

FORECASTS BY PREDICTIVE LIKELIHOOD IN
THRESHOLD AUTOREGRESSIVE MODEL
G. González-Farías, J. Gonzalo and M. G. Russell-Noriega

Comunicación Técnica No I-07-10/14-06-2007
(PE/CIMAT)



Forecasts by Predictive Likelihood in Threshold Autoregressive Model

G. González-Farías (Research Center in Mathematics, México)

J. Gonzalo (Carlos III University, Spain)

M. G. Russell-Noriega (Guanajuato University, México)

June 8, 2007

Abstract

In this paper we study the performances of h-step ahead predictions in TAR models versus those of AR models. We propose the alternative of Predictive Likelihood (PL), based on the principle of likelihood. Unlike other alternatives for prediction generation, PL jointly attacks the problems of estimating, obtaining and evaluating predictions. Which, for non linear models, presents an integrated way to take into account in both processes, the DGP. We also study the asymptotic properties of the maximum likelihood estimators under a known threshold parameter. We implement a wide simulation exercise to compare the predictions under PL and the standard prediction generating method, identified as the recursive Monte Carlo method. Comparison of predictors performances is implemented through the Mean Square Forecast Errors (MSFE). The gain observed in the performances of predictions by means of PL is clear and mainly due to the fact that the predictions for both the threshold variable and the autoregressive variable are obtained simultaneously.

1 Introduction

Non-linear time series models are a useful tool to describe and forecast a great variety of time series. However, several empirical studies have shown that although such models fit well inside the sample, they seldom represent a substantial gain in their forecasts outside the sample, when compared to those stemming from autoregressive models. Some authors like Pesaran and Potter (1997) and Clements and Smith (1999), suggest that nonlinear models may be better performing point forecast competitors as a function of their respective Mean Square Forecast Errors (MSFE's), if the conditional variances for each regime are considered. Clements and Smith (1999) compute MSFE's conditioning on each of the regimes and computing the statistics related to the direction of change as well. The study by Pesaran and Potter(1997) suggests that nonlinear models may be better at predicting high-order moments, implying that the performance of the prediction intervals and of the density predictive is better than their corresponding linear competitor.

Obtaining forecasts in nonlinear models is more difficult than in the linear case, because it involves solving multiple integrals that depend on the number of forecasts to compute. Generally, consider the model,

$$Y_t = g(Y_{t-1}) + \varepsilon_t, \text{ with } \varepsilon_t \sim IID(0, \sigma^2), \quad (1)$$

and $g(\cdot)$, a nonlinear function. Let \mathfrak{F}_T be the σ -algebra generated by the information set up to time T , i.e., $\{Y_1, \dots, Y_T\}$. Using a least squares criterion, the optimal one-step ahead forecast is

$$\hat{Y}_{T+1} = E[Y_{T+1}|\mathfrak{F}_T] = g(Y_T).$$

However, given the nonlinearity of function g , the optimal two-step ahead forecasts is

$$\hat{Y}_{T+2} = E[Y_{T+2}|\mathfrak{F}_T] = E[g(Y_{T+1})|\mathfrak{F}_T] \neq g(E[Y_{T+1}|\mathfrak{F}_T]) = g(\hat{Y}_{T+1}).$$

Or equivalently,

$$\begin{aligned} \hat{Y}_{T+2} &= E[Y_{T+2}|\mathfrak{F}_T] = E[g(Y_{T+1})|\mathfrak{F}_T] = E[g(g(Y_T) + \varepsilon_{T+1})|\mathfrak{F}_T] \\ &= E\left[g\left(\hat{Y}_{T+1} + \varepsilon_{T+1}\right)|\mathfrak{F}_T\right] \neq g\left(\hat{Y}_{T+1}\right). \end{aligned}$$

In a similar way, we can express the h -step ahead forecasts through the following identity:

$$\begin{aligned} \hat{Y}_{T+h} &= E[Y_{T+h}|\mathfrak{F}_T] = E[g(Y_{T+h-1})|\mathfrak{F}_T] \\ &= E\left[g\left(g\left(g\left(\dots g\left(\hat{Y}_{T+1} + \varepsilon_{T+1}\right)\right) + \dots + \varepsilon_{T+h-3}\right) + \varepsilon_{T+h-2} + \varepsilon_{T+h-1}\right)|\mathfrak{F}_T\right], \end{aligned}$$

where function g is evaluated $h - 1$ times.

In nonlinear models, obtaining h -step ahead forecasts generally involves the analytical solution of multiple integrals, and thus the use of alternative numerical integration techniques or simulation methods, Tong (1990). The methods most commonly used in the literature to approach the conditional expectation for the different prediction horizons are: the exact method, the naive method, the Monte-Carlo method and the bootstrap method, discussed in Granger and Teräsvirta (1993); Clements and Smith (1997,1998); Diebold and Mariano (1995); De Gooijer and De Bruin (1998), Davies *et al* (1998), Diebold *et al* (1998); Clements *et al* (2003); Russell-Noriega (2006), among others. From a practical standpoint, the most used method is the Monte Carlo one, given its ease of implementation and properties of convergence for a large enough N .

The Monte Carlo method calculates the conditional expectation through a recursive computational method. The two-step ahead forecasts, for instance, is approximated by the following equation:

$$\hat{Y}_{MC,t+2} = \frac{1}{N} \sum_{j=1}^N g\left(\hat{Y}_{MC,T+1} + Z_j\right)$$

where Z_j , $j = 1, \dots, N$ are extracted from a known distribution D function. Under this scheme, the Monte Carlo forecast for Y_{T+h} with $h > 1$ is

$$\hat{Y}_{MC,T+h} = \frac{1}{N} \sum_{j=1}^N \underbrace{g(g(\dots g(\hat{Y}_{MC,T+1} + Z_j^1) \dots))}_{h-1 \text{ times}} + Z_j^{h-1},$$

with $Z_j^1, Z_j^2, \dots, Z_j^{h-1}$ generated from the distribution D , for $j = 1, 2, \dots, N$.

In practice, function $g(\cdot)$ is unknown and must be specified or estimated through some estimation procedure. The estimated function $\hat{g}(\cdot)$ is substituted into the method considered to obtain the forecasts of interest.

Most of the literature on obtaining and evaluating predictions has focused on point predictions, where the summary of uncertainty of the predictions is measured through their standard errors Clements and Smith (1997), Clements *et al* (2003), Diebold *et al* (1998), Diebold and Mariano (1995), Dacco and Sanchell (1999), among others. Chistosffersen (1998) suggests several forms to assess the performance of the predictions through the evaluation of intervals of conditional predictions, while Diebold *et al* (1998) and Berkowitz (2001) propose some methods to assess the density of the prediction. Clements *et al* (2003) argue that the evaluation methods of the predictive density, under the assumption of normal errors in the DGP, may be viewed as an evaluation method for point predictions, since each of the sequence of normal probabilities is equivalent to the errors of the scaled predictions (See Russell-Noriega (2006)). Regardless of the evaluation method considered, the overall conclusion is that the gain in the prediction is larger in the SETAR model, as long as the prediction and the performance mean used are obtained conditioning on each one of the regimes. Our particular interest lies in economics applications, where it has been observed that autoregressive threshold models are able to capture characteristics commonly seen in economics series, such as: irreversibility in time, asymmetry, persistence, etc.. Also, from an economics standpoint it is more realistic to assume that the changes through time from one regime to another are due to lags or a lag function of an exogenous random variable, rather than to lags in the series itself. Moreover, there are practical situations in which to assume an autoregressive linear model implies the consideration that this phenomenon behaves as a unitary root process, and so the notion of conditioning to obtain and evaluate predictions is not applicable. González and Gonzalo (1998,1999) analyze USA interest rates quarterly data during the period between August 1959 and June 1999, from Citibase, considering inflation changes as the threshold variable. The mathematical representation of the TAR model, with threshold variable Z_{t-d} , is given by:

$$\begin{aligned} Y_t &= [\alpha_1 I(Z_{t-d} \leq \gamma_1) + \dots + \alpha_T I(Z_{t-d} > \gamma_{T-1})] Y_{t-1} + \varepsilon_t \\ &= \delta_t Y_{t-1} + \varepsilon_t, \quad t = 1, 2, \dots, \end{aligned} \quad (2)$$

where $\delta_t = \alpha_1 I(Z_{t-d} \leq \gamma_1) + \dots + \alpha_T I(Z_{t-d} > \gamma_{T-1})$, $I(\cdot)$ is the indicator function, and the processes ε_t and Z_t satisfy the following conditions:

- S1. The processes (ε_t, Z_{t-d}) are strictly stationary and ergodic.
- S2. $E(\varepsilon_t | \mathfrak{F}_{t-1}) = 0$ and $E(\varepsilon_t^2 | \mathfrak{F}_{t-1}) = \sigma^2$.
- S3. For some $\tau > 1$, $E(\varepsilon_t^{2\tau} | \mathfrak{F}_{t-1}) \leq B < \infty$.
- S4. $E(\max(0, \log |\varepsilon_1|)) < \infty$.
- S5. The essential supreme of $|\varepsilon_1| < \infty$, that is,

$$\text{ess sup } |\varepsilon_1| = \inf \{x : \Pr(|\varepsilon_1| > x) = 0\} < \infty.$$

- S6. ε_1 admits a continuous and positive probability density function .

González and Gonzalo (1998) show the ergodicity and stationarity properties for the TAR model given in equation (2). In particular, it is shown the TAR model considered is stationary in covariance if $E(\delta_t) < 1$.

Part of the discussion on the badly performance of predictions in TAR models it relates to the problem of incorrect specification of the regimes. Dacco and Sanchell (1999) illustrate this issue in the Regime-Switching versus random walk models with and without drift. It is concluded that the bad performance of the one-step ahead forecast is due to large values for the probability of misclassification. Russell-Noriega (2006), study the behaviors of the MSFE in TAR and AR models, under incorrect specifications. Bad specification due to the

incorrect classification of the observations in the regimes is first considered, similarly to the case studied in Dacco and Sanchell (1999). Following the incorrect specification of the true DGP, the expressions of the MSFE are obtained under the assumption of a linear model for the forecasts, when the real DGP is a TAR model. The analysis becomes to rewrite the TAR model as a type linear model AR. If the threshold variable Z_{t-1} follows a sequence of independent random variables, it is shown that the error process associated to a linear model follows a volatility process. If the threshold variable Z_{t-1} is given by an AR(1) time series the volatility behavior of the error process is shown through a simulation exercise. The error behavior departs from a white noise process as a function of the values of the autoregressive parameters in the TAR model. The asymmetric behaviors observed in a great variety of real phenomena cannot be described through a linear model, so the residuals inherit the asymmetrical behaviors due to the incorrect specification of the true generating process. However, the increase in variability in the model is counteracted by considering different methods for estimating, obtaining and evaluating, possibly favoring the linear model, and causing the predictions from the TAR model to show poor performances when compared to their linear competitors. Our predictive likelihood proposal is more effective since it jointly considers the processes of estimating, obtaining and evaluating the predictions.

2 Maximum likelihood estimation in TAR models

One of the practical characteristics of threshold models when compared to other nonlinear models is that the estimation procedures are relatively easy to implement computationally. The Ordinary Least Squares (OLS) method is one of the most commonly used estimation procedures in the literature. OLS estimators for SETAR models, as well as their asymptotic properties, are mainly due to the work of Petrucci and Woolford (1984), Chan *et al* (1985), Chen and Tsay (1991), Chan (1993), and Hansen (1997). Russell-Noriega (2006) proves the consistency and asymptotic normality properties of the OLS estimators for the TAR model given in equation (2), under known γ . Estimation of the γ parameter may be performed through OLS, by means a direct search of the parameter in a compact set (Chan (1993) and Hansen (1997)), or through the maximization of the profile log likelihood function discussed in Russell-Noriega (2006).

