

Prediction Model for Minimum Subsistence Allowance Standard Based on Principal Component Analysis-Based Multiple Linear Regression (PMLR)

Xinyi Fang, Jinxuan Guo, Hai Du, Guoxu Li, Xuewen Shen*

School of Media Engineering, Communication University of Zhejiang, Hangzhou, China

Email: *shenxuewen@cuz.edu.cn

How to cite this paper: Fang, X.Y., Guo, J.X., Du, H., Li, G.X. and Shen, X.W. (2025) Prediction Model for Minimum Subsistence Allowance Standard Based on Principal Component Analysis-Based Multiple Linear Regression (PMLR). *Journal of Computer and Communications*, 13, 210-225.
<https://doi.org/10.4236/jcc.2025.134014>

Received: February 20, 2025

Accepted: April 25, 2025

Published: April 28, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing 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

Amid the wave of the digital age, advanced technologies such as big data, artificial intelligence, and cloud computing are driving precise analysis and forecasting across various fields. This paper explores the correlation between economic variables and minimum subsistence allowance standards, taking Zhejiang Province as a case study. Through Principal Component Analysis (PCA) and multiple linear regression models, the comprehensive impact of various economic factors on the minimum allowance standard is thoroughly analyzed, with an assessment of the model's predictive performance. The results indicate that the extracted principal components effectively capture the main influencing factors of the allowance standard, thereby enhancing the model's explanatory power and predictive accuracy. By combining machine learning model predictions with policy analysis, this study provides data support and theoretical basis for establishing scientifically grounded allowance standards and aims to offer new perspectives and references for future social assistance work.

Keywords

Minimum Subsistence Allowance, Social Assistance, Principal Component Analysis

1. Introduction

1.1. Research Background and Significance

1.1.1. Background

With the rapid development of China's socio-economic landscape and the contin-

uous improvement of living standards, the construction of a robust social security system has become one of the nation's focal points. Particularly in safeguarding the basic living standards of vulnerable groups, the minimum subsistence allowance (MSA) system plays a critical role within the social assistance framework, upholding social equity and stability. According to the Interim Measures for Social Assistance in the People's Republic of China (Order No. 46, National Development and Reform Commission, Ministry of Finance, Ministry of Civil Affairs) [1], the social assistance system encompasses not only MSA but also various other forms of assistance for individuals in extreme hardship. The policy mandates that local governments dynamically adjust MSA standards based on factors such as regional economic development, price levels, and household income, emphasizing the scientific and timely nature of these adjustments to ensure that aid standards reflect the current economic landscape and societal trends in a rapidly evolving environment [2]. Furthermore, the Guiding Opinions on Further Improving the Determination and Adjustment of Minimum Living Security Standards [3] further clarifies the principles and 1 procedures for adjusting MSA standards, outlining a systematic process involving data collection, statistical analysis, and social security assessment. This announcement highlights the importance of reliable data and comprehensive analysis during the adjustment process, requiring that government bodies base MSA adjustments on rigorous economic data to ensure policy objectivity and equity. These policy documents provide a solid foundation for this study on adjusting MSA standards and underscore the importance of data-driven decision-making in social assistance.

Against this background, the question of how to systematically analyze factors influencing MSA standards using big data and artificial intelligence, and how to accurately forecast MSA standards based on actual economic data, has become a pressing issue [4]. Traditional manual adjustment methods, though effective in the early stages of MSA policy formation, face challenges in terms of precision and relevance as socio-economic complexity and data volumes increase. Hence, intelligent forecasting and analysis methods leveraging machine learning can offer government agencies more scientifically grounded and efficient support for setting and adjusting MSA standards [5].

This paper aims to investigate the use of statistical analysis techniques, such as Principal Component Analysis (PCA), to analyze and forecast factors affecting MSA standards. Through modeling economic data, we aim to uncover the intrinsic relationships between MSA standards and regional economic development, thereby providing policymakers with data support and theoretical insights to ensure that MSA standards are more scientifically sound and equitable, further promoting social stability and justice [6] [7].

1.1.2. Research Significance

With the rapid advancement of big data and artificial intelligence technologies, machine learning has gradually become an essential tool in the field of social security, particularly in identifying and predicting eligible recipients for subsidies.

Traditional manual review methods are not only inefficient but also prone to misjudgments, failing to meet the needs of modern social assistance efforts. Consequently, intelligent machine learning models are seen as effective solutions for optimizing resource allocation and enhancing the efficiency of social assistance [8].

Setting appropriate minimum subsistence allowance (MSA) standards is crucial for the effective operation of the social security system. With economic development and shifts in the social environment, the factors influencing MSA standards have become increasingly complex. To analyze and predict the impact of these factors on MSA standards, this study selected 21 economic variables, including minimum per capita consumption expenditure and per capita disposable income. While traditional multiple linear regression models can analyze the relationships among these variables to some extent, issues of multicollinearity and high-dimensional variables often limit the models explanatory power.

