Short-Term Price Forecasting For Agro-products Using Artificial Neural Networks

Gan-qiong Li a,b*, Shi-wei Xu a,b, Zhe-min Li a,b

aAgricultural Information Institute, the Chinese Academy of Agricultural Sciences, China
bKey Lab of Digital Agricultural Early Warning Technology, Ministry of Agriculture, China

Abstract

It is well known that short-term market price forecasting has been a difficult problem for a long time because of too many factors which can not be accurately predicted. Conventionally, time series analysis has been often employed in modeling short-term price forecasts. In recent years a new technique of artificial neural networks (ANN) has been proposed as an efficient tool for modeling and forecasting. A feed-forward ANN model has been developed for short-term price forecasting of tomato and in comparison with time series model ARIMA in this study. The data used include daily wholesale price, weekly wholesale price and monthly wholesale price collected between 1996 and 2010. The results showed that ANN model evidently outperformed the time series model in forecasting the price before one day or one week. A good correlation between the modeled and the real prices was observed from the feed-forward ANN model, with a relative error less than 5.0%.

Key words: short-term forecasting, neural networks, regression analysis, time series analysis, model;

1. Introduction

Making accurate predictions for short-term market prices of agro-products is not a trivial task as several influencing factors have to be considered, especially many sudden factors which are difficult to be known in advance have a decided influence on the fluctuation of short-term prices. The short term violent fluctuation of agro-products will have some negative impacts on production, consumption as well as social stability. In China the production of agriculture is not only scattered but also small in scale, making the farmers face great market risk. In most cases the farmers are price takers not price makers, therefore the conditions of market may be opposite to the initial when agro-products had been produced. A well known fact is that short-term market price forecasting has been a challenge for many decades. Nikolay Arcak and Panagiotis G. attempted to model volatility in prediction markets. Generally most researchers prefer to use leaner techniques such as regression models and time series models since its easy using. Nowadays short term forecasting is used popularly in many fields such as electric load, traffic flow, stock exchange, water demand, disaster and whether . The short-term forecasting model has always

*Corresponding author, Tel.: 010-82109761; fax: 010-82106285
E-mail address: lgqsjf@caas.net.cn
been applied to agro-product price forecasting before one month or one quarter, but rarely applied to that before several days or one week.

Short term price forecasting for agro-products is a subject that has been studied extensively since 21 century, and several different methods have been developed. Conventional statistical forecasting methods can be divided into linear regression method and time series method. In the simple or multiple regression model, the relationships between the dependent variables and independent variables are modelled. The parameters are estimated according to the past values of these variables, and the relationships are used to forecast the dependent variable by using the future evaluations of the independent ones. In fact, linear regression models have a significant drawback. The market price of agro-products has multiple influencing factors which have complex inter-relations. In regression models it was assumed that the explanatory variables and dependent variable are linear relationships. However, they are more or less nonlinear. In the time series methods such as persistence method, trends method, the prediction is based on the past values. Taking persistence method for example, the main idea is today equals tomorrow, i.e. if the wholesale price of tomato is 2.50 Yuan/kg today, the persistence method predicts that it will be 2.50 Yuan/kg tomorrow. This method works well when factors of market change very little. However, if market conditions change significantly from day to day, this method usually breaks down. Trends method also has some drawback because the conditions did not continue to move in the same direction simultaneously. If they slow down, speed up, change intensity, or change direction, the trends forecasting probably will not work as well. Paulo cortez [12] employed evolving time series forecasting ARMA models to overcome some shortage mentioned above. Qun zhou et al [13] proposed a two-stage approach for generating simulated price scenarios based on the available price date, and their results indicated that the proposed approach was able to generate price scenarios for distinct seasons with empirically realistic characteristics.

Real-world price data of agro-products and its underlying market changes are often nonlinear in nature. As a result, linear models may not be suitable when market changes frequently. The rapid development of computing power has allowed nonlinear method to become applicable to modelling and forecasting. The artificial neural network (ANN) with a nonlinear dynamics system is based on the distributive storage and parallel co-processing of messages. Recently, numerous artificial neural networks in modelling the complex nonlinear relationships between the market price and its determining factors such as meteorological and seasonal variables, holidays, food safety cases and other emergencies have been reported [14-25].

The main objective of this study is to examine whether ANN model is suitable for short-term forecasting and how its accuracy is compared with time series model. Historical forecasting and future forecasting were used here. The historical forecasting is mainly to evaluate the fitness of the two methods based on historical data, and the future forecasting is to examine the accuracy of forecasting results based on the two models.

