A Comparison between Linear and Nonlinear Forecasts for Nonlinear AR Models

MEIHUI GUO* AND Y. K. TSENG
National Sun Yat-sen University, Taiwan, ROC

ABSTRACT

In this paper the relative forecast performance of nonlinear models to linear models is assessed by the conditional probability that the absolute forecast error of the nonlinear forecast is smaller than that of the linear forecast. The comparison probability is explicitly expressed and is shown to be an increasing function of the distance between nonlinear and linear forecasts under certain conditions. This expression of the comparison probability may not only be useful in determining the predictor, which is either a more accurate or a simpler forecast, to be used but also provides a good explanation for an odd phenomenon discussed by Pemberton. The relative forecast performance of a nonlinear model to a linear model is demonstrated to be sensitive to its forecast origins. A new forecast is thus proposed to improve the relative forecast performance of nonlinear models based on forecast origins. © 1997 John Wiley & Sons, Ltd.

J. forecast, Vol. **16**, 491–508 (1997)

No. of Figures: 13. No. of Tables: 1. No. of References: 8.

KEY WORDS forecast origin; multi-step forecast; SETAR model

INTRODUCTION

Prediction is an important topic in time series analysis. Many stationary phenomena in practice can be described or at least be approximated by stationary linear time series models, e.g. the ARMA(p, q) models (see, for example, Priestly, 1989). However, many nonlinear phenomena such as limit cycles, frequency modulations (Tong, 1990) and animal population cycles (Oster and Ipaktchi, 1978) cannot be described adequately by linear time series models, unless superfluous parameters are involved with. Intuitively, if the true model is a nonlinear time series model, then any statistical inferences using analysis for the nonlinear model, which capture the nonlinear characteristic of the data, should be better than those by using linear approximation.

* Correspondence to: Meihui Guo, Department of Applied Mathematics, National Sun Yat-sen University, Kaohsiung, ROC 80424, Taiwan.

Contract grant sponsor: National Science Council of the Republic of China; Contract grant number: NSC 82-0208-N-110-064.

CCC 0277–6693/97/070491–18$17.50

Received January 1996
Accepted July 1997

However, this is not always the case for the prediction problem. For example, Davies *et al*. (1988) and Pemberton (1989), by numerical simulation, observe the phenomenon that the conditional mean and conditional median forecasts of nonlinear time series models have poor forecast performance compared to those of linear models. Therefore, if the linear forecast is not very much worse than the nonlinear forecast, then it is reasonable to adopt the former forecast rather than the nonlinear forecast, at least for computational reasons. More importantly, in practical applications for a given data set, unless an additional test is done in advance, there is no way of knowing whether the true model is factually linear or nonlinear. Thus, there arises the question: Under what circumstances should the nonlinear predictors be adopted?

The present research is motivated by Davies *et al.* (1988) who conducted three experiments to compare the forecast performance of nonlinear SETAR (Self-Exciting Threshold AutoRegressive) models with linear AR models. In experiment I they carried out the identification and estimation of SETAR models for real time series data sets. In experiments II and III they simulated series for different lengths respectively from each of 50 randomly selected stationary SETAR(2;1,1) models of the form

$$X_n = \begin{cases} aX_{n-1} + \varepsilon_n & \text{if} \quad X_{n-1} \leqslant 0 \\ bX_{n-1} + \varepsilon_n & \text{if} \quad X_{n-1} > 0 \end{cases} \tag{1}$$

where ε_n's are i.i.d. $N(0, 1)$ r.v's. For each set of the data, both nonlinear SETAR models and linear AR models were fitted. Among three experiments the relative frequencies, of the events that the absolute forecast error of an m-step ($m = 1, 2, 3$) nonlinear forecast is smaller than that of the linear AR forecast, range from 48% to 69% when the conditional mean is used as the forecast and range from 47% to 71% when the conditional median is used. These results strongly imply that there is a positive probability that the linear forecast is better than the nonlinear forecast even when the true model is nonlinear.

Pemberton (1989) conducted simulations for model (1) with $a = -b = 0.4$ and 0.9 and fitted generated data by AR models. For the conditional median forecast, the conditional probability that the absolute forecast error of the m-step nonlinear forecast is smaller than that of the AR forecast given $\vec{X}_t = (X_1, \ldots, X_t)$, is $P_m(\vec{X}_t) = \max\{F_m(Q_m(\vec{X}_t) \mid \vec{X}_t), 1 - F_m(Q_m(\vec{X}_t) \mid \vec{X}_t)\}$, where $F_m(\cdot \mid \vec{X}_t)$ is the conditional distribution function of X_{t+m} given $\vec{X}_t$ and $Q_m(\vec{X}_t)$ is the midpoint of the two forecasts. It can be shown that $P_m(\vec{X}_t)$ is not less than 1/2. Pemberton, by plotting $P_m(\vec{X}_t)$ against the values of the stationary density function of the true model, observed that the nonlinear conditional median forecast performs better at those forecast origin points whose probability density has low values than at those points whose probability density has high ones (in the sense that $P_m(\vec{X}_t)$ is far greater than or close to 0.5). Pemberton also conducted his simulation for the simple piecewise constant model:

$$X_n = \begin{cases} \alpha + \varepsilon_n & \text{if} \quad X_{n-1} \leqslant 0 \\ -\alpha + \varepsilon_n & \text{if} \quad X_{n-1} > 0 \end{cases} \tag{2}$$

with $\alpha = 3$ and fitted the data with a linear AR model by least squares method. Pemberton observed that (1) for lags $m = 1, 2, \ldots, 30$, the minima of $P_m(\vec{X}_t)$, RMSE and RMAE appear in a vicinity of the modes of the stationary distribution. The RMSE (RMAE) denotes the Ratio of the Mean Squared (Absolute) Error of the conditional mean (median, correspondingly) forecast by using the nonlinear model to that by using the AR model; (2) when $m \to \infty$, RMSE (decreasingly) and RMAE (not necessarily decreasingly) tend to one whereas $P_m(\vec{X}_t)$ tends to 1/2.

Recently, Tiao and Tsay (1994) compare the post-sample forecast of TAR model versus linear AR(2) model for US quarterly real GNP data by the ratio of mean squared forecast errors. The nonlinear TAR model only outperforms the linear AR model by a substantial margin in short-term forecasts at two regimes of the four forecast regimes.

All the above examples show that for the prediction problem, the nonlinear forecast may not be more accurate than the linear forecast, even when the true model is nonlinear and that both linear and nonlinear forecasts are asymptotically equivalent for long-term prediction. In order to better understand the poor forecast problem of the nonlinear time series models, the properties of the forecast performance criterion $P_m(\vec{X}_t)$ considered by both Davies *et al.* and Pemberton are studied in this paper. Under certain conditions, the criterion $P_m(\vec{X}_t)$ is shown to be an increasing function of the distance between linear and nonlinear forecasts. This result gives a nice explanation for the odd phenomena discussed by Pemberton (1989). In their simulation studies, since the distance between two forecasts are minimized around the modes of the stationary distribution of the process, the criterion $P_m(\vec{X}_t)$ is minimized at forecast regions of high probability density (see Example 1 below). For time series forecasting, usually either conditional means or conditional medians are used as optimal forecasts. However, from the above-mentioned studies and as will also be shown in this paper, the relative forecast performance (for both conditional mean and conditional median) of a nonlinear model to a linear model is rather sensitive to the forecast origins. Given a nonlinear time series, one can theoretically find a linear model which well approximates the true nonlinear models.

The organization of the paper is as follows. In the next section definitions and symbols are presented. In the third section the relation between $P_m(\vec{X}_t)$ and the distance between linear and nonlinear forecasts is exposed. An example is given to illustrate the odd phenomenon discussed by Pemberton. In the fourth section a new method is suggested which selects forecasts from conditional means or conditional medians by comparing their forecast performance with linear models. An example of piecewise constant SETAR model is also given. General conclusions will be given in the fifth section. All the technical proofs are given in the Appendix.

