

# A Comparison of Various Forecasting Methods for Autocorrelated Time Series

Regular Paper

Karin Kandananond\*

Faculty of Industrial Technology, Rajabhat University Valaya-Alongkorn, Thailand

\* Corresponding author E-mail: kandananond@hotmail.com

Received 22 Apr 2012; Accepted 21 June 2012

DOI : 10.5772/51088

© 2012 Kandananond; licensee InTech. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Abstract** The accuracy of forecasts significantly affects the overall performance of a whole supply chain system. Sometimes, the nature of consumer products might cause difficulties in forecasting for the future demands because of its complicated structure. In this study, two machine learning methods, artificial neural network (ANN) and support vector machine (SVM), and a traditional approach, the autoregressive integrated moving average (ARIMA) model, were utilized to predict the demand for consumer products. The training data used were the actual demand of six different products from a consumer product company in Thailand. Initially, each set of data was analysed using Ljung-Box-Q statistics to test for autocorrelation. Afterwards, each method was applied to different sets of data. The results indicated that the SVM method had a better forecast quality (in terms of MAPE) than ANN and ARIMA in every category of products.

**Keywords** Artificial neural network (ANN), Autoregressive integrated moving average (ARIMA), Consumer products, Demand forecasting, Supply chain, Support vector machine (SVM)

## 1. Introduction

Demand forecasting is critical to improving the efficiency of a supply chain system. Since each party in the supply chain will process the order in response to the demand signal, the accuracy of demand forecasts will significantly improve the production scheduling, capacity planning, material requirement planning and inventory management. Without accurate forecasting, this scenario will lead to inefficiency of a supply chain system. Product demand is one of the most challenging types of time series to forecast because of its uncertainty. There were several attempts to identify the structure of this type of data and autocorrelation is one of these structures. Under uncorrelated conditions, the observations have a fixed mean and the fluctuation around the mean is the result of only random shock or white noise. However, when observations are autocorrelated, this scenario can be categorized into two cases: stationary and non-stationary. If process observations vary around a fixed mean and have a constant variance, this type of variability is called stationary behaviour. On the other hand, if a process mean drifts from a fixed value, this behaviour is called non-stationary. Since there are a number of forecasting methods to predict time series data efficiently, it is a good idea for practitioners to have the information regarding

the most appropriate forecasting technique under different situations, i.e., non-autocorrelation and autocorrelation.

## 2. Literature Review

According to the literature, there were three methods which were popular for forecasting time series data. Among these techniques were artificial neural network (ANN), support vector machine (SVM) and a traditional method, the Box Jenkins autoregressive integrated moving average (ARIMA) model. Since the performance of these approaches was still questionable, empirical study was always utilized as a basis to benchmark these techniques. Most popular sources of data used were industrial, financial and electrical demand data.

For industrial data, Bansal, Vadhavkar and Gupta [1] identified the inventory patterns of a large medical distribution organization and elaborated a method to construct and choose an appropriate neural network for optimizing the inventory. The implementation led to the reduction of the total inventory by 50% in the organization while the customer satisfaction level was still high. Hua, Wang, Xu, Zhang and Liang [2] utilized the SVM approach to forecast the demand for spare parts. The data used were spare parts from a petrochemical enterprise in China. It is obvious that the introduced method was better able to forecast the demand for spare parts than the traditional methods. Gutierrez, Solis and Mukhopadhyay [3] applied the ANN method to forecast lumpy demand and compared the performance of ANN to three traditional methods (single exponential smoothing, Croston's method and the Syntetos-Boylan approximation). The results showed that it outperformed those three methods significantly. Another popular type of data was a financial data. Tay and Cao [4] assessed the performance of SVM and ANN to forecast the financial time series. The historical data was based on five real future contracts collected from the Chicago mercantile market. The results indicated that the SVM method performed better than the ANN methods. Another study was conducted by Kim [5] applying the SVM approach to predict the stock price index and compared the performance with the ANN method. The study showed that the SVM approach outclassed the ANN method significantly. Similarly, Huang, Nakamori and Wang [6] also utilized the SVM to predict the NIKKEI 225 stock price index. The results also revealed that SVM was preferred to ANN, linear discriminant analysis and quadratic discriminant analysis.

