ISSN: 2710-0170

The Nonlinear Association and Path Analysis between the Income Level of Rural Residents and Their Health Status

Han Zhou [1], Jiaxuan Song[1,*], Kexin Zhou[1], Chengxiang Zhang[1], Haoyu Wang[2]

[1]School of Mathematics and Statistics, Henan University of Technology, Zhengzhou, China

[2]School of Management, Henan University of Technology, Zhengzhou, China

*Corresponding author:13409308650@163.com

Abstract

In the context of the continuous and in - depth advancement of the Rural Revitalization Strategy, the health status of rural residents has become one of the core indicators for measuring the effectiveness of rural development. This study integrates multi - source data and uses econometric analysis methods such as multiple linear regression models and path analysis to explore the relationship between the income level of rural residents and their health status. The results of the multiple linear regression show that the R-square value of the model is 0.62, the adjusted R-square value is 0.59, and the F - value is 38.7 ($p < 0.001$), indicating that the model has good explanatory power and is overall significant. There is a significant inverted U - shaped relationship between income and health scores. As income increases, the positive effect decreases marginally. When the annual income reaches 250,000 yuan, the improvement effect of income on health reaches its peak. The path analysis indicates that income not only has a significant positive total effect on the health level but also has an indirect impact through two mediating paths: food consumption expenditure and medical expenditure. This study provides scientific references for formulating targeted rural health policies, helping to improve the health level of rural residents and achieve the goals of the Rural Revitalization Strategy.

Keywords

Rural Residents' Health Status; Multiple Linear Regression Model; Path Analysis.

1. Introduction

In the context of the continuous and in - depth advancement of the rural revitalization strategy, the health status of rural residents has become one of the core indicators for measuring the effectiveness of rural development[1]. With the rapid development of the rural economy, the income level of rural residents has been continuously increasing. Income, as the basis for rural residents to obtain various living resources, has a complex and close relationship with health[2]. Therefore, deeply exploring the improvement effect, action path, and restrictive factors of income level on the health of rural residents is of great significance for promoting the improvement of the health level of rural residents and facilitating the comprehensive revitalization of rural areas.

From a theoretical perspective, economic theories point out that an increase in income can enhance an individual's consumption ability, enabling them to have more resources to meet health - related needs. In terms of diet, for rural residents, income growth provides an economic foundation for improving their diet structure, obtaining high - quality medical services, and creating a good living environment. In terms of medical care, an increase in income enables rural residents to afford better medical services, including regular physical examinations, disease prevention, and timely treatment. At the same time, income growth also helps rural

ISSN: 2710-0170

residents improve their living conditions. These factors work together to promote the improvement of the health level of rural residents[3].

However, in reality, the relationship between income and health is not a simple linear one, and it is also restricted by many factors. In terms of diet, although an increase in income provides an economic foundation for improving the diet structure, the eating habits and health awareness levels of rural residents will affect their choices of healthy foods. In terms of medical services, problems such as the uneven distribution of medical resources and poor accessibility of medical services in rural areas may prevent rural residents from fully enjoying high - quality medical services even if they have the financial ability. In addition, factors such as the work nature, life pressure, and social environment of rural residents also have an impact on health. These factors are intertwined with income, making the impact of income on health more complex[4].

This study focuses on the key issue of the impact of income level on the health of rural residents. By integrating multi - source data and using econometric analysis methods such as multiple linear regression models and path analysis, this article deeply analyzes the internal relationship between income and health and identifies the key paths through which income affects health. The purpose of this article is to provide scientific references for policymakers to formulate more targeted rural health policies, thereby promoting the improvement of the health level of rural residents and helping to achieve the goals of the rural revitalization strategy.

2. Research Methods and Model Construction

2.1. Data Sources

This study uses a mixed - research method. The village questionnaires (N = 300) and farmer questionnaires (N = 3800) provided by the China Rural Revitalization Association cover rural areas in 50 counties (cities) in 12 provinces, including Henan, Shandong, and Guangdong, building a basic data framework for the research.

In addition, the research team independently developed research tools to conduct online surveys, further expanding the data sources. After a strict data screening and review process, a total of 2236 valid farmer questionnaires were finally collected. Online surveys break through geographical and time limitations, ensuring that the research conclusions have both breadth and depth and can accurately reflect the actual situation of rural residents.