DEFINITIONS AND NOTATIONS

Consider the nonlinear autoregressive model of order p(NLAR(p))

$$X_n = \lambda_0(X_{n-1}, \ldots, X_{n-p}) + \varepsilon_n \tag{3}$$

where $\lambda_0(\cdot)$ is a measurable function from $\mathbf{R}^p$ to $\mathbf{R}$, ε_n's are i.i.d. r.v's with mean zero, variance σ^2 and a common probability density function $g(\cdot)$. Model (3) includes linear, SETAR and exponentially autoregressive (see Ozaki, 1985) models as special cases. Throughout this paper, the process $\{X_n\}$ defined by model (3) is assumed to be stationary. Assuming the observations $\vec{X}_t = (X_1, X_2, \ldots, X_t)$ (capitals are used for both random variables and realizations) is generated from model (3), the unknown value of X_{t+m} ($m \geqslant 1$) will be predicted by either the conditional mean or the conditional median when $\vec{X}_t$ is given. Since the conditional mean $E(X_{t+m} \mid \vec{X}_t)$ minimizes the mean squared error, it is also called the MMSE forecast. Similarly, the conditional median of X_{t+m} when $\vec{X}_t$ is given will be called the MMAE (Minimum Mean Absolute Error) forecast. In general, there are no closed forms of the MMSE and MMAE forecasts for NLAR(p) models, except some numerical algorithms that may be available. The goal of this paper is to compare the NLAR(p) forecast with the linear AR(p) forecast when the true model is factually NLAR(p). Define

$$P_m(\vec{X}_t) = P(\mid X_{t+m} - NL_m(\vec{X}_t) \mid < \mid X_{t+m} - AR_m(X_t) \parallel, \vec{X}_t) \tag{4}$$

where $NL_m(\vec{X}_t)$ is the nonlinear m-step MMSE or MMAE forecast and $AR_m(\vec{X}_t)$ is the linear AR(p) m-step conditional mean forecast. When $NL_m(\vec{X}_t) \neq AR_m(\vec{X}_t)$, it can be shown that

$$P_m(\vec{X}_t) = F_m(Q_m(\vec{X}_t) \mid \vec{X}_t) I_{[R^-]} + (1 - F_m(Q_m(\vec{X}_t) \mid \vec{X}_t) I_{[R^+]} \tag{5}$$

where $F_m(\cdot \mid \vec{X}_t)$ is the conditional distribution of X_{t+m} given $\vec{X}_t$, $Q_m(\vec{X}_t) = \frac{1}{2}(NL_m(\vec{X}_t) + AR_m(\vec{X}_t))$, $I_{[\cdot]}$ is the indicator function, $R^- = \{\vec{X}_t : NL_m(\vec{X}_t) < AR_m(\vec{X}_t)\}$, and $R^+ = \{\vec{X}_t : NL_m(\vec{X}_t) > AR_m(\vec{X}_t)\}$. The probability $P_m(\vec{X}_t)$ is considered as a criterion to compare the forecast performance, by Davies *et al.* as well as by Pemberton. Finally, we denote $\xi_m(\vec{X}_t) = \mid NL_m(\vec{X}_t) - AR_m(\vec{X}_t) \mid$, the distance between the nonlinear and linear forecasts at the forecast origin $\vec{X}_t$.

FORECAST ACCURACY AND FORECAST DISTANCE

In this section the comparison probability $P_m(\vec{X}_t)$ and the ratio of mean squared forecast errors are shown to be increasing functions of the forecast distance $\xi_m(\vec{X}_t)$ (see Theorems 1 and 5 below). Throughout this paper, $\mu_m(\vec{X}_t)$ denotes the m-step conditional mean and $M_m(\vec{X}_t)$ the m-step conditional median of the NLAR(p) model.

Theorem 1 Suppose that $(X_1, X_2, \ldots, X_t)$ is observed from the NLAR(p) model (3). If we write $F_m(x \mid \vec{X}_t) = F(x - NL_m(\vec{X}_t))$, then

$$P_m(\vec{X}_t) = F(\tfrac{1}{2}\xi_m(\vec{X}_t)) I_{[R^-]} + (1 - F(-\tfrac{1}{2}\xi_m(\vec{X}_t))) I_{[R^+]} \tag{6}$$

This equation shows that for the MMAE forecast, $P_m(\vec{X}_t) \geqslant 1/2$ a.s., if $P(\xi_m(\vec{X}_t) = 0) = 0$. If the distribution $F(\cdot)$ is continuous and symmetric about 0, then $P_m(\vec{X}_t) = F(\frac{1}{2}\xi_m(\vec{X}_t))$.

Corollary 2 Let $G(\cdot)$ denote the distribution function of the innovations. For multi-step forecast (i.e. $m > 1$), the distribution function F in Theorem 1 is given by $F(x) = E[G(x - \lambda_0(X_{t+m-1}, \ldots, X_{t+m-p}) + NL_m(\vec{X}_t)) \mid \vec{X}_t]$. Especially for the one-step MMSE forecast, since $NL_1(\vec{X}_t) = \lambda_0(X_t, \ldots, X_{t+1-p})$, we have $P_1(\vec{X}_t) = G(\frac{1}{2}\xi_1(\vec{X}_t)) I_{[R^-]} + (1 - G(-\frac{1}{2}\xi_1(\vec{X}_t))) I_{[R^+]}$.

The proofs of Theorem 1 and Corollary 2 are immediate and hence omitted.

Note that $P_m(\vec{X}_t)$ is the probability that the nonlinear forecast is better than the linear forecast. Therefore, we may say that the nonlinear forecast is better than the linear forecast if $P_m(\vec{X}_t) > 1/2$. Theorem 1 shows that the nonlinear MMAE forecast is not worse than the linear MMAE forecast. However, it should be noted that this conclusion is not true practically since the theoretic nonlinear forecast $NL_m(\vec{X}_t)$ is factually unknown in real situations. In real applications, the nonlinear AR function $\lambda_0(X_{t+m-1}, \ldots, X_{t+m-p})$ is usually unknown and only its estimates are available. Even if it were known, the forecast $NL_m(\vec{X}_t)$ cannot be exactly computed and only an approximation of it may be available. This shows that the nonlinear forecast may not be actually better than the linear forecast.

For the MMSE forecasts, the nonlinear forecast may not be better than the linear forecast, even theoretically. For example, when the distribution of the innovations satisfies $G(\frac{1}{2}\xi_1(\vec{X}_t)) < 1/2$ and the origin is in the set R^-; or $1 - G(-\frac{1}{2}\xi_1(\vec{X}_t)) < 1/2$ and the origin is in the set R^+. If the error in the computation of the MMSE nonlinear forecast is taken into account, the MMSE nonlinear forecast will further lose its superiority to the linear forecast.

If the regressor λ_0 has an explicit form and the distribution of the innovations is known, sometimes the conditional distribution $F_m(x \mid \vec{X}_t)$ of X_{t+m} can be explicitly expressed. As an example, consider the piecewise constant SETAR model:

$$X_n = \alpha_j + \varepsilon_n \qquad \text{if} \qquad X_{n-1} \in (c_{j-1}, c_j] \tag{7}$$

where $j = 1, 2, \ldots, l$, $-\infty = c_0 < c_1 < \cdots < c_l = \infty$, ε_n's are i.i.d. $N(0, 1)$ random variables.

Theorem 3 Consider the m-step MMSE forecast when the true model is (7). In this case, we have $NL_m(\vec{X}_t) = \mu_m(\vec{X}_t)$. When $X_t \in (c_{j-1}, c_j]$, $j = 1, 2, \ldots, l$, we have

$$P_m(\vec{X}_t) = P_m(X_t) = H_{m,j}\left(\frac{\xi_m(X_t)}{2}\right) I_{[R^-]} + \left(1 - H_{m,j}\left(-\frac{\xi_m(X_t)}{2}\right)\right) I_{[R^+]}$$

where $H_{m,j}(x) = \sum_{i=1}^{l} k_{ij}^{(m)} \Phi(x + \alpha_{ij}^{(m)})$, $\forall j = 1, 2, \ldots, l, k_{ij}^{(m)}$ are elements of the matrix K^{m-1} with $K = (\Phi(c_i - \alpha_j) - \Phi(c_{i-1} - \alpha_j))$ and $\alpha_{ij}^{(m)} = -\alpha_i + \sum_{i=1}^{l} k_{ij}^{(m)} \alpha_1$.

Corollary 4 For model (2), $NL_m(\vec{X}_t) = \mu_m(\vec{X}_t) = -\alpha\beta^{m-1}\,\mathrm{sign}(X_t)$, the comparison probability has the following explicit form:

$$P_m(X_t) = H_m\left(\frac{\xi_m(X_t)}{2}\right) I_{[R^-]} + \left(1 - H_m\left(-\frac{\xi_m(X_t)}{2}\right)\right) I_{[R^+]}$$

where $H_m(x) = k_1^{(m)}\Phi(x - 2\,\mathrm{sign}(X_t)\alpha k_2^{(m)}) + k_2^{(m)}\Phi(x + 2\,\mathrm{sign}(X_t)\alpha k_1^{(m)})$, $k_1^{(m)} = (1 - \beta^{m-1})/2$, $k_2^{(m)} = (1 + \beta^{m-1})/2$, and $\beta = 1 - 2\Phi(\alpha)$.

