

# Regression Mathematical Model of Impact Factors for Social Science Journals Based on Information Reused Analysis Method

Xuexia Wei\*, Xiao Dong, Shiqiang Zhang

Medical Information College, Chongqing Medical University, Chongqing, China  
Email: \*842086448@qq.com

Received 12 January 2016; accepted 26 February 2016; published 1 March 2016

---

## Abstract

A regression mathematical model of impact factors for social science journals was given based on information reused analysis method. This regression mathematical model could evaluate and predict change tendency of the impact factor of the social science journals.

## Keywords

Impact Factor, Regression, Mathematical Model, Social Science

---

## 1. Introduction

In the era of big data, it is important to find out the information hiding behind the data. Journal Citation Reports [1] published by the institute for scientific information is known as the most authoritative journal evaluation system. According to journal citation report published by ISI, we can get the data of impact factors for social science journals. Based on these data, a regression mathematical model of impact factors for social science journals was given based on information reused analysis method [2].

## 2. Information Reused Analysis Method

Hypothesis  $t = \{t_1, t_2, \dots, t_n\}$  indicates time series, the original data sequence corresponding to the time series is  $x^{(0)} = \{x^{(0)}(t_1), x^{(0)}(t_2), \dots, x^{(0)}(t_n)\}$ . The differential operation of time series  $t = \{t_1, t_2, \dots, t_n\}$  is  $\Delta t_k = t_k - t_{k-1}$ , when differential  $\Delta t_k = \text{const}$ , the original data sequence  $x^{(0)} = \{x^{(0)}(t_1), x^{(0)}(t_2), \dots, x^{(0)}(t_n)\}$  is equal-space sequence. When differential  $\Delta t_k \neq \text{const}$ , the original data sequence  $x^{(0)} = \{x^{(0)}(t_1), x^{(0)}(t_2), \dots, x^{(0)}(t_n)\}$  is non-equal-space sequence. According to the grey system theory [3] One-accumulated generate sequence of the original data sequence  $x^{(0)} = \{x^{(0)}(t_1), x^{(0)}(t_2), \dots, x^{(0)}(t_n)\}$  is  $x^{(1)} = \{x^{(1)}(t_1), x^{(1)}(t_2), \dots, x^{(1)}(t_n)\}$ , where in

$$\begin{cases} x^{(1)}(t_1) = x^{(0)}(t_1) \\ x^{(1)}(t_k) = x^{(1)}(t_1) + \sum_{i=2}^k x^{(0)}(t_i) \Delta t_i \\ k = 2, 3, \dots, n \end{cases} \quad (1)$$

---

\*Corresponding author.

When sequence  $x^{(1)} = \{x^{(1)}(t_1), x^{(1)}(t_2), \dots, x^{(1)}(t_n)\}$  was close to nonhomogeneous exponential law change, the solution  $x^{(1)}(t)$  of differential Equation (2) was called the response function.

$$\begin{cases} \frac{dx^{(1)}(t)}{dt} + ax^{(1)}(t) = b \\ x^{(1)}(t_1) = x^{(0)}(t_1) \end{cases} \quad (2)$$

This response function  $x^{(1)}(t)$  was

$$x^{(1)}(t) = (x^{(0)}(t_1) - \frac{b}{a})e^{-a(t-t_1)} + \frac{b}{a}.$$

In response function  $x^{(1)}(t)$ , the constants  $a$  and  $b$  were called uncertain parameters. Discrete response function  $x^{(1)}(t)$  of the differential Equation (2) was

$$\begin{cases} x^{(1)}(t_1) = x^{(0)}(t_1) \\ x^{(1)}(t_k) = (x^{(0)}(t_1) - \frac{b}{a})e^{-a(t_k-t_1)} + \frac{b}{a} \\ k = 2, 3, \dots, n \end{cases} \quad (3)$$

In response function  $x^{(1)}(t)$ , for determining constants  $a$  and  $b$ , we could use differential operation of difference Equation (2). That is

$$\begin{cases} \frac{\Delta x^{(1)}(t_k)}{\Delta t_k} + ax^{(1)}(t_k) = b \\ x^{(1)}(t_1) = x^{(0)}(t_1) \end{cases} \quad (4)$$

In differential operation (4), we could get

$$\frac{\Delta x^{(1)}(t_k)}{\Delta t_k} = \frac{x^{(1)}(t_k) - x^{(1)}(t_{k-1})}{t_k - t_{k-1}} = x^{(0)}(t_k).$$

In differential operation (4), let  $\lambda \in [0, 1]$ , smoothing  $x^{(1)}(t_k)$  of differential Equation (4) by formula  $z^{(1)}(t_k) = \lambda x^{(1)}(t_k) + (1 - \lambda)x^{(1)}(t_{k-1})$ , we could get differential equation as follow

$$\begin{cases} \frac{\Delta z^{(1)}(t_k)}{\Delta t_k} + az^{(1)}(t_k) = b \\ z^{(1)}(t_1) = x^{(0)}(t_1) \end{cases} \quad (5)$$

In above differential Equation (5),  $\lambda$  was called as background parameters and  $z^{(1)}(t_k)$  was called as background value. At present, there is still no optimum getter for background parameters  $\lambda$ , in order to be used simply and easily, we generally take background parameters for 1/2 in reference [3].

