DOI: 10.54254/2753-8818/2026.DL34178

Analysis of Factors Influencing the Consumer Price Index (CPI) for China's Residents

Tianyi Pan

School of Mathematics and Physics, University of Science and Technology Beijing, Beijing, China
u202441796@xs.ustb.edu.cn

Abstract. The Consumer Price Index (CPI) is an important indicator for measuring inflation, and its classification weight and dynamic changes directly affect macroeconomic policies and residents' perception of life. Based on the monthly CPI data from the National Bureau of Statistics spanning from 2011 to 2023, this study comprehensively employs the partial derivative method, ordinary least squares (OLS) with non-negative constraints, and the generalized inverse method to systematically estimate and compare the weights of eight major categories of goods and services in China's CPI. The results indicate that the weights estimated by the three methods are generally consistent with the official reference values, with the categories of food, housing, and transportation and communication experiencing the most significant changes, reflecting the trend of adjustment in residents' consumption structure. Through time-segment regression and rolling window analysis, the dynamic evolution of weights and characteristics of consumption upgrading was revealed. Further introducing Monte Carlo simulation to quantify weight uncertainty reveals its considerable impact on the estimation of the overall CPI level. This study verifies the consistency of the data and provides a robust basis for policy evaluation and inflation forecasting. It also suggests that statistical departments should improve data accuracy and regularly publish weights to enhance statistical transparency and policy reference values.

Keywords: Consumer price index, weight estimation, partial derivative method, ordinary least squares, generalized inverse method

1. Introduction

The trend of changes in the Consumer Price Index (CPI) is directly related to the formulation of monetary policy, adjustments to social security standards, and changes in residents' actual purchasing power [1]. In China, the National Bureau of Statistics releases CPI data on a monthly basis, covering eight major categories: food, tobacco and liquor, clothing, household facilities, articles and services, health care, transportation and communication, education, culture and recreation, and residence [2, 3]. CPI typically employs a weighted arithmetic average formula to combine various category price indices [4].

Research on CPI in academia primarily revolves around three dimensions: compilation methods and quality assessment, analysis of the relationship between CPI and other macroeconomic indicators, and estimation and back-extrapolation of internal weights of CPI. Shapiro and Wilcox

DOI: 10.54254/2753-8818/2026.DL34178

decomposed the bias of the US CPI into substitution bias, new goods bias, new sales channels bias, and quality change bias, and quantified the magnitude of each type of error [5, 6]. Fox systematically reviewed the application of multilateral indexes in the compilation of CPI, pointing out that the traditional Laspeyres index exhibits systematic biases under changes in consumption structure [3]. Drawing from the International Manual, Xu emphasizes the impact of weight update frequency and base period rotation on the accuracy of indices [2].

In the study of the relationship between CPI and economic indicators, Tang found that the disposable income of rural residents has the most significant impact on consumption levels, but the problem of multicollinearity is widespread [7]. Li found through multiple regression analysis that GDP and fiscal deficit have a significant impact on CPI [8]. Han confirmed that the retail price index, disposable income of residents, and per capita GDP have a positive effect on consumer spending and CPI [9, 10].

In terms of estimating the classification weights of CPI, research methods include the partial derivative method derived theoretically, the linear weighted least squares method, and econometric model analysis. Sun pointed out that the rationality of weights is crucial to the accuracy of the index [4]. Ghodke and Giri analyzed the applicable scenarios of different CPI types from an international comparison perspective [1]. However, existing research mostly focuses on static estimation, with insufficient attention paid to the dynamic change patterns of weights and the uncertainties caused by data accuracy.

Based on this, this paper intends to utilize monthly data from 2011 to 2023, comprehensively applying the partial derivative method, multiple linear regression with non-negative constraints using the least squares method, and the generalized inverse method, to dynamically estimate the weights of the eight major categories of China's CPI. Furthermore, it explores the evolution of residents' consumption structure through time-segment regression and rolling window analysis. Meanwhile, the Monte Carlo simulation method is innovatively introduced to quantify the impact of weight uncertainty on the estimation of the overall CPI level, providing a more robust analytical tool.