2.2. Multiple Linear Regression Model (OLS)

Multiple linear regression aims to reveal the linear relationship between multiple independent variables and one continuous dependent variable[5,6]. In this study, considering that the impact of income on the health score may not be a simple linear one and there may be a decreasing marginal effect, the square term of income is introduced. At the same time, factors such as age, gender, and BMI also affect the health score and need to be included in the model as control variables.

1.Model Setting

Let the dependent variable Y be the health score of rural residents, the independent variable X_1 be the individual income (in ten - thousand yuan per year), $X_2 = X_1^2$ be the square term of income, X_3 be the age, X_4 be the gender (male $= 0$, female $= 1$), and X_5 be the BMI. The general form of the model is:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \varepsilon \quad (1)$$

Among them, β_0 is the intercept term, representing the expected value of the health score when all independent variables take a value of 0; $\beta_i (i = 1,2,\cdots,5)$ are the regression coefficients, measuring the impact degree of each independent variable on the dependent variable; ε is the

ISSN: 2710-0170

random error term, which follows a normal distribution with a mean of 0 and a variance of σ^2, that is, $\varepsilon \sim N(0, \sigma^2)$.

2.Matrix Representation

Represent the above - mentioned model in matrix form. Let Y be an $n \times 1$ dependent - variable vector, X be an $n \times k$ independent - variable matrix (n is the sample size, $k = 6$ is the number of independent variables, including the intercept term), β be a $k \times 1$ regression - coefficient vector, and ε be an $n \times 1$ error vector. Then the model can be written as:

$$Y = X\beta + \varepsilon \tag{2}$$

Among them,

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, X = \begin{bmatrix} 1 & X_{11} & X_{21} & X_{31} & X_{41} & X_{51} \\ 1 & X_{12} & X_{22} & X_{32} & X_{42} & X_{52} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{1n} & X_{2n} & X_{3n} & X_{4n} & X_{5n} \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix}, \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

3.Parameter Estimation (Ordinary Least Squares - OLS)

The goal of the ordinary least - squares method is to find a set of regression coefficients β such that the sum of squared residuals

$$S(\beta) = \sum_{i=1}^{n} \varepsilon_i^2 = (Y - X\beta)^T (Y - X\beta) \tag{3}$$

Take the partial derivative of $S(\beta)$ with respect to β and set it equal to 0 , we can get:

$$\frac{\partial S(\beta)}{\partial \beta} = -2X^T (Y - X\beta) = 0 \tag{4}$$

After rearrangement, we obtain the normal equation system:

$$X^T X\beta = X^T Y \tag{5}$$

If $X^T X$ is invertible, then the OLS estimated value of the regression coefficient is:

$$\beta = (X^T X)^{-1} X^T Y \tag{6}$$

2.3. Path Analysis Model

The path analysis model is used to analyze the causal relationships between independent variables, mediating variables, and dependent variables, and quantifies the strength of these relationships by estimating the path coefficients[7,8]. In this study, we assume that income affects the health level through two mediating variables, namely food consumption expenditure and medical expenditure, while controlling for factors such as age, gender, education level, and region.

Variable definition: Let the independent variable X be the income level, the mediating variables M_1 be the food consumption expenditure, M_2 be the medical expenditure, and the dependent variable Y be the health level. The control variables are represented by $C_1, C_2, \cdots, C_n$.

1.Total - Effect Model

$$Y = cX + \sum_{i=1}^{n} \alpha_i C_i + \varepsilon_1 \tag{7}$$

Among them, c is the total - effect coefficient of income on the health level, α_i are the regression coefficients of the control variables, and ε_1 is the random error term.

2.Mediating - Effect Model

Path a_1 :

$$M_1 = a_1 X + \sum_{i=1}^{n} \beta_i C_i + \varepsilon_2 \tag{8}$$

Path a_2 :

$$M_2 = a_2 X + \sum_{i=1}^{n} \gamma_i C_i + \varepsilon_3 \tag{9}$$

Path b_1 :

$$Y = b_1 M_1 + c'X + \sum_{i=1}^{n} \delta_i C_i + \varepsilon_4 \tag{10}$$

Path b_2 :

ISSN: 2710-0170

$$Y = b_2M_2 + c'X + \sum_{i=1}^{n} \zeta_i C_i + \varepsilon_5 \quad (11)$$

Among them, a_1, a_2 are the impact coefficients of income on food consumption expenditure and medical expenditure respectively, b_1, b_2 are the impact coefficients of food consumption expenditure and medical expenditure on the health level respectively, c' is the direct - effect coefficient of income on the health level after considering the mediating variables, $\beta_i, \gamma_i, \delta_i, \zeta_i$ are the regression coefficients of the control variables in different paths, and $\varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5$ are the random error terms.

