Adaptation of model selecting criteria for nonlinear time series forecasting

Myladis Rocio Cogollo-Flórez
mcogollo@eafit.edu.co
Andrea Molina-Alonso
amolin13@eafit.edu.co

EAFIT University
Mathematical Engineering
Department of Mathematical Sciences
June 10, 2016

Abstract

The selection of a time series model is a well studied problem, because of its importance in forecasting problems. Different model selection criteria have been used, and recently, studies use them combined to find an adaptation of the criteria to reach accuracy. This paper implement some multivariate statistic techniques to find a new adaptation of a weighted criterion for model selection of time series.

Keywords: Model selection criteria, Principal Component Analysis, Analytic Hierarchy Process

1. Introduction

Forecasting problems involve predicting events time periods into the future and it can be found in different fields. Most forecasting problems involve the use of time series data [1, 2]. In time series forecasting the interest is to discover, with some margin of error, future values of a signal or a function of time, X_t, based on its past values and considering its randomness or fluctuating properties [3].

Although it is difficult to identify the type of a time series, in practice they are divided into two general groups: linear and nonlinear. The work by

Box and Jenkins [4] in the 70s, gave way to an important effort in the study and application of construction of linear models by mathematical models that represent autoregressive processes (AR), moving average process (MA), and their combination. Many successful practical experiences have shown that this approach can represent the dynamics of many series of real time, which popularized this class of models both academic and professional fields [5, 6]. However, it has also been found that many real time series have non-linear behavior [2, 7, 8], for which the Box and Jenkins approximation is insufficient to represent these dynamics [5].

Formally, a model of pth-order of nonlinear time series is defined as [9, 10]

$$X_t = f(F_{t-1}; \phi) + a_t \quad , \tag{1}$$

where f is a known nonlinear function of past X_t's and ϕ is a $p \times 1$ vector of parameters. Let $\{X_t\}$ be a stationary and ergodic time series, with F_t the σ-field generated by $\{X_t, X_{t-1}, \ldots\}$. The function f is assumed to have continuous second order derivatives almost surely. The noise process $\{a_t\}$ is assumed to be independent, with mean zero, variance σ_a^2, and finite fourth order moment. It is further assumed that (1) is invertible or equivalently, $\{a_t\}$ is measurable with respect to F_t.

While there is no single methodology for modeling phenomena that have only temporary data for forecasting in the time series, they all follow the key steps of specification, estimation, validation and prognosis. Abraham and Ledolter [11] specify that in general modeling methodology of time series for forecasting consists of two stages: Model-building and forecasting (Figure 1).

Stages to construct a time series model for forecasting [11]

Phase 1: Model-Building Phase

A model for forecasting is constructed from measurements of observations and theory (economics, among others) available. In some cases this theory may suggest certain structures of the model; in other cases, this theory can not exist or be incomplete, and the available data should be used to specify an appropriate model. To choose the structure of forecasting model, the

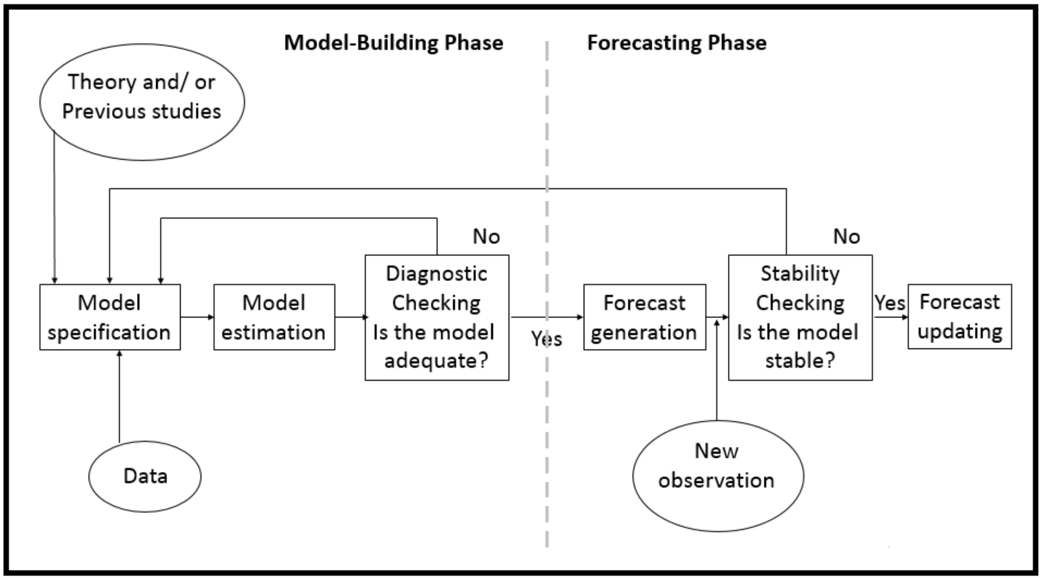

Figure 1: Conceptual framework of a forecasting system. Taken from [11]

following criteria must be keep:

