

Research on the correlation between some factors and GDP in different countries

Chang Liu *

Nanjing University of Posts and Telecommunications

* Corresponding Author Email: Lucy.chang1@outlook.com

Abstract. The global economic situation is extremely not optimistic after the impact of the epidemic. In order to promote GDP development, countries urgently need to promulgate some policies related to GDP development. This paper is devoted to exploring the correlation between population growth rate, unemployment rate and real interest rate and GDP development, and summarizes the relationship between positive and negative correlation of each factor and national conditions, so as to provide references for various countries to choose appropriate policies in these three aspects. As for the selection of data, this paper selects 4 sets of relevant data of 5 developed countries and 5 developing countries for 13 consecutive years for calculation. At the same time, Shapiro-Wilk test was used to judge whether the data fit the normal distribution, but most of the results did not, so Spearman correlation coefficient with loose premise was chosen. Finally, based on the data of significant correlation, we reached the following conclusions: Generally, for developed countries or populous countries, the correlation between population and GDP is generally negative. For developed European countries or countries with large geographical areas, the correlation between unemployment and GDP is generally negative. For countries with extremely high exchange rates, such as the US or the UK, real interest rates tend to have a positive correlation with GDP and a negative one.

Keywords: GDP, correlation analysis, population growth, unemployment, interest rate.

1. Introduction

As we all know, GDP is the standard to measure a country's economic development and has irreplaceable importance. GDP can be said to be an important reference to predict the future economic development trend of a country. The importance of GDP development has become more obvious in the era of the epidemic since then, in order to ensure the future development of countries, countries will take measures to increase GDP as much as possible. According to the data of the World Bank, in the outbreak of COVID-19 in 2020, the GDP growth rate of countries around the world dropped to the lowest in decades -3.1%, and each country urgently needs to implement new policies to recover the huge property lost during the epidemic period.

As for the factors affecting GDP, many scholars have put forward their own views. Zhang Jin et al. [1] pointed out the importance of education and human capital stock to economic growth, and the accumulation of human capital brought about by educational input and the stock of human capital mainly represented by the level of education had a promoting effect on economic growth. It is no surprise that education drives the economy, but how the rate of population growth affects GDP is still a topic of interest.

In addition, we all know that when the economy is depressed, the unemployment rate will also rise. In economics, there is a Phillips curve between the unemployment rate and the inflation rate. Li Xiaowen [2] also used real data to further prove the correctness of Phillips, that is, the unemployment rate rises and the inflation rate falls. The unemployment rate is not the only thing related to the inflation rate. Lin Chong [3] clearly proposed that monetary policy operations should be used to establish flexible inflation targets, reflecting the concept of monetary policy neutrality, and indicating that monetary policies, such as the real interest rate, also have a great impact on inflation. However, the result of inflation is not to blindly increase GDP. Chileshe A [4] proposed in his study on the relationship between economic growth and inflation in Zambia that if inflation exceeds a certain threshold, it will have a negative impact on GDP growth. So, for the relationship between the two, it

still contains a lot of places worth studying. So, the unemployment rate and the real interest rate, which are related to the inflation rate, are also worth exploring directly in relation to GDP. In a word, this paper discusses the relationship between population size, unemployment rate, real interest rate and GDP.

In addition, we all know that when the economy is depressed, the unemployment rate will also rise. In economics, there is a Phillips curve between the unemployment rate and the inflation rate. Li Xiaowen [2] also used real data to further prove the correctness of Phillips, that is, the unemployment rate rises and the inflation rate falls. The unemployment rate is not the only thing related to the inflation rate. Lin Chong [3] clearly proposed that monetary policy operations should be used to establish flexible inflation targets, reflecting the concept of monetary policy neutrality, and indicating that monetary policies, such as the real interest rate, also have a great impact on inflation. However, the result of inflation is not to blindly increase GDP. Chileshe A [4] proposed in his study on the relationship between economic growth and inflation in Zambia that if inflation exceeds a certain threshold, it will have a negative impact on GDP growth. So, for the relationship between the two, it still contains a lot of places worth studying. So, the unemployment rate and the real interest rate, which are related to the inflation rate, are also worth exploring directly in relation to GDP. In a word, this paper discusses the relationship between population size, unemployment rate, real interest rate and GDP.

