Smooth Transition Autoregressive Models - A Survey of Recent Developments*

Dick van Dijk†
Econometric Institute
Erasmus University Rotterdam

Timo Teräsvirta‡
Department of Economic Statistics
Stockholm School of Economics

Philip Hans Franses§
Econometric Institute
Erasmus University Rotterdam

Econometric Institute Research Report EI2000-23/A

June 9, 2000

Abstract

This paper surveys recent developments related to the smooth transition autoregressive [STAR] time series model and several of its variants. We put emphasis on new methods for testing for STAR nonlinearity, model evaluation, and forecasting. Several useful extensions of the basic STAR model, which concern multiple regimes, time-varying nonlinear properties, and models for vector time series, are also reviewed.

Keywords: Regime-switching models, time series model specification, model evaluation, forecasting, impulse response analysis

JEL Classification: C22, C52

*This paper is based on Chapter 2 of the Ph.D thesis of the first author. Part of this paper was written while the first author was visiting the Department of Economics, University of Western Australia and the Department of Economic Statistics, Stockholm School of Economics. The hospitality and stimulating research environment provided are gratefully acknowledged. Financial support from the Jan Wallander's and Tom Hedelius' Foundation for Social Research, Contract No. J99/37, is gratefully acknowledged.

†Erasmus University Rotterdam, Econometric Institute, P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands, `email: djvandijk@few.eur.nl` (corresponding author)

‡Department of Economic Statistics, Stockholm School of Economics, Box 6501, S-113 83 Stockholm, Sweden, `email: Timo.Terasvirta@hhs.se`

§Erasmus University Rotterdam, Econometric Institute, P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands, `email: franses@few.eur.nl`

1 Introduction

Over the last fifteen years, say, the interest in nonlinear time series models has been steadily increasing. In applications to economic time series, models which allow for state-dependent or regime-switching behaviour have been most popular. This paper provides a survey of recent developments related to one of these regime-switching models, that is, the smooth transition model, where we mainly consider variants of the smooth transition autoregressive [STAR] time series model. The discussion is aimed towards practitioners and, therefore, is organized around the empirical modelling cycle for STAR models devised by Teräsvirta (1994). The cycle allows modelling time series with STAR models in an organized fashion. It consists of specification, estimation and evaluation stages and, thus, is similar to the modelling cycle for linear models of Box and Jenkins (1970).

Previous reviews of the smooth transition model include Granger and Teräsvirta (1993), Teräsvirta (1998) and Potter (1999). Compared with these surveys, we put more emphasis on aspects such as model evaluation by means of out-of-sample forecasting and impulse response analysis, and the influence of possible outliers on the analysis of smooth transition type nonlinearity. We also discuss recently introduced extensions of the basic smooth transition model.

The plan of this paper is as follows. In Section 2, representation of the smooth transition model and interpretation of the model parameters are discussed. Three extensions of the basic model, involving multiple regimes, time-varying smooth transition nonlinearity and smooth transition models for vector time series, are discussed in Section 3. Hypothesis testing in the STAR framework is reviewed in Section 4. This concerns both testing linearity against smooth transition nonlinearity and misspecification testing in smooth transition models. The empirical modelling cycle for smooth transition models is outlined in Section 5. The modelling cycle consists of specification, estimation and evaluation stages, which are discussed in more detail in turn. Section 5.1 deals with specification. Estimation of the model parameters is the subject of Section 5.2. Evaluation of estimated smooth transition models by means of diagnostic tests, local spectra and impulse response analysis is addressed in Sections 5.3. Out-of-sample forecasting with smooth transition models is discussed in Section 6. In Section 7, we analyze a monthly US unemployment rate series to illustrate the various elements of the modelling cycle. Finally, Section 8 concludes.

2 Representation of the basic STAR model

The smooth transition autoregressive [STAR] model for a univariate time series y_t, which is observed at $t = 1-p, 1-(p-1), \ldots, -1, 0, 1, \ldots, T-1, T$, is given by

$$\begin{aligned} y_t = {} & (\phi_{1,0} + \phi_{1,1} y_{t-1} + \cdots + \phi_{1,p} y_{t-p}) \left(1 - G(s_t; \gamma, c)\right) \\ & + (\phi_{2,0} + \phi_{2,1} y_{t-1} + \cdots + \phi_{2,p} y_{t-p}) \, G(s_t; \gamma, c) + \varepsilon_t, \qquad t = 1, \ldots, T, \end{aligned} \tag{1}$$

or

$$y_t = \phi_1' x_t (1 - G(s_t; \gamma, c)) + \phi_2' x_t G(s_t; \gamma, c) + \varepsilon_t, \tag{2}$$

where $x_t = (1, \tilde{x}_t')'$ with $\tilde{x}_t = (y_{t-1}, \ldots, y_{t-p})'$ and $\phi_i = (\phi_{i,0}, \phi_{i,1}, \ldots, \phi_{i,p})'$, $i = 1, 2$. It is straightforward to extend the model to allow for exogenous variables $z_{1t}, \ldots, z_{kt}$ as additional regressors. The resultant smooth transition regression [STR] model is discussed at length in Teräsvirta (1998). The ε_t's are assumed to be a martingale difference sequence with respect to the history of the time series up to time $t-1$, which is denoted as $\Omega_{t-1} = \{y_{t-1}, y_{t-2}, \ldots, y_{1-(p-1)}, y_{1-p}\}$, that is, $\mathrm{E}[\varepsilon_t | \Omega_{t-1}] = 0$. For simplicity, we also assume that the conditional variance of ε_t is constant, $\mathrm{E}[\varepsilon_t^2 | \Omega_{t-1}] = \sigma^2$. An extension of the STAR model which allows for (possibly asymmetric) autoregressive conditional heteroscedasticity [ARCH] is considered in Lundbergh and Teräsvirta (1998).

The transition function $G(s_t; \gamma, c)$ is a continuous function that is bounded between 0 to 1. In the STAR model as discussed in Teräsvirta (1994), the transition variable s_t is assumed to be a lagged endogenous variable, that is, $s_t = y_{t-d}$ for certain integer $d > 0$. We do not make this assumption here. Thus, the transition variable can also be an exogenous variable ($s_t = z_t$), or a (possibly nonlinear) function of lagged endogenous variables ($s_t = h(\tilde{x}_t; \alpha)$ for some function h, which depends on the $(q \times 1)$ parameter vector α). Finally, the transition variable can be a (function of a) linear time trend ($s_t = t$), which gives rise to a model with smoothly changing parameters, see Lin and Teräsvirta (1994).

Two interpretations of the STAR model are possible. On the one hand, the STAR model be thought of as a regime-switching model that allows for two regimes, associated with the extreme values of the transition function, $G(s_t; \gamma, c) = 0$ and $G(s_t; \gamma, c) = 1$, where the transition from one regime to the other is smooth. On the other hand, the STAR model can be said to allow for a 'continuum' of regimes, each associated with a different value of $G(s_t; \gamma, c)$ between 0 and 1. In this paper we will use the 'two-regime' interpretation.

The regime that occurs at time t can be determined by the observable variable s_t and the associated value of $G(s_t; \gamma, c)$. Different choices for the transition function $G(s_t; \gamma, c)$ give rise to different types of regime-switching behaviour. A popular choice for $G(s_t; \gamma, c)$

is the first-order logistic function

$$G(s_t; \gamma, c) = (1 + \exp\{-\gamma(s_t - c)\})^{-1}, \qquad \gamma > 0, \tag{3}$$

and the resultant model is called the logistic STAR [LSTAR] model. The parameter c in (3) can be interpreted as the threshold between the two regimes, in the sense that the logistic function changes monotonically from 0 to 1 as s_t increases, and $G(c; \gamma, c) = .5$. The parameter γ determines the smoothness of the change in the value of the logistic function and, thus, the smoothness of the transition from one regime to the other. As γ becomes very large, the change of $G(s_t; \gamma, c)$ from 0 to 1 becomes almost instantaneous at $s_t = c$ and, consequently, the logistic function $G(s_t; \gamma, c)$ approaches the indicator function $I[s_t > c]$, defined as $I[A] = 1$ if A is true and $I[A] = 0$ otherwise. Hence, the LSTAR model (1) with (3) nests a two-regime threshold autoregressive [TAR] model as a special case. In case $s_t = y_{t-d}$, this model is called a self-exciting TAR [SETAR] model. An extensive discussion of (SE)TAR models can be found in Tong (1990). When $\gamma \to 0$, the logistic function becomes equal to a constant (equal to 0.5) and when $\gamma = 0$, the LSTAR model reduces to a linear model.

In the LSTAR model, the two regimes are associated with small and large values of the transition variable s_t relative to c. This type of regime-switching can be convenient for modelling, for example, business cycle asymmetry to distinguish expansions and recessions. If y_t is the growth rate of an output variable, and if the transition variable is the growth rate in the previous period, $s_t = y_{t-1}$, and if $c \approx 0$, the model distinguishes between periods of positive and negative growth, that is, between expansions and contractions. The LSTAR model has been successfully applied by Teräsvirta and Anderson (1992) and Teräsvirta, Tjøstheim and Granger (1994) to characterize the different dynamics of industrial production indexes in a number of OECD countries during expansions and recessions.

In certain applications another type of regime-switching behaviour might be more appropriate. For example, it can be argued that the behaviour of the real exchange rate depends on the size of the deviation from purchasing power parity [PPP]. In particular, the presence of transaction costs, such as costs of transportation and storage of goods, leads to the notion of different regimes in real exchange rates. The profits from commodity arbitrage do not make up for the costs involved in the necessary transactions for small deviations from the equilibrium real exchange rate, which implies the existence of a band around the equilibrium rate in which there is no tendency of the real exchange rate to revert to its equilibrium value. Outside this band, commodity arbitrage becomes profitable, which forces the real exchange rate back towards the band. See Taylor, Peel and Sarno (2000) for a review and discussion of theoretical models that incorporate effects of transaction costs as described above. If regime-switching of this form is to be captured by a STAR model with y_t denoting the real exchange rate and $s_t = y_{t-d}$, it appears more appropriate to specify the transition function such that the regimes are associated with small and large absolute values of s_t. This can be achieved by using, for example, the exponential function

$$G(s_t; \gamma, c) = 1 - \exp\{-\gamma(s_t - c)^2\}, \qquad \gamma > 0. \tag{4}$$

The exponential function has the property that $G(s_t; \gamma, c) \to 1$ both as $s_t \to -\infty$ and $s_t \to \infty$ whereas $G(s_t; \gamma, c) = 0$ for $s_t = c$. The resultant exponential STAR [ESTAR] model has been applied to real exchange rates by Michael, Nobay and Peel (1997) and Taylor, Peel and Sarno (2000) and to real effective exchange rates by Sarantis (1999).

A drawback of the exponential function (4) is that for either $\gamma \to 0$ or $\gamma \to \infty$, the function collapses to a constant (equal to 0 and 1, respectively). Hence, the model becomes linear in both cases and the ESTAR model does not nest a SETAR model as a special case. If this is thought to be desirable, one can instead use the second-order logistic function

$$G(s_t; \gamma, c) = (1 + \exp\{-\gamma(s_t - c_1)(s_t - c_2)\})^{-1}, \qquad c_1 \leq c_2, \gamma > 0, \tag{5}$$

where now $c = (c_1, c_2)'$, as proposed by Jansen and Teräsvirta (1996). In this case, if $\gamma \to 0$, the model becomes linear, whereas if $\gamma \to \infty$ and $c_1 \neq c_2$, the function $G(s_t; \gamma, c)$ is equal to 1 for $s_t < c_1$ and $s_t > c_2$ and equal to 0 in between. Hence, the STAR model with this particular transition function nests a restricted three-regime (SE)TAR model, where the restriction is that the outer regimes are identical. Note that for moderate values of γ, the minimum value of the second-order logistic function, attained for $s_t = (c_1 + c_2)/2$, remains between zero and 1/2, unless $\gamma \to \infty$. In the latter case, the minimum value does equal zero. This has to be kept in mind when interpreting estimates from models with this particular transition function.

Finally, the transition functions (3) and (5) are special cases of the general nth-order logistic function

$$G(s_t; \gamma, c) = (1 + \exp\{-\gamma \prod_{i=1}^{n}(s_t - c_i)\})^{-1}, \qquad c_1 \leq c_2 \leq \ldots \leq c_n, \gamma > 0, \tag{6}$$

which can be used to obtain multiple switches between the two regimes.

3 Recent developments: extensions of the basic STAR model

Recently, several extensions of the basic STAR model as given in (2) have been proposed. Below we discuss extensions which allow for multiple regimes, time-varying properties in conjunction with regime-switching behaviour, and modelling several time series jointly. A model which can be used to describe regime-switching behaviour in both autoregressive dynamics and in seasonal properties is discussed in Franses, de Bruin and van Dijk (2000).

3.1 Multiple regime STAR models

The representation of the STAR model in (2) highlights the basic characteristic of the model, which is that at any given point in time, y_t is determined as a weighted average of two AR models, where the weights assigned to the two models depend on the value taken by the transition function $G(s_t;\gamma,c)$. Hence, it follows that the STAR model cannot accommodate more than two regimes, irrespective of what form the transition function takes. Even though two regimes might be sufficient in many applications, it can be desirable to allow for multiple regimes.

To obtain a STAR model that accommodates more than two regimes, it is useful to distinguish two cases, depending on whether the regimes can be characterized by a single transition variable s_t or by a combination of several variables $s_{1t},\ldots,s_{mt}$, say. In case the prevailing regime is determined by a single variable, one can start with the LSTAR model (2) with (3), rewritten as

$$y_t = \phi_1' x_t + (\phi_2 - \phi_1)' x_t G_1(s_t;\gamma_1,c_1) + \varepsilon_t, \tag{7}$$

where a subscript 1 has been added to the logistic transition function and the parameters contained therein for reasons that will become clear shortly. A three-regime model can be obtained by adding a second nonlinear component to give

$$y_t = \phi_1' x_t + (\phi_2 - \phi_1)' x_t G_1(s_t;\gamma_1,c_1) + (\phi_3 - \phi_2)' x_t G_2(s_t;\gamma_2,c_2) + \varepsilon_t. \tag{8}$$

If it is assumed that $c_1 < c_2$, the autoregressive parameters in this model change smoothly from ϕ_1 via ϕ_2 to ϕ_3 for increasing values of s_t, as first the function G_1 changes from 0 to 1, followed by a similar change of G_2. More generally, one can define a set of $m-1$ smoothness parameters $\gamma_1,\ldots,\gamma_{m-1}$, and a set of $m-1$ location parameters $c_1,\ldots,c_{m-1}$, to arrive at a STAR model with m regimes as

$$y_t = \phi_1' x_t + (\phi_2 - \phi_1)' x_t G_1(s_t) + (\phi_3 - \phi_2)' x_t G_2(s_t) + \cdots + (\phi_m - \phi_{m-1})' x_t G_{m-1}(s_t) + \varepsilon_t, \tag{9}$$

where the $G_j(s_t) \equiv G_j(s_t;\gamma_j,c_j), j=1,\ldots,m-1$, are logistic functions as in (3). In case all smoothness parameters become very large, the STAR model in (9) effectively becomes a SETAR model with m regimes.

Extending the basic STAR model in case the regimes are determined by a combination of different variables is done most easily by building upon the notation used in (2). A 4-regime model can be obtained by 'encapsulating' two different two-regime LSTAR models as follows:

$$\begin{aligned} y_t = {} & [\phi_1' x_t (1 - G_1(s_{1t};\gamma_1,c_1)) + \phi_2' x_t G_1(s_{1t};\gamma_1,c_1)][1 - G_2(s_{2t};\gamma_2,c_2)] \\ & + [\phi_3' x_t (1 - G_1(s_{1t};\gamma_1,c_1)) + \phi_4' x_t G_1(s_{1t};\gamma_1,c_1)] G_2(s_{2t};\gamma_2,c_2) + \varepsilon_t. \end{aligned} \tag{10}$$

The effective relationship between y_t and its lagged values is given by a linear combination of four linear AR models, each associated with a particular combination of $G_1(s_{1t})$ and $G_2(s_{2t})$ being equal to 0 or 1. This so-called Multiple Regime STAR [MRSTAR] model is discussed in detail in van Dijk and Franses (1999). The MRSTAR model as given in (10) allows for a maximum of four different regimes, but it is obvious that by applying the principle of encapsulating repeatedly, the model can be extended to contain 2^m regimes with $m > 2$, at least conceptually.

The MRSTAR model reduces to a (SE)TAR model with multiple regimes determined by multiple sources in case the smoothness parameters γ_1 and γ_2 become arbitrarily large, such that the logistic functions G_1 and G_2 approach indicator functions $I[s_{1t} > c_1]$ and $I[s_{2t} > c_2]$, respectively. The resultant Nested TAR [NeTAR] model is discussed in Astatkie, Watts and Watt (1997). The name *nested* TAR model stems from the fact that the time series y_t can be thought of as being described by a two-regime SETAR model with regimes defined by s_{1t}, and within each of those regimes by a two-regime SETAR model with regimes defined by s_{2t}, or vice versa.

The MRSTAR model also nests the flexible coefficient smooth transition time series model considered in Medeiros and Veiga (2000). This model is obtained by assuming that the transition variables s_{1t} and s_{2t} are linear combinations of lagged dependent variables, that is, $s_{it} = \alpha_i' \tilde{x}_t$, $i = 1, 2$, and imposing the restriction $\phi^*_{4,j} \equiv \phi_{1,j} - \phi_{2,j} - \phi_{3,j} + \phi_{4,j} = 0$ for $j = 0, 1, \ldots, p$. This restriction ensures that the interaction term $\phi_4^{*\prime} x_t G_1(s_{1t}) G_2(s_{2t})$ drops out of the model, which now can be rewritten as

$$y_t = \phi_0^{*\prime} x_t + \phi_1^{*\prime} x_t G_1(\alpha_1' \tilde{x}_t; \gamma_1, c_1) + \phi_2^{*\prime} x_t G_2(\alpha_2' \tilde{x}_t; \gamma_2, c_2) + \varepsilon_t, \tag{11}$$

where $\phi_0^* = \phi_1$, $\phi_1^* = \phi_2 - \phi_1$ and $\phi_2^* = \phi_3 - \phi_1$. Öcal and Osborn (2000) apply a special case of this flexible coefficient model to describe business cycle nonlinearity in UK macroeconomic time series. In their models it is assumed that α_1 and α_2 are $(p \times 1)$ vectors with unity as d-th and e-th element and zeros elsewhere for certain $1 \leq d, e \leq p$, such that $s_{1t} = y_{t-d}$ and $s_{2t} = y_{t-e}$.

Finally, note that the flexible coefficient model in (11) includes a (single hidden layer) artificial neural network [ANN] model as a special case. This ANN is obtained from (11) by leaving α_i, $i = 1, 2$, unspecified and imposing the restrictions $\phi^*_{i,j} = 0$, $i = 0, 1, 2$, $j = 1, \ldots, p$. It is well-known that by incorporating additional nonlinear components or so-called hidden units $\phi^*_{i,0} G_i(\alpha_i' \tilde{x}_t; \gamma_i, c_i)$, $i = 3, 4, \ldots$ in the model, an ANN with a finite number of hidden units can approximate any continuous function to any desired degree of accuracy, see Hornik, Stinchcombe and White (1989, 1990), among others. It follows that the same holds true for the flexible coefficient model and, hence, for the MRSTAR model. For reviews of ANN models, see Kuan and White (1994) and Franses and van Dijk (2000, Chapter 5).

3.2 Time Varying STAR models

Nonlinearity is only one of many different features a time series can possess. Another important characteristic, especially of macro-economic time series observed over long time spans, is structural instability, see, for example, Stock and Watson (1996). Despite a large amount of evidence indicating that both nonlinearity and structural change are relevant for many time series, to date these features have mainly been analyzed in isolation. It is our impression that this dichotomy is due to how time series modelling usually is carried out. Typically, such modelling starts with specifying a linear model. The estimated linear model is routinely subjected to a battery of misspecification tests, including tests of linearity and parameter constancy. If certain misspecification tests indicate that the linear model is inadequate, the model is modified accordingly. The modelling usually ends with estimating this alternative model. Thus, when nonlinearity is found and modelled, parameter constancy of the estimated nonlinear model is rarely tested, and thus even more seldom rejected. Conversely, when parameter constancy is rejected in a linear model, testing the alternative time-varying parameter model for nonlinearity is not normally done.