Now, consider the ratio of the mean squared (absolute) errors of the nonlinear forecasts to linear forecasts under model (3). Define

$$RMSE_m(\vec{X}_t) = \frac{[(X_{t+m} - AR_m(\vec{X}_t))^2 \mid \vec{X}_t]}{MSE(\mu_m(\vec{X}_t) \mid \vec{X}_t)}$$

where

$$\begin{aligned} MSE(\mu_m(\vec{X}_t) \mid \vec{X}_t) &= E[(X_{t+m} - \mu_m(\vec{X}_t))^2 \mid \vec{X}_t] \\ &= \sigma^2 + E[(\lambda_0(X_{t+m-1}, \ldots, X_{t+m-p}) - \mu_m(\vec{X}_t))^2 \mid \vec{X}_t] \end{aligned}$$

and define

$$RMAE_m(\vec{X}_t) = \frac{E[\,|(X_{t+m} - AR_m(\vec{X}_t) \| \vec{X}_t]}{MAE(M_m(\vec{X}_t) \mid \vec{X}_t)}$$

where $MAE(M_m(\vec{X}_t) \mid \vec{X}_t) = E[|\,(X_{t+m} - M_m(\vec{X}_t) \ \| \vec{X}_t]$.

The following theorem shows that both $RMSE_m(\vec{X}_t)$ and $RMAE(\vec{X}_t)$ are increasing functions of $\xi_m(\vec{X}_t)$.

Theorem 5 If model (3) is the true model, then

$$RMSE_m(\vec{X}_t) = 1 + \frac{(\xi_m(\vec{X}_t))^2}{MSE(\mu_m(\vec{X}_t) \mid \vec{X}_t)}$$

and

$$RMAE_m(\vec{X}_t) = 1 + \frac{2 \left| \int_{M_m(\vec{X}_t)}^{AR_m(\vec{X}_t)} (x - AR_m(\vec{X}_t)) \, \mathrm{d}F_m(x \mid \vec{X}_t) \right|}{MAE(M_m(\vec{X}_t) \mid \vec{X}_t)}$$

Next, let us consider the m-step forecast for large m when model (3) is the true model. We say that the time series $\{X_t\}$ defined by model (3) is causal if for all t, X_t can be written as a function of $\varepsilon_t, \varepsilon_{t-1}, \ldots$. In this case, we have $\sigma(X_t, X_{t-1}, \ldots) \subset \sigma(\varepsilon_t, \varepsilon_{t-1}, \ldots)$. When $\{X_t\}$ defined by model (3) is stationary and causal, we have

$$\mu_m(\vec{X}_t) = E(X_{t+m} \mid \vec{X}_t) \stackrel{\mathcal{D}}{=} E(X_0 \mid B^{-t-m}\vec{X}_t) \in \sigma(\varepsilon_{-m}, \varepsilon_{-m-1}, \ldots)$$

where B is the backward shift operator, that is, $BX_t = X_{t-1}$. By Kolmogorov's zero-one law, $\mu_m(\vec{X}_t)$ tends to a constant as $m \to \infty$. It is easy to see that the limit of $\mu_m(\vec{X}_t)$ is the mean of X_t. Similarly, we can show that $M_m(\vec{X}_t)$ tends to the median of X_t and that $AR_m(\vec{X}_t)$ tends to the mean of X_t. Thus, we proved the following theorem.

Theorem 6 In addition to the condition of Theorem 1, we assume that the time series $\{X_t\}$ is stationary and causal. When m is large, for the conditional mean nonlinear forecast, we have

$$\begin{gathered} P_m(\vec{X}_t) \approx F(0) I_{[R^-]} + (1 - F(0)) I_{[R^+]} \\ RMSE_m(\vec{X}_t) \to 1 \end{gathered} \tag{8}$$

and for the conditional median nonlinear forecast,

$$P_m(\vec{X}_t) \to \begin{cases} F(\frac{1}{2} \mid \text{mean} - \text{medium} \mid) & \text{if median} < \text{mean} \\ = \frac{1}{2} & \text{if median} = \text{mean} \\ 1 - F(-\frac{1}{2} \mid \text{mean} - \text{median} \mid) & \text{if median} > \text{mean} \end{cases} \tag{9}$$

$$RMAE_m(\vec{X}_t) \to 1 + \frac{2 \left| \int_{\text{median}}^{\text{mean}} (x - \text{mean}) \, \mathrm{d}F_m(x \mid \vec{X}_t) \right.}{E \mid X_t - \text{median} \mid}$$

Note that the $F(0)$ in model (8) may not be 1/2 since F is centerized at the mean, whereas $F(0) = 1/2$ in model (9) since F is centerized at the median.

In the following, we give an example to illustrate the results in this section.

Example 1 Suppose that $\vec{X}_t = (X_1, \ldots, X_t)$ is observed from model (2) and an AR(1) model is fitted to the data. The autoregressive coefficient ϕ is estimated by the least squares estimate

$$\hat{\phi} = \frac{\sum_{i=2}^{t} X_i X_{i-1}}{\sum_{i=1}^{t} X_i^2}$$

The linear conditional mean forecast is $AR_m(X_t) = \hat{\phi}^m X_t$ and the nonlinear forecast is $\mu_m(X_t) = -\alpha\beta^{m-1}\,\mathrm{sign}(X_t)$, where $\beta = 1 - 2\Phi(\alpha)$. The conditional distribution of X_{t+m} given $\vec{X}_t$ is $F_m(x \mid \vec{X}_t) = \frac{1}{2}[\Phi(x+\alpha)(1+\beta^{m-1}\,\mathrm{sign}(X_t)) + \Phi(x-\alpha)(1-\beta^{m-1}\,\mathrm{sign}(X_t))]$. From this, the conditional median forecast $M_m(\vec{X}_t)$ can be evaluated. Figures 1 and 2 give the plots of $\mu_m(X_t)$, $M_m(\vec{X}_t)$ and $AR_m(X_t)$ versus X_t, with $\alpha = 1.5$. The forecast distance of two conditional means is $\xi_m(X_t) = |\hat{\phi}^m X_t + \alpha\beta^{m-1}\,\mathrm{sign}(X_t)|$. Note that the least squares estimate $\hat{\phi}$ converges to

$$\rho = \frac{E(X_1 X_2)}{E(X_1^2)} = -\frac{\alpha(2\varphi(\alpha) - \alpha\beta)}{1+\alpha^2}$$

as $t \to \infty$, where φ is the density function of the standard normal distribution. If $\alpha = 1.5$, then $\hat{\phi} \cong -0.7$. Figure 3 is the plot of $\xi_m(X_t)$ versus X_t, Figures 4–6 the plots of $P_m(X_t)$ (given by

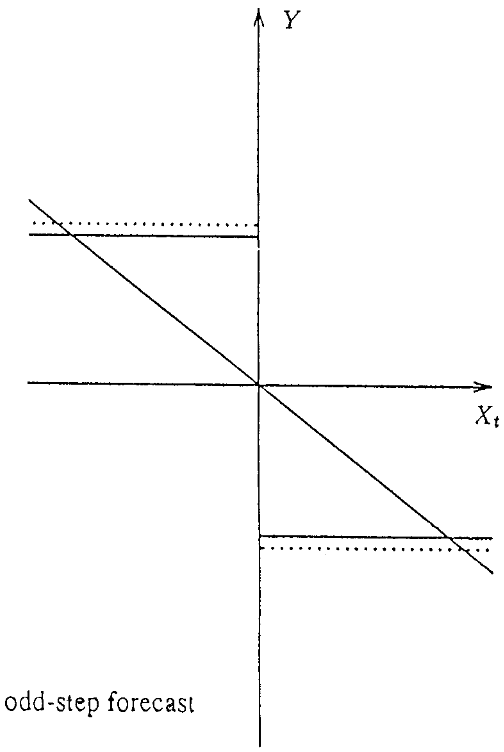

Figure 1. Conditional mean and conditional median forecasts versus origins X_1 for the AR(1) model and model (2) with $\alpha = 1.5$: odd-step forecast. ——— Conditional mean, ······· conditional median

