Stat. Meth. & Appl. (2007) 16:395–410
DOI 10.1007/s10260-007-0044-x

ORIGINAL ARTICLE

Multi-step forecasts from threshold ARMA models using asymmetric loss functions

Marcella Niglio

Accepted: 20 December 2006 / Published online: 1 February 2007

Abstract The forecasts generation from nonlinear time series models is investigated under general loss functions. After presenting the main results and some relevant features of these functions, the Linex loss has been used to generate multi-step forecasts from threshold autoregressive moving average models showing their main properties and some results connected to a proper transformation of the forecast errors. A simulation exercise highlights interesting properties of the proposed predictors, both in terms of their bias and their distribution, further clarifying how the Linex predictor can be helpful in empirical applications.

Keywords Nonlinear prediction · General loss functions · SETARMA · Linex

1 Introduction

The forecasts generation from nonlinear time series models has been differently faced and in some cases their predictive ability, evaluated in terms of forecast accuracy, has been widely discussed. In this context, among the others, Tong (1990), Fan and Yao (2003) present the main features of nonlinear predictors further comparing their properties with those well known in the linear domain.

It is common practice in nonlinear context to make use of quadratic loss functions to generate predictors even when more general functions should be preferred.

M. Niglio (✉)
Dipartimento di Scienze Economiche e Statistiche,
Università degli Studi di Salerno, Via Ponte Don Melillo, 84084 Fisciano (SA), Italy
e-mail: mniglio@unisa.it

The motivation behind this choice is often related to analytical problems that arise when closed-form predictors are generated. In fact they are more easily obtained in the presence of quadratic losses even if the symmetry of these loss functions can disregard some properties of the process upon which the model has been estimated. For example when nonlinear models are used to catch asymmetric effects in the data, even asymmetric loss functions should be selected to generate predictors that differently treat positive and negative forecast errors.

In this paper we discuss the use of general loss functions to generate forecasts from nonlinear time series models. The seminal paper on this topic is quite dated even if the interest and the main results of the literature are relatively recent. In fact Granger (1969) firstly states the limit of the classical prediction theory based only on the use of quadratic loss functions and hopes for more general loss functions.

In particular, after the presentation of some general results and assumptions upon which the paper is founded, in Sect. 2 the main loss functions used in the forecasting literature are presented and compared. Section 3 introduces the nonlinear Self Exciting Threshold Autoregressive Moving Average model whose h-step ahead predictor, obtained though the Linex loss function, is shown and discussed in Sect. 4. Section 5 studies, in a Monte Carlo simulation, the properties of the proposed predictor and describes how it can be of interest in empirical studies.

Given a univariate time series $\mathbf{Y} = \{y_1, y_2, \ldots\}$ denote with $\hat{y}_{t+h|t}$ the h-step ahead predictor conditional to the information set $\Omega_t = \{y_t, \ldots, y_2, y_1\}$. After the selection of a convex and differentiable loss function $L(\cdot)$, such that $L(0) = 0$, $L(x) > 0$, $\forall x \in \mathbb{R} - \{0\}$, with first derivative $L'(x) \geq 0$, for $x > 0$, and $L'(x) \leq 0$ for $x < 0$, $\hat{y}_{t+h|t}$ is obtained minimizing $E_t\left[L(y_{t+h} - \hat{y}_{t+h|t})\right]$:

$$\min_{\hat{y}_{t+h|t}} E_t\left[L\left(y_{t+h} - \hat{y}_{t+h|t}\right)\right], \tag{1}$$

where $E_t[\cdot]$ denotes the conditional expectation to the information set Ω_t, h is the lead time, y_{t+h} is the observed value at time $t+h$ and $y_{t+h} - \hat{y}_{t+h|t} = \hat{e}_{t+h|t}$ is the forecast error.

Under the further assumptions:

A1. $\mathbf{Y}$ is a compact set of $\mathbb{R}$

A2. $E_t\left[L\left(y_{t+h} - \hat{y}_{t+h|t}\right)\right] = \int L(y_{t+h} - \hat{y}_{t+h|t}) f_{t+h|t}(y)dy < \infty$, with $f_{t+h|t}(y)$ the conditional density function of y_{t+h}

A3.

$$\frac{\partial E_t\left[L\left(y_{t+h} - \hat{y}_{t+h|t}\right)\right]}{\partial \hat{y}_{t+h|t}} = E_t\left[\frac{\partial L\left(y_{t+h} - \hat{y}_{t+h|t}\right)}{\partial \hat{y}_{t+h|t}}\right] < \infty,$$

the predictor which minimizes function (1) is obtained from the following first order condition:

$$\int L'(y_{t+h} - \hat{y}_{t+h|t}) f_{t+h|t}(y) dy = 0, \tag{2}$$

whose properties are widely described in Christoffersen and Diedold (1997) and Granger (1999), among the others.

Starting from this general framework different forms can be assumed by $L(\cdot)$ and in some context the use of loss functions different from the Square Errors is even related to the empirical application of the forecasts. For example in finance the choice of $L(x)$ is based on the risk aversion of the operators that, when high, leads to prefer a negative exponential function (Pesaran and Skouras 2002) that gives more weight to negative forecast errors. On the contrary, a positive exponential function can be properly used in hydrology to prevent underprediction of river flows. In other settings Elliott et al. (2005) applies general loss functions to establish forecast rationality so extending, in a wider context, results traditionally based on square losses.

2 General loss functions

In the forecasting literature numerous proposals have been made in the context of loss functions that are called *general* in order to distinguish them from the traditional Square function $L(y_{t+h} - \hat{y}_{t+h|t}) = (y_{t+h} - \hat{y}_{t+h|t})^2$. The motivation is often based on a widely shared need turned to give weight, of different magnitude, to positive and negative forecast errors.

Among them, the most applied in practice are the so called Lin–Lin, Quad–Quad or Linex functions where the first two can be considered as generalizations of the Absolute and Square loss functions respectively.

In particular, the Lin–Lin function, discussed in Christoffersen and Diedold (1997), is:

$$L(y_{t+h} - \hat{y}_{t+h|t}) = \begin{cases} a|y_{t+h} - \hat{y}_{t+h|t}| & \text{if} \quad (y_{t+h} - \hat{y}_{t+h|t}) > 0 \\ b|y_{t+h} - \hat{y}_{t+h|t}| & \text{if} \quad (y_{t+h} - \hat{y}_{t+h|t}) \leq 0, \end{cases} \tag{3}$$

where the parameters $a > 0$ and $b > 0$ regulate the asymmetry of the function.

