Research on the Factors Influencing the Global Unemployment Rate: Taking China as an Example

Yue Liu *

Hangzhou NO.2 High School, Hangzhou, 310051, China

* Corresponding Author Email: nings@usf.edu

Abstract. This article is aiming at finding out the independent variables that can affect the unemployment rate. Twelve independent variables and 572 observations from 1978 to 2021 are chosen. Based on an assumption, these twelve variables do correlate with the unemployment rate. To find out the more accurate relationship between the variables and the unemployment rate, the Multiple Linear Regression Model is built, and in order to avoid the collinearity problem during building the model, the stepwise regression model need to be employed. To get the results, the regression coefficients are compared, especially the VIF value. After identified by the model, only leave four variables. In the end, it shows that the impact of labor force, total imports and exports of goods, number of high schools on the unemployment rate is significant positive, additionally, the impact of the number of libraries on the unemployment rate is significant negative. Beyond that, other factors have no apparent effect on the unemployment rate.

Keywords: Unemployment Rate; Multiple Linear Regression; Stepwise Model.

1. Introduction

The problem that some people who has working ability and who want to have a job cannot find employment is called unemployment, which is a common a social phenomenon, especially during the COVID-19 period, the unemployment rate in the US once reached 14.8% (data from Kaggle) [1]. Unemployment is a quite complicated problem, which can be classified into three types, to be specific, those are Fractional Unemployment, Structural Unemployment, Cyclical Unemployment, and it is one of the biggest concerns in the world [2]. Unemployment has fatal impacts on humans' daily life and social communities, such as the economic, psychological and social well-being aspects [3]. Firstly, unemployment is a vital indicator in the economy field. It is believed that the unemployment rate is the most effective indicators that can measure the performance of the labor market [4]. Then, the unemployment also has negative effects on people's mental health, leading to some disease, such as anxiety disorder, depressive disorder and even suicide [2]. Finally, unemployment would reduce the level of social well-being, without income, many people may live a poor life and crime rate would rise. Above all, it is important to find out the factors which can influence the unemployment to help government make some regulations to reduce the unemployment rate. In this paper, the author would search for the reasons behind global unemployment.

Unemployment is a popular topic and there are many papers to find out the factors of unemployment in different areas and different types of people. A group of economists found the European youth unemployment is related to flexibility, education, migration, family legacies and European Union policy [5]. Another group of experts, including Banda et al., in Africa, started the research about the effects of the economic growth on the unemployment. They found that Gross Domestic Product is positively affected the unemployment rate in long term [6]. In Malaysia, a group of students do research about reasons behind the unemployment among recent graduates, they found that the mismatches of skills between the graduates and the company is a significant factor for unemployment [7]. Gostkowski and Rokicki mentioned that unemployment is the result of the supply-demand imbalance in the labor market [1]. Alam et al. mentioned that the conditions of social economy, demographic structure, the contribution from women are the factors which have one of the most significant impacts on the unemployment in Bangladesh, and they do the research about the influence of GDPs, inflation, population growth, FDI to unemployment in details [8]. Wan did the

research on the relationship between the micro-index, such as GDP, CPI, education appropriations and population, and the employment. However, the author only used 56 figures, it is limited [9]. Wang et al. also did the research about the same factors, using stepwise regression method to establish multiple linear regression model [10]. Another group of experts from Gansu, China, explore not only the external macro influence, but also the regional microscopic factors, such as size and number of companies, salary level and so on [11]. Ashi et al. found that the main factor of the unemployment in China is the rise of organic composition of capital [2]. Blustein et al. analyzed the influencing factors of unemployment in China's transition period, using the model of Okun's Law to explore six factors including economic restructuring, fixed asset investment, current account surplus, household consumption, inflation and minimum wage for empirical research [3].

In summary, the analysis of unemployment factors has attracted many researchers' attention. This article mainly uses the multiple linear regression model to study the impact of eight factors (GDP, CPI, labour force population, education investment, exchange rate, amount of foreign investment, tax revenue and total value of agricultural output) to unemployment, taking China as an example.

2. Methods

2.1. Data Source

The dataset used in this paper is selected from the website of China National Bureau of Statistics (Annually Data about China). There are all kinds of indicators about various fields on this dataset, and this research just listed the figures about unemployment and the eight variables from 1978 to 2021, among 44- year-period.

