

Research Article

A Hybrid Fresh Apple Export Volume Forecasting Model Based on Time Series and Artificial Neural Network

Lihua Yang and Baolin Li

School of Economics and Management, Hubei University of Automotive Technology,
Shiyan 442002, China

Abstract: Export volume forecasting of fresh fruits is a complex task due to the large number of factors affecting the demand. In order to guide the fruit growers' sales, decreasing the cultivating cost and increasing their incomes, a hybrid fresh apple export volume forecasting model is proposed. Using the actual data of fresh apple export volume, the Seasonal Decomposition (SD) model of time series and Radial Basis Function (RBF) model of artificial neural network are built. The predictive results are compared among the three forecasting model based on the criterion of Mean Absolute Percentage Error (MAPE). The result indicates that the proposed combined forecasting model is effective because it can improve the prediction accuracy of fresh apple export volumes.

Keywords: Artificial neural network, forecasting, fresh apple, time series

INTRODUCTION

China has the highest total quantity of apple yield in the world. With the development of market economy in China, the yield advantage and the price advantage led to the export volume of China's apple have being improved in recent years. According to statistics, from 1990 to 2008, the increase of China's fruits export volume averaged annually 14.33%. However, various factors lead to the market of the Chinese fruits has changed from "the long-term in short supply" into "supply exceeds demand", the oversupply market push down the fresh fruit price and depress the increase of farmers revenue, seriously damaging the interests of fruit growers. These have hindered China's fruit industry sustained and healthy development. To solve the problem of fruits slow selling, different media have called on people to buy "love fruits", some governmental officials also take action to help fruit growers promote their products (Yang and Hu, 2013). After joining WTO, it has to face the international market competition pressure. Therefore, it is significant to accurately predict the fresh fruit export volume, which can make reasonable planning of fruits planting; the fruit growers can appropriately determine their planting scale and the supply chain can also be optimized.

Forecasting is very important. An efficient forecasting can reduce inventories, increase profits and enhance the responsive ability to the changes of the environment. A lot of experts and scholars had made some research about the forecasting. Generally, the forecasting methods can be divided into the qualitative

forecasting methods and quantitative forecasting methods. Time series model has been typically used in sales forecasting. In fact, fruit sales or export volume are affected by many factors, such as state policy, price change, income and so on. These complicated factors cause the fluctuation and non-linear characteristics of the historic data and some intelligent algorithms such as ANN (Wang *et al.*, 2010) are applied to solve the problems. Support vector regression was applied to the newspaper or magazines sales forecasting is a superior method (Wang *et al.*, 2010). However, the single model has its own limitation, which can only use some certain effective information. Therefore, the combination forecasting model was proposed by Bates and Granger (1969).

Fresh fruits are more concerned with sales forecasting due to their special characteristics, such as the short shelf-life, the need to maintain high product quality and the uncertainty and fluctuations in consumer demands. As fresh fruit can only be sold for a limited period of time, both shortage and surplus of goods can lead to loss of income (Doganis *et al.*, 2006). Different models are selected for grape sales forecasting aiming at different seasonal features in different months, the given results show that the forecasting model is feasible and valid (Ma *et al.*, 2013). However, most of the existing research is lack of the forecasting of fresh fruits. This shows that the combination forecasting on fresh fruit export volume have important significance. Here, we take fresh apple export volume of China as an example. The statistics analysis software of SPSS16.0 is used for data analysis.

Corresponding Author: Baolin Li, School of Economics and Management, Hubei University of Automotive Technology, Shiyan 442002, China

This work is licensed under a Creative Commons Attribution 4.0 International License (URL: <http://creativecommons.org/licenses/by/4.0/>).

In this study, we apply a novel hybrid model to solve the problem of forecasting the monthly export volume of fresh apple. The new hybrid model combines neural network algorithm and time series method. The radial basis function neural network and seasonal decomposition method for constructing the time series model are applied. Different models are compared based on the MAPE criterion. The results illustrate the efficiency and accuracy of the proposed model.

