Multi-step ahead predictors of SETARMA models

Alessandra Amendola
Di.S.E.S., Università di Salerno
alamendola@unisa.it

Marcella Niglio
Di.S.E.S., Università di Salerno
mniglio@unisa.it

Cosimo Vitale
Di.S.E.S., Università di Salerno
vitale@unina.it

Abstract

The generation and the properties of the multi-step point predictors of the Self Exciting Threshold AutoRegressive Moving Average model has been examined. In the first part we focus the attention on the main properties of this class of models and on their different representations that can be derived extending, to the non linear case, some well know results given in Box and Jenkins (1976). Starting from these results, the forecast generation of the SETARMA models is discussed showing its relation with the threshold delay, $d$. In particular, when the lead time $h$ is less or equal than $d$, the least square predictor can be easily generated whereas when $h > d$ different predictors are proposed and their combination is evaluated.

Keywords: Threshold models, multi-step forecasts, forecast combination, prediction accuracy.

JEL codes: C22, C53

1 Introduction

The use of statistical models to describe the dynamic of real phenomena has given evidence that the linear ARMA family (Box and Jenkins, 1976) is not able to catch the structure of most of them.

In order to overcome this limit a relevant number of nonlinear models has been proposed in literature some of which can be intended as direct generalization of linear structures.
In this context Priestley (1988), Tong (1990), Granger and Teräsvirta (1993), Tjøstheim (1994), Franses and Van Dijk (2000) and more recently Fan and Yao (2003), have organically presented different nonlinear structures highlighting the main statistical features of each of them and their possible applications.

In this paper the attention is focused on the class of Threshold AutoRegressive (TAR) models proposed in Tong between the end of 70’s and the start of 80’s, and later on generalized in a relevant number of contributions.

Different features of the TAR models have been studied: the statistical properties of the generating process, some problems related to the coefficients estimation and the forecasts generation.

The last feature has been mainly faced in the class of the so called SETAR (Self Exciting Threshold AutoRegressive) models. For example De Gooijer and De Bruin (1998) approximate, under well defined assumptions on the process, the closed form expression for the SETAR predictor. Clements and Smith (1997, 1999) have instead evaluated various predictors for this class of models whereas Clements et al. (2003) investigate the accuracy of the one-step ahead point forecasts comparing their forecast ability with respect to the linear case. Recently Amendola and Niglio (2003) have further studied the distribution of the multi-step SETAR predictors, through an extensive set of Monte Carlo simulations, showing their asymmetry and multimodality.

In the following pages, the forecast generation has been investigated for another particular class of threshold nonlinear structure, the Self Exciting Threshold AutoRegressive Moving Average (SETARMA) model (Tong, 1983) where the generation of predictions implies some complications here discussed.

In particular, the aim of the present paper is to derive the exact point multi-step forecasts of the SETARMA model, showing their main features and some problems related to their generation.

Before to go into these questions, in Section 2 the SETARMA model is presented giving some remarks on its different representations; in Section 3, ignoring estimation uncertainty, the best predictor, in term of minimum mean square error, for this class of models is presented highlighting different aspects which heavily affect its generation further distinguishing among the least squares multi-step predictors and the so called plug-in predictors. Using these results, the predictors have been properly evaluated and combined in section 4 in order to limit the volatility of the generated forecasts. In section 5, an example with simulated data highlights the main features of the proposed predictors whereas some concluding remarks are given in the last section.

### 2 The SETARMA model

The Self-Exciting Threshold AutoRegressive Moving Average process (Tong, 1983) of order \( (k; p_1, \ldots, p_k; q_1, \ldots, q_k) \) and delay \( d \in \mathbb{N} \) is defined as:
\[ X_t = \sum_{i=1}^{k} \left[ \phi^{(i)}_0 + \sum_{j=1}^{p_i} \phi^{(i)}_j X_{t-j} + \epsilon^{(i)}_t + \sum_{w=1}^{q_i} \theta^{(i)}_w \epsilon_{t-w} \right] I(X_{t-d} \in R_i) \]  

(1)

where \( \epsilon^{(i)}_t = \sigma^2_i \epsilon_t, \epsilon_t \) are i.i.d. random variables with \( E[\epsilon_t] = 0 \) and \( E[\epsilon_t^2] = 1 \), for \( i = 1, \ldots, k \), \( I(\cdot) \) is an indicator function, \( R_i = [r_{i-1}, r_i) \) forms a partition of the real line such that \(-\infty = r_0 < r_1 < r_2 < \ldots < r_k = +\infty, r_i \) are the threshold values, \( d \) is the threshold delay, \( p_i \) and \( q_i \) are positive integers and \( \phi^{(i)}_j \) and \( \theta^{(i)}_w \) are unknown parameters with \( j = 1, 2, \ldots, p_i \) and \( w = 1, 2, \ldots, q_i \).

