Temporal series and neural networks: a comparative analysis of techniques in the Brazilian retail sales forecast

Claudio Felisoni de Angelo[†]
FEA/USP

Ronaldo Zwicker[Ω]
FEA/USP

Nuno Manoel Martins Dias Fouto[¥]
FEA/USP

Marcos Roberto Luppe[£]
FEA/USP

ABSTRACT: An important economic activity in any society regards the commercialization of assets. The retail consists exactly of the link established between the industry and the final consumer. To predict the sales is essential so that one can manage in a proper way the production and commercialization processes. In the retail, this aspect is even more important. To sale means to harmonize the concerns of those producing with those who buy. Therefore, this paper is intended to exam comparatively the application of two retail sales forecast methods in the Brazilian market: the temporal series and the neural networks. The selection of those two techniques as object of that comparison was aroused by the importance those two conceptions have assumed in the literature. Although the utilization of neural networks has provided the smallest sum of the squares of the residues, one may say that the results using models of the ARIMA type have shown to be practically equivalent.

Keywords: Temporal series; neural networks; sales forecast; retail.

Authors' correspondence*:

[†] Doctor in the area of Economics of the Organizations – FEA/USP.
Link: FEA/USP.
Address: Avenida Luciano Gualberto 908, sala G-173 - Cidade Universitária - São Paulo – SP - Brazil
CEP: 05508-010
E-mail: cfa@usp.br
Telephone: (11) 3091 6045

[Ω] Doctor in the area of Economics of the Organizations – FEA/USP.
Link: FEA/USP.
Address: Avenida Luciano Gualberto 908, sala G-173 – Cidade Universitária - São Paulo – SP - Brazil
CEP: 05508-010.
E-mail: rzwicker@usp.br
Telephone: (11)3091 6045

[¥] Doctor in the area of Economics of the Organizations – FEA/USP.
Link: FEA/USP.
Address: Avenida Luciano Gualberto 908, sala G-173 - Cidade Universitária - São Paulo – SP - Brazil
CEP: 05508-010
E-mail: nfouto@usp.br
 Telephone: (11) 3091 6045

[£] Doctor of Management by FEA/USP.
Link: FEA/USP.
Address: Avenida Luciano Gualberto 908, sala G-173 – Cidade Universitária - São Paulo – SP- Brazil
CEP: 05508-010.
E-mail: mluppe@usp.br
Telephone*:* (11) 3091 6045

1. INTRODUCTION

Selling has constituted an absolutely essential activity in any type of company or organization. Sales generate the flow of resources funding the current expenses, the investments, and of course, it sustains its own profitability.

However, this activity obviously requires a previous work. The production system and the distribution, depending on the type of product or service, require the anticipation of preparatory actions. The industry must complete the orders and once they are processed, the lots of goods or the allocation of the services must be properly directed. Such transactions demand time, effort and resources.

Time regards the full transaction period, that is, from the moment the demand has materialized in the market to the moment in which the referred request was met. The efforts and resources comprise the production factors involved and the respective utilization levels both of persons, and materials and equipment.

The closer the sales volume expectations are from the provided compliance conditions, the more efficient the operation shall be. High volumes estimated without the corresponding accomplishment generate high stocks loading costs. On the other hand, ruptures of supply by scarcity generate real opportunities losses and open unnecessary space for a better positioning of the competition.

If the forecast is a function present in all organizations, the evaluation of future sales may be one of the most important and frequent activities in that field. For this reason, the literature dealing with the matter is both rich and extensive. Many and different are the techniques being currently used in a large number of examples and cases related to the most different segments of the goods and services markets.

Generically speaking, one may divide the predictive sales techniques in two large groups: on one side, the qualitative techniques, and on the other side, the quantitative ones. The qualitative techniques seek to capture the individuals' sensing on the future flows by means of different analytical resources. On the other hand, the quantitative models are based on well defined and objective conceptual structures to accomplish their forecasts.

One may say that if on one hand, the qualitative techniques are more versatile and richer in the sense of incorporating different situations they are, on the other hand, more limited exactly by the lack of objectivity in their processes. The quantitative techniques in

their turn are less flexible, but they allow a cleared discussion on the prior conjectures established on their results.

In other words one may argue that the qualitative models seek to expand the perception horizons, capturing subjectively the consequences arising from the individuals' mental structure (BROCKHOFF, 1983). The quantitative conceptions, in their turn are supported on formal and perfectly defined analytical structures.

One cannot set preferences among the different methods. The literature richness evidences that the techniques do not compete among each other; in fact, they supplement each other. The joint application of the two analytical frameworks has been largely used aiming the enrichment of the quantitative models' severity with the creative flexibility of the qualitative conceptions. Thus, nowadays, the planning areas of the duly organized companies combine different approaches in order to accomplish the forecast activity considering the higher volume available of information, whether they are qualitative or quantitative.

The interaction among the individuals improves the forecast quality (SNIEZEK, 1989). The active participation of a larger number of individuals widens the analytical horizons expanding the critical capacity of those involved. This process may be even more efficient by supporting their interpretations in more objective parameters provided by the quantitative methods. (ANG, O'CONNOR, 1991; FRANSES, 2008).

By recognizing the importance of combining the quantitative and qualitative techniques, this paper is specifically intended to exam comparatively two quantitative techniques applied to the retail sales forecast in Brazil: the temporal series and the neural networks. It was set on those two approaches due to the wide application of the two methods in many similar problems reported in the literature on that matter.

The comparison between forecast techniques has also been widely contemplated in academic papers. Several studies have been made comparing results provided by different theoretical conceptions. This paper is inspired in this scientific production line trying to offer subsidies to the predictive sales process in the Brazilian retail.

