Research Article

Analysis and Application of Grey-Markov Chain Model in Tax Forecasting

Huidi Zhang and Yimao Chen

School of Tax, Jilin University of Finance and Economics, Changchun 130117, Jilin Province, China

Correspondence should be addressed to Huidi Zhang; zhanghuidi@s.jlufe.edu.cn

Tax data is a typical time series data, which is subject to the interaction and influence of economic and political factors and has dynamic and highly nonlinear characteristics. The key to correct tax forecasting is the choice of forecasting algorithm. Traditional tax forecasting methods, such as factor scoring method, factor regression method, and system adjustment method, have a certain guiding role in actual work, but there are still many shortcomings, such as the limitation from the distribution and size of sample data and difficulty of grasping the nonlinear phenomena in economic system. Grey-Markov chain model formed by the combination of grey forecasting and Markov chain forecasting can not only reveal the general developmental trend of time series data, but also predict their state change patterns. Based on the summary and analysis of previous research works, this paper expounds the current research status and significance of tax forecasting, elaborates the development background, current status, and future challenges of the Grey-Markov chain model, introduces the basic principles of grey forecasting model and Markov chain model, constructs the Grey-Markov chain model, analyzes the model's residual error and posteriori error tests, conducts the analysis of Grey-Markov chain model, performs grey forecasting model construction and its state division, implements the calculation of transition probability matrix and the determination of tax forecasting value, discusses the application of the Grey-Markov chain model in tax forecasting, and finally carries out a simulation experiment and its result analysis. The study results show that, compared with separate grey forecasting, Markov chain forecasting, and other commonly used time series forecasting methods, the Grey-Markov chain model increases the accuracy of tax forecasts by an average of 2.3–13.1%. This indicates that the combinative forecasting of Grey-Markov chain model can make full use of the information provided by time series data for tax analysis and forecasting. It can not only avoid the influence of economic, political, and human subjective factors, but also have simple calculations, higher accuracy, and stronger practicality. The study results of this paper provide a reference for further researches on the analysis and application of Grey-Markov chain model in tax forecasting.

1. Introduction

Tax forecasting is a scientific management work that makes a relatively definite judgment on the prospects of future tax revenues. Based on a full grasp of the factors that affect changes in tax revenues and tax historical data, this management work is implemented by using statistical and mathematical methods and through rigorous reasoning and calculations [1]. Therefore, tax forecasting is of great significance for strengthening taxation organization, better fulfilling taxation tasks, and providing services for scientific taxation decision-making and management. The Grey-

Markov chain mode not only considers the dependence of economic phenomena on the time series, but also considers the interference of random fluctuations and has a high accuracy rate for forecasting the short-term trend of economic operation, which involves establishing a model for the target variable in the way of explanatory variable functions. Previous works have also proposed many tax forecasting methods, such as factor scoring method, factor regression method, and system adjustment method [2]. These methods have a certain guiding role in actual work, but there are still many shortcomings, such as the distribution of sample data and restrictions on sample size. However, it is difficult for

traditional linear macroeconomic models to grasp the nonlinear phenomena in the economic system, which will inevitably lead to increased errors [3]. Tax data is a typical time series data and has dynamic and highly nonlinear characteristics, which is subject to the interaction and influence of economic and political factors. Therefore, the key to correct tax forecasting is the choice of forecasting algorithm [4].

Grey forecasting model is a set of mathematical models given by differential equations, which can be used to observe, analyze, and predict the development and changes of the system under study. The advantage of grey forecasting lies in short-term forecasting, but the disadvantage lies in the poor fitting of long-term forecasts and volatile data series [5]. The Grey-Markov chain model effectively reduces the influence of a single weak predictor composed of multiple weak predictors and improves the forecasting accuracy. This forecasting method applies the reduction method of the Grey-Markov chain model to the support vector regression machine to eliminate redundant attributes. Combining grey forecasting and Markov chain to form a Grey-Markov chain forecasting model can not only reveal the general trend of the development and change of the data sequence, but also predict the state of the law [6]. The Markov chain forecasting method is unique in solving the forecasting with no after-effect. The combination of the two not only avoids considering other influencing factors, but also makes the forecasting more scientific and practical [7]. The basic idea of using the Grey-Markov chain model to predict is to first establish a grey model and find its forecasting curve; then use the smooth forecasting curve as a benchmark to divide a number of dynamic state intervals; calculate the Markov transition probability matrix at the points and predict its future state; get the predicted value interval and take the midpoint of the interval; and finally get the predicted value with higher accuracy [8].

Based on the summary and analysis of previous research works, this paper expounds the current research status and significance of tax forecasting, elaborates the development background, current status, and future challenges of the Grey-Markov chain model, introduces the basic principles of grey forecasting model and Markov chain model, constructs the Grey-Markov chain model, analyzes the model's residual error and posteriori error tests, conducts the analysis of Grey-Markov chain model, performs grey forecasting model construction and its state division, implements the calculation of transition probability matrix and the determination of tax forecasting value, discusses the application of the Grey-Markov chain model in tax forecasting, and finally carries out a simulation experiment and its result analysis. The study results of this paper provide a reference for further researches on the analysis and application of Grey-Markov chain model in tax forecasting. The detailed chapters are arranged as follows: Section 2 introduces the basic principles of grey forecasting model and Markov chain model; Section 3 conducts the analysis of Grey-Markov chain model; Section 4 discusses the application of the Grey-Markov chain model in tax forecasting; Section 5 carries out a simulation experiment and its result analysis; Section 6 is conclusion.

2. Method and Principle

2.1. Grey Forecasting Model. For a nonstationary random process with a certain changing trend over time, the boundary and connotation of its state are also changing. For this reason, a dynamic state division criterion should be considered. This criterion should be consistent with the basic temporal change trend of the predicted object. Since the original data column A is a nonstationary random sequence conforming to the characteristics of the Markov chain, it can be divided into n states, and any state a_i is expressed as

$$a_i = \lim_{n \to \infty} \sum_{i=1}^{n} b_i c(A) \left(\frac{1}{d_i} - \frac{1}{e_i} \right), \tag{1}$$

where b_i is a parameter determined according to the tax data to be predicted; $c(A)$ is a function of nonstationary time series; d_i is the inertia factor; e_i is the grey element, which will change with time; that is, the state is dynamic.

When looking at a row in A_i, there are two or more equal and similar maximum values of a_i; the criterion is until each row in the matrix has two or more equal and similar maximum values. For a function conforming to the n-order Markov nonflat tax random sequence, a certain state $f_i(a_i)$ can be expressed as

$$f_i(a_i) = \frac{a_i^{2x} + 1}{a_i^{2x} - 1} + \frac{1}{k} \lim_{n \to \infty} \sum_{i=1}^{n} \frac{g_i(a_i)}{h_i(a_i)}, \tag{2}$$

where x is the step of state transition probability; $g_i(a_i)$ is the number of samples transferred from state a_i to state g_i in one step; $h_i(a_i)$ is the total number of samples in h_i state; k is a coefficient selected according to needs.

