This is a repository copy of *A New Direct Approach of Computing Multi-Step Ahead Predictions for Nonlinear Models*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/83230/

Monograph:
Lee, K.L. and Billings, S.A. (2002) A New Direct Approach of Computing Multi-Step Ahead Predictions for Nonlinear Models. Research Report. ACSE Research Report 457 .
Department of Automatic Control and Systems Engineering

A New Direct Approach of Computing

Multi-step Ahead Predictions for

Nonlinear Models

K.L. Lee   S.A. Billings

Department of Automatic Control and Systems Engineering

University of Sheffield

Mappin Street, Sheffield S1 3JD

United Kingdom

University of Sheffield

A New Direct Approach of Computing Multi-step Ahead Predictions for Nonlinear Models

K.L. Lee and S.A. Billings

Abstract: A new direct approach of computing multi-step ahead predictions for nonlinear time series is introduced. The covariance of the parameter estimates associated with, and the mean squared k-step ahead prediction errors of the new direct approach are smaller than those obtained using the conventional direct approach. Numerical examples are included to illustrate the application of the new direct approach.

1. Introduction

Most prediction algorithms are based on a minimisation of the mean squared one-step ahead prediction errors. When multi-step ahead predictions are required, the linear iterative approach of computing multi-step ahead predictions is often employed, even though the identified models may be nonlinear and the obtained multi-step ahead predictions may not be optimal. However obtaining accurate multi-step ahead predictions is very important in many problems. The task of obtaining optimal one-step ahead predictions is often not trivial, but computing optimal multi-step ahead predictions poses a more challenging problem.

For linear time series models with white noise innovations, the least squares multi-step ahead model depends only on the parameters of the one-step ahead model and past observations. Hence the optimal least squares multi-step ahead predictions can be computed easily using the Box-Jenkins iterative approach if the parameters or good estimates of the parameters of the one-step ahead model are available (Box et al 1994). Unfortunately, the linear model is the only model where the multi-step ahead predictions are independent of the noise distribution (Guo et. al. 1999). To compute the optimal multi-step ahead predictions for a nonlinear model requires a knowledge of the noise probability distribution or an estimate of this and the use of iterated numerical integration, assuming that the noise probability distribution is continuous (Pemberton 1987, Tong 1990). Iterated numerical integration becomes computationally very expensive for many step ahead predictions and complex nonlinear models. Therefore sub-optimal approaches of computing multi-step ahead predictions for nonlinear models have been developed (Gabr and Rao 1981, Findley 1985, Al-Qassem and Lane 1987, Cloarec and Ringwood 1998, Parlos et al. 2000). The two most commonly used approaches are the iterative and the direct approaches.

Atiya et al. (1999) and Zhang et al. (1998) argued that the direct approach could give a better prediction than the iterative approach for nonlinear time series. However Weigend et al. (1992) reported that the direct approach performs significantly worse than the iterative approach for the sunspot data set. In this paper the direct approach will be investigated in detail.

It will be shown that the direct k-step ahead prediction errors are auto-correlated, even if the optimal k-step ahead model is used to obtain the predictions. This can result in a large covariance of the parameter estimates associated with the direct k-step ahead model. It will also be shown that the prediction accuracy of the direct model is directly related to the covariance of the parameter estimates. Therefore the prediction accuracy of the direct approach can be improved by reducing the covariance of the parameter estimates associated with the direct k-step ahead model. Hence a new direct approach will be introduced. The new direct approach can produce parameter estimates with a smaller covariance, and hence a higher prediction accuracy can be expected compared to the direct approach.

The paper is organized as follows. Section 2 briefly introduces basic concepts, algorithms, and the optimal and sub-optimal approaches for computing multi-step ahead predictions for a general Nonlinear AutoRegressive (NAR) model. Detailed analysis of the direct approach is performed in Section 3. The new direct approach is introduced and discussed in Section 4. Simulation results are presented in Section 5 and finally conclusions are given in Section 6.

2. Preliminaries

2.1 Identification

It is well known that input-output descriptions that expand the current output in terms of past inputs and outputs provide models that represent a broad class of nonlinear stochastic control systems, and a unified representation called the Nonlinear AutoRegressive Moving Average model with eXogenous inputs (NARMAX) has been derived under some mild assumptions (Leontaritis and Billings 1985 a, b). This approach has been adopted for time series modelling (Chen and Billings 1989a) and leads to the following general NARMA model

$$y_t = f\left(y_{t-1}, \ldots, y_{t-n_y}, e_{t-1}, \ldots, e_{t-n_e}\right) + e_t \qquad (1)$$

where y_t is the output signal, e_t is an unobserved and unpredictable zero mean noise sequence, $f(\bullet)$ is some nonlinear function and n_y and n_e are the maximum output and noise lags respectively. The NARMA model is about as far as one can go in terms of specifying a general

2

finite dimensional nonlinear relationship (Chen and Billings 1989a). In some approximations, it may be possible to reduce the NARMA model to a simplified form called the NAR model,

$$y_t = f\left(y_{t-1}, \dots, y_{t-n_y}\right) + e_t \tag{2}$$

The model in eqn (2) is obviously less general than the model in eqn (1) but the identification schemes and system analysis of the NAR model are simpler.

The functional form $f(\bullet)$ for a real world system is generally very complex and is usually unknown. Any practical modelling must be based on a chosen model set of known functions. In addition, an efficient identification procedure must be available to select a parsimonious model structure because the dimension of a nonlinear model can easily become extremely large. In this study, linear-in-the-parameters models are considered because rich linear identification techniques can readily be applied to this class of models. A general linear-in-the-parameters model takes the form

$$y_t = \sum_{i=1}^{n_m} f_i\left(\mathbf{x}_{t-1}\right)\theta_i + e_t \tag{3}$$

where θ_i are the parameters, n_m is the number of terms in the model, $f_i\left(\mathbf{x}_{t-1}\right)$ is a linear or nonlinear model term which depends only on $\mathbf{x}_{t-1}$, and $\mathbf{x}_{t-1}$ is the vector

$$\mathbf{x}_{t-1} = \left(y_{t-1}, \dots, y_{t-n_y}, e_{t-1}, \dots, e_{t-n_e}\right)^T \tag{4}$$

where the superscript T denotes transpose. The polynomial model (Chen and Billings 1989a) and the extended model set (Billings and Chen 1989) are examples of linear-in-the-parameters models. For linear-in-the-parameters models, the parameters can be estimated easily using least squares type algorithms. The following cost function

$$J = \sum_{t=1}^{N} e_t^2 = \sum_{t=1}^{N}\left(y_t - \sum_{i=1}^{n_m} f_i\left(\mathbf{x}_{t-1}\right)\theta_i\right)^2 \tag{5}$$

where N is the length of the time series, is minimised to estimate the unknown parameters θ_i. Radial Basis Function (RBF) models are 'universal approximators' and can also be configured to be linear-in-the-parameter models by various methods. In this study the nonlinear function $f_i(\bullet)$ in eqn (3) will be assumed to be a thin-plate-spline radial basis function so that

$$\sum_{i=1}^{n_m} f_i\left(\mathbf{x}_{t-1}\right)\theta_i = \sum_{i=1}^{n_m} \phi\left(\|\mathbf{x}_{t-1} - \mathbf{c}_i\|\right)\theta_i \tag{6}$$

where $\|\bullet\|$ denotes the Euclidean norm, $\phi(v) = v^2 \log(v)$, and $\mathbf{c}_i$ are the centres of the radial basis function. The well-known and efficient forward regression orthogonal least squares

3

algorithm (Chen et al 1989, Chen et al 1991) will be employed to identify the models by selecting n_m important centres from all the data and estimating the associated parameters.