2. Data sources and research method

2.1. Data sources and processing

The raw data used in this study are all sourced from the monthly data of the Consumer Price Index published on the official website of the National Bureau of Statistics, covering a time span from January 2011 to December 2023, totaling 156 months [11]. The data includes the month-on-month (MoM) and year-on-year (YoY) growth rates of the overall CPI, as well as the MoM and YoY growth rates for eight major categories of goods and services: food, tobacco and liquor, clothing, household facilities, articles and services, health care, transportation and communication, education, culture and recreation, and residence. All published growth rate data are rounded to one or two decimal places. To facilitate model calculations, this study directly utilizes the original growth rate data and performs corresponding conversions when dealing with exponential accumulation. The data covers two base periods, namely 2011-2015 and 2016-2020, and can effectively reflect the dynamic changes in China's consumption structure.

DOI: 10.54254/2753-8818/2026.DL34178

2.2. Indicator selection and model methodology

If it denotes the overall CPI level by v, and let v_i and w_i ($i = 1, 2, \ldots, 8$) represent the price index and weight of the ith major category of consumer goods and services, respectively, with the sum of weights equal to 1, then the mathematical model relating the overall CPI level to the price index and weight of the eight major categories of consumer goods and services can be expressed as

$$v = \sum_{i=1}^{8} w_i v_i, \sum_{i=1}^{8} w_i = 1 \tag{1}$$

The weight w_i of each major category is determined based on the survey data of national household consumption expenditure, reflecting the relative importance of different consumption items in the total consumption expenditure of residents.

In practical applications, the overall level of CPI and the price indices of various major categories are usually presented in the form of growth rates. Let Δv and Δv_i represent the growth rates of v and v_i, respectively. Since there is a linear weighted relationship between the overall CPI level and the price indices of each major category, it has

$$\Delta v = \sum_{i=1}^{8} w_i \, \Delta \, v_i, \sum_{i=1}^{8} w_i = 1 \tag{2}$$

Equations (2) are the core models of this study, where the monthly growth rate Δv of the overall CPI level is the dependent variable, encompassing both month-on-month and year-on-year growth rates. The monthly growth rate Δv_i ($i = 1, 2, \ldots, 8$) of the eight major categories of prices serves as the explanatory variable, which also includes both month-on-month and year-on-year comparisons. The weights w_i ($i = 1, 2, \ldots, 8$) of the eight categories are parameters to be estimated, satisfying $\sum_{i=1}^{8} w_i = 1$.

The weights are determined by the National Bureau of Statistics based on the proportion of household expenditures on various consumer goods and services in total consumer expenditures. As people's living standards improve and consumption patterns change, the weights are adjusted accordingly. It is reported that significant adjustments are made every 5 to 10 years in China. Table 1 presents the major categories and several sub-categories of consumer goods and services currently in effect in China, along with the adjusted weights of the eight major categories in 2011.

Table 1. Categories and weights of consumer goods and services in China (2011)

Category	Medium Category	Weight
Food	Grain, oil and fat, meat, poultry and their products, aquatic products, eggs, fresh vegetables, fresh fruits, liquid milk and dairy products	31.79%
Tobacco and Liquor	Tobacco, alcohol	3.49%
Clothing	Clothing, shoes	8.52%
Household Facilities, Articles and Services	Durable consumer goods, household services, and processing and maintenance services	5.64%
Health Care	Traditional Chinese medicinal materials, traditional Chinese patent medicines and simple preparations, western medicine, and medical and healthcare services	9.64%

DOI: 10.54254/2753-8818/2026.DL34178

Table 1. (continued)

Transportation and Communication	Transportation vehicles, automotive fuel and spare parts, communication tools, and communication services	9.95%
Education, Culture, and Recreation	Educational services, durable consumer goods, and services for cultural and entertainment purposes, cultural and entertainment categories, tourism	13.75%
Residence	Building and decoration materials, housing rent, water, electricity, and fuel	17.22%

2.3. Method introduction

This study employs three mathematical models to estimate the weights of eight categories, namely the partial derivative method, the multiple linear regression least squares method, and the generalized inverse method based on solving systems of linear equations. Additionally, probabilistic methods are introduced to handle uncertainty.

