KYBERNETIKA — VOLUME 40 (2004), NUMBER 6, PAGES 691 - 702

ON GEOMETRIC ERGODICITY AND PREDICTION IN NONNEGATIVE NON-LINEAR AUTOREGRESSIVE PROCESSES

PETR ZVÁRA

A non-linear AR(1) process is investigated when the associated white noise is positive. A criterion is derived for the geometric ergodicity of the process. Some explicit formulas are derived for one and two steps ahead extrapolation. Influence of parameter estimation on extrapolation is studied.

Keywords: geometric ergodicity, non-linear autoregression, least squares extrapolation

AMS Subject Classification: 62M10, 62M20

1. INTRODUCTION

When one predicts in parametric AR time series models, usually the autoregressive functional form is assumed to be known, the order and parameters are estimated and the point prediction with a predictive interval are obtained using estimated order and parameters as the true ones. In this paper, we study the effect of uncertainty in parameters on prediction.

The paper is organized as follows. In Section 2 some Markov chain terminology is introduced and conditions for geometric ergodicity are derived in Theorem 2. Section 3 is used to review some estimation methods for nonnegative non-linear AR processes. The least squares extrapolation is studied in Section 4, some explicit formulas for extrapolation and bias of estimated prediction are derived.

Consider a non-linear AR(1) process $\{X_t, t \geq 0\}$ (abbreviated as NLAR(1)) given by

$$X_t = \lambda(X_{t-1} \,|\, \boldsymbol{\theta}) + e_t, \;\; t = 1, \ldots, n, \tag{1}$$

where λ is a Borel measurable function, $\boldsymbol{\theta} \in \mathbb{R}^k$ is a vector of parameters, $\{e_t, t \geq 1\}$ is a sequence of i.i.d. variables with finite variance and a density g, and X_0 is a given random variable independent of $\{e_t\}$. In what follows we assume that the white noise distribution is known and need not be estimated.

The least squares predictor of X_{t+m} given a past history, $X_t, X_{t-1}, \ldots$, is given by

$$K_{m,t}(x) = \mathsf{E}\,[X_{t+m} \,|\, X_t = x]\,. \tag{2}$$

Functions $K_{m,t}(x)$ are independent of t. They satisfy, with subscript t already dropped, the recurrent relationship

$$K_0(x) = x, \quad K_m(x) = \int_{-\infty}^{\infty} K_{m-1}(y) g\left[y - \lambda(x \mid \theta)\right] \mathrm{d}y. \tag{3}$$

A proof can be found in Tong [12] when $\{X_t\}$ is stationary and in Anděl, Dupač [3] for the nonstationary, general autoregressive model. An estimate of $K_m(x)$ is obtained by plugging in the estimated regression parameters $\hat{\theta}_n$ into $\lambda(\cdot \mid \theta)$ in (3). We study the bias of this estimate in Section 4.

2. GEOMETRIC ERGODICITY AND STATIONARY DISTRIBUTION

When we consider the time series models, because of its importance in the statistical inference of the stochastic process, their stationarity properties are to be studied carefully. In many papers, for example, Tjøstheim [11], Bhattacharya and Lee [5], Cline and Pu [6], Lee [9], general non-linear models have been studied and many suitable tools and criteria are given, but the usual assumption for the innovation process $\{e_t\}$ is that it has a positive density function $g(x) > 0, x \in (-\infty, -\infty)$ on the real line with respect to Lebesgue measure. The methods developed here provide modifications of sufficient conditions for geometric ergodicity for non-linear AR(1) models and specific types of distribution of $\{e_t\}$.

We start with some Markov chain terminology drawn from the papers by Tjøstheim [11], Bhattacharya and Lee [5], Lee [9] and the book by Tong [12].

Let $\{X_t, t \geq 0\}$ be a homogeneous Markov chain taking values in $(E, \mathcal{E})$, where $\mathcal{E}$ is a countably generated σ-algebra of subsets of E. In this paper usually $E = \mathbb{R}^1$ and $\mathcal{E}$ is the σ-algebra of Borel sets. Denote the n-step transition probabilities by

$$P^n(x, A) = P\left[X_n \in A \mid X_0 = x\right], \quad x \in E, \ A \in \mathcal{E},$$

with $P(x, A) = P^1(x, A)$.

The Markov chain $\{X_t\}$ is said to be *φ-irreducible* if, for some nontrivial σ-finite measure φ on $(E, \mathcal{E})$,

$$\sum_n P^n(x, A) > 0$$

for all $x \in E$ and every $A \in \mathcal{E}$ with $\varphi(A) > 0$.

