Research Article

Energy Demand Forecasting: Combining Cointegration Analysis and Artificial Intelligence Algorithm

Junbing Huang [ID], Yuee Tang, and Shuxing Chen

School of Economics, Southwestern University of Finance and Economics, Chengdu 611130, China

Correspondence should be addressed to Junbing Huang; 20701432@bit.edu.cn

Energy is vital for the sustainable development of China. Accurate forecasts of annual energy demand are essential to schedule energy supply and provide valuable suggestions for developing related industries. In the existing literature on energy use prediction, the artificial intelligence-based (AI-based) model has received considerable attention. However, few econometric and statistical evidences exist that can prove the reliability of the current AI-based model, an area that still needs to be addressed. In this study, a new energy demand forecasting framework is presented at first. On the basis of historical annual data of electricity usage over the period of 1985–2015, the coefficients of linear and quadratic forms of the AI-based model are optimized by combining an adaptive genetic algorithm and a cointegration analysis shown as an example. Prediction results of the proposed model indicate that the annual growth rate of electricity demand in China will slow down. However, China will continue to demand about 13 trillion kilowatt hours in 2030 because of population growth, economic growth, and urbanization. In addition, the model has greater accuracy and reliability compared with other single optimization methods.

1. Introduction

Energy, which is a vital input for the economic and social development of any economy, has gained special attention. Combined with globalization and industrialization, global energy demand has been increasing continually for decades and is expected to rise approximately 30% from 2015 to 2035 in accordance with the worldwide economic growth [1]. Therefore, energy demand projection should be developed because accurate energy demand forecasts aid policy makers in improving the schedule of energy supply and providing valuable suggestions for planning energy supply system operations.

Given the importance of accurate energy forecasts, extant studies using different estimation methods have been undertaken since the 1970s. In general, these early studies can be classified into two major categories: econometric [2–9] and machine learning (ML) methods [10–23]. The artificial intelligence (AI) energy forecasting model, which is a class of ML method, has gained popularity in recent years because of its superiority in time series processing and its capability to deal with noise data. Several tools, such as artificial

neural networks (ANN), genetic algorithm (GA), ant colony optimization (ACO), and particle swarm optimization, are commonly employed in the model [10–17]. Compared with the conventional econometric energy forecasting method, the AI-based model frequently demonstrates higher prediction accuracy in terms of mean absolute error (MAE), mean square error (MSE), mean absolute percentage error (MAPE), and root mean square error (RMSE) [16, 17]. According to economic theories, the model is feasible for predicting future energy demand by using the historical relationship when the periodical characteristics between energy demand and its explanatory variables will not change in the long term. However, the current AI-based method is referred to as the "black-box" because it predicts energy demand without knowing the internal relationship between energy demand and its affecting factors [23]. In addition, few econometric and statistical evidences are found that can prove the relationship between energy demand and its factors. This relationship may change in the long run based on the current AI-based model.

This study aims to present a more scientific AI-based energy demand forecasting framework that ensures the

reliability of predicted results. The electricity demand of China is forecasted as an example to show the process of implementing this framework. In addition, the predicted results are beneficial for policy makers to perform appropriate measures to bridge the electricity gap and arrange the supply of electricity demand.

The rest of the paper is organized as follows. Section 2 conducts a detailed literature review on the recent developments of energy demand forecasting. Section 3 presents the new framework. Section 4 predicts the electricity demand in China for 2016–2030 under three scenarios. The final section summarizes the main conclusions and presents the policy implications.

2. Literature Review

Energy estimation modeling has attracted wide spread interest among current practitioners and academicians. The commonly used econometric techniques include cointegration analysis, autoregressive integrated moving average (ARIMA) model, partial least square regression (PLSR), and vector error correction model. The ML method mainly refers to the AI model, support vector regression (SVR) method, and Grey forecasting method. Their details are described in the following sections.

2.1. Econometric Method. Cointegration analysis can establish a long-run relationship among variables, and the forecasting results are reliably shown through tests ranging from unit root to cointegration analysis [2, 3]. Early studies such as Chan and Lee [24] and Lin [2] forecasted the total energy and electricity demands in China, respectively. They conducted a series of tests ranging from unit root test to cointegration test to guarantee that a cointegration relationship exists between energy demand and its factors (i.e., the nexus will not change in the medium and long term). The ARIMA model is presented as an appropriate method for long-term projections [4–8, 25]. This model depends on three parameters, including order of moving average, order of differencing, and order of autoregressive scheme. However, ARIMA cannot be employed with missing and nonstationary data; otherwise, the original data should be first transformed by differencing. Recently, Cabral et al. [7] considered the spatiotemporal dynamics in the conventional ARIMA model. Their results confirmed that the new spatiotemporal model improves the electricity demand forecasts in Brazil and is paramount to achieving the goals of the Brazilian electricity sector for a secured electricity supply. Contrary to the ARIMA model, PLSR is a popular statistical tool that can deal with data, especially missing or highly correlated data [26]. However, PLSR was recently discussed in the field of energy demand estimation [26, 27]. For instance, Zhang et al. [26] employed the PLSR model to estimate the transportation energy demand in China on the basis of GDP, urbanization rate, passenger turnover, and freight turnover. Their results demonstrate that the transport energy demand for 2020 will reach a level of 4.3313 billion tons of coal equivalent (BTCE) and 4.6826 BTCE under different scenarios.

2.2. ML Method. Any optimization technique requires information on future scenarios and a search for the best solutions against a test criterion. In this case, ML techniques are superior and are frequently used to solve these two problems. The ML models include several tools, such as the AI, SVR, and Grey forecasting methods. To motivate our research, we focused particularly on the AI-based model.

The concept of SVR is developed from the computation of a linear regression function in a high-dimensional feature space where the input data are mapped via a nonlinear function, which can be found in Vapnik [28] and Vapnik et al. [29]. Dong et al. [19] were the first to employ SVR to predict the monthly energy use of buildings in tropical regions. Local weather data, including monthly average outdoor dry-bulb temperature, relative humidity, and global solar radiation, are selected as the factors affecting energy demand. Their results demonstrate that the relative error rate is less than 4%. Wang et al. [30] applied SVR for predicting hourly electricity use in residences and compared the results with other AI-based methods. They report that SVR improves the prediction accuracy.

Energy Grey forecasting model adopts the essential part of Grey system theory. In energy demand forecasting [18], the basic Grey model (GM (1,1)) was employed. Recently, Kang and Zhao [31] combined the moving average method and Markov model with GM to improve the accuracy of forecasting results. The improved Grey forecasting model demonstrates better performance compared with the conventional GM (1, 1). Xu et al. [32] combine GM and the Autoregressive and moving average model. The result indicates that the improved energy forecasting model has excellent accuracy and a high level of reliability for the case study of Guangdong Province.

AI-based prediction method predicts energy use according to its correlated variables, such as population growth, economic growth, and economic structure [2–6, 15–17]. For instance, Haldenbilen and Ceylan [10] proposed an AI model based on GA using population, GDP, and vehicle-km as affecting factors to forecast the transport energy demand in Turkey. Recently, Günay [23] modeled an electricity demand function for Turkey using the data on population, GDP per capita, inflation percentage, unemployment percentage, average summer temperature, and average winter temperature. Then, ANN is employed to determine the optimal weights that can maximize the accuracy of the function. The aforementioned algorithms can be called the single AI-based method. To eliminate several essential limitations in these algorithms, researchers also propose hybrid methods that integrate at least two AI algorithms, such as the GA-ANN [33] and PSO-GA models [12–16], to improve the prediction accuracy. The hybrid combination of a single AI algorithm shows greater performance compared with other methods.