To address these limitations, this study employs Principal Component Analysis (PCA) to perform dimensionality reduction on the 21 economic variables [9] [10]. PCA reduces multicollinearity among variables while retaining the primary information in the data. Based on the dimensionally reduced data, we use a multiple linear regression model to further analyze the influence of each principal component on MSA standards and to forecast changes in MSA standards. This approach not only enhances the models explanatory power and accuracy but also provides a more scientifically grounded data basis for formulating equitable MSA policies.

1.2. Research Content and Methods

1.2.1. Research Content

The primary research focus of this study is to analyze and predict the factors influencing minimum subsistence allowance (MSA) standards. Utilizing economic data from the Zhejiang Statistical Yearbook (20202022) [11], 21 economic variables were selected, including minimum per capita consumption expenditure, per capita disposable income, value added of large-scale industries, regional gross domestic product, and the consumer price index. First, Principal Component Analysis (PCA) was employed to reduce the dimensionality of these high-dimensional economic variables, extracting principal components that explain the majority of the variance. Next, using a multiple linear regression model, this study analyzed the linear relationships between these principal components and MSA standards, further investigating the comprehensive impact of various economic factors on MSA standards. Finally, model evaluation was conducted to verify the goodness of fit and predictive performance of the model.

1.2.2. Research Methods

The research methods in this study include data preprocessing and feature selection, training and optimization of machine learning models, application of Principal Component Analysis (PCA), multiple linear regression, and analysis and comparison of model evaluation metrics. The use of these methods provides new

insights for the accurate identification of minimum subsistence allowance (MSA) recipients and offers quantitative support and a theoretical basis for establishing MSA standards.

2. Theoretical Background

2.1. Multiple Linear Regression

Multiple linear regression is a statistical method used to analyze the impact of multiple independent variables (explanatory variables) on a single dependent variable (response variable). The basic form of the multiple linear regression model is:

$$Y = X\beta + \varepsilon \quad (1)$$

- Y is the dependent variable (in this study, it is the minimum living standard, unit: 10,000 yuan/year).

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \quad (2)$$

- X_1, X_2, \dots, X_k It is an independent variable (such as per capita disposable income, industrial added value above designated size, and other economic indicators).

$$X = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1k} \\ 1 & X_{21} & X_{22} & \cdots & X_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{nk} \end{bmatrix} \quad (3)$$

- β_0 is the intercept term, which represents the expected value of the dependent variable when all independent variables are zero. $\beta_1, \beta_2, \dots, \beta_k$ is the regression coefficient, which represents the marginal effect of each independent variable on the dependent variable.

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} \quad (4)$$

- ε is the error term, which represents the random error not explained by the model.

$$\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad (5)$$

By using Ordinary Least Squares (OLS) estimation, the regression coefficients

can be obtained, establishing a linear relationship model between the independent and dependent variables.

2.2. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a dimensionality reduction technique that projects data from a high-dimensional space to a lower-dimensional space through linear transformations. This process reduces the complexity of the data while retaining as much variance as possible. PCA extracts principal components by calculating the covariance matrix of the data and solving for its eigenvalues and eigenvectors. The first few principal components (eigenvectors) typically explain a significant portion of the variance in the data. By projecting the data onto these principal components, PCA reduces the number of variables, eliminates redundant information, and facilitates subsequent analysis and modeling. PCA offers significant advantages in handling multicollinearity, high-dimensional data, and improving model efficiency.

2.3. Expansion of Theoretical Framework

This study is grounded in social policy theory, integrating the Multidimensional Dynamic Security Theory [12], which emphasizes the dynamic adaptation of Minimum Subsidy Allowance (MSA) standards to residents' basic needs. Incorporating a digital governance perspective [13], we synergize economic indicators with MSA formulation policies. Aligned with the "14th Five-Year Plan" for Modern Social Security Systems, MSA adjustments necessitate timeliness, achieved through real-time economic data analysis [14]. Drawing on post-pandemic economic resilience theory, this research incorporates indicators such as industrial profits (X_{15}) and consumption expenditure (X_1) to reflect regional economic risk resistance on MSA. Principal Component Analysis (PCA)-extracted core factors quantify these theoretical dimensions. For instance, Principal Component 1 (PC1) may encapsulate "economic resilience-consumption capacity," while PC2 represents "policy regulation-capital flow," thereby translating theoretical frameworks into modelable economic metrics.