The Quad–Quad function (Christoffersen and Diedold 1996) is:

$$L(y_{t+h} - \hat{y}_{t+h|t}) = \begin{cases} a(y_{t+h} - \hat{y}_{t+h|t})^2 & \text{if} \quad (y_{t+h} - \hat{y}_{t+h|t}) > 0 \\ b(y_{t+h} - \hat{y}_{t+h|t})^2 & \text{if} \quad (y_{t+h} - \hat{y}_{t+h|t}) \leq 0, \end{cases} \tag{4}$$

with the two positive constants a and b properly chosen to give different weight to positive and negative forecast errors.

The Linex loss (Varian 1975) is slightly different from (3) and (4):

$$L(y_{t+h} - \hat{y}_{t+h|t}) = b\{\exp[a(y_{t+h} - \hat{y}_{t+h|t})] - a(y_{t+h} - \hat{y}_{t+h|t}) - 1\}, \tag{5}$$

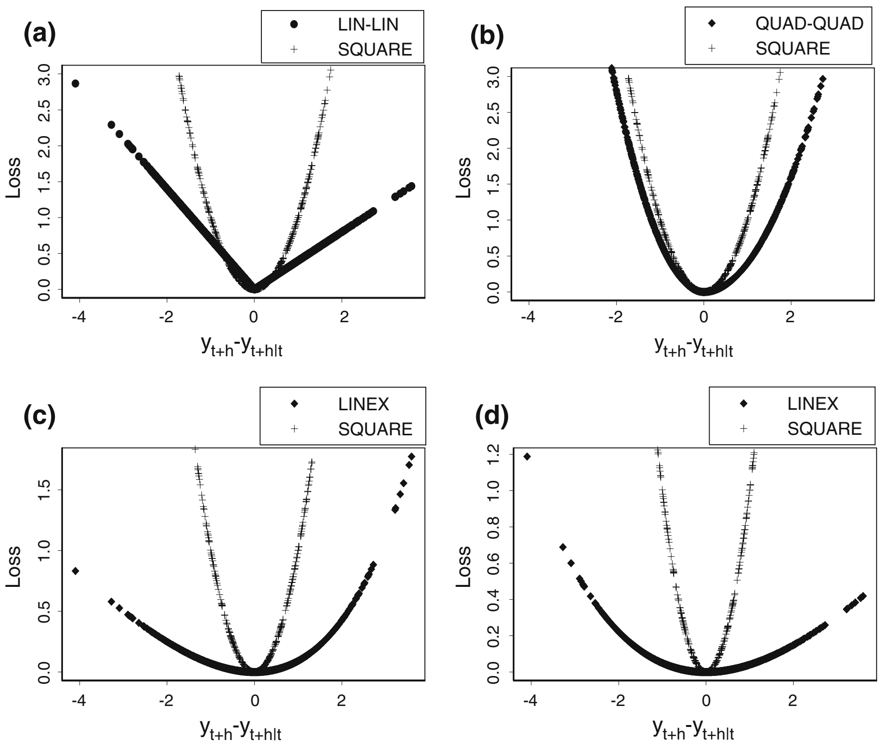

Fig. 1 Lin–Lin, Quad–Quad and Linex loss functions compared with the Square loss. In frame (**a**) the Lin–Lin function (3) is plotted with $a = 0.4$ and $b = 0.7$. The Quad–Quad loss (4) is shown in frame (**b**) with $a = 0.4$ and $b = 0.7$. The last two frames show the Linex function (5) with $b = 1$ and $a = 0.4$ in (**c**) and $a = -0.3$ in (**d**)

with $a \neq 0$ and $b > 0$. This function has been widely discussed in Zellner (1986) and it can be easily observed that b is just a scale parameter whereas the shape (and so the asymmetry) of the function is regulated by the parameter a.

In Fig. 1 is shown the behavior of the three loss functions (3), (4) and (5) to better appreciate their asymmetry and to compare them with the Square loss function.

Using 2000 values generated from a standard Gaussian distribution, the two frames (a) and (b) of Fig. 1 show the Lin–Lin and the Quad–Quad loss functions with $a = 0.4$ and $b = 0.7$ whereas the Linex loss with $b = 1$ is presented in frames (c) and (d) with parameter $a = 0.4$ and $a = -0.3$ respectively. It is interesting to note that small changes of $|a|$ in the Linex case, greatly affect the shape of the function that becomes more flat as $|a|$ decreases.

The problem that can arise when general loss functions are used is related to the generation of closed form predictors obtained from (1). Further, when they exist, they are strongly based on well defined assumptions on the conditional distribution of the process.

For example Christoffersen and Diedold (1997) propose analytic solutions for the predictor $\hat{y}_{t+h|t}$ under Lin–Lin and Linex functions assuming that the conditional distribution of y_{t+h} is Gaussian with conditional mean $\mu_{t+h|t}$ and conditional variance $\sigma^2_{t+h|t}$. In particular they show that the "optimal" Linex predictor is:

$$\hat{y}^{LX}_{t+h|t} = \mu_{t+h|t} + \frac{a}{2}\sigma^2_{t+h|t} \qquad \text{with} \quad b = 1 \quad \text{and} \quad a \neq 0,$$

whereas the "optimal" Lin–Lin predictor is:

$$\hat{y}^{LL}_{t+h|t} = \mu_{t+h|t} + \sigma^2_{t+h|t}\Phi^{-1}\left(\frac{a}{a+b}\right),$$

with $a, b > 0$ and $\mathbf{\Phi}(\cdot)$ the cumulative distribution of the standard Normal random variable.

More recently, Christodoulakis (2005) provides an analytical Linex predictor under non-normality using the class of Gram-Charlier densities, whereas in more general settings, the "optimal" predictor can be instead approximated choosing proper series expansions.

In some cases the Normality assumption can be considered acceptable as shown in the following example.