The current AI-based prediction method is generally composed of four main steps: data collection, data preprocessing, model training, and model testing. With the superiority in time series processing, the AI-based model displays a good performance in predicting future energy demands. However, a spurious regression problem occurs in a wide range of time series analysis in econometrics owing to

its nonstationarity. The current AI-based model cannot avoid this problem. If the selected variables do not satisfy the basic requirements of constructing a cointegration relationship over the sample period, the AI-based forecasting models cannot be employed to make energy demand projections because the nexus between energy demand and its factors will change in the medium and long term. Therefore, the mechanism for predicting energy demand should be reformulated.

3. Methodology

3.1. Introduction to AI-Based Energy Demand Model.
In the precedent AI-based models, the commonly employed independent variables were around population, GDP, urbanization, industrialization, energy price, and energy mix. Three forms of the estimation models, including linear, quadratic, and exponential forms, were then adopted for data training [10, 11, 15–17], which can be expressed as follows:

$$y_{\text{lin}} = w_0 + \sum_{i=1}^{N} w_i X_i, \tag{1}$$

$$y_{\text{qua}} = w_0 + \sum_{i=1}^{N} w_i X_i + \sum_{i=1}^{N} \sum_{j=i+1}^{N} k_{ij} X_i X_j + \sum_{i=1}^{N} w_{N+i} X_i^2, \tag{2}$$

$$y_{\text{exp}} = w_0 + \sum_{i=1}^{N} w_i X_i^{w_{i+t}}, \tag{3}$$

where models (1), (2), and (3) are the linear, quadratic, and exponential forms, respectively. In each model, X_i is the ith energy demand-affecting factor, N is the number of energy demand-affecting factors, and w_i and $k_{i,j}$ are corresponding weights.

The "fittest" weights are finally searched through different AI tools, such as GA, ACO, and hybrid algorithms, based on the fitness function employed to monitor the forecasting accuracy, which aims to minimize the sum of squared error between the actual and estimated values shown as follows:

$$\min f(x) = \sum_{j=1}^{m} \left(E_{\text{actual}} - E_{\text{predicted}} \right)^2, \tag{4}$$

where E_{actual} and $E_{\text{predicted}}$ denote the actual and predicted energy demand values, respectively. m is the number of observations.

After obtaining the optimal weights, the model was applied to forecast the future energy demand under different scenarios. Compared with the traditional econometric energy demand forecasting model, the proposed AI-based model frequently demonstrates higher prediction accuracy. However, according to economic theory, these periodical characteristics of economic variables will not change in the medium and long term when an economy remains in a consistent state. Consequently, their historical relationship between energy demand and factors in the sampling period should be entirely stable. When this relationship was satisfied, it could be used for forecasting energy demand. However, the current AI-based energy demand forecasting

model does not determine this historical relationship through econometric and statistical analysis. This condition can be recognized as a "black-box" without knowing the internal relationship between energy demand and its affecting factors [33]. Accordingly, this model cannot be adopted for energy demand prediction when the historical relationship estimated through the AI-based model will change over time. Therefore, the improved AI-based model framework should be presented to improve the reliability.

3.2. Improved AI-Based Model.
As indicated in the abovementioned conventional AI-based model, the AI tool is directly applied to obtain the optimal weights for the model after preprocessing the original data. Then, the model is employed to forecast future energy demand. However, the prediction results are not reliable when the variables cannot build a stable and long-run relationship or when the parameters will change over time. Therefore, the model stability tests should be performed before proceeding to obtain the fittest weights through the AI tools. The cointegration analysis is widely employed as a key econometric method to forecast mid- and long-run energy demand because it can establish a long-run relationship among variables [3]. Cointegration theory and operations are employed to determine whether a long-run relationship exists between energy demand and its factors. To compare with the precedent AI-based model, our new framework for energy demand forecasting is shown in Figure 1(b) and the original framework described in previous literature is presented in Figure 1(a). As shown in Figure 1, if the energy demand and its factors cannot satisfy the cointegration relationship over the sample period, then this model cannot be adopted to predict future energy demand based on the current AI-based model because the stable relationship between them does not exist in the medium and long term.

3.2.1. Cointegration Test.
According to cointegration theory, the existence of a long-run equilibrium relationship among economic variables is based on the stationary linear combination of a time series. The cointegration relationship over the sampling period can be tested when the economic variables are integrated at $I(n)$ at the same time or at either $I(0)$ or $I(1)$. Hence, the first step to conduct the cointegration analysis is by employing a unit root test approach to check the stationarity of the variables. In empirical studies, the methods, including augmented Dickey–Fuller (ADF) and Phillips–Perron (PP) tests, are commonly employed to test the time series. Cointegration tests such as Engle–Granger [34], Johansen–Juselius [35], and autoregressive distributed lag (ARDL) bound testing approach [36, 37] can be adopted after the probability of building a cointegration relationship among the variables is verified. Engle–Granger method is feasible for testing single equation cointegration when the economic variables are integrated at $I(n)$ simultaneously. Compared with the Engle–Granger method, the Johansen–Juselius method can determine the number of cointegration vectors and test the existence of cointegration among variables. However, the Johansen–Juselius method can be employed when the variables are integrated at $I(n)$ simultaneously. Compared with
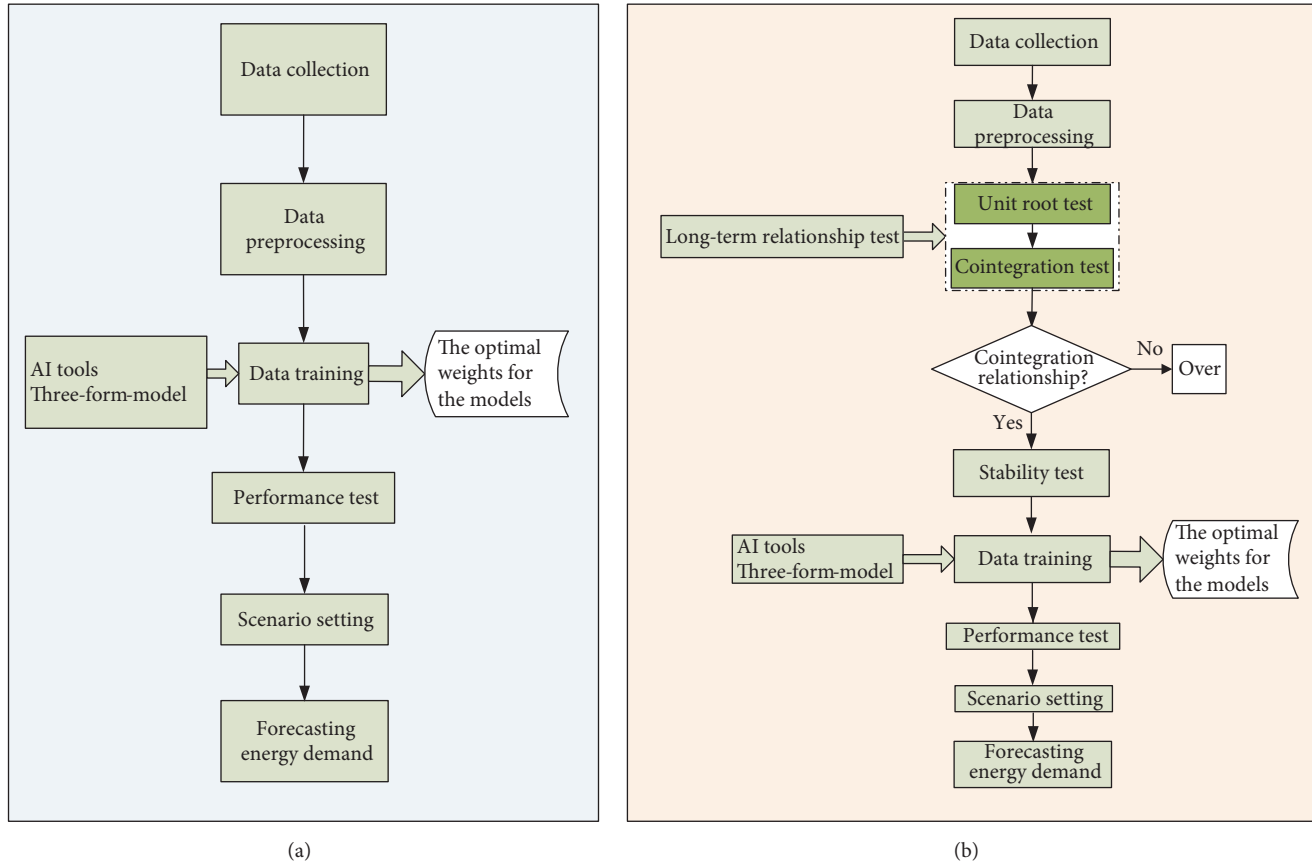
(a)

