Quality Evaluation of GDP Data Based on Structural Matching in the Context of Big Data

Yue Siwei

Department of Public Administration, Chongqing Business Vocational College,81 Middle University City Road, Chongqing, China

Abstract: Based on the connotation of GDP data structure matching, a GDP data quality evaluation model was constructed from three aspects: factor input structure, aggregate structure, and production structure, and the construction model was used to evaluate Chongqing's GDP data quality from 2000 to 2018. It is found that Chongqing 's GDP data is normal from the perspective of factor input structure, Chongqing 's GDP data is of poor quality from the perspective of aggregate structure matching, Chongqing 's GDP data quality is from the perspective of production structure stability. There is still room for improvement. From the consistency of the results of the three assessment models, the internal structure of Chongqing's GDP data quality is still not balanced. Taken together, the overall development trend of Chongqing's GDP data quality is good. The current data quality of GDP in terms of total structure matching is the main factor affecting Chongqing GDP data quality.

1. Introduction

With the development and development of the concept of cloud computing technology, social economy is entering the era of big data applications, and the application of big data technology is infiltrating into various industries. Data has become a "resource" to promote the progress and development of society. However, the premise of data to provide a reasonable and scientific basis for decision-makers is that the data must be reliable and accurate. Therefore, it is particularly important to ensure the quality of statistical data. The quality of macroeconomic data, which is an important information carrier for regional economic and social development, has become increasingly important with economic development. Among them, GDP, as an important macroeconomic data, is an important indicator in macroeconomic research, and its quality directly affects the quality of regional macroeconomic data. Therefore, it is very important and necessary to evaluate the quality of GDP data. However, there are currently some studies on the evaluation of the quality of macroeconomic data, but relatively few evaluations of the quality of GDP data, and the research mainly focuses on several aspects such as parameter reliability, outlier diagnosis, and international standards, but from data structure matching The quality dimension of GDP data needs to be further studied [1-5]. In view of this, the quality of GDP data is evaluated from the perspective of data structure matching, with a view to providing a certain basis for the evaluation of the quality of macroeconomic data in theory and practice.

2. Connotation of GDP data structure matching

Data matching refers to examining the relationship between data from the height of the system. As a system, it is specific and hierarchical. Therefore, to assess the quality of GDP data from the perspective of structural matching, it is a systematic perspective to examine the relationship with GDP-related statistics, including the relationship between input and output, the relationship between the population and the hierarchy, and indicators. And indicators, to characterize the quality of GDP data [6-7]. Therefore, the evaluation of GDP data quality based on data structure matching mainly refers to three aspects. First, the rational data quality of the input structure of GDP elements, that is, the ratio of input and output in the GDP-related production activities is relatively stable. second,

439

GDP Aggregate structure matching data quality, that is, the structure of national economic activity is reflected by the physical quantity structure, and the total amount of GDP matches the structure of national economic activity. The third is the quality of GDP production structure stability data, that is, within a certain period the regional economic production structure is stable.

3. Construction of GDP data quality evaluation model based on structure matching

3.1. Data quality evaluation model of GDP input structure, aggregate structure, and production structure data

According to the connotation of GDP data structure matching, it can be known that the quality of GDP structure matching data can be evaluated from three aspects: factor input structure, aggregate structure, and production structure. The specific evaluation model is as follows:

3.1.1. Data quality evaluation model for the rationality of GDP input structure

The data quality of the rationality of the input structure of GDP factors refers to the evaluation of the data quality of GDP based on the structure of the input-output ratio. Therefore, based on the Cobb Douglas production function, the data quality assessment model for constructing the rationality of GDP input structure is as follows:

$$\ln\text{GDP} = \ln A + \alpha \ln K + \beta \ln L \tag{1}$$

Among them, K it refers to asset input, in order to facilitate data collection, it is replaced with fixed asset input, L it refers to labor force input, in order to facilitate data collection, it is replaced by the total population, E_1 represents the error rate of the evaluation model..

