

Research on the Influencing Factors of Bank-Balance

Zhiqian Liu ¹ and Jiarui Tang ^{2, *}

¹ Cogdel Cranleigh High School, Wuhan, 430000, China

² Chongqing DEPU Foreign Language School, Chongqing, 400000, China

* Corresponding author: maxdongliang@tzc.edu.cn

Abstract. Today, mathematics is everywhere, especially in the field of finance. This paper investigates the main factors affecting income, this paper expounds the theoretical background of linear regression, then uses linear regression to analyze the correlation between the dependent variable and one or more independent variables, uses a linear equation to estimate the dependent variable, and uses a dataset of multiple economic variables to analyze and process the coefficients of linear regression. The dataset contains 32562 groups of data and this paper will use these data to show the factors that affect bank-balance. This paper considers China's GDP development, consumer price level, tax collection ratio, personal ability, and regional development level, and discusses the impact of various cross-cutting factors on income on this basis. Combined with China's development status and China's social situation, this paper gives a reasonable explanation of the influence of variables on personal income in the process of mathematical model exploration. The variables have significant impact is IsAdult, Gender, Job and Marital Status.

Keywords: Influencing factors, correlation, bank balance.

1. Introduction

Since the implementation of China's reform and opening up strategy in 1978, China has undergone earth-shaking changes in all aspects, especially in the economic field. The achievements of the domain are even more gratifying. According to the National Bureau of Statistics, the country's gross domestic product (GDP) in 1979 was 4100 billion yuan, and in 2020, despite the impact of the new crown pneumonia epidemic, China's economy still achieved contrarian growth, and the gross domestic product (GDP) exceeded 1,000,000 trillion yuan for the first time, reaching 1,015,986 trillion yuan, about 248 times that of 1979, which is the lowest in the history of the world economy [1, 2]. There is a great miracle. Along with the rising economic aggregate of the country, there is also the income of ordinary citizens, and in 2020, the per capita disposable income of residents in the whole country.

The income reached 32,189 yuan, of which the per capita disposable income of urban residents was 43,834 yuan and that of rural residents was 17,131 yuan, all of which have increased in different proportions compared with the previous year [3]. There are significant differences in average wages in different industries, such as in urban non-private units. The three industries with the highest average annual wages are: information transmission, software and information technology services (177,544 yuan), scientific research and \$139,851 for technical services and \$133,390 for finance. The three industries with the highest average annual wages for employees in the urban private sector are 10,1281 yuan for information transmission, software and information technology services, 82,930 yuan for financial services, and 72,233 yuan for scientific research and technical services. Overall, the industries with high average wages are relatively consistent [4]. The ultimate goal of economic development is to make people's lives happy, so the influencing factors of income, one of the indicators to examine whether people's lives are happy are very important [5].

Based on the 2013 survey data of the China General Social Survey (CGSS), a multi-level model analysis of the personal income of Chinese residents was carried out based on provinces, regions and municipalities directly under the central government [6]. First, this paper introduces how to use multiple linear regression functions to calculate factor data and the theoretical and practical foundations. Secondly, the idea of data cleansing and variable setting are introduced, and the process

of calculation with different data models is elaborated later, as well as the summary and analysis of the influencing factors of income [7]. This paper considers China's GDP development, consumer price level, tax collection ratio, personal ability, and regional development level, and discusses the impact of various cross-cutting factors on income on this basis. Combined with China's development status and China's social situation, this paper gives a reasonable explanation of the influence of variables on personal income in the process of mathematical model exploration.

2. Methods

2.1. Data Source

This data is quoted from Kaggle website, which is collected by DR. ALOK YADAV AT YBI FOUNDATION. And these data were updated 7 months ago. This dataset contains 32562 groups of data and this paper will use these data to show the factors that affect bank-balance.

2.2. Data Selection

In the present study, the author descends the hierarchical levels and make us use a dataset, which for the analysis incorporates multiple socio-economic variables, which include individual's financial and other personal characteristics [8, 9]. The dataset contains the following variables:

IsAdult (X1): This binary variable gives an indication of whether the individual is an adult, followed by 0 for non-adults. Adults tend to have a higher employment rate and are therefore responsible for their incomes, which the author assumes will improve their bank balances.

marital Status (X2): This categorical variable defining the individual's marital status includes categories such as Married, Unmarried, and Widowed. A person's financial situation is impacted greatly by marriage and divorce, since those events embody an individual's financial obligations and duties for life.

Gender (X3): Gender is also a binary variable here, where male is 1 and female is 0. Gender-related factors believe to be affecting the financial outcomes of individuals because of income as well as job opportunities may be more favorable to some groups compared to others, thus paving different patterns in the bank balances.

Job (X4): It takes this form of categorical variable - people's job type. The jobs include categories like government job, self-employed, and private job. Type of employment is a considerable factor when it comes to income stability and future financial help.

