



Journal of Applied Sciences

ISSN 1812-5654

science
alert

ANSI*net*
an open access publisher
<http://ansinet.com>

Supply Chain Demand Forecasting; A Comparison of Machine Learning Techniques and Traditional Methods

¹J. Shahrabi, ²S.S. Mousavi and ²M. Heydar

¹Department of Industrial Engineering, Amirkabir University of Technology, Tehran, Iran

²Member of Young Researchers, Department of Industrial Engineering, Graduate School, Islamic Azad University, South Tehran Branch, Tehran, Iran

Abstract: In this study, supply chain demand is forecasted with different methods and their results are compared. In this research traditional time series forecasting methods including moving average, exponential smoothing, exponential smoothing with trend at the first stage and finally two machine learning techniques including Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs), are used to forecast the long-term demand of supply chain. By using the data set of the component supplier of the biggest Iranian's car company this research is then implemented. The comparison reveals that the results producing by machine learning techniques are more accurate and much closer to the actual data in contrast with traditional forecasting methods.

Key words: Artificial neural networks, support vector machine, demand forecasting, bullwhip effect

INTRODUCTION

This study is concerned with forecasting an Iranian's car components supplier company, a time-series with trend and seasonal patterns.

Some machine learning techniques, including artificial neural networks and support vector machines, are compared to the more traditional time-series forecasting methods, including moving average, exponential smoothing without and with trend as MAPE (Mean Absolute Percentage Error) index. These methods are chosen because of their ability to model trend and seasonal fluctuations present in suppliers' data. The objectives of this study are three-fold: (1) to show how to forecast wholesaler sales using ANNs and SVMs and (2) to display how various time-series forecasting methods compare in their forecasting accuracy of wholesaler sales.

The reasons for this study are both theoretical and practical. Theoretically speaking, how to improve the quality of forecasts is still an outstanding question (Granger, 1996). For data containing trend or/and seasonal patterns, failure to account for these patterns may result in poor forecasts. Over the last few decades several methods such as moving average, Holt method, Winters exponential smoothing, Box-Jenkins ARIMA model and multivariate regressions have been proposed and widely

used to account for these patterns. ANN is a new contender in forecasting sophisticated trend and seasonal data. Franses and Draisma (1997) suggested that ANNs be used to investigate how and when seasonal patterns change over time.

Industry forecasts are especially useful to wholesalers who may have a greater market share. For the supply chains, Peterson (1993) showed that larger retailers or wholesalers are more likely to use time-series methods and prepare industry forecasts, while smaller retailers emphasize judgmental methods and company forecasts. Better forecasts of wholesaler sales can improve the forecasts of individual retailers because changes in their sales levels are often systematic. For example, around new years, sales of most retailers increase. Moreover, models of forecasting individual store sales will often include assumptions about industry-wide sales and market-share.

Indeed, accurate forecasts of wholesalers' sales have the potential to improve individual stores' sales forecasts, especially of larger retailers who may have a significant market share.

In the past decade, ANNs have emerged as a technology with a great chance for identifying and modeling data patterns that are not easily discernible by traditional statistical methods in many fields of science, such as computer science, electrical engineering and finance. Qi and Maddala (1999) and Qi (1999) showed that

many studies in the finance literature evidencing predictability of stock returns by means of linear regression can be improved by a neural network. Qi (1996) did comprehensive survey of ANN applications in finance.

ANNs have also been increasingly used in management, marketing and retailing. The reader is referred to Krycha and Wagner (1999) for a comprehensive survey of ANN applications in management and marketing. Zhang *et al.* (1998) provided a comprehensive review of ANNs' usage in forecasting.

Jeong *et al.* (2002) have presented a computerized system for implementing the forecasting activities required in SCM. For building a generic forecasting model applicable to SCM, a linear causal forecasting model has proposed and its coefficients have efficiently determined using the proposed genetic algorithms (GA), canonical GA and guided GA (GGA). Compared to canonical GA, GGA adopts a fitness function with penalty operators and uses Population Diversity Index (PDI) to overcome premature convergence of the algorithm. The results obtained from two case studies show that the proposed GGA provides the best forecasting accuracy and greatly outperforms the regression analysis and canonical GA methods.

