International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems

TESTING FOR REMAINING AUTOCORRELATION OF THE RESIDUALS IN THE FRAMEWORK OF FUZZY RULE-BASED TIME SERIES MODELLING

JOSÉ LUIS AZNARTE

Centre for Energy and Processes
MINES ParisTech, France
Jose-Luis.Aznarte@mines-paristech.fr

MARCELO C. MEDEIROS

Department of Economics
Pontifical Catholic University of Rio de Janeiro, Brasil
mcm@econ.puc-rio.br

JOSÉ M. BENÍTEZ

Department of Computer Science and A.I.
CITIC – UGR, University of Granada, Spain
J.M.Benitez@decsai.ugr.es

Received (received date)
Revised (revised date)

In time series analysis remaining autocorrelation in the errors of a model implies that it is failing to properly capture the structure of time-dependence of the series under study. This can be used as a diagnostic checking tool and as an indicator of the adequacy of the model.

Through the study of the errors of the model in the Lagrange Multiplier testing framework, in this paper we derive (and validate using simulated and real world examples) a hypothesis test which allows us to determine if there is some left autocorrelation in the error series. This represents a new diagnostic checking tool for fuzzy rule-based modelling of time series and is an important step towards statistically sound modelling strategy for fuzzy rule-based models.

Keywords: statistical test, fuzzy rule based models, residual analysis, autocorrelation.

1. Introduction

In general, once a time series model is built and estimated, it has to be evaluated. This is true in the Soft Computing framework as well as in the classical Statistics approach. By evaluating a model we aim at finding out if the model satisfies a set of quality criteria that allow us to say if the model is actually succeeding in capturing the interesting characteristics of the system under study.

Notwithstanding, this set of evaluation criteria is heavily dependent on several considerations: the final use that the model is built for, the inner characteristics of the system that are to be captured and whether the emphasis is put on the empirical behaviour of the model or if there are theoretical considerations that are considered to be more important.

This is evident when we consider the evaluation means used in the Soft Computing field as opposed to those used in the statistical approach to time series analysis.

In the usually engineering-oriented Soft Computing framework, there has been an overwhelming preeminence of just one evaluation criterion, and this has been the *goodness of fit*. Generally, evaluation of a model consists on computing the prediction (or classification) error produced when it is faced with a previously unseen problem of the same type of the one used to estimate it. This measure, in its different flavours (mean squared error, mean average error and so on) is affected by some inherent limitations: it is not very meaningful for a single model unless compared against other models, and is usually range-dependent, which makes it difficult to compare the same model applied to different problems represented by data sets with different characteristics.

On the other hand, evaluation in the statistical approach to time series has usually more to do with obtaining an estimate of the probability that the model is effectively capturing the interesting characteristics of the data set, and this is achieved, among other forms, through developing hypothesis tests, also known as misspecification tests.

The inclusion of the error term ε_t in the expression of Fuzzy Rule-Based Models (FRBM) in the context of time series analysis has been suggested[2]. In general, the main assumption behind modelling is that a part of the system under study behaves according to a model but there is another part which cannot be explained by it (and which is described by a known or unknown probability distribution). This is the main idea encoded in the expression of the general time series model

$$y_t = \mathrm{G}(\mathbf{x}_t; \mathbf{\Psi}) + \varepsilon_t, \tag{1}$$

where $\mathbf{x}_t$ is the input of the system, usually a vector of time series lagged values, and $\mathbf{\Psi}$ is a vector of parameters. The idea is also behind the diagnostic checking procedure presented here.

It is interesting to obtain a precise knowledge about the series of the errors of the model, usually referred to as the series of *residuals* and expected to be normal, independent and identically distributed, i.e., $\varepsilon_t \sim \mathrm{NID}(0, \sigma^2)$. For example, we could be interested in determining if its values are autocorrelated. If the residuals were autocorrelated, that would mean that the model is failing to capture an important part of the behaviour of the series, and hence it should be respecified.

The test presented in this paper is developed under the same framework of the Breusch-Godfrey-Pagan test[15,16,17,5] and is a Lagrange Multiplier (LM) applied to the errors of fuzzy rule based models. The test proposed in this paper is similar to the ones discussed in Medeiros and Veiga, 2003[29] and Eirthein and Teräsvirta, 1996[10]. Our main goal is to show that the LM testing framework of Breusch-Godfrey-Pagan can be applied to a fuzzy rule based model as long as the model has been estimated consistently.

Other existing nonlinear correlation tests as the one proposed by Zhang *et al*, 2007[37] and the alternatives reviewed therein are also applicable to FRBM. To our knowledge, so far no attention has been paid to this field of diagnostic checks in the framework of FRBM, and this work should be seen as an invitation to FRBM researchers and practitioners to further study and systematically apply these procedures.

This paper is one more step in the authors' line towards the development and difussion of statistical inferecen methods in the framework of Computational Intelligence. The relevance of this issue has already being recognized by a number of researchers from both fields —i.e. Statistics and Computational Intelligence–, for example, [27,13,11,12]

The structure of the paper is as follows: in Section 2 we will briefly review the fuzzy rule-based models in their application to time series forecasting. Section 3 contains the derivation of the test and a simplified procedure to calculate it. Finally, Section 4 and Section 5 show a Monte Carlo experiment with simulated series and the application of the test to three real world time series, respectively. The paper ends with Section 6, where the main conclusions are gathered.

2. Fuzzy rule-based models for time series forecasting

When dealing with time series problems (and, in general, when dealing with any problem for which precision is more important than interpretability), the Takagi-Sugeno-Kang paradigm is preferred over other variants of FRBM. When applied to model or forecast a univariate time series $\{y_t\}$, the autoregressive rules used by a TSK FRBM are expressed as:

$$\begin{aligned}\text{rule } i: \ & \text{IF } y_{t-1} \text{ IS } A_{i1} \text{ AND } \ldots \text{ AND } y_{t-q} \text{ IS } A_{iq} \\ & \text{THEN } y_t = \mathbf{b}_i'\mathbf{x}_t = b_{i0} + b_{i1}y_{t-1} + \ldots + b_{ip}y_{t-p}. \end{aligned} \tag{2}$$

In this rule, all the variables y_{t-j} are lagged values of the time series.

