Analysis and Time Series, Vol. 1* (Academic Press, London).