

A New Background Value Improvement of Fractional Order

Accumulated FAGM(1,1) Model and Its Application

Abstract: The prediction accuracy of the fractional FAGM(1,1) model mainly depends on the calculation of the background value. To improve the prediction accuracy of the FAGM(1,1) model, a new background value calculation method is proposed. By analyzing the cause of the background value error, considering the regularity of the fractional-order accumulation sequence with non-homogeneous exponential growth, the non-homogeneous exponential curve is used to fit the fractional-order accumulation sequence, combined with the integral theory, to accumulate the actual sequence on the interval. The result of the integration is used as a new background value. The example shows that using the new background value calculation method combined with the Genetic Algorithm to find the optimal order, the fitting and prediction accuracy of the fractional FAGM(1,1) model is obviously improved, And the background value has the characteristics of simple calculation and strong practicability.

Key words: National Tax Revenue; Background value optimization; Genetic Algorithm; Prediction Accuracy

1.Introduction

Tax revenue refers to the income that the state imposes on the taxpayer according to its political power. Since the reform of the tax-sharing system in 1994, the tax revenue has grown at an average rate of 17.13% per year, and has always dominated fiscal revenue [1]. As the most important source of income for public finance, taxation raises a large amount of funds for the country, implements national economic policies, and promotes economic development [2]. On the other hand, taxation is also an important indicator of a country's financial capacity and the ability of the government to function in schools in economic and social activities. Under the new situation, the state and society pay more and more attention to tax revenue, and the problem of fiscal revenue forecast has become an urgent problem in the economic field [3]. Therefore, scientific and accurate prediction of China's tax revenue is the basis for budgeting and decision-making, and will provide an important theoretical basis for the macro-control of the market economy and the economic development of the country [4].

Since the introduction of the Grey System Theory by the famous scholar Deng[5] in 1982, after more than 30 years of development, the grey prediction model has been widely used in economic, agricultural, military, ecological and other fields[6-10]. The grey prediction model shows good performance for solving prediction problems such as “small sample” and “poor information”, but its modeling mechanism still has certain problems. Relevant theory and applied research have confirmed that the accuracy of the GM(1,1) model largely depends on the background value construction. In 2000, Tan[11-13] first proposed the idea of background

value transformation and improved it. This method not only maintains the advantages of GM(1,1) modeling, but also expands GM(1,1). The application range of the model was subsequently studied by a large number of scholars[14-16]. Many scholars used different integral formulas to modify the background value of the GM(1,1) model. Such as Gauss formula, Simpson formula and Newton-Cotes formula[17-21]. Later, Luo [22], Wang[23] and others used the solution of the homogeneous and non-homogeneous exponential forms to transform the background value of the GM(1,1) whitening differential equation, which was widely recognized by the academic circles. However, the classic grey theory is not a perfect theory. Small disturbances in the data will cause disturbances in the solution. To this end, Wu[24-27] extended the integer order accumulation to the fractional order accumulation for the first time, and established a gray prediction model based on fractional order accumulation, which made the disturbance bound of the grey prediction model smaller.

Based on the above research, the author analyzes its shortcomings based on the classical fractional order FAGM(1,1), and proposes a structural background value optimization method that accurately approximates the essence of background values. Using the China tax revenue data from 2000 to 2006 as the model fitting data, the data from 2007 to 2011 is used as the test data to test the prediction accuracy of the model. And compared with the classic GM(1,1), Simpson formula improved GM(1,1), classic FAGM(1,1). The results show that the background value construction method proposed in this paper is superior in prediction accuracy and applicability.

2. Fractional Order Model

2.1 Fractional order accumulation operator and accumulating operator

Set the original sequence $X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))^T$, the r -th order accumulation generation operator (r -AGO) can be defined as follows.

Definition 2.1.1 Let the sequence $X^{(r)}$ ($r > 0$) be the cumulative generator of (r -AGO), where:

$$x^{(r)}(k) = \sum_{i=1}^k x^{(r-1)}(i), k = 1, 2, \dots, n. \quad (1)$$

The expression $X^{(r)} = A^r X^{(0)}$ can be obtained from the matrix operation theory, where A^r represents the r -AGO matrix, and A^r can be expressed as follows:

$$A^r = \begin{bmatrix} \begin{bmatrix} r \\ 0 \end{bmatrix} & 0 & 0 & \dots & L & 0 \\ \begin{bmatrix} r \\ 1 \end{bmatrix} & \begin{bmatrix} r \\ 0 \end{bmatrix} & 0 & \dots & L & 0 \\ \begin{bmatrix} r \\ 2 \end{bmatrix} & \begin{bmatrix} r \\ 1 \end{bmatrix} & \begin{bmatrix} r \\ 0 \end{bmatrix} & \dots & L & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ M & M & M & \dots & O & M \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \begin{bmatrix} r \\ n-1 \end{bmatrix} & \begin{bmatrix} r \\ n-2 \end{bmatrix} & \begin{bmatrix} r \\ n-3 \end{bmatrix} & \dots & L & \begin{bmatrix} r \\ 0 \end{bmatrix} \end{bmatrix}_{n \times n}, \quad (2)$$

where,

$$\begin{bmatrix} r \\ n \end{bmatrix} = \begin{cases} 1 & n = 0 \\ \frac{r(r+1)L(r+n-1)}{n!} & n \in N^+ \end{cases}. \quad (3)$$

