World Academy of Science, Engineering and Technology
International Journal of Industrial and Manufacturing Engineering
Vol:3, No:1, 2009

Validation and Selection between Machine Learning Technique and Traditional Methods to Reduce Bullwhip Effects: a Data Mining Approach

Hamid R. S. Mojaveri, Seyed S. Mousavi, Mojtaba Heydar, and Ahmad Aminian

***Abstract*—**The aim of this paper is to present a methodology in three steps to forecast supply chain demand. In first step, various data mining techniques are applied in order to prepare data for entering into forecasting models. In second step, the modeling step, an artificial neural network and support vector machine is presented after defining Mean Absolute Percentage Error index for measuring error. The structure of artificial neural network is selected based on previous researchers' results and in this article the accuracy of network is increased by using sensitivity analysis. The best forecast for classical forecasting methods (Moving Average, Exponential Smoothing, and Exponential Smoothing with Trend) is resulted based on prepared data and this forecast is compared with result of support vector machine and proposed artificial neural network. The results show that artificial neural network can forecast more precisely in comparison with other methods. Finally, forecasting methods' stability is analyzed by using raw data and even the effectiveness of clustering analysis is measured.

***Keywords*—**Artificial Neural Networks (ANN), bullwhip effect, demand forecasting, Support Vector Machine (SVM).

I. Introduction

THIS study is concerned with forecasting an Iranian's automobile 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.

H. R. S. Mojaveri is a master student of Executive Management at the Faculty of Management and Economics, Islamic Azad University, Science and Research Branch (email: mazisalmani@yahoo.com).

S. S. Mousavi is a master student of Industrial Engineering at Department of Industrial Engineering, Graduate School of Engineering, Islamic Azad University, Tehran South Branch. He is also a member of Young Researchers Club, Tehran, Iran (e-mail: mousavi9293@ yahoo.com).

M. Heydar is a master student of Industrial Engineering at Department of Industrial Engineering, Graduate School of Engineering, Islamic Azad University, Tehran South Branch. He is also a member of Young Researchers Club, Tehran, Iran (corresponding author to provide phone: 98-0912-6938359;e-mail: mh_ie8051@yahoo.com).

A. Aminian is an Industrial Engineer graduated form Islam Azad University, Gachsaran Branch, gachsaran, Iran (email: a.aminian@gmail.com).

These methods are chosen because of their ability to model trend and seasonal fluctuations present in suppliers' data. The objectives of this article are three-fold: (1) to show how we can use data mining techniques in order to preparation of supply chain data (2) to show how to forecast wholesaler sales using ANNs and SVMs, and (3) to display how various time-series forecasting methods are compared in their forecasting accuracy of wholesaler sales.

The reasons for this article are both theoretical and practical. Theoretically speaking, how to improve the quality of forecasts is still an outstanding question [1]. 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, Halt 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. Reference [2] suggested that ANNs can 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, reference [3] 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.

II. Literature Review

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. References [4] and [5], for example, 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 (See [6] for a comprehensive survey of ANN applications in finance).

ANNs have also been increasingly used in management, marketing and retailing. The reader is referred to [7] for a comprehensive survey of ANN applications in management and marketing. In the literature a comprehensive review of ANNs' usage in forecasting is provided [8] provided.

Reference [9] has 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.

Reference [10] 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 and Croston's method.

Forecasting models for the monthly outgoing telephone calls in a University Campus is used [11]. Three different methods, namely the Seasonal Decomposition, Exponential Smoothing Method and SARIMA Method, have been used. Forecasts with 95% confidence intervals were calculated for each method and compared with the actual data.

Despite the great potential of ANNs in time-series forecasting, the empirical findings, thus far, are somewhat mixed. In comparing ANNs and ARIMA models on the 50 M-competition series that are designated as most appropriate for the Box-Jenkins technique, it is found that although the ARIMA model has a superior or equivalent mean absolute percentage error (MAPE) to that of the ANNs, the forecast error for the ANNs is lower when trend and seasonal patterns are in the data [12]. Reference [13] 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, it is found that ANNs do not model the seasonal fluctuations in the data very well [14]. Reference [15] 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. As shown in [16], 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. Using aggregate level data, and reference [17] found that the Winters model forecasts aggregate retail sales more accurately then simple exponential and Holt's models. The Winters model was shown to be a robust model that can accurately forecast individual product sales, company sales, income statement items, and aggregate retail sales. Comparison between artificial neural networks and traditional methods including Winters exponential smoothing, Box-Jenkins ARIMA model and multivariate regression has been also addressed in the literature [18]. 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 [19]. This perception was real- firmed when more recently work showed that the ARIMA model forecasted income statement items more accurately than Census X-11 and random walk models [16].

