



An Impact Factors Model of Resident Income in Henan Province

Shuwen Wang

Henan University Henan province, China, 475004

*Corresponding author: tarows@outlook.com

Abstract

The ultimate goal of economic development is for the well-being of people, that is why the factor of resident income is one of the most significant and crucial indicators of examining whether people are living happily. On this basis, this paper mainly focuses on the factors which are related with the resident income in Henan province. By analyzing the data in Rstudio, the paper found that GDP per capita, tax revenue and resident consumption in Henan province have strongly positive correlation with resident income level in Henan province. The rational income could contribute to the development of the society, but it may stimulate social contradictions once there is no reasonable resident income, which will impede the economic society, so it is a quite meaningful research for both China and Henan province, not only improving the distribution pattern, but also narrow the disparity between the rich and poor.

Keywords-*The consumption and income of resident; GDP; tax income; mathematical modeling*

1.INTRODUCTION

For the grassroots masses, the amount of personal income is the most intuitive reflection of the happiness index of residents [1], and it is necessary to study the information related to the income.

According to the previous research, resident income in Henan province has an unstable growth percentage [2], and so far, most people study the level of income based on the country, while there are very few studies in the cities, which is specified to the provinces. As we all know, resident income is inextricably linked between provinces and countries, while the development and strategies between provinces are quite different [3], also with the fact that Henan province is a large population and an agriculture province in China, especially the implementation of the two-child policy [4], so it is bound to make greater contributions to the blossom of Henan province through the conclusion made in Rstudio.

In order to find whether some significant factors have high correlation with income or not, this paper does the research with the GDP, tax revenue, resident consumption

and the unemployment in Henan province to find the relationship between each of them and resident income through modeling, not only does it offer some references for improving the distribution pattern, but also it could narrow the disparity between the rich and poor [5].

2.METHODOLOGY

2.1. Data Collection

As Table 1 shows, the data belongs to the second-hand data [6], which was got from the statistical yearbooks of local authorities “National Bureau of Statistics - local statistics website - Henan province - Statistical Yearbook”, and they are the per capita income and expenditure of households in each city, GDP per person, the fiscal and tax revenue of each city, and the number of unemployed people, respectively. In this part, this paper uses INCOME represents the per capita income and expenditure of households, TAX REVENUE represents the fiscal and tax revenue and UNEMPLOYMENT represents the number of unemployed people. What’s more, the data was in 2020, which was collected in 2021 and has properties of timeliness and integrity.

TABLE 1. STATISTICAL DATA

Cities in Henan	Income	GDP Per Capita	Consumption	Tax Revenue	Unemployment
Zhengzhou City	37275	96134	22991	870.07	75800

Kaifeng City	22647	49166	16899	102.79	27200
Luoyang City	28096	72872	19016	248.57	75600
Pingdingshan City	24929	49379	15196	123.2	18600
Anyang City	25530	42185	14834	125.62	26900
Hebi City	27110	62736	15496	44.28	13300
Xinxiang City	25497	48229	15878	137.84	67600
Jiaozuo City	28127	60384	18886	99.33	39300
Puyang City	22584	43908	13097	71.82	27900
Xuchang City	26935	78875	16296	117.96	30400
Luoyang City	25585	66239	16125	75.7	10000
Sanmenxia City	24864	71541	16799	84.75	22200
Nanyang City	23481	40315	14862	140.03	45800
Shangqiu City	21117	37439	14507	115.65	40600
Xinyang City	21861	44922	14994	84.01	23900
Zhoukou City	19143	36214	13411	105.41	30800
Zhumadian City	20520	40710	15243	115.45	27400
Jiyuan City	30013	96674	13254	44.47	9200

2.2. Data Analysis

1) *Descriptive Statistical Analysis.* Characteristic statistics of each variable are shown in Table 2. As we can

TABLE 2. CHARACTERISTIC STATISTICS OF EACH VARIABLE

Variables	Sample size	Unit	Mean	Standard variance	Lower quartile	Upper quartile	Minimum	Maximum
Resident income	18	yuan	25295	4170.465	22600	27066	19143	37275
GDP per capita	18	yuan	57662	19190.65	42616	70216	36214	96674
Resident consumption	18	yuan	15988	2399.793	14841	16673	13097	22991
Tax revenue	18	Billion yuan	150.39	185.11	84.19	125.02	44.28	870.07
unemployment	18	people	34028	20459.71	22625	40275	9200	75800

2) *Correlation Analysis.* Considering the inconsistent units, and in order to eliminate the influence of dimension [8],

see, there still have great differences between different standard variances, which also shows it is significant to study the income in Henan province [7].

this paper did some standardized processes. Finally, the data after four decimal places are retained as shown in Table 3.

