DEMAND FORECASTING USING MODELS ARIMA AND ARTIFICIAL NEURAL NETWORKS (ANNS) MODELS

Maria Emilia Camargo
e-mail: Kamargo@terra.com.br
Walter Priesnitz Filho
e-mail: wpfilho@ucs.br
Marcelo Machado Barbosa Pinto
e-mail:mmbpinto@ucs.br
University of Caxias do Sul, CAMVA, Vacaria, RS, Brasil
Angela Isabel dos Santos
Department of Statistical, University of Santa Maria, Santa Maria,RS, Brasil
e-mail: dr_angel@terra.com.br

ABSTRACT

This paper presents the use of times series Auto-Regressive Integrated Moving Average (ARIMA) ARIMA model with interventions, and artificial neural network back-propagation models in analyzing the behavior of sales in a medium size enterprise located in Brazil for the period January 1990 to December 2005. The ARIMA model interventions presented a residual variation of 0.0008, where as the neural network model presented a residual variation of 0.0003. The chosen neural network presented, for the last 12 months, better forecasts was 4.2532 and that of the ARIMA model with interventions was 6.8237. The model obtained by the Neural Network was superior to ARIMA model, in adjustment as well as in forecasting for the data analyzed.

Keywords: Neural Network; ARIMA Model; Sales, Forecasting.

1. INTRODUCTION

Artificial Neural Network is becoming a poweful tool for scientists working in various fields, such as optimization, pattern recognition and forecast. The so called Neural Nets models are formed by a set of non-linear elements arranged in layers which operate in a parallel manner similarly to the human brain. In this paper we show a study of a neural network models applied to the area of time series analysis and forecasting and comparation wiht ARIMA model. The literature on the kind of application considered here stared in the second half of the last decade. The first papers showing the gain in forecasting accuracy with neural network models was published in 1987 by Lapedes & Farber (1988). After his study, various paper containing similar comparisons have always pointed to the same approachs. For a detailed review of forecasting the ARIMA methodology, we recommend Box & Jenkins (1976). This paper application an approach for study the sales data collected from medium size enterprise located in Santa Maria (RS), Brazil for the period January 1990 to December 2005. The methodologies related to this paper is presented in section 2. The empirical analysis and discussion on the results are presented in section 3. Conclusions are highlighted in section 4.

2. METHODOLOGIES

2.1 Artificial neural networks

According to Kohonen (1987) one possible definition to artificial neural Networks could be. Artificial consist of parallel networks of single and adaptive interconnected elements which interacts with objects of the real world as do the biological neural system. The neural network models are formed by associating a set of known inputs to the corresponding set of outputs previously stored. As stated by Rumelhart & Mc Clelland (1986) these models, which aim to represent the human brain and its neurons, has the "Parallet Distributed Processing property and is characterized by following aspects:

- a set of processing units;
- an activation state;

- an output function for each units;
- weights for the connection as between units;
- an weighting propagation rule;
- an activation rule which combines the input of a unit with its current state to produce a new activation level for the unit;
- a learning rule, which alows the changes in the weights;
- an enviroment where the system operates.

2.1.1Backprogation model

The backpropagation model thoroughly used in the current formulation of neural network models is the paradign normally adapted in areas such as, pattern recognition and mainly in the forecasting or time series (Beale, 1990; Refenes, et al, 1997).

Equation (1) to (3) show below the backpropagation algorithm. The subscripts *i* and *j* are used to identify a particular weight and *k* to identify the layer while the superscripts denote the step of adjustment. To start the process, small random numbers are set as weights and then equations (1) to (3) are used to adjust them.

$$\Delta w_{y,k}^{n+I} = \eta \delta_{j,k} y_{i,k-1} \tag{1}$$

$$w_{y,k}^{n+1} = w_{y,k}^{n} + \Delta w_{y,k}^{n+1} \tag{2}$$

$$\delta_{j,k} = \left(\frac{dy}{dz}\right)(y_{i,k} - y_{i,k}^{T}) \tag{3}$$

for *i*=1, 2 ..., I; *j*=1,2 ..., J and *k*=K

where: $w_{y,k}^{n+I}$ is the weight connecting neuron *i* in layer *k-1* to neuron *j* in layer *k* at step *n+1*;

η is the training rate coefficient (usually between 0.001 and 1.0);

$y_{i,k}$ is the output of neuron *i* in layer *k*;

$y_{i,k}^{T}$ is the target value for $y_{i,k}$.

