


RESEARCH

Open Access



A new method of large-scale short-term forecasting of agricultural commodity prices: illustrated by the case of agricultural markets in Beijing

Haoyang Wu^{1*} , Huaili Wu¹, Minfeng Zhu¹, Weifeng Chen² and Wei Chen¹

*Correspondence:

why0902@zju.edu.cn

¹ State Key Lab of CAD&CG, Zhejiang University, No. 866 Yuhangtang Road, Xihu District, Hangzhou, Zhejiang, China

Full list of author information is available at the end of the article

Abstract

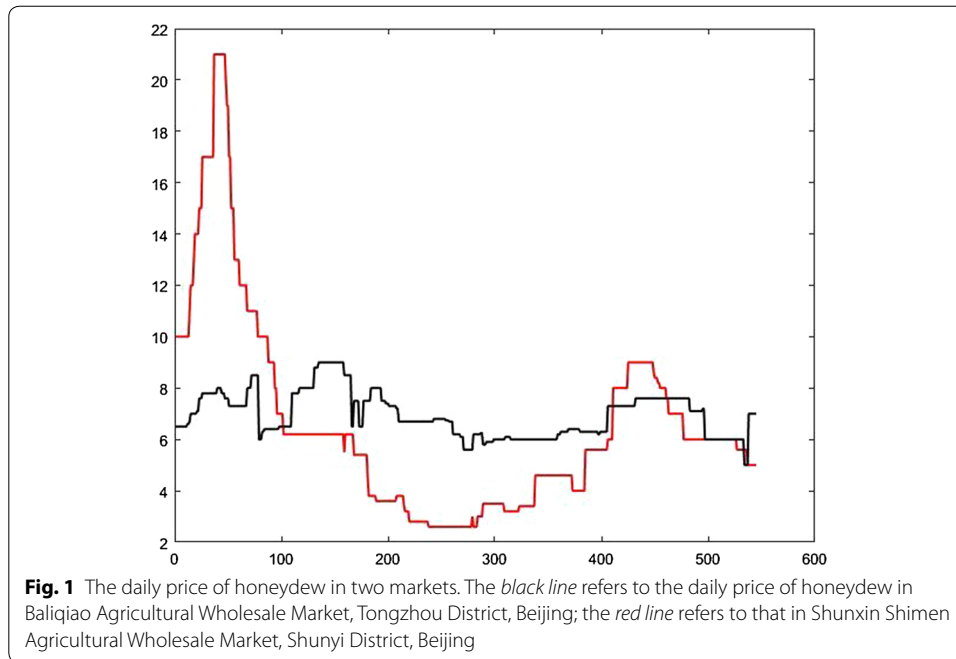
In order to forecast prices of arbitrary agricultural commodity in different wholesale markets in one city, this paper proposes a mixed model, which combines ARIMA model and PLS regression method based on time and space factors. This mixed model is able to obtain the forecasting results of weekly prices of agricultural commodities in different markets. Meanwhile, this paper sets up variables to measure the price changing trend based on the change of exogenous variables and prices, thus achieves the warning of daily price changes using neural networks. The model is tested with the data of several types of agricultural commodities and error analysis is made. The result shows that the mixed model is more accurate in forecasting agricultural commodity prices than each single model does, and has better accuracy in warning values. The mixed model, to some extent, forecasts the daily price changes of agricultural commodities.

Keywords: Change warning, Mixed model, Neural networks, Price forecasting

Background

There is an old saying that “food is the paramount necessity of the people”. The price of agricultural commodity, which is an important necessity, is closely related to people’s lives. The fluctuation of agricultural commodity prices is affected by economic and social factors. Therefore, accurate forecasting of price change trends can instruct people’s consuming behaviors, and has great significance to some heated social issues like predicting macroeconomic trend.

There are various agricultural commodities in the markets. The prices of agricultural commodities can be influenced by many factors, even same commodity can be priced differently in diverse markets. Taking the daily price of honeydew in Beijing as an example, Fig. 1 shows the daily price of honeydew in Baliqiao Agricultural Wholesale Market, Tongzhou District, Beijing, and Shunxin Shimen Agricultural Wholesale Market, Shunyi District, Beijing, from January 2014 to June 2015. It can be seen from Fig. 1 that price trends of two markets have great difference. Consumers and administrative departments certainly would like to have an overall knowledge of forecasting prices in several agricultural markets.



Agricultural commodity prices are influenced by a combination of factors, including supply–demand relationship, weather, policy, etc. These factors cannot be quantified by the same standard, and have different influences on different agricultural commodities in different wholesale markets, which brings great difficulty to the forecasting of agricultural commodity prices [1].

The short-term forecasting, including the weekly price changes and the daily price changes, is challenging because the fluctuation of prices is affected by a combination of uncertain factors. Meanwhile, it is also important to forecast when a drastic price change will happen, as in most cases, agricultural commodity prices are alternately stable and fluctuant [2].

Currently, as following, there are three types of short-term forecasting methods to predict the agricultural commodity prices:

1. Time series methods, including short-term forecasting methods like ARIMA model, GARCH model. These methods are only based on history prices of agricultural commodities while ignoring other factors. Therefore, these models no longer work when the prices are affected by non-seasonal factors.
2. Regression methods, including vector auto-regression model, vector auto-regressive moving average model. These methods take other factors into consideration. However, due to the limitation of the using conditions, it is impossible for a single model to be used to forecast several different kinds of agricultural commodities in the same time.
3. Learning methods, including neural networks. These methods have extensive application scope. However, when forecasting different agricultural commodities, the effects cannot be ensured and overfitting may happen. Thus, these methods are usually used to forecast some specific kinds of agricultural commodities.

Current methods are mostly based on a single model and target on a certain agricultural commodity in a specific market. These methods have not been tested by large-scale data, and can only be used in a small range. Also, most current methods fail to consider exogenous economic variables and interactions between different markets with seasonal factors together, which reduces the accuracy of the forecasting of variation time and variation amplitude [3].

This paper designs a data model with sample tests to solve the problems mentioned above and proposes a new mixed model to forecast agricultural commodity prices. We revise ARIMA model by PLS regression method, taking the influence of other agricultural markets in the same city into consideration. We forecast weekly price changes of agricultural markets by considering the interactions between different markets and seasonal factors.

On the basis of the mixed model of time and space factors, this paper also proposes a price change warning model with a variable “urgency” to quantify the price change trend. We use neural networks to analyze the “urgency” and other exogenous variables, and forecast the value of the coefficient of the “urgency”. Thus, to some extent, we can predict the trend of daily price changes.

The method proposed by this paper has good forecasting effects on over 20 types of agricultural commodities in Beijing agricultural markets. The error analysis and visible result analysis show that the mixed model of this paper has obtained satisfactory forecasting results. The mixed model makes an improvement both on the forecasting accuracy and efficiency compared with any other single models.

The breakthrough of the method proposed by this paper basically includes:

1. The model proposes a daily warning model to quantify and forecast the daily change trend of agricultural commodities.
2. The model can be used to forecast a great many types of agricultural commodities with good effects.
3. The model realizes simultaneous forecasting of agricultural commodities in different markets in one city by considering space factors.

This paper starts from current researches, combines single models and proposes a mixed forecasting model, this model makes forecasting of agricultural commodities prices in different markets simultaneously possible. Also, it provides more stable and accurate results as compared to single models or some other models. Meanwhile, we build a daily price warning model based on neural networks and to some extent realize daily price forecasting of agricultural commodities, which has application value for consumers and relative administrative departments.

Literature review

Forecasting models of agricultural commodity prices are mainly divided into two types. One is structural models, which analyze price factors from economic perspective. On the basis of microeconomics and econometrics, Lord [4] proposed that price was interacted with demand, supply and inventory, therefore built a price model with a time-related equation set.

Another type is nonstructural methods, which ignore economic principle and directly research on the time series of prices. Box and Jenkins [5] proposed autoregressive integrated moving average (ARIMA) model. The modeling, parameter estimation, model testing and forecasting result analysis were based on the assumption that future prices were related to historical prices and random variables. This model ignored the influence of all other factors. Rausser and Carter [6] used ARIMA model to analyze the futures prices of soybean, soybean oil and soybean meal, drawing conclusion that soybean and soybean meal performed better in ARIMA model than in random walk model. Granger [7] pointed out that over-difference happened when using ARIMA model to deal with data which have long-term memory, therefore proposed autoregressive fractionally integrated moving average (ARFIMA) model. Barkoulas et al. [8] computed the fractional difference of futures price of agricultural commodities and found that some futures prices had long-term memory, thus met the requirement of ARFIMA model. ARIMA model ignored the influence of other factors on price. Sims [9] proposed vector autoregression (VAR) model to build time series of a vector. Park [10] used different VAR models to analyze the prices of fodder and cows, drawing conclusions that Bayesian vector auto-regression (BVAR) model and unrestricted vector VAR (UVAR) model generated forecasts which were superior to both a restrict VAR (RVAR) model and a vector auto-regressive moving average (VARMA) model in this case.

