Mathematics

AIMS Mathematics, 8(7): 15600–15617.
DOI: 10.3934/math.2023796
Received date: 27 December 2022
Revised date: 18 March 2023
Accepted date: 19 March 2023
Published date: 27 April 2023

http://www.aimspress.com/journal/ Math

Research article

Forecasting the public financial budget expenditure in Dongguan with an optimal weighted combination Markov model

Dewang Li[1], Daming Xu[2], Meilan Qiu[1] and Shuiping Yang[1,*]

[1] School of Mathematics and Statistics, Huizhou University, Guangdong, Huizhou 516007, China
[2] Hunan college for preschool education, Hunan, Changde 415000, China

* **Correspondence:** Email: ysp@hzu.edu.cn; Tel: +8613138343882.

Abstract: In this work, a novel optimal weighted combination Markov model (OWCMM) is proposed to forecast the public financial budget expenditure of Dongguan, China, from 2016 to 2020. The new model is constructed based on the optimal combination, which includes the fractional grey model, the Fourier function regression model and the autoregressive integrated moving average model (ARIMA), and modifies this optimal combination by the Markov model. The number of the optimal fractional order is determined by particle swarm optimization algorithm. One example is provided to verify the high fitting accuracy of the new model, the results show that the mean absolute percentage error (MAPE) and the root mean square error (RMSE) of the optimal weighted combination Markov model are smaller than that of the quadratic function model (QFM), the classical combinatorial model and its three sub-models, which proves the robustness of the optimal weighted combination Markov model. This work will provide a scientific basis and technical reference for the further research in finance field.

Keywords: optimal weighted combination Markov model; public finance budget expenditure; particle swarm optimization algorithm; MAPE; RMSE
Mathematics Subject Classification: 62P05, 62P20, 62M05, 62M10, 91-10

1. Introduction

The public financial budget expenditure is of great significance for economic development. Budgets is an important activity for calculating and planning revenue and expenditure. The annual public finance budget expenditure plays a key role for economic planning and the development, and it enables the policymakers to carry out the economic decisions more efficient.

Dongguan is one of the central city in Guangdong-Hong Kong-Macao Greater Bay Area, and it locates at the east bank of the Pearl River Delta. At the end of 2020, the per capita GDP of the Dongguan has reached more than 100,000 CNY. At the same time, the growth rate of the expenditure in general public budget has appeared an increasing trend, achieved at 84 billion CNY, which is an increase of 66.7% compared to 1996. Hence, it is very necessary to forecast the fiscal expenditure and explore the relationship with history relevant data. This will provide a scientific guidance for the Dongguan government to make the reasonable economic decisions.

A great deal of effort has been dedicated to develop the relevant theoretical models, such as the time series forecasting model, the grey model, the regression forecasting model and so on. In 1997, Li [1] proposed the financial forecasting model and applied it to research fiscal revenue and expenditure. Chen et al. [2] established the autoregressive single integrated moving average model to predict the China's fiscal expenditure. Chen et al. [3] constructed the time series forecasting model to forecast the fiscal revenue in Inner Mongolia. Zhao et al. [4] applied a grey radial basis function (RBF) neural network in forecasting the fiscal revenue. Hansen [5] proposed a panel threshold model and investigated the nonlinear relationship between the inhabitant financial expenditure and the urban-rural income gap. The above mentioned works have not only made a series of scientific forecasts on the expenditure and revenue in budgets but also provided some corresponding theories in making financial decisions and measures. Unfortunately, the main disadvantage of the classical forecasting model lies in the low prediction accuracy.

The financial and economic system has distinct characteristics such as nonlinearity, correlation, systematization, randomness and chaos, which makes it difficult to describe the change rule of regional fiscal revenues and get better forecast values from a single model [6]. In order to achieve the integration of different types and levels of information and knowledge, it is necessary to integrate the various forecasting methods for mutual learning from each other [7]. Generally, it is a classic approach to combine several prediction models optimally. For example, Li et al. proposed an optimal weighted method to establish the combination forecasting model for grain yields in China [8]. Fang et al. [9] established the combined prediction model of fiscal revenue through the max-min closeness evaluation method. Chen et al. [10] conducted a housing price prediction analysis by establishing an optimal weighted combination model. Fisher et al. [11] revealed an inefficient forecasting process and that there exists substantial serial correlation in errors for forecasting budget revenues in the last two decades, and they improved the efficiency and the prediction accuracy of the tax budget by incorporating the national accounts data on household saving behaviors. By employing the out-of-sample forecasting framework, Rich et al. [12] found that, compared to those models generated from the univariate auto-regression, the coincident indexes can improve significantly tax base forecasts ether in statistic or in economic domains. Li [13] predicted financial time series data by combining the support vector machine model with the convolutional neural network model. In [14], Gan constructed a multiple regression model based on four explanatory variables, the fiscal expenditure, the GDP, the tax revenue and the fixed asset investment. The empirical results showed that the change of the fiscal revenue in Sichuan Province was mainly affected by the tax revenue and the fixed asset investment. In order to forecast the fiscal revenue, Sheng et al. [15] considered the combination of the grey prediction model and the BP neural network after dimension reduction by the Lasso, and they found that this model has a good effect on multiple inputs.

The optimal combination model performs well in the prediction efficiency and accuracy. However, because it is obtained from the weighted combination of the fractional grey prediction model, the Fourier regression model and the time series model, this model is still a random model, and the prediction accuracy is still limited. In order to solve this problem, we propose an optimal weighted combination Markov model to modify the optimal weighted combination model in this paper. Then, we apply this new model to forecast and explore the fiscal budget expenditure of Dongguan. The results show that the fitting accuracy of this model is more than that of the classical combination model itself and the three sub-models, which proves the efficiency of the new model. This work will provide a scientific forecasting method and theoretical basis for the local government to predict the local fiscal expenditure, the future value and the growth trend, which helps the local government to carry out economic management and make the corresponding economic strategies.

2. Optimal weighted combination Markov model

In this part, we construct the OWCMKM model by using Markov to modify the OWCM model constructed by the combination of the FGM model, the Fourier model and the autoregressive integrated moving average (ARIMA) model. We also compare the effect of the new model to that of the quadratic function model (QFM). Finally, we provide two statistic experiments to analyze the results of each model.