The value of p and q determine, respectively, the number of variables present in the consequent and in the antecedent of the rule. We will call $\mathbf{x}_t \in \mathbb{R}^p, \mathbf{x}_t = (y_{t-1}, \ldots, y_{t-p})$ the *consequent variables* and $\mathbf{z}_t \in \mathbb{R}^q$ the *antecedent variables*, being common that $\mathbf{z}_t$ is composed of a subset of the elements of $\mathbf{x}_t$ (thus $q \leq p$).

Concerning the fuzzy reasoning mechanism for TSK rules, the *firing strength* of the ith rule is obtained as the t-norm (usually, multiplication operator) of the membership values of the of the linguistic variables included in the antecedent:

$$f(\mathbf{z}_t; \boldsymbol{\psi}_i) = \prod_{j=1}^{q} \mu_{A_{ij}}(y_{t-j}), \tag{3}$$

where $\boldsymbol{\psi}_i$ is the vector of parameters for the i-th rule. The shape of the membership function of the linguistic terms $\mu_{A_{ij}}$ can be chosen from a wide range of functions. Some of the usual functional forms are the Gaussian bell

$$\mu_{A_{ij}}(y_{t-j}; c_{ij}, \sigma_{ij}) = \exp\left(\frac{-(y_{t-j} - c_{ij})^2}{2\sigma_{ij}^2}\right), \tag{4}$$

the logistic function

$$\mu_{A_{ij}}(y_{t-j}; c_{ij}, \omega_{ij}) = \left(1 + \exp\left((c_{ij} - \omega_{ij}y_{t-j})\right)\right)^{-1}, \tag{5}$$

and also non-derivable functions such as triangular or trapezoidal functions. In this paper, we will consider the case of the logistic function. By substituting (5) in (3) and operating results in a rule firing strength defined as

$$f(\mathbf{z}_t;\boldsymbol{\psi}_i) = \left(1 + \exp\left(\gamma_i(c_i - \boldsymbol{\omega}_i'\mathbf{z}_t)\right)\right)^{-1}, \tag{6}$$

where $\boldsymbol{\psi}_i = (\gamma_i, \boldsymbol{\omega}_i, c_i)$ and γ_i is the norm of $\boldsymbol{\omega}_i$.

The overall output is computed as a weighted average or weighted sum of the rules' outputs. In the case of the weighted sum, the output expression for a fuzzy rule-based model with k rules is:

$$y_t = \mathrm{G}\left(\mathbf{x}_t, \mathbf{z}_t; \boldsymbol{\Psi}\right) + \varepsilon_t = \sum_{i=1}^{k} f(\mathbf{z}_t;\boldsymbol{\psi}_i) \cdot \mathbf{b}_i'\mathbf{x}_t + \varepsilon_t, \tag{7}$$

where G is a general nonlinear autoregressive function with parameters $\boldsymbol{\Psi} = (\boldsymbol{\psi}_1, \mathbf{b}_1, \ldots, \boldsymbol{\psi}_k, \mathbf{b}_k)$. While many TSK FRBMs perform a weighted average to compute the output, additive FRBMs are also a common choice. They have been used in a large number of applications[6,21,22,8].

It has been proved[3] that this specification of the FRBM nests some models from the autoregressive regime switching family. More precisely, it is closely related with the Threshold Autoregressive model (TAR)[35], the Smooth Transition Autoregressive model (STAR)[34], the Linear Local-Global Neural Network (L^2GNN)[33] and the Neuro-Coefficient STAR[30].

This relation gave place to an exchange of knowledge and methods from the statistical framework characterising those models to the fuzzy rule-based modelling of time series. For instance, a linearity test against FRBMs has been developed[2], and other contributions are yet to come.

3. Testing for remaining autocorrelation in the residuals

If we are able to find any remaining autocorrelation in the residuals series $\{\varepsilon_t\}$, we would be able to conclude that our model is failing to capture a part of the inner behaviour of the series, and that it should hence be re-specified.

To study the autocorrelation of the residuals, it might seem a good idea to use the autocorrelation function (ACF) and the partial autocorrelation function (PACF) functions, or the Ljung-Box (LB) test[23]. It should be clear that the LB test, which is based on the estimated autocorrelation function (ACF) of the residuals, is not a valid test in the context of nonlinear models. The reason is that the LB test does not take into account that the estimated model is nonlinear and its asymptotic null distribution is unknown if the test is based on estimated residuals of a nonlinear model [10].

Consider the following FRBM with autocorrelated errors:

$$\begin{array}{l} y_t = \mathrm{G}\left(\mathbf{x}_t, \mathbf{z}_t; \boldsymbol{\Psi}\right) + \varepsilon_t = \sum_{i=1}^{k} f(\mathbf{z}_t;\boldsymbol{\psi}_i) \cdot \mathbf{b}_i'\mathbf{x}_t + \varepsilon_t \\ \varepsilon_t = \boldsymbol{\rho}'\boldsymbol{\nu}_t + u_t \end{array}, \tag{8}$$

where r is the order of the autocorrelation, $\boldsymbol{\rho} = [\rho_1, \rho_2, ..., \rho_r]$ is a vector of parameters, $\boldsymbol{\nu}_t' = [\varepsilon_{t-1}, \varepsilon_{t-2}, ..., \varepsilon_{t-r}]$ and $u_t \sim \mathrm{NID}(0, \sigma^2)$. Note that r is the order of the autoregressor considered for the residuals series.

In the context of this model, we can formulate the null hypothesis of non-autocorrelation of the residuals as $\mathbb{H}_0 : \boldsymbol{\rho} = 0$. We assume that ε_t is a stationary linear process, and furthermore, that under the assumption $\varepsilon_t \sim \mathrm{NID}(0, \sigma^2)$, that is, under $\mathbb{H}_0$, $\{y_t\}$ is stationary and ergodic such that the parameters of (8) can be consistently estimated by nonlinear least squares.

