



The Forecast Research of Linear Regression Forecast Model in National Economy

Yu Xiao¹, Zhezhi Jin^{2*}

¹Department of Mathematics, Yanbian University, Yanji, China

²Department of Economics and Management, Yanbian University, Yanji, China

Email: 1150457896@qq.com, *jinzhezhi@sina.com

How to cite this paper: Xiao, Y. and Jin, Z.Z. (2021) The Forecast Research of Linear Regression Forecast Model in National Economy. *Open Access Library Journal*, 8: e7797.

<https://doi.org/10.4236/oalib.1107797>

Received: July 27, 2021

Accepted: August 20, 2021

Published: August 23, 2021

Copyright © 2021 by author(s) and Open Access Library Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Based on application of Linear Regression Model in national economy. *Statistical Yearbook of China* is a statistical data book published by China. Some different types of data in China have been documented every year. The dissertation is researching some part of national economy in *Statistical Yearbook of China* to research the relationship between GNP (gross national product) and various data, and afterwards to establish model to achieve some rationalization suggestions for the future of the National Economy. The basic method for data mining is prediction. The method proposed in this dissertation will use multiple linear regression model combined with Gray Prediction to research to achieve processing information. This dissertation uses forward stepwise regression or backward stepwise regression in multiple linear regression to make specific data model. Afterwards some fuzzy data will be analysed with Gray prediction model. Finally, the combination of the two realizes the data prediction algorithm. The main calculation tools that will be used in this dissertation are SPSS and MATLAB. The prediction results are based on model which is obtained in this dissertation. After comparing the multiple linear regression and the grey prediction model, the researcher found the algorithm of this dissertation is more accurate, thereby verifying the rationality of the prediction model in this dissertation.

Subject Areas

Statistics

Keywords

Multiple Linear Regression, Gray Model, GM Models, Prediction

1. Introduction

The analysis method of the national economy has always been linear regression or other various methods for regression analysis, but when we use a single method for forecasting or data analysis, there will often be large errors or large limitations. In this case, the combined model can often be used to calculate more accurate results. For example, when using a multiple linear regression model, it can predict data with a linear relationship, but it cannot predict data with an exponential relationship. When using a gray prediction model, can only make fuzzy predictions on the data, and cannot reflect the linear relationship of the data. The model in this article solves the problem that the linear prediction cannot determine the highest power and function, and at the same time solves the problem that the gray prediction model cannot perform linear prediction. Combining the two can be used in a variety of sequence forecasts with both linear and exponential growth trends in the future. It has broad significance and can make better predictions on the data.

2. Multiple Linear Regression and Grey Forecast Theory Basis

2.1. Introduction to Multiple Linear Regression

Suppose there is a random variable y and a general variable x_1, x_2, \dots, x_p . The linear regression model between is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon \quad (1.1)$$

In $\beta_0, \beta_1, \dots, \beta_p$ is $p+1$, unknown parameter β_0 can be called regression constant, β_1, \dots, β_p is regression coefficients. Y is dependent variable, x_1, x_2, \dots, x_p is p general variables that can be accurately measured and controlled at the same time are called independent variables. When $P = 1$, the formula (1) is called a univariate linear regression model. At that time, we can call Equation (1) a multiple linear regression model. ε is random error. Regarding the random error

term, we usually set $\begin{cases} E(\varepsilon) = 0 \\ \text{Var}(\varepsilon) = \sigma^2 \end{cases}$, $E(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$ is the

theoretical regression equation.

For a practical problem, if we have obtained n sets of observation data $(x_{i1}, x_{i2}, \dots, x_{ip}; y_i) (i=1, 2, \dots, n)$ then the above linear regression model Equation (1) can be expressed as

$$\begin{cases} y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \dots + \beta_p x_{1p} + \varepsilon_1 \\ y_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \dots + \beta_p x_{2p} + \varepsilon_2 \\ \vdots \\ y_n = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \dots + \beta_p x_{np} + \varepsilon_n \end{cases} \quad (1.2)$$

It is written in the form of a matrix:

$$y = X\beta + \varepsilon, \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_p \end{bmatrix} \quad (1.3)$$

The second term X is a $n \times (p+1)$ matrix of order, which is called regression design matrix or data matrix. In the design of the experiment, because the elements in X can be set in advance and can be controlled, because there are some subjective factors, so X is called the design matrix [1].

For example, to analyze the relationship between automobile characteristics and automobile sales volume, a multivariate linear regression model is established first. The data used by the model is shown in **Figure 1** [2] [3] [4].

Firstly, it is stepwise regressed by SPSS, then selected, analyzed, regressed and linearized, and obtained (**Tables 1-3**).

sales	resale	type	price	engines	horsepow	wheelbas	width	length
16.919	16.360	0	21.500	1.8	140	101.2	67.3	172.4
39.384	19.875	0	28.400	3.2	225	108.1	70.3	192.9
14.114	18.225	0		3.2	225	106.9	70.6	192.0
8.588	29.725	0	42.000	3.5	210	114.6	71.4	196.6
20.397	22.255	0	23.990	1.8	150	102.6	68.2	178.0
18.780	23.555	0	33.950	2.8	200	108.7	76.1	192.0
1.380	39.000	0	62.000	4.2	310	113.0	74.0	198.2
19.747		0	26.990	2.5	170	107.3	68.4	176.0
9.231	28.675	0	33.400	2.8	193	107.3	68.5	176.0
17.527	36.125	0	38.900	2.8	193	111.4	70.9	188.0
91.561	12.475	0	21.975	3.1	175	109.0	72.7	194.6
39.350	13.740	0	25.300	3.8	240	109.0	72.7	196.2
27.851	20.190	0	31.965	3.8	205	113.8	74.7	206.8
83.257	13.360	0	27.885	3.8	205	112.2	73.5	200.0
63.729	22.525	0	39.895	4.6	275	115.3	74.5	207.2
15.943	27.100	0	44.475	4.6	275	112.2	75.0	201.0
6.536	25.725	0	39.665	4.6	275	108.0	75.5	200.6
11.185	18.225	0	31.010	3.0	200	107.4	70.3	194.8
14.785		1	46.225	5.7	255	117.5	77.0	201.2
145.519	9.250	0	13.260	2.2	115	104.1	67.9	180.9
135.126	11.225	0	16.535	3.1	170	107.0	69.4	190.4
24.629	10.310	0	18.890	3.1	175	107.5	72.5	200.9
42.593	11.525	0	19.390	3.4	180	110.5	72.7	197.9

Figure 1. Relationship between automobile characteristics and automobile sales volume.

