

Research Article

Integrated Feature Selection of ARIMA with Computational Intelligence Approaches for Food Crop Price Prediction

Yuehjen E. Shao  and Jun-Ting Dai

Department of Statistics and Information Science, Fu Jen Catholic University, New Taipei City, Taiwan

Correspondence should be addressed to Yuehjen E. Shao; stat1003@mail.fju.edu.tw

Received 26 September 2017; Revised 22 May 2018; Accepted 3 June 2018; Published 10 July 2018

Academic Editor: Laszlo T. Koczy

Copyright © 2018 Yuehjen E. Shao and Jun-Ting Dai. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Because of global climate change, lack of arable land, and rapid population growth, the supplies of three major food crops (i.e., rice, wheat, and corn) have been gradually decreasing worldwide. The rapid increase in demand for food has contributed to a continuous rise in food prices, which directly threatens the lives of over 800 million people around the world who are reported to be chronically undernourished. Consequently, food crop price prediction has attracted considerable attention in recent years. Recent integrated forecasting models have developed various feature selection methods (FSMs) to capture fewer, but more important, explanatory variables. However, one major problem is that the future values of these important explanatory variables are not available. Thus, predictions based on these variables are not actually possible. Because an autoregressive integrated moving average (ARIMA) can extract important self-predictor variables with future values that can be calculated, this study incorporates an ARIMA as the FSM for computational intelligence (CI) models to predict three major food crop (i.e., rice, wheat, and corn) prices. Other than the ARIMA, the components of the proposed integrated forecasting models include artificial neural networks (ANNs), support vector regression (SVR), and multivariate adaptive regression splines (MARS). The predictive accuracies of ARIMA, ANN, SVR, MARS, and the proposed integrated model are compared and discussed. Experimental results reveal that the proposed integrated model achieves superior forecasting performance for predicting food crop prices.

1. Introduction

1.1. Background. Everyone needs food. However, not everyone has enough food to survive. Food crops are a primary source of human food, with rice, wheat, and corn being the most widely consumed sources of grains around the world. In this study, food crops refer to plants, which provide food for human consumption, cultivated by agriculturists. The demand and consumption for food crops will rapidly increase in the future, propelled by a 2.3 billion person increase in global population and greater per capita incomes anticipated through the midcentury [1, 2]. According to the Food and Agriculture Organization of the United Nations (FAO) [3], the demand for food is expected to grow substantially by 2050. A major factor for this increase is world population growth. Today, world population has reached

7.6 billion, and we may reach 9.7 billion by 2050. This growth, along with rising incomes in developing countries, is driving up global food demand. Consequently, humanity may directly face one difficulty, which is “will enough food crops be produced at affordable prices or will rising prices drive more of the humanity into hunger?”

The three most important food crops, rice, wheat, and corn, directly provide more than 50% of all calories consumed by the world population. While the area harvested for wheat each year is 214 million ha, the areas harvested for rice and corn are 154 million ha and 140 million ha, respectively [4]. Additionally, the rapid rise in food prices has been a burden on the poor in developing countries, who spend roughly half their household income on food. The issue of being able to affordably purchase the aforementioned food crops plays a very important role in human life.

Therefore, from the human point of view, the prediction of prices for rice, wheat, and corn has become a significant research topic.

1.2. Related Work. Owing to various agricultural and environmental factors, meteorological factors, and biophysical factors, it was indicated in [5] that the exports of rice were very difficult to predict. They employed the autoregressive integrated moving average (ARIMA) and artificial neural network (ANN) methodologies to predict Thailand's rice exports. The wheat price in the Chinese market was predicted by using the ARIMA, ANN, and linear combination models [6]. The overall results showed that the ANN technique was the best prediction model. In Africa, around 40% of the rice consumed is imported [7]. This high dependence on rice imports indicates that Africa is highly exposed to international rice market shocks with sometimes grave consequences for its food security and political stability. In addition, models were constructed to predict the quarterly prices of two types of food crops, barley and wheat [8]. In [9], it was reported that the world food prices would rise around 32% by 2022.

They showed the dynamic relationship between acreage response, food crop yield, and price volatility by developing an optimization model [10]. In [11], it was stated that most countries would like to predict their annual food requirements in order to provide food security for the people. They proposed an artificial intelligent support vector regression (SVR) model to predict the output energy in rice production.

Because food crop prices are seasonally affected [5, 6], this study uses the ARIMA to predict the prices of rice, wheat, and corn. However, the assumptions inherent to the linear form of ARIMA may encounter problems in adopting nonlinear relationships for practical data [12]. Computational intelligence (CI) techniques have been widely used in many forecasting applications because of data-driven features and fewer a priori assumptions. Accordingly, in addition to ARIMA modeling, this study uses CI schemes, including ANN, SVR, and multivariate adaptive regression splines (MARS), for predicting the prices of the three food crops because they allow nonlinearity modeling and provide good forecasting characteristics.

However, CI modeling may face difficulties in its training process for designing an optimal topology owing to the use of a high number of input variables. Therefore, feature selection methods (FSMs) have been incorporated in order to reduce the number of explanatory or predictor variables [13, 14]. In this study, feature selection refers to the process of identifying a subset of relevant explanatory or predictor variables for use during forecasting model construction. This subset of variables contains fewer but more important input variables that aid in predicting the outcome.

Feature selection techniques have become a research hot spot in many forecasting applications. For example, in order to accurately predict wind speed, an SVR forecasting model with a feature selection procedure called phase space reconstruction has been proposed [15]. Additionally, a set of general ARIMA models was used for performance comparison. A hybrid forecasting model with a feature selection technique

based on mutual information, extreme learning machines, and bootstrap techniques was proposed to predict day-ahead electricity prices [16]. The authors of [17] proposed a hybrid filter-wrapper approach to predict electrical load and the price of electricity. The feature selection method used in their approach identifies a minimum subset of the most informative features by considering the relevancy, redundancy, and interactions of candidate inputs. In [18], the authors proposed a two-step approach that selects a set of candidate features based on data characteristics. A hybrid ANN-based model was then used to predict the day-ahead price of electricity. A hybrid forecasting methodology was designed to predict electrical load in [19]. The study used a feature selection method that was developed based on entropy and several CI techniques. The prediction performance of the proposed model was compared to various ARIMA models. A regression model was proposed to predict online sales in [20]. Additionally, a feature selection methodology based on a multiobjective evolutionary algorithm was combined with the forecasting model. A combined ANN and Kalman filter approach was proposed to predict wind speed in [21]. An ARIMA feature selection method was used to determine the input structure for their hybrid model.

When reviewing related research, we noticed that, although many studies have focused on using FSMs to obtain accurate forecasts, little attention has been devoted to the use of FSMs for food crop price prediction. Additionally, we encounter practical problems when using FSMs to predict food crop prices. As mentioned earlier, one of the major purposes for using an FSM is to select a subset of explanatory variables for further use in forecasting. However, the future values of the explanatory variables that are selected by the FSM are unknown. Relatively little research has attempted to address this problem. As a consequence, predictions cannot be computed even though the most important explanatory variables have been identified by effective FSMs. We propose an integrated forecasting model to remedy the problems caused by unknown explanatory variables. The proposed integrated model employs an ARIMA as an FSM to capture important self-predictor variables, which then serve as input variables for the ANN, SVR, and MARS models. Because the future values of a self-predictor variable can be computed based on its previous and current values, food crop price predictions can be obtained.

In this study, we propose four single-stage models and three integrated models for predicting food crop prices. A real monthly dataset was obtained, which contains the prices of rice, wheat, and corn from January 1990 to September 2015. This real dataset makes it possible to compare predictions of food crop prices using single-stage models and integrated models. The remainder of the study is organized as follows: Section 2 presents the modeling methodologies of different forecasting schemes. In Section 3, we describe the design structure for our single- and two-stage integrated models. Real data on the prices of three food crops are collected to verify the single-stage models and integrated models. Section 4 contains forecasting comparisons for four single-stage models and three integrated models. Some practical implications and other concerns are addressed in

Section 5. The final section summarizes our research findings and contains our conclusions.

2. Forecasting Methodologies

2.1. Autoregressive Integrated Moving Average. Because seasonal effects are possibly involved in food crop price forecasting, seasonal ARIMA modeling should be developed [22]. A seasonal ARIMA model can be described as follows:

$$\phi_p(B)\Phi_P(B^L)(1-B)^d(1-B^L)^D Z_t = \delta + \theta_q(B)\Theta_Q(B^L)a_t, \quad (1)$$

where

B is the backward shift operator, defined as $B^j y_t = y_{t-j}$;

L is the number of seasons in a year, and $L = 12$ for monthly data;

d is the values of nonseasonal difference transformations;

D is the values of seasonal difference transformations;

Z_t is the working series, which are stationary after fitting a suitable difference transformation from original time series Y_t ;

δ is an unknown constant;

a_t is the white noise at time t , which is independent and identical (iid) with normal distribution;

p is the order of nonseasonal autoregressive (AR) models;

P is the order of seasonal AR models;

q is the order of nonseasonal moving average (MA) models;

Q is the order of seasonal MA models;

$\phi_p(B)$ is a polynomial function for a nonseasonal AR model, defined as $\phi_p(B) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$;

$\Phi_P(B^L)$ is a polynomial function for a seasonal AR model, defined as $\Phi_P(B^L) = (1 - \Phi_{1,L} B^L - \Phi_{2,L} B^{2L} - \dots - \Phi_{P,L} B^{PL})$;

$\theta_q(B)$ is a polynomial function for a nonseasonal MA model, defined as $\theta_q(B) = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q)$;

$\Theta_Q(B^L)$ is a polynomial function for a seasonal MA model, defined as $\Theta_Q(B^L) = (1 - \Theta_{1,L} B^L - \Theta_{2,L} B^{2L} - \dots - \Theta_{P,L} B^{QL})$.

The original nonstationary time series should be transformed into a stationary working series through differencing. Typically, the appropriate transformations can be performed by four combinations of d and D ; that is, $(d, D) = (0, 0)$, $(d, D) = (1, 0)$, $(d, D) = (0, 1)$, and $(d, D) = (1, 1)$, respectively. Once the stationary working series has been attained, we can observe the behavior of the sample autocorrelation function (ACF) and sample partial autocorrelation function (PACF) to determine the values of p , P , q , and Q for the seasonal ARIMA models.

