

## Analysis of the Influencing Factors of Yiwu Cross-Border E-Commerce Transaction Volume Based on Grey Model

Wenbo Peng<sup>1,\*</sup>  
Xuewen Gui<sup>2</sup>  
Yanling Xu<sup>3</sup>  
Ji Luo<sup>4</sup>

\*Corresponding author

<sup>1</sup> Business School, Wuchang University of Technology, Wuhan 430223, Hubei, China. 5861598@qq.com

<sup>2</sup> School of Information Management, Central China Normal University, Hubei, China. 843908025@qq.com

<sup>3</sup> Business School, Wuchang University of Technology, Wuhan 430223, Hubei, China. 21980217@qq.com

<sup>4</sup> Business School, Wuchang University of Technology, Wuhan 430223, Hubei, China. sashagee@163.com

### ABSTRACT

Through the case of Yiwu Comprehensive Pilot Zone, the key factors affecting the development of the city's cross-border e-commerce industry are discussed, and the future cross-border e-commerce transaction volume is predicted. The Pearson coefficient was used to analyze the correlation of the feature data, and the key features were selected by combining the Lasso regression model. Build a grey prediction model to make predictions about future data for selected key features. An SVR vector regression model is established to predict the cross-border e-commerce transaction volume in the next two years. Sorting out the key elements to promote the increase of cross-border e-commerce transaction volume, and promote the high-quality and rapid development of the industry.

**Keywords:** Cross-border e-commerce, industrial competitiveness, gray model.

### INTRODUCTION

Today, cross-border e-commerce, characterized by the digitalization of trade, is becoming an important driving force for the growth of global trade. At present, China has approved the establishment of 165 cross-border e-commerce comprehensive pilot zones. Scholars use methods such as the grey correlation method and the diamond theory model (Bi Lingyan et al., 2019) to measure the competitiveness level of the cross-border e-commerce industry. Some scholars have also discussed the competitiveness level of regional cross-border e-commerce industry and the gap between regions (Wu Jinghong et al., 2022). In terms of specific methods, some scholars have used the grey correlation analysis method to obtain the grey correlation degree and weight of each influencing factor index based on the statistical data of Ningbo over the years (Wu et al., 2022), and tried to construct an index system of influencing factors for the fluctuation of cross-border e-commerce transaction volume in Ningbo (Chen Xiaoxing et al., 2022). In addition, the application of Lasso regression and SVR support vector regression can also provide new ideas for the analysis of cross-border e-commerce observation indicators (E. De Vito et al., 2011).

Considering the systematization of the index system, the integrity of the concept data and the distinction between import and export of cross-border e-commerce transactions, there is still a lot of room for optimization of the above-mentioned analysis models and prediction methods. Taking Yiwu as an example, this paper analyzes the important influencing factors of cross-border e-commerce transactions.

### INDEX SYSTEM AND DATA COLLECTION

#### Indicator System based on Diamond Theory

Considering the typicality of the sample, this paper selects the cross-border e-commerce transaction volume of Yiwu City as the analysis object. Porter's diamond model (Robert Huggins et al., 2015) is introduced, combined with the "motivation-trust-vulnerability (MTV) framework" in the cross-border scenario (Wagner Gerhard et al., 2023), and the indicators are mainly constructed from the dimensions of production factors, market demand, industrial agglomeration, organizational competition and cooperation, and policy environment, as shown in Table 1.