For time series with a long history, Box-Jenkins and ANNs provided comparable results ([20], [21]). Reference [12] also showed that 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 ([22], [23]), has a very strong mathematical foundation, and has been previously applied to time series analysis ([24], [25]).

In the following sections, we use some data mining techniques in order to data preparation at first, and then 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 paper, we examine robustness of presented machine learning models using a new data set and

show the efficiency of them in comparison of traditional time series forecasting methods.

III. Data Preparation

In this step, a two phased approach is applied. In the first phase, regarding the characteristics of available data and objective of forecasting, proper data mining techniques are used for data preparation purposes, and subsequently, these just prepared data are compared using three traditional forecasting technique and machine learning techniques to forecast the future data and finally the result are measured by an index known as MAPE. It is worth mentioning that stability of the forecasting methods has been evaluated using unprepared data.

A. Data Preparation

In data mining projects, data preparation is the most important and time consuming task. The better the data are prepared, the more accurate the result will be. If this task is not well done, the project will be fallen through [26]. In what follows steps needed to prepare data will be briefly stated.

Identifying variables that affect the customer demand. In this step, having considered the data characteristics in recorded data set and by applying experts' suggestions collected using brain storming method, 6 independent variables versus single variable, i.e. costumer demand, are identified, then importance of each variable is measured by ANOVA technique. Finally, 4 independent important variables are identified.

Cluster Analysis. Amongst all identified variables, one variable shows the distance between retailer and wholesaler, that based on data characteristics and experts' suggestions, it is put forwarded that the just-mentioned variable using cluster analysis is considered in modeling process in the form of categorical variable. The characteristics of cluster analysis will be outlined in the following.

The firsts steps is clustering algorithm that derived from Ward's methods in which distance from two clusters are measured based on sum of squares of distance between every two variable in each cluster. In this step sum of squares of clusters distance are minimized and two clusters from previous step are combined.

This method for 9 cases (2 to 10 clusters) has been done using SPSS v.12 and, then, for each case, Pseudo-F index is measured. This index, introduced by reference [27], is defined as ratio of mean sum of square between groups (MSSBG) and mean sum of square within groups (MSSWG). For this purpose, for each 9 cases the two above-mentioned measures are calculated and Pseudo-F index will be given and tabulated in the Table I.

Since this index shows sum of squares between groups to within groups, the higher value shows the less similarity between clusters and vice versa. Based on this result, the optimum number of cluster is 8.

TABLE I
Results of Using Cluster Analysis Technique

No. of Clusters	MSSBG	MSSWG	Pseudo-F
2	835.812	153.378	5.449
3	617.245	95.237	6.481
4	719.853	80.016	8.996
5	615.608	64.554	9.536
6	589.656	45.617	12.926
7	676.436	30.858	21.921
8	624.134	22.478	**27.766**
9	616.996	25.365	24.324
10	581.449	35.453	16.401

IV. Artificial Neural Networks (ANNs)

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 ([28] - [30]). 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 [13]. They 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 [30]. 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

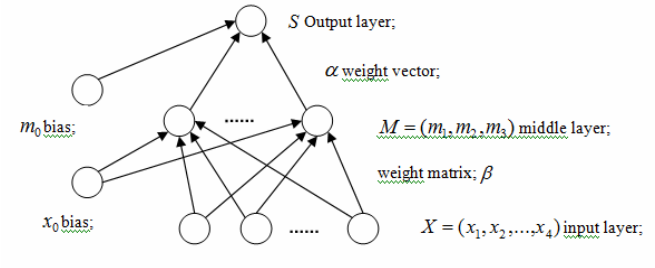

Fig. 1 A three layer feedforward neural network

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

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 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 listed below:

(1) *Initial parameter values*: The initial values of α_t and β_t are generated using a uniform distribution. Because of high records in our data set, it is not very important to use sophisticated method in order to generate initial value for α_t and β_t .

(2) *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 [31].

(3) *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 [32], but we experimented with several different numbers of middle-layer units and number or 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 are tabulated as MAPE index in Table II and are shown in Fig. 2.