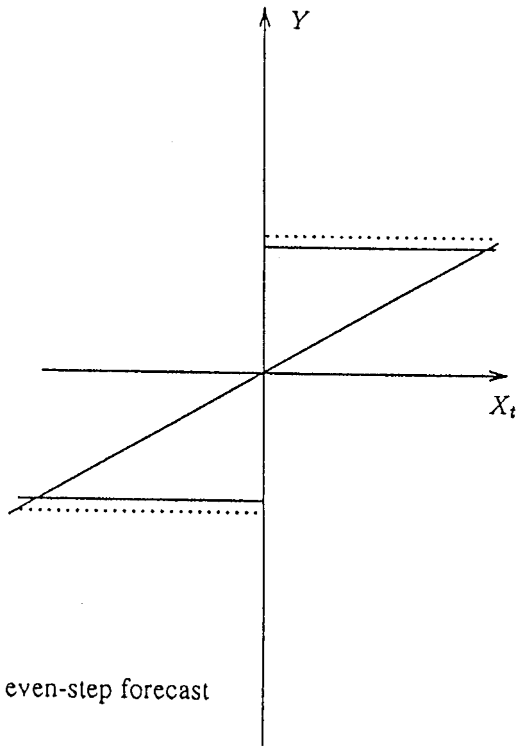

Figure 2. Even-step forecast

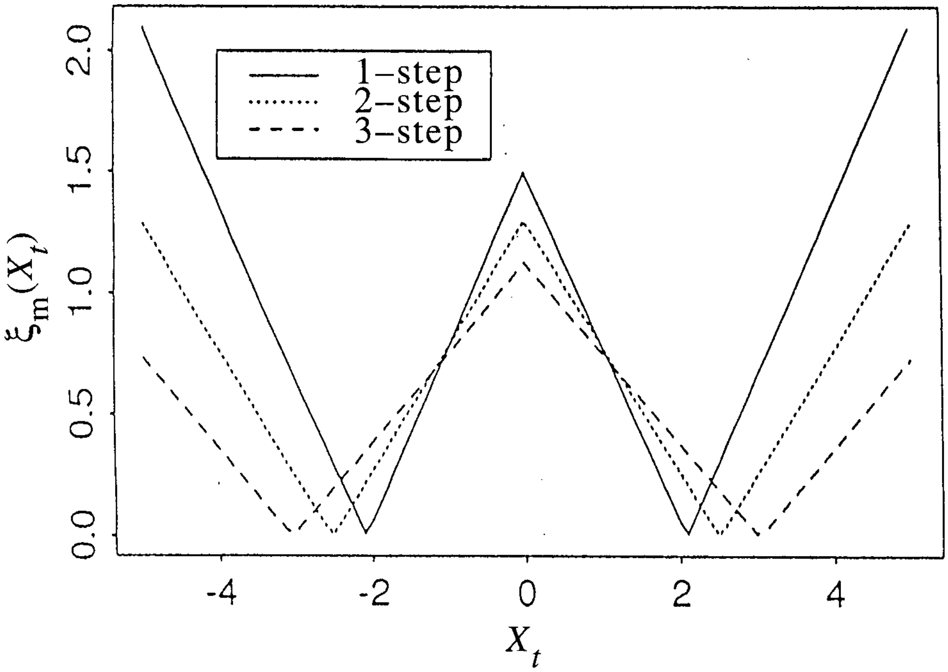

Figure 3

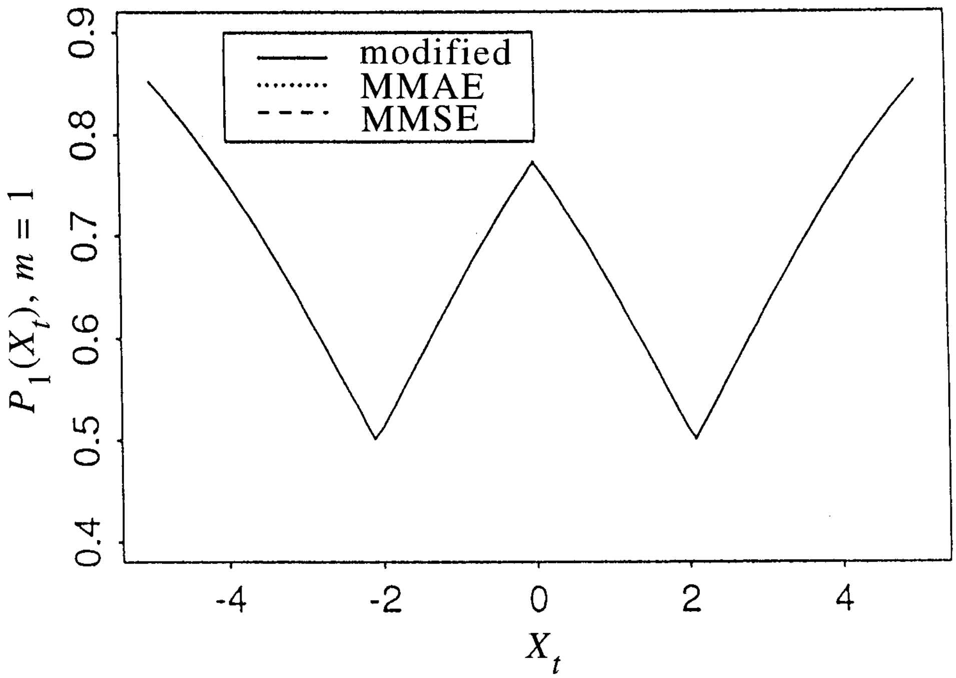

Figure 4

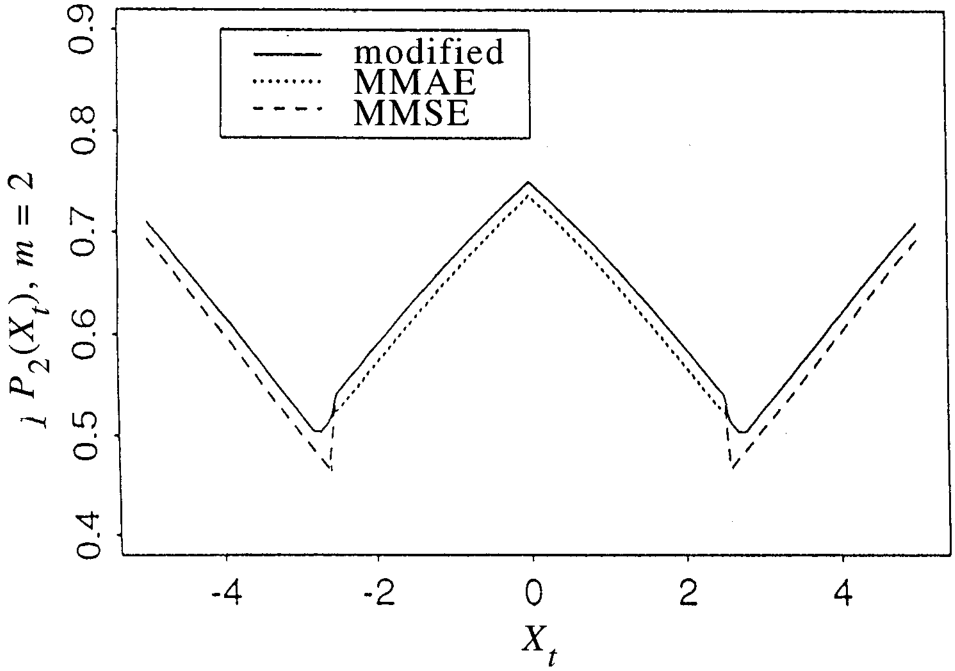

Figure 5

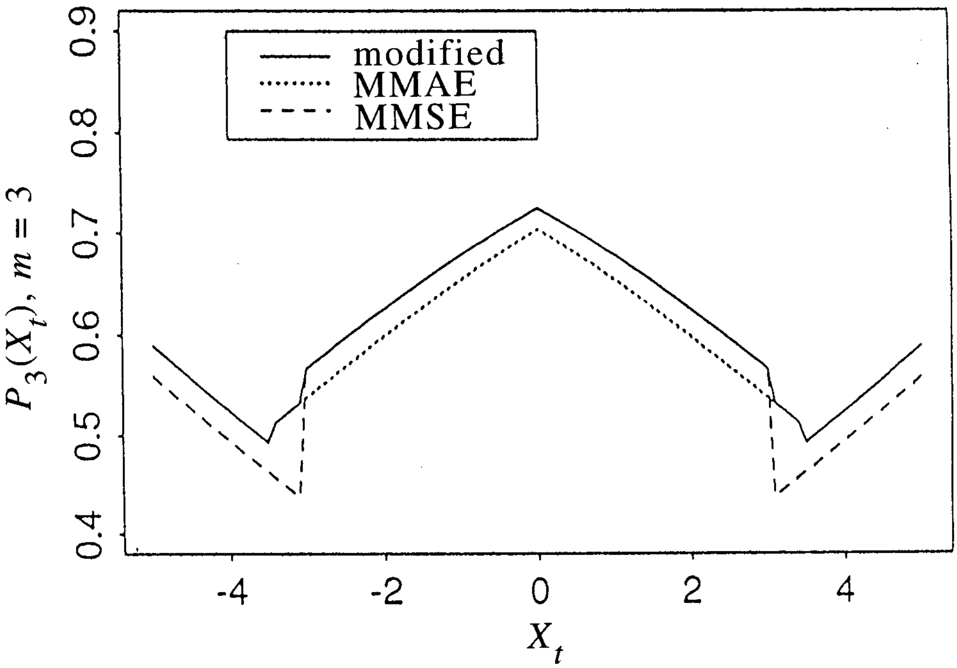

Figure 6

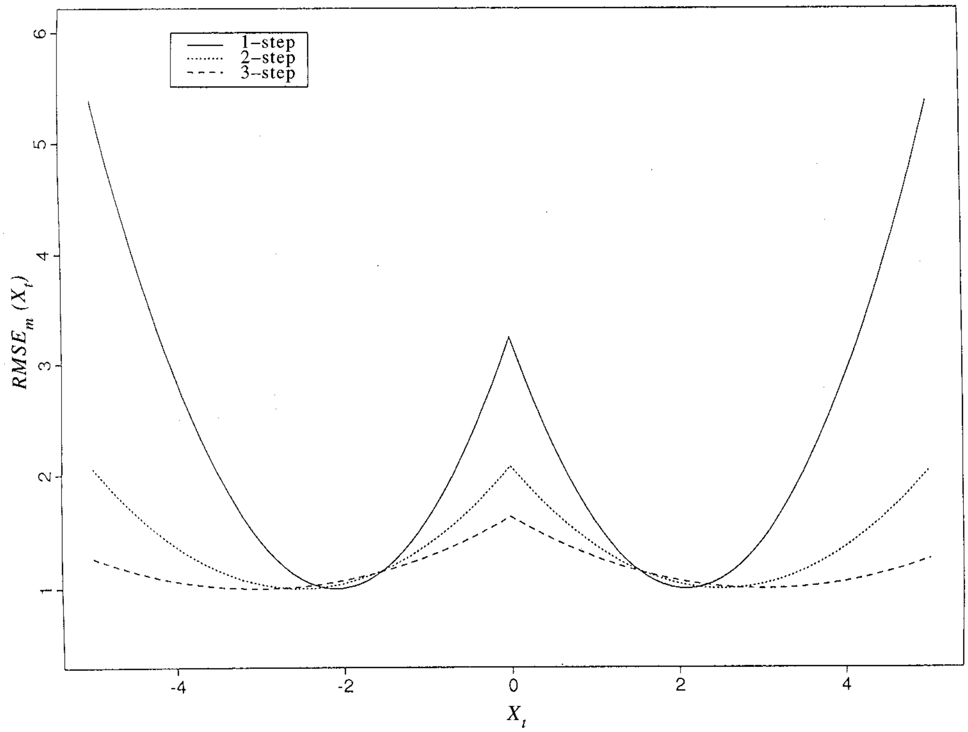

Figure 7. $RMSE_m(X_t)$ versus X_t for the AR(1) model and model (2) with $\alpha = 1.5$

Corollary 4) versus X_t (the definition of the modified forecast is given in Theorem 7) and Figure 7 the plot of $RMSE_m(X_t)$ versus X_t, where

$$RMSE_m(X_t) = 1 + \frac{(\xi_m(X_t))^2}{1 + \alpha^2(1 - \beta^{2m-2})}$$

Clearly, from Figure 7 it can be seen that the $RMSE_m(X_t)$ is not necessarily a decreasing function in m at all origins X_t. This is contrary to Pemberton's observation about $RMSE_m(\vec{X}_t)$ as mentioned earlier. When t is large, the intersection of the two m-step conditional means given X_t (by equating the two conditional means) is given by

$$|X_t| = \frac{-\alpha\beta^{m-1}}{\hat{\varphi}^m} \approx \left| \frac{\alpha}{\beta} \left[\frac{\alpha^2 + 1}{\alpha^2 + \Delta(\alpha)} \right]^m \right| \to \infty$$

as $m \to \infty$ where

$$\Delta(\alpha) = -\frac{2\alpha\varphi(\alpha)}{\beta} \in (0, 1)$$

When $\alpha = 1.5$, the stationary density

$$f(x) = \frac{\varphi(x + \alpha) + \varphi(x - \alpha)}{2}$$

Table I. The intersection points of model (2) with $\alpha = 1.5$ and AR(1) model when $X_t \leqslant 0$

Step m	1	2	3	4	5	6	7	8	9	10
