Omega 41 (2013) 517–524

Contents lists available at SciVerse ScienceDirect

Omega

journal homepage: www.elsevier.com/locate/omega

Carbon price forecasting with a novel hybrid ARIMA and least squares support vector machines methodology

Bangzhu Zhu [a,b,c,*], Yiming Wei [b,c]

[a] School of Economics and Management, Wuyi University, Jiangmen, Guangdong 529020, China
[b] School of Management and Economics, Beijing Institute of Technology, Beijing 100081, China
[c] Center for Energy and Environmental Policy Research, Beijing Institute of Technology, Beijing 100081, China

ARTICLE INFO

Article history:
Received 8 September 2011
Accepted 22 June 2012
Available online 1 July 2012

Keywords:
Least squares support vector machine
ARIMA
Particle swarm optimization
Hybrid models
Time series forecasting
Carbon price

ABSTRACT

In general, due to inherently high complexity, carbon prices simultaneously contain linear and nonlinear patterns. Although the traditional autoregressive integrated moving average (ARIMA) model has been one of the most popular linear models in time series forecasting, the ARIMA model cannot capture nonlinear patterns. The least squares support vector machine (LSSVM), a novel neural network technique, has been successfully applied in solving nonlinear regression estimation problems. Therefore, we propose a novel hybrid methodology that exploits the unique strength of the ARIMA and LSSVM models in forecasting carbon prices. Additionally, particle swarm optimization (PSO) is used to find the optimal parameters of LSSVM in order to improve the prediction accuracy. For verification and testing, two main future carbon prices under the EU ETS were used to examine the forecasting ability of the proposed hybrid methodology. The empirical results obtained demonstrate the appeal of the proposed hybrid methodology for carbon price forecasting.

1. Introduction

Climate change has been a serious challenge in the last few decades. In order to reduce greenhouse gas emissions at the lowest overall cost, the European Union (EU) launched the EU Emissions Trading Scheme (EU ETS) covering around 12,000 installations in 25 countries and six major industrial sectors in 2005. As the first multinational atmospheric greenhouse gas cap-and-trade system, EU ETS is the largest carbon market in the world to date [1], and has proven to be not only an important tool for mankind to address climate change, but also a major choice for investors to decentralize their investment risks [2]. Therefore, the need for more accurate forecasts of carbon prices is driven by the desire to reduce risk and uncertainty.

Recently, although carbon price analysis has become one of the key issues concerned by many academic researchers and business practitioners [2–6], very little literature regarding carbon price forecasting can be found. In fact, carbon price changes over time, which means that it can be treated as a time series process. Therefore, carbon price forecasting is a kind of time series forecasting. During the past few decades, various approaches have been developed for time series forecasting, among which the autoregressive integrated moving average (ARIMA) model has been found to be one of the most popular models due to its statistical properties, as well as the well-known Box–Jenkins methodology in the modeling process. However, the ARIMA model is only a class of linear model and thus it can only capture linear patterns in a time series. Therefore, the ARIMA model cannot effectively capture nonlinear patterns hidden in a time series.

In order to overcome the limitations of the linear models and account for the nonlinear patterns existing in real cases, numerous nonlinear models have been proposed, among which the artificial neural network (ANN) has shown excellent nonlinear modeling capability. Although a large number of successful applications have shown that ANN has been successfully adopted in many forecasting fields [7–10], ANN suffers from some weaknesses, such as locally optimal solutions and over-fitting, which can make the forecasting precision unsatisfactory. In 1995, support vector machine (SVM), a novel ANN, was developed by Vapnik [11]. Established on the structural risk minimization (SRM) principle by minimizing an upper bound of the generalization error, SVM can result in resistance to the over-fitting problem [12]. However, SVM formulates the training process through quadratic programming, which can take much more time. In 1999, Suykens and his colleagues proposed a novel SVM known as least squares support vector machine (LSSVM) [13], which is able to solve linear problems quicker with a more straight forward approach. Until now, LSSVM has been successfully used in pattern recognition and nonlinear regression estimation problems. At the same time, to obtain the optimal LSSVM model, it

* Corresponding author at: School of Economics and Management, Wuyi University, Jiangmen, Guangdong 529020, China. Tel.: +86 750 3299331; fax: +86 750 3299335.
E-mail address: wpzbz@126.com (B. Zhu).

0305-0483/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.omega.2012.06.005

is important to choose a kernel function and determine the kernel parameters [14]. Therefore, we introduce particle swarm optimization (PSO) [15] to optimize the parameters of LSSVM in this study.

Although both LSSVM and ARIMA models have achieved success in their own linear or nonlinear domains, neither is suitable for all circumstances. The approximation of ARIMA models to complex, nonlinear problems, as well as LSSVM to model linear problems, may be inappropriate, let alone in problems that contain both linear and nonlinear patterns. Since it is difficult to completely know the characteristics of a real situation, a hybrid methodology that has both linear and nonlinear modeling capabilities can be a good candidate for practical use [16], which has been demonstrated by numerous studies [16–25]. However, existing hybrid methodologies often combine the traditional ANN [16–21] or SVM [22–25] and ARIMA models. Until now, no LSSVM and ARIMA hybrid model has been found for carbon price forecasting; therefore, this study aims to fill this gap.

The contributions made by this paper are two-fold. Firstly, we establish a novel LSSVM and ARIMA hybrid forecasting methodology to forecast carbon prices. In our proposed methodology, carbon prices are decomposed into two components: a linear component and a nonlinear component. An ARIMA model and a LSSVM model are used to capture the linear and nonlinear components of carbon prices respectively, and their forecasting values are integrated into the final forecasting results. Furthermore, PSO is used to find the optimal LSSVM parameter settings to forecast carbon prices in the future. Secondly, we evaluate the forecasting performance of the single ARIMA, ANN and LSSVM models, the hybrid ARIMA and ANN model, the hybrid ARIMA and SVM model, and the hybrid ARIMA and LSSVM model by forecasting two main future carbon prices under the EU ETS. The empirical results obtained demonstrate that the proposed hybrid model can outperform the single ARIMA, ANN, LSSVM models, the hybrid ARIMA and ANN model, and the hybrid ARIMA and SVM model.