(b)

Figure 1: The frameworks for the conventional and new AI-based energy demand forecasting model.

the Engle–Granger and Johansen methods, ARDL bounds approach was recently applied to test the existence of a long-run equilibrium among the time series because it can establish the long- and short-run relationships among the variables. It also extended the mandatory requirements on the variables and can be applied even when the variables are integrated at $I(0)$, $I(1)$, or fractionally cointegrated. Third, the ARDL procedure is a more powerful approach to determine the cointegration relationship in small samples than the Johansen–Juselius technique. Finally, the problems of serial correlation and endogeneity are not difficult to tackle within the ARDL model.

3.2.2. Performance Test. After verifying the existence of a long-run relationship among variables, the next step is to test the prediction accuracy performance of the model and determine whether the estimated parameters will change with time. Parameter inconsistency may result in poor consequences on inferences and lead to wrong conclusions. For the Johansen–Juselius cointegration test technique, the stability test for the vector autoregressive model should be conducted through the unit root analysis. Meanwhile, for the ARDL bound approach, the cumulative (CUSUM) and cumulative sum of squares (CUSUMSQ) are suitable for the stability test because their statistics are updated recursively and plotted against the break points.

4. Forecasting Electricity Demand in China

In this section, electricity demand forecasting in China is shown as an example based on our new framework. First, we list the electric energy demand-affecting factors and the proxy variables based on existing electricity demand prediction literature. Second, given that the AI-based model does not require many explanatory variables, we employ the cointegration analysis to select the variables that can be contributed toward building a long-run relationship. Third, we employ the adaptive genetic algorithm (AGA), which is superior to conventional AI algorithms, to optimize the model. Based on the above estimation, three scenarios for economic growth, namely, low (Scenario A), business as usual (Scenario B), and high (Scenario A), are set. Finally, the electricity demand projections for the three scenarios are conducted.

4.1. Electricity Demand-Affecting Factors. Electricity demand can be viewed as a causal function of several affecting factors, such as population, GDP, electricity prices, economic structure, urbanization, and life styles [15–17]. In line with Lin [2] and Yu et al. [15], the number of total populations is employed as a reflector of population growth on electricity demand. Secondly, economic growth influences the electricity demand mainly through various ways. Inconsistent with most existing

TABLE 1: Definition and description for the variables.

Variables	Definition	Max	Min	Mean	Std. dev.
E	ln form of total electricity energy consume (10^{12} KWH)	1.7393	−0.8873	0.4418	0.8381
X_1	ln form of total population (10^8)	2.6208	2.3594	2.5208	0.0769
X_2	ln form of gross domestic product (trillion Yuan RMB 10^{12})	2.6540	−0.0944	1.2919	0.8578
X_3	ln form of the ratio of tertiary sector to GDP (%)	3.9160	3.3810	3.6459	0.1506
X_4	ln form of the urban population to the total (%)	4.0271	3.1659	3.5881	0.2809
X_5	ln form of price index for the electricity demand	2.6623	0.0000	1.6674	0.0825

Note. KWH denotes kilowatt hour.

literature, GDP is adopted as a key factor of electric energy demand. Thirdly, in China, the amount of energy required to produce a unit of GDP differs significantly among the three industries. Numerous researchers have noted that secondary industries are the main electricity consumers [2]. Therefore, the economic structure shifted from the secondary industry (e.g., heavy industry) to the tertiary industry (e.g., high-technology industry) may directly reduce electricity consumption [2]. In line with Feng et al. [38] and Li et al. [39], we take the ratio of the tertiary sector in GDP to capture the effect of relative change in economic structure on electricity demand. Fourthly, urbanization in China is one of the key stages exerting an important but complicated impact on electricity demand. In each year, millions of immigrants moved from rural to big cities in search of good job opportunities, which exerted a great influence on electric energy demand. This study utilizes urbanization rate to control the effect of urbanization on electricity demand. Lastly, according to economic theory, the price can affect the demand through income and substitution effects. However, adjusting the electricity price in China is complicated because it has been under the full control of the government for years and is mainly determined by the production cost. In addition, the price among the regions differs significantly, and no practical method can estimate the electricity price in China [2]. Therefore, we employ the price index for fuel and power to denote the electricity price index.

4.2. Data Management. To forecast the electricity demand, we use 31 years of observed data from 1985 to 2015. Electricity consumption in each year is measured in trillion (10^{12}) kilowatt hours (KWH), and population is measured as 100 million (10^8) persons. GDP data are measured in trillion Yuan (10^{12}) RMB and adjusted to the constant price in 1985. Economic structure is denoted by the ratio of output in the tertiary industry to GDP (%). The share of the urban population to the total population is used to substitute the urbanization rate (%). To represent the price for electricity, we use the price index for fuel and power to denote the electricity price. The price index for the base year (1985) is assumed to be 1.00, and the price index for other years is adjusted to the constant price index of 1985. The definition of the variables is shown in Table 1. The trends for electricity demand and its factors in the period of 1985–2015 are shown in Figure 2.

4.3. AGA. As mentioned in Section 2, the conventional AI algorithms also experience low prediction accuracy.
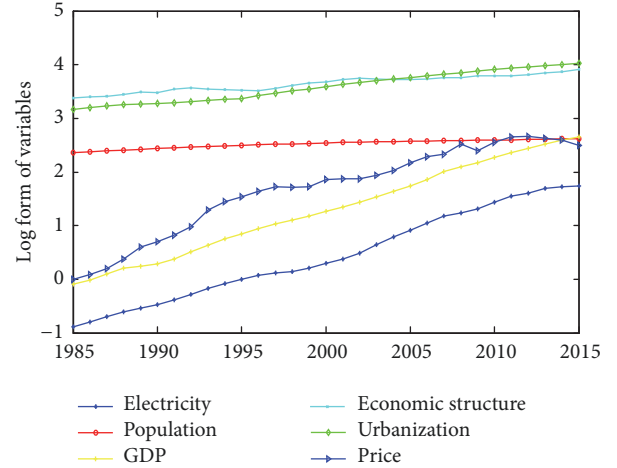

FIGURE 2: The trends of electricity demand and its factors (1985–2015).

However, the hybrid AI algorithms (e.g., PSO-GA and ANN-GA) are complicated. In this study, we employ AGA, which has a more profound intelligent background and yields good efficiency in optimizing global coefficients. The major difference between traditional GA and AGA is the selection of crossover probability p_c and mutation probability p_m. In conventional GA, the two probabilities are randomly determined or based on an inadequate reference, whereas AGA relies on the fitness function to select the optimal p_c and p_m. The flowchart for AGA is shown in Figure 3. In this figure, AGA contains major operations, including initialization, judgement and selection, crossover, and mutation. The detailed descriptions of the operations are given as follows:

(A) Initialization. The parameters, including the number of pop sizes (Pop_size) and the number of generations (Num_generations) are first set. The fitness functions, crossover probability, and mutation probability should also be determined.