2.2 Identification of NARMA Models

The system noise e_t in eqn (1) is generally unobserved and will be replaced by the prediction error or residual sequence ξ_t, $\mathbf{x}_{t-1}$ in eqn (4) will then be modified accordingly

$$\mathbf{x}_{t-1} = \left(y_{t-1}, \ldots, y_{t-n_y}, \xi_{t-1}, \ldots, \xi_{t-n_\xi} \right)^T \tag{7}$$

Initially, the values of the residuals ξ_t are not available and are set to zero. Then the forward regression orthogonal least squares algorithm (Chen et al. 1989) is used to identify the model and the initial residual sequence $\xi_t^{(0)}$ is calculated. An iterative loop is then entered to update the vector $\mathbf{x}_{t-1}^{(k)}$ such that

$$\mathbf{x}_{t-1}^{(k)} = \left(y_{t-1}, \ldots, y_{t-n_y}, \xi_{t-1}^{(k-1)}, \ldots, \xi_{t-n_\xi}^{(k-1)} \right)^T \tag{8}$$

and the forward regression orthogonal least squares algorithm is used to select and update the model. This iterative procedure is terminated when the residuals ξ_t converge, and has been successfully employed to identify NARMA models with polynomial functions (Chen et al. 1989). Unfortunately, it was found in the current study that when a thin-plate-spline radial basis function (or the Gaussian function) was employed to approximate $f(\bullet)$ in eqn (1), in many cases the residuals ξ_t, unlike the polynomial case, do not converge. This may be related to m-invertibility, see Chen and Billings (1989b). Therefore an alternative approach of identifying the NARMA model was employed in this study, on the assumption that the nonlinear time series is invertible. A NAR model with sufficiently long maximum output lag is identified from the data to give

$$y_t = \hat{f}_{nar} \left(y_{t-1}, y_{t-2}, \ldots, y_{t-n_y} \right) + \hat{\xi}_t \tag{9}$$

where $\hat{f}_{nar}(\bullet)$ is the identified NAR model and $\hat{\xi}_t$ is the calculated residual sequence. The unknown residuals in eqn (7) are then replaced by the calculated residuals $\hat{\xi}_t$ to give

$$\mathbf{x}_{t-1} = \left(y_{t-1}, \ldots, y_{t-n_y}, \hat{\xi}_{t-1}, \ldots, \hat{\xi}_{t-n_\xi} \right)^T \tag{10}$$

and hence a NARMA model can then be identified using

$$y_t = \sum_{i=1}^{n_m} f_i(\mathbf{x}_{t-1}) \theta_i + e_t \tag{11}$$

This approach was first adopted by Durbin (1959) to identify linear ARMA model, and was recently investigated by Broersen (2000). The advantage of this approach is that no iteration is required to identify the moving average terms. This approach will be employed to identify NARMA models in this study.

2.3 Model Validation

Once a model has been identified the model should be validated. Billings and Zhu (1994) developed statistical validation methods which check that the residuals, which are computed as

$$\xi_t = y_t - \sum_{i=1}^{n_m} f_i(\mathbf{x}_{t-1})\hat{\theta}_i \tag{12}$$

where $\hat{\theta}_i$ are the estimated parameters, are unpredictable. The model validity tests are defined as

$$\rho_{\xi\xi}(\tau_\rho) = \frac{\sum_{t=1}^{N-\tau_\rho} \xi_t \xi_{t+\tau_\rho}}{\sum_{t=1}^{N} \xi_t^2} = \delta(\tau_\rho) \tag{13}$$

$$\rho_{[y\xi][\xi\xi]}(\tau_\rho) = \frac{\sum_{t=1}^{N-\tau_\rho} [y_t\xi_t - \overline{y_t\xi_t}][\xi_{t+\tau_\rho}^2 - \overline{\xi_t^2}]}{\sqrt{\sum_{t=1}^{N}[y_t\xi_t - \overline{y_t\xi_t}]^2}\sqrt{\sum_{t=1}^{N}[\xi_t^2 - \overline{\xi_t^2}]^2}} = k_\rho \delta(\tau_\rho) \tag{14}$$

where k_ρ is a constant, δ denotes the Kronecker delta and overbar denotes time average. For sufficiently long data sets the 95% confidence intervals are approximately $\pm 1.96/\sqrt{N}$.

2.4 Optimal and Sub-Optimal Multi-Step Ahead Prediction Methods

For ease of discussion initially the time series will be assumed to be generated by a very simple Nonlinear AutoRegressive (NAR) model of order 1 in this section.

$$y_t = f(y_{t-1}) + e_t \tag{15}$$

2.4.1 Optimal Multi-Step Ahead Predictions

The optimal least squares k-step ahead prediction $y_{t+k/t}^{opt}$, given information up to and including time t is known to be

$$y_{t+k/t}^{opt} = E[y_{t+k} / y_t, y_{t-1}, y_{t-2},...] \tag{16}$$

where $E[a/b]$ denotes the conditional expectation of a given b.

The optimal one-step ahead prediction can easily be obtained as

$$y_{t+1/t}^{opt} = E_t\left[f(y_t) + e_{t+1}\right]$$

$$= f(y_t) \tag{17}$$

where $E_t[\bullet]$ denotes $E[\bullet \, / \, y_t, y_{t-1}, ...]$.

The optimal two-step ahead prediction would be

$$y_{t+2/t}^{opt} = E_t\left[y_{t+2}\right]$$

$$= E_t\left[f(f(y_t) + e_{t+1}) + e_{t+2}\right]$$

$$= \int f(f(y_t) + e)dF_n(e) \tag{18}$$

where $F_n(\bullet)$ is the probability distribution of the noise e_t, which must either be known or which has to be approximated before the integral can be solved, either analytically or in most cases numerically. Pemberton (1987) derived a simple recurrence relationship for the optimal multi-step ahead prediction functions for a general NAR model. Assuming that the noise has a known distribution such as a zero mean normal distribution, Pemberton (1987) showed that the solution can be computed using numerical integration techniques. Unfortunately this method becomes computationally very expensive for many step ahead predictions and high order models. Therefore sub-optimal methods are usually employed to compute multi-step ahead predictions for nonlinear models. The two most commonly used approaches of computing multi-step ahead predictions, the iterative and the direct approaches, will be discussed below.

2.4.2 Iterative Approach

In this approach the one-step ahead model is constructed by minimizing the mean squared one-step ahead prediction errors and the identified model is then used to recursively forecast k-steps ahead. At any intermediate step, the model will use some of the forecasts obtained at previous steps as inputs. For example the two-step ahead prediction, using the iterative approach, $\hat{y}_{t+2/t}^{ite}$ for the NAR(1) model would be

$$\hat{y}_{t+2/t}^{ite} = f\left(\hat{y}_{t+1/t}^{ite}\right) = f(f(y_t)) \tag{19}$$

where $(\bullet)^{ite}$ indicates that $(\bullet)$ is computed using the iterative approach. The prediction is obtained as though the error e_{t+1} in eqn (18) is set to the mean value of zero. The iterative approach will usually be biased for nonlinear models because the expected value of a function is usually not equal to the function of the expected value $E[f(\bullet)] \neq f(E[\bullet])$. Brown and Mariano (1989) show that the bias associated with the iterative approach will not go to zero for large

sample sizes, even if the nonlinear model $f(\bullet)$ is known exactly. The advantages of the iterative approach however, are that the method is very easy to implement and fast to use.

2.4.3 Direct Approach

In the direct approach a direct k-step ahead model is constructed to predict k-steps ahead directly. Therefore the two-step ahead prediction using the direct approach $\hat{y}_{t+2/t}^{dir}$ for the NAR(1) model would be

$$\hat{y}_{t+2/t}^{dir} = f_2^{dir}(y_t) \tag{20}$$

where $(\bullet)^{dir}$ indicates that $(\bullet)$ is computed using the direct approach and $f_2^{dir}(\bullet)$ is the identified two-step ahead model constructed by minimizing the direct two-step ahead mean squared prediction errors. Unlike the iterative approach where a single one-step ahead model is used repeatedly to compute many step ahead forecasts, in the direct approach a direct k-step ahead model is identified for each prediction time step. Potential problems with the direct approach are that the direct k-step ahead prediction function may be very complex and hence difficult to identify for large k. Also the k-step ahead prediction errors are usually auto-correlated and this may cause problems in the estimation of the parameters which may become biased.