But smoothing the data by difference cannot be explained from the economic perspective. Engle and Granger [11] analyzed the linear combination of variables based on their co-integration relationship. They proposed vector error correction (VEC) model, thus smoothing the data in a different way. Due to the limitation of premise, a single model usually cannot precisely forecast prices. Yu Le et al. [12] respectively forecasted prices with three exponential smoothing model, simple linear regression model, grey forecast model, and then found the optimal linear combination which had the least error sum of squares.

Scholars have long been researching on long-term trend of agricultural commodity prices which have conspicuous periodicity. Beveridge and Nelson [13] proposed a universal method to smooth nonstationary time series. This method only required that the continuous change of the time series is stationary. Harvey [14] proposed structural time series (STS) model, which consisted of a series of univariate time series models. This method avoided model recognition and successfully separated season factors from the price change. It was economically explainable. Recently, some new methods have been proposed. Davidson et al. [15] used semi-parametric regression method based on wavelet analysis to estimate the variation period and illustrated the potential of this method. The volatility of price is another important research direction. Random noise is usually hard to observe, but it's important in price forecasting. Engle [16] proposed autoregressive conditional heteroscedasticity (ARCH) model. The model believes that the variance of noise is not constant, instead it is affected by past information. Bollerslev [17] proposed generalized autoregressive conditional heteroscedasticity (GARCH) model, an improvement of ARCH model. GARCH model performed better in stimulating time series with long-term memory. Krytsou et al. [18] proposed that long-term forecasting of noisy chaotic return series no longer worked. Instead, Mackey–Glass–GARCH model could be used. Schroeder [19] divided price noise into four categories based on

power-law exponent, specifically white noise, pink noise, brown noise and black noise. Empirical studies using this method by Labys [20] came to the conclusion that most agricultural commodities had black noise, which meant that forecasting the agricultural commodity price was rather difficult.

Neural networks have become a heated method to forecast prices. Lapedes and Farber [21] forecasted prices with neural networks. It can fit an arbitrary curve, and has good generalization ability.

Another forecasting direction is volatility forecasting model. Andersen et al. [22] compared several models including GARCH fluctuation, random fluctuation and multivariate fluctuation. Manfredo et al. [23] forecasted the volatility of the price of corn and cows with volatility model. Kroner et al. [24] forecasted prices of gold, corn, cotton etc. with expectation-variance model. Nowadays, scholars are considering combining structural and nonstructural forecasting methods, making the forecasting results more economically meaningful.

This paper uses time series method based on the periodicity of agricultural commodities, meanwhile uses space model based on the relevance of different markets, and forecasts the weekly prices of agricultural commodities by the integration of two models above. Furthermore, this paper processes exogenous variables and thus achieves the warning of daily prices by neural networks.

Data processing

Data source

Agricultural commodity price data come from the website of commerce department.¹ The data include daily prices of all agricultural commodities in wholesale markets all over China from January 2, 2014 to June 30, 2015. Some data are missing due to holidays or network causes. This paper uses data in Beijing as a sample.

This paper takes weather,² sino-US exchange rate,³ and international crude oil prices⁴ as exogenous variables. Daily weather data, daily sino-US exchange rate data, and daily price of international crude oil are from January 1, 2014 to June 30, 2015. The data of exchange rate and international crude oil are only available on their working days.

The model built in this paper is based on a large data dimension. This paper analyzes and deals with prices of all agricultural commodities in all markets as well as daily data of other variables in the same time and finally obtains forecasting results.

Sample processing

This paper uses the data of the former 80% days as the training set, and forecasts the prices of the latter 20% days. The real prices of the latter 20% days are used to evaluate the forecasting results.

The relationship between agricultural commodity prices of day t and day $t-1$ can be either change or not change. Usually agricultural commodity prices will alternately

¹ http://nc.mofcom.gov.cn/channel/gxdj/jghq/jg_list.shtml.

² <http://en.tutiempo.net/climate/2014/ws-545110.html>.

³ http://www.eia.gov/dnav/pet/pet_pri_spt_s1_d.htm.

⁴ <http://www.chinabgao.com/stat/stats/44589.html>.

change or keep constant. We observe the data and notice that prices usually keep constant for a while before a sudden change. Therefore, we assume in the daily warning model that prices keep constant and change when exogenous variables reach a certain degree. So this paper assigns the last day's price to the data missing day, instead of using linear interpolation. That is to say:

$$price(t) = price(t - 1) \quad \text{if } price(t) \text{ is missing}$$

The missing data of other exogenous variables are assigned in the same way.

The preprocessing method of data is different in different sub-models. The preprocessing method in this paper follows the rule that retaining the price change trend and ignoring huge fluctuation of prices in a rather short period of time because consumers are unable to react to huge fluctuation of prices.

Price forecasting and the warning model

This paper uses a mixed model to deal with different factors, integrate the forecasting results of different factors, and get the final forecasting results.

The mixed model can be divided into two parts: weekly price change forecasting model and price change warning model. Weekly price change forecasting model includes time factor forecasting model (4.1), space factor forecasting model (4.2) and time-space integrated model (4.3), respectively dealing with the season factor, the space factor (the influence of price change in other markets) and the integration of outputs of sub-models [25, 26]. Price change warning model deals with exogenous variables (4.4). This paper uses different data preprocessing methods according to different sub-models to obtain better forecasting results.

The frame of the overall model is shown in Fig. 2.

Time factor forecasting model

Most papers forecast agricultural commodity prices based on time series models. These models do not require data of any other variables and the feasibility has been proved. Therefore time series models are still an important part in the mixed model of this paper.

Data preprocessing

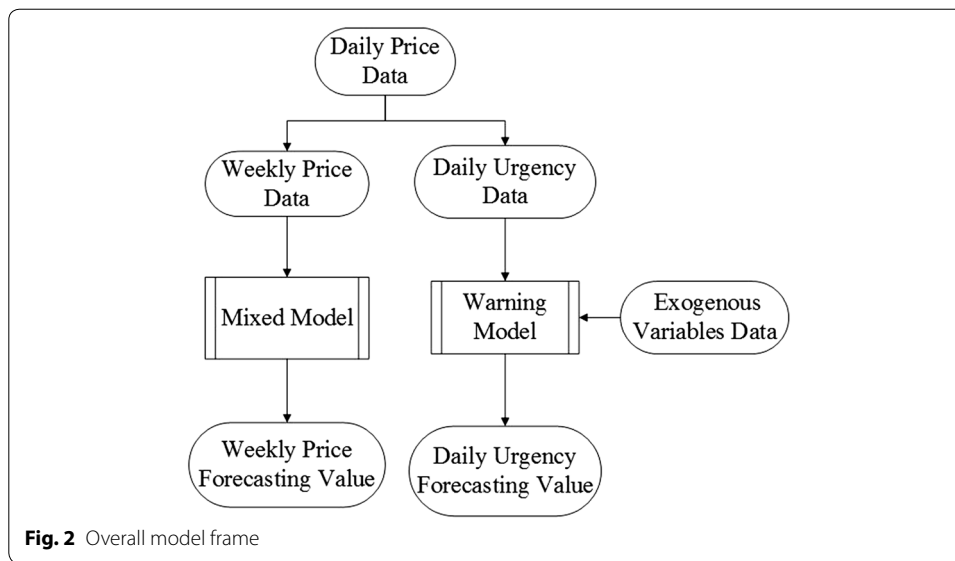
Time series models are good at analyzing and forecasting long-term data, which has clear trend and regular fluctuation. Therefore, this paper uses weekly price in time series models, by calculating the average daily prices in 1 week [27]. The purpose is to raise forecasting accuracy, by avoiding the influence of fluctuation and abnormal amplitude.

ARIMA (p, d, q) model

This paper forecasts agricultural commodity weekly prices with ARIMA model as time factor forecasting model. ARIMA model is a classical and widely-used model. Parameters p, d, q respectively represents the order of auto-regression, the difference time of smoothing the time series, and the order of moving average.