Willemain *et al.* (2004) forecasted the cumulative distribution of intermittent (or irregular) demand, i.e., random demand with a large proportion of zero values, over a fixed lead time using a new type of time series bootstrap. To assess accuracy in forecasting an entire distribution, they adapted the probability integral transformation to intermittent demand. Using nine large industrial datasets, they showed that the bootstrapping method produces more accurate forecasts of the distribution of demand over a fixed lead time than do exponential smoothing method and Croston's method (Croston, 1972).

Hilas *et al.* (2006) used forecasting models for the monthly outgoing telephone calls in a University Campus. In this study, three different forecasting methods, including Seasonal Decomposition, Exponential Smoothing and SARIMA method, have been used. Forecasts with 95% confidence intervals were calculated for each method and compared with the actual data.

Hill *et al.* (1996) showed that ANNs significantly outperform traditional methods of forecasting when forecasting quarterly and monthly data. Although theoretically speaking ANN may improve on the traditional time-series methods in forecasting a series with trend and seasonal patterns, Nelson *et al.* (1994) found that ANNs do not model the seasonal fluctuations in the

data very well. Foster *et al.* (1992) found that exponential smoothing is superior to ANNs in forecasting yearly data and comparable in forecasting quarterly data. Monthly data were not used in their study. Winters exponential smoothing model, in particular, has been found to provide superior forecasts in a variety of contexts. Dugan *et al.* (1994) showed that the Winters model can outperform both the Census X-11 and the random walk models in predicting a variety of income statement items (i.e., sales, earnings before interest and taxes, interest expenses, earnings before taxes, tax expenses and earnings before extraordinary items). This result was derived from a 15-year sample (1971-1985) of 127 manufacturing and retailing firms. Alon *et al.* (2001) compared artificial neural networks and traditional methods including Winters exponential smoothing, Box-Jenkins ARIMA model and multivariate regression. The results indicated that on average ANNs fare favorably in relation to the more traditional statistical methods, followed by the Box-Jenkins model. The derivative analysis has showed that the neural network model is able to capture the dynamic nonlinear trend and seasonal patterns, as well as the interactions between them.

In the 1980s, the overall impression was that for immediate and short-term forecasts ARIMA models provide more accurate forecasts than other econometric models (O'Donovan, 1983). This perception was reaffirmed when more recently, Dugan *et al.* (1994) showed that the ARIMA model forecasted income statement items more accurately than Census X-11 and random walk models.

For time series with a long history, Box-Jenkins and ANNs provided comparable results (Sharda and Pati, 1992; Tang *et al.*, 1990). The ARIMA model is the same or superior to ANNs in terms of MAPE in a variety of applications. Despite its old tradition, the Box-Jenkins approach is a formidable competitor in the forecasting area.

Support Vector Machines (SVM), a more recent learning algorithm that has been developed from statistical learning theory (Vapnik, 1995; Vapnik *et al.*, 1997), has a very strong mathematical foundation and has been previously applied to time series analysis (Mukherjee *et al.*, 1997; Ruping and Morik, 2003).

We use two machine learning techniques, including Artificial Neural Network (ANN) and Support Vector Machines (SVM) for forecasting long-term demand. As benchmarks of comparison to machine learning techniques, we forecast same data with traditional time series forecasting methods, including moving average, exponential smoothing and exponential smoothing with trend.

At the end of this study, Robustness of presented forecasting methods is examined using a new data set. It is shown that machine learning techniques can forecast more accurate than traditional time series forecasting methods.

ARTIFICIAL NEURAL NETWORKS

Artificial Neural Networks (ANNs) are a class of generalized nonlinear nonparametric models inspired by studies of the brain and nerve system. The comparative advantage of ANNs over more conventional econometric models is that they can model complex, possibly nonlinear relationships without any prior assumptions about the underlying data generating process (Hornik *et al.*, 1989, 1990; White, 1990). The data-driven nature of ANNs makes them appealing in time series modeling and forecasting. ANN models overcome the limitations of traditional forecasting methods, including misspecification, biased outliers, assumption of linearity and re-estimation (Hill *et al.*, 1996). ANNs have been shown to be universal approximators, a property which makes them attractive in most forecasting applications. In addition, ANNs are more parsimonious than linear subspace methods such as polynomial and trigonometric series in approximating unknown functions.

Despite the many desirable features of ANNs, constructing a good network for a particular application is a non-trivial task. It involves choosing an appropriate architecture (the number of layers, the number of units in each layer and the connections among units), selecting the transfer functions of the middle and output units, designing a training algorithm, choosing initial weights and specifying the stopping rule.