- The degree of accuracy required.
- The desired forecast horizon.
- The maximum tolerable cost for forecasts.
- The degree of complexity required.
- Data availability.

Moreover, the proposed model generally contains unknown parameters to be estimated in the next step using conventional estimation methods. Finally, it is needed to inspect if the model is appropriate. This should be done to avoid inadequate variables in the model and have an incorrect specification of the functional relationship. If the model is not satisfactory, it must be specified again, and the iterative cycle model specification-estimation-forecasting should be repeated until a satisfactory model is found. This is where the model selection criteria play an important role (Table 1).

Phase 2: Forecasting Phase

At this stage the final model is used for forecasting. The model structure and parameters must remain constant during the forecast period. The stability of the forecasting model can be assessed by checking against new observations. At this point, the forecast error is calculated to detect changes in the model.

Compared to the linear case, the nonlinear time series have been little explored and theory is not sufficient to uncover nonlinearities [12]. One of the most critical issues is to select the appropriate forecasting nonlinear model [2]. The statistical tool used for the evaluation of the accuracy of a selected models are the models selection criteria, which allow given a set of rival models, select the "best" among them [13].

In this research project we propose a weighted criterion for selecting models of nonlinear time series, using statistical techniques that consider the inherent characteristics of the series to determine the weights.

This paper is organized in the following way. Important aspects and some revision of literature is presented in Section 2. The methodology for the obtainment of the data and the use of the statistical techniques are described in Section 3. In Section 4, are presented the results and they are finally discussed in Section 5. Future work is discussed in Section 6.

2. Important aspects

2.1. Model selection criteria

Much of modern scientific enterprise is concerned with the question of model choice, and so it should not come as a surprise that many approaches have been proposed over the years for dealing with this issue [14]. Model selection criteria take account of the goodness-of-fit of a model and the number of parameters used to achieve that fit [15].

While there are in the literature several decision criteria, most based on the distance between the actual values and their respective predicted values, they are based on assumptions that sometimes are not satisfied by the data under study, and also have some disadvantages that are not considered in practice (see Table 1).

Recently, it has been shown that a good choice to take advantage of the different criteria is through a weighted average of them. However the proposed ways to calculate the weights have not been entirely successful [8, 16] and the criteria are applied without validating the assumptions required, leading to bad decisions. In addition, in the literature there are no guidelines on which criteria to use, bearing in mind the inherent behavior of the time series.

2.2. Principal component analysis

Principal Components Analysis (PCA) is a procedure for identifying a smaller number of uncorrelated variables, called "principal components", from a large set of data, reducing the number of variables and avoiding multicollinearity. The goal of principal components analysis is to explain the maximum amount of variance with the fewest number of principal components, which are linear combinations of the original variables [17]. Principal components analysis is commonly used in the social sciences, market research, and other industries that use large data sets [18].

2.3. Analytic hierarchy process

The Analytic Hierarchy Process (AHP) is a theory of measurement through pairwise comparisons and relies on the judgements of experts to derive priority scales. The comparisons are made using a scale of absolute judgements that represents, how much more, one element dominates another with respect to a given attribute [19].

As said before, pairwise comparisons are the fundamental basis of AHP. The method uses a scale of 1 to 9 for assessing the relative preferences between two elements as showed in Table 2.

