Mathematical Statistics
Stockholm University

Comparing forecasts from LSTAR and linear Autoregressive models

Steven Wahlström

Examensarbete 2004:11

Postal address:

Mathematical Statistics
Dept. of Mathematics
Stockholm University
SE-106 91 Stockholm
Sweden

Internet:

http://www.math.su.se/matstat

Mathematical Statistics
Stockholm University
Examensarbete **2004:11**,
http://www.math.su.se/matstat

Comparing forecasts from LSTAR and linear Autoregressive models

Steven Wahlström*

May 2004

Abstract

Financial economics is largely based on time series. One of the major uses for time-series models is to produce forecasts and nonlinear models have particular difficulties in doing this, especially for several steps ahead. Two related classes of models that have been proposed for forecasting returns are regime switching models and artificial neural networks. This work will focus on regime switching models particularly the logistic smooth transition autoregressive (LSTAR) model.

The LSTAR model is a nonlinear model, which makes multi-period forecasting more different than from a related linear model. Obtaining the forecast would require numerical integration and multiple integration would be encountered for longer time horizons. The purpose with this essay is to compare forecasts from the LSTAR model to those from a linear one. In this work, Monte Carlo technique will be used to produce multi-step forecast for a given LSTAR model.

*Postal address: Dept of Mathematical Statistics, Stockholm University, SE-106 91 Stockholm, Sweden. E-mail: steven.wahlstrom@swipnet.se. Supervisor: Timo Teräsvirta.

Acknowledgements

First of all, I would like to thank my supervisor professor *Timo Teräsvirta* at the Department of Economic Statistics, Stockholm School of Economics, for his advice and assistance. I am most grateful for the help he has provided me and for introducing me to the world of nonlinear models. I would also like to thank *Mikael Andersson* at the Department of Mathematics division Mathematical Statistics, Stockholm University, for his support and for the effort of going through the material. Furthermore, I am also greatly indebted to the memory of my father who once introduced me to start studying mathematical statistics. I thank heartily all people who have made this essay possible.

Steven Wahlström
Stockholm, May 2004

Contents

1 Introduction

1.1 Statistical background

Empirical analysis in macroeconomics as well as in financial economics is largely based on times series. Ever since the mid-seventeenth century it has been standard to view economic time series as realizations of stochastic processes. The existence of unexpected shocks or innovations to the economy plus measurement errors, strongly suggest that economic variables are stochastic. This approach allows the model builder to use statistical inference in constructing and testing equations that characterize relationships between economic variables.

Investment functions, production functions and Philips curves are usually specified in nonlinear forms. For the time series analysis area, linear techniques have long dominated macroeconomic model building. There thus seems to be a need for explanatory statistical techniques to produce nonlinear models, perhaps used in conjunction with appropriate theories. Before examining non-linear models, it is necessary to define what is meant by a linear model. So far there is little experience using nonlinear models with economic data and therefore it is often recommended (*Granger and Teräsvirta [1993]*) to use a test of linearity before fitting models and decide which type of model is the most appropriate.

There are many nonlinear models that have been suggested by economic theory. One class that does seem to be potentially relevant involves switching regimes. These models are interesting because many parts of the economic theory include the idea that the economy behaves differently if some variable lies in one region rather in another. The class of switching regression models also includes models that assume a finite number of linear regimes. A special type of these nonlinear models is the smooth transition autoregressive (STAR) model. This model nests a linear part and the extra parameters give the model added flexibility. Once the model is specified we are faced with the problem of making forecasts. (*Lundbergh and Terävirta [2002]*)

There are many ways of producing forecasts such as formal model-based statistical analyses, statistical analyses not based on parametric models, forecasting based on leading indicators, survey analyses, knowledge of econometric systems and time series. One of the best known statistical approaches of forecasting derives from classical regressions analysis. This approach has been applied to time series and it is useful in forecasting analysis. However nonlinear models have particular difficulties to produce forecasts especially for several steps ahead. Alternative ways of producing forecast are necessary. During the recent years, computer-based simulation methods have revolutionized the way we approach statistical analysis.

A forecast might be judged successful if it is close to the outcome but that judgment may also depend on how close it is measured. Depending upon the degree of forecast uncertainty, forecasts may range from being highly informative to utterly useless for the tasks at hand. A measure of forecast uncertainty provides an assessment of the expected or predicted uncertainty of the forecast errors which helps to qualify the forecasts themselves and to give a picture of the expected range of likely outcomes. For example, we could try the ideas for measuring the success or failure of forecasts by testing combinations of forecasting models for encompassing. Forecasting methods can also be compared by Monte Carlo (or stochastic simulation) where an investigator generates artificial data on which the models are compared in repeated trials. (*Ericsson [2002]*)

Despite the many problems that economic forecasts from economic systems confront, these models offer a vehicle for understanding and learning from, failures, as well as consolidating our growing knowledge of economic behaviour.

1.2 Objectives

A recurring argument in practice is that a nonlinear model hardly forecasts better than a linear one. This seems to be the case even if the nonlinear model seems to fit the data better than the linear one when they are built on the same information set. The purpose with this work is thus to compare forecasts from a Logistic Smooth Transition Autoregressive (LSTAR) model to those from a linear one.

1.3 Delimitation and methods

In this work we discuss the forecasting power of a LSTAR model trough simulated examples. Of course, the results cannot therefore be generalized to the whole family of STAR models. We will generate forecasts from each LSTAR model by using Monte Carlo simulation according to chapter three. We do this under the assumption that we have estimated the parameters of a generated LSTAR model. In this way, we don't have to use different linearity test to make sure we have a nonlinear model from the beginning. From the forecasting point of view it would also be important to know the property of the information set available at the forecast step. Different lags will hence be used to examine if all the information about the future of the series is contained in its most recent value. Furthermore, we generate the corresponding forecasts from the related linear autoregressive model. To compare the accuracy of the forecasts we use the Root Mean Square Error (RMSE) measure which will be defined in section 3.1.

2 Nonlinear time series models

This chapter introduces some basic ideas of time series analysis and stochastic processes. Some nonlinear models will be described in the first section. Of particular importance and one of the main objectives with this chapter is the concept of Smooth Transition Autoregressive (STAR) Models. This will be described in section 2.2. Estimating procedures for these Smooth Transition Autoregressive Models will be considered in section 2.3. Once the data have been fitted to a specific STAR model, we are faced with the problem of forecasting.