(B) Judgement and Selection. Population is determined and ranked according to the values of fitness function. The individual with a low fitness value can be selected as the optimal solution for the problem when the number of generations is equal to the maximal number of generations; otherwise, the individual with a low fitness value is selected and the rest of the individuals are replaced by the selected individuals.
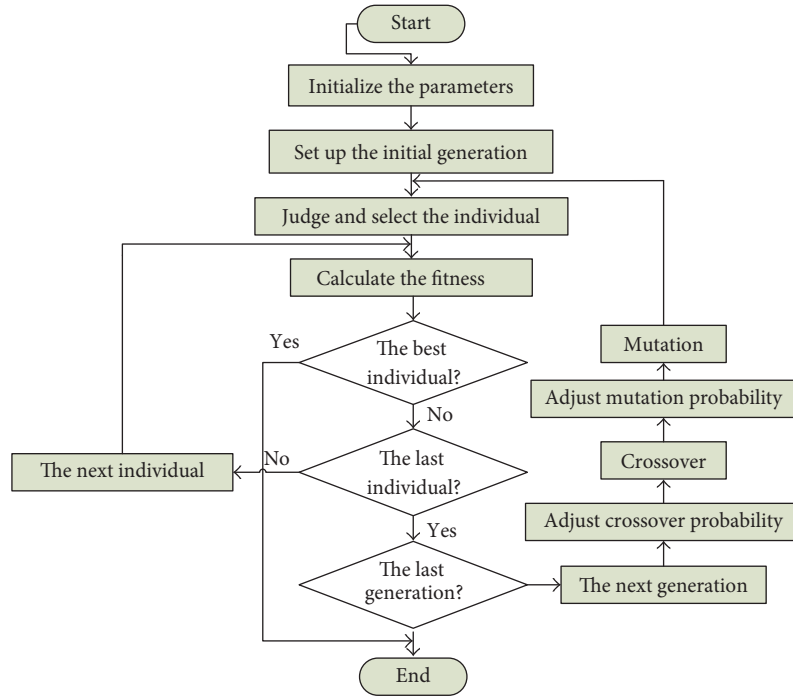
Figure 3: The flowchart of adaptive genetic algorithm.

(C) Crossover. To gain good performance in AGA, the crossover probability is determined by a function compared with a constant in conventional GA. A high crossover probability will be set when the fitness value is less than the average value; otherwise, a high fitness value will lead to a low crossover probability. A crossover probability function employed in our study is shown as follows:

$$
p_c = \begin{cases} p_{c_1} - \dfrac{\left(p_{c_1} - p_{c_2}\right)\left(f' - f_{\text{avg}}\right)}{f_{\max} - f_{\text{avg}}}, & f \geq f_{\text{avg}} \\ p_{c_1} & f < f_{\text{avg}}, \end{cases} \tag{5}
$$

where p_c denotes the crossover probability function and $f_{\max}$ and f_{avg} represent the maximum fitness and average fitness values, respectively. $p_{c1} = 0.9$ and $p_{c2} = 0.6$ are set in step (A).

(D) Mutation. The selection of mutation probability values is the same as that for the selection of crossover probability values. The mutation probability is set to 0.1 when the fitness value is lower than the average value; otherwise, a high fitness value will result in a high crossover probability value. The corresponding mutation probability function is given as

$$
p_m = \begin{cases} p_{m_1} - \dfrac{\left(p_{m_1} - p_{m_2}\right)\left(f_{\max} - f\right)}{f_{\max} - f_{\text{avg}}}, & f \geq f_{\text{avg}} \\ p_{m_1} & f < f_{\text{avg}}, \end{cases} \tag{6}
$$

where p_m refers to the mutation of probability function; $p_{m1} = 0.1$ and $p_{m2} = 0.01$ are determined in the first step; and $f_{\max}$ and f_{avg} represent the maximum and average fitness degrees for each generation, respectively.

5. Estimating Results and Future Projections

5.1. Cointegration Analysis

5.1.1. Unit Root Test. After data collection and preprocess, the first step is to perform the unit root test to determine whether the time series satisfies the basic conditions for constructing the cointegration relationship. Considering that ADF and PP tests are distorted in small sample sizes, the Ng and Perron [40] unit root test is adopted using only the intercept term and the presence of intercept and trend terms in the unit root estimating equation. The corresponding unit root tests are shown in Table 2.

The first six rows in Table 2 are the unit root test results with the presence of an intercept term, while the last six rows are unit root test results with the presence of intercept and trend terms. This result indicates that all the time series are first-difference stationary at the 10% significance level with the presence of an intercept term, a situation that satisfies the necessary requirements for building the cointegration relationship of Johansen–Juselius technique [35] and ARDL bound testing approach [36] simultaneously.

5.1.2. Cointegration Test. Next, we conducted the test to determine the presence of a long-run relationship using the ARDL bounds testing approach. The ordinary least squares (OLS) procedure is first employed for the next equation, which is expressed as

$$
\Delta E_t = \lambda_0 + \sum_{i=1}^{P} \delta_{1i} \Delta E_{t-i} + \sum_{i=0}^{P} \varepsilon_{1i} \Delta X_{1,t-i} + \sum_{i=0}^{P} \phi_{1i} \Delta X_{2,t-i}
$$
$$
+ \sum_{i=0}^{P} \varphi_{1i} \Delta X_{3,t-i} + \sum_{i=0}^{P} \gamma_{1i} \Delta X_{4,t-i} + \sum_{i=0}^{P} \omega_{1i} \Delta X_{5,t-i}
$$

TABLE 2: Results based on Ng-Perron unit root test.

Variables	Level				The first difference				The second difference			
	MZa	MZt	MSB	MPT	MZa	MZt	MSB	MPT	MZa	MZt	MSB	MPT
E	-7.32*	-1.71	0.23	4.02	-19.04***	-2.90	0.15	1.93	-2.13**	-0.93	0.44	10.62
X_1	1.19	1.74	1.46	145.21	-20.03***	-3.10	0.15	1.45	-1.94	-0.97	0.50	12.46
X_2	-13.70***	-2.45	0.18	2.39	-9.97***	-2.17	0.22	2.70	-13.05**	-2.50	0.19	2.08
X_3	-0.75**	0.41	0.55	24.74	-13.92***	-2.56	0.18	2.03	-11.79**	-2.40	0.20	2.17
X_4	-4.68	-1.30	0.28	5.66	-6.68*	-1.81	0.27	3.74	-13.87**	-2.61	0.18	1.85
X_5	-2.23	-0.88	0.39	9.69	-13.02**	-2.38	0.18	2.52	-21.98**	-3.30	0.15	1.16
E	-2.17	-0.97	0.44	37.97	-7.99	-1.79	0.23	11.90	-5.41	-1.62	0.30	16.77
X_1	-3.16	-1.07	0.34	24.85	-0.41	-0.25	0.61	85.70	-9.73	-2.21	0.23	9.36
X_2	-36.73***	-4.25	0.12	2.70	-10.23	-2.19	0.21	9.20	-13.39	-2.58	0.19	6.85
X_3	-16.68*	-2.87	0.17	5.57	-13.87	-2.58	0.19	6.87	-11.60	-2.39	0.21	7.96
X_4	-13.31	-2.57	0.19	6.91	-6.75	-1.82	0.27	13.50	-13.89	-2.62	0.19	6.64
X_5	-0.48	-0.20	0.41	42.25	-14.15	-2.60	0.18	6.83	-22.46**	-3.34	0.15	4.08
	Critical values (intercept)								Critical values (intercept and trend)			
1% level	-13.80	-2.58	0.17	1.78					-3.42	0.14	4.03	
5% level	-8.10	-1.98	0.23	3.17					2.91	0.17	5.48	
10% level	-5.70	-1.62	0.27	4.45					-2.62	0.18	6.67	

Note. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively; E, X_1, X_2, X_3, X_4, and X_5 denote the ln form of electricity demand, population, GDP, economic structure, urbanization, and energy price, respectively.