Model (1) can be considered as direct generalization, in the nonlinear domain, of the ARMA model (Box and Jenkins, 1976). In fact, it assumes different linear ARMA representations in no-overlapping regions of the real space, with \( \bigcup_{i=1}^{k} R_i = \mathbb{R} \), each limited by the threshold variable \( X_{t-d} \in R_i \). Further, using the backward shift operator \( B \), such that \( B^s X_t = X_{t-s} \), it can be equivalently written as:

\[ X_t = \sum_{i=1}^{k} \left[ \phi^{(i)}_0 + \phi^{(i)}_{p_i}(B)X_t + \theta^{(i)}_{q_i}(B)\epsilon^{(i)}_t \right] \cdot I(X_{t-d} \in R_i) \]  

(2)

with the polynomials \( \phi^{(i)}_{p_i}(B) = \sum_{j=1}^{p_i} \phi^{(i)}_j B^j \) and \( \theta^{(i)}_{q_i}(B) = 1 - \sum_{w=1}^{q_i} \theta^{(i)}_w B^w \), for \( i = 1, 2, \ldots, k \).

In order to show in the more simple way the results reached for model (2), in the following pages a SETARMA \((2; p_1, p_2; q_1, q_2)\) model with threshold delay \( d \) and threshold value \( c \), such that \( R_1 = [c, +\infty) \) and \( R_2 = (-\infty, c) \), is considered:

\[ X_t = \left[ \phi^{(1)}_0 + \phi^{(1)}_{p_1}(B)X_t + \theta^{(1)}_{q_1}(B)\epsilon^{(1)}_t \right] I(X_{t-d} \in R_1) \]  

\[ + \left[ \phi^{(2)}_0 + \phi^{(2)}_{p_2}(B)X_t + \theta^{(2)}_{q_2}(B)\epsilon^{(2)}_t \right] \cdot [1 - I(X_{t-d} \in R_1)] \]  

(3)

Extensions of the results presented in the following pages to the general \( k \)-regimes case can be easily considered even if, for our purpose, it implies only heavy notation with no explicative advantage.

The switching between the two regimes in model (3) is regulated by the indicator process \( I(X_{t-d} \in R_1) = I_{t-d} \) that, given its dichotomy, can be written as:

\[ I_{t-d} = \begin{cases} 1 & \text{if} \quad X_{t-d} \geq c \\ 0 & \text{if} \quad X_{t-d} < c \end{cases} \]  

(4)

for \( t = 1, 2, \ldots, \) and \( d > 0 \).

The four main properties of process (4) are:

a) \( E[I_{t-d}] = P(I_{t-d} = 1) = P(X_{t-d} \geq c) \)

b) \( I_{t-d}^2 = I_t \) and \( (1 - I_{t-d})^2 = 1 - I_{t-d} \)

c) \( I_{t-d}(1 - I_{t-d}) = 0 \)
\[ I_{t-d} \cdot I_{t-d-j} = \begin{cases} 1 & \text{if } (X_{t-d} \geq c \text{ and } X_{t-d-j} \geq c) \\ 0 & \text{otherwise} \end{cases} \]

They are strictly related to the dichotomy of \( I_{t-d} \) and, in the next section, they greatly simplify the algebraic formula related to the forecasts generation and the predictors properties. In the following, the strict stationarity and the ergodicity of the joint process \((X_t^{(1)}, X_t^{(2)}, I_{t-d})\) is assumed, where \(X_t^{(i)}\) is the process which governs the \(i\)-th regime \((i = 1, 2)\). Under this assumption, the SETARMA(2; \(p_1, p_2; q_1, q_2\)) model (3) can further be written as:

\[
X_t = \left[ c_0^{(1)} + \frac{\theta_0^{(1)}(B)}{1 - \phi_0^{(1)}(B)} \right] I_{t-d} + \left[ c_0^{(2)} + \frac{\theta_0^{(2)}(B)}{1 - \phi_0^{(2)}(B)} \right] (1 - I_{t-d}) \tag{5}
\]

where the polynomial \(\phi_0^{(i)}(B) = 0\) is such that \(|B| > 1\), and the two constants \(c_0^{(i)}, i = 1, 2\) are:

\[
c_0^{(i)} = \frac{\phi_0^{(i)}}{1 - \phi_0^{(i)}(B)} = \frac{\phi_0^{(i)}}{1 - \sum_{j=1}^{p_i} \phi_j^{(i)}} \text{ when } X_{t-d} \in R_t
\]

The dichotomy of the process \(I_{t-d}\) in (4), which makes the two regimes incompatible over the range of \(X_t\), and the use of the identity principle of algebraic polynomials, allow to write model (5) even as:

\[
X_t = \left[ c_0^{(1)} + \sum_{j=0}^{\infty} \psi_j^{(1)}(B) e_t^{(1)} \right] I_{t-d} + \left[ c_0^{(2)} + \sum_{j=0}^{\infty} \psi_j^{(2)}(B) e_t^{(2)} \right] (1 - I_{t-d}) \tag{6}
\]

with:

- \( \frac{\theta_0^{(i)}(B)}{\phi_0^{(i)}(B)} = \psi_0^{(i)}(B) = \sum_{j=0}^{\infty} \psi_j^{(i)}(B) \)

- \( \sum_{j=0}^{\infty} |\psi_j^{(i)}| < \infty, \text{ for } i = 1, 2. \)

where:

\[
\psi_j^{(i)} = \sum_{s=0}^{j-1} \phi_s^{(i)} \psi_{j-s}^{(i)} - \theta_j^{(i)} \quad \text{for } j \geq 1 \text{ and } i = 1, 2 \tag{7}
\]

with \(\psi_0^{(i)} = 1, \theta_j^{(i)} = 0\) for \(j > q_i\).