Table 1. Model summary table.

Model summary				
Model	R	R-squared	[Adjusted] [R Square]	Standard skewness error
1	0.757 ^A	0.573	0.323	34.524902
2	0.756 ^B	0.572	0.374	33.208118
3	0.752 ^C	0.566	0.411	32.220528
4	0.742 ^D	0.551	0.431	31.660083
5	0.729 ^E	0.532	0.444	31.290444

^APredicted value: (constant), resale, length, horsepower, width, wheelbase, engine, price; ^BPredicted value: (constant), resale, length, horsepower, width, wheelbase, engine; ^CPredicted value: (constant), resale, length, horsepower, wheelbase, engine; ^DPredicted value: (constant), resale, length, wheelbase, engine; ^EPredicted Value: (constant), resale, length, wheelbase.

Table 2. Analysis of variance.

Analysis of Variance ^[A]						
Model	Sum of squares	Df	Mean square	F	Significance	
1	Regression	19,170.934	7	2738.705	2.298	0.098 ^B
	Residual error	14,303.626	12	1191.969		
	Total	33,474.560	19			
2	Regression	19,138.432	6	3189.739	2.892	0.051 ^C
	Residual error	14,336.128	13	1102.779		
	Total	33,474.560	19			
3	Regression	18,940.286	5	3788.057	3.649	0.025 ^D
	Residual error	14,534.274	14	1038.162		
	Total	33,474.560	19			
4	Regression	18,439.147	4	4609.787	4.599	0.013 ^E
	Residual error	15,035.413	15	1002.361		
	Total	33,474.560	19			
5	Regression	17,809.090	3	5936.363	6.063	0.006 ^F
	Residual error	15,665.470	16	979.092		
	Total	33,474.560	19			

^AVariable: Sales; ^BPredicted values: (constant), resale, length, horsepower, width, wheelbase, engine, price; ^CPredicted values: (constant), resale, length, horsepower, width, wheelbase, engine; ^DPredicted values: (constant), resale, length, horsepower, wheelbase, engine; ^EPredicted values: (constant), resale, length, wheelbase, engine; ^FPredicted values: (constant), resale, length, wheelbase.

Table 3. Excluded variables table.

Excluded variables ^A						
Model		Beta in	T	Significance	Partial correlation	[Collinearity] [Statistics]
						Tolerance
2	Price	0.151 ^B	0.165	0.872	0.048	0.042
3	Price	0.109 ^C	0.124	0.903	0.034	0.043
	Width	-0.156 ^C	-0.424	0.679	-0.117	0.242
4	Price	-0.194 ^D	-0.268	0.792	-0.072	0.061
	Width	-0.163 ^D	-0.453	0.658	-0.120	0.243
	Horsepow	-0.456 ^D	-0.695	0.499	-0.183	0.072
5	Price	0.088 ^E	0.141	0.890	0.036	0.080
	Width	-0.120 ^E	-0.338	0.740	-0.087	0.248
	Horsepow	0.098 ^E	0.271	0.790	0.070	0.240
	Engines	0.300 ^E	0.793	0.440	0.201	0.209

^AVariable: Sales; ^BPredicted values in the model: (constant), resale, length, horsepow, width, wheelbas, engineers; ^CPredicted values in the model: (constant), resale, length, horsepow, wheelbas, engineers; ^DPredicted values in model: (constant), resale, length, wheelbas, engineer; ^EPredicted values in the model: (constant), resale, length, wheelbas.

2.2. Grey Pre-Test and Test

Grey prediction method is a kind of prediction method for some systems with uncertain factors. Grey theory holds that although some systematic behavior phenomena are hazy and the data are complex, they are still orderly and have certain overall functions. Therefore, the generation of grey number is to find the law from the clutter, and the grey theory establishes a generated data model instead of the original data model, so the data used in grey prediction is the result of inverse processing of some predicted values obtained from the GM model of generated data. The method of grey prediction is a branch of theory in the field of fuzzy prediction, which is more perfect in theory and method [5].

Model of grey prediction

First of all, we should do some preprocessing to the data.

For example, let's say that the original data sequence is

$$x^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(N)\} = \{6, 3, 8, 10, 7\}$$

Accumulate this set of data

$$x^{(1)}(1) = x^{(0)}(1) = 6$$

$$x^{(1)}(2) = x^{(0)}(1) + x^{(0)}(2) = 6 + 3 = 9$$

$$x^{(1)}(3) = x^{(0)}(1) + x^{(0)}(2) + x^{(0)}(3) = 6 + 3 + 8 = 17$$

$$x^{(1)}(4) = x^{(0)}(1) + x^{(0)}(2) + x^{(0)}(3) + x^{(0)}(4) = 6 + 3 + 8 + 10 = 27$$

$$x^{(1)}(5) = x^{(0)}(1) + x^{(0)}(2) + x^{(0)}(3) + x^{(0)}(4) + x^{(0)}(5) = 6 + 3 + 8 + 10 + 7 = 34$$

