Available online at www.sciencedirect.com
ScienceDirect
Fuzzy Sets and Systems 158 (2007) 2734–2745
FUZZY sets and systems
www.elsevier.com/locate/fss

Smooth transition autoregressive models and fuzzy rule-based systems: Functional equivalence and consequences

José Luis Aznarte M.*, José Manuel Benítez, Juan Luis Castro

Department of Computer Science and Artificial Intelligence, University of Granada, Spain

Received 28 March 2006; received in revised form 1 March 2007; accepted 24 March 2007
Available online 7 April 2007

Abstract

In this work we will explore the theoretical connections existing between fuzzy rule-based systems (FRBS) applied on univariate time series and two statistical reference tools, the autoregressive (AR) models and the smooth transition autoregressive (STAR) model. We will show that a TSK fuzzy rule happens to be a localised AR model and that a STAR model can hence be interpreted as a restricted FRBS. Several consequences derive from this fact, and we will explore some of them, including a statistical inference-based procedure to incrementally build FRBS or the linguistic interpretation of STAR models. Finally an econometric example is given to illustrate the core idea.
© 2007 Elsevier B.V. All rights reserved.

Keywords: AR models; Fuzzy rules; STAR models; Fuzzy rule-based systems; Functional equivalence

1. Introduction

Predicting or forecasting the future is one of the fundamental issues found in a variety of disciplines, ranging from Economics through Physics to Engineering. The traditional approach to this problem was to discover certain laws governing the phenomena under consideration, based either on first principles or on real observations. Soft Computing, as a part of Artificial Intelligence, has proved to be an alternative approach to this problem with some advantages over the traditional one. Hence, there is a growing interest about time series in the Soft Computing research field.

So far, Soft Computing researchers have approached time series modelling problem mainly in two different fashions. On the one hand, some researchers saw in time series a huge collection of datasets that could be used to test new or existing models. This approach is fair, but ignores the special features that distinguish a time series from other sources of data, disregarding at the same time all the scientific knowledge gathered through years for this specific problem. Examples that illustrate this situation are the tests performed with models like ANFIS [10,9,1], EFuNN [12,13] or ANNBFIS [20]. More examples can be found in [8,15,30,22,19].

On the other hand, there are papers that present Soft Computing-based models tailored to model and forecast specific time series. Electric load forecasting with neural or neuro-fuzzy models is one of the most frequently faced problems [11,14,37,5], but there is a great number of other examples: financial forecasting [18,17,43], biological forecasting

* Corresponding author. Tel.: +34 958240467; fax: +34 958240079.
E-mail address: jlaznarte@decsai.ugr.es (J.L. Aznarte M.).

0165-0114/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.fss.2007.03.021

[36,31] etc. These researchers try to model or predict real world cases using generic models and adapting them to some observable features of the data, but still do not make use of the tools provided by statistical time series modelling.

Notwithstanding, time series modelling, as a part of Statistical Sciences, is a wide area of study, in which a great number of researchers worked in the past and are working nowadays. These works have produced a vast collection of models and methodologies, starting from basic tools as graphical summaries (time plots, seasonal plots, scatter plots...) or numerical summaries (univariate statistics, autocorrelation...). Other techniques developed for this problem are time series decomposition (moving averages, local regression smoothing, STL decomposition...), exponential smoothing methods (averaging methods, Holt–Winters' method, Pegels' classification...), regression-based methods, the Box–Jenkins [3] methodology (AR, MA, ARIMA models and advanced applications like regression with ARIMA, dynamic regression, intervention analysis, multivariate autoregressive models...) and nonlinear methods like threshold autoregression, which will be discussed below.

While both statistical time series modelling and Soft Computing have produced good results, there is a lack of communication between the two areas. This situation is starting to change nowadays, with some researchers trying to build a bridge between both disciplines [4,34,39,33]. These works deal mainly with comparisons between methods, but not with the merger of each area's knowledge. The works by Medeiros et al. [23–28] are an exception to this situation, as they do face this problem, dealing mainly with links between artificial neural networks and nonlinear statistical methods. Nevertheless, as part of the core idea behind Soft Computing (to explore the hybridisation of methods and tools coming from different areas), it is interesting to study how we could develop new hybrid methods for time series modelling. These methods could be based on the latest AI developments and should also make extensive use of the knowledge gained by statistical time series modelling.

In this work we present a promising approach towards the aforementioned goal. We will prove how a TSK rule can be seen as a localised autoregressive model, and this basic idea will allow us to think of fuzzy rule-based systems as a generalisation of the regime-switching autoregressive models. The structure of this paper is as follows: in Section 2 we cover briefly the fuzzy rule-based systems (FRBS) and their application to time series modelling, while in Section 3, we review the main concepts behind Box–Jenkins methodology. Section 4 shows how a TSK fuzzy rule relates to linear autoregressive models, in particular to AR models. Section 5 covers the nonlinear regime-switching autoregressive models TAR and STAR, and Section 6 shows how an FRBS relates to them. In Section 7 we explore the consequences of the connection between the models, offering a statistical inference-based methodology to build FRBS incrementally, while Section 8 contains an econometric example of the relation between FRBS and the STAR model. Finally, conclusions and future lines of research are presented in Section 9.

2. Fuzzy rules and fuzzy rule-based systems for time series modelling

The fuzzy rule-based system (FRBS) or fuzzy inference system (FIS) is a popular computing framework based on the concepts of Fuzzy Set Theory, fuzzy *IF-THEN* rules, and fuzzy reasoning [44,16]. It has found successful applications in a wide variety of fields, such as automatic control, data classification, decision analysis, expert systems, robotics, pattern recognition and forecasting.