Following Medeiros and Veiga, 2003[29], the conditional normal log-likelihood, given a fixed set of starting values, has the form

$$l_t = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\times \left\{ y_t - \sum_{j=1}^{r}\rho_j y_{t-j} - \mathrm{G}(\mathbf{x}_t, \mathbf{z}_t; \boldsymbol{\Psi}) + \sum_{j=1}^{r}\rho_j \mathrm{G}(\mathbf{x}_{t-j}, \mathbf{z}_{t-j}; \boldsymbol{\Psi}) \right\}^2 . \tag{9}$$

The information matrix related to (9) is block diagonal such that the element corresponding to the second derivative of (9) forms its own block. The variance σ^2 can thus be treated as a fixed constant in (9) when deriving the test statistic. The first partial derivatives of the normal log-likelihood with respect to $\boldsymbol{\rho}$ and $\boldsymbol{\psi}$ are

$$\begin{aligned} \frac{\partial l_t}{\partial \rho_j} &= \left(\frac{u_t}{\sigma^2}\right)\left\{y_{t-j} - \mathrm{G}(\mathbf{x}_{t-j}, \mathbf{z}_{t-j}; \boldsymbol{\Psi})\right\}, \quad j = 1, ..., r \\ \frac{\partial l_t}{\partial \boldsymbol{\Psi}} &= -\left(\frac{u_t}{\sigma^2}\right)\left\{\frac{\partial \mathrm{G}(\mathbf{x}_t, \mathbf{z}_t; \boldsymbol{\Psi})}{\partial \boldsymbol{\Psi}} - \sum_{j=1}^{r}\rho_j \frac{\partial \mathrm{G}(\mathbf{x}_{t-j}, \mathbf{z}_{t-j}; \boldsymbol{\Psi})}{\partial \boldsymbol{\Psi}}\right\} \end{aligned} . \tag{10}$$

Under the null hypothesis, the consistent estimators of (10) are

$$\left.\frac{\partial \hat{l}_t}{\partial \boldsymbol{\rho}}\right|_{\mathbb{H}_0} = \frac{1}{\hat{\sigma}^2}\hat{\varepsilon}_t \hat{\boldsymbol{\nu}}_t \quad \text{and} \quad \left.\frac{\partial \hat{l}_t}{\partial \boldsymbol{\Psi}}\right|_{\mathbb{H}_0} = -\frac{1}{\hat{\sigma}^2}\hat{\varepsilon}_t \hat{\mathbf{h}}_t, \tag{11}$$

where $\hat{\varepsilon}_t$ are the residuals estimated under the null hypothesis, $\hat{\boldsymbol{\nu}}_t = [\hat{\varepsilon}_{t-1}, \ldots, \hat{\varepsilon}_{t-r}]$, $\hat{\sigma}^2 = (1/T)\sum_{t=1}^{T}\hat{\varepsilon}_t^2$ and $\hat{\mathbf{h}}_t$ is the gradient of the model,

$$\hat{\mathbf{h}}_t = \nabla \mathrm{G}(\mathbf{x}_t, \mathbf{z}_t; \hat{\boldsymbol{\Psi}}) = \frac{\partial \mathrm{G}(\mathbf{x}_t; \hat{\boldsymbol{\Psi}})}{\partial \boldsymbol{\Psi}}. \tag{12}$$

The LM statistic can be thus written as

$$\mathrm{LM} = \frac{1}{\hat{\sigma}^2}\sum_{t=1}^{T}\hat{\varepsilon}_t \hat{\boldsymbol{\nu}}_t{}' \times \left\{\sum_{t=1}^{T}\hat{\boldsymbol{\nu}}_t\hat{\boldsymbol{\nu}}_t{}' - \sum_{t=1}^{T}\hat{\boldsymbol{\nu}}_t\hat{\mathbf{h}}_t' \times \left(\sum_{t=1}^{T}\hat{\mathbf{h}}_t\hat{\mathbf{h}}_t'\right)^{-1} \times \sum_{t=1}^{T}\hat{\mathbf{h}}_t\hat{\boldsymbol{\nu}}_t{}'\right\} \times \sum_{t=1}^{T}\hat{\boldsymbol{\nu}}_t{}'\hat{\varepsilon}_t, \tag{13}$$

where T is the length of the series.

Under the condition that the moments implied by (13) exist, the LM statistic is asymptotically distributed as a χ^2 with r degrees of freedom.

3.1. *The test in three steps*

Although it may look complicated at first sight, the application of this test is straightforward. Imagine we have built and adjusted an FRBM as in Equation (7), that is, assuming that the null hypothesis is true, to model and predict a given time series. Having as inputs the residuals and the gradient of the model, we can perform the test in three stages as follows[29]:

(1) Take the residuals of our model and orthogonalize them by regressing them on the gradient $\hat{\mathbf{h}}_t$. Now compute the residual sum of squares of such regression, $SSR_0 = (1/T)\sum_{t=1}^{T}\tilde{\varepsilon}_t^2$.
(2) Regress the residuals of the regression of step (1), $\tilde{\varepsilon}_t$, on $\hat{\mathbf{h}}_t$ and $\hat{\boldsymbol{\nu}}_t$, that is, on the gradient of the model under the alternative hypothesis. Compute a second residual sum of squares $SSR_1 = (1/T)\sum_{t=1}^{T}\hat{v}_t^2$.
(3) Using the two sums of squares previously calculated, we can compute the χ^2 statistic as

$$\mathrm{LM}_{\chi^2} = T\frac{SSR_0 - SSR_1}{SSR_0}.$$

As mentioned by Teräsvirta, 1994[34], the χ^2 statistic suffers from size problems when the series is short. In such cases, we can still compute an alternative statistic which responds to the F (Fisher-Snedecor) distribution as follows:

$$\mathrm{LM}_F = \frac{(SSR_0 - SSR_1)}{r}\left(\frac{SSR_1}{(T-r-n)}\right)^{-1}.$$

Under $\mathbb{H}_0$, the statistic LM_{χ^2} is asymptotically distributed as a χ^2 with r degrees of freedom and LM_F has approximately an F distribution with r and $T-r-n$ degrees of freedom, being n the number of elements of $\hat{\mathbf{h}}_t$ (i.e. the number of parameters of the model under the null). Fixing a significance value allows us to obtain the p-value and to decide upon accepting or rejecting the null hypothesis.