The motivation for considering either nonlinearity or parameter non-constancy, and not both, as alternative to linearity might be that empirically the two can be close substitutes. For example, it is not difficult to parameterize a nonlinear time series model in such a way that its realizations resemble series that are subject to occasional level shifts, see Granger and Teräsvirta (1999) for an example. Casual inspection of a graph of such a series might suggest that a model with time-varying parameters is an appropriate characterization of its properties. Conversely, as discussed by Timmermann (2000), structural breaks can be described by a nonlinear time series model with infrequent regime shifts. Garcia and Perron (1996) provide an illustrative empirical example of this phenomenon. The 3-regime Markov Switching model which they estimate for the US real interest rate exhibits only 2 regime shifts over the 40-year sample period. Consequently, statistical procedures might have difficulty to distinguish nonlinearity from structural change. For example, Carrasco (1997) and Clements and Smith (1998) find that tests for SETAR type nonlinearity reject the null hypothesis of linearity with high probability when the data in fact are generated by a structural change model, and *vice versa*. In a similar vein, Koop and Potter (2000) use Bayesian techniques to show that a lot of evidence for nonlinearity in economic time series might in fact be due to structural change.

Given the above, nonlinearity, and regime-switching behaviour in particular, and structural change can be regarded as competing alternatives to linearity and it might be difficult to distinguish between the two. Of course, it is also possible that a time series displays both nonlinearity and structural instability.

An interesting special case of the MRSTAR model can be used to allow for both nonlinear dynamics of the STAR-type and time-varying characteristics. This so-called time-varying STAR [TVSTAR] model arises if one of the transition variables in (10) is taken to be time, say, $s_{2t} = t$. The TVSTAR model implies that y_t follows a STAR model at all times, with a smooth change in the autoregressive parameters in both regimes, from ϕ_1 and ϕ_2 to ϕ_3 and ϕ_4 for $G(s_{1t};\gamma_1,c_1) = 0$ and $G(s_{1t};\gamma_1,c_1) = 1$, respectively, which can easily be seen from the alternative representation

$$y_t = \phi_1(t)'x_t(1 - G_1(s_{1t};\gamma_1,c_1)) + \phi_2(t)'x_tG_1(s_{1t};\gamma_1,c_1) + \varepsilon_t, \tag{12}$$

with

$$\phi_1(t) = \phi_1[1 - G_2(t;\gamma_2,c_2)] + \phi_3G_2(t;\gamma_2,c_2), \tag{13}$$

$$\phi_2(t) = \phi_2[1 - G_2(t;\gamma_2,c_2)] + \phi_4G_2(t;\gamma_2,c_2). \tag{14}$$

The TVSTAR model is discussed in detail in Lundbergh, Teräsvirta and van Dijk (2000).

3.3 Vector STAR models

Linear vector AR [VAR] models constitute the most common way of modelling vector time series. In some situations, it could be worthwhile to consider nonlinear models for this purpose. Regime-switching at different phases of the business cycle could serve as an example, see Diebold and Rudebusch (1996), Koop, Pesaran and Potter (1996), Ravn and Sola (1995) and Weise (1999). Conceptually it is straightforward to extend the existing univariate regime-switching models to a multivariate context. However, the interest in multivariate nonlinear modelling has started to develop only very recently and, therefore, the relevant statistical theory is not yet fully developed. Krolzig (1997) and Tsay (1998) consider vector Markov-Switching models and threshold models, respectively. Here we concentrate on vector STAR models.

Let $Y_t = (y_{1t},\ldots,y_{kt})'$ be a $(k \times 1)$ vector time series. A k-dimensional analogue of the univariate 2-regime STAR model (1) can be specified as

$$\begin{aligned} Y_t = (\Phi_{1,0} + \Phi_{1,1}Y_{t-1} + \cdots + \Phi_{1,p}Y_{t-p})(1 - G(s_t;\gamma,c)) \\ + (\Phi_{2,0} + \Phi_{2,1}Y_{t-1} + \cdots + \Phi_{2,p}Y_{t-p})G(s_t;\gamma,c) + \varepsilon_t, \end{aligned} \tag{15}$$

where $\Phi_{i,0}$, $i = 1, 2$, are $(k \times 1)$ vectors, $\Phi_{i,j}$, $i = 1, 2$, $j = 1,\ldots,p$, are $(k \times k)$ matrices, and $\varepsilon_t = (\varepsilon_{1t},\ldots,\varepsilon_{kt})'$ is a k-dimensional vector white noise process with mean zero and $(k \times k)$ positive definite covariance matrix Σ.

Notice that in (15) the regimes are common to the k variables, in the sense that one and the same transition function determines the prevailing regime and the switches between regimes in all k equations of the model. It is straightforward to generalize the model to incorporate equation-specific transition functions $G_1(s_{1t};\gamma_1,c_1),\ldots,G_k(s_{kt};\gamma_k,c_k)$ and thereby allow for equation-specific regime-switching.

Judging from applications of multivariate regime-switching models that are available at present, it seems that a model of particular interest is one in which the components of Y_t are linked by a linear long-run equilibrium relationship, whereas adjustment towards this equilibrium is nonlinear and can be characterized as regime-switching, with the regimes determined by the size and/or sign of the deviation from equilibrium. In linear time series, this type of behaviour is captured by cointegration and error-correction models, see Banerjee, Dolado, Galbraith and Hendry (1993), Johansen (1995), and Hatanaka (1996) for in-depth treatments. Recently, nonlinear extensions of these concepts have been considered. Here we concentrate on incorporating the smooth transition mechanism in an ECM to allow for nonlinear or asymmetric adjustment, see Granger and Swanson (1996) for more general discussion. A smooth transition error-correction model [STECM] is given by

$$\begin{aligned}\Delta Y_t = (\Phi_{1,0} + \alpha_1 z_{t-1} + \sum_{j=1}^{p-1}\Phi_{1,j}\Delta Y_{t-j})(1 - G(s_t;\gamma,c))\\ + (\Phi_{2,0} + \alpha_2 z_{t-1} + \sum_{j=1}^{p-1}\Phi_{2,j}\Delta Y_{t-j})G(s_t;\gamma,c) + \varepsilon_t, \quad (16)\end{aligned}$$

where α_i, $i=1,2$ are $k\times 1$ vectors and $z_t=\beta' Y_t$ for some $k\times 1$ vector β denote the error-correction term, that is, z_t is the deviation from the equilibrium relationship which is given by $\beta' Y_t = 0$. The model can be extended to incorporate multiple equilibrium relationships, $\beta_1' Y_t,\ldots,\beta_r' Y_t$ say, for certain $1\leq r\leq k$, see Swanson (1999).

It appears that relevant forms of nonlinear error correction often concern different adjustment to positive and negative or to large and small deviations from equilibrium. (Of course, other forms of nonlinear error correction, which do not depend directly on the deviation from equilibrium itself, are possible, see Siklos and Granger (1997) for an example.) Both types of asymmetry arise in a natural way when modelling prices of so-called equivalent assets in financial markets, see Yadav, Pope and Paudyal (1994) and Anderson (1997) for elaborate discussions. Equivalent assets in a certain sense represent the same underlying value. Examples of equivalent assets include stocks and futures, and bonds of different maturity. Since they are traded in the same market, or in markets linked by arbitrage-related forces, the prices of equivalent assets should be such that investors are indifferent between holding either one of them. Deviations from the equilibrium create arbitrage opportunities that drive the prices back together. However, market frictions can give rise to asymmetric adjustment of such deviations. Due to short-selling constraints, for example, the response to negative deviations from equilibrium might be different from the response to positive deviations. Alternatively, transaction costs prevent adjustment of equilibrium errors as long as the benefits from adjustment, which equal the price difference, are smaller than those costs. These market frictions suggest that the degree of error correction is a function of the sign and/or size of the deviation from equilibrium.

Asymmetric effects of positive and negative deviations from equilibrium can be obtained by defining $G(s_t; \gamma, c)$ as in (3) and setting $s_t = z_{t-1}$. In the resultant model, the strength of reversion of z_t to its attractor changes monotonically for increasing values of z_{t-1}. The constant c in (3) can be set equal to zero to render the change symmetric around the equilibrium value of zero. Asymmetric behaviour for small and large equilibrium errors can be achieved by taking $G(s_t; \gamma, c)$ to be the exponential function (4) with $s_t = z_{t-1}$, where again c should be set equal to 0 to center the function at the equilibrium, or the quadratic logistic function (5) with $s_t = z_{t-1}$ and $-c_1 = c_2 > 0$. This results in gradually changing strength of adjustment for larger (both positive and negative) deviations from equilibrium.

The smooth transition ECM with the exponential transition function (4) is used by Taylor, van Dijk, Franses and Lucas (2000) to describe the relationship between spot and futures prices of the FTSE100 index in the presence of transaction costs. Other applications of smooth transition error correction models include the term structure of interest rates (Anderson (1997), van Dijk and Franses (2000)), although both consider only single-equation (conditional) error-correction models) and the relationship between money and output (Swanson (1999) and Rothman, van Dijk and Franses (1999)).

The smooth transition error correction model with a quadratic logistic function resembles the threshold error correction model, introduced by Balke and Fomby (1997). The threshold error correction model is obtained from (16) by letting $\gamma \to \infty$ in (5) and imposing the restriction $\alpha_1 = 0$. Intuitively, each element of Y_t then contains a unit root as long as $z_{t-1} \in (c_1, c_2)$ and the component time series y_{it}, $i = 1, \ldots, k$ behave as nonstationary non-cointegrated variables. When z_{t-1} becomes smaller than c_1 or larger than c_2, Y_t is cointegrated. Threshold error correction models are applied by Dwyer, Locke and Yu (1996), Martens, Kofman and Vorst (1998), Tsay (1998), and Forbes, Kalb and Kofman (1999) to describe the relationship between spot and futures prices of the S&P 500 index.

Common nonlinearity

In the case of vector time series, there exists the possibility that nonlinearity is caused by common nonlinear components. Following Anderson and Vahid (1998), the k-dimensional time series Y_t is said to contain s common nonlinear components if there exist $k - s$ linear combinations $\alpha_i' Y_t$, $i = 1, \ldots, k - s$, whose conditional expectations are linear in the past of Y_t. Rewriting model (15) as

$$Y_t = \Phi_{1,0}^* + \Phi_{1,1}^* Y_{t-1} + \cdots + \Phi_{1,p}^* Y_{t-p} + (\Phi_{2,0}^* + \Phi_{2,1}^* Y_{t-1} + \cdots + \Phi_{2,p}^* Y_{t-p}) G(s_t; \gamma, c) + \varepsilon_t,$$

where $\Phi_{1,j}^* = \Phi_{1,j}$ and $\Phi_{2,j}^* = \Phi_{2,j} - \Phi_{1,j}$, $j = 0, 1, \ldots, p$, the existence of s common nonlinear components means that there exists a $(k \times k - s)$ matrix A such that

$$A'(\Phi_{2,0}^* + \Phi_{2,1}^* Y_{t-1} + \cdots + \Phi_{2,p}^* Y_{t-p}) G(s_t; \gamma, c) = 0, \tag{17}$$

for all $Y_{t-1}, \ldots, Y_{t-p}$ and s_t. Anderson and Vahid (1998) develop test statistics for the existence of common nonlinearity based upon canonical correlations.

4 Hypothesis testing in the STAR framework

Before presenting the modelling cycle for STAR models mentioned in the Introduction, we discuss hypothesis testing in the STAR framework. This involves tests of linearity against the alternatives of LSTAR and ESTAR nonlinearity in Section 4.1, and misspecification tests of smooth transition models in Section 4.2. The complications that arise in these testing procedures due to the presence of unidentified nuisance parameters under the null hypothesis are considered in Section 4.1. Finally, in Section 4.3 we review newly developed test procedures which can handle outliers.

4.1 Testing linearity against STAR

Testing linearity against STAR constitutes a first step towards building STAR models. The null hypothesis of linearity can be expressed as equality of the autoregressive parameters in the two regimes of the STAR model in (2). Thus, $H_0 : \phi_1 = \phi_2$, whereas the alternative hypothesis is $H_1 : \phi_{1,j} \neq \phi_{2,j}$ for at least one $j \in \{0, \ldots, p\}$.

The testing problem is complicated by the presence of unidentified nuisance parameters under the null hypothesis. Informally, the STAR model contains parameters which are not restricted by the null hypothesis, but about which nothing can be learned from the data when the null hypothesis holds true. For example, the null hypothesis $\phi_1 = \phi_2$ does not restrict the parameters in the transition function, γ and c, but when the null hypothesis is valid, the likelihood is unaffected by the values of γ and c.

An alternative way to illustrate the presence of unidentified nuisance parameters in this case is to note that the null hypothesis of linearity can be formulated in different ways. Besides equality of the AR parameters in the two regimes, $H_0 : \phi_1 = \phi_2$, the alternative null hypothesis $H_0' : \gamma = 0$ also gives rise to a linear model. For example, if $\gamma = 0$ the logistic function (3) is equal to 0.5 for all values of s_t, and the STAR model (2) reduces to an AR model with parameters $(\phi_1 + \phi_2)/2$. In case H_0' is used, the location parameter c

and the parameters ϕ_1 and ϕ_2 are the unidentified parameters. Under H_0', ϕ_1 and ϕ_2 can take any value as long as their average remains the same.

The problem of unidentified nuisance parameters under the null hypothesis was first considered by Davies (1977, 1987) and occurs in many testing problems, see Andrews and Ploberger (1994), Hansen (1996) and Stinchcombe and White (1998) for recent general accounts. The main consequence of the presence of such nuisance parameters is that the conventional statistical theory is not available for obtaining the asymptotic null distribution of the test statistics. Instead, the test statistics tend to have non-standard distributions for which analytic expressions are most often not available. This implies that critical values have to be determined by means of simulation.

The problem of testing linearity against STAR alternatives was addressed in Luukkonen, Saikkonen and Teräsvirta (1988a). Their proposed solution is to replace the transition function $G(s_t;\gamma,c)$ by a suitable Taylor series approximation. In the reparametrized equation, the identification problem is no longer present, and linearity can be tested by means of a Lagrange Multiplier [LM] statistic with a standard asymptotic χ^2-distribution under the null hypothesis. This approach has two main advantages. First, the model under the alternative hypothesis need not be estimated and, second, standard asymptotic theory is available for obtaining (asymptotic) critical values for the test statistics.

Tests against LSTAR

Consider the LSTAR model (2) with (3), rewritten as

$$y_t = \phi_1' x_t + (\phi_2 - \phi_1)' x_t G(s_t;\gamma,c) + \varepsilon_t, \tag{18}$$

and assume that $\{\varepsilon_t\} \sim$ n.i.d.$(0,\sigma^2)$. In order to derive a linearity test against (18), Luukkonen *et al.* (1988a) suggest to approximate the logistic function $G(s_t;\gamma,c) = 1/(1+\exp\{-\gamma(s_t - c)\})$ with a first-order Taylor approximation around $\gamma = 0$. This results in the auxiliary regression

$$y_t = \beta_0' x_t + \beta_1' x_t s_t + e_t, \tag{19}$$

where $\beta_i = (\beta_{i,0}, \beta_{i,1}, \ldots, \beta_{i,p})'$, $i = 0, 1$, and $e_t = \varepsilon_t + (\phi_2 - \phi_1)' x_t R_1(s_t;\gamma,c)$, with $R_1(s_t;\gamma,c)$ the remainder term from the Taylor expansion. Under the null hypothesis, $R_1(s_t;\gamma,c) \equiv 0$ and $e_t = \varepsilon_t$. Consequently, this remainder term does not affect the properties of the errors under the null hypothesis and, hence, the asymptotic distribution theory. The parameters β_i, $i = 0, 1$, in the auxiliary regression (19) are functions of the parameters in the STAR model (18) such that the restriction $\gamma = 0$ implies $\beta_{0,j} \neq 0$ and $\beta_{1,j} = 0$ for $j = 0, \ldots, p$. Hence, testing the null hypothesis $H_0' : \gamma = 0$ (or $H_0 : \phi_1 = \phi_2$) in (18) is equivalent to testing the null hypothesis $H_0'' : \beta_1 = 0$ in (19). This hypothesis can be tested by a standard variable addition test in a straightforward manner. The test statistic, to be denoted as LM_1, has an asymptotic χ^2 distribution with $p+1$ degrees of freedom under the null hypothesis of linearity. As the LM_1 statistic does not test the original null hypothesis $H_0' : \gamma = 0$ but rather the auxiliary null hypothesis $H_0'' : \beta_1 = 0$, this test is usually referred to as an LM-*type* statistic. The test statistic can also be developed from first principles as a genuine LM statistic, see Granger and Teräsvirta (1993, pp. 71-72). It can be shown that the statistic is in fact the supremum of the pointwise statistics for fixed $\phi_2 - \phi_1$ and c and is thus similar in spirit to the test statistic that is commonly applied to test for (SE)TAR nonlinearity, see Hansen (1997).

Note that in case $s_t = y_{t-d}$ for certain integer $1 \leq d \leq p$, $\beta_{1,0}s_t$ should be dropped from the auxiliary regression (19) to avoid perfect multi-collinearity. As noted by Luukkonen *et al.* (1988a), for this choice of transition variable, the LM_1 statistic does not have power in situations where only the intercept differs across regimes, that is, when $\phi_{1,0} \neq \phi_{2,0}$ but $\phi_{1,j} = \phi_{2,j}$ for $j = 1, \dots, p$. This problem can be solved by approximating the transition function $G(s_t; \gamma, c)$ by a third-order Taylor approximation. This yields the auxiliary regression

$$y_t = \beta_0' x_t + \beta_1' x_t s_t + \beta_2' x_t s_t^2 + \beta_3' x_t s_t^3 + e_t, \tag{20}$$

where $e_t = \varepsilon_t + (\phi_2 - \phi_1)' x_t R_3(s_t; \gamma, c)$, and $\beta_{0,0}$ and the β_i, $i = 1, 2, 3$, again are functions of the parameters ϕ_1, ϕ_2, γ and c. Inspection of the exact relationships shows that the null hypothesis $H_0' : \gamma = 0$ now corresponds to $H_0'' : \beta_1 = \beta_2 = \beta_3 = 0$, which again can be tested by a standard LM-type test. Under the null hypothesis of linearity, the test statistic, to be denoted as LM_3, has an asymptotic χ^2 distribution with $3(p+1)$ degrees of freedom. Again, if $s_t = y_{t-d}$ for certain integer $d \leq p$, the terms $\beta_{i,0}s_t^i$, $i = 1, 2, 3$, should be dropped from the auxiliary regression.

The expressions of β_i, $i = 1, 2, 3$, in (20) in terms of ϕ_1, ϕ_2, γ and c also reveal that, in case s_t is not included in $\tilde{x}_t$, the only parameters that depend on the constants $\phi_{1,0}$ and $\phi_{2,0}$ are $\beta_{1,0}$, $\beta_{2,0}$ and $\beta_{3,0}$. Hence, a parsimonious, or 'economy', version of the LM_3 statistic can be obtained by augmenting the auxiliary regression (19) with regressors s_t^2 and s_t^3, that is,

$$y_t = \beta_0' x_t + \beta_1' x_t s_t + \beta_{2,0} s_t^2 + \beta_{3,0} s_t^3 + e_t, \tag{21}$$

and testing the null hypothesis $H_0'' : \beta_1 = 0$ and $\beta_{2,0} = \beta_{3,0} = 0$. The resultant test statistic, denoted LM_3^e, has an asymptotic χ^2 distribution with $p+3$ degrees of freedom. The advantage of the LM_3^e statistic over the the LM_2 statistic is that it requires considerably less degrees of freedom. In case $s_t = y_{t-d}$ for certain $d \leq p$, the only parameters in the auxiliary regression that are informative about $\phi_{1,0}$ and $\phi_{2,0}$ are $\beta_{2,d}$ and $\beta_{3,d}$, and the LM_3^e statistic is obtained by augmenting the auxiliary regression (19) with y_{t-d}^3 and y_{t-d}^4.

Sometimes the appropriate transition variable s_t under the alternative may not be obvious. In case the choice is between $y_{t-1}, \ldots, y_{t-p}$, one can define the linear combination $s_t = \alpha'\tilde{x}_t$ with $\alpha = (0, \ldots, 0, 1, 0, \ldots, 0)'$, where the position of the only unity element is left unspecified. The LM_1, LM_3 and LM_3^e statistics then become become LM-type tests against LSTAR with $s_t = y_{t-d}$ with the delay parameter d assumed unknown. Notice that in this case the auxiliary regression used in computing the LM_1 statistic becomes

$$y_t = \beta_0' x_t + \sum_{i=1}^{p} \sum_{j=i}^{p} \beta_{1,ij} y_{t-i} y_{t-j} + e_t,$$

and the resultant test in fact is identical to the general linearity test of Tsay (1986).

Tests against ESTAR

Saikkonen and Luukkonen (1988) suggest testing linearity against an ESTAR alternative by using the auxiliary regression

$$y_t = \beta_0' x_t + \beta_1' x_t s_t + \beta_2' x_t s_t^2 + e_t, \tag{22}$$

where $e_t = \varepsilon_t + (\phi_2 - \phi_1)' x_t R_2(s_t; \gamma, c)$. Equation (22) is based on the first-order Taylor series expansion of (2) with (4) [or (5)]. The expressions for β_i, $i = 0, 1, 2$, show that the restriction $\gamma = 0$ corresponds with $\beta_1 = \beta_2 = 0$ in (22). The LM_2 statistic which tests this null hypothesis has an asymptotic χ^2 distribution with $2(p+1)$ degrees of freedom.