In particular, when $r=1$, $x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i)$, $X^{(1)} = AX^{(0)}$ is obtained, where A represents the 1-AGO matrix, which can be expressed as follows:

$$A = \begin{bmatrix} 1 & 0 & 0 & L & 0 \\ 1 & 1 & 0 & L & 0 \\ 1 & 1 & 1 & L & 0 \\ M & M & M & O & M \\ 1 & 1 & 1 & L & 1 \end{bmatrix}_{n \times n}. \quad (4)$$

Definition 2.1.2 Let $x^{(r-1)}(k) = x^{(r)}(k) - x^{(r)}(k-1)$, $k=1,2,\dots,n$ be the gray r -order subtraction sequence (r -IAGO). Similarly, the expression $X^{(0)} = D^{-r}X^{(r)}$ can be obtained from the matrix operation theory, where D^{-r} represents the r -IAGO matrix, which can be expressed as follows:

$$D^{-r} = \begin{bmatrix} \begin{bmatrix} -r \\ 0 \end{bmatrix} & 0 & 0 & L & 0 \\ \begin{bmatrix} -r \\ 1 \end{bmatrix} & \begin{bmatrix} -r \\ 0 \end{bmatrix} & 0 & L & 0 \\ \begin{bmatrix} -r \\ 2 \end{bmatrix} & \begin{bmatrix} -r \\ 1 \end{bmatrix} & \begin{bmatrix} -r \\ 0 \end{bmatrix} & L & 0 \\ M & M & M & O & M \\ \begin{bmatrix} -r \\ n-1 \end{bmatrix} & \begin{bmatrix} -r \\ n-2 \end{bmatrix} & \begin{bmatrix} -r \\ n-3 \end{bmatrix} & L & \begin{bmatrix} -r \\ 0 \end{bmatrix} \end{bmatrix}_{n \times n}, \quad (5)$$

where,

$$\begin{bmatrix} -r \\ n \end{bmatrix} = \begin{cases} 1 & n=0 \\ \frac{(-r)(-r+1)L(-r+n-1)}{n!} & n \in N^+ \end{cases}. \quad (6)$$

In particular, when $r=1$, the 1-IAGO sequence can be represented as $x^{(0)}(k) = x^{(1)}(k) - x^{(1)}(k-1)$, $k=1,2,\dots,n$, and $x^{(0)} = D^{-1}x^{(1)}$ is obtained, where D^{-1} represents the 1-IAGO matrix, which can be expressed as follows:

$$D^{-1} = \begin{bmatrix} 1 & 0 & 0 & L & 0 \\ -1 & 1 & 0 & L & 0 \\ 0 & -1 & 1 & L & 0 \\ M & M & M & O & M \\ 0 & 0 & 0 & L & 1 \end{bmatrix}_{n \times n}. \quad (7)$$

2.2 Fractional FAGM (1,1) model

Definition 2.2.1 Let $X^{(0)} = (x^{(0)}(1), x^{(0)}(2), L, x^{(0)}(n))^T$ be the original sequence, $X^{(r)} = (x^{(r)}(1), x^{(r)}(2), L, x^{(r)}(n))^T = A^r X^{(0)}$, $X^{(r)}$ is an r -order accumulation of $X^{(0)}$ to generate a sequence (r -AGO). Let $Z_1^{(r)} = (z_1^{(r)}(2), z_1^{(r)}(3), L, z_1^{(r)}(n))^T$ be the sequence of the nearest mean (the background value) of $X^{(r)}$. Where,

$$z_1^{(r)}(k) = \frac{x^{(r)}(k) + x^{(r)}(k-1)}{2}, k=2,3,\dots,n, \quad (8)$$

The gray differential equation:

$$x^{(r)}(k+1) - x^{(r)}(k) + a_1 z_1^{(r)}(k) = b_1. \quad (9)$$

is called the r -order cumulative gray FAGM (1,1) model. In particular, when $r=1$, $x^{(r)}(k+1) - x^{(r)}(k) + a_1 z_1^{(r)}(k) = b$ becomes $x^{(0)}(k) + a_1 z^{(1)}(k) = b_1$, which is the mean GM(1,1) model. Where a_1 is the development coefficient and b_1 is the gray action amount.

