Independent Journal of Management & Production (IJM&P) ISSN: 2236-269X
DOI: 10.14807/ijmp.v3i2.43 v. 3, n. 2, July – December 2012

BOX & JENKINS MODEL IDENTIFICATION: A COMPARISON OF METHODOLOGIES

Maria Augusta Soares Machado
IBMEC/RJ - Brazil
E-mail: mmachado@ibmecrj.br

Reinaldo Castro Souza
Pontifícia Universidade Católica do Rio de Janeiro (PUC/RJ) - Brazil
E-mail: reinaldo@ele.puc-rio.br

Ricardo Tanscheit
Pontifícia Universidade Católica do Rio de Janeiro (PUC/RJ) - Brazil
E-mail: ricardo@ele.puc-rio.br

Submission: 19/10/2012
Accept: 10/11/2012

ABSTRACT

This paper focuses on a presentation of a comparison of a neuro-fuzzy back propagation network and Forecast automatic model Identification to identify automatically Box & Jenkins non seasonal models.

Recently some combinations of neural networks and fuzzy logic technologies have being used to deal with uncertain and subjective problems. It is concluded on the basis of the obtained results that this type of approach is very powerful to be used.

Key-words: Neuro-Fuzzy Networks, Box & Jenkins Methodology, Fuzzy Logic

1 Introduction

Artificial neural network applications have shown that this technology has significant capabilities in pattern recognition. The abilities of feed forward back propagation artificial neural networks used together with fuzzy modeling that try to extract the model directly from the experts knowledge, seem to offer a good approach to the problems inherent in the Box & Jenkins ARIMA model identification.

The literature in time series forecasting clearly indicates the properly applied the Box & Jenkins approach to time series forecasting yields forecasts that are superior to those resulting from other standard time series forecasting procedures. As a result, the method has received much attention however, the literature also indicates some reluctance to use this method in practice, due to the difficulties associated with model identification Vandaele(1983) states," identification is the key to time series model building". The task of forecaster is to use basic model identification tools.

2 Application

The algorithm used to determine Box & Jenkins non-seasonal patterns was implemented in seven steps:

Step 1 - Generation of 400 random time series AR(1),MA(1),AR(2),MA(2) and ARMA(1,1) with 700 observations.

AR(1) model:

$z_t = \phi_1 z_{t-1} + a_t$ t=1,...,700;

where:: ϕ_1 = model parameter ; ϕ_1 ~ Uniform (-1,1) ; a_t ~ Normal (0,1)

MA(1) model:

$z_t = a_t - \theta_1 a_{t-1}$ t=1,...,700;

where:: θ_1 = model parameter ; θ_1 ~ Uniform (-1,1) ; a_t ~ Normal (0,1)

AR(2) model:

$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + a_t$ t=1,...,700;

where: ϕ_1 , ϕ_2 = model parameters; ϕ_1 , ϕ_2 ~ Uniform (-2,2) ; a_t ~ Normal (0,1)

MA(2) model:

$z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$ t=1,...700;

where: θ_1 , θ_2 = model parameters ; θ_1 , θ_2 ~ Uniform (-2,2) ; a_t ~ Normal (0,1)

ARMA(`1,1) model:

$z_t = \phi_1 z_{t-1} + a_t - \theta_1 a_{t-1}$ t=1,...,700 ;

where ϕ_1 , θ_2 = model parameters ; ϕ_1 , θ_2 ~ Uniform (-2,2); a_t ~ Normal (0,1)

Step 2 - It was estimated ACF and PACF using the first 10 lags, for each model, which are the neuro-fuzzy inputs. For estimated ACF (model " j " ,j=1,...,400):

$\hat{\rho}_1^{(j)}$, $\hat{\rho}_2^{(j)}$, $\hat{\rho}_3^{(j)}$, $\hat{\rho}_4^{(j)}$, $\hat{\rho}_5^{(j)}$, $\hat{\rho}_6^{(j)}$, $\hat{\rho}_7^{(j)}$, $\hat{\rho}_8^{(j)}$, $\hat{\rho}_9^{(j)}$, $\hat{\rho}_{10}^{(j)}$, where:

$\hat{\rho}_1^{(j)}$ ACF's value of "j" model for lag 1; $\hat{\rho}_2^{(j)}$ ACF's value of "j" model for lag 2; . $\hat{\rho}_9^{(j)}$ ACF's value of "j" model for lag 9; $\hat{\rho}_{10}^{(j)}$ ACF's value of "j " model for lag 10;

For estimated ACF (model " j " ,j=1,...,400): $\hat{\phi}_{11}^{(j)}$, $\hat{\phi}_{22}^{(j)}$, $\hat{\phi}_{33}^{(j)}$, $\hat{\phi}_{44}^{(j)}$, $\hat{\phi}_{55}^{(j)}$,

$\hat{\phi}_{66}^{(j)}$, $\hat{\phi}_{77}^{(j)}$, $\hat{\phi}_{88}^{(j)}$, $\hat{\phi}_{99}^{(j)}$, $\hat{\phi}_{1010}^{(j)}$, where:

$\hat{\phi}_{11}^{(j)}$ PACF's value of "j " model for lag 1; $\hat{\phi}_{22}^{(j)}$ PACF's value of "j " model for lag 2;.