2.1. Fractional grey prediction model

The grey prediction technique is an important branch of the modern prediction theory system, and many researchers have contributed to this method. For example, Xu et al. [16] predicted the water demand in agriculture by adopting the fractional-order cumulative discrete grey model, and they put out that the prediction performance of this model is better than that of the GM (1,1) model. Considering different growth rates, Xie et al. [17] established the grey multivariate convolution (GMCN(1,N)) model of the electricity consumption for China's three major industries based on the analysis of the socio-economic factors. Halis [18] proposed a new exponential grey prediction model, which is called EXGM (1,1). By using this model, new cases, deaths and recovered cases of COVID-19 in Turkey were forecasted. Cai and Ma [19] constructed a novel ensemble learning-based grey model for forecasting the electricity supply in China. Ma et al. [20] developed a novel nonlinear multivariate forecasting grey model based on the Bernoulli equation (NGBMC (1, n)). By using the fractional accumulation generation, the forecast model can not only reduced the randomness and volatility of the original data but also reduced the disturbance boundary of the solution for the grey model [21]. Next, we will establish the fractional grey prediction model, the specific modeling steps are as follows:

First, the original sequence of fiscal expenditure is accumulated in order r, and the following equation can be calculated:

$$y^{(r)} = \left[y^{(r)}(1), y^{(r)}(2), y^{(r)}(3), \cdots, y^{(r)}(n)\right] \ ,0 < \mathrm{r} < 1 \,. \tag{1}$$

Among them,

$$y^{(r)}(k) = \sum_{i=1}^{k} C_{k-i+r-1}^{k-i}\, y^{(0)}(i), k = 1,2,3, \cdots, n \ , \tag{2}$$

and

$$C_{r-1}^{0} = 1, C_{k-1}^{k} = 0, C_{k-i+r-1}^{k-i} = \frac{(k-i+r-1)(k-i+r-1)\cdots(r+1)r}{(k-i)!}.$$

The adjacent mean sequence of $y^{(r)}$ is generated: namely,

$$z^{(r)} = \left[z^{(r)}(2), z^{(r)}(3), z^{(r)}(4), \cdots, z^{(r)}(n)\right], \tag{3}$$

where

$$z^{(r)}(k) = \frac{1}{2}\left(y^{(r)}(k) + y^{(r)}(k-1)\right),\ k = 2,3,4,\cdots,n. \tag{4}$$

Second, the fractional grey prediction differential equation of the fiscal expenditure is created as

$$\frac{dy^{(r)}}{dt} + ay^{(r)} = b. \tag{5}$$

Thus, the prediction equation of the fractional grey prediction model is obtained:

$$\hat{y}^{(r)}(k) = \left(x^{(0)}(1) - \frac{b}{a}\right)e^{-a(k-1)} + \frac{b}{a} \quad ,k = 1,2,3,\cdots,n. \tag{6}$$

The least square method is applied to solve the parameter estimates of a and b, and the following results can be figured out:

$$\begin{bmatrix}\hat{a}\\ \hat{b}\end{bmatrix} = (Q^TQ)^{-1}Q^TY, \tag{7}$$

where

$$Q = \begin{bmatrix} -z^{(r)}(2) & 1\\ -z^{(r)}(3) & 1\\ \vdots & \vdots\\ -z^{(r)}(n) & 1\end{bmatrix} = \begin{bmatrix} -\frac{1}{2}(y^{(r)}(2) + y^{(r)}(1)) & 1\\ -\frac{1}{2}(y^{(r)}(3) + y^{(r)}(2)) & 1\\ \vdots & \vdots\\ -\frac{1}{2}(y^{(r)}(n) + y^{(r)}(n-1)) & 1\end{bmatrix}, \tag{8}$$

$$Y = \begin{bmatrix} y^{(r)}(2) - y^{(r)}(1)\\ y^{(r)}(3) - y^{(r)}(2)\\ \vdots\\ y^{(r)}(n) - y^{(r)}(n-1)\end{bmatrix}. \tag{9}$$

Finally, r order reduction of $\hat{y}^{(r)}$ is computed:

$$\hat{y}^{(0)}(1) = \hat{y}^{(r)}(1), \tag{10}$$

$$\hat{y}^{(0)}(k) = \sum_{i=1}^{k} C_{k-i+1+r-1}^{k-i}\, y^{(r)}(k) - \sum_{i=1}^{k-1} C_{k-1-i+1+r-1}^{k-1-i}\, y^{(r)}(k-1) = \sum_{i=1}^{k} C_{k-i+r}^{k-i}\, y^{(r)}(k) - \sum_{i=1}^{k-1} C_{k-i+r-1}^{k-1-i}\, y^{(r)}(k-1), (k = 2,3,4,\cdots). \tag{11}$$

In order to optimize the new model and improve the prediction accuracy, we choose the universal particle swarm optimization (PSO) algorithm [22] to search for the optimal fractional order, and it is based on the principle of minimizing the error of the forecast values. Then, we obtain the following objective function:

$$f(x) = min\frac{1}{n}\sum_{k=1}^{n}\left|\frac{y^{(0)}(k) - \hat{y}^{(0)}(k)}{y^{(0)}(k)}\right|. \tag{12}$$

Because the accumulated fractional order r could weaken the randomness and the volatility of the original sequence data effectively, it can not only reflect the features of the new information priority but also improve the prediction accuracy of the model.

In order to seek the optimal fractional order r, the following optimization model is constructed:

$$\begin{cases} f(x) = min\frac{1}{n}\sum_{k=1}^{n}\left|\frac{y^{(0)}(k)-\hat{y}^{(0)}(k)}{y^{(0)}(k)}\right|. \\ 0 < r < 1 \end{cases} \tag{13}$$

The particle swarm optimization algorithm solves the fractional order grey prediction model optimal order.