The basic structure of a FRBS consists of three conceptual components: a *rule base*, which contains a selection of fuzzy rules; a *database*, which defines the membership functions used in the fuzzy rules and a *reasoning mechanism*, which performs the inference procedure (*generalised modus ponens*) upon the rules and given facts to derive a reasonable output or conclusion.

Note that the basic FRBS can take either fuzzy or crisp inputs (which are viewed as fuzzy singletons), but the outputs it produces are almost always fuzzy sets. Sometimes it is necessary to have a crisp output, especially in a situation where a fuzzy inference system is used as a controller. Therefore, we need a method of *defuzzification* to extract a crisp value that best represents a fuzzy set.

With crisp inputs and outputs, an FIS implements a nonlinear mapping from a real input space to a real output space. This mapping is accomplished by a number of fuzzy *IF-THEN* rules, each of which describes the local behaviour of the mapping. In particular, the antecedent of a rule defines a fuzzy region in the input space, while the consequent specifies the output in that fuzzy region.

There are several types of FRBS, cf. [29], whose differences mainly fall in the type of the consequents of their fuzzy rules and their reasoning procedures. If the consequent is a fuzzy set, we have a Mamdani fuzzy model. If, on the other hand, the consequent is a crisp function, what we have is called a Takagi–Sugeno–Kang fuzzy model (TSK fuzzy

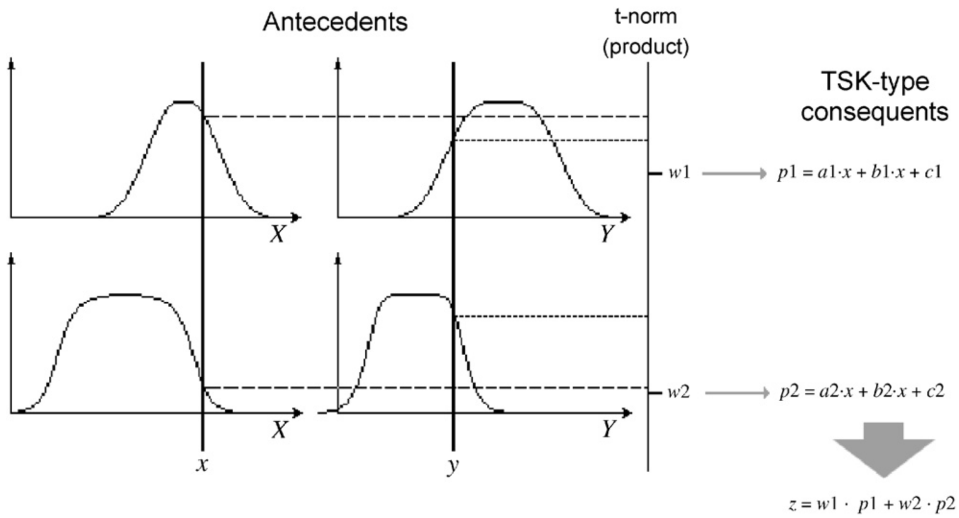

Fig. 1. TSK fuzzy reasoning for multiple rules with multiple antecedents.

model). Finally, if the consequent is represented by a fuzzy set with a monotonical membership function, i.e. a crisp value, the system is called Tsukamoto fuzzy model. In this work we will focus our attention on TSK fuzzy models.

2.1. Takagi–Sugeno–Kang fuzzy rule-based system

A fuzzy rule of type TSK has the following shape:

$$\text{IF } x_1 \text{ IS } A_1 \text{ AND } x_2 \text{ IS } A_2 \text{ AND } \ldots \text{ AND } x_d \text{ IS } A_d$$
$$\text{THEN } y = \mathbf{b}'\mathbf{x} = b_0 + b_1x_1 + b_2x_2 + \cdots + b_dx_d, \tag{1}$$

where x_i are input variables, A_j are fuzzy sets for input variables and y is a linear output function.

The fuzzy reasoning mechanism for two TSK rules is illustrated in Fig. 1. The *firing strength* of the *i*th rule is obtained as the t-norm (usually, multiplication operator) of the membership values of the premise part terms of the linguistic variables:

$$\omega_i(\mathbf{x}) = \prod_{j=1}^{d} \mu_{A_j^i}(x_j), \tag{2}$$

where the shape of the membership function of the linguistic terms $\mu_{A_j^i}$ can be chosen from a wide range of functions. One of the most common is the Gaussian bell,

$$\mu_A(x) = \exp\frac{-(x-c)^2}{2\sigma^2}, \tag{3}$$

but it can also be a logistic function,

$$\mu_A(x) = \frac{1}{1+\exp((c-x)/\sigma^2)}, \tag{4}$$

or even a triangular or trapezoid function.

The overall output is computed as a weighted average or weighted sum of the rules output. In the case of the weighted sum, the output expression is

$$y(\mathbf{x}) = \mathrm{G}(\mathbf{x};\boldsymbol{\psi}) = \sum_{i=1}^{R} \mathbf{b}_i'\mathbf{x}\cdot\omega_i(\mathbf{x}), \tag{5}$$

where G is the general nonlinear function with parameter $\boldsymbol{\psi}$, and R denotes the number of fuzzy rules included in the system.

2.2. TSK rules for univariate time series modelling

When applied to model or forecast a univariate time series $\{y_t\}$, a TSK FRBS usually makes use of the following type of adapted rules:

$$\text{IF } y_{t-1} \text{ IS } A_1 \text{ AND } y_{t-2} \text{ IS } A_2 \text{ AND } \dots \text{ AND } y_{t-p} \text{ IS } A_p \quad \text{THEN } y_t = b_0 + b_1 y_{t-1} + b_2 y_{t-2} + \cdots + b_p y_{t-p}. \tag{6}$$

In this rule, all the variables y_{t-i} are lagged values of the time series, $\{y_t\}$. This means that, in order to apply an optimisation procedure (usually gradient descent-based) to find suitable values for the parameters of the FRBS, we need to build input–output vectors of the form $(y_{t-p}, y_{t-p+1}, \dots, y_{t-1}; y_t)$, which in turn implies that, if we had N samples from the series, we would be able to use $N - p + 1$ data in the parameter tuning process.