Some statistical properties of the SETARMA model have been investigated, under well defined conditions, in Brockwell et al. (1992) and in Liu and Susko (1992) whereas Amendola et al. (2004) have recently examined its moments generation.

### 3 The forecast generation

The complex structure of nonlinear models, such as the SETARMA (1), is able to catch some features of the data generating process, often neglected from linear models. They are
referred as non linear features (Fan and Yao, 2003) related for example to the nonnormality of the errors, nonlinear relation among variables, bimodality of the generating process, multimodality of the predictor distribution, sensitivity of predictions to initial conditions and to the forecast horizon.

The listed aspects have heavy impact on the generation of forecasts from model (2) which are further affected from other complications shown in the following.

Suppose that \( \Omega_t = \{X_1, \ldots, X_t\} \) is the vector of the observations from a time series \( \{X_t\} \) and \( h \) is the lead time. The least square predictor of \( X_{t+h} \) is defined as:

\[
f_{t,h}(x, \Omega_t) = \arg \inf_f E[X_{t+h} - f(x, \Omega_t)]^2
\]

where \( f(\cdot) \) is a measurable function over \( \Omega_t \) and it is easy to show that:

\[
f_{t,h}(x, \Omega_t) = E[X_{t+h}|\Omega_t] \equiv X_t(h)
\]

When \( f(\cdot) \) is a linear function, \( X_t(h) \) has some optimal properties in terms of prediction accuracy and variability as shown in Box and Jenkins (1976).

The same results are not always true when \( f(\cdot) \) has a nonlinear structure. This is due to the fact that the forecasts generation can be differently affected. In particular when the SETARMA model (3) is involved, the forecast \( X_t(h) \) is mainly subjected to the threshold variable, which controls the switching among regimes, and the threshold delay which has relevant implications on the predictor form and on its distribution.

In order to show this dependence between the generation of SETARMA forecasts and the threshold delay, the least squares predictor of model (3) is considered:

\[
X_t(h) = E[(\phi_0^{(1)} + \phi_p^{(1)}(B)X_{t+h} + \phi_{q_1}^{(1)}(B)e_{t+h}^{(1)}) \cdot I_{t+h-d}|\Omega_t] + E[(\phi_0^{(2)} + \phi_{p_2}^{(2)}(B)X_{t+h} + \phi_{q_2}^{(2)}(B)e_{t+h}^{(2)}) \cdot (1 - I_{t+h-d})|\Omega_t]
\]

It can be equivalently written as (see Theorem 5.5.11, p.226, in Ash, Doleans-Dade, 2000):

\[
X_t(h) = X_t^{(1)}(h)E (I_{t+h-d} = 1|\Omega_t) + X_t^{(2)}(h) [1 - E (I_{t+h-d} = 1|\Omega_t)]
\]

where \( X_t^{(i)}(h) \) is the ARMA predictor in regime \( i = 1, 2 \).

The result (10) shows that the forecasts of \( X_{t+h} \) can be computed extending the ARMA prediction results to each regime, even if we need to consider the conditional expectation \( E(I_{t+h-d} = 1|\Omega_t) \), for \( h = 1, 2, \ldots \), which has different forms with respect to the values assumed by \( h \) and \( d \). More precisely, when \( h \leq d \) the predictor is derived in a straightforward way following the results of the linear time series analysis. On the contrary, when \( h > d \) the estimation of the threshold variable implies some relevant complications. In the following sections least squares predictors have been generated ignoring estimation uncertainty and further distinguishing, when \( h > d \), among forecasts computed using informations up to time \( t \) and plug-in forecasts where the predicted values of time \( t + 1, \ldots, t + h - d \) are treated as true values.
3.1 The SETARMA predictor when $h \leq d$

The generation of multi-step forecasts, when $h \leq d$, takes advantage of the Box and Jenkins (1976) results which are easily generalized. In fact, in this case, the threshold variable $X_{t+h-d} \in \Omega_t$ and so the indicator function, $I_{t+h-d}$, has values 1 or 0 when $X_{t+h-d} \geq c$ or $X_{t+h-d} < c$ respectively. Therefore the predictor (10) becomes:

$$X_t(h) = X_t^{(1)}(h)I_{t+h-d} + X_t^{(2)}(h)(1 - I_{t+h-d})$$  \hspace{1cm} (11)

and, using the SETARMA representation (6), it can be even written as:

$$X_t(h) = \left[ c_0^{(1)} + \sum_{j=0}^{\infty} \psi_j^{(1)}E(e_{t+h-j}^{(1)}|\Omega_t) \right] \cdot I_{t+h-d} + \left[ c_0^{(2)} + \sum_{j=0}^{\infty} \psi_j^{(2)}E(e_{t+h-j}^{(2)}|\Omega_t) \right] \cdot (1 - I_{t+h-d})$$  \hspace{1cm} (12)

where $E(e_{t+h-j}^{(i)}|\Omega_t) = 0$, for $j = 1, 2, \ldots, h - 1$ and $E(e_{t+h-j}^{(i)}|\Omega_t) = e_{t+h-j}^{(i)}$, for $j \geq h$.