2.2. Fourier regression model

After analyzing the characteristics of the data, we found that the growth rate of Dongguan public finance budget expenditure is relatively slow in the previous years. From 2000 to 2020, the spending in the general public budget of Dongguan appears to have nonlinear growth, which makes it difficult to fit these data efficiently. In order to find a suitable regression model, we will propose the Fourier regression model to fit the expenditure in general public finance budget of Dongguan from 1996 to 2015. Let years be the independent variables and the fiscal expenditures be the dependent variables and combine with the exponential function. We obtain the Fourier curve as follows [23]:

$$\hat{y} = a_0 + a_1 cos(xw) + a_2 sin(xw). \tag{14}$$

2.3. Autoregressive integrated moving average model

As is known, time series models regard the research object data in time series as a random series. This type model can predict and infer the future behaviors based on the past and present behavior of the data in time series. For the observed value series data, the ARIMA model can be established by transforming the non-stationary original data in time series into the stationary non-white noise time series with a $\boldsymbol{d}$-order difference process.

The sequence with the $\boldsymbol{d}$-order difference can be expressed as

$$\nabla^d x_t = \sum_{i=0}^{d}(-1)^i C_d^i x_{t-i}. \tag{15}$$

The structure of the ARIMA (p, d, q) model is [24]

$$\begin{cases} \Phi(B)\nabla^d x_t = \Theta(B)\varepsilon_t \\ E(\varepsilon_t) = 0, Var(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t\varepsilon_s) = 0, s \neq t \\ E(x_s\varepsilon_t) = 0, \forall\, s < t \end{cases}, \tag{16}$$

where B is the delay operator, $\nabla^d = (1-B)^d$, $\Phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$, $\Theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$.

ARIMA can also be defined as

$$\nabla^d x_t = \frac{\Theta(B)}{\Phi(B)}\varepsilon_t. \tag{17}$$

2.4. Combination model

Due to the different characteristics and the data processing of the single models, there exists a great deviation of the results from each other. In this part, in order to make the more reasonable and efficient prediction for the financial expenditures in Dongguan, we will establish a combined model according to the fundamentals of the above proposed single models. The error at a certain time is calculated by the optimal weighting method.

Let $\hat{y}_i(k)$ be the predicted value of the ith prediction model and $y_i(k)-\hat{y}_i(k)$ be the residual error at the time k. The error between each model is

$$e_{ik}=\frac{1}{n}\sum_k^n\left|\frac{y_i(k)-\hat{y}_i(k)}{y_i(k)}\right|. \tag{18}$$

Then, the prediction error of the ith forecast model with weight is

$$e_i=\sum_i\sum_k w_i e_{ik}. \tag{19}$$

In order to improve the accuracy and minimize the error for forecasting the financial expenditure of this model, we recall the following optimization model [25]:

$$\begin{cases} minQ=\sum_i\sum_k w_i e_{ik} \\ s.t.\sum_i w_i=1 \\ w_i\geq 0 \end{cases}, \tag{20}$$

where

$$w_i=\left(e_i^2\sum_i\frac{1}{e_i^2}\right)^{-1}, \tag{21}$$

and it can be computed by using the least squares method.

2.5. Markov model

A Markov model is a general tool for statistical analysis of data. In this system, the state at a certain moment predicts the latest state according to the transition probability of the state at the previous moment. A Markov process is suitable for processing fluctuating data and has the characteristics of no aftereffect and good short-term prediction effect, which has been widely used in customer assets, intelligent health, remote sensing evaluation and other fields [26–32].

2.5.1. Status division

We use $\mathrm{E}_1, E_2,\cdots, E_m$ to represent the data sequence divided into several different states by the Markov chain. The state transition only occurs at countable moments denoted by $\mathrm{t}_1, t_2,\cdots, t_m$.

$$\mathrm{E}_i=[Q_{i1},Q_{i2}],\ (i=1,2,\cdots,j), \tag{22}$$

where j is the number of divided states, and Q_{i1} and Q_{i2} represent the lower and the upper limits for the relative error of the state intervals, respectively.

2.5.2. State transition probability matrix

The transition probability of the Markov chain from state E_i to state E_j after k steps is denoted by $p_{ij}(k)$, and we have

$$p_{ij}(k) = \frac{m_{ij}(k)}{M_i} \ . \tag{23}$$

M_i represents the total number of occurrences for the state E_i, $m_{ij}(k)$ denotes the number of the state E_i transferred to the state E_j after k steps, and m is the number of divided states. The one-step state transition probability matrix is displayed as follows:

$$P(1) = \begin{bmatrix} p_{11}(1) & \cdots & p_{1m}(1) \\ \vdots & \vdots & \vdots \\ p_{m1}(1) & \cdots & p_{mm}(1) \end{bmatrix}. \tag{24}$$

By using the Chapman-Kolmogorov equation repeatedly, we assume that the initial vector of the variables for the initial state E_i is $V(0)$, and then the k-step transition probability matrix and the state vector are obtained as follows:

$$P(k) = (P(1)^k), \tag{25}$$

$$V(k) = V(0) \cdot (P(1)^k). \tag{26}$$

2.5.3. The determination of the predicted value

Choose the *j*-group data which are closest to the prediction data, and denote the number of steps from near to far by *t* $(1,2,\cdots,j)$. Then, we construct a new matrix by selecting the row vectors of the t-step corresponding state transition matrix to each data, the most probable state of the predicted value is determined by the sum of the column vectors of the new matrix. Once the state is determined, the state interval is determined, and the Markov modification value is equal to the midpoint of the interval, that is, $\frac{1}{2}(Q_{i1} + Q_{i2})$. Finally, the Markov predicted value is obtained,

$$\hat{y}_{OWCMM}(k) = \frac{\hat{y}_{OWCM}(k)}{1+\frac{1}{2}(Q_{i1}+Q_{i2})}. \tag{27}$$

2.6. Quadratic function model

According to the features of the data in general public budget expenditure, we chose the QFM as a comparison model to the OWCMM. The QFM is expressed as

$$y_4(\mathrm{t}) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \varepsilon_t \ , \tag{28}$$

where $y_4(\mathrm{t})$ is the public finance budget expenditure, $\alpha_0, \alpha_1, \alpha_2$ are the regression parameters, t represents the time, and ε_t is the random error term and satisfies $\varepsilon_t \sim N(0, \sigma^2)$. Assuming that $\hat{y}_4(\mathrm{t})$ is an estimation of the $y_4(\mathrm{t})$, we can obtain the least square estimate of the parameters $\alpha_0, \alpha_1, \alpha_2$ under the condition $\sum_{t=1}^{n}(\hat{y}_4(t) - y_4\,(t))^2 = min$ by treating t^2 as a variable.