2. Date Processing and Evaluation Methods of Model

Tomato was used as a target throughout this study because it is one of the most popular and familiar agro-products in China. Historical data employed in the study were available on the website of Ministry of Agriculture (MOA) of China, where the database included tomatoes’ daily price of 511 wholesale markets of agro-products in China since 1996. Because there were short of daily volumes of tomatoes’ trade for some wholesale market, the whole country daily wholesale price of tomatoes was an arithmetical average of 511 wholesale markets. The weekly wholesale price and monthly wholesale price of the whole country were also calculated by the same method.

To evaluate the fitness of different models, we used relative error as an important index to compare them. The relative error is calculated as follows:

\[
P_{\text{error}} = \frac{|X_f - X_{\text{obs}}|}{X_{\text{obs}}} \times 100\%
\]

\(P_{\text{error}}\) means relative error. \(X_f\) is the fit of model simulation. \(X_{\text{obs}}\) is the real observation.

3. Time Series Model

3.1 Simple Description

The time series analysis ARIMA model is popularly applied in short term forecasting since it doesn’t need too much economics knowledge for the users. In the following paragraph we used the method to establish daily price forecasting model, weekly price forecasting model and monthly price forecasting model by PASW Statistic 18.0.
3.2 Historical forecasting and Evaluation

A. Daily Price Forecasting Model

According to the data of 365 daily price in 2009, a suitable model of ARIMA (0,1,7) can be established by PASW Statistic 18.0. Figure 1 shows the results of simulation on daily price. The accuracy distribution of historical prediction is listed below:

1) The relative error between 10% and 15% was 5, indicating that the proportion of prediction accuracy between 85% and 90% was 1.4 percent of the total historical forecasting for daily price.

2) The relative error between 5% and 10% was 22, indicating that the proportion of prediction accuracy between 90% and 95% was 6.0 percent of the total historical forecasting for daily price.

3) The relative error less than 5% was 337, indicating that the proportion of prediction accuracy more than 95% was 92.6 percent of the total historical forecasting for daily price.

![Figure 1: historical forecasting for daily wholesale price using ARIMA model](image)

B. Weekly Price forecasting Model

According to the data of weekly price from 2007 to 2009, a suitable model of ARIMA (1,1,0) can be established by PASW Statistic 18.0. Figure 2 shows the results of simulation on weekly price. The accuracy distribution of historical prediction is listed below:

1) The relative error between 20% and 25% was 5, indicating that the proportion of prediction accuracy between 75% and 80% was 3.2 percent of the total historical forecasting for weekly price.

2) The relative error between 10% and 20% was 25, indicating that the proportion of prediction accuracy between 80% and 90% was 12.7 percent of the total historical forecasting for weekly price.

3) The relative error between 5% and 10% was 42, indicating that the proportion of prediction accuracy between 90% and 95% was 26.6 percent of the total historical forecasting for weekly price.

4) The relative error less than 5% was 91, indicating that the proportion of prediction accuracy more than 95% was 57.5 percent of the total historical forecasting for weekly price.
C. Monthly Price Forecasting Model

According to the data of monthly price from 1996 to 2009, a suitable model of ARIMA \((0,1,2)(1,1,1)^s\) can be established by PASW Statistic 18.0. Figure 3 shows the results of simulation on monthly price. The accuracy distribution of current prediction is listed below:

1) The relative error more than 30% was only 1, indicating that the proportion of prediction accuracy less than 70% was 0.7 percent of the total historical forecasting for monthly price.

2) The relative error between 20% and 30% was 14, indicating that the proportion of prediction accuracy between 70% and 80% was 9.0 percent of the total historical forecasting for monthly price.

3) The relative error between 10% and 20% was 44, indicating that the proportion of prediction accuracy between 80% and 90% was 28.4 percent of the total historical forecasting for monthly price.

4) The relative error between 5% and 10% was 43, indicating that the proportion of prediction accuracy between 90% and 95% was 27.7 percent of the total historical forecasting for monthly price.

5) The relative error less than 5% was 53, indicating that the proportion of prediction accuracy more than 95% was 34.2 percent of the total historical forecasting for monthly price.
4. ANN model

The ANN usually called “neural network” (NN), is a mathematical model or computational model that tries to simulate the structure and/or functional aspects of biological neural networks. It consists of an interconnected group of artificial neurons and processes information using a connectionist approach to computation. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase. Modern neural networks are non-linear statistical data modelling tools. They are usually used to model complex relationships between inputs and outputs or to find patterns in data.