Escribano and Jordá (1999) claim that a first-order Taylor approximation of the exponential function is not sufficient to capture its characteristic features, the two inflexion points of this function in particular. They suggest that a second-order Taylor approximation is necessary, yielding the auxiliary regression,

$$y_t = \beta_0' x_t + \beta_1' x_t s_t + \beta_2' x_t s_t^2 + \beta_3' x_t s_t^3 + \beta_4' x_t s_t^4 + e_t. \tag{23}$$

The null hypothesis to be tested now is $H_0' : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$. The resultant LM-type test statistic, denoted LM_4, has an asymptotic χ^2 distribution with $4(p+1)$ degrees of freedom under the null hypothesis. There is a trade-off between the extra variables in the auxiliary regression and the increase in the dimension of the null hypothesis. Neither one of the tests based on (22) or (23) dominates the other in terms of power.

Computational aspects

In small samples, it is a good strategy to use F-versions of the LM test statistics because these have better size properties than the χ^2 variants, which may be heavily oversized in small samples. Both the χ^2 and F versions can be computed by means of two auxiliary linear regressions. As an example, the LM_3 statistic based on (20) can be computed as follows:

1. Estimate the model under the null hypothesis of linearity by regressing y_t on x_t. Compute the residuals $\hat{\varepsilon}_t$ and the sum of squared residuals $\mathrm{SSR}_0 = \sum_{t=1}^{T} \hat{\varepsilon}_t^2$.

2. Estimate the auxiliary regression of y_t on x_t and $x_t s_t^i$, $i = 1, 2, 3$. Compute the residuals $\hat{e}_t$ and the sum of squared residuals $\mathrm{SSR}_1 = \sum_{t=1}^{T} \hat{e}_t^2$.

3. The χ^2 version of the LM_3 statistic can now be computed as

$$\mathrm{LM}_3 = \frac{T(\mathrm{SSR}_0 - \mathrm{SSR}_1)}{\mathrm{SSR}_0}, \tag{24}$$

whereas the F version can be computed as

$$\mathrm{LM}_3 = \frac{(\mathrm{SSR}_0 - \mathrm{SSR}_1)/3(p+1)}{\mathrm{SSR}_1/(T - 4(p+1))}. \tag{25}$$

Under the null hypothesis, the F version of the test is approximately F distributed with $3(p+1)$ and $T - 4(p+1)$ degrees of freedom.

4.2 Misspecification tests of STAR models

Before an estimated STAR model can be accepted as adequate, it should be subjected to a thorough evaluation, including a number of misspecification tests. Obvious hypotheses which might be tested are no residual autocorrelation, no remaining nonlinearity and parameter constancy. Eitrheim and Teräsvirta (1996) develop LM-type tests for these three hypotheses in the basic two-regime STAR model. In that context, the tests of no remaining nonlinearity and parameter constancy also can be interpreted as tests against the alternatives of a multiple regime STAR model and a TVSTAR model, respectively. It is straightforward to generalize the misspecification tests to MRSTAR and TVSTAR models, see van Dijk and Franses (1999) and Lundbergh *et al.* (2000), respectively. Extensions of the misspecification tests to the vector STAR framework are considered in Anderson and Vahid (1998, Appendix D).

4.2.1 Testing the hypothesis of no residual autocorrelation

Consider the STAR model (2) and denote the so-called skeleton of the model as

$$F(x_t; \theta) = \phi_1' x_t (1 - G(s_t; \gamma, c)) + \phi_2' x_t G(s_t; \gamma, c). \tag{26}$$

An LM-test for q-th order serial dependence in ε_t can be obtained as nR^2, where R^2 is the coefficient of determination from the regression of $\hat{\varepsilon}_t$ on $\nabla F(x_t; \hat{\theta}) = \partial F(x_t; \hat{\theta})/\partial\theta$, with $\theta = (\phi_1, \phi_2, \gamma, c)'$, and q lagged residuals $\hat{\varepsilon}_{t-1}, \ldots, \hat{\varepsilon}_{t-q}$. Hats indicate that the relevant quantities are estimates under the null hypothesis of serial independence of ε_t. The resultant test statistic, to be denoted as $\mathrm{LM_{SI}}(q)$, is asymptotically χ^2 distributed with q degrees of freedom. This test statistic is a generalization of the LM-test for serial correlation in an AR(p) model of Godfrey (1979), which is obtained by setting $F(x_t;\theta) = \phi' x_t$.

Testing the hypothesis of no remaining nonlinearity

Eitrheim and Teräsvirta (1996) develop an LM statistic to test the two-regime LSTAR model (18) against the alternative of an additive STAR model defined in (8). The null hypothesis of a two-regime model can be expressed as either $H_0' : \gamma_2 = 0$ or $H_0 : \phi_3 = \phi_2$. Evidently, this testing problem suffers from a similar identification problem as encountered in testing linearity against a two-regime STAR model, see Section 4.1. The solution to this identification problem again is to replace the transition function $G_2(s_t;\gamma_2,c_2)$ by a Taylor series approximation around $\gamma_2 = 0$. Using a third-order approximation, the resultant approximation to model (8) becomes

$$y_t = \beta_0' x_t + (\phi_2 - \phi_1)' x_t G_1(s_t;\gamma_1,c_1) + \beta_1' x_t s_t + \beta_2' x_t s_t^2 + \beta_3' x_t s_t^3 + e_t, \tag{27}$$

where the parameters $\beta_i,\ i = 0,1,2,3$, are functions of the parameters $\phi_1,\phi_2,\phi_3,\gamma_2$ and c_2. The null hypothesis $H_0' : \gamma_2 = 0$ in (8) translates into $H_0'' : \beta_1 = \beta_2 = \beta_3 = 0$ in (27). The test statistic can be computed as nR^2 from the auxiliary regression of the residuals obtained from estimating the model under the null hypothesis on the partial derivates of the regression function with respect to the parameters in the two-regime model, ϕ_1,ϕ_2,γ_1 and c_1, evaluated under the null hypothesis, and the auxiliary regressors $x_t s_t^i$, $i = 1,2,3$. The resultant test statistic $\mathrm{LM_{AMR,3}}$ has an asymptotic χ^2 distribution with $3(p+1)$ degrees of freedom, where the subscript AMR is used to indicate that this statistic is designed as a test against an additive multiple regime model. Note that in going from (8) to (27), we have implicitly assumed that s_t is not an element of x_t. If it is, the auxiliary regressors $\beta_{i,0} s_t^i$, $i = 1,2,3$ should be omitted from (27).

van Dijk and Franses (1999) derive an LM-type test for testing the null of the two-regime LSTAR model (18) against the MRSTAR alternative given in (10). The null hypothesis can be expressed as either $H_0' : \gamma_2 = 0$ or $H_0 : \phi_1 = \phi_3$ and $\phi_2 = \phi_4$. In case the transition function $G_2(s_{2t};\gamma_2,c_2)$ is replaced with a third-order Taylor series approximation, the corresponding approximation to (10) can be written as

$$\begin{aligned} y_t = \theta_1' x_t + \theta_2' x_t G_1(s_{1t};\gamma_1,c_1) + \beta_1' x_t s_{2t} + \beta_2' x_t s_{2t}^2 + \beta_3' x_t s_{2t}^3 \\ + (\beta_4' x_t s_{2t} + \beta_5' x_t s_{2t}^2 + \beta_6' x_t s_{2t}^3) G_1(s_{1t};\gamma_1,c_1) + e_t. \end{aligned} \tag{28}$$

The parameter vectors $\beta_i = (\beta_{i,1}, \dots, \beta_{i,p})'$, $i = 1, \dots, 6$, in (28) are defined in terms of ϕ_i^*, $i = 1, \dots, 4$, γ_2 and c_2, such that the null hypothesis can be reformulated as $H_0' : \beta_i = 0$, $i = 1, \dots, 6$. The resultant test statistic $\mathrm{LM}_{\mathrm{EMR},3}$ is asymptotically χ^2 distributed with $6(p+1)$ degrees of freedom, where the subscript EMR is used to indicate that this statistic is designed as a test against an 'encapsulated' multiple regime model. Again it is implicitly assumed that s_{2t} is not an element of x_t. If it is, the terms $\beta_{i,0} s_{2t}^i$, $i = 1, 2, 3$, and $\beta_{i,0} s_{2t}^{i-3} G_1(s_{1t}; \gamma_1, c_1)$, $i = 4, 5, 6$, do not appear in (28).

Testing the hypothesis of parameter constancy

By testing the hypothesis $\gamma_2 = 0$ in (12), one can test for parameter constancy in the two-regime STAR model (18), against the alternative of smoothly changing parameters. The appropriate LM-type test statistic based on a third-order Taylor approximation of $G_2(t; \gamma_2, c_2)$, to be denoted as $\mathrm{LM}_{\mathrm{C},3}$, is identical to the $\mathrm{LM}_{\mathrm{EMR},3}$ statistic with $s_{2t} = t$. Note that the asymptotic distribution theory remains the same even if the transition variable is a non-stationary deterministic trend, see Lin and Teräsvirta (1994).

Computational aspects

Eitrheim and Teräsvirta (1996) point out potential numerical problems in the computation of the misspecification tests. In particular, if $\hat{\gamma}_1$ is very large, such that the transition between the two regimes in the model under the null hypothesis is rapid, the partial derivatives of the transition function $G_1(s_{1t}; \gamma_1, c_1)$ with respect to γ_1 and c_1 approach zero functions, with the possible exception of a few 'blips'. The 'blips' in these partial derivatives occur simultaneously, and as a result the moment matrix of the regressors in the auxiliary regressions used in computing the test statistics becomes near-singular. However, because the contributions of the terms involving these partial derivatives are likely to be very small for all $t = 1, \dots, T$ when $\hat{\gamma}_1$ is very large, they contain little information and these terms can simply be omitted from the auxiliary regression without affecting the power properties of the test statistics. Another practical problem is that the residuals $\hat{\varepsilon}_t$ of the two-regime STAR model may not always be exactly orthogonal to the gradient matrix. This may be the case if the model does not fit the data very well, so that the numerical algorithm applied in parameter estimation has difficulty finding an optimum. Eitrheim and Teräsvirta (1996) suggest accounting for this replacing by $\hat{\varepsilon}_t$ with the residuals from the regression of $\hat{\varepsilon}_t$ on the elements of the gradient $\partial F(x_t; \hat{\theta})/\partial\theta$. By construction, these residuals are orthogonal to the gradient.

4.3 Recent developments

The LM-type tests discussed above are sensitive to several kinds of misspecification of the model under the null hypothesis. For example, it is well-known that residual autocorrelation in a linear AR model may lead to spurious findings of nonlinearity. In this section we discuss the effects of two other forms of misspecification, neglected heteroskedasticity and outliers, and ways to robustify the LM-type tests against these effects.

Heteroskedasticity and tests for STAR nonlinearity

The LM-type tests assume constant (conditional) variance. Neglected heteroskedasticity has similar effects on tests for nonlinearity as residual autocorrelation, in that it may lead to spurious rejection of the null hypothesis. Davidson and MacKinnon (1985) and Wooldridge (1990, 1991) develop specification tests which can be used in the presence of heteroskedasticity, without the need to specify the form the heteroskedasticity (which often is unknown) explicitly. Their procedures may be readily applied to robustify the tests against STAR nonlinearity, see also Granger and Teräsvirta (1993, pp. 69-70). For example, a heteroskedasticity robust variant of the LM_3 statistic based upon (20) can be computed as follows:

(i) Regress y_t on x_t and obtain the residuals $\hat{\varepsilon}_t$.

(ii) Regress the auxiliary regressors $x_t s_t^i$, $i = 1, 2, 3$, on x_t and compute the residuals $\hat{r}_t$.

(iii) Regress 1 on $\hat{\varepsilon}_t \hat{r}_t$. The explained sum of squares from this regression is the LM-type statistic.

Lundbergh and Teräsvirta (1998) present simulation evidence which suggests that in some cases this robustification removes most of the power of the linearity tests, so that existing nonlinearity may not be detected. If the objective of the analysis is to find and model nonlinearity in the conditional mean, robustification therefore cannot be recommended. It might be expected that false rejections of the null hypothesis of linearity due to heteroskedasticity are discovered at the estimation or evaluation stages of model building. It may also be added that the standard tests of constant conditional variance against ARCH have power against nonlinearity in the conditional mean; for simulation evidence see, for example, Luukkonen, Saikkonen and Teräsvirta (1988b) and Lee, White and Granger (1993).

Outliers and tests for STAR nonlinearity

STAR models can be parameterized to generate very asymmetric realizations, in the sense that its realizations resemble linear time series with a few outliers. An interesting question then is how the LM-type tests for STAR nonlinearity perform when the data-generating process is a linear model but the observations are contaminated by occasional outliers. van Dijk, Franses and Lucas (1999) show that in the presence of additive outliers, these tests tend to reject the correct null hypothesis of linearity too often, even asymptotically. As a solution to this problem, van Dijk *et al.* (1999) suggest to use outlier-robust estimation techniques (see Huber (1981), Martin (1981), Hampel, Ronchetti, Rousseeuw and Stahel (1986), and Lucas, Franses and van Dijk (2001), among others).

Robust estimators are designed to obtain better parameter estimates in the presence of contamination, by assigning less weight to influential observations such as outliers. For example, a robust estimator for the AR(p) model $y_t = \phi' x_t + \varepsilon_t$ can be defined as the solution of the first order conditions

$$\sum_{t=1}^{T} w_r(r_t) x_t (y_t - \phi' x_t) = 0, \tag{29}$$

where r_t denotes the standardized residual, $r_t \equiv (y_t - \phi' x_t)/(\sigma_\varepsilon w_x(x_t))$, with σ_ε a measure of scale of the residuals $\varepsilon_t \equiv y_t - \phi' x_t$ and $w_x(\cdot)$ and $w_r(\cdot)$ are weight functions that are bounded between 0 and 1. From (29) it can be seen that the robust estimator is a type of weighted least squares estimator, with the weight for the t-th observation given by the value of $w_r(\cdot)$. The functions $w_x(\cdot)$ and $w_r(\cdot)$ should be chosen such that the t-th observation receives a relatively small weight if either the regressor x_t or the standardized residual $(y_t - \phi' x_t)/\sigma_\varepsilon$ becomes unusually large. The weight function $w_r(r_t)$ usually is specified in terms of a function $\psi(r_t)$ as $w_r(r_t) = \psi(r_t)/r_t$ for $r_t \neq 0$ and $w_r(0) = 1$. See Hampel *et al.* (1986) for a discussion of possible specifications for $\psi(r_t)$.

In addition to rendering better estimates of the model under the null hypothesis, robust estimation procedures allow to construct test statistics that are robust to outliers. A robust equivalent to the LM_3 statistic to test $H_0'' : \beta_1 = \beta_2 = \beta_3 = 0$ in (20) is given by nR^2, using the R^2 from the regression of the weighted residuals $\hat{\psi}(\hat{r}_t) = \hat{w}_r(\hat{r}_t)\hat{r}_t$ on the weighted regressors $\hat{w}_x(x_t) \odot (x_t', x_t' s_t, x_t' s_t^2, x_t' s_t^3)'$, where $\odot$ denotes element-by-element multiplication. The weights $\hat{w}_r(\hat{r}_t)$ and $\hat{w}_x(x_t)$ are obtained from robust estimation of the AR(p) model under the null. The resultant LM-type statistic has an asymptotic χ^2-distribution with $3(p+1)$ degrees of freedom. An outlier-robust equivalent of the F-version of the tests can also be computed without difficulty.

Simulation results in van Dijk *et al.* (1999) suggest that the robustified LM-type tests have good size properties in small samples, also in the presence of outliers. As expected, in case no outliers occur, the power of the robust tests is somewhat lower than that of their non-robust counterparts. However, in the presence of outliers, the power of the standard tests decreases dramatically, whereas the power of the robust tests is hardly affected.

5 The modelling cycle

Granger (1993) strongly recommends a 'specific-to-general' strategy for building nonlinear time series models. This implies starting with a simple or restricted model and proceeding to more complicated ones only if diagnostic tests indicate that the maintained model is inadequate. In the present situation, an additional (statistical) motivation for such an approach is that the identification problems discussed above prevent us from starting with a STAR model and reducing its size by, say, a series of likelihood ratio tests. The data-based modelling cycle for STAR models put forward by Teräsvirta (1994) follows this approach and consists of the following steps.

1. Specify a linear AR model of order p for the time series under investigation.
2. Test the null hypothesis of linearity against the alternative of STAR nonlinearity. If linearity is rejected, select the appropriate transition variable s_t and the form of the transition function $G(s_t; \gamma, c)$.
3. Estimate the parameters in the selected STAR model.
4. Evaluate the model, using diagnostic tests and impulse response analysis.
5. Modify the model if necessary.
6. Use the model for descriptive or forecasting purposes.

Steps 2-4 in the modelling cycle are described in detail in the subsections below. Forecasting with STAR models is discussed in Section 6.

The main element involved in the first step is the choice of the appropriate lag order p in the AR(p) model for y_t, that is,

$$y_t = \phi_0 + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t. \tag{30}$$

This lag order should be such that the corresponding residuals are approximately white noise, as the tests for nonlinearity that are used in the second step of the modelling cycle are sensitive to remaining linear residual autocorrelation. The order of the AR model can be selected by conventional methods, such as the Akaike Information Criterion [AIC], the Schwarz Information Criterion [BIC] or the Ljung-Box statistic. It should be kept in mind that if linearity is rejected in the second step of the modelling cycle, the lag order used in the AR model is not necessarily the appropriate lag order in the alternative STAR model, although usually it provides a reasonable first guess.

5.1 Specification

The second stage of the modelling cycle presented above is labeled specification, as the main objectives, besides testing of linearity, are to select the appropriate transition variable in the STAR model and the most suitable form of the transition function.

Selecting the transition variable

Even though the LM_3 statistic was developed as a test against the LSTAR alternative, it has power against ESTAR alternatives as well. An intuitive way to understand this is to note that all auxiliary regressors in the first-order approximation to the ESTAR model in (22) are contained in (20). This suggests that the appropriate transition variable in the STAR model can be determined first, without specifying the form of the transition function, by computing the LM_3 statistic for various candidate transition variables $s_{1t}, \ldots, s_{mt}$, say, and selecting the one for which the p-value of the test is smallest. The rationale behind this procedure is that the test should have maximum power in case the alternative model is correctly specified, that is, if the correct transition variable is used. Simulation results in Teräsvirta (1994) suggest that this approach works quite well. Notice that the significance level of the linearity test is not under control in this selection procedure. This is not problematic however, as the test statistic is used here as a model specification tool rather than as a strict linearity test.

This selection procedure may be preceded by a general test for STAR nonlinearity assuming only that the appropriate transition variable is one of the candidates $s_{1t}, \ldots, s_{mt}$, by computing the LM_3 statistic with transition variable $s_t = \sum_{i=1}^{m} \alpha_i s_{it}$, with $\alpha_d = 1$ for certain $d \in \{1, \ldots, m\}$ and $\alpha_i = 0$ for $i \neq d$.

Selecting the transition function

When linearity is rejected in favor of STAR nonlinearity and the transition variable has been selected, the final decision to be made at this stage concerns the appropriate form of the transition function $G(s_t; \gamma, c)$. In practice, the choice may be limited to that between the first-order logistic function (3) on the one hand and the exponential function (4) or the second-order logistic function (5) on the other. Consider the following sequence of null hypotheses:

$$
\begin{aligned}
H_{03} &: \ \beta_3 = 0, \\
H_{02} &: \ \beta_2 = 0 |\ \beta_3 = 0, \\
H_{01} &: \ \beta_1 = 0 |\ \beta_3 = \beta_2 = 0,
\end{aligned}
$$

in (20), all of which can be tested by LM-type tests. Closer inspection of the expressions of the auxiliary parameters β_1, β_2 and β_3 in terms of parameters of the original STAR model reveals that (i) $\beta_3 \neq 0$ only if the model is an LSTAR model, (ii) $\beta_2 = 0$ if the model is an LSTAR model with $\phi_{1,0} = \phi_{2,0}$ and $c = 0$ but is always nonzero if the model is an ESTAR model, and (iii) that $\beta_1 = 0$ if the model is an ESTAR model with $\phi_{1,0} = \phi_{2,0}$ and $c = 0$ but is always nonzero if the model is an LSTAR model. Combining these three properties of the auxiliary parameters leads to the following decision rule: if the p-value of the test corresponding to H_{02} is the smallest, an ESTAR model should be selected, while in all other cases an LSTAR model is to be the preferred choice.