2.7. Model error test

In this article, the mean absolute percentage error (MAPE) and the root mean square error (RMSE) are offered to validate the prediction accuracy of the proposed new model. The specific formulations are represented as follows, respectively.

The mean absolute percentage error (MAPE)

$$\mathrm{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{\hat{y}_t - y_t}{y_t}\right|. \tag{29}$$

The root mean square error (RMSE)

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(\hat{y}_t - y_t)^2}\ . \tag{30}$$

3. Application of the model

In this part, we will apply the above-mentioned models to forecast and analyze the general public finance budget expenditure of Dongguan from 2016 to 2020. We chose the data from 1996 to 2015 as the sample data, which comes from the Dongguan Statistical Yearbook.

3.1. FGM model prediction

The optimal order $r = 0.00134$ was obtained by particle swarm optimization (PSO) algorithm using R software. Substitute the order number r and the fiscal expenditure series into the above Eqs (1), (3) and (6), and the parameter estimates of the fractional grey prediction model on the fiscal expenditure can be computed:

$$a = -0.1043,\ b = 80596.6079.$$

Thus, it can be known that the prediction equation of the fractional grey prediction model is

$$\hat{y}^{(0.00134)}(k) = 896797.3310e^{0.1043(k-1)} - 772738.3310\,, k = 1,2,3,\dots,n.$$

Finally, according to Eqs (10) and (11), the cumulative reduction is carried out to obtain the predicted value.

Now, we use the FGM model to predict the budget expenditure in general public finance of Dongguan from 2016–2020, and the results are listed in Table 1. As can be seen from Table 1, we found that there is still a larger error between the prediction value and the real value in 2020, which may be related to the factors such as the outbreak of the epidemic in the whole world in recent years.

Table 1. Comparison of results of various models (ten thousand CNY).

		FGM		Fourier		ARIMA		OWCM	
Year	Raw	Predicted value	Relative error	Predicted value	Relative error	Predicted value	Relative error	Predicted value	Relative error
1996	124059	124059	0	167649	35.14%	124004	−0.04%	145291	17.11%
1997	148510	197046	32.68%	158058	6.43%	148622	0.08%	156520	5.39%
1998	179064	280533	56.67%	182369	1.85%	175847	−1.80%	186155	3.96%
1999	212641	376013	76.83%	240573	13.14%	205118	−3.54%	234042	10.06%
2000	336102	485208	44.36%	332641	−1.03%	242155	−27.95%	302829	−9.89%
2001	478646	610088	27.46%	458535	−4.20%	355075	−25.82%	422889	−11.64%
2002	649606	752912	1.59%	618197	−4.84%	552161	−15%	598031	−7.93%
2003	765190	916258	19.74%	811556	6.06%	756648	−1.12%	794291	3.80%
2004	941554	1103077	17.15%	1038527	10.30%	915650	−2.75%	988331	4.96%
2005	1170427	1316745	1.25%	1299008	10.99%	1063687	−9.12%	1195670	2.15%
2006	1478955	1561121	5.56%	1592884	7.70%	1314301	−11.13%	1466956	−0.81%
2007	1930968	1840621	−4.68%	1920023	−0.57%	1657306	−14.17%	1797900	−6.89%
2008	2182626	2160293	−1.02%	2280280	4.47%	2158860	−1.09%	2218168	1.62%
2009	2326216	2525913	8.58%	2673495	14.93%	2563693	10.21%	2614666	12.39%
2010	2898306	2944087	1.58%	3099492	6.94%	2652789	−8.47%	2890456	−0.27%
2011	3519171	3422369	−2.75%	3558082	1.11%	3059623	−13.06%	3327393	−5.44%
2012	3855844	3969401	2.95%	4049061	5.01%	3898111	1.10%	3976575	3.13%
2013	4446589	4595064	3.34%	4572209	2.83%	4410433	−0. 81%	4501893	1.24%
2014	4576816	5310663	16.03%	5127295	12.03%	4843252	5.82%	5013588	9.54%
2015	5812410	6129126	5.45%	5714071	−1.69%	5149673	−11.40%	5491586	−5.51%
		RMSE	212180	RMSE	170780	RMSE	222430	RMSE	152000
		MAPE	17.76%	MAPE	7.56%	MAPE	8.22%	MAPE	6.19%
2016	5992899	7065240	17.89%	6332276	5.66%	5975072	−0.30%	6223484	3.85%
2017	6676462	8135918	21.86%	6981635	4.57%	6888291	3.17%	7018752	5.13%
2018	7654053	9360503	22.29%	7661858	0.10%	7276467	−4.93%	7606290	−0.62%
2019	8630134	10761120	24.69%	8372643	−2.98%	8031930	−6.93%	8383890	−2.85%
2020	8403253	12363074	47.12%	9113673	8.45%	8530462	1.51%	9075800	8.00%
		RMSE	2298400	RMSE	394810	RMSE	335200	RMSE	370290
		MAPE	26.77%	MAPE	4.35%	MAPE	3.37%	MAPE	4.09%

3.2. Analysis results of Fourier model

MATLAB was used to achieve the regression fitting. The parameter estimate of the Fourier model is

$$a_0 = 76310000,\ a_1 = 21000000,\ a_2 = 73200000,\ w = 0.0211.$$

By the Fourier model, the prediction equation is obtained:

$$\hat{y} = 76310000 + 21000000cos(0.0211x) + 73200000sin(0.0211x).$$

In order to test the advantages and disadvantages of the two models for Fourier curve fitting regression and exponential function fitting regression, the goodness-of-fit tests were carried out, and the average MAPE values of the forecast error for the fiscal expenditure forecasting model were calculated.