If $\mathbb{H}_0$ is rejected, this implies that the residuals might be autocorrelated up to the r-th order.

4. Monte Carlo power analysis

In order to assess the effectivity of a hypothesis tests its power must be conveniently analyzed. That is the goal of this section.

A common procedure to study the behaviour and modelling capabilities of statistical models is to use synthetic data sets. Recently, this has been also studied in the framework of Soft Computing[4,2].

Here we propose the use of synthetic data sets in order to empirically validate the usefulness of the aforementioned test. The idea behind the experiment is to simulate two sets of series: one set obtained from a model with autocorrelated residuals and another with non-autocorrelated residuals. Then, apply the test over these sets of series to compare the obtained results with the expected outcome of each simulation.

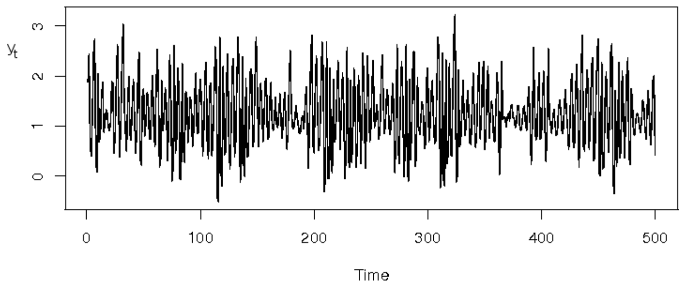

Fig. 1. Synthetic time series obtained by using an FRBM with 3 fuzzy rules (eq. (15)) as data generating process.

In order to do so, we can define a fuzzy rule-based model with the following rules:

$$\begin{aligned} &\text{IF } y_{t-1} \text{ IS } A_{11} \text{ AND } y_{t-2} \text{ IS } A_{12} \text{THEN } y_t = \mathbf{b}_1'\mathbf{x}_t = 0.2 + 0.3y_{t-1} - 0.9y_{t-2} \\ &\text{IF } y_{t-1} \text{ IS } A_{21} \text{ AND } y_{t-2} \text{ IS } A_{22} \text{THEN } y_t = \mathbf{b}_2'\mathbf{x}_t = -0.5 - 1.2y_{t-1} + 0.7y_{t-2}. \\ &\qquad\qquad\text{IN ANY CASE } y_t = \mathbf{b}_3'\mathbf{x}_t = 0.5 + 0.8y_{t-1} - 0.2y_{t-2} \end{aligned} \tag{14}$$

This model can also be written as:

$$\begin{aligned} y_t =&(0.2 + 0.3y_{t-1} - 0.9y_{t-2}) \times f(\mathbf{z}_t; \boldsymbol{\psi}_1)+ \\ &(-0.5 - 1.2y_{t-1} + 0.7y_{t-2}) \times f(\mathbf{z}_t; \boldsymbol{\psi}_2)+ \\ &0.5 + 0.8y_{t-1} - 0.2y_{t-2} + \varepsilon_t, \end{aligned} \tag{15}$$

where the membership functions are logistic (as in Equation (6)) with parameters

$$\begin{aligned} \boldsymbol{\psi}_1 &= [\gamma_1, \boldsymbol{\omega}_1, c_1] = [8.49, (0.7071, -0.7071), -1.0607] \\ \boldsymbol{\psi}_2 &= [\gamma_2, \boldsymbol{\omega}_2, c_2] = [8.49, (0.7071, -0.7071), 1.0607]. \end{aligned}$$

Note that this model has three fuzzy rules, being the third rule a *default rule*, i.e. a rule which applies to the whole input space as its antecedent membership function is always equal to 1.

Now, we can define the residuals of this model as a first order autoregressive process:

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t, \tag{16}$$

where ρ is a parameter defining the autoregression and $u_t \sim NID(0, \sigma^2)$. Throughout this experiment $\sigma = 0.2$, and in order to ensure that ε_t is stationary, we will take $|\rho| < 1$.

By using model (14) with (16) as a data generating process, we can simulate 200 artificial time series with no autocorrelation in the residuals ($\rho = 0$) and 200 series with some autocorrelation in the residuals ($\rho = 0.2$). Figure 1 shows an example of one of the latter. The complexity of these synthetic series is known beforehand, so we can use them to check the properties of the test. In order to do so, we build and train a fuzzy rule-based model for each of those series. The results of the application of the F version of the test at a significance level of 0.05 to the residuals of each model are shown in Table 1.

In this Table we can see how the test is very powerful against type I errors (rejecting the null hypothesis when it is actually true). The first seven rows of the table show the results

Table 1. Results of the F version of the test at the 0.05 significance level for remaining autocorrelation of the residuals using the synthetic series generated by equation (14) with (16).

ρ	r	accept $\mathbb{H}_0$	reject $\mathbb{H}_0$	% correct	average p-value
0.0	1	192	8	96	0.5313
	2	191	9	95.5	0.5032
	3	190	10	95	0.4806
	4	188	12	94	0.4879
	5	189	11	94.5	0.4849
	6	193	7	96.5	0.4945
	7	193	7	96.5	0.5056
0.2	1	7	193	96.5	0.0083
	2	12	188	94	0.0135
	3	17	183	91.5	0.0205
	4	22	178	89	0.0265
	5	30	170	85	0.0332
	6	40	160	80	0.0413
	7	39	161	80.5	0.0483

of the test over the 200 series generated without autocorrelation in the residuals ($\rho = 0$), for the first seven autocorrelation orders ($r = 1, \ldots, 7$). We can see that the null hypothesis was properly accepted in over 94% of the cases for a significance value of 0.05, which states the robustness of the test.