2.2. Variable Selection

The original dataset has a great amount of data, but there are plenty of them have nothing to do with unemployment, and some of them null values, such as the city areas, city population and so on. So finally, this research chose 572 observations, containing twelve variables (GDP, CPI, labor force, education investment, exchange rate, amount of foreign investment, tax revenue and total value of agricultural output, number of industrial enterprises, number of high schools, number of libraries, foreign exchange reserve), and one independent variable (unemployment). The specific meaning of variables is shown in Table 1.

Table 1. List of Variables

Name	Variable	Meaning
X_1	Unemployment Rate	The unemployment rate in China annually
X_2	GDP	Gross Domestic Product in China annually
X_3	CPI	Consumer Price Index in China annually (1978=100).
X_4	Labor force	The population of labor force in China annually.
X_5	Graduate student enrollment	The total number of graduate students annually in China (unit: ten thousand).
X_6	exchange rate	Exchange rate of RMB against US dollars (=100 dollars)
X_7	Total imports and exports of good	Total imports and exports of goods in China annually.
X_8	Tax revenue	The total tax revenue received by Chinese Government annually.
X_9	Total value of agricultural output	Total value of agricultural output in China annually.
X_{10}	Number of industrial enterprises	Number of industrial enterprises in China annually.
X_{11}	Number of High schools	Number of high schools in China annually.
X_{12}	Number of library	Number of libraries in China annually.
X_{13}	Foreign exchange reserve	Foreign exchange reserve in China annually in hundred million dollars.

2.3. Method Introduction

The main method used in this research is the multiple linear regression. The multiple linear regression is applied when many different vital factors have the impacts on the change of a same dependent variable, so it is a wise choice to use two or more influential factors as independent variables to explain the change of dependent variable [10]. Using this method can not only find if there have linear relationships, but also compare the magnitude of different variables. The multiple linear regression can be demonstrated as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon \quad (1)$$

Where Y represents the dependent variable. X_i represents independent variable, β_i represents regression coefficient. β_0 represents the constant term. ε representation error term, is the difference between the true value of the dependent variable and the estimated value, is a random variable.

3. Results and Discussion

3.1. Data Visilisation

Before setting the multiple regression model, it should be examined that whether there are linear regressions. The method for examination is using the scatter diagram. From figure 1, it is found that all these twelve variables have the linear regression with the unemployment rate.

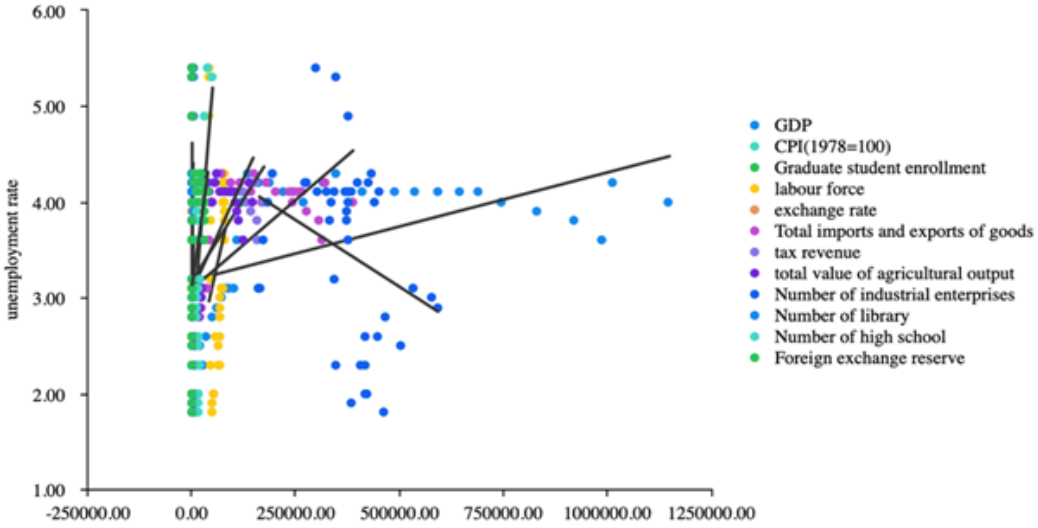

Fig. 1 The scatter diagram of variables

3.2. Multiple Linear Model Results

From table 2, it shows that the twelve independent variables have the linear regression with the unemployment rate. So, the multiple linear regression model can be built.