They prefer using the maximum likelihood (ML) technique to obtain estimates for model parameters [22]. For the ML technique, the likelihood function is maximized via nonlinear least squares using Marquardt's method. Because the ML approach is more computationally expensive than conditional least squares (CLS) estimates, most computer packages employ CLS as a default approach. LS refers to the parameter estimate associated with the smallest sum of

squared errors (SSE). For the CLS approach, we assume that the number of past unobserved errors is zero. The data series y_t can be represented as follows:

$$y_t = \sum_{i=1}^{\infty} \pi_i y_{t-i} + a_t. \quad (2)$$

The π weights are calculated as follows:

$$\frac{\phi(B)}{\theta(B)} = 1 - \sum_{i=1}^{\infty} \pi_i B^i. \quad (3)$$

The CLS approach should produce parameter estimates that can minimize the following:

$$\text{SSE} = \sum_{t=1}^n (a_t^\wedge)^2 = \sum_{t=1}^n \left(y_t - \sum_{i=1}^{\infty} \pi_i^\wedge y_{t-i} \right)^2, \quad (4)$$

where the unobserved past values of y_t are set to 0 and π_i is computed from the estimates of the ARIMA model parameters ϕ and θ at each iteration.

After estimating the model parameters, the diagnostic checks for testing the lack of fit of the seasonal ARIMA model should be conducted. A logical way to check the adequacy of a seasonal ARIMA model is to analyze the residuals obtained from the underlying model. The Ljung-Box test was developed to examine whether the first K sample autocorrelations of residuals indicate adequacy of the model [23]. The null hypothesis for this test is that the first K autocorrelations are jointly zero; that is,

$$H_0 : \rho_1 = \rho_2 = \dots = \rho_k = 0. \quad (5)$$

The Ljung-Box statistics are described as

$$Q = n_d(n_d + 2) \sum_{l=1}^k (n_l - l)^{-1} r_l^2(\hat{a}), \quad n_d = n - d, \quad (6)$$

where

Q is the Ljung-Box statistics, and the asymptotic distribution of Q follows a chi-square distribution with $k - n_p$ degrees of freedom;

n is the number of observations;

d is the degree of nonseasonal differencing;

n_p is the number of parameters other than δ that must be estimated in the ARIMA model under consideration;

$r_l^2(\hat{a})$ is the square of the sample autocorrelation of the residuals at lag l .

The Ljung-Box test rejects the null hypothesis, which indicates that the underlying model has significant lack of fit if

$$Q > \chi_\alpha^2(k - n_p), \quad (7)$$

where $\chi_\alpha^2(k - n_p)$ is the chi-square distribution table value with $k - n_p$ degrees of freedom and significance level α .

2.2. Artificial Neural Networks. ANN is composed of a large number of highly interconnected processing elements, which are referred to as nodes or neurons, working in parallel to

solve a particular problem. While the leftmost layer of the network is referred to as the input layer, the rightmost layer is called the output layer. The middle layer is called the hidden layer. In the ANN structure, each layer consists of a number of nodes connected by links. The ANN contains nodes that are connected to themselves, enabling a node to influence other nodes and itself. ANN includes input variables and output variables. In addition, a certain number of nodes are contained in the hidden layer, and the hidden nodes are nonlinear functions of the input variables. The output variable is a function of the nodes in the hidden layer.

ANN modeling is briefly described as follows. For an ANN model, the relationship between inputs ($\mathbf{I}_1, \mathbf{I}_2, \dots, \mathbf{I}_a$) and output (O) can be represented as

$$O = \alpha_0 + \sum_{j=1}^b \alpha_j g \left(\delta_{0j} + \sum_{i=1}^a \delta_{ij} \mathbf{I}_i \right) + \varepsilon, \quad (8)$$

where α_j ($j = 0, 1, 2, \dots, b$) and δ_{ij} ($i = 0, 1, 2, \dots, a; j = 0, 1, 2, \dots, b$) are the model connection weights; a is the number of input nodes; b is the number of hidden nodes; and ε is the error term. In the hidden layer, the transfer function is often used with a logistic function.

$$g(z) = \frac{1}{1 + \exp(-z)}. \quad (9)$$

Consequently, the ANN transports nonlinear functional mapping from the inputs ($\mathbf{I}_1, \mathbf{I}_2, \dots, \mathbf{I}_a$) to the output O ; namely,

$$O = f(\mathbf{I}_1, \mathbf{I}_2, \dots, \mathbf{I}_a, \mathbf{w}) + \varepsilon, \quad (10)$$

where \mathbf{w} is a vector of model parameters and f is a function determined by the ANN structure and connection weights.

For the ANN structure, the nodes in the input layers receive input signals from an external source and the nodes in the output layers generate the target output signals. The output of each neuron in the input layer is the same as the input to that neuron. The hidden layers adjust the weights of those inputs until the neural network's error is minimized. For ANN processing, backpropagation is one method for computing the error contribution of each neuron after a batch of data is processed. This method can be used to adjust the weight of each neuron, thereby completing the learning process for that case. For each neuron j in the hidden layer and each neuron k in the output layer, the net inputs are given by

$$\begin{aligned} \text{net}_j &= \sum_i w_{ji} \times o_i, \\ \text{net}_k &= \sum_j w_{kj} \times o_j, \end{aligned} \quad (11)$$

where $i(j)$ is a neuron in the previous layer, $o_i(o_j)$ is the output of node $i(j)$, and $w_{ji}(w_{kj})$ is the connection weight from neuron $i(j)$ to neuron $j(k)$. The neuron outputs are given by

$$\begin{aligned} o_i &= \text{net}_i, \\ o_j &= \frac{1}{1 + e^{-(\text{net}_j + \theta_j)}} = f_j(\text{net}_j, \theta_j), \\ o_k &= \frac{1}{1 + e^{-(\text{net}_k + \theta_k)}} = f_k(\text{net}_k, \theta_k), \end{aligned} \quad (12)$$

where $\text{net}_j(\text{net}_k)$ is the input signal from the external source to node $j(k)$ in the input layer and $\theta_j(\theta_k)$ is a bias.

The generalized delta rule is the conventional technique used to derive the connection weights in the network. First, a set of random numbers is assigned to the connection weights. Then for a presentation of a pattern p with a target output vector $\mathbf{t}_p = [t_{p1}, t_{p2}, \dots, t_{pM}]^T$, the sum of squared errors to be minimized is given by

$$E_p = \left(\frac{1}{2} \right) \sum_{j=1}^M t_{pj} - o_{pj}^2, \quad (13)$$

where M is the number of output nodes. By minimizing the error E_p using the gradient descent technique, the connection weights can be updated by applying the following equations:

$$\Delta w_{ji}(p) = \eta \delta_{pj} o_{pj} + \alpha \Delta w_{ji}(p-1), \quad (14)$$

where for output nodes,

$$\delta_{pj} = (t_{pj} - o_{pj}) o_{pj} (1 - o_{pj}), \quad (15)$$

and for hidden nodes,

$$\delta_{pj} = o_{pj} (1 - o_{pj}) \sum_k (\delta_{pk} \times w_{kj}). \quad (16)$$

Note that the learning rate affects the network's generalization ability and learning speed to a great extent.

2.3. Support Vector Regression. SVR is an adaptation of the support vector machine (SVM), one of the most powerful techniques in CI [24, 25]. The basic concept of SVR is to find a model function $f(x)$ to represent the relationship between the features and target. The modeling of SVR can be described as follows. Suppose

$$\mathbf{y} = f(\mathbf{x}) = (\mathbf{w} \cdot \Phi(\mathbf{x})) + b, \quad (17)$$

where \mathbf{x} and \mathbf{y} represent the model input and output vectors, \mathbf{w} is the weight vector, b is a constant, $\Phi(\mathbf{x})$ denotes a mapping function in the feature space, and $(\mathbf{w} \cdot \Phi(\mathbf{x}))$ describes the dot production in the feature space F .

Typical regression modeling estimates the coefficients by minimizing the square error, which can be considered as an empirical risk based on loss function. The ε -insensitivity loss function was introduced [25] and can be described as follows:

$$L_\varepsilon(f(x), y) = \begin{cases} |f(x) - y| - \varepsilon, & \text{if } |f(x) - y| \geq \varepsilon, \\ 0, & \text{otherwise,} \end{cases} \quad (18)$$

where y is the model output and ε is the region of ε -insensitivity. When the predicted value falls into the band area, the loss is zero. However, when the predicted value falls

outside the band area, the loss is defined as the difference between the predicted value and margin.

When both empirical risk and structure risk are considered, the SVR can be designed to minimize the following quadratic programming problem [25]:

$$\begin{aligned} \min \quad & \left(\frac{1}{2}\right)\|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*), \\ \text{subject to} \quad & (y_i - (w \cdot \Phi(x_i))) - b \leq \varepsilon + \xi_i, \\ & (w \cdot \Phi(x_i)) + b - y_i \leq \varepsilon + \xi_i^*, \\ & \xi_i, \xi_i^* \geq 0, \\ & \text{for } i = 1, \dots, n, \end{aligned} \quad (19)$$

where $i = 1, \dots, n$ is the number of training observations; ξ_i, ξ_i^* is the empirical risk; $(1/2)\|w\|^2$ is the structure risk preventing overlearning and lack of applied universality; and C is a modifying coefficient representing the trade-off between empirical and structure risks. With an appropriate modifying coefficient C , band area width ε , and kernel function, the optimum value of each parameter can be solved by the Lagrange procedure. In addition, the general form of the SVR-based regression function can be expressed as follows [25]:

$$f(x, w) = f(x, r, r^*) = \sum_{i=1}^N (r_i - r_i^*) K(x, x_i) + b, \quad (20)$$

where r_i and r_i^* are the Lagrangian multipliers and satisfy the equality $r_i r_i^* = 0$. In addition, $K(x, x_i)$ is the kernel function. The values of the kernel are equal to the inner product of two vectors, \mathbf{x}_i and \mathbf{x}_j , in the feature space $\Phi(\mathbf{x}_i)$ and $\Phi(\mathbf{x}_j)$; that is, $K(x_i, x_j) = \Phi(\mathbf{x}_i) \Phi(\mathbf{x}_j)$. Because the radial basis function (RBF) is the most commonly used kernel function [26], we employ it for the experimental study. The RBF is written as follows:

$$K(x_i, x_j) = \exp\left(\frac{-\|x_i - x_j\|^2}{2\sigma^2}\right), \quad (21)$$

where σ is the width of the RBF.

2.4. Multivariate Adaptive Regression Splines. MARS was developed for solving regression-type problems [27]. It is a nonparametric regression procedure that makes no assumption about the functional relationship between the response and explanatory variables. MARS modeling is based on a divide-and-conquer strategy where training datasets are partitioned into separate regions, each of which is assigned to its own regression equation. Consequently, MARS is appropriate and effective for problems with more than two input variables. Particularly, MARS can select important explanatory variables and relationships for complex data structures that often hide in higher dimensional data series.