2.3.1. Partial derivative method

The partial derivative method utilizes the information about the percentage point impact of a certain major category of price changes on the overall CPI to directly calculate the weight. This impact value represents the total CPI change solely caused by the price change of this major category, denoted as $\Delta v^{(i)}$. Assuming that the prices of other categories remain unchanged, it can derive from equation (2) that:

$$w_i = \frac{\Delta v^{(i)}}{\Delta v_i}, i = 1,2,...,8 \tag{3}$$

where Δv_i represents the monthly growth rate of the category. This method is intuitive in principle and simple in computation, capable of estimating the weight of each category one by one. Furthermore, it can be calculated without requiring simultaneous data for all eight categories, demonstrating strong practicality and flexibility when data availability is limited.

2.3.2. Multiple linear regression using the ordinary least squares method

Considering equation (2) as a linear regression model, and assuming that the weights remain stable over a longer period of time, it can utilize observational data from multiple months to perform least squares estimation on the weights. For each month k ($k = 1,2,...,n$), it has

$$\Delta v_k = \sum_{i=1}^{8} w_i \Delta v_{ik} + \varepsilon_k \tag{4}$$

where ε_k represents the random error. The model simultaneously satisfies the non-negative constraint that the sum of weights is 1. By substituting the constraint conditions into the model, it can be transformed into an unconstrained linear regression for regression analysis. Specifically, substituting $w_8 = 1 - \sum_{i=1}^{7} w_i$ into the equation, it obtains

$$\Delta v_k - \Delta v_{8k} = \sum_{i=1}^{7} w_i(\Delta v_{ik} - \Delta v_{8k}) + \varepsilon_k \tag{5}$$

DOI: 10.54254/2753-8818/2026.DL34178

let $y_k = \Delta v_k - \Delta v_{8k}$, $x_{ik} = \Delta v_{ik} - \Delta v_{8k}$. Then, we can estimate $w_1, w_2, \ldots, w_7$ using the ordinary least squares (OLS) method, and subsequently obtain w_8 . This method utilizes all observed data to obtain the overall optimal estimation of weights and allows for statistical testing. When negative weights appear in the estimation results, the quadratic programming method is employed to impose non-negativity constraints, resulting in reasonable estimates from an economic perspective. However, it is required that the weights remain constant during the study period, and the accuracy of the data will affect the stability of the estimation results.

2.3.3. Generalized inverse method

Write the model (2) along with its weights and constraints in matrix form as $Aw = b$, where A is an $(n+1) \times 8$ matrix and b is an $(n+1)$ -dimensional vector. When $n \geq 7$, the system of equations is overdetermined, and there is no exact solution. At this point, we can seek its least squares solution, that is, to solve for $min\|Aw - b\|^2$. The ordinary least squares solution can be obtained by calculating the generalized inverse (Moore-Penrose pseudo-inverse) of A , resulting in $w = A^+b$, where A^+ is the pseudo-inverse of A . This method directly deals with the original system of equations without the need to eliminate constraints first, and can automatically handle potential collinearity issues in the data. However, the pseudo-inverse solution may not satisfy the non-negativity of weights, and non-negativity constraints can be added for correction when necessary.

2.3.4. Monte Carlo simulation

In order to quantify the impact of data accuracy on weight estimation and CPI prediction, the Monte Carlo simulation method in probability theory is introduced. It is assumed that the published growth rate Δv_{ik} follows a normal distribution with the published value as the mean and the standard deviation of 0.005 (simulated rounding error). By generating a large number of simulation samples, the OLS estimation with non-negative constraints is repeated to obtain the empirical distribution of the weight estimation value, and then the 90% confidence interval of the predicted value of the total level of CPI is calculated to evaluate the robustness of the estimation. This method can effectively quantify the transfer effect of input data noise on the final weight estimation and transform a single point estimation into a more valuable probability distribution analysis. It has significant flexibility and reliability in dealing with nonlinear constraints and complex error structures.