After determining the future state transition of the system, it also determines the relative value change interval of the predicted value. The median of the interval can be used as the relative value $l_i(a_i)$ of the predicted value of the system at the future time:

$$l_i(a_i) = \left[\frac{1}{m(a_i)} - \frac{1}{f_i(a_i)} \right] \lim_{n \to \infty} \sum_{i=1}^{n} \left[o(a_i) - p(a_i) \right], \tag{3}$$

where $m(a_i)$ is the transition probability of state a_i to the state a_{i+m} through m-step transition; $o(a_i)$ is the number of times that state a_i reaches state a_{i+m} through m-step transition; $p(a_i)$ is the time response coefficient.

Because the original data may have random fluctuations and cannot fully reflect the true internal laws, the original data must be preprocessed before modeling to make it show the internal laws. The preparation method of the forecasting table is to select several periods closest to the forecasting date and take the row vector corresponding to the initial state in the transition matrix corresponding to the number of transition steps according to the distance from the forecasting date to form a new probability matrix. Summing the new probability matrix columns and the state corresponding to the maximum number of transition steps is the possible future state. However, for a nonstationary random process with a certain changing trend over time, the boundary and connotation of the state are changing. For this reason, a

dynamic state division criterion should be considered [5]. This criterion should be consistent with the basic temporal change trend of the predicted object. The data columns obtained in actual problem modeling are often nonequally spaced sequences, and the equally spaced grey model is no longer applicable at this time. Therefore, it is of practical significance to study a forecasting model that is suitable for data series with large random volatility and can overcome the difficulty of nonequal spacing. The modeling process for sequences with equal spacing is similar, and it can be better reflected when the given data column is not too long.

2.2. Markov Chain Model. The essence of the grey model forecasting model is to accumulate the original data sequence once, so that the generated data sequence shows a certain law, and the corresponding curve can be approximated by a typical curve. If $q(a_i)$ is the state a_i through m steps, the number of original data samples is transferred to the state a_{i+m} and the state transition matrix $B(a_i)$ is obtained from the original data samples of the state a_i:

$$B(a_i) = \frac{q_i(a_i)}{l_i(a_i)f_i(a_i)} = \frac{\lim_{n \to \infty} \sum_{i=1}^{n} q_i(a_i)}{\sqrt{\lim_{n \to \infty} \sum_{i=1}^{n} l_i(a_i) \lim_{n \to \infty} \sum_{i=1}^{n} f_i(a_i)}}. \tag{4}$$

The state transition matrix $B(a_i)$ reflects the transition law between the states of the system; $\|B(a_i)\|$ reflects the probability that state a_i transfers to state a_{i+m} through m steps.

Because all the data that can be used in the process of evaluating the accuracy of the forecasting model is its accuracy sequence, if the accuracy indicators can completely describe the characteristics of the accuracy sequence, it is considered that these accuracy indicators can reasonably describe forecasting model. According to the knowledge of mathematical statistics, the stability index $C(a_i)$ is determined as follows:

$$C(a_i) = \frac{1}{\lim_{n \to \infty} \sum_{i=1}^{n} B(a_i)f_i(a_i)} \begin{bmatrix} p(a_1) & B(a_1) & f_1(a_1) \\ p(a_2) & B(a_2) & f_2(a_2) \\ & \cdots & \\ p(a_n) & B(a_n) & f_n(a_n) \end{bmatrix}. \tag{5}$$

The stability index $C(a_i)$ reflects the degree of dispersion of the accuracy data distribution and marks the accuracy stability of the forecasting model. If the time series $D(a_i)$ can be expressed as a linear function of the previous period value and the random term:

$$D(a_i) = \frac{p(a_i)}{2} \lim_{n \to \infty} \sum_{i=1}^{n} \frac{1}{m(a_i)} - \lim_{n \to \infty} \frac{1}{n} \left[\sum_{i=1}^{n} B(a_i)f_i(a_i) \cdot \lg \frac{\sum_{i=1}^{n} q_i(a_i)}{\sum_{i=1}^{n} B(a_i)f_i(a_i)} \right]. \tag{6}$$

The time series $D(a_i)$ is called an autoregressive sequence, the above formula is an n-order autoregressive model, and the real parameters are called autoregressive coefficients. The random term a_i is a mutually independent white noise sequence, which obeys a normal distribution with a mean value of 0 and a variance of a.

Grey forecasting model is a set of mathematical models given by differential equations. The grey model can be used to observe, analyze, and predict the development and changes of the system under study. According to the number of predictors, the grey model can be divided into a first-order multivariate forecasting model and a first-order forecasting model. The essence of grey model forecasting model is to accumulate and generate the original data list once, so that the generated data list shows a certain law. The corresponding curve can be approximated by a typical curve, and then the approximated curve can be used as a model to predict the value of the model, which can make a cumulative reduction to predict the system [9]. From the forecasting process of the grey model above, it can be seen that the predicted geometric figure is essentially a smooth curve with a monotonous increase or a monotonic decrease, which can only reflect the overall change trend of the time series. The idea of improving the Markov chain to the grey model is to

first divide a number of dynamic state intervals based on the smooth forecasting curve; secondly, the Markov transition probability matrix is calculated according to the time points falling into each state interval; finally, the middle point of the state interval of the future predicted value is taken to obtain the predicted value with higher final accuracy.

3. Analysis of Grey-Markov Chain Model

3.1. Construction of Grey-Markov Chain Model. The statistical laws of various state transitions of the system are reflected in the state transition probability matrix. By examining the state transition probability matrix, the future development and changes of the system can be predicted. In tax forecasting, the model needs to list the forecasting table first. The preparation method of the table is to select the year closest to the forecasting year according to the distance from the forecasting year. In the transition matrix corresponding to the number of transition steps, the model takes the row vector corresponding to the initial state to form a new probability matrix. The column vectors of the new probability matrix are summed, and the state corresponding to the maximum number of transfer steps is the predicted state of the system. Therefore, how to use scientific methods to sort

out the original data and find the law of its change is a meaningful study to understand the true face of the system. The self-adaptive filtering model weights and averages the time series data according to the given weights, then adjusts the weights according to the forecasting results to reduce errors, continuously filters the influence of noise on the forecasting results, and repeatedly adjusts the weights to reduce errors until the best weight is found. Figure 1 shows the framework for the construction of Grey-Markov chain model in tax forecasting.