2.1 Time series and general nonlinear models

An important part of the analysis for a set of observations is the selection of a suitable probability model. To allow for the possibly unpredictable nature of future observations it is natural to suppose that each observation x_t is a realized value of a certain random variable X_t. A **time series** is a set of observations x_t, each one being recorded at a specific time t.[1] This could be compared with a general **stochastic process** $\{X(t), t \in T\}$ which is a collection of random variables. That is, for each $t \in T$, X(t) is a random variable. The index t is often interpreted as time and, as a result, X(t) is the state of the process at time t. A **time series model** for the observed data $\{x_t\}$ is a specification of the joint distribution of a sequence of random variables $\{X_t\}$ of which $\{x_t\}$ is postulated to be a realization. Having chosen a model (or a family of models) it then becomes possible to estimate parameters, check for goodness-of-fit to the data, and possibly to use the fitted model to enhance our understanding of the mechanism generating the series. Once a satisfactory model has been developed, it could be used to predict future values.

Before considering the specification of the time series, it is necessary to impose the properties of the models. Define $x_{t,m}$ as the section of a series, $x_{t-j}, j = 0,....,m-1$ so that the section contains m consecutive terms and $p_{t,m}(x)$ denote the probability distribution function of $x_{t,m}$ A series x_t is said to be (completely) **stationary** if $p_{t,m}(x)$ is not a function of t for any finite m. Thus, a series is stationary if its generating mechanism is time invariant and if the series is **short-memory**, so that the conditional distribution of $x_{t+h,m}$ given $x_{t-j}, j \geq 0$ is equal to the unconditional distribution, for h large. In general, consider the conditional distribution of x_{t+h} given the information set $I_t : x_{t-j}, j \geq 0$, i.e. $F_h(x) = \text{prob}(x_{t+h} \leq x | I_t)$ The series is said to be **short memory in distribution** (SMD) if: $\lim_h F_h(x) = \bar{F}(x)$ does not depend on I_t. If $\bar{F}(x)$ always depends on I_t, the series may be called **long-memory in distribution** (LMD).[2]

[1] Harvey [1993]

[2] Granger and Teräsvirta [1993]

Let $f_{t,h} = E[x_{t+h}|I_t]$ be the optimum (least-squares) forecast of x_{t+h} using the information in I_t, not necessarily linearly. If: $\lim_h f_{t,h} = D$ where D is some random variable, then x_t is said to be **short-memory in mean** (SMM) if the distribution of D does not depend on I_t. If $f_{t,h}$ continues to depend on I_t as h increases, the series will be called **long-memory in mean** (LMM). However, SMD implies SMM but not vice versa, and LMM implies LMD, but not necessarily vice versa. For example, series can be short-memory in mean but long-memory in distribution.

Suppose we have a situation where there is an output to a system, y_t and a distinct vector input, x_t which is observed with v components. A general model may then be

$$y_t = f(x_t, x_{t-1}, ..., x_{t-q}) + \varepsilon_t \qquad \textit{(2.1.1)}$$

where we have q lags of the input series and ε_t is a zero mean independent and identically distributed (i.i.d) sequence. When the function f is well behaved, the Taylor series around zero will give an expansion:

$$\begin{aligned} y_t = \mu + & \text{ (linear component)} + \text{(quadratic component)} + \\ & + \text{(cubic component)} + ... + \text{(m components)} + \varepsilon_t \end{aligned} \qquad \textit{(2.1.2)}$$

For example, the cubic component contains all triple term $x_{i,t}x_{j,s}x_{k,p}, i, j, k$ ranging over 1 to v, and t, s, p each ranging over 0 to q. The expansion around zero and the notation will quickly be unwieldy and the number of parameters will be very large. A problem with input-output systems is that they are using the available information inefficiently. There is no direct use of lagged dependent variables. [3]

i) An alternative input-output model which does not have this objection is the "**state affine model**", used by Sontag (1979), Guegan (1987), and others. Let $X_{t,r} \equiv (X_t, X_{t-1}, ..., X_{t-r})$ be an input series and its lags and consider a generating mechanism

$$\sum_{k=0}^{q} P_k(X_{t,r}) y_{t-k} = Q(X_{t,r}) \qquad \textit{(2.1.3)}$$

where $P_k(.)$ and $Q(.)$ are polynomials in the components of $X_{t,r}$.

[3] Granger and Teräsvirta [1993]

ii) A related, general class of models, called "**state-dependent models**", was introduced by Priestley (1980). Starting with a relationship of the form $y_t = h(I_{t-1}) + e_t$ where $I_t : y_{t-j}, e_{t-j}\ 0 \le j \le m$ is the information set and the sequence $e_t, t = 1,..., T$ is called white noise if its mean $m = E[e_t]$ is constant, its variance $s^2 = \text{var}(e_t)$ is also a constant and covariances $\text{cov}[e_t, e_s] = 0$ for t≠s. Using first-order terms from a Taylor expansion, Priestly suggests a model

$$y_t = \mu(z_{t-j}) - \sum_{j+1}^{p} \phi_j(z_{t-1}) y_{t-j} + \varepsilon_t + \sum_{j+1}^{q} \theta_j(z_{t-1}) \varepsilon_{t-j} \qquad \textbf{(2.1.4)}$$

where z_t is a vector of "state-variables", which are themselves functions of the contents of I_t. The coefficients $\mu(z_t)$, $\phi_j(z_t)$ and $?_j(z_t)$ will change through time as the state variables evolve and the model is completed by specifying how the state variables are generated.

iii) A related class is the "**doubly stochastic models**" introduced by Tjostheim (1986). An example is the relationship $y_t = \sum \theta_{jt} y_{t-j} + \varepsilon_t$ where θ_{jt} is a stochastic process for each j, usually taken to be independent of ε_t. Various alternatives arise from specifying different generating mechanism for the θ_{jt}. Particular special cases are the random coefficient models, in which $\theta_{jt} = m_j + e_{jt}$, where the m_j are constant and e_{jt} are i.i.d series, as studied by Nichols and Quinn (1982).

The three general classes of model just considered, *the state affine, state dependent* and *doubly stochastic* are all essentially linear with different forms of time-varying parameters, through which nonlinearities are introduced. The simple, specific models, such as nonlinear autoregressive and threshold models are all special cases of these models or extended versions of them.