Besides the industrial and financial demand, electricity load demand was also utilized in the empirical study to compare the performance of these three forecasting methods (SVM, ANN and ARIMA). The above study was

shown in the work by Pai and Hong [7]. The conclusion from this research indicated that the SVM method should be the preferred choice over the traditional ARIMA and ANN approach. Another study related to electricity demand was done to assess the performance of two statistical methods, linear regression and ARIMA, and the ANN model (Prybutok, Yi and Mitchell [8]) to forecast a set of time series. According to their research results, the ANN model outperformed the ARIMA method. Similarly, Ho, Xie and Goh [9] used the simulated failure time of a compressor in the study for determining the most efficient forecasting model. Two methods, ARIMA and ANN, were utilized to forecast the failure of the system.

In addition to the comparison of these methods, another interesting aspect of the study was the utilization of the autocorrelation structure as a basis to compare the performance of different forecasting methods. Lachtermacher and Fuller [10] utilized the Box-Jenkins model in terms of the lag component to specify the complexity of ANN structure. According to their study, each lag of the autocorrelation structure was deployed to represent a unit of input for ANN. Hwang [11] conducted a study to assess the performance of ANN when the process was stationary by using the ARMA model as a benchmark. This study led to a profound understanding of how the ANN performed at the different degrees of autocorrelation.

As a result, most of the studies were conducted empirically to compare the performance of the ANN, SVM and ARIMA methods. However, they did not focus on using a specific pattern of data to choose an appropriate forecasting method. In this research, sets of data with different patterns (non-autocorrelation and autocorrelation) were deployed to compare the performance of three popular methods, ANN, SVM and ARIMA. This aspect was crucial because it might enhance the forecasting capability by utilizing autocorrelation as a basis.

## 3. Methodology

Three methods used in this study are ANN, SVM and ARIMA.

### 3.1 ANN Method

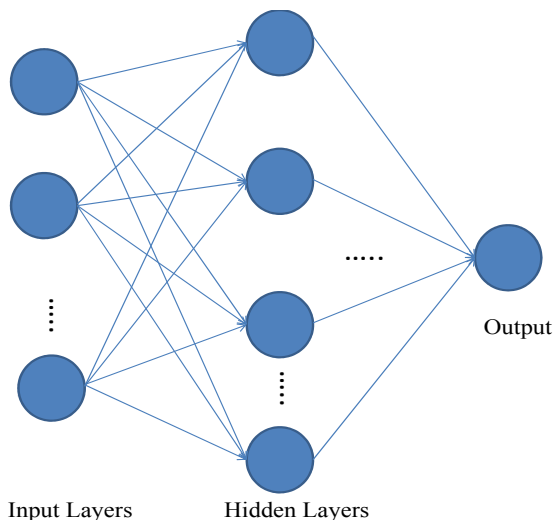
The development of ANN models was based on studying the relationship between input variables and output variables. Basically, the neural architecture consisted of three or more layers, i.e., input layer, output layer and hidden layer as shown in Fig. 1. The function of this network was described as follows:

$$Y_j = f\left(\sum_i w_{ij} X_{ij}\right) \quad (1)$$

where  $Y_j$  is the output of node  $j$ ,  $f(\cdot)$  is the transfer function,  $w_{ij}$  the connection weight between node  $j$  and node  $i$  in the lower layer and  $X_{ij}$  is the input signal from the node  $i$  in the lower layer to node  $j$ .

### 3.2 SVM Method

Support Vector Machine (SVM) was a classification method which was based on the construction of hyperplanes in a multidimensional space. As a result, it allowed different class labels to be differentiated. Normally, SVM was utilized for both classification and regression tasks, and it was able to handle multiple continuous and categorical variables.