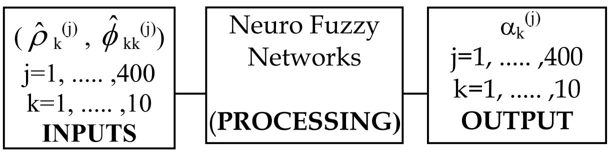
$\hat{\phi}_{99}^{(j)}$ PACF's value of "j " model for lag 9; $\hat{\phi}_{1010}^{(j)}$ PACF's value of "j " model for lag 10;

Step 3 – Determination of pairs.

$(\hat{\rho}_k^{(j)}, \hat{\phi}_{kk}^{(j)})$, j=1,....,400 ; k=1, ... ,10 as neural fuzzy networks inputs

Step 4 – Determination of neural fuzzy networks outputs.

The neural fuzzy networks "Black- Box" is shown next:



where:

$\alpha_1^{(j)}$ - neuro-fuzzy output of model "j" for lag 1; $\alpha_2^{(j)}$ - neuro-fuzzy output of model "j" for lag 2; ..$\alpha_9^{(j)}$ - neuro-fuzzy output of model "j" for lag 9; $\alpha_{10}^{(j)}$ - neuro-fuzzy output of model "j" for lag 10;

Step 5 Determination of a pattern for each structure. The pattern of each structure is:

$\bar{\alpha}_1$, $\bar{\alpha}_2$, $\bar{\alpha}_3$, $\bar{\alpha}_4$, $\bar{\alpha}_5$, $\bar{\alpha}_6$, $\bar{\alpha}_7$, $\bar{\alpha}_8$, $\bar{\alpha}_9$, $\bar{\alpha}_{10}$, where:

$\bar{\alpha}_1$ mean of neuro-fuzzy network for lag 1; $\bar{\alpha}_2$ mean of neuro-fuzzy network for lag 2; ..$\bar{\alpha}_9$ mean of neuro-fuzzy network for lag 9; $\bar{\alpha}_{10}$ mean of neuro-fuzzy network for lag 10;

Step 6 - Determination of weighted Euclidean distances using exponential smoothing

for " lag " j

$$d^{structure}_{weighted\ Euclidean\ mean}\left(\beta(\beta-1)^{j-1}\left|\Sigma\left(\alpha-\alpha_i\right)\right|\right)$$

where:

β = 0.7 for AR(1);β = 0.5 ; for MA(1) ; β = 0.2 for AR(2) ; β = 0.4 for MA(2); β = 0.4 for ARMA(1,1)

These values where determined based on the results of a detailed analysis of networks outputs.

Step 7 – The minimum of weighted Euclidean distances is indicated as the best model to fit the time series being studied.

AR(1) pattern: [0.0191 0.1540 0.0397 0.1358 0.1194 0.1256 0.1220 0.1104 0.1141 0.1042]

MA(1) pattern: [0.4362 0.4443 0.4571 0.4303 0.4517 0.4458 0.4377 0.4492 0.4588 0.4440]

AR(2) pattern: [0.0353 0.0819 0.0749 0.0300 0.0270 0.0301 0.0260 0.0206 0.0256 0.0216]

MA(2) pattern: [0.2840 0.3114 0.3160 0.3157 0.3159 0.3042 0.3015 0.2877 0.3062 0.2947]

ARMA(1,1) pattern: [0.1196 0.3775 0.2944 0.3237 0.3394 0.3306 0.3148 0.3262 0.3243 0.3173]

Independent Journal of Management & Production (IJM&P) ISSN: 2236-269X
DOI: 10.14807/ijmp.v3i2.43 v. 3, n. 2, July – December 2012

3 Results

3.1 - Simulated random AR(1) models

The networks indications were:

Nº Observations	Correct Indication	Incorrect indication	
		AR (2)	ARMA (1,1)
50	92%	6%	2%
100	88%	6%	6%
200	94%	2%	4%
300	96%	2%	2%

Total percentage of right indication: 92,5 %

3.2 - Simulated random MA(1) models

The networks indications were:

Nº Observations	Correct Indication	Incorrect indication		
		MA (2)	AR (2)	ARMA (1,1)
50	56%	20%	12%	12%
100	48%	34%	12%	6%
200	48%	30%	12%	10%
300	58%	30%	6%	6%