3.1.2. Data quality evaluation model of GDP aggregate structure matching

The data quality of GDP structure matching refers to the data quality of GDP based on the structure of GDP and national economic activity. Therefore, we choose to quantify the total GDP structure through industrial output and agricultural output. However, there are relatively many indicators of industrial output value and agricultural output. Therefore, in order to make the evaluation results based on objectivity, a principal component regression analysis model is used to quantify the matching structure of the total GDP. The specific model is as follows:

$$\text{GDP} = \alpha_i G_i + \beta_j N_j + A_0, (i = 1, 2, L, m; j = 1, 2, L, n) \tag{2}$$

Among them, G_i represents the i-th principal component of industrial output value, N_j represents the j-th principal component of agricultural production, and m and n represent the number of principal components of industrial and agricultural production, α_i, β_j, A_0 represents regression coefficient, and the error rate of the evaluation model for E_2.

3.1.3. Data Quality Evaluation Model of GDP Production Structure Stability

The quality of GDP production structure stability data is evaluated based on whether GDP matches regional production structure over time. GDP is closely related to regional industry, agriculture, and social consumption. Therefore, the coupling of three indicators, such as total industrial output value, total agricultural output value, and total social consumption, with total GDP is selected to carry out the quality of GDP production structural data. Evaluation, the specific model is as follows:

$$\text{GDP} = \alpha_1 G_z + \alpha_2 N_z + \alpha_3 X_z + A_0 \tag{3}$$

Among them, G_z represents the total industrial output value, N_z represents the total agricultural output value, X_z represents the total social consumption, E_3 represents the error rate of the evaluation model.

3.2. Comprehensive assessment of GDP data quality based on structural matching

3.2.1. GDP data quality evaluation criteria

According to the connotation of GDP data structure matching, the consistency of the GDP input structure, aggregate structure, and production structure is positively related to the quality of GDP data. The better the consistency of GDP in these three aspects, the better the quality of GDP data. it is good. Therefore, the consistency of the GDP in these three aspects is measured by calculating the error rate, and a critical value of 5% is adopted. A judgment that exceeds the error rate of 5% is an abnormal value, indicating that the quality of GDP data in this year is not good on data structure matching[7- 9].

3.2.2. Comprehensive assessment model of GDP data quality

In order to comprehensively evaluate the data quality of GDP in terms of structural matching, a structure-based matching The comprehensive evaluation model of GDP data quality is as follows:

$$E = w_1E_1 + w_2E_2 + w_3E_3 \tag{4}$$

Among them, w1, w2, w3 represents the weights, according to R^2 of the above three models, a normalization process is performed to determine. E represents the overall error rate.

4. Empirical research: Taking Chongqing as an example

Chongqing is an important node city of the Belt and Road and the Yangtze River Economic Belt, and is the core city of the Chengdu-Chongqing Two-City Economic Circle. Therefore, the quality of Chongqing's macroeconomic data is directly related to the quality of Chongqing's decision-making. It is crucial to sing the "two cities" and build an "economic circle" to create a new situation in Chongqing's high-quality development and contribute to the overall development of the country. Significant influence. Therefore, Chongqing was selected as the evaluation object. Because the statistical yearbook has a certain delay, 2000 to 2018 was selected as the evaluation time period, and the data of each evaluation index was obtained using the "Chongqing Data" website.

4.1. Data Quality Evaluation of the Rationality of Chongqing GDP Input Structure

According to formula (1), using SPSS22.0 for parameter estimation, the Chongqing GDP factor input structure rationality data quality evaluation model is obtained as follows:

$$lnGDP = -17.886 + 0.461lnK + 2.639lnL, R^2 = 0.996 \tag{5}$$

According to the parameter estimation results, the determination coefficient $R^2 = 0.996$ indicates that the model has a good fitting degree. This model can be used to evaluate the data quality of the rationality of Chongqing's GDP input structure. On this basis, the model was used to calculate the error rate between the estimated GDP value and the real value from 2000 to 2018, and the results of the data quality assessment of the rationality of Chongqing's GDP input structure were obtained, as shown in Table 1.

According to Table 1, from the perspective of factor input structure, Chongqing's GDP data from 2000 to 2018 is normal, indicating that the quality of Chongqing GDP data during the 19 years is relatively good.