$-\alpha\beta^{m-1}/\hat{\phi}^m$	2.08	2.51	3.02	3.64	4.38	5.28	6.36	7.66	9.23	11.12

of model (2) has modes around $x = \pm 1.5$ and the intersection points for $x_t \leqslant 0$ are given in Table I. Since the intersection points of the two forecasts are not in the neighbourhood of the modes of the stationary density function, contrary to Pemberton's observation (see above), the poor forecast region is not in the region of high density. See Figures 8–11 for the plots of $P_m(\vec{X}_t)$ of $\mu_m(\vec{X}_t)$ versus the stationary density of X_t. However, if α is large but m is not, then $\beta^{m-1}/\hat{\phi}^m \approx 1$ and $\alpha\beta^{m-1}/\hat{\phi}^m \approx \alpha$. For example, in Pemberton's experiment $\alpha = 3$, the intersection points are close to the modes $X_t = \pm 3$ of the stationary p.d.f. Thus, Pemberton observe a minimum of $P_m(X_t)$ and $RMSE_m(X_t)$ in the vicinity of the modes when m is small. Furthermore, since for all X_t, $\xi_m(X_t) \to 0$ as $m \to \infty$, thus $P_m(X_t) \to 1/2$ and $RMSE_m(X_t) \to 1$ as $m \to \infty$, which implies the multi-step forecast performance of linear model is asymptotically as good as that of the true nonlinear piecewise constant model.

A NEWLY PROPOSED FORECAST

Calculating the nonlinear conditional mean forecast, in general, is much easier than computing the conditional median forecast. However, the former is sometimes worse than the AR predictor whose computation is the easiest among the three. If one wants to use the nonlinear forecast for