The remainder of this study is organized as follows: Section 2 describes the ARIMA, LSSVM and hybrid models; Section 3 elaborates on the PSO–LSSVM model; Section 4 reports the experimental results and Section 5 provides the overall conclusions.

2. Methodology

2.1. ARIMA model

In the ARIMA model, carbon price is a linear function of past values and error terms. An ARIMA (p, d, q) model of degree of AR (p), difference (d) and MA (q) can be mathematically expressed as

$$x_t = u_t + \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \cdots + \varphi_p x_{t-p} - \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}$$

where x_t is the carbon price obtained by differencing d times, ε_t (hypothetical white noise) is assumed to be independently and identically distributed with a mean of zero and a constant variance of σ_ε^2, p and q are the numbers of autoregressive and moving average terms in the ARIMA model, and φ_i ($i=1,2,...,p$) and θ_i ($i=1,2,...,q$) are the model parameters to be estimated.

Fitting an ARIMA model to carbon price involves the following four-step iterative processes: determine the structure of ARIMA model, estimate the parameter values of the ARIMA model, perform ARIMA model tests on the residuals, and predict future carbon prices.

The major advantage of the ARIMA model is that it can capture the linear patterns of carbon prices well and is relatively easy to use. However, ARIMA on its own is not adequate for carbon price forecasting because real carbon prices are often nonlinear and irregular. Therefore, we introduce the LSSVM model to capture nonlinear patterns existing in the carbon price.

2.2. Least squares support vector machine for regression

In contrast to other forecasting approaches, SVM, firstly proposed by Vapnik [11] in 1995 and based on the SRM principle, has been successfully applied to classification and regression. However, SVM training is a time consuming process when analyzing huge data. LSSVM is a modification of the standard SVM developed by Suykens et al. [13] in 1999 to overcome these shortcomings, which results in a set of linear equations instead of a quadratic programming problem. Consider a given training set x_i, y_i, $i=1,2,...,l$ with input data, x_i, and output data, y_i. LSSVM defines the regression function as

$$\min J(w,e) = \frac{1}{2} w^T w + \frac{1}{2}\gamma \sum_{i=1}^{l} e_i^2 \tag{1}$$

subject to

$$y_i = w^T \phi(x_i) + b + e_i, \quad i=1,2,...,l \tag{2}$$

where w is the weight vector, γ is the penalty parameter, e_i is the approximation error, $\phi(\cdot)$ is the nonlinear mapping function and b is the bias term. The corresponding Lagrange function can be obtained by

$$L(w,e,\alpha,b) = J(w,e) - \sum_{i=1}^{l} \alpha_i w^T \phi(x_i) + b + e_i - y_i \tag{3}$$

where α_i is the Lagrange multiplier. Using the Karush–Kuhn–Tucker (KKT) conditions, the solutions can be obtained by partially differentiating with respect to w, b, $e_i$and α_i:

$$\begin{cases} \frac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{i=1}^{l} \alpha_i \phi(x_i) \\ \frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^{l} \alpha_i = 0 \\ \frac{\partial L}{\partial e_i} = 0 \rightarrow \alpha_i = \gamma e_i \\ \frac{\partial L}{\partial \alpha_i} = 0 \rightarrow w^T \phi(x_i) + b + e_i - y_i = 0 \end{cases} \tag{4}$$

By eliminating w and e_i, the equations can be changed into

$$\begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 & I_v^T \\ I_v & \Omega + \gamma^{-1} I \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ y \end{bmatrix} \tag{5}$$

where $y=[y_1,y_2,...,y_l]^T$, $I_v=[1,1,...,1]^T$, $\alpha=[\alpha_1,\alpha_2,...,\alpha_l]^T$, and the Mercer condition [11] has been applied to matrix Ω with $\Omega_{km} = \phi(x_k)^T \phi(x_m)$, $k,m=1,2,...,l$. Therefore, the LSSVM for regression can be obtained from

$$y(x) = \sum_{i=1}^{l} \alpha_i K(x,x_i) + b \tag{6}$$

where $K(x,x_i)$ is the kernel function.

A major advantage of the LSSVM model is that it can capture the nonlinear patterns hidden in the carbon price. However, using the LSSVM model alone to model the carbon prices may produce mixed results [26,27]. Therefore, we can conclude that the relationship between the ARIMA and LSSVM models is complementary, and it is necessary to combine the two for effective carbon price forecasting.

2.3. The hybrid models

In reality, carbon prices are rarely purely linear or nonlinear, but often contain both linear and nonlinear patterns, due to their inherently high complexity. This makes carbon price forecasting

very difficult. Neither ARIMA nor LSSVM can sufficiently model and predict the carbon price since the linear model cannot deal with nonlinear patterns and the nonlinear model alone cannot handle both linear and nonlinear patterns equally well [26,27]. Hybridizing the two models can yield a robust method and result in good forecasting results. Therefore, motivated by the previous studies [16–25], we propose a novel hybrid methodology which integrates the ARIMA and LSSVM models that can further improve forecasting accuracy.

The carbon price $Y_t, t=1,2,...,n$ is considered as the sum of a linear component and a nonlinear component, i.e., $Y_t = L_t + N_t$, where L_t and N_t are the linear and nonlinear components to be estimated, respectively. First, an ARIMA model is used to fit the linear component of the carbon price and generate a series of forecasts defined as $\hat{L}_t$. Next, the residuals containing only the nonlinear patterns in the carbon price, denoted as ε_t, can be generated by comparing the actual value Y_t of the carbon price with forecast value $\hat{L}_t$ of the linear component. That is, $\varepsilon_t = Y_t - \hat{L}_t$, where $\hat{L}_t$ is the forecast value of the ARIMA model at time t. Through modeling the residuals with a LSSVM, nonlinear patterns can be discovered. Finally, we build three various hybrid models with the following inputs.