It is widely accepted that a three-layer feed-forward network with an identity transfer function in the output unit and logistic functions in the middle-layer units can approximate any continuous function arbitrarily well given sufficient amount of middle-layer units (White, 1990). Thus, the network used in this research is a three-layer feed-forward one that is shown in Fig. 1.

The inputs (similar to the regressors used in the multivariate regression model) are connected to the output (similar to the regress and) via a middle layer. The network model can be specified as:

$$S_t = f(X_t, \alpha, \beta) + \varepsilon_t = \alpha_0 + \sum_{j=1}^n \alpha_j F\left(\sum_{i=1}^m \beta_{ij} x_{ij} + \beta_{0j}\right) + \varepsilon_t$$

where, S_t are the wholesaler sales at time t ; X is a vector of regressors that are exactly the same as the ones used in

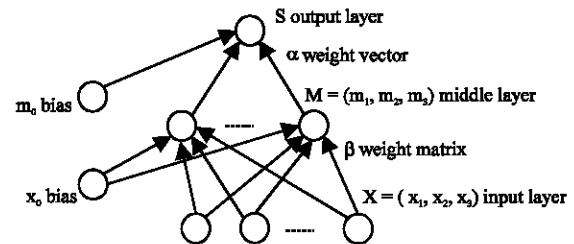


Fig. 1: A three-layer feedforward neural network

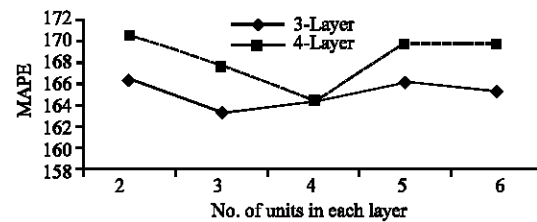


Fig. 2: MAPE index change trend graph

the multivariate regressions. n is the number of units in the middle layer; f is a logistic transfer function $F(a) = 1/(1 + \exp(-a))$, α_t represents a vector of coefficients (or weights) from the middle to output layer units and β_t represents a matrix of coefficients from the input to middle-layer units at time t . The details of the specification and estimation of our ANN model is shown below:

- Initial parameter values: The initial values of α_t and β_t are generated using a uniform distribution. Because of high records in the data set, it is not very important to use sophisticated method in order to generate initial value for α_t and β_t .
- Training algorithm: The ANN network is trained using the Levenberg-Marquardt's algorithm which has been found to be the fastest method for training moderate-sized feed-forward neural networks of up to several hundred weights (Demuth and Beale, 1997)
- Number of middle-layer units

Although we can use Bayesian regularization in the training algorithm in order to measure how many network parameters are being effectively used by the network regardless of the total number of parameters in the network (MacKay, 1992), but we experimented with several different numbers of middle-layer units and number of layers and found that the final number of middle-layer units and final number of layer are set to three and three, respectively. Results of each case MAPE index are shown in Fig. 2.

SUPPORT VECTOR MACHINES

Support Vector Machines (SVMs) are a newer type of universal function approximators that are based on the structural risk minimization principle from statistical learning theory (Vapnik, 1995) as opposed to the empirical risk minimization principle on which neural networks and linear regression, to name a few, are based. The objective of structural risk minimization is to reduce the true error on an unseen and randomly selected test example as opposed to NN and MLR, which minimize the error for the currently seen examples. Support vector machines project the data into a higher dimensional space and maximize the margins between classes or minimize the error margin for regression. Margins are soft, meaning that a solution can be found even if there are contradicting examples in the training set. A complexity parameter permits the adjustment of the number of error versus the model complexity and different kernels, such as the Radial Basis Function (RBF) kernel, can be used to permit non-linear mapping into the higher dimensional space.

The technique corresponds to minimization of the following function:

$$F(f) = \frac{C}{N} \sum_{i=1}^N |Y_i - f(x_i)|_\epsilon + \frac{1}{2} \|f\|^2$$

Where:

$$|Y_i - f(x_i)|_\epsilon = \begin{cases} 0 & \text{if } |Y_i - f(x_i)| < \epsilon \\ |Y_i - f(x_i)| & \text{otherwise} \end{cases}$$

is Vapnik's ϵ Insensitive Loss Function (ILF). An important point is that this function assigns zero loss to errors less than, ϵ thus safeguarding against over-fitting. In other words, this function doesn't fit a crisp value but instead fits a tube with radius ϵ to the data. This is similar to fuzzy description of the function. The other important aspect of this loss function that it minimizes a least modulus but not least squares. The ϵ parameter also plays an important role by providing a sparse representation of the data, as we shall see later.