The estimation procedure considered here is based on the principle of Maximum Likelihood (ML). It is shown that the ML estimators in the TAR model are strongly consistent, under the assumption that the error distribution has a finite fourth moment. In particular, the normal distribution satisfies this assumption, implying that the OLS estimators are strongly consistent since they coincide with the ML estimators under normality distribution of the errors.

The following equivalent representation of the stationary and ergodic TAR model, given in equation (2), is considered

$$\begin{aligned} Y_t &= [\phi_1 I(Z_{t-1} \leq \gamma) + \phi_2 I(Z_{t-1} > \gamma)] Y_{t-1} + \varepsilon_t \\ &= h(Z_{t-1}, Y_{t-1}, \boldsymbol{\theta}) + \varepsilon_t, \quad t = 1, 2, \dots, \end{aligned} \quad (3)$$

where $h(\cdot) = [\phi_1 I(Z_{t-1} \leq \gamma) + \phi_2 I(Z_{t-1} > \gamma)] Y_{t-1}$, $I(\cdot)$ denotes the indicator function, and in general Z_{t-1} is a time series observed up to time t . The threshold variable Z_{t-1} is a stationary time series, since the Y_t values depend on the values taken $I(Z_{t-1} \leq \gamma)$. It is also considered that the behavior of the Z_{t-1} series is not ruled by the behavior of the Y_t series (Z_{t-1} causes Y_t , but Y_t does not cause Z_{t-1}). Two scenarios for the Z_{t-1} process are

considered, Z_{t-1} as an autoregressive stationary process of order one ($Z_{t-1} \sim \text{AR}(1)$) and Z_{t-1} as an independent process with uniform distribution over the interval $(0,1)$, that is, $Z_{t-1} \sim \text{iid}U(0,1)$.

2.1 Obtaining maximum likelihood estimators

Let us assume that the errors $\{\varepsilon_t\}$ in equation (3) are identically distributed, independent random variables with distribution D , *i.e.*, $\varepsilon_t \sim \text{iid}D$. ε_t independent from Y_{t-1}, Y_{t-2}, \dots , $t \geq 1$, and with a D distribution function with mean zero and finite variance. Let process Z_{t-1} , be stationary, given by: $Z_{t-1} = \rho Z_{t-2} + u_t = h_1(Z_{t-2}, \rho) + u_t$, with $u_t \sim \text{iid}N(0, \sigma_u^2)$ and ε_t mutually independent from Z_{t-1} , such that the likelihood function can be written as:

$$\begin{aligned} & L(\phi_1, \phi_2, \sigma^2, \gamma, \rho, \sigma_u^2; \{Y_t, Z_t\}) \\ &= \prod_{t=2}^T f(Y_t | Y_{t-1}, Z_{t-1}; \phi_1, \phi_2, \sigma^2, \gamma) f_1(Z_{t-1} | Z_{t-2}; \rho, \sigma_u^2) g_{\boldsymbol{\theta}}(Y_1, Z_0) \\ &= \prod_{t=2}^T f(Y_t - h(Y_{t-1}, Z_{t-1}, \boldsymbol{\vartheta}); \sigma) f_1(Z_{t-1} - h_1(Z_{t-2}, \rho); \sigma_u) g_{\boldsymbol{\theta}}(Y_1, Z_0), \end{aligned}$$

where $\boldsymbol{\vartheta} = (\phi_1, \phi_2, \gamma)'$, and f and f_1 , are the density functions of ε_t and u_t , respectively. Likewise, the conditional likelihood function of the parameter vector $\boldsymbol{\theta} = (\phi_1, \phi_2, \sigma^2, \gamma, \rho, \sigma_u^2)'$, is

$$L(\boldsymbol{\theta}; \{Y_t, Z_t\}) = \prod_{t=2}^T f(Y_t - h(Y_{t-1}, Z_{t-1}, \boldsymbol{\vartheta}); \sigma) f_1(Z_{t-1} - h_1(Z_{t-2}, \rho); \sigma_u). \quad (4)$$

For an observed sample Y_1, \dots, Y_T , Z_0, \dots, Z_{T-1} and a D a normal distribution with mean zero and variance σ^2 , we obtain a conditional log likelihood function of $\boldsymbol{\theta}$, given the sample, of the form:

$$\begin{aligned} \ell(\boldsymbol{\theta}) &= -(T-1) \log \sigma - \frac{1}{2\sigma^2} \sum_{t=2}^T (Y_t - \delta_t Y_{t-1})^2 - (T-1) \log \sigma_u \\ &\quad - \frac{1}{2\sigma_u^2} \sum_{t=2}^T (Z_{t-1} - \rho Z_{t-2})^2. \end{aligned}$$

From the previous equation, solving $\partial \ell(\boldsymbol{\theta}) / \partial \sigma^2 = 0$, we get the following expression for the estimated variance:

$$\hat{\sigma}^2(\phi_1, \phi_2, \gamma) = \frac{1}{T-1} \sum_{t=2}^T (Y_t - \delta_t Y_{t-1})^2,$$

Likewise, solving the equation $\frac{\partial \ell(\boldsymbol{\theta})}{\partial \phi_i} = 0$, for $i = 1, 2$ we have:

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \phi_i} = -\frac{1}{2\sigma^2} \sum_{t=2}^T (-2I_t Y_{t-1} Y_t + 2\phi_i I_t Y_{t-1}^2) = 0, \quad i = 1, 2.$$

And from here we find the ML estimators for ϕ_1 and ϕ_2 as a function of parameter γ ,

$$\hat{\phi}_1(\gamma) = \frac{\sum_{t=2}^T I_t Y_t Y_{t-1}}{\sum_{t=2}^T I_t Y_{t-1}^2}, \quad \hat{\phi}_2(\gamma) = \frac{\sum_{t=2}^T Y_t Y_{t-1} - \sum_{t=2}^T I_t Y_t Y_{t-1}}{\sum_{t=2}^T Y_{t-1}^2 - \sum_{t=2}^T I_t Y_{t-1}^2},$$

with $I_t = I(Z_t \leq \gamma)$, hence the ML estimator for σ^2 as a function of γ is given by:

$$\hat{\sigma}^2(\gamma) = \hat{\sigma}^2(\hat{\phi}_1, \hat{\phi}_2, \gamma) = \frac{1}{T-1} \sum_{t=2}^T \left(Y_t - \hat{\delta}_t(\gamma) Y_{t-1} \right)^2,$$

with:

$$\hat{\delta}_t(\gamma) = \begin{cases} \hat{\phi}_1(\gamma), & Z_{t-1} \leq \gamma, \\ \hat{\phi}_2(\gamma), & Z_{t-1} > \gamma. \end{cases}$$

Similarly, the ML estimators for ρ and σ_u^2 , are, respectively:

$$\hat{\rho} = \frac{\sum_{t=2}^T Z_{t-1} Z_{t-2}}{\sum_{t=2}^T Z_{t-2}^2}, \quad \hat{\sigma}_u^2(\hat{\rho}) = \frac{1}{T-1} \sum_{t=2}^T (Z_t - \hat{\rho} Z_{t-1})^2.$$

The expressions for the maximum likelihood estimators for coefficients ϕ_1 , ϕ_2 and σ^2 assume a fixed γ value. It is not possible to derive the log likelihood function with respect to parameter γ , because it is not continuous at the points Z_0, \dots, Z_{T-1} , so in this case a search method is recommend to estimate the γ parameter. Once the MLE's for a fixed γ have been calculated, the log likelihood function is evaluated at this values, obtaining the profile log likelihood function for γ , i.e.,

$$\begin{aligned} \ell_p(\gamma) &= \ell \left[\hat{\phi}_1(\gamma), \hat{\phi}_2(\gamma), \hat{\sigma}_\gamma, \hat{\rho}, \hat{\sigma}_u \right] \\ &= -(T-1) \log \hat{\sigma}_\gamma - \frac{1}{2\hat{\sigma}_\gamma^2} \sum_{t=2}^T \left(Y_t - \hat{\delta}_t(\gamma) Y_{t-1} \right)^2 \\ &\quad - (T-1) \log \hat{\sigma}_u^2 - \frac{1}{2\hat{\sigma}_u^2} \sum_{t=2}^T (Z_t - \hat{\rho} Z_{t-1})^2. \end{aligned}$$

Note that $\hat{\rho}$ and $\hat{\sigma}_u$ do not depend functionally on γ , so the effects of maximization respect to parameter γ , maximizing $\ell_p(\gamma)$ is equivalent to maximizing:

$$\begin{aligned} \ell_p^*(\gamma) &= -(T-1) \log \hat{\sigma}_\gamma - \frac{1}{2\hat{\sigma}_\gamma^2} \sum_{t=2}^T \left(Y_t - \hat{\delta}_t(\gamma) Y_{t-1} \right)^2 \\ &= -\frac{T-1}{2} \log \left[\frac{1}{T-1} \sum_{t=2}^T \left(Y_t - \hat{\delta}_t(\gamma) Y_{t-1} \right)^2 \right]. \end{aligned}$$

Likewise, maximizing $\ell_p^*(\gamma)$ in the previous equation is equivalent to minimizing $-\ell_p^*(\gamma)$, with respect to γ . This procedure gives us a large number of γ values; Qian's (1998) recommendation is to take $\hat{\gamma}$ as the value of γ that satisfies the following condition:

$$\inf \arg \min_{\gamma} \frac{T-1}{2} \log \left[\frac{1}{T-1} \sum_{t=2}^T \left(Y_t - \hat{\delta}_t(\gamma) Y_{t-1} \right)^2 \right].$$

Equations for the case in which the variable $Z_{t-1} \sim iidU(0,1)$ simplify since the density function $f_1(Z_{t-1}|Z_{t-2}; \rho, \sigma_u^2) = 1$. Generalization of the likelihood equations for TAR models is direct; however we will use the TAR(2;1,1) model for illustration purposes.

2.2 Asymptotic properties of maximum likelihood estimators

The study of the asymptotic properties of OLS and ML parameter estimators in SETAR models for an unknown γ is found in the papers by Chan (1993), Qian (1998), and Chan and Tsay (1998). Chan (1993) and Qian (1998) consider the case of discontinuous SETAR models, while Chan and Tsay (1998) do so for continuous SETAR.

Definition 1 (Discontinuous autoregressive function, Chan 1993). *Let ϕ_1 and ϕ_2 be coefficients of the model SETAR(2;1,1), it is said that the model has a discontinuous autoregressive function if there exists*

$$Z_* = z_0, \tag{5}$$

such that $(\phi_1 - \phi_2) Z_ \neq 0$ and $z_0 = \gamma$. In this case, the threshold γ becomes the autoregressive function break point. If $(\phi_1 - \phi_2) Z_* = 0$ for all Z_* satisfying (5), it is said that the model has a continuous autoregressive function.*

Note that the continuity definition for the autoregressive function relates to a *smooth regime change* when the value of $Y_{t-1} = \gamma$, for $t = 2, \dots, T$.

Chan's paper (1993) considers the OLS estimation, which is equivalent to the ML method when the distribution of the sequence $\{\varepsilon_t\}$ is Gaussian ($\varepsilon_t \sim iidN(0, \sigma^2)$). Qian (1998) shows the asymptotic properties of the SETAR model parameters MLE when the $\{\varepsilon_t\} \sim iidD$, where D is a distribution with mean zero and finite variance. However, Qian (1998) does not consider the estimation of the parameters of distribution D , while Chan (1993) does estimate and study the properties of the estimator for σ .

It is convenient to highlight that the TAR model considered in this discussion lacks the concept of continuity; we cannot speak of discontinuity in the autoregressive function in the sense of the previous Definition because Y_{t-1} may take the γ value without implying a change of regime. The change of regime occurs when $Z_{t-d} = \gamma$ and if we wished for the change of regime to be continuous we would need to ask that $\phi_1 y_{t-1} = \phi_2 y_{t-1}$ when $Z_{t-d} = \gamma$, which in the continuous case has a zero probability.

