Scientific and Social Research

Research Article

RESEARCH AND PRESS INNOSCIENCE

Multiple Linear Regression Analysis of the Relationship between the Three Industries and GDP Growth

Chenxi Ni

Soochow University, Suzhou 215000, Jiangsu Province, China

Abstract: This essay employs the econometric method of multiple linear regression model based on the quarterly data from 2000 to 2019 in China, to make an analysis of the relationship between the growth rate of three industries and GDP and the effects of adjustments of the industrial structure on economic development. The result shows that there is a statistically significant correlation between the growth rate of three industries and GDP. Specifically, the secondary industry has the most significant influence on China's economy for this period, the primary industry has the least, and the service industry's impact is somewhere in between which has become increasingly crucial.

Keywords: Three industries; GDP growth; Linear regression model

Publication date: December, 2020
Publication online: 31 December, 2020
[*]***Corresponding author:*** Chenxi Ni, Pluto1001xixi@163.com

1 Introduction

Over the past twenty years, China's economy has achieved swift advancement. Each of three industries and the adjustment of industrial structure have a profound impact on GDP growth (Zhao,2018). The first industry stands for agriculture, forestry, animal husbandry and fishery. The second refers to mining industry, manufacturing industry, production and supply of power, heat, gas and water industries and construction industry. The third, namely service industry, presents other industries expect for the primary and the secondary industry (Regulations on Clarification of Three Industries).

Currently, China is in the critical transition stage where the economic expansion has moderated to a "new normal" pace and its goal is to achieve the balanced economic development of high speed and good quality (Xiao,2018). This essay's objective is to analyze the relationships between the growth rate of three industries and GDP which will better adapt to the requirements of economic advancement. Section 2 gives some data and makes a simple descriptive explanation. Section 3 presents relevant theory of industrial structure. Section 4 applies the multiple linear regression model to make a detailed analysis. Section 5 draws the conclusions and points out some deficiencies of this model.

2 Data Collection and Descriptive Statistical Analysis

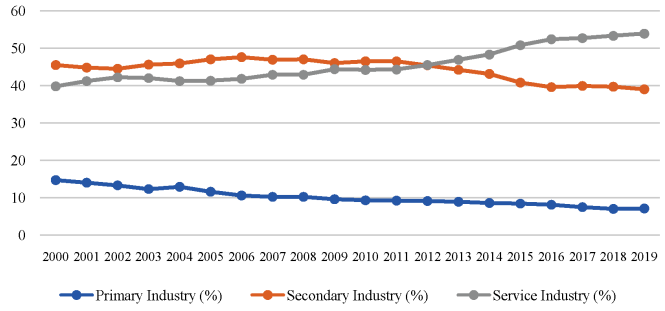

Figure 1. Proportion of the Added Value of Three Industries in GDP from Year 2000 to 2019

Table 1. Descriptive Statistics of the Proportion of Three Industries in GDP

Variable	Time Period (Sample size)	Mean	Standard Deviation	Maximum (Year)	Minimum (Year)
Primary Industry (%)	2000-2019 (20)	10.13	2.30	14.7(2000)	7(2019)
Secondary Industry (%)		44.28	2.87	47.6(2006)	39(2019)
Service Industry (%)		45.60	4.65	53.9(2019)	39.8(2000)

Data source: National Bureau of Statistics of China

Since the turn of 21st century, the proportion of the primary and the secondary industry in GDP is gradually decreasing, while the contribution of the service industry to the whole economy is expanding year by year.

According to the report of the 18th National Congress of the Communist Party of China (2012), promoting the strategic adjustment of economic structure is the main direction of transforming the mode of economic development. The service industry gradually occupies the leading position and becomes the main driving force for economy. It has been an important indicator of a country's comprehensive modernization and economic strength. Since 2012, the proportion of the tertiary industry in GDP (45.5%) has begun to exceed that of the secondary industry (45.4%), and the gap has been continuously widened. Within the 20 years, the mean value of the tertiary industry has even outnumbered that of the secondary industry by 1.32%. Until 2019, the proportion of the first and second industries has reached a minimum level among these 20 samples, while the third industry accounts for as high as 53.9%. The standard deviation of proportion of the primary and secondary industry in GDP are 2.3 and 2.87 respectively, which reflected that they had the relatively stable fluctuations compared with the service industry.