Table 1 Indicators related to cross-border e-commerce transaction value

Level 1 indicators	Level 2 indicators	Data source
Cross-border e-commerce transaction value	Cross-border e-commerce transaction value (100 million yuan)	Yiwu Statistical Bulletin
Factor of production	x1 Number of China-Europe trains (columns)	China-Europe Express website
	x2 Number of boxes sent by China-Europe trains (boxes)	China-Europe Express website
	x3 Freight Traffic: Highway (10,000 tons)	Jinhua Statistical Yearbook
	x4 Total Cargo Traffic: Railway (10 000 tons)	Jinhua Statistical Yearbook
	x5 Total Cargo Traffic: Air (10,000 tons)	Jinhua Statistical Yearbook
	x6 Trade Freight Volume (10,000 tons)	Yiwu Statistical Bulletin
	x7 Trade Freight Volume: Road Freight Volume (10,000 tons)	Yiwu Statistical Bulletin
	x8 Trade Freight Volume: Railway Arrivals and Departures (10,000 tons)	Yiwu Statistical Bulletin
	x9 Trade Cargo Volume: Air Cargo Volume (10,000 tons)	Yiwu Statistical Bulletin
	x10 Permanent population of Yiwu City (10,000 people)	Yiwu Statistical Bulletin
	x11 Registered population of Yiwu City (10,000 people)	Yiwu Statistical Bulletin
	x12 Education expenditure (10,000 yuan)	Yiwu Statistical Bulletin
	x13 New college students (10,000 people)	Yiwu Statistical Bulletin
	x14 Total GDP (100 million yuan)	Yiwu Statistical Bulletin
	x15 Value-added of tertiary industry (100 million yuan)	Yiwu Statistical Bulletin
	x16 Regional R&D investment (100 million yuan)	Yiwu Statistical Bulletin
Market demand	x17 Total Retail Sales of Consumer Goods (100 million yuan)	Yiwu Statistical Bulletin
	x18 Yiwu China Small Commodity Prosperity Index	China Commodity Index website
	x19 Turnover of China Commodity City (100 million yuan)	Yiwu Statistical Bulletin
	x20 Yiwu China Small Commodity Price Index	<a href="http://www.eastmoney.com">http://www.eastmoney.com</a>
	x21 Total Imports and Exports (100 million yuan)	Yiwu Statistical Bulletin
	x22 Imports (100 million yuan)	Yiwu Statistical Bulletin
	x23 Exports (100 million yuan)	Yiwu Statistical Bulletin
	x24 E-commerce Transaction Value (100 million yuan)	Yiwu Statistical Bulletin
	x25 Mobile Internet Users	Yiwu Statistical Bulletin
	x26 Proportion of cross-border e-commerce in total e-commerce (%)	Calculated based on indicators
	x27 Per Capita Disposable Income of Urban Permanent Residents (yuan)	Yiwu Statistical Bulletin
Industry agglomeration	x28 International Mail Exchange Bureau Business Volume (10,000 pieces)	Yiwu Statistical Bulletin
	x29 Express business volume (100 million pieces)	Postal Industry Statistical Bulletin
	x30 Express business revenue (100 million yuan)	National Post Office
	x31 Exports by Market Procurement Trade (100 million yuan)	Yiwu Statistical Bulletin
	x32 Science expenditure (10,000 yuan)	Yiwu Statistical Bulletin
Organize competitive cooperation	x33 Number of patent applications granted	Yiwu Statistical Bulletin
	x34 Foreign-related economic entities	Yiwu Statistical Bulletin
	x35 Entry and exit of foreign businessmen	Yiwu Statistical Bulletin
Policy environment	X36 Yiwu Market Credit Composite Index (YMCI)	<a href="http://www.ywindex.com">http://www.ywindex.com</a>

*Source:* According to the key indicators of the diamond model, combined with the data of Yiwu Municipal Bureau of Statistics and the Ministry of Commerce's "Yiwu Index Network".

### Data Collection

Data collection channels include: from Yiwu (Jinhua) Bureau of Statistics, China Small Commodities Index website, Ministry of Commerce, State Post Bureau, Oriental Fortune Network and other platforms. The data were collected according to the

following conditions: (1) the completeness of the time series, focusing on the collection of characteristic values with relatively complete data from 2018 to 2022. (2) The interpolation method was used to supplement the indicators such as the number of new college students in 2018 and 2019.

### Research Methods and Data Processing

Data analysis was performed in Anaconda's Jupiter Notebook environment through Python's numpy and pandas libraries. (1) The Pearson coefficient method and the Lasso feature selection model were used to exploratory analysis of the correlation of various features, and the correlation coefficient matrix was obtained. Get important characteristics related to cross-border e-commerce transaction value. (2) The grey prediction model and support vector regression prediction model for a single feature were established, and the accuracy was evaluated. (3) The support vector regression prediction model is used to calculate the total amount of Yiwu cross-border e-commerce in the next two years. (4) Evaluate the results of the prediction model.

### Data Preprocessing and Analysis

#### Feature Correlation Analysis

Combined with the existing data sources, this paper selects 36 impact characteristic indicators. The Pearson correlation coefficient was used for preliminary index screening. This is shown in Table 2.

