Multi-step prediction for nonlinear autoregressive models based on empirical distributions

Meihui Guo and Zhidong Bai

Department of Applied Mathematics

National Sun Yat-sen University

Kaohsiung, Taiwan

Hong Zhi An

Institute of Applied Mathematics

Academia Sinica, Beijing, China

Abstract A multi-step prediction procedure for nonlinear autoregressive (NLAR) models based on empirical distributions is proposed. Calculations involved in this prediction scheme are rather simple. It is shown that the proposed predictors are asymptotically equivalent to the exact least squares multi-step predictors, which are computable only when the innovation distribution has a simple known form. Simulation studies are conducted for two- and three-step predictors of two NLAR models.

Key words and phrases: empirical distribution; multi-step prediction; nonlinear autoregressive model.

1 Introduction

The nonlinear autoregressive (NLAR) model is a natural extension of the linear autoregressive (LAR) model and has attracted considerable interest in recent years. Important results on this model can be found in the literature; for example, Tong (1990), Tjϕstheim (1990) and Tiao and Tsay (1994). Most of the important results for the LAR model have been extended to the NLAR model, except for the multi-step prediction, which is one of the most important topics in time series analysis. This lack might be due to the difficulty or even impossibility of calculating the exact least squared multi-step predictors for NLAR models.

For linear models with martingale difference innovations, the least squares (linear) multi-step predictors will depend on the parameters of the models and the historical observations, but not on the innovation distributions (see Box and Jenkins, 1976). Therefore, when the parameters of the LAR models are known or good estimates are available, the multi-step predictors can be computed directly from explicit expressions and thus their limiting properties can be derived from those formulae. However, such is not the case for NLAR models. It will be shown that, under the assumption of i.i.d. innovations, the LAR model is the only time series model whose multi-step prediction is independent of the innovation distribution. This is the cause of the complexity of the multi-step prediction for nonlinear models. It is worth noting that the least squares predictors for more than one step ahead rely not only on the nonlinear regression function, but also on the innovation distribution of the model. Therefore, when the innovation distribution is unknown, it is impossible to calculate the exact least squares multi-step predictors for NLAR models.

Numerical and Monte Carlo methods are proposed in the literature to compute the least squares m-step ($m > 1$) predictors (see Pemberton, 1987 or Tong, p.346,

1990) when both the nonlinear regression function and the innovation distribution are known. However, the use of these approaches is not very realistic since the innovation distribution is unknown in most situations. The purpose of the present study is to propose an empirical scheme to evaluate the least squares multi-step predictors for the NLAR models and to investigate the asymptotics of the newly proposed predictors. We confine ourselves to the discussion of the multi-step prediction when the NLAR model is given but the innovation distribution is unspecified. In Section 2, we propose a simple multi-step prediction method for NLAR models by using the empirical distribution of the innovation. In Section 3, we show that the proposed prediction scheme is asymptotically equivalent to the exact least squares multi-step prediction. In Section 4, simulation results are presented for two- and three-step prediction of two kinds of NLAR models. In Section 5, further discussions and generalizations are given.

2 Prediction based on the empirical distributions

We first consider the following NLAR model of order one

$$x_t = \varphi(x_{t-1}) + \varepsilon_t, \quad t = 0, \pm 1, \pm 2, \cdots, \tag{2.1}$$

where ε_t's are i.i.d. random variables with $E\varepsilon_t = 0$, $E\varepsilon_t^2 < \infty$ and an unknown distribution F, and the innovation ε_t is independent of x_{t-1}. The time series x_t is assumed to be causal, that is, x_t is a function of $\epsilon_t, \epsilon_{t-1}, \cdots$. Throughout this paper, we assume that the regression function φ is known and a set of historical data $x_1, x_2, \cdots, x_n$ are available. It is not difficult to see that all the results of this paper remain true when the regression function φ is replaced by a consistent estimator. The latter is well studied in the literature (Lai and Zhu, 1991) and will not be considered here.

For one-step prediction, the least squares predictor of x_{n+1} with $x_1, x_2, \cdots, x_n$

being given, is

$$\hat{x}_{n+1} = E(x_{n+1}|x_n, \cdots, x_1) = E\{\varphi(x_n) + \varepsilon_{n+1}|x_n\} = \varphi(x_n). \tag{2.2}$$

Under model (2.1), the calculation of $\hat{x}_{n+1}$ is easy and is independent of the distribution of ε_{n+1}. This is an important property of one step ahead prediction for both linear and non-linear AR models.