Since the tax revenue forecasting is related to many factors, in order to clarify the internal relations and development rules of the various factors in the system, this paper adopts the Grey-Markov chain model to optimize the selection of relevant economic indicators that affect tax revenue and uses the grey relational analysis method to analyze the tax revenue. Quantitative analysis and screening of macrofactors are carried out on the influencing factors of income, the effects of many factors in tax revenue are evaluated, and the factors worth paying attention to in future development are sought.

For the training sample set E, a_i is the value of the input variable, q_i is the corresponding output value, and n is the number of training samples. The regression problem is to find a mapping from the input space to the output space. One of the goals of the Grey-Markov chain model is to seek the regression function $E(a_i)$:

$$E(a_i) = \frac{e^{q_i}}{\lim_{n \to \infty} \sum_{i=1}^{n} e^{a_i}} = \frac{1}{r} \lim_{n \to \infty} \sum_{i=1}^{n} \left[\frac{F(a_i)}{G(a_i)} - \frac{H(a_i)}{I(a_i)} \right],$$

(7)

where $F(a_i)$ is the weight of the input variable a_i; $G(a_i)$ is the threshold of the input variable a_i; $H(a_i)$ is the sensitive loss function of the input variable a_i; $I(a_i)$ is the complex weight parameter of the balance model; r is the weight parameter of the training error term.

When the state division is not appropriate, there is no original data in a certain state, $E(a_i) = E(a_{i+m}) = 0$. From this, the state transition probability matrix $J(a_i)$ is to predict the future state turn:

$$J(a_i) = \begin{bmatrix} J_{11}(a_i) & J_{12}(a_i) & \cdots & J_{1n}(a_i) \\ J_{21}(a_i) & J_{21}(a_i) & \cdots & J_{2n}(a_i) \\ \vdots & \vdots & \vdots & \vdots \\ J_{n1}(a_i) & J_{n1}(a_i) & \cdots & J_{nn}(a_i) \end{bmatrix}.$$

(8)

Generally, only one-step transition probability matrix $J(a_i)$ is considered, but when the future turn of the state is difficult to determine, it is necessary to examine the multistep transition probability matrix $J(a_{i+m})$. The transition probability reflects the degree of influence of various random factors, so the Markov chain is suitable for data forecasting problems with large random volatility, and it can just make up for the limitations of grey forecasting.

Grey-Markov chain model analysis is to measure the degree of relevance between factors according to the degree of similarity or difference in the development of factors and
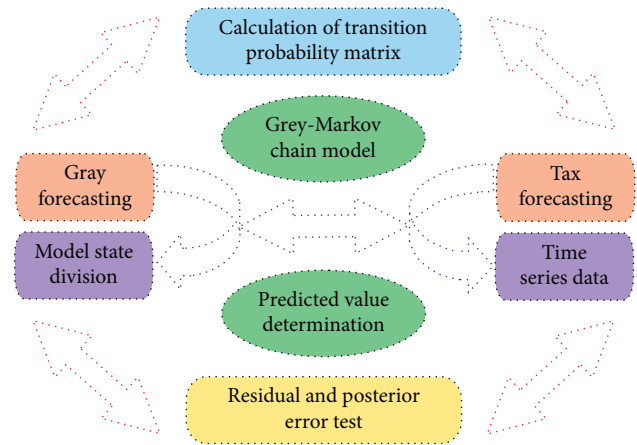

Figure 1: Framework for the construction of Grey-Markov chain model in tax forecasting.

is a quantitative analysis of the development of dynamic processes [10]. But for a specific problem, what kind of forecasting model should be chosen and what kind of sequence operator should be introduced should be based on sufficient qualitative analysis. In fact, in the development of the grey system of total taxation, as time goes by, some random disturbances or driving factors will continue to enter the system, and the development of the system will be successively affected by it. Therefore, by constantly adding new data and removing old data in time, the modeling sequence will better reflect the current characteristics of the system, and the forecasting accuracy will be greater.

3.2. Residual and Posterior Error Test. The above-mentioned characteristics of the Markov chain determine that its forecasting is based on the transition probability between states to infer the future development and changes of the system. The transition probability reflects the degree of influence of various random factors, so the Markov chain is suitable for data forecasting problems with large random volatility, and at this point it can just make up for the limitations of grey forecasting. However, the forecasting object of Markov chain also requires a stable process equal mean value. The forecasting problem of the total tax amount is a nonstationary random process that shows a certain changing trend with time [11]. Using the grey model to fit the time series data of the total amount of cultivated land and find out its change trend can make up for the limitations of Markov chain forecasting. When making long-term forecasts, the forecasting value is often high or low, so the fitting of the data series with large random volatility is poor, and the forecasting accuracy is low (Figure 2). The grey system forecasting model is established by the module generated by the accumulation of time series data. Residual test filters out the random amount that may be mixed in the original sequence, looks for some implicit regularity from the time series that fluctuates up and down, and obtains a new series with weakened randomness and strengthened regularity. The posterior error test can also dig out the inherent
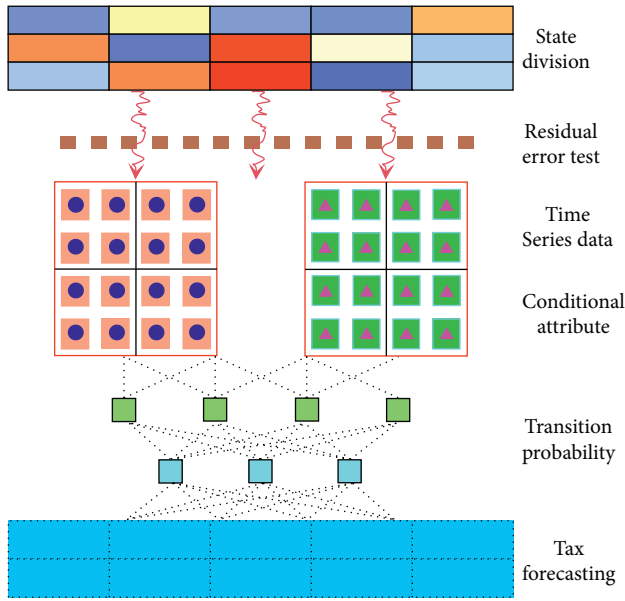
Figure 2: Residual error test and posterior error test of Grey-Markov chain model in tax forecasting.

characteristics of the original sequence, which can be seen from the process of establishing the grey forecasting model.

The research of econometric theory assumes that the time series to be predicted is generated by a random process. If the stochastic characteristics of the stochastic process do not change with time, the time series can be modeled by the deterministic coefficient equation, and the coefficients of the equation can be estimated using the past data of the series and traditional regression techniques. An important concept involved here is the singularity of the time series [12]. If a time series must be transformed into a stationary and reversible time series after difference, it is said that the time series has the singularity.