Example 1 Consider the Threshold Moving Average model (De Gooijer 1998) or order (2;1,1):

$$y_t = \begin{cases} \phi_0^{(1)} + \epsilon_t^{(1)} + \theta_1^{(1)}\epsilon_{t-1}^{(1)} & \text{if } y_{t-1} \geq c \\ \phi_0^{(2)} + \epsilon_t^{(2)} + \theta_1^{(2)}\epsilon_{t-1}^{(2)} & \text{if } y_{t-1} < c, \end{cases} \tag{6}$$

where $|c| < +\infty$ is the threshold parameter, $\phi_0^{(1)}$ and $\phi_0^{(2)}$ are two constants, $\epsilon_t^{(i)} = \sigma_i\epsilon_t$, with $\epsilon_t \sim NID(0,1)$ and $\sigma_i > 0, i = 1,2$. Under the assumptions given on ϵ_t, the distribution of $y_{t+1|t}$ is Normal with conditional mean and conditional variance respectively:

$$\begin{array}{lll} \mu^{(1)}_{t+1|t} = \phi_0^{(1)} + \theta_1^{(1)}\epsilon_t^{(1)}, & \sigma^{2(1)}_{t+1|t} = \sigma^2_{\epsilon^{(1)}_{t+1|t}} + \theta_2^{2(1)}\sigma_1^2 & \text{if } y_t \geq c \\ \mu^{(2)}_{t+1|t} = \phi_0^{(2)} + \theta_1^{(2)}\epsilon_t^{(2)}, & \sigma^{2(2)}_{t+1|t} = \sigma^2_{\epsilon^{(2)}_{t+1|t}} + \theta_2^{2(2)}\sigma_2^2 & \text{if } y_t < c, \end{array}$$

where $\sigma^2_{\epsilon^{(i)}_{t+1|t}} = E_t[(\epsilon^{(i)}_{t+1})^2]$.

These results highlight that the hypothesis under which Christoffersen and Diedold (1997) obtain "optimal" Linex predictors are completely fulfilled from model (6) and so:

$$\hat{y}^{LX}_{t+1|t} = \begin{cases} \mu^{(1)}_{t+1|t} + \frac{a}{2}\sigma^{2(1)}_{t+1|t} & \text{if } y_t \geq c \\ \mu^{(2)}_{t+1|t} + \frac{a}{2}\sigma^{2(2)}_{t+1|t} & \text{if } y_t < c. \end{cases}$$

□

In the following, the Linex loss is used to generate predictors form nonlinear models that belong to the threshold class (Tong 1990). In particular, the attention is focused on the forecasts generation from the Self Exciting Threshold AutoRegressive Moving Average model (SETARMA), introduced in Tong (1983), which can be considered as direct generalization, in nonlinear context, of the ARMA model (Box and Jenkins 1976). The predictor has been derived through the Linex loss function that, with its asymmetric behaviour, takes into account the (asymmetric) effects caught from the SETARMA model, as widely discussed in Amendola et al. (2006).

3 The SETARMA model

Following the notation in Tong (1990) the Self Exciting Threshold Autoregressive Moving Average model of order $(k;p,q)$ can be written as:

$$\mathbf{Y}_t = \boldsymbol{\phi}_0^{(J_t)} + \boldsymbol{\Phi}^{(J_t)}\mathbf{Y}_{t-1} - \boldsymbol{\Theta}^{(J_t)}\mathbf{e}_t, \tag{7}$$

where $J_t : \mathbb{R} \to \mathbb{N}$ identifies the generating process of each observation in $\mathbf{Y}_t = (y_t, y_{t-1}, \ldots, y_{t-p+1})'$ such that $J_t = j$, for $j = 1, \ldots, k$, if $y_{t-d} \in R_j$ with $R_j = (r_{j-1}, r_j] \subseteq \mathbb{R}$ and $-\infty = r_0 \le r_1 \le \cdots \le r_{j-1} \le r_j \le \cdots \le r_k = \infty$, with y_{t-d} the threshold variable, d the threshold delay and r_j the threshold parameter. The other terms in (7):

$$\boldsymbol{\phi}_0^{(j)} = (\phi_0^{(j)}, 0, \ldots, 0)'_{(p\times 1)}, \qquad \mathbf{e}_t = (e_t, e_{t-1}, \ldots, e_{t-q})'_{[(q+1)\times 1]},$$

$$\Phi^{(j)} = \begin{pmatrix} \phi_1^{(j)} & \phi_2^{(j)} & \cdots & \phi_p^{(j)} \\ & \mathbf{I}_{(p-1)} & & \mathbf{0}_{[(p-1)\times 1]} \end{pmatrix}, \quad \Theta^{(j)} = \begin{pmatrix} \theta_0^{(j)} & \theta_1^{(j)} & \cdots & \theta_q^{(j)} \\ & & \mathbf{0}_{[q\times(q+1)]} & \end{pmatrix},$$

with $\theta_0^{(j)} = -1$, $q = p-1$ and $j = 1, \ldots, k$.

Simplifying the notation (7), the SETARMA model can be written in a more simple and, at the same time, more general form:

$$y_t = \phi_0^{(j)} + \boldsymbol{\phi}'^{(j)}\mathbf{Y}_{t-1} - \boldsymbol{\theta}'^{(j)}\boldsymbol{e}_t, \qquad \text{for} \quad y_{t-d} \in R_j, \tag{8}$$

where $\phi_0^{(j)}$ is a constant , $\boldsymbol{\phi}^{(j)} = (\phi_1^{(j)}, \phi_2^{(j)}, \ldots, \phi_{p_j}^{(j)})'$, $\boldsymbol{\theta}^{(j)} = (-1, \theta_1^{(j)}, \ldots, \theta_{q_j}^{(j)})'$ with p_j and q_j two integers that can be different in each regime j and with the vectors $\mathbf{Y}_t = (y_t, \ldots, y_{t-p_j+1})'$ and $\mathbf{e}_t = (e_t, e_{t-1}, \ldots, e_{t-q_j})'$.

In the following pages the notation of model (8) is used under the assumptions that the process y_t is strictly stationary and the errors e_t are Gaussian white noise such that $\{e_t\}$ is a sequence of i.i.d. random variables with $e_t \sim N(0, \sigma^2)$, for $t = 1, 2, \ldots$.

In particular, under these assumptions and following the theoretical background of the previous sections, Linex predictors are generated for model (8).

4 The Linex SETARMA predictor

The generation of forecasts through the Linex loss function has been differently applied: to forecast volatility in the parametric domain of the ARCH models family (Hwang et al. 2001), to generate forecasts from a Gaussian mixture model with constant mean (Patton and Timmermann 2006a) or even to testing forecast rationality (Elliott et al. 2005).

In some context, the prediction of the time series level can be of more interest than to forecast its volatility (e.g. in macroeconomy, hydrology). This is the reason why in the following we focus the attention on the SETARMA predictor generated under function (5), with scale parameter $b = 1$, and assuming the model parameters known, in order to avoid the effect of parameters estimation.

Taking advantage of the local linearity of the SETARMA model, it is further assumed that the conditional distribution:

$$y_{t+h|t} \sim N\left(\mu^{(j)}_{t+h|t}, \sigma^{2(j)}_{t+h|t}\right), \tag{9}$$

with $y_{t+h-d|t} \in R_j$, $0 < h \leq d$ and $j = 1, 2, \ldots, k$.