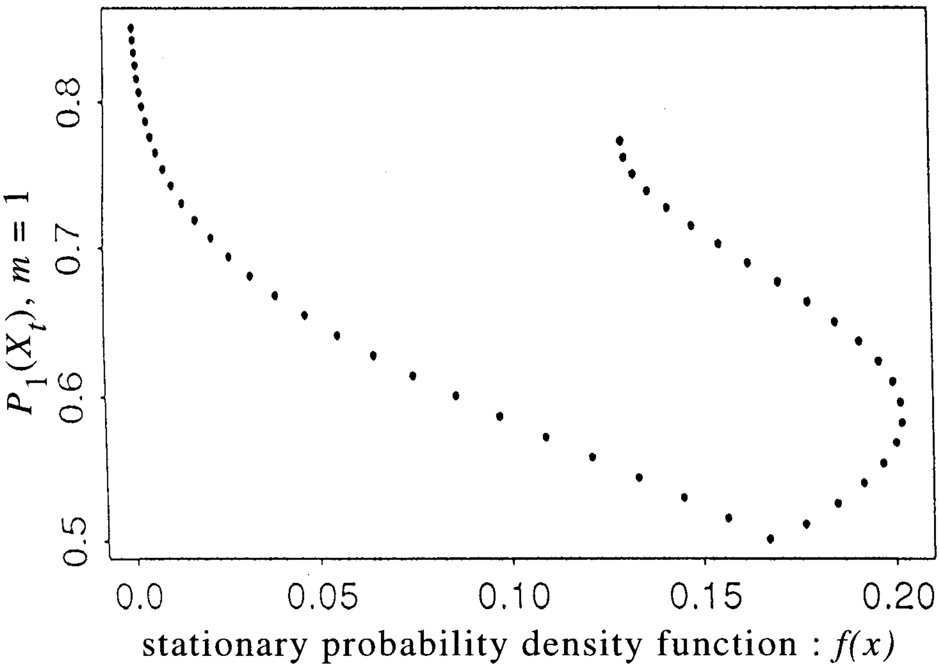

Figure 8

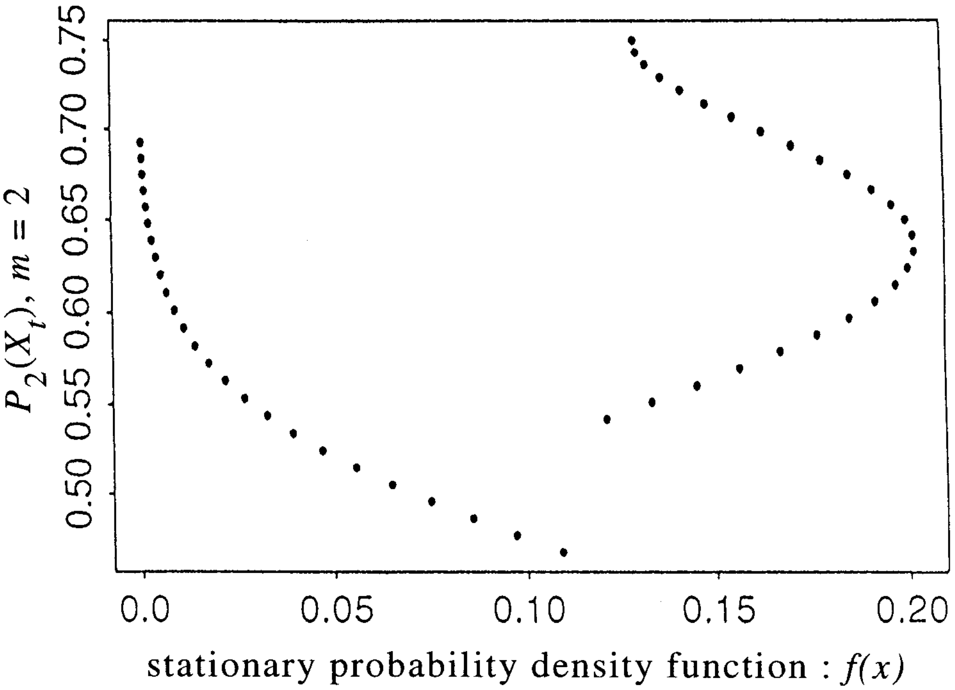

Figure 9

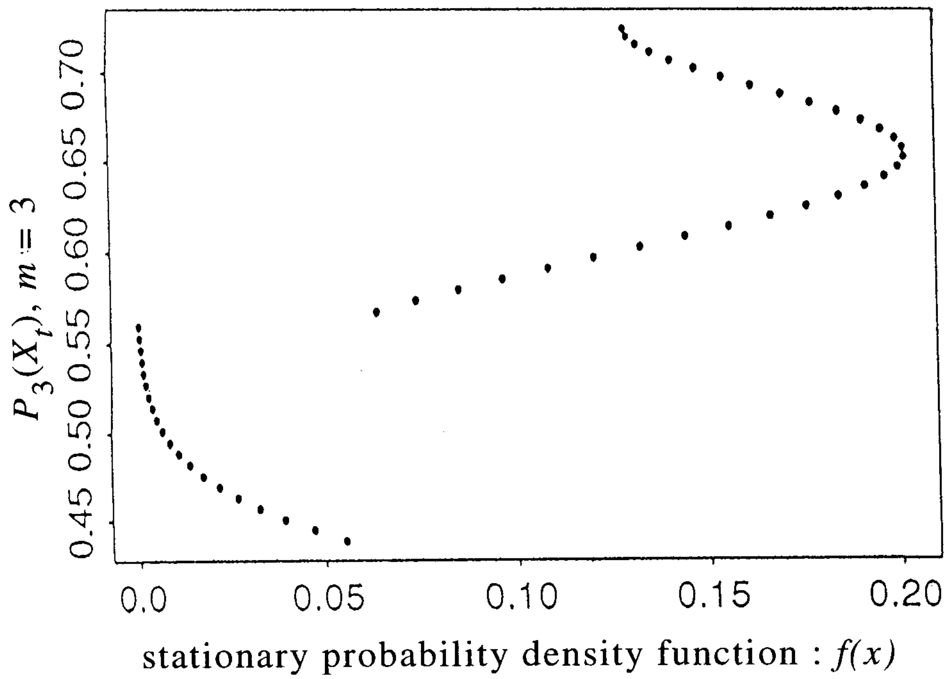

Figure 10

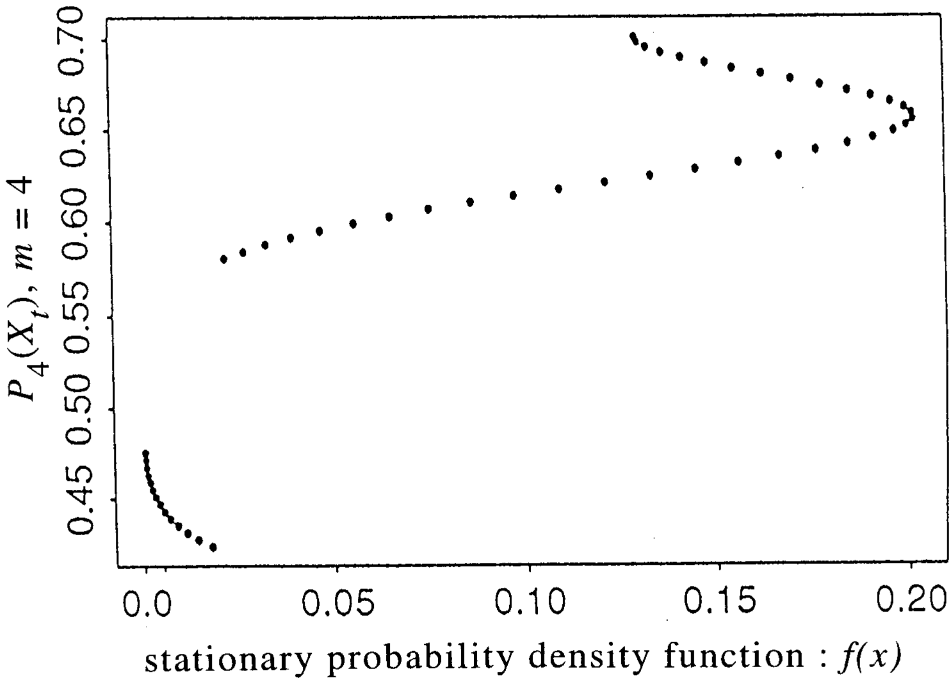

Figure 11

not losing the benefits of the nonlinear modelling, we suggest a new forecast which has the highest comparison probability relative to the AR forecast.

Consider the case when the true model is a nonlinear one. Then the m-step conditional probability distribution function of X_{t+m} given $\vec{X}_t$ is usually skewed. Thus, the MMSE forecast is liable to be influenced by the skewness of the conditional distribution; furthermore, its $P_m(\vec{X}_t)$ is sometimes less than 0.5. For the MMAE forecast, although its comparison probability $P_m(\vec{X}_t)$ is always greater than or equal to 0.5, yet its value is more difficult to compute. Therefore, we suggest an alternative forecast, combining the MMSE forecast, and the MMAE forecast, which has a higher comparison probability $P_m(\vec{X}_t)$. The proposed forecast $MF_m(\vec{X}_t)$ is defined as

$$\begin{aligned} MF_m(\vec{X}_t) &= \mu_m(\vec{X}_t)I_{[R^-]} + M_m(\vec{X}_t)I_{[R^+]} \qquad \text{if} \qquad \mu_m(\vec{X}_t) > M_m(\vec{X}_t) \\ &= M_m(\vec{X}_t)I_{[R^-]} + \mu_m(\vec{X}_t)I_{[R^+]} \qquad \text{if} \qquad \mu_m(\vec{X}_t) \leqslant M_m(\vec{X}_t) \end{aligned}$$

where $R^- = \{\vec{X}_t : \mu_m(\vec{X}_t) \leqslant AR_m(\vec{X}_t)\}$ and $R^+ = \{\vec{X}_t : \mu_m(\vec{X}_t) > AR_m(\vec{X}_t)\}$. If the true model is nonlinear, theoretically we can still find a reasonable approximate linear model. Then, the new forecast is the nonlinear MMSE forecast unless its relative forecast performance is worse than the MMAE forecast, which certainly includes the case when the MMSE performs worse than the linear forecast. In Example 2, the $MF_m(\vec{X}_t)$ forecast of model (2) is shown to be the MMSE forecast for $\vec{X}_t$ lying in a high stationary density region. In Theorem 7, assuming the true model is nonlinear, $P_m(\vec{X}_t)$ of the forecast $MF_m(\vec{X}_t)$ is shown to be the maximum of those of MMSE and MMAE forecasts.

To indicate the predictor used in the comparison probability, let $P_m(\vec{X}_t, \mu_m)$, $P_m(\vec{X}_t, M_m)$ and $P_m(\vec{X}_t, MF_m)$ denote the $P_m(\vec{X}_t)$ of $\mu_m(\vec{X}_t)$, $M_m(\vec{X}_t)$ and $MF_m(\vec{X}_t)$ respectively.

Theorem 7 If model (3) is the true model, then

$$P_m(\vec{X}_t, MF_m) = \max\{P_m(\vec{X}_t, \mu_m), P_m(\vec{X}_t, M_m)\} \geqslant \tfrac{1}{2}$$

Remark It seems that the computation of MF_m is more complicated than computing $M_m(\vec{X}_t)$ for $M_m(\vec{X}_t)$ is involved in the definition of MF_m. However, in many cases, the relation between $\mu_m(\vec{X}_t)$ and $M_m(\vec{X}_t)$ can be determined before computing them (see Example 2 below). Thus, $M_m(\vec{X}_t)$ needs only be computed when $MF_m = M_m$.