Table 2 Pearson Correlation Coefficient Matrix

	y	x1	x2	x3	x4	x5	x6	x7	-	x34	x35	x36
y	1	0.99	0.99	0.97	0.84	-0.57	0.39	0.26	-	0.91	-0.9	0.93
x1	0.99	1	1	0.99	0.84	-0.6	0.31	0.17	-	0.9	-0.92	0.96
x2	0.99	1	1	0.99	0.84	-0.6	0.3	0.16	-	0.9	-0.92	0.96
x3	0.97	0.99	0.99	1	0.84	-0.62	0.19	0.04	-	0.87	-0.93	0.97
x4	0.84	0.84	0.84	0.84	1	-0.52	0.44	0.32	-	0.86	-0.59	0.91
x5	-0.57	-0.6	-0.6	-0.62	-0.52	1	-0.04	0.06	-	-0.22	0.54	-0.51
x6	0.39	0.31	0.3	0.19	0.44	-0.04	1	0.99	-	0.45	0	0.2
x7	0.26	0.17	0.16	0.04	0.32	0.06	0.99	1	-	0.33	0.14	0.06
-	-	-	-	-	-	-	-	-	-	-	-	-
x34	0.91	0.9	0.9	0.87	0.86	-0.22	0.45	0.33	-	1	-0.75	0.92
x35	-0.9	-0.92	-0.92	-0.93	-0.59	0.54	0	0.14	-	-0.75	1	-0.85
x36	0.93	0.96	0.96	0.97	0.91	-0.51	0.2	0.06	-	0.92	-0.85	1

Source: Calculated based on Pearson coefficient.

Through the table matrix analysis, the characteristics of strong correlation with cross-border e-commerce transaction volume are as follows: x1, x2, x3, x4, x8, x10, x11, x12, x13, x14, x15, x16, x17, x18, x19, x21, x22, x23, x24, x27, x29, x31, x32, 33, 34, 36. Characteristics of weak correlation are: x6, x7, x20, x26, x30. Less correlated features are: x5, x9, x20, x25, x28, x35.

#### Feature Selection Based on Lasso Regression

The Lasso method was first proposed by Robert Tibshiran in 1996, and its essence is to seek a process of sparse expression. While ensuring the best fit error, LASSO regression makes the parameters as "simple" as possible, making the model have strong generalization ability. On the basis of least squares method, LASSO regression takes the sum of the absolute values of

the regression coefficients  $\lambda \sum_{j=1}^p |\beta_j|$  as the penalty term, and adjusts the number and size of model parameters through the penalty term to reduce the complexity of the model. By compressing the coefficients of the variables in the regression model, the purpose of variable screening is achieved at the cost of controllable estimation bias.

The Lasso parameter estimate is shown in the formula:

$$\hat{\beta}(\text{lasso}) = \arg \min_{\beta} \left\| y - \sum_{j=1}^p \mathbf{X}_j \beta_j \right\|^2 + \lambda \sum_{j=1}^p |\beta_j| \quad (1)$$

With the increase of  $\lambda$ , the degree of compression of the estimated coefficients of each independent variable increases, and the coefficients of the independent variables that have little influence on the prediction results of the model are compressed to 0, and the number of independent variables decreases gradually. In the Python environment, the number of features based on  $\lambda$  is simulated and calculated, and the value of  $\lambda$  is 70. Combined with the Pearson coefficient correlation coefficient matrix and the Lasso regression key feature ranking, the characteristic values associated with the trend of cross-border e-commerce transaction volume were obtained, namely, x1 (the number of China-Europe trains), x2 (the number of boxes sent by China-Europe trains), x12 (education expenditure), x27 (per capita disposable income of urban permanent residents), and x32 (science expenditure).

Combined with the Pearson coefficient correlation coefficient matrix and the Lasso regression key feature ranking, the characteristic values associated with the trend of cross-border e-commerce transaction volume were obtained, namely, x1 (the number of China-Europe trains), x2 (the number of boxes sent by China-Europe trains), x12 (education expenditure), x27 (per capita disposable income of urban permanent residents), and x32 (science expenditure).

### Key Feature Prediction Based on Grey Prediction Model

The gray prediction model is a prediction method that establishes a mathematical model to make predictions through a small amount of incomplete information, which is based on the past and present development laws of objective things. The first-order accumulation of the historical data of the time series is carried out to obtain the generated series.