3. Data Preprocessing

3.1. Data Source

3.1.1. Minimum Living Standard and Economic Data

The standards for minimum living allowances and economic data in this study are sourced from the "Zhejiang Statistical Yearbook" and the Zhejiang Public Data Open Platform. Data from the past three years have been collected and processed using statistical methods and Python for analysis. Based on innovative principles and the operational feasibility of the data, this study selects the following economic variables as factors influencing the minimum living allowance standards in Zhejiang Province: minimum per capita consumption expenditure (X_1), per capita disposable income (X_2), industrial added value

above designated size (X_3), industrial electricity consumption (X_4), real estate development investment (X_5), sales area of commercial housing (X_6), contracted foreign investment utilization (X_7), actual foreign investment utilization (X_8), total fiscal revenue (X_9), general public budget revenue (X_{10}), general public budget expenditure (X_{11}), total retail sales of consumer goods above designated size (X_{12}), total imports and exports (X_{13}), total exports (X_{14}), total profits of industrial enterprises above designated size (X_{15}), gross domestic product (X_{16}), per capita disposable income of urban permanent residents (X_{17}), per capita disposable income of rural permanent residents (X_{18}), consumer price index (X_{19}), balance of loans from financial institutions in both domestic and foreign currencies (X_{20}), and balance of deposits in financial institutions in both domestic and foreign currencies (X_{21}). These variables are used to model and analyze the macroeconomic factors affecting the minimum living allowance standards (Y) in Zhejiang Province. After standardization, the variables are denoted as ZY and ZX_i .

3.1.2. Theoretical Basis for Variable Selection

Guided by the Opinions on Reforming and Improving the Social Assistance System [15], which advocates “precise identification and dynamic adjustment,” and referencing the Three-Stage Exploratory Data Analysis (EDA) model [16], 21 key indicators were selected to ensure scientific rigor:

Survival Guarantee Core Indicators: Minimum per capita consumption expenditure (X_1) and CPI (X_{19}), directly addressing basic living cost coverage mandated by the Revised Zhejiang Social Assistance Regulations (2022).

Economic Vitality Monitoring Indicators: Industrial profits above designated size (X_{15}) and digital economy core industry value added, reflecting regional economic restructuring’s impact on MSA.

Policy Response Indicators: Social welfare expenditure ratio (X_{11}) and affordable housing investment (X_5), quantifying government resource allocation efficiency via fiscal expenditure models.

3.2. Data Preprocessing

In the application of machine learning, particularly in predicting and classifying beneficiaries of assistance, data preprocessing is a crucial step. Among these processes, removing duplicate data and handling missing data significantly enhance the model’s accuracy and efficiency.

3.2.1. Missing Value Removal Function

Remove the columns with missing values from the economic data, as shown in **Table 1** and **Table 2**.

3.2.2. The Role of Removing Duplicate Data

The role of removing duplicate data is mainly reflected in the following aspects:

1. **Avoid model overfitting:** Duplicate data is equivalent to oversampling of a part of the samples, which may cause the model to pay too much attention to these

Table 1. Example before handling missing values in economic data.

Year	M.L.S. (yuan/year)	T.P & T.O.I.E. (100 M)	...	P.C.D.I. (yuan/year)
2022	14,592	0	...	70,281
2022	12,420	0	...	60,554
⋮	⋮	⋮	⋮	⋮
2022	12,840	0	...	62,626
2022	12,600	0	...	58,080

Notes: M.L.S.: Minimum Living Standard; T.P & T.O.I.E.: Total profits and taxes of industrial enterprises above designated size; 100M: 100 Million Yuan; P.C.D.I.: Per capita disposable income.

Table 2. Example after handling missing values in economic data.

Year	M.L.S. (yuan/year)	M.P.C.C.E. (yuan/year)	...	P.C.D.I. (yuan/year)
2022	14,592	46,440	...	70,281
2022	12,420	38,322	...	60,554
⋮	⋮	⋮	⋮	⋮
2022	12,840	39,146	...	62,626
2022	12,600	38,371	...	58,080

Notes: M.L.S.: Minimum Living Standard; M.P.C.C.E.: minimum per capita consumption expenditure; P.C.D.I.: Per capita disposable income.

duplicate samples during training, thereby increasing the risk of overfitting. Removing duplicate data can reduce this risk and make the model more generalized.

2. **Improve training efficiency:** Duplicate data increases training time and computational cost. Removing this data can shorten training time and improve training efficiency.

3. **Improve model performance:** Duplicate data may distort the sample distribution and affect the model's learning of the real data distribution. Removing duplicate data helps the model learn the real distribution of the data more accurately, thereby improving prediction performance.

3.2.3. The Role of Handling Missing Data

Handling missing data is also crucial, mainly reflected in the following aspects:

1. **Reduce model bias:** Missing data may cause the model to be unable to fully learn the characteristics of the data, thereby generating bias. By properly handling missing data (such as filling, deleting, or interpolating), this bias can be reduced, allowing the model to more accurately reflect the real situation of the data.