Assuming that attribute value of each object in time series data K has been characterized, the equivalent relationship on K can be determined to establish a knowledge system; therefore, the dependence of economic index a_i on tax factors can be calculated based on this:

$$K(a_i) = \frac{1}{\sqrt{\alpha}} \sqrt{\lim_{n \to \infty} \sum_{i=1}^{n} \left[s(a_i) - t(a_i) \right]^2 - \frac{1}{\beta} \lim_{n \to \infty} \sum_{i=1}^{n} |u(a_i) - v(a_i)|},$$

(9)

where $s(a_i)$ is the set of conditional attributes; $t(a_i)$ is the set of decision-making attributes; $u(a_i)$ is the attribute value of the conditional attribute with respect to the object a_i; $v(a_i)$ is the change feature of the tax index; α is the feature set; and β is the weight coefficient.

The moving average process of the Grey-Markov chain model test is unconditionally stable, but it is usually hoped that the input process and the output process can express each other, so the roots of the lag polynomial $L(a_i)$ is required to be outside the unit circle, which can be obtained by derivation:

$$L(a_i) = (1 - w)y_i(n - 1) + (1 - z)M(x, y) + O(x, y), \quad (10)$$

where w is the summation order; y_i is the learning constant; n is the number of weights; z is the smoothing parameter; $M(x, y)$ is the threshold of hidden layer; $O(x, y)$ is the summation subject to autoregressive in nonstationary time series of moving average model.

To judge and test whether a time series is stationary, the model can use the autocorrelation coefficient or unit root test of the time series. The difference sequence of a nonstationary time series is usually stationary, which is why the difference model is traditionally established. However, this method of solving problems has shortcomings, because the difference also smooths out valuable information. The nonstationary of the time series will destroy the normal distribution assumption in traditional regression techniques, and the regression of nonstationary time series will lead to absurd results, which is the false regression problem. Therefore, when establishing an econometric model, the stationary test of the time series is very necessary.

4. Application of Grey-Markov Chain Model in Tax Forecasting

4.1. Grey Model Construction and State Division. The quality of a predictive model is closely related to the accuracy of the predicted object data. It is impossible to determine the quality of a predictive model from a single data case, but it just means that a certain predictive model is not suitable in this case. The predicted system is constantly evolving, and the historical information of the system has a certain effect on the research object, and the information of the period closer to the predicted age is more valuable to the development and change trend of the research system. In the forward transfer process, the input information is passed from the input layer to the output layer through the hidden layer calculation layer by layer, and the state of each layer of neurons only affects the state of the next layer of neurons. If the desired output is not obtained in the output layer, calculate the error change value of the output layer, then turn to backpropagation, and pass the error signal back along the original connection path through the network to modify the weights of the neurons in each layer until the desired goal is reached. Therefore, the basis of classification is attribute value characterization; that is, the attribute value of each attribute is replaced with a characteristic value and data-driven attribute values may get too many eigenvalues (Figure 3). Too many eigenvalues will make the formation of related rules very complicated and slow in calculation, so the new model can combine data information and experience to determine predicted attributes eigenvalues.

Monthly tax data can be regarded as a random time series formed over time. Through the analysis of the randomness, stationarity, and seasonality of the tax value on the time series, the relevance or dependence relationship between these single-month tax values is described by a mathematical model, so as to achieve the purpose of using the past and present tax value information to predict the future tax situation. Once the model is identified, it can
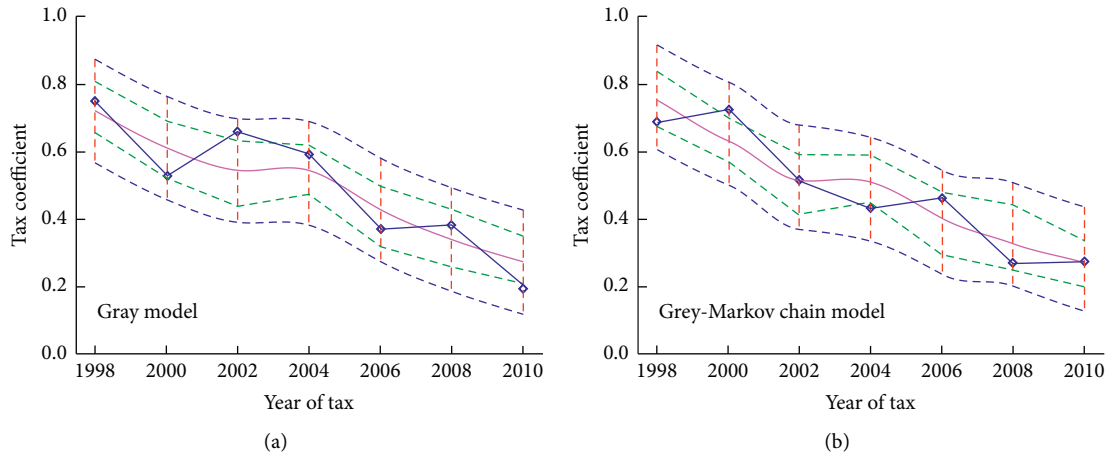
(a)

(b)

FIGURE 3: Tax coefficients of state division in grey model (a) and Grey-Markov chain model (b).

predict future values from the past and present values of the time series. In the process of economic forecasting, the Grey-Markov chain model not only considers the dependence of economic phenomena on the time series, but also considers the interference of random fluctuations and has a high accuracy rate for forecasting the short-term trend of economic operation. This tax forecasting value is based on the current customs supervision level and the level of tax collection and management unchanged or improved, mining the information covered by the tax sample data itself, and using mathematical statistical analysis methods to establish the theoretical forecasting value obtained by the forecasting model. Once the actual external environment and conditions change, such as the country's implementation of macro-control, excessive currency appreciation, and large changes in exchange rates and adjustments in foreign economic policies, it will have a certain impact on the results of tax forecasts [13].

The Grey-Markov chain model is mainly based on the extensional analogy of the direction and degree of the change of the time series, or it directly extends the past to the future, so as to predict the tax level that may be reached in the next period or several periods. From a long-term perspective, the dynamic series of taxation phenomena are constantly rising or falling, which is the objective law of the continuous development of taxation. Predictive modeling involves establishing a model for the target variable in the way of explanatory variable functions; its two types of predictive modeling tasks are classification and regression, which are used to predict discrete target variables and predict continuous target variables, respectively. Association analysis is used to discover patterns that describe strong association features in data. The discovered patterns are usually expressed in the form of implication rules or feature subsets [14]. Since the search space is exponential, the goal of association analysis is to extract the most interesting patterns in an efficient way. Cluster analysis aims to find closely related groups of observations so that observations belonging to the same cluster are as similar to each other as possible compared to observations belonging to different clusters. The goal of the anomaly detection algorithm is to find the real anomaly and to avoid mistakenly marking normal objects as anomalous points. In other words, a good anomaly detector must have a high detection rate and a low false alarm rate.

