Adaptation of model selecting criteria for nonlinear time series forecasting

Research practise 2: Project presentation

Andrea Molina-Alonso[1] Myladis Rocio Cogollo-Flórez[2]

[1]Responsible
amolin13@eafit.edu.co

[2]Tutor
mcogollo@eafit.edu.co

EAFIT University
Mathematical Engineering
Department of Mathematical Sciences

June 10, 2016

Modeling methodology of time series for forecasting

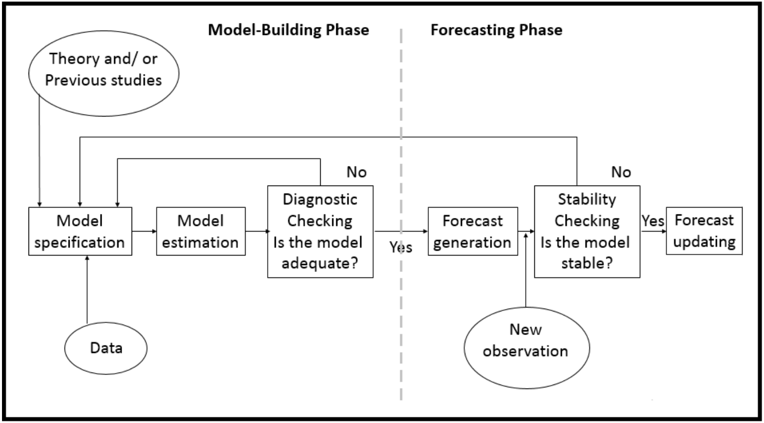


Figure: Conceptual framework of a forecasting system. Taken from [Abraham and Ledolter, 2009]

Types models of time series

A model of pth-order of time series is defined as [Li, 2003, Hwang et al., 1994]

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

where

- f is a known linear or nonlinear function of past X_t's.
- ϕ is a $p \times 1$ vector of parameters.
- The noise process $\{a_t\}$ is assumed to be independent, with mean zero, variance σ_a^2, and finite fourth order moment.

- Compared to the linear case, the nonlinear time series have been little explored and theory is not sufficient to uncover nonlinearities [Anders and Korn, 1999].
- One of the most critical issues is to select the appropriate forecasting nonlineal model [Qi and Zhang, 2001].

Model selection criteria

Model selection criterion	Definition
SSE	$\sum_{i=1}^{T}(y_i - \hat{y}_i)^2$
AIC	$\log\left(\frac{SSE}{T}\right) + \frac{2m}{T}$
AICC	$\log\left(\frac{SSE}{T}\right) + \frac{2m}{T-m-1}$
BIC	$\log\left(\frac{SSE}{T}\right) + \frac{m\log(T)}{T}$

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

Model selection criteria

Model selection criterion	Definition
RMSE	$\sqrt{\frac{1}{T} SSE}$
MAPE	$\frac{1}{T}\sum_{i=1}^{T}\left\lvert\frac{(y_i - \hat{y}_i)}{y_i}\right\rvert$
MAE	$\frac{1}{T}\sum_{i=1}^{T}\lvert(y_i - \hat{y}_i)\rvert$
ME	$\frac{1}{T}\sum_{i=1}^{T}(y_i - \hat{y}_i)$

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

Model selection criteria

Model selection criterion	Definition
DA	$\frac{1}{T}\sum_{i=1}^{T} a_i$, where $a_1 = \begin{cases} 1 & if\,(y_{i+1}-y_i)(\hat{y}_{i+1}-y_i) > 0 \\ 0 & otherwise \end{cases}$
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 $z_1 = \begin{cases} 1 & if\,(y_{i+1})(\hat{y}_{i+1}) > 0 \\ 0 & otherwise \end{cases}$

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

Weighted selection criterion

A weighted selection criteria using optimization was proposed by [Aladag et al., 2010]:

$$AWIC = w_1 RMSE + w_2 MAPE + w_3(1 - DA) + w_4 MDA + 0.1AIC + 0.1BIC \quad (2)$$

- It is not shown a criteria for determining the weights of AIC and BIC.
- There are no guidelines to know which criteria to use, bearing in mind the inherent behavior of the time series.
- Heuristic methods have not been successful in the estimation of weights for the combined methods.
- This method does not consider the time series characteristics.

Objectives I

General Objective

Formulate a criterion for selecting models of nonlinear time series using multivariate analysis techniques and the inherent characteristics of the series.

Objectives II

Specific Objectives

- Identify the different selection criteria formulated in the literature for non-linear time series.
- Determine the multivariate analysis techniques that allow the creation of synthetic indicators according to the characteristics of the data.
- Establish a methodological framework that considers the characteristics of the data and consider the advantages of the proposed selection criteria in literature to date.
- Validate the feasibility of the proposed methodology by experimental data.