Vapnik (1998) showed that the minimizer of the objective function under very general conditions can be written as:

$$f(x) = \sum_{i=1}^N c_i K(x, x_i)$$

where, c_i are the solution of a quadratic problem. X is equal to (x, x_i) is the so-called kernel function that defines the generalized inner product and is a commonly

used tool to perform nonlinear mapping. Several choices for the kernel function are available, such as Gaussian, sigmoid, polynomial, splines. The only item defined by the data is the coefficients c_i which are obtained by maximizing the following quadratic form:

$$\text{Min}(E(c)) = \frac{1}{2} \sum_{i,j=1}^N c_i c_j K(x_i, x_j) - \sum_{i=1}^N c_i Y_i + \epsilon \sum_{i=1}^N |c_i|$$

Subject to the constraints:

$$\sum_{i=1}^N c_i = 0, -\frac{C}{N} \leq c_i \leq \frac{C}{N}, i=1 \dots N$$

where, parameters C and ϵ are regularization parameters that control the flexibility or complexity of the SVM. These parameters should be selected by user using resampling or other standard techniques. However, it should be emphasized that, in contrast to the classical regularization techniques, a clear theoretical understanding of C and ϵ is still missing and is a subject of theoretical as well as experimental efforts.

Here, an appropriate kernel function is found for SVM that it can forecast with least error. In order to achieve to this aim, 4 different kernel function, including linear, polynomial, RBF and sigmoid, are examined. Errors of these forecasts are measured as MAPE index.

Based on result presented in Fig. 3, the best type of kernel function for this kind of data is linear. The output of different kernel functions and their solving time are shown in Fig. 3 and 4 respectively.

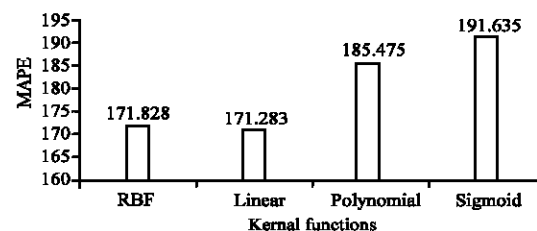


Fig. 3: Output of different kernel functions

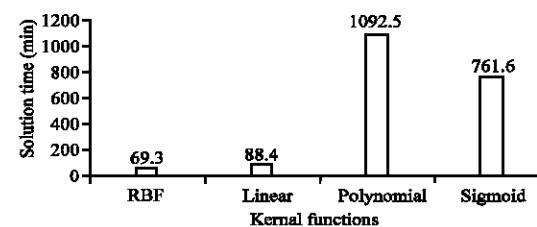


Fig. 4: Solution time of different kernel function

BENCHMARK STATISTICAL METHODS

Moving average: This method considers the average of n previous periods as a forecast for the next period. The problem is determining optimum value for n . In this research, we considered a range of values for n and then determined MAPE index in order to select best value of n . We found 150 as optimum value for n and MAPE =167.7526 consequently. Result of using this method is shown in Table 1.

Exponential smoothing: These models use a weighted average of past values, in which the weights decline geometrically over time to suppress short-term fluctuations in the data. The following formula is used to forecast:

$$F_{t+1} = F_t + \alpha(A_t - F_t)$$

Where:

F_{t+1} = Forecasted demand at time $t+1$

A_t = Real demand at time t

By using this method on data set, the demand was forecasted and found that the best value for α is 0.010.

With this value, forecasting error was measured and determined 167.797 as MAPE index in the best combination (Table 2).

