Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models

Timo Terasvirta

Stable URL:

http://links.jstor.org/sici?sici=0162-1459%28199403%2989%3A425%3C208%3ASEAEOS%3E2.0.CO%3B2-7

Journal of the American Statistical Association is currently published by American Statistical Association.

Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models

Timo TERÄSVIRTA*

This article considers the application of two families of nonlinear autoregressive models, the logistic (LSTAR) and exponential (ESTAR) autoregressive models. This includes the specification of the model based on simple statistical tests: linearity testing against smooth transition autoregression, determining the delay parameter and choosing between LSTAR and ESTAR models are discussed. Estimation by nonlinear least squares is considered as well as evaluating the properties of the estimated model. The proposed techniques are illustrated by examples using both simulated and real time series.

KEY WORDS: Canadian lynx; Linearity testing; Nonlinear autoregression; Nonlinear time series; Univariate time series modeling.

1. INTRODUCTION

In recent years, quite a few different nonlinear time series models have been suggested in the literature. This wealth of alternative models creates a problem for a practitioner, because it generally is difficult to decide which type of model to fit to a given data set. To construct workable procedures for specifying an appropriate model or family of models from the data, it appears necessary to narrow down the choice before attempting at the actual specification. Haggan, Heravi, and Priestley (1984) assumed that the model generating the data is a member of the family of state-dependent models (see also Priestley 1988, pp. 99–131). This family contains several well-known nonlinear models, such as the bilinear, the exponential autoregressive (EAR), the threshold autoregressive (TAR), and the smooth transition autoregressive (STAR) models. The authors constructed a technique that is based on studying the graphs for sequences of parameter estimates of a linear state-space model. The estimates are obtained by recursive estimation under the assumption that the parameters follow a random walk. But interpreting the graphs seems often not easy, which makes the method difficult to apply in practical modeling situations. Tsay (1989) considered the modeling of threshold autoregressive processes; his specification technique was based on the use of misspecification tests and recursive residuals as well as t ratios. This procedure is easier to apply than that of Haggan et al. (1984), but the practitioner must assume at the outset that the only conceivable alternative to the linear autoregressive model is the TAR model.

The specification technique that I discuss in this article is more limited in scope than that of Haggan et al. (1984) but in some respects somewhat less restrictive than the TAR

modeling procedure of Tsay (1989). I assume that if the process is not linear, then it is a STAR model. In this article I consider two different STAR models: the logistic STAR (LSTAR) model and the exponential STAR (ESTAR) model. The LSTAR family contains as a special case the single-threshold TAR model. The ESTAR model is a slight generalization of the EAR model of Haggan and Ozaki (1981). It may also be viewed as a generalization of a special case of a double-threshold TAR model. The present approach is fully parametric. For the use of nonparametric and semiparametric modeling techniques in nonlinear univariate time series analysis, the reader is referred to Auestad and Tjøstheim (1990, 1991) and Lewis and Stevens (1991).

A brief description of the modeling procedure is as follows. I use a linearity test to determine the delay parameter of the nonlinear model as in Tsay (1989). After fixing the delay, a short sequence of F tests is needed to decide between the ESTAR and the LSTAR family of models. Section 2 introduces the models; Section 3, the linearity tests. Section 4 outlines the specification technique. Section 5 discusses the estimation of parameters by conditional least squares and postestimation diagnostic checks. Section 6 contains both simulated and real examples, and Section 7 provides concluding remarks.

2. THE MODELS

Consider a logistic smooth transition autoregressive model of order p (LSTAR(p) model)

$$y_t = \pi_{10} + \pi_1' w_t + (\pi_{20} + \pi_2' w_t)$$
$$\times [(1 + \exp\{-\gamma(y_{t-d} - c)\})^{-1} - 1/2] + u_t, \quad (2.1)$$

where $u_t \sim \text{nid}(0, \sigma_u^2)$, $\pi_j = (\pi_{j1}, \ldots, \pi_{jp})'$, $j = 1, 2$, $\gamma > 0$, and $w_t = (y_{t-1}, \ldots, y_{t-p})'$. Subtracting one-half from the logistic function is useful just in deriving linearity tests; the LSTAR models I estimate do not contain that term. The normality assumption is needed if the specification tests of the article are derived as Lagrange multiplier (LM)-type tests. If they are interpreted simply as tests based on artificial regressions (Davidson and MacKinnon 1990), then a martingale difference assumption is sufficient. Another model of interest is the exponential STAR model of order p (ESTAR(p) model)

$$y_t = \theta_{10} + \theta'_1 w_t + (\theta_{20} + \theta'_2 w_t)$$

$$\times (1 - \exp\{-\gamma^*(y_{t-d} - c^*)^2\}) + v_t, \quad (2.2)$$

where $v_t \sim \text{nid}(0, \sigma_v^2)$, $\theta_j = (\theta_{j1}, \ldots, \theta_{jp})'$, $j = 1, 2$, $\gamma^* > 0$. Note that (2.2) is a generalization of the regular exponential autoregressive (EAR) model where $\theta_{20} = c^* = 0$ (see, for example, Haggan and Ozaki 1981 and Priestley 1988, pp. 83–89). The purpose of the generalization is to make the EAR model location invariant like the LSTAR model. Like the LSTAR model, the EAR model is also a model with a smooth transition from one regime to the other (and back again), with y_{t-d}; to stress this property, I shall call the generalized model the ESTAR model.

Assume for notational simplicity that $d \leq p$ and that there are observations $y_{-p+1}, \ldots, y_0, y_1, \ldots, y_T$ available. Of these, $y_{-p+1}, \ldots, y_0$ are considered fixed. There are several ways of defining linearity of (2.1) and (2.2). For instance, if H_{01}: $\gamma = 0$ holds, then model (2.1) is a linear $AR(p)$ model. If H_{02}: $\gamma^* = 0$ is true, then model (2.2) is linear. Arguing that there are also other restrictions making these models linear is another way of saying that the models are not identified under H_{01} or H_{02}. For example, if H_{01} holds, then π_{20}, π_2, and c in (2.1) can take any value, and if H_{02} is true, then the same holds for θ_{20}, θ_2, and c^* in (2.2). If $\pi_{20} = 0$ and $\pi_2 = 0$, then (2.1) is again $AR(p)$, and γ and c can assume any value; a similar argument holds for (2.2). If (2.1) and (2.2) are linear, then I assume that $Q_1(z) = z^p - \pi_{11}z^{p-1} - \cdots - \pi_{1p} = 0$ and $Q_2(z) = z^p - \theta_{11}z^{p-1} - \cdots - \theta_{1p} = 0$ have their roots inside the unit circle. This guarantees the stationarity and ergodicity of these models under linearity.

3. LM-TYPE LINEARITY TESTS

3.1 Logistic smooth transition autoregression

To lay groundwork for the specification procedure, I first consider Lagrange multiplier (LM)-type tests for H_{01} and H_{02}. Because (2.1) and (2.2) are not identified under these null hypotheses, I follow the suggestion of Davies (1977) and first derive the test statistic keeping the unidentified values fixed. The information matrix of σ^2 and the other parameters is block diagonal, so that σ^2 can be conveniently kept fixed in the derivation of the test. Consider (2.1) with $\gamma = 0$ and define $\tau = (\tau'_1, \tau_2)'$, where $\tau_1 = (\pi_{10}, \pi'_1)'$ and $\tau_2 = \gamma$. Let $\hat{\tau}_1$ be the least squares estimator of τ_1 under H_0 and set $\hat{\tau} = (\hat{\tau}'_1, 0)'$. Furthermore, let $z_t = z_t(\tau) = \partial u_t/\partial \tau$ and $\hat{z}_t = z_t(\hat{\tau}) = (\hat{z}'_{1t}, \hat{z}_{2t})'$, where the partition conforms to that of τ. Then the general form of the LM statistic is

$$\text{LM} = \hat{\sigma}^{-2}(\sum \hat{u}_t \hat{z}_{2t})^2 \{\sum \hat{z}_{2t}^2$$

$$- \sum \hat{z}_{2t}\hat{z}'_{1t}(\sum \hat{z}_{1t}\hat{z}'_{1t})^{-1} \sum \hat{z}_{1t}\hat{z}_{2t}\}^{-1}, \quad (3.1)$$