The article is organized in five items. Item 2 outlines the Brazilian economic panorama which the retail sales projections are based on. Item 3 presents a revision of the literature on the forecast models related to the temporal series and the neural networks. Item 4 sets the models, synthesizes and compares the results achieved. Finally, item 5 explains the main conclusions and marks the possible article extensions.

2. THE BRAZILIAN ECONOMY AND THE CONSUMPTION MARKET

One of the main components of the aggregated demand is without a doubt, the consumption: the expenses of the families in buying products or in the contracting of services. The evolution of that variable depends on the general conditions of the economy outlined mainly by the income evolution, of the general prices level and of the interest rate. The growth of the real income favors the consumption, as well as the most favorable credit conditions expressed in terms of smaller interest rates and/or longer terms

One of the possibilities to express the consumption evolution is to consider the sales of goods intended for the final consumption, that is, the goods retail. Although the consumption is more comprehensive including all expenses made, in this paper the analytical perspective refers exclusively to the commercialization of goods.

The commerce of goods as a result of the high inflation rate and of the resulting loss of acquisitive power which devastated the Country basically from the end of the 1970s on evolved very slowly throughout the 80s and part of the 90s. The income concentration decreased systematically from 1960 to 1990. The 50% poorer who represented in 1960 17.7% of the population felt to a little less than 14% in 1980 and to 12% in the early 90s. (IPEADATA, 2008).

Several attempts were made with the purpose of controlling the pace of the prices expansion and the resulting economy disturbance. However, all initiatives, such as the *Cruzado*, *Bresser* and *Collor* Plans in one way or another, or still, more properly, by a combination of economic and political factors, have failed.

However, in 1994 the *Real* Plan was launched. In fact, for the first time, after several initiatives the inflation was effectively decreased and kept at levels similar to those recorded in the economies of the developed countries. The general price index has fallen from approximately 5,150.00 % in June 1994 to approximately 10% in December 2001. Next, the prices evolution was always much lower than the levels seen before the Bresser Plan issuing. (GIAMBIASI, 2005).

Like other markets, the Brazilian retail commerce also went through big and profound transformations. Until then, that is, until 1994, the operating commerce aspects namely the linking function between the industry and the final consumer have been neglected due to the gains offered by the financial applications. In fact, the profit from the essential commerce

activity, that is, the purchase and sale of goods was much below the results that could be achieved in the financial market. (SESSO FILHO, 2001).

By decreasing the inflation and requiring the companies to regard the respective operations the competitiveness in the commerce of goods increased significantly. The variance of the relative prices was naturally decreased expressing in a most suitable and perennial form the values intrinsically advised by the market. Such higher visibility of the amounts made explicit in the prices has implied a decrease in the margins forced by the higher intensity of the rivalry among the companies.

The impact of the Real Plan can be seen in the income decentralization evidenced from the years 1996 to 2006. At the same time in which the economy grew the portion of the 50% poorer increased its relative weight in the income, from 12.09% to 14.47% (IPEADATA, 2008). In spite of the economic reorganizing promoted by the stabilization plan, the consumption expansion only occurred in a continuous form from the mid 2003 on.

The limitations to the consumption expansion in the first years of the Real currency may be attributed to the low economic growth. All along the first five years of the Real Plan, the financial market went through three important financial crises: Mexican (1994), Asian (1997) and Russian (1998) crises. In those years, although the inflation has been kept in quite low levels, the economy growth was very discrete, only 2.8% a year. (FERRARI-FILHO, DE PAULA, 2003).

As it can be seen in Chart 1, the improvement of the international situation allowed a more accelerated income expansion from early 2003, where the seasonal character of the retail sales in twelve months cycles is also evidenced. Such expansion has associated, or still, sustained, a gradual but continuing process of drops in the internal interest rates and a systematic elongation of the average payments' terms.

The increase in the real income may be illustrated for example, by the situation of the group comprised by the private sector workers with working papers. By the mid 2003 from May 2008 the referred group reported a real earning of eight percent points. In its turn, the interest rate in the funding for consumption has considerably dropped from May 2003, from 98.09% a year to 48.88% at the end of 2007. At the same time, the average term of the funding operations expanded from 249 days to 441 days (BANCO CENTRAL, 2008).
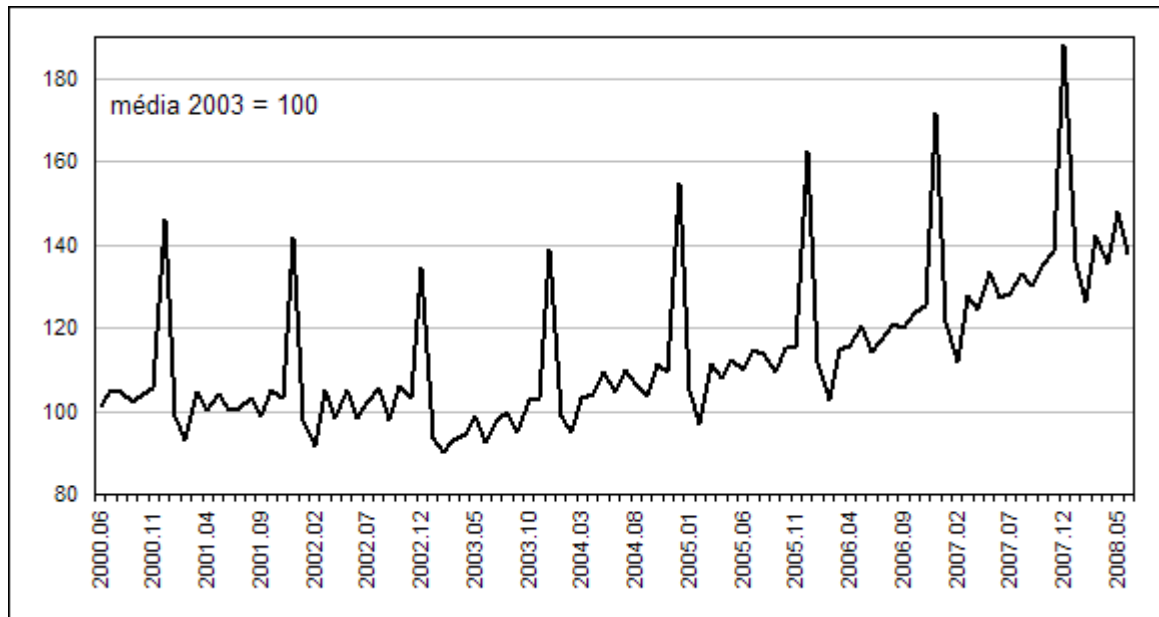
Chart 1 - Real sales of the Brazilian retail
Source: IPEADATA, 2008