A MARS model can be described as follows [27]:

$$f(x) = \beta_0 + \sum_{m=1}^M \beta_m \prod_{j=1}^{J_m} [S_{jm}(x_{v(j,m)} - l_{jm})], \quad (22)$$

where β_0 and β_m are the parameters, M is the number of basis functions (BFs), J_m is the number of knots, S_{jm} takes on values of either 1 or -1 and indicates the right or left sense of the associated step function, $v(j, m)$ is the label of the independent variable, and l_{jm} is the knot location. The optimal MARS model is determined in a two-step procedure. In the first step, a model is grown by adding BFs until an overly large model is obtained. The BFs are then deleted in the order of least contribution to most contribution by using the generalized cross-validation (GCV) criterion in the second step. The measure of variable importance is provided by observing the decrease in the computed GCV values when a variable is removed from the model. The GCV is described as follows:

$$\text{GCV}(M) = \frac{(1/N) \sum_{i=1}^N [y_i - f_M(x_i)]^2}{[1 - (C(M)/N)]^2}, \quad (23)$$

where N is the number of observations and $C(M)$ is the cost penalty measures of a model containing M BF.

3. Food Crop Price Forecasting

3.1. Datasets and Forecasting Criteria. In this study, the price data of the three most important food crops were collected for the period of January 1990 to September 2015 from the web sites of IndexMundi [28–31]. The datasets consist of 309 records for each food crop price. Among them, the first 297 data records were used to develop the different forecasting models, while the remaining 12 data records were used to perform model validation. This study has presented an experiment based on a larger population of data and has foreseen obtaining accurate forecasts of food crop price, with the forecast horizon for the out-of-sample forecasting experiment of 12. Good prediction of prices may assist in planning agricultural supply, but the investigated data does not contain any information on population.

The forecasting capability of the models was compared using three forecasting accuracy criteria, including mean absolute percentage error (MAPE), root mean square error (RMSE), and mean absolute error (MAE). These forecasting measurements are expressed as follows:

$$\begin{aligned} \text{MAPE} &= \left(\frac{1}{n}\right) \sum_{t=1}^n \frac{|e_t|}{Y_t} \times 100\%, \\ \text{MAE} &= \left(\frac{1}{n}\right) \sum_{t=1}^n |e_t|, \\ \text{RMSE} &= \sqrt{\left(\frac{1}{n}\right) \sum_{t=1}^n (e_t)^2}, \end{aligned} \quad (24)$$

where e_t is the value of the residual at time t and Y_t is the observation at time t .

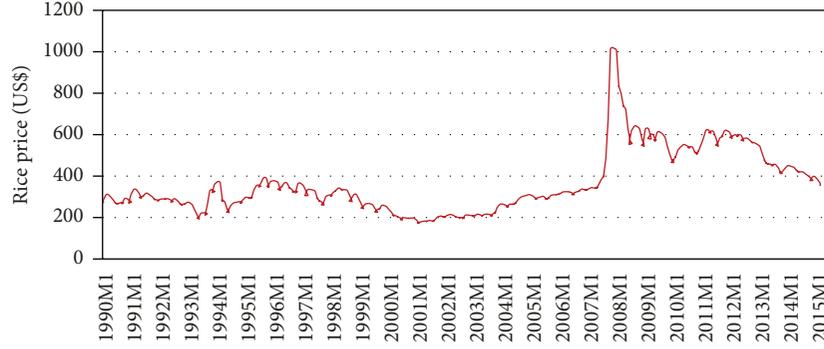


FIGURE 1: Time plot for rice price (unit: US dollars per metric ton).

Obviously, it can be noted that the lower the MAPE, M AE, and RMSE values, the closer the forecasted values to the actual values.

3.2. Forecast Modeling of Rice Prices. Figure 1 shows the original time plot for the rice price data series. This study uses an SAS package to run the ARIMA modeling. Table 1 shows the parameter estimates, and all the estimated parameters are significantly different from zero. In Table 1, the notations “AR1,1,” “AR1,2,” “AR1,3,” and “AR1,4” correspond to the parameters ϕ_1 , ϕ_2 , ϕ_7 , and ϕ_{13} of the AR model. The ARIMA model presented in Table 1 was a subset model. Since the parameters, ϕ_1 , ϕ_2 , ϕ_7 , and ϕ_{13} , are significantly different from zero, this study refers to the model in Table 1 as an AR (1, 2, 7, and 13) model.

In addition, Table 2 shows the Ljung-Box test results. As mentioned earlier, the Ljung-Box statistic can be used to check the fit of an ARIMA model. A simple method to derive the hypothesis testing result is described as follows. We should reject the model under consideration by setting the type I error equal to α if and only if the p value is less than α . Generally, α is set equal to 0.05. The SAS package computes the Ljung-Box statistics Q and their associated p values for values of K equal to 6, 12, 18, 24, 30, and 36. In Table 2, the first column contains the values of K , the second column contains the values of Q , and the fourth column contains the associated p values. By studying Table 2, we observe that all the associated p values are greater than α . Accordingly, a conclusion can be made concerning the fit of the underlying ARIMA model. Thus, the model expressed in (25) is suitable for modeling rice prices.

$$Z_t = 0.46514Z_{t-1} - 0.502Z_{t-2} - 0.17055Z_{t-7} - 0.11264Z_{t-13} + a_t, \quad (25)$$

where

$$Z_t = (y_t - y_{t-1}). \quad (26)$$

For ANN designs, there is no fixed mode to decide the number of hidden nodes. Too few hidden nodes confine the network generalization capability, whereas too many hidden nodes may lead to overtraining difficulties. Therefore, in this study, we consider the hidden nodes to be set from $(2n - 2)$ to $(2n + 2)$ if $n \leq 8$, where n represents the number of input

TABLE 1: Parameter estimates for rice prices.

Parameter	The ARIMA procedure				Lag
	Estimate	Standard error	t value	Approx Pr > t	
AR1,1	0.46514	0.05586	8.33	<0.0001	1
AR1,2	-0.23278	0.05559	-4.19	<0.0001	2
AR1,3	-0.17055	0.05183	-3.29	0.0011	7
AR1,4	-0.11264	0.05170	-2.18	0.0301	13

variables. When $n > 8$, we consider the hidden nodes to be set from $(n - 2)$ to $(n + 2)$. In this study, we denote the term $\{n_i - n_h - n_o\}$ for ANN parameter settings, where n_i is the number of neurons in the input layer, n_h stands for the number of neurons in the hidden layer, and n_o represents the number of neurons in the output layer, respectively. Additionally, this study employs the learning rate for all ANN models at the default value (i.e., 0.01) to ensure consistency [32]. The network topology with the minimum MAPE is considered as the optimal network topology, due to the fact that the MAPE is one of the most important performance measurements of forecasting capability.

In this study, because there are no explanatory variables available, the inputs of ANN modeling for food crop price forecasting becomes unfeasible. Consequently, this study uses self-predictor variables to serve as the inputs for ANN. In here, the self-predictor variable is defined as the lags or past values of a variable. This study proposes two input designs for the modeling of CI techniques. Because seasonal effects may influence food crop price forecasts, the first design rationally selects the preceding 12 observations to forecast food crop prices at time t . Accordingly, the first design used 12 self-predictor variables (i.e., y_{t-1} , y_{t-2} , y_{t-3} , ..., and y_{t-12}) to serve as input variables and employed a single variable (i.e., y_t) to serve as the output for ANN modeling. That is, the first design considered 12 input nodes and one output node for the ANN structures. This design is denoted as ANN₁.

The second input design for CI technique modeling is performed using FSM. For this proposed modeling, this study incorporated FSMs with ANN, SVR, and MARS to develop forecasting models. The proposed input design used ARIMA as an FSM to extract important self-predictor

TABLE 2: Ljung-Box test results for ARIMA modeling of rice prices.

To Lag	Chi-square	DF	Autocorrelation check of residuals						
			Pr > ChiSq	Autocorrelations					
6	4.61	2	0.0995	-0.002	0.007	-0.024	-0.073	0.006	-0.096
12	6.34	8	0.6097	0.009	0.007	0.016	0.027	-0.031	0.059
18	13.10	14	0.5190	-0.023	0.028	0.028	-0.09	0.061	-0.087
24	14.40	20	0.8096	0.059	0.012	0.008	-0.006	-0.012	-0.015
30	19.29	26	0.8242	-0.102	-0.045	-0.025	-0.002	0.026	0.036
36	20.10	32	0.9494	0.018	-0.022	-0.002	-0.024	-0.003	0.032

TABLE 3: Forecasting measurements for two ANN designs (rice price forecasting).

ANN ₁ topology	MAE	RMSE	MAPE (%)	ARIMA-ANN topology	MAE	RMSE	MAPE (%)
{12-10-1}	15.438	16.648	3.939	{4-6-1}	10.384	12.053	2.654
{12-11-1}	14.737	16.022	3.765	{4-7-1}	10.365	12.037	2.650
{12-12-1}	15.001	16.262	3.825	{4-8-1}	10.374	12.044	2.652
{12-13-1}	15.182	16.499	3.869	{4-9-1}	10.379	12.062	2.653
{12-14-1}	13.915	15.320	3.557	{4-10-1}	10.389	12.039	2.655

variables, which serve as the inputs for the ANN, SVR, and MARS models. Because the ARIMA model for rice price forecasting is obtained in (25), it indicates that Z_{t-1} , Z_{t-2} , Z_{t-7} , and Z_{t-13} will influence Z_t . Consequently, this study selected Z_{t-1} , Z_{t-2} , Z_{t-7} , and Z_{t-13} to serve as self-predictor, or input, variables for the proposed integrated modeling on the rice price data series. In this study, this second input design is denoted as ARIMA-ANN.

To forecast the rice prices, the two ANN models contain twelve and four input nodes for the first and second designs, respectively. The hidden nodes were selected as 10, 11, 12, 13, and 14 for the first design and 6, 7, 8, 9, and 10 for the second design. After performing the ANN modeling, the ANN₁ design showed that the {12-14-1} structure provided the best results and minimum testing MAPE for rice prices. For the ARIMA-ANN design, the best structure was {4-7-1}. Three forecasting measurements for different settings of the ANN topologies for the two designs are shown in Table 3.

For modeling SVR on rice prices, this study adopted the input structure as ANN modeling. SVR modeling employed two designs for the input variables. The first design used 12 variables, y_{t-1} , y_{t-2} , y_{t-3} , \dots , and y_{t-12} as the input variables and y_t as the output variable. The second design used four self-predictor variables, Z_{t-1} , Z_{t-2} , Z_{t-7} , and Z_{t-13} to serve as the input variables and Z_t as the output variable. The first and second SVR designs are denoted as SVR₁ and ARIMA-SVR, respectively. Because the parameter settings of C and ϵ often affect the performance of SVR modeling, an analytic parameter selection approach and grid search are used in this study [26]. Accordingly, we denote the term $\{\epsilon, C\}_{\text{SVR}}$ as the best SVR parameter settings. After performing the SVR modeling, the SVR₁ design reported that the parameter settings of $\{2^{-4}, 2^{11}\}_{\text{SVR}}$ provided MAPE=6.912% for rice price forecasting. For the second design, ARIMA-SVR, the parameter settings of $\{2^{-5}, 2^9\}_{\text{SVR}}$ provided MAPE=2.057%.