Example 2 (Example 1, continued) The m-step conditional density of X_{t+m} given X_t for model (2) is

$$f_m(x \mid X_t) = k_1^{(m)} \varphi(x - \text{sign}(X_t)\alpha) + k_2^{(m)} \varphi(x + \text{sign}(X_t)\alpha)$$

where $\beta = 1 - 2\Phi(\alpha)$, $k_1^{(m)}$ and $k_2^{(m)}$ are given in Corollary 4. The skewness of $f_m(x \mid x_t)$ is affected by both m and β which determine the weights of the linear combination of $\varphi(x + \alpha)$ and $\varphi(x - \alpha)$. See Figures 12 and 13 for the plots of $f_m(x \mid x_t \leqslant 0)$ versus x when $\alpha = 1, 2$, $m = 1, 2, 3, 4, 5$. It can be proved (see the Appendix) that

(1) $\mu_m(X_t) \leqslant M_m(X_t)$ if $X_t \leqslant 0$ and m is odd or $X_t > 0$ and m is even
(2) $\mu_m(X_t) > M_m(X_t)$ if $X_t \leqslant 0$ and m is even or $X_t > 0$ and m is odd.

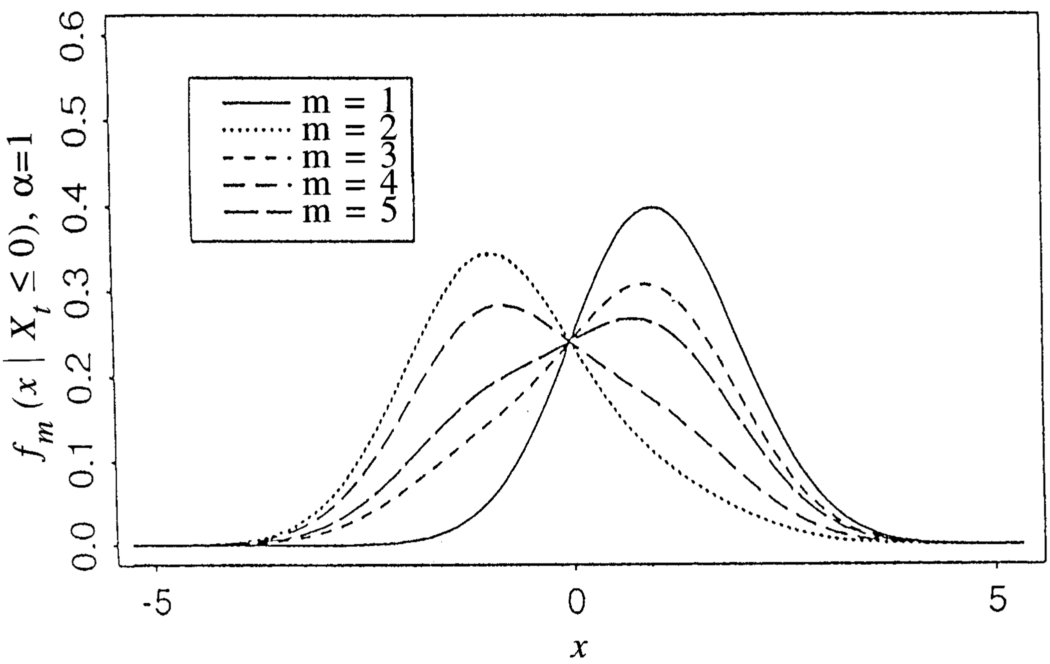

Figure 12

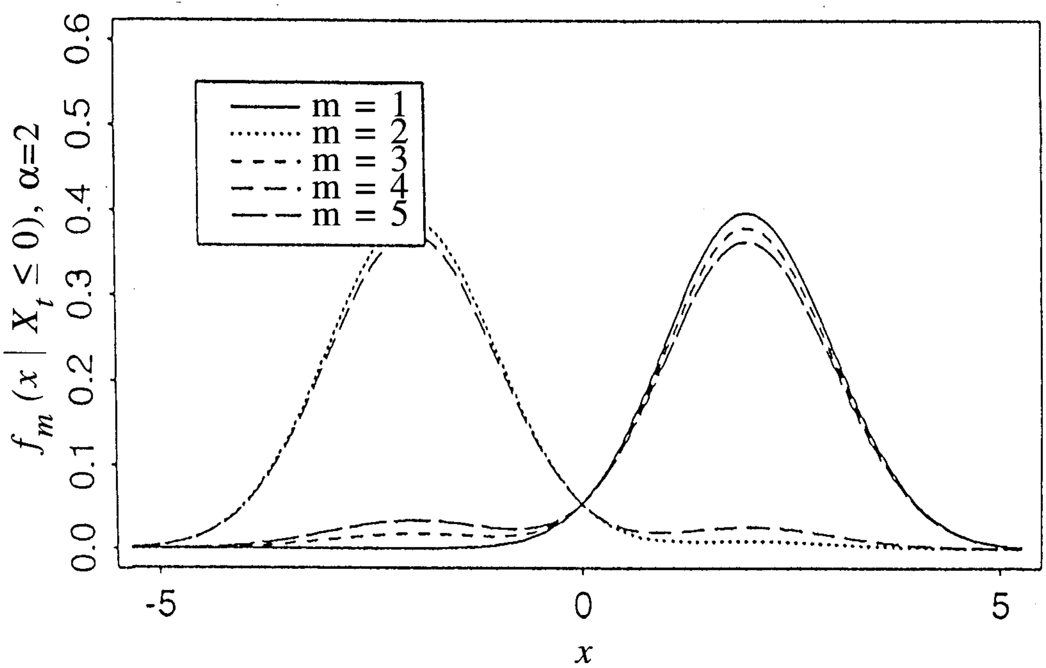

Figure 13

The new forecast of model (2) with $\alpha = 1.5$ can be simplified as $MF_m(X_t) = \mu_m(X_t)I_{[K_m]} + M_m(X_t)I_{[K_m^c]}$, where $\mu_m(X_t) = -\alpha\beta^{m-1}\,\text{sign}(X_t)$, $K_m = \{X_t : |X_t| \leqslant |\alpha\beta^{m-1}/\hat{\phi}^m|\}$, $K_m^c = \{X_t : |X_t| > |\alpha\beta^{m-1}/\hat{\phi}^m|\}$ and $\hat{\phi} \cong -0.7$. See Table I in Example 1 for the values of $\alpha\beta^{m-1}/\hat{\phi}^m$ for $m = 1, 2, \ldots, 10$. Therefore the proposed forecast $MF_m(X_t)$ is the m-step conditional mean for origins X_t in high stationary probability density regions and is the m-step conditional median for origins X_t in low stationary density regions. The plots of $P_m(X_t)$ versus X_t for $MF_m(X_t)$, $\mu_m(X_t)$ and $M_m(X_t)$, $m = 1, 2, 3$ are given in Figures 4–6.

DISCUSSION AND CONCLUSIONS

If the true model is completely known or partially known (or known functional form with unknown parameters) and the criterion is chosen as the RMSE (RMAE), then the best forecast is the conditional mean (median) forecast. However, if the comparison probability $P_m(\vec{X}_t)$ is chosen as the criterion, the case becomes complicated. In the theorems in the third section, the expressions of $P_m(\vec{X}_t)$ are given. It has been proven that $P_m(\vec{X}_t)$ is an increasing function of the forecast distance $\xi_m(\vec{X}_t)$ for origin $\vec{X}_t$ in disjoint regions. To determine which predictor should be used, one has to compute the two predictors in comparison. This is seemingly not a good deal because the purpose is to use the linear forecast to save computation time. However, there may be alternative ways to determine which predictor should be used. As illustrated in Example 2, when computing the $MF_m(\vec{X}_t)$, one can determine whether $M_m(\vec{X}_t)$ is needed. Also, as noted by

Pemberton, when $\vec{X}_t$ is in the region of high density values, the distance $\xi_m(\bar{X}_t)$ generally (although not always) has smaller values and hence the linear and nonlinear forecasts are comparable.

The newly proposed forecast provides an alternative way of selecting conditional mean or conditional median forecast when the m-step conditional distribution is skewed. However, when the m-step conditional distribution is not of a location scale family or when the model is more complicated than the piecewise constant SETAR model the property of $P_m(\vec{X}_t)$ still needs further studies.