Substituting one-accumulated generate sequence of original sequence  $x^{(0)} = \{x^{(0)}(t_1), x^{(0)}(t_2), \dots, x^{(0)}(t_n)\}$  into above differential equation, with matrix equation  $a$  and  $b$  could be  $[a \ b]^T = (B^T B)^{-1} B^T Y$  determined, inside  $Y = [x^{(0)}(t_2), x^{(0)}(t_3), \dots, x^{(0)}(t_n)]^T$ , and

$$B^T = \begin{bmatrix} -z^{(1)}(t_2) & -z^{(1)}(t_3) & \dots & -z^{(1)}(t_n) \\ 1 & 1 & \dots & 1 \end{bmatrix}.$$

Substituting obtained parameters  $a$  and  $b$  into differential operation (4), we could get traditional GM(1.1) model of sequence  $x^{(0)} = \{x^{(0)}(t_1), x^{(0)}(t_2), \dots, x^{(0)}(t_n)\}$ :

$$\begin{cases} \hat{x}^{(0)}(t_1) = x^{(0)}(t_1) \\ \hat{x}^{(0)}(t_k) = \frac{\hat{x}^{(1)}(t_k) - \hat{x}^{(1)}(t_{k-1})}{t_k - t_{k-1}} \\ \hat{x}^{(1)}(t_k) = (x^{(0)}(t_1) - \frac{b}{a})e^{-a(t_k-t_1)} + \frac{b}{a} \end{cases} \quad (6)$$

Fitting and forecast precision sometimes of the traditional GM(1.1) model was poor. We put forward a information reused analysis method. Firstly, we could call traditional GM (1.1) model (6) of original sequence  $x^{(0)}$

$\{x^{(0)}(t_1), x^{(0)}(t_2), \dots, x^{(0)}(t_n)\}$  as rough model. Then we rewrote third formulas of rough model (6) as follow, wherein,  $a$  and  $b$  were new uncertain parameters.

$$\hat{x}^{(1)}(t_k) = \alpha e^{-a(t_k-t_1)} + \beta \quad k = 2, 3, \dots, n \tag{7}$$

Third, using the accumulated generate sequence and corresponding time series of original sequence again, substituting the accumulated generate sequence  $x^{(1)} = \{x^{(1)}(t_1), x^{(1)}(t_2), \dots, x^{(1)}(t_n)\}$  and corresponding time series of original sequence  $t = \{t_1, t_2, \dots, t_n\}$  into above formula (7), we could determine uncertain parameters  $\alpha$  and  $\beta$  with matrix equation [4]  $[\alpha \beta]^T = (B^T B)^{-1} B^T Y$ , inside  $Y = [x^{(1)}(t_1), x^{(1)}(t_2), \dots, x^{(1)}(t_n)]^T$  and

$$B^T = \begin{bmatrix} 1 & e^{-a(t_2-t_1)} & \dots & e^{-a(t_n-t_1)} \\ 1 & 1 & \dots & 1 \end{bmatrix}.$$

Substituting parameters  $\alpha$  and  $\beta$  into Equation (7), we could get the reducing value of original sequence  $x^{(0)} = \{x^{(0)}(t_1), x^{(0)}(t_2), \dots, x^{(0)}(t_n)\}$

$$\begin{cases} \hat{x}^{(1)}(t_k) = \alpha e^{-a(t_k-t_1)} + \beta & (k = 1, 2, \dots, n) \\ \hat{x}^{(0)}(t_k) = \frac{\hat{x}^{(1)}(t_k) - \hat{x}^{(1)}(t_{k-1})}{t_k - t_{k-1}} & (k = 2, 3, \dots, n) \end{cases} \tag{8}$$

The formula (8) was the regression mathematical model based on information reused analysis method.