Also, as shown in Equation 2, the priority matrix A must satisfy:

- The diagonal is 1
- The lower triangular matrix is filled using $a_{ji} = 1/a_{ij}$

$$A = \begin{bmatrix} 1 & a_{12} & \cdots & a_{1n} \\ 1/a_{12} & 1 & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 1/a_{1n} & 1/a_{2n} & \cdots & 1 \end{bmatrix} \tag{2}$$

3. Methodology

3.1. Simulation study

In order to obtain experimental data used in this study, three time series models were considered:

- Nonlinear Autoregressive Model (NAR) [20]: $y_t = 0.7y_{t-1} - 0.017y_{t-1}^2 + \varepsilon_t$
- Linear Autoregressive Model (AR): $y_t = 0.67y_{t-1} - 0.41y_{t-2}\varepsilon_t$
- Generalized Autoregressive Conditional Heteroskedastic (GARCH) [21]: $y_t = \sqrt{h_t \varepsilon_t}$, with $h_t^2 = 0.00002281 + 0.0593y_{t-1}^2 + 0.901h_{t-1}^2$

The parameters of each model were modified 14 times[1], so that different distance measurements between the estimate and the parameter were considered, as well as various measures to the selection criteria. Figures 2, 3 and 4 show the original trajectory of the time series and some of the modifications.

After evaluating each of the estimated variations with the model selection criteria presented in Table 1, a database with i rows (i models or variations) and eight columns (each corresponding to a criterion) is obtained. Then the values are standardized for avoid the problem of the effect of the different units of measure. To standardize the following criteria were used:

[1]For each model the first variations are characterized by being closer estimations of the parameters.

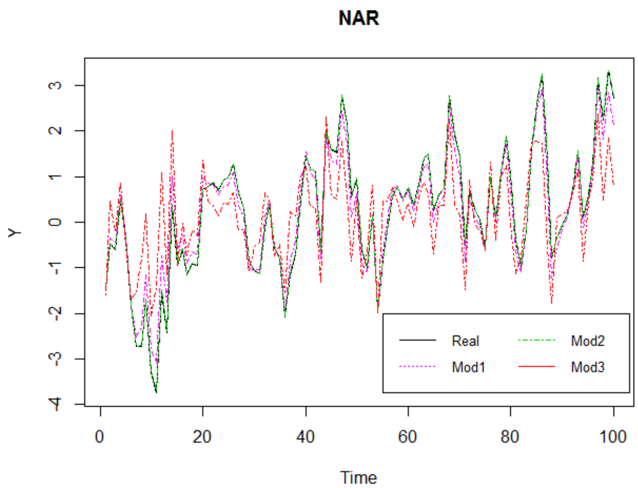

Figure 2: Variations of the model $y_t = 0.7y_{t-1} - 0.017y_{t-1}^2 + \varepsilon_t$

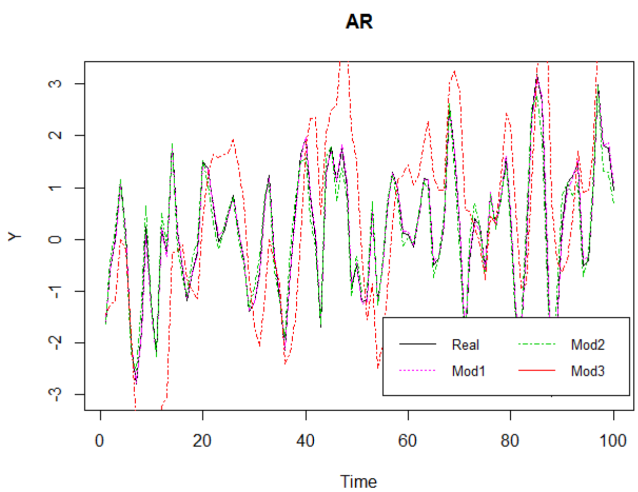

Figure 3: Variations of the model $y_t = 0.67y_{t-1} - 0.41 * y_{t-2}\varepsilon_t$

$$C_i^* = \frac{C_i - \min(C)}{\max(C) - \min(C)} \tag{3}$$

where C is any of the eight criteria shown in Table 1, C_i is the criterion

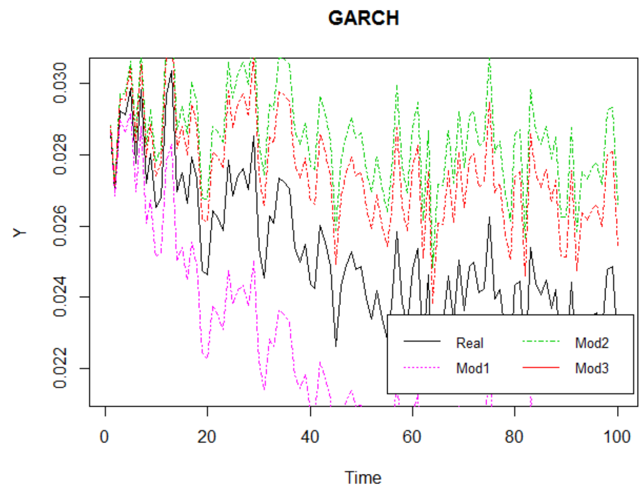

Figure 4: Variations of the model $y_t = \sqrt{h_t \varepsilon_t}$ with $h_t^2 = 0.00002281 + 0.0593 y_{t-1}^2 + 0.901 h_{t-1}^2$

value in the ith model, $\max(C)$ and $\min(C)$ denote the maximum and the minimum value obtained for the criterion.