You can get a new sequence of data

$$x^{(1)} = \{6, 9, 17, 27, 34\}$$

The above formula can be $x^{(1)}(i) = \left\{ \sum_{j=1}^i x^{(0)}(j) \mid i = 1, 2, \dots, N \right\}$ summed up as,

then we can call the sequence represented in this data an accumulation generation of the original data sequence, and we can also call it an accumulation $x^{(1)}(1) = x^{(0)}(1)$ generation, obviously. Therefore, it can be assumed that an exponential curve or even a straight line is used to approximate this accumulated generated $x^{(1)}$ sequence. In order to return this accumulated sequence to the original sequence, it is necessary to carry out post-subtraction operation, which can also be called subtraction generation, that is, the difference between the two data before and after, as in the above example

$$\Delta x^{(1)}(5) = x^{(1)}(5) - x^{(1)}(4) = 34 - 27 = 7$$

$$\Delta x^{(1)}(4) = x^{(1)}(4) - x^{(1)}(3) = 27 - 17 = 10$$

$$\Delta x^{(1)}(3) = x^{(1)}(3) - x^{(1)}(2) = 17 - 9 = 8$$

$$\Delta x^{(1)}(2) = x^{(1)}(2) - x^{(1)}(1) = 9 - 6 = 3$$

$$\Delta x^{(1)}(1) = x^{(1)}(1) - x^{(1)}(0) = 6 - 0 = 6$$

Summarizing the above formulas, we can get the following results:

$$\Delta x^{(1)}(i) = x^{(1)}(i) - x^{(1)}(i-1) = x^{(0)}(i) \quad (1.4)$$

Among them $i = 1, 2, \dots, N$, $x^{(0)}(0) = 0$.

Modeling principle of grey prediction model.

Give a column of observation data $x^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(N)\}$ Add it up once and get it. $x^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(N)\}$ Let $x^{(1)}$ the first order differential equation be $\frac{dx^{(1)}}{dt} + ax^{(1)} = u$ satisfied, and this equation also satisfies

the initial conditions when $t = t_0$, $x^{(1)} = x^{(1)}(t_0)$ The solution of is

$$x^{(1)}(t) = \left[x^{(1)}(t_0) - \frac{u}{a} \right] e^{-a(t-t_0)} + \frac{u}{a}.$$

The method of grey modeling is to estimate the constants and by using the least square method through a a cumulative u sequence. Because $x^{(1)}(1)$ they are used as initial values, they are $x^{(1)}(2), x^{(1)}(3), \dots, x^{(1)}(N)$ all brought into the equation separately, $x^{(1)} = x^{(1)}(t_0)$ and the difference is used instead of the differential, and because they are sampled at equal $\Delta t = (t+1) - t = 1$ intervals, they can be obtained

$$\frac{\Delta x^{(1)}(2)}{\Delta t} = \Delta x^{(1)}(2) = x^{(1)}(2) - x^{(1)}(1) = x^{(0)}(2) \quad (1.5)$$

So there are similar $\frac{\Delta x^{(1)}(3)}{\Delta t} = x^{(0)}(3), \dots, \frac{\Delta x^{(1)}(N)}{\Delta t} = x^{(0)}(N)$ ones,

The $x^{(i)} = x^{(i)}(t_0)$ term $\begin{cases} x^{(0)}(2) + ax^{(1)}(2) = u \\ x^{(0)}(3) + ax^{(1)}(3) = u \\ \vdots \\ x^{(0)}(N) + ax^{(1)}(N) = u \end{cases}$ is then $ax^{(i)}(i)$ moved to

the right and written as the quantity product of the vector because it

$$\begin{cases} x^{(0)}(2) = [-x^{(1)}(2), 1] \begin{bmatrix} a \\ u \end{bmatrix} \\ x^{(0)}(3) = [-x^{(1)}(3), 1] \begin{bmatrix} a \\ u \end{bmatrix} \\ \vdots \\ x^{(0)}(N) = [-x^{(1)}(N), 1] \begin{bmatrix} a \\ u \end{bmatrix} \end{cases} \text{ also } \frac{\Delta x^{(i)}}{\Delta t} \text{ involves accumulating the values } x^{(i)}$$

of the two times in the column, $x^{(i)}(i)$ so it is more appropriate to take the average of the two times before and after, that is, $x^{(i)}(i)$ it will $\frac{1}{2}[x^{(i)}(i) + x^{(i)}(i-1)]$, ($i = 2, 3, \dots, N$) be replaced by, so the formula in the form of the quantity product of the vector before will be rewritten into a matrix expression

$$\begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(N) \end{bmatrix} = \begin{bmatrix} -\frac{1}{2}[x^{(1)}(2) + x^{(1)}(1)] & 1 \\ -\frac{1}{2}[x^{(1)}(3) + x^{(1)}(2)] & 1 \\ \vdots & \vdots \\ -\frac{1}{2}[x^{(1)}(N) + x^{(1)}(N-1)] & 1 \end{bmatrix} \begin{bmatrix} a \\ u \end{bmatrix}, \tag{1.6}$$

Order, $y = (x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(N))^T$ where T means transposition, order

$$B = \begin{bmatrix} -\frac{1}{2}[x^{(1)}(2) + x^{(1)}(1)] & 1 \\ -\frac{1}{2}[x^{(1)}(3) + x^{(1)}(2)] & 1 \\ \vdots & \vdots \\ -\frac{1}{2}[x^{(1)}(N) + x^{(1)}(N-1)] & 1 \end{bmatrix} \text{ Then } U = \begin{bmatrix} a \\ u \end{bmatrix} \text{ the original matrix expres-}$$

sion can also be written as, $Y = BU$ it is estimated by the least square method

$$\hat{U} = \begin{bmatrix} \hat{a} \\ \hat{u} \end{bmatrix} = (B^T B)^{-1} B^T y \text{ as, and the } \hat{a} \text{ estimated } \hat{u} \text{ values}$$