Now, we adopt the goodness-of-fit tests and calculate the MAPE of the errors to validate the advantages and the disadvantages of the Fourier curve fitting regression model and the exponential function fitting regression model. The results are listed as follows.

Table 2. Comparison results of regression prediction models.

Test indicators	Exponential function	Fourier curve
R^2	0.9782	0.9954
$R^2_{Adjusted}$	0.9773	0.9948
SSE	4064000000000	852700000000
$RMSE$	420400	201500
$MAPE$	20.92%	6.92%

Based on the results in Table 2, we can see that the determination coefficient of the Fourier curve fitting model is larger than that of the exponential function model, while the mean square error (RMSE) of the former is smaller than that of the latter. This indicates that the fitting effect of the Fourier curve fitting model is better than that of the exponential function model. As can be seen from the results shown in Tables 1 and 2, the Fourier curve fitting model can describe the change rule of the historical data well. However, it is not suitable for a long-term prediction since this model omitted the impact of other factors such as market rule on the fiscal data.

3.3. Application of ARIMA model

According to the fiscal expenditure budget system and the actual situation of the regional fiscal expenditure, we can regard the general public finance budget expenditure data of Dongguan as the random time series. Due to the volatile exponential growth trend of the fiscal expenditure data, we can judge this time series to be non-stationary data. In fact, as is shown in Table 3, the p value of the ADF test is larger than 0.05, which indicates the existence of the unit root, and this proves that the fiscal expenditure series is a non-stationary series.

Table 3. Unit root test results of ARIMA.

ADF unit root test	P
Before the difference	0.99
2 order difference	0.01

In order to get more accurate prediction results, we transform this fiscal expenditure series into a stationary series by using the differential transformation. Compared with the original fiscal expenditure series, it has no obvious features of exponential growth after the second order difference. Moreover, it is still a stable series since the p value of the ADF test is 0.01. Next, we use the ARIMA model to forecast this fiscal expenditure series.

Now, we construct the LB statistic to validate the pure randomness of the above fiscal expenditure series, and the results are listed in Table 4.

Table 4. Test results of pure randomness of fiscal expenditure series after difference.

LB	P	DF
7.7254	0.005445	1

According to the results in Tables 3 and 4, it can be seen that the fiscal expenditure series after the second-order difference satisfies the establishing conditions of the ARIMA model. Next, we present four ARIMA models, ARIMA(0,2,0), ARIMA(0,2,1), ARIMA(1,2,1) and ARIMA(2,2,1), based on the above time series. The AIC values of these models were calculated to select the most optimal ARIMA model, and all the results are listed in Table 5.

Table 5. AIC values of each ARIMA model.

$ARIMA$	AIC
$ARIMA$(0,2,0)	509.8608
$ARIMA$(0,2,1)	505.8489
$ARIMA$(1,2,1)	505.3887
$ARIMA$(2,2,1)	508.7512

According to the AIC criterion, we obtain that the most optimal time series model is ARIMA(1,2,1). The second order difference equation of the ARIMA(1,2,1) model can be calculated by the conditional least squares estimation as follows:

$$(1-B)^2 x_t = \frac{1+0.4522B}{1+0.6995B}\varepsilon_t, \tag{31}$$

that is,

$$x_t = 1.3005x_{t-1} + 0.399x_{t-2} - 0.6995x_{t-3} + \varepsilon_t + 0.4522\varepsilon_{t-1}. \tag{32}$$

In order to carry out the White noise test, we need to predict the residual of the ARIMA model by constructing the statistics once the p, d, q are determined. The results are shown in Table 6.

Table 6. Residual white noise test of ARIMA (1,2,1).

Statistic	P	DF
0.96406	0.3262	1

From the P=0.3262>0.05 in Table 6, we know that this model passes the residual white noise test. Hence, the residual of the time series forecasting model based on the fiscal expenditures is a white noise series.

According to the above ARIMA model, the general public finance budget expenditure of

Dongguan from 2016 to 2020 is predicted, and the prediction results are shown in Table 1. There is no significant deviation between the predicted expenditure and the actual expenditure, which indicates that the established ARIMA (1,2,1) model has a small error in forecasting the general public finance budget expenditure in Dongguan. This illustrates that the ARIMA model has a high accuracy in forecasting the financial expenditure data. By combining MAPE and RMSE, it can be seen that the forecasting accuracy of the ARIMA model is more improved compared to the two former forecasting models.

3.4. Analysis of the optimal weighted combination model

In this part, we establish the optimal weighted combination model by combining the fractional grey prediction model, the Fourier regression model and the ARIMA model. We obtain the optimal weighting coefficient by using the following equation:

$$\begin{cases} minQ = \sum_{i=1}^{3}\sum_{k=1}^{20} w_i e_{ik} \\ s.t. \sum_{i=1}^{3} w_i = 1 \\ w_i \geq 0 \end{cases} . \tag{33}$$

The weights are obtained by the least squares method,

$$w_i = (0.06808, 0.48765, 0.44427).$$

The optimal weighted combination model is

$$\hat{y}(k) = 0.06808\hat{y}_{1k} + 0.48765\hat{y}_{2k} + 0.44427\hat{y}_{3k}, \tag{34}$$

where $\hat{y}_{1k}, \hat{y}_{2k}$ and $\hat{y}_{3k}$ are the prediction values of the fractional grey model, the Fourier regression model and the ARIMA model at the time k, respectively. The prediction results of the fiscal budget expenditures with this combination model are listed in Table 1.