3. Linear autoregressive models

The most popular class of linear time series models consists of autoregressive moving average (ARMA) models, including purely autoregressive models (AR) and purely moving-average (MA) models as special cases. ARMA models are frequently used to model linear dynamic structures, to depict linear relationships among lagged variables, and to serve as vehicles for linear forecasting. We will center our attention in the autoregressive type.

3.1. AR model

An *autoregressive model* of order $p \geqslant 1$ is defined as

$$y_t = b_0 + b_1 y_{t-1} + \cdots + b_p y_{t-p} + \varepsilon_t, \tag{7}$$

where $\{\varepsilon_t\} \sim \text{NID}(0, \sigma^2)$, usually known as *white noise* or *random shocks*. For this model we write $\{y_t\} \sim \text{AR}(p)$, and the time series $\{y_t\}$ generated from this model is called the AR(p) process.

Model (7) represents the current state y_t through its immediate p past values $y_{t-1}, \dots, y_{t-p}$ in a linear regression form. It explicitly specifies the relationship between the current value and its past values. The model is easy to implement and therefore is arguably the most popular time series model in practice.

4. TSK fuzzy rules are local AR models

Once we have briefly reviewed the Box–Jenkins models and the TSK fuzzy inference systems, it is easy to prove that

Proposition 1. *When used for time series modelling, a TSK fuzzy rule can be seen as a local AR model, applied on the state-space fuzzy subset defined by the rule's antecedent.*

Comparing the expression of a fuzzy rule applied on a time series, Eq. (6), and the expression for an autoregressive model, Eq. (7), it is clear that the consequent of the rule is exactly an AR model. The antecedent of the rule defines, in fuzzy terms, the necessary conditions for the rule to be applied, that is, the state-space fuzzy subset where the rule is applicable.

The only formal difference is ε_t, the stochastic error term (also called *disturbances* or *random shock series* in the econometric literature), which is absent from the expression of the fuzzy rule. It may be argued, though, that because a time series is a stochastic process, the error term should be present on any model developed for it, even if it makes use of fuzzy logic. This may be considered as further evidence supporting some of the positions held in the long debate between Fuzzy Logic and Probability Theory (see, for example [35] and responses).

In this case, the absence of ε_t comes from the fact that many of the problems usually solved by means of fuzzy rules (and those for which fuzzy rule-based systems were originally designed) are not of a stochastic nature, e.g. control

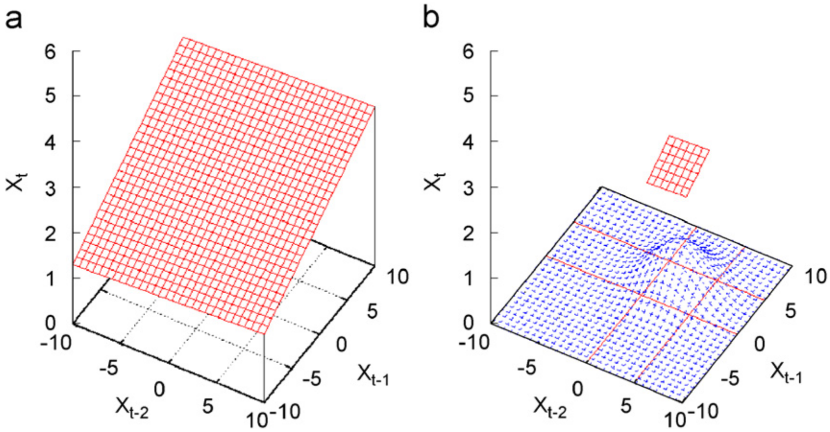

Fig. 2. On the left, the plane defined by the AR(2) model from Section 4.1. On the right, the graphical representation of the fuzzy rule which makes use of that AR(2) model.

problems. Hence it is reasonable to omit the error term in the original formulation of the model. But if we intend to use fuzzy rules for time series modelling (or other stochastic problems), ε_t should be present at least on the notation, if not on the whole modelling procedure (for example, to evaluate it *a posteriori* studying the statistical properties of the residuals, as we shall see in Section 7.1).

Anyway, this connection between the two models opens the possibility of an exchange of knowledge from one field to the other, enabling us to apply what we know about AR models to fuzzy rules and vice versa.

From the point of view of the Box–Jenkins models, this kind of fuzzy rules represents a local AR model which is applied only when some conditions hold. These conditions are given by the terms in the rule's antecedent, and are expressed in *linguistic terms* as the fuzzy membership degree of the lagged variables to some fuzzy sets describing parts of the state-space domain. The linguistic expression has an immediate advantage: we can interpret it in human language terms. Furthermore, this scheme is closely related to the structure of a family of models called *regime-switching autoregressive* models, which include TAR (*threshold autoregressive model*) [40] and STAR (*smooth threshold autoregressive model*) [38] models. The relationship with those models will be studied in Section 6.

Regarding fuzzy rules, its link with autoregressive models may allow us to use the knowledge gathered through years about identification and estimation of those models to develop new, more appropriate methods to set the structure of the consequent of fuzzy rules. For example, it might be interesting to study the proposals concerning variable selection for AR models, which are usually based on selection criteria such as the Akaike information criteria (AIC) or the Schwarz's Bayesian information criterion (BIC), and apply them to the selection of linear terms used in TSK fuzzy rules. As well, we might use other linear models as the consequent of the rules, like for example MA or ARMA models, which opens a wide field of possible innovations for fuzzy rules.