From Table 2, the equation of the model can be gotten:

$$X_1 = 3.729 + 0.002 * X_3 + \cdots - 0.002 * X_{12} \quad (2)$$

The regression coefficient value of X_2 is -0.000(t=-2.466, p=0.019<0.05) and that value of X_{12} is -0.002(t=-2.546, p=0.016<0.05), which means that X_2, X_{12}will have a significant negative impact on the unemployment rate.

The regression coefficient value of X_3, X_4, X_5, X_6, X_7, X_8, X_{10}, X_{13} is 0.002(t=1.015, p=0.318>0.05), 0.000(t=0.946, p=0.352>0.05), 0.022(t=1.090, p=0.284>0.05), 0.000(t=0.385, p=0.703>0.05), 0.000(t=0.145, p=0.885>0.05), 0.000(t=0.532, p=0.599>0.05), -0.000(t=-0.659, p=0.515>0.05), -0.000(t=-0.351, p=0.728>0.05) respectively, which means that X_3, X_4, X_5, X_6, X_7, X_8, X_{10}, X_{13} don't affect the unemployment rate.

The regression coefficient of X_9 is 0.000(t=2.074, p=0.046<0.05), and that value of X_{11} is 0.000(t=2.972, p=0.006<0.01). It means that the impact of X_9, X_{11} on unemployment rate is significant positive.

R^2 =0.898, means these twelve variables can explain 89.8% of the change in the unemployment rate. Besides, F=22.645, p=0.000<0.05, meaning that among these twelve variables, at least one variable would affect the unemployment rate. Moreover, it shows that VIF>10, so the collinearity problem exists, the stepwise regression should be employed to solve collinearity problem.

Table 2. Multiple Linear Regression coefficient

	B	S.E.	Beta	t	p	VIF	tolerance
Constant	3.729	2.002	-	1.862	0.072	-	-
X2	-0.000	0.000	-5.210	-2.466	0.019*	1351.360	0.001
X3	0.002	0.002	1.762	1.015	0.318	911.510	0.001
X4	0.022	0.021	0.803	1.090	0.284	164.262	0.006
X5	0.000	0.000	0.387	0.946	0.352	50.585	0.020
X6	0.000	0.001	0.113	0.385	0.703	26.148	0.038
X7	0.000	0.000	0.101	0.145	0.885	144.597	0.007
X8	0.000	0.000	1.185	0.532	0.599	1501.229	0.001
X9	0.000	0.000	2.700	2.074	0.046*	512.959	0.002
X10	-0.000	0.000	-0.046	-0.659	0.515	1.501	0.666
X11	0.000	0.000	0.554	2.972	0.006**	10.527	0.095
X12	-0.002	0.001	-1.001	-2.546	0.016*	46.819	0.021
X13	-0.000	0.000	-0.236	-0.351	0.728	137.202	0.007
R^2	0.898						
Adj R^2	0.858						
F	F(12,31)=22.645,p=0.000						
D-W	1.248						

Dependent variable= Unemployment Rate
*p<0.5,**p<0.01

3.3. Stepwise Linear Regression Model

What makes stepwise regression different from the regression analysis is that the stepwise regression model will automatically identify the significant independent variable (X), and the non-significant X will automatically move out of the model (table 3).

Table 3. Stepwise Linear Regression coefficient

	B	S.E.	Beta	t	p	VIF	tolerance
constant	1.392	1.811	-	0.769	0.447	-	-
labor force	0.000	0.000	1.244	6.833	0.000**	7.948	0.126
Total imports and exports of goods	0.000	0.000	0.670	3.930	0.000**	6.963	0.144
Number of high school	0.000	0.000	0.624	3.243	0.002**	8.863	0.113
Number of library	-0.002	0.001	-1.098	-2.906	0.006**	34.229	0.029
R 2	0.837						
Adj R 2	0.821						
F	F (4,39)=50.193,p=0.000						
D-W	1.133						