3. Detailed Analysis of the Direct Approach

For multi-step ahead predictions for nonlinear time series, the direct approach can in most cases yield a better prediction performance than the iterative approach for nonlinear models, see Stoica and Nehorai (1989), Lin and Granger (1994), Zhang et al (1998) and Atiya et al (1999). Therefore the direct approach of computing multi-step ahead predictions will be investigated further in this section.

3.1 Identification of Multi-step Ahead Models using the Direct Approach

The direct k-step ahead model is of the form

$$y_{t+k} = \sum_{i=1}^{n_m} f_i(\mathbf{x}_t)\theta_i^{dir} + e_{t+k/t}^{dir} \tag{21}$$

where $e_{t+k/t}^{dir}$ is the direct k-step ahead prediction error sequence. The following cost function

$$J = \sum_{t=1}^{N-k}\left(e_{t+k/t}^{dir}\right)^2 = \sum_{t=1}^{N-k}\left(y_{t+k} - \sum_{i=1}^{n_m} f_i(\mathbf{x}_t)\theta_i^{dir}\right)^2 \tag{22}$$

is usually minimised to estimate the unknown parameters in the direct k-step ahead model. Least squares type algorithms such as the forward regression orthogonal least squares algorithm (Chen et al. 1989) can readily be used to identify the direct k-step ahead model. After the model has been identified, the direct k-step ahead residuals can be computed as

$$\xi_{t+k/t}^{dir} = y_{t+k} - \sum_{i=1}^{n_m} f_i(\mathbf{x}_t)\hat{\theta}_i^{dir} \tag{23}$$

where $\hat{\theta}_i^{dir}$ are the estimated parameters.

3.2 Autocorrelation of the Direct k-step Ahead Prediction Errors

It will be shown in this section that the direct k-step ahead prediction errors are auto-correlated. First linear time series will be considered and this is followed by the nonlinear case. Consider a simple linear $AR(n_y)$ time series represented by

$$A(z^{-1})y_t = e_t \tag{24}$$

where $A(z^{-1}) = 1 - a_1 z^{-1} - \dots - a_{n_y} z^{-n_y}$ \hfill (25)

and z^{-1} is the backward shift operator. Let the polynomials

$$G_k(z^{-1}) = g_o^{(k)} + g_1^{(k)} z^{-1} + \dots + g_{n_y-1}^{(k)} z^{-ny+1} \tag{26}$$

$$\psi_k(z^{-1}) = 1 + \psi_1 z^{-1} + \dots + \psi_{k-1} z^{-k+1} \tag{27}$$

be uniquely defined by the identity

$$1 = A(z^{-1})\psi_k(z^{-1}) + z^{-k}G_k(z^{-1}) \tag{28}$$

where $\psi_k(z^{-1})$ is the polynomial of order $(k-1)$ obtained from the long division of 1 by $A(z^{-1})$, and $z^{-k}G_k(z^{-1})$ is the remainder. Substituting $A(z^{-1})$ from eqn (28) into eqn (24) with t replaced by $t+k$ yields

$$y_{t+k} = G_k(z^{-1})y_t + \psi_k(z^{-1})e_{t+k} \tag{29}$$

It follows that the optimal k-step ahead prediction (Soderstrom and Stoica 1989) is

$$y_{t+k/t}^{opt} = G_k(z^{-1})y_t \tag{30}$$

and the optimal k-step ahead prediction error sequence is

$$e_{t+k/t}^{opt} = \psi_k(z^{-1})e_{t+k} \tag{31}$$

which is auto-correlated until lag k. The same phenomena can be observed for the nonlinear case. Consider a time series generated by a nonlinear AutoRegressive (NAR) model as described in eqn (2). At time $t+1$, eqn (2) becomes

$$y_{t+1} = f\left(y_t, ..., y_{t-n_y-1}\right) + e_{t+1} \tag{32}$$

Similarly at time $t + 2$

$$y_{t+2} = f\left(y_{t+1}, ..., y_{t-n_y+2}\right) + e_{t+2} = f\left(f\left(y_t, ..., y_{t-n_y+1}\right) + e_{t+1}, ..., y_{t-n_y+2}\right) + e_{t+2} \tag{33}$$

From the discussion in Section 2.4.1, the optimal two-step ahead prediction is

$$y_{t+2/t}^{opt} = \int f\left(f\left(y_t, ..., y_{t-n_y+1}\right) + e, ..., y_{t-n_y+2}\right) dF_n(e) = f_2^{opt}\left(y_t, ..., y_{t-n_y+1}\right) \tag{34}$$

where $f_2^{opt}(\bullet)$ is some complex nonlinear function. The two-step ahead prediction error is

$$e_{t+2/t}^{opt} = y_{t+2} - y_{t+2/t}^{opt} = g_2^{opt}\left(y_t, ..., y_{t-n_y+1}, e_{t+1}\right) + e_{t+2} \tag{35}$$

where $g_2^{opt}(\bullet)$ is some complex nonlinear function. In general the optimal prediction at time $t + k$ is

$$y_{t+k/t}^{opt} = f_k^{opt}\left(y_t, ..., y_{t-n_y+1}\right) \tag{36}$$

where $f_k^{opt}(\bullet)$ is some complex nonlinear function whose analytical form is usually not available. The optimal k-step ahead prediction errors are

$$e_{t+k/t}^{opt} = g_k^{opt}\left(y_t, ..., y_{t-n_y+1}, e_{t+1}, ..., e_{t+k-1}\right) + e_{t+k} \tag{37}$$

Therefore similar to the linear case, the direct k-step ahead prediction errors will also be auto-correlated until lag k for nonlinear time series.

3.3 Covariance of the Parameter Estimates Associated with the k-step Ahead Model Identified Using the Direct Approach

It has been shown that the direct k-step ahead prediction errors are auto-correlated. This may influence the application of the direct approach. The optimal prediction function $f_k^{opt}(\bullet)$ in eqn (36) can be parameterised to give

$$y_{t+k} = y_{t+k/t}^{opt} + e_{t+k/t}^{opt} = \sum_{i=1}^{n_m} f_i(\mathbf{x}_t)\theta_i^{(k)} + e_{t+k/t}^{opt} \tag{38}$$

In matrix form eqn (38) can be written as

$$Y_k = P\Theta + \Xi_k \tag{39}$$

where $Y_k = \begin{bmatrix} y_{k+1} & y_{k+2} & \cdots & y_N \end{bmatrix}^T$, $\Theta = \begin{bmatrix} \theta_1^{(k)} & \theta_2^{(k)} & \cdots & \theta_{n_m}^{(k)} \end{bmatrix}^T$, $P = \begin{bmatrix} P_1 & P_2 & \cdots & P_{n_m} \end{bmatrix}$,

$P_i = \begin{bmatrix} f_i(\mathbf{x}_1) & f_i(\mathbf{x}_2) & \cdots & f_i(\mathbf{x}_{N-k}) \end{bmatrix}^T$ and $\Xi_k = \begin{bmatrix} e_{k+1/1}^{opt} & e_{k+2/2}^{opt} & \cdots & e_{N/N-k}^{opt} \end{bmatrix}^T$.