Once more, those favorable conditions have implied also in income decentralization. In a recent research ordered by the French group BNP Paribas financing company, Cetelem, in partnership with the Research Institute Ipsos, on the Brazilian population distribution per consumption class, class C, representing 36% in 2006, went to 45% in 2007, reaching 86 million people. As to the classes D/E, which up to 2006 rated higher than C, presented a drop from 46% to 39%, falling to 73 million people, in 2007. The survey also shows that there was a decrease in the income inequality, with a slight average income drop of the A/B classes, growth of a large contingent to the class C and a small increase in the average income of classes D/E (DE CHIARA, 2008).

It is in the economic environment experienced in Brazil from June 2000 that one intends to investigate the sales performance in the Brazilian retail, investigating comparatively the utilization of the temporal series and the neural networks in the forecast of the marketed volumes. With this purpose the next section reviews those two techniques, which shall be subsequently applied to the data available. The estimation period starts in June 2000 and ends in June 2007. The data referring to the period from July 2007 to June 2008 shall be used to compare the adjustment degree of the two forecast techniques examined in this article.

3. TEMPORAL SERIES AND NEURAL NETWORKS

This item was split into two different topics. In the first one, a retrospective is made of the econometrics development of the temporal series, emphasizing the flatting methods and the stochastic models. In the second one, a synthesis of the neural networks' argument and their recent development is made. One seeks to explain the nature of this method, highlighting its applicability in the forecasting problems,

3.1 Econometrics of the temporal series

The temporal series constitute a completely separated chapter of the econometrics. Differently from the econometric models, the study of the temporal series has the essential purpose of making forecasts. There is no concern in setting the causal mechanisms; one intends only to perform accurate forecasts at the most.

The most traditional approach consists of the application of the flatting techniques. Holt in 1957 and subsequently Winters in 1960 conceived the model which came to be known in the literature as Holt-Winters, in which three parameters are identified: permanent, tendentious and seasonal component. In 1969, Pegel expands the contributions of the previous authors considering additive and multiplicative specifications (DE GOOIJER, HYNDMAN, 2006), explained below:

$$\hat{Y}_{t+k} = \alpha + b \cdot K + c_{t+k} \qquad (1)$$

$$\hat{Y}_{t+k} = (a + b \cdot K) \cdot c_{t+k} \qquad (2)$$

The value of "a" corresponds to the permanent component, "b" to the representative parameter of the trend and "c_t" to the factor associated to the additive or multiplicative seasonal behavior. The values of those parameters are given by the following expressions:

$$a(t) = a \cdot [y_t/c_t(t-s)] + (1-\alpha) \cdot [c.(t-1) + b.(t-1)] \qquad (i)$$

$$b(t) = \beta \cdot [a(t) - a.(t-1)] + (1 - \beta) \cdot b \cdot (t - 1) \qquad (ii)$$

$$c(t) = \gamma \cdot [y_t/a(t)] + (1 - \gamma) \cdot c_t \cdot (t-s) \qquad (iii)$$

In order for this softening to be achieved one needs, therefore, to consider the components permanence, trend and seasonality. The softening parameters are attributed in order to achieve the best possible adjustment to the data. The statistic programs compute those parameters interactively and automatically in order to minimize the errors. An example of that procedure can be found in Segura and Vercher (2001). The authors use Solver to set the most suitable parameters given a certain series.

A major improvement in the econometrics of the temporal series is due to the utilization of stochastic models. Such modeling starts from the presupposition that the series are stationary. In case the series are not stationary in level, one should take the first differences. If the series continue expressing a non-stationary behavior, as evidenced by specific tests, the second differences should be sought. Obviously this process goes on until the series becomes stationary. However, the practice has shown that the economic variables become stationary at most in the second difference.

A stationary series may be modeled in different ways considering basically two processes: the self-regressive and the moving averages processes. The modeling aims to reproduce the values of the "Y" concern variable from the two processes, either separately or in a combined form. The self-regressive model expresses the current "Y" value as a function of the values of the variable recorded in the past. The structure of the moving averages performs the representation in terms of the errors incurred in the previous periods, the expression (3) below expresses a self-regressive "Y" generation model.

$$(Y_t - \delta) = \alpha_1 (Y_{t-1} - \delta) + \alpha_2 (Y_{t-2} - \delta) + ... + \alpha (Y_{t-p} - \delta) + u_t \qquad (3)$$

Where δ is the Y variable average. One says that (3) describes a self-regressive process of a "p" order, or in other words, an AR(p) process. Otherwise, as it was pointed out in the previous paragraph, the values of Y may also be generated by linear combinations of error terms (white noise), that is:

$$Y_t = \mu + \beta_0 u_t + \beta_1 u_{t-1} + \beta_2 u_{t-2} + ... + \beta_3 u_{t-q} \qquad (4)$$

The expression (4) shows a model of moving averages of the "q" order, or even a MA(q) process. One may finally presume a generation process combining self-regressive and moving average terms. In that case, one says that the model is ARIMA (p,q). By adding the differentiation degree to make the series stationary, one obtains the ARIMA models (CHU, ZHANG, 2003).