Let $\{X_t\}$ be a *φ-irreducible* chain. A set $B \in \mathcal{E}$ with $\varphi(B) > 0$ is said to be *small* (with respect to φ) if for all A, $\varphi(A) > 0$

$$\inf_{x \in B} \sum_{n=1}^{m} P^n(x, A) > 0$$

for some $m \geq 1$. It can be shown (Tjøstheim [11]) that in the model (1) with λ a measurable function and $\{e_t\}$ having a positive density g on the real line with respect to Lebesgue measure l_1, $\{X_t\}$ is *l_1-irreducible* and a compact set is small if λ is bounded on it.

A measure π on $\mathcal{E}$ is called the *invariant* for the process $\{X_t\}$ if, for all sets $B \in \mathcal{E}$,

$$\pi(B) = \int P(y, B)\pi(\mathrm{d}y).$$

A φ-irreducible Markov process is said to be *geometrically ergodic* if there exists a probability measure π and $\rho \in (0, 1)$, such that

$$\sup_{B \in \mathcal{E}} \rho^{-n} |P^n(x, B) - \pi(B)| \to 0, \quad \text{as } n \to \infty, \forall\, x \in E.$$

Suppose $\{X_t\}$ is geometrically ergodic. Then, if the initial distribution of X_0 is π, $\{X_t\}$ is strictly stationary (Tong [12]).

The following lemma provides the so called h-step criterion of geometric ergodicity (Tjøstheim [11]), which will be used later in this section.

Lemma 1. Assume that $\{X_t\}$ is an aperiodic φ-irreducible chain, and let v be a nonnegative measurable function on E (test function). The chain $\{X_t\}$ is geometrically ergodic if there exist a positive integer h, a small set $K \in \mathcal{E}$ with complement K^C, and $\epsilon > 0, M < \infty, R > 1$, such that

$$R\mathsf{E}[v(X_{t+h}) \mid X_t = x] \leq v(x) - \epsilon, \quad x \in K^C, \tag{4}$$

and

$$\mathsf{E}\left[v(X_{t+h})I(X_{t+h} \in K^C) \mid X_t = x\right] \leq M, \quad x \in K, \tag{5}$$

where $I(\cdot)$ is the indicator function.

Taking $v(x) = |x|$ and $h = 1$, one can prove the following theorem (Tjøstheim [11]):

Theorem 1. Let $\{X_t\}$ be given by (1), where density g of $\{e_t\}$ is positive everywhere. If $\mathsf{E}|e_t| < \infty$ and if the function λ of (1) is bounded on compact sets, and if there exists an $r > 0$ such that

$$\sup_{|x|>r} \left|\frac{\lambda(x)}{x}\right| < 1, \tag{6}$$

then $\{X_t\}$ is geometrically ergodic.

In the next sections, NLAR(1) models with positive values and positive innovation terms will be studied. Therefore, the conditions on g need to be relaxed. We will prove following theorem:

Theorem 2. Let $\{X_t\}$ given by (1) be a nonnegative process, that is $X_t \geq 0$ a.s., $t = 0, 1, \ldots$. If the function λ of (1) is measurable on $[0, \infty)$, bounded on compact sets, $\lambda(x) \geq 0$ for $x \geq 0$, and if there exists an $r > 0$ such that λ satisfies (6), then each of the following conditions is sufficient for geometric ergodicity of $\{X_t\}$:

1) Density g of $\{e_t\}$ is positive on $(0, \infty)$.

2) λ is concave, strictly increasing and has continuous derivative on $[0, \infty)$ and g is positive on some (a, b), where $0 \leq a < b < \infty$.

Proof. Assume that condition 1) is satisfied.

First we verify irreducibility of the process. Let φ be the Lebesgue measure restricted to the interval $[r, \infty)$, $\varphi([a, b]) = \max(b, r) - \max(a, r)$, where r is from (6). Since we assume now $g(x) > 0$, $x > 0$, then $\lambda(x_0) \leq r$ implies that $P(x_0, A) > 0$ for every $A, \varphi(A) > 0$. Assume $x_0 > r$, define $\alpha = \sup_{|x|>r} \left|\frac{\lambda(x)}{x}\right|$ and let $x_1 = \lambda(x_0)$, $x_k = \lambda(x_{k-1}) = \lambda^{(k)}(x_0)$. Then we have $x_h \leq \alpha^h x_0$, if $x_k > r$, $0 \leq k \leq h$. Hence, for every $x \geq 0$, there exists a finite m, such that $\lambda^{(m)}(x) < r$ and therefore $P^m(x, A) > 0$ for every A, $\varphi(A) > 0$ and thus irreducibility is verified.