The model has the variable Bank-Balance (a continuous variable that indicates the amount of money in an individual's bank account) as the dependent variable. In these recent studies, the author applied the method of one-hot encoding to convert categorical variables (such as MaritalStatus, Gender, and Job) into either 1 or 0. The author can utilize this transformation to scale the impact of the various independent variables on the bank balance. The specific description of this dataset is shown in Table 1:

Table 1. List of Variables

Variable	Logogram
IsAdult	x_1
MaritalStatus	x_2
Gender	x_3
Job	x_4

2.3. Method Introduction

To analyze the correlation between a dependent variable and one or more independent variables, linear regression, a commonly used statistical method, comes into play. The major endeavor of linear regression is to formulate a linear equation, wherein the dependent variable can be estimated based on the values of the independent variables [10]. The most common version of the linear regression equation is:

$$Bank = \beta_0 + \beta_1(IsAdult)x_1 + \beta_2(MaritalStatus)x_2 + \beta_3(Gender)x_3 + \beta_4(Job)x_4 + \varepsilon \quad (1)$$

In this equation: β_0 stands for intercept, that means the expected bank balance if the values of independent variables are equal to zero. β_1 , β_2 , β_3 , and β_4 denote coefficients of independent variables, which reflect the magnitude and sign of their influence on the dependent variable. ε , or the error term, which describes variations in the bank balance arising from independent variables. Linear regression, on the other hand, is based on several main assumptions, which ensure the validity and reliability of the process, as follows: Linearity: The association between the dependent variable and each individual independent variable should be linear. Independence: The assumptions should be self-sufficient for the variable. Homoscedasticity: The variance of the errors should remain consistent over the range of the independent variable. Normality: The errors of the model (residuals) should be normally distributed. Failing to meet these assumptions might lead to inaccurate or unsuitable values, which will decrease the ability of the model to predict.

3. Results and Discussion

3.1. Scattered Point Diagram

Firstly, this paper uses several scattered point diagrams to show the relationship of dependent variable and four independent variables. Figure 1 shows the scattered point of bank-balance and Marital Status, figure 2 shows the scattered point of bank-balance and Job.

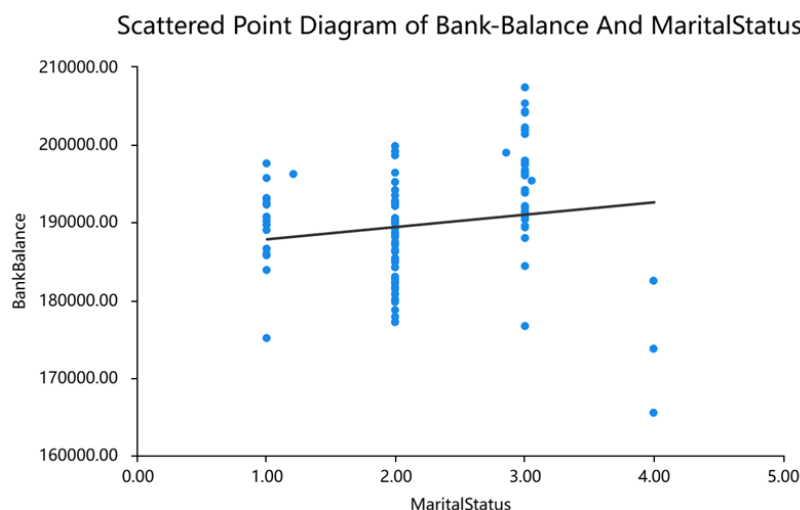


Figure 1. Scattered point diagram of bank-balance and Marital Status

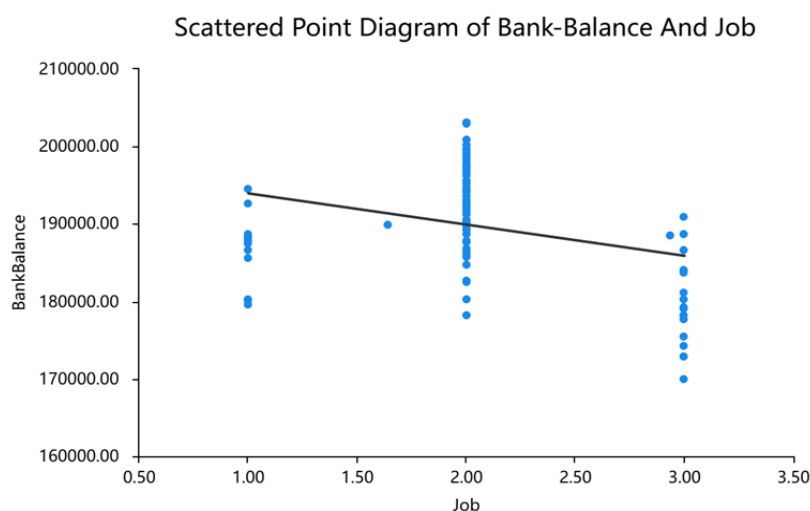


Figure 2. Scattered point diagram of bank-balance and Job

3.2. Correlation Analysis

By observing the table 2, this paper uses Pearson correlation coefficient to research the relationship between BankBalance and IsAdult, Job, MaritalStatus, Gender, use Pearson correlation coefficient to represent the strength of correlation.