4.2. Calculation of Transition Probability Matrix and Determination of Predicted Value. The average absolute percentage error of the Grey-Markov chain model is lower than the average absolute percentage error of the previous three individual tax forecasting models; the mean square percentage error of the Grey-Markov chain model is lower than the average square percentage error of the previous three individual tax forecasting models. This result shows that the Grey-Markov chain model has a good tax forecasting effect under the two test indicators of average absolute percentage error and mean square percentage error (Figure 4). The combined forecasting model has better stability than a single forecasting model and can better adapt to different sleeping sound forecasting environments [15]. Different single forecasting models often have great differences in their accuracy, and a single forecasting model cannot cover all the effective information of tax revenue. But if some methods are simply discarded, some useful conclusions will be lost. Generally speaking, combined forecasting can make full use of the useful information of each individual forecasting model to improve the accuracy and comprehensiveness of forecasting. Establishing an accurate and dynamic tax forecasting model based on a reliable model is a prerequisite for ensuring that tax policies are becoming scientific. It will also provide a scientific and rational basis for the formulation of fiscal budgets and macroeconomic regulation.

In the process of tax data forecasting, many attributes are often used, but not all attributes are equally important to the forecasting result. Too many redundant attributes will lead to too much calculation and the parametric method is mostly used in practice. The parametric method requires the construction of an explicit function form and generally implies the assumption of a linear relationship, while the

2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	Grey-Markov chain model
													State division
													Residual error test
													Posterior error test
													Time Series data
													Transition probability
													Tax forecasting
													Conditional attribute
													Weight coefficient
													Summation subject

FIGURE 4: Calculation of transition probability matrix and determination of predicted value in the application of Grey-Markov chain model in tax forecasting.

actual problem is mostly a nonlinear relationship. In Grey-Markov chain model, the tax forecast accuracy of nonequally spaced sequence decreases from 0.98 in 2015 to 0.05 in 2020, while that of nonstationary random process decreases from 0.89 in 2015 to 0.50 in 2020 (Figure 5). The dependence and importance of the attributes in the grey model process the sample set to obtain a series of reduced attributes. Data forecasting uses the obtained training model to predict the test sample and obtain the forecasting accuracy. This forecasting method applies the reduction method of the Grey-Markov chain model to the support vector regression machine to eliminate redundant attributes. Among them, the dependence between conditional attributes and decision attributes is measured by the information entropy between them, so that they can be compared. It accurately reflects the degree of interdependence between conditional attributes and decision-making attributes, so as to more accurately judge whether an attribute is an important attribute. Through the processing of attributes, certain attributes are reduced and the reliability of the training model is also guaranteed.

The Grey-Markov chain model usually divides the obtained comprehensive indicators into principal components, of which the one with the largest proportion is called the first principal component, the one with second largest proportion is called the second principal component, and so on. Each principal component is a linear combination of the original indicators and the principal components are not related to each other. Through principal component analysis, the relationship between multiple indicators and dependent variables can be obtained, without losing too much information and obtaining more new and new information. It emphasizes the similarities and differences between data and is a general method for processing high-dimensional data. For the original data, the Grey-Markov chain model uses some transformations to extract the inherent characteristics between the data, one of which is to achieve it through linear transformation [16]. It is usually necessary to find the tax base or agency tax base corresponding to each tax type and the forecasting value of the corresponding economic index must be obtained. Even for the correct fit model, the accuracy of income forecasting depends on the accuracy of forecasting independent variables such as economic indicators. If, in the past period, the economic structure and taxation policies have been undergoing major changes, it will be more difficult to adjust the historical data, and it will directly affect the accuracy of the forecasting.

5. Simulation Experiment and Result Analysis

5.1. Simulation Experiment Design. This paper selects the tax revenue of a certain province in China from 2000 to 2020 as the experimental data and uses the Grey-Markov chain model to conduct a confirmatory experiment of tax forecasting. According to the size of the influencing factors, the comparability of data, and the requirements of forecasting models, the following 10 indicators (each indicator unit is 10,000 yuan) are selected as the indicator system that affects taxation: primary industry added value, secondary industry increase value, added value of the tertiary industry, total agricultural output value, total industrial output value, fixed asset investment, total retail sales of consumer goods, total sales of goods, total foreign trade import and export, and balance of urban and rural residents' savings deposits. According to the statistical yearbook to obtain the data of each indicator, the data is divided into two parts, the data from 2000 to 2015 is used as training data, and the data from 2015 to 2020 is used as test data.

The Grey-Markov chain models identify patterns or relationships in data. Unlike predictive models, descriptive models provide a way to explore the properties of the analyzed data, rather than predicting new properties. In a time
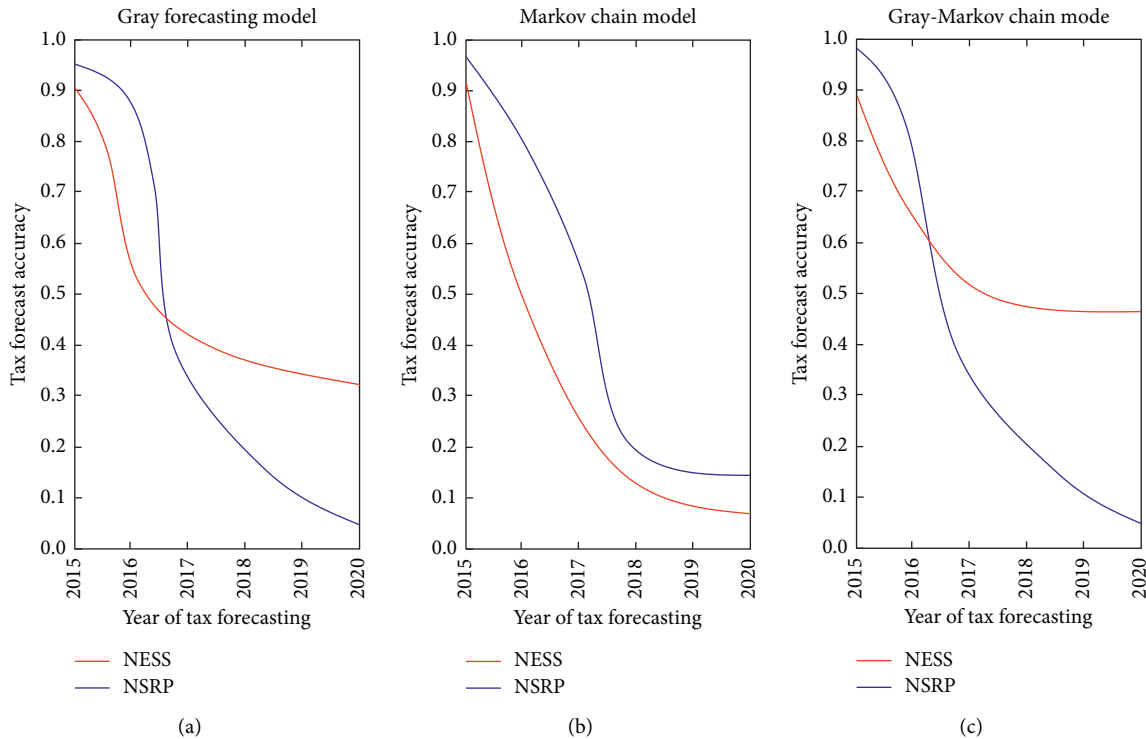
FIGURE 5: Tax forecast accuracies in grey forecasting model (a), Markov chain model (b), and Grey-Markov chain model (c). *Note*: NESS, nonequally spaced sequence; NSRP, nonstationary random process.