There is also an important variable, which is that the increase or decrease of different factors has different effects on development in different types of countries. For example, Menendez Gonzalez [5] mentioned in his paper that in developing countries, low-technology products are often manufactured by high-technology manpower, and such inequality is common in developing countries. Therefore, due to the general environment, the same factors do not play the same role in different countries.

Therefore, in the end, this paper selects five countries in each of the developing and developed countries to analyze the relationship between the factors mentioned above and GDP, so as to draw a more credible conclusion, and provide ideas on population growth rate, unemployment rate and real interest rate for the developed and developing countries trapped in the low GDP development in the post-epidemic era.

2. Related works

2.1. Data Preparation

The first is data preparation. In order to ensure that the data is accurate and diverse, this paper collects all relevant data from ten countries on the World Bank website, five of which are developed countries, including the United States, the United Kingdom, Australia, Japan, and the Republic of Korea. The other five are developing countries: South Africa, the Republic of Congo, Pakistan, Brazil and China. If the website does not have the latest partial year data, it will automatically push forward some years to ensure that the final four data sets of GDPS, population growth rate, unemployment rate and real interest rate have the latest data for each country for 13 years.

2.2. Prerequisites for the use of different correlation coefficient

In addition, for the choice of correlation coefficient, the popular choice is Pearson correlation coefficient, Spearman correlation coefficient or Kendall correlation coefficient, but each coefficient is applicable to different environments. Cheng Juanjuan [6] explained the use of three coefficients in her paper: The three correlation coefficients are suitable for studying data sets with different properties, among which Pearson correlation coefficient (product difference correlation coefficient) is suitable for correlation analysis of continuous variables, and Spearman correlation coefficient is suitable for correlation analysis of categorical variables or variables with obvious abnormal distribution or unknown distribution. This requires us to judge the normal distribution of the data first and then choose which correlation coefficient to use.

De R R S et al. [7] suggest that the graphical method for judging normal distributions, while very useful, may not be ideal for drawing conclusions about violations of normality. Therefore, numerical methods, mainly normality tests, are more accurate in this respect, and the Shapiro-Wilk test using small samples is widely recommended. Therefore, Shapiro-Wilk test is used to determine whether the distribution is normal.

Although Pearson correlation coefficient is the most popular one, but the premise of use must be linear relationship and meet the normal distribution. It should be noted that for most randomly selected data, it is difficult for them to meet the strict condition of normal distribution. Moreover, compared with Pearson correlation coefficient, Spearman correlation coefficient is not affected by data dimension and is not sensitive to abnormally large numbers [8]. Therefore, for the small sample in this paper, it is more likely to choose the Spearman correlation coefficient with less stringent requirements in the end.

Of course, we need to be strict with the data, so we must first determine whether the obtained data meets the normal distribution, and then choose the coefficient, and then make the correlation analysis.

3. Models and methods

3.1. Shapiro-Wilk test

Estrada G et al. [9] proposed the calculation method of Shapiro-Wilk test for judging the normal distribution in their study:

$$W_X = \frac{(\sum_{i=1}^n a_i X_{(i)})^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

Where $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ and the random variables $X_{(1)} < \dots < X_{(n)}$ denote the order statistics corresponding to $X_1, \dots, X_n$. The vector $(a_1, \dots, a_n)'$ is a function of the expectation and covariance matrix of the order statistics vector for a standard normal random sample of size n .

Using the Shapiro-Wilk test, we can decide whether to use the Pearson correlation coefficient or Spearman correlation coefficient.