For MARS modeling, this study also adopted the same input structure of ANN or SVR modeling. The first and second MARS designs are denoted as MARS₁ and ARIMA-MARS, respectively. Table 4 lists the variable selection results and BFs after modeling rice price data using MARS₁. In Table 4, the first column contains the variables that should be included in the model, the second column contains the relative importance of the variables that are listed in the first column, the third column contains the various BFs, and the final column contains the estimated coefficient values for the BFs that are listed in the third column. Based on the results in Table 4, it can be seen that three input variables (i.e., y_{t-12} , y_{t-11} , and y_{t-10}) played important roles in building the MARS forecasting models. Their relative importance values (%) are 100, 13.1, and 12.1. The construction of the BFs can be expressed as follows. Consider the variable y_{t-12} as an example. We have two BFs (i.e., BF₁ and BF₂) to be considered. The observed values of BF₁ are determined by the values larger than 0 or $y_{t-12} - 367$. The corresponding coefficient is estimated to be 0.985. The observed values of BF₂ are determined by the values larger than 0 or $552.09 - y_{t-12}$. The corresponding coefficient is estimated to be -1.258. Accordingly, the MARS forecasting model for rice prices can be expressed as follows:

$$\begin{aligned}
Y = & 542.484 + 0.985(\text{BF}_1) - 1.258(\text{BF}_2) \\
& + 1.539\text{BF}_3 - 1.880(\text{BF}_4) + 0.524(\text{BF}_5) \\
& - 2.410(\text{BF}_6) + 1.783(\text{BF}_7) + 0.454(\text{BF}_8) \\
& - 0.234(\text{BF}_9) + 0.346(\text{BF}_{10}).
\end{aligned} \tag{27}$$

For ARIMA-MARS modeling, the variable selection results and BFs are summarized in Table 5. In addition, it can be seen that three input variables (i.e., Z_{t-13} , Z_{t-7} , and Z_{t-1}) played important roles in building the ARIMA-MARS

TABLE 4: The results of MARS₁ for rice price forecasting.

Variable selection results			
Variable	Relative importance (%)	BF	Coefficient
Intercept			542.484
y_{t-12}	100.0	$BF_1 = \text{Max}$ (0, $y_{t-12} - 367$)	0.985
		$BF_2 = \text{Max}$ (0, $552.09 - y_{t-12}$)	-1.258
		$BF_3 = \text{Max}$ (0, $y_{t-11} - 356$)	1.539
y_{t-11}	13.1	$BF_4 = \text{Max}$ (0, $y_{t-11} - 472.48$)	-1.880
		$BF_5 = \text{Max}$ (0, $585.95 - y_{t-11}$)	0.524
		$BF_6 = \text{Max}$ (0, $y_{t-10} - 363$)	-2.410
y_{t-10}	12.1	$BF_7 = \text{Max}$ (0, $y_{t-10} - 472.48$)	1.783
		$BF_8 = \text{Max}$ (0, $y_{t-10} - 543.14$)	0.454
		$BF_9 = \text{Max}$ (0, $623 - y_{t-10}$)	-0.234
		$BF_{10} = \text{Max}$ (0, $y_{t-10} - 623$)	0.346

forecasting models. The MARS forecasting model for rice prices can be described as follows:

$$Y = -13.404 + 0.407(BF_1) + 23.543(BF_2) - 73.468(BF_3) + 49.716(BF_4) + 0.479(BF_5), \quad (28)$$

where

$$Z_t = (y_t - y_{t-1}). \quad (29)$$

3.3. Forecast Modeling of Wheat Prices. Figure 2 displays the original time plot for wheat prices. After modeling the wheat price data series with the ARIMA procedure, the parameter estimates are shown in Table 6. In Table 6, the notations “MA1,1,” “MA1,2,” “MA1,3,” and “MA1,4” correspond to the parameters θ_1 , θ_7 , θ_{12} , and θ_{13} of the MA model. Table 7 displays the Ljung-Box test results, and a conclusion is made on the appropriateness of the underlying ARIMA model. Accordingly, (30) is suitable for modeling wheat prices.

$$Z_t = a_t - 0.2402a_{t-1} + 0.14629a_{t-7} + 0.11171a_{t-12} + 0.13721a_{t-13}, \quad (30)$$

where

$$Z_t = y_t - y_{t-1}. \quad (31)$$

In addition, a_{t-j} stands for the white noise at time $t - j$.

For ANN modeling on wheat prices, this study also used two ANN designs, same as the structures for modeling rice prices. The first design used 12 self-predictor variables (i.e.,

TABLE 5: The results of ARIMA-MARS for rice price forecasting.

Variable selection results			
Variable	Relative importance (%)	BF	Coefficient
Intercept			-13.404
Z_{t-13}	100.0	$BF_1 = \text{Max}$ (0, $Z_{t-13} + 35$)	0.407
		$BF_2 = \text{Max}$ (0, $Z_{t-7} - 0$)	23.543
Z_{t-7}	45.4	$BF_3 = \text{Max}$ (0, $Z_{t-7} + 1.30$)	-73.468
		$BF_4 = \text{Max}$ (0, $Z_{t-7} + 1.87$)	49.716
Z_{t-1}	28.5	$BF_5 = \text{Max}$ (0, $-39.85 - Z_{t-1}$)	0.479

y_{t-1} , y_{t-2} , y_{t-3} , ..., and y_{t-12}) to serve as input variables and a single variable (i.e., y_t) to serve as the output variable. The second design considers the ARIMA as an FSM to extract important self-predictor variables, which serve as the inputs for the ANN models. The ARIMA model for wheat price forecasting is obtained in (30) and indicates that a_{t-1} , a_{t-7} , a_{t-12} , and a_{t-13} will influence Z_t .

Because $a_{t-1} = Z_{t-1} - Z_{t-1}^{\wedge}$, where Z_{t-1}^{\wedge} is the forecast at time $t - 1$, we observe that Z_{t-1} will influence Z_t . Therefore, Z_{t-1} should be selected as an input variable as long as the component a_{t-1} is in the MA model. Following the same logic, because the model (e.i., (30)) contains a_{t-1} , a_{t-7} , a_{t-12} , and a_{t-13} , we conclude that Z_{t-1} , Z_{t-7} , Z_{t-12} , and Z_{t-13} will influence Z_t . Accordingly, we have selected Z_{t-1} , Z_{t-7} , Z_{t-12} , and Z_{t-13} to serve as input variables for the second proposed ANN modeling design for the wheat price data series.

After performing ANN modeling on wheat price data, Table 8 shows the corresponding forecast validity measures for different settings of ANN topologies for the two designs. As shown in Table 8, we observe that the ANN₁ design with the {12-11-1} structure has the smallest MAPE. For the second, or ARIMA-ANN design, the {4-8-1} structure was associated with the smallest MAPE.

For modeling SVR on wheat prices, this study adopted the input structure as ANN modeling. The first design employed 12 variables, y_{t-1} , y_{t-2} , y_{t-3} , ..., and y_{t-12} as the input variables and y_t as the output variable. The second design employed four self-predictor variables, Z_{t-1} , Z_{t-7} , Z_{t-12} , and Z_{t-13} to serve as the input variables and Z_t as the output variable. After performing SVR modeling on the wheat price data series, we obtained the parameter settings of $\{2^{-8}, 2^{10}\}_{\text{SVR}}$ and associated MAPE = 4.661% for the first design. In addition, we had parameter settings of $\{2^{-7}, 2^{-1}\}_{\text{SVR}}$ and associated MAPE = 4.347% for the second design.

For MARS modeling with the first design, Table 9 presents the variable selection results and BFs. As shown in Table 9, we notice that three input variables (i.e., y_{t-12} , y_{t-6} , and y_{t-11}) played important roles in building the MARS forecasting models. In addition, the MARS forecasting model for wheat prices can be expressed as follows:

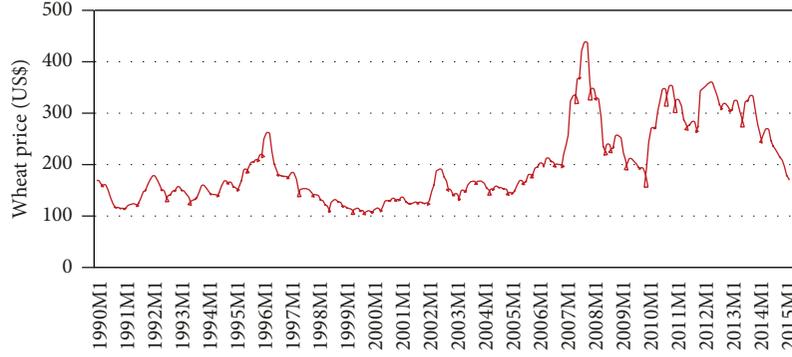


FIGURE 2: Time plot for wheat price (unit: US dollars per metric ton).

TABLE 6: Parameter estimates for wheat prices.