where $\hat{\sigma}^2 = (1/T)\sum_1^T \hat{u}_t^2$ and $\hat{u}_t = y_t - \hat{\tau}'_1\bar{w}_t$, $\bar{w}_t = (1, w'_t)'$, $t = 1, \ldots, T$ (see, for example, Saikkonen and Luukkonen 1988). When the model is a LSTAR(p) model, $\hat{z}_{1t} = -\bar{w}_t$ and $\hat{z}_{2t} = -(1/4)\{\pi_{20}(y_{t-d} - c) - c\pi'_2 w_t + \pi'_2 w_t y_{t-d}\}$. I write $\hat{z}_{2t}(\pi) = \hat{z}_{2t}$, where $\pi = (\pi_{20}, \pi'_2, c)'$, to stress the dependence of $\hat{z}_{2t}$ on π_{20}, π_2, and c. Consider the auxiliary regression

$$\hat{u}_t = \hat{z}'_{1t}\tilde{\beta}_1 + \hat{z}_{2t}(\pi)\tilde{\beta}_2 + \varepsilon_t(\pi), \quad t = 1, \ldots, T, \quad (3.2)$$

where $\tilde{\beta}_1 = (\tilde{\beta}_{11}, \ldots, \tilde{\beta}_{1,p+1})'$ and $\varepsilon_t(\pi)$ is an error term. Let $\text{SSR}(\pi)$ be the sum of squared residuals of model (3.2) and let SSR_0 be the corresponding sum under H_0: $\tilde{\beta}_2 = 0$ ($\gamma = 0$). Following the arguments of Harvey (1990, pp. 172–174) and Saikkonen and Luukkonen (1988), the LM statistic (3.1) is equivalent to

$$\text{LM}(\pi) = (\text{SSR}_0 - \text{SSR}(\pi))/\hat{\sigma}^2. \quad (3.3)$$

Statistic (3.3) has an asymptotic $\chi^2(1)$ distribution under H_0: $\tilde{\beta}_2 = 0$, but the distribution is dependent on π. To remove this dependence, Davies (1977) suggested a conservative statistic:

$$\text{LM}_1 = \sup_\pi \text{LM}(\pi) = (\text{SSR}_0 - \inf_\pi \text{SSR}(\pi))/\hat{\sigma}^2. \quad (3.4)$$

The distribution of (3.4) under the null hypothesis is generally unknown. Nevertheless, in some cases it is a standard one (see, for example, Godfrey 1988, pp. 87–91 for a discussion). Fortunately, the present example yields a standard χ^2 distribution. Consider an auxiliary regression

$$\hat{u}_t = \hat{z}'_{1t}\tilde{\beta}_1 + \beta'_2 w_t y_{t-d} + \varepsilon_t, \quad (3.5)$$

where ε_t is an error term, $\tilde{\beta}_1 = (\beta_{10}, \beta'_1)'$, and

$$\beta_{10} = \pi_{10} - (c/4)\pi_{20}, \quad (3.6)$$

$$\beta_1 = \pi_1 - (c/4)\pi_2 + (1/4)\pi_{20}e_d, \quad (3.7)$$

and

$$\beta_2 = \pi_2/4. \quad (3.8)$$

Furthermore, $e_d = (0, 0, \ldots, 0, 1, 0, \ldots, 0)'$, with the dth element equal to unity. If (3.6)–(3.8) apply, then (3.5) reduces to (3.2). Thus, minimizing the sum of squared errors $\sum_1^T \varepsilon_t^2$ with respect to β_1 and β_2 yields $\inf_\pi \text{SSR}$, and under H_{01}, LM_1 has an asymptotic $\chi^2(p)$ distribution if $Eu_t^4 < \infty$.

An important thing to notice is that while linearity implies $\beta_2 = 0$ in (3.5), β_2 is only a function of π_2 and independent of π_{20}. Thus in a situation where the nonlinearity is mainly due to π_{20}, a LM-type test based on (3.5) has low power against the alternative. Luukkonen, Saikkonen, and Teräsvirta (1988) encountered the same problem in a slightly more general setting. Their solution is also available here. Replace $F(z) = (1 + \exp\{-z\})^{-1} - .5$ in (2.1), where $z = \gamma(y_{t-d} - c)$, by its third-order Taylor approximation $T_3(z) = g_1 z + g_3 z^3$, where $g_1 = \partial F/\partial z|_{z=0}$ and $g_3 = (1/6)\partial^3 F/\partial z^3|_{z=0}$. This yields the following approximation:

$$y_t = \pi_{10} + \pi'_1 w_t$$

$$+ (\pi_{20} + \pi'_2 w_t)T_3(\gamma(y_{t-d} - c)) + u'_t. \quad (3.9)$$

Deriving a LM-type test statistic for linearity leads to an auxiliary regression,

$$\hat{u}_t = \hat{z}'_{1t}\tilde{\beta}_1 + \sum_{j=1}^p \beta_{2j}y_{t-j}y_{t-d} + \sum_{j=1}^p \beta_{3j}y_{t-j}y_{t-d}^2$$

$$+ \sum_{j=1}^p \beta_{4j}y_{t-j}y_{t-d}^3 + v'_t \quad (3.10)$$

(see Luukkonen et al. 1988 for details). The null hypothesis is

$$H'_0: \quad \beta_{2j} = \beta_{3j} = \beta_{4j} = 0, \qquad j = 1, \ldots, p. \quad (3.11)$$

When a linear autoregressive model holds and $Eu_t^8 < \infty$, the test statistic, $LM_2 = (SSR_0 - SSR)/\hat{\sigma}^2$ where SSR is the sum of squared estimated errors from the full regression (3.10), has an asymptotic $\chi^2(3p)$ distribution. Statistic LM_2 may also be derived from the artificial model

$$y_t = \gamma_0 + \gamma'_1 \mathbf{w}_t + \gamma'_2(\mathbf{w}_t y_{t-d}) + \gamma'_3(\mathbf{w}_t y_{t-d}^2)$$
$$+ \gamma'_4(\mathbf{w}_t y_{t-d}^3) + v''_t, \quad (3.12)$$

where $v''_t \sim \text{nid}(0, \sigma_{v''}^2)$, $\gamma_j = (\gamma_{1j}, \ldots, \gamma_{1p})'$, and $j = 1, \ldots, 4$, when the null hypothesis is $\gamma_2 = \gamma_3 = \gamma_4 = 0$. In terms of (3.9), one obtains

$$\gamma_2 = g_1 \gamma \pi_2 + 3 g_3 \gamma^3 c^2 \pi_2 - 3 g_3 \gamma^3 c \pi_{20} \mathbf{e}_d, \quad (3.13)$$

$$\gamma_3 = -3 g_3 \gamma^3 c \pi_2 + g_3 \gamma^3 \pi_{20} \mathbf{e}_d, \quad (3.14)$$

and

$$\gamma_4 = g_3 \gamma^3 \pi_2. \quad (3.15)$$

Parameter π_{20} appears in only one element of γ_3, namely γ_{3d}, whereas γ_4 is independent of π_{20}. Thus a more parsimonious test of $H_{01}: \gamma = 0$ in (2.1) sensitive to π_{20} could be obtained by merely augmenting (3.5) by the term $\beta_{3d} y_{t-d}^3$.

Note that if d is assumed unknown, $d \leq p$, then the test statistics must be generalized to account for this. The generalization of (3.5) is the linearity test of Tsay (1986), and generalizing (3.10) yields the statistic S_2 in Luukkonen et al. (1988).