The second part of the Table (rows 8 to 14) show the results of the test when the series did contain some autocorrelation in the residuals ($\rho = 0.2$). As we can see, the test shows also good power against type II errors (not rejecting the null hypothesis when it is actually false). For the different values of the autocorrelation order r, the test properly rejects the null hypothesis in over 80% of the cases.

It is important to note that the power of the test is lower for higher values of r, and this is coherent with the fact that the simulated series were obtained using a first order residual autocorrelation. As expected, the influence of the first order autocorrelation is weaker for higher values of r. In general, the power of the test decreases with r because for first order models the autocorrelation of lag r is $|\rho|^r$, i.e., exponentially decreasing for $|\rho| < 1$.

As usual in hypothesis testing, when the models are closer to the null hypothesis, the discrimination power of the test is reduced. In our case, when the absolute values of ρ approach 0, there is less autocorrelation in the residuals and we can expect a reduction on the power of the test. In order to study this phenomenon, we might be interested in seeing how the test behaves for values of $|\rho|$ close to zero.

In Figure 2, we can see the effect of $|\rho|$ on the power of the test. We repeated the simulations using four different values of ρ, and the graphs show the power (in percentage of proper rejections of the null hypothesis) for different values of the significance and for

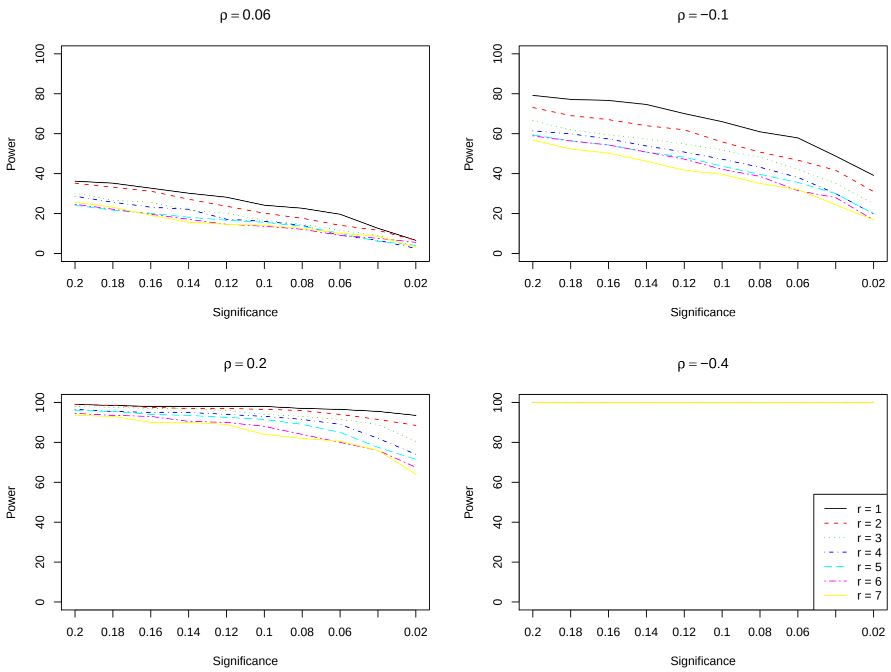

Fig. 2. Power of the test for different values of ρ with respect to the significance level. On each graph, the different lines correspond to different values of r (as shown in the legend of the fourth graph).

different autocorrelation orders. When $\rho = 0.06$ (top left graph), the test does not show a big power, with a maximum of 40% of rejections of the null hypothesis for a significance of 0.2. However, we can see that for a value of $\rho = 0.4$, the test has full power, as it already rejects the null hypothesis for all values of r at a level of significance as high as 0.2.

5. Application examples

After the Monte Carlo experiments, the good properties of the test are clear and we turn our attention to some real world applications. In order to show the usefulness of the test, we chose some widely studied series for which the autoregressive order is known either by experience or by definition. In each case, we proceeded by building a model and applying the test to its residual series.

5.1. *Annual sunspot numbers*

One of the most common benchmarking series is the annual sunspot numbers for the period 1700-1998[36]. To remain coherent with previous studies[30,18,14], we used the transformed data $y_t = 2\left(\sqrt{1+N_t} - 1\right)$ where N_t is the registered number of sunspots at year t.

Table 2. For the annual sunspot numbers, p-values of the test for remaining autocorrelation of the residuals and root mean squared error for models with different orders of lagged variables. Bold faces indicate acceptation of the null hypothesis at a 0.05 significance level. Last row shows the root mean squared error for each model.

	p-value			
	model M1	model M2	model M3	model M4
$r = 1$	0.0084	0.0025	3.0628e-05	**0.2060**
$r = 2$	0.0055	0.0014	0.0001	**0.4168**
$r = 3$	0.0309	0.0021	0.0004	**0.5768**
$r = 4$	0.0001	5.4906e-06	0.0008	**0.7373**
$r = 5$	1.7193e-05	1.0024e-06	0.0028	**0.6727**
$r = 6$	4.0849e-05	6.5735e-06	0.0079	**0.3452**
$r = 7$	1.1870e-05	2.6182e-07	0.0007	0.0129
$r = 8$	3.0858e-06	1.4268e-07	0.0001	0.0028
RMSE	4.7421	4.6966	4.3890	4.0703

We used an FRBM to model this series, and we first chose to use 4 variables in the consequent, corresponding to the first four lags of the series, $\mathbf{x}_t = (y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4})$, and only one antecedent variable, $\mathbf{z}_t = y_{t-1}$, noting this model as M1. When applying the linearity test[2] we rejected the hypothesis of linearity against a fuzzy rule-based model, so we proceeded to build our model using the statistical identification and estimation procedure suggested for neural networks by Medeiros *et al.*, 2006[26], which resulted in a very simple model with two rules and a single linguistic label A:

$$\begin{array}{r} \text{IF } y_{t-1} \text{ IS } A \text{ THEN } y_t = -0.1125 + 0.1182 y_{t-1} - 0.4028 y_{t-2} + 0.7967 y_{t-3} - 0.2057 y_{t-4} \\ \text{IN ANY CASE } y_t = 0.2034 - 0.1166 y_{t-1} + 1.4805 y_{t-2} - 0.8619 y_{t-3} - 0.2058 y_{t-4} \end{array} . \quad (17)$$

The parameters of the logistic firing strength function were determined to be $\psi = (\gamma, \omega, c) = (20.9, 1, 0.0839)$.