Moreover, as stated above, the addition of the probabilistic error term ε_t to fuzzy rules applied to time series modelling (and its theoretical and practical consequences) should also be considered.

4.1. An example

Let us consider the following AR process:

$$y_t = 2.1 + 0.01 y_{t-1} - 0.1 y_{t-2} + \varepsilon_t, \tag{8}$$

which can be seen as a definition of the relationship between the "output" variable y_t and the "input" variables (y_{t-1} and y_{t-2}). This relationship can be displayed graphically as shown in the left-hand side of Fig. 2.

We might build a fuzzy rule whose consequent is the aforementioned AR model:

$$\text{IF } y_{t-2} \text{ IS } A_1 \text{ AND } y_{t-1} \text{ IS } A_2 \text{ THEN } y_t = 2.1 + 0.01 y_{t-1} - 0.1 y_{t-2} + \varepsilon_t, \tag{9}$$

where (ignoring multiplicative constants)

$$A_i(x) = \exp \frac{-(x-\mu_i)^2}{2\sigma_i^2}, \quad i = 1, 2 \tag{10}$$

are the membership functions of the fuzzy variables of the rule's antecedent. A graphical representation of this fuzzy rule is shown in the right-hand side of Fig. 2, in which $\mu_1 = \mu_2 = 2.5$ and $\sigma_1 = \sigma_2 = 2.0$.

From the graphical representation shown in Fig. 2, it is fairly clear that the application of the fuzzy rule amounts to the application of the AR(2) model in the state-space subset defined by the membership of the "input" variables to the membership functions defined for its antecedent. It must also be noted that this state-space subset is fuzzy, and hence its borders are not crisp.

5. Regime-switching autoregressive models

It is clear that a global linear law is inappropriate to tackle a nonlinear problem, and most of real world problems happen to be nonlinear. A natural alternative would be to break a global linear approximation into several, each one on a subset of the state-space. In an attempt to solve the limitations of the linear approach, Tong [40] presented a "piecewise" linear approximation which consisted in partitioning a state-space into several subspaces.

A *threshold autoregressive* (TAR) model with k ($k \geqslant 2$) regimes is defined as

$$y_t = \sum_{i=1}^{k} \{b_{i,0} + b_{i,1} y_{t-1} + \cdots + b_{i,p_i} y_{t-p_i}\} I(y_{t-d} \in A_i) + \varepsilon_t, \tag{11}$$

where $\{\varepsilon_t\} \sim IID(0, \sigma^2)$, d, $p_1, \ldots, p_k$ are some unknown positive integers, $b_{i,j}$ are unknown parameters, and $\{A_i\}$ forms a partition of $(-\infty, \infty)$ with $\bigcup_{i=1}^{k} A_i = (-\infty, \infty)$ and $A_i \cap A_j = \emptyset, \forall i \neq j$.

When the threshold variable is one of the lagged variables (as in (11)), the model is known as *self-exciting* — yielding the acronym SETAR. In this model, we fit on each subset A_i a linear autoregressive form. The partition is dictated by the threshold variable y_{t-d}, and d is called *delay parameter*. It is often the case that $A_i = (r_{i-1}, r_i]$, with $-\infty = r_0 < r_1 < \cdots < r_k = \infty$, where r_i's are called thresholds.

A key feature of TAR models is the discontinuous nature of the AR relationship as the threshold is passed. If one believes that Nature is generally continuous, one might choose an alternative model called *smooth threshold autoregressive* or *smooth transition autoregressive* (STAR) proposed by Teräsvirta [38], where there is a smooth continuous transition from one linear AR to another, rather than a sudden jump.

In STAR models and variants (cf. [42]), we change the indicator function $I(x)$ in (11) from a *step* function that takes the value zero below the threshold and one above it, to a smooth function with similar characteristics. The STAR model with k regimes [1] is defined as

$$y_t = \sum_{i=1}^{k} \{b_{i,0} + b_{i,1} y_{t-1} + \cdots + b_{i,p_i} y_{t-p_i} + \varepsilon_t\} \mathrm{G}(y_{t-d}; \gamma_i, c_i). \tag{12}$$

The transition is defined by a lagged endogenous variable, y_{t-d}, and the transition function, $\mathrm{G}(y_{t-d}; \gamma, c)$, is a continuous function that is bounded between 0 and 1. Different choices for the transition function give rise to different types of regime-switching behaviour. A popular choice for $\mathrm{G}(s_t; \gamma, c)$ is the first-order logistic function,

$$\mathrm{G}(\gamma(y_{t-d} - c)) = (1 + \exp(\gamma(y_{t-d} - c)))^{-1}, \tag{13}$$

and the resultant model is called the logistic STAR (LSTAR).

[1] It can be noted that in the original proposal $k = 2$, but the model has since been extended in the context of real world cases where more regimes were required. This generalisation does not affect any of its properties, and we will assume it here.

In the LSTAR model, we redefine the transition function $\mathrm{G}(y_{t-d}; \gamma, c)$ of expression (12) as

$$\mathrm{F}_i(y_{t-d}; \gamma_i, c_i) = \begin{cases} 1 - \mathrm{G}(y_{t-d}; \gamma_i, c_i) & \text{if } i = 1, \\ \mathrm{G}(y_{t-d}; \gamma_i, c_i) - \mathrm{G}(y_{t-d}; \gamma_{i+1}, c_{i+1}) & \text{if } 1 < i < k, \\ \mathrm{G}(y_{t-d}; \gamma_i, c_i) & \text{if } i = k, \end{cases} \tag{14}$$

where $\mathrm{G}(y_{t-d}; \gamma_i, c_i)$ is defined as in Eq. (13).