TABLE 3. NORMALIZED DATA

Cities in Henan	Income	GDP Per Capita	Consumption	Tax Revenue	Unemployment
Zhengzhou City	2.8725	2.0047	2.9182	3.8879	2.0417
Kaifeng City	-0.635	-0.427	0.3796	-0.2571	-0.3337
Luoyang City	0.6716	0.7926	1.2618	0.5304	2.0319
Pingdingshan City	-0.0878	-0.4316	-0.33	-0.1469	-0.7514

Anyang City	0.0563	0.8065	-0.4809	-0.1338	-0.3484
Hebi City	0.4352	0.2644	-0.205	-0.5732	-1.0131
Xinxiang City	0.0484	-0.4916	-0.0458	-0.0678	1.6409
Jiaozuo City	0.679	0.1418	1.2076	-0.2758	0.2577
Puyang City	-0.6501	0.7167	-1.2047	-0.4244	-0.2995
Xuchang City	0.3932	1.1054	-0.1283	-0.1752	-0.1773
Luoyang City	0.0695	0.4469	-0.0571	-0.4035	-1.1744
Sanmenxia City	-0.1034	0.7232	-0.3379	-0.3546	-0.5781
Nanyang City	-0.435	-0.9039	-0.4692	-0.0559	0.5754
Shangqiu City	-1.0019	-1.0538	-0.6171	-0.1877	0.3212
Xinyang City	-0.8235	-0.6639	-0.4142	-0.3586	-0.495
Zhoukou City	-1.4752	-1.1176	-1.0739	-0.243	-0.1578
Zhumadian City	-1.145	-0.8834	-0.3104	-0.1887	-0.3239
Jiyuan City	-1.1312	2.0328	-1.1393	-0.5722	-1.2135

After that, there are four scatter plots about relationship between each variable and the resident

income in the Figure 1 below, which also added some auxiliary lines.

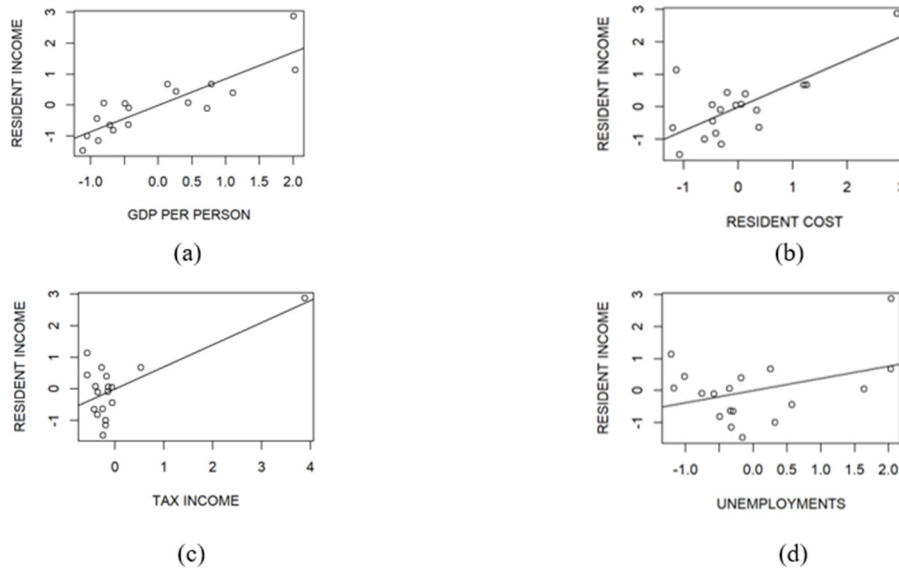


Figure 1. Scatter plots

From the above four related scatterplots, it can be seen that the linear relationship between per capita GDP or resident consumption level and resident income level is very significant, while the relationship between tax revenue, the number of unemployed and resident income

is not very significant. At this time, further correlation is needed.