TABLE II
OUTPUT OF EXAMINED ANNS AS MAPE INDEX

		Number of unit in each hidden layer				
		2	**3**	**4**	**5**	**6**
Number of layer	**3**	166.417	**163.310**	164.374	166.172	165.386
	4	170.528	167.560	164.401	169.778	169.752

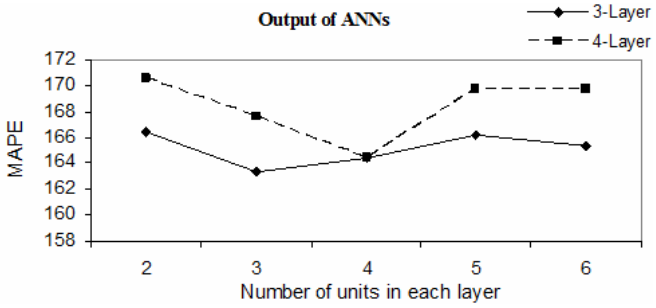

Fig. 2 MAPE index change trend graph

V. SUPPORT VECTOR MACHINES (SVMS)

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 as opposed to the empirical risk minimization principle on which neural networks and linear regression, to name a few, are based [22]. 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:

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

Where

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

is Vapnik's $\varepsilon - Insensitive\ Loss\ Function\ (ILF)$. 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. In [33] it is shown that the minimizer of the objective function under very general conditions can be written as:

$$f(x) = \sum_{i=1}^{N} c_i K(x, x_i) \quad (3)$$

where c_i is the solution of a quadratic problem. $K(x, x_i)$ 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:

$$\text{MIN} \quad (E(c)) = \frac{1}{2}\sum_{i=1}^{N} c_i c_j K(x_i, x_j) - \sum_{i=1}^{N} c_i Y_i + \varepsilon \sum_{i=1}^{N} |c_i| \quad (4)$$

Subject to

$$\sum c_i = 0, -\frac{C}{N} \le c_i \le \frac{C}{N}, i = 1, \ldots, N \quad (5)$$

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.

The aim of this section of research is finding an appropriate kernel function 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. Results of using these kernel functions are tabulated in Table III.

TABLE III
RESULTS OF USING DIFFERENT KERNEL FUNCTIONS IN SVM

Kernel function	MAPE
RBF	171.828
Linear	171.283
Polynomial	185.475
Sigmoid	191.635

Based on presented result in above table, 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.

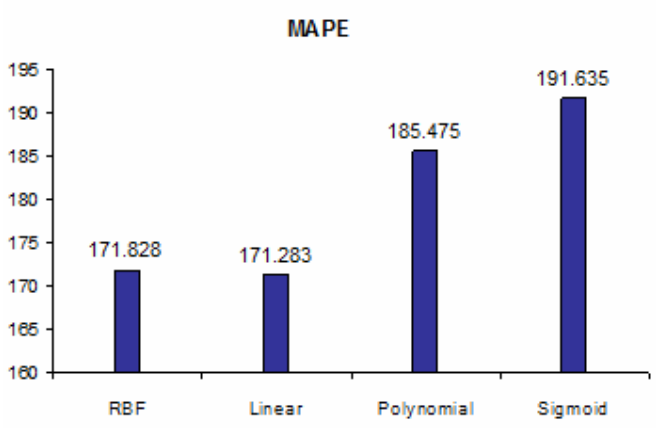

Fig. 3 Output of different kernel function

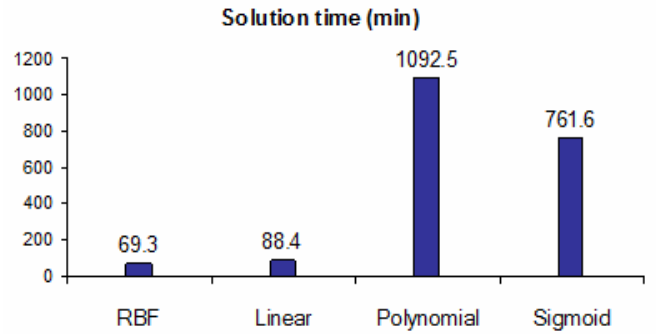

Fig. 4 Solution time of different kernel function

TABLE IV
RESULTS OF USING MOVING AVERAGE TECHNIQUE

n	MAPE	n	MAPE
2	179.248	100	167.954
3	178.923	**150**	**167.753**
4	176.980	200	168.004
5	175.320	250	167.990
10	171.989	300	168.631
20	172.1426	400	168.978
30	170.838	500	169.527
40	169.419	600	169.255
50	168.760	700	168.849
60	167.9443	800	169.35
70	168.077	900	170.012
80	168.107	1000	170.023

VI. BENCHMARK STATISTICAL METHODS

A. 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 article, 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. Results of using this method are shown in the following table, Table IV.