MATERIALS AND METHODS

Data selection: In the study, the monthly export volume of fresh apple is selected from agricultural statistical data (Yu *et al.*, 2013). The forecasting models are constructed based on monthly export volume from January 1992 to June 2002, a total of 126 samples. In order to compare different models, the models are tested using the monthly data, from July 2002 to December 2002, a total of 6 samples. Time series are commonly plotted via line charts. The line charts of monthly export volume are shown in Fig. 1.

In the Fig. 1, we can know that the time series exhibits obviously seasonal fluctuations and long-term growth trends. The annual regularity is from low to high and from high to low. Therefore, the time series probably has an annual periodicity. We can also notice that this output appears to grow with the upward series trend, which implies a multiplicative model should be adopted. To obtain a more quantitative conclusion about the underlying periodicity, we can calculate the autocorrelations and partial autocorrelations of a time series.

In the Fig. 2, we can find that the autocorrelation function of first difference shows a significant peak at a lag of 12 and a lag of 24, which exceed the confidence interval. The output implies that the underlying periodicity is 12.

Time series model: A time series is a sequence of data points, typically consisting of successive measurements made over a time interval, which is used in many field, such as statistics, signal processing, pattern recognition, econometrics, mathematical finance, weather forecasting and so on.

Time series model can be classified into linear or nonlinear models depending on the nature of the data. From the above analysis, we know that the monthly export volumes of fresh apple exhibit obviously seasonal fluctuations and long-term growth trends. Therefore, the seasonal decomposition model of time series can be applied. In commonly, a time series can be decomposed into four components: the long term Trend (T), the Seasonal component (S), the cyclical component (C) and the Irregular component (I). Seasonal decomposition model can be divided into multiplicative model and additive model. According to the relations between the four components, multiplicative model and additive model can be respectively represented as follows:

$$Y = T + S + C + I, Y = TSCI$$

Here, the multiplicative model is applied. The procedure of Seasonal Decomposition is as follows:

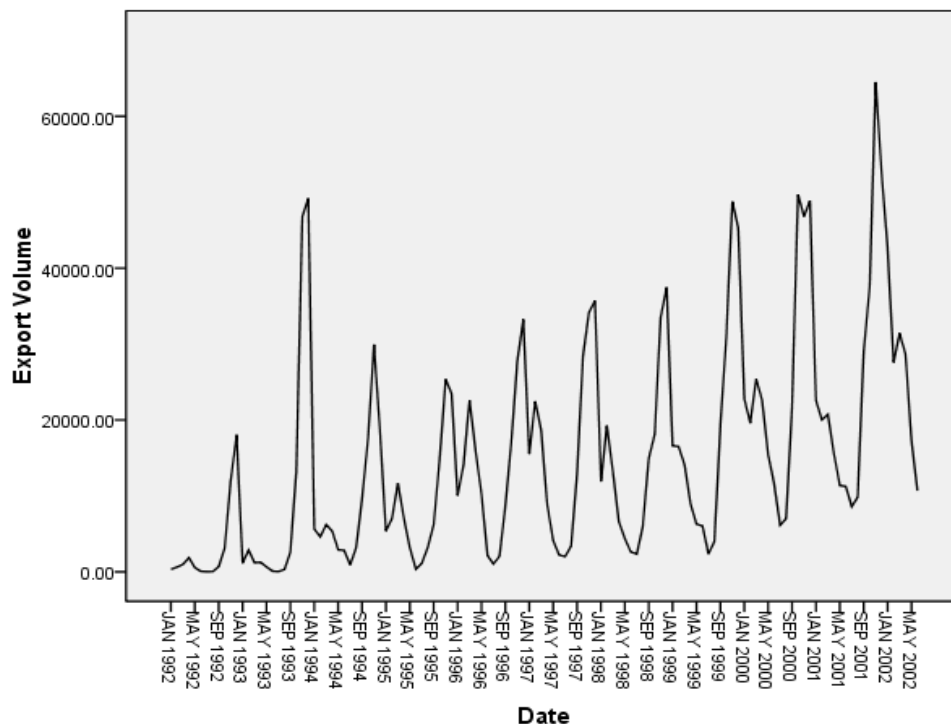


Fig. 1: Time series of monthly export volume

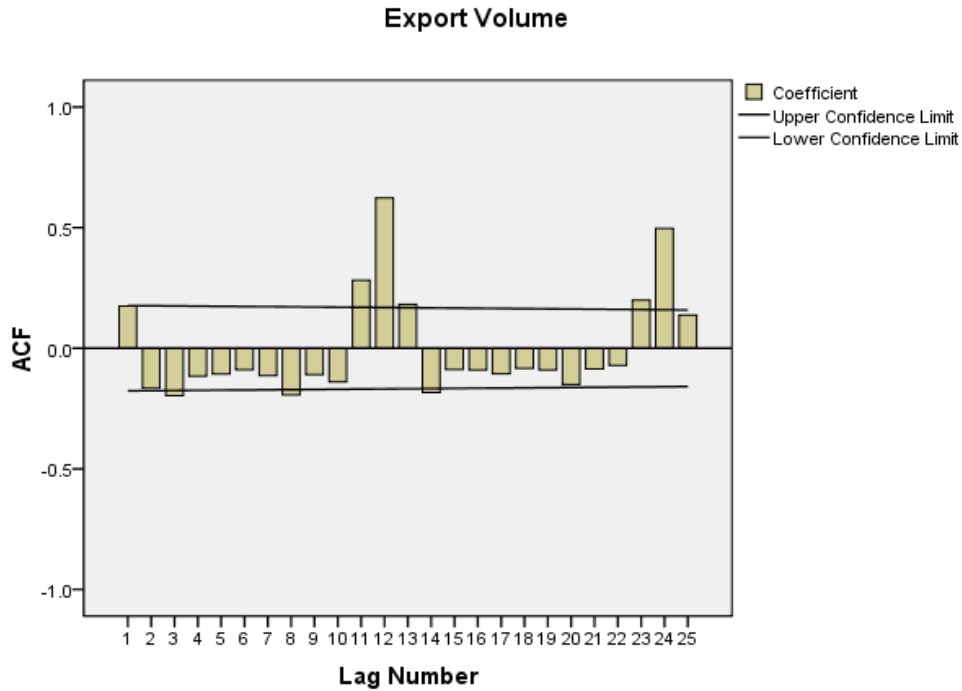


Fig. 2: Autocorrelation analysis of first difference

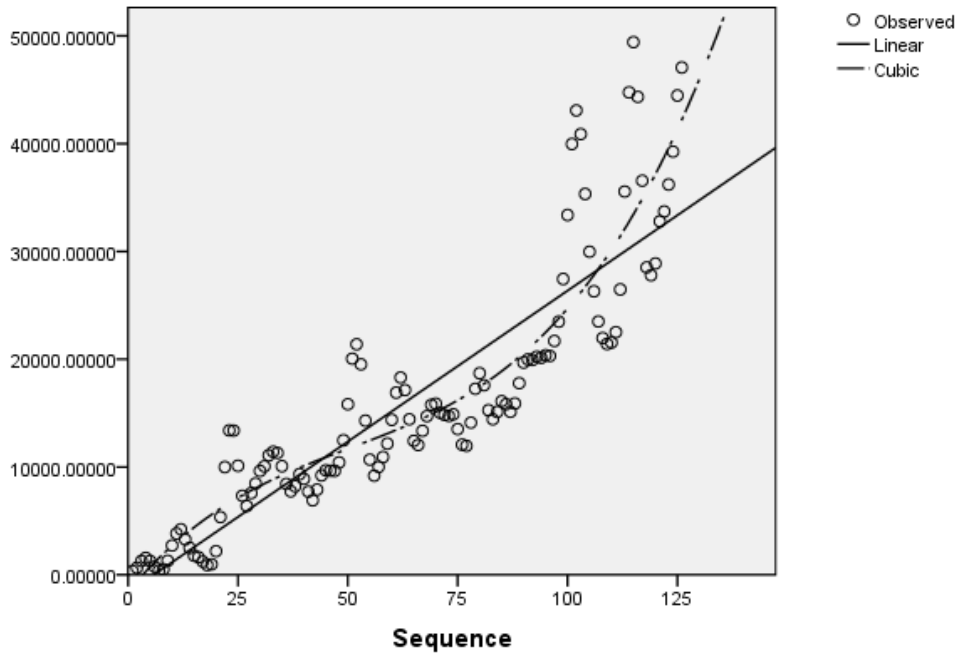


Fig. 3: Curve fit chart

Seasonal adjustment: We can get the moving average series from the original series by using the moving average method, which weight is all points equal. The seasonal decomposition procedure can creates four new variables: Seasonal Factor (SAF), Seasonally Adjusted Series (SAS), Smoothed Trend-Cycle Series (STC) and Irregular or Error component (ERR). Where SAF represents seasonal variation; SAS represents the

original series removing seasonal variations; STC represents a smoothed series that shows both trend and cyclic components; ERR is the residual component of the series.

Curve estimation: Select STC as the dependent variable and select time as the independent variable. We find that the relationship between the dependent

variable and the independent variable is not necessarily linear, so the curve fitting is applied. The curve fit chart gives us a visual assessment of the fit of each model to the actual values. From this plot, it appears that the cubic model better than linear model, then we select cubic as model to estimate, the R square of cubic model is 0.833. The curve fit chart is as follows (Fig. 3):

Artificial Neural network model: Artificial Neural Network (ANN) can simulate human brain's nerve cells by using network nodes, which is distributed and parallel. ANN has the ability of nonlinear processing, self-adaption and self-learning. Back Propagation (BP) neural network is relatively mature and commonly used in different fields. RBF neural network is better than BP neural network in approximation capability, fault tolerance and adaptation.

RBF neural network is a special artificial neural network, the architecture of it consist of three layers, which are an input layer, hidden layer and an output layer, each layer contains a number of neurons. The input layer is used to connect the network to its environment. The hidden layer adopts a nonlinear transformation to