3 Relevant Theories and Researches

According to Petty-Clerk Theorem, with the improvement of economy and national income, the labor force would first move from the primary industry to the secondary industry, and then moves to the third industry when economy further develops. Later, the adjusted industrial structure would further promote economy to grow (Clark, 1940). Baumol (1967) considered the distribution of labor force in industries, deeply discussed the evolution process of industrial structure as the economy grows and concluded that resources and labor force gradually tend to enter into the secondary and service industries. ROM (2000) and Hartwig (2012) pointed out that different industrial structures derive different production functions which ultimately determine the speed and stability of long-term economic growth.

4 Linear Regression Analysis and Hypothesis Testing

Now, we can infer that the structure of three industries has a deep relation with GDP growth and each of three industries has a different and important influence on economy. The problems are to consider whether there exists a significant linear relationship between the growth rate of three industries and GDP and which one of the three industries has the most significant effect. Therefore, relevant data should be processed by establishing a linear regression model.

Use Y to represent the quarter-on-quarter growth rate of GDP. Use X_1, X_2 and X_3 to represent the quarter-on-quarter growth rate of the primary, secondary and service industry, respectively.

4.1 Collect Data

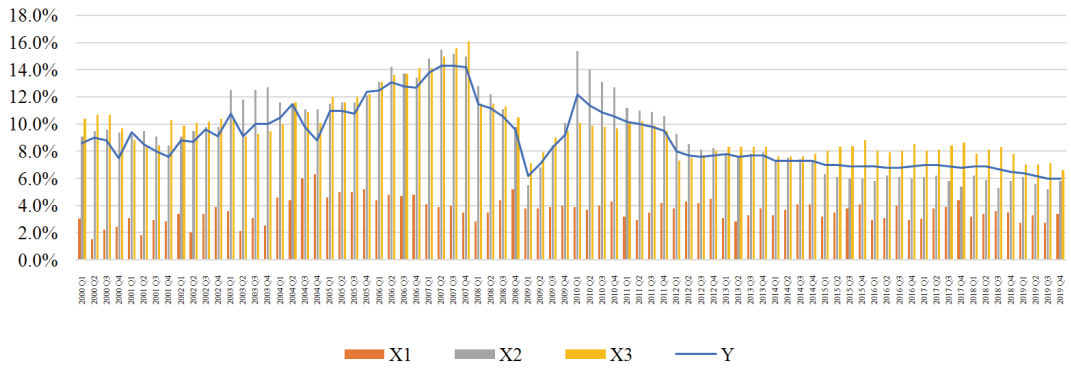

Figure 2. Quarterly Growth Rate of GDP and Three Industries

4.2 Set the Multiple Regression Model

A multiple linear regression model can be employed to analyze the effect of the growth rate of three industries on GDP development. The basic model can be established and expressed as the following term:

$Y_t=\alpha+\beta_1 X_1+\beta_2 X_2+\beta_3 X_3+\mu_t$, (Equation 1)

where Y_t, the regressand, represents the quarterly growth rate of GDP compared with the corresponding quarter of last year; X_1, X_2 and X_3, serving as regressors, represent the quarterly growth rate of the first, second and third industries respectively. The intercept, α, represents the inherent economic growth rate when other factors remain unchanged. β_1, β_2, β_3 show how the quarterly growth rate of GDP increases, on average, as that of the primary, secondary and service industry increases respectively, other things being equal. μ_t, the random disturbance term, represents other factors affecting Y_t that are omitted from the model. The associated sample regression function would be:

$Y_t=a+b_1 X_1+b_2 X_2+b_3 X_3+e_t$, (Equation 2)

where a, b_1, b_2 and b_3 are sample estimates for α, β_1, β_2 and β_3 respectively. The goal is to find values of a, b_1, b_2 and b_3 that would minimize the residual sum of squares ($\sum_t^n=1e_t^2$) and give the best fitting equation for the relevant data.