However, for multi-step prediction, this is true only for linear models. Indeed, if φ is linear, say $\varphi(x) = ax + b$, then from (2.2), we have

$$\begin{aligned} \hat{x}_{n+2} &= \int [a(ax_n + b + \varepsilon) + b] dF(\varepsilon) \\ &= (a^2 x_n + ab + b) + a \int \varepsilon dF(\varepsilon) \\ &= (a^2 x_n + ab + b) = a\hat{x}_{n+1} + b. \end{aligned} \tag{2.3}$$

By induction, the l-step ahead predictor of the linear model can also be computed as follows

$$\hat{x}_{n+l} = a\hat{x}_{n+l-1} + b = \cdots = a^l x_n + a^{l-1} b + \cdots + ab + b. \tag{2.4}$$

It is seen that formula (2.3) does not require any knowledge of the distribution F, except for its first two moments.

For non-linear models, the situation becomes more complicated. For example, the two-step least squares predictor for model (2.1) is

$$\begin{aligned} \hat{x}_{n+2} &= E(x_{n+2}|x_n, \cdots, x_1) = E\{\varphi(x_{n+1}) + \varepsilon_{n+2}|x_n\} \\ &= E\{\varphi(\varphi(x_n) + \varepsilon_{n+1})|x_n\} = \int \varphi(\varphi(x_n) + \varepsilon) dF(\varepsilon), \end{aligned} \tag{2.5}$$

where $F(\cdot)$ is the distribution of ε_t. It is seen that the right hand side of (2.5) depends on F. A question is whether (2.5) can be further reduced to become not depending on F, as in (2.3). Unfortunately, the answer is negative. We will show in the following that linear models are the only models to possess the property that the multi-step prediction is independent of the innovation distribution.

In fact, if the two-step prediction is independent of F, then the right hand side of (2.5) should remain the same when F is the binary distribution at a and $-a$ with equal probabilities, for any x and $a \geq 0$. Thus, by taking $a > 0$ and $= 0$ respectively, we obtain,

$$\varphi(x+a) + \varphi(x-a) = 2\varphi(x).$$

This, together with the measurability of φ, implies that φ is a linear function.

The objective of this paper is to propose a prediction procedure when the innovation distribution is unknown. In most practical situations, a set of historical records $x_1, x_2, \cdots, x_n$ of the NLAR model (2.1) is available, and in general, the regression function φ is either known or can be well estimated. Thus, throughout this paper, we assume that φ is known. Hence, to compute the multi-step predictors, one needs a good estimator of the innovation distribution. Note that, when φ is given, the innovations $\{\varepsilon_t\}$ can be calculated exactly from the data set by the relation $\varepsilon_k = x_k - \varphi(x_{k-1}), \quad k = 2, 3, \cdots, n$. Therefore, the empirical distribution F_n of the innovations $\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n$ can be taken as an estimate of the innovation distribution, where

$$F_n(x) = \frac{1}{n-1} \sum_{k=2}^{n} I(\varepsilon_k < x), \tag{2.6}$$

and $I(\varepsilon_k < x)$ denotes the indicator function of the set $(\varepsilon_k < x)$. Thus, as an estimate of (2.5) we propose to use

$$\hat{x}^*_{n+2} = \frac{1}{n-1} \sum_{k=2}^{n} \varphi(\varphi(x_n) + \varepsilon_k) \tag{2.7}$$

as our two-step predictor.

The prediction procedure (2.7) can be easily extended to the l-step prediction for $l > 2$. For instance, for the third-step prediction, the exact least squares predictor is

$$\hat{x}_{n+3} = \int \int \varphi(\varphi(\varphi(x_n) + \varepsilon) + \varepsilon')dF(\varepsilon)dF(\varepsilon'). \tag{2.8}$$

Similar to (2.6), we introduce

$$\hat{x}^*_{n+3} = \frac{1}{(n-1)(n-2)} \sum\sum_{2\le k\ne j\le n} \varphi(\varphi(\varphi(x_n)+\varepsilon_k)+\varepsilon_j) \tag{2.9}$$

as the third-step predictor. In general, the exact l-step least squares predictor is

$$\hat{x}_{n+l} = \int \cdots \int \varphi(\varphi(\cdots(\varphi(x_n)+\varepsilon_1)+\cdots)+\varepsilon_{l-1}) dF(\varepsilon_1)\cdots dF(\varepsilon_{l-1})$$

and our proposed predictor is

$$\hat{x}^*_{n+l} = \frac{(n-l)!}{(n-1)!} \sum_{(l-1,n)} \varphi(\varphi(\cdots(\varphi(x_n)+\varepsilon_{i_1})+\cdots)+\varepsilon_{i_{l-1}}), \tag{2.10}$$

where the summation $\sum_{(l-1,n)}$ runs over all possible $(l-1)$-ples of distinct $i_1,\cdots,i_{l-1}$.