Through the Rstudio programming code “cor.test”, we can see that the correlation coefficients of each variable

with the residents' income below:

- X1 (GDP per capita) and Y: 0.8567
- X2 (Resident consumption) and Y: 0.7256
- X3 (Tax revenue) and Y: 0.7002
- X4 (Unemployment) and Y: 0.3806

3.MODELING AND TEST

3.1. Preparations

The analysis results above shows that the correlation coefficient between the unemployment and resident income is relatively low, so the significance test -- ANOVA is carried out in Table 4, which will able to identify whether controllable factors (which can be quantitatively controlled or directly measured) have a significant impact on the test results and the magnitude of the impact [9].

TABLE 4. ANOVA RESULTS

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
data\$ GDP per capita	1	217022276	217022276	69.829	1.38e-06 ***
data\$ resident consumption	1	29258877	29258877	9.414	0.00898 **
data\$ tax revenue	1	8760209	8760209	4.819	0.01703 *
data\$ unemployment	1	233189	233189	0.075	0.78845
Residuals	13	40402659	3107897		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

As shown in Table 4 above, at the significance level of 0.05, only the P value of the unemployed is more than 0.05, so the null hypothesis is accepted [10], indicating that the relationship between the variable unemployed and residents' income is not significant.

To sum up, through various analyses, it can be found that X4, that is, the relationship between the number of unemployed and residents' income is very insignificant, so this factor is excluded, and the other three influencing factors (per capita GDP, consumption level, and tax revenue) are retained for further research and modeling analysis.

3.2. Modeling

It can be seen that there is a strong correlation between resident income and per capita GDP (X1), consumption level (X2), tax revenue (X3), and build the multiple linear regression model (below)

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \tag{1}$$

By doing linear regression in Rstudio, it found the intercept is almost close to 0, and the final model is:

$$Y = 0.6432X_1 + 0.1580X_2 + 0.2812X_3 \tag{2}$$

X1: per capita GDP

X2: resident consumption

X3: tax revenue

It can be seen that per capita GDP, resident consumption, and tax revenue is all positively correlated with resident income.

3.3. Model testing

1) *Heteroskedasticity Test.* Through the heteroskedasticity test, the result shows p-value=0.6996, which is greater than 0.05, and the null hypothesis is accepted, so it is considered that the model does not have heteroscedasticity and it is a homoscedastic model.

2) *Multicollinearity Test.* There will need to download the "car" package and use the variance inflation factor in the package to test. It can be seen from the test results that the VIF values of the three variables are independently 1.411776 (X1), 2.966311 (X2), 2.668259 (X3), all of them are less than 5 and far less than 10, so it is reasonable to think that there is no Multicollinearity among the three variables.

3) *Residual test.* > plot(lm,which = 1:4),

Through the regression in the research, the final four plots of the residual test are in the Figure 2.