However, an important, simple model is the nonlinear autoregressive of order one, where y_t has the form $y_t = f(y_{t-1}) + \varepsilon_t$ ***(2.1.5)***
where ε_t is zero mean i.i.d. A Markov process has the property that the conditional distribution of y_{t+1}given all $y_{t-j}, j \ge 0$ is the same as the conditional distribution of y_{t+1} given just y_t. Thus, all the information about the future of the series is contained in its most recent value. Hence, the above process, *(2.1.5)* is a Markov process. Another alternative form is the deterministic process

$$\bar{y}_t = g(\bar{y}_{t-1})\bar{y}_{t-1} \qquad \textbf{(2.1.6)}$$

Notice that if $\bar{y}_s = 0$ then $\bar{y}_t = 0$ for t > s. An obvious generalisation to p lag is the model

$$y_t = f(y_{t-j}, j = 1,....., p) + \varepsilon_t \qquad \textbf{(2.1.7)}$$

which is not Markov if p>1. Finally, models *(2.1.5)-(2.1.7)* have in common that they are all ***univariate*** models.

2.2 Smooth Transition Autoregressive Models

Many parts of economic theory include the idea that the economy behaves differently if some variable lies in one region rather than in another. These models can be grouped under the heading 'regime-switching' theories. This yields a variety of non-linear time-series models. A simple single-equation form might be:

$$y_t = \alpha_0 + \alpha_1 x_{t-1} + (\beta_0 + \beta_1 x_{t-1}) F(x_{t-1} - \mu) + e_t \qquad \textbf{(2.2.1)}$$

where e_t is i.i.d. Here, F(x) is a continuous function which may be either even or odd, x_t are explanatory variables, α the time-delay and the explanatory variable x_{t-1} for function F(x) is called the transition variable. The function, F(x) may be an even function with $F(\pm\infty) = 0, F(0) = 1$ and in that case F may equal the density function of a $N(\mu, \sigma^2)$ variable. If F is odd and monotonically increasing, $|x_{t-1} - \mu|$ is large and if $x_{t-1} < \mu$, y_t is effectively generated by the linear model

$$y_t = \alpha_0 + \alpha_1 x_{t-1} + e_t \qquad \textbf{(2.2.2)}$$

If $|x_{t-1} - \mu|$ is large and $x_{t-1} > \mu$, y_t is virtually generated by

$$y_t = (\alpha_0 + \beta_0) + (\alpha_1 + \beta_1) x_{t-1} + e_t \qquad \textbf{(2.2.3)}$$

Bacon and Watts (1971) and Maddala (1977) were early proponents, when F(x) was an odd, monotonically increasing function with $F(-\infty) = 0, F(+\infty) = 1$. Such models are called smooth transition regression (STR) models. If x_{t-1} is replaced by y_{t-1}, *(2.2.1)* is a special case of the Smooth-Transition Autoregressive (STAR) model.

When F is even, y_t is practically generated by *(2.2.3)* whenever $|x_{t-1} - \mu|$ is large and by *(2.2.2)* if $x_{t-1} \approx \mu$. If x_{t-1} is replaced by y_{t-1}, then *(2.2.1)* become univariate and is hence another special case of the STAR model. It is often convenient to assume that the function F is logistic if it is odd and a general situation is then achieved if $F(y_{t-1})$ is replaced by $F(y_{t-d})$ for some d>0, i.e.[4]

$$F(\gamma, c; y_t) = \frac{1}{1 + \exp(-\gamma(y_{t-d} - c))}, \quad \gamma > 0 \qquad \textbf{(2.2.4)}$$

[4] Teräsvirta [1998]

Model (2.2.1) is then called a logistic STR (LSTR) model. A more general model of a **logistic smooth transition autoregressive model** of order p (LSTAR(P) model) with a vector w_t of p explanatory variables, can be written using the vectors

$$\mathbf{w}_t = (1, y_{t-1},..., y_{t-p})', \quad \mathbf{?} = (\theta_0,....., \theta_p)', \quad \mathbf{f} = (\phi_0,....., \phi_p)'$$

and a model

$$y_t = G(z_t, y_t; \psi) + \varepsilon_t = \mathbf{f}'\mathbf{w_t} + (\mathbf{?}'\mathbf{w_t}) * F(\gamma, c; y_{t-d}) + \varepsilon_t \qquad \textbf{(2.2.5)}$$

where

$$F(\gamma, c; y_{t-d}) = (1 + \exp\{-\gamma(y_{t-d} - c)\})^{-1}$$

and ε_t is a zero mean sequence of normal independent variables. The coefficient $\gamma, \gamma > 0$ is the **smoothness parameter** and the scalar c is the location parameter and d is known as the **delay parameter** when $F(x_{t-1})$ is replaced by $F(y_{t-d})$. The variable y_{t-d} is then called the **transition variable** for some d>0 in *model (2.2.4).* [5]

The transition function *(2.2.4)* is a monotonically increasing function of y_{t-d}. The slope parameter γ indicates how rapid the transition from zero to unity is for a function of y_{t-d} and the local parameter c determines where the transition occurs. If $\gamma \to \infty$ in *model (2.2.4),* then *(2.2.1)* becomes a two-regime switching regression model with the switching variable y_{t-d} In this special case, $y_{t-d} = c$ is the switchpoint between the regimes

$$y_t = \alpha_0 + \alpha_1 x_{t-1} + e_t$$

$$y_t = \alpha_0 + \beta_0 + x_{t-1}(\alpha_1 + \beta_1) + e_t$$

The slope or smoothness parameter γ controls the slope of the transition function *(2.2.4).*

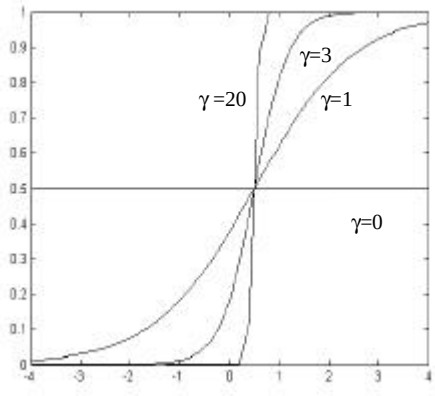

Figure 2.2.1: Graphs of the logistic transition function (2.2.4) for c=0.5

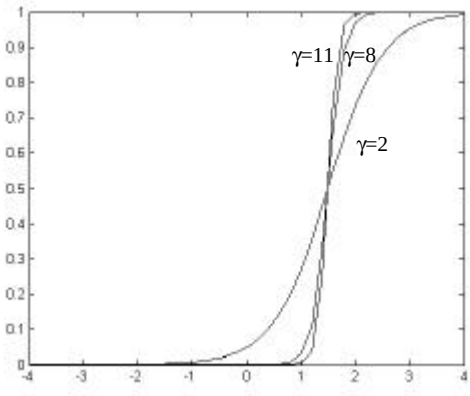

Figure 2.2.2: Graphs of the logistic transition function (2.2.4) for c=1.5

[5] Teräsvirta [1998]

When $\gamma \to \infty$ then *(2.2.4)* becomes a step function and the STAR *model (2.2.1)* becomes a threshold autoregressive (TAR) model with two regimes. [6] Note that if $\gamma \to 0$ the model becomes linear as *(2.2.4)* becomes a constant. This is illustrated in *figure 2.2.1* and *figure 2.2.2* where the transition function increases monotonically from zero to unity with y_{t-d}. As a result, the LSTAR model could be applied for modelling asymmetric business cycles. This is because the dynamics of the model are different in the expansion from the recession.