2. **Improve data quality:** Missing data is one of the manifestations of poor data quality. Handling missing data can improve the completeness and consistency of the dataset and provide higher quality data for model training.

3. **Enhance model robustness:** In practical applications, data often contains

various noises and outliers. Processing missing data is part of data cleaning, which helps to enhance the robustness of the model to noise and outliers, and improve the stability and reliability of the model.

3.3. Data Normalization

Data standardization is a crucial step in the process of data analysis and modeling. Data with different features often have different dimensions and units. Standardization converts features with different units and dimensions to the same scale, so that the model can treat all features fairly and avoid some features dominating the model output due to their large dimensions. Unstandardized data may lead to numerical instability or difficulty in interpreting model parameters, which will cause the model to assign unreasonable weights to different features when processing these data. For example, a variable with a large range of values (such as per capita disposable income) may dominate the model, while another variable with a small range of values (such as year) may be ignored. Before performing principal component analysis (PCA), data standardization is essential because PCA is very sensitive to the scale of variables. Without standardization, the difference in the dimensions of the features may cause the PCA results to favor features with larger variance, thus affecting the dimensionality reduction effect. Therefore, before inputting the data into the model, the data needs to be standardized to ensure that all features are on the same scale.

Data normalization is a common data preprocessing technique used to convert data to a uniform scale range to eliminate the dimensional differences between different variables. Min-Max normalization is one of the methods, which maps the data to the range of [0, 1]. The specific formula is as follows:

$$x_{\text{norm}} = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (6)$$

Where,

- x is a value in the original data.
- $\min(x)$ is the minimum value in the original data set.
- $\max(x)$ is the maximum value in the original data set.
- x_{norm} is the normalized data value.

Through this formula, any original data value x will be converted to a value x_{norm} between 0 and 1. This method is particularly suitable for situations where the data range is known and relatively stable, such as pixel value normalization in image processing.

4. Models and Methods

4.1. Multiple Linear Regression Model Construction

This article uses three regression model selection methods to explore the impact of various economic indicators on the subsistence allowance standard.

- Forward Selection: Gradually add variables with high significance.

- Backward Elimination: Gradually eliminate variables with low significance.
- Bidirectional stepwise regression (Stepwise Selection): combines forward screening and backward elimination.

Regression analysis was performed through three methods to obtain the following regression equation.

4.1.1. Forward Selection

The results of the forward screening include four independent variables (minimum per capita consumption expenditure (10,000 yuan/year) (ZX_1), contracted use of foreign capital (10,000 yuan/year) U.S. dollars) (ZX_7), real estate development investment (10,000 yuan) (ZX_5), total industrial profits above designated size (100 million yuan) (ZX_{15}), The final regression equation is:

$$ZY = X\beta + \varepsilon \tag{7}$$

$$X = [ZX_1 \quad ZX_7 \quad ZX_5 \quad ZX_{15}], \beta = \begin{bmatrix} 0.3206 \\ 0.2110 \\ -0.9721 \\ 0.9103 \end{bmatrix}, \varepsilon = [0.1873]$$

The model has an adjusted R^2 of 0.9569, an AIC of -67.7615 , and a mean square error (MSE) of 0.0026. The regression coefficient shows that the lowest

Per capita consumption expenditure (10,000 yuan/year) (ZX_1), contracted use of foreign capital (ZX_7) and total industrial profits above designated size (ZX_{15})

It has a positive impact on the subsistence allowance standard, while real estate development investment (ZX_5) has a negative impact. The stepwise regression process is detailed in **Table 3**.

Table 3. Stepwise regression process.

Step	Variables Added/Removed	Model Variables	Adjusted R^2	AIC	MSE
1	Add ZX_1	$[ZX_1]$	0.5972	-14.3700	0.0281
2	Add ZX_7	$[ZX_1, ZX_7]$	0.9303	-57.3408	0.0046
3	Add ZX_5	$[ZX_1, ZX_7, ZX_5]$	0.9400	-60.2403	0.0038
4	Add ZX_{15}	$[ZX_1, ZX_7, ZX_5, ZX_{15}]$	0.9569	-67.7615	0.0026

4.1.2. Backward Elimination

The backward elimination method selects more independent variables, and the final regression equation is:

$$ZY = X\beta + c \tag{8}$$

$$X = \begin{bmatrix} ZX_1 & ZX_2 & ZX_3 & ZX_4 & ZX_5 & ZX_8 \\ & ZX_9 & ZX_{10} & ZX_{11} & ZX_{13} & ZX_{14} \\ & & ZX_{15} & ZX_{18} & ZX_{19} & ZX_{20} & ZX_{21} \end{bmatrix}$$

$$\beta = \begin{bmatrix} 3.6556 \\ -9.9724 \\ -4.0832 \\ 0.5333 \\ -4.0552 \\ 0.5247 \\ -4.1769 \\ 6.6800 \\ -1.9687 \\ 1.9278 \\ -1.2151 \\ -0.9974 \\ 8.0779 \\ 1.5105 \\ 4.7571 \\ -0.5327 \end{bmatrix}, c = [0.1159]$$

The adjusted R^2 of the model is 0.9849, AIC is -92.8482 , and MSE is 0.0004. Although this model contains more variables, due to the possible multicollinearity between the variables, the regression coefficients of some variables are negative.