Once model M1 was trained, we studied the residuals by using the test for remaining autocorrelation. The null hypothesis was rejected for $r = 1, \ldots, 8$, which indicates that the residuals are still autocorrelated and hence that this model is failing to capture all the information available in the series. The first column of Table 2 shows the p-values obtained by the test for M1. As we can see, the rejections are very strong in the sense that they were associated to small p-values.

In an attempt to reduce the remaining autocorrelation of the residuals, we decided to include more lagged variables in the consequent of the rules. After testing for linearity, we trained new models which used time lags up to y_{t-5} (model M2) and y_{t-6} (model M3) in $\mathbf{x}_t$, but the test kept indicating that remaining autocorrelation was present. See columns 3 and 4 of Table 2 for the p-values, which were even smaller.

However, when including the lagged variables up to the seventh lag in the consequent of the rules (model M4), the results of the test changed significatively. The null hypothesis was then accepted for $r \leq 7$, which indicated that the model, when using 7 lagged variables, succeeds in capturing all the autocorrelation information present in the data. The p-values are shown in the fifth column of Table 2, where we can see that they are clearly different from the previous ones. The findings of the test are also in agreement with the Root Mean Squared Errors (RMSE) showed in the last row of Table 2.

This is coherent with the findings of Medeiros and Veiga, 2005[30], who through the use of information criteria proposed to select the first seven lags as input variables for this series. This is a further validation of the performance of our method.

5.2. *Canadian lynx series*

The second series that we used is also known as a standard benchmarking problem, corresponding to the 10-based logarithm of the number of lynx trapped in a particular zone of Northwest Canada from 1821 to 1934. It has been widely studied in the past.

A first time series model of the Canadian lynx data was fitted by P.A.P Moran in 1953[31]. He proposed an $\mathrm{AR}(2)$ model considering the sample correlogram. Second order autoregression was also chosen by Campbell, 1977[7] in a harmonic-autoregressive combined model and by Medeiros and Veiga, 2005[30] for the NCSTAR model. Here we fix the order of the linear consequents of our model also to 2, i.e., we will use two lagged variables for input.

With $\mathbf{x_t} = \mathbf{z}_t = (y_{t-1}, y_{t-2})$ as inputs, the linearity test[2] recommended the use of an FRBM, rejecting the hypothesis of linearity. We built and trained the model, and applied the remaining autocorrelation test to its residuals.

Through the observation of Table 3, the null hypothesis of absence of remaining autocorrelation in the residuals was rejected for 2nd order models of $r = 1, 2, 3, 4, 5$. This indicates that the model was failing to capture the autoregressive structure of the series.

In the spirit of the preceding example, we added y_{t-3} to the model. In this case we added the new variable both to $\mathbf{z}_t$ (the antecedent) and to $\mathbf{x}_t$ (the consequent), resulting in the following rule base:

$$\begin{aligned}
&\text{IF } y_{t-1} \text{ IS } A_1 \text{ AND } y_{t-2} \text{ IS } A_2 \text{ AND } y_{t-3} \text{ IS } A_3 \\
&\qquad\qquad \text{THEN } y_t = -0.1125 + 0.1182 y_{t-1} - 0.4028 y_{t-2} + 0.7967 y_{t-3} \\
&\qquad \text{IN ANY CASE } y_t = 0.2034 - 0.1166 y_{t-1} + 1.4805 y_{t-2} - 0.8619 y_{t-3}
\end{aligned}$$

with $\boldsymbol{\psi} = [\gamma, \boldsymbol{\omega}, c] = [98.13, (0.6070, -0.1805, 0.7738), 3.73]$.

For such an FRBM, the test does not reject the null hypothesis for all the values of r, as we can see in the second column of Table 3. Hence we assume that by including y_{t-4} the model is capable of capturing all the autoregressive information contained in the series.

5.3. *Mackey-Glass chaotic series*

The Mackey-Glass series, based on the Mackey-Glass differential equation[25] is widely regarded as a benchmark for comparing the generalization ability of different

Table 3. For the Canadian lynx series, p-values of the test for remaining autocorrelation of the residuals and root mean squared error for different orders of lagged variables. Bold faces indicate acceptation of the null hypothesis at a 0.05 significance level and the last row shows the root mean squared error of each model.

	p-value	
	2nd order	3rd order
$r = 1$	0.0162	**0.3312**
$r = 2$	0.0221	**0.1227**
$r = 3$	0.0427	**0.0756**
$r = 4$	0.0309	**0.1226**
$r = 5$	**0.0625**	**0.0838**
$r = 6$	**0.1240**	**0.0799**
$r = 7$	**0.1516**	**0.0663**
$r = 8$	**0.1586**	**0.0780**
RMSE	4.7421	4.6966

methods[20,19,24,9,1]. This series is a chaotic time series generated from the following time-delay ordinary differential equation:

$$\frac{dy_t}{dt} = -by_t + a\frac{y_{t-\tau}}{1 + y_{t-\tau}^{10}}. \tag{18}$$

Following the majority of studies, the series has been generated using a fixed set of values for the parameters: $a = 0.2$, $b = 0.1$ and $\tau = 17$. Usually, the task is to predict the value of the time series at point y_{t+P} from the lagged values $\mathbf{x}_t = (y_t, y_{t-6}, y_{t-12}, y_{t-18})$.