Further we discuss small sets. For $s > r$ choose a set $K = \{x : r < x < s\}$. Since λ is bounded on K, for some finite M is $\lambda(x) \leq M$, $x \in K$. There exists an m large enough, such that $\alpha^m M < r$. For such an m, for every $x \in K$, is $\lambda^{(h)}(x) < r$ for some $1 \leq h \leq m$. It follows that for every A, $\varphi(A) > 0$ is $\sum_{n=1}^{m} P^n(x, A) > 0$, $\forall x \in K$ for some $m \geq 1$, which shows that K is small set.

Finally, we deal with h-step criterion. Take $v(x) = x$, $h = 1$, and choose $K = \{x : r < x < s\}$. For $x \in K^C$, we have

$$\mathsf{E}[X_{t+1} \mid X_t = x] = \lambda(x) + \mathsf{E}\, e_t \leq \alpha x + \gamma,$$

where $\gamma = \mathsf{E}\, e_t$. Choose $R > 1$ such that $R\alpha < 1$. Then

$$\mathsf{E}[X_{t+1} \mid X_t = x] \leq x + (R\alpha - 1)x + R\gamma,$$

and it follows that (4) is satisfied by taking a small set K with s large enough.

For $x \in K$, we have

$$\mathsf{E}\left[X_{t+1} I(X_{t+1} \in K^C) \mid X_t = x\right] \leq \lambda(x) + \gamma \leq M + \gamma,$$

which shows (5) and completes geometric ergodicity of $\{X_t\}$.

Now, assume that condition 2) is satisfied. Anděl [2] has studied the model $X_t = \omega X_{t-1}^q + e_t$, where $\omega > 0$, $q \in (0, 1)$, and e_t has a rectangular distribution e_t with parameters a, b, $0 \leq a < b < \infty$. Our situation is a generalized version of this model.

Again, first we verify irreducibility of the process. We show, that for each $z > 0$ the equation $x = \lambda(x) + z$ has a unique positive root x_z and x_z is an increasing function of z.

Since $\lambda(0) \geq 0$ and λ is continuous, there exists a $\delta \in (0, z)$, such that $\lambda(x) + z - x > 0$ for $0 \leq x \leq \delta$. Further, $\lambda(x) \leq \alpha x$ for $x > r$, where $1 > \alpha = \sup_{x>r} \left|\frac{\lambda(x)}{x}\right|$.

Since $\alpha - 1 < 0$, there exists $M > 0$ such that $(\alpha - 1)M + z < 0$. Thus $\lambda(x) + z - x < (\alpha - 1)x + c < 0$ for $x > M$.

Hence for each $z > 0$ we have found $0 \leq \delta < M < \infty$ such that $\lambda(x) + z - x > 0$ for $0 \leq x < \delta$ and $\lambda(x) + z - x < 0$ for $x > M$. Therefore there exists $x = x_z$ such that $x_z = \lambda(x_z) + z$.

Since λ is increasing and concave, the slope of the λ curve at x_z must be from interval $(0, \frac{\pi}{4})$ (that is $0 < \lambda(x_z)' < 1$) and x_z is unique.

According to implicit function theory, $x_z' = -\frac{1}{\lambda'(x_z)-1} > 0$, thus x_z is an increasing function of z.

Let φ be the Lebesgue measure restricted to the interval $[x_a, x_b)$, $\varphi([u, v]) = \min(\max(v, x_a), x_b) - \min(\max(u, x_a), x_b)$. If $x < x_a$ and $x > \lambda(x)$, then there exists a positive number $c < a$, such that $x = \lambda(x) + c$, since x_z is an increasing function of z, and we have $x = \lambda(x) + c < \lambda(x) + a$. Hence, with positive probability, $X_{t+1} > X_t + (b - a)/2$, if $X_t < x_a$. Similarly, $X_{t+1} < X_t - (b - a)/2$, if $X_t > x_b$. Therefore X_t reaches (x_a, x_b) in a finite number of steps with positive probability.

If $x_a < x < x_b$, then there exists a $d \in (a, b)$, such that $x = \lambda(x) + d$. Thus we have

$$x_a = \lambda(x_a) + a < \lambda(x) + a < x = \lambda(x) + d < \lambda(x_b) + b = x_b.$$

This implies that if $X_t \in (x_a, x_b)$, then $X_{t+1} \in (x_a, x_b)$ with probability 1 and there is positive probability that $X_{t+1} < X_t$ and that $X_{t+1} > X_t$. Therefore $\{X_t\}$ is φ-irreducible.