Table 1: Result of using moving average technique

n	MAPE	n	MAPE
2	179.2480	100	167.954
3	178.9230	150	167.753
4	176.9800	200	168.004
5	175.3200	250	167.990
10	171.9890	300	168.631
20	172.1426	400	168.978
30	170.8380	500	169.527
40	169.4190	600	169.255
50	168.7600	700	168.849
60	167.9443	800	169.350
70	168.0770	900	170.012
80	168.1070	1000	170.023

Table 2: Result of using exponential smoothing technique

α	MAPE	α	MAPE
0.001	167.850	0.150	170.510
0.002	168.092	0.200	170.775
0.003	168.045	0.250	171.092
0.005	167.850	0.300	171.438
0.010	167.797	0.350	171.793
0.020	168.098	0.400	172.161
0.030	168.492	0.450	172.553
0.040	168.841	0.500	172.953
0.050	169.154	0.550	173.333
0.100	170.105	0.600	173.700

Exponential smoothing with trend: The simple exponential smoothing method may provide an adequate future forecast if no trend, seasonal or cyclical effects exit. If a trend is present, the method can be extended to adjust to variable, the average level and the trend level. With assumption of existence of trend in data series, formula is used and then measured MAPE index in order to determining α^* and β^* .

$$T_{t+1} = \beta(F_{t+1} - F_t) + (1 - \beta)T_t, F_{t+1} = F_t + \frac{1}{\alpha}T_{t+1}$$

Where:

T_{t+1} = Demand trend at period t_{t+1}

F_{t+1} = Forecasted demand at time t_{t+1}

A_t = Real demand at time t

F_{t+1} = Forecasted demand with trend consideration at period t_{t+1}

Using this method, the best value for α and β were found 0.02 and 0.05 respectively. In this combination, forecasting error is determined 170.081 as MAPE index (Table 3).

Comparisons and models validity: In order to assess stability of proposed methods, they are tested on raw data sets and MAPE index is calculated for their results again. The least amount of MAPE which is resulted from each method is shown in Table 4. Results show that proposed ANN in this research can be applied in more efficient way than previous classic methods for forecasting demand in supply chain.

Table 3: Result of using exponential smoothing with trend technique

α	β				
	0.05	0.10	0.15	0.20	0.25
0.01	170.116	171.228	174.940	180.873	187.036
0.02	170.081	173.813	177.830	181.227	186.638
0.03	170.950	174.481	178.418	184.865	191.736
0.04	171.617	175.199	180.652	186.577	191.122
0.05	172.122	176.193	181.492	186.049	190.122
0.10	173.253	176.452	179.242	181.738	184.775
0.15	173.131	175.317	177.489	180.427	184.093
0.20	173.020	174.889	177.332	180.467	183.921
0.25	173.149	175.046	177.605	180.744	184.042

Table 4: Comparison and models validity by MAPE index

Forecasting techniques	Testing data set (rank)	Training data set (rank)
Moving average	167.753 (2)	180.604 (3)
Exponential smoothing	167.797 (3)	179.791 (4)
Exponential smoothing with trend	170.081 (4)	187.051 (5)
Support vector machines	171.282 (5)	178.540 (2)
Artificial neural network	163.310 (1)	169.141 (1)

CONCLUSIONS

In this study, some machine learning techniques, including artificial neural network and support vector machines, are used to forecast demand in supply chain. The process contained two steps. In the first step, an artificial neural network with three layers and three middle units was trained by using sensitivity analysis, then 4 different kernel function was used for finding best kernel function and parameter combination in SVM algorithm, then three traditional forecasting methods was used to forecast. Forecasting errors measured by using MAPE index in all methods. Results showed that artificial neural network can forecast precisely better than other methods. Best parameter combination of each method is used to comparison and model validity in the next step. In the second step, efficiency of proposed models measured with raw data. Results showed again that artificial neural network can forecast precisely in comparison with other traditional forecasting methods and SVM.

Rank of each method in both training and testing steps shows that results of SVM method are improved so that its rank is 2, after ANN method. This finding shows that structure of data in testing set has changed probably so that linear kernel function could forecast better than past.