Finally, we briefly describe the conditional variance of x_{n+2} given $x_1,\cdots,x_n$, which is useful for interval prediction of x_{n+2}. Note that

$$Var(x_{n+2}|x_1,\cdots,x_n) = Var(x_{n+2}|x_n) = \int \varphi^2(\varphi(x_n)+\varepsilon)dF(\varepsilon) - \hat{x}^2_{n+2} + \int \epsilon^2 dF(\epsilon). \tag{2.11}$$

Following the same idea as in the construction of the predictors, we may replace the unknown F in (2.11) by the empirical distribution F_n and obtain the following estimator of $Var(x_{n+2}|x_n)$

$$Var^*(x_{n+2}|x_n) = \frac{1}{n-1}\sum_{k=2}^{n} \varphi^2(\varphi(x_n)+\varepsilon_k) - (\hat{x}^*_{n+2})^2 + \frac{1}{n-1}\sum_{k=2}^{n} \epsilon_k^2. \tag{2.12}$$

3 Asymptotic equivalence of $\hat{x}^*_{n+l}$ and $\hat{x}_{n+l}$

In this section, we show that the proposed l-step predictor $\hat{x}^*_{n+l}$ is asymptotically equivalent to the exact least squares predictor $\hat{x}_{n+l}$. Reviewing formulae (2.7), (2.9) and (2.10), and since the innovations ε_t $(t \le n)$ are correlated with the variable x_n, at first glance it seems implausible that the proposed multi-step predictors are asymptotically equivalent to the exact least squares predictors. However, we will prove in the following that it is really the case.

As an illustration, we first consider the LAR model, say $\varphi(x) = ax + b$. By (2.6) and (2.3),

$$\begin{aligned}\hat{x}^*_{n+2} &= \frac{1}{n-1}\sum_{k=2}^{n}\{a(ax_n + b + \varepsilon_k) + b\} \\ &= a^2 x_n + ab + b + \frac{1}{n-1}\sum_{k=2}^{n} a\varepsilon_k = \hat{x}_{n+2} + \frac{1}{n-1}\sum_{k=2}^{n} a\varepsilon_k. \qquad (3.1)\end{aligned}$$

It is trivial that $\hat{x}^*_{n+2} - \hat{x}_{n+2} \to 0$ a.s. by the strong law of large numbers. That is, the proposed predictor $\hat{x}^*_{n+2}$ is asymptotically equivalent to $\hat{x}_{n+2}$.

For the NLAR models, we can hardly find a formula similar to (3.1). Furthermore, as mentioned earlier, the variable x_n is dependent on the innovations $\varepsilon_2, \varepsilon_3, \cdots, \varepsilon_n$. Thus, the summands $\varphi(\varphi(x_n)+\varepsilon_k)$, $k = 2, 3, \cdots, n$, in (2.7) are no longer independent of each other. This reveals the difficulty in understanding the asymptotic equivalence of $\hat{x}^*_{n+2}$ and $\hat{x}_{n+2}$. Fortunately, the following lemma, which is interesting in its own right, enables us to establish the equivalence.

Lemma 3.1. *Suppose that $\{\varepsilon_1, \varepsilon_2, \ldots,\}$ is a sequence of iid. random variables and X is a random p-vector which may be dependent on $\{\varepsilon_1, \varepsilon_2, \ldots,\}$ in an unspecified manner. Let $\psi(x, e_1, \ldots, e_k)$ be a measurable function in R^{p+k} which is uniformly equi-continuous in x, that is, for each given $\eta > 0$, there is a constant $\delta > 0$ such that*

$$\sup_{e_1,\cdots,e_k} |\psi(x_1, e_1, \ldots, e_k) - \psi(x_2, e_1, \ldots, e_k)| < \eta, \qquad (3.2)$$

when $\|x_1 - x_2\| \le \delta$, where $\|\cdot\|$ can be any norm equivalent to the Euclidean norm. Assume that for each fixed x, $\mathrm{E}|\psi(x, \varepsilon_1, \ldots, \varepsilon_k)| < \infty$.
Then, with probability one,

$$\frac{(n-k)!}{n!}\sum_{(k,n)} \psi(x, \varepsilon_{i_1}, \ldots, \varepsilon_{i_k}) \to \int\ldots\int \psi(x, e_1, \ldots, e_k)F(de_1)\ldots F(de_k), \qquad (3.3)$$

where the summation $\sum_{(k,n)}$ is defined as in (2.10) and F is the common distribution of the random variables ε's.