The LSTAR model can be (and usually is) consequently rewritten as

$$y_t = \mathbf{b}_1' \mathbf{x}_t + \sum_{i=2}^{k} \mathbf{b}_i' \mathbf{x}_t \mathrm{G}(y_{t-d}; \gamma_i, c_i) + \varepsilon_t. \tag{15}$$

Each of the parameters c_i in (15) can be interpreted as the threshold between two regimes, in the sense that the logistic function changes monotonically from 0 to 1 as y_{t-d} increases and $\mathrm{G}(c_i; \gamma_i, c_i) = 0.5$. The parameter γ_i determines the smoothness of the transition from one regimen to another. As γ_i becomes very large, the logistic function approaches the indicator function $I(\cdot)$ and hence the change of $\mathrm{G}(y_{t-d}; \gamma_i, c_i)$ from 0 to 1 becomes instantaneous at $y_{t-d} = c$. Consequently, the LSTAR nests threshold autoregressive (TAR) models as a special case. When $c_i \to 0$ the LSTAR model reduces to a linear AR model.

In the LSTAR model, the regime switches are associated with small and large values of the transition variable y_{t-d} relative to c. In certain applications it may be more appropriate to specify the transition function such that the regimes are associated with small and large absolute values of y_{t-d} (again relative to c). This can be achieved by using, for example, the exponential function, in which case the model may be named ESTAR. Other frequently used function is the normal distribution, which yields the acronym NSTAR.

6. STAR models and TSK fuzzy rule-based systems

At this point, we are able to go a step further in the exploration of the connections between nonlinear time series statistical models and fuzzy logic-based models. On the one hand, we have seen AR models as good linear models applicable to prediction problems. As well, we know that a TAR model is basically a set of local AR models, and that it allows for some nonlinearity in its computations. On the other hand, we have seen how a fuzzy rule relates to an AR model, in Proposition 1. Knowing that fuzzy rule-based systems contain sets of fuzzy rules, we may be interested in considering the link existing between TAR models and FRBS.

It is rather clear that there is some parallelism between the two aforementioned models. At a high level, both models are composed of a set of elements (AR-fuzzy rules) which happen to be closely related, as stated above. On a much lower level, both models rely on building a hyper-surface on the state-space which tries to model the relationship between the lagged variables of a time series. Moreover, both models define this hyper-surface as the composition of hyper-planes which apply only in certain parts of the state-space. This can be seen clearly in Fig. 3, which shows the graphical representation of the fuzzy inference system or the STAR model.

Indeed, recalling the expression of a STAR model, Eq. (12), and the expression of the inference mechanism of an additive TSK FIS, Eq. (5), it is trivial to prove the following:

Theorem 1. *TSK fuzzy inference systems with only one linguistic variable in the rule's antecedents are functionally equivalent to STAR models.*

This result entails important implications that may affect the way STAR and FRBS are understood and used. Since every STAR model can be expressed as a restricted fuzzy rule-based system, the properties and tools of this Soft Computing approach may be also applicable to it. On the other hand, given that an FRBS with just one linguistic variable on the antecedent can be seen as a STAR model, there is a great volume of statistical knowledge that could be used on this fuzzy framework. Below we will explore some of these possibilities.

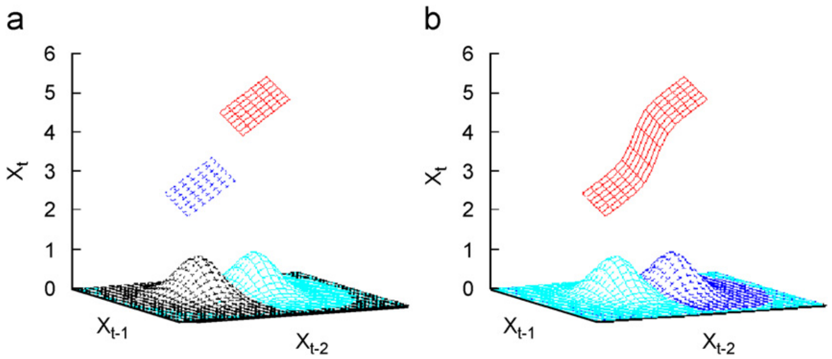

Fig. 3. Two local AR models (or two fuzzy rules) on the left side. On the right side, the STAR model (or the fuzzy inference system) derived from the two AR (or rules) shown in the left.

7. Some consequences

During many years, statisticians have studied linear models for time series developing a vast knowledge of their properties and tools. Later on, systems combining two or more linear models have been proposed, and some of them are still being studied. The behaviour or dynamics of a system composed of just two linear terms is far more complex in statistical terms than the behaviour of each of its elements taken aside, and this makes its study very difficult.

As an example, in an AR(1) model, the first-order autocorrelation of y_t is given by b_1, the coefficient of y_{t-1}. In a STAR model (Eq. (12)) composed of two AR(1) elements, the first order autocorrelation changes gradually from $b_{1,1}$ to $b_{2,1}$ as y_{t-d} increases. This results in a large variety of dynamic patterns being generated by this simple two-regime STAR, making its analytic study much more complicated (in fact, it is usually performed through deterministic simulations). Nevertheless, when approaching a STAR with multiple rules, the increase in complexity makes the development of such an analytic study very difficult if not unfeasible.

In fact, this type of "behavioural" study is normally unfeasible in regular FRBS, which usually contain many rules, and we have not found any attempt in the literature. Notwithstanding, Theorem 1 suggests a new look into FRBS as *data generating processes*, and then concepts as *fixed point*, *equilibrium*, *attractor* or *limit cycle* may be used to better describe and understand the models.

Another interesting question is the selection of the order of the linear models (the consequents of the rules) in an FRBS. So far, when an FRBS was applied to a time series, the order of the linear terms was given by the number of inputs of the system, as in its original formulation the consequents are defined as a linear combination *of the system's inputs*. A look at the TAR and STAR literature unveils some alternatives to this. One is the use of the (partial) autocorrelation to select the order of the linear terms. An alternative procedure, which is proved to be consistent, is the use of an information criterion as the AIC or the BIC. There is even a specific criterion proposed in [41] for a 2-regime model, which should be studied in the FRBS framework.

