Testing for Serial Independence of the Residuals in the Framework of Fuzzy Rule-based Time Series Modeling

José Luis Aznarte M.
Centre for Energy and Processes
Ecole des Mines de Paris
Sophia Antipolis, France
jose-luis.aznarte@mines-paristech.fr

Antonio Arauzo
Area of Project Engineering
University of Cordoba
Crdoba, Spain
arauzo@uco.es

José Manuel Benítez Sánchez
Department of Computer Science and I.A.
University of Granada
Granada, Spain
J.M.Benitez@decsai.ugr.es

Abstract

In this paper, we propose a new diagnostic checking tool for fuzzy rule-based modelling of time series. Through the study of the residuals in the Lagrange Multiplier testing framework we devise a hypothesis test which allows us to determine if there is some left autocorrelation in the error series. This is an important step towards a statistically sound modelling strategy for fuzzy rule-based models.

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 understand to find out if the model satisfies a set of quality criteria that allow us to say if the interesting characteristics of the system under study are actually being captured by it or not.

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 through developing hypothesis tests, also known as misspecification tests.

The inclusion of the error term ε_t in the expression of FRBM in the context of time series analysis has been suggested [1]. 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 is usually considered to be white noise. This is the main idea encoded in the expression of the general model

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

and it is also behind the diagnostic checking procedure presented here.

It is interesting to obtain a precise knowledge about the series of the residuals, $\{\varepsilon_t\}$, for example determining if

978-0-7695-3872-3/09 $26.00 © 2009 IEEE
DOI 10.1109/ISDA.2009.249

its values are independent and normally distributed. If the residuals were not independent, 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.

2. Fuzzy Rule-based Models for Time Series Analysis

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 rules of a TSK FRBM are expressed as:

$$\begin{aligned} &\text{IF } y_{t-1} \text{ IS } A_1 \text{ AND } y_{t-2} \text{ IS } A_2 \text{ AND } \ldots \text{ AND } y_{t-p} \text{ IS } A_p \\ &\text{THEN } y_t = b_0 + b_1 y_{t-1} + b_2 y_{t-2} + \ldots + b_p y_{t-p}. \end{aligned} \quad (2)$$

In this rule, all the variables y_{t-i} are lagged values of the time series, $\{y_t\}$.

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 premise part terms of the linguistic variables:

$$\omega_i(\mathbf{x}) = \prod_{j=1}^{d} \mu_{A_j^i}(x_j), \quad (3)$$

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, although it can also be a logistic function and even non-derivable functions as a triangular or trapezoidal 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:

$$y_t = \mathrm{G}\left(\mathbf{x}_t; \boldsymbol{\psi}\right) = \sum_{i=1}^{R} \omega_i(\mathbf{x}_t) \cdot \mathbf{b}_i \mathbf{x}_t, \quad (4)$$

where G is the general nonlinear function with parameters $\boldsymbol{\psi}$, and R denotes the number of fuzzy rules included in the system. While many TSK FRBS perform a weighted average to compute the output, additive FRBS are also a common choice. They have been used in a large number of applications, for example [4, 7, 8, 14].

It has been proved [1] 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) [13], the Smooth Transition Autoregressive model (STAR) [12], the Linear Local-Global Neural Network (L^{2}GNN) [11] and the Neuro-Coefficient STAR [9].

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 FRBM has been developed [2], and other contributions are yet to come.

3. Testing for independence of 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 in capturing a part of the inner behaviour of the series, and that it should hence be re-specified.