The mathematical form of ARIMA model is:

$$\theta_i^{(1)}(t) = \mu + \rho(B)\varphi(B)^{-1}\varepsilon(t)$$



$\theta_i^{(1)}(t)$ is a series of random variables, in this model is the weekly price change of time t . t represents for time. μ represents for mean value. B is backward shift operator, $B(W(t)) = W(t - 1)$. $\rho(B)$ is moving average operator, $\rho(B) = 1 - \rho_1(B) - \dots - \rho_q(B)$. $\phi(B)$ is auto-regression operator, $\phi(B) = 1 - \phi_1(B) - \dots - \phi_p(B)$. $\varepsilon(t)$ is independent disturbance, or random error.

In this model, we first put the data set to ADF stationarity test (augmented DF stationarity test). If the data set fails the test, difference the data set until it can pass the test or abandon this group of data [28]. In fact, most agricultural commodity prices can pass ADF test within one order of difference, therefore we assign $d = 1$. The values of p and q are chosen by AIC (Akaike information criterion) test. Set the range of p and q within 1 to 10. Then put the training set to AIC test, and find out p and q of the least AIC value. It takes a long time to figure out p and q for each agricultural commodity and an alternative solution is to directly take $p = 10$, $q = 8$. The forecasting results are accurate.

Space factor forecasting model

This paper forecasts prices of all agricultural markets in one city. In this part, the paper mainly considers the influence of the price changes in other agricultural markets. The consideration of this factor is based on consumers' behavior that price changes will affect consuming behavior in the same city.

Data preprocessing

Consumers will not react to price changes within the same day. Therefore there is a time lag in the influence of price changes in other markets. This paper takes weekly average value of price difference, in this way to retain the trend of price changes, and leave enough time for reaction time lag.

Besides the time lag, the relevance between different wholesale markets is another difficulty in model designing, as most methods in regression analysis require variables to be mutually independent. The purpose of the model designing in this part is to evaluate the

influence intensity between agricultural markets. Therefore, this paper uses partial least squares (PLS) method to forecast prices based on the space factor [29].

PLS model

Partial least squares method includes one procedure which is similar to principal component analysis (PCA), therefore can be used on variables with multiple correlations. For an agricultural commodity in market i , we want to forecast weekly price change $\theta_i(t)$ at time t . Independent variables are the price changes of other markets ($\theta_1(t-1), \dots, \theta_{i-1}(t-1), \theta_{i+1}(t-1), \dots, \theta_n(t-1)$) at time $t-1$. We preprocess the training set with procedures above and put it in PLS model, and obtain regression relations between the price changes of target market and the price changes of other markets at the last time point. Finally we get the forecasting value $\theta_i^{(2)}(t)$ of space model by the regression relations.

Through PLS model, we can obtain regression coefficients between each pair of agricultural markets, which to some extent reflect influential relationship between agricultural markets.

Furthermore, here we use PLS instead of directly using multivariate ARIMA model because multivariate ARIMA model require variables to be co-integrated. However, the price changes of agricultural markets in China indicate that the price change series in different agricultural markets have different stationarity. Therefore, different markets, failing in co-integration test, are not co-integrated. So we consider about using PLS method, a more general method, which can deal with all types of multivariate series.

Mixed forecasting model of weekly prices

After the preprocessing of two models above, we can get two groups of data, which is forecasting difference of time and space model of the next week (of the last week in the training set). Based on the analysis above, we've already known that weekly price changes are influenced both by seasonal factor and space factor, yet we don't know the detail how two factors work together. There are two ways to figure out the relationship of the two factors: one is by economic analysis, the other is to test several possible model with historical data and choose the best one.

The integration method in this paper is to stimulate the forecasting of two models based on the training set, then put the forecasting results as independent variables and real weekly price changes into a linear regression. Linear model is an effective and relatively simple model, besides it can reveal the weight of each factor in the relationship. Also, it is proper as currently no relative research about weight of two factors is published and the amount of sample we have is small. Finally we can get the regression relations of two factors affecting price changes in different weights,

$$\theta_i(t) = \alpha_1 \theta_i^{(1)}(t) + \alpha_2 \theta_i^{(2)}(t)$$

$\theta_i^{(1)}(t)$ is the forecasting value of ARIMA model of market i . $\theta_i^{(2)}(t)$ is the forecasting value of PLA regression model of market i . We obtain α_1 and α_2 through regression of historical data. We can put the forecasting results of two sub-models into the regression equations and get the final forecasting values of weekly price changes.

Warning model

As is mentioned above, agricultural commodity prices tend to change after keeping constant for a while. No apparent rule is observed, thus the exact moment of the price change is quite hard to predict. The solution of this paper is to preprocess the data and obtain weekly prices. It is important for consumers to know the possible price changes of each single day [30]. Therefore, this paper proposes a price change warning model to quantify the intensity of possible price changes by the output values.

Hypothesis of price fluctuation

First, this model proposes a hypothesis that besides fluctuation around the mean value, all price changes are caused by the change of exogenous variables.

The agricultural commodity is a component of market economy. Its price is irreversibly influenced by other economic variables and exogenous variables including weather and price changes. This kind of change is definitely not a fluctuation around the mean value [31–33]. Therefore, it's a reasonable hypothesis.

The influence brought by exogenous variables will accumulate as time goes by. Due to the uncertainty of the influence, analysis of the influence at a single moment has a huge error. Therefore the next section will propose several methods to deal with exogenous variables, in this way to synchronize the price changes with the accumulation of exogenous variables.

Definition of urgency and sample calculation

The preprocessing of price data in the warning model follows hypotheses raised above, meanwhile we expect to gain daily data with the trend maintained, which means to keep relevant information of every single day. Therefore we use the following way to deal with data:

1. Smoothing. Smooth daily price data, thus we can keep the price trend and eliminate meaningless fluctuation. We use moving average smoothing method on the historical data. Take the parameter value as 15. The price $\theta_i(t)$ of an agricultural commodity in the market i at the moment t is:

$$\theta_i(1) = \theta_i(1).$$

$$\theta_i(2) = \frac{\theta_i(1) + \theta_i(2) + \theta_i(3)}{3}$$

...

$$\theta_i(8) = \frac{\sum_{i=1}^{15} \theta_i(i)}{15}$$

...

$$\theta_i(t) = \frac{\sum_{i=t-7}^{t+7} \theta_i(i)}{15}$$

2. Clustering. In order to synchronize price changes with the accumulative changes of exogenous variables, meanwhile ignoring slight fluctuation, this paper uses cluster

analysis in the data preprocessing. Here we use K-means unidimensional clustering. Set c to be the cluster number [34]:

$$c = \min\{\text{the number of different prices}, 7\}$$

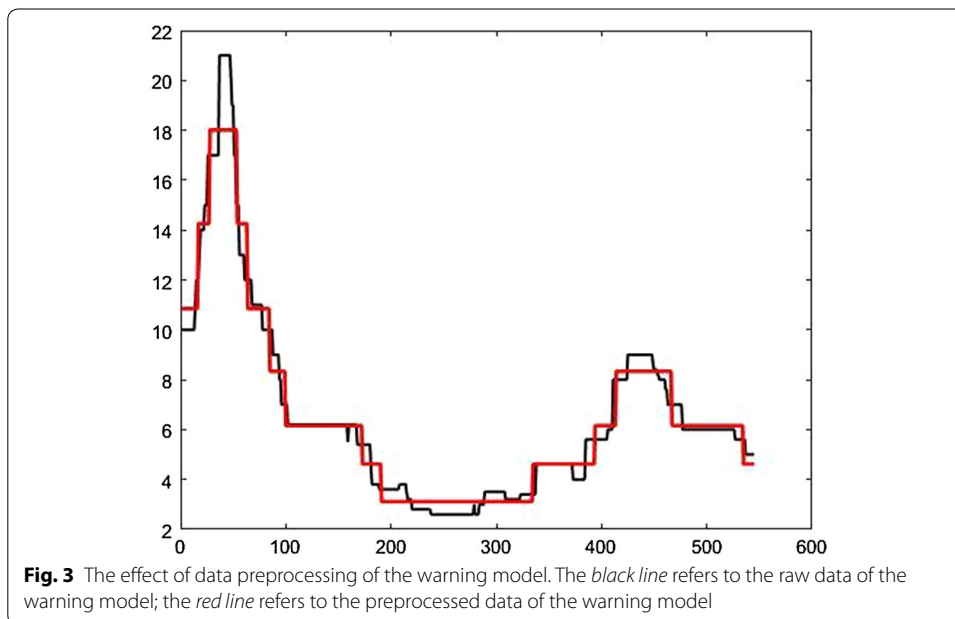
We set c no more than 7, thus we can divide one cluster to be the median value, three clusters higher and three clusters lower, to reflect the stability and huge fluctuation of the price, as showed in Fig. 3. There will be no more than 7 different values of agricultural commodity prices after clustering.