3.5. Markov model modification

3.5.1. Interval division and establishment of state transition matrix

According to the relative error of the optimal weighted combination model (OWCM), the state interval is divided. Table 1 shows that the minimum relative error of the first 20 fitted data of this model is −11.64%, and the maximum is 17.11%. Therefore, five state intervals are divided according to the equal spacing rule and listed as follows:

$$E_1(-11.64\%, -5.89\%], E_2(-5.89\%, -0.14\%], E_3(-0.14\%, 5.61\%],$$

$$E_4(5.61\%, 11.36\%], E_5(11.36\%, 17.11\%].$$

According to the state of each data and the probability of transferring from the current state to the next state, we obtain the following transition probability matrixes of the 1, 2, 3, 4 and 5 steps state as well as the final state vector.

$$P(1)=\begin{bmatrix}0.66&0&0.34&0&0\\0.33&0.33&0.34&0&0\\0&0&0.57&0.29&0.14\\0.5&0.5&0&0&0\\0&0.5&0.5&0&0\end{bmatrix},\ P(2)=\begin{bmatrix}0.4356&0&0.4182&0.0986&0.0476\\0.3267&0.1089&0.4182&0.0986&0.0476\\0.1450&0.2150&0.3949&0.1653&0.0798\\0.4950&0.1650&0.3400&0&0\\0.1650&0.1650&0.4550&0.1450&0.0700\end{bmatrix},$$

$$P(3)=\begin{bmatrix}0.3368&0.0731&0.4103&0.1213&0.0585\\0.3009&0.1090&0.4103&0.1213&0.0585\\0.2493&0.1935&0.3874&0.1145&0.0553\\0.3811&0.0545&0.4182&0.0986&0.0476\\0.2359&0.1620&0.4065&0.1319&0.0637\end{bmatrix},\ P(4)=\begin{bmatrix}0.3070&0.1140&0.4025&0.1190&0.0547\\0.2952&0.1259&0.4025&0.1190&0.0547\\0.2857&0.1488&0.3990&0.1123&0.0542\\0.3188&0.0911&0.4103&0.1213&0.0585\\0.2751&0.1513&0.3988&0.1179&0.0569\end{bmatrix},$$

$$P(5)=\begin{bmatrix}0.2998&0.1258&0.4013&0.1167&0.0563\\0.2959&0.1298&0.4013&0.1167&0.0563\\0.2938&0.1324&0.4023&0.1157&0.0559\\0.3011&0.1200&0.4025&0.1190&0.0574\\0.2904&0.1373&0.4008&0.1157&0.0558\end{bmatrix},$$

$$V(k)=(0.2965,0.1289,0.4018,0.1165,0.0563)(k\to\infty).$$

3.5.2. Forecast the public finance budget expenditure of Dongguan

A new state transition matrix is constructed by using the recent data sets. The statuses of 2016 are listed in Table 7, and the results are shown in Table 8.

According to Table 7, the most likely state of Dongguan's public financial budget expenditure in 2016 is E_3, because E_3 has the largest value in the total. The predicted value of the optimal weighted combination model in 2016 is 6223484 million CNY. According to Eq (27), the predicted value of the OWCMM model is 6057803 million CNY. Using the same method, the predicted value of the Markov model from 2017 to 2020 can be gained. The specific results are shown in Table 8. The estimated and predicted values of each sub-model, optimal weighted model and Markov model are displayed in Figure 1.

Table 7. Forecast the status of 2016.

Year	Initial state	Transfer steps	p_{ij}	E_1	E_2	E_3	E_4	E_5
2015	2	1	p_{12}	0.33	0.33	0.34	0	0
2014	4	2	p_{24}	0.4950	0.1650	0.34	0	0
2013	3	3	p_{33}	0.2493	0.1935	0.3874	0.1145	0.0553
2012	3	4	p_{43}	0.2857	0.1488	0.3990	0.1123	0.0542
2011	2	5	p_{52}	0.2959	0.1298	0.4013	0.1167	0.0563
Total				1.6559	0.9671	1.8677	0.3435	0.6855

Table 8. Comparison of OWCM, QFM and OWCMM model results (ten thousand CNY).

Year	Raw	OWCM			QFM		OWCMM	
		Predicted value	Relative error	Station value	Predicted value	Relative error	Predicted value	Relative error
1996	124059	145291	17.11%	5	179834	44.96%	127185	2.52%
1997	148510	156520	5.39%	3	161733	8.90%	152353	2.59%
1998	179064	186155	3.96%	3	176856	−1.23%	181198	1.19%
1999	212641	234042	10.06%	4	225203	5.91%	215736	1.46%
2000	336102	302829	−9.89%	1	306774	−8.73%	331920	−1.24%
2001	478646	422889	−11.64%	1	421569	−11.92%	463516	−3.16%
2002	649606	598031	−7.93%	1	569588	−12.32%	655483	0. 9%
2003	765190	794291	3.80%	3	750831	−1.88%	773144	1.04%
2004	941554	988331	4.96%	3	965298	2.52%	962019	2.17%
2005	1170427	1195670	2.15%	3	1212989	3.64%	1163839	−0.56%
2006	1478955	1466956	−0.81%	2	1493904	1.01%	1512559	2.27%
2007	1930968	1797900	−6.89%	1	1808043	−6.37%	1970625	2.05%
2008	2182626	2218168	1.62%	3	2155406	−1.25%	2159115	−1.08%
2009	2326216	2614666	12.39%	5	2535993	9.02%	2288847	−1.61%
2010	2898306	2890456	−0.27%	2	2949804	1.78%	2980312	2.83%
2011	3519171	3327393	−5.44%	2	3396839	−3.48%	3432832	−2.45%
2012	3855844	3976575	3.13%	3	3877098	0.55%	3870711	0.39%
2013	4446589	4501893	1.24%	3	4390581	−1.26%	4382044	−1.45%
2014	4576816	5013588	9.54%	4	4937288	7.88%	4621457	0.98%
2015	5812410	5491586	−5.51%	2	5517219	-5.08%	5662304	−2.58%
		RMSE	152000		RMSE	125480	RMSE	49289
		MAPE	6.19%		MAPE	6.98%	MAPE	1.73%
2016	5992899	6223484	3.85%	3	6130374	2.29%	6057803	1.08%
2017	6676462	7018752	5.13%	3	6776753	1.50%	6831899	2.33%
2018	7654053	7606290	−0.62%	2	7456356	−2.58%	7403796	−3.27%
2019	8630134	8383890	−2.85%	2	8169183	−5.34%	8160694	−5.44%
2020	8403253	9075800	8%	4	8915234	6.09%	8834185	5.13%
		RMSE	370290		RMSE	329440	RMSE	315300
		MAPE	4.09%		MAPE	3.56%	MAPE	3.45%