Consider the following FRBM with autocorrelated errors:

$$\begin{aligned} y_t &= \mathrm{G}(\mathbf{x}_t; \boldsymbol{\psi}) + \varepsilon_t = \textstyle\sum_{i=1}^{r} \mathbf{b}_i \mathbf{x}_t \cdot \mu_i\left(\mathbf{x}_t; \boldsymbol{\psi}_{\mu_i}\right) + \varepsilon_t \\ \varepsilon_t &= \boldsymbol{\pi}' \boldsymbol{\nu}_t + u_t \end{aligned} \quad (5)$$

where the $\boldsymbol{\pi}' = [\pi_1, \pi_2, ..., \pi_s]$ is a vector of parameters, $\boldsymbol{\nu}_t = [\varepsilon_{t-1}, \varepsilon_{t-2}, ..., \varepsilon_{t-s}]$ and $u_t \sim \mathrm{NID}(0, \sigma^2)$. We assume that ε_t is stationary, and furthermore, that under the assumption $\varepsilon_t \sim \mathrm{NID}(0, \sigma^2)$, that is, $\pi = \mathbf{0}$, $\{y_t\}$ is stationary and ergodic such that the parameters of (5) can be consistently estimated by nonlinear least squares.

In the context of this model, we can formulate the null hypothesis of serial independence of the residuals as $\mathrm{H}_0 : \boldsymbol{\pi} = \mathbf{0}$.

The conditional normal log-likelihood, given the fixed starting values, has the form

$$l_t = -\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln \varsigma^2 - \frac{1}{2\varsigma^2} \left\{ y_t - \sum_{j=1}^{s} \pi_j y_{t-j} - \mathrm{G}(\mathbf{x}_t; \boldsymbol{\psi}) + \sum_{j=1}^{s} \pi_j \mathrm{G}(\mathbf{x}_{t-j}; \boldsymbol{\psi}) \right\}^2. \quad (6)$$

The information matrix related to (6) is block diagonal such that the element corresponding to the second derivative of (6) forms its own block. The variance ς^2 can thus be treated as a fixed constant in (6) when deriving the test statistic. The first partial derivatives of the normal log-likelihood with respect to $\boldsymbol{\pi}$ and $\boldsymbol{\psi}$ are

$$\frac{\partial l_t}{\partial \pi_j} = \left(\frac{u_t}{\sigma^2}\right) \left\{ y_{t-j} - \mathrm{G}(\mathbf{x}_{t-j}; \boldsymbol{\psi}) \right\}, j = 1, ..., s \quad (7)$$

$$\frac{\partial l_t}{\partial \boldsymbol{\psi}} = -\left(\frac{u_t}{\sigma^2}\right) \left\{ \frac{\partial \mathrm{G}(\mathbf{x}_t; \boldsymbol{\psi})}{\partial \boldsymbol{\psi}} - \sum_{j=1}^{s} \pi_j \frac{\partial \mathrm{G}(\mathbf{x}_{t-j}; \boldsymbol{\psi})}{\partial \boldsymbol{\psi}} \right\} \quad (8)$$

Under the null hypothesis, the consistent estimators of

(7) are

$$\left.\frac{\partial \hat{l}_t}{\partial \pi_j}\right|_{\mathrm{H}_0} = \frac{1}{\hat{\sigma}^2}\hat{\varepsilon}_t \hat{\boldsymbol{\nu}}_t \quad \text{and} \quad \left.\frac{\partial \hat{l}_t}{\partial \boldsymbol{\psi}}\right|_{\mathrm{H}_0} = -\frac{1}{\hat{\sigma}^2}\hat{\varepsilon}_t \hat{\mathbf{h}}_t, \qquad (9)$$

where $\hat{\boldsymbol{\nu}}_t = [\hat{\varepsilon}_{t-1}, \hat{\varepsilon}_{t-2}, ..., \hat{\varepsilon}_{t-s}]$, $\hat{\varepsilon}_{t-j} = y_{t-j} - \mathrm{G}(\mathbf{x}_{t-j}; \boldsymbol{\psi})$, $j = 1, ..., s$, $\hat{\mathbf{h}}_t = \nabla \mathrm{G}(\mathbf{x}_t; \hat{\boldsymbol{\psi}})$ and $\hat{\sigma}^2 = (1/T)\sum_{t=1}^{T} \hat{\varepsilon}_t$.