3.2 Exponential smooth transition autoregression

Saikkonen and Luukkonen (1988) derived a LM-type test for testing linearity against (2.2) where $\theta_{20} = c^* = 0$. Only minor modifications are needed to make their procedure applicable to this slightly more general situation. If the model is an ESTAR(p) model, then in (3.1) $\hat{z}_{1t} = -\bar{\mathbf{w}}_t$ as before, and, letting $\theta = (\theta_{20}, \theta'_2, c^*)'$,

$$\hat{z}_{2t}(\theta) = -(y_{t-d} - c^*)^2 (\theta_{20} + \theta'_2 \mathbf{w}_t)$$
$$= -(\theta'_2 \mathbf{w}_t y_{t-d}^2 + \theta_{20} y_{t-d}^2 - 2 c^* \theta'_2 \mathbf{w}_t y_{t-d}$$
$$+ c^{*2} \theta'_2 \mathbf{w}_t - 2 c^* \theta_{20} y_{t-d} + c^{*2} \theta_{20}).$$

This yields the following auxiliary regression:

$$\hat{v}_t = \tilde{\beta}'_1 \hat{z}_{1t} + \beta'_2 \mathbf{w}_t y_{t-d} + \beta'_3 \mathbf{w}_t y_{t-d}^2 + e'_t, \quad (3.16)$$

where e'_t is an error term and $\tilde{\beta}_1 = (\beta_{10}, \beta'_1)'$, with $\beta_{10} = \theta_{10} - c^{*2} \theta_{20}$ and $\beta_1 = \theta_1 - c^{*2} \theta_2 + 2 c^* \theta_{20} \mathbf{e}_d$.
Furthermore,

$$\beta_2 = 2 c^* \theta_2 - \theta_{20} \mathbf{e}_d \quad (3.17)$$

and

$$\beta_3 = -\theta_2. \quad (3.18)$$

Thus the null hypothesis of linearity is $H_0: \beta_2 = \beta_3 = 0$, and when it is valid, the test statistic

$$LM_3 = (SSR_0 - SSR)/\hat{\sigma}^2, \quad (3.19)$$

where SSR is the sum of squared estimated errors from (3.16), has an asymptotic $\chi^2(2p)$ distribution. The test of Saikkonen and Luukkonen (1988) is a special case where $\beta_2 = 0$ in (3.16) as $\theta_{20} = c^* = 0$, so that (3.19) has an asymptotic $\chi^2(p)$ distribution under $H_0: \beta_3 = 0$.

Note that this test already has power against the case where the nonlinear part of the model consists solely of an intercept multiplied by the transition function. Thus there is no pressing reason such as there was earlier to consider the fourth-order Taylor expansion of the transition function, the use of which would double the dimension of the null hypothesis. At least in small samples, when the maximum lag of the linear model is not very short, the usefulness of such an extension remains doubtful.

The three aforementioned LM-type tests can also be carried out as F tests. The true significance level of the test may then be reasonably close to its nominal value, whereas the power may often be higher than that of the asymptotic χ^2 test (see, for example, Harvey 1990, pp. 174–175, for discussion). In fact, when the maximum lag p is relatively large while the time series is short, the χ^2 test is likely to suffer from size problems and should not be applied. The nominal significance levels of the tests remain free for the model builder to choose. A general advice is to be more critical if the time series is long and let the significance levels decline with the sample size.

4. MODEL SPECIFICATION

The first problem a prospective builder of nonlinear time series models must solve is determining whether a linear model is an adequate representation of the process generating the data. In this article an adequate linear model simply means a linear autoregressive model. If the answer is negative, then the next step will be the selection of an appropriate nonlinear model. Here the alternatives are the LSTAR and the ESTAR model, and the specification procedure as a whole may be viewed as a sequence consisting of the following three steps:

1. Specifying a linear autoregressive model
2. Testing linearity for different values of d, the delay parameter, and if it is rejected determining d in (2.1) and (2.2)
3. Choosing between the LSTAR and the ESTAR by testing a sequence of nested hypotheses within (3.10) or, equivalently, (3.12).

The choice between LSTAR and ESTAR may also be postponed until after both types of models are estimated and made using postestimation model evaluation criteria.

4.1 Specification of autoregression

In most practical situations, the length of the autoregression, p, and the lags present in the model are unknown even if the model is linear. Steps 2 and 3 in the specification procedure build on the assumption of knowing the autoregressive structure, so that it must be determined first. A common technique is to use an order selection criterion like AIC

(Akaike 1974) or SBIC (Rissanen 1978; Schwarz 1978) to select a proper subset of lags. (For an overview of such criteria see, for example, de Gooijer, Abraham, Gould, and Robinson 1985 or Teräsvirta and Mellin 1986.) My experience in the present context is that SBIC, which is dimension-consistent, sometimes leads to too parsimonious a model in the sense that the estimated residuals of the selected model are not free from serial correlation. Thus the use of any model selection procedure should be accompanied by a proper test for residual autocorrelation, like the portmanteau test of Ljung and Box (1978). This is important, because omitted autocorrelation as such may also cause rejection of the linearity hypothesis. In that respect the LM-type tests of this article are no different from many other misspecification tests that may have power against a wide variety of alternatives to the null hypothesis.

4.2 Determining the delay parameter

So far I have assumed that the delay parameter d is known, but in applications it needs to be determined from the data. Tsay (1989) discussed the same problem in the specification of TAR models. His idea was to first select the order of autoregression p and then determine d by varying it and choosing the value minimizing the p value of his linearity test. I shall apply the same idea here. If $p_T(d)$ is the p value of the F test of (3.11) within (3.10), then I choose the delay parameter $\hat{d}$ such that $p_T(\hat{d}) = \min_{1 \le d \le D} p_T(d)$. This procedure is motivated as follows. When d is selected correctly, (3.10) is the appropriate auxiliary regression against the nonlinear alternative. If another d is selected, then (3.10) is misspecified. The power of the corresponding test against the misspecified nonlinear model thus can hardly be expected to be systematically higher than the power of the test based on the correctly specified auxiliary regression.

To illustrate the asymptotic situation, assume that there are two values, d_1 and d_2, to choose among. The former is selected if $p_T(d_1) < p_T(d_2)$ or $t_T(d_1) > t_T(d_2)$, where t_T is the chi-squared test statistic. Assume that the moment conditions holding under linearity also hold when the alternative is true. Then $\text{plim}_{T \to \infty} T^{-1} t_T(d)$, which is the approximate slope of the test (Geweke 1981), exists. Thus, asymptotically, d_1 is selected when the approximate slope of the linearity test against the alternative with d_1 is greater than that of the test against the alternative with d_2. The procedure is thus consistent if the test with the correct d has the greatest approximate slope. Because the model is nonlinear under the alternative, showing this is difficult. But I considered the procedure numerically by the LSTAR(2) model

$$y_t = 1.8 y_{t-1} - 1.06 y_{t-2}$$
$$+ (\pi_{20} - .9 y_{t-1} + .795 y_{t-2}) F(z_t) + u_t, \quad (4.1)$$

where $\pi_{20} = .02$ and

$$F(y_{t-d}) = (1 + \exp\{-\gamma(y_{t-d} - c)\})^{-1}, \quad \gamma > 0, \quad (4.2)$$

$c = .02$, $u_t \sim \text{nid}(0, .02^2)$, and $\gamma = 20, 100$. I simulated 1,000 replications from (4.1) with (4.2) using the random number generator in Gauss 2.0. The sample sizes were 100 and 300; the first 100 observations from each sequence were discarded to avoid the initialization effects. The delay parameter d was varied from 1 to 5, and the minimum d was recorded if the corresponding p value did not exceed .01. The results appear in Table 1. The empirical size of the overall linearity test is about 3–4%. The larger the γ (the greater the slope of the transition function at $y_{t-1} = .02$), the more powerful the test. The uncertainty about d also decreases with the power of the procedure. The results with $T = 300$ suggest that the procedure may be consistent. For $T = 300$ and $\gamma = 100$, the null hypothesis is strongly rejected for all values of d considered, but the rejection for $d = 1$ is strongest in all except one of the 1,000 replications.