Escribano and Jordá (1999) propose an alternative transition function selection procedure, which makes use of LM_4 as a test for general STAR nonlinearity. Their decision rule for choosing between the LSTAR and ESTAR alternatives is based on the observation that, assuming $\phi_{1,0} = \phi_{2,0}$ and $c = 0$ in (18), the properties of β_1 and β_2 given above also apply to β_3 and β_4 in (23), respectively. Therefore they suggest to test the hypotheses

$$H_{0E}: \ \beta_2 = \beta_4 = 0,$$
$$H_{0L}: \ \beta_1 = \beta_3 = 0,$$

in (23) and to select an LSTAR (ESTAR) model if the minimum p-value is obtained for H_{0L} (H_{0E}).

This is a neat idea in that it corrects an asymmetry in the original selection rule. When the true model is an ESTAR model behaving almost like an LSTAR one (see Teräsvirta, 1994, for discussion), the original rule often tends to choose the LSTAR model, at least in small samples. The Escribano-Jordá rule does not have this property. On the other hand, if the true model is an LSTAR model or an ESTAR model which can not be approximated adequately with an LSTAR model, both rules lead to selecting this model with high probability. In general, neither procedure dominates the other.

Recent increases in computational power have made these decision rules less important in practice. It is now easy to estimate a number of both LSTAR and ESTAR models and to choose between them at the evaluation stage by misspecification tests. It is also possible to develop non-nested hypothesis tests for distinguishing between these two families of models. Nevertheless, the two decision rules seem to work well in practice, and carrying out the tests may be recommended even if the actual decision were postponed to the evaluation stage of the modelling cycle.

5.2 Estimation

Once the transition variable s_t and the transition function $G(s_t; \gamma, c)$ have been selected, the next stage in the modelling cycle is estimation of the parameters in the STAR model.

The discussion below is framed in terms of the basic two-regime model, but the issues that are addressed also apply to the MRSTAR and TVSTAR models.

Estimation of the parameters in the STAR model (2) is a relatively straightforward application of nonlinear least squares [NLS], that is, the parameters $\theta = (\phi_1', \phi_2', \gamma, c)'$ can be estimated as

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \; Q_T(\theta) = \underset{\theta}{\operatorname{argmin}} \; \sum_{t=1}^{T} \left(y_t - F(x_t;\theta)\right)^2, \tag{31}$$

where $F(x_t;\theta)$ is the skeleton of the model given in (26). Under the additional assumption that the errors ε_t are normally distributed, NLS is equivalent to maximum likelihood. Otherwise, the NLS estimates can be interpreted as quasi maximum likelihood estimates. Under certain regularity conditions, which are discussed in Wooldridge (1994) and Pötscher and Prucha (1997), among others, the NLS estimates are consistent and asymptotically normal, that is,

$$\sqrt{T}(\hat{\theta} - \theta_0) \to N(0, C), \tag{32}$$

where θ_0 denotes the true parameter values. The asymptotic covariance-matrix C of $\hat{\theta}$ can be estimated consistently as $\widehat{A}_T^{-1}\widehat{B}_T\widehat{A}_T^{-1}$, where $\widehat{A}_T$ is the Hessian evaluated at $\hat{\theta}$

$$\widehat{A}_T = -\frac{1}{T}\sum_{t=1}^{T} \nabla^2 q_t(\hat{\theta}) = \frac{1}{T}\sum_{t=1}^{T}\left(\nabla F(x_t;\hat{\theta})\nabla F(x_t;\hat{\theta})' - \nabla^2 F(x_t;\hat{\theta})\hat{\varepsilon}_t\right), \tag{33}$$

with $q_t(\hat{\theta}) = (y_t - F(x_t;\hat{\theta}))^2$, and $\widehat{B}_T$ is the outer product of the gradient

$$\widehat{B}_T = \frac{1}{T}\sum_{t=1}^{T} \nabla q_t(\hat{\theta})\nabla q_t(\hat{\theta})' = \frac{1}{T}\sum_{t=1}^{T} \hat{\varepsilon}_t^2 \nabla F(x_t;\hat{\theta})\nabla F(x_t;\hat{\theta})'. \tag{34}$$

The estimation can in principle be performed using any conventional nonlinear optimization procedure, see Quandt (1983), Hamilton (1994, Section 5.7) and Hendry (1995, Appendix A5) for surveys. Issues deserving particular attention here include concentrating the sum of squares function, the choice of starting-values for the parameters, and the estimation of the smoothness parameter γ in the transition function.

Concentrating the sum of squares function

As suggested by Leybourne, Newbold and Vougas (1998), the estimation problem can be simplified by concentrating the sum of squares function. Note that when the parameters γ and c in the transition function are known and fixed, the STAR model is linear in the autoregressive parameters ϕ_1 and ϕ_2. Conditional upon γ and c, estimates of $\phi = (\phi_1', \phi_2')'$

can be obtained by ordinary least squares [OLS] as

$$\hat{\phi}(\gamma, c) = \left(\sum_{t=1}^{T} x_t(\gamma, c)x_t(\gamma, c)'\right)^{-1} \left(\sum_{t=1}^{T} x_t(\gamma, c)y_t\right), \tag{35}$$

where $x_t(\gamma, c) = (x_t'(1 - G(s_t; \gamma, c)), x_t'G(s_t; \gamma, c))'$ and the notation $\phi(\gamma, c)$ is used to indicate that the estimate of ϕ is conditional upon γ and c. Thus, the sum of squares function $Q_T(\theta)$ can be concentrated with respect to ϕ_1 and ϕ_2 as

$$Q_T(\gamma, c) = \sum_{t=1}^{T} (y_t - \phi(\gamma, c)'x_t(\gamma, c))^2.$$

This reduces the dimensionality of the NLS estimation problem considerably, as $Q(\gamma, c)$ needs to be minimized with respect to the two parameters γ and c only.

Starting values

From the conditional linearity of the STAR model, it immediately follows that sensible starting-values for the nonlinear optimization can be easily obtained by a two-dimensional grid search over γ and c. Replacing the transition function (36) by

$$G(s_t; \gamma, c) = (1 + \exp\{-\gamma \prod_{i=1}^{n} (s_t - c_i)/\hat{\sigma}_{s_t}^n\})^{-1}, \tag{36}$$

where $\hat{\sigma}_{s_t}$ is the sample standard deviation of s_t, makes γ approximately scale-free. This helps in determining a useful set of grid values for this parameter. A meaningful set of grid values for the location parameter c may be defined as sample percentiles of the transition variable s_t. This guarantees that the values of the transition function contain enough sample variation for each choice of γ and c. If the transition function remains almost constant in the whole sample, the moment matrix of the regression (35) is ill-conditioned, and the estimation fails.

It should be noted that if the logistic function (6) is used with $n > 1$, or if an MRSTAR model (10) or a TVSTAR model (12) is estimated, the dimension of the grid increases. Still, the grid search is worth undertaking, as it likely renders starting-values which are reasonably close to the optimum. This reduces the burden on the nonlinear optimization considerably. Furthermore, if analytical second derivatives are used in computing the Hessian (as in the Newton-Raphson method, for example), good starting-values are absolutely necessary for convergence of the algorithm.

The estimate of γ

It is difficult to obtain a very accurate estimate of the smoothness of the transition between the two regimes, characterized by γ, when this parameter is large. This is due to the fact that for such large values of γ, the STAR model is similar to a threshold model, as the transition function comes close to a step function. To obtain an accurate estimate of γ, one then needs many observations in the immediate neighborhood of c, because even large changes in γ only have a small effect on the shape of the transition function. The estimate of γ may therefore be rather imprecise and often appear to be insignificant when judged by its t-statistic, see Bates and Watts (1988, p.87) for discussion in a more general context. This should, however, not be interpreted as evidence for weak nonlinearity, as the t-statistic does not have its customary asymptotic t-distribution under the hypothesis that $\gamma = 0$, due to the identification problems discussed in Section 4.1. In this situation, the causes of a large standard error estimate are purely numerical. Besides, for the very reason that large changes in γ have only a minor effect on the transition function, high accuracy in estimating γ is not necessary.

5.3 Evaluation

After estimating the parameters in a STAR model, the next stage in the modelling cycle is evaluation of the properties of the fitted model. Next to 'common sense' diagnostics, such as examining the properties of the skeleton and inspecting the regimes that are implied by the model, the model should be subjected to misspecification tests such as the ones discussed in Section 4.2. Rejection of one or more of the null hypotheses should lead to reconsidering the specification of the model. Other methods to evaluate the properties of estimated STAR models include local or sliced spectra and impulse response analysis, which are discussed below.

5.3.1 Sliced spectra

Parameter estimates generally do not provide much information about the dynamics of an estimated STAR model. To characterize local dynamic behaviour, one can compute the roots of the characteristic polynomial of the model for given values of the transition function $G(s_t;\gamma,c)$, as in Teräsvirta and Anderson (1992) and Teräsvirta (1994), among others. Skalin and Teräsvirta (1999) suggest a more economic way to summarize the local dynamics by using the local or sliced spectrum of the STAR model (1). This is defined as

$$\begin{aligned} f_{yy}(\omega;s_t) = \frac{1}{2\pi}\Bigg[& \left\{1-\sum_{j=1}^{p}\phi_{1,j}(1-G(s_t;\gamma,c))+\phi_{2,j}G(s_t;\gamma,c)\exp^{-ij\omega}\right\} \\ & \times\left\{1-\sum_{j=1}^{p}\phi_{1,j}(1-G(s_t;\gamma,c))+\phi_{2,j}G(s_t;\gamma,c)\exp^{ij\omega}\right\}\Bigg]^{-1}, \quad (37) \end{aligned}$$

for $-\pi \leq \omega \leq \pi$, see Priestley (1981, Section 4.12). Obviously, the spectrum is defined only for points where the estimated STAR model is locally stationary, that is, only for those values of $G(s_t;\gamma,c)$ for which the roots of the lag polynomial

$$1-\sum_{j=1}^{p}(\phi_{1,j}(1-G(s_t;\gamma,c))+\phi_{2,j}G(s_t;\gamma,c))L^j,$$

where L is the lag operator, have modulus larger than unity.

The 'global' dynamics of an estimated STAR model are better characterized by a 'model' spectrum. As this cannot be computed analytically, it has to be obtained by simulation.

5.3.2 Impulse response functions

Another useful way of considering the dynamic behaviour of an estimated STAR model is to examine the effects of the shocks ε_t on the future patterns of the time series y_t. Impulse response functions are a convenient tool to carry out such an analysis, as they provide a measure of the response of y_{t+h}, $h=1,2,\ldots$ to a shock or impulse δ at time t.

The impulse response commonly used in the analysis of linear models is defined as the difference between two realizations of y_{t+h} which start from identical histories of the time series up to time $t-1$, denoted as ω_{t-1}. In one realization, the process is 'hit' by a shock of size δ at time t, while in the other realization no shock occurs at time t. All shocks in intermediate periods between t and $t+h$ are set equal to zero in both realizations. That is, the traditional impulse response function [TI] is given by

$$\begin{aligned}\mathrm{TI}_y(h,\delta,\omega_{t-1}) = \mathrm{E}[y_{t+h}|\varepsilon_t=\delta,\varepsilon_{t+1}=\ldots=\varepsilon_{t+h}=0,\omega_{t-1}] - \\ \mathrm{E}[y_{t+h}|\varepsilon_t=0,\varepsilon_{t+1}=\ldots=\varepsilon_{t+h}=0,\omega_{t-1}],\end{aligned} \tag{38}$$

for $h=0,1,2,\ldots$. The second conditional expectation usually is called the benchmark profile.

The traditional impulse response function as defined above has some simple properties when the underlying model is linear. First, the TI is *symmetric*, in the sense that a shock of size $-\delta$ has an effect exactly opposite to that of a shock of size $+\delta$. Furthermore, it might be called *linear*, as the impulse response is proportional to the size of the shock. Finally, the impulse response is *history independent* as its shape does not depend on the particular history ω_{t-1}. For example, in the AR(1) model $y_t=\phi_0+\phi_1 y_{t-1}+\varepsilon_t$, it follows easily that the impulse response $\mathrm{TI}_y(h,\delta,\omega_{t-1})=\phi_1^h\delta$, $h=0,1,2,\ldots$.

These properties do not carry over to nonlinear models, including the STAR model. In nonlinear models, the impact of a shock depends on the history of the process, as well as on the sign and the size of the shock. Furthermore, if the effect of a shock on the time series $h > 1$ periods ahead is to be analyzed, the assumption that no shocks occur in intermediate periods may give a misleading picture of the propagation mechanism of the model. Pesaran and Potter (1997) provide an example.

The Generalized Impulse Response Function [GI], introduced by Koop *et al.* (1996) offers a useful generalization of the concept of impulse response to nonlinear models. The GI for a specific shock $\varepsilon_t = \delta$ and history ω_{t-1} is defined as

$$\mathrm{GI}_y(h, \delta, \omega_{t-1}) = \mathrm{E}[y_{t+h}|\varepsilon_t = \delta, \omega_{t-1}] - \mathrm{E}[y_{t+h}|\omega_{t-1}], \tag{39}$$

for $h = 0, 1, 2, \ldots$. In the GI, the expectation of y_{t+h} given that a shock δ occurs at time t is conditioned only on the history and this shock. Put differently, the problem of dealing with shocks occurring in intermediate time periods is handled by averaging them out. Given this choice, the natural benchmark profile for the impulse response is the expectation of y_{t+h} conditional only on the history of the process ω_{t-1}. Thus, in the benchmark profile the current shock is averaged out as well. It is easily seen that for linear models the GI is equivalent to the TI.

The GI is a function of δ and ω_{t-1} that are realizations of the random variables ε_t and Ω_{t-1}. Koop *et al.* (1996) point out that $\mathrm{GI}_y(h, \delta, \omega_{t-1})$ itself is a realization of a random variable, defined as

$$\mathrm{GI}_y(h, \varepsilon_t, \Omega_{t-1}) = \mathrm{E}[y_{t+h}|\varepsilon_t, \Omega_{t-1}] - \mathrm{E}[y_{t+h}|\Omega_{t-1}]. \tag{40}$$

Definition (40) allows a number of conditional versions of potential interest. For example, one might consider only a particular history ω_{t-1} and treat the GI as a random variable in terms of ε_t, that is,

$$\mathrm{GI}_y(h, \varepsilon_t, \omega_{t-1}) = \mathrm{E}[y_{t+h}|\varepsilon_t, \omega_{t-1}] - \mathrm{E}[y_{t+h}|\omega_{t-1}]. \tag{41}$$

It is equally possibly to reverse the roles of the shock and the history by fixing the shock at $\varepsilon_t = \delta$ and defining the GI to be a random variable with respect to the history Ω_{t-1}. In general, one might consider the GI to be random conditional on particular subsets A and B of shocks and histories respectively, that is, $\mathrm{GI}_y(h, A, B)$. For example, one might condition on all histories in a particular regime and consider only negative shocks.

In case of the STAR model, analytic expressions for the conditional expectations involved in the GI are not available for $h > 1$. Stochastic simulation has to be used to obtain estimates of the impulse response measures. See Koop *et al.* (1996) for a detailed description of the relevant techniques.

Measuring persistence of shocks

A shock $\varepsilon_t = \delta$ is said to be *transient* at history ω_{t-1} if $\mathrm{GI}_y(h, \delta, \omega_{t-1})$ becomes equal to 0 as

$h \to \infty$. If this is not the case, the shock is said to be persistent. It is intuitively clear that if a time series process is stationary and ergodic, the effects of all shocks eventually converge to zero for all possible histories of the process. Hence, the distribution of $\mathrm{GI}_y(h, \varepsilon_t, \Omega_{t-1})$ collapses to a spike at 0 as $h \to \infty$. By contrast, for nonstationary time series the dispersion of the distribution of $\mathrm{GI}_y(h, \varepsilon_t, \Omega_{t-1})$ is positive for all h. Potter (1995) and Koop *et al.* (1996) suggest that the dispersion of the distribution of $\mathrm{GI}_y(h, \varepsilon_t, \Omega_{t-1})$ at finite horizons can be interpreted as a measure of persistence of shocks. Conditional versions of the GI are particularly suited to assess this persistence. For example, one can compare densities of GIs conditional on positive and negative shocks to find out whether, say, negative shocks are more persistent than positive ones, or vice versa. The notion of second-order stochastic dominance might be a useful measure of dispersion in this context, see Potter (2000).

Measuring asymmetric impulse response

Another use of the GI is to assess the significance of asymmetric effects over time. Potter (1994) defines a measure of asymmetric response to a particular shock $\varepsilon_t = \delta$ given a particular history ω_{t-1} as the sum of the GI for this particular shock and the GI for the shock of the same magnitude but with opposite sign, that is,

$$\mathrm{ASY}_y(h, \delta, \omega_{t-1}) = \mathrm{GI}_y(h, \delta, \omega_{t-1}) + \mathrm{GI}_y(h, -\delta, \omega_{t-1}). \tag{42}$$

By taking into account parameter uncertainty as an additional source of randomness, $\mathrm{ASY}_y(h, \delta, \omega_{t-1})$ can still be interpreted as a random variable. Potter (1995) uses a straightforward simulation procedure to assess whether the asymmetry measure is significantly different from zero or not.

Alternatively, one could consider the distribution of the random asymmetry measure

$$\mathrm{ASY}_y(h, \varepsilon_t^+, \Omega_{t-1}) = \mathrm{GI}_y(h, \varepsilon_t^+, \Omega_{t-1}) + \mathrm{GI}_y(h, -\varepsilon_t^+, \Omega_{t-1}) \tag{43}$$

where $\varepsilon_t^+ = \{\varepsilon_t | \varepsilon_t > 0\}$ indicates the set of all possible positive shocks. If positive and negative shocks have exactly the same effect (with opposite sign), $\mathrm{ASY}_y(h, \varepsilon_t^+, \Omega_{t-1})$ should be equal to zero almost surely. More generally, we say that shocks have a symmetric effect (on average) when $\mathrm{ASY}_y(h, \varepsilon_t^+, \Omega_{t-1})$ has a symmetric distribution with mean equal to zero. The dispersion of this distribution might be interpreted as a measure of the asymmetry in the effects of positive and negative shocks.

6 Forecasting

Forecasting with nonlinear models is more involved than forecasting with linear models, see Tong (1990, Chapter 6) and Granger and Teräsvirta (1993, Section 8.1) for general reviews. In this section we discuss several issues related to out-of-sample forecasting with STAR models, see also Lundbergh and Teräsvirta (2001). Techniques for constructing point and interval forecasts are considered in Sections 6.1 and 6.2, respectively. In Section 6.3 the questions of how to evaluate forecasts from STAR models and how to compare forecasts from linear and STAR models in particular are addressed.

6.1 Point forecasts

Consider the case where y_t is described by the STAR model (1) with $s_t = y_{t-1}$, that is

$$y_t = F(x_t; \theta) + \varepsilon_t, \tag{44}$$

where $F(x_t; \theta)$ is given by

$$F(x_t; \theta) = \phi_1' x_t (1 - G(y_{t-1}; \gamma, c)) + \phi_2' x_t G(y_{t-1}; \gamma, c), \tag{45}$$

with $x_t = (1, y_{t-1}, \ldots, y_{t-p})'$. Denote the optimal point forecast of y_{t+h} made at time t as $\hat{y}_{t+h|t} = \mathrm{E}[y_{t+h}|\Omega_t]$, and the associated forecast or prediction error as $e_{t+h|t} = y_{t+h} - \hat{y}_{t+h|t}$.

Using the fact that $\mathrm{E}[\varepsilon_{t+1}|\Omega_t] = 0$, the optimal 1-step ahead forecast of y_{t+1} is easily obtained as $\hat{y}_{t+1|t} = \mathrm{E}[y_{t+1}|\Omega_t] = F(x_{t+1}; \theta)$, which is equivalent to the optimal 1-step ahead forecast in case the model $F(x_t; \theta)$ is linear. When the forecast horizon is larger than 1 period, things become more complicated however. For example, the optimal 2-step ahead forecast follows from (44) as

$$\hat{y}_{t+2|t} = \mathrm{E}[y_{t+2}|\Omega_t] = \mathrm{E}[F(\hat{x}_{t+2|t}; \theta)|\Omega_t], \tag{46}$$

where $\hat{x}_{t+2|t} = (1, \hat{y}_{t+1|t} + \varepsilon_{t+1}, y_t, \ldots, y_{t-(p-2)})'$. The exact expression for (46) is

$$\hat{y}_{t+2|t} = \int_{-\infty}^{\infty} F(\hat{x}_{t+2|t}; \theta) f(\varepsilon) d\varepsilon, \tag{47}$$

where f denotes the density of ε_{t+1}. As an analytic expression for the integral (47) is not available, it needs to be approximated using numerical techniques. Even though such numerical integration is not complicated, the dimension of the integral grows with the forecast horizon, which makes forecasting rather time-consuming and possibly inaccurate. Several methods to obtain forecasts while avoiding numerical integration have been developed.