the input variables by using a radial basis function. The output is usually a simple linear function. RBF network is supervised in that the predicted results can be compared with actual values of the samples. RBF training algorithm is to determine the structure and the parameters of the network, so that the error between the actual values and the predicted output values in the training set is minimized. The minimization optimization problem can be expressed as follows:

$$J(L, c_j, w_j) = \sum_{i=1}^k (y_i - \hat{y}_i)^2$$

where,

$$\hat{y}_i = \sum_{j=1}^L w_j f(\|x_i - c_j\|), i = 1, 2, \dots, K$$

In the above equations

f = The radial basis function

L = The number of nodes in the hidden layer

c_j = The center of the j th hidden node

w_j = The connection weight between the j th hidden node and the output node

$\|\cdot\|$ = The Euclidean norm

From the above analysis, we can know that seasonal periodicity of the monthly export volume is 12 months. Monthly data from January 1992 to December 1992 are selected as the input nodes of the first sample in training datasets. The consequently monthly data of January 1993 is selected as corresponding output node

of the first sample in training datasets. Similarly, the other data samples in the training datasets can be determined on the basis of this method. Therefore, the training datasets consists of 114 records, monthly export volumes from July 2002 to December 2002 are assigned to the holdout sample. The holdout sample is another independent set of data records can be used to verify the final RBF neural network. The architecture of RBF is 12 neurons in the input layer according to the number of independent variables, likewise, there is 1 unit in the output layer and there are 2 units in the hidden layer. The best number of hidden units is automatically computed according to the smallest error in the training dataset, the activation function for the hidden layer is normalized radial basis function and activation function for the output layer is identity function.

RESULTS AND DISCUSSION

Single models forecasting: The predictive results of two single models are shown in Table 1.