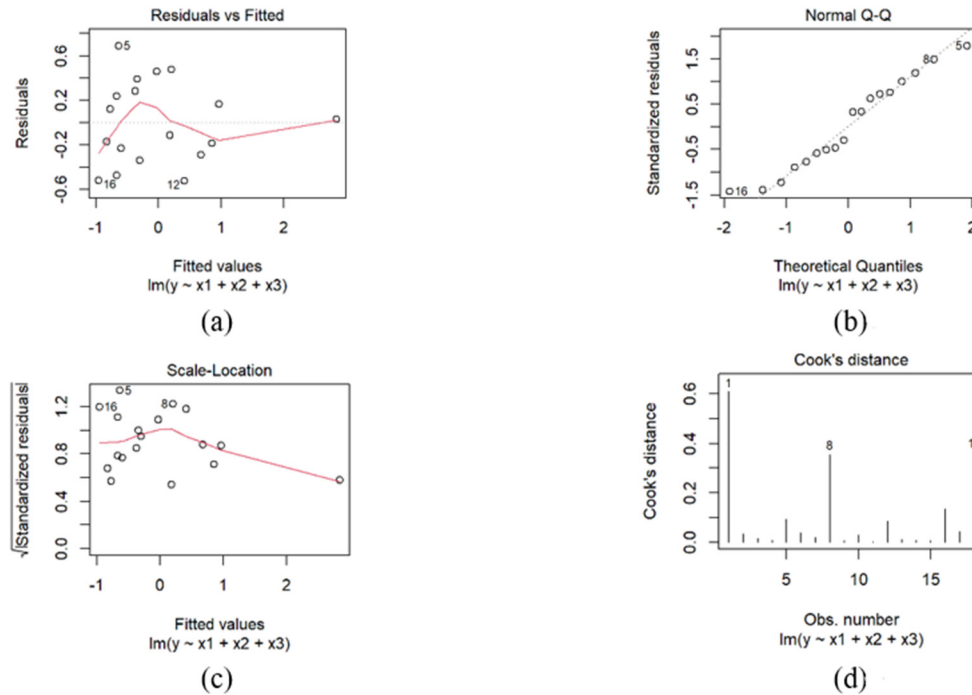


Figure 2. Model Diagnostic Plots

From Figure 2 (a), it shows that the residuals are not proportional to the estimated values, which means the residuals are basically irrelevant to the estimated values; From Figure 2 (b), it can be seen that the points and lines of normal Q-Q test plot are basically coincide, indicating that the residuals are approximately Normal distribution; From Figure 2 (c), it shows that the variance of the residual does not increase and change with the change of others, and the correlation is not significant; From Figure 2 (d), it can be seen that the cook distance is at most 0.6, indicating that there are no large outliers, which could

accept. To sum up, the model estimation of residuals is great, which means the residuals in the model have better properties.

4.MODEL PREDICTION

Find the data in 2021 through the same path, predicting the income level of residents in 2021 according to the model (Table 5), and compare with the real resident income in 2021(Table 6).

TABLE 5. PREDICT DATA

	Cities	fit	lwr	Upr
1	Zhengzhou City	37154.40	32044.39	42264.41
2	Kaifeng City	24056.26	20151.06	27961.45
3	Luoyang City	28874.54	24896.54	32862.55
4	Pingdingshan City	23747.73	19968.32	27527.13
5	Anyang City	22658.10	18816.00	26500.20
6	Hebi City	25197.18	21367.43	29026.94
7	Xinxiang City	23866.98	20080.62	27653.33
8	Jiaozuo City	26147.93	21838.56	30457.31
9	Puyang City	22081.21	18102.75	26059.68
10	Xuchang City	28139.44	24200.65	32078.23
11	Luoyang City	26058.56	22218.42	29898.70
12	Sanmenxia City	27042.05	23119.81	30964.29
13	Nanyang City	22495.69	18624.59	26366.78
14	Shangqiu City	21841.78	17942.82	25740.73
15	Xinyang City	22821.02	19020.50	26621.54
16	Zhoukou City	21304.75	17285.45	25324.06

17	Zhumadian City	22499.80	18658.78	26340.83
18	Jiyuan City	29326.58	24538.39	34114.76

TABLE 6. REAL DATA

Cities	Resident Income
Zhengzhou City	37275
Kaifeng City	22647
Luoyang City	28096
Pingdingshan City	24929
Anyang City	25530
Hebi City	27110
Xinxiang City	25497
Jiaozuo City	28127
Puyang City	22584
Xuchang City	26935
Luoyang City	25585
Sanmenxia City	24864
Nanyang City	23481
Shangqiu City	21117
Xinyang City	21861
Zhoukou City	19143
Zhumadian City	20520
Jiyuan City	30013

The comparison in Table 5 and Table 6 shows that the model prediction is roughly accurate, and it can be seen that the model have a good effect.