Suppose the observed values for a sequence of behavioral features in the system are:

$$X^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)\} \quad (2)$$

(1) AGO accumulation sequence. The  $x^{(0)}$  is accumulated, and through the accumulation process, the development trend of the ash accumulation development process can be seen. It is possible to obtain an accumulation AGO sequence in one time:

$$X^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)\} \quad (3)$$

$$x^{(1)}(k) = \sum_{m=1}^k x^{(0)}(m), k = 1, 2, \dots, n,$$

Among them,

(2) MEAN series. Mean generation is to average the adjacent numbers in the AGO sequence to obtain the generation sequence. The mean is generated in the following sequence:

$$Z^{(1)} = \{z^{(1)}(2), z^{(1)}(3), \dots, z^{(1)}(n)\} \quad (4)$$

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

Among them,

(3) Grayscale model. The basic form of the GM(1,1) model is as follows:

$$x^{(0)}(k) + az^{(1)}(k) = b \quad (5)$$

Establish a first-order linear differential equation for  $x^{(1)}$ :

$$\frac{dx^{(1)}(t)}{dt} + ax^{(1)}(t) = b \quad (6)$$

Solve the differential equations to get the prediction model, as shown in the equation:

$$\hat{x}^{(1)}(k+1) = \left[ x^{(0)}(1) - \frac{b}{a} \right] e^{-ak} + \frac{b}{a} \quad (7)$$

Since the GM(1,1) model obtained a cumulative value, the data obtained by the GM(1,1) model were subtracted and reduced to  $\hat{x}^{(0)}(k+1)$ . The grey prediction model can use the posteriori to test the accuracy of the model, and the discriminant rules using the posteriori test method are shown in Table 3.

Table 3 Reference table for post-difference test discrimination

P	C	Model accuracy
>0.95	<0.35	Good
>0.80	<0.5	qualified
>0.70	<0.65	Barely qualified
<0.70	>0.65	Unqualified

In the table, C and P are calculated as follows:

$$C = \frac{\sigma(\text{delta})}{\sigma(X^{(0)})} \quad (8)$$

$$P = \frac{S}{L} \quad (9)$$

Among them,  $\text{delta} = |X^{(0)} - X^{(1)}|$ ,  $\sigma$  represents the standard deviation, S represents the number  $|\text{delta} - \text{mean}(\text{delta})| < 0.6745 \cdot \sigma(X^{(0)})$ ,  $\text{mean}(\text{delta})$  represents the average value of delta, L represents the length of  $X^{(0)}$ .

Using the grey prediction model and referring to the time series characteristics (Xu Peng et al. 2022; Orieb Abu Alghanam et al, 2022; Danyang Cao et al., 2023), the historical data of various eigenvalues from 2018 to 2022 was analyzed, and the transaction volume from 2023 to 2024 was predicted, and the results are shown in Table 4.

Table 4 Prediction Results based on the Grey Model

	x1	x2	x12	x27	x32	y
--	----	----	-----	-----	-----	---

2018	320	25060	261000	71207	49000	654.7
2019	528	42286	310039	77150	59327	753.98
2020	974	80392	324695	80137	62363	870.88
2021	1277	105292	343485	86628	71802	1013.57
2022	1569	129300	365426	86975	82596	1083.5
2023	2183.53	180998.2	384831.2	92031.2	91543.62	-
2024	2949.9	245414.1	406697.6	96092.22	102916	-
Model accuracy	good	good	good	good	good	-

Source: Obtained based on data analysis in this study

### Regression Prediction Based on Support Vector Machine Model

Support Vector Regression (SVR) uses the idea of support vectors to perform regression analysis on data when doing fitting. By constructing a support vector machine regression prediction model (Y. Geetha Reddy et al., 2022; Hu Yili et al., 2022), substituting the grey prediction structure into the vector machine model, and calling the LinearSVR() function in Python, the total cross-border e-commerce transaction volume in Yiwu in 2023 and 2024 is predicted, as shown in Figure 1.