Note that the algorithm above can only be applied to the last layer (*k*=K) of neurons, since it requires the knowledge of the target output value. For the hidden layers there is no target output and, therefore, $\delta_{j,k}$ has to be obtained in another fashion. Backpropagation derives its name from the fact that it propagates the value of $\delta_{j,k}$ backwards throughout the networks. For the hidden layers equation (3) has to be modified as:

$$\delta_{j,k} = \left(\frac{dy}{dz}\right)_{i,k} \left(\sum_{j} \delta_{j,k} w_{y,k}\right); \qquad \text{for } k=1,2 \ldots, K-1. \tag{4}$$

2.2 ARIMA(p,d,q) model

Their approach consists essentially in fitting a parametric linear stochastic model to the stationary time series. They argue that most real processes, although non-stationary, exhibit some consistency in their behaviour. In particular, it is often the case that the first or second difference of the process is stationary. Thus although $\{Z_t\}$ may be non-stationary, $\{\nabla^d Z_t\}$ is stationary. It having obtained a stationary process $\{\omega_t\}$ say, by differencing if necessary, they attempt to identify it as a mixed autoregressive moving average (ARIMA) process, i.e. it satisfies an equation of the form

$$\phi(B)\omega_t = \theta(B)a_t \quad (5)$$

where:

$$\omega_t = \begin{cases} \Delta^d Z_t, d > 0 \\ Z_t, d = 0 \end{cases}$$

B: backward shift operator, is defined as $B^k Z_{t-1} = Z_{t-k}$

$\Delta^d Z_t$: difference operator, is defined as $\Delta^d Z_t = \omega_t = Z_t - Z_{t-d}$
Alternatively, we can rewrite this model in level as form

$Z_t = \frac{\theta(B)}{\phi(B)(1-B)^d} a_t$	(6)

$\phi(B) = (1 - \phi_1 B - ... - \phi_p B_p)$, is an autoregressive polynomial of order p;
$\theta(B) = (1-\theta_1 B - ... - \theta_q B_q)$, is a moving average polynomial of order q.

The roots of $\phi(B) = 0$ must lie outside the unit circle in order to guarantee stationarity (of Z_t) and to ensure uniqueness of representation, the roots of $\theta(B) = 0$ must also lie outside the unit circle.

a_t: white noise process, normally and independently distributed with mean zero, constant variance σ_a^2 (NID(0, σ_a^2)), and independent of Z_{t-1}, that is,

$E(a_t) = 0$

$$E(a_t a_s) = \begin{cases} \sigma_a^2, t = s \\ 0, t \neq s \end{cases}$$

$E(a_t, Z_{t-1}) = 0$

This paper the construction of ARMA (p,d, q) model was obtained through the iterative cycle of Box-Jenkins methodology: identification, parameter estimation and diagnostic checking, Box & Jenkins (1976).

i) Model identification defines the (p,d, q) orders of the AR and MA components, nonseasonal. In this step, fundamental analytical tool is the autocorrelation function (ACF) and partial autocorrelation function (PACF) .

The ACF and PACF are very important for the definition of the internal structure of the analyzed series. The ACF ρ_k at lag k of the Z_t series is the linear correlation coefficient between Z_t and Z_{t-1}, calculated for k =0, 1, 2...

$\rho_k = \frac{\gamma_k}{\gamma_o} = \frac{Cov[Z_t, Z_{t+k}]}{\sqrt{Var(Z_t) Var(Z_{t+k})}}$	(7)

where:

Var (Z_t) = Var $(Z_{t+k}) = \gamma_0$ = variance of process

$\rho_o = 1$ e, $\rho_k = \rho_{-k}$.