3.2. Pearson correlation coefficient

If most of the data do not meet the normal distribution, Spearman correlation coefficient formula can be used to calculate the correlation coefficient between GDP and the other three factors; otherwise, Pearson's correlation coefficient is used. Pearson correlation coefficient is actually easy to calculate. When Zhang Xiaobin et al. [10] calculated the correlation factors affecting smog in Beijing, the calculation formula of Pearson coefficient was as follows:

$$r = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{s_X} \right) \left(\frac{Y_i - \bar{Y}}{s_Y} \right)$$

Where $\bar{x}, \bar{y}$ are the mean of the two variables X and Y , which is called expectation; s_X, s_Y are the standard variances of X and Y .

The reason why the variance s in the formula is not the more common variance σ , is because s is actually the sample variance, while σ is the population variance. Zhao Jian et al. [11] pointed out that when using sample variance for calculation, what should be divided is not n but $n-1$, because if only the degree of dispersion between sample data is described, the divisor in the sample variance calculation formula should be n . If the variance of the population is estimated, the sample variance formula with the divisor $n-1$ needs to be selected. Obviously, detailed values are needed in the calculation of Pearson correlation coefficient, and part of the data we selected must belong to samples rather than huge whole data, so the result of this formula is accurate, and this method of calculating correlation coefficient is adopted in this paper.

3.3. Spearman correlation coefficient

Spearman correlation coefficient is a correlation coefficient used in the case of non-normal distribution. Ji Peng [12] used the formula of Spearman correlation coefficient in his study:

$$r = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}}$$

Where N is the total number, and x and y are the data for both variables.

3.4. Hypothesis Testing

After the correlation coefficient is obtained, these correlation conclusions need to be tested, and this step often uses hypothesis testing. Ma Fengming et al. [13] explained the steps of hypothesis testing: the null hypothesis and alternative hypothesis are proposed, and the confidence interval is 95% by default. The observed value of the test statistic is calculated according to the sample observations, and the null hypothesis is accepted or rejected according to whether it is greater than or less than the boundary value. The final judgment will be easier to see using P-value, so we further refine on this basis and use the size relationship between P-value and 0.05 to reach a conclusion.

In the case of judging whether it is normal distribution, H0 is subject to normal distribution, and the P-value is less than 0.05, even if it is in line with normal distribution. In the case of correlation calculation, H0 is the correlation coefficient r is equal to 0, and the final calculated P-value must be less than 0.05 to be considered as correlation.

To sum up, the overall model should first use Shapiro-Wilk test to judge whether the normal distribution is satisfied. If it is satisfied, Pearson correlation coefficient is used to calculate the result; if it is not satisfied, Spearman correlation coefficient is used. Finally, the P-value is calculated by hypothesis test to determine whether the correlation is valid.

4. Correlation coefficient and significance of different countries

Table 1 shows the correlation coefficients for 10 countries, and then the Shapiro Wilk test is used to determine whether GDP and population growth rate, unemployment rate, and real interest rate follow a normal distribution. From Table 2, it can be seen that many data points have P-values greater than 0.05, including key GDP, which is not a normal distribution. We can only use the Spearman correlation coefficient for subsequent calculations.

Table 1. ten countries' correlation coefficients

r	America	Britian	Australia	Japan	Korea	South Africa	Cong, Dem Rep	Pakistan	Brazil	China
population	-0.68	-0.644	-0.193	0.063	-0.828	-0.019	-0.548	-0.399	0.836	0.713
unemployment	-0.78	-0.474	-0.374	0.481	0.556	-0.208	0.771	0.852	-0.789	0.284
interest rate	0.646	0.061	-0.168	-0.08	-0.726	-0.381	-0.404	-0.89	-0.008	-0.799

Table 2. P-value of ten countries' correlation coefficients

P-value	America	Britian	Australia	Japan	Korea	South Africa	Cong, Dem Rep	Pakistan	Brazil	China
population	0.011	0.036	0.528	0.838	0	0.95	0.063	0.372	0	0.006
unemployment	0.002	0.181	0.209	0.096	0.048	0.495	0.03	0	0.001	0.347
interest rate	0.017	0.843	0.584	0.794	0.005	0.199	0.171	0	0.979	0.001

In Table 3, correlation coefficients with practical significance can be selected based on P-values.