Proof. Write

$$g_n(x) = \frac{(n-k)!}{n!} \sum_{(k,n)} \psi(x, \varepsilon_{i_1}, \ldots, \varepsilon_{i_k})$$

and

$$g(x) = \int \ldots \int \psi(x, e_1, \ldots, e_k) F(de_1) \ldots F(de_k).$$

For any given $\eta > 0$, by the assumption on ψ, there is a constant $\delta > 0$ such that (3.2) is true. For any large but fixed $M > 0$, we may split the p-dimensional ball $B(M) = \{x,\ \|x\| \le M\}$ into disjoint subsets A_i with $a_i \in A_i$, $i = 1, 2, \cdots, m$, such that for any point $x \in B(M)$, $\|x - a_i\| < \delta$ for some integer $i \le m$. For brevity, we shall also denote the indicator function $I(X \in A_i)$, $i = 1, 2, \ldots, m$ by A_i. For convenience, write $A_0 = I(X \notin B(M))$.

Then, by (3.2), we have

$$\sum_{i=1}^{m} A_i |g_n(X) - g_n(a_i)| < \eta$$

and

$$\sum_{i=1}^{m} A_i |g(X) - g(a_i)| < \eta.$$

Therefore,

$$|g_n(X) - g(X)| \le 2\eta + \infty \cdot A_0 + \sum_{i=1}^{m} A_i |g_n(a_i) - g(a_i)|.$$

Write $h_i(e_1, \ldots, e_k) = (k!)^{-1} \sum_{(\pi)} \psi(a_i, e_{\pi(1)}, \ldots, e_{\pi(k)})$, where the summation $\sum_{(\pi)}$ runs over all permutations of $\{1, 2, \ldots, k\}$. Then, $g_n(a_i)$ can be viewed as a U-statistic with the kernel h_i. Therefore, we have $g_n(a_i) \to g(a_i)$ *a.s.*, by the strong law of large numbers for U-statistics (see Chow and Teicher, 1988, p.389). Consequently, with probability one,

$$\limsup |g_n(X) - g(X)| \le 2\eta + \infty \cdot A_0.$$

Note that η can be chosen arbitrarily small, and M can be arbitrarily large. We finally obtain

$$\lim |g_n(X) - g(X)| = 0, \ \ a.s.$$

$\square$

Without loss of generality, we establish the following theorem for the two-step predictors.

Theorem 3.2. *Suppose that the function* φ *and the distribution of* ε_t *in model* (2.1) *satisfy the following conditions,*

(*i*) φ *is uniformly equi-continuous and for some constants* $0 < \rho < 1$, $M > 0$,

$$|\varphi(x)| \leq \rho |x|, \quad |x| \geq M;$$

(*ii*) *the innovations* ε_t *have zero mean, finite variance and a density which is positive everywhere.*

Then, we have

$$\lim_{n\to\infty} (\hat{x}^*_{n+2} - \hat{x}_{n+2}) = 0, \quad \textit{in probability} \tag{3.4}$$

and

$$\lim_{n\to\infty} (Var^*(x_{n+2}|x_n) - Var(x_{n+2}|x_n)) = 0, \quad \textit{in probability.} \tag{3.5}$$

Proof. By An and Huang (1996), under conditions (i) and (ii), model (2.1) has a unique stationary solution which is geometrically ergodic. Then, by Chen and An (1997), the condition $E\varepsilon_t^2 < \infty$ implies $Ex_t^2 < \infty$. Hence, by condition (i), $E\varphi^2(x_t) < \infty$ and $E\varphi^2(\varphi(x_n) + \varepsilon) < \infty$. By stationarity, to prove (3.4) it is sufficient to show that

$$\frac{1}{n}\sum_{i=1}^{n} \varphi(\varphi(x) + \varepsilon_i) \to \int \varphi(\varphi(x) + \varepsilon) dF(\varepsilon), \quad \text{as } n \to \infty, \tag{3.6}$$

in probability. In fact, (3.6) is a direct consequence of Lemma 3.1 by taking $k = 1$ and $\psi(x, e) = \varphi(\varphi(x) + e)$. Conclusion (3.5) can be proved similarly. $\square$

It is straightforward to establish the asymptotic equivalence of the proposed l-step predictor and the exact least squares predictors along the same lines as in the proof of Theorem 3.2. Moreover, the restrictions on φ and the distribution of ε_t can be relaxed to a certain extent. This is discussed in Remarks 3.1 and 3.2. Stronger versions of Theorem 3.2 are conjectured in Remark 3.3.

Remark 3.1. The positiveness of the innovation density is used only to guarantee the existence of a stationary solution to model (2.1). Therefore, if the function φ is bounded, say $|\varphi(x)| \leq K$, then the innovation density can be weakened to be positive on the interval $(-K, K)$, see An and Huang (1996).