$$\begin{cases} x^{(0)}(2) + ax^{(1)}(2) = u \\ x^{(0)}(3) + ax^{(1)}(3) = u \\ \vdots \\ x^{(0)}(N) + ax^{(1)}(N) = u \end{cases} \text{ are substituted into such a time response equation}$$

$$\hat{x}^{(i)}(k+1) = \left[x^{(i)}(1) - \frac{\hat{u}}{\hat{a}} \right] e^{-\hat{a}k} + \frac{\hat{u}}{\hat{a}}, \text{ when } k = 1, 2, \dots, N-1 \text{ What is calculated}$$

from the above $\hat{x}^{(1)}(k+1)$ formula is a fitting value; $k \geq N$ At $\hat{x}^{(1)}(k+1)$ that time, it was a predicted value, which was the fitted value $x^{(1)}$ relative to a cumulative sequence, and then restored by post-subtraction operation. At $k=1,2,\dots,N-1$ that time, the fitted $x^{(0)}$ value could $\hat{x}^{(0)}(k+1)$ be $k \geq N$ obtained; At that time, the forecast $x^{(0)}$ value of the original series can be obtained.

This chapter introduces the theoretical basis of this paper, introduces the concept of multiple linear regression model and its application examples, and introduces the concept of grey theory and the modeling basis of grey theory, as well as the simple modeling method and calculation method of grey theory [6] [7].

3. Prediction Model Based on Multiple Linear Regression and Grey Prediction

3.1. Grey Forecast and Its Advantages

3.1.1. GM (1, 1) Model

GM series models are the basic models in grey prediction theory, especially GM (1, 1) model is widely used. The model used in this paper is GM (1, 1) model. There are four basic forms of GM (1, 1) model, including mean GM (1, 1) model, original difference GM (1, 1) model, mean difference GM (1, 1) model and discrete GM (1, 1) model. The definitions of these four models are explained as follows:

There is $X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))$ a $x^{(0)}(k) \geq 0$, $k=1,2,\dots,n$ sequence, $X^{(1)}$ wherein; $X^{(0)}$ 1-AGO sequence of: $x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i)$, $k=1,2,\dots,n$ where; Weighing

$$x^{(0)}(k) + ax^{(1)}(k) = b \quad (2.1)$$

It is the original form of GM (1, 1) model, and the original form of GM (1, 1) model is essentially a difference equation. The parameter $\hat{a} = [a, b]^T$ vector in Equation (2.1) can also be estimated by least square method

$$\hat{a} = (B^T B)^{(-1)} B^T Y \quad (2.2)$$

where Y and B are respectively

$$Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}, \quad B = \begin{bmatrix} -x^{(1)}(2) & 1 \\ -x^{(1)}(3) & 1 \\ \vdots & \vdots \\ -x^{(1)}(n) & 1 \end{bmatrix} \quad (2.3)$$

Therefore, based on the original form of GM (1, 1) model and Formula (2.2), the parameters of the model are estimated, and the solution of the original difference Equation (2.1) is directly used as the time response formula, and the obtained model is called the original difference form of GM (1, 1) model [8].

Suppose, $X^{(0)}$ as $X^{(1)}$ shown in Equation (2.1)
 $Z^{(1)} = (z^{(1)}(2), z^{(1)}(3), \dots, z^{(1)}(n))$, where $Z^{(1)}(k) = \frac{1}{2}(x^{(1)}(k) + x^{(1)}(k+1))$,
 $k = 2, 3, \dots, n$ call

$$x^{(0)}(k) + az^{(1)}(k) = b \quad (2.4)$$

is in the form of the mean value of GM (1, 1) model. The mean value of GM (1, 1) model is essentially a difference equation. The parameter $\hat{a} = [a, b]^T$ vectors in Equation (2.4) can also be estimated using Equation (2.2), where it is important to note that some elements in Matrix B differ from Equation (2.3)

$$B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix} \quad (2.5)$$

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b \quad (2.6)$$

is called an albino differential equation in mean form of GM (1, 1) model. The matrix B in Equation (2.2) is replaced by Equation (2.5). The parameter vectors in Equation (2.6) are estimated according to the least square estimation method. $\hat{a} = [a, b]^T$ With the help of the solution of the above-mentioned whitening differential equation Equation (2.6), the mixed model of difference and differential of the time response equation of GM (1, 1) is constructed, which is called the mean mixed form of GM (1, 1) model, which can be called the mean GM (1, 1) model. The parameter of the mean GM (1, 1) $-a$ model is the development coefficient, and B is the grey action. The development $-a$ coefficient reflects $\hat{x}^{(1)}$ the $\hat{x}^{(0)}$ development trend of as well. The estimated model parameters in the form of mean value of GM (1, 1) model can be directly taken the solution of mean difference Equation (2.4) as the model obtained in the time response formula, which is called the mean difference form of GM (1, 1) model, that is, GM (1, 1) model called mean difference. The $x^{(1)}(k+1) = \beta_1 x^{(1)}(k) + \beta_2$ discrete form called GM (1, 1) model is called discrete GM (1, 1) model [5].

3.1.2. Advantages and Disadvantages of GM (1, 1) Model

GM (1, 1) model can only use the behavior data sequence of the system to establish the prediction model, which belongs to a relatively simple and practical single sequence modeling method. In the case of time data, it only involves some regular time variables, and in the case of horizontal sequence data, it only involves some regular object serial number variables, but does not involve other explanatory variables. It is simple and can discover some valuable development and change information at the same time, so it has been widely used. However, the practice of forecasting shows that GM (1, 1) model is sometimes very effective, sometimes it is very unsatisfactory, and sometimes it will lose its utility. In order to make up for this small deviation, sometimes some combination models can be used to make up for the deficiency that GM (1, 1) model cannot describe

some linear relationships among variables [9].