The optimal h-step ahead predictor of model (8) which minimizes the conditional expectation (1) under Linex loss:

$$\min_{\hat{y}_{t+h|t}} E_t\{\exp[a(y_{t+h} - \hat{y}_{t+h|t})] - a(y_{t+h} - \hat{y}_{t+h|t}) - 1\}, \tag{10}$$

is presented in the following proposition.

Proposition 1 *Given the strictly stationary SETARMA model* (8) *with conditional distribution* $y_{t+h|t} \sim N\left(\mu^{(j)}_{t+h|t}, \sigma^{2(j)}_{t+h|t}\right)$, *the optimal Linex predictor is:*

$$\begin{aligned}\hat{y}_{t+h|t} &= \mu^{(j)}_{t+h|t} + \frac{a}{2}\sigma^{2(j)}_{t+h|t} = \\ &= \phi^{(j)}_0 + \sum_{i=1}^{h-1}\phi^{(j)}_i E_t(y_{t+h-i}) + \sum_{i=h}^{p_j}\phi^{(j)}_i y_{t+h-i} + \sum_{i=h}^{q_j}\theta^{(j)}_i e_{t+h-i} + \frac{a}{2}\sigma^{2(j)}_{t+h|t},\end{aligned} \tag{11}$$

with $y_{t+h-d|t} \in R_j$, *for* $j = 1, 2, \ldots, k$, *and* $0 < h \leq d$.

Proof See Appendix. □

The result (11) shows that the Linex predictor of model (7) is the sum of the conditional mean of y_{t+h} with a term related to the conditional variance of the process. This completely agrees with Proposition 1 of Christoffersen and Diedold (1997):

Proposition 2 (Christoffersen and Diedold 1997) *If $y_{t+h|t} \sim N(\mu_{t+h|t}, \sigma^2_{t+h|t})$ is a conditionally Gaussian process and $L(\hat{e}_{t+h|t})$ is any loss function defined on the h-step ahead prediction error $\hat{e}_{t+h|t}$, then the optimal predictor is of the form $\hat{y}_{t+h|t} = \mu_{t+h|t} + \alpha_{t+h|t}$, where $\alpha_{t+h|t}$ depends only on the loss function and the conditional prediction error variance $\sigma^2_{t+h|t} = var(y_{t+h}|\Omega_t) = var(e_{t+h}|\Omega_t)$.*

Coming back to the Linex predictor (11), the conditional mean $\mu^{(j)}_{t+h|t}$ is explicitly presented whereas the conditional variance, $\sigma^{2(j)}_{t+h|t}$, needs to be properly estimated.

Corollary 1 *Given the strictly stationary SETARMA model (8) with conditional distribution $y_{t+h|t} \sim N\left(\mu^{(j)}_{t+h|t}, \sigma^{2(j)}_{t+h|t}\right)$, the conditional variance $\sigma^{2(j)}_{t+h|t}$ is:*

$$\begin{aligned}
\sigma^{2(j)}_{t+h|t} &= \sum_{i=1}^{h-1}\left(\phi_i^{(j_h)}\right)^2 \sigma^{2(j)}_{t+i|t} + \sum_{i=0}^{h-1}\left(\theta_i^{(j_h)}\right)^2 \sigma^2 \\
&\quad + 2\sum_{i=1}^{h-2}\sum_{w=i+1}^{h-1} E_t[\tilde{e}_{t+h-i|t}\tilde{e}_{t+h-w|t}]\phi_i^{(j_h)}\phi_w^{(j_h)} \\
&\quad - 2\sum_{i=1}^{h-1}\sum_{w=i}^{h-1} E_t[\tilde{e}_{t+h-i|t}e_{t+h-w}]\phi_i^{(j_h)}\theta_w^{(j_h)}, \qquad (12)
\end{aligned}$$

with $0 < h \le d$, j_h is the regime from which the h-step ahead forecast is generated, $\theta_0^{(j_h)} = 1$ and $\tilde{e}_{t+h-i} = y_{t+h-i} - \mu^{(j)}_{t+h-i|t}$.

Proof See Appendix. □

Example 2 In order to illustrate how $\sigma^{2(j)}_{t+h|t}$ is computed, consider the three following cases.
Case 1: $h = 1$.
This is the simplest case where the conditional variance is equal to the variance of the error e_t:

$$\begin{aligned}
\sigma^{2(j)}_{t+1|t} &= E_t\{[y_{t+1} - E_t(y_{t+1})]^2\} = \\
&= E_t[(y_{t+1} - \mu^{(j)}_{t+1|t})^2] = \\
&= E_t[\tilde{e}^2_{t+1}] = E_t[e^2_{t+1}] = \sigma^2, \qquad (13)
\end{aligned}$$

where $y_{t+1-h} \in R_j, j = 1, 2, \ldots, k$, $\tilde{e}_{t+1} = y_{t+1} - \mu^{(j)}_{t+1|t}$ that, given the definition of $\mu^{(j)}_{t+h|t}$ in (11), corresponds to e_{t+1}.

Case 2: $h = 2$.
In this case we have:

$$\sigma^{2(j)}_{t+2|t} = \left[\left(\phi^{(j_2)}_1 - \theta^{(j_2)}_1\right)^2 + 1\right]\sigma^2 \qquad \text{with } y_{t+2-h} \in R_j, \tag{14}$$

for $j = 1, 2, \ldots, k$.
Case 3: $h = 3$.
The form of the conditional variance becomes more complex as h grows. When $h = 3$:

$$\begin{aligned}\sigma^{2(j)}_{t+3|t} &= E_t[(y_{t+3} - y_{t+3|t})^2] \\ &= E_t[(\phi^{(j_3)}_1 \tilde{e}_{t+2|t} + \phi^{(j_3)}_2 \tilde{e}_{t+1|t} + a_{t+3} - \theta^{(j_3)}_1 a_{t+2} - \theta^{(j_3)}_2)^2] \\ &\qquad \text{(after eliminating all terms with null conditional expected value)} \\ &= \sum_{i=1}^{2}\left(\phi^{(j_3)}_i\right)^2 \sigma^{2(j)}_{t+i|t} + \sum_{i=0}^{2}\left(\theta^{(j_3)}_i\right)^2 \sigma^2 + 2\phi^{(j_3)}_1 \phi^{(j_3)}_2 E_t[\tilde{e}_{t+2|t}\tilde{e}_{t+1|t}] \\ &\quad - 2\phi^{(j_3)}_1 \theta^{(j_3)}_1 E_t[\tilde{e}_{t+2|t} e_{t+2}] - 2\phi^{(j_3)}_1 \theta^{(j_3)}_2 E_t[\tilde{e}_{t+2|t} e_{t+1}] \\ &\quad - 2\phi^{(j_3)}_2 \theta^{(j_3)}_2 E_t[\tilde{e}_{t+1|t} e_{t+1}],\end{aligned} \tag{15}$$

where:

- $E_t[\tilde{e}_{t+2|t}\tilde{e}_{t+1|t}] = \left[\phi^{(j_2)}_1 - \theta^{(j_2)}_1\right]\sigma^2$;
- $E_t[\tilde{e}_{t+i|t} e_{t+i}] = \sigma^2$, for $i = 1, 2$;
- $E_t[\tilde{e}_{t+2|t} e_{t+1}] = E_t[\tilde{e}_{t+2|t}\tilde{e}_{t+1|t}]$ because $\tilde{e}_{t+1|t} = e_{t+1}$ in model (8).

□

The forecast error of the Linex predictor $\hat{y}_{t+h|t}$, denoted $\hat{e}_{t+h|t}$, is given as:

$$\begin{aligned}\hat{e}_{t+h|t} &= y_{t+h} - \hat{y}_{t+h|t} \\ &= \sum_{i=1}^{h-1} \phi^{(j)}_i y_{t+h-i} - \sum_{i=1}^{h-1} \phi^{(j)}_i E_t(y_{t+h-i}) - \sum_{i=0}^{h-1} \theta^{(j)}_i e_{t+h-i} - \frac{a}{2}\sigma^{2(j)}_{t+h|t},\end{aligned} \tag{16}$$

with expected value:

$$E[\hat{e}_{t+h|t}] = -\frac{a}{2}\sigma^{2(j)}_{t+h|t}, \tag{17}$$

that highlights the biasedness of the Linex predictor (11), whose amount depends on the conditional variance.

This is not new in the context of asymmetric loss functions where, as shown in Patton and Timmermann (2006a), some properties established in the MSE framework cannot be recognized.

Proper transformations of the forecast error can be used to obtain null expected value for (16). Granger (1999) calls these transformed errors as *generalized forecast errors*, $\hat{e}^g_{t+h|t}$, that fulfill the classical properties of the prediction errors and let properly compare forecasts. It is obtained from the minimization of the loss function $L\left(y_{t+h} - \hat{y}_{t+h|t}; a\right)$, where $\hat{y}_{t+h|t}$ is the optimal predictor generated from (10) and with the generalized forecast error, $\hat{e}^g_{t+h|t}$, obtained from:

$$\hat{e}^g_{t+h|t} = \frac{\partial L\left(y_{t+h} - \hat{y}_{t+h|t}; a\right)}{\partial \hat{y}_{t+h|t}}, \quad \text{such that} \quad E_t[\hat{e}^g_{t+h|t}] = 0. \tag{18}$$

It can be shown that the conditions (18) are met, with the Linex SETARMA predictor, when:

$$\hat{e}^g_{t+h|t} = a - a \cdot \exp(a\hat{e}_{t+h|t}), \tag{19}$$

whose conditional and unconditional expectations are null. In fact:

$$\begin{aligned}
E_t[\hat{e}^g_{t+h|t}] &= E_t[a - a\exp(a\hat{e}_{t+h|t})] \\
&= E_t[a - a\exp(ay_{t+h} - a\hat{y}_{t+h|t})] \\
&= a - aE_t[\exp(ay_{t+h})]\exp(-a\hat{y}_{t+h|t}) \\
&= a - a\exp\left(a\mu^{(j)}_{t+h|t} + \frac{a^2}{2}\sigma^{2(j)}_{t+h|t}\right)\exp(-a\hat{y}_{t+h|t}) \\
&\quad \text{(from the definition of } \hat{y}_{t+h|t}) \\
&= a - a = 0,
\end{aligned}$$

whereas $E[\hat{e}^g_{t+h|t}] = 0$ from the law of iterated expectation.

The variance of $\hat{e}^g_{t+h|t}$ is given in the following proposition.

Proposition 3 *The unconditional variance of the generalized prediction error* (19), *obtained from the Linex predictor* (11) *of the SETARMA model* (8), *is:*

$$Var(\hat{e}^g_{t+h|t}) = a^2\left[\exp(a^2\sigma^{2(j)}_{t+h|t}) - 1\right] \qquad \text{with} \quad a \neq 0. \tag{20}$$

Proof See Appendix. □

The positivity of the variance (20) is is always guaranteed (because $\sigma^{2(j)}_{t+h|t} > 0$ and $a \neq 0$) whereas its behaviour is related to that one of $\sigma^{2(j)}_{t+h|t}$. In fact, $Var(\hat{e}^g_{t+h|t})$ is monotonically increasing with h when this property is even fulfilled by $\sigma^{2(j)}_{t+h|t}$.

5 Forecast evaluation

The point forecasts just presented need to be evaluated and compared with forecasts generated from different models or alternative loss functions.

In a similar context where the classical predictors features cannot be always recognized, some indexes of forecast accuracy such as mean square error, mean absolute error, cannot be helpful. On the contrary the study of the predictors distribution can give potentially more relevant results when, for example, interval or density forecasts are of interest.

This implies the estimation of the distribution of $\hat{y}_{t+h|t}$ that for nonlinear time series is not easy to perform analytically and so computer intensive methods can be applied.

In the following, empirical distributions of the SETARMA Linex and Least Squares predictors are obtained through a Monte Carlo exercise in order to present some features of the predictors introduced in the previous section and to compare them to the traditional Least Squares predictors widely discussed in Amendola et al. (2005). Given the SETARMA(2;1,1) model:

$$y_t = \begin{cases} 0.5y_{t-1} + e_t - 0.4e_{t-1} & y_{t-2} \geq 0 \\ -0.41y_{t-1} + e_t - 0.12e_{t-1} & y_{t-2} < 0, \end{cases} \tag{21}$$

the empirical distributions of $\hat{y}^{LX}_{t+h|t}$ and $\hat{y}^{LS}_{t+h|t}$ (where LX refers to Linex predictors and LS to Least Squares predictors) are obtained from 1000 Monte Carlo replications with series length 500, $e_t \sim N(0, 1.5)$, $h = 1, 2$ and further considering, for the Linex case, four different parameters of asymmetry $a = 0.3, 0.7, -0.3, -0.7$.