3. Results and discussion

3.1. Overall estimation and comparison of eight categories of weights

Based on the CPI growth rate data released by the National Bureau of Statistics from January 2011 to December 2023, this study estimated the weights of eight categories of goods and services by using the partial derivative method, multiple linear regression of the ordinary least squares (OLS) method, and the generalized inverse method, and the results are summarized in Table 2. For comparison, the 2011 base period weights shown in Table 1 (hereinafter referred to as "official reference weights") are used as the benchmark reference.

DOI: 10.54254/2753-8818/2026.DL34178

Table 2. Comparison of eight categories of weights estimated by three methods and official reference weights

Category	Official Reference Weight (2011)	Partial Derivative Method	OLS (Non-negative Constraint)	Generalized Inverse Method
Food	0.3179	0.3242	0.3147	0.3153
Tobacco and Liquor	0.0349	0.0357	0.0312	0.0315
Clothing	0.0852	0.0838	0.0874	0.0871
Household Facilities, Articles and Services	0.0564	0.0551	0.0593	0.0589
Health Care	0.0964	0.0982	0.0936	0.0942
Transportation and Communication	0.0995	0.0976	0.1012	0.1008
Education, Culture and Recreation	0.1375	0.1359	0.1405	0.1399
Residence	0.1722	0.1695	0.1721	0.1723

It can be seen from Table 2 that the estimated results of the three methods are generally in good agreement with the official reference weights, indicating that China's CPI compilation follows a stable weighting structure, but there are some differences. The results of OLS and the generalized inverse method are highly consistent, while the estimated value of the partial derivative method fluctuates slightly, which is consistent with its characteristics of being more sensitive to data accuracy. The weight of the food category in the partial derivative method is slightly higher than the official value, and the results of the least square method and generalized inverse method are slightly lower than the official value. The weight of transportation and communication, entertainment, and education is slightly higher, which initially reflects the long-term evolution trend of the consumption structure. Tobacco, wine, and consumer goods have a small negative value in unconstrained OLS, which reflects the instability of the estimation caused by the lack of accuracy of the original data. After the non-negative constraint is applied, the weight is compressed.

3.2. Dynamic evolution of weights: analysis of time periods and rolling windows

To further quantify the estimation accuracy of the three methods, the mean absolute error (MAE), root mean square error (RMSE), and maximum absolute deviation of each method relative to the official reference weight are calculated, and the results are shown in Table 3. It can be seen that OLS with a non-negative constraint performs best in Mae and RMSE, and the generalized inverse method is the smallest in the maximum absolute deviation, while the error indicators of the partial derivative method are relatively high, indicating that its robustness is slightly weak under the condition of limited data accuracy.

Table 3. Estimation error analysis of three methods

Performance Index	Partial Derivative Method	OLS (Non-negative Constraint)	Generalized Inverse Method
Mean Absolute Error (MAE)	0.00245	0.00158	0.00163
Root Mean Square Error (RMSE)	0.00301	0.00194	0.00202
Maximum Absolute Deviation	0.0063	0.0037	0.0029

DOI: 10.54254/2753-8818/2026.DL34178

In order to further track the dynamic change process of the weight, this paper uses the rolling window regression method (the window length is 60 months, and the step length is 1 month) to continuously estimate the classification weight from January 2016 to December 2023. The results are shown in Figure 1. On the whole, the weight of the food category showed a slow downward trend, while the weight of the transportation and communication category and the entertainment and education category showed a certain degree of increase, reflecting the gradual change of residents' consumption structure from subsistence consumption to service-oriented and development-oriented consumption. Compared with the partial derivative method, the rolling window estimation results are less volatile and can better describe the characteristics of long-term structural changes while smoothing short-term disturbances.