3.2. Model selection criterion using PCA

The method used to estimate the optimal weights of the proposed criterion for selecting time series models under the PCA is as follows:

1. Apply a PCA to the database obtained in Section 3.1.
2. Define the number of partnerships between selection criteria using the Biplot chart.
3. Apply the PCA to each found association and determine the weights associated with each selection criterion.
4. Calculate the ratio between the index and its ideal value. This applies only in cases where the ideal value is nonzero.

3.3. Model selection criterion using AHP

According to (2) and Table 2, was created a matrix of importance for the model selection criteria according to the frequency of use of each method in the literature to obtain the weighted model:

$$
\begin{array}{c}
\begin{array}{ccccccccc}
SSE & RMSE & MAPE & MAE & ME & DA & MDA & Sign & AIC
\end{array} \\
\begin{pmatrix}
1 & 7 & 5 & 3 & 7 & 5 & 9 & 10 & 2 \\
1/7 & 1 & 1/3 & 1/5 & 1/2 & 5 & 9 & 7 & 1/9 \\
1/5 & 3 & 1 & 1/3 & 3 & 5 & 7 & 9 & 1/5 \\
1/3 & 5 & 3 & 1 & 2 & 2 & 3 & 3 & 1/3 \\
1/7 & 2 & 1/3 & 1/2 & 1 & 1/2 & 3 & 5 & 1/10 \\
1/5 & 1/5 & 1/5 & 1/2 & 2 & 1 & 5 & 1/3 & 1/9 \\
1/9 & 1/9 & 1/7 & 1/3 & 1/3 & 1/5 & 1 & 1/2 & 1/10 \\
1/10 & 1/7 & 1/9 & 1/3 & 1/5 & 3 & 2 & 1 & 1/7 \\
1/2 & 9 & 5 & 3 & 10 & 9 & 10 & 7 & 1
\end{pmatrix}
\begin{array}{c}
SSE \\ RMSE \\ MAPE \\ MAE \\ ME \\ DA \\ MDA \\ Sign \\ AIC
\end{array}
\end{array}
$$

4. Results

Before applying the statistical methods to the obtained data, a correlation analysis was performed between the selection criteria and it was shown that the AIC, BIC, AICC are extremely correlated, therefore we just take in account only the AIC criterion.

4.1. Model selection criterion using PCA

Initially all variations of the three models were considered in the same database and the PCA was applied. In the graph Biplot groups defined by the types of AR, NAR and GARCH were obtained (see Figure 5).

Based on this, was decided to apply the methodology proposed in Section 3.2 to each model under study. The results obtained were as follows.

4.1.1. NAR Model

For this nonlinear model it was found that MDA and ME criteria are not significant, because they have a very low weight in the components. In addition, two groups of associations were identified (see Figures 6 and 7).

Group 1: SSE, RMSE, MAPE, MAE and AIC

Group 2: DA and Sign.

which led to the following weighted selection criteria:

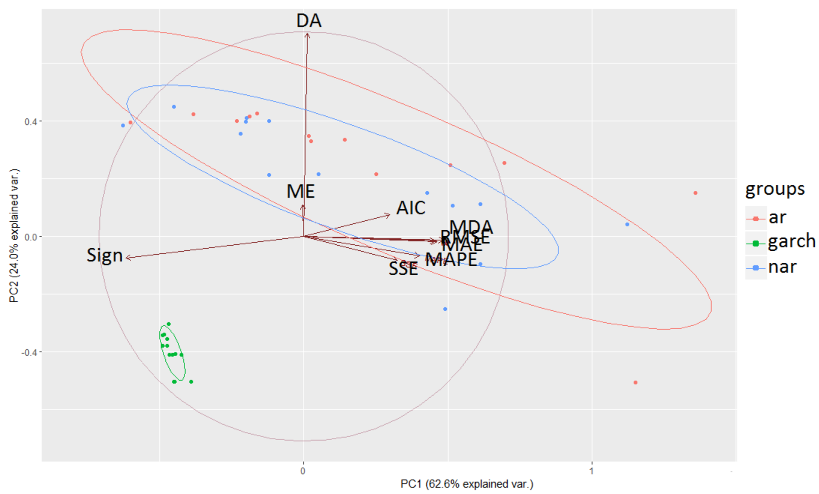

Figure 5: PCA general

	PC1	PC2	PC3	PC4
SSE	0.36842267	-0.02299887	-0.50664876	-0.20932938
RMSE	0.38428547	-0.04702933	-0.08712538	-0.05286600
MAPE	0.36838790	-0.27896182	-0.06205923	0.07378212
MAE	0.38894523	-0.10435533	-0.09531094	0.04451500
ME	-0.02518079	-0.70598204	-0.11918015	-0.45730021
DA	-0.36715807	0.24287362	-0.15669143	-0.63806274
MDA	0.30081839	0.50921407	-0.12944001	-0.40502644
Sign	-0.35676085	-0.29630947	-0.09606673	-0.10552733
AIC	0.27639149	-0.06366367	0.81119728	-0.39310435

Figure 6: PCA rotations of NAR model

$$
\begin{aligned}
PC_1 &= 0.4744SSE + 0.4765RMSE + 0.5084MAPE + 0.4897MAE + 0.2225AIC \\
PC_2 &= 0.5201DA + 0.4799Sign
\end{aligned}
\tag{4}
$$

Note that the first criterion PC_1 consists of measures associated with the error squared and PC_2 by criteria of direction of models.

4.1.2. AR Model

For this linear model it was found that MDA criteria, opposite to the result obtained with the NAR model, is significant, but ME criterion also have a very low weight in the components. In addition, two groups of associations

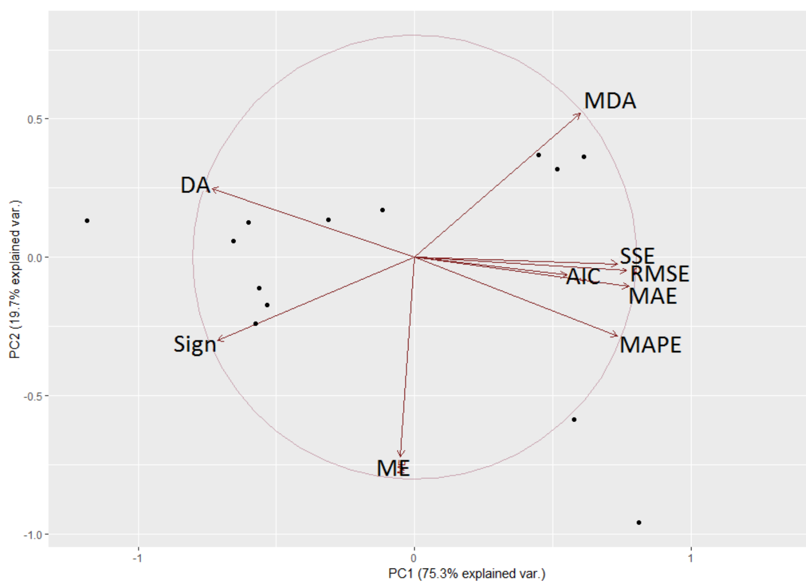

Figure 7: PCA NAR

were identified (see Figures 8 and 9).

Group 1: SSE, RMSE, MAPE, MAE, MDA and AIC

Group 2: DA and Sign.

	PC1	PC2	PC3	PC4
SSE	0.3266987	-0.04138381	-0.426870813	0.4240521
RMSE	0.3596530	-0.08178399	-0.169611387	0.1333069
MAPE	0.3885879	0.10113556	-0.028554145	0.0604021
MAE	0.3612480	-0.07924623	-0.161126110	0.1207773
ME	-0.1144983	-0.81184820	-0.293831555	0.1252542
DA	-0.3515510	-0.38948746	0.003745842	-0.1785826
MDA	0.3894182	-0.08620407	0.039724455	-0.4088894
Sign	-0.4044117	0.28835932	-0.205385011	0.5677995
AIC	0.1747105	-0.27149617	0.795062434	0.4969974