In TAR models it is impossible to use this property because their continuity or discontinuity cannot be determined.

2.2.1 Strong consistency of maximum likelihood estimators

A demonstration for the general case where the error sequence $\{\varepsilon_t\}$ has an arbitrary distribution with mean zero and finite fourth moment will be made.

Let be the TAR(2; p, p) model, with initial conditions $Y_1, \dots, Y_p, Z_0, \dots, Z_d$, given in the following equation

$$Y_t = \begin{cases} \phi_{01} + \sum_{i=1}^p \phi_{i1} Y_{t-i} + \varepsilon_t, & Z_{t-d} \leq \gamma \\ \phi_{02} + \sum_{i=1}^p \phi_{i2} Y_{t-i} + \varepsilon_t, & Z_{t-d} > \gamma \end{cases},$$

or,

$$Y_t = h(\mathbf{Y}_{t-1}, Z_{t-d}, \phi) + \varepsilon_t, \quad t \geq 1 \quad (6)$$

for some $\phi = (\phi'_1, \phi'_2, \gamma, d) \in \mathfrak{R}^{2p+3} \times \{1, 2, \dots, p\}$, where $\mathbf{Y}_{t-1} = (Y_{t-1}, Y_{t-2}, \dots, Y_{t-p})'$, $\phi_j = (\phi_{0j}, \phi_{1j}, \dots, \phi_{pj})' \in \mathfrak{R}^{p+1}$, $j = 1, 2$ and for $\mathbf{y} \in \mathfrak{R}^p$ and $\mathbf{z} \in \mathfrak{R}^d$,

$$h(\mathbf{y}, \mathbf{z}, \phi) = \left(\phi_{01} + \sum_{k=1}^p \phi_{k1} y_k \right) I(z_d \leq \gamma) + \left(\phi_{02} + \sum_{k=1}^p \phi_{k2} y_k \right) I(z_d > \gamma).$$

The errors $\{\varepsilon_t\}$ in (6) are *iidN* $(0, \sigma^2)$.

Theorem 2 (Consistency of ML estimators). *Assume that $\{Y_t, Z_t\}$ of the model (6) is stationary and ergodic and that the SC1-SC4 conditions in Technical Appendix are met. Then $\hat{\boldsymbol{\theta}}_T$ is a strongly consistent estimator for $\boldsymbol{\theta}$.*

The Theorem proof is given in the Technical Appendix.

3 Predictions in TAR models through predictive likelihood

The problem of predicting unobserved values or future values from random variables is fundamental in Statistics. One form of solving it is to use the concept of predictive likelihood. From a Bayesian perspective, the problem of predicting unobserved values is directly solved by finding the posterior predictive density for the unobserved variables given the data. Chen and Lee (1995), Chen (1998) and Sáfadi and Morettin (2000) study the problem of estimating and obtaining predictions in SETAR models under this scope. However, we were unable to find references dealing with the problem of prediction in TAR models through the predictive likelihood approach studied here. Geisser (2002) presents a general overview of predictive inference considering both Bayesian and non Bayesian approaches.

The predictive likelihood proposal to order the plausibility of future values was developed by Lauritzen (1974), Hinkley (1979), and generalized by Butler (1986), although the initial idea was introduced by Fisher (1956) for the binomial model.

Let $\mathbf{X} = \mathbf{x}$ be the observed value of the \mathbf{X} random variable, the problem is to predict unobserved values \mathbf{x}^* from \mathbf{X}^* , and the inference is performed on the \mathbf{x}^* values. Assume that $(\mathbf{X}, \mathbf{X}^*)'$ has a probability density with regard to the Lebesgue mean, denoted by $f_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{x}^*)$, or equivalently denoted by $f(\mathbf{x}, \mathbf{x}^*; \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is the unknown parameters vector. Likewise, we denote the conditional density function as $f(\mathbf{x}^* | \mathbf{x}; \boldsymbol{\theta})$. Let $\hat{\boldsymbol{\theta}}$ be the

ML estimator for $\boldsymbol{\theta}$ based on the \mathbf{x} data, and $\hat{\boldsymbol{\theta}}_{\mathbf{w}}$ the ML estimator based on $\mathbf{w} = (\mathbf{x}, \mathbf{x}^*)'$. Let us suppose that $\mathbf{X} = (Y_1, \dots, Y_T)'$ contains the observed sample of size T and $\mathbf{X}^* = (Y_{T+1}, \dots, Y_{T+h})'$ the unobserved sample for which we desire to make inferences.

The basic prediction problem is the following: two unknown quantities, \mathbf{x}^* and $\boldsymbol{\theta}$, are considered, and the primary interest is to obtain information about \mathbf{x}^* with $\boldsymbol{\theta}$ playing the role of nuisance parameter. The principle of likelihood for prediction, formulated by Berger and Wolpert (1984), stems from the fact that all evidence about $(\mathbf{x}^*, \boldsymbol{\theta})$ is contained in the joint likelihood function,

$$L(\mathbf{x}^*, \boldsymbol{\theta}; \mathbf{x}) = f(\mathbf{x}, \mathbf{x}^*; \boldsymbol{\theta}). \quad (7)$$

We start from the base likelihood for the observed \mathbf{x} , and strive to develop a likelihood function for \mathbf{x}^* , say $L(\mathbf{x}^*|\mathbf{x})$, eliminating $\boldsymbol{\theta}$ from the equation (7). $L(\mathbf{x}^*|\mathbf{x})$, with the previously described characteristics, is known as the predictive likelihood function. Due to the fact that there are different ways to eliminate the nuisance parameter $\boldsymbol{\theta}$, several proposals for predictive likelihood arise. Bjørnstad (1990) counts fourteen proposals for predictive likelihood; however, some of them are completely similar, and all of them are based on one of the following three operations on $L(\mathbf{x}^*, \boldsymbol{\theta}; \mathbf{x})$: integration, maximization, or conditioning. The Bayesian approach is equivalent to integrating $L(\mathbf{x}^*, \boldsymbol{\theta}; \mathbf{x})$ with respect to the a priori distribution for $\boldsymbol{\theta}$. That is, the posterior predictive density, $f_0(\mathbf{x}^*|\mathbf{x})$ may be thought of as the marginal integrated likelihood such that $f_0(\mathbf{x}^*|\mathbf{x}) \propto \int L(\mathbf{x}^*, \boldsymbol{\theta}; \mathbf{x}) d\boldsymbol{\theta}$.

In this paper we consider the *Profile Predictive Likelihood* of the future \mathbf{x}^* observations, introduced by Mathiasen (1979) and recommended by Bjørnstad (1990), which eliminates the nuisance parameter $\boldsymbol{\theta}$ through the maximization of the following likelihood function:

$$L_p(\mathbf{x}^*|\mathbf{x}) = \sup_{\boldsymbol{\theta}} f(\mathbf{x}, \mathbf{x}^*; \boldsymbol{\theta}) = L(\mathbf{x}^*, \hat{\boldsymbol{\theta}}_{\mathbf{w}}; \mathbf{x}). \quad (8)$$

intuitively, the motivation for the predictive likelihood function L_p is as follows: for \mathbf{x}^* being the vector of parameters of interest and $\boldsymbol{\theta}$ the vector of nuisance parameters, the most likely values for $\boldsymbol{\theta}$ are obtained given $(\mathbf{x}, \mathbf{x}^*)'$, resulting in the likelihood function given in equation (8). In parametric inference this likelihood corresponds to the profile likelihood, introduced by Kalbfleisch and Sprott (1970) and hence the name of profile predictive likelihood.

Russell-Noriega (2006) illustrates the implementation of the predictive likelihood for the case of the linear AR(1) model. The fact that predictions for the AR(1) case through predictive likelihood and the recursive equations in terms of the autoregressive coefficient are equal, is highlighted by Clements and Hendry (1998).

4 Predictive Likelihood for the TAR Model

The implementation of the predictive likelihood for the TAR(k;p,p) model is direct. However, in order to clarify it, the equations for the TAR(2; 1, 1) model are presented in equation (3). Given a size T observed sample, the following equation is satisfied:

$$Y_t^* = \delta_t^* Y_{t-1}^* + \varepsilon_t, \quad t = T+1, \dots, T+h,$$

with $\varepsilon_t \sim iidN(0, \sigma^2)$, Z_t^* a stationary AR(1) process with gaussian noise $(0, \sigma_u^2)$ and $\delta_t^* = \phi_1 I(Z_{t-1}^* \leq \gamma) + \phi_2 I(Z_{t-1}^* > \gamma)$. Let also be $\mathbf{y} = (y_1, \dots, y_T)'$, $\mathbf{y}^* = (y_{T+1}^*, \dots, y_{T+h}^*)'$, $\mathbf{z} = (z_0, \dots, z_T)'$, and $\mathbf{z}^* = (z_{T+1}^*, \dots, z_{T+h}^*)'$ such that $\mathbf{w} = (\mathbf{y}', \mathbf{y}^{*\prime})'$ and $\mathbf{v} = (\mathbf{z}', \mathbf{z}^{*\prime})'$, then the TAR(2; 1, 1) model for $t = 1, 2, \dots, T, T+1, \dots, T+h$ may be written as:

$$w_t = \delta_t w_{t-1} + \varepsilon_t, \quad (9)$$

for which the δ_t variable is given by: $\delta_t^* = \phi_1 I(v_{t-1} \leq \gamma) + \phi_2 I(v_{t-1} > \gamma)$. The likelihood function for $(\mathbf{y}^*, \mathbf{z}^*, \boldsymbol{\theta})$, with $\boldsymbol{\theta} = (\phi_1, \phi_2, \sigma, \gamma, \rho, \sigma_u)'$ depending on (\mathbf{y}, \mathbf{z}) , with T large enough, and equivalent to the likelihood (4), is:

$$L(\boldsymbol{\theta}; \mathbf{y}, \mathbf{y}^*, \mathbf{z}, \mathbf{z}^*) = \frac{1}{\sigma^{T+h-1}} \exp \left[-\frac{1}{2\sigma^2} \sum_{t=2}^{T+h} (w_t - \delta_t w_{t-1})^2 \right] \\ \times \frac{1}{\sigma_u^{T+h-1}} \exp \left[-\frac{1}{2\sigma_u^2} \sum_{t=2}^{T+h} (v_t - \rho v_{t-1})^2 \right].$$

Thus the log likelihood function for (\mathbf{w}, \mathbf{v}) is given by:

$$\ell(\boldsymbol{\theta}; \mathbf{y}, \mathbf{y}^*, \mathbf{z}, \mathbf{z}^*) = -(T+h-1) \log \sigma - \frac{1}{2\sigma^2} \sum_{t=2}^{T+h} (w_t - \delta_t w_{t-1})^2 \\ - (T+h-1) \log \sigma_u - \frac{1}{2\sigma_u^2} \sum_{t=2}^{T+h} (v_t - \rho v_{t-1})^2.$$

Equivalently to obtaining the ML estimators in the previous section, we have the following identities:

$$\hat{\phi}_1(\mathbf{y}^*; \gamma) = \frac{\sum_{t=2}^{T+h} I_t w_t w_{t-1}}{\sum_{t=2}^{T+h} I_t w_{t-1}^2}, \quad \hat{\phi}_2(\mathbf{y}^*; \gamma) = \frac{\sum_{t=2}^{T+h} (1-I_t) w_t w_{t-1}}{\sum_{t=2}^{T+h} (1-I_t) w_{t-1}^2},$$

with $I_t = I(v_t \leq \gamma)$. As well as:

$$\hat{\rho}(\mathbf{z}^*) = \frac{\sum_{t=2}^{T+h} v_t v_{t-1}}{\sum_{t=2}^{T+h} v_{t-1}^2} \text{ and } \hat{\sigma}_u^2(\mathbf{z}^*; \hat{\rho}) = \frac{1}{T+h-1} \sum_{t=2}^{T+h} (v_t - \hat{\rho} v_{t-1})^2.$$