Stationarity of the generated series is another issue considered by statisticians when dealing with AR and STAR models. As before, it might be desirable to know *a priori* under which conditions an FRBS will generate a stationary series.

7.1. Statistical inference-based specification procedure

In the nonlinear time series modelling framework, it is widely recommended to use a specific-to-general procedure to build the models. Following [7], we propose an inference-based strategy to build an FRBS (with only one antecedent) for time series modelling. We suggest proceeding through the following steps:

(i) *Specify a linear model or default rule*. First we want to specify an appropriate linear AR model of order p for the whole state-space of the time series under investigation. If we can prove that such a default rule is able to capture the complexity of the series, i.e. that the series is linear, we do not need an FRBS.

(ii) *Test the linearity hypothesis*. We have to test the null hypothesis of linearity against the alternative of making the model nonlinear by adding another rule to the default one. Such a linearity test is presented in [38] for a STAR model. It is a Lagrange Multiplier (LM) type test that makes use of a Taylor expansion of the nonlinear term as proposed in [21] to avoid the unidentifiability of the model under the null hypothesis. After Theorem 1, this test is applicable to FRBS with single antecedent.
In this step we must also set the appropriate transition variable that will control the system's behaviour. Teräsvirta [38] suggests using the same LM-type test for this purpose, computing the statistic for several candidate transition variables and choosing one according to the results.

(iii) *Find the sufficient rules*. Starting from an FRBS composed of a default rule (from step (i)) plus a local rule (from step (ii)), we will iteratively apply the aforementioned LM-type test in order to find the number of rules that are required for the series under study.
For generality, assume that we already have an FRBS with r rules obtained through this procedure, and we want to check if a rule $r + 1$ is needed. We write a null hypothesis stating that the model developed so far is complex enough, i.e. that the effect of rule $r + 1$ should be null. The alternative hypothesis will be that we do need another fuzzy rule. If the null hypothesis is accepted, we proceed to the next step, otherwise we add rule $r + 1$ and repeat this step to check if we need to further add more rules.

(iv) *Evaluate the model using diagnostic tests*. Once we have determined the model's structure and parameters, by looking carefully at the residuals ε_t, we can show if it is effectively capturing the characteristics of the time series under study. More precisely, we might want to test for residual autocorrelation, homoscedasticity, normality, and so on.
The LM approach is also used in [6] to develop three important diagnostic checks for SETAR and STAR models. Applying this same approach to FRBS we can test for serial correlation and remaining nonlinearity of the residuals. Moreover, we can check if the parameters of the model remain constant through time.
If the diagnostic tests show that the model is not performing well on the data, we have a chance to modify it prior to its use.

In steps (ii) and (iii) an estimation method is required. Usually, in the STAR framework the cost function is the sum of the squared error and a nonlinear least squares (NLS) algorithm is advocated. When the errors are Gaussian, NLS is equivalent to maximum likelihood. Otherwise, NSL estimation should be seen as a quasi-maximum likelihood estimator.

7.2. Consequences for STAR models

Regarding STAR models, the possibility of seeing them as FRBS opens a wide field of study. The more pragmatic approach used by Soft Computing researchers have resulted in a myriad of FRBS building and optimisation methods (cf. for example [2]) that could now be applied to STAR models. Although some of those methods do not have their statistical properties fully studied, they are proved to be successful in many applications.

In particular, the neuro-fuzzy and evolutive approaches have proved very successful in the development and tuning of FRBS. They can be readily applied to STAR models.

In addition, the interpretative capabilities of fuzzy rules must be exploited in the STAR environment. Interpretation in linguistic terms of already built STAR models might help practitioners to understand and modify them. For example, the introduction of newly gained knowledge about a problem (an econometric one, for example) faced via a STAR model is much easier if a linguistic description of it is available. Notwithstanding, as raised by one of the referees, the additive nature of STAR models difficulties is its linguistic interpretability. STAR models with weighted average of the contributions of each linear model should be studied in the framework of regime switching models, analyzing how the change in the aggregation mechanism could affect their statistical properties. These and other possibilities are to be studied in further works.

8. An econometric example

In [32] a STAR model is built to model the consumer's expenditure in the UK sampled quarterly from 1955 to 1994. The authors apply the methodology outlined above and proposed by [38] to incrementally build a model which ends

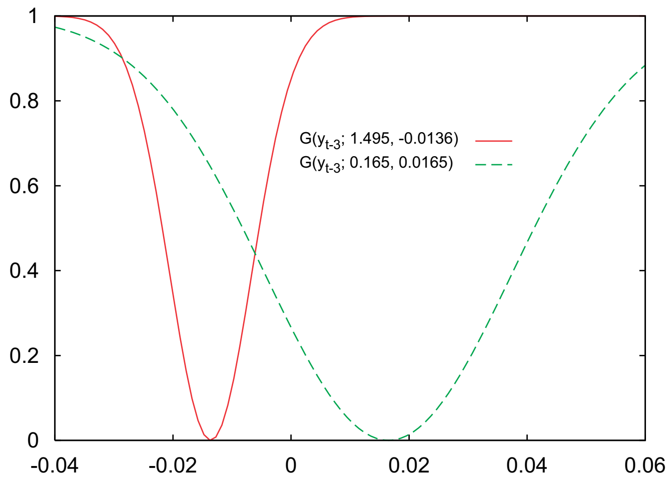

Fig. 4. Exponential transition functions proposed by [32] for the STAR model describing the consumer's expenditure in the UK.

up being composed of three linear models and two exponential transition functions:

$$\begin{aligned} y_t &= 0.0048 + 2.684 y_{t-1} + 0.347 y_{t-2} \\ &\quad + (-2.587 y_{t-3}) G(y_{t-3}; 1.495, -0.0136) \\ &\quad + (-1.500 y_{t-1} - 0.980 y_{t-2}) G(y_{t-3}; 0.165, 0.0165) + \varepsilon_t. \end{aligned} \tag{16}$$

Fig. 4 shows the shape of the transition functions estimated for this model (the exponent of the transition function has been standardised using the sample variance of the dependent variable, 0.000145, as proposed in [32]).