Further we discuss small sets. Let $[u, v]$ be a subset of $[x_a, x_b]$. Let $A = [x_a, x_a + \epsilon]$, where ϵ is small positive number. Define $v_1 = \lambda(v) + a$, $v_k = \lambda(v_{k-1}) + a$. Clearly, $\{v_k\}$ is a decreasing sequence and $v_k \to x_a$ as $k \to \infty$. There exists an m such that $v_m < x_a + \epsilon$. Hence

$$\inf_{x \in [u,v]} \sum_{n=1}^{m} P^n(x, A) = \sum_{n=1}^{m} P^n(v, A) > 0,$$

which implies that any interval $[u, v]$ is small set.

Finally, proof of the h-step criterion is the same as in 1). □

3. PARAMETER ESTIMATION

Let $\boldsymbol{X}^n = (X_1, X_2, \ldots, X_n)^T$ denote a sample of n consecutive observations from the series $\{X_t : t \in \mathbb{Z}\}$. Assume that $\boldsymbol{X}^n$ has a probability density $p_n(x_1, \ldots, x_n \mid \boldsymbol{\theta}, x_0)$, which depends on $\boldsymbol{\theta} \in \mathbb{R}^k$ and $X_0 = x_0 \in \mathbb{R}^1$. Let $L_n(\boldsymbol{\theta})$ denote the conditional log-likelihood, that is $\log p_n(x_1, \ldots, x_n \mid \boldsymbol{\theta}, x_0)$. In the model (1) is

$$L_n(\boldsymbol{\theta}) = \sum_{t=1}^{n} \log g\left[X_t - \lambda(X_{t-1} \mid \boldsymbol{\theta})\right]. \tag{7}$$

We define the maximum likelihood estimator $\hat{\boldsymbol{\theta}}_n$ as a global maximizer of (7),

$$\hat{\boldsymbol{\theta}}_n = \arg\max_{\boldsymbol{\theta}} \sum_{t=1}^{n} \log g\left[X_t - \lambda(X_{t-1} \mid \boldsymbol{\theta})\right]. \tag{8}$$

Distributional properties of $\hat{\theta}_n$ are known in some situations.

Assume that

(i) $L_n(\theta)$ is differentiable w.r.t. θ,

(ii) $\mathsf{E}_\theta \left[\frac{\partial L_n}{\partial \theta}(\theta)\frac{\partial L_n}{\partial \theta}(\theta)^T\right] < \infty$ for each n, and

(iii) $\int p_n(x_1,\ldots,x_n;\theta)\,\mathrm{d}x^n$ can be differentiated twice w.r.t. θ under the integral sign.

Define $L_0(\theta) = 0$. Let $u_i(\theta)$ be the column vector of the first partial derivatives of $L_i(\theta) - L_{i-1}(\theta)$. Further define $I_n(\theta)$ by $\sum_{i=1}^n \mathsf{E}_\theta \left[u_i(\theta)u_i(\theta)^T \mid X_{i-1}\right]$. Then, under some additional technical conditions on $I_n(\theta)$, we have (Tong [12], p. 295)

$$\hat{\theta}_n \xrightarrow[n\to\infty]{\mathrm{d}} N(\theta, I_n(\theta)^{-1}).$$

However, there is a bias of order n^{-1}, which will be discussed later.

Example 1. Consider the model discussed by Anděl [2]

$$X_t = \omega X_t^q + e_t,\ t = 1,\ldots,n,\ 0 < q < 1 \text{ known}, \tag{9}$$

where $\omega \geq 0$ and e_t is a gamma distributed sequence of i.i.d. variables with known parameters α and β, i.e. with density $g(y) = \frac{1}{\Gamma(\alpha)\beta^\alpha} y^{\alpha-1}\exp\{-\frac{y}{\beta}\}$, $y > 0$. Then we have $k = 1$, $\theta = \omega$. Calculations give $u_i(\omega) = X_{i-1}^q\left(\frac{1}{\beta} - \frac{\alpha-1}{X_i - \omega X_{i-1}^q}\right)$ and

$$I_n(\omega) = \begin{cases} n\dfrac{\mathsf{E}\,X_1^{2q}}{(\alpha-2)\beta^2}, & \alpha > 2, \\ \text{does not exist}, & \alpha \leq 2. \end{cases} \tag{10}$$

Expectation is taken with respect to the stationary density of $\{X_t\}$, which exists according to Theorem 2. Hence $\hat{\omega}_n$ which is solution of the equation

$$\sum_{i=1}^n X_{i-1}^q\left(\frac{1}{\beta} - \frac{\alpha-1}{X_i - \omega X_{i-1}^q}\right) = 0 \tag{11}$$

is asymptotically normally distributed with mean ω and variance

$$\mathsf{var}_{as}\hat{\omega}_n = \frac{(\alpha-2)\beta^2}{n\mathsf{E}\,X_1^{2q}},$$

if $\alpha > 2$.