And similarly, if $\hat{\delta}_t(\mathbf{y}^*; \gamma) = \hat{\phi}_1(\mathbf{y}^*; \gamma) + \hat{\phi}_2(\mathbf{y}^*; \gamma) I(Z_{t-1} > \gamma)$, then:

$$\hat{\sigma}^2(\mathbf{y}^*; \gamma) = \hat{\sigma}^2(\mathbf{y}^*; \hat{\phi}_1, \hat{\phi}_2, \gamma) \\ = \frac{1}{T+h-1} \sum_{t=2}^{T+h} (w_t - \hat{\delta}_t(\mathbf{y}^*; \gamma) w_{t-1})^2,$$

So maximizing the profile predictive log likelihood for $(\mathbf{y}^*, \mathbf{z}^*)$, $\ell_p(\mathbf{y}^*, \mathbf{z}^*; \gamma | \mathbf{y}, \mathbf{z})$ is equivalent to minimizing $-\ell_p(\mathbf{y}^*, \mathbf{z}^*; \gamma | \mathbf{y}, \mathbf{z})$. that is, to minimize

$$\frac{T+h-1}{2} \log \left[\frac{1}{T+h-1} \sum_{t=2}^{T+h} (w_t - \hat{\delta}_t(\mathbf{y}^*; \gamma) w_{t-1})^2 \right] \\ + \frac{T+h-1}{2} \log \left[\frac{1}{T+h-1} \sum_{t=2}^{T+h} (v_t - \hat{\rho}(\mathbf{y}^*) v_{t-1})^2 \right], \quad (10)$$

with respect to $y^*, z^*; \gamma$. If the threshold parameter γ is known, we must minimize the target function (10) with respect to (y^*, z^*) to obtain the corresponding predictions, say \hat{y}^* , and \hat{z}^* . The values \hat{y}^*, \hat{z}^* are the predictive likelihood predictions for y^* and z^* , respectively, for a fixed γ .

If the threshold variable $Z_{t-1} \sim iidU(0, 1)$ the predictive likelihood predictions for y^* are obtained under the assumption that the future values for the z^* threshold variable are known. The function to minimize with respect to y^* and γ is,

$$\frac{1}{T+h-1} \sum_{t=2}^{T+h} \left(w_t - \hat{\delta}_t(\mathbf{y}^*; \gamma) w_{t-1} \right)^2, \quad (11)$$

due to the fact that the logarithmic function is an increasing function.

The generalization for the TAR model of more than two regimes and an autoregressive order $p > 1$ for each regime is written directly, through the adequate redefinition of the δ_t variable.

5 Simulation study

A comparison of the performance of the h -step ahead forecasts in the TAR model versus the performance of the forecasts in the AR model is made. We compute the ratios of the MSFE from the AR model divided by the MSFE from the TAR model for each of the forecast horizons. If the ratio is higher than one, it follows that the MSFE in horizon h for the TAR model is lowest. The predictive likelihood approach is compared to the recursive Monte Carlo alternative. For the AR linear model both alternatives result in the same expressions for the computation of predictions (Russell-Noriega (2006)). Implementation of the recursive method in the TAR model implies knowledge of the future behavior of the threshold variable, or of the value of the indicator variable, that determine the change or regime. The simulation exercise consists of a large variety of important scenarios around the problem studied.

5.1 Description of the simulation algorithm

The simulation exercise begins with the generation of 500 time series of size 255, considering the model associated to the threshold variable, and then to simulate the 500 series from the TAR(2; 1, 1) model. The first 250 observations in each series are utilized for the parameter estimation stage, while the last five observations are used for the predictors comparison stage. The simulated scenarios are given with consideration to the values of the model parameters and the probabilities of dominance in each one of the regimes. The analyzed situations scope a large variety of possible situations, as will be seen in the Table 1. Consider the case in which the threshold variable Z_{t-1} is a stationary AR(1) time series, with Gaussian white noise. The case where $Z_{t-1} \sim iidU(0, 1)$ is discussed in Russell-Noriega (2006) and it is mainly useful at the time of manipulating and reducing the resulting theoretical expressions. The forecasts for $h = 1, 2, 3, 4$, and 5 steps ahead are obtained through methods identified as recursive method (RM), and predictive likelihood method (PL), under the assumption that the threshold parameter γ is known. The RM method is to approximate the conditional expectation given in the following equation:

$$\hat{Y}_{t+h} = E(Y_{t+h} | \mathfrak{F}_t) = \phi_1 E[I(Z_{t+h-1} \leq \gamma) Y_{t+h-1} | \mathfrak{F}_t] + \phi_2 E[I(Z_{t+h-1} > \gamma) Y_{t+h-1} | \mathfrak{F}_t]$$

through a recursive Monte Carlo procedure (Granger and Teräsvirta, (1993), Russell-Noriega (2006)). The first step is to predict the values of $Z_{T+1}, \dots, Z_{T+h-1}$ from the model associated to Z_{t-1} , or else to predict the values of $I(Z_{t-1} \leq \gamma)$ with fixed γ . The indicator variable is predicted following the ideas of Kedem and Fokianos (2002), as detailed in Russell-Noriega (2006).

Given a set of T observations we have that the $T+1$ forecast, conditional on the information up to time T , for the TAR(2; 1, 1) model, is easily arrived at since the regime is known with certainty and the one-step ahead forecast by RM is simply:

$$\hat{Y}_{T+1}^{RM} = [\hat{\phi}_1 I(Z_T \leq \gamma) + \hat{\phi}_2 I(Z_T > \gamma)] Y_T$$

with $\hat{\phi}_1$ and $\hat{\phi}_2$, the estimator of the autoregressive coefficients in the TAR(2; 1, 1) model, obtained through likelihood equations under the assumption of normality in the errors. At time $T+2$, the regime prediction is determined by the value of Z_{T+1} , so we substitute the corresponding prediction, generated from the model that rules the threshold variable dynamics. We have the same result if substitute the predicted value of the indicator $I(Z_{T+1} \leq \gamma)$. Then, the prediction for $Y_{T+2}, Y_{T+3}, \dots, Y_{T+h}$ for one iteration of the error process $\zeta_{i,j} \sim N(0, \hat{\sigma}^2)$, where $\hat{\sigma}^2$ is the estimator of the error variance under the consideration of the TAR(2; 1, 1) model, is defined as:

$$\hat{Y}_{T+h}^{PRZt_j} = [\hat{\phi}_1 I(\tilde{Z}_{T+h-1} \leq \gamma) + \hat{\phi}_2 I(\tilde{Z}_{T+h-1} > \gamma)] \hat{Y}_{T+h-1}^{PRZt_j} + \zeta_{h,j} \quad (12)$$

or

$$\hat{Y}_{T+h}^{PRIZt_j} = [\hat{\phi}_1 \tilde{I}(Z_{T+h-1} \leq \gamma) + \hat{\phi}_2 \tilde{I}(Z_{T+h-1} > \gamma)] \hat{Y}_{T+h-1}^{PRIZt_j} + \zeta_{h,j}, \quad (13)$$

where $PRZt_j$ refers to the recursive method using the prediction for the Z_t variable, while $PRIZt_j$ indicates that the recursive method uses the prediction of the indicator variable $I(Z_{T+h-1} \leq \gamma)$. Each of the h -step ahead forecasts is performed j times iteratively, in such a way that the final prediction is generated by averaging the values for each of the h periods over the $j = 1, \dots, N$ iterations. We denote \hat{Y}_{T+h}^{PRZt} or \hat{Y}_{T+h}^{PRIZt} as the h -step ahead forecasts for Y_t by means the recursive method depending on whether we use the prediction of the threshold variable or that of the indicator variable and given by the sample average:

$$\hat{Y}_{T+h}^{PRZt} = \frac{1}{N} \sum_{j=1}^N \hat{Y}_{T+h}^{PRZt_j}, \quad \hat{Y}_{T+h}^{PRIZt} = \frac{1}{N} \sum_{j=1}^N \hat{Y}_{T+h}^{PRIZt_j}, \quad h > 1.$$

This alternative to generate forecasts, as well as its variations, have been thoroughly studied for the SETAR models by De Gooijer and De Bruin (1998), and Clements and Smith (1997,1999). Inherent to the process of approximation of the predictions, we gather in each iteration the corresponding forecast errors and the MSFE for each of the 500 series simulated. To complete the forecasts for the TAR(2; 1, 1) model under this method, we generate the h -step ahead forecasts under the AR(1) model, by means of the corresponding recursive equation, and obtain the valued for the MSFE loss functions as well as those for the TAR model.

the predictive likelihood predictors for the TAR(2; 1, 1) model are obtained by means of this procedure: from the likelihood principles, the predictors for the h unobserved values are those whose values minimize the negative of the profile log likelihood function, given in equation (10). The optimization process for each of the simulated series is performed, jointly predicting the Y_t and Z_{t-1} variables by PL, i.e., the predictive likelihood function is optimized as a function of the values Y_{T+1}, \dots, Y_{T+h} and $Z_T, Z_{T+1}, \dots, Z_{T+h-1}$. Later, the prediction errors are computed considering these predictions and the last five values

simulated for Y_t . The MSFE reported is the average of the 500 MSFE obtained from each one of the simulated series and each one of the prediction horizons. The same procedure is followed for the alternative AR(1) method.

6 Presentation and analysis of the results

Table 1 presents the parameter values for each of the simulated scenarios, as well as the averages for the 500 parameter values estimated by maximum likelihood under the TAR(2;1,1) model and the AR(1) model, respectively. The regime dominance probabilities considered are : $p = \Pr(Z_{t-1} \leq \gamma) = 0.5$ and 0.75 , with variance $\sigma^2 = 1$.

Table 1. Values of the parameters used in the simulation exercise and averages of their estimated values under the TAR an AR models, respectively, by ML

| Case | ϕ_1 | ϕ_2 | p | $\hat{\phi}_1$ | $\hat{\phi}_2$ | $\hat{\sigma}_{TAR}^2$ | $\hat{\phi}_{AR}$ | $\hat{\sigma}_{AR}^2$ |
|------|----------|----------|------|----------------|----------------|------------------------|-------------------|-----------------------|
| 1-1 | -0.1 | 1 | 0.50 | -0.096 | 0.99 | 0.994 | 0.54 | 1.71 |
| 1-2 | -0.1 | 1 | 0.75 | -0.099 | 0.989 | 0.984 | 0.239 | 1.366 |
| 1-3 | 1 | -0.1 | 0.75 | 0.989 | -0.096 | 0.995 | 0.712 | 1.888 |
| 2-1 | 1 | 0.1 | 0.50 | 0.988 | 0.101 | 0.988 | 0.631 | 1.467 |
| 2-2 | 1 | 0.1 | 0.75 | 0.992 | 0.104 | 0.991 | 0.816 | 1.632 |
| 2-3 | 0.1 | 1 | 0.75 | 0.096 | 0.983 | 0.990 | 0.380 | 1.253 |
| 3-1 | -0.8 | -0.6 | 0.50 | -0.786 | -0.594 | 0.996 | -0.697 | 1.019 |
| 4-1 | -0.8 | 0.6 | 0.50 | -0.791 | 0.597 | 0.994 | -0.138 | 1.973 |
| 4-2 | -0.8 | 0.6 | 0.75 | -0.79 | 0.596 | 0.991 | -0.471 | 1.792 |
| 4-3 | 0.6 | -0.8 | 0.75 | 0.595 | -0.792 | 0.989 | 0.227 | 1.658 |
| 5-1 | -0.4 | -0.6 | 0.50 | -0.400 | -0.591 | 0.990 | -0.499 | 1.006 |
| 6-1 | -0.4 | 0.6 | 0.50 | -0.399 | 0.591 | 0.995 | 0.114 | 1.333 |
| 6-2 | -0.4 | 0.6 | 0.75 | -0.396 | 0.581 | 0.993 | -0.133 | 1.236 |
| 6-3 | 0.6 | -0.4 | 0.75 | 0.589 | -0.400 | 0.993 | 0.352 | 1.255 |
| 7-1 | -0.4 | 0.4 | 0.50 | -0.403 | 0.396 | 0.986 | -0.002 | 1.181 |
| 7-2 | -0.4 | 0.4 | 0.75 | -0.398 | 0.392 | 0.988 | -0.202 | 1.130 |
| 7-3 | 0.4 | -0.4 | 0.75 | 0.394 | -0.381 | 0.988 | 0.201 | 1.125 |
| 8-1 | -0.95 | 0.95 | 0.50 | -0.941 | 0.943 | 0.990 | -0.0006 | 9.664 |
| 8-2 | -0.95 | 0.95 | 0.75 | -0.941 | 0.941 | 0.989 | -0.468 | 7.528 |
| 8-3 | 0.95 | -0.95 | 0.75 | 0.941 | -0.942 | 0.988 | 0.006 | 9.762 |
| 9-1 | -0.9 | 1 | 0.50 | -0.893 | 0.992 | 0.990 | 0.096 | 10.576 |
| 9-2 | -0.9 | 1 | 0.75 | -0.894 | 0.988 | 0.994 | -0.3871 | 5.981 |
| 9-3 | 1 | -0.9 | 0.75 | 0.993 | -0.894 | 0.992 | 0.541 | 14.43 |
| 10-1 | 0.9 | 1 | 0.50 | 0.893 | 0.993 | 0.997 | 0.945 | 1.027 |
| 11-1 | 0.9 | 0.7 | 0.50 | 0.895 | 0.692 | 0.989 | 0.802 | 1.023 |