Starting from the results of Amendola et al. (2005) that show the unbiasedness of the Least Square predictor, the empirical distributions of $\hat{y}^{LX}_{t+h|t}$ and $\hat{y}^{LS}_{t+h|t}$ are presented in Fig. 2 where the bias of the Linex predictor is emphasized. It can be appreciated in frames (a), (c), (e), (g) when $h = 1$ and more deeply in the remaining frames when $h = 2$.

The empirical distributions of the Linex and Least Squares prediction errors are instead given in Fig. 3. They have symmetric shapes when $h = 1$ and, as expected in nonlinear context, they become increasingly asymmetric when $h = 2$ and as $|a|$ grows.

To better clarify the advantages of the Linex predictor, and making use of the results of the 1000 Monte Carlo replications, in Fig. 4 the mean values of the generated y_{t+h}, $\hat{y}^{LS}_{t+h|t}$ and $\hat{y}^{LX}_{t+h|t}$ (with $a = 0.3, 0.7, -0.3, -0.7$) are plotted, for $h = 1, 2$.

The straight line represents $\bar{y}_{t+h}$ whereas the other dotted and dashed lines are $\bar{\hat{y}}^{LS}_{t+h|t}$ and $\bar{\hat{y}}^{LX}_{t+h|t}$. The Linex forecasts can be intended as lower and upper bounds for the Least Squares forecasts that can be used to evaluate the results of different weights, given to positive and negative forecast errors, on the generated predictions.

In other terms, when $\hat{y}^{LX}_{t+h|t}$ is greater than $\hat{y}^{LS}_{t+h|t}$ (and consequently $a > 0$) the forecaster believes to have advantages from overpredictions whereas in the opposite case (with $a < 0$) y_{t+h} is underestimated and so it can be of interest to forecasters with more cautious attitudes.

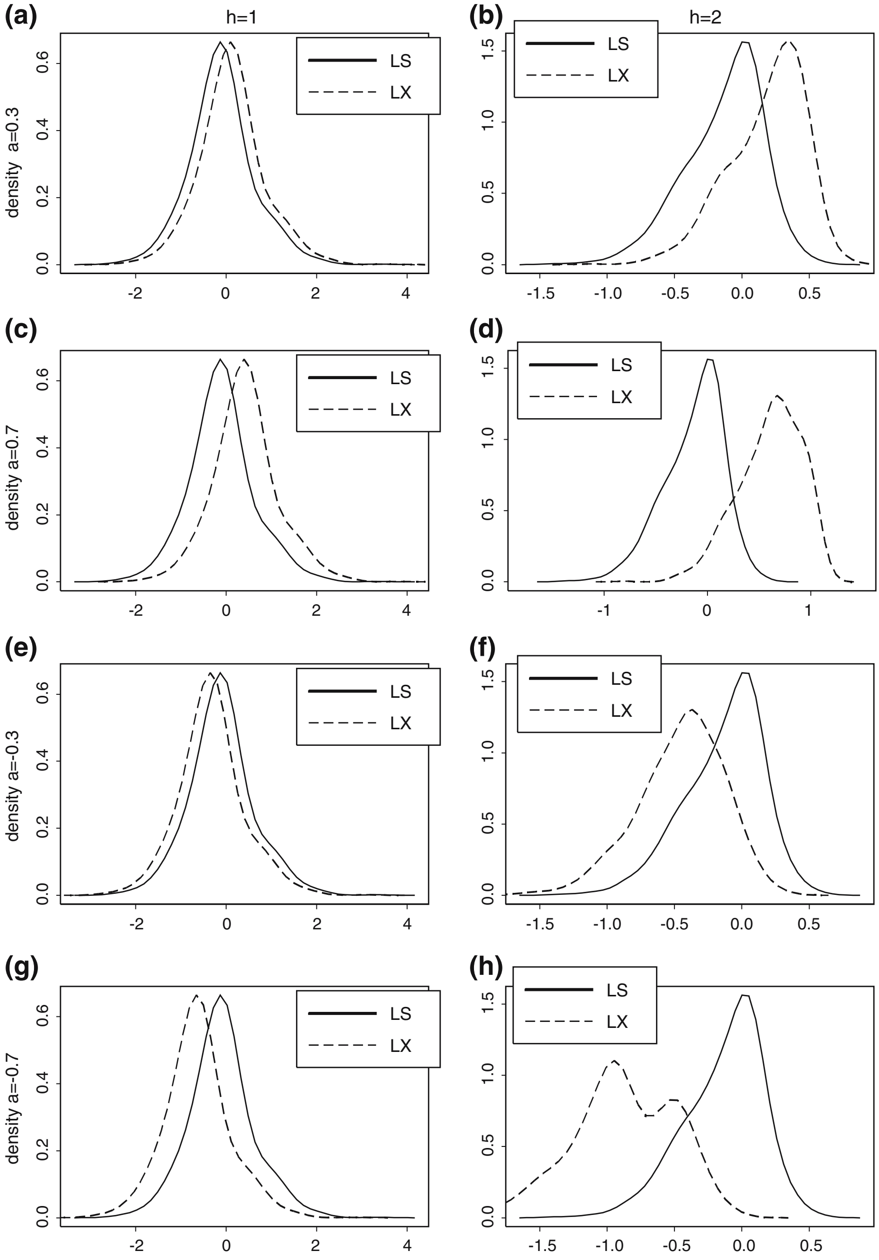

Fig. 2 Empirical distribution of Linex (LX) and Least Squares (LS) predictors from model (21) with $h = 1$ and $h = 2$. The parameter a of function (5) is $a = 0.3$ in frames (**a**)–(**b**); $a = 0.7$ in frames (**c**)–(**d**); $a = -0.3$ in frames (**e**)–(**f**); $a = -0.7$ in frames (**g**)–(**h**)