APPENDIX

Proof of Theorem 3 Since the conditional distribution of X_{t+m} given $X_t \in (c_{j-1}, \; c_j]$ is $F_{m,j}(x) = \sum_{i=1}^{l} k_{ij}^{(m)} \Phi(x - \alpha_i)$ (see Pemberton, 1990),

$$F_{m,j}(Q_m(X_t)) = \left[\sum_{i=1}^{l} k_{ij}^{(m)} \Phi\left(\frac{\xi_m(X_t)}{2} + \mu_m(X_t) - \alpha_i\right)\right] I_{[R^-]} + \left[\sum_{i=1}^{l} k_{ij}^{(m)} \Phi\left(-\frac{\xi_m(X_t)}{2} + \mu_m(X_t) - \alpha_i\right)\right] I_{[R^+]}$$

Given $X_t \in (c_{j-1}, c_j]$, $\mu_m(X_t) = \sum_{i=1}^{l} k_{ij}^{(m)} \alpha_i$. Therefore, the result is obtained by setting $\alpha_{ij}^{(m)} = \mu_m(X_t) - \alpha_i$ and by model (5). QED

Proof of Corollary 4 The conditional distribution of X_{t+m} given X_t is

$$F_m(x) = k_2^{(m)} \Phi(x + \text{sign}(X_t)\alpha) + k_1^{(m)} \Phi(x - \text{sign}(X_t)\alpha)$$

Therefore, by model (5), for X_t in R^-

$$\begin{aligned} P_m(X_t) &= F_m(Q_m(X_t)) \\ &= k_2^{(m)} \Phi(Q_m(X_t) + \text{sign}(X_t)\alpha) + k_1^{(m)} \Phi(Q_m(X_t) - \text{sign}(X_t)\alpha) \\ &= k_2^{(m)} \Phi(\tfrac{1}{2}\xi_m(X_t) - \alpha\beta^{m-1} \text{ sign}(X_t) + \text{sign}(X_t)\alpha) \\ &\quad + k_1^{(m)} \Phi(\tfrac{1}{2}\xi_m(X_t) - \alpha\beta^{m-1} \text{ sign}(X_t) - \text{sign}(X_t)\alpha) \\ &= k_2^{(m)} \Phi(\tfrac{1}{2}\xi_m(X_t) + 2\alpha \text{ sign}(X_t) k_1^{(m)}) + k_1^{(m)} \Phi(\tfrac{1}{2}\xi_m(X_t) - 2\alpha \text{ sign}(X_t) k_2^{(m)}) \\ &= H_m(\tfrac{1}{2}\xi_m(X_t)) \end{aligned} \tag{A1}$$

Similarly, one can prove for X_t in R^+

$$P_m(X_t) = 1 - F_m(Q_m(X_t)) = 1 - H_m(-\tfrac{1}{2}\xi_m(X_t)) \tag{A2}$$

QED

Proof of Theorem 5 Since

$$E[(X_{t+m} - AR_m(\vec{X}_t))^2 \mid \vec{X}_t] = MSE(\mu_m(\vec{X}_t) \mid \vec{X}_t) + (AR_m(\vec{X}_t) - \mu_m(\vec{X}_t))^2$$

it is easy to derive model (8). Furthermore, since

$$\begin{aligned} E|x-a| - E|x-b| &= (b-a)(1-2F(a)) - 2\int_a^b (b-x)\,\mathrm{d}F(x), \quad \text{if} \quad b > a \\ &= (a-b)(2F(a)-1) - 2\int_b^a (x-b)\,\mathrm{d}F(x), \quad \text{if} \quad b < a \end{aligned}$$

thus model (9) is derived by letting $a = M_m(\vec{X}_t)$ and $b = AR_m(\vec{X}_t)$.

Proof of Theorem 7 First consider the case when $\mu_m(\vec{X}_t) > M_m(\vec{X}_t)$.

(1) If $AR_m(\vec{X}_t) \geqslant \mu_m(\vec{X}_t)$, then $MF_m(\vec{X}_t) = \mu_m(\vec{X}_t)$ and

$$\begin{aligned} P_m(\vec{X}_t, MF_m) &= F_m(Q_m(\mu_m) \mid \vec{X}_t) \\ &> F_m(Q_m(M_m) \mid \vec{X}_t)(\geqslant \tfrac{1}{2}) \\ &= P_m(\vec{X}_t, M_m) \end{aligned}$$

where $F_m(\cdot \mid \vec{X}_t)$ is the conditional distribution of X_{t+m} given $\vec{X}_t$, $Q_m(\mu_m) = \frac{1}{2}(\mu_m(\vec{X}_t) + AR_m(\vec{X}_t))$ and $Q_m(M_m) = \frac{1}{2}(M_m(\vec{X}_t) + AR_m(\vec{X}_t))$.

(2) If $M_m(\vec{X}_t) \leqslant AR_m(\vec{X}_t) < \mu_m(\vec{X}_t)$, then $MF_m(\vec{X}_t) = M_m(\vec{X}_t)$ and

$$\begin{aligned} P_m(\vec{X}_t, MF_m) &= F_m(Q_m(M_m) \mid \vec{X}_t) \\ &\geqslant F_m(M_m(X_t) \mid \vec{X}_t) = 1 - F_m(M_m(X_t) \mid \vec{X}_t)(= \tfrac{1}{2}) \\ &> 1 - F_m(Q_m(\mu_m) \mid \vec{X}_t) = P_m(\vec{X}_t, \mu_m) \end{aligned}$$

(3) If $AR_m(\vec{X}_t) < M_m(\vec{X}_t)$, then $MF_m(\vec{X}_t) = M_m(\vec{X}_t)$ and

$$\begin{aligned} P_m(\vec{X}_t, MF_m) &= 1 - F_m(Q_m(M_m) \mid \vec{X}_t)(\geqslant \tfrac{1}{2}) \\ &> 1 - F_m(Q_m(\mu_m) \mid \vec{X}_t) \\ &= P_m(\vec{X}_t, \mu_m) \end{aligned}$$

When $\mu_m(\vec{X}_t) \leqslant M_m(\vec{X}_t)$, the result can be obtained similarly. QED

Proof of (1) and (2) in Example 2 Since $\varphi(x)$ is decreasing for $x \geqslant 0$,

$$\frac{1}{k}\int_0^{2\alpha k} \varphi(x)\,\mathrm{d}x = \frac{1}{k}\left(\Phi(2\alpha k) - \frac{1}{2}\right)$$

is decreasing in k. First consider the case for $X_t \leqslant 0$. Let $F_m(\cdot \mid X_t)$ denote the conditional distribution of X_{t+m} given X_t, then

$$\begin{aligned} F_m(\mu_m(X_t) \mid X_t) - \frac{1}{2} &= k_1^{(m)}\Phi(2\alpha k_2^{(m)}) + k_2^{(m)}\Phi(-2\alpha k_1^{(m)}) - \frac{1}{2} \\ &= k_1^{(m)} k_2^{(m)} \left[\frac{1}{k_2^{(m)}}\left(\Phi(2\alpha k_2^{(m)}) - \frac{1}{2}\right) - \frac{1}{k_1^{(m)}}\left(\Phi(2\alpha k_1^{(m)}) - \frac{1}{2}\right)\right] \\ &\geqslant 0 \quad \text{if and only if } k_2^{(m)} \leqslant k_1^{(m)} \end{aligned}$$

Thus, $\mu_m(X_t) \geqslant M_m(X_t) \Leftrightarrow F_m(\mu_m(X_t) \mid X_t) \geqslant \frac{1}{2} \Leftrightarrow k_2^{(m)} \leqslant k_1^{(m)} \Leftrightarrow m$ is even, since $\beta < 0$.

The case when $X_t > 0$ can be similarly proved. QED

ACKNOWLEDGEMENTS

The authors are grateful to Zhidong Bai for his helpful comments and suggestions. The work presented in this paper was supported by the National Science Council of the Republic of China, Grant No. NSC 82-0208-N-110-064.

REFERENCES

Davies, N., Petruccelli, J. D. and Pemberton, J., 'An automatic procedure for identification, estimation and forecasting self exciting threshold autoregressive models', *Statistician*, **37** (1988), 199–204.

Oster, G. and Ipaktchi, A., 'Population cycles', *Theor. Chem. Periodicity*, in Eyring, H. and Henderson, D. (eds), *Chem & Biol*, New York: Academic Press, 111–32.

Ozaki, T., 'Nonlinear time series models and dynamical systems', in Hannan, E. J., Krishnaiah, P. R. and Rao, M. M. (eds), *Handbook of Statistics 5*, Amsterdam: North-Holland, 1985, 25–83.

Pemberton, J., 'Forecast accuracy of nonlinear time series models', *Tech. Rep. 77*, Statistics Research Center, Graduate School of Business, University of Chicago, 1989.

Pemberton, J., 'Piecewise constant autoregressive models for univariate time series', *Tech. Rep. MCS-90-04*, Department of Mathematics & Computer Science, University of Salford, 1990.

Priestly, M. B., *Spectral Analysis and Time Series*, New York: Academic Press, 1989.

Tiao, G. C. and Tsay, R. S., 'Some advances in nonlinear and adaptive modeling in time series', *J. Forecast.*, **13** (1994), 109–31.

Tong, H., *Non-linear Time Series: A dynamical systems approach*, Oxford: Oxford University Press, 1990.

Authors' address:
Meihui Guo and **Y. K. Tseng**, Department of Applied Mathematics, National Sun Yat-sen University, Kaohsiung, ROC 80424.