**Figure 1.** The architecture of a neural network

The purpose of the regression task of SVM was to find a function  $f$  (such that  $y = f(x) + \text{noise}$ ) which was able to predict new cases. This was achieved by training the SVM model on a sample set, i.e., training set, a process that involved the sequential optimization of an error function. There were two types of SVM models for the regression purpose, type 1 and 2. For regression type 1, the objective function was the minimization of the error function.

$$\min \frac{1}{2} w^T w + C \sum_{i=1}^N \xi_i + \sum_{i=1}^N \xi_i^*$$

s.t.

$$\begin{aligned} w^T \phi(x_i) + b - y_i &\leq \varepsilon + \xi_i^* \\ y_i - w^T \phi(x_i) - b_i &\leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* &\geq 0, i = 1, \dots, N, \varepsilon \geq 0 \end{aligned}$$

Similarly, objective function of the regression type 2 was

$$\min \frac{1}{2} w^T w - C \left[ v\varepsilon + \frac{1}{N} \sum_{i=1}^N (\xi_i + \xi_i^*) \right]$$

The regression type 2 also shared the same constraint as the regression type 1. For the SVM model, there were four types of kernels ( $\phi$ ), linear, polynomial, radial basis function (RBF) and sigmoid. Among these kernels, RBF were the most frequently used kernels because of their localized and finite responses across the entire range of the real  $x$ -axis. The functions of these kernels were shown as follows:

$$\phi = \begin{cases} x_i * y_i & \dots \text{Linear} \\ (\gamma x_i X_j + \text{coefficient } t)^d & \dots \text{Polynomial} \\ \exp(-\gamma |X_i - x_j|^2) & \dots \text{RBF} \\ \tanh(\gamma x_i X_j + \text{coefficient } t) & \dots \text{Sigmoid} \end{cases}$$

### 3.3 ARIMA Method

For time series analysis, the ARIMA model was a stochastic difference equation that was frequently utilized to model stochastic disturbances. The general form of the ARIMA model is shown in equation (2).

$$\begin{aligned} \Delta_d Y_t &= \mu + \phi_1 \Delta_d Y_{t-1} + \phi_2 \Delta_d Y_{t-2} + \dots \\ &+ \phi_p \Delta_d Y_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \end{aligned} \quad (2)$$

The order of the ARIMA model was normally identified in the form of  $(p, d, q)$ ;  $p$  indicated the order of the autoregressive part while  $d$  was for the amount of difference and  $q$  for the order of the moving average part. Some specific forms of the ARIMA model were utilized to represent autocorrelated disturbances, e.g., autoregressive order one, ARIMA (1, 0, 0) or AR (1) for stationary disturbances, while for integrated moving average, ARIMA (0, 1, 1) or IMA (1, 1) was used to represent non-stationary disturbances.

## 4. Research Procedures

The actual data used in the empirical study was the monthly data of six different consumer products, cooking aids brand A, shower gel brand B, body lotion brand C, dishwashing liquid brand D, deodorant brand E and fabric detergent brand F, from January 2009 to August 2011 (32 cases) as shown in Fig. 2-7. After obtaining the data, they were analysed using the Ljung-Box-Q test. The analysis showed that all sets of data were categorized into two types, non-autocorrelation and autocorrelation. Additionally, the in-depth details of the autocorrelation analysis were elaborated in Table 1. According to the analysis, the time series data of three products (cooking aids, shower gel and body lotion) were autocorrelated, while another three categories (dishwashing liquid, deodorant and fabric detergent) showed no sign of autocorrelation structure. It should also be highlighted that the demand data for body lotion possessed the highest degree of autocorrelation.