4.2. Quality Data Evaluation of Chongqing GDP Aggregate Structure Matching Data

According to the industrial output value and agricultural output data of Chongqing from 2000 to 2018, SPSS 22.0 was used to conduct a principal component analysis of industrial output value and agricultural output, respectively. According to the principle of selecting principal components for principal component analysis, three principal components are selected for industrial output and recorded as G1, G2, and G3, five principal components are selected for agricultural output and recorded as N1, N2, N3, N4, and N5. According to formula (2), a principal component regression analysis was performed, and the quality evaluation model of Chongqing GDP aggregate structure

matching data was obtained as follows:

$$GDP = 8735.041 + 71.07?\,G_1 - 111.646?\,G_2 - 10.539?\,G_3 + 292.067?\,N_1 - 216.388?\,N_2$$
$$+ 66.023?\,N_3 - 183.649?\,N_4 - 529.326?\,N_5, R^2 = 0.996 \tag{6}$$

Table 1 Data quality evaluation results of the rationality of Chongqing's GDP input structure.

Years	E_2	Evaluation result
2000	-0.76%	normal
2001	-0.39%	normal
2002	-0.28%	normal
2003	-0.05%	normal
2004	-0.43%	normal
2005	-0.19%	normal
2006	0.14%	normal
2007	0.19%	normal
2008	-0.31%	normal
2009	0.61%	normal
2010	0.95%	normal
2011	0.12%	normal
2012	0.07%	normal
2013	0.10%	normal
2014	0.02%	normal
2015	-0.11%	normal
2016	-0.59%	normal
2017	-1.52%	normal
2018	-1.80%	normal

Table 2 Results of quality evaluation of Chongqing GDP aggregate structure matching data.

Years	E_2	Evaluation result
2000	25.01%	Abnormal (overestimated)
2001	-17.12%	Abnormal (underestimated)
2002	-21.08%	Abnormal (underestimated)
2003	5.40%	Abnormal (overestimated)
2004	12.59%	Abnormal (overestimated)
2005	-7.26%	Abnormal (underestimated)
2006	6.71%	Abnormal (overestimated)
2007	4.15%	normal
2008	-4.63%	normal
2009	-0.02%	normal
2010	8.94%	Abnormal (overestimated)
2011	-4.26%	normal
2012	-6.16%	Abnormal (underestimated)
2013	-1.63%	normal
2014	2.45%	normal
2015	2.88%	normal
2016	0.27%	normal
2017	-0.83%	normal
2018	-0.76%	normal

According to the results of regression analysis, the judgment coefficient $R^2 = 0.996$ indicates that the model has a good fitting degree. This model can be used to assess the quality of the matching data of Chongqing's GDP structure. On this basis, the model was used to calculate the

error rate between the estimated GDP value and the real value from 2000 to 2018, and the result of the quality evaluation of Chongqing GDP aggregate structure matching data was obtained, as shown in Table 2.

According to Table 2, from the perspective of the matching of the aggregate structure, Chongqing's GDP data was abnormal for 9 years from 2000 to 2018, of which 5 years were overvalued and 4 years were undervalued. It shows that 47% of the years of Chongqing GDP data from 2000 to 2018 have data quality errors, indicating that from the perspective of the overall structure matching, Chongqing GDP data quality is relatively poor.

4.3. Quality Evaluation of Chongqing GDP Production Structure Stability Data

According to formula (3), using SPSS22.0 for parameter estimation, the Chongqing GDP production structure stability data quality assessment model is obtained as follows:

$$GDP = 0.005G_z + 0.39N_z + 0.607X_z, R^2 = 0.999 \tag{7}$$

According to the parameter estimation results, the determination coefficient $R^2 = 0.999$ indicates that the model has a good fitting degree. This model can be used to evaluate the quality of Chongqing GDP production structure stability data. On this basis, the model was used to calculate the error rate between the estimated GDP and the true value from 2000 to 2018, and the results of the quality evaluation of Chongqing GDP production structure stability data were obtained.

Table 3 Data quality evaluation results of Chongqing GDP production structure stability.