Remark 3.2. An important time series model is the *Threshold Autoregressive* (TAR) model (see Tong 1990, Chan 1993), for which, unfortunately, the φ function is discontinuous and hence does not satisfy the continuity condition of Theorem 3.2. However, this continuity condition can be weakened to (3.2′) below to cover the TAR case.

For each $\eta > 0$ and each compact sub-set $E \subset R^k$, there exists a constant $\delta > 0$ such that for any $x \in R^p$, there are two points a_1 and a_2 in the ball $B(x, \delta) = \{y,\ \|y - x\| \leq \delta\}$ such that for any $y \in B(x, \delta)$ and any $(e_1, \cdots, e_k) \in E$,

$$|\varphi(y, e_1, \cdots, e_k) - \varphi(a_j, e_1, \cdots, e_k)| < \eta, \tag{3.2′}$$

is true either for $j = 1$ or 2.

Remark 3.3. In the linear case, the convergence of $\hat{x}^*_{n+2} - \hat{x}_{n+2}$ is in the strong version. It is natural to conjecture that (3.4) can be strengthened to the a.s. version. Also, when φ is linear, by the iid assumption and $E\varepsilon_t^2 = \sigma^2 < \infty$, from (3.1) we know that $\sqrt{n}(\hat{x}^*_{n+2} - \hat{x}_{n+2}) \to N(0, a^2\sigma^2)$. This implies that $\hat{x}^*_{n+2} - \hat{x}_{n+2} = O_p(1/\sqrt{n})$. We therefore conjecture that this should still be true for the nonlinear case under the same conditions of Theorem 3.2. It might also be of interest to establish the asymptotic normality, under even stronger conditions than those proposed in Theorem 3.2.

4 Simulation Results

In this section, two examples of NLAR models are discussed and simulated to compare the proposed two- and thee-step predictors with the exact least squares predictors.

The first NLAR model used in the simulation is

$$x_t = \frac{-x_{t-1}}{1 + x_{t-1}^2} + \varepsilon_t, \quad t = 1, 2, \cdots, n, \tag{4.1}$$

with the autoregressive function $\varphi(x) = -x/(1 + x^2)$.

If $\{\varepsilon_t\}$ is uniformly distributed on the interval $[-1, 1]$, then, by (2.3), the exact least squares predictor for the second step is given by

$$\begin{aligned} \hat{x}_{n+2} &= \int \varphi(\varphi(x_n) + \varepsilon) dF(\varepsilon) \quad (\text{set} \quad c = \varphi(x_n)) \\ &= \frac{1}{2} \int_{-1}^{1} \varphi(c + \varepsilon) d\varepsilon = \frac{1}{2} \int_{c-1}^{c+1} \varphi(x) dx \\ &= \frac{1}{2} \int_{c-1}^{c+1} \frac{-x}{1 + x^2} dx = \frac{1}{4} \log \frac{1 + [\varphi(x_n) - 1]^2}{1 + [\varphi(x_n) + 1]^2}. \end{aligned} \tag{4.2}$$

The second NLAR model used in the simulation is the threshold NLAR model:

$$x_t = \begin{cases} 1 - 0.5x_{t-1} + \varepsilon_t, & x_{t-1} < 0; \\ 0.5x_{t-1} + \varepsilon_t, & x_{t-1} \geq 0. \end{cases} \tag{4.3}$$

If ε_t has the double exponential density $f(\varepsilon) = \frac{1}{2} e^{-|\varepsilon|}$, then by (2.3) and further calculations, we have

$$\hat{x}_{n+2} = \begin{cases} 1 - \varphi(x_n)/2, & \text{if} \quad \varphi(x_n) < 0; \\ \varphi(x_n)/2 + e^{-\varphi(x_n)}, & \text{if} \quad \varphi(x_n) \geq 0. \end{cases} \tag{4.4}$$

The analytic form of the higher multi-step predictors for models (4.1) and (4.3) is rather involved and thus difficult to present.

In the following we will compare $\hat{x}_{n+k}$ with $\hat{x}_{n+k}^*$, $k = 2, 3$ for NLAR models (4.1) and (4.3). First we generate $x_1, x_2, \cdots, x_n$ from model (4.1) with uniform innovation and from model (4.3) with double exponential innovation, respectively. The exact two-step predictors $\hat{x}_{n+2}$ for models (4.1) and (4.3) are calculated directly by formulae

(4.2) and (4.4), respectively, while the three-step predictors $\hat{x}_{n+3}$ is calculated by using formulae (2.8) via numerical integration. Whereas, the proposed predictors $\hat{x}^*_{n+2}$ and $\hat{x}^*_{n+3}$ are computed by (2.7) and (2.9), respectively. The sample sizes (n) in our simulation studies are chosen as 100, 200 and 400 with 1,000 replications. For the ith replication, $i = 1, \cdots, 1000$, let $d_{k,i} = \hat{x}_{n+k,i} - \hat{x}^*_{n+k,i}$ for $k = 2, 3$. Let $\bar{d}_k = \frac{1}{1000}\sum_{i=1}^{1000} d_{k,i}$ be the sample mean of the k-th step prediction errors and MSE_k be the sample mean squared error of $\hat{x}^*_{n+k}$ for k =2,3, respectively.