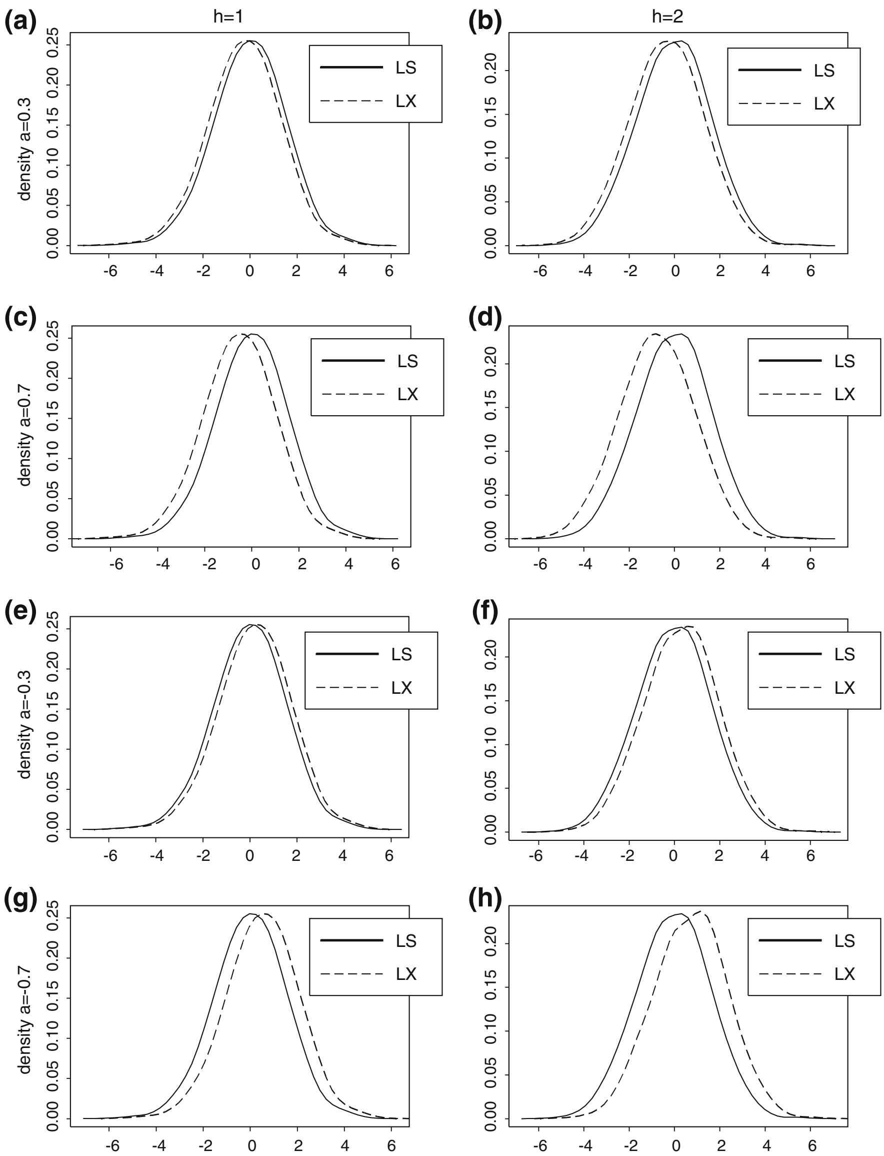

Fig. 3 Empirical distribution of Linex (LX) and Least Squares (LS) prediction errors from model (21) with $h = 1$ and $h = 2$. The parameter a of function (5) is $a = 0.3$ in frames (**a**)–(**b**); $a = 0.7$ in frames (**c**)–(**d**); $a = -0.3$ in frames (**e**)–(**f**); $a = -0.7$ in frames (**g**)–(**h**)

The bias combined with the skewness of the two predictors (and even multimodality in frame (h) of Fig. 2), makes no longer pertinent the use of classical tools, widely considered with linear time series models, to construct interval forecasts or to estimate predictive density.

In order to cope with these features, predictive sets with minimum length (in terms of Lebesgue measure) proposed in Polonik and Yao (2000) can be

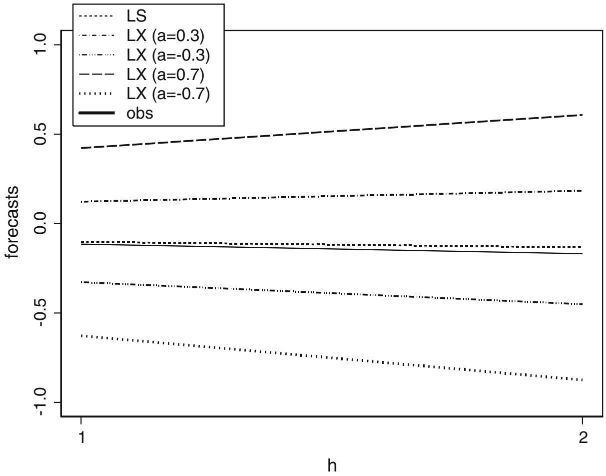

Fig. 4 Mean values of 1000 Monte Carlo replications generated from model (21) for y_{t+h}, $\hat{y}^{LS}_{t+h|t}$ and $\hat{y}^{LX}_{t+h|t}(a)$, with $a = 0.3, 0.7, -0.3, -0.7$ and $h = 1, 2$

used, among the others, or high density regions (Hyndman 1995) can be built in presence of multimodality.

The study of forecast evaluation in presence of asymmetric loss functions is only at its first steps. In fact, the results developed in literature to asses forecast accuracy, such as tests or statistical indexes, are often based on the use of quadratic losses and so have to be properly revised to help even in this more general context. Some early results can be found in Patton and Timmermann (2006b) whose extension can be a very interesting object of future research.

Acknowledgements The author would like to thank the editor and two anonymous referees for their valuable comments which lead to the improvement of the paper. All deficiencies are the responsibility of the author.

Appendix

Proof of Proposition 1

The predictor (11) can be derived using the first order condition of (10) and further making use of the moment generating function of $y_{t+h|t} \sim N(\mu^{(j)}_{t+h|t}, \sigma^{2(j)}_{t+h|t})$, with $y_{t+h-d} \in R_j,\ j = 1, \dots, k$.

In particular:

$$E_t[L(y_{t+h} - \hat{y}_{t+h|t})] = E_t\{\exp[a(y_{t+h} - \hat{y}_{t+h|t})] - a(y_{t+h} - \hat{y}_{t+h|t}) - 1\}$$
$$= \exp(-a\hat{y}_{t+h|t}) \exp\left(a\mu^{(j)}_{t+h|t} + \frac{a^2}{2}\sigma^{2(j)}_{t+h|t}\right) - a\mu^{(j)}_{t+h|t} + a\hat{y}_{t+h|t} - 1. \quad (22)$$

The first order condition with respect to $\hat{y}_{t+h|t}$ of (22) is:

$$\exp(-a\hat{y}_{t+h|t})\exp\left(a\mu^{(j)}_{t+h|t}+\frac{a^2}{2}\sigma^{2(j)}_{t+h|t}\right)=1, \tag{23}$$

with $\mu^{(j)}_{t+h|t}=\phi^{(j)}_0+\sum_{i=1}^{h-1}\phi^{(j)}_i E_t(y_{t+h-i})+\sum_{i=h}^{p_j}\phi^{(j)}_i y_{t+h-i}+\sum_{i=h}^{q_j}\theta^{(j)}_i e_{t+h-i}$.