B. Exponential Smoothing

Historically, exponential smoothing models were used by approximately 13% of industry [34]. 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.

We used the following formula to forecast:

$$F_{t+1} = F_t + \alpha(A_t - F_t) \quad (6)$$

where F_{t+1} is defined as forecasted demand at time $t+1$, and A_t is real demand at time t. After using this method on

data set, we forecasted demand value and found that the best value for α is 0.010.

With this value, we measured forecasting error and determined 167.797 as MAPE index in the best combination. Result of using this method is shown in the following table, Table V.

TABLE V
RESULTS OF 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

C. 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 tow variable, the average level and the trend level. With assumption of existence of trend in data series, we used following formulas and then measured MAPE index in order to determining α^* and β^*.

$$T_{t+1} = \beta(F_{t+1} - F_t) + (1-\beta)T_t \quad (7)$$

$$FT_{t+1} = F_{t+1} \frac{1}{\alpha} T_{t+1} \quad (8)$$

where T_{t+1} is demand trend at period $t+1$, FT_{t+1} is forecasted demand with trend consideration at period $t+1$, and F_{t+1} and A_t are defined as before.

By using this method, we found that the best value for α and β is 0.02 and 0.05 respectively. In this combination, forecasting error is determined 170.081 as MAPE index. Result of using this method is shown in the following table, Table VI.

TABLE VI
RESULTS OF EXPONENTIAL SMOOTHING WITH TREND TECHNIQUE

		β				
		0.05	**0.10**	**0.15**	**0.20**	**0.25**
α	**0.01**	170.116	171.228	174.94	180.873	187.036
	0.02	**170.081**	173.813	177.83	181.227	186.638
	0.03	170.95	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	173.02	174.889	177.332	180.467	183.921
	0.25	173.149	175.046	177.605	180.744	184.042

VII. 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 VII. 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 VII
COMPARISON AND MODEL 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)

VIII. CONCLUSION

In this paper, some machine learning techniques, including artificial neural network and support vector machines, are used to forecast demand in supply chain. The process contained three steps. In the first step, data mining appropriate techniques selected based on nature of data and forecasting object and then data prepared for entering into second step. In the second step, an artificial neural network with three layers and three middle units was selected by using sensitivity analysis, then 4 different kernel function was used for finding best kernel function and parameter combination in SVM algorithm, then 3 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 third 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

[1] C. W. J. Granger, "Can we improve the perceived quality of economic forecasts?", *J. Appl. Econom.*,. Vol. 11, 1996, pp. 455-473.
[2] P. H. Franses, G. Draisma, "Recognizing changing seasonal patterns using artificial neural networks", *J. Econometrics*, Vol. 81, 1997, pp. 273-280.
[3] R. T. Peterson, "Forecasting practices in the retail industry", *J. Bus. Forec.*, vol. 12, 1993, pp. 11-14.