Russell-Noriega (2006) discusses the importance of the cases in Table 1, considering a higher value of the error variance. All cases analyzed in Table 1 satisfy the second order stationarity condition ($E(\delta_t) < 1$); however, several of them are close to the non-stationarity condition. The discussion of the results is illustrated by two interesting cases. The first one, the 1-1 case, has a unitary root in regime 2, under a condition of stationarity $E(\delta_t^2) = 0.505$, which is much lower than one. This model is identified in literature like Threshold Unit Root model (TUR model). The second case is the 8-1, where $E(\delta_t^2) = 0.903$ is close to one, with

autoregressive coefficients in the regimes equal in magnitude but with opposite sign. Both examples have the same dominance in each regime. The results of the analysis extend to the rest of the cases listed in Table 1.

The analysis summary concentrates on Figures 1, 2 and 3. Figures 1 and 2 show the variability of the 500 prediction errors for the TAR and AR models under the alternatives of the recursive and the predictive likelihood method. Notice that the error variability under predictive likelihood for the TAR model around zero is lower in both cases. The MSFE under the recursive method will be affected by the values of errors that exceed the interquartile interval.

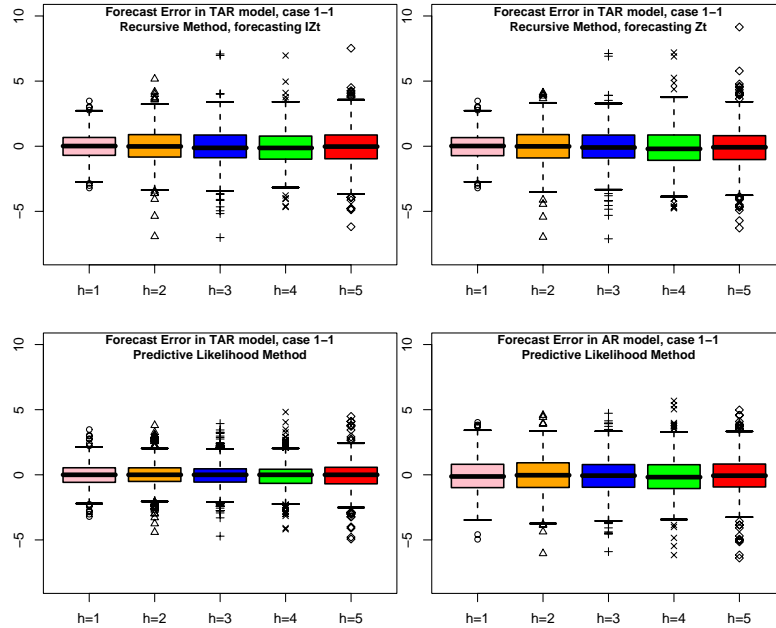


Figure 1. Box-plot for the forecast error, case 1-1.

Figure 3 shows the averages for the MSFE under each prediction horizon, for each method. It highlights the behavior of the MSFE for the TAR model under predictive likelihood, making it clear that their performance is superior to that of the forecasts under the AR model. The same result is obtained when comparing the TAR model forecasts under the recursive method.

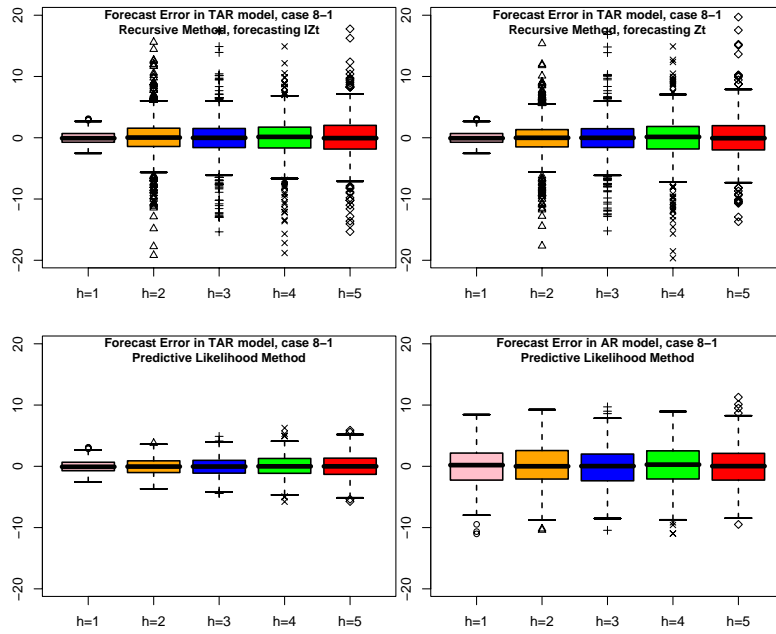


Figure 2. Box-plot for the forecast error, case 8-1.

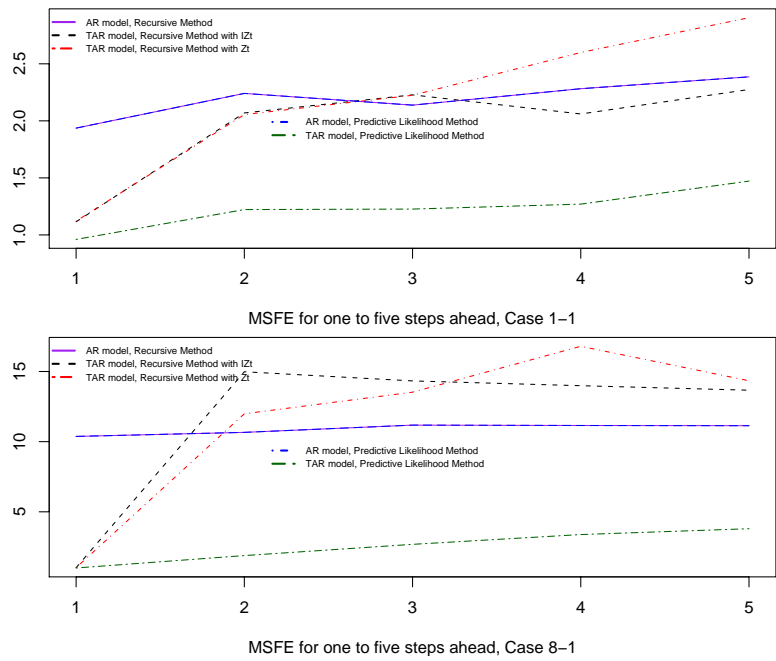


Figure 3. Mean squares forecast error

7 Conclusions

One characteristic of TAR models is that under certain scenarios (as a function of the magnitudes and signs of the autoregressive parameters), the linear model provides a good approximation to the nonlinear model, hence in view of the simplicity of the prediction-generating linear model this last one is often and wrongly considered as the best one. However, Russell-Noriega (2006) showed that under an incorrectly specified model, the process of associated errors is ruled by a volatility model. The observed variability through time in that process is directly identified as a function of the first two moments of the variable δ_t , because they determine the stationarity condition of the TAR model. This variability is directly inherited by the error variance under the incorrect specification, i.e., the farther we are from the non stationarity condition the better the linear model approximation will become. Even in those cases, the predictive likelihood method shows a good performance against its linear rival, AR.

The increased variability identified under the misspecified model, as well as the distributional lack of fit of the respective residuals, is overtaken upon consideration of the different methodologies for estimating and obtaining predictions that have been discussed in most of the literature. Predictive likelihood approaches the problem of estimating and obtaining the parameters under the same statistical methodology and, because of this, at the time of evaluating the predictions performance under the actual DGP, they show better performances than those under an incorrect model.

Under the normality assumption, OLS and ML estimators are the same; however, the prediction methods in TAR models associated to those estimation methods are not the same, while in the linear case they actually are. Predictive likelihood jointly considers all the information, using the probabilistic properties of the threshold variable, resulting in a more efficient optimization process. The recursive method assumes a known threshold variable, and the process for generating the predictions for this variable is performed as previously to the generation of the predictions in the TAR model. Russell-Noriega (2006) comments, as part of the results of the simulation exercise, that the TAR predictions through the recursive method are more efficient if the indicator variable predictions are used instead of the threshold variable ones. This is due to the misclassification probabilities in the regimes under the predictions of the threshold variable.

In order to solve the problem that originated this paper it was necessary to undertake a detailed study of the TAR models parameter estimation process, showing the consistency and asymptotic distribution properties of the MLE, with a known γ . For an unknown γ only the consistency of the MLE was shown. The particular probabilistic structure of the TAR model must be determined before applying the results from the SETAR case to the TAR case. The demonstrations in this paper are contributions to the non linear models literature.

8 Future work

In order to study the robustness of predictive likelihood it is necessary to exhaustively study the effect of non-normal distributions over the errors.

The proofs for asymptotic distributions in OLS and ML estimators in SETAR models rely heavily on the continuity or discontinuity concept for those models. The MLE asymptotic

distributions with unknown γ were not studied in this paper, since the concept of continuity does not have a clear extension to the TAR case.

It is also tempting to extend the predictive likelihood methodology to the SETAR models, an effort not yet undertaken in the literature and which is quite direct.

One of our results, that of the residuals following a volatility process under an incorrectly specified model, allows us to set as a future task that of finding and comparing asymmetric function methods to quantify the performances, as well as to explore new methods, robust for departures from normality.

Technical appendix: Proof of the MLE consistency property

Notation

Function f is the ε_2 density function and function F is the distribution function for f .

The expectation under $\boldsymbol{\vartheta}$ is denoted by E .

Let \mathfrak{R} be the real line $(-\infty, \infty)$ and $\bar{\mathfrak{R}} = \mathfrak{R} \cup \{-\infty, \infty\}$, then $\bar{\mathfrak{R}}$ is compact under the $d(\cdot, \cdot)$ metrics, defined by $d(x, y) = |\arctan x - \arctan y|$.

A function φ is Lip(1) if $\forall x, y \in \mathfrak{R}$, there exists an $L \geq 0$, such that $|\varphi(x) - \varphi(y)| \leq L|x - y|$.

The complement of A is denoted by A^c .

The inner product of \mathbf{x} and \mathbf{y} is denoted by $\mathbf{x}'\mathbf{y}$.

The subscript $t = 2, \dots, T$ will be omitted unless otherwise specified.