Table2: Summary Output

Table 2.1: **Regression Statistics**	
Multiple R	0.985868
R Square	0.971935
Adjusted R Square	0.970827
Standard Error	0.003788
Observations	80

Table 2.2: **Analysis of Variance (ANOVA)**				
	df	SS	F	Significance F
Regression	3	0.037766	877.3367	7.43E-59
Residual	76	0.001091		
Total	79	0.038857		

Table 2.3: **Regression Parameters**						
	Coefficients	Std. Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	0.002099	0.002404	0.873198	0.385306	-0.00269	0.006888
X_1	0.110876	0.051219	2.164725	0.033548	0.008864	0.212888
X_2	0.461512	0.026075	17.6991	1.12E-28	0.409578	0.513446
X_3	0.417298	0.037074	11.25579	7.41E-18	0.343458	0.491137

From Table 2.3, we can see the intercept a is approximately 0.002 (rounded to three decimal places). 0.111,0.462 and 0.417 are the OLS estimates for b_1, b_2 and b_3 respectively. Standard errors, $SE(b_1)$, $SE(b_2)$ and $SE(b_3)$, are 0.051, 0.026 and 0.037 respectively. Therefore, the estimated sample regression function is

$(\hat{Y}t)=0.002+0.111X_1+0.462X_2+0.417X_3$ (Equation 3)

(0.002) (0.051) (0.026) (0.037)

According to Equation 3, the signs of b_1, b_2 and b_3 are all positive, i.e., the growth rates of three industries are positively proportional to GDP. Besides, these values are between [0,1], which conforms to the law of economic development. The coefficient for the quarterly growth rate of the first industry, b_1, equals 0.111, which tells that the quarterly growth rate of GDP is likely to increase by approximately 0.111% for every 1% increase in the quarterly growth rate of the primary industry while other factors remain unchanged. The same applies to b_2 and b_3.

Therefore, the model implies that the primary industry gives a relatively weak impetus to GDP growth. The impact of the secondary industry on the national economy is more obvious and significant, as the tertiary industry follows very closely since China has been vigorously developing it to promote the economic restructuring.

4.3 Hypothesis Testing

Goodness of Fit Test (R^2 Test)

Table 2.1 offers R^2, calculated as approximately 0.972, which can be used to evaluate the overall goodness of fit. R^2 is very close to 1, indicating the correlation is high and the model offers a good fit to data that around 97.2% of the variation of growth rate of GDP can be explained by X_1, X_2 and X_3 based on 80 observations. Further, the adjusted R^2 considers the loss of degrees of freedom due to extra variables. 0.971 is approaching to 1, proving that the model has a particularly good fit.

F Test of Significance

Set the null and alternative hypothesis:

H_0:$\beta_1=0$ and $\beta_2=0$ and $\beta_3=0$

H_1:$\beta_1\neq0$ or $\beta_2\neq0$ or $\beta_3\neq0$

The null hypothesis is that all the coefficients to the X variables are truly and jointly equal to 0, meaning that the model as a whole has no ability to explain the behavior of Y (Kahane,2014). As for the alternative hypothesis, it would be explained that one or more coefficients are not simultaneously equal to 0. In this case, it means the quarterly growth rates of the primary, secondary and service industries have significant influences on that of GDP.

Using a significance level α=5% (or a confidence interval 1-α=95%), the critical value of the F

distribution from Table 2.2, with 3 and 76 degrees of freedom, is 2.725. Because $F_{STAT}=877.377>F_{0.05}(3,76)=2.725$ or because Significance $F=0.000<\alpha=0.05$, we should reject the null hypothesis H_0.That is, the regression is statistically significant, and the quarterly growth rates of the primary, secondary and service industries have significant impacts on that of GDP.

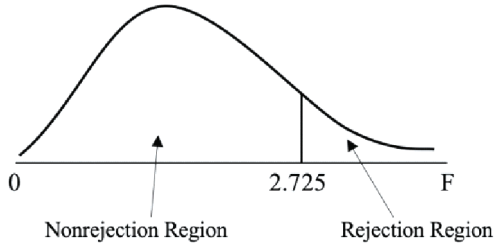

Figure 3. Regions of rejection and nonrejection at the significance level 0.05, with 3 and 76 degrees of freedom

T Test of Significance

Set the null and alternative hypothesis:

$$H_0:\beta_1=0;$$

$$H_1:\beta_1\neq0.$$

Using a significance level $\alpha=5\%$, the critical value of t with 76 degrees of freedom is 1.992. Because $T_{STAT}=2.168>T_{0.025}(76)=1.992$, we should reject the null hypothesis H_0 and get to the result that there is a significant linear relationship between X_1 and Y.