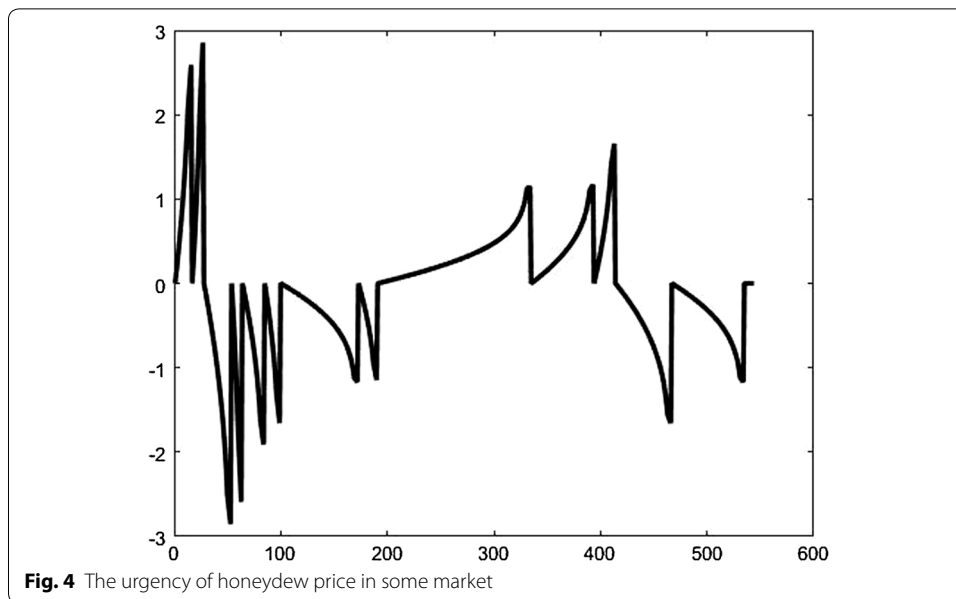
3. Raising the dimension. To the price $p_i(t)$ of an agricultural commodity in the market i at the moment t , set the nearest future price change to be $\delta_i(t)$, which happens $N_i(t)$ days from now, therefore we can get a new daily data $(\delta_i(t), N_i(t))$.
4. Obtaining new variables. After last three steps, we get $(\delta_i(t), N_i(t))$. Now we define some new variables. This paper expects to quantify the range of possible price changes from the values of $\delta_i(t)$ and $N_i(t)$, therefore defines a variable of urgency $U_i(t)$. Suppose that price $\theta_i(t)$ lasts for time $T_i(t)$. Based on experimental effect and quantification purpose, we define the $U_i(t)$ as:

$$U_i(t) = \frac{(T_i(t) - N_i(t)) \cdot \delta_i(t)}{T_i(t) \cdot N_i(t)^{\frac{1}{4}}}$$

If $N_i(t) < 3$, take $N_i(t) = 3$, in order to prevent the urgency from sudden change which makes training and forecasting difficult.

From the definition of $U_i(t)$, we can see that the bigger the price rise is or the sooner the change happens, the stronger the urgency is. So $U_i(t)$ can quantify the urgency degree of price changes and send warning messages. The urgency change of honeydew price in some market is shown as Fig. 4.





The transformation of exogenous variables and sample calculation

Some exogenous variables have their own change trends, therefore showing no conspicuous relationship with the urgency change trends of agriculture commodities. Meanwhile these variables are random. So it is inappropriate to directly use the daily data of these exogenous variables in fitting and forecasting. Because we cannot avoid the random volatility of exogenous variables and the influences caused by their own features.

Therefore we need to find out factors to better reflect how price changes are influenced. Based on this consideration, this paper processes exogenous variables in following four steps:

1. Averaging. We take the average values of last 2 months.
2. Accumulating. We take the accumulating values since the last price change. When the price changes, all these variables are set to 0.
3. Taking the value of that day. We directly assign the real values to the variables.
4. Recording the maximum/minimum values. We assign the maximum/minimum values since the last price change to the variables.

This paper takes urgency as the independent variable, respectively takes the accumulating value of temperature change, whether snowy, foggy or stormy, takes the accumulated maximum values, average values and each day's values of crude oil prices and the exchange rate and thus obtain 14 derivative exogenous variables.

Warning model based on neural networks

This paper builds a BP neural network model [35] to research on exogenous variables and urgency.

The choice of BP neural networks is based on 2 considerations:

1. The relationship between 14 exogenous variables remains unknown, and no research has been conducted about quantifying the exact relationship of exogenous variables and price changes of agricultural commodities. Neural networks have flexible function form consisting of linear and non-linear functional relationship, thus have unique advantage in the forecasting required in this paper. Multi-factor analysis based on neural networks turned out to be effective in some applications in [25].
2. The relationship between exogenous variables and agricultural commodity prices may fluctuation as time goes by. Neural networks model can be updated according to up-to-date historical data.

Set the number of hidden layers to be 1. We choose the node number of hidden layer by mean square error (MSE), and choose LM method as the training algorithm. After the parameters are determined, we can train the training set using neural networks [36, 37].

In fact, the purpose of urgency is to reflect the accumulated effect of exogenous variables. From the definitions of 14 derivative variables, we can see that some of them are monotone as time goes by, and some of them are accumulating.

Trained neural networks can adjust to the urgency every day. The definition of urgency indicates that urgency measures the trend of price changes. High urgency doesn't indicate a certain price change. Instead, it indicates a wider range of price change (if the price really changes).

Considering the asynchronism of price changes and the accumulating of exogenous variables, this paper is conservative with the forecasting value of urgency. We consider the urgency value of the last week's forecasting. To the forecasting value $U_i(t)$ at time t , the adjusted value $U'_i(t)$ is defined as:

$$U'_i(t) = \frac{\text{med}\{U_i(s), s = t - 6, \dots, t - 1, t\} + U_i(t)}{2}$$

Here, $\text{med}\{U_i(s), s = t - 6, \dots, t - 1, t\}$ is the median of the urgency values from day $t - 6$ to day t .

Results and error analysis

We finally obtain two groups of values from the model: the forecasting value of the weekly price change $\theta'_i(t)$ and the daily adjusted price warning urgency value $U'_i(t)$ in the market i at day t . Compare these two groups of values with the true values $\theta_i(t)$ and the adjusted price warning urgency values $U_i(t)$ of true prices.

The following figures compare the forecasting values with the true values, including weekly price forecasting values and urgency forecasting values.

Introduction of sample and results

We've tested over 20 types of agricultural commodities in Beijing based on the prices data from January 2014 to June 2015, including beef and eggs of meat and egg category, we ever and blunt-snout bream of aquatic product category, cowpea and Chinese yam of vegetable category, sweet orange of fruit category, and rice of grain and oil category. We trained the data of former 60 weeks (from January 9, 2014 to March 5, 2015) and tried

to forecast the price changes from week 61 to week 75 (from March 6, 2015 to June 19, 2015). The forecasting results are good.

Error calculation

This paper uses the ratios of mean square error (denoted by MSE_{θ}) and mean absolute error (denoted by MAE_{θ}) to the mean price to measure the error of the forecasting values of weekly price changes. Here we take the price into consideration because the price also determines the growing rate. The higher the price is, the wider the possible growing range is. Therefore, the ratio of the error and the price is a better way to evaluate the result [38]. That is to say:

$$MSE_{\theta} = \frac{\sum_{t=61}^{75} \theta'_i(t) - \theta_i(t)^2}{15 \cdot m}$$

$$MAE_{\theta} = \frac{\sum_{t=61}^{75} |\theta'_i(t) - \theta_i(t)|}{15 \cdot m}$$

$$m = \frac{\sum_{t=61}^{75} \theta_i(t)}{15}$$

As for the forecasting values of the urgency, this paper compares the forecasting urgency values with the urgency values of price changes of the latter 105 days and calculate their mean square error (denoted by MSE_U) and mean absolute error (denoted by MAE_U). Since there is no evaluation standard for the urgency (like the price change to the price), we define the formula in the following way:

$$MSE_U = \frac{\sum_{t=421}^{525} U'_i(t) - U_i(t)^2}{105}$$

$$MAE_U = \frac{\sum_{t=421}^{525} |U'_i(t) - U_i(t)|}{105}$$

Forecasting result

Taking the same agricultural commodity in different markets as samples, we analyze the errors of the time model, the space model, the mixed model in the same time. Meanwhile, we compare the mixed model with some other forecasting models including AR model, grey prediction model and GARCH model. Here we choose some typical time series forecasting methods. AR model is the simplest one. Grey prediction model has good results when only having a little amount of data. GARCH model is frequently used to forecast the variance of time series. As we mentioned in part II, these three models were all used in forecasting price changes of agricultural markets before. By comparing MSE and MAE, we draw conclusion that the mixed model has better forecasting results in most cases. In some other cases when the market is mainly influenced by either time factor or space factor, the forecasting results of the mixed model might be worse than that of ARIMA model or PLS model.