2.3.1. Model 1: ARIMALSSVM1

$$\hat{N}_t = f(\varepsilon_{t-1}, \varepsilon_{t-2}, ..., \varepsilon_{t-p}) \tag{7}$$

where $\hat{N}_t$ is the forecast value of nonlinear component of the carbon price at time t, f is a nonlinear regression function determined by the LSSVM model and p is determined by the ARIMA model. Therefore, the final forecasting results will be calculated as

$$\hat{Y}_t = \hat{L}_t + \hat{N}_t. \tag{8}$$

2.3.2. Model 2: ARIMALSSVM2

$$\hat{Y}_t = f(Y_{t-1}, Y_{t-2}, ..., Y_{t-p}, \varepsilon_t) \tag{9}$$

where f is a nonlinear regression function determined by the LSSVM model and p is determined by the ARIMA model.

2.3.3. Model 3: ARIMALSSVM3

$$\hat{Y}_t = f(Y_{t-1}, Y_{t-2}, ..., Y_{t-p}, \hat{L}_t) \tag{10}$$

where f is a nonlinear regression function determined by the LSSVM model and p is determined by the ARIMA model.

3. The optimal LSSVM model by particle swarm optimization

In the LSSVM, choosing inappropriate parameters can result in over-fitting or under-fitting [23,24]. To build a LSSVM model efficiently, model parameters must be accurately determined [28]. These parameters include:

(1) *The kernel function*: The kernel function is used to construct a nonlinear decision hyper-surface on the LSSVM input space. In general, using the RBF kernel function, $K(x,y) = \exp(\frac{-\|x-y\|^2}{2\sigma^2})$, can yield a good prediction [29]. Therefore, we adopt the RBF kernel function as the kernel function of the LSSVM model.
(2) The regularization parameter, γ, balances the complexity and approximation accuracy of the LSSVM model.
(3) The bandwidth of the kernel function, σ, describes the variance of the RBF kernel function.

PSO is one of the modern heuristic algorithms developed by Kennedy and Eberhart in 1995 [15]. PSO uses several particles to constitute a swarm, and each particle represents a potential solution. Each particle flies around within the search space to seek the global optimal solution. In this study, we propose a new model known as PSO–LSSVM to optimize both the LSSVM parameters in order to improve the prediction accuracy. Fig. 1 illustrates the processes of the proposed PSO–LSSVM model, which are described in detail as follows:.

Step 1: *Data preparation*: The carbon price data are separated into training and testing sets known as Tr and Te, respectively.
Step 2: *Particle initialization and PSO parameter setting*: The two LSSVM parameters, γ and σ, are directly coded with real values to randomly generate a swarm of initial particles. The PSO parameters are set, including the number of particles (m), particle dimension (D), maximal iterations ($t_{\max}$), position

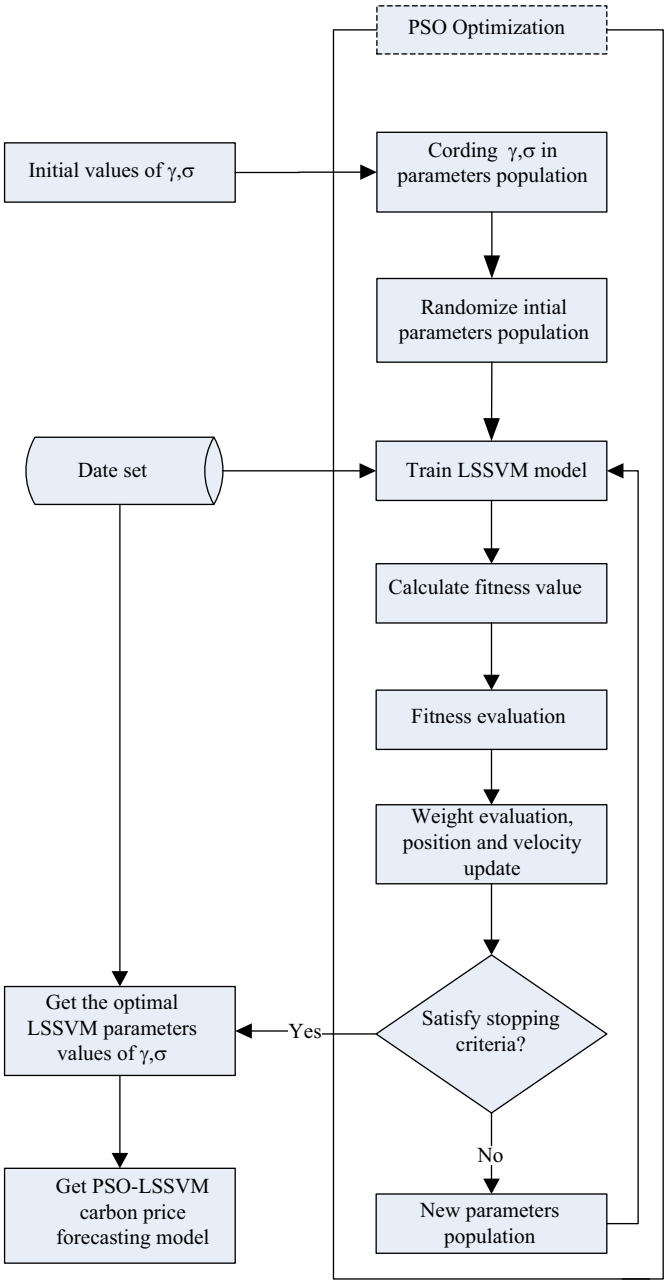

Fig. 1. PSO–LSSVM model.