A feed-forward neural network is one of ANN where connections between the units do not form a directed cycle. In this network, the information moves in only one direction, forward, from the input nodes, through the hidden nodes (if any) and to the output nodes. There are no cycles or loops in the network. Now feed-forward neural network is the most popular and most widely used model in many practical applications. In this study, feed-forward neural network is used as the forecasting network.

4.1 Algorithm Steps

The basic flow chart of ANN is shown as below:

The steps of price forecasting are listed as follows:

1) Data Pre-processing: Since poor or insufficient data were not suitable to develop good models, the initial data should be pre-processed before applying neural network algorithm. Here we only carry on data normalization in the pre-processing step \( y = \frac{x - \text{mean}}{\sigma} \), \( \sigma \) is the standard deviation, and \( x \) is the input data). Usually we can divide the initial data into two parts: sample data (used in network training) and validation data (used in network evaluation).

2) Neural Network Creation: In this study a three-layer feed-forward neural network including input layer, hidden layer and output layer was used. The number of neurons in input and output layer depends on practical application. In our study, the neuron number of input layer and output layer should be set as \( n \) and \( m \) if we want to forecast the price of next \( m \) days based on the price of past \( n \) days. And according to that, the neuron number of hidden layer should be set as \( 2 \times n + 1 \) based on Kolmogorov Theorem.

3) Neural Network Training: During the neural network training step, firstly we should determine the value of two important parameters. One is the Mean Squared Error (MSE), which represents the average squared error between the network's output and the target value. The other is the maximum number of learning epochs, up to which the training will be stopped no matter whether the mean-squared error target has been achieved. Now we can train the neural network with the sample data generated in Step 1. The training phase will stop when the mean-squared error is achieved or the maximum number of learning epochs is reached.

4) Forecasting based on Trained Neural Network: Using the trained neural network to forecast the future
agricultural commodity price.

4.2 Historical Forecasting and Evaluation

A. Daily Price Forecasting

The data of daily price from 1996 to 2008 were used as sample date (network training) and the data of 2009 were used as historical prediction model evaluation (network evaluation). Figure 4 shows the results of simulation on daily price. The accuracy distribution of historical prediction is listed below:

1) The proportion of prediction accuracy between 85% and 90% was 2.1 percent of the total historical forecasting for daily price.

2) The proportion of prediction accuracy between 90% and 95% was 29.9 percent of the total historical forecasting for daily price.

3) The proportion of prediction accuracy more than 95% was 68.0 percent of the total historical forecasting for daily price.

![Figure 4: Forecasting for daily wholesale price of tomatoes using ANN](image)

B. Weekly Price Forecasting

The data of weekly price from 1996 to 2006 were used as sample data (network training) and the weekly data between 2007 and 2009 were used as historical prediction model evaluation (network evaluation). Figure 5 shows the results of simulation on weekly price. The accuracy distribution of historical prediction is listed below:

1) The proportion of prediction accuracy between 75% and 80% was 2.4 percent of the total historical forecasting for weekly price.

2) The proportion of prediction accuracy between 80% and 90% was 16.8 percent of the total historical forecasting for weekly price.

3) The proportion of prediction accuracy between 90% and 95% was 19.2 percent of the total historical forecasting for weekly price.

4) The proportion of prediction accuracy more than 95% was 61.6 percent of the total historical forecasting data of weekly price.
C. Monthly Price Forecasting

The data of monthly price from 1996 to 2009 were used as sample data (network training) and historical prediction evaluation (network evaluation). Figure 6 shows the results of simulation on monthly price. The accuracy distribution of historical prediction is listed below:

1) The proportion of prediction accuracy less than 70% was 8.1 percent of the total historical forecasting for monthly price.

2) The proportion of prediction accuracy between 70% and 80% was 14.3 percent of the total historical forecasting for monthly price.

3) The proportion of prediction accuracy between 80% and 90% was 24.2 percent of the total historical forecasting for monthly price.

4) The proportion of prediction accuracy between 90% and 95% was 21.7 percent of the total historical forecasting for monthly price.

5) The proportion of prediction accuracy above 95% was 31.7 percent of the total historical forecasting for monthly price.
5. Future Forecasting and Conclusions

5.1 Future Forecasting Results

Applying the above models of ARIMA and ANN, the forecasting results ahead of seven days for 2010 are listed in table 1. Table 2 is the results ahead of four weeks and table 3 is the results ahead of three months.

As we can seen from table 1, ARIMA model and ANN model are both effective for forecasting daily price and the accuracy is more than 90%. The accuracy of ANN model declines with the longer forecasting cycles.