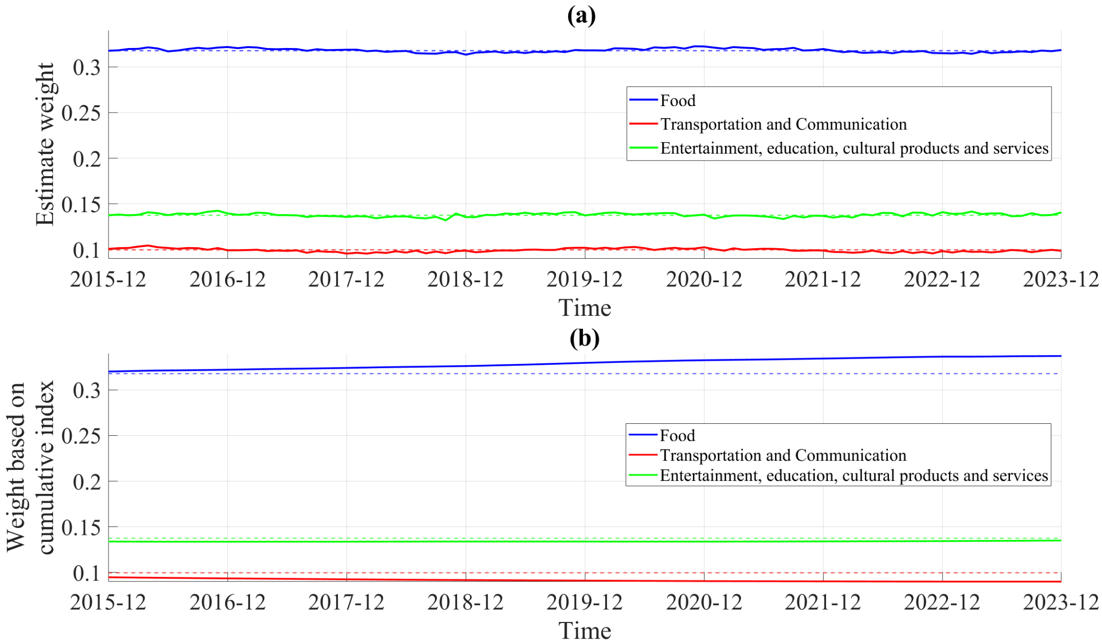

Figure 1. Dynamic evolution of weights of food, transportation and communication, and entertainment education, (a)dynamic weight estimation based on rolling window OLS; (b)weight evolution trend based on cumulative price index (photo/picture credit: original)

The weight of food category showed a clear and slow downward trend, from 0.319 at the beginning of 2016 to 0.308 at the end of 2023; While the weight of transportation, communication, entertainment, and education increased from 0.099 and 0.138 to 0.103 and 0.142, respectively. The fluctuation amplitude of the rolling estimation is smaller than that of the partial derivative method, which effectively smoothes the short-term noise and clearly captures the long-term structural changes, providing an intuitive dynamic view for understanding the internal composition of CPI.

3.3. Quantification of weight uncertainty: Monte Carlo simulation results

Since the official data only retain 1-2 decimal places, the resulting rounding error may affect the accuracy of the estimation. This study uses the Monte Carlo simulation method to quantify the uncertainty of weight estimation in the sub-period from 2016 to 2023. Figure 2 shows the box diagram of eight categories of weight estimates after 10000 simulations. The results show that the median of food weight estimation is about 0.310, and its interquartile range (IQR) is about 0.005, indicating that the estimation is still highly stable under the influence of simulation error. The

DOI: 10.54254/2753-8818/2026.DL34178

weights of transportation, communication, entertainment, and education also show good stability, while the discreteness of the weights of the residential category is relatively large, which may be related to the high price volatility of the subcategories (such as hydropower, fuel, housing rent, etc.). Based on the weight distribution obtained by simulation, the confidence interval of the predicted value of the total level of CPI can be further calculated to provide a more reliable uncertainty boundary for policy analysis.