In the third and fourth columns of Table 1 and Table 2, we list $\bar{d}_k$ and MSE_k for $k = 2, 3$ and n =100, 200 and 400 for models (4.1) and (4.3), respectively. Furthermore, in order to investigate the effect of assuming false innovation density, we also generate n observations from model (4.1) when $\{\varepsilon_t\}$ has the following exponential mixture density:

$$f(x) = \frac{7}{8}(7e^{7x}I_{[x\leq 0]}) + \frac{1}{8}(e^{x}I_{[x>0]}).$$

First, we compute the exact predictors $\hat{x}_{n+k}(k = 2, 3)$ from (2.5) and (2.8) by numerical integration and the proposed predictors $\hat{x}^*_{n+k}$ as before. The simulation results are listed on the fifth and sixth columns of Table 1. Next, we calculate the pseudo k-step predictors $\widehat{wnp}^{(1)}_{n+k,i}$ with a false density of $N(0, \frac{2}{7})$ at the i-th repetition , $k = 2, 3$ and $i = 1, 2, \cdots, 1000$, from (2.5) and (2.8) via numerical integration. Let $d^{(1)}_{wnp_{n+k,i}} = \hat{x}_{n+k,i} - \widehat{wnp}^{(1)}_{n+k,i}$, $k = 2, 3$ and $i = 1, 2, \cdots, 1000$ and $\bar{d}^{(1)}_{wnp}$ be its sample mean and $MSE^{(1)}_k$ be the sample mean squared error of $\widehat{wnp}^{(1)}_{n+k}$. The results are listed on the 7th and 8th columns of Table 1. Similarly, for model (4.2) with double exponential innovation, we compute the pseudo k-th step predictors at the i-th iteration $\widehat{wnp}^{(2)}_{n+k,i}$, $k = 2, 3$, and $i = 1, 2, \cdots, 1000$, from (2.2) and (2.8) with a false density of $N(0, 2)$, and compute $\bar{d}^{(2)}_{wnp}$ (the sample mean of the prediction errors of $\widehat{wnp}^{(2)}_{n+k}$'s) and $MSE^{(2)}_k$ (the sample mean squared error of $\widehat{wnp}^{(2)}_{n+k}$'s). The results are listed on column 5 and 6 of Table 2.

The results in Table 1 and 2 reveal the following phenomena.

(1) From columns 3 and 5 of Table 1 and column 3 of Table 2, the sample means of the two- and three-step prediction errors decrease as sample size n increases. This shows that the proposed predictors are asymptotically equivalent to the exact least squares predictors.

(2) From columns 4 and 6 in Table 1 and column 4 in Table 2, the two- and three-step mean squared errors of our predictors are approximately inverse-proportional to the sample size n. This result is in accordance with our conjecture in Remark 3.3 that $\hat{x}^*_{n+2} - \hat{x}_{n+2} = O_p(\frac{1}{\sqrt{n}})$. Furthermore, the normal probability plot of $d_{3,i}$ (see Fig. 1) for model (4.1) with uniform innovation and n=400 strongly supports our asymptotic normality conjecture.

(3) The accuracy of the proposed predictor is irrelevant to the knowledge of the innovation density. However, the accuracy of $\hat{x}_{n+2}$ and $\hat{x}_{n+3}$ computed from (2.5) and (2.8) was greatly influenced by the correctness of the innovation distribution $F(x)$ (see columns 7 and 8 in Table 1 and columns 5 and 6 in Table 2). For example, for model (4.1) with exponential mixture innovation, $\bar{d}_3 = 10^{-3}\bar{d}^{(1)}_{wnp_2}$ and $MSE_3 \doteq 0.05 MSE_3^{(1)}$ when $n = 400$.

5 Further discussions

An advantage of the multi-step prediction scheme proposed in this paper is its simplicity of calculation. Although a disadvantage is to assume that the autoregressive function φ of the NLAR model is known, this disadvantage can be easily overcome by replacing the unknown autoregressive function φ with a good estimator. It is easy to see from the proof of Theorem 3.2 that this replacement does not alter the asymptotic equivalence provided the estimator of the autoregressive function is consistent. An example of such replacement can be found in Lai and Zhu (1991).