series, the attribute value of the data is constantly changing over time. Under normal circumstances, these data can be obtained in equal time intervals. Time series analysis has three basic functions: first, use distance metrics to determine the similarity of different time series; second, examine the structure of time series graphs to determine the behavior of time series; third, use historical time series graphs to predict future values of data [17]. Classification is used to predict the category of data objects, that is, use existing information to predict unknown events. For example, predict whether the taxpayer will evade or evade tax or the level of tax evasion (Figure 6). The time series is generally composed of two basic elements: one is the period of the variable reflecting the economic phenomenon, and the other is the index reflecting the economic phenomenon and its index value. One feasible method is to collect a large amount of tax-related data in different periods. In addition to tax data, the following data that has an impact on tax revenue should also be collected and combined with the Grey-Markov chain model; it can be carried out based on various previous data.

5.2. Result Analysis. Different models select the rectangular features that best distinguish between positive and negative samples. For each feature, the weak classifier is given an optimal threshold of the classification function, so that the fewest samples are misclassified. The advantage of the Grey-Markov chain model is that it uses weighted training data instead of randomly selected training samples, combines weak classifiers, and uses a weighted voting mechanism instead of an average

voting mechanism. The forecasting error value of the strong predictor is smaller than the average forecasting error value of the weak predictor, and the forecasting result of the strong predictor is better [18]. It shows that the Grey-Markov chain model effectively reduces the influence of a single weak predictor composed of multiple weak predictors and improves the forecasting accuracy. During training, the Grey-Markov chain model is trained with the training set data. After training, the system automatically inputs the sample data in the validation set data into the neural network for verification. After the validation set is input, an error will be obtained. The system will judge whether this error has stabilized after multiple iterations. If it does not decrease or increase, the model will stop training on the training set to prevent overlearning. Therefore, the strong predictor tax forecasting method of the Grey-Markov chain model has a good forecasting effect and is a feasible tax forecasting method.

The actual economic system is a discrete nonlinear system. The nonlinear chaotic dynamics model is more suitable for studying economic system, and it can more accurately reveal the operating law of economic system. Compared with the statistical method of measurement, the Grey-Markov chain model does not need to know the mathematical equation describing the independent variable and the dependent variable in advance and can approximate any continuous function with arbitrary precision [19]. The fault tolerance of the Grey-Markov chain model also allows noise in the input samples, which is of great significance for economic forecasting problems where the collection of samples is restricted by various conditions.
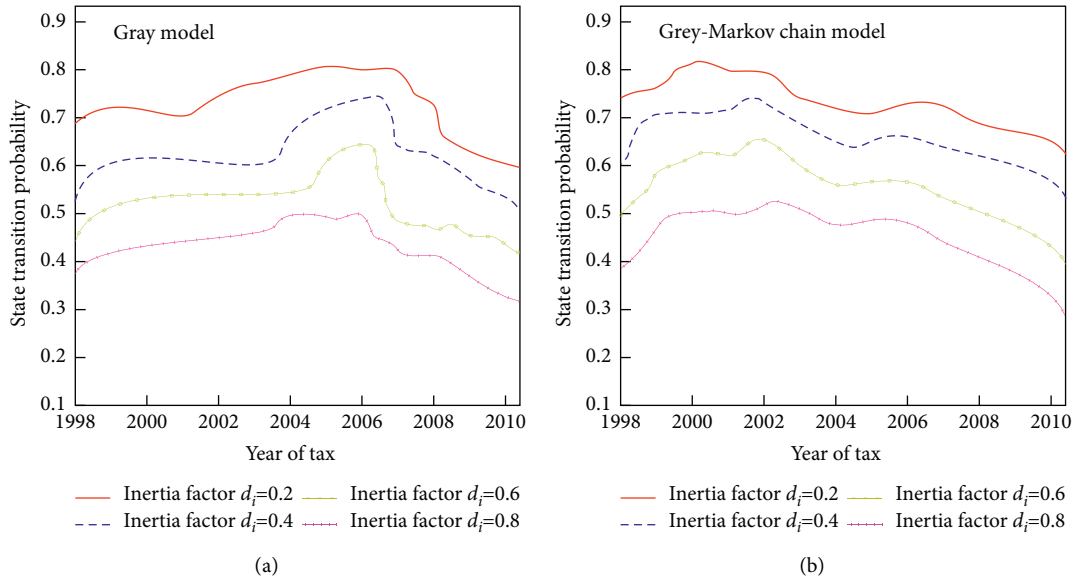
FIGURE 6: State transition probabilities with different inertia factors (d_i = 0.2, 0.4, 0.6, and 0.8) in the grey model (a) and Grey-Markov chain model (b).
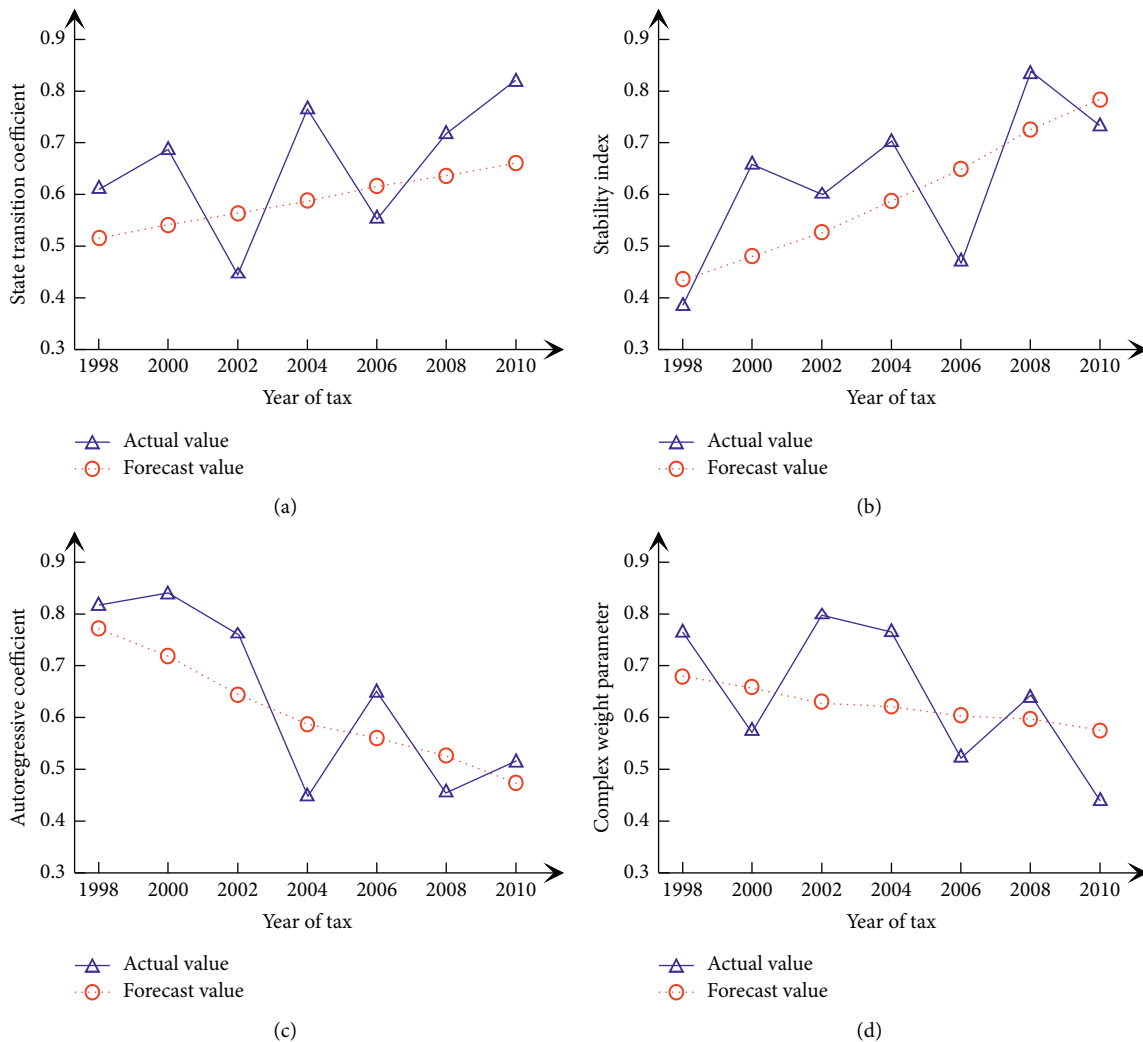