It is well known that minimising the sum of the squared errors yields the least squares estimates

$$\hat{\Theta} = \left(P^T P\right)^{-1} P^T Y_k \tag{40}$$

9

Substituting Y_k from eqn (39) into eqn (40) gives

$$\hat{\Theta} = \Theta + \left(P^T P\right)^{-1} P^T \Xi_k \tag{41}$$

Ξ_k would consist of future noise terms such as $\left(e_{t+1}, \dots, e_{t+k}\right)$ and would not be correlated with P, hence $E\left[P^T \Xi_k\right] = 0$. Therefore the parameter estimates associated with the direct k-step ahead model would be unbiased even though the prediction errors are auto-correlated. The covariance of the parameter estimates can be computed as

$$\mathrm{cov}\left(\hat{\Theta}\right) = E\left[\left(\hat{\Theta} - \Theta\right)\left(\hat{\Theta} - \Theta\right)^T\right] = \left(P^T P\right)^{-1} P^T E\left[\Xi_k \Xi_k^T\right] P\left(P^T P\right)^{-1} \tag{42}$$

It is known from Section 3.2 that the direct k-step ahead prediction errors are auto-correlated until lag k, hence

$$E\left[\Xi_k \Xi_k^T\right] \ge E\left[\Xi_1 \Xi_1^T\right], \quad k \ge 2 \tag{43}$$

Due to the auto-correlation of the direct k-step ahead prediction errors, the covariance of the parameter estimates associated with the direct k-step ahead model is greater than that for one-step ahead model. A simple linear time series generated by an AR(1) model is used to further illustrate this point. Assuming that the time series is obtained as

$$y_t = a y_{t-1} + e_t \tag{44}$$

The optimal k-step ahead prediction is

$$y_{t+k/t}^{opt} = a^k y_t \tag{45}$$

and the prediction obtained using the direct approach is

$$y_{t+k/t} = \left(a^{dir}\right)^{(k)} y_t \tag{46}$$

where $\left(a^{dir}\right)^{(k)} = \dfrac{\displaystyle\sum_{i=1}^{N-k} y_i y_{i+k}}{\displaystyle\sum_{i=1}^{N} y_i^2}$. The variance of $\left(a^{dir}\right)^{(k)}$ can be obtained as, see Findley (1985)

$$\mathrm{cov}\left(\hat{a}^{dir}\right)^{(k)} = E\left[\left(\left(\hat{a}^{dir}\right)^{(k)} - a^k\right)\left(\left(\hat{a}^{dir}\right)^{(k)} - a^k\right)^T\right] = \left[\frac{\left(1 + a^2\right)\left(1 - a^{2k}\right)}{1 - a^2} - 2ka^{2k}\right]\Big/ N \tag{47}$$

From the above eqn (47), as k increases the variance of the parameter estimates associated with the direct k-step ahead model increases. The variance of the parameter estimates associated with the k-step ahead model obtained using the direct approach has been shown to be much higher than that obtained using the iterative approach, see Findley (1985).

3.4 Prediction Accuracy of the Direct Approach

The effect of the covariance of the parameter estimates associated with the direct approach on prediction accuracy will be discussed in this section. For a general NAR model, the optimal multi-step ahead prediction can be obtained by $y_{t+k/t}^{opt} = P\Theta$, see eqn (39) and that obtained using the direct approach is $y_{t+k/t}^{dir} = P\hat{\Theta}^{dir}$, where $\hat{\Theta}^{dir}$ are the estimated parameters. A measure of the prediction accuracy can be obtained by computing the mean squared k-step ahead prediction errors

$$E\left[y_{t+k} - y_{t+k/t}^{dir}\right]^2 = E\left[y_{t+k} - P\hat{\Theta}^{dir}\right]^2 = E\left[\left(y_{t+k} - P\Theta\right) + \left(P\Theta - P\hat{\Theta}^{dir}\right)\right]^2 \tag{48}$$

Because $P\Theta$ is the optimal prediction, hence $\left(y_{t+k} - P\Theta\right)$ contains new information (noise) from time $t+1$ to $t+k$, and should not be correlated with $\left(P\Theta - P\hat{\Theta}^{dir}\right)$, which is obtained based on information up to time t. Hence eqn (48) becomes

$$E\left[y_{t+k} - y_{t+k/t}^{dir}\right]^2 = E\left[y_{t+k} - P\Theta\right]^2 + E\left[P\Theta - P\hat{\Theta}^{dir}\right]^2$$

$$= E\left[y_{t+k} - P\Theta\right]^2 + PE\left[\left(\Theta - \hat{\Theta}^{dir}\right)\left(\Theta - \hat{\Theta}^{dir}\right)^T\right]P^T \tag{49}$$

The first term on the r..h.s of eqn (49) is the mean squared prediction errors that would have been obtained if the optimal k-step ahead model was used to obtain the predictions. The second term on the r.h.s of eqn (49) is associated with the covariance of the parameter estimates of the direct approach. Therefore the accuracy of the prediction is directly related to the covariance of the parameter estimates associated with the direct model. For the simple linear AR(1) time series studied in Section 3.3, the prediction performance of the iterative approach is better than the direct approach because the variance of the parameter estimate associated with the iterative approach is smaller than that for the direct approach. The large covariance of the parameter estimates associated with the direct approach occurs for nonlinear models as well. Kabaila (1981) shows that for a general NAR model, the parameter estimates based on the one-step ahead prediction error approach (iterative approach) provides a better estimate of the model parameters than that obtained by minimizing the direct k-step ahead prediction errors (direct approach). This problem motivated the search for a new direct approach with uncorrelated k-step ahead prediction errors, such as $e_{t+k/t} = e_{t+k}$. The covariance of the parameter estimates associated with this new approach would then be small and hence a better prediction accuracy may be obtained.

11

4. A New Direct Approach for Multi-Step Ahead Predictions

Consider a general $\mathrm{NAR}(n_y)$ time series described in eqn (2). To construct the k-step ahead model using the conventional direct approach, a general $\mathrm{NAR}(n_y)$ model is employed to give

$$y_{t+k} = f_k^{dir}\left(y_t, ..., y_{t-n_y+1}\right) + e_{t+k/t}^{dir} \tag{50}$$

where $f_k^{dir}(\bullet)$ is the direct k-step ahead model. Even if the optimal direct k-step ahead model is identified as in eqn (36), the obtained k-step ahead prediction errors are auto-correlated, see eqn (37). This is because at time $t+k$, eqn (2) becomes

$$y_{t+k} = f_k\left(y_t, ..., y_{t-n_y+1}, e_{t+1}, ..., e_{t+k-1}\right) + e_{t+k} \tag{51}$$

where $f_k(\bullet)$ is some complex nonlinear function and e_{t+k} is the noise at time $t+k$. Hence the correct model structure to be employed to identify the direct k-step ahead model should be a $\mathrm{NARMA}(n_y, k-1)$ model

$$y_{t+k} = f_k^{dir-noise}\left(y_t, ..., y_{t-n_y+1}, e_{t+1}^{dir-noise}, ..., e_{t+k-1}^{dir-noise}\right) + e_{t+k}^{dir-noise} \tag{52}$$

where $f_k^{dir-noise}(\bullet)$ is some complex nonlinear function and $e_{t+k}^{dir-noise}$ is the one-step ahead prediction error at time $t+k$. Strictly speaking $f_k^{dir-noise}(\bullet)$ is just a one-step ahead model because information up to time $t+k-1$ is used to predict the output at time $t+k$. However the model $f_k^{dir-noise}(\bullet)$ will be employed to make k-step ahead predictions and hence will still be called the direct k-step ahead model and also $e_{t+k}^{dir-noise}$ is accordingly named as the k-step ahead prediction error. Comparing eqns (51) & (52), it can be seen that the prediction error sequence $e_{t+k}^{dir-noise}$ can be made to be an uncorrelated sequence if the identified k-step ahead model $f_k^{dir-noise}(\bullet)$ in eqn (52) is a good approximation to the actual k-step ahead prediction function $f_k(\bullet)$ in eqn (51). This new approach of identifying the direct k-step ahead model will be called the direct-future-noise approach. The new direct k-step ahead model in eqn (52) can alternatively be expressed in the form

$$y_{t+k} = \sum_{i=1}^{n_m} f_i^{dir-noise}\left(\mathbf{x}_{t+k-1}\right)\theta_i^{dir-noise} + e_{t+k}^{dir-noise} \tag{53}$$

where $\mathbf{x}_{t+k-1} = \left(y_t, ..., y_{t-n_y+1}, e_{t+1}^{dir-noise}, ..., e_{t+k-1}^{dir-noise}\right)^T$. The following cost function

$$J = \sum_{i=1}^{N-k}\left(e_{t+k}^{dir-noise}\right)^2 = \sum_{t=1}^{N-k}\left(y_{t+k} - \sum_{i=1}^{n_m} f_i\left(\mathbf{x}_{t+k-1}\right)\theta_i^{dir-noise}\right)^2 \tag{54}$$

can be minimised to estimate the unknown parameters in the new direct k-step ahead model. Least square type algorithms such as the well-known forward regression orthogonal least algorithm (Chen et al. 1989) can be used to identify the new direct k-step ahead model.