The prediction error is:

$$e_t(h) = \left[ \sum_{j=0}^{h-1} \psi_j^{(1)}e_{t+h-j}^{(1)} \right] I_{t+h-d} + \left[ \sum_{j=0}^{h-1} \psi_j^{(2)}e_{t+h-j}^{(2)} \right] (1 - I_{t+h-d})$$  \hspace{1cm} (13)

which is a linear combination of the errors with weights $\psi_j^{(i)}$, for $j = 0, 1, 2, \ldots, h - 1$ and $i = 1, 2$.

From the prediction error (13), it can be shown that $X_t(h)$ is an unbiased estimate for $X_{t+h}$ as stated in the following proposition.

**Proposition 1** Given the strictly stationary and ergodic SETARMA model (3), with known coefficients, when the lead time is less than the threshold delay ($h \leq d$) and the i.i.d errors $e_t^{(i)}$ have $E[e_t^{(i)}] = 0$ and $\text{var}(e_t^{(i)}) = \sigma_i^2$, for $i = 1, 2$, the least square predictor (12) is an unbiased estimator for $X_{t+h}$, such that:

$$E[X_t(h)] = X_{t+h} \quad h = 1, 2, \ldots, d$$

**Proof:** see Appendix.

Under the hypothesis of proposition 1, the variance of the predictor error (13) is:

$$\sigma^2(h) = \sum_{j=0}^{h-1} \left[ \left( \psi_j^{(1)} \right)^2 \sigma_1^2 I_{t+h-d} + \left( \psi_j^{(2)} \right)^2 \sigma_2^2 (1 - I_{t+h-d}) \right]$$  \hspace{1cm} (14)

which is monotonically nondecreasing with respect to the lead time $h$.

The previous results highlight that the preserved deterministic nature of the indicator function, $I_{t+h-d}$, allows to extend, to the SETARMA model, the Box and Jenkins (1976) forecast theory.

The results completely change when the lead time $h$ is greater than the threshold delay $d$, as shown in the next section.

6
3.2 SETARMA predictor when $h > d$

The strict relation of the SETARMA forecasts with the lead time $h$ and the threshold delay $d$, is clearly highlighted when $h > d$. In section 3.1 the presentation of the case $h \leq d$ has shown no relevant difficulties in the forecasts generation that on the contrary arise now. In particular, given model (3), its predictor (10) implies the estimation of $E(I_{t+h-d} = 1|\Omega_t) = P(X_{t+h-d} \geq c|\Omega_t)$. 

In fact, in this case, $X_{t+h-d} \notin \Omega_t$ and so $I_{t+h-d}$ becomes a Bernoulli random variable:

$$i_{h-d} = \begin{cases} 1 & \text{with } P(X_{t+h-d} \geq c|\Omega_t) \\ 0 & \text{with } P(X_{t+h-d} < c|\Omega_t) \end{cases}$$

for $h = d + 1, d + 2, \ldots$ (15)

where

$$P(X_{t+h-d} \geq c|\Omega_t) = E[i_{h-d}|\Omega_t] = p_{(h-d)}$$

and further the stationarity assumptions on the process $X_t$ allows to write the probability $p_{(h-d)}$ with no dependence from the time $t$.

Using the classical properties of the conditional expected value, the predictor (10) in this case becomes:

$$X_t(h) = X_t(2)(h) + p_{(h-d)} \cdot \left( X_t(1)(h) - X_t(2)(h) \right)$$

which can be even written as:

$$X_t(h) = c_0^{(2)} + \sum_{j=h}^{\infty} \psi_j^{(2)} e_{t+h-j}^{(1)} + \left[ c_0^{(1)} - c_0^{(2)} + \sum_{j=h}^{\infty} \left( \psi_j^{(1)} - \psi_j^{(2)} \right) e_{t+h-j}^{(2)} \right] p_{(h-d)}$$

whereas the prediction error $e_t(h)$, is given as:

$$e_t(h) = e_t(2)(h) + I_{t+h-d} \cdot \left[ e_t(1)(h) - e_t(2)(h) \right] + [I_{t+h-d} - p_{(h-d)}] \cdot [X_t(1)(h) - X_t(2)(h)]$$

where $e_t(i)(h) = \sum_{j=0}^{h-1} \psi_j^{(i)} e_{t+h-j}^{(i)}$ and $X_t(i)(h) = \sum_{j=h}^{\infty} \psi_j^{(i)} e_{t+h-j}^{(i)}$ are the forecast errors and the prediction generated from regime $i$ respectively ($i = 1, 2$).