4.3 Choosing between LSTAR and ESTAR models

To discriminate between LSTAR and ESTAR models, all I may need is a sequence of ordinary F tests. To set up the sequence, consider equations (3.10) [or (3.12)] and (3.16). First, comparing (3.10) and (3.16), it is seen that the latter does not contain fourth-order terms. Thus, after rejecting linearity by LM_2 or its F version, which I shall call F_L, one may test

$$H_{01}: \gamma_4 = 0 \quad (4.3)$$

in (3.12) against $H_{11}: \gamma_4 \ne 0$ with an ordinary F test (F_4). A rejection of (4.3) can in principle be interpreted as a rejection of the ESTAR model family. Second, compare (3.14) and (3.18), and (3.13) and (3.17). Note that (3.18) cannot equal 0 if the model is an ESTAR model, unless it is a very special case with $\theta_2 = 0$. On the other hand, (3.14) may equal 0 if the true model is a LSTAR model. This happens if $c = \pi_{20} = 0$. Thus one may test

$$H_{02}: \gamma_3 = 0 \,|\, \gamma_4 = 0 \quad (4.4)$$

against $H_{12}: \gamma_3 \ne 0 \,|\, \gamma_4 = 0$ using another F test (F_3). If (4.4) is not rejected, then this is taken as evidence in favor of a LSTAR model. A rejection is not very informative one way or the other. The last F test (F_2) in the sequence is the test of

$$H_{03}: \gamma_2 = 0 \,|\, \gamma_3 = \gamma_4 = 0 \quad (4.5)$$

Table 1. Empirical Power of the Linearity Test Based on (3.10) for Different Values of Delay Estimated From 1,000 Trials with 100 and 300 Observations Each From the LSTAR Model (4.1) With (4.2)

	Delay					
γ	1	2	3	4	5	EP[a]
$T = 100$						
0[b]	.24	.14	.21	.29	.14	.029
20	.70	.14	.08	.03	.05	.24
100	.90	.06	.02	.01	.01	.86
$T = 300$						
0	.22	.29	.17	.10	.22	.041
20	.93	.05	.01	.01	.01	.94
100	1.00	.00	—	—	—	1

[a] EP = empirical power of the linearity test, in which linearity is rejected if the p value of at least one individual test lies below .01.
[b] The selection frequencies for $\gamma = 0$ are based on only 29 ($T = 100$) and 41 ($T = 300$) observations.

against H_{13}: $\gamma_2 \neq 0 | \gamma_3 = \gamma_4 = 0$. If the true model is a LSTAR model, then it is seen from (3.13) that (4.5) is generally rejected. Thus, rejecting H_{03} after accepting H_{02} supports the choice of a LSTAR model. On the other hand, (3.17) shows that it is quite possible to accept H_{03} if the data-generating process is an ESTAR model and $c^* = \theta_{20} = 0$. Thus, accepting H_{03} after rejecting H_{02} often points at an ESTAR model.

The preceding may suggest the following procedure. Carry out the three F tests and note which hypotheses are rejected. Interpret the results as previously and use that information to choose between the ESTAR and LSTAR models. This procedure may cause problems, however. As an example, consider a fourth-order Taylor expansion of the transition function when the model is (2.2). In that case (3.17) will equal 0 only if $c^* = \theta_{20} = 0$. Suppose that the true model is an ESTAR model such that at least one of these parameters does not equal 0. Then H_{01} may be rejected if the significance level of F_4 is not very low and/or the sample size is large.

This would mean erroneously selecting a LSTAR model. It is again better to compare the relative strengths of the rejections. If the model is a LSTAR model, then typically H_{01} and H_{03} are rejected more strongly than H_{02}. For an ESTAR model, the situation may be expected to be the opposite. I propose the following decision rule. After rejecting the general null hypothesis, carry out the three F tests. If the p value of F_3 (the test of H_{02}) is the smallest of the three, select an ESTAR model; if not, choose a LSTAR model.

I simulated this rule with the LSTAR model (4.1) with (4.2) for $d = 1$, $\gamma = 100$, and various values of c and π_{20}. I also considered the case where the true model was ESTAR and replaced (4.2) by

$$F(y_{t-1}) = 1 - \exp\{-1{,}000(y_{t-1} - c^*)^2\}. \quad (4.6)$$

The results, found in Table 2, are based on the number of cases in which the overall linearity test F_L rejected linearity at the 5% significance level in 1,000 trials. It is seen that the selection procedure works very well when the true model is a LSTAR model. It performs best when the number of observations below c is about the same as above it but deteriorates only slightly when this is no longer true. The performance of the rule improves with the sample size.

The results for the ESTAR model with transition function (4.6) are interesting. When the observations are symmetri-cally distributed about c^* (i.e., when $c^* = 0$ in (4.6) and $\theta_{20} = 0$), the ESTAR model is selected correctly almost without exception. On the other hand, the performance of the decision rule changes when the distribution becomes asymmetric. In the last experiment for $T = 300$, only 6% of the observations on the average are less than c, and an ESTAR model is selected in just 40% of the cases. It is important to note that in this case, the ESTAR and LSTAR models are close substitutes for each other. If most of the observations lie on one side of c, then (4.2) with suitable parameter values can be a very good approximation to the relevant part of (4.6). The seemingly unsatisfactory performance of the selection procedure just reflects the substitutability of the two models. On the other hand, if $c^* = \theta_{20} = 0$, then no LSTAR model can adequately approximate the ESTAR model that generated the data. This is because the transition function of the LSTAR model is monotonically increasing, whereas the range of the observations stretches out on both tails of the transition function of the ESTAR model. In that case it is crucial to distinguish between the two models, and the decision rule then also performs very well.

If a LSTAR model is always well approximated by an ESTAR model, then the former family of models may be redundant. But this may be true only when the first derivative of the transition function of the LSTAR model at c is not too large. If the function grows rapidly from 0 to 1, then the corresponding LSTAR model cannot be approximated satisfactorily by an ESTAR model, and it is useful to have a procedure to choose between them.

The idea of the aforementioned selection technique is to save time and effort by avoiding the estimation of nonlinear models as long as possible, but it need not be applied strictly. In practice the model-builder can compare the three p values. If the ones for F_4 and F_3 or for F_3 and F_2 are very near each other in relative terms, then I recommend postponing the choice and proceeding by estimating both kinds of models. Of course, this can always be done whatever the test results. The final choice should then be based on results of the post-estimation model evaluation.

5. ESTIMATION AND EVALUATION

After specifying the family of models and determining the delay parameter, I can estimate the parameters of the model by conditional least squares. Assume that the model is

Table 2. Relative Frequencies of Choosing the Type of Model (LSTAR or ESTAR) Correctly by the Minimum p Value Rule When the Data Were Generated by Model (4.1) With Either a LSTAR (4.2) or an ESTAR (4.6) Transition Function at Sample Sizes 100 and 300

				Average							Average		
			Selection							Selection			
Sample size	c	π_{20}	frequency	min(y)	max(y)	freq{$y \leq c$}	Sample size	c^*	θ_{20}	frequency	min(y)	max(y)	freq{$y \leq c$}
100	0	0	.971	−.180	.089	.547	100	0	0	.902	−.086	.086	.502
	0	.02	.904	−.134	.124	.166		0	.02	.653	−.068	.123	.221
	.02	.04	.898	−.104	.178	.093		.02	.04	.608	−.037	.175	.095
300	0	0	.991	−.213	.105	.549	300	0	0	.991	−.099	.099	.499
	0	.02	.986	−.156	.137	.142		0	.02	.741	−.078	.137	.259
	.02	.04	.930	−.100	.191	.038		.02	.04	.405	−.043	.191	.064

Header spanning: Model: LSTAR (columns Sample size through freq{$y \leq c$}) and Model: ESTAR (columns Sample size through freq{$y \leq c$}).

NOTE: The rightmost columns contain the estimated average maxima and minima for {y_t} and the average relative frequency of y values at most equal to c.

$$y_t = g(\alpha, F_{t-1}) + u_t,$$

where $\alpha = (\alpha_1, \ldots, \alpha_\gamma)'$ is the parameter vector, $F_t = \sigma(y_t, y_{t-1}, \ldots, y_{-p+1})$ (σ-algebra), $\{y_t\}$ is a stationary and ergodic process, $E(u_t|F_{t-1}) = 0$, and $\text{var}(u_t|F_{t-1}) = \sigma^2$. The problem is to minimize

$$Q_T(\alpha) = \sum_{t=1}^{T} \{y_t - g(\alpha, F_{t-1})\}^2 \tag{5.1}$$

with respect to α. Under certain regularity conditions, Klimko and Nelson (1978) discussed how the sequence $\{\hat{\alpha}_T\}$ of estimators from (5.1) converges to the true parameter vector α as $T \to \infty$. They also presented the conditions for the limiting estimator to be multivariate normal. Tjøstheim (1986) showed that if the EAR model is stationary and ergodic, then the results of Klimko and Nelson (1978) apply. Tong (1990, pp. 299–302) proved the same result for the stationary and ergodic LSTAR model.