After Theorem 1, this STAR model can be seen as an FRBS with three rules. The shape of the membership functions is such that it could be understood as a fuzzy logic negation of the most commonly used Gaussian membership function. If we assume the membership functions ($G(y_{t-3}; 1.495, -0.0136)$ and $G(y_{t-3}; 0.165, 0.0165)$) to mean *not low* and *not high*, respectively, the rule base might be written as

IF y_{t-3} IS *not low* THEN $y_t = -2.587 y_{t-3}$
IF y_{t-3} IS *not high* THEN $y_t = -1.500 y_{t-1} - 0.980 y_{t-2}$
IN ANY CASE $y_t = 0.0048 + 2.684 y_{t-1} + 0.347 y_{t-2}$.

Or, according to the econometric semantics atributed by the authors to the regimes of the STAR,

IF y_{t-3} IS *recessive* THEN $y_t = -2.587 y_{t-3}$
IF y_{t-3} IS *expansive* THEN $y_t = -1.500 y_{t-1} - 0.980 y_{t-2}$
IN ANY CASE $y_t = 0.0048 + 2.684 y_{t-1} + 0.347 y_{t-2}$.

9. Conclusions and future work

In this work we have shown how a fuzzy rule with linear consequent (TSK rule) is equivalent to a localised Box–Jenkins autoregressive (AR) model when applied to time series modelling. This preliminary result entails us to study widely the links existing between some Soft Computing models and the statistical time series approach.

As a first consequence of this result, we have proved that the smooth transition autoregressive (STAR) model is functionally equivalent to a restricted fuzzy rule-based system (FRBS). This implies that the tools and theoretical results developed in each one of the two areas can be applied to the other area. This has been illustrated with some examples.

A formal procedure to build FRBS, based on its equivalent on the STAR framework, has been outlined. This procedure allows the practitioner to apply statistical inference to fix the number of rules needed and to know under which conditions the model will present desirable properties (as could be the absence of serial correlation of the residuals, for example). Finally we have given an econometric example of the equivalence relation between FRBS and STAR models.

While it may be argued that this test applies to a limited class of FBRS, it certainly sets an interesting path for further developments. It is a promising point of departure to extend this test — or others already applied in STAR and other nonlinear time series model development — to the framework of fuzzy systems.

It is clear that Theorem 1 represents a distinctive piece to build bridges connecting Soft Computing and Statistics applied to time series modelling. This link leads to the exchange of insightful views and knowledge between the two areas. The theorem provides a good tool to further elaborate in the merger of both area allowing for the development of new hybrid methods for time series modelling and forecasting.

In future works, we will explore in-depth how the results shown here affect the way we understand and use STAR models as well as fuzzy inference systems. Moreover, new hybrid models that make use of the advantages of each approach might be devised after the functional equivalence proved here.