Even in this case, it can be shown that the predictor (17) is unbiased for $X_{t+h}$, as stated in the following proposition:

**Proposition 2** The least square predictor (17) of the strictly stationary and ergodic SETARMA model (3), with know coefficients, is an unbiased estimator for $X_{t+h}$, that is:

$$E[X_t(h)] = X_{t+h} \quad h = d + 1, \ldots$$

**Proof:** see Appendix.
where, using the properties of $I_{t+h-d}$ in section 2 and the assumptions on the errors distribution $\epsilon_t^{(i)}$, the two forecast variances are $\sigma^2_{t,X}(h) = \sigma^2_t \sum_{j=h}^{\infty} \left( \psi_j^{(i)} \right)^2$, for $i = 1, 2$, the forecast covariance is $\sigma_{12,X}(h) = \sigma_1 \sigma_2 \sum_{j=h}^{\infty} \psi_j^{(1)} \psi_j^{(2)}$ (with the infinite sums truncated at a suitable finite value) and $p$ is the unconditional expected value of $I_{t+h-d}$.

The predictor (16) so obtained can be seen as weighted mean of the two forecasts, $X_t^{(1)}(h)$ and $X_t^{(2)}(h)$, generated from the two regimes whose weights are related to the probability $p_{(h-d)}$. It can be estimated using the informations up to time $t$, $\Omega_t$:

$$\hat{p}_{(h-d)} = \frac{\#(X_{t-d} \in R_1)}{n}$$

(21)

where $\#(X_{t-d} \in R_1)$ is the cardinality of the observations which belong to the first regime and $n$ is the total number of observed data.

### 3.3 Plug-in forecasts

The predictor (17) shows that $X_t(h)$ has threshold variable $X_t(h-d)$ which has been predicted it selves. This suggests the definition of a new augmented set $\Omega_t(h-d) = \{X_1, \ldots, X_t, X_t(1), \ldots, X_t(h-d)\}$ by means of to generate the forecasts which are now obtained from the expectation of $X_{t+h}$ conditional to $\Omega_t(h-d)$:

$$X_t(h) = E[X_{t+h}|\Omega_t(h-d)]$$

(22)

In particular in this case the predictions $X_t(1), \ldots, X_t(h-d)$ which belong to $\Omega_t(h-d)$ are treated as true values whereas the indicator function $I_{t+h-d}$ becomes:

$$i_t(h-d) = [I_{t+h-d}|\Omega_t(h-d)] \begin{cases} 1 & \text{if } X_t(h-d) \geq c \\ 0 & \text{if } X_t(h-d) < c \end{cases}$$

(23)

where the value of $i_t(h-d)$ is related to the forecasts generated at the previous steps.

The predictor (22), called plug-in and denoted $X_t^{PI}(h)$, in order to distinguish it from the least square predictor (17)(denoted $X_t^{LS}(h)$ in the following), is so given as:

$$X_t^{PI}(h) = E[X_{t+h}|\Omega_t(h-d)] = X_t^{(2)}(h) + i_t(h-d) \cdot \left( X_t^{(1)}(h) - X_t^{(2)}(h) \right)$$

(24)

with forecast error:

$$\epsilon_t^{PI}(h) = \epsilon_t^{(2)}(h) + I_{t+h-d} \cdot [\epsilon_t^{(1)}(h) - \epsilon_t^{(2)}(h)] + [I_{t+h-d} - i_t(h-d)] \cdot [X_t^{(1)}(h) - X_t^{(2)}(h)]$$

(25)
where $X_t^{(i)}(h)$ and $e_t^{(i)}(h)$ are defined in the previous section.

It is interesting to note, that even in this case the predictor (24) is an unbiased estimator for $X_{t+h}$.

**Proposition 3** The plug-in predictor $X_t^{PI}(h)$ generated from the strictly stationary and ergodic model (3), with known coefficients, is unbiased for $X_{t+h}$, then:

$$E[X_t^{PI}(h)] = X_{t+h}$$

**Proof:** see Appendix.

The variance of the prediction error (25) is instead given as:

$$
\sigma_{PI,e}^2(h) = \sigma_{1,e}^2(h) + p \cdot [\sigma_{1,e}^2(h) - \sigma_{2,e}^2(h)] + [p + i_t(h - d) - 2p \cdot i_t(h - d)] \cdot [\sigma_{1,X}^2(h) + \sigma_{2,X}^2(h) - 2\sigma_{12,X}(h)]
$$

(26)

that, even in this case, is related to the variance of the forecast errors and of the predictors generated from the two regimes.

## 4 Forecasts evaluation and their combination

When two or more predictors are involved, it needs to find a reasonable criteria which allows to compare them and that gives a measure of their forecast performance. If the selected criteria is the mean square forecast error (MSFE) the comparison of the plug-in and least squares predictors, in the linear context, shows that:

$$E[(X_{t+h} - X_t^{PI}(h))^2 | \Omega_t] \geq E[(X_{t+h} - X_t^{LS}(h))^2 | \Omega_t]$$

and, according to the results of proposition 3.4 in Fan and Yao (2003), it is equivalent to state that $\sigma_{PI,e}^2(h) > \sigma_{LS,e}^2(h)$, hence leading to prefer the least square predictor.