The work by Box and Jenkins (1976) is essentially a criterion of structuring and composing the model. The method is split into four different stages. The first part consists of the proper specification for the stationary series of the self-regressive terms and those representing the moving averages. The second stage refers to the estimation process. Based on the values achieved for the parameters, one goes on to the third stage, which is the result investigation stage, mainly regarding residues. An indication of the model adequacy is that such results must be qualified as white noise. Finally in the fourth and last stage the forecast is made.

De Gooijer and Hyndman (2006) present an excellent summary about the utilization of the temporal series techniques in forecast models over the last 25 years, until 2006.

3.2 Artificial Neural Networks (ANN's)

The artificial neural networks are inspired in the behavior of biologic neural networks. They are comprised of knots named neurons lied to each other by connections having a relationship with the connections formed by axons and dendrites comprising biologic neural cells. Generally, an ANN is comprised by a large number of knots organized in layers and which are connected to other knots by means of connections in which signals flow just as electric nature signals flow among biologic neurons. Each connection of a knot to other knot typically has a weight associated to it that, respecting the due differences, may be understood as representing a coupling degree existing between two biologic neurons.

The task accomplished by a knot is typically a simple one. Initially, it consists of receiving the signals from other knots by the input connections, weighed by the corresponding weights of the connections and the summation of which shall correspond to the total knot input signal. Next, the neuron activation threshold represented in figure 1 by a_j is added to that total. That threshold is typically a negative value that works as a point from which the (total of) input signal shall determine an output signal. Weak input signals shall be inhibited by such threshold. Then, a function is applied to the input signal, computing an yl value corresponding to the knot output signal and that is transmitted as input to the other knots the first one is connected to. The function is denominated activation function and in general, it is of a non-linear nature.
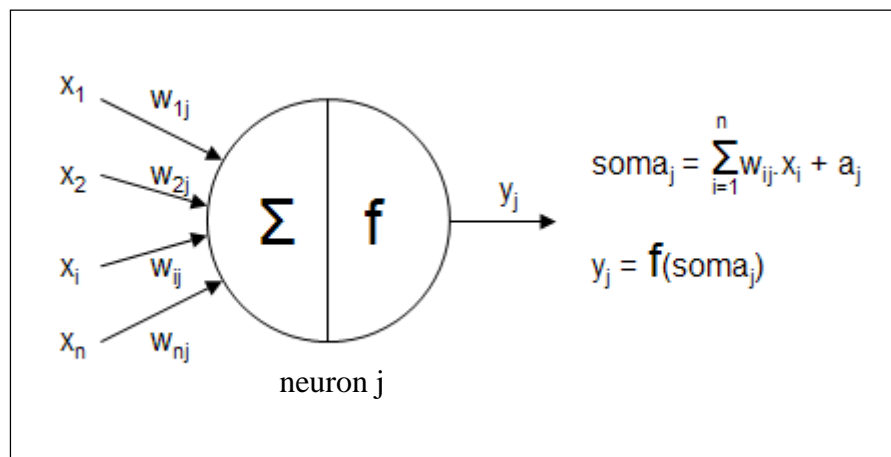

Figure 1 – How the artificial neuron works
Source: Made by the authors

The most disseminated model of the network is the MLP (*Multi Layer Perceptron*) model, which is also the most used model in the foreseeing of temporal series. Networks of that kind are comprised of a layer of input knots, one or more intermediate layers and an output layer. According to figure 3, the input signals are transmitted from the input layer to the output layer through the intermediate layers and there is no retro feeding of the intermediate or output signals (the feed is said to be feed forward). From the statistic point of view each input knot represents an independent variable and each output knot represents a depending variable that is being foreseen by means of the network (PALMER *et al*., 2006).
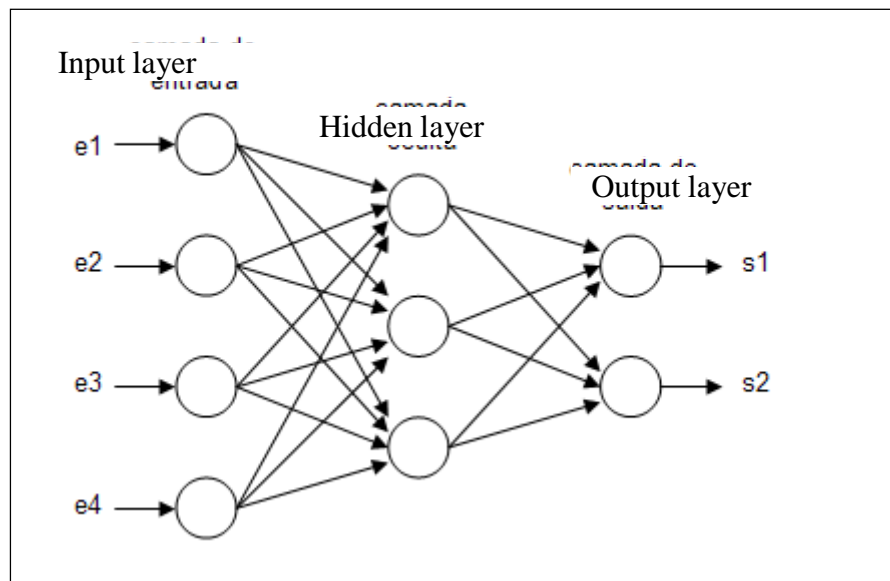