Dependent variable= Unemployment Rate
*p<0.5, **p<0.01

Automatically identified by the model, only leave X4, X7, X11, X12, these four variables. R2=0.837, means these four variables can explain 83.7% of the change in unemployment rate. Besides, F=50.193, p=0.000<0.05, meaning the model is effective. The equation of the model is:

$$X_1 = 1.392 + 0.000 * X_4 + 0.000 * X_7 + 0.000 * X_{11} - 0.002 * X_{12} \quad (3)$$

In addition, after testing the multicollinearity of the model, the result shows that in this model, the value of VIF is larger than 10. As a result, there is a collinearity problem in such model. It is suggested to check the independent variables with close correlation and re-analyze after eliminating the independent variables with close correlation. In the end, the concrete analysis found the results shown that:

The regression coefficient of labor force is 0.000(t=6.833, p=0.000<0.01), which means that labor force will positively affect the unemployment rate. The regression coefficient of total imports and exports of goods is 0.000(t=3.930, p=0.000<0.01), meaning that total imports and exports of goods have a significant positive impact on the unemployment rate. The regression coefficient of number of high schools is 0.000(t=3.243, p=0.002<0.01), meaning that the impact of the number of high schools on the unemployment rate is significant positive. The regression coefficient of number of libraries is -0.002(t=-2.906, p=0.006<0.01), giving the meaning that the number of libraries would negatively affect the unemployment rate and such impact is significant at the same time.

Although the method is correct, the results may not accurate and correct, as China is an example country in the study and different countries may have different situations. What's more, the factors selected only by the author themselves, so it may be a little subjective. Besides, the amount of data samples may be too small to find out the real regulation. To find out the real factors, the more country examples may be chosen and the data need to be more.

4. Conclusion

In this study, 12 independent variables are chosen with the 572 samples of the data from the large, official database. The method applied in the research is effective and doable, it is setting the multiple linear regression model initially and using the stepwise regression to deal with the collinearity problem.

From the working process, it shows that labor force, total imports and exports of goods, the impact of the number of high schools on the unemployment rate is significant and positive. Besides, the number of libraries have a significant negative impact on unemployment rate. However, others do not affect the unemployment rate. From the results, it really inspired the government about the measures which can be taken to reduce the unemployment rate nowadays. Nowadays, it is really important for government to take measures to reduce the unemployment rate, so this study may give some guidance and has some importance.

References

[1] Gostkowski M, Rokicki T. Forecasting the unemployment rate: application of selected prediction methods. European Research Studies, 2021, 24(3): 985-1000.

[2] Ashi H A, Al-Maalwi R M, Al-Sheikh S. Study of the unemployment problem by mathematical modeling: Predictions and controls. The Journal of Mathematical Sociology, 2022, 46(3): 301-313.

[3] Blustein D L, et al. Unemployment in the time of COVID-19: A research agenda. Journal of vocational behavior, 2020, 119: 103436.

[4] Feng Y, Lagakos D, Rauch J E. Unemployment and development. The Economic Journal, 2024, 134(658): 614-647.

[5] O'Reilly J, et al. Five characteristics of youth unemployment in Europe: Flexibility, education, migration, family legacies, and EU policy. Sage Open, 2015, 5(1).

[6] Banda H, Ngirande H, Hogwe F. The impact of economic growth on unemployment in South Africa: 1994-2012. Investment Management and Financial Innovations, 2016, 246-255.

[7] Hossain M I, et al. Factors influencing unemployment among fresh graduates: A case study in Klang Valley, Malaysia. International Journal of Academic Research in Business and Social Sciences, 2018, 8(9): 1494-1507.

[8] Alam J, Nur Alam Q, Hoque M T. Impact of GDP, inflation, population growth and FDI on unemployment: A study on Bangladesh economy. African Journal of Economics and Sustainable Development, 2020, 3(3): 67-79.

[9] Tong Ruo Wan. Empirical study on influencing factors of unemployment rate in China. Modern Economic Information, 2015, 15: 3.

[10] Chen J. Multivariate linear analysis of factors affecting unemployment rate. Financial News, 2018, 19: 2.

[11] Wang Liping, Chen Zhen, Wang Xiaowen, Bai Lifei. Empirical analysis of influencing factors of unemployment rate: A case study of Baiyin City, Gansu Province. Gansu Science and Technology, 2015, 44(12): 4.