limitation ($p_{\max}$), velocity limitation ($v_{\max}$), inertia weight limitation [$w_{\min}$,$w_{\max}$,], acceleration coefficients c_1 and c_2, γ limitation [$\gamma_{\min}$,,$\gamma_{\max}$], and σ limitation [$\sigma_{\min}$,$\sigma_{\max}$,]. Therefore, the position and velocity of the ith particle can be treated as $x_i=(x_{i1},x_{i2},...,x_{iD})$ and $v_i=(v_{i1},v_{i2},...,v_{iD})$, respectively. The ith particle has its own best position $p_{best}=(p_{i1},p_{i2},...,p_{iD})$ where the best fitness encountered by the particle so far, and the global best position is denoted by $g_{best}=(p_{g1},p_{g2},...,p_{gD})$ for all particles in the current generation. The iterative variable is set to t=0 and the training process can begin.

Step 3: *Fitness definition*: Because the training error can be measured by the root mean squared error (RMSE) built with a given set of parameters γ,σ, we can define the fitness function as follows:

$$\mathrm{Min}F_{fitness}=RMSE(\gamma,\sigma) \tag{11}$$

$$RMSE(\gamma,\sigma)=\sqrt{\frac{1}{n}\sum_{i=1}^{n}[y_i-\phi(x_i,\gamma,\sigma)]^2} \tag{12}$$

where n is the number of the training set; y_i is the actual value corresponding to x_i, and $\phi(\cdot)$ is the nonlinear regression function determined by the LSSVM model. Therefore, the solution with a smaller RMSE has a larger probability of surviving in successive generations.

Step 4: *Fitness evaluation*: Compute the fitness function value for each particle with Eqs. (11) and (12), and find the best position p_{best} of each particle and the global best position g_{best} of all the particles in the current generation.

Step 5: *Weight evaluation, position and velocity update*: Evaluate the inertia weight value with Eq. (13) [30], compute and update the position and velocity of each particle with Eqs. (14) and (15), respectively:

$$w(t)=w_{\max}-\frac{w_{\max}-w_{\min}}{t_{\max}}\times t \tag{13}$$

$$v_{id}(t+1)=\begin{cases}-v_{\max}, & v_{id}<-v_{\max}\\ w(t)v_{id}(t)+c_1r_1[p_{id}(t)-x_{id}(t)]+c_2r_2[p_{gd}(t)-x_{id}(t)], & -v_{\max}\le v_{id}\le v_{\max}\\ v_{\max}, & v_{id}>v_{\max}\end{cases} \tag{14}$$

$$x_{id}(t+1)=\begin{cases}-p_{\max}, & x_{id}<-p_{\max}\\ x_{id}(t)+v_{id}(t+1), & -p_{\max}\le x_{id}\le p_{\max}\\ p_{\max}, & x_{id}>p_{\max}\end{cases} \tag{15}$$

where $1\le i\le m$, $1\le d\le D$, $v_{id}(t)$ and $x_{id}(t)$ are the current velocity and position of the particle i at iteration t, respectively, p_{id} is the best previous position of the particle i, and p_{gd} is the best position among all particles.

Step 6: *Stopping criteria check*: Stop the search process if the termination criterion such as maximum iteration is met, and output the optimal solution. At this point the best LSSVM model is obtained. Otherwise, go to step 7.

Step 7: $t=t+1$, return to step 4.

4. Forecasting of carbon prices

4.1. Data description

The European Climate Exchange (ECX) located in London, is the largest carbon market under the EU ETS, since its daily carbon trading volume generally accounts for over 80% of the total carbon trading volume contributed by major EU carbon exchanges. It goes without saying that the state of ECX can reflect the overall state of EU ETS to a great extent.

As is well known, there are a great number of carbon prices in the ECX. In this study, two main carbon future prices with maturity in December, 2010 (DEC10) and December, 2012 (DEC12) were chosen as experimental samples. The main reason for selecting these two carbon price indicators is that they are the most famous benchmark prices which have been traded on the market since the inception of EU ETS in April 2005. They are the longest two future contracts covering the entire operating segment of EU ETS and are widely used as the basis of many carbon price formulae. The data of the two carbon prices used in this study are daily data which are freely available from the ECX website (http://www.theice.com).

For DEC10, we take daily data from April 22, 2005 to December 20, 2010, excluding public holidays, with a total of 1449 observations. For DEC12, we take daily data from April 22, 2005 to June 27, 2011, excluding public holidays, with a total of 1583 observations. For the convenience of LSSVM modeling, the data from April 22, 2005 to December 31, 2009 are used as the in-sample training sets (DEC10 and DEC12:1201 observations), and the remainder are used as the out-of-sample testing sets (DEC10:248 observations and DEC12:382 observations), which are used to check the forecasting ability based on evaluation criteria. Fig. 2 describes the curve of daily carbon prices for DEC10 and DEC12 in units of Euros/Ton and shows that carbon prices have highly uncertain,

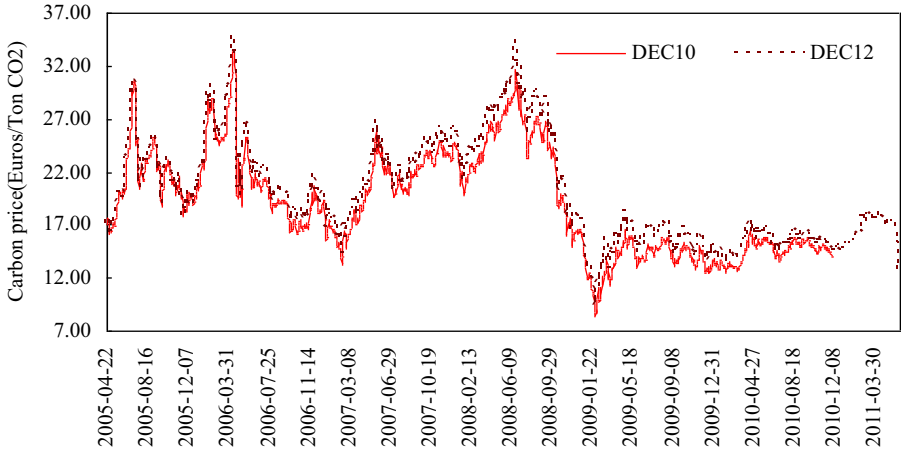

Fig. 2. ECX DEC10 and DEC12 carbon prices.