Consider first the estimation of the ESTAR model (2.2). Haggan and Ozaki (1981) pointed out that joint estimation of $\{\gamma^*, \theta_{10}, \theta_1, \theta_2\}$ is difficult. The reason for this is that the estimator of γ^* tends to be very heavily negatively correlated with that of θ_2, so that a nonlinear optimization may not converge at all. Haggan and Ozaki suggested that the estimation be carried out by keeping γ^* fixed and using a grid of values to estimate it. In the present case there are two more parameters, c^* and θ_{20}, to be estimated. This does not seem to make the estimation more complicated if the model is well specified, probably because the ESTAR model is more flexible than the EAR model. My proposal is as follows. First, standardize the exponent of F by dividing it by $\hat{\sigma}^2(y)$, the sample variance of y_t. This makes it easier to select a reasonable starting value for (standardized) γ^*. Quite often $\gamma^* = 1$ is a proper starting value. Then, simply estimate the whole set of parameters by nonlinear least squares. If the algorithm does not converge, carry out the estimation using a grid for γ^*. However, when estimating ESTAR models to specify the lag structure, it may be practical to keep γ^* completely fixed ($\gamma^* = 1$ is again a reasonable choice after the exponent has been standardized) and estimate it only after a satisfactory specification has been found. This is because even fairly large changes in γ^* often seem to have rather minor effects on the estimates of the other parameters. Furthermore, estimating γ^* may turn out to be difficult if the ESTAR model is not sufficiently well specified.

As for LSTAR models, joint estimation of $\{\gamma, c, \pi_{10}, \pi_{20}, \pi_1, \pi_2\}$ is not always an easy task either, because estimating γ and c sometimes may pose a problem. Bates and Watts (1988, p. 87) and Seber and Wild (1989, pp. 480–481) touched on this issue, considering the estimation of γ. When γ is large, the slope of the transition function at c is steep, and a large number of observations in the neighborhood of c would be needed to estimate γ accurately. Even relatively large changes in γ then have only a minor effect on the shape of F. As a result, the sequence of estimates for γ may converge rather slowly. This uncertainty is also reflected in the estimated standard deviation of γ, which tends to be large for large values of γ. Here the situation is even more intricate, because even c is an unknown parameter. If γ is large and

c at the same time is sufficiently close to 0, then a negative definite Hessian may not be obtained for numerical reasons even when convergence is achieved. A workable solution is to rescale the parameters before estimation by shrinking γ and inflating c; see Example 4. That γ is large is often noticed early during the specification of the lag structure. In that case I propose that γ be fixed and estimated only after the final specification has been found. Even when γ may not be large, finding a good starting value for it is important. The exponent of F thus should be rescaled as in the estimation of ESTAR models; for LSTAR models, it is natural to divide the exponent by $\hat{\sigma}(y)$.

If convergence is reached in the estimation of an LSTAR or ESTAR model, then the validity of the model must be evaluated. Because of the existence of local minima, at least when the time series are relatively short, the first task is to check whether the estimates look reasonable. For instance, if the estimate of c or c^* is far outside the observed range of $\{y_t\}$, then the estimated model is not satisfactory. Excessively large standard deviation estimates for coefficient estimators (other than $\hat{\gamma}$) suggest that the model contains redundant parameters. All of the parameters estimated with large standard deviations need not be redundant; usually, a subset is. Experimenting with more parsimonious models reveals which variables can actually be omitted from the model. It is useful to consider the long-term properties of the model. This can be done only numerically, using a sequence of observed values of the time series as starting values and extrapolating the series without adding noise. If the process diverges, then the model is rejected. Several sets of starting values should be applied. Postsample forecasting may also be used to evaluate the behavior of the estimated model, although the results may be inconclusive if the observations in the prediction period cover only a part of the relevant values of the transition function. Finally, model evaluation should include an inspection of the residuals and residual autocorrelations. Examples of tests that come into the question are given in the next section.

6. EXAMPLES

In this section I consider a few examples to illustrate the application of the aforementioned procedures to data. The time series include a realization from a LSTAR model, another from an ESTAR model, and two real data sets.

Example 1. The time series A is a realization of 100 observations from the LSTAR(2) model (4.1) with (4.2), $d = 1$, and $\gamma = 100$. The "lower regime" ($F = 0$) of this LSTAR process is such that the roots of the characteristic polynomial $g(z) = z^2 - 1.8z + 1.06$ are a complex pair with the modulus 1.03 (see Table 5), so that the regime is explosive. The roots of $g(z) = z^2 - .92z + .265$ corresponding to the "upper regime" ($F = 1$) are also a complex pair with the modulus .51, so that this regime is not explosive. In fact, the oscillations are very weak; they have half-lives of only one (time) unit. As to the long-term behavior, the model has a unique stable stationary point (Ozaki 1985), $y_\infty = .055$. The graph of series A is shown in Figure 1. When the order of the autoregressive model was determined, both AIC and SBIC suggested an
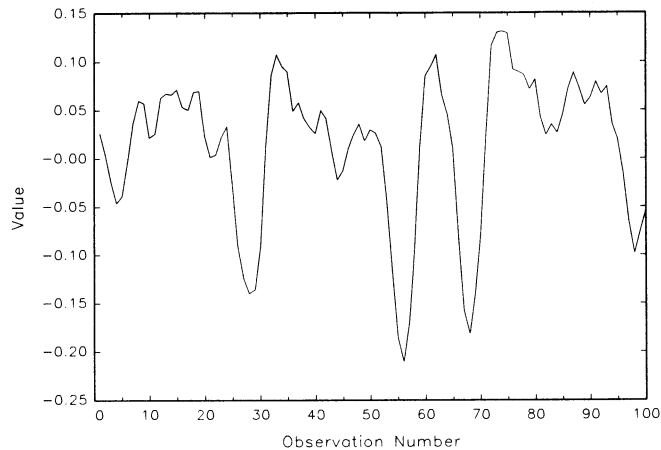
Figure 1. A Realization of 100 Observations From the Logistic Smooth Transition Autoregressive Model (4.1) With Transition Function (4.2) and the Slope Parameter $\gamma = 100$.

Table 4. p Values of the Fourth-Order, Cubic, and Quadratic Linearity Tests at a Given Value of the Delay Parameter for Time Series A–D

Time series	Lag length in AR model	d	p Values			
			$p(F_L)$	$p(F_4)$	$p(F_3)$	$p(F_2)$
A	2	1	3.0×10^{-4}	.374	.055	1.3×10^{-4}
B	2	1	7.7×10^{-4}	.832	.0022	.0048
C	11	3	3.9×10^{-5}	.0084	.015	.0038
D	9	4	.0041	.021	.120	.056

AR(2) model. Judging by the Ljung–Box statistic, the AR(2) model did not seem underspecified, and I accepted it as a basis for linearity tests. Table 3 presents the results of testing linearity against the general null hypothesis. The linearity is rejected most strongly at $d = 1$, so that the choice $\hat{d} = 1$ is unambiguous. Table 4 shows the results of the tests of the conditional null hypotheses. The p value of F_2 is very small; thus I selected a LSTAR model. Estimating the parameters of (4.1) from A yields

$$y_t = 1.80 y_{t-1} - 1.08 y_{t-2} + (.0060 - .74 y_{t-1} + .82 y_{t-2})$$
$$\quad (.086) \qquad (.090) \qquad (.0077) \quad (.17) \qquad (.16)$$
$$\times (1 + \exp\{-8.9 \times 13(y_{t-1} - .021)\})^{-1} + \hat{u}_t \quad (6.1)$$
$$\quad (6.7) \qquad\qquad (.0068)$$

$$s = .0209, \qquad ML(4) = 2.4(.66), \qquad sk = -.58,$$

$$ek = .28, \qquad LJB = 5.8(.056), \qquad AIC = -7.663,$$

$$SBIC = -7.786,$$

where $s^2 = (T - r)^{-1} \sum_1^T \hat{u}_t^2$ is the residual variance, ML(k) is the McLeod and Li (1983) test of no autoregressive conditional heteroscedasticity (ARCH) of order k, sk is skewness, ek is extra kurtosis, and LJB is the test of Lomnicki (1961) and Jarque and Bera (1980) of the normality of errors. The figures in parentheses following the values of the test statistics are p values. The exponent of F has been standardized by multiplying by $1/\hat{\sigma}(y) = 13.0$. The corresponding γ ($=100/13.0 = 7.69$) is somewhat overestimated. This agrees with

the simulation results of Chan and Tong (1986) and Luukkonen (1990). The standard deviation of the estimate is relatively large, which might be expected as $\partial F/\partial y = 25$ at $y = .02$ for $\gamma = 7.69$. The roots of the lower regime are estimated very accurately (see Table 5). The estimated roots of the upper regime equal .68 and .38. Note the amount of skewness in the residuals. The nonstationary local dynamics for low values of y_{t-1} ultimately lead to higher values of y. On the other hand, there is nothing inherent in the model to push a realization down to low values, because the upper regime is stationary and very weakly oscillatory. A large negative shock is needed for that. Because such a shock is exogenous, it thus remains largely unexplained, and these shocks show up as negative skewness of the residuals.