Now, we assumed that we could model this series taking into account just one variable in the antecedent of the rules, and we took it to be $\mathbf{z}_t = y_t$. The linearity test indicates that indeed we need an FRBM to model this series, and the aforementioned identification and estimation algorithm proposed a model with, again, 2 fuzzy rules:

$$\begin{array}{r} \text{IF } y_t \text{ IS } A_\psi \text{ THEN } y_t = \mathbf{b}_1'\mathbf{x}_t \\ \text{IN ANY CASE } y_t = \mathbf{b}_2'\mathbf{x}_t \end{array}, \tag{19}$$

where

$$\mathbf{b}_1 = (-0.2365, 0.3444, -0.2238, 0.9367, 0.8276)$$
$$\mathbf{b}_2 = (0.1265, 0.6533, 0.3387, -0.6337, 0.4476).$$

For this model, the test for remaining autocorrelation in the residuals produced the p-values shown in Table 4. As we can see, the test indicates that there is no first order autocorrelation in the residuals, but this is not the case for higher orders, as for $2 \leq r \leq 8$ the small p-values indicate strong rejections of the null. This model is clearly insufficiently complex as to be used for this series.

As the lagged input structure of the consequents is fixed by the definition of the model, in order to build a model which could capture the autoregressive information

Table 4. For the Mackey-Glass series, p-values of the test for remaining autocorrelation of the residuals and root mean squared error for the two considered models. Bold faces indicate acceptation of the null hypothesis at a 0.05 significance level and the last row shows the root mean squared error of each model.

	p-value	
	model (19)	model (20)
$r = 1$	**0.9835**	**0.3923**
$r = 2$	0.0039	**0.5132**
$r = 3$	0.0001	**0.6683**
$r = 4$	0.0001	**0.6052**
$r = 5$	0.0003	**0.5830**
$r = 6$	3.2395e-05	**0.6305**
$r = 7$	5.2512e-06	**0.7870**
$r = 8$	0.0	**0.8791**
RMSE	0.1750	0.0498

of the series, we added some more elements to the antecedents of the rules, making $\mathbf{z}_t = (y_t, y_{t-6}, y_{t-12}, y_{t-18})$. The new model obtained used the following rules:

$$\begin{aligned}
&\text{IF } y_t \text{ IS } A_{\psi_1} \text{ AND } y_{t-6} \text{ IS } A_{\psi_1} \text{ AND } y_{t-12} \text{ IS } A_{\psi_1} \text{ AND } y_{t-18} \text{ IS } A_{\psi_1} \text{ THEN } y_t = \mathbf{b}_1'\mathbf{x}_t \\
&\text{IF } y_t \text{ IS } A_{\psi_2} \text{ AND } y_{t-6} \text{ IS } A_{\psi_2} \text{ AND } y_{t-12} \text{ IS } A_{\psi_2} \text{ AND } y_{t-18} \text{ IS } A_{\psi_2} \text{ THEN } y_t = \mathbf{b}_2'\mathbf{x}_t \\
&\text{IN ANY CASE } y_t = \mathbf{b}_3'\mathbf{x}_t
\end{aligned} \tag{20}$$

with the consequent parameters

$$\begin{aligned}
\mathbf{b}_1 &= (0.7134, -0.7593, 0.2231, -1.6839, 0.9724) \\
\mathbf{b}_2 &= (-0.3172, -0.2163, 0.3228, -0.1707, -1.2815) \\
\mathbf{b}_3 &= (0.2809, -0.3744, 0.2153, -1.1782, -2.0252464)
\end{aligned}$$

and the antecedent parameters $\boldsymbol{\psi}_1 = (25.02, (0.55, -0.46, -0.68, -0.09), -1.95)$ and $\boldsymbol{\psi}_2 = (7.34, (0.07, 0.79, -0.59, -0.12), 0.09)$.

Column 2 of Table 4 show the p-values of the remaining autocorrelation test applied to the residuals of this FRBM. As we can see, the null hypothesis is accepted in all the cases, so, again, we conclude that this model succeeds in properly capturing the autoregressive component of this series.

These examples show how the test can be effectively used in the common application of FRBMs to time series forecasting. When facing a real world series using an FRBM, investigators should apply the test immediately after finishing the training process, to verify that the basic assumption of their model is effectively being observed. In case that the test rejected the null hypothesis, they should restart the modelling process.

These experiments were carried out using the publicly available open source implementation of this test contained in the `R` package `tsDyn` [32] `http://cran.r-project.org/web/packages/tsDyn/index.html`.

6. Conclusions

In this paper we have shown how to apply hypothesis testing against linear independence of the residuals of an FRBM, when used in the framework of time series modelling and analysis.

The application of the proposed test allows the practitioner to gain a deeper insight about the goodness of his/her model, and to discard it if it fails to capture the underlying autoregressive information of the data. The use of the test complements the use of other common error measures as it gives a different type of information about the performance of a given model.

Whereas time series prediction can be considered as a special case of regression, the extension of the results presented in this paper to any regression problem must first solve some open issues as the meaning of the autocorrelation of the residuals if the data is not chronologically ordered. The same holds for the extension to the classification problem, which is nonetheless very different to the regression or autoregression problem.

With respect to the applicability of the test to other computational intelligence methodologies, as long as the asymptotic distribution of the model is known, and as long as the model is properly specified, it should be possible. In fact, the works by Medeiros *et al.*[29,28,27,26] have proved that Neural Networks can be dealt with in the LM testing framework.

This test is an important result which is framed in an on-going effort to provide the fuzzy rule-based modelling of time series with a statistically sound background and with useful statistical methods and procedures.

Acknowledgements

The research included in this work is part of the first author's PhD dissertation at the Department of Computer Science and Artificial Intelligence of the University of Granada. It has been partially supported by the Ministerio de Ciencia e Innovación of the Spanish Government through Research Grant Ref. TIN2009-14575.

References

1. J.L. Alexandrino, C. Zanchettin, and E.C.B.C. Filho. Hybrid intelligent immune system using radial basis function applied to time series analysis. In *Proceedings of the International Joint Conference on Neural Networks*, pages 94–101, 2009.
2. José Luis Aznarte, Marcelo C. Medeiros, and José M. Benítez. Linearity testing for fuzzy rule-based models. *Fuzzy Sets and Systems*, 161(13):1836–1851, July 2010.