Figure 2 – Example of a neural network
Source: Made by the authors

Generally, a MLP network goes through an initial stage named training in order for it to be executed next, which in the forecast case, corresponds to the task of effectively computing forecasts. The connections weights are computed in the training stage from a set of input-output pairs, pertaining to the known values of the independent variables and corresponding values of the depending values. During that stage, from the known values of the input signals, the corresponding values of the output signals are computed, which in their turn are compared to the known values of the depending variables. In short, the difference between the known and computed values determines the interactive adjustment in the connections' weights, in order to minimize that difference or error. During the accomplishment stage the values of the input variables are presented to the trained network and it computes the output variables' values or foreseen values. In both stages, the computation of outputs from the inputs of each knot is made according to the previously specified operation of the artificial neuron.

The attainment of the difference between the known values and the computed ones is analog to the learning supervised by a teacher, where the result computed by the apprentice is systematically evaluated by the teacher and the difference with the wanted result determines changes in the apprentice behavior. On the other hand, the adjustment of the weights keep a relationship with the animal learning physiology which, in an extremely simplified way, leads to an approximation or departure of the connections (synapses) among biologic neurons. The bigger the weight is, the closest the connection.

The analogy between the ANN's learning and animals may be glimpsed considering the conditioning of a dog. The conditioning uses a very simple procedure: compensating the desired behavior and punishing the undesired one. As the dog reacts properly to an order to sit, we should compensate it for example, with a simple caress. We can understand that such caress makes the neural connections, which go from the neurons responsible for capturing the sound of the order to the neural terminals actuating the muscles and folding its rear legs, have their connections approximated in order to facilitate the input signal flow (sound) to the terminals actuating the muscles. Thus, upon receiving a new order to sit, the dog tends to sit more easily. Inversely, if the dog reacts unduly to the order to sit (for example, by lying down), we punish it, (for example, with a slight pull in its leash) and so we are increasing the distance between the neural connections conducting the input signal to the undue muscles.

The "artificial" designation in the networks' designation refers to the fact that those analytical constructions are only inspired in the biologic systems, particularly in the human brain study. The research in ANN's has shown that they have significant capacities of patterns' classification and recognition. Keeping the due differences, the ANN's have acquired their learning and generalization capacity from the experience analog to the human beings. According to Widrow *et al.* (1994), ANN's are used in several applications in the business, industry and science areas in a very successful way.

One of ANN's applications refers to general forecasting procedures. Without wanting to disqualify already well established forecasting procedures, the ANN's eventually offer interesting and attractive alternatives for those who study and makes forecasting.

According to Zhang *et al.* (1998), several ANN's characteristics make them attractive in the scope of forecasting procedures. Firstly, opposite to the traditional methods the ANN's constitute methods directed by data and self-adaptive points to the effect that they require little premises regarding the models representing the problems under study. This means that they can learn from examples by capturing subtle relationships among the available data, even

if such relationships are beforehand unknown or hard to describe. Putting it differently, an ANN is used when one does not know the precise nature of the relationship among inputs and outputs, and in case the relationship was known, then one would model it directly. In short, ANN's are indicated to help in the resolution of problems needing a knowledge that is difficult to specify, provided that there are sufficient data or remarks.

One may say that the ANN'S learning capacity from the experience constitutes in a very useful way of tacking problems for which there are data without however, one has more knowledge about the processes generating the referred information. There are clearly many situations where it is easier to obtain data than to obtain good theoretical models regarding the problem under study. This situation is similar to that addressed previously when the econometric models of temporal series were formulated. Also as an example of such analytical conceptions, the modeling based in the neural networks may be very limited when one has little observation.

Another characteristic mentioned in the literature on the ANN's is that once a stable network is established, one may make inferences or generalizations. Networks so structured and generated by the learning process are in principle, able to properly infer results related to data not used explicitly during the training also when they are affected by noise. That inference capacity has a particular concern in the forecast where the association of future data observed with past data constitutes a training course so that it can then perform generalizations by means of the association of data observed with non-observed future data.

Such associations constitute universal approximations of functions. According to Haykin (2001) it was demonstrated that certain ANN's types may approximate any continuing function to any precision level. Generally, ANN's constitute functional forms that are more general and flexible than the forms with which the conventional statistic methods can work.

In general, one presumes that there is some relationship, either known or unknown, among inputs (past variables or values in the forecast situation) and outputs (future variables or values). Typically the conventional forecast statistic methods have limitations to estimate such relationship or function and in principle, the ANN's may contribute to excel those limitations.

Thus, the ANN's are intrinsically able to capture non-linearities while the traditional forecasting methods typically have linear models behind them. For example, the models generated by the Box-Jenkins method as it was seen, presume that the analyzed time series is

generated from linear processes. It is a recognized fact that linear models have the great advantage of being able to be understood and reviewed in depth and detail, besides being easy to explain and implement. On the other hand, it can be widely questionable to presume that the model behind the data is a linear one, in fact it can be seen that this is totally inappropriate when the data result from a non-linear process, which is not unusual in the practice.

Even though there are non-linear conventional statistic methods presuming a preset non-linear model, they are intrinsically restricted when such model is presumed without a higher knowledge of the mechanisms generating the data at issue. Upon formulating a model by the conventional methods, we are typically limiting the possible generating mechanisms; clearly that model may be insufficient to capture all the non-linear characteristics of the data.