Preliminaries to the maximum likelihood estimation method

Let us assume that ε_2 has a known distribution and that σ is involved in the model as a scale modification, i.e., the TAR(2;p,p) model may be written as:

$$Y_t = h(\mathbf{Y}_{t-1}, \mathbf{Z}_{t-d}, \boldsymbol{\phi}, \gamma) + \sigma\varepsilon_t, \quad t \geq 2, \quad (14)$$

for some $\boldsymbol{\vartheta} = (\boldsymbol{\phi}'_1, \boldsymbol{\phi}'_2, \gamma) \in \mathfrak{R}^{2p+3}$, $d \in \{1, 2, \dots, p\}$, where $\mathbf{Y}_{t-1} = (Y_{t-1}, \dots, Y_{t-p})'$, $\mathbf{Z}_{t-d} = (Z_{t-d}, \dots, Z_{t-d-p+1})'$, $\boldsymbol{\phi}_j = (\alpha_{0j}, \alpha_{1j}, \dots, \alpha_{pj})' \in \mathfrak{R}^{p+1}$, $j = 1, 2$ and for $\mathbf{y} \in \mathfrak{R}^p$ and $\mathbf{z} \in \mathfrak{R}^p$,

$$h(\mathbf{y}, \mathbf{z}, \boldsymbol{\vartheta}) = \left(\phi_{01} + \sum_{k=1}^p \phi_{k1} y_k \right) I(z_d \leq \gamma) + \left(\phi_{02} + \sum_{k=1}^p \phi_{k2} y_k \right) I(z_d > \gamma).$$

The errors $\{\varepsilon_t\}$ in (14) are $iidN(0, 1)$ and $Z_t = \rho Z_{t-1} + u_t$, where $\{u_t\}$ are $iidN(0, \sigma_u^2)$.

The proof of the asymptotic properties of the MLE for $\boldsymbol{\phi}, \gamma, d, \sigma, \rho, \sigma_u$ will be done assuming that $\sigma = 1$, for convenience in the expressions we proof Lemma 1 and Theorem 2. In Lemma 2 the general case $\sigma > 0$ will be proved.

Because Z_t has its own dynamics, estimators for ρ and σ_u satisfy the properties of AR(1) models: they are strongly consistent and asymptotically normal (see Theorem 8.2.1. in Fuller (1996)).

We will assume that $\boldsymbol{\vartheta} = (\boldsymbol{\phi}', \gamma, d)'$ is an inner point in the parametric space $\mathfrak{R}^{2p+2} \times \bar{\mathfrak{R}} \times \{1, 2, \dots, p\}$.

Since $\boldsymbol{\vartheta}$ is an inner point, it follows that there exists a compact set $K \subset \mathfrak{R}^{2p+2}$ such that $\boldsymbol{\vartheta}$ is an inner point of $\boldsymbol{\Lambda} = K \times \mathfrak{K} \times \{1, 2, \dots, p\}$, $\boldsymbol{\Lambda}$ is a compact set.

Let $\boldsymbol{v} = (\boldsymbol{\beta}', r, q)'$ be an inner point of $\boldsymbol{\Lambda}$. Let $(\mathbf{Y}'_t, \mathbf{Z}'_t)' = (Y_t, \dots, Y_{t-p+1}, Z_t, \dots, Z_{t-p+1})'$, $\{\mathbf{Z}_t\}$ and $\{\mathbf{Y}_t, \mathbf{Z}_t\}$ be Markov chains. Let $g_{\boldsymbol{\vartheta}}(\mathbf{Y}_1, \mathbf{Z}_{2-d})$ be the initial density of \mathbf{Y}_1 and \mathbf{Z}_{2-d} under $\boldsymbol{\vartheta}$. The one-step transitional function, initiating at $(\mathbf{Y}'_1, \mathbf{Z}'_{2-d})'$ is

$$f(Y_t - h(\mathbf{Y}_{t-1}, \mathbf{Z}_{t-d}, \boldsymbol{\vartheta})) f_1(Z_{t-d} - h_1(\mathbf{Z}_{t-d-1}, \rho)), t \geq 2.$$

The one-step transitional function, initiating at \mathbf{Z}'_{2-d} is $f_1(Z_t - h_1(\mathbf{Z}_{t-1}, \rho)) w(Z_t | Z_{t-1})$, $t \geq 2$.

Let f_1 be the density function of u_t , $g_1(\mathbf{Z}_{2-d})$ is the initial density of \mathbf{Z}_{2-d} under $(\rho, \sigma_u^2)'$, If we observe $(\mathbf{Y}'_1, Y_2, \dots, Y_T, \mathbf{Z}_{2-d}, Z_{2-d+1}, \dots, Z_{T-d+1})'$, it follows the the likelihood function under $\boldsymbol{\vartheta}$ is

$$\prod_{t=2}^T f(Y_t - h(\mathbf{Y}_{t-1}, \mathbf{Z}_{t-d}, \boldsymbol{\vartheta})) f_1(Z_{t-d} - h_1(\mathbf{Z}_{t-d-1}, \rho)) g_{\boldsymbol{\vartheta}}(\mathbf{Y}_1, \mathbf{Z}_{2-d}).$$

Estimation of $\boldsymbol{\vartheta}$ is not affected by the estimation of ρ , so we will consider the conditional likelihood function of $\boldsymbol{\vartheta}$ given by

$$L_n(\boldsymbol{\vartheta}) = \prod_{t=d}^T f(Y_t - h(\mathbf{Y}_{t-1}, \mathbf{Z}_{t-d}, \boldsymbol{\vartheta})). \quad (15)$$

Let $\hat{\boldsymbol{\vartheta}}_T = (\hat{\boldsymbol{\alpha}}', \hat{\gamma}, \hat{d})'$ be any measurable function of $(\mathbf{Y}_1, Y_2, \dots, Y_T, \mathbf{Z}_{2-d}, Z_{2-d+1}, \dots, Z_{T-d+1})'$ from \mathfrak{R}^{T+2p} to $\boldsymbol{\Lambda}$ such that $\hat{\boldsymbol{\vartheta}}_T$ maximizes over $\boldsymbol{\Lambda}$, the conditional likelihood function given in (15).

Assumptions

AC1. $f(y) > 0$, for all $y \in \mathfrak{R}$ and is absolutely continuous. f' exists almost everywhere, $\varphi = f'/f$ and $I(f) = \int_{-\infty}^{\infty} \varphi^2(y) f(y) dy < \infty$.

AC2. φ is Lip(1).

AC3. φ is and the derivative of φ' is Lip(1).

AC4. $E|\varepsilon_2|^4 < \infty$ and $E|u_2|^4 < \infty$.

We will use the fact that $E|\varepsilon_2|^k < \infty$, for $k = 1, 2, 3, 4$, implies that $E|\mathbf{Y}_{d-1}|^k < \infty$.

Let us prove the strong consistency of the MLE for $\hat{\boldsymbol{\vartheta}}_T$. Let ℓ_T be the conditional log likelihood ratio:

$$\ell_T(\boldsymbol{v}) = \ell_T(\boldsymbol{\beta}, r, q) = \frac{1}{T} \sum \ln \frac{f(Y_t - h(\mathbf{Y}_{t-1}, \mathbf{Z}_{t-q}, \boldsymbol{\beta}, r))}{f(Y_t - h(\mathbf{Y}_{t-1}, \mathbf{Z}_{t-d}, \boldsymbol{\phi}, \boldsymbol{\gamma}))}, \quad \boldsymbol{v} \in \boldsymbol{\Lambda} \quad (16)$$

The invariant equation

$$g_{\boldsymbol{\vartheta}}(\mathbf{y}, \mathbf{z}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(y - h(u, v, \boldsymbol{\vartheta})) f_1(z - h_1(u, \rho)) g_{\boldsymbol{\vartheta}}(u, v) dudv$$

and assumption AC1 imply that g_θ is bound outside of $[0, \infty]^2$ over compact sets. $I(f) < \infty$ implies that f is bounded.

$$\psi(Y_{t-1}, Z_{t-q}, \varepsilon_t, \beta, r) = \ln \frac{f(\varepsilon_t + h(\mathbf{Y}_{t-1}, \mathbf{Z}_{t-d}, \phi, \gamma)) - h(\mathbf{Y}_{t-1}, \mathbf{Z}_{t-q}, \beta, r)}{f(\varepsilon_t)}, \quad 1 \leq t \leq n.$$

Notice that

$$\ell_T(\mathbf{v}) = \frac{1}{T} \sum \psi(\mathbf{Y}_{t-1}, \mathbf{Z}_{t-q}, \varepsilon_t, \beta, r), \quad \mathbf{v} \in \Lambda.$$

let $\mathbf{W} = (1, \mathbf{y}')'$ and

$$\begin{aligned} \nabla_\phi h(\mathbf{y}, \mathbf{z}_q; \beta, s) &= \frac{\partial}{\partial \beta} h(\mathbf{y}, \mathbf{z}_q; \beta, r) \\ &= (\mathbf{W}' I(z_d \leq r), \mathbf{W}' I(z_d > r))'. \end{aligned}$$

Hence

$$|\nabla_\phi h(\mathbf{y}, \mathbf{z}_q; \beta, r)| = \sqrt{1 + |\mathbf{y}|^2}. \quad (17)$$

We have

$$|h(\mathbf{y}, \mathbf{z}_q; \beta, r)| \leq |\beta| \sqrt{1 + |\mathbf{y}|^2}, \quad (18)$$

since

$$h(\mathbf{y}, \mathbf{z}_q; \beta, s) = \beta' \nabla_\beta h(\mathbf{y}, \mathbf{z}_q; \beta, s). \quad (19)$$

Additionally, for all $s \in \bar{\mathfrak{R}}, r \in \bar{\mathfrak{R}}, q \in \{1, \dots, p\}$

$$\begin{aligned} |\nabla_\phi h(\mathbf{y}, \mathbf{z}_q; \beta, r) - \nabla_\phi h(\mathbf{y}, \mathbf{z}_q; \beta, s)| &\leq \sqrt{2(1 + |\mathbf{y}|^2)} I(\min(s, r) < z_q < \max(s, r)) \\ &\leq \sqrt{2(1 + |\mathbf{y}|^2)} I(|z_q - s| \leq |r - s|). \end{aligned} \quad (20)$$

Lemma 1. *Under the assumptions of Theorem 2, for all $\mathbf{v} = (\beta', r, q)' \in \Lambda$ with $\Lambda = K \times \bar{\mathfrak{R}} \times \{1, 2, \dots, p\}$ and an open vicinity $U_{\mathbf{v}}$,*

$$E \left[\sup_{\mathbf{v}^* \in U_{\mathbf{v}}} |\psi(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \varepsilon_2, \beta^*, r^*) - \psi(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \varepsilon_2, \beta, r)| \right] \longrightarrow 0, \quad (21)$$

when $U_{\mathbf{v}}$ contracts to \mathbf{v} .