Forecasting results of the same agricultural commodity in different markets

We forecast the prices of cowpea and watermelon in four markets in Beijing. From week 61 to 75, several price fluctuations of cowpea happened with long intervals and wide range. And price fluctuations of watermelon last for short time. Two agricultural commodities have quite different price change trends.

The forecasting results of cowpea prices in Xinfadi Agricultural Wholesale Market, Fengtai District, Beijing and Dayanglu Agricultural and Sideline Products Wholesale Market, Chaoyang District, Beijing are showed in Figs. 5, 6, 7 and 8.

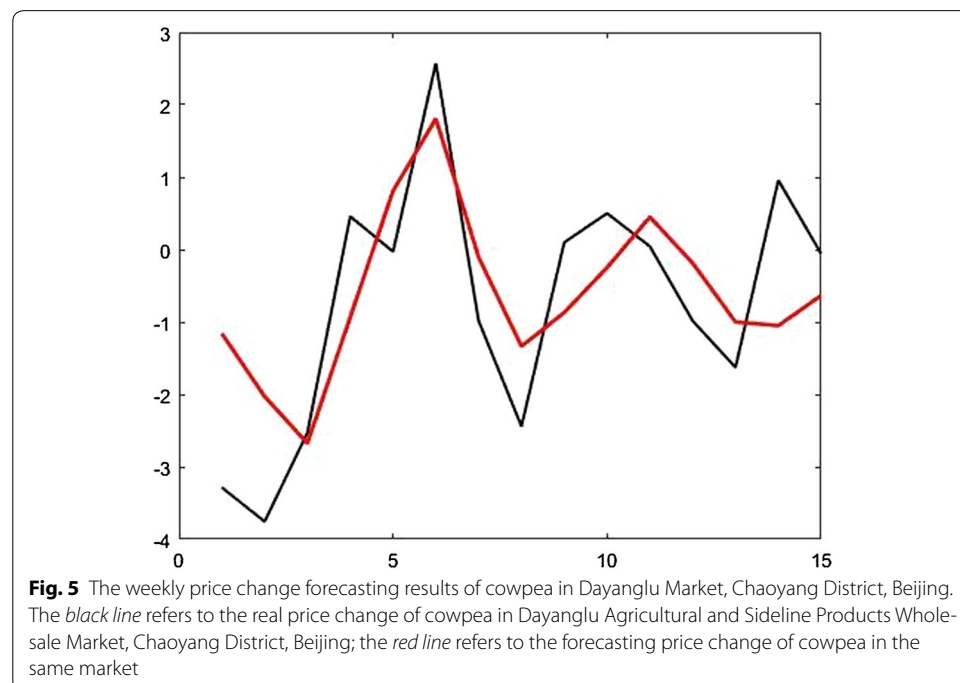
We can see from the figures that the forecasting results of weekly prices trend are nearly consistent with the real data. The forecasting results of the warning values are quite satisfactory, too. We almost precisely forecasted the trend of the price change warning values and rising amplitude.

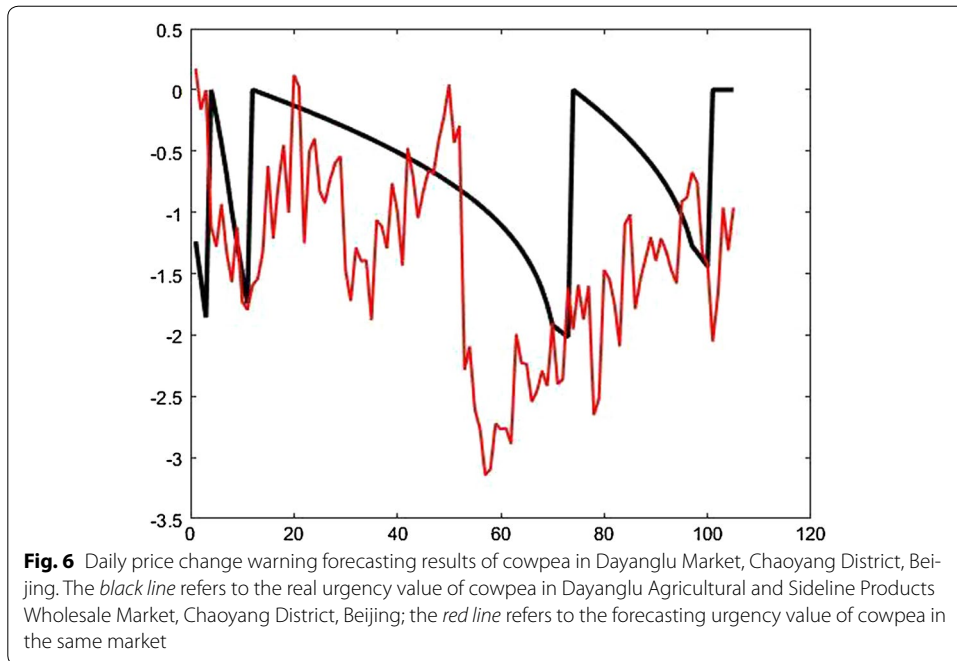
The error analysis of sub-models, the mixed model and urgency values of the cowpea prices forecasting in 4 markets are listed in Table 1.

We can see that the error of the mixed model is relatively smaller, which means that we can decrease the errors by combining the time factor model and the space factor model. The error values are all within 0.4. Cowpea prices are relatively cheap and change in a wide range, so the forecasting effect of weekly prices are good, which we can also see from the figures.

Furthermore, the forecasting results of the weekly prices of cowpea are quite precise in the trend forecasting (whether the price goes up/down), but are less precise in the forecasting of a sudden rising or declining. Time series models have limitation in forecasting a sudden change.

The forecasting results of the watermelon prices in two markets in Beijing are shown in Figs. 9, 10, 11 and 12.





The prices of watermelon have sudden fluctuation in short time periods from week 61 to 75. From the forecasting results of weekly prices, the mixed model has better forecasting results of price changes. In the warning model, as exogenous variables change, the change trend of the forecasting warning values are nearly consistent with real situation.

The error analysis of sub-models, the mixed model and urgency value of the price forecasting of watermelon in 4 markets are list in Table 2.