3. J.L. Aznarte M. and J.M. Benítez. Neural-autoregressive time series models with fuzzy equivalences. *IEEE Transactions on Neural Networks*, 2010. Submitted.
4. M. Basu and T.K. Ho, editors. *Data Complexity in Pattern Recognition*. Springer, 2006.
5. T. S. Breusch and A. R. Pagan. A simple test for heteroscedasticity and random coefficient variation. *Econometrica*, 47(5):1287–1294, 1979.
6. H.S. Byun and K.H. Lee. A decision support system for the selection of a rapid prototyping process using teh modified topsis method. *Intern. Journal of Advanced Manufacturing Technology*, 26(11–12):1338–1347, 2005.
7. M.J. Campbell and A.M. Walker. A survey of statistical work on the McKenzie river series of annual Canadian lynx trappings for the years 1821 - 1934, and a new analysis. *J. Roy. Statist. Soc. A*, 140:411 – 431, 1977.
8. F. V. Coito, L. B. Palma, and R. N. da Silva. Robust fault diagnosis approach using analytical and knowledge based techniques applied to a water tank system. *International journal of engineering intelligent systems for electrical engineering and communications*, 13(4):237 – 244, 2005.
9. J. Contreras and O. Acuña. Generating dynamic fuzzy models for prediction problems. In *Annual Conference of the North American Fuzzy Information Processing Society - NAFIPS*, 2009.
10. O. Eitrheim and T. Teräsvirta. Testing the adequacy of smooth transition autoregressive models. *Journal of Econometrics*, 74:59–76, 1996.
11. S. García, A. Fernández, J. Luengo, and F. Herrera. A study of statisticial techniques and performance measures for genetics-based machine learning: Accuracy and interpretability. *Soft Computing*, 13(10):959–977, 2009.
12. S. García, A. Fernández, J. Luengo, and F. Herrera. Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power. *Information Sciences*, 180:2044–2064, 2010.
13. S. García and F. Herrera. An extension on 'statistical comparisons of classifiers over multiple datasets' for all pairwise comparisons. *Journal of Machine Learning Research*, 9(2677–2694), 2008.
14. A. Gholipour, C. Lucas, B.N. Araabi, M. Mirmomeni, and M. Shafiee. Extracting the main patterns of natural time series for long-term neurofuzzy prediction. *Neural Computing and Applications*, 16(4-5):383–393, 2007.
15. L. G. Godfrey. Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables. *Econometrica*, 46(6):1293–1301, 1978.
16. L. G. Godfrey. Testing for higher order serial correlation in regression equations when the regressors include lagged dependent variables. *Econometrica*, 46(6):1303–1310, 1978.
17. L. G. Godfrey. *Misspecification Tests in Econometrics*. Number 9780521424592 in Cambridge Books. Cambridge University Press, 1991.
18. E. Gómez-Ramírez, K. Najim, and E. Ikonen. Forecasting time series with a new architecture for polynomial artificial neural network. *Applied Soft Computing Journal*, 7(4):1209–1216, 2007.
19. Y.-C. Hwang, Q. Song, and N. Kasabov. Mufis: A neuro-fuzzy inference system using multiple types of fuzzy rules. In *IEEE International Conference on Fuzzy Systems*, pages 1411–1414, 2008.
20. J.-S.R. Jang. ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Transactions on systems, man and cibernetics*, 23(3), May-June 1993.
21. R.I. John and P.R. Innocent. Modeling uncertainty in clinical diagnosis using fuzzy logic. *Systems, Man, and Cybernetics, Part B, IEEE Transactions on*, 35(6):1340–1350, Dec. 2005.
22. I. Lee, B. Kosko, and W. F. Anderson. Modeling gunshot bruises in soft body armor with an adaptive fuzzy system. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 35(6):1374–1390, 2005.
23. G. M. LJUNG and G. E. P. BOX. On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303, 1978.

24. C. Luque, J.M. Valls, and P. Isasi. Time series prediction evolving Voronoi regions. *Applied Intelligence*, in press, 2009.
25. M.C. Mackey and L. Glass. Oscillation and chaos in physiological control systems. *Science*, 197:287–289, July 1977.
26. M. C. Medeiros, T. Teräsvirta, and G. Rech. Building neural network models for time series: a statistical approach. *Journal of Forecasting*, 25(1):49–75, 2006.
27. M.C. Medeiros and T. Teräsvirta. Statistical methods for modelling neural networks. *Engineering Intelligent Systems*, 9:227–235, 2001.
28. M.C. Medeiros and A. Veiga. A hybrid linear-neural model for time series forecasting. *Neural Networks, IEEE Transactions on*, 11(6):1402–1412, 2000.
29. M.C. Medeiros and A. Veiga. Diagnostic checking in a flexible nonlinear time series model. *Journal of Time Series Analysis*, 24:461–482, 2003.
30. M.C. Medeiros and A. Veiga. A flexible coefficient smooth transition time series model. *IEEE Transactions on Neural Networks*, 16(1):97–113, January 2005.
31. P.A.P. Moran. The statistical analysis of the Canadian lynx cycle. I: structure and prediction. *Aust. J. Zool.*, 1:163 – 173, 1953.
32. A. F. di Narzo, J. L. Aznarte, and M. Stigler. *tsDyn: Time series analysis based on dynamical systems theory*, 2008. R package version 3.1-89.
33. M. Suarez-Farinas, C. E. Pedreira, and M. C. Medeiros. Local global neural networks: A new approach for nonlinear time series modeling. *Journal of the American Statistical Association*, 99:1092–1107, December 2004.
34. T. Teräsvirta. Specification, estimation and evaluation of smooth transition autoregresive models. *J. Am. Stat. Assoc.*, 89:208–218, 1994.
35. H. Tong. On a threshold model. *Pattern Recognition and Signal Processing*, 1978.
36. H. Tong. *Non-linear Time Series: a dynamical system approach*. Oxford Statistical Science Series, Oxford: Clarendon Press, 1990.
37. L. F. Zhang, Q. M. Zhu, and A. Longden. A set of novel correlation tests for nonlinear system variables. *Intern. J. Syst. Sci.*, 38(1):47–60, 2007.