Proof. Let

$$U_{\mathbf{v}}(\eta) = \left\{ \mathbf{v}^* = (\beta^{*'}, r^*, q)' \in \Lambda : |\beta^* - \beta| < \eta, d(r^*, r) < \eta \right\}, \quad \eta > 0.$$

Without loss of generality, we will prove the lemma for

$$E \sup_{\mathbf{v}^* \in U_{\mathbf{v}}(\eta)} |\psi(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \varepsilon_2, \mathbf{v}^*) - \psi(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \varepsilon_2, \beta, r)| \longrightarrow 0 \text{ when } \eta \rightarrow 0.$$

Let $\varepsilon_2(\mathbf{v}) = Y_2 - h(Y_1, \mathbf{Z}_{2-q}, \boldsymbol{\beta}, r)$ and $\delta(\mathbf{y}, \mathbf{z}_q, \boldsymbol{\beta}^*, r^*) = h(\mathbf{y}, \mathbf{z}_q, \boldsymbol{\beta}, r) - h(\mathbf{y}, \mathbf{z}_q, \boldsymbol{\beta}^*, r^*)$. For all $\mathbf{v} = (\boldsymbol{\beta}', r, q)'$ and for all $\mathbf{y} \in \mathfrak{R}^p$, by (17) and (19) we conclude that

$$|h(\mathbf{y}, \mathbf{z}_q, \boldsymbol{\beta}, r) - h(\mathbf{y}, \mathbf{z}_q, \boldsymbol{\beta}^*, r^*)| \leq |\boldsymbol{\beta} - \boldsymbol{\beta}^*| \sqrt{1 + |\mathbf{y}|^2}, \quad (22)$$

and from equations (18) and (20)

$$\begin{aligned} |h(\mathbf{y}, \mathbf{z}_q, \boldsymbol{\beta}^*, r) - h(\mathbf{y}, \mathbf{z}_q, \boldsymbol{\beta}^*, r^*)| &\leq |\boldsymbol{\beta}^*| \sqrt{2(1 + |\mathbf{y}|^2)} I(\min(r, r^*) < z_q < \max(r, r^*)) \\ &\leq |\boldsymbol{\beta}^*| \sqrt{2(1 + |\mathbf{y}|^2)} I(|z_q - r| \leq |r^* - r|). \end{aligned}$$

Then for $\mathbf{v}^* \in U_{\mathbf{v}}(\eta)$ and for $r \in \mathfrak{R}$, $\mathbf{y} \in \mathfrak{R}^p$,

$$\begin{aligned} |\delta(\mathbf{y}, \mathbf{z}_q, \mathbf{v}^*)| &= |h(\mathbf{y}, \mathbf{z}_q, \boldsymbol{\beta}, r) - h(\mathbf{y}, \mathbf{z}_q, \boldsymbol{\beta}^*, r^*)| \\ &= |h(\mathbf{y}, \mathbf{z}_q, \boldsymbol{\beta}, r) - h(\mathbf{y}, \mathbf{z}_q; \boldsymbol{\beta}, r^*) + h(\mathbf{y}, \mathbf{z}_q; \boldsymbol{\beta}, r^*) - h(\mathbf{y}, \mathbf{z}_q, \boldsymbol{\beta}^*, r^*)| \\ &\leq |h(\mathbf{y}, \mathbf{z}_q; \boldsymbol{\beta}, r) - h(\mathbf{y}, \mathbf{z}_q; \boldsymbol{\beta}, r^*)| + |h(\mathbf{y}, \mathbf{z}_q; \boldsymbol{\beta}, r^*) - h(\mathbf{y}, \mathbf{z}_q; \boldsymbol{\beta}^*, r^*)| \\ &\leq \left[\sqrt{2} |\boldsymbol{\beta}| I(|z_q - r| \leq |r^* - r|) + |\boldsymbol{\beta} - \boldsymbol{\beta}^*| \right] \sqrt{1 + |\mathbf{y}|^2} \\ &\leq \left[\sqrt{2} |\boldsymbol{\beta}| I(|z_q - r| \leq |r_0(\eta) - r|) + \eta \right] \sqrt{1 + |\mathbf{y}|^2}, \end{aligned} \quad (23)$$

where $r_0(\eta)$ is such that $d(r_0(\eta), r) = \eta$.

Let us define

$$\Delta(\mathbf{y}, \mathbf{z}_q, \eta) \equiv \left[\sqrt{2} |\boldsymbol{\beta}| I(|z_q - r| \leq |r_0(\eta) - r|) + \eta \right] \sqrt{1 + |\mathbf{y}|^2}. \quad (24)$$

Assumptions AC1 and AC2, $E|\mathbf{Y}_{d-1}| < \infty$ and equation (18) imply that there exists a constant L such that for any $\mathbf{v} \in \boldsymbol{\Lambda}$,

$$\begin{aligned} E\varphi^2(\varepsilon_2(\mathbf{v})) &= E\{[\varphi(\varepsilon_2 + h(\mathbf{Y}_1, \mathbf{Z}_{2-d}, \boldsymbol{\phi}, \gamma) - h(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \boldsymbol{\beta}, r)) - \varphi(\varepsilon_2)] + \varphi(\varepsilon_2)\}^2 \\ &= E\left\{[\varphi(\varepsilon_2 + h(\mathbf{Y}_1, \mathbf{Z}_{2-d}, \boldsymbol{\phi}, \gamma) - h(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \boldsymbol{\beta}, r)) - \varphi(\varepsilon_2)]^2 + \varphi^2(\varepsilon_2)\right. \\ &\quad \left.+ 2[\varphi(\varepsilon_2 + h(\mathbf{Y}_1, \mathbf{Z}_{2-d}, \boldsymbol{\phi}, \gamma) - h(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \boldsymbol{\beta}, r)) - \varphi(\varepsilon_2)]\varphi(\varepsilon_2)\right\} \\ &\leq 2E\left[\varphi^2(\varepsilon_2) + L|h(\mathbf{Y}_1, \mathbf{Z}_{2-d}, \boldsymbol{\phi}, \gamma) - h(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \boldsymbol{\beta}, r)|^2\right] \\ &\leq 2I(f) + 4L(|\boldsymbol{\phi}|^2 + |\boldsymbol{\beta}|^2)E(1 + |\mathbf{Y}_1|^2) < \infty \end{aligned} \quad (25)$$

From assumption AC1 and equation (23) it follows that $\ln f$ is absolutely continuous, and hence

$$|\psi(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \varepsilon_2, \mathbf{v}^*) - \psi(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \varepsilon_2, \mathbf{v})| \leq \int_{-\Delta(\mathbf{y}, \mathbf{z}_q, \eta)}^{\Delta(\mathbf{y}, \mathbf{z}_q, \eta)} |\varphi(\varepsilon_2(\mathbf{v}) + v)| dv, \quad (26)$$

where $\Delta(\mathbf{y}, \mathbf{z}_q, \eta)$ is defined in (24).

Therefore, from equations (25) and (26) and the Cauchy-Bunyakovskii-Schwarz inequality, it follows that

$$\begin{aligned} & E \left[\sup_{\mathbf{v}^* \in U_{\mathbf{v}}} |\psi(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \varepsilon_2, \mathbf{v}^*) - \psi(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \varepsilon_2, \mathbf{v})| \right] \\ & \leq E \{ [2\varphi(\varepsilon_2(\mathbf{v})) + L\Delta(\mathbf{y}, \mathbf{z}_q, \eta)] \Delta(\mathbf{y}, \mathbf{z}_q, \eta) \} \\ & \leq 2 (E\varphi^2(\varepsilon_2(\mathbf{v})))^{1/2} (E\Delta^2(\mathbf{y}, \mathbf{z}_q, \eta))^{1/2} + LE\Delta^2(\mathbf{y}, \mathbf{z}_q, \eta). \end{aligned}$$

Because

$$\begin{aligned} E\Delta^2(\mathbf{y}, \mathbf{z}_q, \eta) &= E \left\{ (1 + |\mathbf{y}|^2) \left[\sqrt{2} |\boldsymbol{\beta}| I(|z_q - r| \leq |r_0(\eta) - r|) + \eta \right] \right\} \\ &\xrightarrow{\eta \rightarrow 0} 0. \end{aligned} \tag{27}$$

From (25) and (27) it follows (21).

If $r = \infty$, we have a result similar to (23)

$$\begin{aligned} |h(\mathbf{y}, \mathbf{z}_q, \boldsymbol{\beta}^*, r) - h(\mathbf{y}, \mathbf{z}_q, \boldsymbol{\beta}^*, r^*)| &\leq |\boldsymbol{\beta}^*| \sqrt{2(1 + |\mathbf{y}|^2)} I(\min(\infty, r^*) < z_q < \max(\infty, r^*)) \\ &\leq |\boldsymbol{\beta}^*| \sqrt{2(1 + |\mathbf{y}|^2)} I(z_q > r^*). \end{aligned}$$

$$\begin{aligned} |\delta(\mathbf{y}, \mathbf{z}_q, \mathbf{v}^*)| &= |h(\mathbf{y}, \mathbf{z}_q, \boldsymbol{\beta}, r) - h(\mathbf{y}, \mathbf{z}_q, \boldsymbol{\beta}^*, r^*)| \\ &\leq |h(\mathbf{y}, \mathbf{z}_q; \boldsymbol{\beta}, r) - h(\mathbf{y}, \mathbf{z}_q; \boldsymbol{\beta}, r^*)| + |h(\mathbf{y}, \mathbf{z}_q; \boldsymbol{\beta}, r^*) - h(\mathbf{y}, \mathbf{z}_q; \boldsymbol{\beta}^*, r^*)| \\ &\leq \left[\sqrt{2} |\boldsymbol{\beta}| I(z_q > r^*) + |\boldsymbol{\beta} - \boldsymbol{\beta}^*| \right] \sqrt{1 + |\mathbf{y}|^2} \\ &\leq \left[\sqrt{2} |\boldsymbol{\beta}| I(z_q > r_0(\eta)) + \eta \right] \sqrt{1 + |\mathbf{y}|^2}, \end{aligned}$$

where $d(r_0(\eta), \infty) = \eta$. Again, if $\Delta_1(\mathbf{y}, \mathbf{z}_q, \eta) = \left[\sqrt{2} |\boldsymbol{\beta}| I(z_q > r_0(\eta)) + \eta \right] \sqrt{1 + |\mathbf{y}|^2}$ it may be proved that

$$E\Delta_1^2(\mathbf{y}, \mathbf{z}_q, \eta) \longrightarrow 0 \text{ when } \eta \longrightarrow 0.$$

If $r = -\infty$

$$\begin{aligned} |h(\mathbf{y}, \mathbf{z}_q, \boldsymbol{\beta}^*, r) - h(\mathbf{y}, \mathbf{z}_q, \boldsymbol{\beta}^*, r^*)| &\leq |\boldsymbol{\beta}^*| \sqrt{2(1 + |\mathbf{y}|^2)} I(\min(-\infty, r^*) < z_q < \max(-\infty, r^*)) \\ &\leq |\boldsymbol{\beta}^*| \sqrt{2(1 + |\mathbf{y}|^2)} I(z_q < r^*). \end{aligned}$$

$$\begin{aligned} |\delta(\mathbf{y}, \mathbf{z}_q, \mathbf{v}^*)| &= |h(\mathbf{y}, \mathbf{z}_q, \boldsymbol{\beta}, r) - h(\mathbf{y}, \mathbf{z}_q, \boldsymbol{\beta}^*, r^*)| \\ &\leq \left[\sqrt{2} |\boldsymbol{\beta}| I(z_q < r^*) + |\boldsymbol{\beta} - \boldsymbol{\beta}^*| \right] \sqrt{1 + |\mathbf{y}|^2} \\ &\leq \left[\sqrt{2} |\boldsymbol{\beta}| I(z_q < r_0(\eta)) + \eta \right] \sqrt{1 + |\mathbf{y}|^2}, \end{aligned}$$

Lastly, if $\Delta_2(\mathbf{y}, \mathbf{z}_q, \eta) = [\sqrt{2}|\boldsymbol{\beta}| I(z_q < r_0(\eta)) + \eta] \sqrt{1 + |\mathbf{y}|^2}$ it may be proved that

$$E\Delta_2^2(\mathbf{y}, \mathbf{z}_q, \eta) \longrightarrow 0 \text{ when } \eta \longrightarrow 0.$$

□

Proof for Theorem 2 with $\sigma = 1$.