2.3 Estimation of Smooth Transition Autoregressive models

After specifying the model, the parameters can be estimated. If p is known, the coefficient can be estimated by nonlinear least squares (NLS) or maximum likelihood (ML) and some optimisation procedure. [7] In the case where $\varepsilon_t \sim N(0,\sigma^2)$, both methods are equivalent. Hence the parameter vector ψ of *(2.2.1)* with the logistic function

$$F(\gamma(y_{t-d}-c)) = \frac{1}{1+\exp(-\gamma(y_{t-d}-c))} \quad \text{is estimated as}$$

$$\hat{\psi} = \arg\min_{\psi} Q_T(\psi) = \arg\min_{\psi} \sum_{t=1}^{T} (y_t - G(z_t, x_t; \psi))^2$$

Under some regularity conditions the estimates are consistent and asymptotically normal, that is $\sqrt{T}\left(\hat{\psi}-\psi^*\right) \to N(0,C)$ where ψ^* is the true parameter vector and C is the covariance matrix of estimates. The parameters of the STAR model will hence be estimated by nonlinear least squares, and to do that, a suitable iterative optimisation algorithm is needed. The optimisation problem in nonlinear least squares is conditional on the starting values $y_0, y_{-1}, ..., y_{-p+1}$ and consists of finding the minimum of the criteria

$$q(\boldsymbol{\theta}) = (\mathbf{y} - f(\mathbf{y};\boldsymbol{\theta}))'(\mathbf{y} - f(\mathbf{y};\boldsymbol{\theta})) \qquad \textbf{(2.3.1)}$$

with respect to the parameter vector $\boldsymbol{\theta}$. To find a starting-point for the iteration procedure we approximate *(2.3.1)* with a second-order Taylor expansion about θ_n which yields

$$q(\boldsymbol{\theta}) \cong q(\boldsymbol{\theta}_n) + \mathbf{h}_n'(\boldsymbol{\theta} - \boldsymbol{\theta}_n) + \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_n)'\mathbf{H}_n(\boldsymbol{\theta} - \boldsymbol{\theta}_n) \qquad \textbf{(2.3.2)}$$

where the gradient evaluated at $\boldsymbol{\theta}_n$, $\mathbf{h_n} = \mathbf{h}(\boldsymbol{\theta}_n) = \frac{\partial q(\boldsymbol{\theta}_n)}{\partial \boldsymbol{\theta}}$ and the Hessian evaluated at $\boldsymbol{\theta}_n$ $\mathbf{H}_n = \mathbf{H}(\boldsymbol{\theta}_n) = \frac{\partial^2 q(\boldsymbol{\theta}_n)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}$. The first-order conditions for the minimum i.e $\frac{\partial q(\theta)}{\partial \theta} = 0$ will be obtained by differentiating *(2.3.2)* with respect to $\boldsymbol{\theta}$ and the basis for iteration yields [8]:

[6] Lundbergh and Teräsvirta [2002]

[7] Medeiros and Veiga [2000]

[8] Granger and Terävirta [1993]

$$\mathbf{h}_n + \mathbf{H}_n(\boldsymbol{?} - \boldsymbol{?}_n) = \mathbf{0} \qquad \textit{(2.3.3)}$$

If we have fixed $\boldsymbol{?}_n$ and know how to compute $\mathbf{H}_n$ and $\mathbf{h}_n$, equation *(2.3.3)* yields the search direction for the next value of $\boldsymbol{?} : \boldsymbol{?} = \boldsymbol{?}_{n+1}$ which is obtained from $\boldsymbol{?} = \boldsymbol{?}_n + k_n \mathbf{H}_n^{-1}\mathbf{h}_n$ where k_n is the step-length. The value forms the starting-point of the next iteration. This is the Newton-Raphson method. In general the optimisation algorithm is very sensitive to the choice of the starting values of the parameters. The implemented program is estimating all the models according to the Levenberg-Marquardt algorithm with cubic interpolation linear search. Concerning the selection of the starting values, the following algorithm is used:
Rewrite *model (2.2.5)* as $y = W\lambda + \varepsilon,$ where

$$y' = [y_1, y_2, ..., y_T], e' = [e_1, e_2, ..., e_T], \lambda' = [\varphi, \boldsymbol{?}], \ \tilde{\mathbf{w}}_t = F(\gamma(y_{t-d} - c))\mathbf{w}_t$$

$$W = \begin{pmatrix} \mathbf{w}_1' & \tilde{\mathbf{w}}_1' \\ \cdots & \cdots \\ \cdots & \cdots \\ \cdots & \cdots \\ \mathbf{w}_T' & \tilde{\mathbf{w}}_T' \end{pmatrix}$$

Once γ, **d** and c have been determined, the parameter vector λ can be estimated by $\hat{\lambda} = (W'W)^{-1}Wy$

Estimating the slope parameter γ in the transition function causes sometimes problems. The reason is that the slope parameter is not scale invariant. This makes it difficult to find a good starting-value. While all the other iterative parameter estimates converge rapidly, the slope parameter $\hat{\gamma}$ converges very slowly. This may be the case when the true parameter is relatively large. This is because a large set of γ-values yields almost the same F. The transition functions corresponding to these γ-values deviate noticeably from each other only in a small neighbourhood of the transition value c. When γ is large, rescaling it becomes important. A method may be obtained by rescaling the argument of F by dividing by $\hat{\sigma}(x_t)$, which is the standard deviation of the transition variable. A reasonable starting value for iterative nonlinear least squares estimation would then be $\gamma = 1$. Rescaling the argument will make the choice of a starting-value for γ more easier. Moreover, when γ is large then 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. However this uncertainty is reflected in the estimated standard deviation of γ, which tends to be large for large values of γ. A workable solution would therefore be to rescale the parameter before doing estimation. [9]

The selection of the parameters ?, d and c is divided in the following steps:
1. Draw K possible values for d such as $\delta_1 \in (0,1)$ and $\delta_j \in [-1,1]$, j = 2,…,q and call them v_k, k=1,…,K. The values for δ_1 will be from a uniform random distribution over the interval (0,1] and for δ_j , j = 2,…,q from a uniform random distribution over the interval [-1,1].