Years	E_3	Evaluation result
2000	-2.76%	normal
2001	-2.06%	normal
2002	-1.34%	normal
2003	1.16%	normal
2004	-1.51%	normal
2005	-1.27%	normal
2006	6.78%	Abnormal (overestimated)
2007	2.20%	normal
2008	4.72%	normal
2009	-2.45%	normal
2010	-5.95%	Abnormal (underestimated)
2011	-8.68%	Abnormal (underestimated)
2012	0.03%	normal
2013	-0.28%	normal
2014	-2.81%	normal
2015	-2.19%	normal
2016	4.19%	normal
2017	0.84%	normal
2018	-0.64%	normal

According to Table 3, from the perspective of production structure stability, Chongqing's GDP data was abnormal for three years from 2000 to 2018, including one year overestimation and two year underestimation. It shows that 16% of the years of Chongqing GDP data from 2000 to 2018 have data quality errors, indicating that from the perspective of production structure stability, the quality of Chongqing GDP data needs to be improved.

4.4. Comprehensive evaluation of Chongqing GDP data quality based on structural matching

Based on the value of the judgment coefficient R^2 of the GDP element input structure, aggregate structure, and production structure data quality evaluation model, the values are normalized to obtain the error rate of GDP in the three input model of factor input structure,

aggregate structure, and production structure. The weights are: $w_1 = 0.333$, $w_2 = 0.334$, $w_3 = 0.333$.
Therefore, the comprehensive evaluation model of Chongqing GDP data quality based on structural matching is obtained:

$$E = 0.333E_1 + 0.334E_2 + 0.333E_3 \tag{8}$$

According to the comprehensive evaluation model, the comprehensive evaluation results of Chongqing GDP data quality based on structural matching are obtained, as shown in Table 4.

Table 4 Comprehensive evaluation results of Chongqing GDP data quality based on structural matching.

Years	E	Evaluation result
2000	7.18%	Abnormal (overestimated)
2001	-6.53%	Abnormal (underestimated)
2002	-7.58%	Abnormal (underestimated)
2003	2.17%	normal
2004	3.56%	normal
2005	-2.91%	normal
2006	4.55%	normal
2007	2.18%	normal
2008	-0.08%	normal
2009	-0.62%	normal
2010	1.32%	normal
2011	-4.27%	normal
2012	-2.02%	normal
2013	-0.60%	normal
2014	-0.11%	normal
2015	0.20%	normal
2016	1.29%	normal
2017	-0.50%	normal
2018	-1.07%	normal

First of all, according to Table 4, in terms of structural matching, Chongqing 's GDP data has anomalies in three years between 2000 and 2018 (1 year overestimation and 2 years underestimation), indicating that Chongqing has There are data errors in 16% of the years, but the outliers are concentrated in 2000-2002, indicating that the quality of Chongqing GDP data has improved significantly.

Secondly, specifically, based on the data in Table 1, Table 2, Table 3, and Table 4, make a line chart to get a line chart of the Chongqing structural matching assessment results, as shown in Figure 1. According to Figure 1, from 2000 to 2018, the results of the three evaluation models were consistent for 6 years, and the remaining 13 years were inconsistent. Overestimation indicates that the consistency of the error rate of the single model is poor, which indicates that the internal structure of Chongqing GDP data quality is not well balanced. Moreover, as a whole, from 2000 to 2011, no matter from the evaluation results of factor input structure, total volume structure, production structure, or comprehensive evaluation results, the GDP data quality error fluctuated significantly, after 2011 GDP data quality error has improved significantly. Moreover, according to Figure 1, it can be seen that the current impact on the quality of Chongqing's GDP data is mainly due to the fact that the data quality of GDP in terms of aggregate structure still needs to be improved.
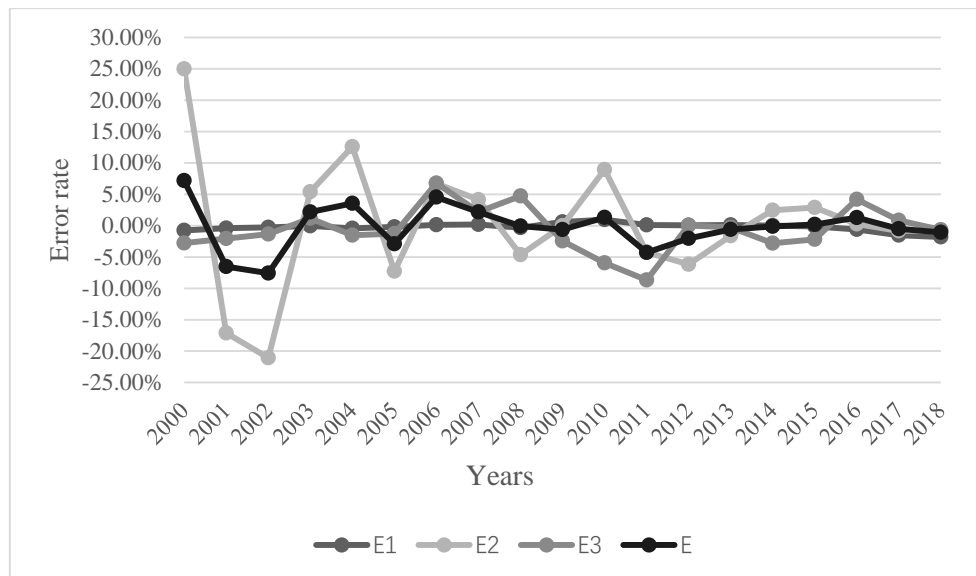
Figure 1 Line chart of Chongqing structural matching assessment results.