Let $\alpha(\mathbf{v}) = E\psi(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \varepsilon_2, \mathbf{v})$ for $\mathbf{v} \in \Lambda$. Assumptions AC1 and AC2, the mean value Theorem, the independence of ε_2 and \mathbf{Y}_1 and the Cauchy-Bunyakovskiï-Schwarz inequality imply that $E|(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \varepsilon_2, \mathbf{v})| < \infty$. Thus, α is well defined. $\alpha(\boldsymbol{\vartheta}) = 0$ and $\ln y \leq y - 1$, $y \neq 1$ are both satisfied. For any open vicinity V of $\boldsymbol{\vartheta}$ in Λ and any $\mathbf{v} \in V^c$, a conditional argument gives

$$\begin{aligned} \alpha(\mathbf{v}) &= E\psi(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \varepsilon_2, \mathbf{v}) \\ &= E \ln \frac{f(\varepsilon_t + h(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \boldsymbol{\vartheta}) - h(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \mathbf{v}))}{f(\varepsilon_t)} \\ &= E \left\{ E \left[\ln \frac{f(\varepsilon_t + h(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \boldsymbol{\vartheta}) - h(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \mathbf{v}))}{f(\varepsilon_t)} \middle| \mathbf{Y}_1, \mathbf{Z}_{2-q} \right] \right\} \\ &< E \left\{ \int [f(y + h(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \boldsymbol{\vartheta}) - h(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \mathbf{v})) - f(y)] dy \right\} \\ &= 0. \end{aligned}$$

For Lemma 1, function α is continuous and from here, by the compacity of V^c , there exists a $\mathbf{v}_0 \in V^c$, such that

$$\sup_{\mathbf{v} \in V^c} \alpha(\mathbf{v}) = \alpha(\mathbf{v}_0) < 0.$$

Let $\delta_0 = -\alpha(\mathbf{v}_0)/3$. For any $\mathbf{v} \in V^c$, by Lemma 1 we have that there exists a $\eta_0 > 0$ such that

$$\begin{aligned} E \sup_{\mathbf{v}^* \in U_{\mathbf{v}}(\eta_0)} \psi(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \varepsilon_2, \mathbf{v}^*) &\leq E\psi(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \varepsilon_2, \mathbf{v}) + \delta_0 \\ &\leq \alpha(\mathbf{v}_0) + \delta_0 \\ &= -2\delta_0. \end{aligned} \tag{28}$$

The compacity of V^c implies that there exists a finite number M of vicinities $U_{\mathbf{v}_j}(\eta_0)$, $\mathbf{v}_j \in V^c$, $j = 1, 2, \dots, M$ such that

$$\bigcup U_{\mathbf{v}_j}(\eta_0) = V^c.$$

Thus, by the ergodicity Theorem and by (28), there exists a T_0 such that for any $T \geq T_0$, $1 \leq j \leq M$,

$$\begin{aligned} \sup_{\mathbf{v}^* \in U_{\mathbf{v}_j}(\eta_0)} \ell_T(\mathbf{v}^*) &\leq \frac{1}{T} \sum \sup_{\mathbf{v}^* \in U_{\mathbf{v}_j}(\eta_0)} \psi(Y_{t-1}, Z_{t-d}, \varepsilon_t, \boldsymbol{\beta}^*, r^*) \\ &\leq E \sup_{\mathbf{v}^* \in U_{\mathbf{v}_j}(\eta_0)} \psi(\mathbf{Y}_1, \mathbf{Z}_{2-q}, \varepsilon_2, \mathbf{v}^*) + \delta_0 \\ &\leq -\delta_0, \quad \text{c.s.} \end{aligned}$$

But

$$\sup_{\mathbf{v} \in V} \ell_T(\mathbf{v}) \geq \ell_T(\boldsymbol{\vartheta}) = 0.$$

Then for any vicinity V of $\boldsymbol{\vartheta}$ in $\mathbf{\Lambda}$, there exists a T_0 such that for all $T \geq T_0$,

$$\begin{aligned} \sup_{\mathbf{v}^* \in V^c} \ell_T(\mathbf{v}^*) &\leq \max_{1 \leq j \leq M} \sup_{\mathbf{v}^* \in U_{\mathbf{v}_j}(\eta_0)} \ell_T(\mathbf{v}^*) \\ &\leq -\delta_0 \\ &< 0 \\ &\leq \sup_{\mathbf{v} \in V} \ell_T(\mathbf{v}). \end{aligned}$$

Which implies that

$$\hat{\boldsymbol{\vartheta}}_T \in V, \quad \text{c.s. for all } V \text{ and for all } T \geq T_0.$$

Since V is arbitrary, it follows that $\hat{\boldsymbol{\vartheta}}_T \rightarrow \boldsymbol{\vartheta}$ c.s. □

Proof of *Theorem 2* with $\sigma > 0$.

Let $\boldsymbol{\vartheta} = (\boldsymbol{\phi}', \gamma, d, \sigma)$ and $\mathbf{v} = (\boldsymbol{\beta}', r, q, \tau)$.

The proof of Lemma 1. with $\sigma > 0$ may be obtained in a similar way to the proof already described. The proof for the case $\sigma = 1$ may also be generalized to the case $\sigma > 0$ with the following modifications.

Define

$$\ell_T(\mathbf{v}) = \ell_T(\boldsymbol{\beta}, r, q, \tau) = \frac{1}{T} \sum \ln \frac{\frac{1}{\tau} f\left(\frac{Y_t - h(Y_{t-1}, Z_{t-q}, \boldsymbol{\beta}, r)}{\tau}\right)}{\frac{1}{\sigma} f\left(\frac{Y_t - h(Y_{t-1}, Z_{t-d}, \boldsymbol{\phi}, \gamma)}{\sigma}\right)}, \quad \mathbf{v} \in \mathbf{\Lambda} \times \mathfrak{R}$$

and

$$\psi(Y_{t-1}, Z_{t-q}, \varepsilon_t, \boldsymbol{\beta}, r) = \ln \frac{\frac{1}{\tau} f(\sigma \varepsilon_t + h(Y_{t-1}, Z_{t-d}, \boldsymbol{\phi}, \gamma) - h(Y_{t-1}, Z_{t-q}, \boldsymbol{\beta}, r))}{\frac{1}{\sigma} f(\sigma \varepsilon_t)}, \quad 1 \leq t \leq n.$$

Given that if $\varepsilon_t^* = \sigma \varepsilon_t$ and f is the density function for ε_t , then the density function for ε_t^* is given by $f^* = \frac{1}{\sigma} f(y/\sigma)$, f^* defined in this way it satisfies assumptions AC1-AC4. □

References

- Berger, J. O. and Wolpert, R. L. (1984). *The Likelihood Principle*. IMS, Hayward, Calif.
- Berkowitz, J. (2001). Testing Density Forecast with Applications to Risk Management. *Journal of the Royal Statistical Society. Series B (Methodological)*. **48**, No.1, 1-38.
- Butler, R. W. (1986). Predictive Likelihood Inference with Applications. *Journal of the Royal Statistical Society. Series B (Methodological)*. **48**, No.1, 1-38.
- Björnstad, J. F. (1990). Predictive Likelihood: A Review. *Statistical Science*. **5**, No. 1, 242-265.

- Clements, M. P., Frances, P. H., Smith, J. and Van Dijk D. (2003). On SETAR Non-linearity and Forecasting. *Journal of Forecasting*. **22**, 359-375.
- Clements, M. P. and Hendry, D. F. (1998). *Forecasting Economic Time Series*. Cambridge University Press.
- Clements, M. P. and Smith, J. (1997). The performance of alternative forecasting methods for SETAR models. *International Journal of Forecasting*. **13**, 463-475.
- Clements, M. P. and Smith, J. (1999). A Monte Carlo Study of the Forecasting Performance of Empirical SETAR Models. *Journal of Applied Econometrics*. **14**, 123-141.
- Chan, K. S. (1993). Consistency and limiting distribution of the least square estimation of a threshold autoregressive model. *The Annals of Statistics*. **21**, No. 1, 520-533.
- Chan, K. S., Petrucci, J. D, Tong, H. and Woolford, S. W. (1985). A Multiple Threshold AR(1) Model. *Journal Applied Probability*. **22**, 267-279.
- Chan, K. S. and Tong, H. (1985). On the use of the deterministic lyapunov functions for the ergodicity of stochastic difference equations. *Advances in Applied Probability*. **17**, 666-678.
- Chan, K. S. and Tsay, R. S. (1998). Limiting Properties of the Least Square Estimator of a Continuous Threshold Autoregressive Model. *Biometrika*. **85**, No. 2, 413-426.
- Chen, C. W. S. and Lee J. C. (1995). Bayesian Inference of Threshold Autoregressive Models. *Journal of Time Series Analysis*. **16**, No. 5, 483-492.
- Chen, C. W. S. (1998). A Bayesian analysis of generalized threshold autoregressive models. *Statistics & Probability Letters*. **40**, 15-22.
- Chen, R. and Tsay, R. S. (1991). On the Ergodicity of Tar(1) Processes. *The Annals of Applied Probability*. **1**, No. 4, 613-634.
- Christoffersen P.F. (1998). Evaluating interval forecasts. *International Economic Review*, **39**: 841-862.
- Dacco, R. and Satchell, S. (1999). Why do Regime-Switching Models Forecast so Badly. *Journal of Forecasting*. **18**, 1-16.
- De Gooijer, J. G. and De Bruin, P. T. (1998). On Forecasting SETAR processes. *Statistics & Probability Letters*. **37**, 7-14.
- Diebold, F. X. and Mariano, R. S. (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*. **13**, Num. 3.
- Diebold, F. X., Gunther, T. A. and Tay, A.S. (1998). Evaluating Density Forecast with Applications to financial Risk Management. *International Economic Review*. **39**, 863-883.
- Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer Series in Statistics.

- Fisher, R. A. (1956). *Statistical Methods and Scientific Inference*. Oliver and Boyd, London.
- Geisser, S. (2002). *Predictive Inference: An Introduction*. Monographs on Statistics and Applied Probability 55, Chapman & Hall.
- González, M. and Gonzalo, J. (1998). Threshold Unit Root Models. *Working paper, Universidad Carlos III*.
- González, M. and Gonzalo, J. (1999). Threshold Unit Root (TUR) Models for Interest Rates. *Working paper, Universidad Carlos III*.
- Granger, C. W. J. and Teräsvirta, T. (1993). *Modelling Nonlinear Economic Relationships*. Advanced Texts in Econometrics.
- Hansen, B. E. (1997). Inference in TAR Models. *Studies in Nonlinear Dynamics and Econometrics*. **2**(1), 1-14.
- Hinkley, D. V. (1979). Predictive Likelihood. *Annals of Statistics*. **7**, 718-728. Corrigendum **8** 694.
- Kalbfleisch, J. D. and Sprott, D. A. (1970). Applications of likelihood methods to models involving large numbers of parameters (with discussion). *Journal Royal Statistics Society Serie B*. **32**, 175-208.
- Kedem, B. and Fokianos, K. (2002). *Regression Models for Time Series Analysis*. Wiley Series in Probability and Statistics.
- Lauritzen, S. L. (1974). Sufficiency, prediction and extreme models. *Scand. Journal Statistics*. **1**, 128-134.
- Mathiasen, P. E. (1979). Prediction functions. *Scand. Journal Statistics*. **6**, 1-21.
- Pesaran, M. H. and Potter, S. M. (1997). A floor and ceiling model of US output. *Journal of Economic Dynamics and Control*. **21**, 661-695.
- Petrucelli J. D. and Woolford S. W. (1984). A Threshold AR(1) Models. *Journal of Applied Probability*. **21**, No. 2, 270-286.
- Qian L. (1998). On maximum likelihood estimators for a threshold autoregression. *Journal of Statistical Planning and Inference*. **75**, 21-46.
- Russell-Noriega, M. G. (2006). Pronósticos en modelos autorregresivos con umbral. *Ph.D. Thesis dissertation*, Centro de Investigación en Matemáticas, A. C., Guanajuato, Guanajuato, México.
- Sáfadi, T. and Morettin, P. A. (2000). Bayesian Analysis of Threshold Autoregressive Moving Average Models. *Sankhya: The Indian Journal of Statistics*. **62**, Series B, Pt. 3, 353-371.
- Tong, H. (1990). *Non-linear Time Series, A Dynamical System Approach*. Oxford Science Publications.