[9] Teräsvirta[1994]

2. For k = 1,…,K:

(a) Normalize each vector v_k and compute the projection $x_{/v_k}$ of x ($x = [x_1, ..., x_T]$) in the direction of v_k.

(b) Compute the median of $x_{/v_k}$ and call it b_k.

(c) Draw a grid of N possible values for the slope and call them γ_n, n=1,….,N.

3. For k = 1,…,K and n = 1,…,N, set ? = γ_n , d = v_k and c = b_k and compute the value of . $Q_T(\psi)$. The values of the parameters will be selected such that they minimize the objective function and we call them δ^* ,v_k^* and b^*.

4. Set ? = γ^*, d = v_k^* and c = b^* and use them as starting-values for the Levenberg-Marquardt algorithm..

3 Forecasting

This chapter will begin by defining a suitable measure of forecast accuracy. The measure will later on be used to compare one forecast method with another. Section 3.2 provide a method of forecasting with an alternative model to the Logistic STAR model. An application of estimation and forecasting the alternative model is then described in section 3.3. The predictions will be tested against predictions of the Logistic STAR model. Particular Logistic STAR models have difficulties in doing predictions, especially for several steps ahead. Section 3.4 will describe this and a method of producing one-, two- and three-steps forecasts. An application of the prediction methods for Logistic STAR models is presented in section 3.5. The comparison between results and the accuracy of the predictions are finally presented in section 3.6.

3.1 Measuring forecast accuracy

A forecast might be judged "successful" if it is close to the outcome, but that judgment may also depend on how "close" it is measured. One of the key feature of measuring forecast accuracy is the separation of the data into separate parts, one of which we refer as the hold-out sample, which is not used at all during the estimation process. Instead it will be used to evaluate performance after fitting is complete. When making forecasting comparisons, the time origin of the forecast may be progressively shifted through the hold-out sample. This provides a set of rolling forecasts from which we calculate the measure. To simulate this, the selection of forecast method will require repeated productions of forecast steps. The general procedure is as follows:

1. T+H known data points are split into one T period for estimation and one H period for accuracy and forecasting.

2. At each T+h, we do a forecast on T+h+1, T+h+2 and T+h+3 for h=0,1,2,……,H-1

Thus, the procedure will generate a set of H three-step-ahead forecasts which will be compared with the hold-out sample. Defining a suitable error measure based on the hold-out sample is as follows:

$$\mathrm{RMSE}_j = \sqrt{\frac{1}{H}\sum_{h=0}^{H-1}\left(\hat{y}_{T+h+j|T+h} - y_{T+h+j}\right)^2}, \quad j = 1,\ 2,\ 3 \qquad \textbf{(Root Mean Square Error)}$$

where j indexes the forecast step. Hence, the forecast error is the difference between the true outcome y_{T+h+j} and its forecast $\hat{y}_{T+h+j}$.

3.2 Predictions with an alternative linear model

One of the purposes with LSTAR models is to give an adequate characterization of nonlinear features in the data. This characterization is then used to produce forecasts. However, it is often of particular interest to find out how successful the estimated model is in this respect. One way to do this is to form an alternative linear model. When $\gamma = 0$ in the transition function *(2.2.4)*, the STAR model becomes linear and the transition function becomes a constant. A possible null hypothesis is $H_0 : \gamma = 0$ and the alternative $H_1 : \gamma > 0$. Hence, we can specify a linear relationship. Let us define the additive model as

$$y_t = \sum_{j=1}^{p} \phi_j y_{t-j} + u_t \qquad \textit{(3.2.1)}$$

for all t, where $\{u_t\}$ is a sequence of uncorrelated random variables with constant mean of zero and constant variance[10] and $\{\phi_j\}$ is a sequence of constants with $\sum_{j=-\infty}^{\infty} |\phi_j| < \infty$. The class of linear time series models provides a general framework for studying stationary processes. The time series $\{y_t\}$ is a **linear process** and when u_t is uncorrelated with y_s for s<t we define this as a **p order autoregressive process** or **AR(p) process**.[11]

Estimating parameters of the linear autoregressive model (3.2.1*)* is carried out by least-squared method which is equivalent to the conditional maximum likelihood estimation in the case of normal errors.[11] Now we suppose that the observed time series $y_1, y_2, \ldots, y_T$ is a realization of a stationary process admitting an autoregressive representation:

$$\sum_{j=0}^{p} a_j y_{t-j} = u_t, \; a_0 = 1 \qquad \textit{(3.2.2)}$$

where $\{u_t\}$ is a sequence of independent random variables, each with mean 0, variance σ^2 and the a_j are absolutely summable real coefficients. A new autoregressive model could be selected and fitted for each step of prediction, h, by ordinary linear least squares, OLS, regression procedure in which y_{t+h} is regressed on $y_t, y_{t-1}, .., y_{t-k+1}$.

[10] A process of this kind is called *white noise* (WN) and $\{u_t\}$ will be denoted by $u_t \sim WN(0, \sigma^2)$

[11] Harvey [1993]

3.3 Estimation and forecasting with autoregressive models

The basic nonlinear model considered is a logistical smooth transition autoregressive model of order two and the model which is used to generate data is defined as follows[12] :

$$y_t = 1.8y_{t-1} - 1.06y_{t-2} + (0.02 - 0.9y_{t-1} + 0.795y_{t-2})F(y_{t-d}) + \varepsilon_t$$

$$F(y_{t-d}) = (1 + \exp\{-\gamma(y_{t-d} - 0.02)\})^{-1} \qquad \textit{(3.3.1)}$$

The time series in *figure 3.3.1* is a realization of 100 observations of the LSTAR(2) *model (3.3.1)* with the delay parameter d = 1, slope parameter $\gamma = 100$ and ε_t ~NID(0,1).

.