This result is not always true when the nonlinearity is involved. In particular, given the forecast variances (20) and (26), it can be shown that:

$$\frac{\sigma_{LS,e}^2(h)}{\sigma_{PI,e}^2(h)} \geq 1 \quad \text{if} \quad X_t(h - d) \geq c$$

$$\frac{\sigma_{LS,e}^2(h)}{\sigma_{PI,e}^2(h)} \leq 1 \quad \text{if} \quad X_t(h - d) < c$$

(27)

so highlighting that the ratio between the variances (and so the related forecast performances) of the least-squares and plug-in forecast errors changes with respect to the regime involved from the threshold variable.

**Remark 1:** The inequalities (27) are based on the ratio $\frac{\sigma_{LS,e}^2(h)}{\sigma_{PI,e}^2(h)} = \frac{p + \rho^2(h - d) - 2p \cdot \rho(h - d)}{p + i_t(h - d) - 2p \cdot i_t(h - d)}$ and on the result that for $n$ enough large $p(h - d)$ converges to $p$, where an estimate for $p(h - d)$ is given in (21).
To take advantage of the result (27), the unbiased predictors (17) and (24) can be combined as follows:

\[ X_t^C(h) = i_t(h - d)X_t^{PL}(h) + [1 - i_t(h - d)]X_t^{LS}(h) \]

\[ X_t^{LS}(h) + i_t(h - d) \left[ X_t^{PL}(h) - X_t^{LS}(h) \right] \]

(28)

where it is trivial to show that \( X_t^C(h) \) is an unbiased forecast for \( X_{t+h} \), when \( c_0^{(1)} = c_0^{(2)} \), and the variance of the forecast error \( e_t^C(h) \) is such that:

\[ \sigma_{C,e}^2(h) \leq \sigma_{LS,e}^2(h) \quad \text{and} \quad \sigma_{C,e}^2(h) \leq \sigma_{PL,e}^2(h) \]

(29)

The combination (28) so generates forecasts picking the ”best” (in term of MSFE) predictor selected according to the indicator function \( i_t(h - d) \) and conditional to the augmented set \( \Omega_t(h - d) \).

5 Simulation results

In order to show the main features of the theoretical results presented a simulation study has been performed using SETARMA models of different orders.

In particular, starting from the following SETARMA(2; 1,1; 1,1) model:

\[
X_t = \begin{cases} 
\phi_0^{(1)} - 0.87X_{t-1} + e_t^{(1)} + 0.69e_{t-1}^{(1)} & X_{t-1} \geq 0 \\
\phi_0^{(2)} - 0.83X_{t-1} + e_t^{(2)} + 0.52e_{t-1}^{(2)} & X_{t-1} < 0 
\end{cases}
\]

(30)

1000 time series of length 500 have been simulated and forecasts from 1 to 5 steps ahead have been generated in order to evaluate the properties of the three predictors Least Squares, Plug In and Combined discussed in the previous pages. The errors of the simulated models are \( e_t^{(i)} \sim WNN(0, \sigma_i^2) \) with \( \sigma_1^2 = 1 \) and \( \sigma_2^2 = 3 \) whereas the autoregressive and moving average coefficients of the two regimes are selected in order to fulfill the stationary conditions and to allow an equally distribution of the simulated values between the two regimes.

Starting from the case \( h \leq d \) the forecasts have been generated from model (30) using the predictor (12) and subsequently the forecast errors \( e_t(h) = X_{t+h} - X(h) \) have been computed. In Figure 1 the empirical densities of the forecast errors have been shown fixing \( h = d = 1 \) and with \( \phi_0^{(1)} = \phi_0^{(2)} = 0 \) (which correspond to \( c_0^{(1)} = c_0^{(2)} \)) in frame (a) and \( \phi_0^{(1)} \neq \phi_0^{(2)} \) in frame (b).

In both figures a dashed line is inserted at \( e_t(1) = 0 \) in order to highlight, even empirically, the result in proposition 1 and further to show the behavior of the correspondent prediction errors distribution.

As expected, the two densities do not present remarkable differences. This is not new in the context of the nonlinear time series models with changes in regimes and confirms the results available in the literature. For this reason in the following the attention will be
mainly addressed to multi-step forecasts with \( h > d \). Under this condition, in Section 3 three predictors have been presented and briefly called \( X_t^{LS}(h) \), \( X_t^{PI}(h) \) and \( X_t^C(h) \). The unbiasedness of all three has been shown when \( \phi_0^{(1)} = \phi_0^{(2)} \) whereas this property is not fulfilled from \( X_t^C(h) \) when \( \phi_0^{(1)} \neq \phi_0^{(2)} \). This result is even presented in Figure 2 where the empirical densities of the forecast errors from model (30) are shown for \( h = 2 \), \( \phi_0^{(1)} = 0.3 \) and \( \phi_0^{(2)} = -0.34 \). More precisely, the unbiasedness of \( X_t^{LS}(h) \) and \( X_t^{PI}(h) \) is confirmed whereas slight negative biasedness is presented for \( X_t^C(h) \). The biasedness of \( X_t^C(h) \) disappears when \( \phi_0^{(1)} = \phi_0^{(2)} = 0 \) as shown in Figure 3 where the empirical density of \( e_t^C(2) \) is presented.