The LM statistic is

$$\mathrm{LM} = \frac{1}{\hat{\sigma}^2}\sum_{t=1}^{T} \hat{\varepsilon}_t \hat{\boldsymbol{\nu}}_t{}' \times \left\{ \sum_{t=1}^{T} \hat{\boldsymbol{\nu}}_t \hat{\boldsymbol{\nu}}_t{}' - \sum_{t=1}^{T} \hat{\boldsymbol{\nu}}_t \hat{\mathbf{h}}_t' \times \left(\sum_{t=1}^{T} \hat{\mathbf{h}}_t' \hat{\mathbf{h}}_t \right)^{-1} \times \sum_{t=1}^{T} \hat{\mathbf{h}}_t \hat{\boldsymbol{\nu}}_t{}' \right\} \times \sum_{t=1}^{T} \hat{\boldsymbol{\nu}}_t{}' \hat{\varepsilon}_t \qquad (10)$$

where $\hat{\mathbf{h}}_t = \nabla \mathrm{G}(\mathbf{x}_t; \hat{\boldsymbol{\psi}})$ and $\hat{\nu}_t = [t\mathbf{x}_t', t\mathbf{x}_t'\mu_1(\mathbf{x}_t; \boldsymbol{\psi}_{\mu_1}), ..., t\mathbf{x}_t' \ \mu_s(\mathbf{x}_t; \boldsymbol{\psi}_{\mu_s})]'$.

Under the condition that the moments implied by (10) exist, LM is asymptotically distributed as a χ^2 with s degrees of freedom.

The test can be performed in three stages as follows:

1. Estimate model (4) under the assumption of uncorrelated errors and and compute the residuals $\hat{\varepsilon}_t$. Orthogonalize the residuals by regressing $\hat{\varepsilon}_t$ on $\hat{\mathbf{h}}_t$, and compute the residual sum of squares $SSR_0 = (1/T)\sum_{t=1}^{T} \tilde{\varepsilon}_t^2$.

2. Regress $\tilde{\varepsilon}_t$ on $\hat{\mathbf{h}}_t$ and $\hat{\nu}_t$. Compute the residual sum of squares $SSR_1 = (1/T)\sum_{t=1}^{T} \hat{v}_t^2$.

3. Compute the χ^2 statistic

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

or the F version of the test

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

Under H_0, $\mathrm{LM}_{\chi^2}^{si}$ is asymptotically distributed as a χ^2 with s degrees of freedom and LM_F^{si} has approximately an F distribution with s and $T-s-n$ degrees of freedom.

Upon rejection of H_0, we know that the residuals might be autocorrelated up to order $s = \|\boldsymbol{\pi}\|$, which means that the FRBM has failed to capture the lagged structure of the data.

4. An example: modelling the Lynx time series

The Canadian lynx data set is a commonly used series, corresponding to the annual record of the number of the Canadian lynx ”trapped” in the Mackenzie River district of the North-West Canada for the period 1821 to 1934. These data are actually the total fur returns, or total sales, from the London archives of the Hudson’s Bay Company in the years of 1821 to 1891 and 1887 to 1913; and those for 1915 to 1934 are from detailed statements supplied by the Company’s Fur Trade Department in Winnipeg; those for 1892 to 1896 and 1914 are from a series of returns for the MacKenzie River District; those for the years 1863 to 1927 were supplied by Ch. French, then Fur Trade Commissioner of the Company in Canada. By considering the time lag between the year in which a lynx was trapped and the year in which its fur was sold at auction in London, these data were converted in [6] into the number that were presumably caught in a given year for the years 1821 to 1934 as shown in Figure 1(a).

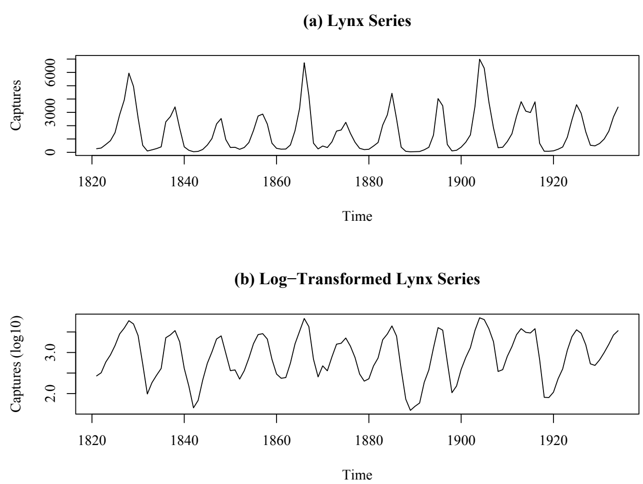

Figure 1. Number of lynx caught in the Mackenzie River district of the North-West Canada from year 1821 to 1934.