In principle one may argue that ANN's by the fact that they are non-linear approaches solely sustained by the data, are able of non-linear modeling without any beforehand knowledge of the mechanisms linking the input variables to the output variables. As discussed by Zhang *et al.* (1998), ANN's constitute a more general and flexible modeling resource for the forecast tasks.

The utilization of ANN's in forecasts has been disseminated from the introduction of the retro propagation algorithm for training the multiple layers' networks circa 1986. In short, this was the algorithm that initially set the non-linear capacity of the ANN's. Since that time, many writers have made comparative tests between the effectiveness of the networks and statistic resources. A recent extensive study using a standardized and publicly used data base for tests can be found in Zhang and Kline (2007).

Thus, the next chapter discusses the estimation models used herein. The models resulting from the smoothing techniques' applications, the model roused by the Box-Jenkins method and the one resulting from the utilization of neural networks are presented.

4. ESTIMATION AND RESULTS MODELS

This section of the paper is reserved to the presentation of the results. Initially, the models and the respective forecasts are addressed. The period from July 2000 to June 2007 was used for setting the models. The period from July 2007 to June 2008 was taken into account to investigate the comparative adjustment of the different methods. The data refer to the real sales' index of the Brazilian retail (June 2000 = 100) published in the website www.ipeadata.gov.br (IPEADATA, 2008). At the end, a comparison of the methods is made using as adjustment criterion the summation of the square of the residues, that is, the quadratic

differences among the real values and the foreseen ones, according to the alternative formulations.

4.1 Additive and multiplicative exponential smoothing

The procedures, the models and the corresponding results are presented in this part of the paper. Starting from the smoothing methods one estimated the Holt-Winters model in the additive and multiplicative form.

The results referring to the two models' parameters are found in Table 1. Based on the corresponding functions the projections for the period from July 2007 to June 2008 were made. The results are presented in Table 4 showing the summation of the squares of the residues of all models.

TABLE 1 – RESULTS OF THE HOLT-WINTERS MODEL

Parameters	Additive	Multiplicative
A	0,4100	0,4600
B	0,1000	0,1400
γ	0,0000	0,0000

Source: Drafted by the authors

4.2 ARIMA Model

The second estimated model was based on the Box-Jenkins technique to obtain ARIMA type predictive functions. The seasonality of the sales data generates a non-stationary series. As shown by Enders (1995, p.227), the series may show seasonal unitary roots.

That was exactly the situation of the data confirmed by the Dickey – Fuller test. Aiming to make the series stationary, the difference between the current value (index logarithm) and the twelve-month out of phase sales ($DY_{12} = Y_t - Y_{t-12}$) was taken. Applying the Dickey- Fuller test again, it was seen that the series thus considered showed a stationary pattern.

Once the stationarity of the variable of concern has been obtained, we went on the estimation step, considering the sales discrepancies, the distribution of the errors in time, the economicity criterion in the representation of the stochastic processes (Akaike-Schwarz) and the significance of the estimated parameters is obvious. The best results were obtained working with the model as presented in Table 2. The stationarity of the model residues was also seen rejecting the assumption of the unitary root existence.

TABLE 2 – RESULTS OF THE ARIMA MODEL			
Variables	*Coefficients*	*Standard error*	*P*
C	0.034189	1.061767	0.2922
AR(1)	0.275941	2.330029	0.0229
AR(2)	0.350662	3.104815	0.0028
AR(3)	0.306025	2.609613	0.0112
R^2	72,26%		
Adjusted R^2	71%		
Akaike	-5.830982		
Schwarz	-5.702497		
Durbin-Watson	1.863138		
F	57.30172		0.0000
Jarque-Bera	6,260		5%

Source: Drafted by the authors

4.3 Neural networks

The application of ANN's in forecast problems is not a common task because many aspects must be taken into account and many decisions must be made. The suitable selection of the ANN architecture is very important which involves the selection of the intermediate layers' number, the number of knots in each layer and the interconnection of those knots. Also to be taken into account are the training algorithm, the activation functions, the data standardization and pre-processing, the selection of the data sets for training, checking and test and the adjustment quality measures. Zhang *et al. (*1998) make a synthesis of those issues and of the reference that gives them empirical treatment.

There is no set method for determining the multiple parameters appearing in the ANN's application problem and forecast problems. It is not possible to set an optimum solution but there are directives and practical rules that in general lead to satisfactory solutions.

For forecast problems, networks with only one intermediate layer are typically adopted and the number of knots in that layer is determined by experimentation with the suggestion that this number is around the number of input knots. In forecast problems of temporal series the number of input knots corresponds in general to the number of out of phase observations deemed necessary to unveil the series' behavior and forecast pattern. For example, for univariate annual monthly series the practice is to adopt the seasonal cycle in months as the number of input knots and the number of output knots is typically the forecast horizon (generally equal to one). Besides, the interconnection among the knots is complete, that is, the knots of a layer are connected to all knots of the next layer in a network scheme of the feedforward type.

In forecast problems the logistic function is normally used as an activation function for the intermediate knots and the logistic or linear function for the output knots. Zhang *et al.* (1998) warn that the linear function restricts the network capacity to model temporal series having a trend, and suggest the previous data processing to remove such effect.

Regarding the training algorithm the most disseminated method is the retro propagation algorithm with the specification of its parameters of learning rate and momentum in an empiric form. The utilization of more efficient methods may be indicated, such as the Levenberg-Marquardt or Quickprop algorithm, tending to converge more rapidly and avoiding to become arrested to local minimums.