4.1.3. Stepwise Selection

The two-way stepwise regression finally selected two variables (real estate development investment (10,000 yuan) (X_5), total industrial profit above designated size (100 million yuan) (X_{15}), and the regression equation is:

$$ZY = X\beta + b$$

$$X = [ZX_5 \quad ZX_{15}], \beta = \begin{bmatrix} -1.2053 \\ 1.3972 \end{bmatrix}, b = [0.3342]$$

The adjusted R^2 of the model is 0.9569, AIC is -67.7615 , and MSE is 0.0026. Although the model is simple, it can still explain part of the variation in the minimum living standard.

5. Experimental Results and Analysis

5.1. Minimum Living Standard Prediction Model

5.1.1. Pearson Correlation Coefficient between Minimum Living Standard and Economic Indicators

In this study, we employed the Pearson correlation coefficient to measure the correlation between various economic indicators and the minimum subsistence allowance standard. **Figure 1** displays the heatmap of the Pearson correlation coefficients among all variables, while **Figure 2** illustrates the ranking of these economic indicators in relation to the subsistence allowance standard.

From **Figure 1**, it can be observed that the minimum per capita consumption expenditure (ten thousand yuan/year) (X_1), industrial revenue above designated size (hundred million yuan), and total industrial profits above designated size

(hundred million yuan) (X_{15}) exhibit strong positive correlations with the minimum subsistence allowance standard (ten thousand yuan/year) (Y), with correlation coefficients approaching or exceeding 0.6. Conversely, the consumer price index shows a weaker correlation.

In **Figure 2**, we can clearly see that the minimum per capita consumption expenditure (ten thousand yuan/year) (X_1) has the highest correlation with the subsistence allowance standard, followed by industrial revenue above designated size and total industrial profits above designated size (X_{15}).

These results indicate that the minimum per capita consumption expenditure (ten thousand yuan/year) (X_1) may be one of the significant factors influencing