First, by extrapolating the skeleton (45) a 2-step ahead forecast can be obtained as

$$\hat{y}^{(n)}_{t+2|t} = F(\hat{x}_{t+2|t}; \theta). \tag{48}$$

This is called the 'naïve' approach (Granger and Teräsvirta, 1993), as it effectively boils down to interchanging the (*linear*) conditional expectation operator E with the (*nonlinear*) operator in (46). As $\mathrm{E}[F(\hat{x}_{t+2|t}; \theta)] \neq F(\mathrm{E}[\hat{x}_{t+2|t}; \theta])$, this approach renders biased forecasts.

An alternative approach is to use Monte Carlo or bootstrap methods to approximate the conditional expectation (46). The 2-step ahead Monte Carlo forecast is given by

$$\hat{y}^{(mc)}_{t+2|t} = \frac{1}{k}\sum_{i=1}^{k} F(\hat{x}^{(i)}_{t+2|t};\theta), \tag{49}$$

where k is some large number and the values of ε_{t+1} in $\hat{x}^{(i)}_{t+2|t}$ are drawn from the presumed distribution of ε_t. The bootstrap forecast $\hat{y}^{(b)}_{t+2|t}$ is very similar, the only difference being that the ε_{t+1} are drawn with replacement from the residuals from the estimated model, $\hat{\varepsilon}_t, t = 1, \ldots, T$. The advantage of the bootstrap over the Monte Carlo method is that no assumptions need to be made concerning the distribution of ε_t.

Clements and Smith (1997) compare various methods of obtaining multiple-step ahead forecasts for SETAR models, respectively. Their main findings are that the Monte Carlo and bootstrap methods compare favorably to the exact and naïve methods. An additional advantage of the Monte Carlo and bootstrap methods is that the individual realizations $F(\hat{x}^{(i)}_{t+2|t};\theta)$ effectively form a forecast density, which can be used to construct interval forecasts as discussed below.

Note that in the above we have assumed that the parameters in the STAR model are known. In practice the parameters of course have to be estimated, which leads to additional forecast uncertainty. This sampling uncertainty can be taken into account by extending the Monte Carlo or bootstrap forecast with an additional averaging over different parameter values as

$$\hat{y}^{(mc)}_{t+2|t} = \frac{1}{kr}\sum_{i=1}^{k}\sum_{j=1}^{r} F(\hat{x}^{(i)}_{t+2|t};\theta^{(j)}), \tag{50}$$

where the $\theta^{(j)}$ are drawn from the large-sample distribution of the parameter estimates $\hat{\theta}$.

6.2 Interval forecasts

Point forecasts may be combined with confidence intervals. For forecasts obtained from linear models, the standard forecast confidence region is taken to be a symmetric interval around the point forecast. This is the case because the conditional distribution $g(y_{t+h}|\Omega_t)$ of a linear time series is symmetric around $\mathrm{E}[y_{t+h}|\Omega_t]$ (which is estimated by $\hat{y}_{t+h|t}$). For nonlinear models, the conditional distribution $g(y_{t+h}|\Omega_t)$ need not be symmetric, and it can even contain multiple modes. This is possible for STAR models as well. How to construct forecast confidence regions in this situation is discussed in detail in Hyndman (1995). He lists three methods of defining a $100(1-\alpha)\%$ forecast region for $\hat{y}_{t+h|t}$:

1. An interval symmetric around the point forecast

$$S_\alpha = (\hat{y}_{t+h|t} - w, \hat{y}_{t+h|t} + w),$$

where $w > 0$ is such that $P(y_{t+h} \in S_\alpha|\Omega_t) = 1 - \alpha$.

2. The interval defined by the $\alpha/2$ and $(1-\alpha/2)$ quantiles of the forecast distribution, denoted $q_{\alpha/2}$ and $q_{1-\alpha/2}$, respectively,

$$Q_\alpha = (q_{\alpha/2}, q_{1-\alpha/2}).$$

3. The highest-density region [HDR]

$$HDR_\alpha = \{y_{t+h}|g(y_{t+h}|\Omega_t) \geq g_\alpha\}, \tag{51}$$

where $g(\cdot)$ is the density of its argument and g_α is such that $P(y_{t+h} \in HDR_\alpha|\Omega_t) = 1 - \alpha$.

For symmetric and unimodal distributions, these three regions are identical. For asymmetric or multimodal distributions they are not. Hyndman (1995) argues that the HDR is the most natural choice, for the following reasons. First, HDR_α is the smallest of all possible $100(1-\alpha)\%$ forecast regions. Second, every point inside the HDR has conditional density $g(y_{t+h}|\Omega_t)$ at least as large as every point outside the region. Furthermore, only the HDR will immediately reveal features such as asymmetry or multimodality of the conditional distribution $g(y_{t+h}|\Omega_t)$. HDRs are straightforward to compute when the Monte Carlo or bootstrap methods are used to compute the point forecast $\hat{y}_{t+h|t}$. Let $y^i_{t+h|t}$, $i = 1, 2, \ldots$, denote the i-th element used in computing the Monte Carlo forecast (49). Note that the $y^i_{t+h|t}$ can be thought of as being realizations drawn from the conditional distribution of interest $g(y_{t+h}|\Omega_t)$. Estimates $g_i \equiv g(y^i_{t+h|t}|\Omega_t)$, $i = 1, \ldots, k$, then can be obtained by using a standard kernel density estimator, that is

$$g_i = \frac{1}{k}\sum_{j=1}^{k} K([y^i_{t+h|t} - y^j_{t+h|t}]/b), \tag{52}$$

where $K(\cdot)$ is a kernel function such as the Gaussian density and $b > 0$ is the bandwidth. An estimate of g_α in (51) is given by $\hat{g}_\alpha = g_{(\lfloor \alpha k \rfloor)}$, where the $g_{(i)}$ are the ordered g_i and $\lfloor \cdot \rfloor$ denotes integer part. See Hyndman (1996) for more details. Finally, it should be noted that HDR's are also useful in summarizing sets of GI densities, see Section 7 and Skalin and Teräsvirta (1999, 2000) for applications to STAR models.

6.3 Evaluating forecasts

It is good practice to evaluate the quality of forecasts from a time series model. Relative forecast performance can also be used as a model selection criterion, as an alternative or complement to an in-sample comparison of different models. Out-of-sample forecasting thus can also be considered as a way to evaluate estimated models. Especially comparison of the forecasts from nonlinear models with those from a benchmark linear model might enable one to determine the added value of the nonlinear features of the model.

Traditionally, forecasts are evaluated using the mean squared prediction error [MSPE], MSPE $= \frac{1}{m}\sum_{j=0}^{m-1}(\hat{y}_{T+h+j|T+j} - y_{T+h+j})^2$, where m is the number of h-steps ahead forecasts, and related criteria. Models with smaller MSPE have a better forecast performance. Diebold and Mariano (1995) discuss various statistics that can be used to examine whether the MSPEs of two alternative models are significantly different from each other.

Experience with (empirical) forecasting with STAR models is rather limited. Teräsvirta and Anderson (1992) obtain mixed results for forecasts of quarterly industrial production series. In some cases the STAR model yields better (1-step ahead) point forecasts than a linear model, in other cases the forecasts are worse. Sarantis (1999) uses STAR models to forecast monthly real effective exchange rates of major industrialized countries with similar results, in that STAR models do not consistently produce more or less accurate forecasts than linear models. The STAR model is found to outperform a Markov-Switching model though. Finally, Stock and Watson (1999) apply STAR models to forecast a large number of monthly US macroeconomic time series and find that on average forecasts from STAR models do not improve upon forecasts from linear models, although they do improve upon forecasts from neural networks.

In general, the fact that a nonlinear model describes the features of a time series within the estimation sample better than a linear model is no guarantee that the nonlinear model also renders better out-of-sample forecasts. In fact, it is reported quite often that, even though a nonlinear model appears to describe certain characteristics of the time series at hand much better than a linear model, the forecasting performance of a linear model is no worse or even better than that of the nonlinear model, see de Gooijer and Kumar (1992) among others. Clements and Hendry (1998) discuss reasons for this, see also Diebold and Nason (1990). For example, the nonlinearity may be 'spurious', in the sense that other features of the time series, such as heteroskedasticity, structural breaks or outliers, suggest the presence of nonlinearity. Even though one may successfully estimate a nonlinear model for such a series, it is not obvious that this will result in improved forecasts compared to ones from linear models.

Another possible cause for the (relatively) poor forecast performance of nonlinear models is that traditional criteria such as the MSPE might not do full justice to the nonlinear model. As noted by Tong (1995), 'how well we can forecast depends on where we are.' In case of a STAR model, it might very well be that the forecastability of the time series is different in the different regimes corresponding with $G(s_t; \gamma, c) = 0$ and $G(s_t; \gamma, c) = 1$. One therefore might evaluate the forecasts for each regime separately to investigate whether the nonlinear model is especially useful to obtain forecasts in a particular regime or state, see Tiao and Tsay (1994) and Clements and Smith (1999) for applications to SETAR models.

Also, MSPE and related criteria focus on the quality of *point* forecasts. It might very well be that the real strength of a nonlinear model lies in producing superior interval and density forecasts, see Pesaran and Potter (1997) for an example. Thus, to assess the merits of a nonlinear model it might be worthwhile to evaluate interval and density forecasts as well, using techniques recently developed in Christoffersen (1998) and Diebold, Gunther and Tay (1998).

Finally, a nonlinear model may not render better forecasts simply because the nonlinearity does not show up during the forecast period. In case of the STAR model, for example, it might be that only one of the regimes is realized during the forecast period. Hence, empirical forecasts do not always allow to assess the forecasting quality of the STAR model completely. A potential solution to this problem is to perform a simulation experiment in which an estimated STAR model is used to generate artificial time series and an out-of-sample forecasting exercise is performed on each of those series using AR and STAR models. In this controlled environment one can make sure that forecasts in each of the regimes are involved. See Clements and Smith (1999,2000) for applications of this approach with SETAR models. This simulation approach can also be applied to compare the forecast performance of alternative nonlinear models by using each of the alternatives as DGP in turn, see Clements and Krolzig (1998).

As a final remark, notice that the second conditional expectation in the right-hand side of (39) is the optimal point forecast of y_{t+h} at time $t-1$, whereas the first conditional expectation can be interpreted as the optimal forecast of y_{t+h} at time t in case $\varepsilon_t = \delta$. Therefore the GI can be interpreted as the change in forecast of y_{t+h} at time t relative to time $t-1$, given that a shock δ occurs at time t. This also suggests that if the density of the conditional GI (41) (or other versions of the GI) effectively is a spike at zero for certain $h \geq m$, the nonlinear model is not useful for forecasting more than m periods ahead.

7 Modelling US unemployment with STAR models

As discussed in Section 2, STAR models have been successfully applied to describing the behaviour of various macro-economic time series, such as output and (un)employment, at different phases of the business cycle. In this section, we analyze a US unemployment rate series to illustrate the modelling cycle for STAR models.

The series we consider represents the seasonally unadjusted unemployment rate among US males aged 20 and over, at the monthly frequency covering the period June 1968 until December 1999 (379 observations). The series is constructed by taking the ratio of the unemployment level and civilian labor force of this population group, which are obtained from the *Bureau of Labor Statistics*.

- insert Figure 1 about here -

The series, which is shown in Figure 1, contains three dominant features: asymmetric behaviour over the business cycle, large persistence, and a distinct seasonal pattern. The behaviour of the unemployment rate over the business cycle can be characterized as steep increases during recessions, followed by slow(er) declines during expansions. Several theories, such as asymmetric labor adjustment costs of enterprises, insider-outsider relationships, and asymmetries in capital destruction and reconstruction have been developed to explain this asymmetry in the dynamic behaviour of the unemployment rate. Plenty of evidence of this type of nonlinearity has been compounded over the years. Using nonparametric methods, Neftçi (1984) found that increases in the unemployment rate are steeper than decreases, see also Sichel (1989) and Rothman (1991). Furthermore, Neftçi (1993) showed that conventional linear models are able to replicate the observed patterns in the unemployment rate only with very small probability. Various parametric nonlinear time series models have also been fitted to a number of unemployment rates, see Peel and Speight (1996), Hansen (1997), Bianchi and Zoega (1998), Montgomery, Zarnowitz, Tsay and Tiao (1998), Rothman (1998), Brännäs and Ohlsson (1999), Koop and Potter (1999), Caner and Hansen (2000), and Skalin and Teräsvirta (2000), among others. In general, it is found that nonlinear models improve upon linear models both in describing the in-sample properties of the unemployment rate and in out-of-sample forecasting.

From Figure 1 it is also clear that the unemployment rate is very persistent. In fact, the persistence of the unemployment rate has received much more attention than its asymmetry properties. The two competing viewpoints are the 'natural rate' hypothesis and the hysteresis hypothesis of Blanchard and Summers (1987). Under the natural rate hypothesis, the unemployment rate is mean-reverting, whereas it is non-stationary under the hysteresis hypothesis. Thus, the two hypotheses imply that different transformations (levels and first differences, respectively) of the unemployment rate are appropriate. Here we follow Bianchi and Zoega (1998) and Skalin and Teräsvirta (2000) by assuming that the unemployment rate is globally stationary but possibly nonlinear and locally nonstationary.

Finally, Figure 1 also shows that the unemployment rate contains a pronounced seasonal pattern. Typically, unemployment is above average during the winter (January-March) and below average during the late summer and fall (August-October). We assume that the systematic component of seasonality can be adequately captured by monthly dummy variables, which are denoted as $D_{i,t}$, $i = 1, \ldots, 11$, where $D_{i,t} = 1$ if observation t corresponds to month i and $D_{i,t} = 0$ otherwise.

We use the series up to December 1989 for estimation and testing and reserve the final 10 years for out-of-sample forecasting. Following the modelling cycle as outlined in Section 5, we start by specifying a linear AR model. In order to anticipate the structure of the STAR model, the AR model is parameterized in first differences, including a single level term at the first lag. We allow for a maximum of $p_{\max} = 18$ lagged first differences, such that the effective estimation sample runs from January 1970 until December 1989 (240 observations). Both AIC and BIC indicate that an AR(p) model with 3 lagged first differences is appropriate. Upon estimation we find that this model is too parsimonious, as it leaves considerable autocorrelation in the residuals. This problem is solved only once the number of lagged first differences is increased to 15. As quite a few of the lagged first differences in the unrestricted model are insignificant and do not contribute to the explanatory power of the model, we remove lagged first differences and seasonal dummies for which the t-statistic of the corresponding parameter is less than 1 in absolute value. We finally arrive at the following linear model:

$$\begin{aligned}
\Delta y_t = \; & \underset{(0.069)}{0.385} - \underset{(0.009)}{0.021}\, y_{t-1} + \underset{(0.060)}{0.143}\, \Delta y_{t-1} + \underset{(0.059)}{0.279}\, \Delta y_{t-2} + \underset{(0.063)}{0.169}\, \Delta y_{t-3} \\
& - \underset{(0.053)}{0.090}\, \Delta y_{t-6} - \underset{(0.057)}{0.090}\, \Delta y_{t-8} + \underset{(0.051)}{0.093}\, \Delta y_{t-11} - \underset{(0.062)}{0.069}\, \Delta y_{t-14} - \underset{(0.061)}{0.086}\, \Delta y_{t-15} \\
& + \underset{(0.074)}{0.639}\, D_{1,t} - \underset{(0.111)}{0.425}\, D_{2,t} + \underset{(0.143)}{0.786}\, D_{3,t} - \underset{(0.126)}{0.818}\, D_{4,t} - \underset{(0.068)}{0.462}\, D_{5,t} \\
& - \underset{(0.084)}{0.146}\, D_{7,t} - \underset{(0.082)}{0.394}\, D_{8,t} - \underset{(0.103)}{0.553}\, D_{9,t} - \underset{(0.069)}{0.252}\, D_{10,t} + \hat{\varepsilon}_t, \qquad (53)
\end{aligned}$$

$\hat{\sigma}_\varepsilon = 0.201$, SK $= 0.71$, EK $= 1.00$,JB $= 30(3.0 \times 10^{-7})$, ARCH(1) $= 5.76(0.02)$, ARCH(4) $=$ $8.69(0.07)$, $\text{LM}_{\text{SI}}(4) = 0.15(0.96)$, $\text{LM}_{\text{SI}}(8) = 0.31(0.96)$, $\text{LM}_{\text{SI}}(12) = 0.31(0.99)$, AIC $=$ -3.049, BIC $= -2.774$,

where OLS standard errors are given in parentheses below the parameter estimates, $\hat{\varepsilon}_t$ denotes the regression residual at time t, $\hat{\sigma}_\varepsilon$ is the residual standard deviation, SK is skewness, EK excess kurtosis, JB the Jarque-Bera test of normality of the residuals, ARCH(q) is the LM test of no ARCH effects up to order q, and $\text{LM}_{\text{SI}}(j)$ is the Breusch-Godfrey test for no residual autocorrelation up to and including lag j. The numbers in parentheses following the test statistics are p-values.

The linear model appears adequate in that the errors seem serially uncorrelated, whereas the excess kurtosis and apparent heteroskedasticity are caused entirely by large positive residuals in January 1975 and April 1980. The skewness of the errors is a more serious problem, as it does not appear to be due to only a few aberrant residuals.

The next stage is to test linearity against STAR nonlinearity using the LM-type statistics discussed in Section 4.1. As we are concerned with the behaviour of the unemployment rate over the business cycle, the transition variable in the STAR model should reflect the property that recession and expansion regimes are sustained periods of increase and decline in the unemployment rate, respectively. This makes the monthly change in the unemployment rate unsuitable as an indicator of the business cycle regime as it is too noisy. Furthermore, using the monthly (or any other intra-year) change as transition variable is impractical due to the seasonal fluctuations in the unemployment rate. Following Skalin and Teräsvirta (2000), we therefore consider the twelve-month difference as transition variable, that is, $s_t = \Delta_{12} y_{t-d} \equiv y_{t-d} - y_{t-d-12}$, $d = 1, \ldots, d_{\max}$. We set the maximum value of the delay parameter $d_{\max}$ equal to 6.

The upper three blocks of Table 1 contain p-values of the standard, heteroskedasticity robust and outlier robust LM_1, LM_3, LM_3^e and LM_4 tests with $\Delta_{12} y_{t-d}$, $d = 1, \ldots, 6$, as transition variable. LM-type tests against the alternative of smoothly changing parameters, where $s_t = t$, are given as well. The tests are based on an AR model with 15 lagged first differences under the null hypothesis, that is, the 'holes' in the model given in (53) are ignored. Linearity is tested for the monthly dummy variables and lagged dependent variables jointly and separately.

First concentrate on the tests of linearity of either all regressors or the intercept and lagged dependent variables only. The p-values of the standard LM_3 and LM_4 tests indicate that linearity can be rejected at the 10% significance level only if $\Delta_{12} y_{t-2}$ is used as transition variable. The p-values for the LM_3^e statistic indicate that $\Delta_{12} y_{t-1}$ and $\Delta_{12} y_{t-3}$ may also be considered as transition variable. The results of the robust tests suggest that the evidence for nonlinearity might perhaps be due to neglected heteroskedasticity, but not to neglected outliers. The observations that are down-weighted in the robust estimation of the linear model are indicated by whiskers on the horizontal axis in Figure 1, where the height of the whiskers is equal to one minus the weight $w_r(\cdot)$. It is seen that only very few observations receive weights smaller than 1. Note that the observations for January 1975 and April 1980, which had large residuals in the OLS estimation of the restricted linear model (53), both receive weight equal to 0. Admittedly, the statistical evidence of nonlinearity is not strong. This may be due to the use of an unrestricted AR model with 15 lagged first differences under the null hypothesis, which may reduce the power of the linearity tests. The results of the LM-type tests applied to the restricted model (53) are qualitatively similar though and do not lead to more convincing rejections of linearity.

The large p-values for the tests of linearity of the intercept and monthly dummies suggest that the seasonal pattern in the unemployment rate is constant over the business cycle and over time. This in fact allows us to investigate a possible cause for the weak evidence for nonlinearity, namely the dominant periodic features of the series. The regression of the first difference of the unemployment rate on a constant and monthly dummies already yields $R^2 = 0.75$. Hence, any nonlinearity present in the series is relatively subtle, and accounting for it would only lead to a comparatively small improvement in fit. Thus the linearity tests may not reject the null very strongly (if at all). To shed light on this issue, we compute the linearity tests for a seasonally adjusted series, which is obtained as the residuals from the regression of Δy_t on a constant and $D_{i,t}$, $i = 1, \ldots, 11$. The p-values of these tests are shown in the lower block of Table 1. It is seen that the standard tests now consistently reject linearity for $s_t = \Delta_{12} y_{t-d}$ for $d = 1, 2, 3$ at the 10% significance level, whereas the LM_1 and LM_3^e tests suggest that $\Delta_{12}y_{t-d}$ with $d = 4, 5$ or 6 may also be considered as transition variable.