FIGURE 7: State transition coefficients (a), stability indexes (b), autoregressive coefficients (c), and complex weight parameters of actual and forecast tax values in the Grey-Markov chain model.
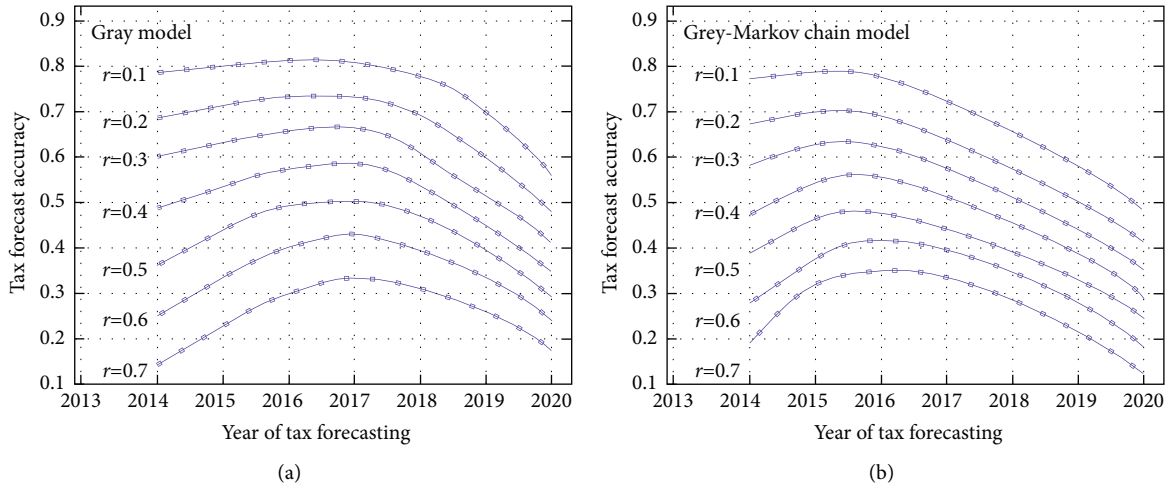
FIGURE 8: Tax forecast accuracies with different weight parameters of training error terms in the grey model (a) and Grey-Markov chain model (b).

The state division is based on the predicted value $Y(k)$, and four parallel curves are made on the upper and lower sides, and the area between each two adjacent curves is a state. The nonstationary random sequence conforming to the Markov chain is divided into 4 states:

$$
\begin{aligned}
Y_i(a_i) &= Y(a_i) + 600, \\
Y_i(a_i) &= Y(a_i) + 300, \\
Y_i(a_i) &= Y(a_i) - 300, \\
Y_i(a_i) &= Y(a_i) - 600.
\end{aligned}
\qquad (11)
$$

Due to the long time span, in order to eliminate the impact of price index and inflation, the original data must be converted according to the current year's price. In order not to affect the forecasting effect of the model, the customs taxation and securities transaction stamp duty are deducted from the original data. Regarding whether the traditional feedforward network model can be used for effective time series forecasting, the three-layer forward network and the Grey-Markov chain model represent the nonlinear autoregressive model and the nonlinear autoregressive moving average model, respectively (Figure 7). Therefore, when modeling tax forecasts, there is a problem of choosing a network model. The model cannot blindly adopt a three-layer forward network. From the perspective of time series analysis, it is more reasonable to use a feedback network [20].

The difference in the choice of dependent variables for tax forecasts makes tax forecasts divide into total forecasts and individual forecasts. The total forecasting uses the total tax revenue as the dependent variable, while the individual forecasts use turnover tax, income tax, and wealth tax as dependent variables. The difference in the choice of the dependent variable will also cause a change in the relationship between the independent variable and the dependent variable, because the correlation between any relevant economic factor as an independent variable and different dependent variables is inherently different. Choosing

different ranges of tax revenues as dependent variables makes tax forecasts divide into comprehensive forecasts and partial or regional forecasts [21]. The relationship between the independent variable and the dependent variable in the model is also different between comprehensive forecasting and regional forecasting and between different regional forecasts. In the ideal Grey-Markov chain model, the selected variables should be highly correlated with tax revenue, the selected variables are independent of each other, and the remainder of the model tends to zero; that is, all relevant factors are basically summarized in the model (Figure 8). It is basically impossible to require the selected variables to be completely independent. However, when selecting, the correlation between the variables can be required to be as low as possible. Only in this way can the uniqueness and reliability of the explanatory variable explaining the dependent variable be guaranteed.