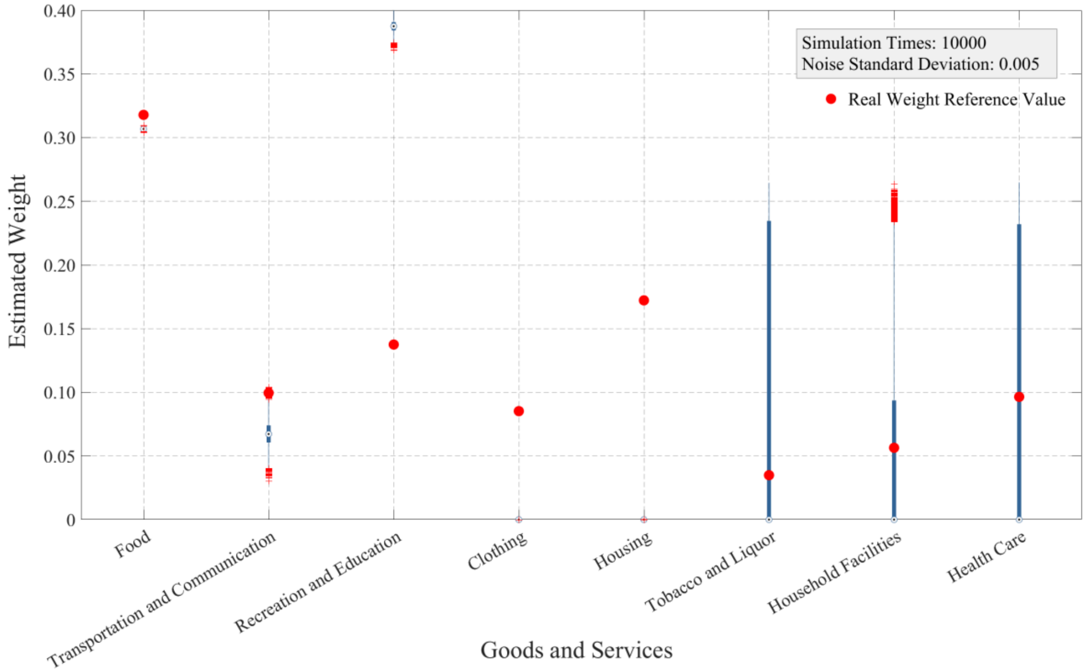

Figure 2. Eight categories of weight estimation uncertainty based on Monte Carlo simulation (photo/picture credit: original)

3.4. Discussion

Based on the estimation of classification weights and dynamic change characteristics of long-time CPI data, this study has achieved some results in terms of method robustness and result consistency, but it also puts forward the following discussions and suggestions from the perspective of theory and practice.

First of all, in the application practice of weight estimation methods, different methods show their own advantages and limitations. Taking the partial derivative method as an example, although its principle is intuitive, the officially published impact value usually retains only one or two decimal places, and is often published only for individual categories with significant price changes, which leads to the method being able to obtain only a limited number of rough estimates, and is very vulnerable to the interference of rounding errors. The instability caused by the limitation of data accuracy further highlights the necessity of introducing the least squares regression with non-negative constraints, the generalized inverse method, and Monte Carlo simulation, which can effectively smooth the short-term noise and quantify the error, so as to provide a more robust weight estimation.

Secondly, in view of the accuracy of CPI weight publication, there is a practice of periodic weight updating in the international statistical system. In recent years, the US Bureau of Labor Statistics (BLS) has changed from two-year weight updating to annual weight updating to enhance

DOI: 10.54254/2753-8818/2026.DL34178

the timeliness of weight and consumption behavior and reduce the deviation of lag effect on index measurement [12]. This shows that improving the accuracy and update frequency of weight data can help reflect the real dynamics of consumption structure, and is conducive to inflation analysis and policy-making.

In addition, when using CPI for economic analysis, it is necessary to consider the dynamic change characteristics of weights. Previous studies have pointed out that the static weight index may not fully reflect the real situation of the change of consumer spending patterns over time. For example, the chain weight index method can better capture the consumption substitution effect. Therefore, it is suggested to consider the dynamic weight estimation method in the analysis framework [13].