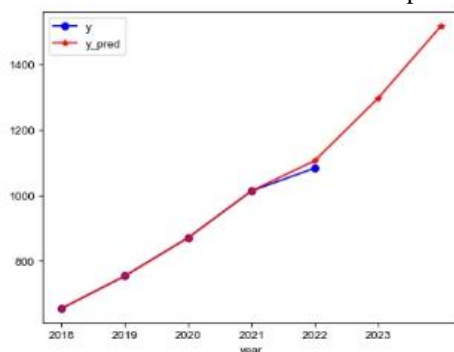


Figure 1: Cross-border e-commerce transaction volume prediction based on SVR model

As can be seen from the table, the forecasts for 2023 and 2024 have high accuracy and can be used as a reference for predicting the key characteristics of cross-border e-commerce transaction volume, as shown in Table 5.

Table 5 2023-2024 cross-border e-commerce transaction volume forecast

	x1	x2	x12	x27	x32	y	y_pred
2018	320	25060	261000	71207	49000	654.7	654.7
2019	528	42286	310039	77150	59327	753.98	753.9848
2020	974	80392	324695	80137	62363	870.88	870.8783
2021	1277	105292	343485	86628	71802	1013.57	1013.54
2022	1569	129300	365426	86975	82596	1083.5	1106.032
2023	2183.53	180998.2	384831.2	92031.2	91543.62	-	1296.624
2024	2949.9	245414.1	406697.6	96092.22	102916	-	1518.208

Source: Obtained based on data analysis in this study

### Performance Metrics

The performance metrics of the regression model are used to measure the performance of the cross-border e-commerce prediction model, as shown in Figure 2.

```
Mean absolute error: [3.76153251]
Median absolute error: 0.0032572849905250223
Explainable variance: [0.99845387]
R-squared value: [0.99814554]
```

Figure 2: Performance measurement results of the prediction model

Source: Obtained based on data analysis in this study

From the results, it can be seen that the mean absolute error and the median absolute error are small, and the explanatory variance is very close to 1 with the R square value. At the same time, combined with the results of Figure 2 and Table 4, it can be seen that the predicted value is basically consistent with the real value, indicating that the established support vector regression model has excellent fitting effect and high model accuracy, which can be used to predict the statistical value of cross-border e-commerce transaction data in the city and be applied to practical work guidance.

## CONCLUSION

The development of the cross-border e-commerce industry is a systematic improvement project, and with the improvement of economic vitality in the post-epidemic era, cross-border transactions such as manufacturing, exhibition and trading, warehousing and logistics, financial credit, and market management will be more complete, and the online product supply chain system will be more perfect. Through the analysis of various characteristic values and the exploration of the influencing factors of index changes, we can help the region to implement precise policies, promote the steady increase of cross-border e-commerce transaction volume, and promote the cross-border e-commerce industry to move towards a higher end of the value chain.

## ACKNOWLEDGMENT

This work is partially supported by the Funded projects: Humanities and Social Sciences Fund of the Ministry of Education, "Research on the Mechanism and Path of E-commerce Empowering Rural Industry Revitalization" (22YJA630023).

This paper is also the general project "Research on the Impact of Generative Artificial Intelligence on Teachers' Professional Development" (jsjy202407) in the 2024 open project of "Hubei Teacher Education Research Center", a key research base of humanities and social sciences in Hubei Province, and the research and development fund project of Wuchang University of Technology, "Research on the Path of Improving Digital Literacy of Applied Undergraduate Teachers in the Era of Digital Intelligence" (Project No.: X2024YS002).