3.2. Combining Multiple Linear Regression with Grey Prediction

3.2.1. The Inadequacy of Multiple Linear Regression

Although multivariate linear regression is widely used, it also has its own shortcomings. When we can't determine what kind of function conditions the data meet, multivariate linear can't determine its specific index, and can't describe the exponential growth trend. This grey prediction can just complement it [10] [11].

3.2.2. Shortcomings of Grey Prediction

Grey prediction sometimes depends too much on historical data, and does not consider the relationship between various factors, which is too much error for medium and long-term prediction results.

3.2.3. Combining Multiple Linear Regression with Grey Prediction

The grey linear regression combination model can make up for the deficiency of the original linear regression model that can't describe the exponential growth trend and the GM (1) model that can't describe the linear relationship among variables, and can predict both the linear growth trend and the exponential growth trend, so the data analysis has higher accuracy.

3.3. Grey Linear Combination Model

Using the method of combining multiple linear regression model with grey forecasting GM (1, 1) model, this paper puts forward that grey linear regression combination model can be used for forecasting, that is, the sum between linear regression $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon$ equation and exponential $y = C_1 e^{-ak} + C_2$ equation can be used $X^{(0)}$, 1-AGO $X^{(1)}$ for fitting.

This chapter plays an important role in this paper, This chapter introduces the advantages and disadvantages of the traditional multiple linear regression model, and introduces the grey GM (1, 1) model, and some advantages and disadvantages of the grey GM (1, 1) model. After analysis, it puts forward a model combining the two models, which makes comprehensive use of their advantages and makes them complement each other at the same time, so as to achieve better prediction data. At the same time, this chapter also introduces the research significance of the new model proposed in this paper [12].

4. Application of Forecasting Model in National Economic Data

4.1. National Economic Index Data and Preprocessing

4.1.1. National Economic Indicators

The national economic indicators used in this paper come from the data of China Statistical Yearbook 2017 in recent ten years, that is, the data of gross national product from 2007 to 2016.

The data in **Figure 2** is the data that needs to be referenced in this article.

4.1.2. Sample Data Preprocessing

If we want to establish a linear relationship between the gross national product and other values, and hope to achieve the purpose of prediction, then we must establish a multiple linear relationship between the attributes of the current year and the gross national product of the next year.

For example:

$$\text{GDP in 2001} = \text{Constant quantity} + k_1 \cdot \text{Gross National Product Income in 2000} \\ + k_2 \cdot \text{Primary industry in 2000} + \dots + k_5 \cdot \text{GDP per capita}$$

By analogy, the linear relationship established in this way is the multiple linear relationship between the current year's gross national income, the primary industry, the secondary industry, the tertiary industry and the per capita GDP and the next year's GDP, that is to say, the current year's value is used to predict the next year's GDP. Import the data in the above table into SPSS as shown in **Figure 3**.

index	2016	2015	2014	2013	2012	2011	2010	2009	2008	2007
Gross National Income (100 million yuan)	74,598.70	686,449.60	644,791.10	590,422.40	539,116.50	484,753.20	411,265.20	348,498.50	321,500.50	270,844.00
GDP (100 million yuan)	743,585.50	689,052.10	643,974.00	595,244.40	540,367.40	489,300.60	413,030.30	349,081.40	319,515.50	270,232.30
Added value of primary industry (100 million yuan)	63,672.80	60,862.10	58,343.50	55,329.10	50,902.30	46,163.10	39,362.60	34,161.80	32,753.20	27,788.00
Added value of the secondary industry (100 million yuan)	296,547.70	282,040.30	277,571.80	261,956.10	244,643.30	227,038.80	191,629.80	160,171.70	149,956.60	126,633.60
Added value of tertiary industry (100 million yuan)	383,365.00	346,149.70	308,058.60	277,959.30	244,821.90	216,098.60	182,038.00	154,747.90	1,366,805.80	115,810.70
GDP per capita (yuan)	53,935.00	50,251.00	47,203.00	43,852.00	40,007.00	36,403.00	30,876.00	26,222.00	24,121.00	20,505.00

Note: The difference between gross national income and gross domestic product after 1980 is foreign net factor income

Figure 2. National economic data table 2007-2016.

	index	Gross National Income (100 million yuan)	GDP (100 million yuan)	Added value of primary industry (100 million yuan)	Added value of the secondary industry (100 million yuan)	Added value of tertiary industry (100 million yuan)	GDP per capita (yuan)
1	2016	74,598.70	743,585.50	63,672.80	296,547.70	383,365.00	53,935.00
2	2015	686,449.60	689,052.10	60,862.10	282,040.30	346,149.70	50,251.00
3	2014	644,791.10	643,974.00	58,343.50	277,571.80	308,058.60	47,203.00
4	2013	590,422.40	595,244.40	55,329.10	261,956.10	277,959.30	43,852.00
5	2012	539,116.50	540,367.40	50,902.30	244,643.30	244,821.90	40,007.00
6	2011	484,753.20	489,300.60	46,163.10	227,038.80	216,098.60	36,403.00
7	2010	411,265.20	413,030.30	39,362.60	191,629.80	182,038.00	30,876.00
8	2009	348,498.50	349,081.40	34,161.80	160,171.70	154,747.90	26,222.00
9	2008	321,500.50	319,515.50	32,753.20	149,956.60	1,366,805.80	24,121.00
10	2007	270,844.00	270,232.30	27,788.00	126,633.60	115,810.70	20,505.00

Figure 3. SPSS national economic data chart.

4.2. Multiple Linear Regression Forecasting Model Predicts National Economy

Import the data used into SPSS and then do the following (Figure 4).

Introduce “next year’s gross national product” into dependent variables, and “GDP, primary industry, secondary industry, tertiary industry and per capita GDP” into independent variables. By using the entry method, the multiple linear relationship between them is obtained. The specific operation method is shown in the following figure (Figure 5).

The results are obtained as in Tables 4-7.