6. Conclusions

This paper constructs the Grey-Markov chain model, analyzes the model's residual error and posteriori error tests, conducts the analysis of Grey-Markov chain model, performs grey forecasting model construction and its state division, implements the calculation of transition probability matrix and the determination of tax forecasting value, discusses the application of the Grey-Markov chain model in tax forecasting, and finally carries out a simulation experiment and its result analysis. The dependence and importance of the attributes in the grey model process the sample set to obtain a series of reduced attributes. Data forecasting uses the obtained training model to predict the test sample and obtain the forecasting accuracy. The Grey-Markov chain model is mainly based on the extensional analogy of the direction and degree of the change of the time series, or it directly extends the past to the future, so as to predict the tax level that may be reached in the next period. Regarding whether the traditional feedforward network model can be

used for effective time series forecasting, the three-layer forward network and the Grey-Markov chain model represent the nonlinear autoregressive model and the nonlinear autoregressive moving average model, respectively. The tax data uses the obtained training model to predict the test sample to obtain the forecasting accuracy. The study results show that, compared with separate grey forecasting, Markov chain forecasting, and other commonly used time series forecasting methods, the Grey-Markov chain model increases the accuracy of tax forecasts by an average of 2.3–13.1%. This indicates that the combinative forecasting of Grey-Markov chain model can make full use of the information provided by time series data for tax analysis and forecasting. It can not only avoid the influence of economic, political, and human subjective factors, but also have simple calculations, higher accuracy, and stronger practicality. The study results of this paper provide a reference for further researches on the analysis and application of Grey-Markov chain model in tax forecasting.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

References

[1] L. X. Liu, Y. Q. Zhuang, and X. Y. Liu, "Tax forecasting theory and model based on SVM optimized by PSO," *Expert Systems with Applications*, vol. 38, no. 1, pp. 116–120, 2011.

[2] J. R. Barnard and W. T. Dent, "State tax revenues-new methods of forecasting," *The Annals of Regional Science*, vol. 13, no. 3, pp. 1–14, 1979.

[3] B. Jochimsen and R. Lehmann, "On the political economy of national tax revenue forecasts: evidence from OECD countries," *Public Choice*, vol. 170, no. 3-4, pp. 211–230, 2017.

[4] A. Musayev and K. Aliyev, "Modelling oil-sector dependency of tax revenues in a resource rich country: evidence from Azerbaijan," *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, vol. 65, no. 3, pp. 1023–1029, 2017.

[5] X. J. Wang, S. L. Yang, and J. Ding, "Simulation of orthogonalization prediction based on grey Markov chain for electricity consumption," *Journal of System Simulation*, vol. 22, no. 10, pp. 2253–2256, 2010.

[6] L.-H. Chen and T.-Y. Guo, "Forecasting financial crises for an enterprise by using the Grey Markov forecasting model," *Quality and Quantity*, vol. 45, no. 4, pp. 911–922, 2011.

[7] Z.-Q. Jia, Z.-F. Zhou, H.-J. Zhang, B. Li, and Y.-X. Zhang, "Forecast of coal consumption in Gansu Province based on Grey-Markov chain model," *Energy*, vol. 199, Article ID 117444, 2020.

[8] U. Kumar and V. K. Jain, "Time series models (Grey-Markov, Grey Model with rolling mechanism and singular spectrum analysis) to forecast energy consumption in India," *Energy*, vol. 35, no. 4, pp. 1709–1716, 2010.

[9] D. Streimikiene, R. Raheem Ahmed, J. Vveinhardt, S. P. Ghauri, and S. Zahid, "Forecasting tax revenues using time series techniques - a case of Pakistan," *Economic Research-Ekonomska Istraživanja*, vol. 31, no. 1, pp. 722–754, 2018.

[10] W. Song, X. Ming, and Z. Xu, "Integrating Kano model and grey-Markov chain to predict customer requirement states," *Proceedings of the Institution of Mechanical Engineers—Part B: Journal of Engineering Manufacture*, vol. 227, no. 8, pp. 1232–1244, 2013.

[11] A. Hannon, E. Leahy, and R. O'Sullivan, "An analysis of tax forecasting errors in Ireland," *Economic and Social Review*, vol. 47, no. 3, pp. 391–423, 2016.

[12] L. H. Chen and Y. J. Cheng, "Integrating the event study method and the Grey Markov forecasting model to evaluate the effect of announcements regarding the establishment of FHCs," *International Journal of Management and Decision Making*, vol. 9, no. 4, pp. 415–428, 2008.

[13] B. Bratten, C. A. Gleason, S. A. Larocque, and L. F. Mills, "Forecasting taxes: new evidence from analysts," *The Accounting Review*, vol. 92, no. 3, pp. 1–29, 2017.

[14] J. C. Hambor, M. R. Norman, and R. R. Russell, "A tax revenue forecasting model for the State of Hawaii," *Public Finance Quarterly*, vol. 2, no. 4, pp. 432–450, 1974.

[15] T. M. Fullerton Jr, "A composite approach to forecasting state government revenues: case study of the Idaho sales tax," *International Journal of Forecasting*, vol. 5, no. 3, pp. 373–380, 1989.

[16] S. Bretschneider and W. Gorr, "Economic, organizational, and political influences on biases in forecasting state sales tax receipts," *International Journal of Forecasting*, vol. 7, no. 4, pp. 457–466, 1992.

[17] Z. Yu, C. Yang, Z. Zhang, and J. Jiao, "Error correction method based on data transformational GM (1,1) and application on tax forecasting," *Applied Soft Computing*, vol. 37, pp. 554–560, 2015.

[18] Q. Yu, W. M. Li, and M. X. Shen, "Application of gray sequence prediction in tax forecasting," *Journal of System Simulation*, vol. 8, no. 18, pp. 971-972, 2006.

[19] Q. Chang and Q. Liu, "Forecasting model of tax based on SVM," *Computer Engineering and Design*, vol. 28, no. 7, pp. 1653-1654, 2007.

[20] Z.-X. Wang, "An optimized Nash nonlinear grey Bernoulli model for forecasting the main economic indices of high technology enterprises in China," *Computers and Industrial Engineering*, vol. 64, no. 3, pp. 780–787, 2013.

[21] Y. Chang, B. Liu, and M. Gao, "Mechanism and prediction of China-Russia timber trade from the perspective of sustainable development," *International Journal of Sustainable Development and Planning*, vol. 15, no. 4, pp. 559–567, 2020.