## REFERENCES

- Huggins, R., & Izushi, H. (2015). The competitive advantage of nations: Origins and journey. *Competitiveness Review*, 25(5), 458-470. <https://doi.org/10.1108/CR-06-2015-0044>
- Wagner, G., Fota, A., Schramm-Klein, H., & Steinmann, S. (2023). Development of a motivation–trust–vulnerability (MTV) framework for cross-border online shopping: A cross-national application to Chinese and German consumers. *Schmalenbach Journal of Business Research*, 75(3), 389-414. <https://doi.org/10.1007/s41471-023-00170-2>
- Xue, C., Wen, W., & Li, S. (2022). Enterprise strategy analysis of synergy between cross-border e-commerce and logistics in a dynamic environment. *Systems Science & Control Engineering*, 10(1), 459-467. <https://doi.org/10.1080/21642583.2022.2063200>
- Ndubuisi, G. (2020). Fettered cross-border capital flows, external finance dependence, and international trade. *International Economics*, 164, 206-216. <https://doi.org/10.1016/j.inteco.2020.10.002>
- Ni, W. (2022). Online and offline integration development of Yiwu cross-border e-commerce in digital economy era. *Procedia Computer Science*, 202, 307-312. <https://doi.org/10.1016/j.procs.2022.04.041>
- Reddy, Y. G., & Prasanth, Y. (2022). A hybrid quartile deviation-based support vector regression model for software reliability datasets. *International Journal of Advanced Computer Science and Applications*, 13(6). <https://doi.org/10.14569/IJACSA.2022.0130694>
- Hu, Y., & Wang, Z. (2022). Support vector machine-based nonlinear model and its application in regional economic forecasting. *Wireless Communications and Mobile Computing*, 2022(1), 3387543. <https://doi.org/10.1155/2022/3387543>
- Cao, D., Ma, J., Sun, L., & Ma, N. (2023). A Time-series Prediction Algorithm Based on a Hybrid Model. *Recent Advances in Computer Science and Communications (Formerly: Recent Patents on Computer Science)*, 16(1), 3-17. <https://doi.org/10.2174/2666255815666220316154957>
- Alghanam, O. A., Al-Khatib, S. N., & Hiari, M. O. (2022). Data mining model for predicting customer purchase behavior in e-commerce context. *International Journal of Advanced Computer Science and Applications*, 13(2). <http://doi.org/10.14569/IJACSA.2022.0130249>
- Karine, H. A. J. I. (2021). E-commerce development in rural and remote areas of BRICS countries. *Journal of Integrative Agriculture*, 20(4), 979-997. [https://doi.org/10.1016/S2095-3119\(20\)63451-7](https://doi.org/10.1016/S2095-3119(20)63451-7)
- Zhou, J., Jiang, Y., Shen, Y., Pantelous, A. A., Liu, Y., Huang, C., & Mei, L. (2022). Intermodal hub-and-spoke logistic network design with differentiated services: The case of China railway express. *Information Sciences*, 612, 796-815. <https://doi.org/10.1016/j.ins.2022.08.083>
- Peng, X., Li, X., & Yang, X. (2022). Analysis of circular economy of e-commerce market based on grey model under the background of big data. *Journal of Enterprise Information Management*, 35(4/5), 1148-1167. <https://doi.org/10.1108/JEIM-01-2021-0015>
- Bi, L. Y., Li, Z. Y., & Li, D. D. (2019). Evaluation and cluster analysis of the competitiveness of China's provincial cross-border e-commerce Industry. *Journal of Commercial Economics*, 14, 78-81.
- Wu, J. & Hu, Y. (2022). Evaluation and improvement strategy of cross-border e-commerce industry competitiveness in Anhui province: An analysis based on the sample data of cross-border e-commerce comprehensive pilot zone in the Yangtze river delta. *Journal of Jiujiang University (Natural Science Edition)*, 3, 50-56+64.
- Chen, X. & Zhou, Z. (2022). Research on the influencing factors of cross-border e-commerce transaction volume in Ningbo. *Journal of Zhejiang Wanli University*, 4, 8-14.

**APPENDIX A: Computer Code (Part)**

```

def GM11(x0): # Define the GM11() function,x0 is in matrix form
import numpy as np
x1 = x0.cumsum() # AGO sequence
z1 = (x1[:len(x1) - 1] + x1[1:]) / 2.0
z1 = z1.reshape((len(z1), 1))
B = np.append(-z1, np.ones_like(z1), axis = 1)
Yn = x0[1:].reshape((len(x0)-1, 1))
[[a], [b]] = np.dot(np.dot(np.linalg.inv(np.dot(B.T, B)), B.T), Yn)
f = lambda k: (x0[0] - b / a) * np.exp(-a * (k - 1))-(x0[0]-b/a) * np.exp(-a * (k - 2))
delta = np.abs(x0 - np.array([f(i) for i in range(1, len(x0) + 1)]))
C = delta.std() / x0.std()
P = 1.0 * (np.abs(delta - delta.mean()) < 0.6745 * x0.std()).sum() / len(x0)
return f, a, b, x0[0], C, P # Returns grey prediction function, a, b, first term, variance ratio, and small residual probability

```