The results shown for model (30) are confirmed with \( h = 3, 4, 5 \) even if, given the model stationarity, they cannot always be clearly distinguished as in the previous figures. Interesting results are even obtained through more complicated models. For example two SETARMA(2; 2,1; 1,1) models, whose parameters have been chosen following the same criteria of model (30), given as:

\[
X_t = \begin{cases} 
-0.87X_{t-1} - 0.75X_{t-2} + e_t^{(1)} + 0.69e_{t-1}^{(1)} & X_{t-1} \geq 0 \\
-0.83X_{t-1} + e_t^{(2)} + 0.52e_{t-1}^{(2)} & X_{t-1} < 0 
\end{cases}
\]

(31)

\[
X_t = \begin{cases} 
0.7 - 0.87X_{t-1} - 0.75X_{t-2} + e_t^{(1)} + 0.69e_{t-1}^{(1)} & X_{t-1} \geq 0 \\
-0.45 - 0.83X_{t-1} + e_t^{(2)} + 0.52e_{t-1}^{(2)} & X_{t-1} < 0 
\end{cases}
\]

(32)

have been used to further compare the three predictors (when \( h \geq d \)).

Under the same assumptions fixed for model (30) on the errors \( e_t^{(i)}, i = 1, 2 \), a time series of length 500 has been simulated from each model and \( h \) steps ahead forecasts, with \( h = 2, \ldots, 5 \) from (17), (24) and (28) have been generated under the assumption that the parameters of the model are known.

Figure 1: Empirical densities of the forecast errors \( e_t(1) \) from the predictor (12) when \( \phi_0^{(1)} = \phi_0^{(2)} = 0 \) (frame (a)) and \( \phi_0^{(1)} \neq \phi_0^{(2)} \) (frame (b))
The criterion used to evaluate the forecast accuracy is the Mean Square Forecast Error:

\[ MSFE = \frac{1}{h} \sum_{i=1}^{h} [X_{t+i} - \hat{X}_t(i)]^2 \]

and in Table 1, in order to simplify the comparison of the plug-in and combined forecasts versus the least squares, the ratios of the \(MSFE(PI)/MSFE(LS)\) and \(MSFE(C)/MSFE(LS)\) are presented, for \(h = 2, \ldots, 5\).

<table>
<thead>
<tr>
<th>Model (31)</th>
<th></th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(h)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(PI/LS)</td>
<td></td>
<td>1.0502</td>
<td>1.0342</td>
<td>1.0443</td>
<td>1.0396</td>
</tr>
<tr>
<td>(C/LS)</td>
<td></td>
<td>0.9879</td>
<td>0.9944</td>
<td>0.9911</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model (32)</th>
<th></th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(h)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(PI/LS)</td>
<td></td>
<td>1.4483</td>
<td>1.1612</td>
<td>1.0126</td>
<td>1.0038</td>
</tr>
<tr>
<td>(C/LS)</td>
<td></td>
<td>0.8506</td>
<td>0.8751</td>
<td>0.8762</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Ratios of the mean square forecast errors of the plug-in and combined forecasts with respect to the least squares forecasts

In all cases the combination seems to be preferred further confirming the effectiveness of...
Figure 3: Empirical densities of the prediction errors generated from the forecasts Combination when $h = 2$ and $\phi_0^{(1)} = \phi_0^{(2)} = 0$

(28) in increasing the forecast accuracy. Further for both models at $h = 2$ the ratio $MSFE(C)/MSFE(LS)$ is one because the selected predictor from (28) is the LS.

6 Conclusions

The multi-step predictors of the SETARMA model have been extensively presented. In particular the attention has been focused on the key role assumed by the threshold delay $d$ when the predictor needs to be selected. In fact, when $h \leq d$ the results of the linear time series models can be used within each regime taking advantage of the local linearity of the model. When $h > d$, two different predictors, called least squares and plug-in, have been generated and after comparing their accuracy in term of MSFE, they have been properly combined and evaluated.

Interesting results have been illustrated in a forecasting exercise which strongly supports the forecast combination. This last result needs more investigation and a wider simulation study which will be the object of future research.