Same as the steps shown above, set the null and alternative hypothesis for β_2 and β_3. Figure 5 illustrates the rejection and nonrejection regions. The similar conclusions can be got that the T_{STAT} lies in the rejection region so that H_0 should be rejected. The growth rates of the secondary and service industries have a significant linear relationship with that of GDP.

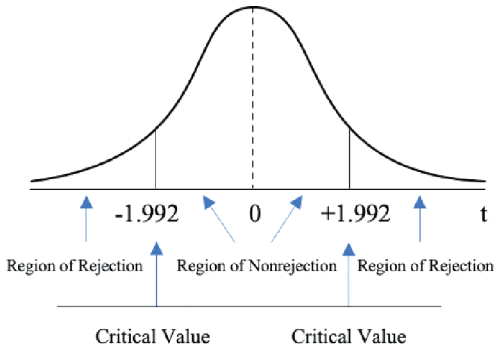

Figure 4. Regions of rejection and nonrejection at the significance level 0.05, with 76 degrees of freedom

The confidence interval approach can also be applied to achieve the same results. From table 2.3, the confidence interval of β_1, β_2 and β_3 is [0.009, 0.213], [0.410, 0.513], [0.343, 0.491], respectively. While the three β^*, being equal to 0, are not included in the intervals. Thus, H_0 should be rejected and there is indeed a significant linear relationship.

4.4 Substitute the Latest Factual Data

Put the data of quarter-on-quarter growth rate of the primary, secondary and service industries in 2020 Q1 into Equation 4:

$$\hat{Y}t=0.002+0.111X_1+0.462X_2+0.417X_3 \text{(Equation 4)}$$

$$X_1=-3.2\%,\ X_2=-9.6\%,\ X_3=-5.2\%$$

We can get the value $\hat{Y}t$ is -0.0676 or -6.76%. That's to say, with the primary, secondary and service industries grow by approximately -3.2%, -9.6% and -5.2% in the first quarter in 2020, the quarterly growth rate of GDP would be -6.76% in prediction. In reality, the whole economy has decreased by 6.8%. Therefore, the estimation and the regression model are suitable and reasonable to some extent.

5 Conclusion

This essay establishes multiple linear regression model, make analysis and get results of the linear correlation between the growth rate of three industries and GDP. The model is reasonable though, it still has some deficiencies. The sample size is 80 which is not large enough. Expanding it will increase the reliability of the model and the accuracy of the analysis results. Besides, the precondition is that there does not exist any perfect linear relationship between the independent variables X_1, X_2 and X_3. That is, we assume that there is no perfect multicollinearity. However, in this model, there are some correlations among the growth rate of the three industries. Therefore, when these three are considered as explanatory variables simultaneously, multicollinearity problems are likely to occur.

Despite some of deficiencies, we can conclude that there exists a significant linear relationship between the quarter-on-quarter growth rates of three industries and that of GDP. In other words, China's GDP has been expanding rapidly from 2000 to 2019, benefiting from the development of three major industries. Particularly, with the secondary industry the leading factor, the contribution of both the secondary and tertiary industries to GDP is prominent, while that of the primary industry is increasingly insignificant.

With the support of policies, the tertiary industry, being equipped with new creativity and more rapid growth, will make more outstanding contributions to the improvement of industrial structure and China's economic growth in the future.

References

[1] Baumol W.J., 1967, 'Macroeconomics of Unbalanced Growth the Anatomy of Urban Crisis', The American Economic Review, 57(3), pp.415-426.

[2] Clark C., 1940, 'The Conditions of Economic Progress', International Affairs Review Supplement, 19(1), pp.36–37

[3] Hartwig J.K., 2012, 'Testing the Growth Effects of Structural Change', Structural Change and Economic Dynamics, 23(1), pp.11-24.

[4] Kahane L.H., 2014, Regression Basics, Thousand Oaks: SAGE Publications, Inc.

[5] Romer D., 2000, Advanced Macroeconomics, Boston: McGraw Hill.

[6] Xiao W., 2018, 'China's Industrial Structure Upgrade in the "New Normal": Empirical Test and Determinants', Singapore Economic Review, 63(4), pp.1037-1058.

[7] Zhao J.F., 2018, 'Industrial Structure Change and Economic Growth: A China-Russia Comparison', China Economic Review, 47, pp.219-233.