Figure 8: PCA rotations of AR model

$$PC_1 = 0.3997SSE + 0.4323RMSE + 0.4607MAPE + 0.4338MAE + 0.4587MDA + 0.2062AIC$$
$$PC_2 = 0.4584DA + 0.5416Sign \tag{5}$$

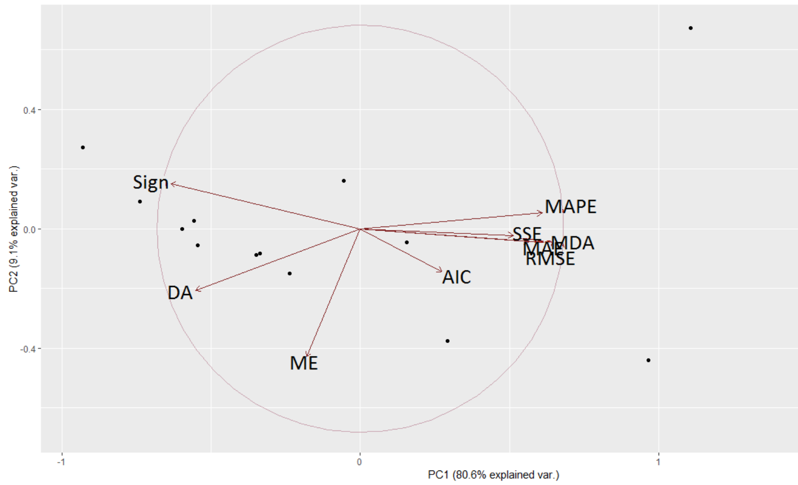

Figure 9: PCA AR

For the AR model is significant the distance measure MDA, and is associated with the selection criteria SSE, MAPE, RMSE and AIC. However the second indicator, as in the case of NAR, is also formed by the DA and the Sign criterion; of course with different weights to the NAR nonlinear model.

4.1.3. GARCH Model

For this nonlinear model it was found the same behavior as the NAR nonlinear model. And two groups of associations were identified (see Figures 10 and 11).

Group 1: SSE, RMSE, MAPE, MAE and AIC

Group 2: DA and Sign.

$$
\begin{aligned}
PC_1 &= 0.4055SSE + 0.4498RMSE + 0.4511MAPE + 0.4544MAE + 0.4726AIC \\
PC_2 &= 0.5912DA + 0.4088Sign
\end{aligned} \tag{6}
$$

4.2. Model selection criterion using AHP

Based on the matrix established in Section 3.2, the AHP methodology was applied. With this measure is possible to set a weight for each selection criteria considered in this paper. The weighted selection criterion obtained

	PC1	PC2	PC3	PC4
SSE	0.3423886	-0.165238288	-0.215855949	0.31464542
RMSE	0.3886483	-0.001139202	0.053896595	0.35381221
MAPE	0.3672652	-0.121318665	0.004204236	0.09831616
MAE	0.3859716	-0.027341777	0.055146954	0.29073122
ME	-0.1901388	0.591439835	0.035506793	0.58326632
DA	-0.3774784	-0.038918146	-0.465663113	0.43263077
MDA	0.2373385	0.649418871	-0.530652303	-0.38821312
Sign	-0.2631608	0.269311291	0.565158793	0.04038546
AIC	0.3789972	0.334125842	0.358250344	-0.04376153

Figure 10: PCA rotations of GARCH model

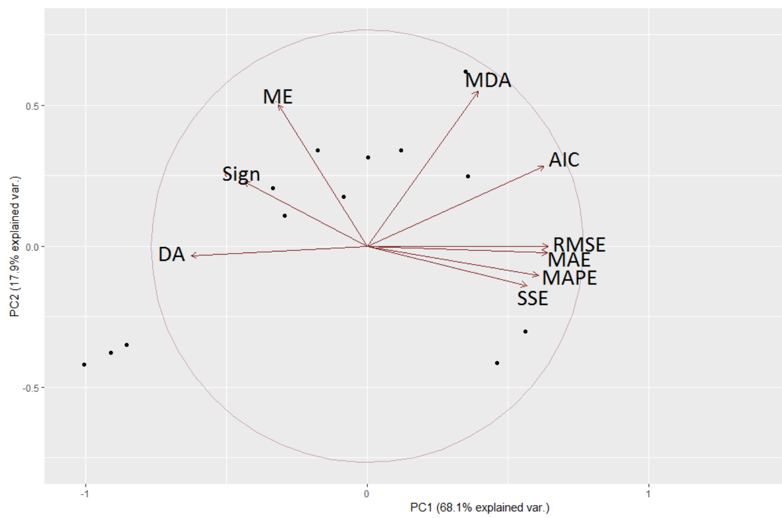

Figure 11: PCA GARCH

is as follows:

$$C_{AHP} = 0.3098SSE + 0.0573RMSE + 0.1046MAPE + 0.1101MAE \\ + 0.0483ME + 0.0339DA + 0.0162MDA + 0.0248Sign + 0.2950AIC \quad (7)$$

In this case the weights of the criteria are directly proportional to their frequency of use.