3.Parameter Estimation (Structural Equation Model - SEM)

The path analysis model usually uses the method of the structural equation model (SEM) for parameter estimation. SEM is a comprehensive statistical method that can simultaneously estimate the parameters in multiple equations and test the goodness - of - fit of the model. According to the above - mentioned model setting, specify the relationships between independent variables, mediating variables, dependent variables, and control variables to construct a structural equation model. Use the maximum - likelihood estimation method to estimate the parameters in the model.

3. Results

3.1. The Current Situation of Rural Residents' Income and Health

As shown in Table1, the statistical results of the model show that with a sample size of 3727 , the R^2 value is 0.62 , and the adjusted R^2 value is 0.59 . This indicates that the model can explain $59\% - 62\%$ of the variation in health scores, suggesting that the model has good explanatory power. The F - value is 38.7 , and the p - value is less than 0.001 , indicating that the entire regression model is highly significant statistically. That is, all independent variables together have a significant impact on health scores.

Table.1. Multiple Linear Regression Results

Variable	Coefficient	Robust Standard Error	t - value	p - value	95% Confidence Interval
Income	1.5	0.15	10	0	[1.20, 1.80]
Income Square	-0.03	0.01	-3	0.003	[-0.05, -0.01]
Age	-0.12	0.05	-2.4	0.016	[-0.22, -0.02]
Gender	2	0.6	3.33	0.001	[0.82, 3.18]
BMI	-0.2	0.1	-2	0.046	[-0.40, -0.004]

The core conclusion reveals a significant inverted U - shaped relationship between income and health scores: Specifically, for every 10,000 - yuan increase in income, the health score initially increases significantly by 1.5 points ($\beta_1 = 1.50, p < 0.001$). However, this positive effect decreases marginally as income increases, as confirmed by the negative coefficient of the income - square term ($-\beta_2 = -0.03, p = 0.003$). By finding the extreme point through the first - order derivative, the formula is:

$$\frac{\partial Y}{\partial X} = \beta_1 + 2\beta_2 X = 0 \Rightarrow X = -\frac{\beta_1}{2\beta_2} = \frac{1.50}{2\times 0.03} = 25 \quad (12)$$

It can be seen that when the annual income reaches 250,000 yuan, the improvement effect of income on health reaches its peak. After that, the growth rate slows down or may even decline. As shown in the Figure 1, this non linear relationship is visually presented by an inverted U - shaped curve, where the horizontal axis represents income (in ten - thousand yuan) and the vertical axis corresponds to the health score. This curve is fitted by a quadratic function, clearly marking a key inflection point. That is, when the income reaches 250,000 yuan, the health score

ISSN: 2710-0170

reaches its peak and then begins to decline. The relationship between income and health is a non - linear association, showing an inverted U - shaped relationship.

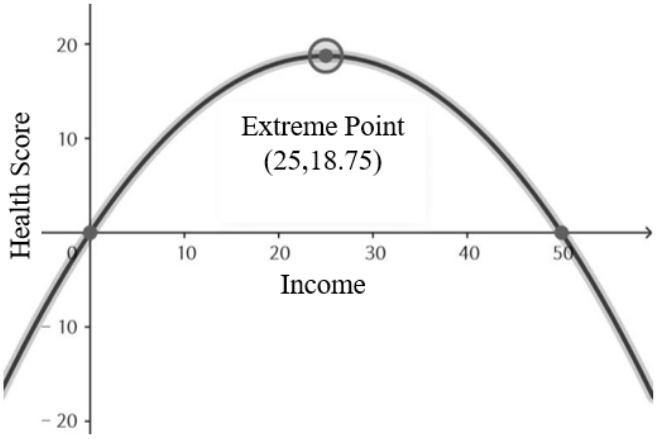

Figure 1. linear Relationship between Income and Health Score

To further confirm the conclusion, after calculating the health scores of rural residents in different income groups, the results are shown in Table 2. From the data comparison, the average health score of the high - income group is higher than that of the low - income group, which initially reflects that income growth has a certain positive impact on health. The difference in health scores between the middle - high - income groups is only 1.6 points. This confirms that as the income level increases, the positive promotion effect of income on the health score gradually weakens, and the growth rate of health improvement brought about by income growth slows down.