Selected models I

Nonlinear Autoregressive Model (NAR) [Aras and Kocakoç, 2016]:

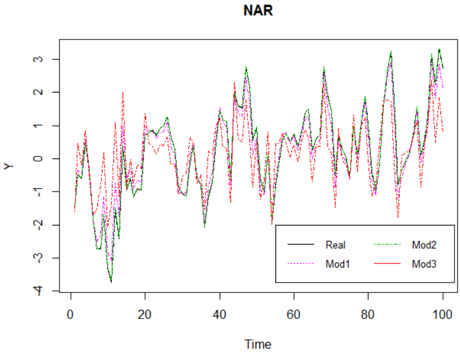


Figure: Variations of the model $y_t = 0.7y_{t-1} - 0.017y_{t-1}^2 + \varepsilon_t$

Selected models II

Generalized Autoregressive Conditional Heteroskedastic (GARCH) [Ennio and Pablo, 2011]:

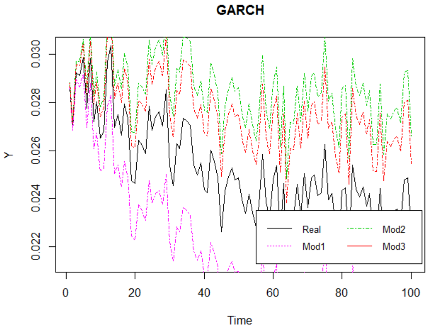


Figure: Variations of the model $y_t = \sqrt{h_t \varepsilon_t}$ with $h_t^2 = 0.00002281 + 0.0593 y_{t-1}^2 + 0.901 h_{t-1}^2$

Selected models III

Autoregressive Model (AR):

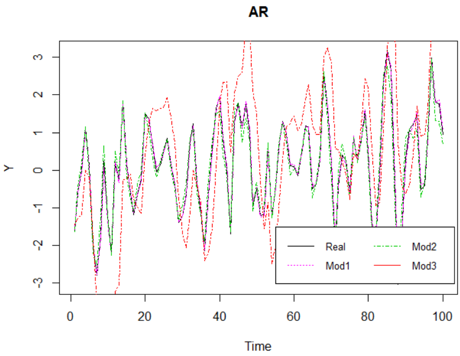
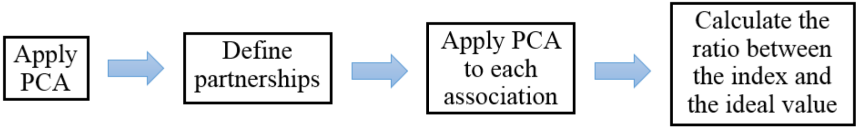


Figure: Variations of the model $y_t = 0.67 y_{t-1} - 0.41 * y_{t-2} \varepsilon_t$

Methodology model selection criterion using PCA

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



Methodology model selection criterion using AHP

According to the frequency of use of each method in the literature was created a matrix of importance for the model selection criteria to obtain the weighted model with the Analytic Hierarchy Process (AHP):

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

Results I

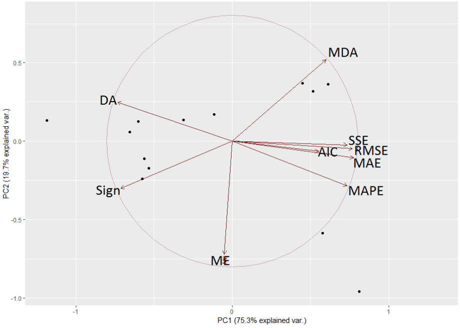
Before applying the statistical methods to the obtained data, a correlation analysis was performed between the selection criteria.

	SSE	RMSE	MAPE	MAE	ME	DA	MDA	Sign	AIC	BIC	AICC
SSE	1.00	0.942	0.46	0.939	-0.10	-0.104	0.69	-0.85	0.574	0.574	0.574
RMSE	0.94	1.000	0.59	0.999	-0.23	-0.076	0.68	-0.89	0.766	0.766	0.766
MAPE	0.46	0.585	1.00	0.576	-0.82	-0.432	0.62	-0.19	0.498	0.498	0.498
MAE	0.94	0.999	0.58	1.000	-0.23	-0.087	0.67	-0.89	0.766	0.766	0.766
ME	-0.10	-0.233	-0.82	-0.234	1.00	0.354	-0.26	-0.13	-0.220	-0.220	-0.220
DA	-0.10	-0.076	-0.43	-0.087	0.35	1.000	-0.49	-0.15	0.067	0.067	0.067
MDA	0.69	0.684	0.62	0.672	-0.26	-0.491	1.00	-0.49	0.506	0.506	0.506
Sign	-0.85	-0.889	-0.19	-0.891	-0.13	-0.145	-0.49	1.00	-0.685	-0.685	-0.685
AIC	0.57	0.766	0.50	0.766	-0.22	0.067	0.51	-0.69	1.000	1.000	1.000
BIC	0.57	0.766	0.50	0.766	-0.22	0.067	0.51	-0.69	1.000	1.000	1.000
AICC	0.57	0.766	0.50	0.766	-0.22	0.067	0.51	-0.69	1.000	1.000	1.000

Results II

For the nonlinear NAR 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:

- Group 1: SSE, RMSE, MAPE, MAE and AIC
- Group 2: DA and Sign.