Where the unit of export volume is tons, From the Table 1, we can see that different model can obtain different predicted value. Here, we use MAPE as the criteria to evaluate the forecasting precision of different models. The formula is expressed as follows:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{\hat{y}_t - y_t}{y_t} \right|$$

where,

\hat{y}_t = The predicted value of the t time period

y_t = The actual value of the t time period,

n = The total number of prediction samples

Combined forecasting: Because single models can take advantage of different information of the data, Bates and Granger (1969) firstly proposed the combined forecasting model. The research about combined forecasting has become the hotspot in the forecasting field. The key step of constructing combined forecasting is how to allocate appropriate weight to single model, many scholars proposed different calculation method, which can be divided into fixed weight and dynamic weight. In the fixed weight calculation process, the combined forecasting model can be expressed as followings:

$$Y = W_1 Y_1 + W_2 Y_2$$

where,

Y_1 = The predictive value of SD model

Y_2 = The predictive value of RBF model

Y = The predictive value of combined forecasting model

Table 1: The forecasting results of three single models

Date	Export volume	S.D.	RBF
200207	11199.6	3916.77	15676.69
200208	18003.17	7889.30	16400.08
200209	40486.26	25638.14	17559.35
200210	60584.71	47888.09	17829.84
200211	77359.71	82516.39	16331.14
200212	73044.94	82526.51	15052.72

S.D.: Standard deviation

Table 2: The calculation of the weights

Model	MAPE	Weight
SD	0.33	0.63
RBF	0.56	0.37

The weight can be calculated via the following formula:

$$W_i = M_i^{-1} / \sum_{i=1}^m M_i^{-1}, i = 1, 2, \dots, m$$

where,

i = The number of model

W_i = The weight of the i th model

M_i = The mean absolute percentage error of the i th model

In the traditional inverse variance method, M_i represents the sum of square error of the i th model.

In the Table 2, we can know that the model of the smaller mean absolute percentage error should have a higher weight. Based on the above weights, we can derived the formula of combined forecasting is as follows:

$$Y = 0.63Y_1 + 0.37Y_2$$

Then, the predictive results of combined forecasting model can be obtained via the above formula. The MAPE of combined model is 0.32, which is lower than the two single models. Therefore, the accuracy of combined model is the highest and accuracy of SD model is better than RBF neural network.

CONCLUSION

In this study, firstly, we analyzed the regularity of real data. Secondly, the Seasonal Decomposition model of time series and Radial Basis Function model of artificial neural network are constructed based on the actual export volume of fresh apple. Thirdly, a hybrid

forecasting model based on the above two single models is proposed, the weight calculation and the comparison of different models are in terms of the criterion of MAPE. The MAPE values of combined model is the smallest, the corresponding values of RBF neural network is the largest, so the new combined model based on seasonal decomposition and RBF neural network are feasible and valid. The results indicate that the proposed combined model is applicable for forecasting fresh fruits export volume and it is also significant to forecasting in other fields.

ACKNOWLEDGMENT

The authors are very grateful to the reviewers for their detailed comments and suggestions that improved the quality and presentation of this study.

The paper has been supported by the project for the Department of Education, Hubei Province of China (NO. B20092306) and by the project of Hubei University of Automotive Technology (No. 2012 X Q02).

REFERENCES

- Bates, J.M. and C.W.J. Granger, 1969. The combination of forecasts. *Oper. Res. Quart.*, 20(4): 451-468.
- Doganis, P., A. Alexandridis, P. Patrinos and H. Sarimveis, 2006. Time series sales forecasting for short shelf-life food products based on artificial neural networks and evolutionary computing. *J. Food Eng.*, 75: 196-204.
- Ma, K.P., D.Q. Zhang, Y.L. Zhang, 2013. A study on short-term forecasting of grape sales. *Appl. Mech. Mater.*, 278: 2319-2323.
- Wang, J.H., Y. Xing, F.H. Qin, T. Ma and H. Liang, 2010. Grey-neural network combination forecast model of the world food consumption. *Proceeding of the International Conference on Information Networking and Automation (ICINA, 2010)*, 1: 129-132.
- Yang, H.X. and J. Hu, 2013. Forecasting of fresh agricultural products demand based on ARIMA model. *Adv. J. Food Sci. Technol.*, 5(7): 855-858.
- Yu, X.D., Z.Q. Qi and Y.M. Zhao, 2013. Support vector regression for newspaper/magazine sales forecasting. *Proc. Comput. Sci.*, 17: 1055-1062.