For simplicity, in previous sections we only discussed the approximation of the multi-step prediction for NLAR models of order one. The approach can be easily extended to higher orders. For example, consider the following NLAR model of order p:

$$x_t = \varphi(x_{t-1}, \cdots, x_{n-p}) + \varepsilon_t. \tag{5.1}$$

When φ is given and $x_1, x_2, \cdots, x_n$ are available,

$$\varepsilon_t = x_t - \varphi(x_{t-1}, \cdots, x_{n-p}), t = p+1, \cdots, n.$$

Consequently, the two-step predictor $\hat{x}_{n+2}$ can be approximated by

$$\hat{x}^*_{n+2} = \frac{1}{n-p} \sum_{k=p+1}^{n} \varphi(\varphi(x_n, x_{n-1}, \cdots, x_{n-p+1}) + \varepsilon_k, x_n, \cdots, x_{n-p+2}). \tag{5.2}$$

The multi-step predictors can be approximated in a similar manner, and one can easily establish the asymptotic equivalence of the proposed and exact predictors. The details are omitted.

Another direction of generalization is to consider NLAR models with conditional heteroscedasticity (see Chen and An, 1997). As an example, we herewith consider the following model

$$x_t = \varphi(x_{t-1}) + \varepsilon_t \sigma(x_{t-1}), \quad t = 1, 2, \cdots, n, \tag{5.3}$$

where the functions $\varphi(\cdot)$ and $\sigma(\cdot)$ are known, ε_t's are assumed to satisfy $E\varepsilon_t^2 = 1$ for the identifiability of model (5.3). When $x_1, x_2, \cdots, x_n$ are obtained,

$$\varepsilon_t = \{x_t - \varphi(x_{t-1})\}/\sigma(x_{t-1}), t = 2, 3, \cdots, n.$$

Hence, the two-step predictor can be obtained by employing the empirical distribution of ε_t, i.e.

$$\hat{x}^*_{n+2} = \frac{1}{n-1} \sum_{k=2}^{n} \varphi(\varphi(x_n) + \varepsilon_k \sigma(x_n)). \tag{5.4}$$

Finally, we emphasize that this approach can be adapted to the so-called autoregressive conditional heteroscedasticity (ARCH) models (see Engle, 1982), an example of which is as follows

$$\begin{cases} x_t &= \varphi(x_{t-1}) + \varepsilon_t h_t^{1/2} \\ h_t &= \alpha_0 + \alpha_1\{x_{t-1} - \varphi(x_{t-2})\}^2, \end{cases} \tag{5.5}$$

where the ε_t's are assumed to be the same as in model (5.3). Substituting the second equation into the first in (5.5), we have

$$x_t = \varphi(x_{t-1}) + \varepsilon_t\{\alpha_0 + \alpha_1[x_{t-1} - \varphi(x_{t-2})]^2\}^{\frac{1}{2}}, \tag{5.6}$$

which turns out to be an NLAR model with conditional heteroscedasticity

$$\sigma^2(x_{t-1}, x_{t-2}) = \alpha_0 + \alpha_1\{x_{t-1} - \varphi(x_{t-2})\}^2.$$

Therefore, the multi-step predictors based on the empirical distribution of ε_t's can be obtained easily.

Fig 1. Normal Probability plot, n=400

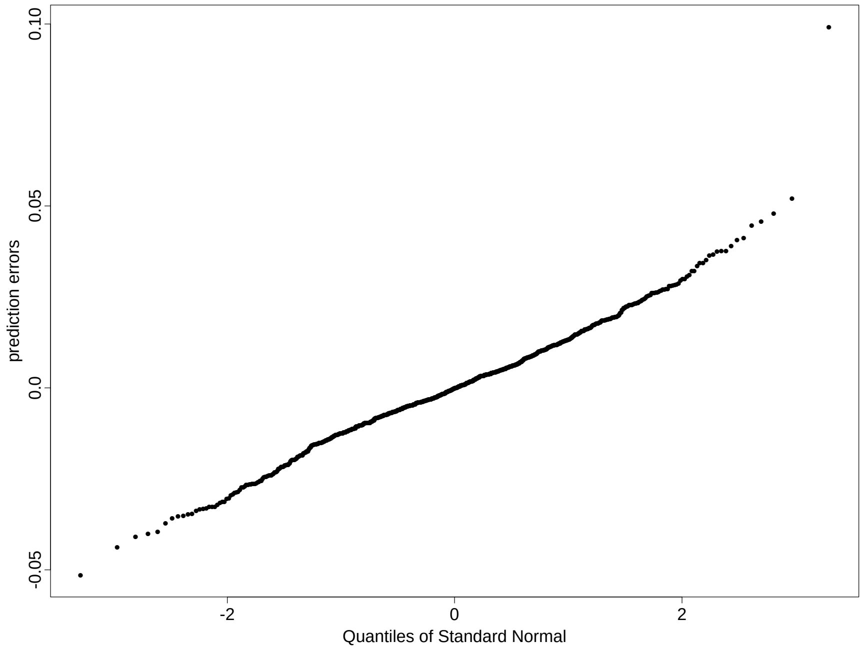

Table 1. Compare $\hat{x}_{n+k}$ with $\hat{x}^*_{n+k}$ for model (4.1)