The model parameters a_1 and b_1 can be solved by the least squares method. Let $\hat{u}_1 = [a_1, b_1]^T$, by least squares method:

$$\hat{u}_1 = (B_1^T B_1)^{-1} B_1^T Y_1, \quad (10)$$

where,

$$Y_1 = \begin{bmatrix} x^{(r)}(2) - x^{(r)}(1) \\ x^{(r)}(3) - x^{(r)}(2) \\ \vdots \\ x^{(r)}(n) - x^{(r)}(n-1) \end{bmatrix}, \quad B_1 = \begin{bmatrix} -z^{(r)}(2) & 1 \\ -z^{(r)}(2) & 1 \\ \vdots & \vdots \\ -z^{(r)}(2) & 1 \end{bmatrix}. \quad (11)$$

Definition 2.2.2 Let the model parameter's be as defined above, then

$$\frac{dx^{(r)}(t)}{dt} + a_1 x^{(r)}(t) = b_1, \quad (12)$$

is called the whitening differential equation of the r-order cumulative gray GM (1,1) model.

To solve equation (12), let $\hat{x}^{(0)}(1) = x^{(0)}(1)$ get the time response sequence as:

$$\hat{x}^{(r)}(k) = \left(x^{(0)}(1) - \frac{b_1}{a_1} \right) e^{-a(k-1)} + \frac{b_1}{a_1}, \quad k = 2, 3, \dots, n. \quad (13)$$

After the restore, the predicted value of $X^{(0)}$ is:

$$\hat{X}^{(0)} = D^{(-r)} \hat{X}^{(r)}, \quad (14)$$

where $\hat{X}^{(0)} = (\hat{x}^{(0)}(1), \hat{x}^{(0)}(2), \dots, \hat{x}^{(0)}(n))^T$, $\hat{X}^{(r)} = (\hat{x}^{(r)}(1), \hat{x}^{(r)}(2), \dots, \hat{x}^{(r)}(n))^T$.

2.3 Model test

The error of the model is an important indicator for evaluating the reliability and practicability of a model. In this paper, in order to test the prediction accuracy of the model, the root mean square error of the fitted data ($RMSEOTFD$) is calculated, the root mean square error of the extrapolated predicted value ($RMSEOPV$) and the total root mean square error ($TRMSE$) are used to represent the calculation error of the model. Calculated as follows:

$$RMSEOTFD = \frac{1}{N} \sum_{k=1}^N \left(x^{(0)}(k) - \hat{x}^{(0)}(k) \right)^2, \quad (15)$$

$$TRMSE = \frac{1}{n} \sum_{k=1}^n \left(x^{(0)}(k) - \hat{x}^{(0)}(k) \right)^2, \quad (16)$$

$$RMSEOPV = \frac{1}{n-N} \sum_{k=N+1}^n \left(x^{(0)}(k) - \hat{x}^{(0)}(k) \right)^2. \quad (17)$$

Where n represents the total number of sample data, and N represents the number of data used to fit the model. Obviously, the smaller the $RMSEOTFE$, $RMSEOPV$ and $RTMSE$, the better the prediction of the model.

3. Model Error Analysis

From the time response sequence of FAGM(1,1), the prediction accuracy of the FAGM(1,1) model depends on the parameters a_1 and b_1 , whereas the parameter a_1 , b_1 is mainly determined by the background value $z^{(r)}(k)$. Therefore, the FAGM(1,1) model error is mainly derived from the calculation of the background value $z^{(r)}(k)$. Integrate the whitening

differential equation (12) in the interval $[k-1, k]$:

$$I = \int_{k-1}^k dx^{(r)}(t) + a_1 \int_{k-1}^k x^{(r)}(t) dt = b_1 \int_{k-1}^k dt, \quad k = 2, 3, \dots, n, \quad (18)$$

$$I = x^{(r)}(k) - x^{(r)}(k-1) + a_1 \int_{k-1}^k x^{(r)}(t) dt = b_1, \quad k = 2, 3, \dots, n. \quad (19)$$

Compare the above formula with $x^{(r)}(k+1) - x^{(r)}(k) + a_1 z_1^{(r)}(k) = b_1$ to get:

$$z_1^{(r)}(k) = \int_{k-1}^k x^{(r)}(t) dt. \quad (20)$$

It can be known from the above formula (20) that the geometrical meaning of the background value in the GM(1,1) model is the area of the curved trapezoid surrounded by the curve $x^{(r)}(t)$ and the t axis in the interval $[k-1, k]$, as shown in Fig. 1.