In a simulation study we set $q = 1/2$, $\omega = 0.1, 1$, and 10. The innovations have mean 1 and variance $1/3$, i.e. $\alpha = 3, \beta = 1/3$. Properties of the estimator for sample size $n = 100$ are summarized in Table 1. The bias and variance of the estimator are computed by Monte Carlo simulations based on 10,000 replications. The solution

of equation (11) was obtained numerically using globally convergent combination of bisection and Newton–Raphson method (Numerical Recipes in C). The stationary density of $\{X_t\}$ used for computation of $\mathrm{var}_{as}\hat{\omega}_n$ was calculated numerically as the limit of conditional densities (Moeanaddin and Tong [10]).

Table 1. Bias and variance of the parameter estimate in the model $X_t = \omega\sqrt{X_{t-1}} + \Gamma(3, 1/3),\ t = 1, \ldots, 100.$

ω	$\mathsf{E}[\hat{\omega}_n - \omega]$	$100 * \mathrm{var}\,\hat{\omega}_n$	(Asymptotic)
0.1	0.005	0.119	(0.101)
1	0.003	0.052	(0.043)
10	0.0005	0.001	(0.001)

Datta, Mathew and McCormick [7] studied models of the form

$$X_t = \sum_{i=0}^{p} \phi_i f_i(X_{t-1}) + e_t, \tag{12}$$

where $\phi = (\phi_0, \ldots \phi_p)^T$ is an unknown parameter vector, f_i are known functions and e_t are nonnegative i.i.d. innovations. Their linear programming estimator (LPE) of ϕ is essentially the maximum likelihood estimator when the innovation distribution is exponential. Let $U(t) = (f_0(X_{t-1}), \ldots, f_p(X_{t-1}))^T$ and consider the set of ϕ values

$$F_n = \{\phi \in \mathbb{R}^{p+1} : X_t - \phi^T U(t) > 0, 1 \leq t \leq n\}.$$

Furthermore, let $\bar{f}(n) = (\bar{f}_0(n), \ldots, \bar{f}_p(n))^T$ be the vector of means, with $\bar{f}_i(n) = n^{-1}\sum_{t=1}^{n} f_i(X_{t-1})$, $0 \leq i \leq p$. Then the LPE of ϕ maximizes the objective function $\phi^T \bar{f}(n)$ over F_n, i. e.

$$\hat{\phi}_n = \arg\max_{\phi \in F_n} \phi^T \bar{f}(n). \tag{13}$$

The limiting distribution of $\hat{\phi}_n$ is rather complicated and explicit formula for its density is not known to our knowledge. Only random samples can be taken from this distribution. It requires knowledge of expectations $\nu_i = \mathsf{E} f_i(X_t)$ taken with respect to the stationary distribution. Details can be found in Datta, Mathew and McCormick [7].

Example 2. Consider the model

$$X_t = \omega X_{t-1}^q + e_t,\ t = 1, \ldots, n,\ 0 < q < 1 \text{ known}, \tag{14}$$

where $\omega \geq 0$ and e_t is exponentially distributed with a known parameter a, i. e. with density $g(y) = a^{-1}\exp\{-\frac{y}{a}\}, y > 0$. In this setting $U(t) = f_0(X_{t-1}) = X_{t-1}^q$ and $\hat{\omega}_n = \min \frac{X_t}{X_{t-1}^q}$.

For the simulation we set $q = 1/2$ and $\omega = 0.1, 1$, and 10. The innovations have mean 1 and variance 1, i. e. $a = 1$. Properties of the estimator for sample size $n = 100$ are summarized in Table 2. The bias and mean square error are calculated by Monte Carlo simulation based on 10,000 replications. The asymptotic counterparts are computed by simulating 2,000 values from the asymptotic distribution. Datta, Mathew and McCormick [7] have observed that their results are close to the ones predicted by limit theory for $n = 200$ and higher. According to our simulation for $n = 100$, the bias and mean square error of estimator is underestimated by asymptotic theory when $\omega = 0.1$. For higher ω, i. e. $\omega = 1$ and $\omega = 10$ are the results in close agreement with asymptotic theory.