A first time series model of the Canadian lynx data was fitted by P.A.P Moran in [10]. He observed that the cycle is very asymmetrical with a sharp and large peak and a relatively smooth and small trough. The log transformation gives a series which appears to vary symmetrically about the mean. As the actual population of lynx is not exactly proportional to the number caught, a better representation would perhaps be obtained by incorporating an additional ”error of observation” in the model, thereby resulting in a more complicated model. The log transformation substantially reduces the effect of ignoring this error of observation;

Table 1. Results of misspecification tests for the lynx problem.

Test for q-order serial correlation		
	AR(2)	FRBM
q	p-value	p-value
1	0.949	0.411
2	0.921	0.622
3	0.813	0.587
4	0.946	0.733
5	0.998	0.234
6	0.953	0.834
7	0.995	0.532
8	0.947	0.424
9	0.944	0.672
10	0.722	0.562
11	0.975	0.623
12	0.826	0.789

therefore, after Moran, nearly all the time series analysis of the lynx data in the literature have used the log-transformed data. Figure 1(b) shows the transformed data.

The aforementioned study, [10], proposed an AR(2) model considering the sample correlogram, and second order autoregression was also chosen by [5] in a harmonic-autoregressive combined model and by [9] for the NCSTAR model. We fix the order of our model also to 2, for these reasons.

The linearity test against a FRBM with sigmoid transition function threw a p-value of 0.000259, while the test against a Gaussian-based FRBM obtained a 0.000115. Both tests indicate that the series is nonlinear and suggest the use of more than one rule.

The modelling cycle ended in both cases when the second regime was added, so the estimated models have just two rules given by

$$y_t = 0.9599 + 1.2514 y_{t-1} - 0.3398 y_{t-2} + (2.5466 + 0.3764 y_{t-1} - 0.7973 y_{t-2})\mu_{\mathsf{S}}(\mathbf{x}_t;\boldsymbol{\psi}_{\mathsf{S}}) + \varepsilon_t \tag{11}$$

in the sigmoid case, with $\boldsymbol{\psi}_{\mathsf{S}} = (\gamma, \boldsymbol{\omega}, c) = (103.1266, [0.4630, 0.8863], 9.4274)$, and

$$y_t = 0.8749 + 1.2302 y_{t-1} - 0.3074 y_{t-2} + (2.0084 + 0.2961 y_{t-1} - 0.6486 y_{t-2})\mu_{\mathsf{G}}(\mathbf{x}_t;\boldsymbol{\psi}_{\mathsf{G}}) + \varepsilon_t \tag{12}$$

in the Gaussian case, where $\boldsymbol{\psi}_{\mathsf{G}} = (\gamma, \mathbf{c}) = (11.0129, [5.8417, 3.6653])$.

The first model obtained a residual standard deviation of $\hat{\sigma}_{\varepsilon,\mathsf{S}} = 0.196$, while the second obtained a value of $\hat{\sigma}_{\varepsilon,\mathsf{G}} = 0.207$. The value obtained for the AIC were $\mathrm{AIC}_{\mathsf{S}} = -314$ and $\mathrm{AIC}_{\mathsf{G}} = -306$ respectively, while the median average percentage error was $\mathrm{MAPE}_{\mathsf{S}} = 5.94\%$ and $\mathrm{MAPE}_{\mathsf{G}} = 6.31\%$.

Once both models were estimated through the standard procedure, we applied a metaheuristic to fine-tune the parameters. After [1]s, a Genetic Algorithm was chosen and applied.