5. Conclusion

Based on the structural matching, a GDP data quality evaluation model was constructed from three aspects: factor input structure, aggregate structure, and production structure. Using the construction model, the GDP data quality of Chongqing in the 19 years from 2000 to 2018 was evaluated. Based on the findings of the study, the following conclusions were reached:

First of all, from the perspective of factor input structure, Chongqing 's GDP data is normal, indicating that the quality of Chongqing 's GDP data is relatively good during these 19 years. From the perspective of total structure matching, Chongqing 's GDP data is available for 9 years Abnormally, 47% of these 19 years have data quality errors. Therefore, Chongqing GDP is relatively poor in terms of data structure matching. From the perspective of production structure stability, Chongqing GDP data is abnormal for 3 years, that is, 16% of the 19 years have data quality errors. Therefore, the data quality of Chongqing GDP in terms of production structure stability needs to be improved.

Secondly, from the consistency of the results of the three evaluation models, there was consistency in 6 years, and inconsistency in the remaining 13 years. The consistency of the model's error rate is poor, which indicates that the internal structure of Chongqing GDP data quality is not well balanced.

Third, According to the comprehensive evaluation results, in terms of structural matching, Chongqing GDP data has abnormal data for three years during the 19 years from 2000 to 2018. From 2000 to 2011, the quality error of Chongqing GDP data fluctuated significantly. Data quality errors have improved significantly. In addition, the current data quality that affects Chongqing 's GDP is mainly related to the data quality of GDP in terms of aggregate structure that has yet to be improved.

In summary, from the perspective of structural matching, the overall development trend of Chongqing GDP data quality is good, but the quality of GDP data needs to be improved in terms of the overall structural matching and the stability of the production structure. The supervision of various aspects of data collection, release and evaluation ensure the authenticity and validity of the data quality, thereby improving the overall quality of GDP data. At the same time, we should actively construct a corresponding guarantee system to ensure the data quality of GDP, so that it can more and more truly, objectively and effectively reflect the state of social development.

References

[1] Wang Jing. The diagnostic method and demonstration of the quality of China's macroeconomic

statistics data. Statistics and Decision, 2018, 34 (04): 34-37.

[2] Yang Shaolang, Luo Qinglan. Theoretical and practical exploration of current macroeconomic statistical monitoring. China Statistics, 2017 (10): 6-9.

[3] Zhang Qing. Research on the Quality of China's GDP Statistics. Yanshan University, 2015.

[4] Wang Yongkai. Empirical Study on the Quality of China's GDP Data. Capital University of Economics and Business, 2015.

[5] Dang Wei, Long Huijun. Empirical test of the quality of government statistics data. Statistics and Decision, 2015 (03): 33-35.

[6] Wang Hang. Evaluation of GDP data quality in various regions of China based on robust regression. Tianjin University of Finance and Economics, 2013.

[7] Wei Ping, Li Xiao. Quantitative diagnosis method and application of timeliness of macroeconomic data. Forum on Statistics and Information, 2014, 29 (06): 30-33.

[8] Zhu Yibo. Empirical Study on Quality Evaluation of Zhejiang Macroeconomic Statistics Data——Based on the Compatibility of Statistical Data. Statistical Science and Practice, 2013 (12): 20-22.

[9] Yang. An assessment on the quality of China's preliminary data of quarterly GDP announcements. Applied Economics,2017,49(54).