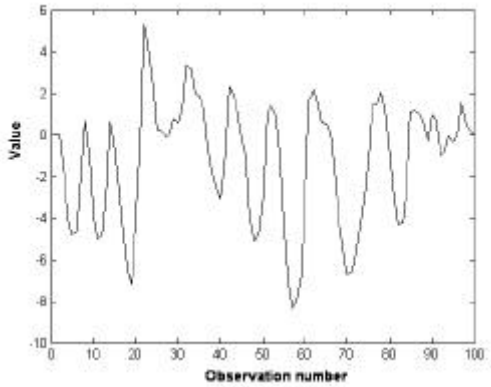

Figure 3.3.1

In most circumstances there are several possible models that can be used to make forecasts. One way is to produce a related linear forecast instead of a nonlinear forecast. When $\gamma = 0$ in *model (2.2.4)*, then the LSTAR(2) model becomes an autoregressive model of order 2. In section 2.2 *model (2.2.5)* is called the logistic STAR (LSTAR) model. Using the least-square method we can estimate the parameters of the linear model which is related to the time series in *figure 3.3.1*.

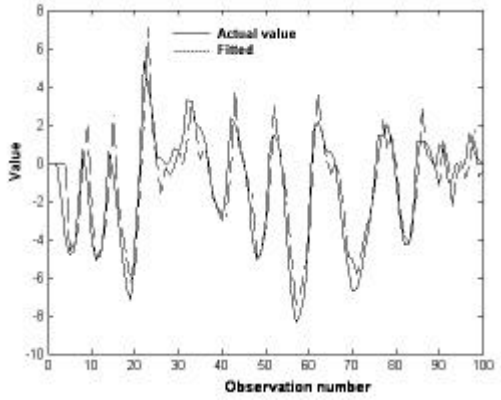

Figure 3.3.2

The figure shows the fitted autoregressive model of order 2 and the generated data from a LSTAR(2) model.

[12] This LSTAR model has been simulated in Teräsvirta [1994]

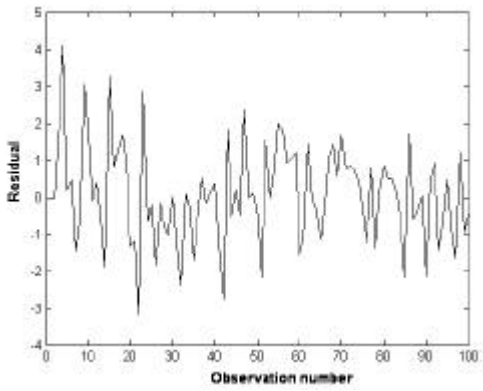

Figure 3.3.3

The residuals of the generated LSTAR(2) model and the linear AR(2) model are graphed in figure 3.3.3. A comparison between the graphs indicates that they are quite similar.

The forecasts can then be expressed as a weighted linear combination of past and current innovations with j=1, 2, 3:

$$E[y_{T+j} \mid y_{T+j-1}, y_{T+j-2}] = -0.347 + 1.419y_{T+j-1} - 0.734y_{T+j-2} \qquad \textbf{(3.3.2)}$$

To investigate the effects of excluding lags and forecasting directly without iteration, we consider the third-order autoregressive model. Estimating the parameters yields the two-step forecast

$$\hat{y}_{T+2|T,T-1} = -0.708 + 1.212y_T - 1.059y_{T-1} \qquad \textbf{(3.3.3)}$$

In this case, the two-step forecast in period T will be depending on lagged values y_{T-1} and y_T This will be compared with the nonlinear case where forecasting two-step ahead is not as easy. Mainly because the exact forecast has to be obtained by numerical integration where multiple integration would be encountered for longer forecast horizons than two periods.

In a similar way, instead of model *(3.3.2)* the forecast made at T for three-step ahead may be obtained as

$$\hat{y}_{T+3|T,T-1} = -1.079 + 0.706y_T - 0.999y_{T-1} \qquad \textbf{(3.3.4)}$$

Note that the expressions for one-, two- and three step ahead don´t depend on the previous forecast as in model (3.3.2). Thus, both model *(3.3.3)* and *(3.3.4)* will be depending on lags y_T and y_{T-1} instead of the previous forecast and hence period T is the last available information.

For comparison, we compute the root mean square error. The results are summarized in the following table:

T / AR	*100*	*200*	*400*
Forecast 1 step ahead			
Model (3.3.2)	1.163	1.187	1.201
Forecast 2 step ahead			
Model (3.3.2)	1.202	1.225	1.245
Model (3.3.3)	1.301	1.312	1.322
Forecast 3 step ahead			
Model (3.3.2)	1.243	1.267	1.281
Model (3.3.4)	1.328	1.332	1.337

Table 3.3.1: RMSE of the estimated models with forecast period, H=100 and estimation periods T=100, 200 and 400

It is seen that RMSE of model (3.3.2) with estimation period T=100 are less than for periods T=200 and 400. Also forecast two- and three-step ahead have a higher value of RMSE than one-step ahead. In the first column, we observe that step one yields the smallest value and step three yields the highest value of RMSE for *model (3.3.2).* In contrast of model (3.3.2), we observe that model (3.3.3) yields a higer value of RMSE than model (3.3.2) for two-step ahead. When forecasting three-step ahead, model (3.3.4) also yields a higher value of root mean square error.

3.4 Forecasting with general nonlinear models

One of the major uses for time-series models is to produce forecasts and nonlinear models have particular difficulties in doing this, especially for several steps ahead. To illustrate the methods available for forecasting we use the model

$$y_t = g(x_{t-1}) + \varepsilon_t \qquad \textbf{(3.4.1)}$$

Using a least-squares criterion, the optimal one-step forecast is

$$f_{t,1}^{y} = E[y_{t+1}|I_t] = g(x_t) \qquad \textbf{(3.4.2)}$$

where $I_t : x_{t-j}, y_{t-j}, j \geq 1$ is the information set available at time t. Thus if, one knows g(.) or has an acceptable approximation for it, one-step forecasts can be achieved with no difficulty. The two-step case is not as easy. The optimum two-step forecast is

$$f_{t,2}^{y} = E[y_{t+2}|I_t] = E[g(x_{t+1})|I_t] \qquad \textbf{(3.4.3)}$$

As x_{t+1} is not usually known at time t, it's necessary to specify the generating mechanism for x_t. Hence, we suppose that a good approximation to this mechanism is available. Let this be the model $x_t = \alpha x_{t-1} + e_t$. This gives a one-step OLS forecast $f^x_{t,1} = \alpha x_t$. The two-step forecast using *(3.3.3)* would then be in the form

$$f^y_{t,2} = E[g(f^x_{t,1} + e_{t+1})|I_t] \qquad \textbf{(3.4.4)}$$

To compute this forecast we use **Monte Carlo technique**: $fm^y_{t,2} = \frac{1}{N}\sum_{j=1}^{N} g(f^x_{t,1} + z_j)$ ***(3.4.5)***

Compare this with the **exact solution**: $fe^y_{t,2} = \int_{-\infty}^{\infty} g(f^x_{t,1} + z)d\Phi(z)$ ***(3.4.6)***

For N large enough in the Monte Carlo technique, *(3.4.5)* and *(3.4.6)* should be virtually identical.[13] A **naïve approach** would be, if the presence of e_{t+1} in *(3.4.4)* is ignored by putting its value to zero. The different forecasts are quite different and have advantages and disadvantages.