[4] M. Qi, G. S. Maddala, "Economic factors and the stock market: a new perspective", *J. Forec.*, vol. 18 no. 3, 1999, pp. 151-166.
[5] M. Qi, "Nonlinear predictability of stock returns using financial and economic variables", *J. Bu. Econom. Statist.*, vol. 17, no. 4, 1999, pp. 419-429.
[6] M. Qi, "Financial applications of artificial neural networks", in *Handbook of Statistics, Statistical Methods in Finance*, vol. 4, G. S. Maddala, C. R. Rao, Ed. North-Holland, Elsevier Science Publishers, Amsterdam, 1996, pp. 529-552.
[7] K. A. Krycha, U. Wagner, "Applications of artificial neural networks in management science: a survey", *J. Retail. Cons. Ser.*, vol. 6, 1999, pp. 185-203.
[8] G. P. Zhang, B. E. Patuwo, M. Y. Hu, "Forecasting with artificial neural networks: the state of the art", *Intl. J. Forec.*, vol. 14, no. 1, 1998, pp. 35-62.
[9] B. Jeong, H. Jung, N. Park, "A computerized causal forecasting system using genetic algorithms in supply chain management", *J. Sys. Software*, vol. 60, 2002, pp. 223–237.
[10] T. R. Willemain, C. N. Smart, H. F. Schwarz, "A new approach to forecasting intermittent demand for service parts inventories", *Intl. J. Forec.*, vol. 20, 2004, pp. 375– 387.
[11] C. S. Hilas, S. K. Goudos, J. N. Sahalos, "Seasonal decomposition and forecasting of telecommunication data: A comparative case study", *Technol. Forec. Soc. Change*, vol. 73, 2006, pp. 495–509.
[12] S. Kang, "An investigation of the use of feedforward neural networks for forecasting", Ph.D. Dissertation, Kent State University, Kent, Ohio, 1991.
[13] T. Hill, M. O'Connor, W. Remus, "Neural network models for time series forecast", *Manag. Sci.*, vol. 42, no. 7, 1996, pp. 1082-1092.
[14] M. Nelson, T. Hill, B. Remus, M. O'Connor, "Can neural networks be applied to time series forecasting learn seasonal patterns: an empirical investigation", in *Proc. 27th Ann. Int. Conf. System Sciences*, Hawaii, 1994, pp. 649-655.
[15] B. Foster, F. Collopy, L. Ungar, "Neural network forecasting of short, noisy time series", *Comput. Chem. Eng.*, vol. 16, no. 12, 1992, pp. 293-297.
[16] M. Dugan, K. A. Shriver, A. Peter, "How to forecast income statement items for auditing purposes", *J. Bus. Forec.*, vol. 13, 1994, pp. 22-26.
[17] I. Alon, "Forecasting aggregate retail sales: the Winters' model revisited", in The 1997 Annual Proceedings, J. C. Goodale, Ed., Midwest Decision Science Institute, 1997, pp. 234-236.
[18] I. Alon, Q. Min, R. Sadowski, "Forecasting aggregate retail sales: a comparison of artificial neural networks and traditional method", *J. Retail. Cons. Ser.*, vol. 8, 2001, pp. 147-156.
[19] T. M. O'Donovan, *Short Term Forecasting: An Introduction to the Box-Jenkins Approach.* NY: New York, Wiley, 1983.
[20] R. Sharda, R. Patil, "Connectionist approach to time series prediction: an empirical test", *J. Intell. Manufac.*, vol. 3, 1992, pp.317-323.
[21] Z. Tang, C. de Almeida, P. Fishwick, "Time series forecasting using neural networks vs. Box-Jenkins methodology", *Simulation*, vol. 57, no. 5, 1990, pp.303-310.
[22] V. Vapnik, The Nature of Statistical Learning Theory. New York, Springer, 1995.
[23] V. Vapnik, S. E. Golowich, A. Smola, "Support vector method for function approximation, regression estimation, and signal processing", *Adv. Neural inf. Process. Syst.*, vol. 9, 1997, pp. 287–291.
[24] S. Mukherjee, E. Osuna, F. Girosi, "Nonlinear prediction of chaotic time series using support vector machines", in *Proc. IEEE NNSP*, 1997.
[25] S. Ruping, K. Morik, "Support Vector Machines and Learning about Time", in *Proc. of ICASSP*, 2003.
[26] J. Han, M. Kamber, *Data Preprocessing, Data Mining, concepts and techniques*, 2nd ed., Morgan Kaufman Publisher, 2006, ch2.
[27] T. Calinski, J. Harabasz, "A Dendrite Method for Cluster Analysis", *Commun. Statist.*, Vol. 3, 1974, pp. 1-27.
[28] K. Hornik, M. Stinchcombe, H. White, "Multilayer feedforward networks are universal approximators", *Neural Net.*, vol. 2, 1989, pp. 359-366.
[29] K. Hornik, M. Stinchcombe, H. White, "Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks", *Neural Net.*, vol. 3, 1990, pp. 551-560.
[30] H. White, "Connectionist nonparametric regression: multilayer feedforward networks can learn arbitrary mappings", *Neural Net.*, vol. 3, 1990, pp. 535-549.
[31] H. Demuth, M. Beale, *Neural Network Toolbox User's Guide, Version 3.0*. The Math Works, Inc., 1997, pp. 5-35.
[32] D. J. C. MacKay, "Bayesian interpolation", *Neural Comput.*, vol. 4, 1992, pp. 415-447.
[33] V. N. Vapnik, *Statistical Learning Theory*. NY: New York, John Wiley & Sons, 1998.
[34] G. L. Lilien, P. Kotler, *Marketing Decision Making: A Model Building Approach*. NY: New York, Harper and Row Publishers, 1983.