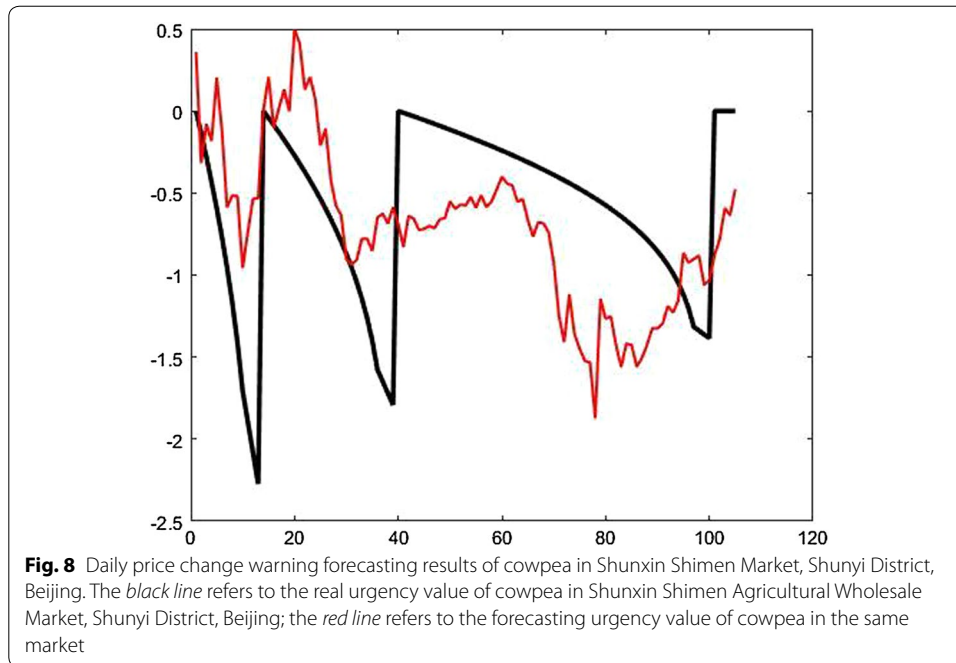
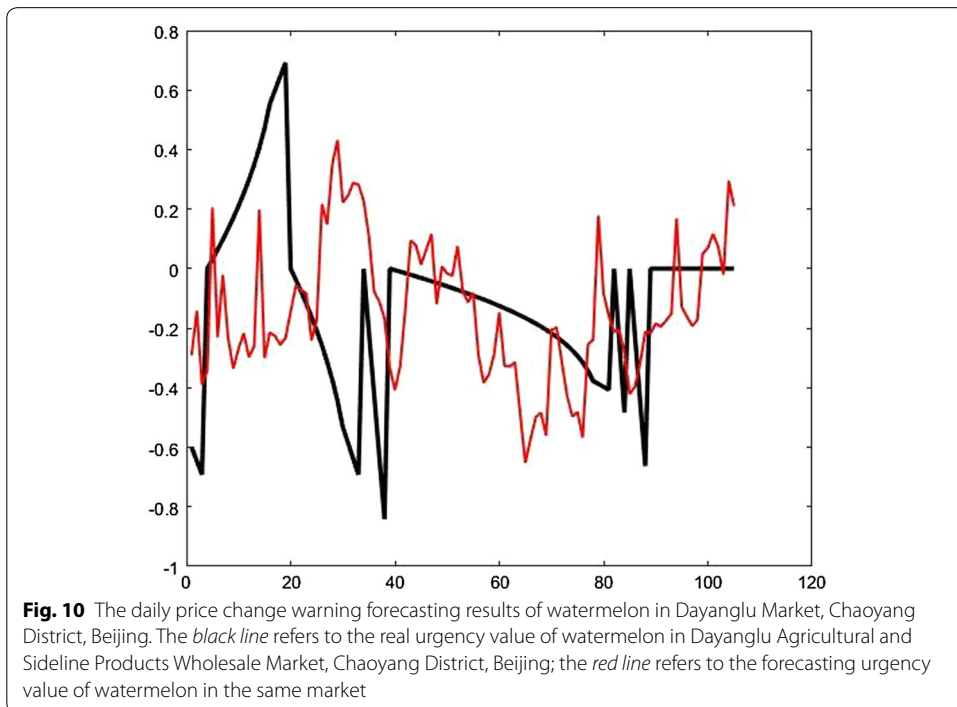
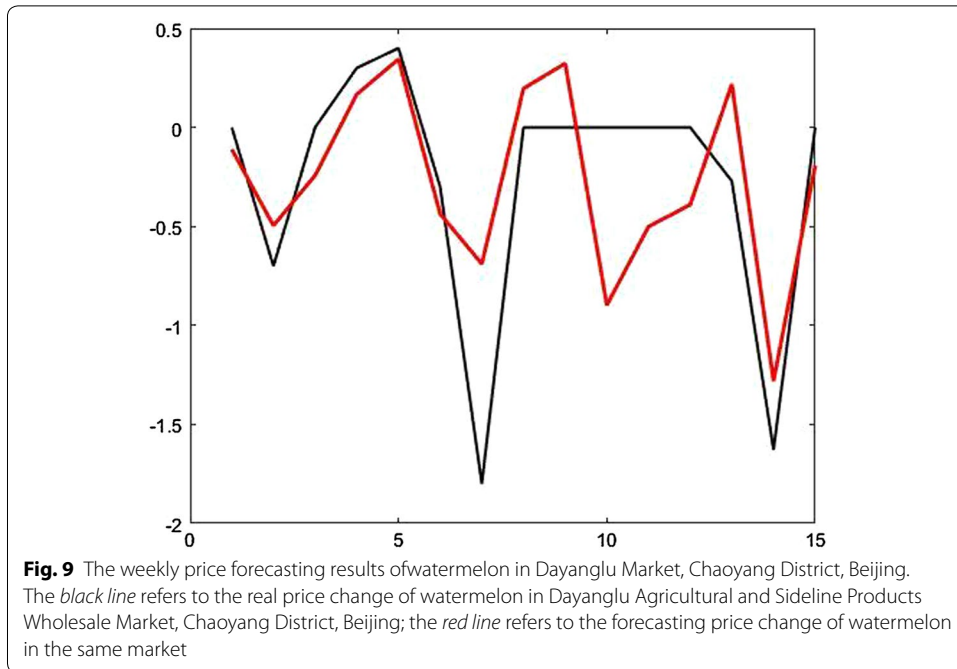


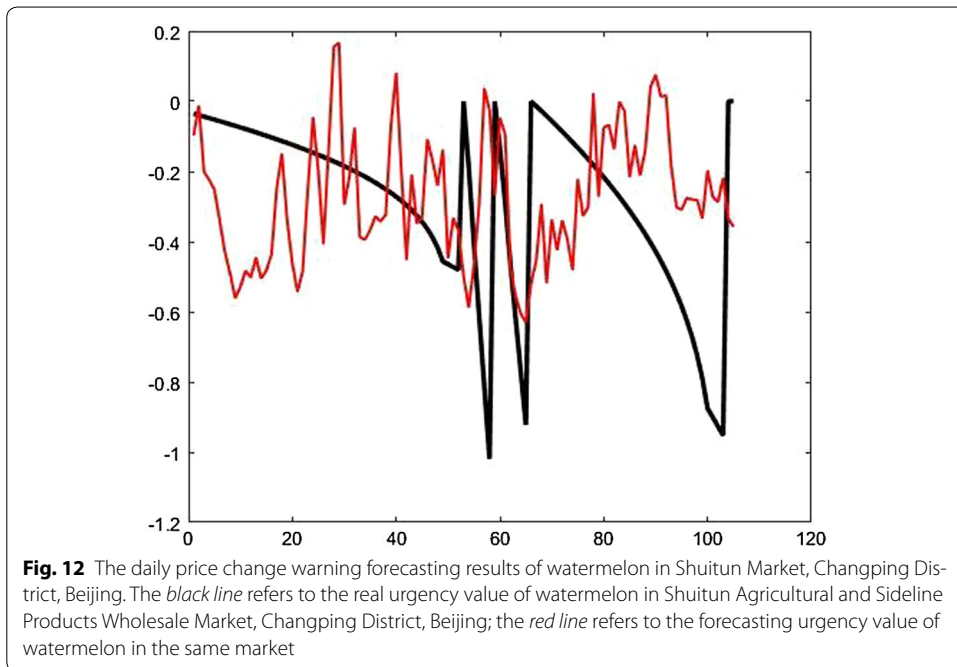
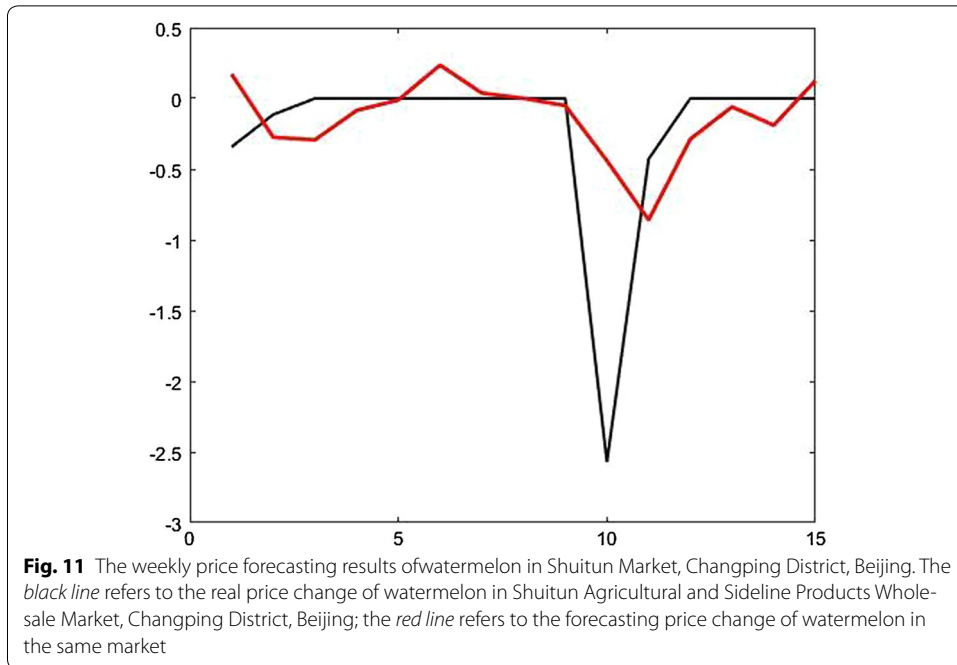
Table 1 The error analysis of the price forecasting of cowpea