4.1 Prediction Using the Direct-Future-Noise Approach for Nonlinear models

The new k-step ahead direct-future-noise model in eqn (52) contains future noise terms because noise terms such as $\left(e_{t+1}^{dir-noise},...,e_{t+k-1}^{dir-noise}\right)$ are unknown at time t. Hence the k-step ahead predictions cannot be computed exactly, except for the linear model case where future noise terms can be set to the mean value of zero. Here a simple approximation will be employed for the nonlinear case. For ease of explanation and clarity, a polynomial nonlinear function is used initially to describe the approximation method. Assuming that the three-step ahead direct-future-noise polynomial model in eqn (52) is identified as

$$\hat{f}_3^{dir-noise}(\bullet) = 0.5 y_t + 0.5 y_t e_{t+1}^{dir-noise} + 0.2\left(e_{t+2}^{dir-noise}\right)^2 \tag{55}$$

Therefore the three-step ahead predictions can be approximated by

$$\begin{aligned}
\hat{y}_{t+3/t}^{dir-noise} &= E_t\left[0.5 y_t + 0.5 y_t e_{t+1}^{dir-noise} + 0.2\left(e_{t+2}^{dir-noise}\right)^2\right] \\
&= 0.5 y_t + 0.5 y_t E_t\left[e_{t+1}^{dir-noise}\right] + 0.2 E_t\left[\left(e_{t+2}^{dir-noise}\right)^2\right] \\
&\approx 0.5 y_t + 0.5 y_t\left(\frac{1}{N_{train}-1}\sum_{i=0}^{N_{train}-1}\xi_{i+1}\right) + 0.2\left(\frac{1}{N_{train}-2}\sum_{i=0}^{N_{train}-2}\xi_{i+2}^2\right) \\
&\approx \frac{1}{N_{train}-2}\sum_{i=0}^{N_{train}-2}\left(0.5 y_t + 0.5 y_t \xi_{i+1} + 0.2\xi_{i+2}^2\right) \tag{56}
\end{aligned}$$

where $\xi_i = y_i - \hat{f}_1\left(y_i,...,y_{i-n_y+1}\right)$, $i=1,...,N_{train}$ are the realised one-step ahead residuals in the training data set with data length N_{train} and $\hat{f}_1(\bullet)$ is the identified one-step ahead model. This means that $e_{t+k}^{dir-noise}$ to the power of l at time t is approximated by the average of ξ_{t+k}^l. This approximation is extended to the RBF case. Assuming that the identified k-step ahead model for the RBF case is

$$\hat{f}_k^{dir-noise}(\bullet) = \sum_{i=1}^{n_m}\phi\left(\left\|\mathbf{x}_{t+k-1} - \mathbf{c}_i\right\|\right)\hat{\theta}_i^{dir-noise} \tag{57}$$

where $\mathbf{x}_{t+k-1} = \left(y_t,...,y_{t-n_y+1},e_{t+1}^{dir-noise},...,e_{t+k-1}^{dir-noise}\right)^T$ and $\hat{\theta}_i^{dir-noise}$ is the estimated parameters. The k-step ahead predictions can be obtained as

$$\hat{y}_{t+k/t}^{dir-noise} = \frac{1}{N_{train}-k}\sum_{j=0}^{N_{train}-k}\left(\sum_{i=1}^{n_m}\phi\left(\left\|\mathbf{x}_{j,t+k} - \mathbf{c}_i\right\|\right)\hat{\theta}_i^{dir-noise}\right) \tag{58}$$

13

where $\mathbf{x}_{j,t+k} = \left(y_t, \dots, y_{t-n_y+1}, \xi_{j+1}, \dots, \xi_{j+k} \right)^T$. Simulation studies will be performed in Section 5 to compare the prediction performance of the direct-future-noise approach using this approximation with the iterative and the direct approaches for nonlinear time series.

4.3 Procedures for the Three Different Multi-step Ahead Prediction Methods

In this section, the procedure associated with each of the three different multi-step ahead prediction methods will be summarised based on an RBF modelling framework. These are the iterative approach discussed in Section 2.4.2, the direct approach discussed in Section 3 and the new direct-future-noise approach introduced in Section 4.

4.3.1 The Iterative Approach

i) Set the prediction step $k=1$ and the vector $\mathbf{x}_t = \left(y_t \quad \cdots \quad y_{t-n_y+1} \right)^T$.

ii) Set the initial number of centres $n_m = 5$.

iii) Use the forward regression orthogonal least squares algorithm (Chen et al. 1989) to identify the one-step ahead model in eqn (6) by selecting n_m significant centres from all the training data and estimating the associated parameters.

iv) Calculate the mean squared one-step ahead prediction errors over the validation data set.

v) Increase n_m by one and repeat steps (iii) to (v). Go to step (vi) when $n_m = 60$.

vi) Record the number of centres with the minimum mean squared one-step ahead prediction errors calculated over the validation data set.

vii) Re-identify the one-step ahead model from the training data set with the recorded number of centres in step (vi).

viii) Compute the mean squared one-step ahead prediction errors over the testing data set using the identified model in step (vii).

ix) Increase the prediction step k by one and update $\mathbf{x}_t = \left(\hat{y}_{t+1/t}^{ite} \quad y_t \quad \cdots \quad y_{t-n_y+2} \right)^T$.

x) Compute the 2-step ahead prediction using the identified one-step ahead model in step (vii) but with updated $\mathbf{x}_t$. Calculate the mean squared 2-step ahead prediction errors over the testing data set.

xi) Repeat steps (ix) to (x) until the required number of prediction steps has been obtained.

4.3.2 The Direct Approach

Steps (i) to (viii) are the same as the iterative approach in Section 4.3.1.

(ix) Increase the prediction step k by one. Use the same arguments as in one-step ahead model

$$\mathbf{x}_t = \begin{pmatrix} y_t & \cdots & y_{t-n_y+1} \end{pmatrix}^T.$$

(x) Repeat steps (ii) to (vii) to identify the direct k-step ahead model described in eqn (21).

(xi) Compute the mean squared k-step ahead prediction errors over the testing data set.

(xii) Repeat step (ix) to (xi) until the required number of prediction steps has been obtained.

4.3.3 The Direct-Future-Noise Approach

Steps (i) to (viii) are the same as the iterative approach in Section 4.3.1.

(ix) Compute the residuals ξ_t of the training data set using the identified one-step ahead model in step (vii).

(x) Increase the prediction step k by one and update the vector

$$\mathbf{x}_{t+k-1} = \begin{pmatrix} y_t, ..., y_{t-n_y+1}, \xi_{t+1}, ..., \xi_{t+k-1} \end{pmatrix}^T.$$

(xi) Repeat steps (ii) to (vii) to identify the direct k-step ahead model described in eqn (53).

(xii) Compute the mean squared k-step ahead prediction errors over the testing data set, using the approximation method described in Section 4.1.

(xiii) Repeat step (ix) to (xi) until the required number of prediction steps has been obtained.

5. Simulation Studies

5.1 Experiments on AR(1) Model

In this experiment, the parameter estimation and the prediction accuracy, of the iterative, the direct, and the new direct-future-noise approaches, are studied based on a simple AR(1) time series described in eqn (44). Two different values of a in eqn (44) were considered in the experiment, $a = 0.5$ and $a = 0.9$, and for each value of a a data set of 600 data points was generated. The first 300 data points were used as the training data set and the next 300 points were used as the testing data set. The parameters $\hat{a}^{ite}$, $\left(\hat{a}^{dir}\right)^{(k)}$ and $\left(\hat{a}^{dir-noise}\right)^{(k)}$ associated with the iterative, the direct and the direct-future-noise approaches were estimated from the training data set and the k-step ahead predictions were computed on the testing data set. For prediction using the direct-future-noise approach, the future noise terms were simply set to zero. This is reasonable for the linear case since $E_t[e_{t+k}] = 0,\ k > 0$.