- insert Table 1 about here -

Table 2 presents p-values of the LM-type statistics which test the sub-hypotheses in the specification procedures of Teräsvirta (1994) and Escribano and Jordá (1999), applied to the seasonally adjusted series. Based on the decision rule of the procedure of Teräsvirta (1994), all three variants of the tests (standard, heteroskedasticity- and outlier-robust) suggest that an LSTAR model is most appropriate for all candidate transition variables. The results from the statistics used in the Escribano-Jordá procedure generally confirm this suggestion, although sometimes they are less conclusive.

- insert Table 2 about here -

The combined evidence in Tables 1 and 2 suggests that the lagged 12-month change in the unemployment rate $\Delta_{12}y_{t-d}$ may indeed be a transition variable in an LSTAR model where only the lagged dependent variables enter nonlinearly. The appropriate value of the delay parameter d cannot be uniquely determined from the test results. For that reason we estimate LSTAR models with $s_t = \Delta_{12}y_{t-d}$ for $d = 1, 2$ and 3 and defer the choice of the delay parameter until the evaluation stage. We find that, whereas the three models provide a comparable in-sample fit, the model with $d = 1$ performs much better in terms of out-of-sample forecasting. In the following, we therefore present results for that model only.

Starting with an unrestricted AR model with 15 lagged first differences in both regimes, we sequentially remove the lagged first difference with the lowest t-statistic (in absolute value), until all parameters of the remaining lagged first differences have t-statistics exceeding 1 in absolute value. The final model is estimated as

$$
\begin{aligned}
\Delta y_t = \; & \underset{(0.064)}{0.479} + \underset{(0.065)}{0.645}\, D_{1,t} - \underset{(0.097)}{0.342}\, D_{2,t} - \underset{(0.081)}{0.680}\, D_{3,t} - \underset{(0.102)}{0.725}\, D_{4,t} - \underset{(0.090)}{0.649}\, D_{5,t} \\
& - \underset{(0.081)}{0.317}\, D_{6,t} - \underset{(0.086)}{0.410}\, D_{7,t} - \underset{(0.079)}{0.501}\, D_{8,t} - \underset{(0.086)}{0.554}\, D_{9,t} - \underset{(0.066)}{0.306}\, D_{10,t} \\
& + [- \underset{(0.008)}{0.040}\, y_{t-1} - \underset{(0.068)}{0.146}\, \Delta y_{t-1} - \underset{(0.062)}{0.101}\, \Delta y_{t-6} + \underset{(0.063)}{0.097}\, \Delta y_{t-8} \\
& - \underset{(0.063)}{0.123}\, \Delta y_{t-10} + \underset{(0.063)}{0.129}\, \Delta y_{t-13} - \underset{(0.057)}{0.103}\, \Delta y_{t-15}] \times [1 - G(\Delta_{12}y_{t-1}; \gamma, c)] \\
& + [- \underset{(0.008)}{0.011}\, y_{t-1} + \underset{(0.078)}{0.225}\, \Delta y_{t-1} + \underset{(0.076)}{0.307}\, \Delta y_{t-2} - \underset{(0.067)}{0.119}\, \Delta y_{t-7} \\
& - \underset{(0.089)}{0.155}\, \Delta y_{t-13} - \underset{(0.085)}{0.215}\, \Delta y_{t-14} - \underset{(0.085)}{0.235}\, \Delta y_{t-15}] \times G(\Delta_{12}y_{t-1}; \gamma, c)
\end{aligned} \tag{54}
$$

$$
G(\Delta_{12}y_{t-1}; \gamma, c) = (1 + \exp\{- \underset{(20.68)}{23.15}\, (\Delta_{12}y_{t-1} - \underset{(0.05)}{0.27})/\sigma_{\Delta_{12}y_{t-1}}\})^{-1} \tag{55}
$$

$\hat{\sigma}_\varepsilon = 0.185$, $\hat{\sigma}_{\text{STAR}}/\hat{\sigma}_{\text{AR}} = 0.92$, SK $= 0.64$, EK $= 0.56$, JB $= 19.5(5.9 \times 10^{-5})$, ARCH(1) $=$ $0.86(0.35)$, ARCH(4) $= 1.32(0.86)$, AIC $= -3.162$, BIC $= -2.785$,

where $\hat{\sigma}_{\text{STAR}}$ and $\hat{\sigma}_{\text{AR}}$ denote the residual standard deviations in the estimated STAR and AR models, respectively. The residual variance of the LSTAR model (54) is about 8% smaller than that of the AR model (53). Nevertheless, this is enough to compensate for the increase in the number of parameters (from 19 to 27) for the LSTAR model to be preferred over the AR model by both AIC and BIC. Both the skewness and excess kurtosis are reduced in the LSTAR model, although normality of the errors is still rejected. The tests against ARCH do not reject the null hypothesis any longer. Finally, results of the diagnostic tests in Table 3 suggest that the model is adequate as there is no evidence for remaining residual autocorrelation, time-variation in the parameters or remaining nonlinearity.

- insert Table 3 about here -

Figure 2 shows the negative of the sum of squares function $Q_T(\gamma, c)$ in the neighborhood of the NLS estimate $(\hat{\gamma}, \hat{c}) = (23.15, 0.27)$. The negative of Q_T yields a more instructive graph than Q_T itself. It is seen that the sum of squares function is essentially flat in the direction of γ for fixed values of c. This illustrates the previous discussion on large NLS estimates of γ. For large values of γ, the value of the logistic transition function changes from 0 to 1 almost instantaneously at c, and even large changes in γ have only little effect on the shape of the function. This is reflected in the standard error for $\hat{\gamma}$.

- insert Figure 2 about here -

The transition function $G(\Delta_{12}y_{t-1}; \gamma, c)$, given in (55), is shown in Figure 3, both over time and against the transition variable $\Delta_{12}y_{t-1}$. The estimates of the parameters γ and c are such that the change of the logistic function $G(\Delta_{12}y_{t-1}; \gamma, c)$ from 0 to 1 takes place for values of $\Delta_{12}y_{t-1}$ between 0 and 0.5. The bottom panel of Figure 3 also contains the (rescaled) seasonally adjusted unemployment rate obtained as described above, where circles indicate individual peaks and troughs as dated by the NBER. These peaks and troughs differ from the reference business cycle turning points, as the unemployment rate is, on average, leading at peaks and lagging at troughs. The two regimes in the LSTAR model correspond reasonably close to the contractions and expansions as identified by these turning points simply because $G(\Delta_{12}y_{t-1})$ is a monotonic transformation of the transition variable $\Delta_{12}y_{t-1}$. As the transition variable is the change in the unemployment rate over the previous year, the switches between the regimes do not coincide exactly with the peaks and troughs of the unemployment rate but usually take place a few months later.

- insert Figure 3 about here -

The dominant root of the characteristic polynomial for $G(\Delta_{12}y_{t-1};\gamma,c)=0$ is real with modulus equal to 0.97. As the value of the transition function increases, the modulus of the dominant root increases monotonically and is greater than unity for $G(\Delta_{12}y_{t-1};\gamma,c)>0.93$. At the same time, however, for $G(\Delta_{12}y_{t-1};\gamma,c)>0.83$ the dominant root becomes a complex pair, which means that the series recovers from a recession quite quickly. This is the source of nonlinearity here. Note in particular that transitions from low to high unemployment do not occur in the same fashion. This finding is confirmed by deterministic extrapolation of the LSTAR model, starting from an arbitrary point in the history of the series. Doing this for the estimated two-regime LSTAR model reveals that, irrespective of the starting point, the extrapolated series converges to a unique seasonal pattern represented by the seasonal dummy variables, where the unemployment rate varies between 4.95% in January and 3.4% in September. The extrapolated series starting from December 1989, the last month of the estimation period, is shown in Figure 4.

- insert Figure 4 about here -

To gain further insight in the dynamic properties of the estimated STAR model, we assess the propagation of shocks by computing several generalized impulse response functions. We compute history- and shock-specific GIs as defined in (39) for all observations in the estimation sample and values of the normalized initial shock equal to $\delta/\hat{\sigma}_\varepsilon=\pm3,\pm2.9,\ldots,\pm0.2,\pm0.1$, where $\hat{\sigma}_\varepsilon$ denotes the estimated standard deviation of the residuals from the LSTAR model. For each combination of history and initial shock, we compute $GI_{\Delta y}(h,\delta,\omega_{t-1})$ for horizons $h=0,1,\ldots,N$ with $N=60$. The conditional expectations in (39) are estimated as the means over 1000 realizations of Δy_{t+h}, obtained by iterating on the LSTAR model, with and without using the selected initial shock to obtain Δy_t and using randomly sampled residuals of the estimated LSTAR model elsewhere. Impulse responses for the level of the unemployment rate are obtained by accumulating the impulse responses for the first differences, that is $GI_y(h,\delta,\omega_{t-1})=\sum_{i=0}^{h}GI_{\Delta y}(i,\delta,\omega_{t-1})$.

The GI's for specific histories and shocks are used to estimate the density of $GI_y(h,A,B)$, where A and B denote sets of selected shocks and histories, respectively. The set of shocks A is the set of all negative or positive shocks, whereas the set B consists of the histories for which the value of the transition function $G(\Delta_{12}y_{t-1};\gamma,c)$ in (55) is greater ('recession') and less ('expansion') than 0.5. The densities are obtained with a standard Nadaraya-Watson kernel estimator, using $\phi(\delta/\hat{\sigma}_\varepsilon)$ as weight for $GI_y(h,\delta,\omega_{t-1})$, where $\phi(z)$ denotes the standard normal probability distribution. The reason for using this weighting scheme is that the standardized shocks $\delta/\hat{\sigma}_\varepsilon$ then effectively are sampled from a discretized normal distribution and the resulting distribution of $GI_y(h,\varepsilon_t,\Omega_{t-1})$ should resemble a normal distribution if the effect of shocks is symmetric and proportional to their magnitude (as is the case in linear models). Finally, the highest density regions are then estimated using the density quantile method outlined in Hansen (1996).

- insert Figure 5 about here -

Figure 5 shows HDRs for distributions of $\mathrm{GI}_y(h, A, B)$ for $h = 0, 3, 6, \ldots, 60$. It appears that several interesting asymmetries in the impulse responses exist. First, shocks occurring during recessions tend to be magnified during the first 6 months, after which their effect declines gradually towards zero. Shocks occurring during expansions reach their maximum effect only after 12 months, where it should also be noted that the effect of (especially) negative shocks appears to become smaller initially (during the first 3 months). Second, the effect of positive shocks during expansions is much larger than the effect of negative shocks during the first 3 years after impact. On the other hand, there does not appear to be much asymmetry between the impulse responses for positive and negative shocks occurring during recessions. The latter observations are confirmed by the measure of asymmetric impulse response $\mathrm{ASY}_y(h, \delta, \omega_{t-1})$ defined in (42). Table 4 contains means and standard deviations of the random asymmetry measures $\mathrm{ASY}_y(h, A^+, B)$ for $h = 12, 24, 36, 48$ and 60, for different sets of shocks A defined as $\mathrm{A(ll)} = \{\varepsilon_t\}$, $\mathrm{S(mall)} = \{\varepsilon_t | 1 \geq |\varepsilon_t/\hat{\sigma}_\varepsilon| > 0\}$, $\mathrm{M(edium)} = \{\varepsilon_t | 2 \geq |\varepsilon_t/\hat{\sigma}_\varepsilon| > 1\}$ and $\mathrm{L(arge)} = \{\varepsilon_t | 3 \geq |\varepsilon_t/\hat{\sigma}_\varepsilon| > 2\}$. The set B consists of all histories ('unconditional') or only of those histories for which the transition function $G(\Delta_{12} y_{t-1}; \gamma, c)$ in (55) is larger ('recession') and smaller ('expansion') than 0.5. To judge whether the mean of $ASY_Y(h, A^+, B)$ is significantly different from zero, we use $\sigma_{ASY_Y(h,A^+,B)}/\sqrt{n_A}$, where $\sigma_{ASY_Y(h,A^+,B)}$ is the standard deviation of $ASY_Y(h, A^+, B)$ and n_A is the number of shocks δ in the set A for which $ASY_Y(h, \delta, \omega_{t-1})$ is computed, as standard error for the mean. The reason for dividing by n_A is that different realizations $ASY_Y(h, \delta, \omega_{t-1})$ are not independent across histories ω_{t-1} but are independent across shocks δ. It is seen that symmetry can almost never be rejected for impulse responses for shocks occurring during recessions, while asymmetry for shocks occurring during expansions is found for all sizes of shocks at all horizons considered.

- insert Table 4 about here -

The final 10 years of data, from January 1990 until December 1999, are used to evaluate the forecast performance of the estimated AR and LSTAR models. For each point from December 1989 up to December 1998, we compute 1 to 12-steps ahead forecasts of the unemployment rate from the AR model given in (53) and the LSTAR model as given in (54)-(55). To obtain the forecasts from the LSTAR model we use the bootstrap method outlined in Section 6.1. We thus obtain 109 1- to 12-step ahead forecasts. The parameters are not updated as new observations become available. Table 5 contains several forecast evaluation criteria, based upon the entire forecast period and conditional upon the regime that is realized at the forecast origin. That is, the forecasts $y_{t+h|t}$ are grouped depending on whether the transition function $G(\Delta_{12}y_{t-1};\gamma,c)$ in (55) is smaller or larger than 0.5. Interestingly, MPE and MedPE suggest that both models render biased forecasts in both regimes. When the unemployment rate is declining, that is, during periods of expansions, both models are overly pessimistic and predict too high unemployment rates on average. When the unemployment rate is increasing, it is consistently under-predicted by the AR model. In this case, the STAR model overpredicts but seems closer to the mark than the AR model. Comparing the MSPE for the AR and STAR models, it is seen that the nonlinear models offer improved forecast performance at short forecast horizons during expansions and at long horizons during recessions, where a reduction of up to 30% in the MSPE is attained.

- insert Table 5 about here -

This example shows that nonlinear STAR models can yield informative inference on a macroeconomic time series, and that it also may forecast well.

8 Concluding remarks

In this paper we have surveyed recent developments related to the STAR model, including several novel extensions of the basic 2-regime model and recently designed model and forecast evaluation techniques. So far the STAR model has mainly been applied to macroeconomic time series. Applications in other areas, such as finance and marketing, may therefore be a major area of future research. While there has been some work on vector STAR models, more research is needed to investigate the properties of such models. Finally, incorporating smooth transitions in panel data models is another challenging new area. A recent paper by Johansen (1999) is one of the first attempts in this direction.

References

Anderson, H.M. (1997), Transaction costs and nonlinear adjustment towards equilibrium in the US Treasury Bill market, *Oxford Bulletin of Economics and Statistics* **59**, 465–484.

Anderson, H.M. and F. Vahid (1998), Testing multiple equation systems for common nonlinear components, *Journal of Econometrics* **84**, 1–36.

Andrews, D.W.K. and W. Ploberger (1994), Optimal tests when a nuisance parameter is present only under the alternative, *Econometrica* **62**, 1383–1414.

Astatkie, T., D.G. Watts and W.E. Watt (1997), Nested threshold autoregressive NeTAR models, *International Journal of Forecasting* **13**, 105–116.

Balke, N.S. and T.B. Fomby (1997), Threshold cointegration, *International Economic Review* **38**, 627–646.

Banerjee, A., J.J. Dolado, J.W. Galbraith and D.F. Hendry (1993), *Co-integration, Error-correction, and the Econometric Analysis of Nonstationary Data*, Oxford: Oxford University Press.

Bates, D.M. and D.G. Watts (1988), *Nonlinear regression and its applications*, New York: John Wiley.

Bianchi, M. and G. Zoega (1998), Unemployment persistence: does the size of the shock matter?, *Journal of Applied Econometrics* **13**, 283–304.

Blanchard, O.J. and L.H. Summers (1987), Hysteresis in unemployment, *European Economic Review* **31**, 288–295.

Box, G.E.P. and G.M. Jenkins (1970), *Time Series Analysis; Forecasting and Control*, San Francisco: Holden-Day.

Brännäs, K. and H. Ohlsson (1999), Asymmetric time series and temporal aggregation, *Review of Economics and Statistics* **81**, 341–344.

Caner, M. and B.E. Hansen (2000), Threshold autoregression with a unit root, unpublished manuscript, University of Wisconsin, Madison.

Carrasco, M. (1997), Misspecified structural change, threshold and Markov switching models, unpublished manuscript, Ohio State University.

Christoffersen, P.F. (1998), Evaluating interval forecasts, *International Economic Review* **39**, 841–862.

Clements, M.P. and D.F. Hendry (1998), *Forecasting Economic Time Series*, Cambridge: Cambridge University Press.

Clements, M.P. and H.-M. Krolzig (1998), A comparison of the forecast performance of Markov-switching and threshold autoregressive models of US GNP, *Econometrics Journal* **1**, C47–C75.

Clements, M.P. and J. Smith (1997), The performance of alternative forecasting methods for SETAR models, *International Journal of Forecasting* **13**, 463–475.

Clements, M.P. and J. Smith (1998), Testing self-exciting threshold autoregressive models against structural change models, unpublished manuscript, University of Warwick.

Clements, M.P. and J. Smith (1999), A Monte Carlo study of the forecasting performance of empirical SETAR models, *Journal of Applied Econometrics* **14**, 123–142.

Clements, M.P. and J. Smith (2000), Evaluating forecasts from SETAR models for exchange rates, *Journal of International Money and Finance*, to appear.

Davidson, R. and J.G. MacKinnon (1985), Heteroskedasticity-robust tests in regression directions, *Annales de l'INSEE* **59/60**, 183–218.

Davies, R.B. (1977), Hypothesis testing when a nuisance parameter is present only under the alternative, *Biometrika* **64**, 247–254.

Davies, R.B. (1987), Hypothesis testing when a nuisance parameter is present only under the alternative, *Biometrika* **74**, 33–43.

de Gooijer, J.G. and K. Kumar (1992), Some recent developments in non-linear time series modelling, testing and forecasting, *International Journal of Forecasting* **8**, 135–156.

Diebold, F.X. and G.D. Rudebusch (1996), Measuring business cycles: a modern perspective, *Review of Economics and Statistics* **78**, 67–77.

Diebold, F.X. and J.A. Nason (1990), Nonparametric exchange rate prediction, *Journal of International Economics* **28**, 315–332.

Diebold, F.X. and R.S. Mariano (1995), Comparing predictive accuracy, *Journal of Business & Economic Statistics* **13**, 253–263.

Diebold, F.X., T.A. Gunther and A.S. Tay (1998), Evaluating density forecasts with applications to financial risk management, *International Economic Review* **39**, 863–883.

Dwyer, G.P., P. Locke and W. Yu (1996), Index arbitrage and nonlinear dynamics between the S&P 500 futures and cash, *Review of Financial Studies* **9**, 301–332.

Eitrheim, Ø. and T. Teräsvirta (1996), Testing the adequacy of smooth transition autoregressive models, *Journal of Econometrics* **74**, 59–76.

Escribano, A. and O. Jordá (1999), Improved testing and specification of smooth transition regression models, in P. Rothman (ed.), *Nonlinear Time Series Analysis of Economic and Financial Data*, Boston: Kluwer, pp. 289–319.

Forbes, C.S., G.R.J. Kalb and P. Kofman (1999), Bayesian arbitrage threshold analysis, *Journal of Business & Economic Statistics* **17**, 364–372.

Franses, P.H. and D. van Dijk (2000), *Nonlinear Time Series Models in Empirical Finance*, Cambridge: Cambridge University Press.

Franses, P.H., P. de Bruin and D. van Dijk (2000), Seasonal smooth transition autoregressive models, Econometric Institute Report 2000-06/A, Erasmus University Rotterdam.

Garcia, R. and P. Perron (1996), An analysis of the real interest rate under regime switches, *Review of Economics and Statistics* **78**, 111–125.

Godfrey, L.G. (1979), Testing the adequacy of a time series model, *Biometrika* **66**, 67–72.

Granger, C.W.J. (1993), Strategies for modelling nonlinear time-series relationships, *The Economic Record* **69**, 233–238.

Granger, C.W.J. and N.R. Swanson (1996), Future developments in the study of cointegrated variables, *Oxford Bulletin of Economics and Statistics* **58**, 537–553.

Granger, C.W.J. and T. Teräsvirta (1993), *Modelling Nonlinear Economic Relationships*, Oxford: Oxford University Press.

Granger, C.W.J. and T. Teräsvirta (1999), A simple nonlinear time series model with misleading linear properties, *Economics Letters* **62**, 161–165.

Hamilton, J.D. (1994), *Time Series Analysis*, Princeton: Princeton University Press.

Hampel, H.R., E.M. Ronchetti, P.J. Rousseeuw and W.A. Stahel (1986), *Robust Statistics - The Approach based on Influence Functions*, New York: John Wiley.