Table 2. Bias and variance of the parameter estimate in the model $X_t = \omega\sqrt{X_{t-1}} + \mathrm{Exp}(1),\ t = 1, \ldots, 100.$

ω	$\mathsf{E}[\hat{\omega}_n - \omega]$	(Asymptotic)	$100 * \mathsf{E}[(\hat{\omega}_n - \omega)^2]$	(Asymptotic)
0.1	0.011	(0.008)	0.023	(0.014)
1	0.006	(0.006)	0.008	(0.007)
10	0.001	(0.001)	0.0002	(0.0002)

Consider an AR(1) process $\{X_t\}$ defined by

$$X_t = bX_{t-1} + e_t,\ t = 1, \ldots, n, \tag{15}$$

where $0 \leq b < 1$, $e_1, \ldots, e_n$ are i.i.d. $\mathrm{Exp}(a)$ variables and $X_0 \sim \mathrm{Exp}(a/(1-b))$. Then the expectation is preserved and the maximum likelihood estimator is $\hat{b} = \min\left(\frac{X_1}{X_0}, \ldots, \frac{X_n}{X_{n-1}}\right)$. The exact distribution of $\hat{b}$ is according to Anděl [1]

$$\begin{aligned} P(\hat{b} < v) &= 0 && (v \leq b) \\ P(\hat{b} < v) &= 1 - (1-b)[\{v + (1-b)\}\{v^2 + (1-b)(1+v)\} \ldots \\ &\quad \times \{v^{n-2} + (1-b)(1 + v + \cdots + v^{n-3})\} \\ &\quad \times \{v^{n-1} + (1-b)(1 + v + \cdots + v^{n-2}) - b\}]^{-1} && (v > b). \end{aligned} \tag{16}$$

Example 3. Consider the model (15) with $a = 1$. We calculate theoretical bias and variance of the estimator $\hat{b}$ using Gauss-Legendre quadrature formula. We simulated 10,000 replications of the model (15) with $n = 100$ for values $b = 0.1, 0.5$ and 0.9. The empirical bias and variance together with their theoretical counterparts are summarized in Table 3. The simulated quantities reflect well the theoretical ones except the mean square error of $\hat{b}$ when $b = 0.9$. This might be caused by the error of numerical quadrature method, because the integrand is almost a singular function.

Table 3. Bias and variance of the parameter estimate in the model $X_t = bX_{t-1} + \text{Exp}(1)$, $X_1 \sim \text{Exp}(1/(1-b))$, $t = 1, \ldots, 100$.

b	$\mathsf{E}\hat{b} - b$	(Theoretical)	$100 * \mathsf{E}[(\hat{b} - b)^2]$	(Theoretical)
0.1	0.009	(0.009)	0.017	(0.016)
0.5	0.005	(0.005)	0.005	(0.005)
0.9	0.001	(0.001)	0.0002	(0.003)

4. INFLUENCE OF ESTIMATED PARAMETERS ON LEAST SQUARES PREDICTION

Let $K_m(x)$ be the true unknown least squares predictor of X_{t+m} given $X_t = x$ in the NLAR(1) model (1). Define the estimate $\hat{K}_m(x)$ of $K_m(x)$ by plugging in the estimated regression coefficients $\hat{\boldsymbol{\theta}}_n$ into $\lambda(\cdot \mid \boldsymbol{\theta})$ in (3) instead of unknown $\boldsymbol{\theta}$. For $m = 1$, we have

$$\hat{K}_1(x) = \lambda(x \mid \hat{\boldsymbol{\theta}}_n) + \gamma, \tag{17}$$

where γ is the expectation of the innovations. We state the formula for bias of $\hat{K}_1(x)$ in the following theorem.

Theorem 3. Let $\hat{K}_1(x)$ be the estimated 1-step prediction (17) in the process (1). Let the parameter $\hat{\boldsymbol{\theta}}_n$ be any estimate of $\boldsymbol{\theta}$ in (8) such that

$$\mathsf{E}\big[(\hat{\theta}_{ni} - \theta_i)(\hat{\theta}_{nj} - \theta_j)(\hat{\theta}_{nl} - \theta_l)\big] = O(n^{-3/2}). \tag{18}$$

Then we have

$$\begin{aligned} \mathsf{E}[\hat{K}_1(x) - K_1(x)] &= \sum_{i=1}^{k} \mathsf{E}[\hat{\theta}_{ni} - \theta_i] \frac{\partial \lambda}{\partial \theta_i}(x \mid \boldsymbol{\theta}) \\ &\quad + \frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{k} \mathsf{E}\big[(\hat{\theta}_{ni} - \theta_i)(\hat{\theta}_{nj} - \theta_j)\big] \frac{\partial^2 \lambda}{\partial \theta_i \partial \theta_j}(x \mid \boldsymbol{\theta}) + O(n^{-3/2}) \end{aligned} \tag{19}$$

provided $\lambda(x \mid \boldsymbol{\theta})$ has bounded third derivatives w.r.t. $\boldsymbol{\theta}$.