4.3. Grey Forecasting Model Predicts National Economy

Establish a grey prediction model to set a $X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))$ sequence, which; $x^{(0)}(k) \geq 0, k = 1, 2, \dots, n$ 1-AGO $X^{(1)}$ sequence $X^{(0)}$ of: where; then the $x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i), k = 1, 2, \dots, n$ parameter $x^{(0)}(k) + ax^{(1)}(k) = b$ vector is $\hat{a} = [a, b]^T$ estimated by the least square $\hat{a} = (B^T B)^{-1} B^T Y$ method, where Y and B are

$$Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}, \quad B = \begin{bmatrix} -x^{(1)}(2) & 1 \\ -x^{(1)}(3) & 1 \\ \vdots & \vdots \\ -x^{(1)}(n) & 1 \end{bmatrix}$$

Import the model data into MATLAB, and get the following graphics (Figure 6).

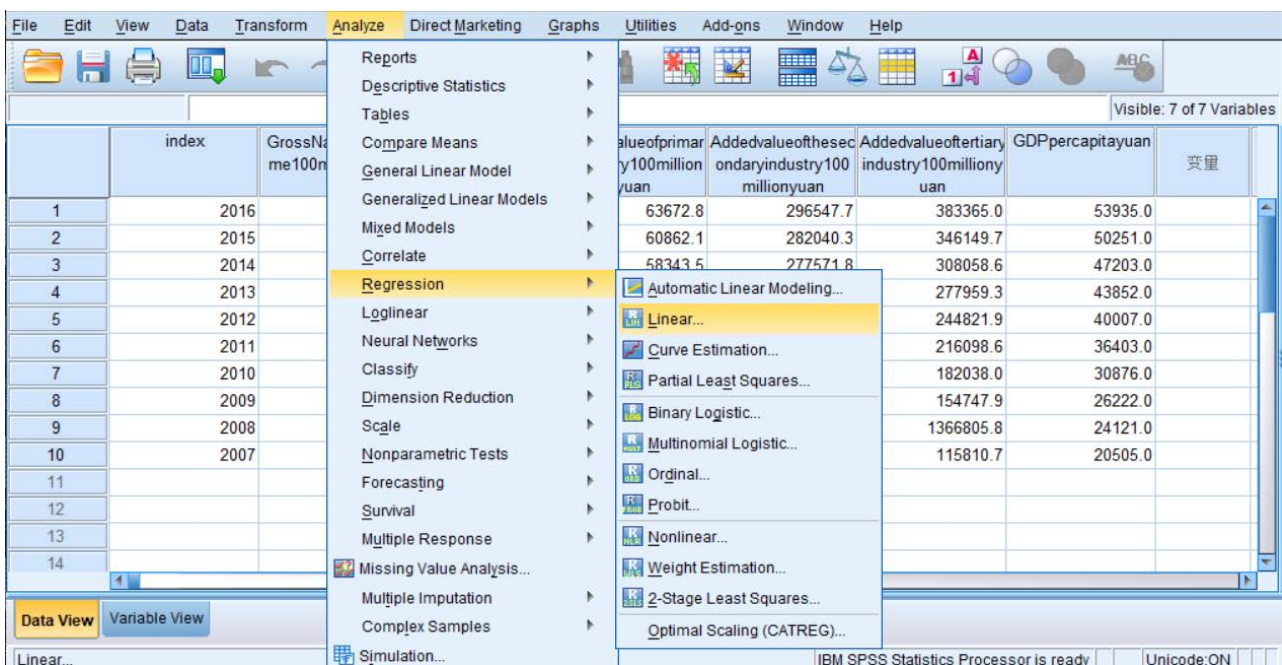


Figure 4. SPSS linear regression operation.

Table 4. Variables entered into table.

Variable entered ^A removed			
Model	Variable entered	Variable removed	Method
1	Per capita GDP (yuan), added value of secondary industry (100 million yuan), added value of primary industry (100 million yuan), gross national income (100 million yuan) [b]	.	Enter

^AVariable: Gross national product for the next year (100 million yuan); ^BHas reached the tolerance = 0.000 limit.

Table 5. Model summary table.

Model summary				
Model	R	R-squared	[Adjusted] [R Square]	Standard skewness error
1	0.999 ^A	0.998	0.996	10,182.58580

^AValue: (constant), per capita GDP (yuan), added value of secondary industry (100 million yuan), added value of primary industry (100 million yuan), gross national income (100 million yuan).

Table 6. Coefficient table.

Coefficient ^A					
Model	Non-normalized coefficient		Normalization coefficient	T	Significance
	B	Standard error	Beta		
(Constant number)	103,420.523	35,875.638		2.883	0.034
Gross national income (100 million yuan)	-1.165	1.751	-1.118	-0.665	0.535
1 Added value of primary industry (100 million yuan)	-21.677	6.912	-1.637	-3.136	0.026
Added value of secondary industry (100 million yuan)	0.612	1.108	0.220	0.553	0.604
Per capita GDP (yuan)	51.591	27.283	3.529	1.891	0.117

^ANumber of changes: Gross National Product of the next year (100 million yuan).

Table 7. Excluded variables table.

Exclude variables						
Model	Beta in	T	Significance	Partial correlation	Collinear statistical data	
					Tolerance	
1 Added value of tertiary industry (100 million yuan)	-2.091 ^b	-0.299	0.780	-0.148	1.002E-5	

^AVariable: Gross national product for the next year (100 million yuan); ^BForecast value in the model: (constant), per capita GDP (yuan), added value of secondary industry (100 million yuan), added value of primary industry (100 million yuan), gross national income (100 million yuan).