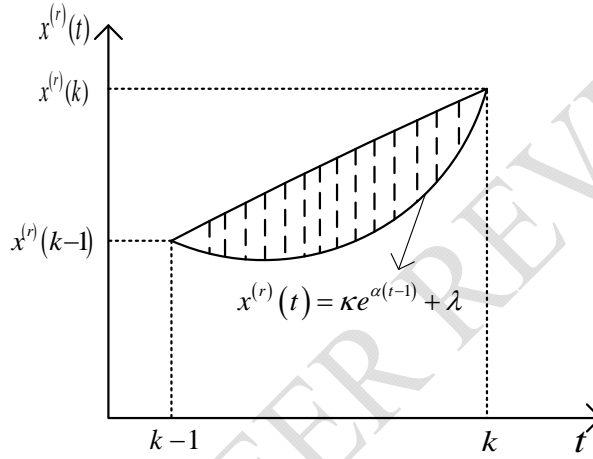


Fig.1 Error analysis chart

It can be seen from the background value calculation formula

$$z_1^{(r)}(k) = \frac{x^{(r)}(k) + x^{(r)}(k-1)}{2}, \quad k = 2, 3, \dots, n$$

that the area calculated by the formula is actually the average value of $x^{(r)}(k)$ and $x^{(r)}(k-1)$, that is, the trapezoidal area enclosed by the t axis, instead of the integral value of the integral curve $x^{(r)}(t)$ on the interval $[k-1, k]$.

When the time interval of the sequence is small and the data changes are relatively flat, it is more reasonable to use the trapezoidal formula to calculate the background value. However, when the data interval is large and the value changes drastically, the trapezoidal formula is used to approximate the background value. This will increase the calculation error of the background value. Therefore, the improvement of the background value is particularly important. As can be seen from Fig. 1, the area calculated by the trapezoidal formula is smaller than the integral value of the integral curve $x^{(r)}(t)$ in the interval $[k-1, k]$. The shaded part of Fig. 1 is the source of the error.

4. Improved FAGM (1,1) model with background values

In order to improve the fitting precision and prediction accuracy of FAGM(1,1) model, the application range of FAGM(1,1) model is wider. Based on the above error analysis, this paper proposes to calculate the background value of the optimized FAGM(1,1) model. It can

be seen from the time response sequence $\hat{x}^{(r)}(k) = \left(x^{(0)}(1) - \frac{b_1}{a_1} \right) e^{-a(k-1)} + \frac{b_1}{a_1}, \quad k = 2, 3, \dots, n$ of

the FAGM(1,1) model that $x^{(r)}(t)$ is a non-homogeneous exponential curve, so $x^{(r)}(k) = \kappa e^{\alpha(k-1)} + \lambda$ is set. Where κ, λ, α is the undetermined coefficient and needs to satisfy:

$$x^{(r)}(k) = \kappa e^{\alpha(k-1)} + \lambda, k = 2, 3, \dots, n. \quad (21)$$

It is also known from the nature of the sequence (r -AGO) generated by the r -order accumulation, when $k = 1$, $x^{(r)}(1) = x^{(0)}(1)$, and:

$$\begin{aligned} x^{(r-1)}(k) &= x^{(r)}(k) - x^{(r)}(k-1) \\ &= \kappa e^{\alpha(k-1)} - \kappa e^{\alpha(k-2)} \\ &= \kappa e^{\alpha(k-1)}(1 - e^{-\alpha}). \end{aligned} \quad (22)$$

Also because:

$$\frac{x^{(r-1)}(k)}{x^{(r-1)}(k-1)} = \frac{\kappa e^{\alpha(k-1)}(1 - e^{-\alpha})}{\kappa e^{\alpha(k-2)}(1 - e^{-\alpha})} = e^{\alpha}. \quad (23)$$

From the above equation (23):

$$\alpha = \ln \frac{x^{(r-1)}(k)}{x^{(r-1)}(k-1)}. \quad (24)$$

According to the above equation (22), the undetermined coefficient κ can be solved:

$$\begin{aligned} \kappa &= \frac{x^{(r-1)}(k)}{e^{\alpha(k-1)}(1 - e^{-\alpha})} = \frac{x^{(r-1)}(k)}{e^{\frac{\ln \frac{x^{(r-1)}(k)}{x^{(r-1)}(k-1)}(k-1)}(1 - e^{-\frac{\ln \frac{x^{(r-1)}(k-1)}{x^{(r-1)}(k)})}})} \\ &= \frac{x^{(r-1)}(k)}{\left(\frac{x^{(r-1)}(k)}{x^{(r-1)}(k-1)}\right)^{(k-1)} \left(1 - \frac{x^{(r-1)}(k-1)}{x^{(r-1)}(k)}\right)}. \end{aligned} \quad (25)$$

It is known from the initial conditions:

$$x^{(r)}(1) = x^{(0)}(1) = \kappa e^{\alpha(1-1)} + \lambda. \quad (26)$$

From the above equation (26):

$$\lambda = x^{(0)}(1) - \kappa. \quad (27)$$

For $x^{(r)}(t)$, integrate the interval $[k-1, k]$ to obtain the background value $z_2^{(r)}(k)$, namely:

$$z_2^{(r)}(k) = \int_{k-1}^k x^{(r)}(t) dt. \quad (28)$$

Bringing $x^{(r)}(k) = \kappa e^{\alpha(k-1)} + \lambda$ into the above equation (28) is:

$$z_2^{(r)}(k) = \int_{k-1}^k (\kappa e^{\alpha(t-1)} + \lambda) dt. \quad (29)$$