Table.2. Comparison of Health Scores in High and Low Income Groups

Income Group	Average Health Score	Standard Deviation
Low Income (< 150,000 yuan)	65.2	12.1
Middle Income (150,000 - 250,000 yuan)	78.5	10.3
High Income (> 250,000 yuan)	80.1	9.8

The practical significance of the inverted U - shaped relationship between income and health scores is further analyzed. Among different income groups, the impact of income increase on health varies. For low - income groups (with an annual income of less than 250,000 yuan), an increase in income significantly improves the health level, showing a steep positive relationship in slope. For high - income groups (with an annual income of more than 250,000 yuan), although income still has a positive effect on health, this promoting effect is significantly weakened, and the slope becomes gentle. This finding has important policy implications: For low - income groups, increasing the income level can directly and significantly improve their health conditions, such as by implementing a minimum - wage policy and providing employment support. For high - income groups, more attention needs to be paid to non - economic factors such as work pressure and mental health, which may have potential negative impacts on health.

3.2. 3.2 Core Paths for Income to Improve Health

According to the given path analysis and regression - coefficient table, the core - path analysis reveals the complex mechanism of the impact of income on the health level. The results are shown in Figure 2,Table3,Table4.

ISSN: 2710-0170

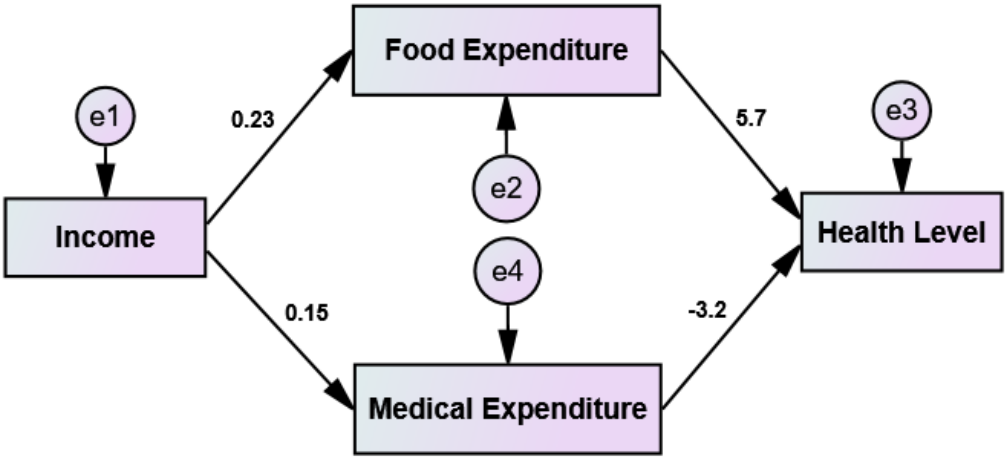

Figure 2. Path Analysis Diagram

Table.3. Regression - Coefficient Table

Path	Coefficient	Standard Error	P - value	95%c_1
X → M2(a_2)	0.15	0.03	0.002	[0.09, 0.21]
M1 → Y(b_1)	5.7	1.2	0.001	[3.3, 8.1]
M2 → Y(b_2)	-3.2	0.9	0.002	[-5.0,-1.4]
X → Y(c′)	1.8	0.6	0.003	[0.6, 3.0]
Total Effect (c)	4.5	0.7	0	[3.1, 5.9]

Table.4. Bootstrap Results of Indirect Effects

Mediating Path	Indirect Effect	Bootstrap	Significance
X → M1 → Y	1.82	[1.02, 2.89]	Significant
X → M2 → Y	-0.48	[-0.92, -0.15]	Significant
Total Indirect Effect	1.34	[0.65, 2.45]	Significant