The naïve forecast is easy to use but will usually be biased because generally $E[f(\varepsilon)] \neq f(E[\varepsilon])$ The exact method can be too difficult to implement. The Monte Carlo forecast can be biased if an incorrect distribution D is selected, but otherwise should be the best technique. Clearly,

Monte Carlo technique can be used for multi-step forecast. In fact the exact three-step forecast has the form

$$f^y_{t,3} = E[g(g(f^x_{t,z} + z_{t+2}) + z_{t+3})|I_t]$$

The naïve forecast just ignores e, but the exact forecast now involves a double integral, and the Monte Carlo requires draws from a bivariate distribution, but with independent components. The three-step forecast is

$$f^y_{t,3} = \frac{1}{n-2}\sum_{j=1}^{n-2} g(g(f^x_{t,2} + z_j) + z_{j+1})$$

As before, if D is well known, the exact and Monte Carlo methods will be the best.

[13] Granger and Teräsvirta [1993]

3.5 Estimation and forecasting with LSTAR models

Now we consider the particular logistical STAR model. The random numbers generated are the same as in section 3.3. As we know, the time series in *figure 3.3.1* is a realization from a LSTAR model of order two with 100 observations for d = 1 and $\gamma = 100$. Estimating the parameters for the LSTAR(2) model yields [14] :

$$y_t = 1.82y_{t-1} - 0.95y_{t-2} + (0.01 - 0.72y_{t-1} + 0.757y_{t-2})F(y_{t-1}) + \hat{e}_t$$

$$F(y_{t-1}) = (1 + \exp\{-98.567(y_{t-1} - 1.42)\})^{-1} \qquad \textbf{(3.5.1)}$$

To illustrate the properties of the estimated LSTAR(2) model, consider the following figure which displays the generated data and the estimated *model (3.5.1).* The difference between the graphs is quite small. Note that the peaks of the generated data are sometimes higher than the peaks of the estimated model.

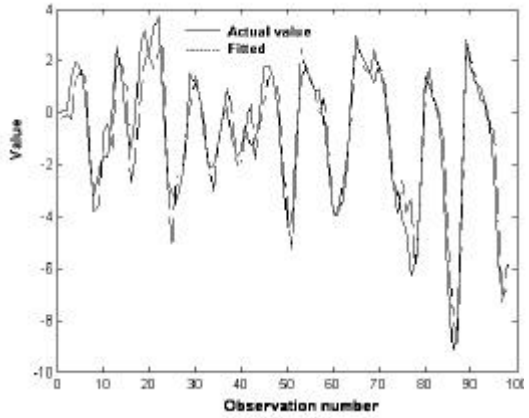

Figure 3.5.1

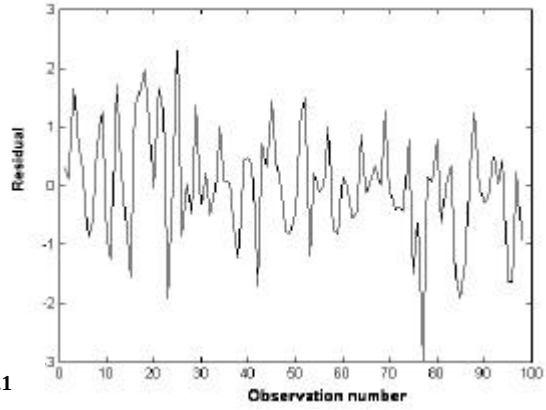

Figure 3.5.2

As discussed before (section 3.4), the dynamics of the nonlinear *model (3.5.1)* will be used to estimate one-, two- and three-step forecasts. At time T, let the model produce one-, two- and three-step forecast given all the information set available at time, T. For prediction two- and three-step ahead we apply the Monte Carlo procedure according to section 3.4. More generally, forecasts with estimation period T will be computed recursively as

$$f^{y}_{T+h+1|T+h} = E[y_{T+h+1}|I_{T+h}]$$

$$f^{y}_{T+h+2|T+h} = E[y_{T+h+2}|I_{T+h}]$$

$$f^{y}_{T+h+3|T+h} = E[y_{T+h+3}|I_{T+h}]$$

where $I_T : y_{T-j}, j \geq 1$ is the information set available at time T and the forecast period is given by h = 0,……, H =1

[14] Note: The problem of selecting reasonable starting values for ? and c is discussed in section 2.3

Having done this, we compute the RMSE. This gives us an idea how much the available information affects the forecasts accuracy.

T *LSTAR(2)*	*100*	*200*	*400*	*Forecast step*
Model 3.5.1:				
RMSE	1.171	1.185	1.208	1
	1.315	1.341	1.358	2
	1.342	1.382	1.402	3

Table 3.5.1: RMSE for the estimated LSTAR(2) model 3.5.1 with estimation periods, T=100, 200, 400 and forecast period, H=100.

In the first column, we observe that one-step ahead yields the smallest value of RMSE and three-step ahead yields the highest value of RMSE. The results also show that RMSE with estimation period T=100 are less than period T=200 and 400.

3.6 Comparison between prediction techniques

Calculations of predictable uncertainty serve important roles, both statistically and economically. Depending upon the degree of forecast uncertainty, forecasts may range from being highly informative to utterly useless for the tasks at hand. A measure of forecast uncertainty provides an assessment of the expected or predicted uncertainty of forecast error and helping to give a picture of the excepted range of likely outcomes.[15] We now move on to our comparisons between the forecasts from a LSTAR model to those from a linear one.