Total percentage of right indication: 52,5 %

3.3 - Simulated random AR(2) models

The networks indications were:

Nº Observations	Correct indications	Incorrect indications	
		AR(1)	ARMA(1,1)
50	38%	62%	
100	14%	74%	12%
200	14%	80%	6%
300	16%	72%	12%

Total percentage of right indication: 20,5 %

3.4 - Simulated random MA(2) models

The networks indications were:

Nº Observations	Correct Indication	Incorrect indication		
		MA (2)	AR (2)	ARMA (1,1)
50	34%	48%	14%	4%
100	34%	52%	12%	2%
200	32%	44%	16%	8%
300	34%	54%	8%	4%

Total percentage of right indication: 33,5 %

Independent Journal of Management & Production (IJM&P) ISSN: 2236-269X
DOI: 10.14807/ijmp.v3i2.43 v. 3, n. 2, July – December 2012

3.5 – Simulated random ARMA(1,1) models

The networks indications were:

Nº Observations	Correct indications	Incorrect indications	
		MA(1)	AR(1)
50	22%	2%	76%
100	5%	3%	84%
200	18%	2%	80%
300	8%	2%	90%

Total percentage of right indication: 14,5 %

3.6 - Comparison of Neuro-Fuzzy Networks Identification and Forecast automatic model Identification

For simulated time series of 50 observations:

	Percentage of right indication	
	Neuro-Fuzzy Network	FORECAST-PRO
AR(1)	92	76
MA(1)	56	18
AR(2)	38	22
MA(2)	34	16
ARMA(1,1)	22	26

For simulated time series of 100 observations:

	Percentage of right indication	
	Neuro-Fuzzy Network	FORECAST-PRO
AR(1)	88	53
MA(1)	48	31
AR(2)	14	18
MA(2)	34	25
ARMA(1,1)	5	11

For simulated time series of 200 observations:

	Percentage of right indication	
	Neuro-Fuzzy Network	FORECAST-PRO
AR(1)	94	31
MA(1)	48	21
AR(2)	14	10
MA(2)	32	19
ARMA(1,1)	18	15

For simulated time series of 300 observations:

	Percentage of right indication	
	Neuro-Fuzzy Network	FORECAST-PRO
AR(1)	96	33
MA(1)	58	41
AR(2)	16	10
MA(2)	34	15
ARMA(1,1)	8	13

A total of 200 random simulated time series from each structure was used to validate the methodology presented in this paper. The total average percentage of right neuro-fuzzy networks indications were:

Structure	Total average percentage of right Identification
AR(1)	98
MA(1)	77
AR(2)	67
MA(2)	78.5
ARMA(1,1)	59

4 Conclusions

The neuro-fuzzy networks make good identification; when using them is recommended to consider their first indication as “over fitted “ . The second indication of their outputs must be considered as possible Box & Jenkins Model .

References

AZOFF, E. M. (1994) *Neural Network Time Series Forecasting of Financial Markets*. Chicester- John Wiley & Sons Ltd., Baffins Lane.

BRAGA, M.J.,BARRETO, J.M., MACHADO, M.A. (1995) *Conceitos da Matemática Nebulosa na Análise de Risco*, Artes & Rabiscus.

DHAR, V. STEIN, R. (1996) *Raising Organizational IQ: Strategies for* ----Knowledge Intensive Decision Support, Prentice Hall.

JANG, J.- S.R., Sun, C.T. (1997) Mizutani, E., *Neuro-Fuzzy and Soft Computing* - A Computational Approach to Learning and Machine Intelligence, Prentice Hall Inc., 1997.

LANGAR and ZADEH (Eds.). (1995) *Industrial Applications of Fuzzy,Logic and Intelligent Systems*, Piscata,Bay, NJ: IEEE Press.

REYNOLDS, B., STEVENS, MELLICHAMP, SMITH M.J. E. (1995) *Box-JenkinsForecast Model Identification*, A.I. Expert June 1995.

SCHWARTZ , The.Fuzzv Systems Come to Life in Japan,*IEEE Expert*, Vol. 5(1), p. 77-78.

SOUZA, C.R.., CAMARGO, M.E. (1996) *Análise e Previsão de Séries Temporais*: os Modelos ARIMA, Sedigraf.

TSOUKALAS , L. H., UHRIG , R. E. (1997) *Fuzzy and Neural Approaches in Engineering*, John Wiley & Sons INC.

VON A., C., Fuzzy Logic Applications Langar~ and Zadeh (Eds.), (1995) *Industrial Applications of Fuzzy,Logic and Intelligent Systems*, Piscata,Bay, NJ: IEEE Press.