Further simplification of equation (29):

$$\begin{aligned}
z_2^{(r)}(k) &= \int_{k-1}^k (\kappa e^{\alpha(t-1)} + \lambda) dt = \kappa \int_{k-1}^k e^{\alpha(k-1)} dt + \lambda \int_{k-1}^k dt \\
&= \frac{1}{\alpha} (\kappa e^{\alpha(k-1)} - \kappa e^{\alpha(k-2)}) + \lambda \\
&= \frac{1}{\alpha} [x^{(r)}(k) - x^{(r)}(k-1)] + \lambda \\
&= \frac{1}{\alpha} x^{(r-1)}(k) + \lambda.
\end{aligned} \tag{30}$$

When $x^{(r)}(k-1)$ and $x^{(r)}(k)$ are equal in interval $[k-1, k]$, there are:

$$z_2^{(r)}(k) = \lim_{x^{(r)}(k) \rightarrow x^{(r-1)}(k-1)} \left(\frac{1}{\alpha} x^{(r-1)}(k) + \lambda \right) = x^{(r)}(k-1). \tag{31}$$

Equation (31) above can be verified from the background value geometry. It is known from the above equation (30) that $z_2^{(r)}(k) = \frac{1}{\alpha} x^{(r-1)}(k) + \lambda$ is an improved background value calculation formula, thereby obtaining an optimized FAGM (1, 1) model. The model is predicted by the improved background value calculation formula, so that $\hat{u}_2 = [a_2, b_2]^T$, by least squares method has:

$$\hat{u}_2 = (B_2^T B_2)^{-1} B_2^T Y_2. \tag{32}$$

Among them:

$$Y_2 = \begin{bmatrix} x^{(r)}(2) - x^{(r)}(1) \\ x^{(r)}(3) - x^{(r)}(2) \\ \vdots \\ x^{(r)}(n) - x^{(r)}(n-1) \end{bmatrix}, \quad B_2 = \begin{bmatrix} -z_2^{(r)}(2) & 1 \\ -z_2^{(r)}(2) & 1 \\ \vdots & \vdots \\ -z_2^{(r)}(2) & 1 \end{bmatrix}. \tag{33}$$

The whitening differential equation of the gray FAGM (1,1) model is:

$$\frac{dx^{(r)}(t)}{dt} + a_2 x^{(r)}(t) = b_2. \tag{34}$$

Let $\hat{x}^{(0)}(1) = x^{(0)}(1)$, get the time response sequence as:

$$\hat{x}^{(r)}(k) = \left(x^{(0)}(1) - \frac{b_2}{a_2} \right) e^{-a(k-1)} + \frac{b_2}{a_2}, k = 2, 3, \dots, n. \tag{35}$$

The predicted value of $X^{(0)}$ after restoration is:

$$\hat{X}^{(0)} = D^{(-r)} \hat{X}^{(r)}. \tag{36}$$

5. Determination of the optimal order

When modeling with the fractional FAGM(1,1) model, the optimal order r needs to be determined so that the fitting accuracy and prediction accuracy of the model are the highest. In this paper, the root mean square error (*RMSEOTFD*) of the fitted data is used as the fitness function, and the optimal order r is solved by the genetic algorithm. Its calculation form is as follows:

$$\min RMSEOTFD(r) = \frac{1}{N} \sum_{k=1}^N \left(\hat{x}^{(0)}(k) - x^{(0)}(k) \right)^2. \tag{37}$$

$$s.t. \begin{cases} \hat{X}^{(0)} = D^{-r} \hat{X}^{(r)}, \\ k = 1, 2, \dots, n. \end{cases} \quad (38)$$

Where N represents the number of data to fit. The fitness function of the above equation (37) has a high degree of nonlinearity, so solving (37) to achieve optimal is a complex optimization problem. If a conventional optimization algorithm is used to solve this optimization problem, not only the calculation amount is large, but also the calculation complex, not conducive to modeling prediction of fractional order FAGM (1,1). Therefore, other algorithms are needed to solve this problem.

The genetic algorithm is an adaptive global optimization search algorithm developed by Professor J.H. Holland [28] for the first time to simulate the biological evolution process of nature. The algorithm simulates the evolutionary rules of natural selection and survival of the fittest through mechanisms such as natural selection, inheritance, and cross-variation. Genetic algorithm has the advantages of high efficiency, practicability and robustness. It has been widely used in the fields of machine learning, pattern recognition, optimization control, etc. [29-31], and it has developed extremely rapidly. Because it can effectively solve the optimal extremum problem of NP problem, highly nonlinear and multi-peak function, this paper uses genetic algorithm to find the optimal order r of the above formula (37).