An estimate of ρ_k can be calculed using the formula:

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_o} = \frac{\sum_{t=1}^{n-k}(Z_t - \overline{Z})(Z_{t+k} - \overline{Z})}{\sum_{t=1}^{n}(Z_t - \overline{Z})^2}, k = 0, 1, 2, ... \quad (8)$$

The PACF is defined as the linear correlation between Z_t and Z_{t-1}, controlling for possible effects of linear relationships among values at intermediate lags.

$$\mathrm{Cor}\left(Z_t, Z_{t+k} \middle| Z_{t+1}, \ \ldots, \ Z_{t+k-1}\right) \tag{9}$$

Theoretically, both an AR (p) process and an MA (q) process should be associated with well-defined patterns of ACF and PACF, usually decreasing exponential or alternate in sign or decreasing sinusoidal patterns. A precise correspondence between ARMA (p,q) processes and defined ACF and PACF patterns is more difficult to recognize. When the order of at least one of the two components (AR or MA) is clearly detectable, however, the other can be identified by attempts in the following step of parameter estimation.

ii) Parameter estimates are usually obtained by maximum likelihood, which is asymptotically correct for time series. Estimators are usually sufficient, efficient, and consistent for Gaussian distributions and are asymptotically normal and efficient for several non-Gaussian distribution families.

iii) Validation of the goodness of fit of an ARMA model can be developed according to the following steps:

1) Evaluation of statistical significance of parameters by the usual comparison between the parameter value and the standard deviation of its estimate. For a test statistic that is valid only asymptotically, a parameter whose value exceeds twice its standard error can be considered significant.

2) The stage of verification of the choice of the model, affected in the previous item, consists in evaluating if the residuals of that model forms a process of white noise.

The verification can be made through the autocorrelation of the residuals, or either, the inspection of the graph $\hat{\rho}_k(\hat{a})$. If the model is adjusted, the autocorrelations $\hat{\rho}_k(\hat{a})$ must practically be all inside of the limits of ± 2 standard deviation. If the verification of the diagnosis accuses inadequacy of the model, it is necessary to find a new model for study. If model inadequate, repeat procedure

2.3 ARIMA Model with Intervention

The intervention model we would modify equation (6) as follows:

$$Z_t = \psi(B) I_t^T \frac{\theta(B)}{\phi(B)(1-B)^d} a_t \tag{10}$$

with $\psi(B) = \dfrac{\omega(B)}{\delta(B)}$ or, alternatively, as

$$Z_t = \omega(B) I_t^T + N_t \tag{11}$$

with

$$N_t = \frac{\theta(B)}{\phi(B)(1-B)^d} a_t \tag{12}$$

and where $\psi(B) I_t^T$ represents the intervention model. I_t^T could be a step function S_t^T or a pulse input variable P_t^T

2.4 Performance Criteria

We measure the performance of both models by the out-of-sample Root Mean Ssquared Error (RMSE), defined as:

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{Z}_{t+1} - Z_{t+1})^2}$$

where $\hat{Z}_t$ t is the model prediction at time t and T is the size of the out-of-sample training set.

The Mean Absolute Percentage Error (MAPE) is measure of accuracy in a fitted time series value . It usually expresses accuracy as a percentage.

$$MAPE = \left[\frac{1}{T}\sum_{t=1}^{T}\left|\frac{Z_t - \hat{Z}_t}{Z_t}\right|\right]$$

3. EMPIRICAL ANALYIS AND RESULTS

The sales data collected from medium size enterprise located in Santa Maria (RS), Brazil for the period January 1990 - December 2005 have been analyzed using the ARIMA model with intervention and Neural Network back-propagation model.

3.1 ARIMA Model with Intervention

The adjusted model is with interventions ARIMA (1,2,1). The estimated parameters and statistics of the model are presented in Table1.

Table 1 – Estimated parameters and statistics "t" for the univariate model with interventions for the time series sales data

Parameter	Estimated Value	Statistic "t"
ϕ_1	-0.2553	-2.02
θ_1	-0.3987	-3.38
ξ_1	-0.1256	-3.78
ξ_2	-0.3765	-4.10
ξ_3	-0.1857	-3.97
ξ_4	0.1074	2.18
ξ_5	0.8476	1.85

The adjusted statistic and noise statistics for the model are:

$R^2 = 0.9546$: Mean ≅ 0.0000; and Variance = 0.00008

The identified interventions for the sales data during the period analyzed are presented in Table 2 considering a level of significance of 5%.