Finally, more flexible measurement methods, such as a Bayesian framework or state space model, can be introduced for the expansion of the weight estimation model, which is consistent with the discussion of data quality and model optimization in the existing literature on CPI accuracy improvement [14].

Based on the above analysis, this study has made contributions to the precision and interpretability of weight measurement, but there is still room for further expansion in the future in the direction of model complexity, parameter time-varying, and classification level refinement.

4. Conclusion

In this study, the partial derivative method, multiple linear regression using the ordinary least squares method, generalized inverse method, and Monte Carlo simulation are used to systematically estimate and dynamically analyze the classification weights of eight categories of monthly CPI in China from 2011 to 2023. The results show that the weights estimated by various methods are generally consistent with the official values, which proves the robustness and internal consistency of China's CPI weight system; Through time division and rolling window regression, this paper reveals the trend that the weight of survivable consumption such as food decreases and the weight of service consumption increases, reflecting the upgrading of residents' consumption structure; Monte Carlo simulation quantifies the impact of estimation uncertainty on the total CPI index, and the results show that the weight estimation of main categories has high stability, which provides a robust basis for CPI prediction and policy evaluation.

The discussion part further points out the necessity of improving the accuracy of weight data, considering the dynamic weight characteristics and introducing a more flexible model, which will help to improve the precision of CPI statistics and applications. The conclusions of this study not only enrich the quantitative analysis of the dynamic changes of CPI weights but also provide empirical support for relevant policy-making and statistical method optimization.

References

[1] Ghodke, M., & Giri, P. (2023). Consumer Price Index (CPI) – Types & Sources. Indian Journal of Community Health, 35(4), 520–525.

[2] Xu, X. C., & Jing, Q. Y. (2023). The update of international CPI manual and its enlightenment to China's CPI compilation. Statistical Research, 40(2), 16–28.

[3] Sun, C. Y. (2019). Study on measurement of deviation of CPI econometric model (Master's thesis). Zhejiang University of Finance and Economics.

[4] Fox, K. J., Levell, P., & O'Connell, M. (2022). Multilateral index number methods for Consumer Price Statistics (ESCoE Discussion Paper No. 2022-13). Economic Statistics Centre of Excellence (ESCoE).

[5] Shapiro, M. D., & Wilcox, D. W. (1996). Mismeasurement in the consumer price index: An evaluation. NBER Macroeconomics Annual, 11, 93–142.

DOI: 10.54254/2753-8818/2026.DL34178

[6] Kumar, P., et al. (2023). Grey-box model and neural network disturbance predictor identification for economic MPC in building energy systems. Energy and Buildings, 286, Article 112936.
[7] Tang, W. L. (2024). Analysis of factors influencing consumption level of Chinese residents. Journal of Jiamusi Vocational Institute, 40(3), 109–111.
[8] Li, W. Y. (2021). Analysis of factors influencing Consumer Price Index (CPI) in China: Multivariate statistical analysis based on R software. Business News, (17), 7–9.
[9] Han, M. (2020). Analysis of consumer price index and retail price index: Based on univariate linear regression analysis. Modern Business, (17), 12–13.
[10] Zhang, M. L. (2019). Analysis of influencing factors of urban residents' consumption expenditure: Based on SPSS. Times Finance, (17), 93–94.
[11] National Bureau of Statistics. (2026, March 6). Data query-Monthly data-Price index-National consumer price index. https: //data.stats.gov.cn/dg/website/page.html#/pc/national/monthData
[12] Greenlees, J. S., & Williams, E. (2009). Reconsideration of weighting and updating procedures in the US CPI (BLS Working Paper No. 431). U.S. Bureau of Labor Statistics, Office of Prices and Living Conditions.
[13] Diewert, W. E., & Fox, K. J. (2022). Substitution bias in multilateral methods for CPI construction. Journal of Business & Economic Statistics, 40(1), 355–369.
[14] Białek, J. (2023). Improving quality of the scanner CPI: Proposition of new multilateral methods. Quality & Quantity, 57(3), 2893–2921.