There are three negative residuals greater than two residual standard errors in absolute value. They correspond (in descending order, with the residuals in parentheses) to observations 25 ($-.062$), 66 ($-.060$), and 53 ($-.043$); compare this information with Figure 1.

I also fitted an ESTAR(2) model to the data. The estimated model is

$$y_t = 1.91 y_{t-1} - 1.18 y_{t-2} + (.0076 - 1.07 y_{t-1} + 1.18 y_{t-2})$$
$$\quad (.095) \qquad (.095) \qquad (.0070) \quad (.14) \qquad (.095)$$
$$\times (1 - \exp\{-.29 \times 169(y_{t-1} + .086)^2\}) + \hat{u}_t \quad (6.2)$$
$$\qquad\qquad (.11) \qquad\qquad (.017)$$

$$s = .0212, \qquad ML(4) = 2.1(.71), \qquad sk = -.69,$$

$$ek = .57, \qquad LJB = 9.1(.011), \qquad AIC = -7.651,$$

$$SBIC = -7.755.$$

The restriction $\theta_{12} = -\theta_{22}$ is imposed, because it is strongly supported by the data. As a whole, the estimates of the AR parameters do not change very much compared to those of the LSTAR model (6.1), but $\hat{c}^* = -.086$. Of 100 observations in the series, 86 exceed $\hat{c}^*$. This is not unexpected, because the transition function now must approximate that of a

Table 3. p Values of the Fourth-Order Linearity Test F_L at Different Values of the Delay Parameter for Time Series A–D

Time series	Delay								
	1	2	3	4	5	6	7	8	9
A	3.0×10^{-4}	.0031	.0040	.072	.18	.21	.0024	.099	.88
B	7.7×10^{-4}	9.2×10^{-4}	.24	.73	.52	.64	.30	.86	.064
C	.94	.050	3.9×10^{-5}	.033	.76	.79	.21	.010	.15
D	.27	.32	.44	.0041	.27	.11	.034	.61	.68

Table 5. Roots of the Characteristic Polynomials for F = 0 and F = 1
in the True Model (4.1) and Estimated LSTAR
and ESTAR Models (6.1)-(6.5)

Model	Regime	Root	Modulus	Period
(4.1) LSTAR with γ = 100 (or ESTAR with γ^* = 1,000)	Lower (Mid)	.90 ± .50i	1.03	12.4
	Upper (Outer)	.45 ± .25i	.51	12.4
(6.1)	Lower	.90 ± .52i	1.04	11.9
	Upper	.68	.68	
		.38	.38	
(6.2)	Mid	.95 ± .52i	1.09	12.6
	Outer	.83		
(6.3)	Mid	.87 ± .66i	1.10	9.5
	Outer	.56	.56	
		.34	.34	
(6.4)	Lower	1.17		
	Upper*	.72 ± .57i	.92	9.5
		−.28 ± .93i	.97	3.4
(6.5)	Lower*	.81 ± .65i	1.04	9.2
	Upper*	.85 ± .31i	.90	18.2
		.87	.87	

* Only dominant roots are shown.

LSTAR model. The true transition function and its estimates from (6.1) and (6.2) are found in Figure 2. The residual variance of (6.2) is only about 2% higher than that of (6.1), but AIC and SBIC correctly select (6.1). The ESTAR model (6.2) characterizes the dynamics of (4.1) fairly well. The roots of the characteristic polynomials for $F = 0$ and $F = 1$ in Table 5 show that the mid-regime is explosive and the outer one is stationary. The roots for $F = 0$ also have approximately the correct period length. The three largest residuals of (6.2) correspond to the same observations as those in the LSTAR model. This is an example of a situation in which a true LSTAR model is reasonably well approximated by an appropriate ESTAR model.

Example 2. The time series B contains 100 observations from the ESTAR(2) model (4.1) with (4.6), $c^* = .02$, $\pi_{20} = .02$, and $u_t \sim$ nid$(0, .02^2)$. The random numbers generated are the same as in Example 1. The mid-regime is explosive, whereas the upper and lower regimes are not. The model has a stable equilibrium point, $y_\infty = .0087$. Table 3 indicates that I am able to correctly specify the delay by applying the minimum p value rule, although $\hat{d} = 2$ might also be considered. Table 4 shows that the same rule applied to the model choice problem yields an ESTAR model. The estimated ESTAR(2) model for series B is

$$y_t = 1.74y_{t-1} - 1.21y_{t-2} + (.018 - .85y_{t-1} + 1.02y_{t-2})$$
$$\quad (.20) \quad (.17) \quad \quad (.0055) \quad (.23) \quad \quad (.20)$$

$$\times (1 - \exp\{-2.59 \times 473(y_{t-1} - .024)^2\}) + \hat{u}_t \quad (6.3)$$
$$\quad (1.18) \quad \quad \quad (.0036)$$

$$s = .0202, \quad \text{ML}(4) = 4.53(.34), \quad \text{sk} = .27,$$

$$\text{ek} = -.53, \quad \text{LJB} = 2.29(.32).$$

Parameter γ^* is overestimated, but only slightly. Table 5 contains the estimated roots of the characteristic polynomial for the mid-regime ($F = 0$) and the outer regime ($F = 1$).

For the former, they are $.87 \pm .68i$, which is an explosive pair with modulus 1.10. The latter polynomial has two small real roots, indicating a strong drift towards the center as in the original model. Furthermore, $\hat{y}_\infty = .0094$. The tests of the errors do not reveal anything unusual. In particular, in this model the local dynamics become stationary when one moves away from the mid-regime ($F = 0$). This quickly dampens large fluctuations originating around $y = .02$. Thus no skewness or excess kurtosis may be expected, and the kurtosis of the estimated residuals is in fact less than that of the normal distribution.

In fitting a LSTAR(2) model to the data, it turned out that three of the parameter estimates had very large standard deviations. Setting $\pi_{21} = 0$ allowed more accurate estimation. The residuals of the estimated model did not reveal any sign of misspecification, and the model had a unique stable stationary point. But its residual variance was about 14% higher than that of (6.3), and both AIC and SBIC clearly favored (6.3).

Example 3. Next I discuss the specification of a STAR model for series C, the Canadian lynx data. Lim (1987) summarized the work done in analyzing this time series; see also the work of Tsay (1989) and Tong (1990, chap. 7). The time series comprises 114 annual observations from 1821 to 1934; following other workers, I shall analyze its 10-based logarithms. Tong (1977) found that AIC selected an AR(11) model, and I test linearity using that model as the basis for the test. Earlier work made me expect a rejection of linearity, and Table 3 shows this to be the case here as well. Biological facts suggest a delay of 2 or 3 years (see Tong 1990, p. 377); here the minimum p value rule yields $\hat{d} = 3$.

The results in Table 4 point to a LSTAR model, because the rejection of (4.4) is much less strong than that of the other two hypotheses. On the other hand, Priestley (1988, pp. 128-129) found an EAR model to be an appropriate choice for fitting to the lynx series when he applied the specification procedure of Haggan et al. (1984).