nonlinear, dynamic and complicated characteristics. Therefore, it is difficult to accurately forecast the carbon prices. Note that before training and testing, all the carbon price data are normalized in the range between 0 and 1 by Eq. (16), which can be helpful for improving the forecasting accuracy [31]:

$$\frac{x_t}{100} \tag{16}$$

where x_t is the carbon price at time t. The main reason for selecting 100 as the denominator is that according to the rules, EUETS will penalize enterprise at 40 Euros/Ton and 100 Euros/Ton for its own over-quota during the first phase from 2005 to 2007 and second phase from 2008 to 2012, respectively.

4.2. Forecasting evaluation criteria

To measure the forecasting performance, we use two main criteria to evaluate the level prediction and directional forecasting. First, the root mean squared error (RMSE) is adopted as the evaluation criterion of level prediction, which is defined by

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}[\hat{x}(t)-x(t)]^2}$$

Secondly, the directional prediction statistic (D_{stat}) [32] is adopted as the evaluation criterion of directional forecasting, which is expressed as

$$D_{stat} = \frac{1}{n}\sum_{t=1}^{n} a_t \times 100\%$$

where $x(t)$ is the actual value, $\hat{x}(t)$ is the predicted value, n is the number of predictions, and $a_t = 1$ if $[x(t+1)-x(t)][(\hat{x}(t+1)-x(t)] \geq 0$, otherwise $a_t = 0$.

For comparing the prediction capacity of the proposed ARIMA and LSSVM hybrid forecasting models with other widely used forecasting approaches, we employ the single ARIMA, ANN and LSSVM models as benchmark models. Moreover, two variants of the hybrid models, the hybrid ARIMA and ANN model (ARIMAANN) and the hybrid ARIMA and SVM model (ARIMASVM), are also used to predict carbon prices for the purpose of comparison.

4.3. Parameters determination of three models

4.3.1. ARIMA model

The DEC10 and DEC12 carbon price data are varied with daily changes, and show characteristics of tendency and instability. The ARIMA model is fitted to a stationary time series, and the daily data require regular differencing to become stationary. In this study, the ARIMA model is implemented via the Eviews statistical software package, produced by Quantitative Micro Software Corporation. The Akaike Information Criterion (AIC) is used to identify the best model. By trial and error in the two single models, both the optimal models generated from the DEC10 and DEC12 carbon prices are ARIMA (1, 1, 0), i.e.,$p = 1$,$d = 1$ and $q = 0$, and the final parameters estimation results are shown in Tables 1 and 2 for DEC 10 and DEC12, respectively.

Table 1
Parameters estimation results of the ARIMA for DEC10.

Variable	Coefficient	Std. error	*t*-statistic	Prob.
C	−0.004401	0.019040	−0.231126	0.8173
AR(1)	0.121241	0.028678	4.227618	0.0000
R^2	0.014712	Mean dependent var		−0.004312
Adjusted R^2	0.013888	S.D. dependent var		0.583409
S.E. of regression	0.579343	Akaike info criterion		1.747823
Sum squared resid	401.7592	Schwarz criterion		1.756312
Log likelihood	−1045.820	Hannan-Quinn criter.		1.751021
F-statistic	17.87276	Durbin–Watson stat		2.006763
Prob (*F*-statistic)	0.000025			

Table 2
Parameters estimation results of the ARIMA for DEC12.

Variable	Coefficient	Std. error	*t*-statistic	Prob.
C	−0.003269	0.019617	−0.166645	0.8677
AR(1)	0.094826	0.028763	3.296833	0.0010
R^2	0.008999	Mean dependent var		−0.003203
Adjusted R^2	0.008171	S.D. dependent var		0.617383
S.E. of regression	0.614856	Akaike info criterion		1.866808
Sum squared resid	452.5227	Schwarz criterion		1.875297
Log likelihood	−1117.151	Hannan-Quinn criter.		1.870006
F-statistic	10.86911	Durbin–Watson stat		2.003584
Prob (*F*-statistic)	0.001007			

Table 3
Optimal parameters for each hybrid model.

Model	DEC10		DEC12	
	γ	σ	γ	σ
ARIMALSSVM1	521.7613	1	947.9399	0.2885
ARIMALSSVM2	411.1023	1	992.9860	1
ARIMALSSVM3	521.7613	1	278.6803	0.9553

4.3.2. LSSVM model

For the LSSVM model, we use the LSSVMLab software package toolbox developed by Suykens et al. for MATLAB 2010b platform to build the LSSVM models. The RBF function is chosen as the kernel function of the LSSVM models. There is no standard procedure for determining the parameters, γ and σ, for the LSSVM model. Thus, we apply the PSO to synchronously optimize both of them. The inputs of the LSSVM models are determined in the same manner as the proposed hybrid models. The PSO is applied to seek and obtain the LSSVM optimal parameter sets when the RMSE is at its minimum. The searching process of optimal parameters is operated with real code, 100 initial particles, 500 generations, $\gamma \in [1,1000]$, $\sigma \in (0,1]$, $c_1 = c_2 = 2$, $w \in [0.1,0.9]$, $p_{\max} = 0.05$ and $v_{\max} = 50$, as well as fitness function $F_{fitness} = RMSE(\gamma,\sigma)$. Therefore, the PSO results in optimal parameters, which are $\gamma = 452.1705$, and $\sigma = 0.8232$ for DEC10 and $\gamma = 150.5224$, and $\sigma = 0.9817$ for DEC12. Both of these optimal parameter sets are used to construct the LSSVM models.