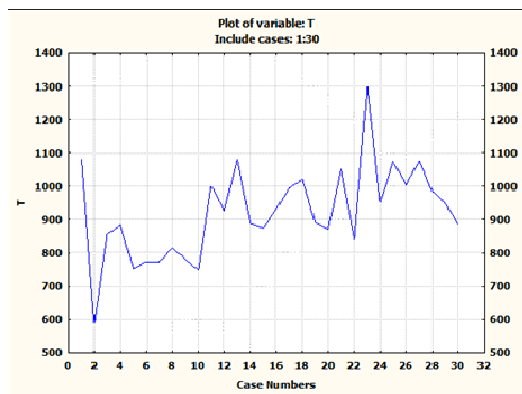


Figure 2. Cooking aid brand A demand

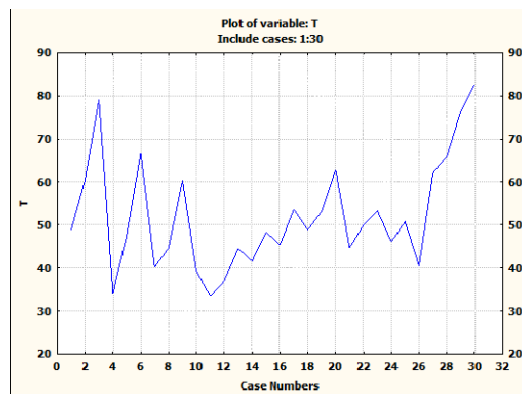


Figure 6. Deodorant brand E demand

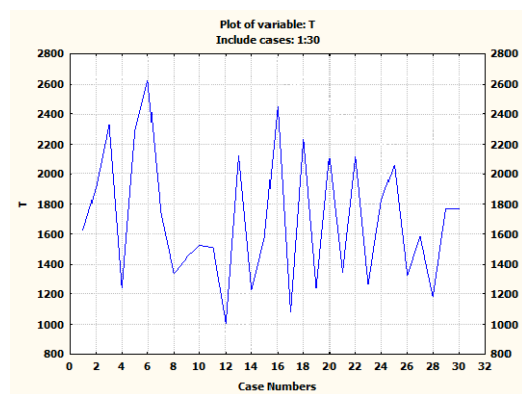


Figure 3. Shower gel brand B demand

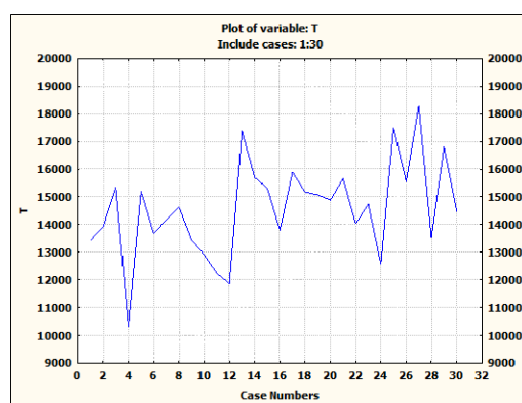


Figure 7. Fabric detergent brand F demand

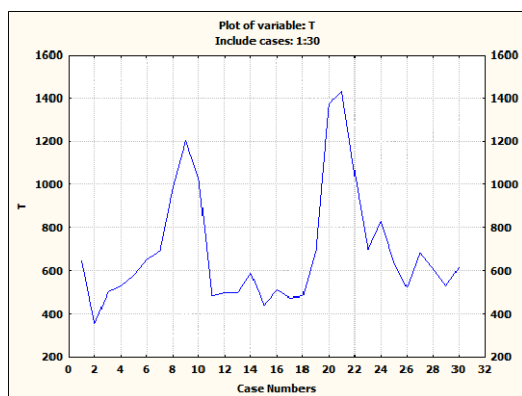


Figure 4. Body lotion brand C demand

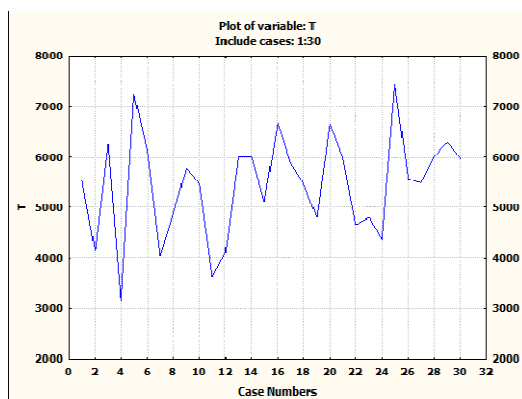


Figure 5. Dishwashing liquid brand D demand

Product Category	Correlation/Patterns
-Cooking aids brand A	Highly positive correlated (lag 2)
-Shower gel brand B	Highly positive correlated (lag 7)
-Body lotion brand C	-Highly positive and negative correlated (lag 1,5,6,7,8,11,12,13) -Seasonal pattern (cyclic)
-Dishwashing liquid brand D	Not correlated
-Deodorant brand E	Not correlated
-Fabric detergent brand F	Not correlated

Table 1. Data characteristics

After the tested data was chosen, three proposed methods, ANN, SVM and ARIMA, were utilized to construct models to forecast the demand for these six products using two statistical packages, STATISTICA and StatGraphics. The performance of these approaches towards the autocorrelated structure was justified by considering their error measurement, mean absolute percentage error (MAPE).