Proof. Taylor expansion of $\hat{K}_1(x)$ around $K_1(x)$ up to third order terms gives

$$\begin{aligned} \hat{K}_1(x) - K_1(x) &= \sum_{i=1}^{k} [\hat{\theta}_{ni} - \theta_i] \frac{\partial \lambda}{\partial \theta_i}(x \mid \boldsymbol{\theta}) \\ &\quad + \frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{k} [(\hat{\theta}_{ni} - \theta_i)(\hat{\theta}_{nj} - \theta_j)] \frac{\partial^2 \lambda}{\partial \theta_i \partial \theta_j}(x \mid \boldsymbol{\theta}) \\ &\quad + \frac{1}{6} \sum_{i=1}^{k} \sum_{j=1}^{k} \sum_{l=1}^{k} [(\hat{\theta}_{ni} - \theta_i)(\hat{\theta}_{nj} - \theta_j)(\hat{\theta}_{nl} - \theta_l)] \frac{\partial^3 \lambda}{\partial \theta_i \partial \theta_j \partial \theta_l}(x \mid \boldsymbol{\theta}^*), \end{aligned}$$

where $\boldsymbol{\theta}^* = \boldsymbol{\theta} + \psi(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})$, $0 < \psi < 1$. Since the expectation of the last term on the right side is of order $O(n^{-3/2})$, the theorem follows. □

Assumption (18) is valid for example in linear Gaussian AR models (Bhansali [4]). Explicit expressions for bias of the estimated higher-step predictions are available only in special cases.

In the model (9) of Example 1, we have $\hat{K}_1(x) = \hat{\omega}_n x^q + \alpha\beta$, where α and β are characteristics of the innovation distribution and $\hat{\omega}_n$ is the maximum likelihood estimate, i.e. solution of equation (11). Hence we have

$$\mathsf{E}[\hat{K}_1(x) - K_1(x)] = \mathsf{E}[\hat{\omega}_n - \omega]x^q. \tag{20}$$

We see that the higher is x, the more biased is the one-step prediction. The extent of this bias depends on how precisely we estimate ω.

It can be shown after long and tedious calculation that the two-step ahead prediction in the model (9) is

$$\begin{aligned} K_2(x) &= \alpha\beta + \frac{\omega\beta^q \exp\left\{\frac{\omega x^q}{\beta}\right\}}{(\alpha-1)!} \\ &\quad \times \sum_{k=0} \alpha - 1 \binom{\alpha-1}{k} \left[-\frac{\omega x^q}{\beta}\right]^{\alpha-1-k} \Gamma(k+q+1)[1 - \Gamma^*_{\frac{\omega x^q}{\beta}}(k+q+1)], \end{aligned} \tag{21}$$

when $\alpha \in \mathbb{N}^+$. Function $\Gamma^*_x(m) = \frac{1}{\Gamma(m)} \int_0^x t^{m-1} e^{-t}\, \mathrm{d}t$ denotes the incomplete gamma function. The estimate $\hat{K}_2(x)$ is obtained by plugging in $\hat{\omega}_n$ into (21).

Table 4. Bias of the one- and two-step prediction estimate in the model $X_t = \omega\sqrt{X_{t-1}} + \Gamma(3, 1/3)$, $t = 1, \ldots, 100$.

ω	x	$K_1(x)$	$\mathsf{E}[\hat{K}_1(x) - K_1(x)]$	(As.)	$K_2(x)$	$\mathsf{E}[\hat{K}_2(x) - K_2(x)]$
0.1	0.522	1.072	0.004	(0.004)	1.100	0.006
	1.101	1.105	0.006	(0.005)	1.102	0.006
	1.679	1.130	0.007	(0.006)	1.103	0.007
1	1.993	2.412	0.005	(0.004)	2.543	0.007
	2.601	2.613	0.005	(0.005)	2.607	0.007
	3.210	2.792	0.006	(0.005)	2.663	0.007
10	98.50	100.25	0.005	(0.005)		
	101.88	101.94	0.005	(0.005)		
	105.26	103.60	0.006	(0.006)		