The calculation steps of the genetic algorithm are as follows:

Step1: Initialize the population and generate a random sequence of determined length as the initial population;

Step2: Calculate individual fitness;

Step3: Perform selection, crossover, and mutation calculations to generate offspring populations and prepare for the next genetic operation;

Step4: Conditional judgment. End the genetic operation if the termination condition is met; otherwise, return to Step2

Step5: Output the optimal order r , and the calculation ends.

6. Example calculation

In order to verify the practicability and effectiveness of the above-mentioned background value optimized FAGM (1,1) model, this paper obtains the national tax revenue data from 2000 to 2011 of the China Statistical Yearbook as the basic data. The data from 2000 to 2006 was used as the model fitting data, and the data from 2007 to 2011 was used as the model predictive verification data. The national tax revenue data from 2000 to 2006 is shown in Table 1 below:

Table 1: National Tax Revenue Statistics data from 2000 to 2006 (Unit: 100 million yuan)

Years	2000	2001	2002	2003	2004	2005	2006
National tax revenue	12581.51	15301.38	17636.45	20017.31	24165.68	28778.54	34804.35

Based on the data in Table 1 above, the GM(1,1) model, the GM(1,1) model improved by the Simpson formula, the FAGM(1,1) model, and the FAGM(1,1) model with improved background values are respectively established. Compare and analyze the modeling effects of different models. The optimal order of the FAGM(1,1) model with fractional FAGM(1,1) model and background value improved by genetic algorithm are respectively -0.12 and 1.189, and is predicted 5 years later. The fitting effect and prediction accuracy of each model are compared and analyzed. The fitting results of the four models are shown in Table 2 below:

Table 2: Fitting of different models(Unit: 100 million yuan)

Years	Actual value	GM(1,1)	Simpson GM(1,1)	FAGM(1,1)	Improve FAGM(1,1)
2000	12581.51	12581.51	12581.51	12581.51	12581.51
2001	15301.38	14633.79	14580.29	15190.82	15421.04
2002	17636.45	17338.67	17275.36	17651.58	17301.91
2003	20017.31	20543.51	20468.60	20502.09	20256.99
2004	24165.68	24340.72	24252.09	24059.85	24100.17
2005	28778.54	28839.80	28734.93	28655.57	28904.89
2006	34804.35	34170.48	34046.40	34698.75	34827.75
<i>RMSEOTFD</i>		422.14	453.22	201.75	170.92

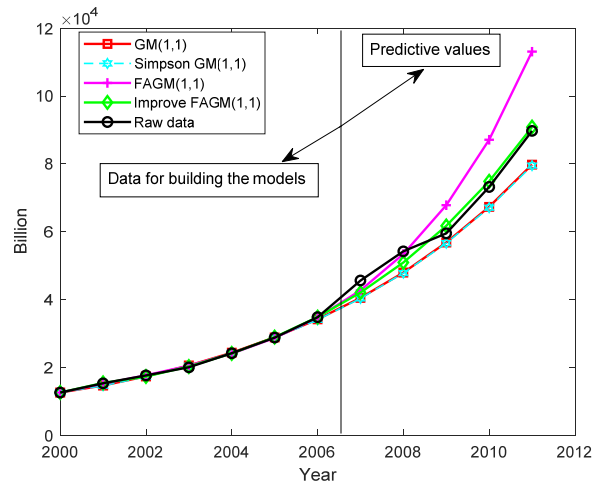


Fig. 1. Fit and prediction effects of different models

It can be seen from Table 2 that the *RMSEOTFD* of the improved FAGM(1,1) model is 170.92, while the classical GM(1,1) model, the Simpson formula improved GM(1,1) and the FAGM(1,1) model *RMSEOTFD* are respectively 422.14, 453.22 and 201.75, both are significantly higher than the improved FAGM (1,1) model. It can be seen from Fig. 1 that the improved FAGM(1,1) model has higher prediction accuracy than the GM(1,1) model and the Simpson modified GM(1,1) and FAGM(1,1) models. It can be seen that the FAGM(1,1) model with improved background value has better fitting effect and prediction accuracy than GM(1,1) model, Simpson improved GM(1,1) and FAGM(1,1). model. Table 3 below shows the extrapolated predictions for different models.