The market	Dayanglu Agricultural and Sideline Products Wholesale Market, Chaoyang District, Beijing		Shunxin Shimen Agricultural Wholesale Market, Shunyi District, Beijing	
	MSE	MAE	MSE	MAE
AR model	2.871	1.255	1.918	1.029
Grey prediction model	3.336	1.357	1.067	0.703
Garch model	5.430	1.785	1.649	0.897
Time model	1.476	1.070	1.013	0.673
Space model	1.735	1.735	0.683	0.683
Mixed model	1.321	1.008	0.879	0.670
Warning model	0.884	0.876	0.764	0.727
The market	Chengbei Huilongguan commodity transaction market		Baliqiao Agricultural Wholesale Market, Tongzhou District, Beijing	
	MSE	MAE	MSE	MAE
AR model	1.849	1.058	2.545	1.312
Grey prediction model	4.673	1.705	2.460	1.210
Garch model	19.282	3.742	3.605	1.537
Time model	1.849	1.058	1.163	0.868
Space model	2.811	2.811	1.774	1.774
Mixed model	1.730	1.013	1.179	0.897
Warning model	1.815	1.791	0.482	0.482

We can see that the mixed model of this paper is superior to each sub-model from the error analysis. The mixed model decreased the maximum values of each single model. The forecasting results are more stable. So the mixed model is superior to each sub-model.



From the forecasting results we can see that the mixed model proposed in this paper is good in forecasting weekly price changes and whole trends. The mixed model cannot precisely predict a huge rising or declining. Daily warning values of urgency are a good supplementation to weekly prices forecasting. When the price stays constant for a rather long time, neural networks can precisely forecast the urgency. When prices have huge



fluctuation in a short time, forecasting results will have bigger error as variables do not accumulate.

As for the forecasting of urgency, from a consumer’s perspective, the sooner the price changes, the more important the accuracy of the warning value is. Therefore some errors are allowed to the forecasting results of the warning values when the next price change is still far away.

Table 2 The error analysis of the price forecasting of watermelon

The market	Dayanglu Agricultural and Sideline Products Wholesale Market, Chaoyang District, Beijing		Xinfadi Agricultural Wholesale Market, Fengtai District, Beijing	
	MSE	MAE	MSE	MAE
AR model	5.870	2.142	6.033	2.246
Grey prediction model	0.453	0.360	0.358	0.530
Garch model	0.453	0.360	1.269	1.725
Time model	0.270	0.377	0.407	0.503
Space model	0.363	0.429	0.169	0.335
Mixed model	0.209	0.355	0.346	0.464
Warning model	0.478	0.485	0.196	0.197

The market	Dayanglu Agricultural and Sideline Products Wholesale Market, Chaoyang District, Beijing		Xinfadi Agricultural Wholesale Market, Fengtai District, Beijing	
	MSE	MAE	MSE	MAE
AR model	4.773	1.892	7.297	2.374
Grey prediction model	0.449	0.284	0.537	0.489
Garch model	1.165	1.810	2.060	1.265
Time model	0.374	0.343	0.537	0.489
Space model	0.393	0.280	0.558	0.527
Mixed model	0.354	0.308	0.499	0.472
Warning model	0.104	0.101	0.127	0.127

The prices of the agricultural commodities chosen by this paper often stayed constant for a long while. Thus the forecasting results of weekly prices are slightly fluctuant around the zero value. Therefore some errors are allowed, and the error values are usually small.

Forecasting results of different agricultural commodities in the same market

The commerce department divide agricultural commodities into five categories: food and oil, fruits, vegetables, livestock and aquatic products. This paper forecasts polished round-grained rice, banana, crowndaisy chrysanthemum, beef and grass carp, typical commodities of each category, in Baliqiao Agricultural Wholesale Market, Tongzhou District, Beijing. The forecasting results are shown in Table 3. We notice that the relative errors of the mixed model are all less than 5% except crowndaisy chrysanthemum, the errors of the warning model are all <0.05. From the prices data we can observe that the fluctuation of crowndaisy chrysanthemum’s prices are relatively intensive, which meets the results we analyzed above. Meanwhile, small errors of multiple commodities demonstrate that models in this paper works in multiple commodities.

Table 3 The forecasting errors of five agricultural commodities in Baliqiao Agricultural Wholesale Market, Tongzhou District, Beijing

Commodity	Polished round-grained rice		Beef		Grass carp		Banana		Crowndaisy chrysanthemum	
	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
Mixed model	0.0018	0.0348	0.4031	0.43	0.0534	0.1637	0.0144	0.0959	0.4934	0.5667
Warning model	0.0021	0.0021	0.0358	0.0349	0.1155	0.1159	0.094	0.094	0.699	0.7147

The forecasting results are almost the same as the real situation. Time series models and regression analysis only deal with data set itself, and the warning model deals with exogenous variables. This indicates that agricultural commodity prices are influenced by seasonal factors, prices of all markets and economical variables.

Analysis of special cases

When the prices intensely fluctuate, both models have huge errors. The mixed model has huge errors because sudden changes of a single market has no direct relationship with price changes of other markets. The feature of ARIMA model makes forecasting of huge fluctuation difficult. Neural networks model learns the relationship between exogenous variables and price changes. The feature that huge fluctuations are different from historical data model makes forecasting difficult. Quick changes make exogenous variables hard to accumulate and obtain the best forecasting results.

In fact, the phenomenon mentioned above hardly happens. When price fluctuations are intense, the mixed model precisely predicts the trend. Taking blunt-snout bream in Baliqiao Agricultural Wholesale Market, Tongzhou District, Beijing as an example, its price declined over 10 yuan for once, thus the mixed model is unable to forecast such huge decline. MSE between forecasting values and real values is 7.62. Two errors of the urgent are more than 15. The forecasting of trend, however, is correct, as the model successfully forecasts the trend for four times out of total six times. Furthermore, these six changes can be regarded as three 2-week continuous changes. In this way, the mixed model forecasts correctly for all three times. The forecasting result is shown in Fig. 13.

Advantages of the model

The forecasting method proposed in this paper has some advantages compared with traditional methods:

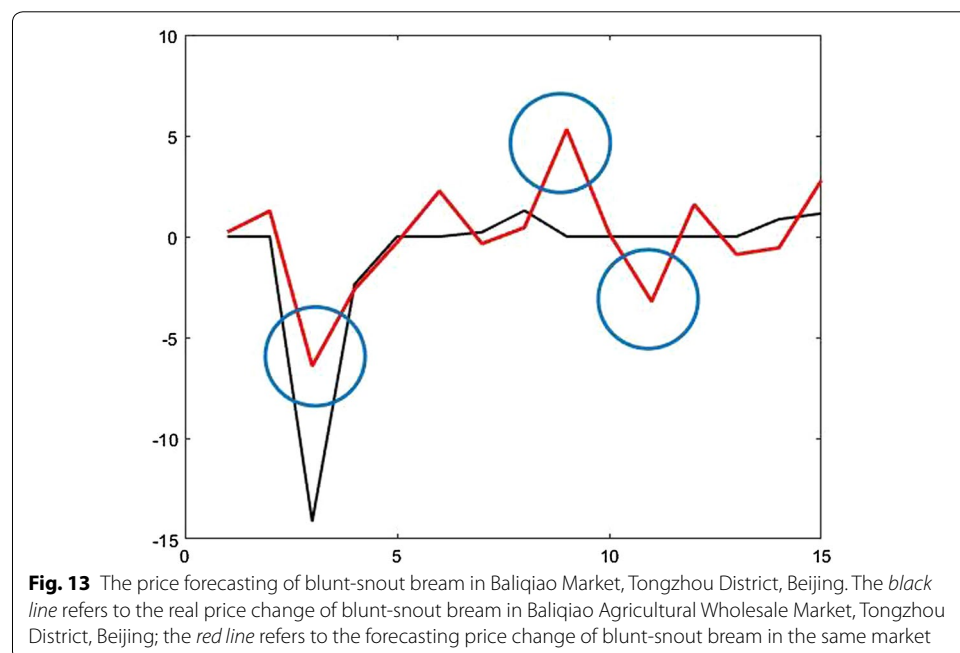


Fig. 13 The price forecasting of blunt-snout bream in Baliqiao Market, Tongzhou District, Beijing. The *black line* refers to the real price change of blunt-snout bream in Baliqiao Agricultural Wholesale Market, Tongzhou District, Beijing; the *red line* refers to the forecasting price change of blunt-snout bream in the same market

1. We notice the weak points of time series models like ARIMA model, then add space factor into the mixed model and modify the model with PLS regression method and analyze non-seasonal factors by price changes of other markets. Forecasting results are more accurate than single time series model and some other typical forecasting models.
2. We design new variable in the warning model, precisely forecast daily price changes of agricultural commodities in a large scale. The forecasting results can provide consumers with meaningful information of the trend of agricultural commodity price change. And the warning model is accurate several days before the price change, which is valuable in application.
3. Unlike traditional methods, the method proposed in this paper can be used for all agricultural commodities in all markets. This method considers several factors and simplify the model designing of price forecasting.