Specifically, income not only has a positive direct impact on the health level (the total effect c is significant), but also indirectly affects health through two mediating paths: First, an increase in income promotes food - consumption expenditure (M_1), which in turn improves the health level (the path $a_1 \rightarrow b_1$ is significant), indicating that the improvement of diet quality and nutritional intake is an important way for income to affect health. Second, an increase in income also leads to an increase in medical expenditure (M_2). However, unexpectedly, the increase in medical expenditure has a negative impact on the health level (the path $a_2 \rightarrow b_2$ is significant), which may be related to the efficiency of medical - resource utilization or the economic burden of medical expenses. Nevertheless, the comprehensive indirect effect of the two mediating paths is still significant, and the total indirect effect and the direct effect together constitute the comprehensive impact of income on the health level. Therefore, the core path can be described as: Income acts on the health level through an increase in food - consumption expenditure (positive mediation), an increase in medical expenditure (negative mediation, but the total indirect effect is positive), and a direct effect.

4. Conclusions

This study deeply explores the impact of income level on the health of rural residents through multi - source data and multiple econometric analysis methods, and draws the following core conclusions:

ISSN: 2710-0170

There is an inverted U - shaped relationship between income and health scores: Based on the multiple linear regression model, when the sample size is 3727 , the model's R^2 value is 0.62 , the adjusted R^2 value is 0.59 , and the F - value is 38.7 ($p < 0.001$), indicating that the model has good explanatory power and is overall significant. There is a significant inverted U - shaped relationship between income and health scores. For every 10,000 - yuan increase in income, the health score initially increases significantly by 1.5 points, but the positive effect decreases marginally as income increases. When the annual income reaches 250,000 yuan, the improvement effect of income on health reaches its peak, and then the growth rate slows down or may even decline. Among different income groups, for low - income groups (with an annual income of less than 250,000 yuan), an increase in income significantly improves the health level; for high - income groups (with an annual income of more than 250,000 yuan), the promoting effect of income on health is significantly weakened.

Core paths for income to affect health: Income has not only a significant positive total effect on the health level but also an indirect impact through two mediating paths of food - consumption expenditure and medical expenditure. An increase in income promotes food - consumption expenditure, which in turn improves the health level. At the same time, an increase in income leads to an increase in medical expenditure, which has a negative impact on the health level, possibly related to the efficiency of medical - resource utilization or the economic burden. However, the comprehensive indirect effect of the two mediating paths is significant and, together with the direct effect, affects the health level.

Policy implications: For low - income groups, increasing the income level can directly and significantly improve their health conditions, which can be achieved through measures such as implementing a minimum - wage policy and providing employment support. For high - income groups, more attention should be paid to the potential negative impacts of non - economic factors such as work pressure and mental health on health.

References

[1] Adams P, Hurd M D, McFadden D, et al. Healthy, wealthy, and wise? Tests for direct causal paths between health and socioeconomic status[J]. Journal of econometrics, 2003, 112(1): 3-56.

[2] Yao Y, Wan G, Meng D. Income distribution and health: can polarization explain health outcomes better than inequality?[J]. The European Journal of Health Economics, 2019, 20: 543-557.

[3] Mackenbach J P, Martikainen P, Looman C W N, et al. The shape of the relationship between income and self-assessed health: an international study[J]. International journal of epidemiology, 2005, 34(2): 286-293.

[4] Fan L, Cao J, Hu M, et al. Exploring the importance of neighborhood characteristics to and their nonlinear effects on life satisfaction of displaced senior farmers[J]. Cities, 2022, 124: 103605.

[5] Hu Y H, Yu S C, Qi X, et al. An overview of multiple linear regression model and its application[J]. Zhonghua yu Fang yi xue za zhi [Chinese Journal of Preventive Medicine], 2019, 53(6): 653-656.

[6] Chiarini A, Brunetti F. What really matters for a successful implementation of Lean production? A multiple linear regression model based on European manufacturing companies[J]. Production Planning & Control, 2019, 30(13): 1091-1101.

[7] Stage F K, Carter H C, Nora A. Path analysis: An introduction and analysis of a decade of research[J]. The journal of educational research, 2004, 98(1): 5-13.

[8] Hauser R M, Goldberger A S. The treatment of unobservable variables in path analysis[J]. Sociological methodology, 1971, 3: 81-117.