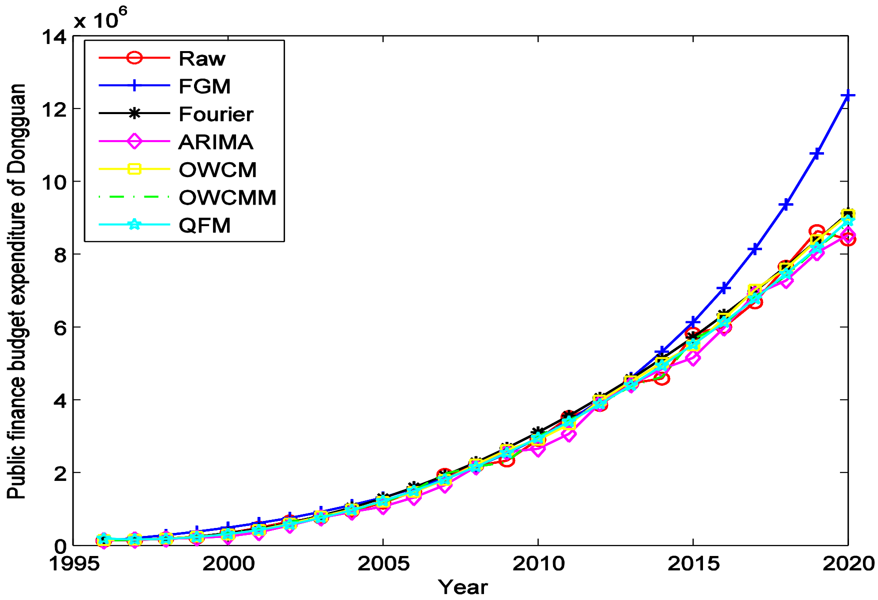

Figure 1. The forecast results of different models for public finance budget expenditure of Dongguan.

3.6. Empirical analysis of QFM model

By using the data from 1996 to 2015 and Eq (28), we obtain the estimated model,

$$\hat{y}_4(\mathrm{t}) = 231159 - 67937t + 16612t^2. \tag{35}$$

The goodness of fit is $R^2 = 0.9945$, and we can calculate estimates (t =1,2,⋯,20) and predictions (t =21,22,⋯,25). See Table 8 for specific values.

3.7. Comparative analysis of the results of each model

As can be seen from Table 1, the results predicted by the OWCM model are closer to the real value, and the relative error is smaller than that of other models. We can also get a smaller value of the RMSE and the MAPE when we forecast the test data and estimate the training data by using the OWCM model.

In Table 8, the results predicted by OWCMM model are closer to the real value, and the relative error is smaller than that of OWCM. By using the OWCMM model to forecast the data (2016–2020) and estimate the data (1996–2015), we also can obtain a smaller value of the RMSE as well as the MAPE. No matter in the terms of the estimation or the prediction, we can see that the MAPE and RMSE values of the OWCMM model are still smaller than those of the QFM model.

The results in Figure 1 show that the curve of the OWCMM model is closer to the true values than that of the OWCM model.

4. Conclusions

In this paper, we established the new optimal weighted combination Markov model by modifying an optimal weighted combination model using the Markov chain. The optimal weighted combination model is constructed by the combination of the fractional grey model, the Fourier function regression model and the autoregressive integrated moving average model. We also constructed the QFM model as a comparison of our proposed model. We applied the new model to predict the fiscal expenditure of Dongguan from 1996 to 2015. The results show that the prediction accuracy is greatly improved, and the error was reduced effectively compared to the three single common prediction models. The RMSE value and the MAPE value of the OWCMM model are also smaller than those of the QFM model, the single model and their combination model. This illustrated that the optimal weighted combination Markov model is more reliable and robustness compared to the classical models. This study has made a significant contribution in predicting the fiscal expenditure of Dongguan.

Acknowledgments

This research was funded by the NSF of Huizhou University (Grant No. hzu201806), the Project of Guangdong Provincial Department of Education (Grant No. 2021ZDJS080). This work was also supported by the Ministry of Education “Blue Fire Program” (Huizhou) Industry-University-Research Joint Innovation Fund 2018 Annual Project: High-power Intelligent Dimming Control System (CXZJHZ201812), Huizhou University School Level Undergraduate Teaching Quality Engineering Project (X-JYJG2021045), Higher Education Teaching Reform Project of Huizhou University ([2022]163), Huizhou Philosophy and Social Sciences Discipline Co-Construction Project (2022ZX046), and Horizontal Project of Huizhou University (2022HX057, 2022HX058) as well as the “One hundred excellent young teachers training project” of Huizhou University.

Conflicts of interest

The authors declare that they have no conflicts of interest.

References

1. H. X. Li, Y. W. Yi, Research and application of financial forecasting model, *Inform. Cont.*, **26** (1997), 56–61.
2. Y. Chen, W. Zhao, X. M. Yan, Application of autoregressive integrated moving average model in fiscal expenditure forecast, (Chinese), *Economic Research Reference*, **33** (2014), 53–62. https://doi.org/10.16110/j.cnki.issn2095-3151.2014.33.020
3. S. J. Chen, L. G. Zhou, M. D. Tan, B. Liang, Fiscal revenue forecast of inner mongolia during the 13th five-year plan period—Based on comparative analysis of coal price and Keqiang Index, *Sub. Nat. Fisc. Res.*, **6** (2016), 58–61.
4. H. H. Zhao, Multi-factor fiscal revenue forecasting model based on Grey RBF neural network, (Chinese), *Statistics & Decision*, **13** (2016), 79–81. https://doi.org/10.13546/j.cnki.tjyjc.2016.13.021
5. B. E. Hansen, Threshold effects in non-dynamic panels: estimation, testing and inference, *J. Econometrics*, **93** (1999), 345–368. https://doi.org/10.1016/S0304-4076(99)00025-1
6. C. H. Guo, H. W. Tang, Research on macroeconomic forecasting model system, *Operations Research and Management Science*, **10** (2001), 4.