Table 3. effective correlation coefficients

r	America	Britian	Korea	Cong, Dem Rep	Pakistan	Brazil	China
population	-0.68	-0.644	-0.828			0.836	-0.713
unemployment	-0.78		0.556	0.771	0.852	-0.789	
interest rate	0.646		-0.726		-0.89		-0.799

5. Analysis and conclusion

From the table 1, it can be seen that the three factors in developed countries and developing countries have different impacts on GDP.

For developed countries, three fifths of the population growth rate is related to GDP, and the correlation is negative. Unemployment is related to GDP in two out of five countries, but the correlation is positive and negative; Real interest rates are related to GDP in two fifths of countries, and the correlation is sometimes negative.

And in developed countries, the polarization of correlation is more obvious, either the three factors are basically not related to GDP, or all three factors are related to GDP. In most developing countries, only one or two factors correlate with GDP.

Taken together, we can roughly infer some reasons for the different correlations between GDP and these three factors. It can be seen from the results that the following conclusion is not applicable to every country, but it is possible for countries that meet the relevant characteristics.

For developed countries and populous countries (China), population growth rate often has a negative effect on GDP, but for some developing countries, population growth may promote GDP development. In terms of unemployment rate, the unemployment rate of some countries in the two types of countries has positive and negative correlation with GDP. However, a careful study of the countries with negative correlation will find that Brazil and the United States are both countries with large land area, and naturally there will be more jobs, so the decline of unemployment rate will promote GDP growth. Even in advanced countries (Korea), it is difficult to guarantee that a small number of unemployed people will boost GDP growth. In terms of the real interest rate, we all know that the real interest rate refers to the real interest rate that savers or investors can receive after deducting the inflation rate, which is related to monetary policy. In fact, it is related to the value of currencies. In recent years, the value of the US dollar has only increased, while the value of other currencies has more or less declined. As a result, a rise in interest rates almost exclusively increases GDP in the United States, while a rise in interest rates in other countries increases the amount of money stored, but does not boost GDP.

With these inferred conclusions, we can try to deduce the correlation between three factors and GDP in other countries. For example, in Russia, after calculation, the correlation between interest rate and GDP is significant, with a P-value of 0.039 and a correlation coefficient of -0.577, which is in line with our inference. For example, in Germany, there is a significant difference between unemployment rate and GDP, but it is -0.657, which is different from our previous conclusion, indicating that there is still some deficiency in our previous thinking, which ignores the differences in social models between different regions. Because European countries and Asian countries have different working thinking and environment, after improvement, our conclusion about the unemployment rate can be changed to: for the developed countries in Europe and the larger countries in other regions, the unemployment rate and GDP have a negative correlation.

However, there are always some exceptions, such as Serbia, where the population growth rate has a significant correlation with GDP, with a P-value of 0.002 and a correlation coefficient of -0.762. Since it is not a developed country at present, it is contrary to our conclusion, but Serbia was a country with good economic development before this. This shows that we cannot accurately judge this kind of relationship with complex causes only by some simple definitions, so finding a deep relationship becomes an important task in the future

6. Conclusion

Faced with the challenges of the post-pandemic era, every country wants to find factors related to GDP and make good use of them to promote national development. According to the research of some scholars, we selected three factors that have a great relationship with GDP: population growth rate, unemployment rate and real interest rate.

We first obtained 13 years of relevant data for 10 different countries, and then calculated and tested whether the data conformed to the normal distribution by using Shapiro-Wilk test and hypothesis test. Because it does not match, we choose Spearman correlation coefficient as the coefficient of correlation calculation, and finally calculate the final result and significance. Except for some countries with special national conditions, our conclusions are relatively reliable. For developed countries or populous countries, the correlation between population and GDP is generally negative, and vice versa. For developed European countries or countries with large geographical areas, the correlation between unemployment and GDP is generally negative, and vice versa. For countries with extremely high exchange rates, such as the US or the UK, real interest rates tend to have a positive correlation with GDP and a negative one.