Table 3: The prediction of different models. (Unit: 100 million yuan):

Years	Actual value	GM(1,1)	Simpson GM(1,1)	FAGM(1,1)	Improve FAGM(1,1)
2007	45621.97	40486.48	40339.65	42724.52	42083.44
2008	54223.79	47969.90	47796.18	53445.43	50943.96
2009	59521.59	56836.54	56631.00	67817.12	61745.69
2010	73210.79	67342.07	67098.88	87125.19	74900.79
2011	89738.39	79789.42	79501.67	113101.60	90912.32
<i>RMSEOPV</i>			6629.0		
		6420.56		12784.89	2547.91
<i>TRMSE</i>		4156.97	4292.98	8254.05	1649.84

As can be seen from Table 3 above, the GM(1,1) model, the Simpson modified GM(1,1)

model, the FAGM(1,1) model, and the improved FAGM(1,1) model $RMSEOPV$ are respectively 64250.56, 6629.0, 12784.89 and 2549.91; $TRMSE$ are respectively 4156.97, 4292.98, 8254.05 and 1649.84. Among them, the improved FAGM (1,1) model has smaller $RMSEOPV$ and $TRMSE$ than the other three models.

Table 4 Relative error of the four models

Years	GM(1,1)	Simpson GM(1,1)	FAGM(1,1)	Improve FAGM(1,1)
2000	0.0000%	0.0000%	0.0000%	0.0000%
2001	-4.3629%	-4.7126%	-0.7226%	0.7820%
2002	-1.6884%	-2.0474%	0.0858%	-1.8969%
2003	2.6287%	2.2545%	2.4218%	1.1974%
2004	0.7243%	0.3576%	-0.4379%	-0.2711%
2005	0.2129%	-0.1515%	-0.4273%	0.4391%
2006	-1.8212%	-2.1778%	-0.3034%	0.0672%
2007	-11.2566%	-11.5784%	-6.3510%	-7.7562%
2008	-11.5335%	-11.8539%	-1.4355%	-6.0487%
2009	-4.5110%	-4.8564%	13.9370%	3.7366%
2010	-8.0162%	-8.3484%	19.0059%	2.3084%
2011	-11.0866%	-11.4073%	26.0348%	1.3082%
Mean relative error	-4.2259%	-4.5435%	4.3173%	-0.5112%

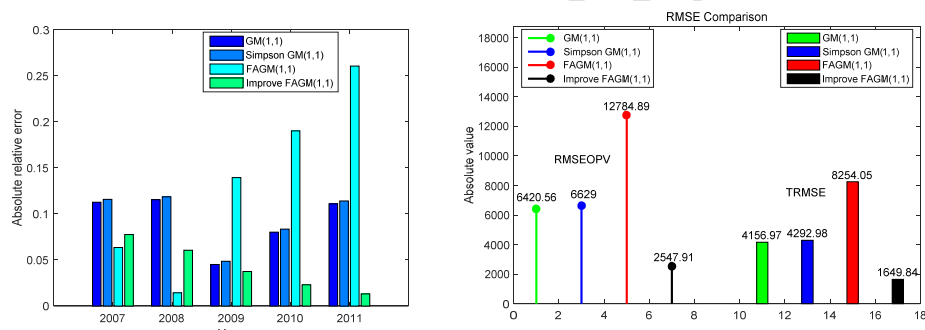


Fig. 2 Error comparison of four gray models

Further, in order to more intuitively see the prediction accuracy of each model, the error of each model is shown graphically, as shown in Fig. 2. In comparison, the FAGM(1,1) model with improved background values has higher prediction accuracy than the GM(1,1) and Simpson formula improved GM(1,1) and FAGM(1,1) models.

7. Conclusion

Based on the original fractional grey theory, this paper analyzes the error sources of FAGM(1,1) model. After theoretical derivation, an improved background value calculation method is proposed, and the genetic algorithm is used to solve the model parameters to improve the prediction accuracy of the model. Finally, this paper applies the improved FGAM(1,1) model with background value calculation to the prediction of China's fiscal tax revenue data, and the improved GM(1,1), Simpson formula GM(1,1), FAGM (1,1) A comparative analysis shows that the absolute relative error of the GM(1,1) and FAGM(1,1) models improved by the classical GM(1,1) and Simpson formulas in the actual prediction exceeds the absolute value. 4%, while the improved FAGM(1,1) model has an average relative error of -0.5112%. The validity and superiority of the model are fully verified. At the same time, it has improved the foresight of China's fiscal tax, laid the foundation for the objective requirements of scientific decision-making and scientific management, and

provided a theoretical background for the macro-control of China's socialist market economy.