T	AR Model			LSTAR Model	Forecast Step
	3.3.2	3.3.3	3.3.4	3.5.1	
100	1.163	-	-	1.171	1
100	1.202	1.301	-	1.315	2
100	1.243	-	1.328	1.342	3
200	1.187	-	-	1.185	1
200	1.225	1.312	-	1.341	2
200	1.267	-	1.332	1.382	3
400	1.201	-	-	1.208	1
400	1.245	1.322	-	1.358	2
400	1.281	-	1.337	1.402	3

Table 3.6.1: RMSE of the estimated models with forecast period, H=100 and estimation periods T=100, 200 and 400

Root mean square error (RMSE) of the LSTAR model is represented with the related linear autoregressive model in *table 3.6.1*. The results are mixed. In some cases the STAR model yield slightly better one-step ahead forcasts than the linear model. In other cases (T=100 and T=400) the situation is the reverse.

[15] Ericsson [2002]

4 Summary and conclusions

In this work the emphasis is on showing how LSTAR models can be applied for forecasting in time series econometrics. It is demonstrated how the estimated model is carried out in a systematic fashion through a generated LSTAR model. The basic nonlinear model considered is a logistic smooth transition autoregressive model of order two. When the smoothness parameter equals zero, then the LSTAR(2) model becomes an autoregressive model of order two. The parameters are estimated by using the least-square method. The forecasts can then be expressed as a weighted linear combination of past and current innovations. Third- and fourth-order autoregressive models are also considered to investigate the effects of excluding lags and forecasting directly without iteration. It is seen that the iterative linear method (model 3.3.2) are more accurate than the direct method (model 3.3.3 and 3.3.4).

Also the iterative method (model 3.3.2) has the lowest root mean square error for estimation period T=100 than the other estimation periods (T=200 and 400). Forecasts from LSTAR are obtained by using Monte Carlo technique. The one-step ahead forecast yields the smallest value of root mean square error and three-step ahead yields the highest value of root mean square error. The results also show that RMSE for period T=100 are smaller than period T=200 and 400. In addition, the simulation study suggests that the linear forecasts are more accurate than those from a LSTAR model.

This may appear surprising at first because data were generated from the LSTAR model at the beginning. However, the linear AR-model 3.3.2 is a weighted linear combination of past and current innovations and it consider the effects of including lags and forecasting with iteration. In contrast, at time T the LSTAR model 3.5.1 is producing one-, two- and three-step forecast given all the information set available at time, T. For comparing the accuracy of forecasts the linear AR-model 3.3.3 and 3.3.4 will not depend on the previous forecast as in model 3.3.2. But even in this case RMSE is lower than the corresponding RMSE for LSTAR-model 3.5.1. As we see in section 3.3 and 3.5 the estimation of the parameters for the linear AR-model and LSTAR model is sufficiently accurate. An argument could then be that the lower value of RMSE for LSTAR-model 3.5.1 depends on the forecast method. The Monte Carlo forcasts are biased if an incorrect distribution D is selected (section 3.4). Another method to produce forecasts is using Bootstrap. This method use the values of the residuals observed over the sample period instead of the random numbers drawn from a normal distribution which is used in this essay. As we see from table 3.6.1 in section 3.6 the forecast accuracy will be lower as the estimation period increase at the same time that the estimation of the LSTAR is sufficiently accuracy. A reason to this behaviour may be explained by the selection of the normal distribution when producing forecasts with Monte Carlo method.

The dynamics will also be of interesting when forecasting with nonlinear models. The estimate of γ is high (i.e. 98.586) and this implies that a change between recessionary and expansionary regimes will be quite rapidly. Furthermore, the "lower regime" (when $F(y_{t-d}) = 0$ in model 3.3.1) of the LSTAR process is such that the roots of the characteristics polynomial are a complex pair with a modulus such that the regime is explosive. On the other hand, the "upper regime" (when $F(y_{t-d}) = 1$ in model 3.3.1) corresponding also a complex pair but with a modulus such that this regime is not explosive. In fact, the oscillations are very weak (see further Teräsvirta [1994] for a discussion of the long-term behaviour of the model). In addition, the value of $\hat{\gamma}$ indicates that the transition from one regime to the other is very quick so that the model is very similar to a threshold AR model. It is a model which the parameters of the linear model change through time due to a switching rule, which also could depend on an earlier value of the series.

Hence, there is no guarantee that the estimated LSTAR model will produce superior forecasts. A necessary condition for that to happen would seem to be that the forecasting period contains "nonlinear features". For instance a nonlinear model may be expected to be superior to a linear one when the forecasting period contains the aftermath of a large negative shock. If that is not the case a linear autoregressive model is likely to perform as well as a nonlinear one.

5 References

Bacon, D. W. and Watts, D. G. [1971], Estimating the transition between two intersecting straight lines, Biometrika, 58: 525-34

Ericsson [2002], Predictable Uncertainty in Economic Forecasting, in: M.P. Clements and D.F. Hendry, eds. A companion to economic forecasting, 19-45. Oxford: Blackwell.

Granger, C.W.J. and T. Teräsvirta [1993], Modelling nonlinear economic relationships, Oxford: Oxford University Press.

Guegan, D. [1987], Different representations for bilinear models, Journal of Time Series Analysis, 8: 389-408

Harvey, A. C. [1993], Time Series Models, Great Britain: Hartnolls, Bodmin, Cornwall

Lundbergh, S. and T. Teräsvirta [2002], Forecasting with smooth transition autoregressive models, in: M.P. Clements and D.F. Hendry, eds. A companion to economic forecasting, 485-509. Oxford: Blackwell.

Maddala, G. S. [1977], Econometrics, New York, McGraw-Hill

Nicholls, D. F. and Quinn, B. G. [1982], Random Coefficient Autoregressive Models: An introduction, Lecture Notes in Statistics, No. 11, New York, Springer-Verlag

Priestley, M. B. [1980], State-dependent models: a general approach to non-linear time series analysis, Journal of Time Series Analysis, 1: 47-71

Sontag, E. D. [1979], Realization theory of discrete time nonlinear systems I: the bounded case, IEEE Transactions Circuits and Systems, 26: 342-56

Teräsvirta, T. [1994], Specification, Estimation and Evaluation of Smooth Transition Autoregressive Models, Journal of the American Statistical Association, Vol. 89, No. 425

Teräsvirta, T. [1998], Modeling economic relationships with smooth transition regressions, in: A. Ullah and D.E.A. Giles, eds. Handbook of applied economic statistics, 507-552. New York: Dekker.

Tjostheim, D. [1986], Estimation in nonlinear time series models, Stochastic Processes and their Applications, 21: 251-73