5.CONCLUSION

According to a series of Rstudio tests, this paper finally decided to exclude the variable “unemployment”, because it has a weak correlation with the explained variables, and finally established a model of a ternary linear regression equation, and the relationship among them is: resident income and per capita GDP, resident consumption, the tax revenue

$$Y=0.6432X1+0.1580X2+0.2812X3 \quad (3)$$

X1: GDP per capita;

X2: Resident consumption;

X3: Tax revenue

According to the obtained results, it is reasonable for Henan Province to choose GDP, resident consumption, and tax revenue to analyze resident income apart from some other uncertain factors, that means, when other factors remain unchanged, every 1 unit growth of GDP will increase 0.1580 units of resident income on average; for every 1 unit increase in resident consumption, resident income will increase by 0.6432 units on average; and for every 1 unit increase in tax revenue, resident income will increase by 0.2812 units on average.

However, as it known to all, there are definitely more than these factors related to residents' income. Such as urbanization, industrial progress, mechanism and economy also have magnificent impacts [11]. Therefore, there is still much space for improvement in this paper, such as developing new related factors, establishing nonlinear models [12], better fitting income models and so on, but it is undeniable that the indicators studied and the model established in this paper are bound to make a certain contribution to the improvement of the income level of residents in Henan Province, and this paper is refined from the national level to the provinces and cities, which not only realizes the transformation from “macro” to “micro”, but also becomes a major significance of this research.

ACKNOWLEDGEMENTS

In the process of writing, I encountered a lot of difficulties. Fortunately, all the teachers have given me great help. Thanks to the thesis guide, Ms. Wang Cuihong; the tutorial guide, Ms. Chaya, and the professors who gave me lectures. Thank them all for their support, without their support, I will not have the achievements I have now, and I will keep on going hard.

REFERENCES

- [1] Yangpeng Wu. Economic Analysis of the Impact of Residents' Income on Happiness Based on the Data

- of China Family Panel Studies [J]. World Scientific Research Journal, 2019, 5(9). Econometric Models [D] Harbin Institute of Technology, 2018.
- [2] Huijuan Li & Ruijun Duan, The influence factors of farmers' income and the strategies of income increasing in Henan [J]. Southern Agriculture, 2021, 15(29):148-152. DOI: 10.19415/j.cnki.1673-890x.2021.29.067.
- [3] Yan-fang Zhang, Xiao-zheng Pan, The Impact Analysis of Resident Income and Urbanization on Fertility in Shandong Province [C] Proceedings of 2019 International Conference on Education Reform, Management Innovation and Social Science (ERMIS 2019). 2019, 136-139.
- [4] Zilian Li, Income and Fertility: An Explanation of Changes in China's Fertility Rate [J]. Economic dynamics, 2016(05):37-48.
- [5] Xude Zhang & Jie Zhang. Increase property income and narrow the gap between rich and poor. Jiangnan Forum, 2008 (02),17+21.
- [6] Yongzhong Zhang, Application of Second-hand Data from the Internet in Statistical Surveys [J]. Journal of Luohe Vocational and Technical College (Comprehensive Edition), 2005(04):114-115.
- [7] Shuli Li.(2020). Research on Markowitz Portfolio Models with Mean and Variance Changes. Economic Outlook of the Bohai Rim (02),185-186. doi:10.16457/j.cnki.hbhjllw.2020.02.138.
- [8] Wei-hua Zhang & Ming-jun Zhao, The Influence of Undimensionalization on the Reliability of Comprehensive Evaluation Results and an Empirical Analysis [J]. Statistics and Information Forum, 2005(03):33-36.
- [9] Chongshuang Chen & Jiayin Tang & Ping He, Linear Regression Model Restructuring of Variance Analysis [J] statistic and decision, 2018, 34(07):71-75. DOI:10.13546/j.cnki.tjyjc.2018.07.016.
- [10] Malone Helen Evelyn, Coyne Imelda. Complementing the P-value from null-hypothesis significance testing with a Bayes factor from null-hypothesis Bayesian testing [J]. Nurse researcher, 2020, 28(4).
- [11] Xiaoling Song & Hongyue Guo. Influence Factors of the Urban-rural Residents' Income Gap: a Restudy with the Digital Inclusive Finance [C] Proceedings of 2017 4th International Conference on Business, Economics and Management (BUSEM 2017), 2017, pp.180-186.
- [12] Yanhong Chen. Research on Parameter Estimation Method for a Class of a Class of Nonlinear

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