4.3.3. Hybrid models

In this investigation, we propose three hybrid models fundamentally derived from the LSSVM model. Therefore, the determining process for parameters is similar to the above-mentioned LSSVM model. The obtained optimal parameters for each hybrid model are shown in Table 3.

4.4. Experimental results

First, the two main carbon price datasets, DEC10 and DEC12, are applied to test these three hybrid models, which integrate the ARIMA and LSSVM models and are built with different inputs. In the meantime, the RMSE and D_{stat} calculated on the testing sets are applied to evaluate the forecasting performance of these three hybrid models. Only if a model has a smaller RMSE and/or larger D_{stat}, will it signify a better model. Besides the RMSE and D_{stat}

measurements, we also use the *t*-value to test the hypothesis that these three hybrid models have the same means of absolute forecast errors. When the hypothesis is statistically significant, we have demonstrated a better model. After the optimal parameter sets of LSSVM are gained by the PSO, these forecasting models are built. The optimal results are reported in Table 4. From Table 4, we can see that the ARIMALSSVM2 model (input nodes Y_{t-1} and residual value ε_t) has the best out-of-sample forecasting accuracy due to the smallest RMSE and the largest D_{stat} among the three hybrid models. *T*-tests also indicate a rejection of the hypothesis that the RMSE of the ARIMALSSVM2 model is the same as those of the ARIMALSSVM1 and ARIMALSSVM3 models, which demonstrates that the ARIMALSSVM2 model is a preferable hybrid model.

Then, we employ the single ARIMA, ANN and LSSVM models as benchmark models. Two variants of the hybrid model, the ARIMAANN and ARIMASVM models, are also used to predict carbon price for the purpose of comparison. The ANN and ARIMAANN models are established with the neural network toolbox (Version 5.0) of the Matlab software package, produced by the Mathworks Laboratory Corporation. ANN has one input node (Y_{t-1}), 20 hidden layer nodes and one output (Y_t). ARIMAANN has two input nodes (Y_{t-1} and residual value ε_t), 20 hidden layer nodes and one output (Y_t). The SVM and ARIMASVM models are established with the LS-SVMlab 1.5 Toolbox of the Matlab software package, produced by Suykens et al. SVM has also one input node (Y_{t-1}) and one output (Y_t), and RBF kernel function is selected as the kernel function. ARIMASVM has two input nodes (Y_{t-1} and residual value ε_t) and one output (Y_t). At the same time, these models are trained by the PSO with the same parameters as Section 4.3.2. The out-of-sample forecasting results are presented in Table 5 and Fig. 3.

Table 4
The out-of-sample forecasting comparison of different hybrid models.

Model	DEC10			DEC12		
	RMSE	D_{stat}	t value	RMSE	D_{stat}	t value
ARIMALSSVM1	0.2503	46.77	−17.816*	0.2721	46.34	−19.529*
ARIMALSSVM2	**0.0311**	**97.98**	**–**	**0.0310**	**96.60**	**–**
ARIMALSSVM3	0.2509	51.61	−17.176*	0.2810	48.95	−18.804*

Null hypothesis of the existence in the same means of the forecasted RMSE is generated by the hybrid model and other models.

* To be rejected at 5% significance level.

Table 5
Out-of-sample forecasting comparison of different models.

Model	DEC10			DEC12		
	RMSE	D_{stat}	*t* value	RMSE	D_{stat}	*t* value
ARIMALSSVM2	**0.0311**	**97.98**	**–**	**0.0309**	**96.60**	**–**
ARIMASVM	0.0636	86.96	−5.486*	0.0628	86.12	−5.712*
ARIMAANN	0.0967	68.25	−8.632*	0.0963	66.16	−9.032*
LSSVM	0.2473	50.81	−17.467*	0.2676	52.09	−18.023*
ANN	0.2568	47.58	−17.472*	0.2712	47.32	−18.436*
ARIMA	0.2474	44.76	−17.484*	0.2678	45.81	−18.639*

Null hypothesis of the existence in the same means of the forecasted RMSE is generated by the hybrid model and other models.

* To be rejected at 5% significance level.

Moreover, we compare the level prediction abilities of the hybrid model (ARIMALSSVM2), the single models (ARIMA, ANN and LSSVM models) and the variant models (ARIMAANN and ARIMASVM) in terms of the RMSE indicator. The ARIMALSSVM2 model performs the best in all cases, followed by ARIMASVM, ARIMAANN, LSSVM, ARIMA and ANN models, in that order. Interestingly, the hybrid models perform much better than the single models. One possible reason could be that the hybrid strategy impacts the forecasting performance. In addition, we also find that the single LSSVM model outperforms the single ARIMA model, while the single ANN model is not better than the single ARIMA model. The possible explanations could be two-fold: on one hand, the LSSVM model is a class of nonlinear forecasting model, which can capture nonlinear patterns hidden in the carbon prices, while the single ARIMA model is a class of linear models, which is not suitable for predicting nonlinear carbon prices. On the other hand, although ANN is also a class of nonlinear forecasting model, ANN cannot capture nonlinear patterns existing in the carbon price with high volatility and irregularity. Therefore, for the out-of-sample error comparisons of DEC10 and DEC12, the hybrid ARIMALSSVM2 model is superior to the single ARIMA, ANN and LSSVM models and to the variant models (ARIMAANN and ARIMASVM). *T*-tests also show a rejection of the hypothesis that the RMSE of the ARIMALSSVM2 model is the same as that of the other forecasting models, which reveals that the ARIMALSSVM2 model is better than the other forecasting models listed in this study.