4.2.1. Calculation of the weighted selection criteria

Note that in general the selection criteria proposed using PCA have the following decision rule:

Group 1: If $PC_1 \longrightarrow 0$ then the estimates obtained are good.

Group 2: If $PC_2 \longrightarrow 1$ then the estimates obtained are good.

Also, for the AHP rule it is as follows: If $C_{AHP} \longrightarrow 0$ then the estimates obtained are good.

Moreover, using standardized database and the expressions obtained in Sections 4.1 and 4.2, Table 3 is formed.

5. Discussion and Conclusion

Were proposed two methodologies for selection criteria of linear models and nonlinear time series using PCA and AHP techniques. This allows us to state the following:

- It is important to highlight that AIC, BIC and AICC criteria are extremely correlated, so when using more than one of them a multicollinearity problem is committed. For this reason it is not necessary to use more than one of them in the weighted methods. This leads to examine in the literature some proposals for development of new weighted selection criteria that do not consider this fact (see for example [8, 16]).

- When analyzing the associations made by groups after performing principal component analysis by type of model, you get that nonlinear models GARCH and NAR show the same configurations in the groups although the weights of each criterion vary between them.

- While analyzing the structure of the obtained groups, it can be noted that the criteria belonging to the first group are related to measurement of the error between the actual and the estimated model and those belonging to the second group are related to the analysis of direction of the models.

- While the AHP technique is easy to use, the PCA technique yields better results in terms of the interpretation of results and group formation.

6. Future work

Additional work will be required to generalize this groups according to the linearity of the studied data. It must be proved with different time series models.

7. References

[1] D. C. Montgomery, C. L. Jennings, M. Kulahci, Introduction to time series analysis and forecasting, John Wiley & Sons, 2015.

[2] M. Qi, G. P. Zhang, An investigation of model selection criteria for neural network time series forecasting, European Journal of Operational Research 132 (2001) 666–680.

[3] A. R. Camina, J. Janacek, Mathematics for seismic data processing and interpretation, Springer Science & Business Media, 2012.

[4] M. Khashei, S. Reza Hejazi, M. Bijari, A new hybrid artificial neural networks and fuzzy regression model for time series forecasting, Fuzzy Sets and Systems 159 (2008) 769–786.

[5] M. Cogollo, J. D. Velasquez, Methodological advances in artificial neural networks for time series forecasting, IEEE Latin America Transactions 12 (2014) 764–771.

[6] R. Tsay, Analysis of Financial Time Series, John Wiley & Sons, 2010.

[7] R. E. Cumby, D. M. Modest, Testing for market timing ability: A framework for forecast evaluation, Journal of Financial Economics 19 (1987) 169–189.

[8] E. Egrioglu, C. H. Aladag, S. Gunay, A new model selection strategy in artificial neural networks, Applied Mathematics and Computation 195 (2008) 591–597.

[9] W. K. Li, Diagnostic checks in time series, CRC Press, 2003.

[10] S. Y. Hwang, I. Basawa, J. Reeves, The asymptotic distributions of residual autocorrelations and related tests of fit for a class of nonlinear time series models, Statistica Sinica 4 (1994) 107–125.

[11] B. Abraham, J. Ledolter, Statistical methods for forecasting, volume 234, John Wiley & Sons, 2009.

[12] U. Anders, O. Korn, Model selection in neural networks, Neural Networks 12 (1999) 309–323.

[13] G. Kapetanios, Model selection in threshold models, Journal of Time Series Analysis 22 (2001) 733–754.

[14] J. B. Kadane, N. A. Lazar, Methods and criteria for model selection, Journal of the American statistical Association 99 (2004) 279–290.

[15] S. L. Sclove, Application of model-selection criteria to some problems in multivariate analysis, Psychometrika 52 (1987) 333–343.