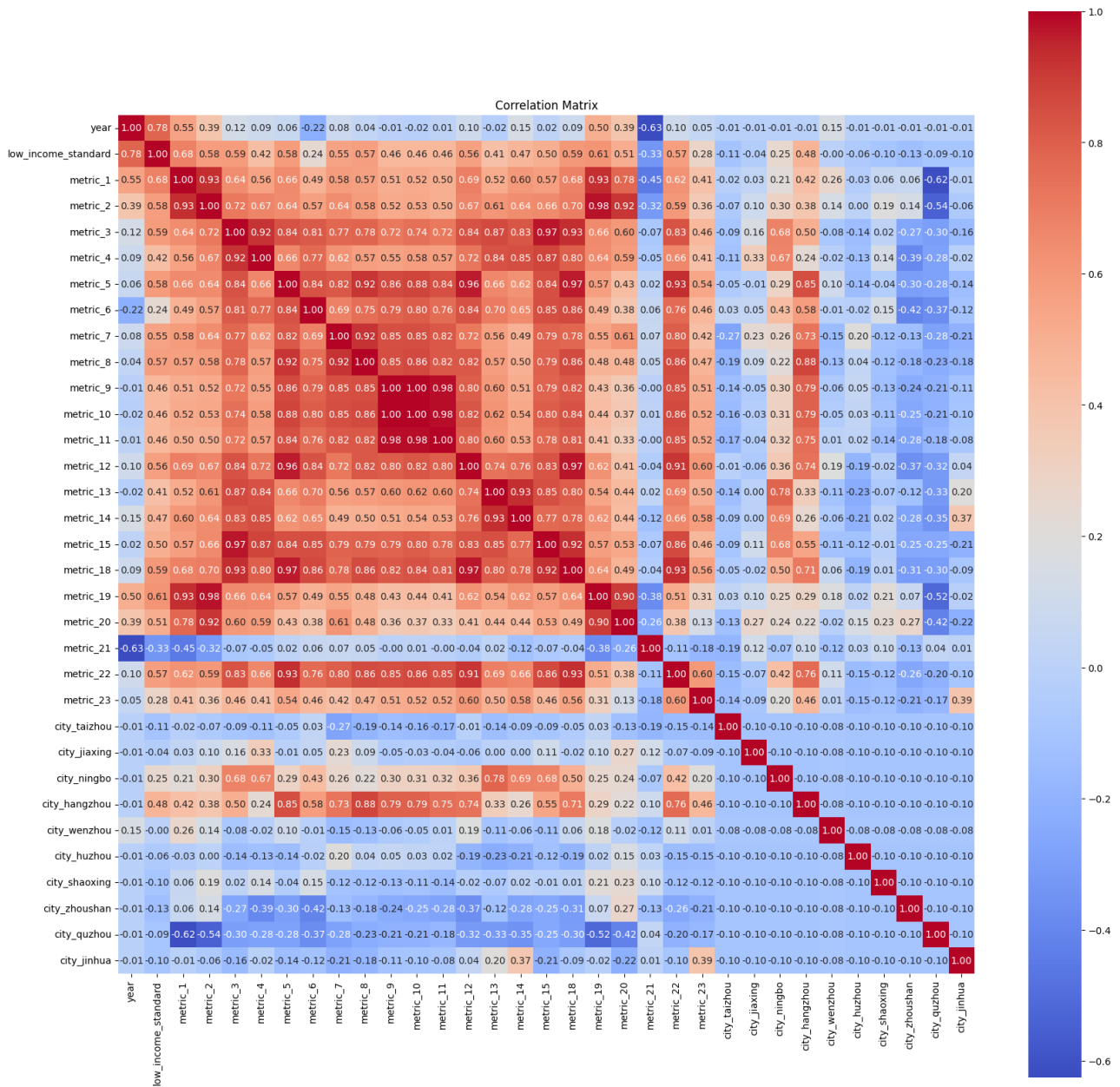


Figure 1. Pearson correlation coefficient heat map.

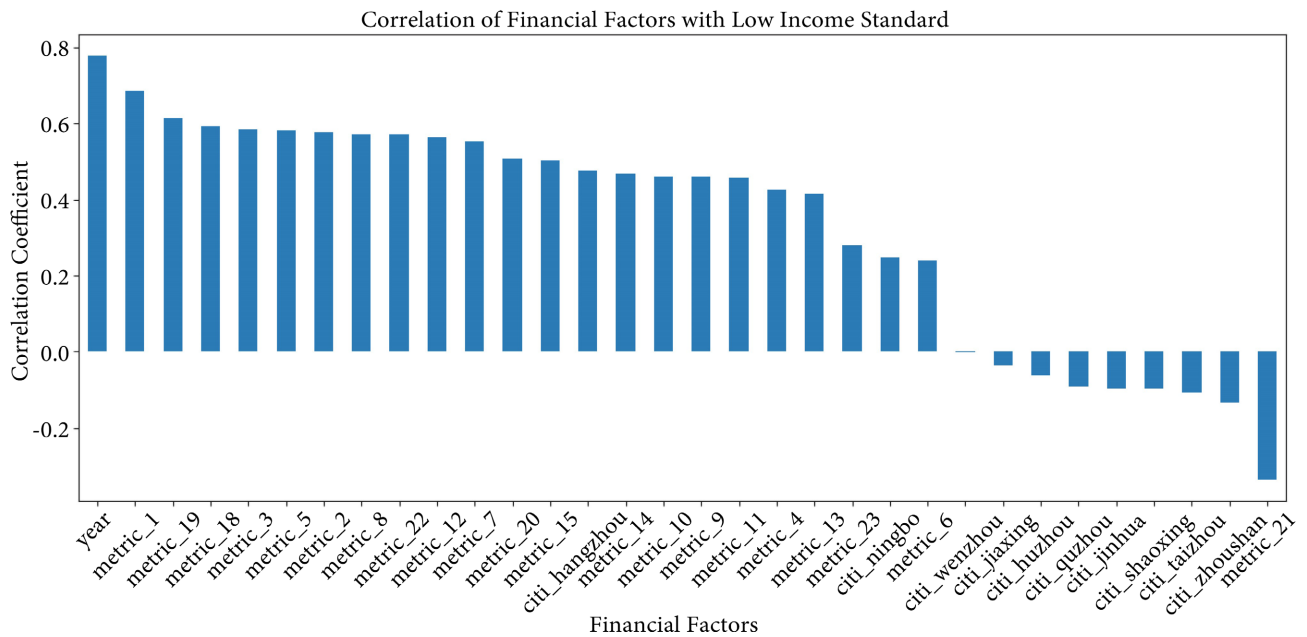


Figure 2. Correlation ranking chart between economic indicators and minimum living standard.

the subsistence allowance standard. The findings from this correlation analysis provide important support for the subsequent multiple linear regression modeling and help us understand the direction and strength of the impact of various economic indicators on the subsistence allowance standard.