8. References

- [1] Zhang F. Study on the tax forecast based on bp neural network [D]. Chang'an University, 2016.Chinese.
- [2] Li YY. Revenue forecasting models and their application to the VAT [D]. Southwestern University of Finance and Economics, 2014.Chinese.
- [3] Zhang XY. The forecast and analysis of financial general revenue in Sichuan province [D]. Sichuan University, 2007.Chinese.
- [4] Liu X. Research and application of prediction and analysis method for fiscal and tax categories [D]. University of Chinese Academy of Sciences (College of Engineering Management and Information Technology, Chinese Academy of Sciences), 2017.Chinese.
- [5] Deng JL. Control problems of grey systems[J]. Systems & Control Letters, 1982, 1(5): 288-294.
- [6]Fei ZH, Jiang ZY, He ZF. Discrete GM(1,1) model and its application for forecasting of real estate prices[C]// International Conference on Management Science & Industrial Engineering. 2011.
- [7]Truong DQ, Ahn KK. An accurate signal estimator using a novel smart adaptive grey model SAGM(1,1)[J]. Expert Systems with Applications, 2012, 39(9):7611-7620.
- [8] Jun LL, Rui QL, Xiao CC. Generalized accumulated GM(1,1) cosine model and its application[J]. Grey Systems Theory & Application, 2011, 2(2):242 - 245.
- [9] Ou SL. Forecasting agricultural output with an improved grey forecasting model based on the genetic algorithm[J]. Computers and Electronics in Agriculture, 2012, 85(none):33-39.
- [10] Hsu LC, Wang CH. Forecasting integrated circuit output using multivariate grey model and grey relational analysis[J]. Expert Systems with Applications, 2009, 36(2):1403-1409.
- [11] Tan GJ. The structure method and application of background value in grey system GM(1,1) model(I)[J].Systems Engineering-Theory &Practice,2000(04):98-103.Chinese.
- [12] Tan GJ. The structure method and application of background value in grey system GM(1,1) model(II)[J]. Systems Engineering-Theory &Practice,2000(05):125-127+132. Chinese.
- [13] Tan GJ. The structure method and application of background value in grey system GM(1,1) model (III). Systems Engineering-Theory &Practice,2000(06):70-74. Chinese.
- [14]Chen YG, Yang DY, Dai WZ. Modling research of modeling GM(1,1) background value-building based on gauss-legendre quadrature and its application [J].Journal of Zhejiang Sci-Tech University,2007(04):444-447+460. Chinese.
- [15] Jiang N, Liu XY. A new methodof background value-building and application in GM(1,1) model based on gauss formula[J].Mathematics in Practice and Theory,2008(07):90-94.
- [16]Qiu SF, Wang ZW. Improvement of background value calculation of grey GM(1,1) model[J].Statistics & Decision,2007(03):129-131.Chinese.
- [17]Dong K, Lv WY. Background value optimization of GM(1,1) model based on complex Simpson formula[J]. Mathematics in Practice and Theory,2015,45(18):180-185. Chinese.
- [18]Wu FF, Shi KB, Yiziteliopu N, Dong SK. Based on Simpson formula improved non-interval GM(1,1) model and application[C]// Control Conference. 2013.
- [19]Shen Y, Sun H. Optimization of background vlues of GM (1, N) model and its application[J]. International Journal of Information Processing & Management, 2013,

4(1):58-64.Chinese.

[20]Li JF, Dai WZ. New method and application of background value construction of GM(1,1) model based on interpolation and Newton-Cores formula[J].Systems Engineering-Theory & Practice, 2004(10):122-126.Chinese.

[21]He MX, Wang Q. Constructing the background value for GM(1,1) model based on Simpson formula[J]Mathematics in Economics,2011,28(04):101-104. Chinese.

[22]Luo D, Liu SF, Dang YG. Grey model GM(1,1) optimization[J].Engineering Science,2003(08):50-53. Chinese.

[23]Wang ZX, Dang YG, Liu SF. An optimal GM(1,1) based on the discrete function with exponential law[J].Systems Engineering-Theory & Practice,2008(02):61-67.Chinese.

[24]Wu LF, Liu SF, Li J. GM(1,1) model based on fractional order accumulating method and its stability[J]Control and Decision,2014,29(05):919-924.Chinese.

[25]Wu LF, Liu SF, Yao LG. Discrete grey model based on fractional order accumulate[J]. Systems Engineering-Theory & Practice, 2014,34(07):1822-1827.Chinese.

[26]Wu LF, Liu SF, Yao LG. Grey system model with the fractional order accumulation[J]. Communications in Nonlinear Science & Numerical Simulation, 2013, 18(7):1775-1785.Chinese

[27]Wu LF, Liu SF, Fang ZG, Xu HY. Properties of the GM(1,1) with fractional order accumulation[J]. Applied Mathematics & Computation, 2015, 252(C):287-293.Chinese

[28]Holland J H . Adaptation in Natural and Artificial System[M]// Adaptation in natural and artificial systems. MIT Press, 1992.

[29]Zhao TW. Studies of face recognition based on machine learning methods[D].Shanghai Jiao Tong University,2008.Chinese.

[30] Wang XF, Yang RY, Zhang XJ, Yang WL. Application of genetic algorithm in pattern recognition [J]. Journal of Chinese computer systems,1997(10):33-37,Chinese.

[31]Zhao LJ, Fan JY. Multi-objective optimization design of roadheader's cutting head based on the GA[J]. Journal of Mechanical Strength,2018,40(06):1349-1355.Chines.