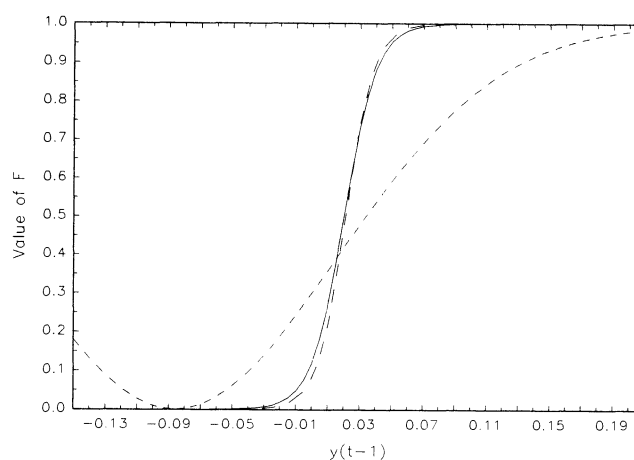The final estimated LSTAR(11) model is as follows:

Figure 2. The Transition Function of the LSTAR(2) Model (4.1) With (4.2), d = 1 and γ = 100 (Solid Line), Its Estimate From LSTAR(2) Model (6.1) (Long-Dashed Line) and Its Estimate From ESTAR(2) Model (6.2) (Short-Dashed Line).

Table 6. Residual Autocorrelations and Their Ljung–Box Statistics for Models (6.4) and (6.5)

| Time series | Model | Residual autocorrelation at lag | | | | | | | | | | | | |
		1	2	3	4	5	6	7	8	9	10	11	12	$T^{-1/2}$
C	(6.4)	.08	−.04	−.03	−.04	−.06	.10	.11	−.03	−.02	−.03	−.01	−.12	.10
D	(6.5)	−.00	−.00	.03	−.00	−.11	.10	.05	−.11	−.02	−.00	−.11	−.19	.10

$$y_t = 1.17 y_{t-1} + (-.92 y_{t-2} + 1.00 y_{t-3} - .41 y_{t-4}$$
$$\quad (.035) \qquad (.20) \qquad (.31) \qquad (.16)$$

$$\quad + .27 y_{t-9} - .21 y_{t-11})$$
$$\qquad (.055) \qquad (.079)$$

$$\times (1 + \exp\{-1.73 \times 1.8(y_{t-3} - 2.73)\})^{-1} + \hat{u}_t \quad (6.4)$$
$$\qquad (.65) \qquad\qquad\qquad (.13)$$

$$s = .187, \quad ML(4) = 4.77(.31), \quad sk = -.024,$$

$$ek = -.17, \quad LJB = .13(.94).$$

The exponent again has been scaled by dividing it by $\hat{\sigma}(y) = 1/1.8$. Table 6 shows that the residual autocorrelations are appreciably small. Note that the McLeod–Li test for the residuals of the AR(11) model strongly rejects the null hypothesis: for ML(4), $p = .0082$. Now it seems that the rejection is not due to ARCH but rather a general indication of nonlinearity in the original process. The residuals look normal: the Lomnicki–Jarque–Bera statistic has a large p value; also note the values for skewness and excess kurtosis.

The estimated model has a limit cycle of 77 years. Within this cycle are eight subcycles, of lengths 9, 10, 10, 9, 10, 9, 10, and 10 years. This is not unreasonable, because the estimated spectrum of the logarithmic lynx series has a distinct peak at the frequency corresponding to approximately 9.5 years (see, for example, Tong 1983, p. 176). For comparison, the TAR(3; 1, 7, 2) model of Tsay (1989) has a limit cycle of 9 years. The TAR(2; 8, 3) model of Tong and Lim (1980, model 9.1) comes closer to the present LSTAR model in that it has a limit cycle of 19 years with 9- and 10-year subcycles (cf. Tong 1983, p. 189).

The lower regime of (6.4) is explosive as in the TAR model of Tsay (1989). On the other hand, the roots of $g(z) = z^{11} - 1.17z^{10} + .92z^9 - 1.00z^8 + .41z^7 - .27z^2 + .21$ lie inside the unit circle (see Table 5). Thus, according to (6.4), when the lynx population has decreased long enough, it ultimately rebounds explosively, but the stationary upper regime brings the resulting growth under control. For comparison, the TAR(2; 8, 3) model ($\hat{d} = 2$) of Tong and Lim (1980) and the subset TAR(2; 11, 11) model ($\hat{d} = 2$) of Thanoon (1990, model 3.1) have explosive upper regimes. The TAR(3; 1, 7, 2) model ($\hat{d} = 2$) of Tsay (1989) is different from those two models in that both the lowest and the highest regime of the three contain explosive roots. The AR polynomial of the lowest regime has a single root that is explosive like in (6.4). The roots of the corresponding polynomial in the highest regime are a complex pair with modulus greater than unity.

When I fit an ESTAR model with the same lags in (6.4) to the lynx series, the results resemble those in Example 1. The estimates of the lag coefficients of the ESTAR model

are very close to those in (6.4). The residual variance is just 1% higher than that of (6.4), and $\hat{c}^* = 1.78$. Only four observations in the series are lower than this value. This is another example of an ESTAR model being a close substitute for a LSTAR model with a relatively slowly increasing transition function.

Example 4. The time series D consists of 104 observations, four-quarter differences of the quarterly logarithmic volume of industrial production in West Germany for 1961–1986. The source is the *OECD Main Economic Indicators*. The graph of the series is shown in Figure 3. AIC selects an AR(9) model, and from Table 3 it is seen that linearity is rejected and $\hat{d} = 4$ is a straightforward choice for the delay estimate. The test sequence (see Table 4) points at a LSTAR model, because the p value of the F test for (4.3) is smallest of the three.

The estimated LSTAR(9) model is

$$y_t = .82 y_{t-1} - .47 y_{t-4} + .43 y_{t-7}$$
$$\quad (.070) \quad (.087) \quad (.14)$$

$$\quad + (.35 y_{t-2} + .24 y_{t-5} - .43 y_{t-7} - .36 y_{t-8} + .30 y_{t-9})$$
$$\qquad (.10) \qquad (.12) \qquad (.15) \qquad (.12) \qquad (.11)$$

$$\times (1 + \exp\{-1,600 \times 33.5(y_{t-4} - .0070)\})^{-1} + \hat{u}_t \quad (6.5)$$
$$\qquad (49,000) \qquad\qquad (.0070)$$

$$s = .0250, \quad ML(4) = 2.59(.63), \quad sk = -.19,$$

$$ek = .30, \quad LJB = .92(.63).$$

The most conspicuous detail of (6.5) is the enormous standard deviation of the estimate for γ. This is the situation
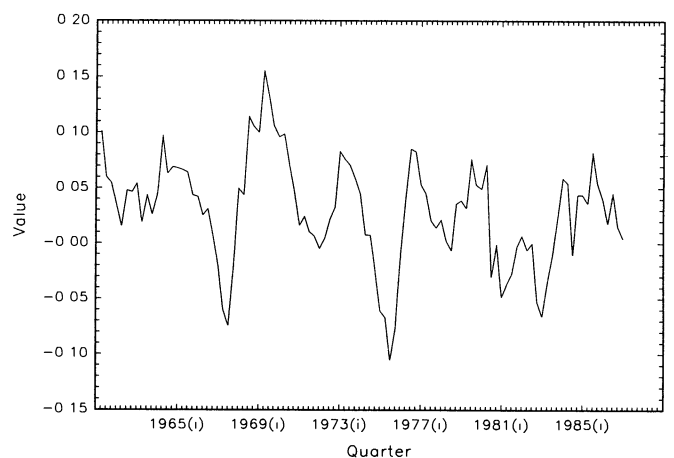

Figure 3. Four-Quarter Differences of the Quarterly Logarithmic Volume of West German Industrial Production, 1961(i)–1986(i).
Source: OECD Economic Indicators.

discussed earlier where γ is large, making joint estimation of the two parameters in F difficult. Although the estimation converged from several sets of starting values, a negative definite Hessian was obtained only after rescaling γ by dividing it by 1,000 and rescaling c by multiplying it by 10. (The exponent was customarily standardized by dividing it by $\hat{\sigma}(y) = 1/33.5$).

The estimates of both γ and c, and their standard deviations, are sensitive to rescaling. They also vary with the estimation algorithm used and the stopping rule. This is not crucial, however, as the general message is clear: γ is very large, and joint estimation of γ and c thus is uncertain. Moreover, these uncertainties do not affect the other estimates.

The best-fitting model ESTAR model I managed to obtain for series D had the same number of parameters as (6.5), but its residual variance was about 7% higher than that of the LSTAR model. Furthermore, its long-term properties were unsatisfactory.