	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

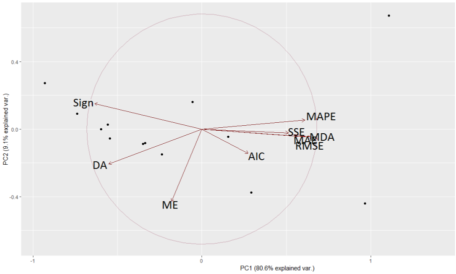
Figure: PCA rotations of NAR model

Results III

For the AR 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 were identified:

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

Group 2: DA and Sign.



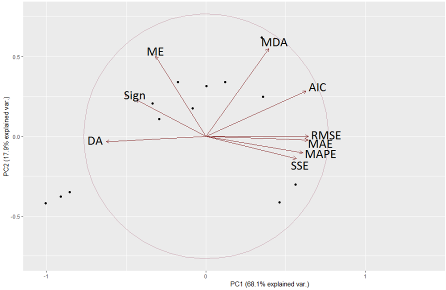
	PC1	PC2	PC3	PC4
SSE	0.3266987	-0.04138381	-0.426870813	0.4240521
RMSE	0.3596530	-0.08178399	-0.169611387	0.1333069
MAPE	0.3885879	0.10113556	-0.028554145	0.0604021
MAE	0.3612480	-0.07924623	-0.161126110	0.1207773
ME	-0.1144983	-0.81184820	-0.293831555	0.1252542
DA	-0.3515510	-0.38948746	0.003745842	-0.1785826
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: PCA rotations of AR model

Results IV

For the GARCH model it was found the same behavior as the NAR nonlinear model. And two groups of associations were identified:

- Group 1: SSE, RMSE, MAPE, MAE and AIC
- Group 2: DA and Sign.



	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: PCA rotations of GARCH model

The obtained weighted selection criteria obtained with the three models were:

NAR

$$\begin{aligned} PC_1 &= 0.4744SSE + 0.4765RMSE + 0.5084MAPE + 0.4897MAE + 0.2225AIC \\ PC_2 &= 0.5201DA + 0.4799Sign \end{aligned} \tag{3}$$

AR

$$\begin{aligned} PC_1 &= 0.3997SSE + 0.4323RMSE + 0.4607MAPE + 0.4338MAE + 0.4587MDA \\ &\quad + 0.2062AIC \\ PC_2 &= 0.4584DA + 0.5416Sign \end{aligned} \tag{4}$$

GARCH

$$\begin{aligned} PC_1 &= 0.4055SSE + 0.4498RMSE + 0.4511MAPE + 0.4544MAE + 0.4726AIC \\ PC_2 &= 0.5912DA + 0.4088Sign \end{aligned} \tag{5}$$

Weighted selection criterion with AHP

Based on the matrix established before, 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 is as follows:

$$\begin{aligned} C_{AHP} &= 0.3098SSE + 0.0573RMSE + 0.1046MAPE + 0.1101MAE \\ &+ 0.0483ME + 0.0339DA + 0.0162MDA + 0.0248Sign \\ &+ 0.2950AIC \end{aligned} \tag{6}$$

	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

Table: Values obtained with the methodologies

	PCA		AHP
	Group 1	Group 2	
	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
AR	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

Table: Values obtained with the methodologies

	PCA		AHP
	Group 1	Group 2	
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: Values obtained with the methodologies

Conclusions

- 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.

Conclusions

- 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.
- 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.

Conclusions

- 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.
- 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.
- 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.

Conclusions

- 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.
- 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.
- 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.

References

Abraham, B. and Ledolter, J. (2009).
Statistical methods for forecasting, volume 234.
John Wiley & Sons.

Aladag, C. H., Egrioglu, E., Gunay, S., and Basaran, M. A. (2010).
Improving weighted information criterion by using optimization.
Journal of computational and applied mathematics, 233(10):2683–2687.

Anders, U. and Korn, O. (1999).
Model selection in neural networks.
Neural Networks, 12(2):309–323.

References

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

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

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

References

Li, W. K. (2003).
Diagnostic checks in time series.
CRC Press.

Qi, M. and Zhang, G. P. (2001).
An investigation of model selection criteria for neural network time series forecasting.
European Journal of Operational Research, 132(3):666–680.

Thanks for your attention!!

Adaptation of model selecting criteria for nonlinear time series forecasting.