The ARIMA procedure					
Conditional least squares estimation					
Parameter	Estimate	Standard error	t value	Approx Pr > $ t $	Lag
MA1,1	-0.2402	0.05601	-4.29	<0.0001	1
MA1,2	0.14629	0.05754	2.54	0.0115	7
MA1,3	0.11171	0.0599	1.86	0.0632	12
MA1,4	0.13721	0.06018	2.28	0.0233	13

$$Y = 360.973 + 0.765(\text{BF}_1) - 1.272(\text{BF}_2) - 0.185(\text{BF}_3) + 0.298(\text{BF}_4) - 0.218(\text{BF}_5). \quad (32)$$

For MARS modeling with the second design, the variable selection results and BFs are summarized in Table 10. Furthermore, the MARS forecasting model for wheat prices is expressed as follows:

$$Y = -13.665 - 1.105(\text{BF}_1) + 8.333(\text{BF}_2) - 13.064(\text{BF}_3) + 4.955(\text{BF}_4) - 0.155(\text{BF}_5). \quad (33)$$

3.4. Forecast Modeling of Corn Prices. Figure 3 presents the original time plot for corn prices. Table 11 presents the parameter estimates after performing ARIMA modeling. In Table 11, the notations “AR1,1,” “AR1,2,” “AR1,3,” “AR1,4,” “AR1,5,” and “AR1,6” correspond to as parameters ϕ_1 , ϕ_3 , ϕ_6 , ϕ_{12} , ϕ_{14} , and ϕ_{24} of the AR model. In Table 11, we still contain the parameters of “AR1,2” and “AR1,4” in the model, although the absolute t values are less than or equal to 2 (i.e., the typical type I error is chosen as 0.05). The main reason is that those two parameters are important, and the Ljung-Box statistics indicate that the model is not appropriate if those two parameters are not involved in the underlying ARIMA model. Additionally, Table 12 demonstrates the Ljung-Box test results and a conclusion is made on the appropriateness of the ARIMA model. Thus, (34) is a suitable ARIMA model for modeling the corn price data series.

$$Z_t = 0.17626Z_{t-1} + 0.09344Z_{t-3} - 0.145Z_{t-6} - 0.0925Z_{t-12} - 0.1476Z_{t-14} - 0.2149Z_{t-24} + a_t, \quad (34)$$

where

$$Z_t = (y_t - y_{t-1}). \quad (35)$$

For ANN modeling on corn prices, while the first design used 12 input variables, the second design used six input variables extracted by using the FSM of ARIMA modeling. After performing ANN modeling, Table 13 presents the corresponding forecast validity measures for different ANN topology settings for the two designs. From Table 13, we observe that the ANN₁ design with the {12-14-1} structure has the smallest MAPE. For the second design, the {6-13-1} structure has the smallest MAPE.

In the modeling of SVR for the corn price data series, this study adopted the same input structure as ANN modeling. The first design employed 12 variables (i.e., y_{t-1} , y_{t-2} , y_{t-3} , ..., and y_{t-12}) as the input variables and y_t as the output variable. The second design employed six self-predictor variables (i.e., Z_{t-1} , Z_{t-3} , Z_{t-6} , Z_{t-12} , Z_{t-14} , and Z_{t-24}), which were selected by using the FSM of ARIMA modeling to serve as the input variables and Z_t as the output variable. After performing SVR modeling, we obtained the parameter settings of $\{2^{-9}, 2^7\}_{\text{SVR}}$ and associated MAPE = 4.060% for the first design. In addition, we obtained parameter settings of $\{2^{-11}, 2^{11}\}_{\text{SVR}}$ and associated MAPE = 4.215% for the second design.

For MARS modeling on corn price data with the first design, Table 14 illustrates the variable selection results and BFs. As shown in Table 14, we notice that two variables (i.e., y_{t-12} and y_{t-11}) play important roles in building the MARS forecasting models. Consequently, the MARS forecasting model for corn prices can be described as follows:

$$Y = 262.148 - 1.354(\text{BF}_1) + 0.969(\text{BF}_2) + 0.342(\text{BF}_3). \quad (36)$$

For MARS modeling with the second design, Table 15 lists the variable selection results and BFs. The MARS forecasting model for the corn prices is expressed as follows:

TABLE 7: Ljung-Box test results for ARIMA modeling of wheat prices.

To Lag	Chi-square	DF	Autocorrelation check of residuals						
			Pr > ChiSq	Autocorrelations					
6	4.4	2	0.111	0.019	-0.045	0.024	-0.036	0.069	0.074
12	11.07	8	0.1977	-0.03	-0.126	0.008	-0.055	0.042	-0.001
18	14.41	14	0.4197	-0.035	-0.084	0.021	-0.038	0.021	0.006
24	19.99	20	0.4589	0.067	0.081	-0.033	-0.03	0.012	-0.064
30	30.52	26	0.2465	-0.007	-0.041	-0.11	-0.132	-0.009	0.026
36	33.96	32	0.3733	-0.025	0.025	-0.071	0.011	0.054	0.029

TABLE 8: Forecast validity measures for two ANN designs (rice price forecasting).

ANN ₁ topology	MAE	RMSE	MAPE (%)	ARIMA-ANN topology	MAE	RMSE	MAPE (%)
{12-10-1}	12.121	13.521	5.448	{4-6-1}	11.833	14.336	5.374
{12-11-1}	11.752	13.230	5.285	{4-7-1}	11.865	14.353	5.388
{12-12-1}	12.084	13.515	5.423	{4-8-1}	8.984	10.864	4.077
{12-13-1}	11.984	13.429	5.385	{4-9-1}	8.984	10.883	4.078
{12-14-1}	12.120	13.516	5.457	{4-10-1}	11.845	14.346	5.379

TABLE 9: The results of MARS₁ for wheat price forecasting.

Variable selection results			
Variable	Relative importance (%)	BF	Coefficient
Intercept			360.973
y_{t-12}	100.0	$BF_1 = \text{Max}$ (0, $y_{t-12} - 321.18$)	0.765
		$BF_2 = \text{Max}$ (0, $321.81 - y_{t-12}$)	-1.272
		$BF_3 = \text{Max}$ (0, $y_{t-6} - 178.32$)	-0.185
y_{t-6}	7.2	$BF_4 = \text{Max}$ (0, $178.32 - y_{t-6}$)	0.298
		$BF_5 = \text{Max}$ (0, $y_{t-11} - 269.03$)	-0.218

$$Y = 3.787 - 0.277(BF_1) - 0.745(BF_2) - 0.466(BF_3) + 0.175(BF_4). \quad (37)$$

4. Forecasting Comparison

Several forecasting models were proposed to forecast the three major food crop prices in this study. These models include four single models (i.e., ARIMA, ANN₁, SVR₁, and MARS₁) and three integrated models with FSMs (i.e., ARIMA-ANN, ARIMA-SVR, and ARIMA-MARS). Tables 16–18 present the forecasting results, as well as the MAE, RMSE, and MAPE values of the forecasting models for rice, wheat, and corn price forecasting, respectively. Low MAE, RMSE, or MAPE values are associated with better forecasting accuracy.

In comparison to the forecasting performance of the first design, or single models, in Tables 16–18, we observe that the three CI models demonstrated better performance than the ARIMA models. The possible reason may be that ARIMA

TABLE 10: The results of ARIMA-MARS for wheat price forecasting.

Variable selection results			
Variable	Relative importance (%)	BF	Coefficient
Intercept			-13.665
Z_{t-7}	100.0	$BF_1 = \text{Max}$ (0, $Z_{t-7} - 26.58$)	-1.105
		$BF_2 = \text{Max}$ (0, $Z_{t-13} + 26.91$)	8.333
Z_{t-13}	63.3	$BF_3 = \text{Max}$ (0, $Z_{t-13} + 22.55$)	-13.064
		$BF_4 = \text{Max}$ (0, $Z_{t-13} + 17.56$)	4.955
Z_{t-1}	14.4	$BF_5 = \text{Max}$ (0, $Z_{t-1} + 10.65$)	-0.155

modeling is difficult for capturing nonlinear features in the food crop price data series. By reviewing these tables, we found that there is no best model for food crop price forecasting. For example, while the ANN₁ model seems to possess better forecasting accuracy for rice price forecasting, the SVR₁ model seems to have better forecasting accuracy for wheat and corn price forecasting.

It can be clearly seen in Tables 16–18 that the proposed integrated models possess better forecasting accuracy than the single models for most cases. For example, by reviewing Table 16, the proposed integrated ARIMA-ANN model is associated with MAE, RMSE, and MAPE values of 10.365%, 12.037%, and 2.650%, respectively, for rice price forecasting. These three performance values are smaller than the corresponding performance values for any one of the four single models. Take the proposed integrated ARIMA-SVR model as another example. By reviewing Table 17, the ARIMA-SVR model is associated with MAE, RMSE, and MAPE values of 9.779%, 11.371%, and 4.347%, respectively, for wheat price forecasting. These three performance values

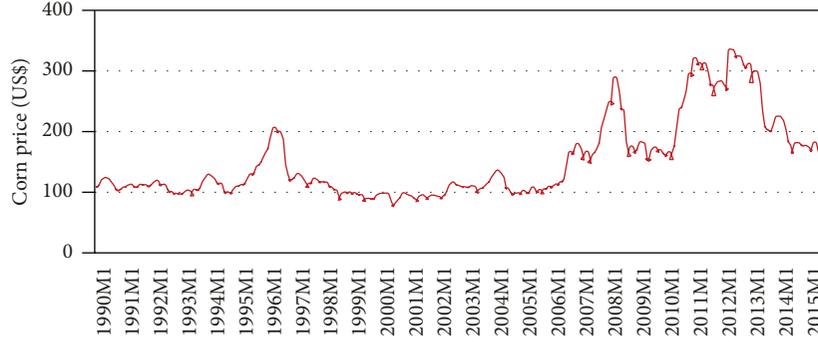


FIGURE 3: Time plot for corn price (unit: US dollars per metric ton).

TABLE 11: Parameter estimates for corn prices.

Parameter	The ARIMA procedure Conditional least squares estimation				Lag
	Estimate	Standard error	t value	Approx Pr > $ t $	
AR1,1	0.17626	0.05909	2.98	0.0031	1
AR1,2	0.09344	0.0651	1.44	0.1524	3
AR1,3	-0.145	0.06527	-2.22	0.0272	6
AR1,4	-0.0925	0.06619	-1.4	0.1636	12
AR1,5	-0.1476	0.06719	-2.2	0.029	14
AR1,6	-0.2149	0.07024	-3.06	0.0024	24

are also smaller than the corresponding performance values for any one of the four single models in Table 17. Thus, our proposed models, which were integrated with FSM, provide more accurate forecasting results than the single models.

In addition, Table 19 presents a comparison with respect to the overall percentage improvements (PIs) of forecasting accuracy for the proposed integrated models over the single models. The PIs of the MAE, RMSE, and MAPE are defined as follows:

$$\begin{aligned}
 MAE_{PI} &= \frac{[MAE(\text{single model}) - MAE(\text{integrated model})]}{MAE(\text{integrated model})} * 100\%, \\
 RMSE_{PI} &= \frac{[RMSE(\text{single model}) - RMSE(\text{integrated model})]}{RMSE(\text{integrated model})} * 100\%, \\
 MAPE_{PI} &= \frac{[MAPE(\text{single model}) - MAPE(\text{integrated model})]}{MAPE(\text{integrated model})} * 100\%.
 \end{aligned}
 \tag{38}$$

From Table 19, it is obvious that positive PIs can be achieved by using the proposed integrated models. For example, in the rice price forecasting, the $MAPE_{PI}$ of the proposed ARIMA-ANN model over the two single models, ARIMA and ANN_S , were 376.038% and 34.226%, respectively. The corresponding MAE_{PI} and $RMSE_{PI}$ were 369.822% and 34.250% and 340.945% and 27.273%, respectively. Apart from the rice price forecasting, most of the MAE_{PI} , $RMSE_{PI}$, and $MAPE_{PI}$ are positive large numbers for wheat and corn price forecasting. Accordingly, considerable forecasting accuracy improvements are achieved through the use

of the proposed integrated models. Furthermore, we also found five negative PIs in Table 19.