Conclusion and outlook

This paper separates the daily price forecasting problem into ARIMA model, PLS regression and neural networks, obtains weekly price forecasting and daily price change urgency after necessary data processing. The model can be used for a large number of agricultural commodities and the results obtained are accurate, and valuable in consumers' daily lives.

In fact, large-scale forecasting of agricultural commodity prices is a challenging problem. The key to this problem is to quantify various factors that might have influence on the agricultural commodity prices, and to combine the factors with forecasting models. The uncertainty of real data is a challenge that cannot be avoided.

In future, the research of the agricultural commodity prices may can focus on the following aspects:

1. Build and optimize mixed models of multi-factors and make quantitative analysis of price change relationship between related agricultural commodities. Build models of price change relationship of agricultural commodities and use it into forecasting.
2. Make more analysis about consumers and the market, for example to collect and analyze data like the turnovers of agricultural commodity wholesale markets, thus more significant results will come out.
3. Quantify the influence of policies. The policy is an important factor that influences the price in microeconomics. If we can combine the price forecasting with policy quantification, we might be able to forecast the price more precisely.

Authors' contributions

HW1) carried out the conception and design of the research, participated in the statistical analysis of data and data's economy background, meanwhile tested the model and drafted the manuscript. HW2) participated in the statistical analysis and model design of the research, made substantial contribution to draft the manuscript. MZ participated in interpretation of data and helped revising manuscript. WC4) made substantial contribution to the conception and design of the research and participated in critically revising the manuscript. WC5) conceived of the study and participated in its design and helped to draft the manuscript, was involved in revising the manuscript. All authors read and approved the final manuscript.

Author details

¹ State Key Lab of CAD&CG, Zhejiang University, No. 866 Yuhangtang Road, Xihu District, Hangzhou, Zhejiang, China.

² Zhejiang University of Finance and Economy, Hangzhou 310058, China.

Competing interests

The authors declare that they have no competing interests.

Received: 25 September 2016 Accepted: 25 December 2016
Published online: 09 January 2017

References

- Li J. Agriculture price fluctuation analysis of influence factors and countermeasures. Shijiazhuang: Hebei University of Economics and Business, Dissertation; 2013 **(in Chinese)**.
- Li Z, Li G. The short-term price forecasting of meat and egg. *Food Nutr China*. 2010;6:36–40 **(in Chinese)**.
- Wang S. Short-term price analysis and forecasting methods selection of agricultural products—take Apple on Beijing Xin Fadi wholesale market as example. Dissertation, Chinese Academy of Agricultural Sciences; 2009 **(in Chinese)**.
- Lord MJ. Imperfect competition and international commodity trade: theory, dynamics, and policy modelling. *Econ J*. 1992;102(415):1554–6.
- Box G, Jenkins G. Time series analysis: forecasting and control. 5th ed. Hoboken: Wiley; 2015.
- Rausser GC, Carter C. Futures market efficiency in the soybean complex. *Rev Econ Stat*. 1983;65(3):469–78.
- Granger CWJ, Joyeux R. An introduction to long-memory time series models and fractional differencing. *J Time*. 1980;1(1):15–29.
- Barkoulas JT, Labys WC, Onochie JI. Long memory in futures prices. *Financ Rev*. 1999;34(1):91–100.
- Sims CA. Macroeconomics and reality. *Econometrica*. 1980;48(1):1–48.
- Park T. Forecast evaluation for multivariate time-series models: the U.S. cattle market. *West J Agric Econ*. 1990;15(1):133–43.
- Engle RF, Granger CWJ. Co-integration and error correction: representation, estimation, and testing. *Econometrica*. 1987;55(2):251–76.
- Ye L, Li Y, Liu Y, et al. Research on the optimal combination forecasting model for vegetable price in Hainan[M]. Berlin Heidelberg: Springer; 2014.
- Beveridge S, Nelson CR. A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle'. *J Monet Econ*. 1981;7(2):151–74.
- Harvey AC. Time series models. 2nd ed. Cambridge: MIT Press; 1993.
- Davidson R, Labys WC, Lesourd JB. Wavelet analysis of commodity price behavior. *Comput Econ*. 1998;11(1–2):103–28.
- Engle RF. Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica*. 1982;50(4):987–1007.
- Bollerslev T. Generalized autoregressive conditional heteroskedasticity. *J Econom*. 1986;31(3):307–27.
- Kyrtsou C, Labys WC, Terraza M. Noisy chaotic dynamics in commodity market. *Empir Econ*. 2004;29(3):489–502.
- Schroeder MR. Fractals, chaos, power laws: minutes from an infinite paradise. New York: W.H. Freeman; 1991.
- Labys WC. Modeling and forecasting primary commodity prices. London: Routledge; 2006.
- Lapedes AS, Farber RF. Nonlinear signal processing using neural networks: prediction and system modeling//1. San Diego: IEEE international conference on neural networks; 1987.
- Andersen Torben. Volatility and correlation forecasting. *Handb Econ Forecast*. 2015;1(05):777–878.
- Manfredo MR, Leuthold RM, Irwin SH. Forecasting cash price volatility of fed cattle, feeder cattle, and corn: time series, implied volatility, and composite approaches. *Ssrn Electr J*. 1999;33(3):523–38.
- Kroner KF, Kneafsey KP, Claessens S. Forecasting volatility in commodity markets. *J Forecast*. 1995;14(1226):77–95.
- Zheng Yu, Yi Xiuwen, Li Ming, et al. Forecasting fine-grained air quality based on big data. *ACM SIGKDD Int Conf*. 2015;2015:2267–76.
- Xiong T, Li C, Bao Y, et al. A combination method for interval forecasting of agricultural commodity futures prices. *Knowl Based Syst*. 2015;77:92–102.
- Jin G. Data analysis and statistical modeling. Beijing: National Defense Industry Press; 2013 **(in Chinese)**.
- He S. Applied time series analysis. Beijing: Peking University Press; 2003 **(in Chinese)**.
- Giordano FR, Fox WP, Horton SB, et al. A first course in mathematical modeling. 4th ed. Boston: Cengage Learning; 2008.
- Guo C. Farm price data mining and tendency forecast model research. Dissertation, Jinan: Shandong University; 2009 **(in Chinese)**.
- Koirala KH, Mishra AK, D'Antoni JM, et al. Energy prices and agricultural commodity prices: testing correlation using copulas method. *Energy*. 2015;81(3):430–6.
- Gargano A, Timmermann A. Forecasting commodity price indexes using macroeconomic and financial predictors. *Int J Forecast*. 2014;30(3):825–43.
- Harri A, Nalley L, Hudson D. The relationship between oil, exchange rates, and commodity prices. *J Agric Appl Econ*. 2009;41(2):501–10.
- Gao H. Applied multi-variate statistical analysis. Beijing: Peking University Press; 2005 **(in Chinese)**.
- Haykin SS. Neural networks and learning machines. 3rd ed. New Jersey: Pearson Education; 2008.
- Jha GK, Sinha K. Agricultural price forecasting using neural network model: an innovative information delivery system. *Agric Econ Res*. 2013;26(26):229–39.
- Lozano M, Rodriguez FJ, García-Martínez C. A two-stage constructive method for the unweighted minimum string cover problem. *Knowl-Based Syst*. 2015;31(77):103–13.
- Clements MP, Hendry DF. On the limitations of comparing mean square forecast errors: a reply. *J Forecast*. 1993;12(8):669–76.