Furthermore, we evaluate the directional forecasting abilities of the hybrid model (ARIMALSSVM2), the single models (ARIMA, ANN and LSSVM models) and the variant models (ARIMAANN and ARIMASVM) in terms of the D_{stat} indicator, which is more

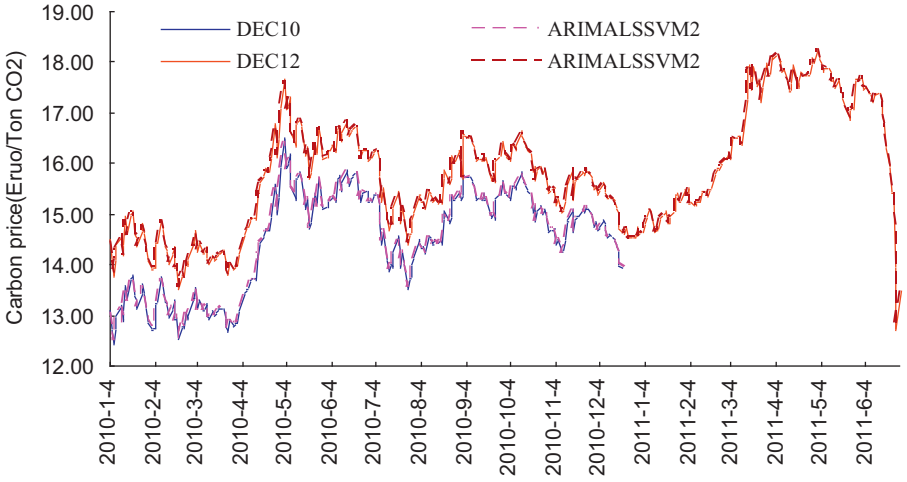

Fig. 3. Out-of-sample forecasting for DEC10 and DEC12 from 2010 to 2011.

important for business investors than the RMSE indicator [32]. The larger D_{stat} is, the greater the directional forecasting ability. For out-of-sample precision comparisons of DEC10 and DEC12, the D_{stat} of the ARIMALSSVM2 models are the largest at 97.98% and 96.60% respectively, followed by ARIMASVM, ARIMAANN, LSSVM, ANN and ARIMA models. The main reason for this phenomenon is the poor directional change detectability of the single models in comparison to the increased overall forecasting accuracy of the hybrid models. Interestingly, in contrast to the RMSE indicator, both the single ANN and LSSVM models perform much better than the single ARIMA model in terms of the D_{stat} indicator. As has already been discussed, the likely explanation is that the carbon price has highly nonlinear and complex characteristics. Therefore, the hybrid ARIMALSSVM2 model has an excellent ability to forecast turning points, which shows that the hybrid ARIMALSSVM2 model can yield significant forecast improvements. The gains in forecast accuracy may result from the ability of the integrated ARIMA, PSO and LSSVM models to capture both linear and nonlinear patterns existing in the carbon price.

In general, according to the experimental results of carbon price forecasting for ECX DEC10 and DEC12 presented in this study, we can draw some conclusions: (1) The experimental results show that the ARIMALSSVM2 model can exceed the single ARIMA, ANN and LSSVM models, as well as the hybrid ARIMALSSVM1, ARIMALSSVM3, ARIMASVM and ARIMAANN models, for the testing sets of the two main carbon future prices in terms of level prediction and directional prediction statistics, measured by RMSE and D_{stat}, as seen from Tables 4 and 5. (2) Not all of the hybrid models are superior to the single models, which indicates that the hybrid principles cannot effectively improve the prediction performance in every situation. (3) The ARIMALSSVM2 hybrid forecasting model can significantly improve forecasting accuracy, namely, the ARIMALSSVM2 model outperforms other forecasting models based on RMSE and D_{stat}, which results in the fourth conclusion. (4) The hybrid ARIMALSSVM2 model is suitable for global carbon prices forecasting.

5. Conclusions

In this paper we have presented several novel hybrid models incorporating ARIMA with LSSVM models for carbon price forecasting. The main contribution of this study is the proposition of several new hybrid models as well as some simple approaches for a robust prediction of nonlinear carbon prices. The ARIMA model is applied to capture linear patterns hidden in the carbon prices, whilst the LSSVM is used to capture nonlinear patterns existing in the carbon prices, resulting in a hybrid methodology which can improve forecasting accuracy. All the hybrid models consist in the evolutionary training of an ANN, a SVM or a LSSVM using the PSO algorithm. Once the optimal model parameters have been determined by the PSO, we can efficiently predict carbon prices of the future. This study has evaluated and compared the hybrid ARIMALSSVM2 model with the single ARIMA, ANN and LSSVM models, as well as the hybrid ARIMALSSVM1, ARIMALSSVM3, ARIMASVM and ARIMAANN models, using RMSE and D_{stat} as the criteria based on two well-known carbon price datasets from the ECX market.

The empirical results show that the hybrid model (ARIMALSSVM2) can produce the lowest RMSE and the highest D_{stat} in the carbon price datasets. It exceeds the single ARIMA, ANN and LSSVM models, as well as the hybrid ARIMALSSVM1, ARIMALSSVM3, ARIMASVM and ARIMAANN models. However, not all the hybrid models are superior to the single models in terms of both level prediction and directional prediction. That is to say, compared with the single models, not all the hybrid models can consistently achieve superior predictive performances, which has also been verified by previous studies [17,26,27]. Most of the single models and hybrid models evaluated show poor ability to detect directional change, as evident from the D_{stat} indicator. The hybrid model (ARIMALSSVM2) can effectively overcome this problem. Besides having excellent turning points delectability, the hybrid model (ARIMALSSVM2) can achieve superior forecasting performances and show good results. Therefore, the proposed hybrid model (ARIMALSSVM2) is very suitable for prediction with highly nonlinear and complex carbon price data, and is a very promising methodology for carbon price forecasting.