TABLE 3: Results of bounds testing approach based on SBC.

Electricity demand function	F-statistics	Lag order	Cointegration
$E_1 = f(X_1, X_2, X_3, X_4, X_5)$	2.5629	1	Inconclusive
$E_2 = f(X_1, X_2, X_3, X_4)$	4.2134*	2	Yes
$E_3 = f(X_1, X_2, X_3, X_5)$	2.5507	2	Inconclusive
$E_4 = f(X_1, X_2, X_4, X_5)$	1.8554	2	No
$E_5 = f(X_1, X_3, X_4, X_5)$	2.5478	2	Inconclusive
$E_6 = f(X_2, X_3, X_4, X_5)$	3.9873*	2	Yes
Critical values			
	$k = 5$ (lower-upper)		$k = 4$ (lower-upper)
1% level	4.134	5.761	4.280	5.840
5% level	2.910	4.193	3.058	4.223
10% level	2.407	3.517	2.525	3.560

Note. (1) E denotes the "ln" form of electricity demand; X_i ($i = 1 \cdots 5$) stand for the "ln" form of population, GDP, economic structure, urbanization, and energy price, respectively; (2) the critical values are taken from the appendix in Narayan [37]; (3) $*$ denotes significance at the 10% level; (4) since there are only thirty-one samples, the max lag order p is 1 for E_1 and 2 for other electric energy demand functions.

$$+ \lambda_{11}E_{t-1} + \lambda_{12}X_{1,t-1} + \lambda_{13}X_{2,t-1} + \lambda_{14}X_{3,t-1}$$
$$+ \lambda_{15}X_{4,t-1} + \lambda_{16}X_{5,t-1} + u_{1t}, \tag{7}$$

where E denotes the "ln" form of electricity demand, and X_i ($i = 1 \cdots 5$) refers to the ln form of population, GDP, economic structure, urbanization rate, and price of electricity, respectively. λ_0 is a constant parameter, and u_t denotes the white-noise process. Δ represents the first-difference operator.

To obtain the optimal lag length for the equation, the ARDL bounds approach should estimate $(p + 1)^k$ times of regressions (p and k, resp, denote the maximum number of lags and the number of variables). The Schwarz–Bayesian criteria (SBC) or Akaike information criterion can then be adopted to determine the optimal lag for this regression. The bounds testing procedure based on the joint F-statistics and Wald statistics is illustrated as follows.

The null hypothesis in the equation is $H_0 : \lambda_i = 0$ against the alternative of $H_1 : \lambda_i \neq 0$, $i = 1, 2, \ldots, 5$. Two sets of critical values are reported in Pesaran et al. [36] and Narayan [37]. The bound statistics in Pesaran et al. [36] are only applicable for a sample size with more than 80 observations; otherwise, Narayan [37] is appropriate. Considering that our sample size is 31 (from 1985 to 2015), the critical values from Narayan [37] for the bounds F-test are more suitable than those from Pesaran et al. [36] to establish the reliable inferences on cointegration.

Null should be rejected when the calculated F-statistics exceeds the upper bound, suggesting that cointegration relationship exists between electricity demand and its factors. No cointegration is found when the calculated F-statistics is below the low critical value. However, few approximate conclusions can be drawn without knowing the order of integration of the underlying regressors when the statistics are located between the bounds. The corresponding results are shown in Table 3. As mentioned by Canyurt and Öztürk [11], the AI-based model did not require many factors to estimate future energy demands. The cointegration tests

(Table 3) are presented using both five input variables and four variables, respectively.

As shown in Table 3, the existence of two cointegration relationships among the variables is confirmed. Hence, the electricity demand functions, including E_2 and E_6, can be applied to estimate future electricity demands. In the following, we employ the electricity demand function E_2 in which population, GDP, economic structure, and urbanization are used as the independent variables to predict future electricity demands.

5.1.3. Stability Test. CUSUM and CUSUMSQ are applied to show the stability of the model, as shown in Figure 4.

In the figure, the plots of CUSUM and CUSUMSQ are located within the critical bounds at the 5% significance level, which suggests that the model is stable. Accordingly, the cointegration relationship between electric energy demand and its factors is reliable.

5.2. Estimating Results. After the long-run equilibrium relationship among the variables is verified, AGA is employed to optimize the coefficients of (1)-(2) (since we employ the ln form of variables, after carrying out the weights of (1)-(2), the electric energy demand can be obtained using $Y = e^{y_{\text{lin}}}$ or $Y = e^{y_{\text{qua}}}$). The electricity demand function, which uses four factors as the economic indicators, namely, GDP, population, economic structure, and urbanization, is selected in the model to predict the future energy demand in China.

To estimate the coefficients for the linear and quadratic forms, the observed data from 1985 to 2015 are used. The linear form for the optimal model is written as follows (for simplicity, we do not present the results in exponential form):

$$E_{\text{lin}} = -7.1855 + 1.3754X_1 + 0.3273X_2 - 1.1367X_3$$
$$+ 2.2540X_4, \tag{8}$$

where E denotes the "ln" form of electricity demand and X_i ($i = 1, 2, 3, 4$) stands for the "ln" form of population, GDP, economic structure, and urbanization. Population growth, economic growth, and urbanization are the leading forces

The straight lines represent critical bounds
at 5% significance level

(a)

The straight lines represent critical bounds
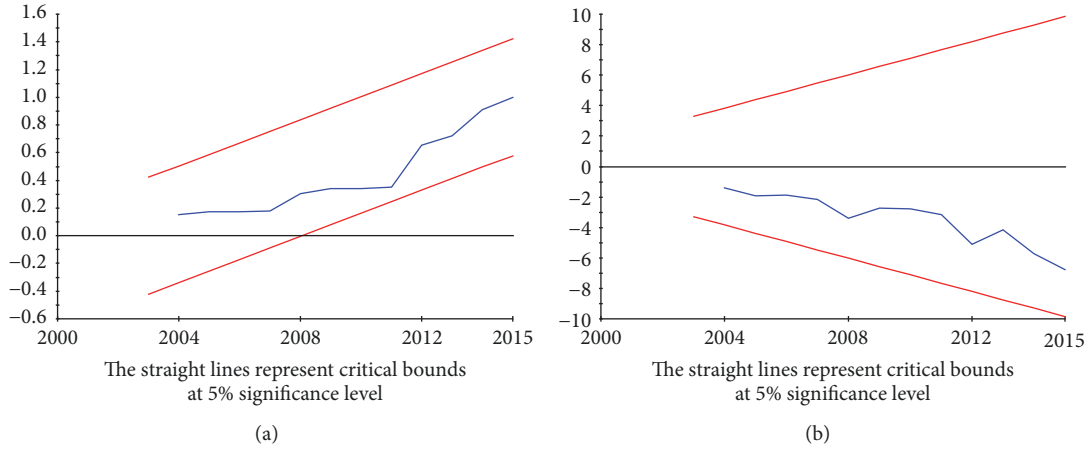at 5% significance level

(b)

FIGURE 4: Plot of CUSUM (a) and CUSUMSQ (b).

contributing to the increase of electricity demand, a finding that is consistent with our expectations. Elasticity coefficients show that a 1% increase in population, economic growth, and urbanization will produce respective increases of 1.3754%, 0.3273%, and 2.254% in electricity consumption. By contrast, a 1% increase in the ratio of tertiary sector to GDP will produce a 1.1367% decline of electric energy consumption.