5.1.2. Domain Interpretation of Principal Components

The first three principal components (PCs) extracted via PCA cumulatively explain 89.7% of variance. Combining factor loadings (Figure 1) with the Zhejiang Common Prosperity Demonstration Zone Construction Plan (2021 – 2025), the PCs are interpreted as follows:

- **PC1 (Digitally-Driven Economic Factor):** High-load variables include digital economy core industry value added (0.91), per capita disposable income (0.87), and industrial profits (0.84), capturing Zhejiang’s digital-real economy integration under the “14th Five-Year Plan”.
- **PC2 (Livelihood-Capital Equilibrium Factor):** Negative loading for real estate development investment (−0.76) juxtaposed with positive loadings for social welfare expenditure (0.68) and CPI (0.63), revealing governance challenges where capital concentration may crowd out essential livelihood investments.
- **PC3 (Open-Economy Resilience Factor):** Cross-border e-commerce imports/exports (0.89) and high-tech industry FDI proportion (0.82), aligning with Zhejiang’s “dual circulation” strategy for enhancing external economic resilience. Future studies may develop intelligent governance targeting models to align “data-policy-outcome” feedback mechanisms, offering actionable regulatory dimensions for policymakers.

5.1.3. Performance Comparison of Different Regression Models

The model performance was evaluated using adjusted R^2 , AIC and MSE. The re-

sults are shown in **Table 4**.

Table 4. Performance comparison of different regression models.

Method	Adjusted R ²	AIC	MSE
Forward Selection	0.6511	-15.4615	0.0211
Backward Elimination	0.9849	-92.8482	0.0004
Stepwise Selection	0.6449	-16.6384	0.0237

Model evaluation indicator description:

When comparing the performance of different regression models, common metrics are typically used to assess the quality of the models, including adjusted R², AIC, and MSE.

- **Adjusted R²:** R² represents the model's ability to explain the variability of the dependent variable, with values ranging from 0 to 1; a higher value indicates stronger explanatory power. However, adding variables can increase R², which may lead to overfitting. Therefore, the adjusted R², which accounts for the model's complexity, corrects the R² value and serves as an appropriate metric for comparing models with different numbers of variables. A higher adjusted R² indicates better explanatory power of the model for the data.
- **AIC (Akaike Information Criterion):** A criterion used to balance the complexity and goodness of fit of a model; a lower value indicates a better model. AIC helps to determine whether a model is overly complex, thereby avoiding overfitting.
- **MSE (Mean Squared Error):** A metric that measures the difference between predicted values and actual values; a smaller value indicates greater accuracy in the model's predictions.

The model obtained through the backward elimination method exhibits the highest adjusted R² value and the lowest AIC and MSE, indicating that it performs best in terms of explanatory power and goodness of fit.

6. Conclusion and Outlook

6.1. Conclusion

In the study of low-income standards, this paper employs a multiple linear regression model to analyze the impact of various economic indicators on these standards. The research demonstrates that the model obtained through the backward elimination method exhibits optimal performance in terms of adjusted R², AIC, and MSE metrics. However, some variables have negative regression coefficients, which may indicate complex interactions or multicollinearity among the variables. These phenomena warrant further investigation in future research, and it may be beneficial to consider employing more complex models or regularization techniques (such as Lasso regression or Ridge regression) to enhance the stability and predictive accuracy of the model.

In summary, this paper effectively combines the advantages of multiple linear regression models and PCA, leveraging the capabilities of high-dimensional data processing and nonlinear modeling to provide an accurate and stable solution for predicting the factors influencing low-income standards. This approach offers robust data support for government departments in social assistance decision-making and provides a theoretical and technical foundation for the future optimization of social security systems and the identification of aid recipients.

6.2. In-Depth Analysis of Negative Coefficients

The negative coefficient for real estate development investment (ZX_5 , $\beta = -0.9721$) may stem from two mechanisms:

Resource Misallocation Effect: Excessive reliance on real estate in certain counties distorts fiscal structures, compressing social welfare budgets. For example, in 2022, real estate taxes accounted for over 40% in a Zhejiang county, yet MSA increased by only 2.3%.

Statistical Coupling Bias: In high-dimensional causal inference models, traditional regression may amplify spurious negative correlations if multicollinear variables coexist. Real estate investment and industrial profits (ZX_{15}) exhibit a hedging effect after Principal Component Analysis (PCA).

To validate these hypotheses, future research could employ elastic net regression to address collinearity and variable selection simultaneously or introduce a “policy shock-effect lag model” to analyze long-term MSA impacts of housing purchase restrictions.

6.3. Outlook

Although this study has achieved certain results, there are still many aspects that warrant further exploration and improvement. Future work can be optimized in the following two areas:

****Optimization of Feature Selection:**** Feature selection has a critical impact on model performance. In the future, more advanced feature selection algorithms can be explored, such as tree-based feature importance evaluation and LASSO regression, to enhance the model’s explanatory power and predictive accuracy.