For wheat price forecasting, the $RMSE_{PI}$ of the ARIMA-MARS over $MARS_1$ was -1.029%. For corn price forecasting, the $RMSE_{PI}$ of the ARIMA-ANN over ANN_1 was -10.973%. The three remaining negative values occurred in corn price forecasting when using the ARIMA-SVR model, and the MAE_{PI} , $RMSE_{PI}$, and $MAPE_{PI}$ were -2.601%, -5.525%, and -3.677%, respectively. A negative PI value implies that the forecasting performance declined. However, the magnitudes of those five negative values are small, and they have minor effects on forecasting performance. Additionally, the associated $MAPE_{PI}$ for those five negative PIs were only one negative. That is, if a forecasting model has the minimum MAPE, this does not mean it has the minimum MAE or RMSE.

5. Discussion

While the aforementioned description focuses on accuracy comparisons for each individual food crop price forecast, the following discussion provides the comparison regarding overall performance for food crop price forecasting as a whole.

Table 20 lists the average PIs of the proposed integrated models over their common element, the ARIMA models. Figure 4 presents the average PIs of MAE, RMSE, and MAPE by employing the proposed integrated models over single ARIMA models. As shown in Figure 4, we notice that considerable accuracy improvements can be reached by using our proposed approaches. Additionally, Table 21 reports the average PIs of MAE, RMSE, and MAPE by using the proposed integrated models over another element, other than ARIMA element. Figure 5 displays the satisfied average PIs by using the proposed integrated models.

In addition, Figure 6 displays the forecasts of rice prices, obtained by using ARIMA, ANN_1 , and the proposed ARIMA-ANN models, for the twelve testing records. One can see that the forecasts of the ARIMA-ANN model are closest to the actual observations. The forecasts of the basic ARIMA model are relatively far away from the actual observations. Additionally, the prediction accuracy of ANN_1 is better than that of the ARIMA model. Regarding the ANN_1 design, we observe that the input vectors I_1, I_2, \dots, I_{12} are equivalent to $y_{t-1}, y_{t-2}, y_{t-3}, \dots, y_{t-12}$ and that the output O

TABLE 12: Ljung-Box test results for ARIMA modeling of corn prices.

To Lag	Chi-square	DF	Autocorrelation check of residuals						
			Pr > ChiSq	Autocorrelations					
6	.	0	.	-0.011	0.082	0.026	0.002	-0.063	-0.027
12	6.71	6	0.3481	0.013	-0.032	0.001	0.012	0.091	-0.014
18	14.3	12	0.2822	-0.069	-0.036	0.042	0.113	0.038	-0.046
24	16.43	18	0.5624	0.014	-0.033	-0.015	0.061	-0.034	-0.015
30	29.11	24	0.2159	0.035	-0.073	0.103	-0.079	-0.025	-0.12
36	30.86	30	0.4225	-0.036	0	0.06	0.005	0.015	-0.001

TABLE 13: Forecast validity measures for two ANN designs (corn price forecasting).

ANN ₁ topology	MAE	RMSE	MAPE (%)	ARIMA-ANN topology	MAE	RMSE	MAPE (%)
{12-10-1}	7.536	9.698	4.387	{6-10-1}	7.304	10.773	4.267
{12-11-1}	7.464	9.635	4.343	{6-11-1}	7.307	10.853	4.269
{12-12-1}	7.444	9.638	4.331	{6-12-1}	7.291	10.805	4.259
{12-13-1}	7.487	9.633	4.357	{6-13-1}	7.286	10.771	4.256
{12-14-1}	7.433	9.589	4.326	{6-14-1}	7.290	10.809	4.259

TABLE 14: The results of MARS₁ for corn price forecasting.

Variable selection results			
Variable	Relative importance (%)	BF	Coefficient
Intercept			262.148
y_{t-12}	100.0	BF ₁ = Max (0, 268.79 - y_{t-12})	-1.354
		BF ₂ = Max (0, y_{t-12} - 268.79)	0.969
y_{t-11}	7.2	BF ₃ = Max (0, 295.29 - y_{t-11})	0.342

TABLE 15: The results of ARIMA-MARS for corn price forecasting.

Variable selection results			
Variable	Relative importance (%)	BF	Coefficient
Intercept			3.787
Z_{t-24}	100.0	BF ₁ = Max (0, 15.08 - Z_{t-24})	-0.277
Z_{t-12}	74.0	BF ₂ = Max (0, Z_{t-12} - 13.9)	-0.745
Z_{t-1}	44.8	BF ₃ = Max (0, Z_{t-1} - 11.59)	-0.466
Z_{t-6}	17.7	BF ₄ = Max (0, Z_{t-6} + 7.36)	0.175

can be obtained by performing a nonlinear functional mapping, as shown in (10). Regarding the ARIMA-ANN mechanism, we observe that the input vectors I_1, I_2, I_3 , and I_4 are characterized by $Z_{t-1}, Z_{t-2}, Z_{t-7}$, and Z_{t-13} and that the output O can be obtained by performing a nonlinear functional mapping, as shown in (10).

Figure 7 presents the forecasts of rice prices, obtained by using ARIMA, SVR₁, and the proposed ARIMA-SVR

models, for the twelve testing records. Regarding the SVR₁ design, we observe that the input vectors I_1, I_2, \dots, I_{12} are characterized by $y_{t-1}, y_{t-2}, y_{t-3}, \dots, y_{t-12}$ and that the output O can be obtained by performing a nonlinear functional mapping, as shown in (17). Regarding the ARIMA-SVR mechanism, we observe that the input vectors I_1, I_2, I_3 , and I_4 are characterized by $Z_{t-1}, Z_{t-2}, Z_{t-7}$, and Z_{t-13} and that the output O can be obtained by performing a nonlinear functional mapping, as shown in (17).

Figure 8 shows the forecasts of rice prices, obtained by using ARIMA, MARS₁, and the proposed ARIMA-MARS models, for the twelve testing records. Regarding MARS₁ modeling, we observe that the input vectors I_1, I_2, \dots, I_{12} are characterized by $y_{t-1}, y_{t-2}, y_{t-3}, \dots, y_{t-12}$ and that the output O can be obtained by performing a nonlinear functional mapping, as shown in (22). Regarding the ARIMA-MARS design, we observe that the input vectors I_1, I_2, I_3 , and I_4 are characterized by $Z_{t-1}, Z_{t-2}, Z_{t-7}$, and Z_{t-13} and that the output O can be obtained by performing a nonlinear functional mapping, as shown in (22). By observing Figures 7 and 8, we can see that the forecasts of the ARIMA-SVR and ARIMA-MARS models are closest to the actual observations. Both figures illustrate that the forecasts of the ARIMA models are relatively far away from the actual observations.

Similar implications were obtained for the cases of wheat and corn price predictions. That is, the forecasts of the proposed integrated models are closer to the actual observations. The forecasts of the ARIMA model are far away from the actual observations. These findings can be observed in Figures 9–14.

In this study, we collected 309 records to build various models for the prediction of the prices of rice, wheat, and corn. One assumption for building the ARIMA models is that their structures will remain unchanged over time. Therefore, even if more sample data become available, there is no need to rebuild the ARIMA model for our proposed integrated

TABLE 16: Performance comparison of various models for rice price forecasting.

	Inputs	MAE (%)	RMSE (%)	MAPE (%)
(1) Single models				
ARIMA		48.697	53.077	12.615
ANN ₁	$y_{t-1} \sim y_{t-12}$	13.915	15.320	3.557
SVR ₁	$y_{t-1} \sim y_{t-12}$	27.994	34.514	6.912
MARS ₁	$y_{t-10}, y_{t-11}, y_{t-12}$	20.392	25.090	5.196
(2) Integrated models				
ARIMA-ANN	$Z_{t-1}, Z_{t-2}, Z_{t-7}, Z_{t-13}$	10.365	12.037	2.650
ARIMA-SVR	$Z_{t-1}, Z_{t-2}, Z_{t-7}, Z_{t-13}$	7.958	9.786	2.057
ARIMA-MARS	$Z_{t-1}, Z_{t-7}, Z_{t-13}$	11.244	15.566	2.916

TABLE 17: Performance comparison of various models for wheat price forecasting.

	Inputs	MAE (%)	RMSE (%)	MAPE (%)
(1) Single models				
ARIMA		33.520	42.250	16.651
ANN ₁	$y_{t-1} \sim y_{t-12}$	11.752	13.230	5.285
SVR ₁	$y_{t-1} \sim y_{t-12}$	10.203	13.586	4.661
MARS ₁	$y_{t-6}, y_{t-11}, y_{t-12}$	12.804	13.794	5.667
(2) Integrated models				
ARIMA-ANN	$Z_{t-1}, Z_{t-7}, Z_{t-12}, Z_{t-13}$	8.984	10.883	4.078
ARIMA-SVR	$Z_{t-1}, Z_{t-7}, Z_{t-12}, Z_{t-13}$	9.779	11.371	4.347
ARIMA-MARS	$Z_{t-1}, Z_{t-7}, Z_{t-13}$	10.643	13.937	4.809

TABLE 18: Performance comparison of various models for corn price forecasting.

	Inputs	MAE (%)	RMSE (%)	MAPE (%)
(1) Single models				
ARIMA		14.783	20.047	8.784
ANN ₁	$y_{t-1} \sim y_{t-12}$	7.433	9.589	4.326
SVR ₁	$y_{t-1} \sim y_{t-12}$	7.003	9.058	4.060
MARS ₁	y_{t-11}, y_{t-12}	7.015	9.337	4.110
(2) Integrated models				
ARIMA-ANN	$Z_{t-1}, Z_{t-3}, Z_{t-6}, Z_{t-12}, Z_{t-14}, Z_{t-24}$	7.286	10.771	4.256
ARIMA-SVR	$Z_{t-1}, Z_{t-3}, Z_{t-6}, Z_{t-12}, Z_{t-14}, Z_{t-24}$	7.190	9.588	4.215
ARIMA-MARS	$Z_{t-1}, Z_{t-6}, Z_{t-12}, Z_{t-24}$	6.495	9.102	3.785

models. This is a significant advantage over existing models. In other words, the ARIMA feature selection procedure only needs to be performed once. Even if more sample data becomes available, we can still use the same input variables for our proposed ARIMA-ANN, ARIMA-SVR, and ARIMA-MARS models. Some feature selection modeling processes may need to be performed repeatedly when more sample data become available, which may be a time-consuming task. It is true that, after performing feature selection with the ARIMA, we can increase the

accuracy of our forecasts by performing ANN, SVR, and MARS modeling with the additional sample data.