Table 1: Comparison of Future Forecasting Results of Tomatoes’ Daily Wholesale Price between ARIMA and ANN

<table>
<thead>
<tr>
<th>Daily observed</th>
<th>ARIMA(0,1,7)</th>
<th>ANN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>fit</td>
<td>P_{error}</td>
</tr>
<tr>
<td>2010-01-01</td>
<td>3.27</td>
<td>3.40</td>
</tr>
<tr>
<td>2010-01-02</td>
<td>3.35</td>
<td>3.39</td>
</tr>
<tr>
<td>2010-01-03</td>
<td>3.38</td>
<td>3.37</td>
</tr>
<tr>
<td>2010-01-04</td>
<td>3.41</td>
<td>3.37</td>
</tr>
<tr>
<td>2010-01-05</td>
<td>3.47</td>
<td>3.36</td>
</tr>
<tr>
<td>2010-01-06</td>
<td>3.47</td>
<td>3.35</td>
</tr>
<tr>
<td>2010-01-07</td>
<td>3.45</td>
<td>3.36</td>
</tr>
</tbody>
</table>

As we can seen from table 2, the accuracy of ANN model in forecasting weekly price is lower than that of ARIMA model. The accuracy of ANN model declines rapidly with the longer forecasting cycles.

Table 2: Comparison of Future Forecasting Results of Tomatoes’ Weekly Wholesale Price between ARIMA and ANN

<table>
<thead>
<tr>
<th>Weekly observed</th>
<th>ARIMA(1,1,0)</th>
<th>ANN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>fit</td>
<td>P_{error}</td>
</tr>
<tr>
<td>2010-1</td>
<td>3.40</td>
<td>3.37</td>
</tr>
<tr>
<td>2010-2</td>
<td>3.39</td>
<td>3.37</td>
</tr>
<tr>
<td>2010-3</td>
<td>3.51</td>
<td>3.36</td>
</tr>
<tr>
<td>2010-4</td>
<td>3.55</td>
<td>3.36</td>
</tr>
</tbody>
</table>

As we can seen from table 3, ARIMA model and ANN model are both not very ideal comparing with daily price forecasting and weekly price forecasting.

Table 3: Comparison of Future Forecasting Results of Tomatoes’ Monthly Wholesale Price between ARIMA and ANN

<table>
<thead>
<tr>
<th>Monthly observed</th>
<th>ARIMA(0,1,2)(1,1,1)</th>
<th>ANN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>fit</td>
<td>P_{error}</td>
</tr>
<tr>
<td>2010-01</td>
<td>3.46</td>
<td>3.70</td>
</tr>
<tr>
<td>2010-02</td>
<td>3.47</td>
<td>3.87</td>
</tr>
<tr>
<td>2010-03</td>
<td>3.26</td>
<td>3.86</td>
</tr>
</tbody>
</table>

5.2 Conclusions

- The time interval of data is shorter, the accuracy is higher. Results showed the accuracy of daily price forecasting is the best. Next is weekly price forecasting.
- The ANN model is more suitable for forecasting ahead of one cycle. Time series model and ANN model are equally effective in accuracy for forecasting ahead of one day, one week and one month, but ANN model is better than time series model.
- The accuracy of ANN model for three types of forecasting is more than 80%, and daily price forecasting is even more than 90%.
6. Discussions

Although ANN is an effective tool for short term forecast and widely used in many fields nowadays, there are still some problems to be further solved.

- **Disadvantages**
  
  ANN is in a sense the ultimate ‘black boxes’. Apart from defining the general architecture of a network and perhaps initially seeding it with a random numbers, the user has no other role than to feed it input and watch it train and await the output. In fact, it has been said that with ANN, “you almost don't know what you're doing”. The final product of this activity is a trained network that provides no equations or coefficients defining a relationship (as in regression) beyond it's own internal mathematics. The network 'IS' the final equation of the relationship.

- **Potential Improvements**
  
  Firstly, we only carried out normalization on the initial data in Step 1. Considering the difficulty of obtaining the price data of each day, we can carry out data smoothing on the price-time curve to gain the missing day’s data.

  Secondly, feed-forward neural network is the most widely used models, but it has the simplest network architecture. So if the price-time curve presents a more sophisticated shape, we can attempt to apply more complicated types of network on this problem.

  Finally, now we only train the network once before using it to forecast price. In order to minimize the instability of neural network, we can carry on a few times of training and forecasting, and use the average value of each time as the final predicted output.

**Acknowledgement**

This work is supported by Key Project of National Key Technology R&D Program—“Support System and Demonstration of Key Technologies of Intelligent-Analysis and Early-Warning for Security of Agro-Products” of China with Project 2009BADA9B01.

**References**