In addition, the quadratic form for the optimal model is expressed as

$$
\begin{aligned}
E_{\text{qua}} = {} & -1452.9642 + 1085.0132X_1 - 161.0172X_2 \\
& - 62.6839X_3 + 162.8816X_4 + 58.5949X_1X_2 \\
& + 22.8037X_1X_3 - 68.1260X_1X_4 \\
& - 3.9987X_2X_3 + 11.0148X_2X_4 + 5.1778X_3X_4 \\
& - 196.5157X_1^2 - 4.4961X_2^2 - 1.2152X_3^2 \\
& - 3.2263X_4^2.
\end{aligned}
\tag{9}
$$

5.3. Performance Tests. To evaluate the performance of the prediction model, the model must be compared with other forecasting optimal models (GA, ACO, GM, and OLS) using MAE, MSE, MAPE, and RMSE. The corresponding definitions of MAE, MSE, MAPE, and RMSE are shown as follows:

$$
\begin{aligned}
\text{MAE} &= \frac{\sum_{i=1}^{N} \left| y_i - y_i' \right|}{N}, \\
\text{MSE} &= \frac{\sum_{i=1}^{N} \left(y_i - y_i' \right)^2}{N}, \\
\text{MAPE} &= \frac{\sum_{i=1}^{N} \left| \left(y_i - y_i' \right) / y_i \right|}{N}, \\
\text{RMSE} &= \sqrt{\frac{\sum_{i=1}^{N} \left(y_i - y_i' \right)^2}{N}},
\end{aligned}
\tag{10}
$$

where y_i and y_i' represent the actual and fitted values, respectively, and N is the number of observations. All these approaches use the population, GDP, industrial proportion, and urbanization rate of China as the independent variables. The training data (1985–2010) are employed to fit the historical relationship between electricity demand and its factors, and the testing data (2011–2015) are adopted to test the performance. The comparison of these criteria for the testing data among various optimal electricity demand models is reported in Table 4. The table demonstrates that MAPE of the quadratic form is the lowest compared with the linear forms optimized by AGA and other methods (GA, ACO, GM, and OLS). In addition, the proposed linear form of AGA achieves better prediction accuracy than other forecasting optimal models. The actual and simulated data for the optimal model from 1985 to 2015 are also shown in Figure 5, which reveals that the proposed model fits the historical data well. Thus, the discussed AGA algorithm effectively enhances the estimating precision of the model.

5.4. Future Projection. The abovementioned framework is applied to forecast electricity demand from 2016 to 2030 based on three scenarios. The trends of the affecting factors are described as follows:

GDP. A declining trend is observed for the GDP growth of China in recent years. For instance, the annual GDP average growth rates of China during 2005–2010 and 2010–2015 are 11.36% and 7.86%, respectively. Today, China has entered a stage of development called "new normal," which indicates that significant uncertainties may occur in the economic development of China. Hence, the possible impacts of different economic growth rates on electricity consumption should be considered. We set three scenarios for economic growth similar to Lin and Ouyang [3]: high (Scenario A), business as usual (Scenario B), and low (Scenario C). In Scenario A, the average growth rate of GDP between 2016 and 2020 is set as 7%. In Scenario B, the average growth rate is 6.5% because China has to fulfill its national goal established in the 13th Five-Year Plan. In Scenario C, the average growth rate

TABLE 4: Prediction accuracy test for the optimal model.

Method	Criteria	MAE	MSE	MAPE (%)	RMSE
AGA (linear)		0.3122	0.1023	2.01	0.3198
AGA (quadric)		0.2704	0.0932	1.73	0.3052
GA (linear) Canyurt and Öztürk [11]		0.4212	0.1652	2.95	0.4064
GA (quadric)		0.3408	0.1206	2.44	0.3473
ACO (Toksari [13])		0.5874	0.3165	4.15	0.5626
GM (Hsu and Chen [18])		0.6731	0.4239	5.83	0.6511
OLS		0.8019	0.4832	8.68	0.6951

TABLE 5: Hypothesis of variables for the three different scenarios (unit: %).

Period	Growth rate of GDP			Growth rate of population	Growth rate of economic structure	Growth rate of urbanization
	Scenario A	Scenario B	Scenario C			
2016–2020	7.0	6.5	6.0	0.60	2.0	1.5
2021–2025	6.5	6.0	5.5	0.70	2.2	0.8
2026–2030	6.0	5.5	5.0	0.75	2.5	0.4
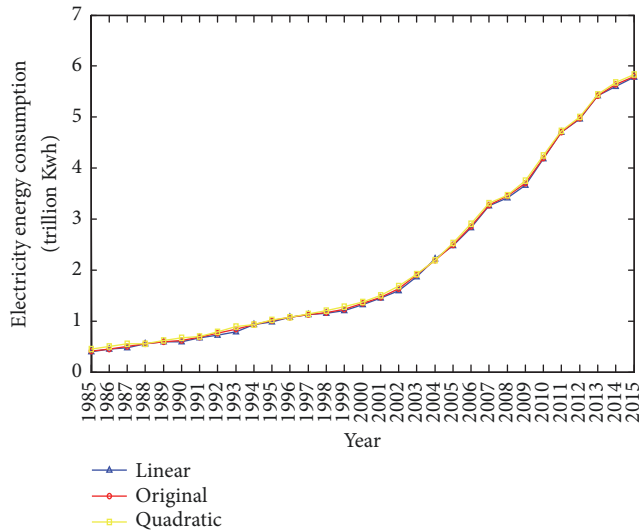
FIGURE 5: The original and prediction time series of electric energy consumption.

is assigned as 6% in 2016–2020, the lowest among the three scenarios. Additional details can be found in Table 5.

Population. The annual growth rate of population in 2010–2015 was approximately 0.5%. We assume that the implementation of the current "two-child policy" will promote population growth. Therefore, we hypothesize that the annual growth rates of population for the periods 2016–2020, 2021–2025, and 2026–2030 are 0.60%, 0.70%, and 0.75%, respectively.

Economic Structure. According to the 13th Five-Year Plan of China, the share of tertiary industries to GDP will be over 56% in 2020, indicating that the annual growth rate of tertiary industries to GDP will be at least 2.09% during 2016–2020. Currently, the economy of China is transitioning

from primary to secondary to tertiary industries. Therefore, the annual growth rates of tertiary industries to GDP for the periods of 2021–2025 and 2026–2030 are assumed to be 2.2% and 2.5%, respectively.

Urbanization. The rapid urbanization process of China is expected to end in 2020 [41]. Additionally, the urbanization process of China will follow an s-curve track, which is similar to the historical experience of most developed countries [3]. We assume that the average annual growth rate for urbanization will decelerate in 2020 and will further decrease to 1.5% in 2016–2020, 0.8% in 2021–2025, and 0.4% in 2026–2030. In summary, the projected GDP growth rates, population, economic structure, and urbanization for the three scenarios are shown in Table 5.

The electricity demand of China can be forecasted after the assumptions of the factors are established. In Figures 6, 7, and 8, the national electricity demand estimates under three scenarios with the linear and quadratic forms are shown, respectively.