6. Conclusion

Humans survive based on food crops. We eat significant quantities of rice, wheat, corn, and other simple crops to maintain energy and good health. Accordingly, food crop price forecasting is very important and has drawn considerable attention in recent decades.

TABLE 19: Improvement of the proposed integrated models over single models.

	MAE _{PI} (%)	RMSE _{PI} (%)	MAPE _{PI} (%)
(1) Rice price forecasting			
(ii) Integrated ARIMA-ANN			
ARIMA	369.822	340.945	376.038
ANN ₁	34.250	27.273	34.226
(ii) Integrated ARIMA-SVR			
ARIMA	511.925	442.372	513.272
SVR ₁	251.772	252.688	236.023
(iii) Integrated ARIMA-MARS			
ARIMA	333.093	240.977	332.613
MARS ₁	81.359	61.186	78.189
(2) Wheat price forecasting			
(i) Integrated ARIMA-ANN			
ARIMA	273.108	288.223	308.313
ANN ₁	30.810	21.562	29.598
(ii) Integrated ARIMA-SVR			
ARIMA	242.775	271.562	283.046
SVR ₁	4.336	19.483	7.223
(iii) Integrated ARIMA-MARS			
ARIMA	214.949	203.152	246.247
MARS ₁	20.304	-1.029	17.842
(3) Corn price forecasting			
(i) Integrated ARIMA-ANN			
ARIMA	102.896	86.124	106.391
ANN ₁	2.018	-10.973	1.645
(ii) Integrated ARIMA-SVR			
ARIMA	105.605	109.089	108.399
SVR ₁	-2.601	-5.525	-3.677
(iii) Integrated ARIMA-MARS			
ARIMA	127.606	120.253	132.074
MARS ₁	8.006	2.579	8.587

TABLE 20: PIs of the whole food crop prices by using the proposed integrated models over the single ARIMA models.

	MAE (%)	RMSE (%)	MAPE (%)
(1) ARIMA-ANN over ARIMA			
Rice	369.822	340.945	376.038
Wheat	273.108	288.223	308.313
Corn	102.896	86.124	106.391
Average PIs	248.609	238.431	263.581
(2) ARIMA-SVR over ARIMA			
Rice	511.925	442.372	513.272
Wheat	242.775	271.562	283.046
Corn	105.605	109.089	108.399
Average PIs	286.768	274.341	301.572
(3) ARIMA-MARS over ARIMA			
Rice	333.093	240.977	332.613
Wheat	214.949	203.152	246.247
Corn	127.606	120.253	132.074
Average PIs	225.216	188.127	236.978

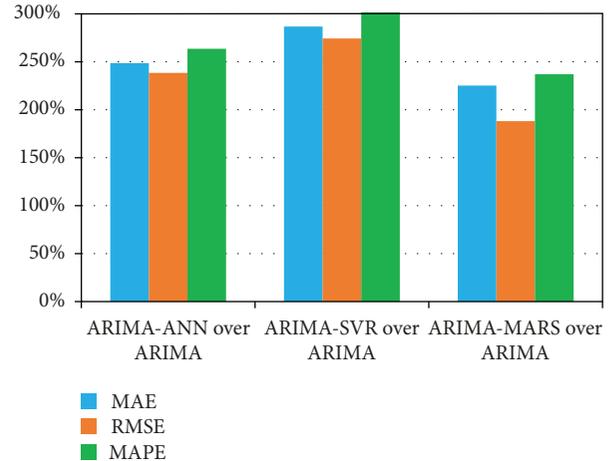


FIGURE 4: Average PIs by using the proposed integrated models over single ARIMA models.

Typical CI forecasting techniques require proper explanatory variables to make predictions. However, proper explanatory variables are difficult to capture, and it is infeasible to

TABLE 21: PIs of the whole food crop prices by using the proposed integrated models over another element, other than ARIMA element.

	MAE (%)	RMSE (%)	MAPE (%)
(1) ARIMA-ANN over ANN			
Rice	34.250	27.273	34.226
Wheat	30.810	21.562	29.598
Corn	2.018	-10.973	1.645
Average PIs	22.359	12.621	21.823
(2) ARIMA-SVR over SVR			
Rice	251.772	252.688	236.023
Wheat	4.336	19.483	7.223
Corn	-2.601	-5.525	-3.677
Average PIs	84.502	88.882	79.856
(3) ARIMA-MARS over MARS			
Rice	81.359	61.186	78.189
Wheat	20.304	-1.029	17.842
Corn	8.006	2.579	8.587
Average PIs	36.556	20.912	34.873

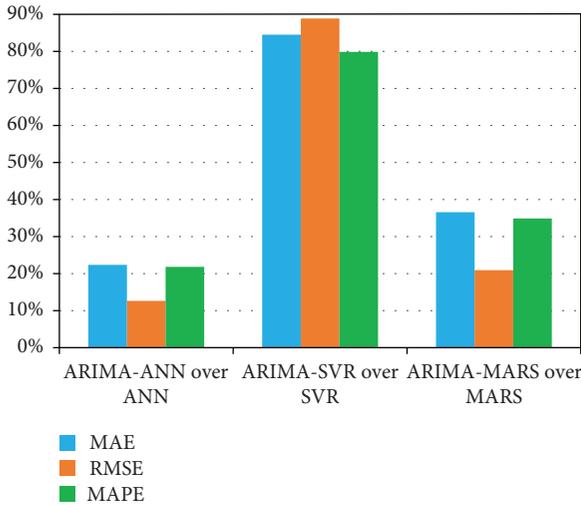


FIGURE 5: Average PIs by using the proposed integrated models over single ANN, SVR, and MARS models.

obtain the future values of these variables. Therefore, we proposed integrated ARIMA-ANN, ARIMA-SVR, and ARIMA-MARS models in order to perform forecasting for three important food crop prices. The role of the ARIMA element is as an FSM that can capture important self-predictor variables. Rather than using unavailable explanatory variables, the self-predictor variables serve as the inputs for ANN, SVR, and MARS modeling. The experimental results reveal that the proposed integrated models are desirable alternatives for food crop price forecasting, because they all have excellent forecasting performance. Most importantly, the main contribution of the proposed models is their ability to provide

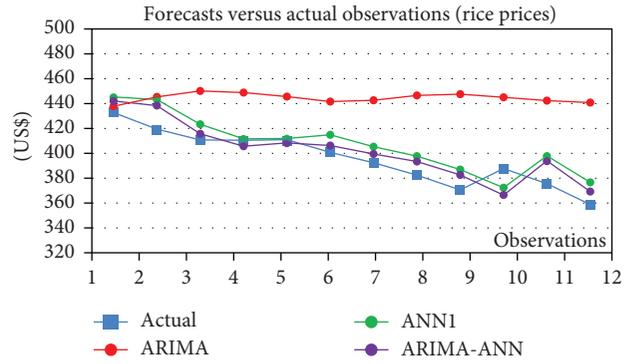


FIGURE 6: Forecasts of rice prices versus actual observations for the 12 testing records (with the use of ARIMA, ANN₁, and the proposed ARIMA-ANN models).

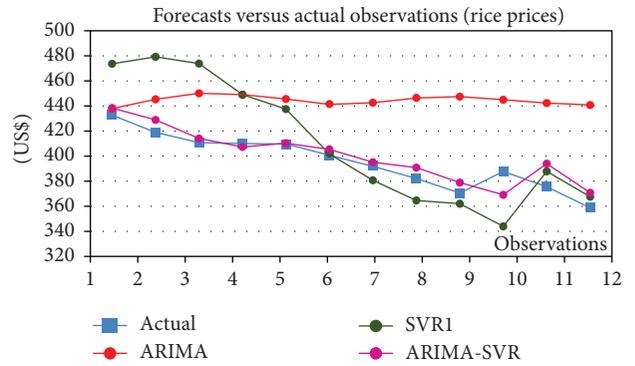


FIGURE 7: Forecasts of rice prices versus actual observations for the 12 testing records (with the use of ARIMA, SVR₁, and the proposed ARIMA-SVR models).

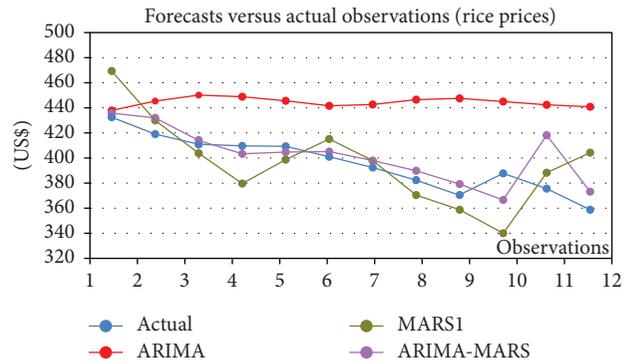


FIGURE 8: Forecasts of rice prices versus actual observations for the 12 testing records (with the use of ARIMA, MARS₁, and the proposed ARIMA-MARS models).

predictions of food crop prices without requiring extensive effort to obtain the future values of explanatory variables.

One limitation of our models is that the techniques for identifying the correct ARIMA model from the variety of possible models may be unintuitive and computationally expensive. However, the modeling in this study can be used as a guideline for developing forecasting models for other

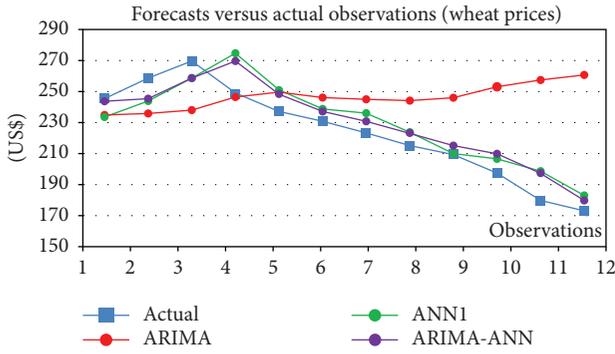


FIGURE 9: Forecasts of wheat prices versus actual observations for the 12 testing records (with the use of ARIMA, ANN₁, and the proposed ARIMA-ANN models).

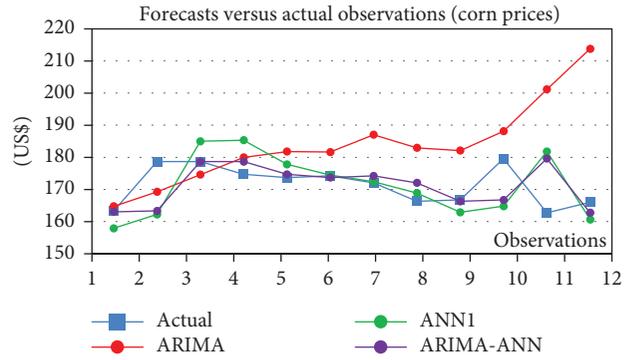


FIGURE 12: Forecasts of corn prices versus actual observations for the 12 testing records (with the use of ARIMA, ANN₁, and the proposed ARIMA-ANN models).