n		$\varepsilon_t \sim U[-1,1]$		$\varepsilon_t \sim$ Exponential Mixture			
						Pseudo Predictor	
		$\bar{d}_k$	MSE_k	$\bar{d}_k$	MSE_k	$\bar{d}^{(1)}_{wnp_k}$	$MSE^{(1)}_k$
100	2nd step	10.9×10^{-4}	11.3×10^{-4}	13.7×10^{-4}	3.21×10^{-4}	-4.34×10^{-2}	5.32×10^{-3}
	3rd step	9.52×10^{-4}	3.34×10^{-4}	3.13×10^{-4}	1.46×10^{-4}	-2.56×10^{-2}	7.18×10^{-3}
200	2nd step	3.80×10^{-4}	5.71×10^{-4}	9.49×10^{-4}	1.72×10^{-4}	-4.38×10^{-2}	5.27×10^{-3}
	3rd step	6.11×10^{-4}	1.66×10^{-4}	-1.18×10^{-4}	0.73×10^{-4}	-2.60×10^{-2}	7.25×10^{-3}
400	2nd step	1.05×10^{-4}	2.67×10^{-4}	2.86×10^{-4}	0.81×10^{-4}	-4.45×10^{-2}	5.32×10^{-3}
	3rd step	0.58×10^{-4}	0.82×10^{-4}	-0.27×10^{-4}	0.34×10^{-4}	-2.59×10^{-2}	7.26×10^{-3}

Table 2. Compare $\hat{x}_{n+k}$ with $\hat{x}^*_{n+k}$ for model (4.3)

n		$\varepsilon_t \sim$ Double Exponential			
				Pseudo Predictor	
		$\bar{d}_k$	MSE_k	$\bar{d}^{(2)}_{wnp_k}$	$MSE^{(2)}_k$
100	2nd step	6.99×10^{-4}	41.3×10^{-4}	6.64×10^{-2}	10.1×10^{-3}
	3rd step	5.71×10^{-4}	40.5×10^{-4}	5.99×10^{-2}	7.66×10^{-3}
200	2nd step	-6.85×10^{-4}	20.9×10^{-4}	6.50×10^{-2}	7.83×10^{-3}
	3rd step	-3.99×10^{-4}	21.2×10^{-4}	5.89×10^{-2}	5.62×10^{-3}
400	2nd step	1.05×10^{-4}	9.48×10^{-4}	6.58×10^{-2}	6.65×10^{-3}
	3rd step	-0.48×10^{-4}	9.71×10^{-4}	5.92×10^{-2}	4.52×10^{-3}

REFERENCES

An, H.Z. and F. C. Huang (1996). The geometrical ergodicity of nonlinear autoregressive models. *Statistica Sinica* **4**, 943-956.

Box, G. E. P. and G. E. Jenkins (1976). *Time Series Analysis, Forecasting and Control.* Holden-Day, San Franciso.

Chan, K. S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *Ann. Statist.* **1**, 520-533.

Chen, M. and H. Z. An (1997). The probabilistic properties of the non-linear autoregressive model with conditional heteroscedasticity. To appear in *ACTA. Math. Appl. Sinica*

Chow, Y. S. and Teicher, H. (1988). *Probability Theory : independence, interchangeability, martingales.* Second ed., Springer Verlag, New York.

Engle, R. E. (1982). Autoregressive heteroscedasticity with estimates of the variance of U. K. inflation. *Econometrica* **50**, 987-1008.

Lai, Z. L. and G. Zhu. (1991). Adaptive prediction in non-linear autoregressive models and control systems. *Statistica Sinica* **1**, 309-334.

Pemberton, J. (1987). Exact least squares multi-step prediction from non-linear autoregressive models. *J. Time Ser. Anal.* **4**, 443-448.

Tiao, G. C. and R. S. Tsay. (1994). Some advance in non-linear and adaptive modeling in time series. *J. of Forecasting* **13**, 109-131.

Tjϕstheim, D. (1990). Non-linear time series and Markov chains. *Adv. Appl. Probab.* **22**, 587-611.

Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach.* Oxford Univ. Press, New York.