## 5. Results

The assessment of all methods was divided into three cases based on the methodology used:

### 5.1 ANN Method

The two most popular neural network architectures, multilayer perceptrons (MLP) and radial basis function (RBF), were utilized for the regression purpose. The inputs for training were the historical demand at  $t-1$ ,  $t-2$ , ...,  $t-10$ , while the top performing five networks were retained for each type of product. The network with the best performance was kept to forecast the demand for each category (time:  $t$ ). The results after applying ANN model are shown in Tables 2 and 3.

Product Category	Algorithm	Hidden Activation	Output Activation
-Cooking aids brand A	RBFT	Gaussian	Identity
-Shower gel brand B	BFGS 31	Exponential	Identity
-Body lotion brand C	BFGS 5	Logistic	Logistic
-Dishwashing Liquid brand D	BFGS 20	Tanh	Identity
-Deodorant brand E	RBFT	Gaussian	Identity
-Fabric detergent brand F	BFGS 3	Identity	Logistic

**Table 2.** Analysis results for ANN model

Product Category	ANN Model	MAPE
-Cooking aids brand A	RBF 10-6-1	0.069
-Shower gel brand B	MLP 10-12-1	0.019
-Body lotion brand C	MLP 10-9-1	0.23
-Dishwashing Liquid brand D	MLP 10-9-1	0.088
-Deodorant brand E	RBF 10-6-1	0.13695
-Fabric detergent brand F	MLP 10-5-1	0.07

**Table 3.** MAPE for ANN model

The results in Tables 2 and 3 indicate that the number of hidden layers ranged from 5 to 12 layers. According to MAPE, the MLP architecture might be suitable for autocorrelated and non-autocorrelated conditions. The results revealed that the appropriated ANN algorithm for most products (shower gel, body lotion, dishwashing liquid and fabric detergent) was the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm (with the number of cycles used to train the model ranged from 3 to 31). On the other hand, the RBF was preferred only for cooking aids and deodorant categories. The hidden neuron activation functions of the retained five networks were Gaussian, tangent hyperbolic (tanh), logistic, identity and exponential, while the exponential, identity (the activation of the neuron is passed on directly as the output) and logistic were assigned to the output neuron activation.

### 5.2 SVM Method

Similar to the ANN case, the indicators used for SVM application were the historical data at  $t-1$ ,  $t-2$ , ...,  $t-10$  to predict the demand at time  $t$ . The selected forecasting model based on the SVM approach was the regression type 1 with  $C=10.0$ ,  $\epsilon=0.1$  and the kernel was the radial basis function with  $\gamma=0.1$ . The number of support vectors and MAPE from the prediction for each category of products is illustrated in Table 4.

According to Table 4, the data with the highest level of autocorrelation (body lotion) needed fewer support vectors than the ones with fewer or no autocorrelation.

Product Category	Number of vectors	MAPE
-Cooking aids brand A	13	0.045
-Shower gel brand B	12	0.049
-Body lotion brand C	10	0.187
-Dishwashing liquid brand D	13	0.077
-Deodorant brand E	14	0.11436
-Fabric detergent brand F	12	0.05

**Table 4.** Analysis results and MAPE for the SVM model

### 5.3 ARIMA Method

A statistical package, StatGraphics Centurion version 10, was utilized to select the most appropriate ARIMA model for forecasting the demand for each product. The optimal models with their MAPEs are shown in Table 5.