The electricity demand of China will continue growing in the medium and long term regardless of the adjustments in economic structure. Under the high-growth scenario (Scenarios A), the electricity demand will still increase rapidly because of the economic growth, urbanization process, and population growth of China. However, the annual growth rate of electricity demand will decrease to 5.8% during the 2016–2030 period in Scenario A owing to the decline in annual growth rate of economic growth and the adjustment in economic structure. This value is much lower compared with that in the period of 2000–2015. In 2020, 2025, and 2030, the electricity demand of China will be 8.2585, 11.139, and 13.821 trillion KWH according to the quadratic form of this model (Figure 6). The minimum predicted values for 2020, combined with 2025 and 2030, were obtained by using the linear form of the optimal model, which result in 8.0322, 10.758, and 13.302 trillion KWH, respectively, under Scenario A. The maximal predicted electricity demand in Scenario
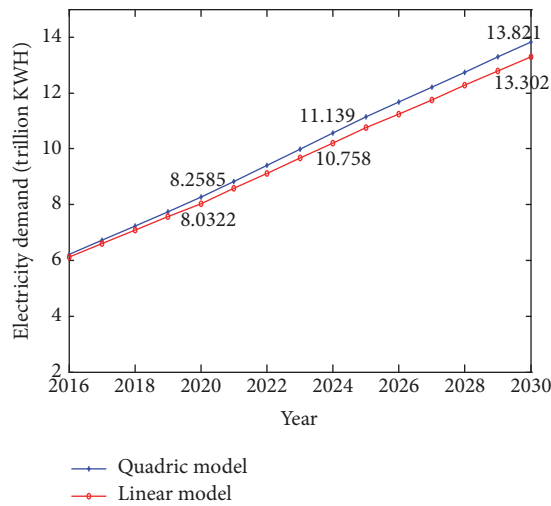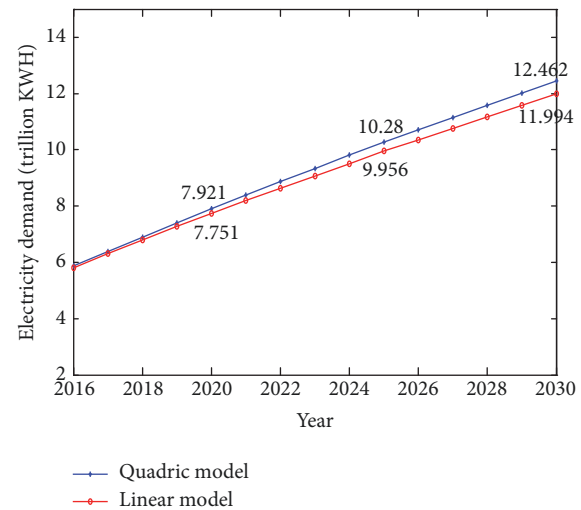
FIGURE 6: Electricity demand in Scenario A.

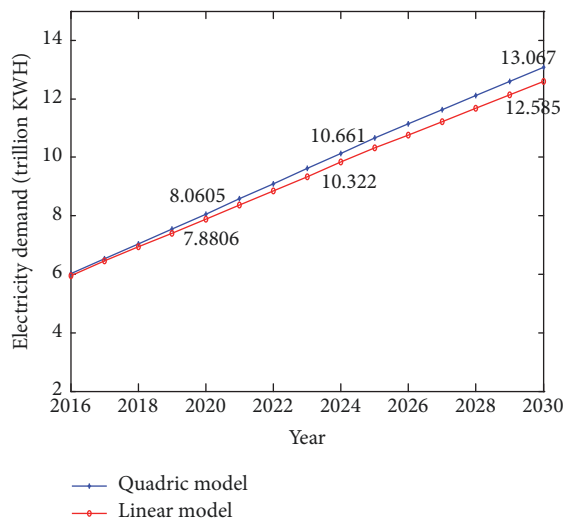FIGURE 7: Electricity demand in Scenario B.

FIGURE 8: Electricity demand in Scenario C.

B in 2020 is about 8 trillion KWH (Figure 7), whereas the minimum is 7.8806 trillion KWH according to the linear form of the model. By 2025, the electricity demand consumption in Scenario B will save about 0.4 trillion KWH compared with Scenario A due to the lower growth rates in GDP. In 2030, the electricity demand in Scenario B will consume less than 0.6 trillion KWH compared with Scenario A. As shown in Figure 8, the electricity demands of China are small under Scenario C. In this scenario, the electricity demand will consume about 7.9, 10.3, and 12.4 trillion KWH in 2020, 2025, and 2030, respectively.

6. Results and Policy Conclusions

In this study, we develop a new framework to predict energy demand based on the conventional AI models and cointegration theory. To develop energy forecasts, we emphasize the use of appropriate data and econometric techniques rather than several computer packages for demand estimation techniques provided by previous studies. In this new framework, the energy demand-affecting factors, which are used as the independent variables in the prediction model, are determined based on theoretical analysis and selected by statistical and econometric analysis or tests. Finally, the future electricity demands of China from 2016 to 2030 are predicted as an example for the new model by using the modified AI-based model. Compared with several previous AI-based literatures, we prove that the present forecasting model demonstrates exceptional performance in forecasting electric energy demand.

The prediction results of electricity demand indicate that population growth, economic growth, and urbanization are the leading forces contributing to the increase of electricity demand, whereas economic structure adjustment is responsible for the decline of electricity consumption. Several specific results are listed below: an electricity demand growth is observed in China in the following years (i.e., 2016–2030). However, the future annual growth rate is lower compared with the last decades. Based on our analysis, electricity demand will still continue to increase at an annual average rate of about 5.5% and will be about 13 trillion KWH in 2030. This value corresponds to nearly two times compared with the 2015 level. The forecasts would be valuable for policy makers in China in planning future energy policies.
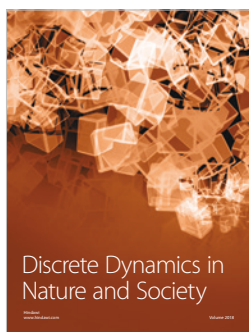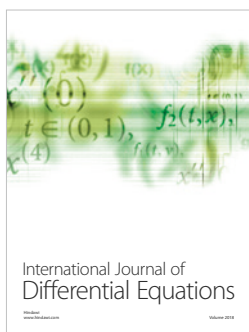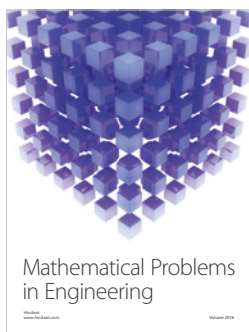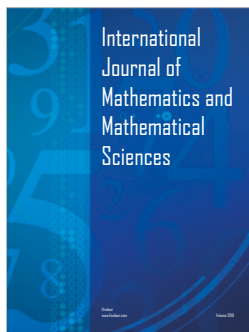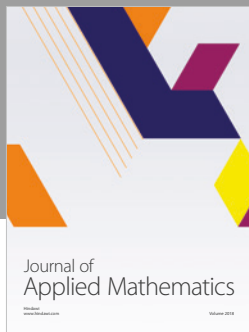
Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] BP. World Energy outlook 2017 ed. Paris, Francis; 2017.

[2] B. Q. Lin, "Structural changes, efficiency improvement and electricity demand forecasting," *Economic Research*, vol. 5, pp. 57–65, 2003 (Chinese).

[3] B. Lin and X. Ouyang, "Energy demand in China: Comparison of characteristics between the US and China in rapid

urbanization stage," *Energy Conversion and Management*, vol. 79, pp. 128–139, 2014.

[4] J. W. Taylor, "Short-term electricity demand forecasting using double seasonal exponential smoothing," *Journal of the Operational Research Society*, vol. 54, no. 8, pp. 799–805, 2003.