Table 2 – Types of detected interventions

Type of intervention	Instant	Period
$I_{1t} \rightarrow 1$	04	APRIL/90
$I_{2t} \rightarrow 1$	45	SEPT/93
$I_{3t} \rightarrow 2$	56	AUG/94
$I_{4t} \rightarrow 3$	60, 72,…	DEC/94
$I_{5t} \rightarrow 2$	65	MAY/95

The types of interventions occurred are: impulse, step, and seasonal impulse. It observed that the estimated coefficients of the intervention variables ξ_i have their expected signals. That is ξ_1, ξ_2 and ξ_3 have negative signals, when ξ_4 and ξ_5 have positive signals:

1. The first intervention represent the reflection due to Color Plan that imposed freezing of prices, which was in vigor from March of 1990:
2. The intervention of Sptember 1993, is the reflection of heterodox shock of Real Plan.
3. The intervention occurred in August 1994 is due to the price increase in consequence to inflationary memory;
4. The increase in sales in Dec 1994 is characterized by the seasonal effect;
5. Finally X_{5t} represents the level change.

3.2 Neural Network

Concerning the Neural Network model, based on the linear dependence structure identified in the Box & Jenkins model, the selected network was the (4,2,1).

1. four units in the input layer;
2. two units in the hidden layer;
3. one unit in the output layer Z_{t-1}.

Training: The sales series was trained 1400 times, updating the weight for every 30 repetitions. The learning constant was maintained at 0.12 and in the last 300 repetitions, a memory loss term of 0.4 was used. This term was used to provide more weight for the most recent observations. The momentum term used was 0.7. The varying interval size of the weight was 4.

To help comparing the two approaches we calculate the Mean Absolute Percentage Error (MAPE) and the Root Mean Square Error (RMSE).. The observation from January to December 2005 were used to analyze the forecasting performance of the fitted models. The performance of the two approaches is summarized below in Table3.

Table3. Performance of the two approaches

Forecasting model	RMSE	MAPE
ARIMA	0.1548	0.7216
Neural Network	0.1126	0.5843

The results show that the Neural Network model adjust well to the sales data, and provide acceptable forecast for the period analyzed with basis in the RMSE and MAPE, respectively. Hence, the Neural Network model is more adequate in study the sales data.

4. CONCLUSIONS

We presented in this paper two approaches for the study the sales data collected from medium size enterprise located in Santa Maria (RS), Brazil for the period January 1990 - December 2005. Box and Jenkins (1976) formalized the ARIMA modeling framework by defining three steps to be carried out in the analysis: identify the model, estimate the coefficients and verify the model. The intervention analysis revelead 5 significant events ($p<0.05$).

The question concerning the identification of optimal architecture of the network has not been answered. However, as we showed, some clues on the nature of the nature of the input units could be obtained via an exploratory data analysis, including plot of series, autocorrelation and partial autocorrelation functions.

Neural Network model was more efficient that ARIMA model for study of sales data.

REFERENCES

1. Alexander I. & Morton H. (1990), An introduction to neural computing. Chapman and Hall, London.

2. Beale, R. & Jackson, T. (1990), Neural Computing - An Introduction. Adam Hilger Publ.

3. Box, G. E. P. & Jenkins, G. M. (1976),. Time Series Analysis: forecasting and control. San Francisco, Holden-Day.

4. Kohonen, T. (1987), State of Art in Neural Computing. Proceedings of the IEEE. International Conference on Neural Networks, vol 1, San. Diego.

5. Lapedes, A. & Farber, R. (1988), How Neural Nets Works. In Neural Information Processing Systems, ed. Dana Z. Anderson, pp 442-456; American Institute of Physics, New York.

6. Refenes, A. N., Burgess, A.N. And Bentz, Y. (1997), Neural networks in financial engineering: a study in metodology, IEEE vol.8, no. 6, November.

7. Rumelhart, D. & MC Clelland, J. (1986), Parallel Distributed Processing: Explorations in the Microstructure of Cognition, vol 1,2 & 3. Cambridge; Bradford Book - Mit Press.