****Real-time Prediction and Application:**** Future research can apply the model to actual assistance work to achieve real-time prediction and decision support. This will help validate the model’s performance in practical operations and allow for adjustments and optimizations based on real-time data.

In summary, with the continuous development of data science and machine learning technologies, there will be more research and application opportunities in the field of low-income standard prediction. By continuously optimizing models and incorporating more data sources, the accuracy and reliability of predictions can be further enhanced, providing more precise data support for social assistance work, ultimately aiming to facilitate the mutual promotion of improved living conditions.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] State Council of the People's Republic of China (2014) Interim Measures for Social Assistance. https://www.gov.cn/gongbao/content/2014/content_2629930.htm
- [2] State Council of the People's Republic of China (2024) Guiding Opinions on Further Improving the Determination and Adjustment of Minimum Living Security Standards (Minfa [2024] No. 16). <https://www.mca.gov.cn/gdnps/mobile/content.jsp?id=1662004999980000352>
- [3] General Office of the Central Committee of the Communist Party of China, General Office of the State Council (2020) Opinions on Reforming and Improving the Social Assistance System. https://www.gov.cn/zhengce/2020-08/25/content_5537371.htm
- [4] Yuan, J.C., Liu, Z. and Gao, Y.S. (2019) Innovative Research on Calculation Methods of Minimum Living Standard for Urban and Rural Residents. *China Collective Economy*, No. 17, 88-92.
- [5] He, Y.Z. and Fan, C.J. (2022) Application of Big Data in Social Assistance. *Reform and Opening*, No. 14, 53-57. <https://doi.org/10.16653/j.cnki.32-1034/f.2022.014.008>
- [6] Jiang, B. and Zhang, Y.Y. (2022) Social Work Improves the Precision, Warmth, and Brightness of Social Assistance. *China Social Work*, No. 25, 29-30.
- [7] Zhou, J. (2022) Application Value and Dilemma Relief of Intelligent Algorithm in Social Assistance. *Journal of Hunan Radio and Television University*, No. 2, 69-75. <https://doi.org/10.19785/j.cnki.hnndxb.2022.02.011>
- [8] Hu, F.X. and Yang, L.J. (2025) Visual Analysis of Research on Minimum Living Security System Based on Citespace. *Journal of Yulin University*, **35**, 98-107. <https://doi.org/10.16752/j.cnki.jylu.2025.01.016>
- [9] Zhang, T. (2023) Analysis and Prediction of Factors Influencing Residents' Income Based on PCA and Regression Combination Model: A Case Study of Tibet Autonomous Region. *Business Exhibition Economy*, No.18, 18-21. <https://doi.org/10.19995/j.cnki.CN10-1617/F7.2023.18.018>
- [10] Liao, A.H., Huang, D.L. and Wang, E.W. (2020) Research on Minimum Living Standard in the Pearl River Delta Based on Multiple Linear Regression and Cluster Analysis. *Practice and Understanding of Mathematics*, **50**, 308-314.
- [11] Zhejiang Provincial Bureau of Statistics (n.d.) Zhejiang Statistical Yearbook. <https://tjj.zj.gov.cn/col/col1525563/index.html>
- [12] Ma, G.Y. (2025) On the Multidimensional Dynamic Model of the Framework System of "Artificial Intelligence Law." *Oriental Law*, No. 1, 186-200. <https://doi.org/10.19404/j.cnki.dffx.2025.01.002>
- [13] Zhao, C. (2025) The Logical Approach and Optimization Path of Building a Social Governance Community in the Digital Era. *Journal of Tianjin Normal University (Social Science Edition)*, No. 2, 56-64. <https://doi.org/10.20247/j.issn1671-1106.2025.02.006>
- [14] Wang, H.B. (2025) Using Data as the Tool, Specialization as the Core: The Practical Mechanism of Digital Social Work in the Digital Society. *Study and Practice*, No. 2, 53-63. <https://doi.org/10.19624/j.cnki.cn42-1005/c.2025.02.012>
- [15] General Office of the Central Committee of the Communist Party of China, General Office of the State Council (2020) Opinions on Reforming and Improving the Social

Assistance System. https://www.gov.cn/zhengce/2020-08/25/content_5537371.htm

- [16] Hu, Y.M., Wu, Y.J. and Li, T. (2019) Fiscal Decentralization, Promotion Incentives, and the Efficiency of Fiscal Social Security Expenditure: An Empirical Study Based on a Three-Stage DEA Model. *Journal of Hunan University of Finance and Economics*, 35, 24-33. <https://doi.org/10.16546/j.cnki.cn43-1510/f.2019.02.003>