REFERENCES

- Alon, I., Q. Min and R. Sadowski, 2001. Forecasting aggregate retail sales: A comparison of artificial neural networks and traditional method. *J. Retailing Consumer Serv.*, 8: 147-156.
- Croston, J.D., 1972. Forecasting and stock control for intermittent demands. *Operat. Res. Q.*, 23: 289-303.
- Demuth, H. and M. Beale, 1997. *Neural Network Toolbox User's Guide, Version 3.0*. The Math Works, Inc., pp: 5-35.
- Dugan, M., K.A. Shriver and A. Peter, 1994. How to forecast income statement items for auditing purposes. *J. Business Forecast.*, 13: 22-26.
- Foster, B., F. Collopy and L. Ungar, 1992. Neural network forecasting of short, noisy time series. *Comput. Chemical Eng.*, 16: 293-297.
- Franses, P.H. and G. Draisma, 1997. Recognizing changing seasonal patterns using artificial neural networks. *J. Econometrics*, 81: 273-280.
- Granger, C.W.J., 1996. Can we improve the perceived quality of economic forecasts. *J. Applied Econometrics*, 11: 455-473.
- Hilas, C.S., S.K. Goudos and J.N. Sahalos, 2006. Seasonal decomposition and forecasting of telecommunication data: A comparative case study. *Technological Forecast. Social Change*, 73: 495-509.
- Hill, T., M. O'Connor and W. Remus, 1996. Neural network models for time series forecasts. *Manage. Sci.*, 42: 1082-1092.
- Horik, K., M. Stinchcombe and H. White, 1989. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2: 359-366.
- Horik, K., M. Stinchcombe and H. White, 1990. Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. *Neural Network*, 3: 551-560.
- Jeong, B., H. Jung and N. Park, 2002. A computerized causal forecasting system using genetic algorithms in supply chain management. *J. Syst. Software*, 60: 233-237.
- Krycha, K.A. and U. Wagner, 1999. Applications of artificial neural networks in management science: A survey. *J. Retailing Consumer Serv.*, 6: 185-203.
- MacKay, D.J.C., 1992. Bayesian interpolation. *Neural Computation*, 4: 415-447.
- Mukherjee, S., E. Osuna and F. Girosi, 1997. Nonlinear prediction of chaotic time series using support vector machines. Paper presented at the Proceedings of IEEE NNSP.
- Nelson, M., T. Hill, B. Remus and M. O'Connor, 1994. Can neural networks be applied to time series forecasting learn seasonal patterns: An empirical investigation. Proceedings of the 27 Annual Hawaii International Conference on System Sciences, Jan 4-7, Information Systems: Decision Support and Knowledge-Based Systems, pp: 649-655.
- O'Donovan, T.M., 1983. *Short Term Forecasting: An Introduction to the Box-Jenkins Approach*. 1st Edn., Wiley, New York.
- Peterson, R.T., 1993. Forecasting practices in the retail industry. *J. Business Forecast.*, 12: 11-14.
- Qi, M., 1996. Financial Applications of Artificial Neural Networks. In: *Handbook of Statistics, Statistical Methods in Finance*, Maddala, G.S., Rao, C.R. (Eds.). North-Holland, Elsevier Science Publishers, Amsterdam, pp: 529-552.
- Qi, M., 1999. Nonlinear predictability of stock returns using financial and economic variables. *J. Business Econ. Statis.*, 17: 419-429.
- Qi, M. and G.S. Maddala, 1999. Economic factors and the stock market: A new perspective. *J. Forecast.*, 18: 151-166.
- Ruping, S. and K. Morik, 2003. Support vector machines and learning about time. Paper presented at the Proceedings of ICASSP.
- Sharda, R. and R. Patil, 1992. Connectionist approach to time series prediction: An empirical test. *J. Intelligent Manufact.*, 3: 317-323.

- Tang, Z., C. De Almeida and P. Fishwick, 1990. Time series forecasting using neural networks vs. Box-Jenkins methodology. *Simulation*, 57: 303-310.
- Vapnik, V., 1995. *The Nature of Statistical Learning Theory*. 1st Edn., Springer, New York.
- Vapnik, V., S.E. Golowich and A. Smola, 1997. Support vector method for function approximation, regression estimation and signal processing. *Adv. Neural Inform. Process. Syst.*, 9: 287-291.
- Vapnik, V.N., 1998. *Statistical Learning Theory*. 1st Edn. John Wiley and Sons, New York, ISBN: 0471030031.
- White, H., 1990. Connectionist nonparametric regression: Multilayer feedforward networks can learn arbitrary mappings. *Neural Networks*, 3: 535-549.
- Willemain, T.R., C.N. Smart and H.F. Schwarz, 2004. A new approach to forecasting intermittent demand for service parts inventories. *Int. J. Forecast.*, 20: 375-387.
- Zhang, G.P., B.E. Patuwo and M.Y. Hu, 1998. Forecasting with artificial neural networks: The state of the art. *Int. J. Forecast.*, 14: 35-62.