Hansen, B.E. (1996), Inference when a nuisance parameter is not identified under the null hypothesis, *Econometrica* **64**, 413–430.

Hansen, B.E. (1997), Inference in TAR models, *Studies in Nonlinear Dynamics and Econometrics* **2**, 1–14.

Hatanaka, M. (1996), *Time-Series-Based Econometrics, Unit Roots and Cointegration*, Oxford: Oxford University Press.

Hendry, D.F. (1995), *Dynamic Econometrics*, Oxford: Oxford University Press.

Hornik, K., M.B. Stinchcombe and H. White (1989), Multilayer feedforward networks are universal approximators, *Neural Networks* **2**, 359–366.

Hornik, K., M.B. Stinchcombe and H. White (1990), Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks, *Neural Networks* **3**, 551–560.

Huber, P.J. (1981), *Robust Statistics*, New York: John Wiley.

Hyndman, R.J. (1995), Highest-density forecast regions for nonlinear and nonnormal time series, *Journal of Forecasting* **14**, 431–441.

Hyndman, R.J. (1996), Computing and graphing highest-density regions, *American Statistician* **50**, 120–126.

Jansen, E.S. and T. Teräsvirta (1996), Testing parameter constancy and super exogeneity in econometric equations, *Oxford Bulletin of Economics and Statistics* **58**, 735–768.

Johansen, K. (1999), Nonlinear wage responses to internal and external factors, unpublished manuscript, Norwegian University of Science and Technology.

Johansen, S. (1995), *Likelihood-based inference in cointegrated vector autoregressive models*, Oxford: Oxford University Press.

Koop, G., M.H. Pesaran and S.M. Potter (1996), Impulse response analysis in nonlinear multivariate models, *Journal of Econometrics* **74**, 119–147.

Koop, G. and S.M. Potter (1999), Dynamic asymmetries in US unemployment, *Journal of Business & Economic Statistics* **17**, 298–313.

Koop, G. and S.M. Potter (2000), Nonlinearity, structural breaks or outliers in economic time series?, in W. A. Barnett, D. F. Hendry, S. Hylleberg, T. Teräsvirta, D. Tjøstheim and A. H. Würtz (eds.), *Nonlinear Econometric Modeling in Time Series Analysis*, Cambridge: Cambridge University Press.

Krolzig, H.-M. (1997), *Markov-Switching Vector Autoregressions - Modelling, Statistical Inference and Applications to Business Cycle Analysis*, Lecture Notes in Economics and Mathematics **454**, Berlin: Springer-Verlag.

Kuan, C.-M. and H. White (1994), Artificial neural networks: an econometric perspective, *Econometric Reviews* **13**, 1–143 (with discussion).

Lee, T-H., H. White and C.W.J. Granger (1993), Testing for neglected nonlinearity in time series models, *Journal of Econometrics* **56**, 269–290.

Leybourne, S., P. Newbold and D. Vougas (1998), Unit roots and smooth transitions, *Journal of Time Series Analysis* **19**, 83–97.

Lin, C-F.J. and T. Teräsvirta (1994), Testing the constancy of regression parameters against continuous structural change, *Journal of Econometrics* **62**, 211–228.

Lucas, A., P.H. Franses and D. van Dijk (2001), *Outlier Robust Analysis of Economic Time Series*, Oxford: Oxford University Press, to appear.

Lundbergh, S. and T. Teräsvirta (1998), Modelling economic high-frequency time series with STAR-GARCH models, Working Paper Series in Economics and Finance No. 291, Stockholm School of Economics.

Lundbergh, S. and T. Teräsvirta (2001), Forecasting with smooth transition autoregressive models, in M.P. Clements and D.F. Hendry (eds.), *Companion to Economic Forecasting*, Oxford: Basil Blackwell, to appear.

Lundbergh, S., T. Teräsvirta and D. van Dijk (2000), Time-varying smooth transition autoregressive models, Working Paper Series in Economics and Finance No. 376, Stockholm School of Economics.

Luukkonen, R., P. Saikkonen and T. Teräsvirta (1988a), Testing linearity against smooth transition autoregressive models, *Biometrika* **75**, 491–499.

Luukkonen, R., P. Saikkonen and T. Teräsvirta (1988b), Testing linearity in univariate time series models, *Scandinavian Journal of Statistics* **16**, 161–175.

Martens, M., P. Kofman and A.C.F. Vorst (1998), A threshold error correction for intraday futures and index returns, *Journal of Applied Econometrics* **13**, 245–263.

Martin, R.D. (1981), Robust methods for time series, in D.F. Findley (ed.), *Applied Time Series Analysis*, New York: Academic Press, pp. 6843–759.

Medeiros, M.C. and A. Veiga (2000), A flexible coefficient smooth transition time series model, Working Paper Series in Economics and Finance No. 360, Stockholm School of Economics.

Michael, P., A.R. Nobay and D.A. Peel (1997), Transaction costs and nonlinear adjustment in real exchange rates: an empirical investigation, *Journal of Political Economy* **105**, 862–879.

Montgomery, A.L., V. Zarnowitz, R.S. Tsay and G.C. Tiao (1998), Forecasting the US unemployment rate, *Journal of the American Statistical Association* **93**, 478–493.

Neftçi, S.N. (1984), Are economic time series asymmetric over the business cycle?, *Journal of Political Economy* **92**, 307–328.

Neftçi, S.N. (1993), Statistical analysis of shapes in macroeconomic time series: is there a business cycle?, *Journal of Business & Economic Statistics* **11**, 215–224.

Öcal, N. and D. Osborn (2000), Business cycle nonlinearities in UK consumption and production, *Journal of Applied Econometrics* **15**, 27–43.

Peel, D.A. and A.E.H. Speight (1996), Is the US business cycle asymmetric? Some further evidence, *Applied Economics* **28**, 405–415.

Pesaran, M.H. and S.M. Potter (1997), A floor and ceiling model of US output, *Journal of Economic Dynamics and Control* **21**, 661–695.

Pötscher, B.M. and I.V. Prucha (1997), *Dynamic Nonlinear Econometric Models - Asymptotic Theory*, Berlin: Springer-Verlag.

Potter, S.M. (1994), Asymmetric economic propagation mechanisms, in W. Semmler (ed.), *Business Cycles: Theory and Empirical Methods*, Boston: Kluwer, pp. 313–330.

Potter, S.M. (1995), Nonlinear models of economic fluctuations, in K. Hoover (ed.), *Macroeconometrics - Developments, Tensions and Prospects*, Boston: Kluwer, pp. 517–560.

Potter, S.M. (1999), Nonlinear time series modelling: an introduction, *Journal of Economic Surveys* **13**, 505–528.

Potter, S.M. (2000), Nonlinear impulse response functions, *Journal of Economic Dynamics and Control*, to appear.

Priestley, M.B. (1981), *Spectral Analysis and Time Series*, New York: Academic Press.

Quandt, R. (1983), Computational problems and methods, in Z. Griliches and M.D. Intriligator (eds.), *Handbook of Econometrics I*, Amsterdam: Elsevier Science, pp. 699–746.

Ravn, M.O. and M. Sola (1995), Stylized facts and regime changes: are prices procyclical?, *Journal of Monetary Economics* **36**, 497–526.

Rothman, P. (1991), Further evidence on the asymmetric behavior of unemployment rates over the business cycle, *Journal of Macroeconomics* **13**, 291–298.

Rothman, P. (1998), Forecasting asymmetric unemployment rates, *Review of Economics and Statistics* **80**, 164–168.

Rothman, P., D. van Dijk and P.H. Franses (1999), A multivariate STAR analysis of the relationship between money and output, Econometric Institute Report 9945/A, Erasmus University Rotterdam.

Saikkonen, P. and R. Luukkonen (1988), Lagrange multiplier tests for testing non-linearities in time series models, *Scandinavian Journal of Statistics* **15**, 55–68.

Sarantis, N. (1999), Modeling non-linearities in real effective exchange rates, *Journal of International Money and Finance* **18**, 27–45.

Sichel, D.E. (1989), Are business cycles asymmetric? A correction, *Journal of Political Economy* **97**, 1255–1260.

Siklos, P.L. and C.W.J. Granger (1997), Temporary cointegration with an application to interest rate parity, *Macroeconomic Dynamics* **1**, 640–657.

Skalin, J. and T. Teräsvirta (1999), Another look at Swedish business cycles, *Journal of Applied Econometrics* **14**, 359–378.

Skalin, J. and T. Teräsvirta (2000), Modelling asymmetries and moving equilibria in unemployment rates, *Macroeconomic Dynamics*, to appear.

Stinchcombe, M.B. and H. White (1998), Consistent specification testing with nuisance parameters present only under the alternative, *Econometric Theory* **14**, 295–325.

Stock, J.H. and M.W. Watson (1996), Evidence on structural instability in macroeconomic time series relations, *Journal of Business & Economic Statistics* **14**, 11–30.

Stock, J.H. and M.W. Watson (1999), A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series, in R.F. Engle and H. White (eds.), *Cointegration, Causality, and Forecasting: A Festschrift in Honor of Clive W.J. Granger*, Oxford: Oxford University Press, pp. 1–44.

Swanson, N.R. (1999), Finite sample properties of a simple LM test for neglected nonlinearity in error-correcting regression equations, *Statistica Neerlandica* **53**, 76–95.

Taylor, M.P., D.A. Peel and L. Sarno (2000), Nonlinear mean-reversion in exchange rate rates: towards a solution to the purchasing power parity puzzles, *International Economic Review*, to appear.

Taylor, N., D. van Dijk, P.H. Franses and A. Lucas (2000), SETS, arbitrage activity, and stock price dynamics, *Journal of Banking and Finance*, to appear.

Teräsvirta, T. (1994), Specification, estimation, and evaluation of smooth transition autoregressive models, *Journal of the American Statistical Association* **89**, 208–218.

Teräsvirta, T. (1998), Modelling economic relationships with smooth transition regressions, in A. Ullah and D.E.A. Giles (eds.), *Handbook of Applied Economic Statistics*, New York: Marcel Dekker, pp. 507–552.

Teräsvirta, T., D. Tjøstheim and C.W.J. Granger (1994), Aspects of modelling nonlinear time series, in R.F. Engle and D.L. McFadden (eds.), *Handbook of econometrics, vol. IV*, Amsterdam: Elsevier Science, pp. 2917–2957.

Teräsvirta, T. and H.M. Anderson (1992), Characterizing nonlinearities in business cycles using smooth transition autoregressive models, *Journal of Applied Econometrics* **7**, S119–S136.

Tiao, G.C. and R.S. Tsay (1994), Some advances in non-linear and adaptive modelling in time-series, *Journal of Forecasting* **13**, 109–140 (with discussion).

Timmermann, A. (2000), Moments of Markov Switching models, *Journal of Econometrics* **96**, 75–111.

Tong, H. (1990), *Non-Linear Time Series: A Dynamical Systems Approach*, Oxford: Oxford University Press.

Tong, H. (1995), A personal overview of non-linear time series analysis from a chaos perspective, *Scandinavian Journal of Statistics* **22**, 399–445.

Tsay, R.S. (1986), Nonlinearity tests for time series, *Biometrika* **73**, 461–466.

Tsay, R.S. (1998), Testing and modeling multivariate threshold models, *Journal of the American Statistical Association* **93**, 1188–1202.

van Dijk, D. and P.H. Franses (1999), Modeling multiple regimes in the business cycle, *Macroeconomic Dynamics* **3**, 311–340.

van Dijk, D. and P.H. Franses (2000), Nonlinear error-correction models for interest rates in the Netherlands, in W. A. Barnett, D. F. Hendry, S. Hylleberg, T. Teräsvirta, D. Tjøstheim and A. H. Würtz (eds.), *Nonlinear Econometric Modeling in Time Series Analysis*, Cambridge: Cambridge University Press.

van Dijk, D., P.H. Franses and A. Lucas (1999), Testing for smooth transition nonlinearity in the presence of additive outliers, *Journal of Business & Economic Statistics* **17**, 217–235.

Weise, C.L. (1999), The asymmetric effects of monetary policy, *Journal of Money, Credit and Banking* **31**, 85–108.

Wooldridge, J.M. (1990), A unified approach to robust, regression-based specification tests, *Econometric Theory* **6**, 17–43.

Wooldridge, J.M. (1991), On the application of robust, regression-based diagnostics to models of conditional means and conditional variances, *Journal of Econometrics* **47**, 5–46.

Wooldridge, J.M. (1994), Estimation and inference for dependent processes, in R.F. Engle and D.L. McFadden (eds.), *Handbook of econometrics, vol. IV*, Amsterdam: Elsevier Science, pp. 2639–2738.

Yadav, P.K., P.F. Pope and K. Paudyal (1994), Threshold autoregressive modeling in finance: the price difference of equivalent assets, *Mathematical Finance* **4**, 205–221.

Table 1: LM-type tests for STAR nonlinearity for monthly US unemployment rate

Transition	Standard tests				Heterosked. rob. tests				Outlier robust tests			
variable s_t	LM_1	LM_3	LM_3^e	LM_4	LM_1	LM_3	LM_3^e	LM_4	LM_1	LM_3	LM_3^e	LM_4
					All regressors							
$\Delta_{12}y_{t-1}$	0.039	0.150	0.044	0.532	0.017	0.316	0.023	0.810	0.062	0.098	0.063	0.541
$\Delta_{12}y_{t-2}$	0.100	0.037	0.126	0.093	0.068	0.373	0.088	0.628	0.080	0.070	0.100	0.576
$\Delta_{12}y_{t-3}$	0.129	0.162	0.160	0.326	0.130	0.737	0.162	0.592	0.078	0.269	0.098	0.416
$\Delta_{12}y_{t-4}$	0.308	0.665	0.354	0.745	0.244	0.755	0.291	0.657	0.133	0.523	0.158	0.379
$\Delta_{12}y_{t-5}$	0.404	0.662	0.503	0.886	0.250	0.677	0.330	0.561	0.197	0.463	0.276	0.691
$\Delta_{12}y_{t-6}$	0.378	0.588	0.430	0.306	0.130	0.603	0.185	0.899	0.191	0.315	0.256	0.262
t	0.386	0.978	0.434	0.864	0.319	0.966	0.362	0.548	0.447	0.924	0.503	0.698
					Intercept and monthly dummies							
$\Delta_{12}y_{t-1}$	0.530	0.753	0.539	0.791	0.530	0.505	0.528	0.418	0.502	0.728	0.533	0.874
$\Delta_{12}y_{t-2}$	0.738	0.760	0.628	0.862	0.681	0.804	0.698	0.796	0.637	0.799	0.564	0.947
$\Delta_{12}y_{t-3}$	0.881	0.913	0.783	0.934	0.832	0.822	0.806	0.722	0.774	0.882	0.657	0.940
$\Delta_{12}y_{t-4}$	0.482	0.966	0.812	0.966	0.850	0.821	0.807	0.894	0.857	0.892	0.715	0.953
$\Delta_{12}y_{t-5}$	0.941	0.946	0.834	0.961	0.891	0.873	0.676	0.875	0.889	0.921	0.801	0.983
$\Delta_{12}y_{t-6}$	0.803	0.857	0.781	0.856	0.798	0.824	0.719	0.853	0.897	0.880	0.890	0.925
t	0.107	0.616	0.144	0.847	0.145	0.731	0.193	0.775	0.152	0.442	0.192	0.649
					Intercept and lagged dependent variables							
$\Delta_{12}y_{t-1}$	0.025	0.073	0.021	0.135	0.009	0.186	0.013	0.331	0.061	0.110	0.053	0.287
$\Delta_{12}y_{t-2}$	0.053	0.019	0.066	0.064	0.028	0.279	0.041	0.505	0.091	0.061	0.110	0.237
$\Delta_{12}y_{t-3}$	0.050	0.158	0.064	0.268	0.022	0.283	0.033	0.517	0.063	0.171	0.078	0.258
$\Delta_{12}y_{t-4}$	0.161	0.842	0.176	0.710	0.055	0.598	0.077	0.594	0.093	0.617	0.091	0.412
$\Delta_{12}y_{t-5}$	0.208	0.460	0.284	0.476	0.083	0.491	0.133	0.686	0.099	0.273	0.140	0.308
$\Delta_{12}y_{t-6}$	0.289	0.057	0.319	0.119	0.043	0.183	0.075	0.306	0.123	0.083	0.165	0.128
t	0.662	0.976	0.715	0.991	0.303	0.936	0.331	0.916	0.639	0.962	0.703	0.963
					Seasonally adjusted series							
$\Delta_{12}y_{t-1}$	0.009	0.062	0.010	0.086	0.003	0.135	0.005	0.220	0.013	0.054	0.013	0.126
$\Delta_{12}y_{t-2}$	0.011	0.009	0.016	0.008	0.008	0.111	0.012	0.174	0.016	0.017	0.025	0.086
$\Delta_{12}y_{t-3}$	0.007	0.065	0.011	0.168	0.007	0.280	0.011	0.325	0.007	0.050	0.011	0.142
$\Delta_{12}y_{t-4}$	0.029	0.501	0.041	0.500	0.016	0.381	0.023	0.511	0.014	0.243	0.019	0.194
$\Delta_{12}y_{t-5}$	0.058	0.270	0.095	0.450	0.027	0.418	0.054	0.777	0.023	0.100	0.038	0.186
$\Delta_{12}y_{t-6}$	0.087	0.046	0.094	0.120	0.015	0.265	0.026	0.402	0.031	0.065	0.043	0.104
t	0.619	0.963	0.685	0.966	0.302	0.934	0.357	0.805	0.585	0.947	0.649	0.942

p-values of F variants of the LM-type tests for STAR nonlinearity of the monthly US unemployment rate, January 1970-December 1989. The tests are applied in an AR(15) model for the first differences, including a lagged level term and monthly seasonal dummies. The LM_1, LM_3, LM_3^e and LM_4 statistics are based on the auxiliary regression models given in (19), (20), (21) and (23), respectively.

Table 2: STAR model selection for monthly US unemployment rate

Transition variable s_t	Teräsvirta H_{03}	H_{02}	H_{01}	Escribano-Jorda H_{0L}	H_{0E}
			Standard tests		
$\Delta_{12}y_{t-1}$	0.495	0.347	0.009	0.107	0.612
$\Delta_{12}y_{t-2}$	0.014	0.769	0.011	0.002	0.079
$\Delta_{12}y_{t-3}$	0.150	0.889	0.007	0.091	0.618
$\Delta_{12}y_{t-4}$	0.657	0.976	0.029	0.371	0.676
$\Delta_{12}y_{t-5}$	0.408	0.751	0.058	0.262	0.595
$\Delta_{12}y_{t-6}$	0.047	0.538	0.087	0.025	0.159
t	0.917	0.880	0.619	0.878	0.885
			Heteroskedasticity robust tests		
$\Delta_{12}y_{t-1}$	0.052	0.359	0.003	0.268	0.237
$\Delta_{12}y_{t-2}$	0.019	0.526	0.008	0.303	0.201
$\Delta_{12}y_{t-3}$	0.243	0.720	0.007	0.028	0.194
$\Delta_{12}y_{t-4}$	0.448	0.814	0.016	0.263	0.416
$\Delta_{12}y_{t-5}$	0.159	0.478	0.027	0.109	0.542
$\Delta_{12}y_{t-6}$	0.070	0.263	0.015	0.275	0.119
t	0.845	0.938	0.302	0.387	0.752
			Outlier robust tests		
$\Delta_{12}y_{t-1}$	0.199	0.304	0.013	0.286	0.566
$\Delta_{12}y_{t-2}$	0.021	0.511	0.017	0.220	0.494
$\Delta_{12}y_{t-3}$	0.134	0.705	0.007	0.157	0.520
$\Delta_{12}y_{t-4}$	0.479	0.884	0.014	0.154	0.372
$\Delta_{12}y_{t-5}$	0.143	0.579	0.023	0.040	0.174
$\Delta_{12}y_{t-6}$	0.159	0.505	0.031	0.060	0.202
t	0.896	0.771	0.585	0.880	0.780

p-values of F variants of the LM-type tests used in the specification procedures of Teräsvirta (1994) and Escribano and Jordá (1999), applied to the monthly US unemployment rate, January 1970-December 1989. The tests are applied in an AR model for the seasonally adjusted series with 15 lagged first differences and including a lagged level term. The hypotheses H_{01}, H_{02}, H_{03}, H_{0L} and H_{0E} are discussed in Section 5.1.