7. M. Fan, W. R. Shi, Y. L. Liang, H. Y. Hua, Application of a combination forecasting model in local financial revenue forecasting, *Journal of Chongqing University*, **5** (2008), 536–540.
8. B. R. Li, Q. Wu, Y. Liu, Application of optimal weighted group method in grain yield forecasting in China, (Chinese), *Statistics & Decision*, **19** (2010), 34–38. https://doi.org/10.13546/j.cnki.tjyjc.2010.19.032
9. B. Fang, L. He, Fiscal revence prediction about the ARMA-BP neural network combination model, (Chinese), *Journal Of Mathematics*, **35** (2015), 709–713. https://doi.org/10.13548/j.sxzz.20140511010
10. J. T. Chen, J. K. Zuo, C. Chen, Application of combined forecasting model based on optimal weighting method in forecast of housing price in Haikou city, (Chinese), *Statistics and Application*, **7** (2018), 569–579. https://doi.org/10.12677/SA.2018.76066
11. L. A. Fisher, G. Kingston, Improved forecasts of tax revenue via the permanent income hypothesis, *The Australian Economic Review*, **50** (2017), 21–31. http://doi.org/10.1111/1467-8462.12198
12. R. Rich, J. Bram, A. Haughwout, J. Orr, R. R. Sela, Using regional economic indexes to forecast tax bases: evidence from New York, *Rev. Econ. Stat.*, **87** (2005), 627–634. https://doi.org/10.2307/40042881
13. X. H. Li, Application of neural networks in financial time series forecasting models, *J. Funct. Space.*, **2022** (2022), 7817264. https://doi.org/10.1155/2022/7817264
14. M. L. Gan, Empirical analysis on influencing factors of financial revenue in sichuan province, *Journal of Economics and Public Finance*, **8** (2022), 68–74. https://doi.org/10.22158/JEPF.V8N2P68
15. Y. F. Sheng, J. J. Zhang, W. W. Tan, J. Wu, H. J. Lin, G. Sun, et al., Application of grey model and neural network in financial revenue forecast, *Comput. Mater. Con.*, **69** (2021), 4043–4059. https://doi.org/10.32604/CMC.2021.019900
16. Y. H. Xu, H. D. Wang, N. L. Hui, Prediction of agricultural water consumption in 2 regions of China based on fractional-order cumulative discrete grey model, *J. Math.*, **2021** (2021), 3023385. https://doi.org/10.1155/2021/3023385
17. Y. H. Xie, Y. F. Yang, L. F. Wu, Power consumption forecast of three major industries in China based on fractional grey model, *Axioms*, **11** (2022), 407–407. https://doi.org/10.3390/AXIOMS11080407
18. H. Bilgil, New grey forecasting model with its application and computer code, *AIMS Mathematics*, **6** (2021), 1497–1514. https://doi.org/10.3934/math.2021091
19. Y. B. Cai, X. Ma, A novel ensemble learning-based grey model for electricity supply forecasting in China, *AIMS Mathematics*, **6** (2021), 12339–12358. https://doi.org/10.3934/math.2021714
20. X. Ma, Z. B. Liu, Y. Wang, Application of a novel nonlinear multivariate grey Bernoulli model to predict the tourist income of China, *J. Comput. Appl. Math.*, **347** (2019), 84–94. https://doi.org/10.1016/j.cam.2018.07.044
21. L. F. Wu, S. F. Liu, L. G. Yao, Discrete grey model based on fractional order accumulate, *Systems Engineering-Theory & Practice*, **34** (2014), 1822–1827.
22. J. Kennedy, R. Eberhart, Particle swarm optimization, In: *Proceedings of ICNN'95—International Conference on Neural Networks*, Perth, WA, Australia, 1995, 1942–1948. https://doi.org/10.1109/ICNN.1995.488968
23. X. Q. Li, P. Y. Liu, G. F. Yin, H. H. Jiang, Weld defect detection by X-ray images method based on Fourier fitting surface, *Trans. Chin. Weld. Inst.*, **35** (2014), 61–64.

24. D. H. Yi, Y. Wang, *Applied time series analysis*, *(Chinese)*, 5 Eds., Beijing: China Renmin University Press, 2019.
25. D. Wang, Q. Cai, Application of optimum weighted combination method in electric power short-term load forecasting, (Chinese), *Computer and Modernization*, **203** (2012), 188–191. https://doi.org/10.3969/j.issn.1006-2475.2012.07.052
26. J. Zhao, F. J. Zhang, S. S. Chen, C. Y. Wang, J. Y. Chen, H. Zhou, et al., Remote sensing evaluation of total suspended solids dynamic with Markov model: a case study of inland reservoir across administrative boundary in South China, *Sensors*, **20** (2020), 6911. https://doi.org/10.3390/s20236911
27. A. Nadeem, A. Jalal, K. Kim, Accurate physical activity recognition using multidimensional features and Markov model for smart health fitness, *Symmetry*, **12** (2020), 1766. https://doi.org/10.3390/sym12111766
28. J. Y. Cai, X. Wang, Y. P. Cai, The prediction of water demand in Beijing City based on unbiased grey-Markov chain model, (Chinese), *Ecology and Environmental Monitoring of Three Gorges*, **7** (2022), 85–96. https://doi.org/10.19478/j.cnki.2096-2347.2022.03.11
29. E. Kappe, A. S. Blank, W. S. Desarbo, A random coefficients mixture hidden Markov model for marketing research, *Int. J. Res. Mark.*, **35** (2018), 415–431. https://doi.org/10.1016/j.ijresmar.2018.07.002
30. N. B. Zhao, J. L. Yang, S. Y. Li, Y. W. Sun, A GM (1,1) Markov chain-based aeroengine performance degradation forecast approach using exhaust gas temperature, *Math. Probl. Eng.*, **2014** (2014), 832851. http://doi.org/10.1155/2014/832851
31. K. Matsuoka, A framework for variance analysis of customer equity based on a Markov chain model, *J. Bus. Res.*, **129** (2021), 57–69. http://doi.org/10.1016/j.jbusres.2021.02.039
32. W. C. Fan, Y. Jiang, S. Y. Huang, W. G. Liu, Research and prediction of opioid crisis based on BP neural network and Markov chain, *AIMS Mathematics*, **4** (2019), 1357–1368. https://doi.org/10.3934/math.2019.5.1357