### 3. Regression Mathematical Model of Impact Factors for Social Science Journals

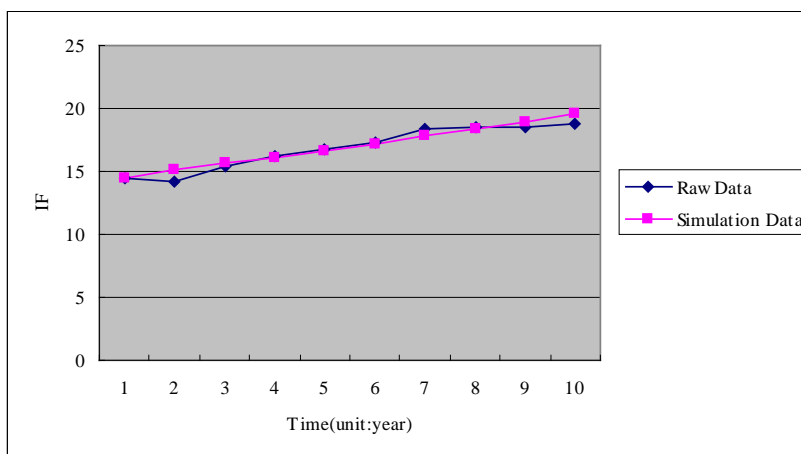
According to the average impact factor (IF) data [1] of the top 20 social science journals, that is  $\{(2004, 14.47655), (2005, 14.2287), (2006, 15.3813), (2007, 16.2098), (2008, 16.7604), (2009, 17.3631), (2010, 18.34185), (2011, 18.5654), (2012, 18.53725), (2013, 18.7618)\}$ , the regression mathematical model based on information reused analysis method as follow

$$\hat{x}^{(1)}(t_k) = 460.835e^{-0.03229(t_k-t_1)} - 447.04 \quad (k = 1, 2, \dots, \text{and } n)$$

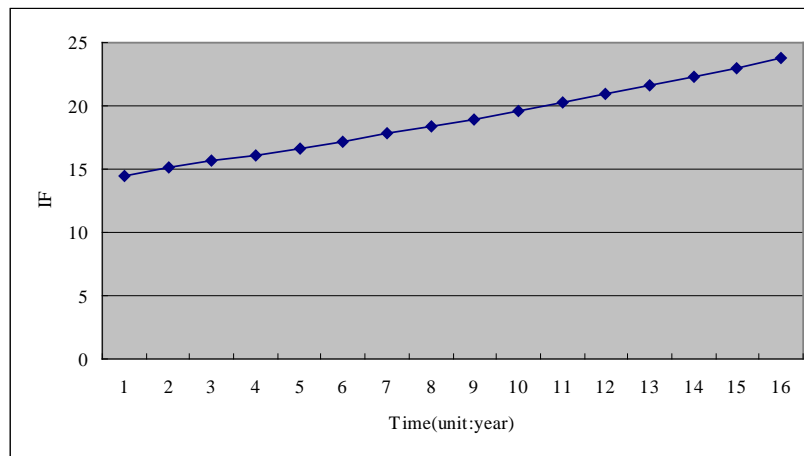
**Figure 1** is a diagram of the curve of original data for social science journals and its simulation data based on information reused analysis method from 2004 to 2013.

**Figure 2** is a diagram of the curve of simulation data of the average impact factor (IF) for social science journals form 2004 to 2019.

From 2004 to 2019 the simulation data of the average impact factor (IF) for social science journals over time trend graph can be seen that the number of the average impact factor (IF) for social science journals is gradually slowly increase trend. The simulation data of the average impact factor (IF) for social science journals form 2015 to 2019 could be forecasted as follow:  $\{(2015, 20.88736246), (2016, 21.57282944), (2017, 22.2807916), (2018, 23.01198718), (2019, 23.76717862)\}$ .



**Figure 1.** The tendency curve of the average impact factor with time.



**Figure 2.** The curve of simulation data of the average impact factor (IF) form 2004 to 2019.

## 4. Conclusions

This paper gives a modeling way based on information reused analysis method. On the one hand, this way based on information reused analysis method greatly improved GM (1,1) model's fitting precision and prediction accuracy; on the other hand, it maintains the advantage of the traditional modeling method which is simple. The model based on information reused analysis method evaluates and predicts the average impact factor (IF) trend of the top 20 social science journals. Case analysis verified the validity and usefulness of the model based on information reused analysis method. The way based on information reused analysis method can provide reference for social science researchers.

## References

- [1] Institute Science Information (ISI). Journal Citation Reports, JCR. <http://admin.isiknowledge.com/JCR>
- [2] Zhang, S.Q. (2009) Modeling Method of Grey System GM(1.1) Model Based Information Reused and It's Application. *Mathematics in Practice and Theory*, **39**, 97-104.
- [3] Deng, J.L. (1982) Control Problems of Grey Systems. *Systems & Control Letters*, **1**, 288-294. [http://dx.doi.org/10.1016/S0167-6911\(82\)80025-X](http://dx.doi.org/10.1016/S0167-6911(82)80025-X)
- [4] Zhang, S.Q. and Jiang, Z. (2010) Modeling Method of Logistic Model Based Grey System. *Mathematics in Practice and Theory*, **40**, 144-148.