[5] J. W. Taylor, "An evaluation of methods for very short-term load forecasting using minute-by-minute British data," *International Journal of Forecasting*, vol. 24, no. 4, pp. 645–658, 2008.

[6] E. Erdogdu, "Electricity demand analysis using cointegration and ARIMA modelling: a case study of Turkey," *Energy Policy*, vol. 35, no. 2, pp. 1129–1146, 2007.

[7] J. D. A. Cabral, L. F. L. Legey, and M. V. D. Freitas Cabral, "Electricity consumption forecasting in Brazil: A spatial econometrics approach," *Energy*, vol. 126, pp. 124–131, 2017.

[8] C. Deb, F. Zhang, J. Yang, S. E. Lee, and K. W. Shah, "A review on time series forecasting techniques for building energy consumption," *Renewable & Sustainable Energy Reviews*, vol. 74, pp. 902–924, 2017.

[9] L. Frías-Paredes, F. Mallor, M. Gastón-Romeo, and T. León, "Assessing energy forecasting inaccuracy by simultaneously considering temporal and absolute errors," *Energy Conversion and Management*, vol. 142, pp. 533–546, 2017.

[10] S. Haldenbilen and H. Ceylan, "Genetic algorithm approach to estimate transport energy demand in Turkey," *Energy Policy*, vol. 33, no. 1, pp. 89–98, 2005.

[11] O. E. Canyurt and H. K. Öztürk, "Three different applications of genetic algorithm (GA) search techniques on oil demand estimation," *Energy Conversion and Management*, vol. 47, no. 18-19, pp. 3138–3148, 2006.

[12] Z. W. Geem and W. E. Roper, "Energy demand estimation of South Korea using artificial neural network," *Energy Policy*, vol. 37, no. 10, pp. 4049–4054, 2009.

[13] M. Duran Toksari, "Ant colony optimization approach to estimate energy demand of Turkey," *Energy Policy*, vol. 35, no. 8, pp. 3984–3990, 2007.

[14] M. D. Toksarı, "Estimating the net electricity energy generation and demand using the ant colony optimization approach: case of Turkey," *Energy Policy*, vol. 37, no. 3, pp. 1181–1187, 2009.

[15] S. Yu, K. Zhu, and X. Zhang, "Energy demand projection of China using a path-coefficient analysis and PSO-GA approach," *Energy Conversion and Management*, vol. 53, no. 1, pp. 142–153, 2012.

[16] S. Yu and K. Zhu, "A hybrid procedure for energy demand forecasting in China," *Energy*, vol. 37, no. 1, pp. 396–404, 2012.

[17] L. Liu, J. Huang, and S. W. Yu, "Prediction of primary energy demand in China based on AGAEDE optimal model," *Chinese Journal of Population Resources and Environment*, vol. 14, no. 1, pp. 16–29, 2016.

[18] C. Hsu and C. Chen, "Applications of improved grey prediction model for power demand forecasting," *Energy Conversion and Management*, vol. 44, no. 14, pp. 2241–2249, 2003.

[19] B. Dong, C. Cao, and S. E. Lee, "Applying support vector machines to predict building energy consumption in tropical region," *Energy and Buildings*, vol. 37, no. 5, pp. 545–553, 2005.

[20] Y. S. Lee and L. I. Tong, "Forecasting nonlinear time series of energy consumption using a hybrid dynamic model," *Applied Energy*, vol. 94, pp. 251–256, 2012.

[21] L. Tang, L. Yu, S. Wang, and J. Li, "A novel hybrid ensemble learning paradigm for nuclear energy consumption forecasting," *Applied Energy*, vol. 93, pp. 432–443, 2012.

[22] Y. T. Chae, R. Horesh, Y. Hwang, and Y. M. Lee, "Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings," *Energy and Buildings*, vol. 111, pp. 184–194, 2016.

[23] M. E. Günay, "Forecasting annual gross electricity demand by artificial neural networks using predicted values of socioeconomic indicators and climatic conditions: Case of Turkey," *Energy Policy*, vol. 90, pp. 92–101, 2016.

[24] H. L. Chan and S. L. Lee, "Forecasting the demand for energy in China," *The Energy Journal*, vol. 17, no. 1, pp. 19–30, 1996.

[25] P. Chujai, N. Kerdprasop, and K. Kerdprasop, "Time series analysis of household electric consumption with ARIMA and ARMA models," in *Proceedings of the International MultiConference of Engineers and Computer Scientists 2013, IMECS 2013*, pp. 295–300, Hong Kong, China, March 2013.

[26] M. Zhang, H. Mu, G. Li, and Y. Ning, "Forecasting the transport energy demand based on PLSR method in China," *Energy*, vol. 34, no. 9, pp. 1396–1400, 2009.

[27] M. Meng, D. Niu, and W. Sun, "Forecasting monthly electric energy consumption using feature extraction," *Energies*, vol. 4, no. 10, pp. 1495–1507, 2011.

[28] V. N. Vapnik, "An overview of statistical learning theory," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 10, no. 5, pp. 988–999, 1999.

[29] V. Vapnik, S. E. Golowich, and A. Smola, "Support vector method for function approximation, regression estimation, and signal processing," in *Proceedings of the 10th Annual Conference on Neural Information Processing Systems (NIPS '96)*, pp. 281–287, December 1996.

[30] Z. Wang, R. S. Srinivasan, and J. Shi, "Artificial intelligent models for improved prediction of residential space heating," *Journal of Energy Engineering*, vol. 142, no. 4, Article ID 04016006, 2016.

[31] J. Kang and H. Zhao, "Application of improved grey model in long-term load forecasting of power engineering," *Systems Engineering Procedia*, vol. 3, pp. 85–91, 2012.

[32] W. Xu, R. Gu, Y. Liu, and Y. Dai, "Forecasting energy consumption using a new GM-ARMA model based on HP filter: the case of Guangdong Province of China," *Economic Modelling*, vol. 45, pp. 127–135, 2015.

[33] Z. Wang and R. S. Srinivasan, "A review of artificial intelligence based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models," *Renewable & Sustainable Energy Reviews*, vol. 75, pp. 796–808, 2017.

[34] R. F. Engle and C. W. J. Granger, "Co-integration and error correction: representation, estimation, and testing," *Econometrica*, vol. 55, no. 2, pp. 251–276, 1987.

[35] S. Johansen and K. Juselius, "Maximum likelihood estimation and inference on cointegration—with applications to the demand for money," *Oxford Bulletin of Economics and Statistics*, vol. 52, no. 2, pp. 169–210, 1990.

[36] M. H. Pesaran, Y. Shin, and R. J. Smith, "Bounds testing approaches to the analysis of level relationships," *Journal of Applied Econometrics*, vol. 16, no. 3, pp. 289–326, 2001.

[37] P. K. Narayan, "The saving and investment nexus for China: evidence from cointegration tests," *Applied Economics*, vol. 37, no. 17, pp. 1979–1990, 2005.

[38] T. W. Feng, L. Y. Sun, and Y. Zhang, "The relationship between energy consumption structure, economic structure and energy intensity in China," *Energy Policy*, vol. 37, no. 12, pp. 5475–5483, 2009.

[39] Y. Li, L. Sun, T. Feng, and C. Zhu, "How to reduce energy intensity in China: A regional comparison perspective," *Energy Policy*, vol. 61, pp. 513–522, 2013.

[40] S. Ng and P. Perron, "Lag length selection and the construction of unit root tests with good size and power," *Econometrica*, vol. 69, no. 6, pp. 1519–1554, 2001.

[41] The World Bank, 2012. China 2030: Building a modern, harmonious, and creative high-income society. USA.