For the simulation we use the same parameter values as in Example 1. Biases of $\hat{K}_1(x)$ and $\hat{K}_2(x)$ are summarized in Table 4. The quantities are computed by Monte Carlo simulation based on 10,000 replications. We calculate $K_i(x)$ and

$\mathsf{E}[\hat{K}_i(x) - K_i(x)]$, $i = 1, 2$ at three different x: stationary mean − stationary standard deviation, stationary mean and stationary mean + stationary standard deviation. The asymptotic approximations of $\mathsf{E}[\hat{K}_1(x) - K_1(x)]$ are obtained by plugging in the numerical values of $\mathsf{E}[\hat{\omega}_n - \omega]$ from Table 1 into (20).

The results are in very good agreement with the theory. All quantities are close to their asymptotic counterparts. Note that the bias in prediction is negligible, it accounts in maximum for 0.4% of the absolute predicted value. We did not calculate the two-step ahead predictions in the model with $\omega = 10$, because of singularity of the required integrands.

Consider the zero mean Gaussian AR(1) model (22).

$$X_t = \omega X_{t-1} + e_t, \ t = 1, \ldots, n, \tag{22}$$

Here we have $K_m(x) = \omega^m x$ and $\hat{K}_m(x) = \hat{\omega}_n^m x$ for $m \geq 1$. Taylor expansion up to the second order term around ω and results (23) (Bhansali [4])

$$\begin{aligned} \mathsf{E}[\hat{\omega}_n - \omega] &= -\frac{2\omega}{n} + O(n^{-3/2}), \\ \mathsf{var}\,\hat{\omega}_n &= \frac{1 - \omega^2}{n} + O(n^{-3/2}) \end{aligned} \tag{23}$$

yield

$$\mathsf{E}[\hat{K}_m(x) - K_m(x)] = -\frac{2m\omega^m}{n} + \frac{m(m-1)(1-\omega^2)\omega^{m-2}}{2n} + O(n^{-3/2})$$

for $m \geq 2$.

It is interesting to note that predictor $\hat{K}_m(x)$ in the linear Gaussian model (22) is biased estimator of $K_m(x)$, however it is unbiased estimator of future value X_{t+m}, i. e. $\mathsf{E}[X_{t+m} - \hat{K}_m(X_t)] = 0$ (Fuller and Hasza [8]).

ACKNOWLEDGEMENT

The author thanks Professor Jiří Anděl for helpful comments on earlier drafts.

(Received January 27, 2004.)

REFERENCES

[1] J. Anděl: Non-negative autoregressive processes. J. Time Ser. Anal. *10* (1989), 1–11.

[2] J. Anděl: On least-squares and naive extrapolations in a non-linear AR(1) process. Test *6* (1997), 91–100.

[3] J. Anděl and V. Dupač: Extrapolations in non-linear autoregressive processes. Kybernetika *35* (1999), 383–389.

[4] R. J. Bhansali: Effects of not knowing the order of an autoregressive process on the mean squared error of prediction – I. J. Amer. Statist. Assoc. *76* (1981), 588–597.

[5] R. Bhattacharya and Ch. Lee: On geometric ergodicity of non-linear autoregressive models. Statist. Probab. Lett. *22* (1995), 311–315.

[6] D. B. H. Cline and H. Pu: Verifying irreducibility and continuity of a non-linear time series. Statist. Probab. Lett. *40* (1998), 139–148.
[7] S. Datta, G. Mathew, and W. P. McCormick: Nonlinear autoregression with positive innovations. Austral. & New Zealand J. Statist. *40* (1998), 229–239.
[8] W. A. Fuller and D. P. Hasza: Predictors for the first-order autoregressive process. J. Econometrics *13* (1980), 139–157.
[9] O. Lee: On probabilistic properties of non-linear ARMA(p, q) models. Statist. Probab. Lett. *46* (2000), 121–131.
[10] R. Moeanaddin and H. Tong: Numerical evaluation of distributions in non-linear autoregression. J. Time Ser. Anal. *11* (1990), 33–48.
[11] D. Tjøstheim: Non-linear time series and Markov chains. Adv. in Appl. Probab. *22* (1990), 587–611.
[12] H. Tong: Non-linear Time Series. Clarendon Press, Oxford 1990.

Petr Zvára, Charles University, Faculty of Mathematics and Physics, Department of Probability and Mathematical Statistics, Sokolovská 83, 186 75 Praha 8. Czech Republic.
e-mail: zvara@stemmark.cz