Product Category	ARIMA model	MAPE
Cooking aids brand A	ARIMA (0, 2, 2)	0.0918
Shower gel brand B	ARIMA (0, 2, 2)	0.0926198
Body lotion brand C	ARIMA (2, 1, 1)	0.212748
Dishwashing liquid brand D	ARIMA (0, 1, 1)	0.172718
Deodorant brand E	ARIMA (2, 1, 0)	0.161938
Fabric detergent brand F	ARIMA (1, 1, 2)	0.0790694

**Table 5.** Analysis results and MAPE for ARIMA model

According to Table 5, the ARIMA model seemed to work really well for forecasting the demand for some products, i.e., cooking aids, shower gel and fabric detergent. However, it was important to note that the minimization of MAPE might not be related to the degree of autocorrelation. For example, the MAPE for body lotion was the highest even though the test indicated that its data was highly autocorrelated.



## 6. Conclusions

The results from the above section are summarized in Table 6. The conclusions indicate that the SVM outperformed the other two methods in almost every category of product (except shower gel where the ANN method dominated). Moreover, they also signified that the autocorrelation structure of data has no effect on the performance of the SVM or ANN method. Although the ARIMA model was based on the autocorrelation structure, it still had lower MAPE than the other two methods. However, the autocorrelation might affect the algorithm of the SVM method since the highest degree of autocorrelation caused the lowest number of supporting vectors.

Product Category	MAPE		
	ANN	SVM	ARIMA
Cooking Aids brand A	0.069	0.045	0.0918012
Shower Gel brand B	0.019	0.049	0.0926198
Body Lotion brand C	0.23	0.187	0.212748
Dishwashing Liquid brand D	0.088	0.077	0.172718
Deodorant brand E	0.13695	0.11436	0.161938
Fabric Detergent brand F	0.07	0.05	0.0790694

**Table 6.** Result comparison for the three models

## 7. References

- [1] K. Bansal, S. Vadhavkar and A. Gupta, "Neural Networks Based Data Mining Applications for Medical Inventory Problems," *Data Mining and Knowledge Discovery*, vol. 2(1), pp. 97-102, 1998.
- [2] Z. Hua, Y. Wang, X. Xu, B. Zhang and L. Liang, "Predicting Corporate Financial Distress on Integration of Support Vector Machine and Logistic Regression," *Expert Systems with Applications*, vol. 33(2), pp. 434-440, 2006.
- [3] R. S. Gutierrez, A. O. Solis and S. Mukhopadhyay, "Lumpy Demand Forecasting Using Neural Network," *International Journal of Production Economics*, vol. 111(2), pp. 409-420, 2008.
- [4] F. E. H. Tay and L. Cao, "Application of Support Vector Machines in Financial Time Series Forecasting," *Omega*, vol. 29, pp. 309-317, 2001.
- [5] K. Kim, "Financial Time Series Forecasting Using Support Vector Machines," *Neurocomputing*, vol. 56(1-2), pp. 307-319, 2003.
- [6] W. Huang, Y. Nakamori and S. Wang, "Forecasting Stock Market Movement Direction with Support Vector Machine," *Computer & Operations Research*, vol. 32(10), pp. 2513-2522, 2005.
- [7] P. Pai and W. Hong, "Forecasting Regional Electricity Load Based on Recurrent Support Vector Machines with Genetic Algorithms," *Electric Power Systems Research*, vol. 74(3), pp. 417-425, 2005.
- [8] V. R. Prybutok, J. Yi and D. Mitchell, "Comparison of Neural Network Models with ARIMA and Regression Models for Prediction of Houston's Daily Maximum Ozone Concentrations," *European Journal of Operational Research*, vol. 122(1), pp. 31-40, 2000.
- [9] S. L. Ho, M. Xie and T. N. Goh, "A Comparative Study of Neural Network and Box Jenkins ARIMA Modeling in Time Series Prediction," *Computer & Industrial Engineering*, vol. 42(2-4), pp. 371-375, 2002.
- [10] G. Lachtermacher and J. D. Fuller, "Back Propagation in Time-Series Forecasting," *Journal of Forecasting*, vol. 14(4), pp. 381-393, 1995.
- [11] H. B. Hwarng, "Insights into Neural-Network Forecasting of Time Series Corresponding to ARMA (p, q) Structures," *Omega*, vol. 29(3), pp. 273-289, 2001.