The mean squared k-step ahead prediction errors of the iterated, the direct and the direct-future-noise approaches were computed on the testing data set. In addition the values of

15

$$\Delta \hat{a}^{ite} = \left(a^k - \left(\hat{a}^{ite}\right)^k\right)^2, \quad \Delta \hat{a}^{dir} = \left(a^k - \left(\hat{a}^{dir}\right)^{(k)}\right)^2 \quad \text{and} \quad \Delta \hat{a}^{dir-noise} = \left(a^k - \left(\hat{a}^{dir-noise}\right)^{(k)}\right)^2 \quad \text{for the}$$

iterated, the direct and the direct-future-noise approaches were computed. Three hundred replications of these simulations were performed and the results are reported in Table 1. The results in Table 1 clearly show that the variance of the parameter estimate associated with the direct approach is higher than that for the iterative approach and the direct-future-noise approach. This resulted in higher mean squared prediction errors for the direct approach compared to both the iterative approach and the direct-future-noise approach.

5.2 Multi-step Ahead Residuals for the New Direct-Future-Noise Approach

This experiment is to show that the k-step ahead residuals of the new direct-future-noise approach can be an unpredictable sequence if the k-step ahead model in eqn (52) is a good approximation to the complex k-step ahead prediction function in eqn (51). This was verified by simulation using the following equation

$$y_t = \left(0.8 - 0.5\exp\left(-y_{t-1}^2\right)\right)y_{t-1} - \left(0.3 + 0.9\exp\left(-y_{t-1}^2\right)\right)y_{t-2} + 0.1\sin(\pi y_{t-1}) + e_t \qquad (59)$$

where e_t was a zero mean unpredictable sequence with variance 0.04. This equation was used to generate 500 data points of $y_t, t = 1...500$. A thin-plate-spline radial basis function was employed to construct the one-step ahead model with $n_y = 2$, so that. $\mathbf{x}_t = \left(y_t, y_{t-1}\right)^T$.

Different models were identified by varying the number of important centres n_m to be selected by the forward regression orthogonal least squares algorithm (Chen et al 1989). It was found that when $n_m = 17$, the one-step ahead residuals pass the correlation conditions stated in eqns (13) & (14). The correlation plots of the one-step ahead residuals are shown in Figure 1. To construct the 2-step ahead direct-future-noise model, the arguments in the radial basis function network were updated as $\mathbf{x}_{t+1} = \left(y_t, y_{t-1}, \xi_{t+1}\right)^T$ where ξ_t were the realized residuals calculated using the one-step ahead model identified earlier with $n_m = 17$. Again different 2-step ahead models were identified using the forward regression orthogonal least squares algorithm (Chen et al. 1989) with different numbers of terms n_m. When $n_m = 20$, it was found that the two-step ahead residuals $\xi_{t+2}^{dir-noise}$ were unpredictable by the correlation plots in Figure 2. This procedure was repeated for 3,4 & 5-steps ahead direct-future-noise models. The required numbers of n_m used to construct the

Table 1. The k-step ahead mean squared prediction errors and the sample variance of the parameter estimates for the iterative, the direct and the direct-residual approaches for an AR(1) model. mse_i represents the mean squared prediction errors on the testing data set of replication i. $\Delta\hat{a}^{ite}$, $\Delta\hat{a}^{dir}$ and $\Delta\hat{a}^{dir-noise}$ are the values of $\left(a^k - \left(\hat{a}^{ite}\right)^k\right)^2$, $\left(a^k - \left(\hat{a}^{dir}\right)^k\right)^2$ and $\left(a^k - \left(\hat{a}^{dir-noise}\right)^{(k)}\right)^2$ respectively.

Prediction Step	$\frac{1}{300}\sum_{i=1}^{300}(mse_i)$ of iterative approach	$\frac{1}{300}\sum_{i=1}^{300}(mse_i)$ of direct approach	$\frac{1}{300}\sum_{i=1}^{300}(mse_i)$ of direct-future-noise approach	$\frac{1}{300}\sum_{i=1}^{300}\left(\Delta\hat{a}^{ite}_i\right)$ of iterative approach	$\frac{1}{300}\sum_{i=1}^{300}\left(\Delta\hat{a}^{dir}_i\right)$ of direct approach	$\frac{1}{300}\sum_{i=1}^{300}\left(\Delta\hat{a}^{dir-noise}_i\right)$ of direct-future-noise approach
Case 1 $a = 0.5$						
$k=1$	0.9986	0.9986	0.9986	0.0023	0.0023	0.0023
$k=2$	1.2479	1.2498	1.2498	0.0022	0.0040	0.0040
$k=3$	1.3093	1.3136	1.3119	0.0013	0.0046	0.0032
$k=4$	1.3235	1.3303	1.3272	0.0006	0.0047	0.0026
$k=5$	1.3275	1.3337	1.3302	0.0002	0.0049	0.0026
Case 2 $a = 0.9$						
$k=1$	1.0012	1.0012	1.0012	0.0006	0.0006	0.0006
$k=2$	1.8155	1.8175	1.8174	0.0020	0.0021	0.0021
$k=3$	2.4734	2.4797	2.4760	0.0035	0.0042	0.0038
$k=4$	3.0021	3.0146	3.0056	0.0049	0.0063	0.0051
$k=5$	3.4327	3.4519	3.4358	0.0060	0.0085	0.0063

respective k-step ahead model were 24,30 & 40, such that the 3,4 & 5-steps ahead residuals were unpredictable according to the model validity tests. The correlation plots for the 5-step ahead residuals are shown in Figure 3. The experiment was repeated but this time the conventional direct approach discussed in Section 3 was used to construct the multi-step ahead model. It was found that the multi-step ahead residuals obtained using the direct approach could not satisfy the correlation tests regardless of the number of centres chosen in the radial basis function networks. This example shows that the multi-step ahead residuals of the direct-future-noise approach can be an unpredictable sequence if the structure of the k-step ahead model, in this case the number of centres of radial basis function network, is chosen properly.
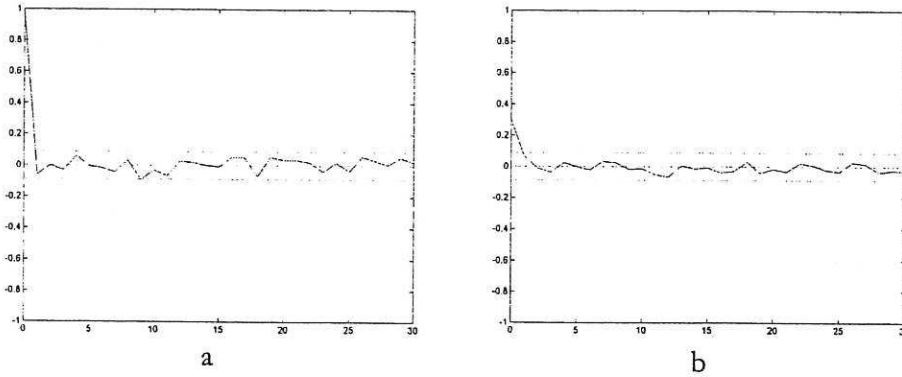