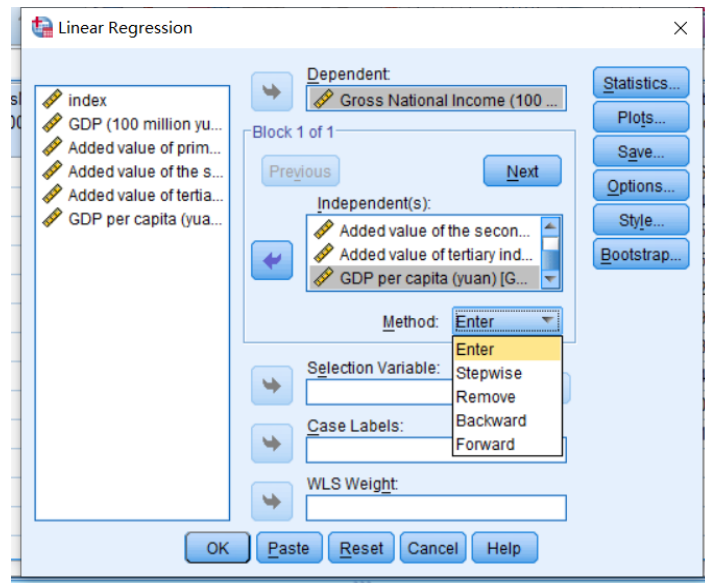


Figure 5. Spss linear regression operation.

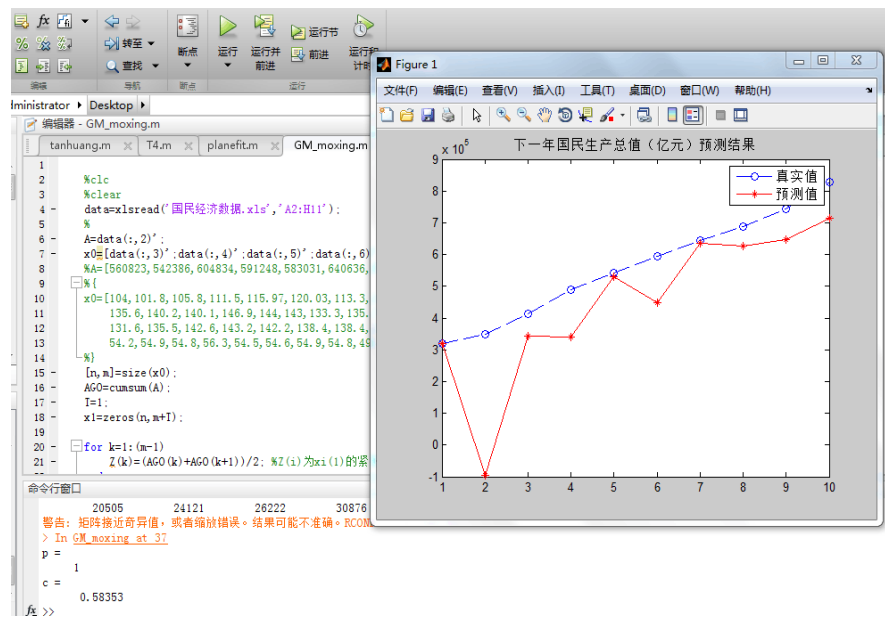


Figure 6. MATLAB analysis chart.

4.4. Grey Linear Combination Model Predicts National Economy

Because there are many sequences of this model, the model of 2.3 is further improved here, and it is $X^{(0)} = \{x_1^0, x_2^0, \dots, x_n^0\}$ assumed that it is accumulated to $X^{(1)} = \{x_1^1, x_2^1, \dots, x_n^1\}$ obtain, in which $X_i^1 = \sum_{j=1}^i x_j^0$ It $\hat{x}_k^1 = C_1 e^{-vk} + C_2 k + C_3$ is called grey linear regression combination prediction model, where v and C_1 are C_2 undetermined C_3 parameters.

$$\text{Orde } z_k = \hat{x}_{k+1}^1 - \hat{x}_k^1 = C_1 e^{-vk} (e^v - 1) + C_2,$$

$$y_m(k) = z_{k+m} - z_k = C_1 e^{-vk} (e^{vm} - 1)(e^v - 1)$$

Then, $\frac{y_m(k+1)}{y_m(k)} = e^v$ from this, because $\hat{v}_m(k) = \ln\left(\frac{y_m(k+1)}{y_m(k)}\right)$ all $y_m(k)$

the data in are predicted \hat{x}_k^1 values, in order to facilitate v fitting \hat{x}_k^1 parameters, all of them are replaced by x_k^1 sequential accumulation sequences, so that the fitting v values of $\hat{v}_m(k)$ parameters to be m identified can be determined, Considering that different values are taken, $\hat{v}_m(k)$ the values of the obtained fitting parameters will be $m=1,2,\dots,n-3$ different, so all $\hat{v}_m(k)$ the obtained selected mean values $\hat{v}_m(k)$ can be selected and calculated separately, and the

finally obtained mean values can be taken v as the fitting $\tilde{V} = \frac{\sum_{m=1}^{n-3} \sum_{k=1}^{n-2-m} \hat{v}_m(k)}{(n-2)(n-3)/2}$

values of v the parameters \tilde{V} to be identified: When the fitting values of the parameters are determined, $\hat{x}_k^1 = C_1 e^{-\tilde{v}k} + C_2 k + C_3$ the grey linear combination

prediction model is as follows: $X^{(1)} = \begin{bmatrix} x_1^1 \\ x_2^1 \\ \vdots \\ x_n^1 \end{bmatrix}$, $C = \begin{bmatrix} C_1 \\ C_2 \\ C_3 \end{bmatrix}$, $A = \begin{bmatrix} e^{-\tilde{v}} & 1 & 1 \\ e^{-\tilde{v}} & 2 & 1 \\ \vdots & \vdots & \vdots \\ e^{-\tilde{v}} & n & 1 \end{bmatrix}$ Then,

the matrix $X^{(1)} = AC$ estimation formula of the parameters to be identified C_1, C_2, C_3 can be obtained, and the prediction of $C = (A^T A)^{-1} A^T X^{(1)}$ this model can be carried out by the above formula.