Table 2. Pearson Correlation

	BankBalance
Job	-0.019
MaritalStatus	-0.064
IsAdult	-0.225*
Gender	-0.114

The value of Pearson correlation coefficient between BankBalance and IsAdult is -0.225, and the value greater than 0.05. So BankBalance and IsAdult have significance negative correlation. The value of Pearson correlation coefficient between BankBalance and Job is -0.019, which is close to 0, and the value of $p:0.848 > 0.05$, So there is no correlation between BankBalance and Job. The value of Pearson correlation coefficient between BankBalance and marital Status is -0.064, which is close to 0, and the value of $p:0.527 > 0.05$. So, there is no correlation between BankBalance and marital Status. The value of Pearson correlation coefficient between BankBalance and Gender is -0.114, which is close to 0, and the value of $p:0.258 > 0.05$. So, there is no correlation between BankBalance and Gender.

3.3. Model Results

The result in this paper shows that there are many factors influencing Bank-Balance. As table 3 shows:

Table 3. Result of linear regression

	β	Std.Error	t	p	VIF	F
Constant	409551.6663	88049.817	4.651	0.000**	-	$F(4,95)=1.756, p=0.144$
x_1 (IsAdult)	-6531.539	18552.726	-0.35	0.726	1.014	$F(4,95)=1.756, p=0.144$
x_2 (MaritalStatus)	-149174.558	63405.940	-2.35	0.021*	1.150	$F(4,95)=1.756, p=0.144$
x_3 (Gender)	-8753.508	24992.673	-0.35	0.727	1.182	$F(4,95)=1.756, p=0.144$
x_4 (Job)	-19002.525	17452.724	-2.08	0.279	1.247	$F(4,95)=1.756, p=0.144$

First this paper will decode the value of VIF. VIF value is used for colinear judgement. Colinear means independent variables are related to each other during the process linear regression. The reason of colinear could be multiple variables already have strong dependent, or the number of samples collected is not enough, or setting wrong virtual variable during linear regression. As the table 3 shows, there are no colinear. Secondly this paper will decode the value of D-W. The immediately interfere items are independent with each other, is one of the hypotheses of multiple linear regression. Random interference items are errors caused by the uncertainty of the data itself. If the random interference terms of the model violate the basic assumption that they are independent of each other, it is called existence self-correlation. Self-correlation can be analysed by D-W test. From the table 3, the value of D-W is 2.139, which is close to 2. So, model doesn't have correlation, structure is good.

By bringing in the coefficient to the fomula, this paper can get the relationship of bank-balance and IsAdult, MaritalStatus, Gender, Job, which is:

$$Bank - balance = 409551.6 + (-6531.539)(IsAdult)x_1 + \dots + (-19002.525)(Job)x_4 \quad (2)$$

By looking at the value of coefficient, these values are negative. So that dependent variables have negative correlation with these four independent variables. Among them, MaritalStatus has largest absolute value and IsAdult has smallest absolute value. So, the ascending order of impact is IsAdult, Gender, Job and MaritalStatus.

4. Conclusion

This paper considers China's GDP development, consumer price level, tax collection ratio, personal ability, and regional development level, and discusses the impact of various cross-cutting factors on income on this basis. Combined with China's development status and China's social situation, this paper gives a reasonable explanation of the influence of variables on personal income in the process of mathematical model exploration. The result in this paper shows that there are many factors influencing Bank-Balance. The variables have significant impact is IsAdult, Gender, Job and Marital Status.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

- [1] Marvin G. The impact of monetary policy on bank balance sheets: A comment. Carnegie Rochester Conference, 1995.
- [2] Vasco Cúrdia, et al. The central-bank balance sheet as an instrument of monetary policy. Journal of Monetary Economics, 2011.
- [3] Chen Q, Filardo A, He D, et al. international spillovers of central bank balance sheet policies. Social Science Electronic Publishing, 2012, 66: 220 - 264.
- [4] Whelan K. TARGET2 and central bank balance sheets. Social Science Electronic Publishing, 2014, 29 (77): 79 - 137.
- [5] Kumar T G, Leonard P A. Bank Balance-Sheet Management: An Alternative Multi-Objective Model. J Oper Res Soc, 1988.
- [6] Berriel T C, Bhattarai S. Monetary Policy and Central Bank Balance Sheet Concerns. Journal of Macroeconomics, 2009, 9 (1): 1770 - 1770.
- [7] D'Erasmo P, Durdu B, Boz E. Sovereign Risk and Bank Balance Sheets: The Role of Macroprudential Policies. Society for Economic Dynamics. Society for Economic Dynamics, 2014.
- [8] Rule G. Understanding the central bank balance sheet. Garreth Rule, 2015.
- [9] Zhang L, Bezemer D J. How the credit cycle affects growth: the role of bank balance sheets. Research Report, 2014, 39 (39): 33 - 63.
- [10] Cúrdia Vasco, Woodford M. The Central-Bank Balance Sheet as an Instrument of Monetary Policy. Journal of Monetary Economics, 2010, 58.