Figure 1 Correlation plots of the 1-step ahead residual sequence ξ_t, (a) $\rho_{\xi\xi}(\tau_\rho)$ and (b) $\rho_{[y\xi][\xi\xi]}(\tau_\rho)$, for the nonlinear example in eqn (59).
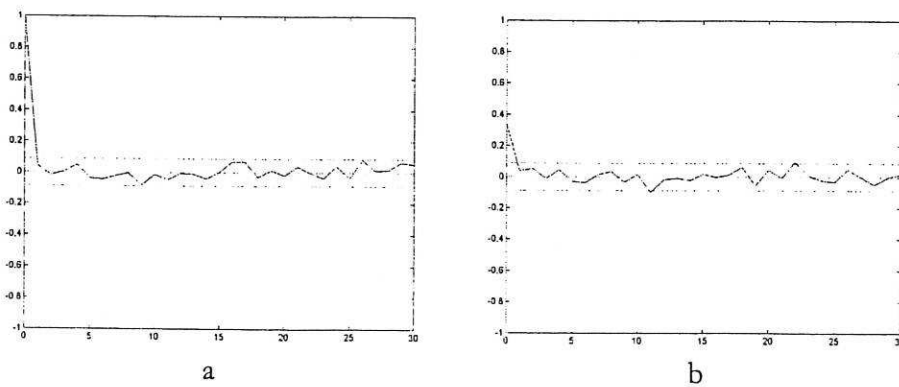

Figure 2 Correlation plots of the 2-step ahead residual sequence $\xi_{t+2}^{dir-noise}$, (a) $\rho_{\xi\xi}(\tau_\rho)$ and (b) $\rho_{[y\xi][\xi\xi]}(\tau_\rho)$, for the nonlinear example in eqn (59).
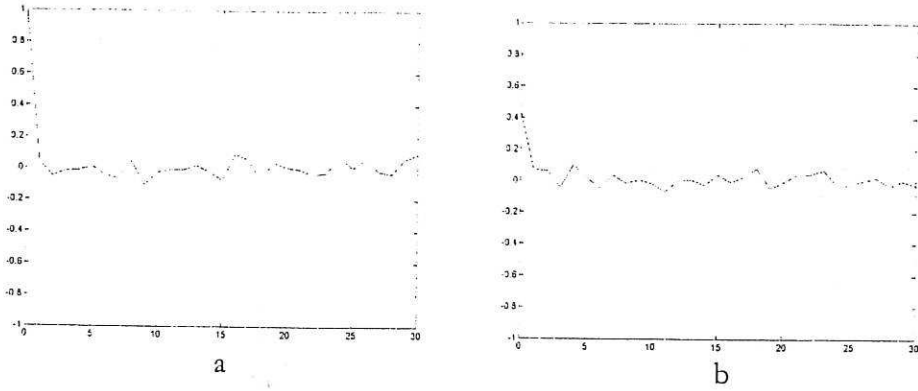
Figure 3 Correlation plots of the 5-step ahead residual sequence $\xi_{t+5}^{dir-noise}$, (a) $\rho_{\xi\xi}(\tau_\rho)$ and (b) $\rho_{[y\xi][\xi\xi]}(\tau_\rho)$, for the nonlinear example in eqn (59).

5.3 Simulation Studies on Prediction Accuracy for the Three Different Prediction Methods Applied to Nonlinear Time Series

The procedures outlined in the preceding Section 4.3 for the three prediction methods will be illustrated by means of three nonlinear examples. The first example is a simple polynomial function of the form

$$y_t = 0.3y_{t-2} - 0.5y_{t-1} - 0.1y_{t-1}y_{t-2} + e_t \tag{60}$$

where e_t was an unpredictable zero mean sequence with variance 1. The second example is a rational model of the form

$$y_t = \frac{0.8y_{t-1} + y_{t-1}y_{t-3}}{1 + y_{t-2}^2} + e_t \tag{61}$$

where e_t was an unpredictable zero mean sequence but with variance 0.04. The third example was generated from eqn (59). For each example, a data set of 1500 data points was generated. The first 500 data points were used as the training data set, the next 500 data points were used as the validation data set and the last 500 data points were used as the testing data set. The maximum number of output lags n_y chosen for the three examples were set to the correct values of 2,3 and 2 respectively so that the results reflect a fair comparison of the prediction effects only. The procedures outlined in Section 4.3 were used to compute the mean squared k-step ahead prediction errors of the iterative, the direct and the direct-future-noise approaches.

Fifty replications of the above simulation were performed. The results obtained for the three cases are presented in Table 2,3 and 4. Only ten step ahead predictions were computed for examples one and two. This is because the mean squared prediction errors at the 10th step were

19

almost equal to the variance of the time series and hence using the mean of the time series as the prediction would give almost the same predictive accuracy. Much higher prediction time steps were computed for example three because the mean squared prediction error at the 20th step was still much less than the variance of the time series. It can clearly be seen from the three tables that the direct-future-noise approach gives the best performance for almost all the prediction time steps considered. The high prediction accuracy of the direct-future-noise approach could be due to the smaller variance in the estimated parameters associated with this approach. Also note that unlike the linear case presented in Table 1, the iterative approach performs worse than the direct approach and the direct-future-noise approach in all the three examples. This indicates clearly that the iterative approach will in most cases yield unsatisfactory results for multi-step ahead predictions for nonlinear models. However it is still not conclusive to state that the direct or direct-future-noise approach will always perform better than the iterative approach for multi-step ahead predictions for nonlinear models. Factors such as the effect of setting the future noise terms to zero and the complexity of the k-step ahead prediction functions will affect the prediction accuracy of the iterative and the direct approach respectively. These factors are clearly case dependent.

Table 2. The multi-step ahead mean squared prediction errors of the different approaches for the first example described in eqn (60) where mse_i represents the mean squared prediction errors on testing data set of replication i. The values in brackets represent the percentage improvement in prediction accuracy over the iterative approach.

Prediction Step	$\frac{1}{50}\sum_{i=1}^{50}(mse_i)$ of iterative approach	$\frac{1}{50}\sum_{i=1}^{50}(mse_i)$ of direct approach	$\frac{1}{50}\sum_{i=1}^{50}(mse_i)$ of direct-future-noise approach
$k=1$	1.0527	1.0527 (0.0%)	1.0527 (0.0%)
$k=2$	1.3441	1.3427 (0.1%)	1.3354 (0.6%)
$k=3$	1.6560	1.6298 (1.6%)	1.6308 (1.5%)
$k=4$	1.8433	1.8221 (1.2%)	1.8145 (1.6%)
$k=5$	1.9961	1.9545 (2.1%)	1.9468 (2.5%)
$k=6$	2.0978	2.0540 (2.1%)	2.0390 (2.8%)
$k=7$	2.1728	2.1185 (2.5%)	2.1033 (3.2%)
$k=8$	2.2223	2.1721 (2.3%)	2.1496 (3.3%)
$k=9$	2.2681	2.1974 (3.1%)	2.1785 (4.0%)
$k=10$	2.3060	2.2229 (3.6%)	2.1983 (4.7%)

Table 3. The multi-step ahead mean squared prediction errors of the different approaches for the second example described in eqn (61) where mse_i represents the mean squared prediction errors on testing data set of replication i. The values in brackets represent the percentage improvement in prediction accuracy over the iterative approach.

Prediction Step	$\frac{1}{50}\sum_{i=1}^{50}(mse_i)$ of iterative approach	$\frac{1}{50}\sum_{i=1}^{50}(mse_i)$ of direct approach	$\frac{1}{50}\sum_{i=1}^{50}(mse_i)$ of direct-future-noise approach
$k=1$	0.0419	0.0419 (0.0%)	0.0419 (0.0%)
$k=2$	0.0746	0.0749 (-0.4%)	0.0743 (0.4%)
$k=3$	0.0905	0.0908 (-0.3%)	0.0896 (1.0%)
$k=4$	0.1038	0.1011 (2.6%)	0.0999 (3.6%)
$k=5$	0.1175	0.1116 (5.0%)	0.1100 (6.4%)
$k=6$	0.1303	0.1207 (7.4%)	0.1199 (8.0%)
$k=7$	0.1429	0.1286 (10.0%)	0.1274 (10.8%)
$k=8$	0.1549	0.1346 (13.1%)	0.1333 (13.9%)
$k=9$	0.1659	0.1399 (15.7%)	0.1386 (16.4%)
$k=10$	0.1754	0.1462 (16.6%)	0.1428 (18.6%)

Table 4. The multi-step ahead mean squared prediction errors of the different approaches for the third example described in eqn (59) where mse_i represents the mean squared prediction errors on testing data set of replication i. The values in brackets represent the percentage improvement in prediction accuracy over the iterative approach.