As to the data standardization, it should be seen that when non-linear activation functions are used in the output knots, the desired output values must be changed for the output interval that the network provides. In general, the standardization of all data is recommended in the case of temporal series, also if the output activation function is a linear one. The matter of data sets selection for training, checking and test is peculiar of ANN's. As much as possible they should be disjointed sets where the training set is used solely for the determination of the weights and the checking set is used to check if the output error (for that set) has increased again showing that the network is "memorizing" the input data and losing its capacity of generalization.

As to the test set, it is used to evaluate the network forecast capacity and allow the comparison among networks and different methods. The need to use checking and test sets generally affects the network performance, particularly when the series is constituted with little observations. In general, the performance measuring it made based on the difference between the value foreseen by the network and the desired value. Several methods are used for example, the sum of the square of the computed differences for the test set.

Finally, according to Zhang and Qi (2005), particular attention should be given to the issue of the presence of seasonal effects and trend in temporal series submitted to neural networks. While some authors consider that the ANN's are fully capable of modeling such effects directly, others realize exactly the opposite. The empiric analysis made by the two mentioned authors concludes for the effectiveness of the temporal series pre-processing aiming the previous and joint removal of the seasonal effect and the trend.

The approach by neural networks to the temporal series deemed herein was not intended to seek solutions that are deemed optimum. The software *Statistica Neural Networks*

4.0B using the resource *Intelligent Problem Solver* (IPS) was used. The deemed seasonal period was 12 observations, also for the series with the seasonality removed, which sets also the number of input knots to 12.

With the help of that resource, several networks architectures were tested varying the number of intermediate knots and identifying the candidate networks with better performance based on the criterion of increasing the checking set error. Just as it was reported by Faraway and Chatfield (1998), the capture effect for local minimums is recurring along the procedure. That leads to multiple networks with the same number of intermediate knots, but with different checking errors as a function of the network's initial weights.

In the processing made the checking set consisted of 12 observations distributed along the data interval effectively used in the training and the test set was constituted by the last 12 series observations. The data from those two sets did not attend the training. The utilization of those two sets implies that less data are used in the training and mainly, the checking set affects the determination of the "best" trained network.

Further, that "best" network is not necessarily the network that provides the best adjustment to the test set. In order to help in the selection process the distribution of the forecast errors all throughout the series was visually reviewed, considering that these errors should not be focused on certain sections of the series.

Table 3 reports the best identified results. The *transfor* column shows the previous transformation of the series data: original data (ori), removal of variability and trend (varten) and removal of variability, trend and seasonality (vartensaz). The *nós* column shows the number of intermediate knots of the network remarking that all networks had 12 input knots and one output knot. The *correlação* column shows the relation among the original data and the forecast for the training set (*ter*), checking (*ver*) and test (*tes*).

The *REQM* column shows the root of the average square error among the original data and the forecast for the same sets. The next column presents the sum of the offsets square among the original data of the test and forecast set, serving as comparison with the other forecast methods hereof. Column *n* reports the number of remarks (or even better, the output data) effectively used in the training considering the loss determined by the checking and training sets, the loss resulting from the previous transformation of the data and the data "ignored" from the series start (because they do not constitute outputs).

TABLE 3 – RESULTS OF THE NEURAL NETWORKS' MODELS

Transfor	nós	correlaçãɔ			REQM			$\Sigma desv^2$	n
		tre	ver	tes	Ter	ver	tes		
ori	12	0.99	0.94	0.99	2.41	4.32	2.45	71.8	61
	12	0.96	0.99	-	4.39	1.37	-	22.5	73
varten	10	0.99	0.99	0.99	2.68	3.16	1.73	36.0	60
	12	0.99	0.99	-	2.73	1.17	-	16.4	72
vartensaz	12	0.96	0.98	0.99	8.33	2.90	2.15	55.2	48
	13	0.97	0.99	-	3.63	1.59	-	30.5	60

Source: Drafted by the authors

Table 3 reports the three transformations suggested by the reference. The adjustment to the original data is deemed satisfactory by many authors and that is confirmed in the present case. The removal of trend is deemed important by others, and apparently, it has produced some improvement.

Finally, other authors suggest the removal of the variability, trend and seasonality which apparently is also confirmed in a smaller scale. For each transformation, two networks were reported: (1) the network with the smallest checking EQM where the test set was replaced with the checking set, at the end of the series, in order to provide 12 more observations for the training.

From the brief analysis made one can infer that the use of the original series may be satisfactory but the trend removal seems to be the most indicated, confirming the reference. The widest transformation of the data possibly is also shown, observing that it is the situation where the highest data loss occurs, which is deemed relevant in the case, because the original series is already a short one. In short, all reported networks and many others identified along the tests point to models that are comparatively as good as, or shortly better than the traditional methods.

TABLE 4 – COMPARISON OF THE MODELS

Models	$\Sigma desv^2$	Rating
Additive exponential smoothing	123.46	4
Multiplicative exponential smoothing	116.87	3
ARIMA	37.45	2
RNAs	16.40	1

Source: Drafted by the authors

5. CONCLUSIONS

Predicting is one of the most important and challenging activities in all areas of the human knowledge. It is important because well done and credible forecasts allow anticipating future situations in order to prepare in advance the systems for the most appropriate answers.

In the management of the companies and organizations, the administrators or the persons in charge for preparing and implanting policies are permanently, tacitly or explicitly trying to peer into the future. Based on their conjectures or certainties, decisions are made with wide or restricted developments, depending on the nature and scope of what one predicts.