Acknowledgments

We sincerely thank the two anonymous referees for insightful comments that led to a greatly improved version of the article. Meanwhile, our heartfelt thanks should be given to the National Science Foundation of China (NSFC) (Grant nos. 71020107026 and 70733005) and the Ministry of Education of China (Grant no. 11YJC630304) for funding support.

References

[1] Zhang YJ, Wei YM. An overview of current research on EU ETS: evidence from its operating mechanism and economic effect. Applied Energy 2010;87(6):1804–1814.
[2] Wei YM, Wang K, Feng ZH, et al. Carbon finance and carbon market: models and empirical analysis .Beijing: Science Press; 2010.
[3] Seifert J, Marliese UH, Michael W. Dynamic behavior of CO_2 spot prices. Journal of Environmental Economics and Management 2008;56:180–194.
[4] Bataller MM, Pardo A, Valor E. CO_2 prices, energy and weather. The Energy Journal 2007;28(3):73–92.
[5] Beat H. Allowance price drivers in the first phase of the EUETS. Journal of Environmental Economics and Management 2010;59:43–56.
[6] Keppler JH, Bataller MM. Causalities between CO_2, electricity, and other energy variables during phase I and phase II of the EU ETS. Energy Policy 2010;38:3329–3341.
[7] Shafie-khah M, Parsa Moghaddam M, Sheikh-El-Eslami MK. Price forecasting of day-ahead electricity markets using a hybrid forecast method. Energy Conversion and Management 2011;52:2165–2169.
[8] Firat M, Gungor M. Generalized regression neural networks and feed forward neural networks for prediction of scour depth around bridge piers. Advances in Engineering Software 2009;40(8):731–737.
[9] Zhang Y, Wu L. Stock market prediction of S&P 500 via combination of improved BCO approach and BP neural network. Expert Systems with Applications 2009;36(5):8849–8854.
[10] Ioannou K, Arabatzis G, Lefakis P. Predicting the prices of forest energy resources with the use of Artificial Neural Networks (ANNs): the case of conifer fuel wood in Greece. Journal of Environmental Protection and Ecology 2009;10(3):678–694.
[11] Vapnik VN. The nature of statistical learning theory. New York: Springer; 1995.
[12] Vapnik VN, Golowich S, Smola A. Support vector method for function approximation, regression estimation and signal processing.Advance in neural information processing system. Cambridge, MA: MIT Press; 1997 vol. 9: p. 281–7.
[13] Suykenns JAK, Vandewalle J. Least squares support vector machine. Neural Processing Letter 1999;9(3):293–300.
[14] Liu LX, Zhuang YQ, Liu XY. Tax forecasting theory and model based on SVM optimized by PSO. Expert Systems with Applications 2011;38:116–120.
[15] Kennedy J, Eberhart RC. Particle swarm optimization. In: Proceedings of the IEEE conference on neural networks. Perth: Piscataway; 1995 vol. 4 p. 1942–48.
[16] Zhang GP. Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 2003;50:159–175.
[17] Mehdi K, Mehdi B. A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. Applied Soft Computing 2011;11:2664–2675.
[18] Ashu J, Avadhnam MK. Hybrid neural network models for hydrologic time series forecasting. Applied Soft Computing 2007;7:585–592.
[19] Omer FD. A hybrid neural network and ARIMA model for water quality time series prediction. Engineering Applications of Artificial Intelligence. 2010;23(4):586–594.
[20] Koutroumanidis T, Ioannou K, Arabatzis G. Predicting fuel wood prices in Greece with the use of ARIMA models, artificial neural networks and a hybrid ARIMA–ANN model. Energy Policy 2009;37:3627–3634.

[21] Kong F, Wu XJ. Time series forecasting model with error correction by structure adaptive support vector machine. In: Proceedings of the conference on computer science and software engineering. 2008. p. 1067–70.
[22] Pai PF, Lin CS. A hybrid ARIMA and support vector machines model in stock price forecasting. Omega 2005;33:497–505.
[23] Chen KY, Wang CH. A hybrid SARIMA and support vector machines in forecasting the production values of the machinery industry in Taiwan. Expert Systems with Applications 2007;32:254–264.
[24] Cagdas HA, Erol E, Cem K. Forecasting nonlinear time series with a hybrid methodology. Applied Mathematics Letters 2009;22:1467–1470.
[25] Tseng FM, Yu HC, Tzeng GH. Combining neural network model with seasonal time series ARIMA model. Technological Forecasting and Social Change 2002;69:71–87.
[26] Taskaya T, Casey MC. A comparative study of autoregressive neural network hybrids. Neural Networks 2005;18:781–789.
[27] Yu L, Wang SY, Lai KK. A novel nonlinear ensemble forecasting model incorporating GLAR and ANN for foreign exchange rates. Computers & Operations Research 2005;32:2523–2541.
[28] Kavaklioglu K .Modeling and prediction of Turkey's electricity consumption using Support Vector Regression. Applied Energy 2011;88:368–375.
[29] Smola AJ. Learning with kernels. PhD thesis, Department of Computer Science, Germany: Technical University Berlin; 1998.
[30] Liu XY, Shao C, Ma HF, Liu RX. Optimal earth pressure balance control for shield tunneling based on LS–SVM and PSO. Automation in Construction 2011;20:321–327.
[31] Chang CC, Lin CJ. LIBSVM: a library for support vector machines. Retrieved May 20, 2004, from National Taiwan University, Department of Computer Science and Information Engineering. Available from 〈http://www.csie.edu.tw/∼cjlin/papers/libsvm.pdf〉.
[32] Yu L, Wang SY, Lai KK, et al. A multiscale neural network learning paradigm for financial crisis forecasting. Neurocomputing 2010;73:716–725.