[16] C. H. Aladag, E. Egrioglu, S. Gunay, M. A. Basaran, Improving weighted information criterion by using optimization, Journal of Computational and Applied Mathematics 233 (2010) 2683–2687.

[17] I. Jolliffe, Principal Component Analysis, Wiley Online Library, 2002.

[18] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, Chemometrics and intelligent laboratory systems 2 (1987) 37–52.

[19] T. L. Saaty, Decision making with the analytic hierarchy process, International journal of services sciences 1 (2008) 83–98.

[20] S. Aras, İ. D. Kocakoç, A new model selection strategy in time series forecasting with artificial neural networks: IHTS, Neurocomputing 174 (2016) 974–987.

[21] A. P. Ennio, O. Pablo, A Switching Threshold Model for Oil Prices, Systems Engineering Procedia 1 (2011) 490–498.

Model selection criterion	Definition
SSE	$\sum_{i=1}^{T}(y_i - \hat{y}_i)^2$
RMSE	$\sqrt{\frac{1}{T}SSE}$
AIC	$\log\left(\frac{SSE}{T}\right) + \frac{2m}{T}$
AICC	$\log\left(\frac{SSE}{T}\right) + \frac{2m}{T - m - 1}$
BIC	$\log\left(\frac{SSE}{T}\right) + \frac{m\log(T)}{T}$
MAPE	$\frac{1}{T}\sum_{i=1}^{T}\left\|\frac{(y_i - \hat{y}_i)}{y_i}\right\|$
MAE	$\frac{1}{T}\sum_{i=1}^{T}\|(y_i - \hat{y}_i)\|$
ME	$\frac{1}{T}\sum_{i=1}^{T}(y_i - \hat{y}_i)$
DA	$\frac{1}{T}\sum_{i=1}^{T}a_i$, where $a_1 = \begin{cases} 1 & \text{if } (y_{i+1} - y_i)(\hat{y}_{i+1} - y_i) > 0 \\ 0 & \text{ioc} \end{cases}$
MDA	$\frac{\sum_{i=1}^{T-1}D_i}{T - 1}$, where $D_i = (A_i - F_i)^2$
Sign	$\frac{1}{T}\sum_{i=1}^{T}z_i$, where $z_1 = \begin{cases} 1 & \text{if } (y_{i+1})(\hat{y}_{i+1}) > 0 \\ 0 & \text{ioc} \end{cases}$

Table 1: Model selection criteria, where m is the number of parameters and T the number of observations.

Intensity of Importance	Definition
1	Equal importance
3	Moderate importance
5	Strong importance
7	Very strong or demonstrated importance
9	Extreme importance

Table 2: Scale of importance for the AHP

	PCA		AHP
	Group 1	Group 2	
NAR	0	1	0.0838
	0.4332	0.8263	0.3867
	0.1690	1	0.2841
	0.4518	0.8416	0.4012
	0.6614	0.7976	0.4777
	0.6232	0.7212	0.4237
	0.4447	0.8435	0.3959
	2.0119	0.2172	0.8666
	2.1664	0.1224	0.9386
	1.7677	0.4086	0.4800
	1.2595	0.2882	0.6269
	1.0868	0.4008	0.5701
	1.8446	0.2533	0.7558
	1.1722	0.3282	0.5979
AR	0	1	0.0967
	0.1961	0.9672	0.3174
	0.3343	0.8503	0.3694
	0.3024	0.8880	0.3586
	0.5190	0.7761	0.4133
	0.3337	0.8812	0.3609
	0.5110	0.7847	0.4113
	1.5483	0.1447	0.5933
	0.6844	0.6400	0.4220
	0.5455	0.6746	0.4024
	0.9595	0.4147	0.4891
	1.6876	0.3287	0.7287
	2.3914	0.0155	0.9265
	0.8950	0.4712	0.4518
GARCH	0.1895	0.7850	0.1543
	0.2913	0.8388	0.2660
	0	0.8925	0.0778
	0.1150	0.8925	0.1668
	0.9564	0.6775	0.5617
	0.0261	0.7850	0.0966
	0.1641	1	0.2082
	0.3009	0.6775	0.2630
	0.5420	0.6775	0.3810
	0.7811	0.4088	0.4400
	0.8511	0.4088	0.4687
	2.9490	0	0.8930
	0.9731	0.6775	0.5839
	0.9920	0.6775	0.5884

Table 3: Values obtained with the methodologies