References


Appendix

Proof of Proposition 1
Starting from:

\[ X_{t+h} = X_t(h) + e_t(h) \]

with expected value:

\[ E[X_t(h)] = E[X_{t+h} - e_t(h)] = X_{t+h} - E[e_t(h)] \] (33)

from (13), \( E[e_t(h)] \) is given as:

\[
E[e_t(h)] = E \left[ \sum_{j=0}^{h-1} \psi_j^{(1)} e_{t+h-j}^1 \cdot I_{t+h-d} \right] + E \left[ \sum_{j=0}^{h-1} \psi_j^{(2)} e_{t+h-j}^2 \cdot (1 - I_{t+h-d}) \right] = \\
= \sum_{j=0}^{h-1} \psi_j^{(1)} E \left[ e_{t+h-j}^1 \cdot I_{t+h-d} \right] \cdot E[I_{t+h-d}] + \sum_{j=0}^{h-1} \psi_j^{(2)} \cdot E \left[ e_{t+h-j}^2 \cdot (1 - I_{t+h-d}) \right] \cdot E[1 - I_{t+h-d}] \] (34)

Remarking that when \( h \leq d \) the value \( X_{t+h-d} \in \Omega_t \), in this case we have that:

\[ E[I_{t+h-d}] = I_{t+h-d} \begin{cases} 
1 & \text{if } X_{t+h-d} \geq c \\
0 & \text{if } X_{t+h-d} < c 
\end{cases} \]

and so, observing that \( E(e_{t+h-j}^{(i)}|\Omega_t) = 0 \) for \( j = 1, 2, \ldots, h-1 \) and \( i = 1, 2 \), the expected value (34) becomes:

\[
E[e_t(h)] = \sum_{j=0}^{h-1} \psi_j^{(1)} E \left[ e_{t+h-j}^1 \cdot I_{t+h-d} = 1 \right] + \sum_{j=0}^{h-1} \psi_j^{(2)} E \left[ e_{t+h-j}^2 \cdot I_{t+h-d} = 0 \right] = 0 \] (35)

From this last result, the expected value (33) is such that:

\[ E[X_t(h)] = X_{t+h} \]

which completes the proof.
Proof of Proposition 2
Even in this case, the proof of the unbiasedness of $X_t(h)$ can be equivalently demonstrated showing that $E[e_t(h)] = 0$. In particular, given the prediction error $e_t(h)$ in (18), it can be even written as:

$$ e_t(h) = \sum_{j=0}^{h-1} \psi_j^{(2)} e_{t+h-j} + I_{t+h-d} \left[ \sum_{j=0}^{h-1} \psi_j^{(1)} e_{t+h-j} - \sum_{j=0}^{h-1} \psi_j^{(2)} e_{t+h-j} \right] + (I_{t+h-d} - p_{h-d}) \left( \sum_{j=h}^{\infty} \psi_j^{(1)} e_{t+h-j} - \sum_{j=h}^{\infty} \psi_j^{(2)} e_{t+h-j} \right) $$

(36)

Observing that for $h > d$, $I_{t+h-d}$ is a random variable with distribution (15) and with $E(i_{h-d}) = P(X_{t+h-d}|\Omega_t) = p(h-d)$, the expected value of (36) is:

$$ E[e_t(h)] = \sum_{j=0}^{h-1} \psi_j^{(2)} E[e_{t+h-j}] + p(h-d) \left[ \sum_{j=0}^{h-1} \psi_j^{(1)} E[e_{t+h-j}] - \sum_{j=0}^{h-1} \psi_j^{(2)} E[e_{t+h-j}] \right] |I_{t+h-d} = 0 $$

which implies that $E[X_t(h)] = E[X_{t+h} + e_t(h)] = X_{t+h}$.

Proof of Proposition 3
The proof follows the same steps of proposition 2.
Starting from:

$$ e_t^{PI}(h) = X_{t+h} - X_t^{PI}(h) $$

(37)

the prediction error $e_t^{PI}(h)$ is:

$$ e_t^{PI}(h) = \sum_{j=0}^{h-1} \psi_j^{(2)} e_{t+h-j} + I_{t+h-d} \left[ \sum_{j=0}^{h-1} \psi_j^{(1)} e_{t+h-j} - \sum_{j=0}^{h-1} \psi_j^{(2)} e_{t+h-j} \right] + (I_{t+h-d} - i_t(h - d)) \left( \sum_{j=h}^{\infty} \psi_j^{(1)} e_{t+h-j} - \sum_{j=h}^{\infty} \psi_j^{(2)} e_{t+h-j} \right) $$

with unconditional expected value:
\[ E[e_t^{PI}(h)] = \sum_{j=0}^{h-1} \psi_j^{(2)} E[e_{t+h-j}^{(2)}] + p(h-d) \left[ \left( \sum_{j=0}^{h-1} \psi_j^{(1)} E[e_{t+h-j}^{(1)}] - \sum_{j=0}^{h-1} \psi_j^{(2)} E[e_{t+h-j}^{(2)}] \right) |I_{t+h-d}| \right] + \\
+ p(h-d) \left[ \left( \sum_{j=h}^{\infty} \psi_j^{(1)} E[e_{t+h-j}^{(1)}] - \sum_{j=h}^{\infty} \psi_j^{(2)} E[e_{t+h-j}^{(2)}] \right) |I_{t+h-d}| \right] + \\
- i_t(h-d) \left( \sum_{j=h}^{\infty} \psi_j^{(1)} E[e_{t+h-j}^{(1)}] - \sum_{j=h}^{\infty} \psi_j^{(2)} E[e_{t+h-j}^{(2)}] \right) = 0 \\
\]

This implies that the expectation of (37) is:

\[ E[X_t+h - X_t^{PI}(h)] = 0 \]

and so \( X_{t+h} = E[X_t^{PI}(h)] \).