References

[1] E.M. Abdelrahim, T. Yahagi, A new transformed input-domain ANFIS for highly nonlinear system modelling and prediction, IEICE Trans. Fundamentals E84-A (8) (2001) 1981–1985.
[2] R. Alcalá, J. Alcalá-Fdez, J. Casillas, O. Cordón, F. Herrera, Hybrid learning models to get the interpretability-accuracy trade-off in fuzzy modelling, Soft Computing, 2005, in press.
[3] G.E.P. Box, G.M. Jenkins, Time Series Analysis: Forecasting and Control, Holden–Day, San Francisco, CA, 1970.
[4] C. Couvreur, P. Couvreur, Neural networks and statistics: a naive comparison, 1997.
[5] P.K. Dash, A.C. Liew, S. Rahman, S. Dash, Fuzzy and neuro-fuzzy computing models for electric load forecasting, Eng. Appl. Artificial Intelligence 8 (4) (1995) 423–433.
[6] O. Eitrheim, T. Teräsvirta, Testing the adequacy of smooth transition autoregressive models, J. Econometrics 74 (1996) 59–76.
[7] P.H. Franses, D. van Dijk, Non-Linear Time Series Models in Empirical Finance, Cambridge University Press, Cambridge, 2000.
[8] A. Gaweda, J. Zurada, Data-driven linguistic modelling using relational fuzzy rules, IEEE Trans. Fuzzy Systems 11 (1) (2003) 121–134.
[9] J.-S. Roger Jang, C.-T. Sun, Predicting chaotic time series with fuzzy if-then rules, In: Proc. of IEEE Internat. Conf. Fuzzy Systems, 1993.
[10] J.-S.R. Jang, ANFIS: adaptive-network-based fuzzy inference system, IEEE Trans. Systems Man and Cybernet. 23 (3) (1993) 665–684.
[11] K. Kalaitzakis, G.S. Stavrakakis, E.M. Anagnostakis, Short-term load forecasting based on artificial neural networks parallels implementation, Electric Power Systems Res. 63 (2002) 185–196.
[12] N. Kasabov, Evolving fuzzy neural networks for supervised/unsupervised online knowledge-based learning, IEEE Trans. Systems Man and Cybernet. Part B 31 (6) (2001) 902–918.
[13] N. Kasabov, On-line learning, reasoning, rule extraction and aggregation in locally optimized evolving fuzzy neural networks, Neurocomputing 41 (2001) 25–45.
[14] C. Kim, I. Yu, Y.H. Song, Kohonen neural network and wavelet transform based approach to short-term load forecasting, Electric Power Systems Res. 63 (2002) 169–176.
[15] J. Kim, N. Kasabov, HyFIS: adaptive neuro-fuzzy inference systems and their application to nonlinear dynamical systems, Neural Networks 12 (1999) 1301–1319.
[16] G.J. Klir, B. Yuan, Fuzzy Sets and Fuzzy Logic, Prentice-Hall, Englewood Cliffs, NJ, 1995.
[17] R.J. Kuo, P. Wu, C.P. Wang, An intelligent sales forecasting system through integration of artificial neural networks and fuzzy neural networks with fuzzy weight elimination, Neural Networks 15 (7) (2002) 909–925.
[18] R.J. Kuo, K.C. Xue, A decision support system for sales forecasting through fuzzy neural networks with asymmetric fuzzy weights, Decision Support System 24 (2) (1998) 105–126.
[19] A. Lendasse, M. Verleysen, E. de Bodt, M. Cottrell, P. Grégoire, Forecasting time-series by Kohonen classification, In: European Symposium on Artificial Neural Networks, Brussels, April 1998.
[20] J. Leski, E. Czogala, A new artificial neural network based fuzzy inference system with moving consequents in if-then rules and selected applications, Fuzzy Sets and Systems 108 (1999) 289–297.
[21] R. Luukkonen, P. Saikkonen, T. Teräsvirta, Testing linearity against smooth transition autoregressive models, Biometrika 75 (1998) 491–499.
[22] L.P. Maguire, B. Roche, T.M. McGuinnity, L.J. McDaid, Predicting a chaotic time series using a fuzzy neural network, Inform. Sci. 112 (1–4) (1998) 125–136.
[23] M.C. Medeiros, C.E. Pedreira, What are the effects of forecasting linear time series with neural networks, Eng. Intelligent Systems Electric. Eng. Commun. 9 (2001) 237–242.
[24] M.C. Medeiros, T. Teräsvirta, Statistical methods for modelling neural networks, Eng. Intelligent Systems Electric. Eng. Commun. 9 (2001) 227–235.
[25] M.C. Medeiros, T. Teräsvirta, G. Rech, Building neural network models for time series: a statistical approach, J. Forecasting, forthcoming.
[26] M.C. Medeiros, A. Veiga, A hybrid linear-neural model for time series forecasting, IEEE Trans. Neural Networks 11 (6) (2000) 1402–1412.
[27] M.C. Medeiros, A. Veiga, Diagnostic checking in a flexible nonlinear time series model, J. Time Series Anal. 24 (2003) 461–482.
[28] M.C. Medeiros, A. Veiga, A flexible coefficient smooth transition time series model, IEEE Trans. Neural Networks 16 (1) (2005) 97–113.
[29] O. Nelles, Nonlinear System Identification: From Classical Approaches to Neural Networks and Fuzzy Models, Springer, Berlin, 2001.

[30] J. Nie, Nonlinear time-series forecasting: a fuzzy-neural approach, Neurocomputing 16 (1997) 63–76.

[31] D. Nieto, J.L. Aznarte, F. Alba, J.M. Benítez, C. Díaz de la Guardia, C. De Linares, Modelling and forecasting *Olea Europaea* L. airborne pollen concentrations in Granada (Southern Spain) using Soft Computing, In: Polen, vol. 14, 2004, pp. 372–373.

[32] N. Ocal, D.R. Osborn, Business cycle non-linearities in UK consumption and production, J. Appl. Econometrics 15 (2000) 27–43.

[33] G. Rech, Forecasting with artificial neural network models, Working Paper Series in Economics and Finance 491, Stockholm School of Economics, February 2002, available at ⟨http://www.ideas.repec.org/p/hhs/hastef/0491.html⟩.

[34] W.S. Sarle, Neural networks and statistical models, In: Proc. of the Nineteenth Annual SAS Users Group International Conference, April 1994, Cary, NC, SAS Institute, 1994, pp. 1538–1550.

[35] N.D. Singpurwalla, J.M. Booker, Membership functions and probability measures of fuzzy sets, J. Amer. Statist. Assoc. 99 (467) (2004) 867–877.

[36] J.A. Sánchez-Mesa, C. Galán, J.A. Martínez Heras, C. Hervás Martínez, The use of a neural network to forecast daily grass pollen concentration in a Mediterranean region: the southern part of the Iberian Peninsula, Clinical Experimental Allergy 32 (2002) 1606–1612.

[37] M. Tamimi, R. Egbert, Short term electric load forecasting via fuzzy neural collaboration, Electric Power Systems Res. 56 (2000) 243–248.

[38] T. Teräsvirta, Specification, estimation and evaluation of smooth transition autoregresive models, J. Amer. Statist. Assoc. 89 (1994) 208–218.

[39] T. Teräsvirta, D. van Dijk, M. Medeiros, Linear models, smooth transition autoregressions, and neural networks for forecasting macroeconomic time series: a re-examination, Technical Report 561, Stockholm School of Economics, July 2004.

[40] H. Tong, On a threshold model, Pattern Recognition Signal Process., 1978.

[41] H. Tong, Non-linear time series, A Dynamical System Approach, Oxford Statistical Science Series, Clarendon Press, Oxford, 1990.

[42] D. van Dijk, T. Teräsvirta, P.H. Franses, Smooth transition autoregresive models — a survey of recent developments, Econometric Rev. 21 (1) (2002) 1–47.

[43] J. Vázquez-Abad, F. Fdez-Riverola, J.M. Corchado, Forecasting economic cycles with conectionist models, In: I Internat. Meeting on Economic Cycles, 2000.

[44] L.A. Zadeh, Fuzzy sets, Inform. and Control 3 (8) (1965) 338–353.