The properties of (6.5) resemble those of the LSTAR(2) model (6.1). The model has a unique single stationary point, $\hat{y}_\infty = 0$. As seen in Table 5, the roots of the characteristic polynomial of the lower regime ($F = 0$) include an explosive complex pair. In the upper regime all the roots are stationary. In particular, the pair with the largest modulus (.90) has a fairly long period: 18 quarters. The economic interpretation is that German industrial output will accelerate strongly again some time after entering a period of stagnation. On the other hand, the recoveries tend to last longer. Manufacturing moves into a recession more slowly than it moves out of it unless there is a large negative shock to start a recession; compare this with model (6.1). Teräsvirta and Anderson (1992) recently modeled other international industrial output series using STAR models, some of them with results rather similar to the ones reported here for the West German series.

7. CONCLUSIONS

The specification technique I have discussed in this article requires rather strict assumptions: if linearity is rejected, then the alternative can only be a STAR model. On the other hand, quite a few nonlinear time series may be adequately described using either LSTAR or ESTAR models. The technique is simple to use, and the examples in the article show that it works reasonably well already in rather small samples. Note that the same method also works when the true nonlinear model is a single-threshold TAR model, because the F tests also have power against this borderline case. But in many applications a smooth transition model may be more attractive than a threshold model. For instance, macroeconomic time series are most often the results of decisions made by a large number of economic agents. Even if one assumes that the agents make only dichotomous decisions or change their behavior discretely, it is unlikely that they do this simultaneously. Thus if only an aggregated process is observed, then the regime changes in that process may be more accurately described as being smooth rather than discrete. In animal ecology, it also seems plausible to think that a possible change in population dynamics from a state with a small population to overpopulation and vice versa may be continuous rather than discrete.

REFERENCES

Akaike, H. (1974), "A New Look at the Statistical Model Identification," *IEEE Transactions in Automatic Control*, AC-19, 716–723.

Auestad, B., and Tjøstheim, D. (1990), "Identification of Nonlinear Time Series: First-Order Characterization and Order Determination," *Biometrika*, 77, 669–687.

—— (1991), "Functional Identification in Nonlinear Time Series," in *Nonparametric Functional Estimation and Related Topics*, G. Roussas (ed.), Amsterdam: Kluwer Academic Publishers, pp. 493–507.

Bates, D. M., and Watts, D. G. (1988), *Nonlinear Regression Analysis & Its Applications*, New York: John Wiley.

Chan, K. S., and Tong, H. (1986), "On Estimating Thresholds in Auto-Regressive Models," *Journal of Time Series Analysis*, 7, 179–190.

Davidson, R., and MacKinnon, J. G. (1990), "Specification Tests Based on Artificial Regressions," *Journal of the American Statistical Association*, 85, 220–227.

Davies, R. B. (1977), "Hypothesis Testing When a Nuisance Parameter is Present Only Under the Alternative," *Biometrika*, 64, 247–254.

de Gooijer, J. G., Abraham, B., Gould, A., and Robinson, L. (1985), "Methods for Determining the Order of an Autoregressive-Moving Average Process: A Survey," *International Statistical Review*, 53, 301–329.

Engle, R. F. (1982), "Autoregressive Conditional Heteroscedasticity With Estimates of the Variance of UK Inflation," *Econometrica*, 50, 987–1008.

Geweke, J. (1981), "The Approximate Slopes of Econometric Tests," *Econometrica*, 49, 1427–1455.

Godfrey, L. G. (1988), *Misspecification Tests in Econometrics*, Cambridge, U.K.: Cambridge University Press.

Haggan, V., Heravi, S. M., and Priestley, M. B. (1984), "A Study of the Application of State-Dependent Models in Non-Linear Time Series Analysis," *Journal of Time Series Analysis*, 5, 69–102.

Haggan, V., and Ozaki, T. (1981), "Modeling Nonlinear Random Vibrations Using an Amplitude-Dependent Autoregressive Time Series Model," *Biometrika*, 68, 189–196.

Harvey, A. C. (1990), *The Econometric Analysis of Time Series* (2nd ed.), New York: Philip Allan.

Jarque, C. M., and Bera, A. K. (1980), "Efficient Tests for Normality, Homoscedasticity, and Serial Independence of Regression Residuals," *Economics Letters*, 6, 255–259.

Klimko, L. A., and Nelson, P. I. (1978), "On Conditional Least Squares Estimation for Stochastic Processes," *The Annals of Statistics*, 6, 629–642.

Lewis, P. A. W., and Stevens, J. G. (1991), "Nonlinear Modeling of Time Series Using Multivariate Adaptive Regression Splines (MARS)," *Journal of the American Statistical Association*, 86, 864–877.

Lim, K. S. (1987), "A Comparative Study of Various Univariate Time Series Models for Canadian Lynx Data," *Journal of Time Series Analysis*, 8, 161–176.

Ljung, G. M., and Box, G. E. P. (1978), "On a Measure of Lack of Fit in Time-Series Models," *Biometrika*, 65, 297–303.

Lomnicki, Z. A. (1961), "Tests for Departure From Normality in the Case of Linear Stochastic Processes," *Metrika*, 4, 37–62.

Luukkonen, R. (1990), "Estimating Smooth Transition Autoregressive Models by Conditional Least Squares," in *On Linearity Testing and Model Estimation in Nonlinear Time Series Analysis* by R. Luukkonen, Helsinki: Finnish Statistical Society.

Luukkonen, R., Saikkonen, P., and Teräsvirta, T. (1988), "Testing Linearity Against Smooth Transition Autoregressive Models," *Biometrika*, 75, 491–499.

McLeod, A. I., and Li, W. K. (1983), "Diagnostic Checking ARMA Time Series Models Using Squared Residual Autocorrelations," *Journal of Time Series Analysis*, 4, 269–273.

Ozaki, T. (1985), "Non-Linear Time Series Models and Dynamical Systems," in *Handbook of Statistics* (Vol. 5), eds. E. J. Hannan, P. R. Krishnaiah, and M. M. Rao, Amsterdam: Elsevier, pp. 25–83.

Priestley, M. B. (1988), *Non-Linear and Non-Stationary Time-Series Analysis*, London and San Diego: Academic Press.

Rissanen, J. (1978), "Modeling by Shortest Data Description," *Automatica*, 14, 465–471.

Saikkonen, P., and Luukkonen, R. (1988), "Lagrange Multiplier Tests for Testing Non-Linearities in Time Series Models," *Scandinavian Journal of Statistics*, 15, 55–68.

Schwarz, G. (1978), "Estimating the Dimension of a Model," *The Annals of Statistics,* 4, 461–464.

Seber, G. A. F., and Wild, C. J. (1989), *Nonlinear Regression,* New York: John Wiley.

Teräsvirta, T., and Anderson, H. M. (1992), "Characterizing Nonlinearities in Business Cycles Using Smooth Transition Autoregressive Models," *Journal of Applied Econometrics,* 7, S119–S136.

Teräsvirta, T., and Mellin, I. (1986), "Model Selection Criteria and Model Selection Tests in Regression Models," *Scandinavian Journal of Statistics,* 13, 159–171.

Thanoon, B. Y. (1990), "Subset Threshold Autoregression With Applications," *Journal of Time Series Analysis,* 11, 75–87.

Tjøstheim, D. (1986), "Estimation in Nonlinear Time Series Models," *Stochastic Processes and Their Applications,* 21, 251–273.

Tong, H. (1977), "Some Comments on the Canadian Lynx Data," *Journal of the Royal Statistical Society,* Ser. A, 140, 432–436.

——— (1983), *Threshold Models in Non-Linear Time Series Analysis,* New York: Springer-Verlag.

——— (1990), *Non-Linear Time Series: A Dynamical System Approach,* Oxford, U.K.: Oxford University Press.

Tong, H., and Lim, K. S. (1980), "Threshold Autoregression, Limit Cycles and Data" (with discussion) *Journal of the Royal Statistical Society,* Ser. B, 42, 245–292.

Tsay, R. (1986), "Non-Linearity Tests for Time Series," *Biometrika,* 73, 461–466.

——— (1989), "Testing and Modeling Threshold Autoregressive Processes," *Journal of the American Statistical Association,* 84, 231–240.