As shown in **Figure 7**, so

$$X^{(0)} = \{270232.3, 319515.5, 349081.4, 413030.3, 489300.6, 540367.4, 595244.4, 643974, 689052.1, 743585.5\}$$

Carry out an accumulation generation

$$X^{(1)} = \{270232.3, 589747.8, 938829.2, 1351859.5, 1841160.1, 2381527.5, 2976771.9, 3620745.9, 4309798, 5053384.5\}$$

Therefore, from the formula provided in this paper, we can see **Figure 8** and **Figure 9**.

Therefore, $\tilde{V} = -0.127$.

Therefore $C = \begin{bmatrix} 1.3433 \\ 2.4962 \\ -1.3839 \end{bmatrix}$, so $\hat{x}_k^1 = 1.3433e^{0.127k} + 2.4962k - 1.3839$,

$$\tilde{V} = \frac{\sum_{m=1}^{n-3} \sum_{k=1}^{n-2-m} \hat{v}_m(k)}{(n-2)(n-3)/2} \quad error = \frac{1}{10} \sum_{k=1}^{10} \frac{|x_k^1 - \hat{x}_k^1|}{x_k^1}$$

time	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
GDP (100 million yuan)	270,232.30	319,515.50	349,081.40	413,030.30	489,300.60	540,367.40	595,244.40	643,974.00	689,052.10	743,585.50

Figure 7. Gross domestic product 2007-2016.

$y_m(k)$	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$	$k = 8$
$m = 1$	29,565.9	63,948.9	76,270.3	51,066.8	54,877	48,729.6	45,078.1	54,533.4
$m = 2$	93,514.8	140,219.2	127,337.1	105,943.8	103,606.6	93,807.4	99,611.5	
$m = 3$	169,785.1	191,286	182,214.1	154,673.4	148,684.7	148,341.1		
$m = 4$	220,851.9	246,163	230,943.7	199,751.5	203,218.1			
$m = 5$	275,728.9	294,892.6	276,021.8	254,284.9				
$m = 6$	324,458.5	339,970.7	330,555.2					
$m = 7$	369,536.6	394,504.1						

Figure 8. $y_m(k)$ and the relationship between m and k .

$\hat{y}_m(k)$	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$
$m = 1$	-1.538	0.176	-0.401	0.072	-0.119	-0.078	0.190
$m = 2$	0.405	-0.096	-0.184	-0.022	-0.099	0.060	
$m = 3$	0.119	-0.049	-0.164	-0.039	-2.314		
$m = 4$	0.109	-0.064	-0.145	0.017			
$m = 5$	0.067	-0.066	-0.082				
$m = 6$	0.047	-0.028					
$m = 7$	0.065						

Figure 9. $\hat{y}_m(k)$ and the relationship between m and k .

4.5. Comparison of Results

From several results, the prediction model used in this paper is more accurate and closer to the predicted value, which better shows the advantages of prediction, and combines the advantages of linear regression with the advantages of grey prediction.

This chapter is the practical part of the data in this paper, using the three models mentioned in this paper to fit the data respectively, and comparing the results obtained, it is found that the model used in this paper is more accurate.

5. Conclusion

In this article, based on the idea of combining linear regression with grey prediction, determines the model of combining the two, and compares it with the separate models of the two. The linear regression model solves the problem that the highest power and function cannot be determined by linear prediction. At the same time, it solves the problem that the grey prediction model can't make linear prediction. By combining the two models, it can be used in the series prediction of various existing linear trends and exponential growth trends in the future, which not only has extensive significance, but also can make better prediction

for data.

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] He, X. and Liu, W. (2015) Applied Regression Analysis. 4th Edition, Renmin University of China Press, Beijing.
- [2] Wang, H. and Meng, J. (2007) Predictive Modeling Method of Multiple Linear Regression. *Journal of Beihang University*, **33**, 500-504.
- [3] Zhong, L. and Gao, S. (2017) Application of Multiple Linear Regression Model in Housing Price Trend and Forecast. *Science Entrepreneurship Monthly*, No. 9, 94-96.
- [4] Leng, J., Gao, X. and Zhu, J. (2016) Application of Multiple Linear Regression Statistical Prediction Model. *Statistics and Decision-Making*, No. 7, 82-85.
- [5] Liu, S.F., *et al.* (2017) Grey System Theory and Its Application (8th Edition). Science Press, Beijing.
- [6] Wu, S. (2009) Grey Rough Set Model and Its Application. Science Press, Beijing.
- [7] Deng, J. (2007) Grey Mathematical Resources] Introduction to Science. Huazhong University of Science and Technology Press, Wuhan.
- [8] Lu, Y. (2017) Financial Forecasting Method Based on Grey Linear Combination *Statistics and Decision*, No. 10, 91-93.
- [9] Wang, B. (2012) Forecasting Application Based on Linear Regression and Neural Network in National Economic Data. Jilin University, Changchun.
- [10] Fisher, R.A. (1936) The Use of Multiple Measures in Taxonomic Problems. *Annals of Eugenics*, **7**, 179-188. <https://doi.org/10.1111/j.1469-1809.1936.tb02137.x>
- [11] Bagirov, A.M., Julien, U. and Dean, W. (2011) Fast Modified Global k-Means Algorithm for Incremental Cluster Construction. *Pattern Recognition*, **44**, 866-876. <https://doi.org/10.1016/j.patcog.2010.10.018>
- [12] Zheng, C.L., Fan, J. and Fei, M.R. (2009) PID Neural Network Control Research Based on Fuzzy Neural Network Model. 2009 *International Conference on Computational Intelligence and Software Engineering*, Wuhan, 11-13 December 2009, 1-4. <https://doi.org/10.1109/CISE.2009.5363758>