Table 3: Diagnostic tests of LSTAR model estimated for monthly US unemployment rate

	Tests for q-th order serial correlation					
q	2	4	6	8	10	12
p-value	0.386	0.454	0.693	0.859	0.814	0.927
	Tests for parameter constancy					
	Constant and dummies			Lagged dependent variables		
	$\text{LM}_{\text{C},1}$	$\text{LM}_{\text{C},2}$	$\text{LM}_{\text{C},3}$	$\text{LM}_{\text{C},1}$	$\text{LM}_{\text{C},2}$	$\text{LM}_{\text{C},3}$
p-value	0.565	0.638	0.894	0.536	0.730	0.847
	Tests for remaining nonlinearity					
Transition	Constant and dummies			Lagged dependent variables		
variable s_{2t}	$\text{LM}_{\text{EMR},1}$	$\text{LM}_{\text{EMR},2}$	$\text{LM}_{\text{EMR},3}$	$\text{LM}_{\text{EMR},1}$	$\text{LM}_{\text{EMR},2}$	$\text{LM}_{\text{EMR},3}$
$\Delta_{12}y_{t-1}$	0.628	0.717	0.828	0.603	0.858	0.923
$\Delta_{12}y_{t-2}$	0.379	0.623	0.686	0.629	0.822	0.225
$\Delta_{12}y_{t-3}$	0.505	0.787	0.584	0.821	0.980	0.293
$\Delta_{12}y_{t-4}$	0.586	0.664	0.777	0.684	0.975	0.825
$\Delta_{12}y_{t-5}$	0.441	0.571	0.845	0.739	0.947	0.880
$\Delta_{12}y_{t-6}$	0.221	0.483	0.648	0.747	0.921	0.984

Diagnostic tests for estimated LSTAR model for the monthly US unemployment rate. $\text{LM}_{\text{EMR},j}$ and $\text{LM}_{\text{C},j}$, $j = 1, 2, 3$, denote the LM-type tests for no remaining nonlinearity and parameter constancy, respectively, based on (28) including auxiliary regressors $\beta_i' x_t s_{2t}^i$ and $\beta_{3+i}' x_t s_{2t}^i G_1(s_{1t}; \gamma_1, c_1)$ for $i = 1, \ldots, j$.

Table 4: Asymmetry measures for impulse responses in LSTAR model

	Unconditional				Recession				Expansion			
h	A	S	M	L	A	S	M	L	A	S	M	L
12	−0.00	−0.04*	0.06*	0.26*	0.00	0.00	0.00	0.02	−0.00	−0.06*	0.09*	0.38*
	(0.15)	(0.08)	(0.16)	(0.40)	(0.07)	(0.05)	(0.09)	(0.23)	(0.18)	(0.09)	(0.18)	(0.41)
24	0.00	−0.03*	0.05*	0.19*	0.00	−0.01	0.02*	0.06*	0.00	−0.04*	0.07*	0.26*
	(0.11)	(0.06)	(0.12)	(0.26)	(0.06)	(0.04)	(0.07)	(0.20)	(0.12)	(0.07)	(0.13)	(0.26)
36	0.00	−0.01*	0.02*	0.06*	0.00	−0.00	0.00	0.02	0.00	−0.01*	0.03*	0.08*
	(0.06)	(0.04)	(0.06)	(0.14)	(0.05)	(0.03)	(0.05)	(0.14)	(0.06)	(0.04)	(0.06)	(0.14)
48	−0.00	0.00	−0.00	−0.01	0.00	0.00	0.00	0.00	−0.00	0.00	−0.00	−0.02*
	(0.03)	(0.03)	(0.03)	(0.07)	(0.03)	(0.03)	(0.04)	(0.08)	(0.03)	(0.03)	(0.03)	(0.06)
60	−0.00*	0.00	−0.01*	−0.02*	0.00	0.00	−0.00	−0.00	−0.00*	0.00	−0.01*	−0.03*
	(0.03)	(0.02)	(0.03)	(0.05)	(0.02)	(0.02)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.05)

Mean and standard deviation (in parentheses) of distribution of asymmetry measure $\text{ASY} =_y (h, A^+, B)$ in estimated LSTAR model for the monthly US unemployment rate. Means larger than two times $\sigma_{ASY_Y(h,A^+,B)}/\sqrt{n_A}$ are marked with an asterisk, where $\sigma_{ASY_Y(h,A^+,B)}$ is the standard deviation of $ASY_Y(h, A^+, B)$ and n_A is the number of shocks δ for which $ASY_Y(h, \delta, \omega_{t-1})$ is computed. The different sets of shocks are defined as A(ll) = $\{\varepsilon_t\}$, S(mall) = $\{\varepsilon_t | 1 \geq |\varepsilon_t/\hat{\sigma}_\varepsilon| > 0\}$, M(edium) = $\{\varepsilon_t | 2 \geq |\varepsilon_t/\hat{\sigma}_\varepsilon| > 1\}$ and L(arge) = $\{\varepsilon_t | 3 \geq |\varepsilon_t/\hat{\sigma}_\varepsilon| > 2\}$. Recession and expansion relate to histories for which the value of the transition function $G(\Delta_{12}y_{t-1}; \gamma, c)$ is smaller and larger than 0.5, respectively.

Table 5: Forecast evaluation of AR and LSTAR models for monthly US unemployment rate

Forecast	Unconditional					Recession					Expansion				
	MPE		MSPE			MPE		MSPE			MPE		MSPE		
horizon	AR	LSTAR	AR	LSTAR	DM	AR	LSTAR	AR	LSTAR	DM	AR	LSTAR	AR	LSTAR	DM
1	−0.01	−0.02	0.05	0.05	−0.13	0.05	−0.01	0.05	0.05	0.49	−0.04	−0.03	0.05	0.04	−0.53
2	−0.02	−0.04	0.08	0.07	−0.90	0.09	−0.03	0.10	0.11	0.47	−0.07	−0.05	0.07	0.06	−2.65
3	−0.03	−0.07	0.12	0.10	−1.74	0.15	−0.04	0.16	0.14	−0.30	−0.12	−0.09	0.11	0.08	−3.23
4	−0.04	−0.11	0.17	0.13	−2.27	0.22	−0.04	0.23	0.18	−0.83	−0.16	−0.13	0.15	0.10	−3.50
5	−0.06	−0.14	0.24	0.18	−2.08	0.29	−0.04	0.33	0.26	−0.92	−0.22	−0.19	0.19	0.15	−2.99
6	−0.07	−0.18	0.28	0.24	−1.36	0.35	−0.06	0.44	0.37	−0.72	−0.26	−0.24	0.21	0.18	−1.63
7	−0.08	−0.22	0.35	0.30	−1.07	0.42	−0.05	0.53	0.43	−0.82	−0.31	−0.30	0.27	0.24	−0.77
8	−0.10	−0.26	0.40	0.36	−0.77	0.48	−0.04	0.62	0.51	−0.81	−0.36	−0.35	0.30	0.29	−0.11
9	−0.11	−0.30	0.44	0.42	−0.29	0.53	−0.04	0.66	0.53	−0.90	−0.40	−0.41	0.33	0.37	0.97
10	−0.13	−0.34	0.48	0.47	−0.16	0.56	−0.05	0.70	0.55	−1.00	−0.44	−0.47	0.38	0.44	1.23
11	−0.15	−0.39	0.55	0.56	0.19	0.60	−0.05	0.75	0.57	−1.07	−0.49	−0.55	0.46	0.55	1.86
12	−0.17	−0.44	0.61	0.66	0.62	0.61	−0.07	0.78	0.60	−1.06	−0.53	−0.61	0.54	0.68	2.28

Forecast evaluation of AR and LSTAR models for the monthly US unemployment rate. The forecast period runs from January 1990 until December 1999 (109 forecasts). MPE = Mean Prediction Error, MSPE = Mean Squared Prediction Error. The column headed DM contains the forecast comparison statistic of Diebold and Mariano (1995) based on squared prediction errors. Recession and expansion relate to forecasts for which the value of the transition function $G(\Delta_{12}y_{t-1};\gamma,c)$ in (55) is larger and smaller than 0.5 at the forecast origin, respectively (34 and 75 observations).

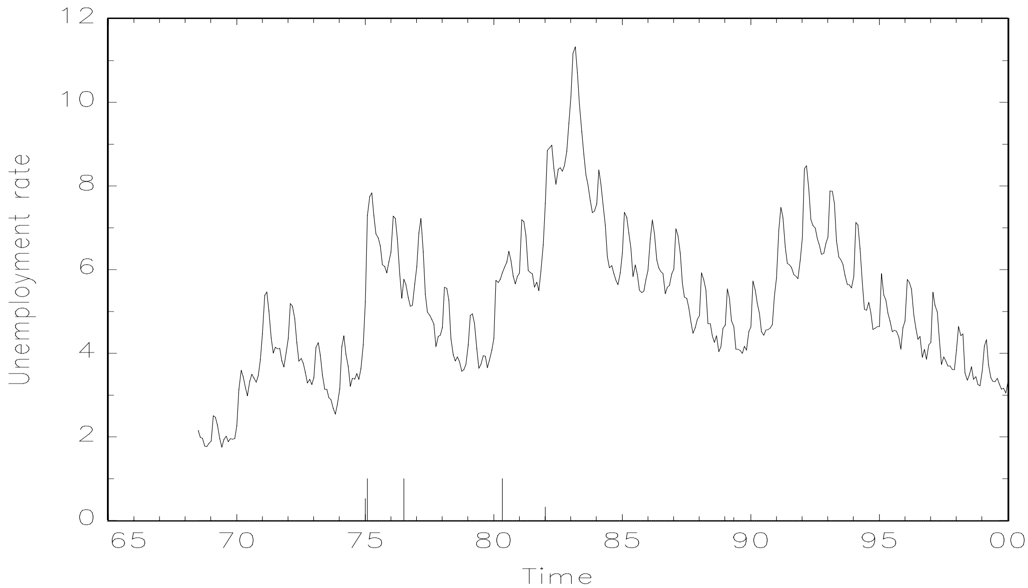


Figure 1: Monthly seasonally unadjusted US unemployment rate, males aged 20 and above, June 1968-December 1999.

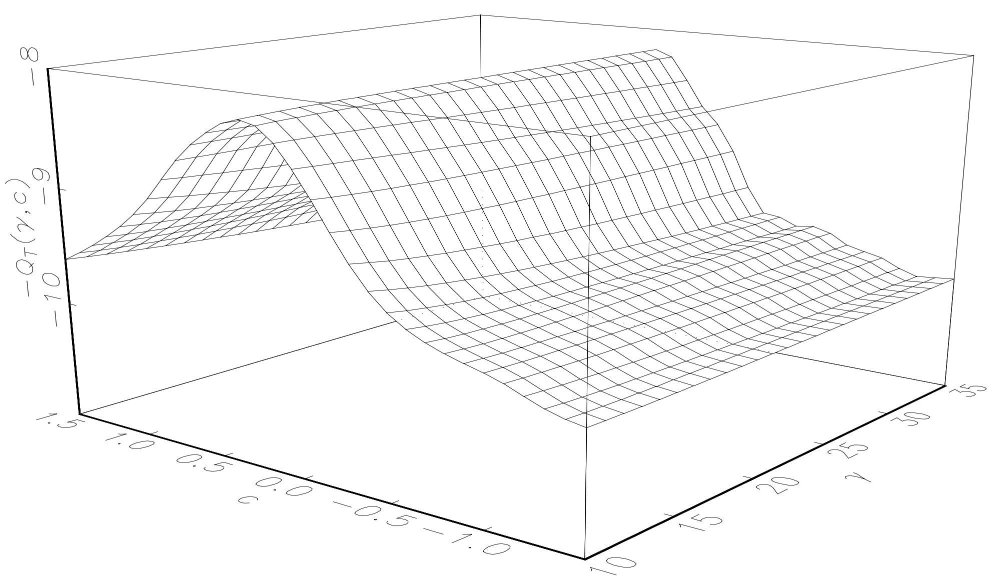
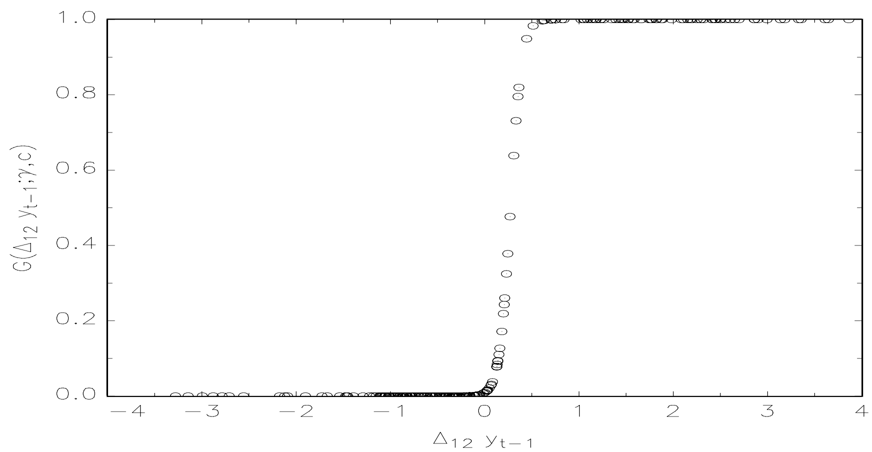
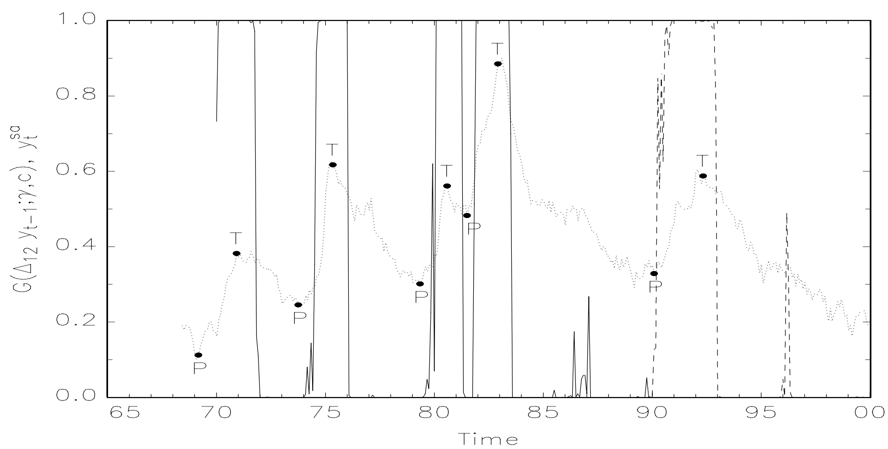


Figure 2: Negative of the sum of squares function $Q_T(\gamma, c)$ of the LSTAR model for the monthly US unemployment rate in the neighborhood of the NLS estimate $(\hat{\gamma}, \hat{c}) = (23.15, 0.27)$.



(a) Transition function versus $\Delta_{12}y_{t-1}$



(b) Transition function versus time

Figure 3: Transition function in LSTAR model for monthly seasonally unadjusted US unemployment rate against the transition variable $\Delta_{12}y_{t-1}$ and over time, during the estimation period (solid line) and forecasting period (dashed line). The dotted line represents the (rescaled) monthly seasonally adjusted unemployment rate. Solid circles indicate NBER-dated unemployment peaks (P) and troughs (T).

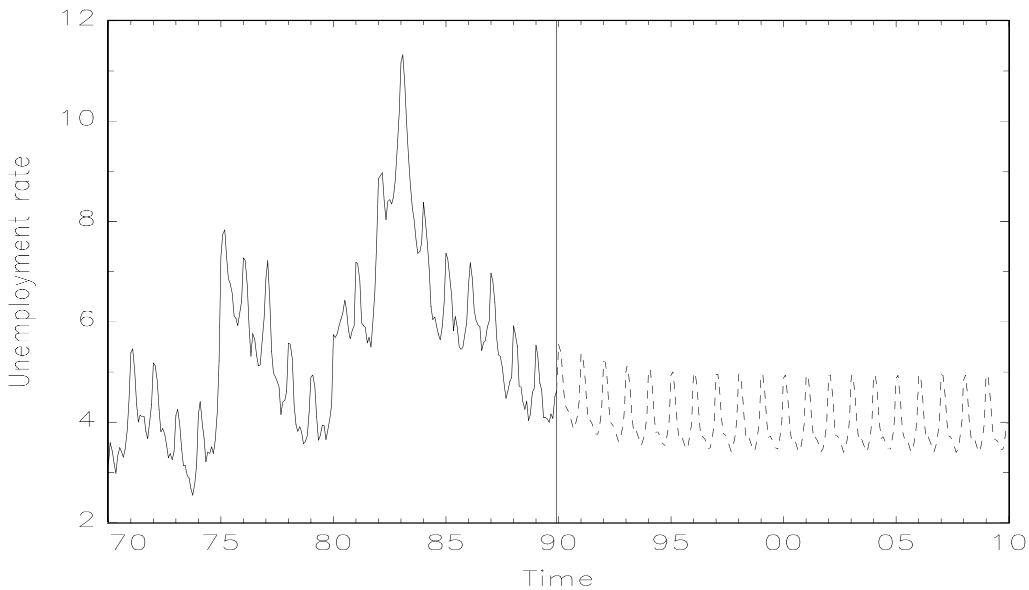
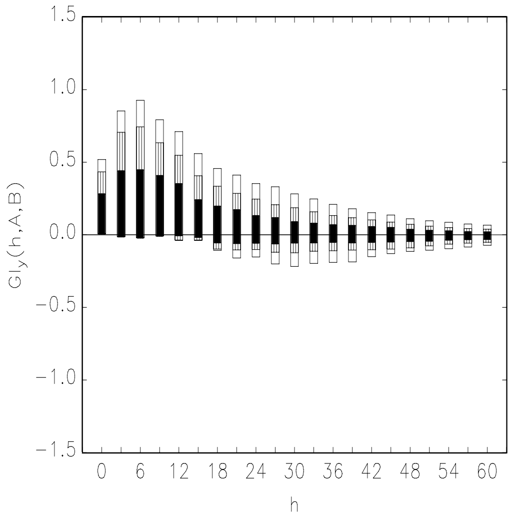
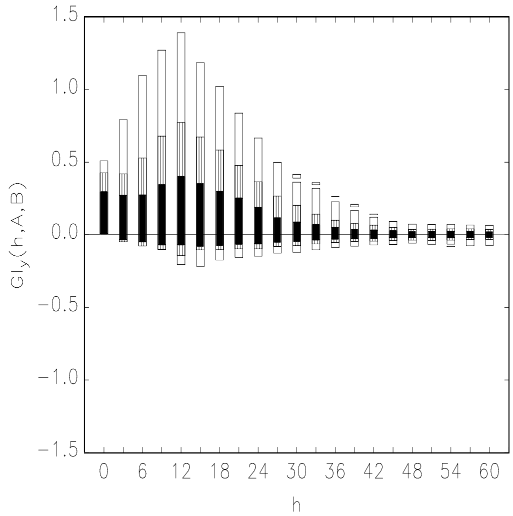


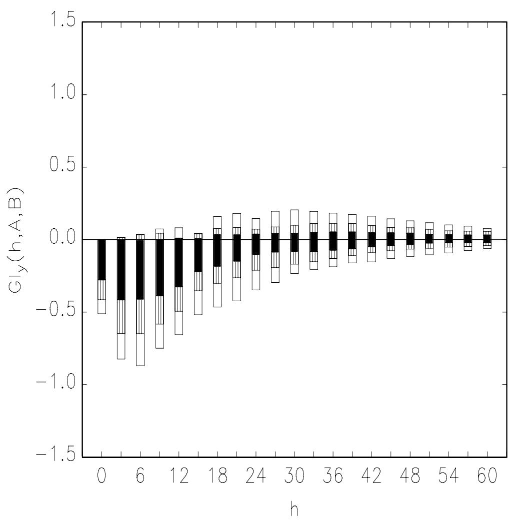
Figure 4: Deterministic extrapolation of LSTAR model for the monthly US unemployment rate.



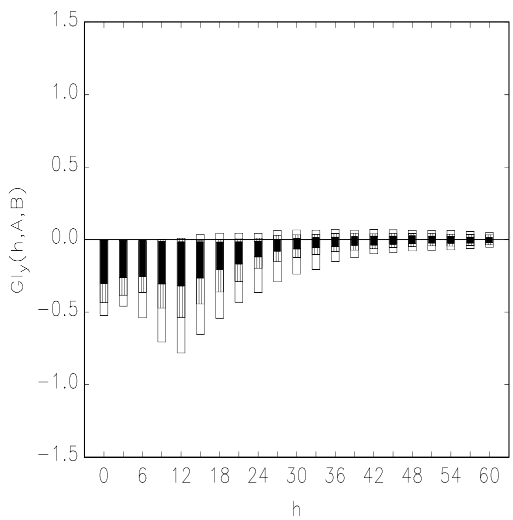
(a) positive shocks, recession



(b) positive shocks, expansion



(c) negative shocks, recession



(d) negative shocks, expansion

Figure 5: 50% (black), 75% (hatched) and 90% (white) highest density regions for generalized impulse responses in the LSTAR model for the monthly US unemployment rate. Recession and expansion relate to histories for which the value of the transition function $G(\Delta_{12}y_{t-1};\gamma,c)$ in (55) is larger and smaller than 0.5, respectively.