Prediction Step	$\frac{1}{50}\sum_{i=1}^{50}(mse_i)$ of iterative approach	$\frac{1}{50}\sum_{i=1}^{50}(mse_i)$ of direct approach	$\frac{1}{50}\sum_{i=1}^{50}(mse_i)$ of direct-future-noise approach
$k=1$	0.0425	0.0425 (0.0%)	0.0425 (0.0%)
$k=2$	0.0804	0.0827 (-2.9%)	0.0809 (-0.6%)
$k=3$	0.0901	0.0903 (-0.2%)	0.0889 (1.3%)
$k=4$	0.1316	0.1297 (1.4%)	0.1274 (3.2%)
$k=5$	0.1533	0.1514 (1.2%)	0.1477 (3.7%)
$k=6$	0.1735	0.1677 (3.3%)	0.1636 (5.7%)
$k=7$	0.2162	0.2031 (6.1%)	0.1989 (8.0%)

$k = 8$	0.2326	0.2180 (6.3%)	0.2116 (9.0%)
$k = 9$	0.2607	0.2395 (8.1%)	0.2344 (10.1%)
$k = 10$	0.2959	0.2670 (9.8%)	0.2603 (12.0%)
$k = 11$	0.3141	0.2797 (10.9%)	0.2731 (13.1%)
$k = 12$	0.3515	0.3053 (13.1%)	0.2986 (15.0%)
$k = 13$	0.3779	0.3252 (13.9%)	0.3192 (15.5%)
$k = 14$	0.4007	0.3395 (15.3%)	0.3308 (17.4%)
$k = 15$	0.4375	0.3627 (17.1%)	0.3556 (18.7%)
$k = 16$	0.4580	0.3736 (18.4%)	0.3687 (19.5%)
$k = 17$	0.4852	0.3912 (19.4%)	0.3821 (21.2%)
$k = 18$	0.5146	0.4108 (20.2%)	0.4019 (21.9%)
$k = 19$	0.5357	0.4181 (21.9%)	0.4120 (23.1%)
$k = 20$	0.5642	0.4371 (22.5%)	0.4249 (24.7%)

6. Conclusions

The direct approach of computing multi-step ahead predictions for nonlinear autoregressive models has been analysed in detail. It has been shown that the direct k-step ahead prediction errors of the direct approach are auto-correlated, even if the optimal direct k-step ahead model is used to compute the predictions. This auto-correlation has been shown to contribute to a higher covariance of the parameter estimates associated with the direct approach. It was also shown that the prediction accuracy of the direct approach is directly related to the covariance of the parameter estimates, hence higher mean squared prediction errors were obtained for this case. A new direct-future-noise approach has been proposed and a nonlinear example was employed to show that the direct k-step ahead prediction errors of this new approach could be an unpredictable sequence. Linear and nonlinear time series examples were then considered to compare the prediction performance of the iterative, the direct and the direct-future-noise approaches. The iterative approach produces the best prediction performance for linear time series but the direct-future-noise approach gives the best results for all the nonlinear time series examples studied. In both linear and nonlinear time series cases the new direct-future-noise approach performed slightly better than the direct approach.

Acknowledgements

References

Al-Qassam, M.S. and Lane, L.A. 1989. Forecasting exponential autoregressive models of order (1), Journal of Time Series Analysis, vol. 10, no. 2, pp95-113.

Atiya, A.F., Suzan, M., El-Shoura, S., Shabeen, I. And El-Sherif, M.S. 1999. A comparison between neural-network forecasting techniques-case study: river flow forecasting. IEEE Transaction on Neural Networks, vol 10, no.2 , pp402-409.

Billings, S.A. and Chen, S. 1989. Extended model set, global data and threshold model identification of severely non-linear systems. International Journal of Control, vol. 50, no. 5, pp1897-1923.

Billings, S.A. and Zhu, Q.M. 1994. Nonlinear model validation using the correlation tests, International journal of Control, vol. 60, pp1107-1120.

Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. 1994. Time series analysis 3rd edition. Prentice Hall, Englewood Cliffs, New Jersey.

Broersen, P.M.T. 2000. Autoregressive model orders for Durbin's MA and ARMA estimators, IEEE Transaction on Signal Processing, vol. 48, no. 8, pp2454-2457.

Brown, B. and Mariano, R. 1989. Predictors in dynamic non-linear models: large sample behaviour, Econometric Theory, 5, pp430-452.

Chen, S. and Billings, S.A. 1989a. Modelling and analysis of non-linear time series, International Journal of Control, vol. 50, no. 6, pp2151-2171; 1989b. Recursive prediction error parameter estimator for non-linear models, International Journal of Control, vol. 49, no. 2, pp569-594.

Chen, S., Billings, S.A. and Luo, W. 1989. Orthogonal least squares method and their applications to non-linear system identification, International journal of Control, vol. 50, pp1873-1896.

Chen, S., Cowan, C.F.N. and Grant, P.M. 1991. Orthogonal least squares learning algorithm for radial basis function network, IEEE Transaction of Neural Networks, vol. 2, no. 2, pp302-309.

Chen, S., Billings, S.A., Cowan, C.F.N. and Grant, P.M. 1990. Practical identification of NARMAX models using radial basis functions. International Journal of Control, vol. 52, no. 6, pp1327-1350.

Cloarec, G. and Ringwood, J. 1998. Incorporation of statistical methods in multi-step neural network prediction models. Proceedings of the 1998 IEEE International Joint Conference on Neural Networks, pp2513-2518, Anchorage, AK: USA.

Durbin, J. 1959. Efficient estimation of parameters in moving average models, Biometrika, vol. 46, pp306-316.

Findley, D.F. 1985. Model selection for multi-step-ahead forecasting in H.A. Barker and P.C. Yound (eds) Prep. Of 7th IFAC/IFORS Symp. On identification and System Parameter Estimation, York, 1985, pp1039-1044.

Gabr, M.M. and Rao, T.S. 1981. The estimation and prediction of subset bilinear time series models with applications. Journal of Time Series Analysis, vol. 9, no. 3, pp11-20.

Guo, M., Bai, Z. and An, H.Z. 1999. Multi-step prediction for nonlinear autoregressive models based on empirical distributions. Statistica Sinica, 9, pp559-570.

Kabaila, P.V. 1981. Estimation based on one-step ahead prediction versus estimation based on multi-step ahead prediction, Stochastics, vol. 6, pp43-55.

Leontaritis, I.J. and Billings, S.A. 1985a. Input-output parametric models for nonlinear systems part I: deterministic nonlinear systems. International Journal of Control, 41, p303-328; (1985b) Input-output parametric models for nonlinear systems part II: stochastic nonlinear systems. International Journal of Control, 41, p329-344.

Lin J.L. and Granger, C.W.J. 1994. Forecasting from non-linear models in practice. Journal of Forecasting, vol. 13, pp1-9.

Parlos, A.G., Rais, O.T. and Atiya, A.F. 2000. Multi-step-ahead prediction using dynamic recurrent neural networks. Neural Networks, 13, pp765-786.

Pemberton J. 1987. Exact least squares multi-step prediction from nonlinear autoregressive models. Journal of Time Series Analysis, vol. 8, no. 4, pp443-448.

Söderström, T. and Stoica, P. 1989. System Identification, Prentice Hall.

Stoica, P. and Nehorai, A. 1989. On multistep prediction error methods for time series models. Journal of Forecasting, vol.8, pp357-368.

Tong, H. 1990. Nonlinear Time Series, A Dynamical System Approach, Clarendon Press, Oxford.

Weigend, A.S., Huberman, B.A., Rumelhart, D.E. 1992. Predicting sunspots and exchange rates with connectionist networks, in Casdagli, M. and Eubank, S. (eds), Nonlinear Modeling and Forecasting, Addison-Wesley, Redwood City, CA, pp395-432.

Zhang, G., Patuwo, B.E. and Hu, M.Y. 1998. Forecasting with artificial neural networks: The state of the art. International Journal of Forecasting, 14, pp35-62.