7. Future works

Finding the correlation between population growth rate, unemployment rate and real interest rate of different types of countries and GDP can help different countries find a way out in the post-epidemic era, make adjustment policies in these three aspects according to their national conditions, and improve the development of GDP. However, it is obvious that the basis for our judgment is too simple, which makes some boundaries blurred. For example, every developed country has a different system, so it is obvious that relying on “developed country” to judge will be flawed; Another example is the number of countries with large land area. Therefore, more in-depth research is needed to find the exact limit and interpretation basis.

However, in addition to hoping to find factors related to GDP, as an important indicator to measure a country's development, people often hope to find out factors related to GDP growth. Therefore, we also try to use the model method in this paper to explore whether the population growth rate, unemployment rate and real interest rate are correlated with GDP. But even in countries where the correlation should be strongest, the correlation between these three factors and GDP growth is far from clear. Therefore, although there is a certain relationship between GDP and GDP growth, this model cannot be applied to GDP growth.

Therefore, if you want to further study the GDP growth rate, you may have to select another indicator, or perform a function transformation of the obtained indicator to find the correlation.

References

- [1] Zhang Jin, Fan Hongwei. Study on the effects of education input and human capital stock on economic growth [J]. China Price, 2023 (07): 35 - 38.
- [2] Li Xiaowen. Research on the relationship between unemployment rate and Inflation Rate in China based on Phillips Curve [J]. Business News, 2023 (12): 135 - 138.
- [3] Lin Song. External shocks, monetary policy dilemmas and inflation targeting [J]. Contemporary Economics, 2019, 40 (02): 95 - 100.
- [4] Chileshe A. The relationship between inflation and economic growth in Zambia [D]. Zhejiang industry and commerce learn, 2023. DOI: 10.27462 / , dc nki. GHZHC. 2023.000480.
- [5] Menendez Gonzalez, I., Owen, E., & Walter, S. (2023). Low-Skill Products by High-Skill Workers: The Distributive Effects of Trade in Emerging and Developing Countries. Comparative Political Studies, 56 (11), 1724 - 1759.
- [6] Cheng Juanjuan. The relationship between research and teaching in universities: An analysis based on Pearson correlation Coefficient [J]. China university of science and technology, 2022 (10): 46 to 52. DOI: 10.16209 / j.carol carroll nki cust. 2022.10.016.
- [7] De R R S, Marcos T, Chuquel A M, et al Sample size and Shapiro-Wilk test: An analysis for soybean grain yield [J]. European Journal of Agronomy, 2023, 142.
- [8] Yang Ye, Pan Chen. Research on on-line monitoring Technology of transmission line tower grounding state [J]. Electronic Paradise, 2019 (10): 263 [9] Estrada G, Elizabeth, Villasenor, et al Shapiro-Wilk test for multivariate skew-normality [J]. Computational Statistics, 2022 (republish).

- [9] Estrada G, Elizabeth, Villasenor, et al Shapiro-Wilk test for multivariate skew-normality[J]. Computational Statistics, 2022 (prepublish).
- [10] Zhang Xiaobin, Yu Bo. Urban smog control strategy in Beijing based on Pearson correlation analysis and BP neural network [J]. Systems Engineering, 2019, 41 (02): 26 - 34.
- [11] Zhao Jian, Guo Liangdong. Sample variance definition analysis [J]. Journal of Normal Science, 2016, 36 (07): 61 - 62+69.
- [12] JI Peng, Chen Fangfang, Xu Tianqi, et al. Research on High Resistance Grounding Fault Line Selection Method based on Spearman Correlation Coefficient Method and Active Component Method [J]. Shandong Electric Power Technology, 2022, 49 (12): 8 - 13+31.
- [13] Ma Fengming, Wang Zhongli. Analysis and Application of hypothesis Testing Methods [J]. Journal of Changchun University, 2012, 22 (02): 188 - 192+196.