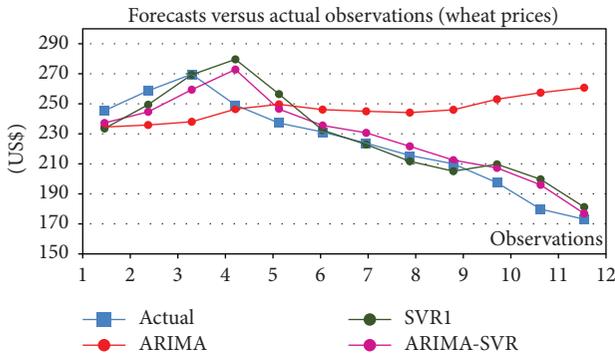


FIGURE 10: Forecasts of wheat prices versus actual observations for the 12 testing records (with the use of ARIMA, SVR₁, and the proposed ARIMA-SVR models).

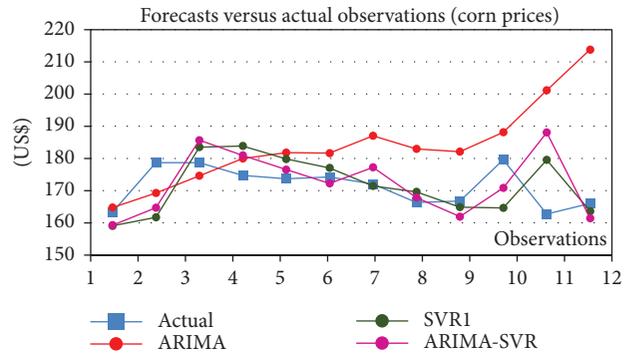


FIGURE 13: Forecasts of corn prices versus actual observations for the 12 testing records (with the use of ARIMA, SVR₁, and the proposed ARIMA-SVR models).

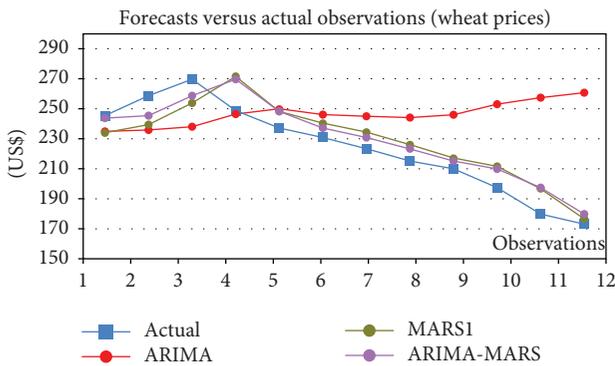


FIGURE 11: Forecasts of wheat prices versus actual observations for the 12 testing records (with the use of ARIMA, MARS₁, and the proposed ARIMA-MARS models).

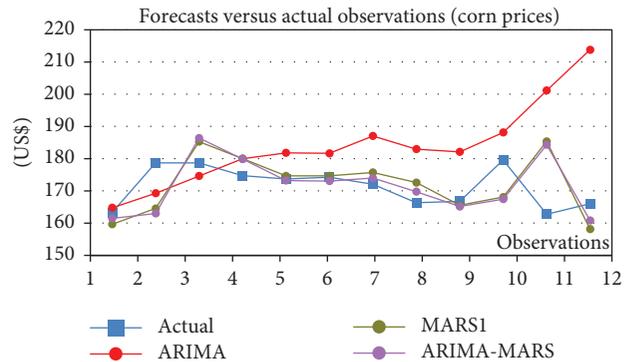


FIGURE 14: Forecasts of corn prices versus actual observations for the 12 testing records (with the use of ARIMA, MARS₁, and the proposed ARIMA-MARS models).

food crop price data series. Additionally, because MARS is effective for selecting important variables for predicting response variables, attempting to extend the FSM of MARS modeling may be a valuable future research direction. Finally, the integrated models may also be combined with other CI techniques, such as extreme learning machines, time

delay neural networks, or artificial immune systems, which may be worthy of investigation in the future.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments

This work is partially supported by the Ministry of Science and Technology of the Republic of China, Grant no. MOST 106-2221-E-030-010-MY2.

References

- [1] H. C. J. Godfray, J. R. Beddington, I. R. Crute et al., "Food security: the challenge of feeding 9 billion people," *Science*, vol. 327, no. 5967, pp. 812–818, 2010.
- [2] D. Tilman, C. Balzer, J. Hill, and B. L. Befort, "Global food demand and the sustainable intensification of agriculture," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 50, pp. 20260–20264, 2011.
- [3] Food and Agriculture Organization of the United Nations, "How to feed the world in 2050," 2009, <http://www.fao.org/wsfs/forum2050/wsfs-forum/en/>.
- [4] International Development Research Centre, "Facts & figures on food and biodiversity," 2017, <https://www.idrc.ca/en/article/facts-figures-food-and-biodiversity>.
- [5] H. C. Co and R. Boosarawongse, "Forecasting Thailand's rice export: statistical techniques vs. artificial neural networks," *Computers & Industrial Engineering*, vol. 53, no. 4, pp. 610–627, 2007.
- [6] H. F. Zou, G. P. Xia, F. T. Yang, and H. Y. Wang, "An investigation and comparison of artificial neural network and time series models for Chinese food grain price forecasting," *Neurocomputing*, vol. 70, no. 16–18, pp. 2913–2923, 2007.
- [7] P. A. Seck, E. Tollens, M. C. S. Wopereis, A. Diagne, and I. Bamba, "Rising trends and variability of rice prices: threats and opportunities for sub-Saharan Africa," *Food Policy*, vol. 35, no. 5, pp. 403–411, 2010.
- [8] A. Jumah and R. M. Kunst, "Seasonal prediction of European cereal prices: good forecasts using bad models?," *Journal of Forecasting*, vol. 27, no. 5, pp. 391–406, 2008.
- [9] U. Chakravorty, M. H. Hubert, M. Moreaux, and L. Nøstbakken, "Long-run impact of biofuels on food prices," *The Scandinavian Journal of Economics*, vol. 119, no. 3, pp. 733–767, 2017.
- [10] R. Cai, J. D. Mullen, M. E. Wetzstein, and J. C. Bergstrom, "The impacts of crop yield and price volatility on producers' cropping patterns: a dynamic optimal crop rotation model," *Agricultural Systems*, vol. 116, pp. 52–59, 2013.
- [11] M. Yousefi, B. Khoshnevisan, S. Shamshirband et al., "Support vector regression methodology for prediction of output energy in rice production," *Stochastic Environmental Research and Risk Assessment*, vol. 29, no. 8, pp. 2115–2126, 2015.
- [12] M. Khashei and M. Bijari, "An artificial neural network (p, d, q) model for timeseries forecasting," *Expert Systems with Applications*, vol. 37, no. 1, pp. 479–489, 2010.
- [13] Y. E. Shao and C. D. Hou, "Change point determination for a multivariate process using a two-stage hybrid scheme," *Applied Soft Computing*, vol. 13, no. 3, pp. 1520–1527, 2013.
- [14] Y. E. Shao, C. D. Hou, and C. C. Chiu, "Hybrid intelligent modeling schemes for heart disease classification," *Applied Soft Computing*, vol. 14, pp. 47–52, 2014.
- [15] G. Santamaria-Bonfil, A. Reyes-Ballesteros, and C. Gershenson, "Wind speed forecasting for wind farms: a method based on support vector regression," *Renewable Energy*, vol. 85, pp. 790–809, 2016.
- [16] R. Tahmasebifar, M. K. Sheikh-El-Eslami, and R. Kheirollahi, "Point and interval forecasting of real-time and day-ahead electricity prices by a novel hybrid approach," *IET Generation, Transmission & Distribution*, vol. 11, no. 9, pp. 2173–2183, 2017.
- [17] O. Abedinia, N. Amjady, and H. Zareipour, "A new feature selection technique for load and price forecast of electrical power systems," *IEEE Transactions on Power Systems*, vol. 32, no. 1, pp. 62–74, 2017.
- [18] A. R. Gollou and N. Ghadimi, "A new feature selection and hybrid forecast engine for day-ahead price forecasting of electricity markets," *Journal of Intelligent & Fuzzy Systems*, vol. 32, no. 6, pp. 4031–4045, 2017.
- [19] S. Jurado, A. Nebot, F. Mugica, and N. Avellana, "Hybrid methodologies for electricity load forecasting: entropy-based feature selection with machine learning and soft computing techniques," *Energy*, vol. 86, pp. 276–291, 2015.
- [20] F. Jimenez, G. Sanchez, J. M. Garcia, G. Sciavicco, and L. Miralles, "Multi-objective evolutionary feature selection for online sales forecasting," *Neurocomputing*, vol. 234, pp. 75–92, 2017.
- [21] O. B. Shukur and M. H. Lee, "Daily wind speed forecasting through hybrid KF-ANN model based on ARIMA," *Renewable Energy*, vol. 76, pp. 637–647, 2015.
- [22] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco, CA, USA, 1970.
- [23] G. M. Ljung and G. E. P. Box, "On a measure of lack of fit in time series models," *Biometrika*, vol. 65, no. 2, pp. 297–303, 1978.
- [24] V. N. Vapnik, "An overview of statistical learning theory," *IEEE Transactions on Neural Networks*, vol. 10, no. 5, pp. 988–999, 1999.
- [25] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, NY, USA, 2000.
- [26] V. Cherkassky and Y. Ma, "Practical selection of SVM parameters and noise estimation for SVM regression," *Neural Networks*, vol. 17, no. 1, pp. 113–126, 2004.
- [27] J. H. Friedman, "Multivariate adaptive regression splines," *The Annals of Statistics*, vol. 19, no. 1, pp. 1–67, 1991.
- [28] 2017, <https://www.indexmundi.com/>.
- [29] Rice, <https://www.indexmundi.com/commodities/?commodity=rice&months=360>.
- [30] Wheat, <https://www.indexmundi.com/commodities/?commodity=wheat&months=360>.
- [31] Corn, <https://www.indexmundi.com/commodities/?commodity=corn&months=360>.
- [32] Y. E. Shao and C. C. Chiu, "Applying emerging soft computing approaches to control chart pattern recognition for an SPC–EPC process," *Neurocomputing*, vol. 201, pp. 19–28, 2016.