The logarithm of (23) allows to obtain the Linex SETARMA predictor:

$$\hat{y}_{t+h|t}=\phi^{(j)}_0+\sum_{i=1}^{h-1}\phi^{(j)}_i E_t(y_{t+h-i})+\sum_{i=h}^{p_j}\phi^{(j)}_i y_{t+h-i}+\sum_{i=h}^{q_j}\theta^{(j)}_i e_{t+h-i}+\frac{a}{2}\sigma^{2(j)}_{t+h|t}.$$

Proof of Corollary 1

Starting from the definition of conditional variance:

$$\sigma^{2(j)}_{t+h}=E_t[y_{t+h}-E_t(y_{t+h})]^2=E_t[y_{t+h}-\mu^{(j)}_{t+h})]^2 \tag{24}$$

the result (12) is obtained after a quite long, but easy, algebra using in (24) the model definition (8) and the conditional mean $\mu^{(j)}_{t+h}$ explicitly given in (11).

Proof of Proposition 2

The unconditional variance of the generalized prediction error is strictly related to the assumptions given on the conditional distribution of y_{t+h}.

In particular, given $\hat{e}^g_{t+h|t}=a-a\cdot\exp(a\hat{e}_{t+h|t})$, with $E[\hat{e}^g_{t+h|t}]=0$, it follows that:

$$\begin{aligned}Var(\hat{e}^g_{t+h|t})&=E[(a-a\exp(a\hat{e}_{t+h|t}))^2]\\&=a^2\cdot E[1+\exp(2a\hat{e}_{t+h|t})-2\exp(a\hat{e}_{t+h|t})].\end{aligned} \tag{25}$$

From the error $\hat{e}_{t+h|t}=y_{t+h}-\hat{y}_{t+h|t}$ and using the law of iterated expectations:

$$\begin{aligned}E[\exp(2a\hat{e}_{t+h|t})]&=E[E_t(\exp(2a\hat{e}_{t+h|t}))]\\&=E\{E_t[\exp(2a(y_{t+h}-\mu^{(j)}_{t+h|t}-\frac{a}{2}\sigma^{2(j)}_{t+h|t}))]\}\\&=E\left\{\frac{E_t[\exp(2ay_{t+h})]}{\exp[2a(\mu^{(j)}_{t+h|t}+\frac{a}{2}\sigma^{2(j)}_{t+h|t})]}\right\}\\&=\exp[2a\mu^{(j)}_{t+h|t}-a^2\sigma^{2(j)}_{t+h|t}]^{-1}\exp[2a\mu^{(j)}_{t+h|t}+2a^2\sigma^{2(j)}_{t+h|t}]\\&=\exp(a^2\sigma^{2(j)}_{t+h|t}),\end{aligned} \tag{26}$$

whereas

$$E[\exp(a\hat{e}_{t+h|t})] = E\left\{E_t[\exp(a(y_{t+h} - \mu^{(j)}_{t+h|t} + \frac{a}{2}\sigma^{2(j)}_{t+h|t}))]\right\}$$
$$= E\left\{\frac{\exp[a\mu^{(j)}_{t+h|t} + \frac{a^2}{2}\sigma^{2(j)}_{t+h|t}]}{\exp[a\mu^{(j)}_{t+h|t} + \frac{a^2}{2}\sigma^{2(j)}_{t+h|t}]}\right\} = 1. \qquad (27)$$

From the combination of the results (26) and (27) in (25), it follows that:

$$Var(\hat{e}^g_{t+h|t}) = a^2[\exp(a^2\sigma^{2(j)}_{t+h|t}) - 1]. \qquad (28)$$

References

Amendola A, Niglio M, Vitale C (2005) Multi-step ahead predictors of SETARMA models. Manuscript (available from the authors)

Amendola A, Niglio M and Vitale C (2006) The moments of SETARMA model. Statist Probab Lett 76:625–633

Box GEP, Jenkins GM (1976) Time series analysis, forecasting and control. Holden-Day, San Francisco

Christodoulakis GA (2005) Financial forecasts in presence of asymmetric loss aversion, skewness and excess kurtosis. Financ Res Lett 2:227–233

Christoffersen PF, Diedold FX (1996) Further results on forecasting and model selection under asymmetric loss. J App Econom 11:561–571

Christoffersen PF, Diedold FX (1997) Optimal prediction under asymmetric loss. Econom Theory 13:808–817

De Gooijer J (1998) On threshold moving-average models. J Time Series Anal 19:1–18

Elliott G, Komunjer I, Timmermann A (2005) Estimation and testing of forecast rationality under flexible loss. Rev Econo Stud 72:1107–1125

Fan J, Yao Q (2003) Nonlinear time series. nonparametric and parametric methods. Springer New York

Granger CWJ (1969) Prediction with a generalized cost of error function. Oper Res Q 20:199–207

Granger CWJ (1999) Outline of forecast theory using generalized cost functions. Spanish Econ Rev 1:161–173

Hyndman RJ (1995) Highest-density forecast regions for non-linear and non-normal time series models. J Forecasting 14:431–441

Hwang S, Knight J, Satchell SE (2001) Forecasting nonlinear functions of returns using LINEX loss functions. Ann Econ Financ 2:187–213

Patton AJ, Timmermann A (2006a) Properties of optimal forecast under asymmetric loss and nonlinearity. J Econ doi: 10.1016/j.jeconom.2006.07.018

Patton AJ, Timmermann A (2006b) Testing forecast optimality under unknown loss. J Am Statist Assoc (forthcoming)

Pesaran MH, Skouras S (2002) Decision-based methods for forecast evaluation. In: Clements MP, Hendry DF (eds) A companion to economic forecasting. Blackwell, pp 241–267

Polonik W, Yao Q (2000) Conditional minimum volume predictive regions for stochastic processes. J Am Statis Assoc 95:509–519

Tong H (1990) Nonlinear time series: a dynamical system approach. Oxford University Press, Oxford

Varian H (1974) A bayesian approach to real estate assessment. In: Feinberg SE, Zellner A (eds) Studies in Bayesian econometrics and statistics in honor of L.J. Savage. North-Holland, Amsterdam pp 195–208

Zellner A (1986) Bayesian estimation and prediction using asymmetric loss functions. J Am Statist Assoc 81:446-451