Using the GA to fine tune the parameters, left us with the following two models:

$$y_t = 0.3978 + 1.2560 y_{t-1} - 0.3359 y_{t-2} + (1.0193 + 0.3744 y_{t-1} - 0.7736 y_{t-2})\mu_{\mathsf{S}}(\mathbf{x}_t;\boldsymbol{\psi}_{\mathsf{S}}) + \varepsilon_t \tag{13}$$

in the sigmoid case, with $\boldsymbol{\psi}_{\mathsf{S}} = (\gamma, \boldsymbol{\omega}, c) = (38.9935, [0.4969, 0.8678], 4.1306)$, and

$$y_t = 0.4023 + 1.2224 y_{t-1} - 0.3103 y_{t-2} + (0.8099 + 0.3751 y_{t-1} - 0.7074 y_{t-2})\mu_{\mathsf{G}}(\mathbf{x}_t;\boldsymbol{\psi}_{\mathsf{G}}) + \varepsilon_t \tag{14}$$

in the Gaussian case, where $\boldsymbol{\psi}_{\mathsf{G}} = (\gamma, \mathbf{c}) = (10.000, [2.576, 6.831])$. For these models tuned with the GA, the obtained residual standard deviation was $\hat{\sigma}_\varepsilon = 0.191$ for the sigmoid and $\hat{\sigma}_\varepsilon = 0.205$ for the Gaussian membership function. The value obtained for the AIC were $\mathrm{AIC}_{\mathsf{S}} = -313$ and $\mathrm{AIC}_{\mathsf{G}} = -307$ respectively, while the median average percentage error was $\mathrm{MAPE}_{\mathsf{S}} = 5.90\%$ and $\mathrm{MAPE}_{\mathsf{G}} = 6.26\%$.

As we can see in Table 1, the FRBM model managed to capture most of the autocorrelation of the data, as up to 12th order the null hypothesis of linear independence of the residuals is not rejected. For comparison, we show the p-values of the test for the AR(2) model, which show that the null hypothesis might be rejected in most of the orders for this model. That means that (as the linearity test already show) an AR(2) model fails to capture the inner behaviour of the series, and hence that the series is nonlinear.

5. Conclusions

In this paper we have shown how to apply hypothesis testing against linear independence of the residuals of a FRBM, when used in the framework of time series modeling 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 possible nonlinearity of the data. The use of the test complements the use of other common error measures as the RMSE or the MAE as it gives a different type of information about the performance of a given model.

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

References

[1] J. L. Aznarte M. *Modelling time series through fuzzy rule-based models: a statistical approach*. PhD thesis, Universidad de Granada, 2008.

[2] J. L. Aznarte M., M. Medeiros, and J. M. B. Snchez. Linearity testing against a fuzzy rule-based model. *Fuzzy Sets and Systems*, (submitted).

[3] M. Bulmer. A statistical analysis of the 10-year cycle in Canada. *J. Anim. Ecol.*, 43:701 – 715, 1974.

[4] H. Byun and K. 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.

[5] M. Campbell and A. 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.

[6] C. Elton and M. Nicholson. The ten-year cycle in numbers of the lynx in Canada. *J. Anim. Ecol.*, 11, 1942.

[7] R. John and P. Innocent. Modeling uncertainty in clinical diagnosis using fuzzy logic. *Systems, Man, and Cybernetics, Part B, IEEE Transactions on*, 35(6):1340–1350, Dec. 2005.

[8] 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.

[9] M. Medeiros and A. Veiga. A flexible coefficient smooth transition time series model. *IEEE Transactions on Neural Networks*, 16(1):97–113, January 2005.

[10] P. Moran. The statistical analysis of the Canadian lynx cycle. i: structure and prediction. *Aust. J. Zool.*, 1:163 – 173, 1953.

[11] 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.

[12] T. Tersvirta. Specification, estimation and evaluation of smooth transition autoregresive models. *J. Am. Stat. Assoc.*, 89:208–218, 1994.

[13] H. Tong. On a threshold model. *Pattern Recognition and Signal Processing*, 1978.

[14] C. F. Vieira, 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.