The set of initiatives covered by the sales forecast is certainly one of the areas where efforts are frequently made in order to identify the future evolution profile. In retail, that activity is absolutely essential, since it is a connection activity between the industry and the final consumer, the efficiency and the efficacy of the commercial operations depend directly on the retail capacity to adjust its purchases and its stocks as a function of the consumption desires' pace and profile.

This article was essentially intended to discuss comparatively two quantitative forecast methods using aggregated sales data from the Brazilian retail market. On one side, the models resulting from the temporal series' econometrics and on the other side, the predictive structures based on the ideas aroused by the neural networks. The selection of those two approaches was based on a literature revision that has evidenced the marking presence of these two approaches in the quantitative studies of sales forecast.

Taking as comparative measure the sum of the square of the resides generated by the differences between the foreseen sales and the sales actually made, one may conclude that the series models named smoothing have shown to be substantially less accurate than the ARIMA temporal series models and the Neural Networks models to deal with the recent aggregated series of retail sales in Brazil. The same could not be concluded regarding the ARIMA models and Neural Networks, because the difference among those models may be deemed little expressive. Even though the ANN's show better results, it should be seen that their use incorporates difficulties, the main being the need of using checking sets to interrupt the calculation process. Putting it in a more general way, the question is if the location of good ANN's can do without previous knowledge of the ARIMA models' results. As pointed out by Faraway and Chatfield (1998), there is no general procedure for such.

This paper employed aggregated data. It would be interesting to investigate the behavior of the methods used herein with disaggregated information. It would be interesting also to investigate series that divert much from the linearity, series with a higher volume of observations and the use of other discrepancy structures for which possibly the use of ANN's may involve less difficulties and present better results.

REFERENCES

ANG, S.; O'CONNOR, M. The effect of group interaction processes on performance in time series extrapolation. **International Journal of Forecasting**, v.7, n.2, p. 141-149, 1991.

BOX, G. E. P.; JENKINS, G. M. **Time Series Analysis: forecasting and control**. San Francisco: Holden Day, 1976.

BANCO CENTRAL. **Séries compiladas pelo *Banco Central do Brasil*.** Available at: http://www.bcb.gov.br. Accessed on: Aug. 15, 2008.

BROCKHOFF, K. Group processes for forecasting. **European Journal of Operational Research**, v.13, n.2, p.115-127, 1983.

CHU, C.; ZHANG, G. P. A comparative study of linear and nonlinear models for aggregate retail sales forecasting. **International Journal of Production Economics**, v.86, n.3, p.217-231, 2003.

DE CHIARA, M. Com 86 milhões de pessoas, classe C já é maioria da população brasileira. **O Estado de São Paulo**, São Paulo, 27-03-2008, Caderno de Economia e Negócios. Available at: http://www.estado.com.br/editorias/2008/03/27/eco-1.93.4.20080327.31.1.xml. Accessed on: Apr. 30, 2008.

DE GOOIJER, J. G.; HYNDMAN, R. J. 25 years of time series forecasting. **International Journal of Forecasting**, v.22, n.3, p. 443-473, 2006.

ENDERS, W. **Applied econometric time series.** New York: John Wiley & Sons, Inc., 1995.

FARAWAY, J.; CHATFIELD, C. Time series forecasting with neural networks: a comparative study using the airline data. **Applied Statistics**, v.47, p.231–250, 1998.

FERRARI-FILHO, F.; DE PAULA, L. F. The legacy of the real plan and an alternative agenda for the Brazilian economy. **Investigación Económica**, v. 63, n. 244, p. 57-92, 2003.

FRANSES, P. H. Merging models and experts. **International Journal of Forecasting**, v.24, n.1, p.31-33, 2008.

GIAMBIASI, F. Estabilização, reformas e desequilíbrios macroeconômicos: os anos FHC (1995-2002). In: GIAMBIASI, F. et al (Org.). **Economia brasileira contemporânea (1945-2004)**. Rio de Janeiro: Editora Campus, 2005, p.166-195.

HAYKIN, S. **Redes neurais: princípios e prática** (translated). 2ª ed., Porto Alegre, Bookman, 900 p., 2001.

IPEADATA. **Séries compiladas pelo Instituto de Pesquisa Econômica e Aplicada**. Available at: http://www.ipeadata.gov.br. Accessed on: Aug. 14, 2008.

PALMER, A.; MONTAÑO, J. J.; SESÉ, A. Designing an artificial neural network for forecasting tourism time series. **Tourism Management**, v.27, p.781-790, 2006.

SEGURA, J. V.; VERCHER, E. .A spreadsheet modeling approach to the Holt-Winters optimal forecasting. **European Journal of Operational Research**, v.131, n.2, p.375-388, 2001.

SESSO FILHO, U. A. Crescimento e desempenho de redes de supermercado na década de 90. **Varejo Competitivo**, São Paulo, v.6, p.21-44, 2001.

SNIEZEK, J. A. An examination of group process in judgmental forecasting. **International Journal of Forecasting**, v.5, n.2, p.171-178, 1989.

WIDROW, B.; RUMELHART, D. E.; LEHR, M. A. Neural networks: applications in industry, business and science. **Communications of the ACM**, v.37, n.3, p.93-105, 1994.

ZHANG, G. P.; KLINE, D. M. Quarterly time series forecasting with neural networks. **IEEE Transactions on Neural Networks**, v.18, n.6, p.1800-1814, 2007.

ZHANG, G. P.; PATUWO, E.; HU, M. Y. Forecasting with artificial neural networks: the state of the art. **International Journal of Forecasting**, v.14, p.35-62, 1998.

ZHANG, G. P.; QI, M. Neural network forecasting for seasonal and trend time series. **European Journal of Operational Research**, v.160, p.501-514, 2005.