Dean&Francis
ISSN 2959-6157

Application of Linear Regression on GDP and International Foreign Exchange Earnings

Yu Lu

Guangzhou Ulink International School, Guangzhou, China

Corresponding author: yulu2518@ulinkcollege.com

Abstract:

GDP is the most significant measure for assessing a nation's or region's economic standing and level of development, which is also the core indicator of national accounting. The rapidly developing tourism industry contributes significantly to GDP, and as China's international status continues to improve, foreign exchange earnings from international tourism also continue to rise. This paper uses The National Bureau of Statistics' data release to measure the four important variables associated with GDP and the three variables associated with international tourism foreign exchange earnings. By constructing a multiple linear regression model, eliminating multiple covariance and using software correction, it is found that there is a positive correlation between GDP and total social consumer goods and net exports, and a positive correlation between foreign exchange earnings from international tourism and the number of international routes of civil aviation. Relevant suggestions are also put forward, such as enhancing the country's economic strength and globalization, building a strong national brand and promoting consumer upgrades by expanding trade with foreign countries, thus improving product quality and meeting consumer demand for a high quality of life. This will not only help to attract more foreign tourists, but will also contribute to sustained GDP growth.

Keywords: GDP; International foreign exchange earnings; Multiple linear regression.

1. Introduction

In the 17th century, gambling games were popular in Europe. French mathematicians Blaise Pascal and Pierre de Fermat discussed gambling problems through correspondence, which became the beginning of probability theory [1]. In the next few centuries, many outstanding mathematicians such as Jacob Bernoulli and Simeon Poisson continuously put forward relevant theories to make probability theory gradually develop into a discipline with a rigorous mathematical foundation. The theory of probability has gradually developed into a discipline with a rigorous mathematical foundation, and has be-

ISSN 2959-6157

come an important branch of modern mathematics and an important tool for solving uncertainty problems. Modern probability theory has been deeply penetrated into various scientific and engineering fields, widely used in finance, statistics, computer science, physics, genetics and other fields [2].

The linear regression is a smaller branch under probability theory, which is widely used in forecasting, trend analysis, causality analysis and so on. Guo Yali empirically analyzed the forecasting of agricultural logistics demand in Chengdu based on multiple linear regression [3]. Liu Dong proposed a mathematical model for predicting urban water use in Chifeng City [4]. It was established by choosing GDP per capita, fixed asset investment, industrial added value, total urban population, industrial water use, daily water uses per capita as independent variables and total water use as dependent variable. Relevant conclusions were obtained to make future predictions. Zou Yuping and Sun Weiguo collected the data of 32 cases of economic residential landscape garden project works in Guangdong, and established a limit design model through factor analysis and regression analysis to accurately and reliably control the price of landscape garden projects [5].

This article will first introduce linear regression, the basic concepts and equations of linear regression, coefficient estimation of the equations. Next, it brings in the use of R-square to determine the effectiveness of the model's fit, as well as the application of linear regression on GDP and international foreign exchange earnings to analyze and predict the results of the study.

2. Linear Regression Theory and Equations

To begin with, the author shall introduce the concepts relating to linear regression method. If an equation is linear, it means that the equation satisfies two properties, namely additivity and odd order. When the equation has additivity $f(X+Y)=f(X)+f(Y)$. When the equation satisfies odd-order property $f(ax)=af(x)$, where a is a constant independent of x. Regression, on the other hand, actually identifies variable relationships between multiple variables that are interdependent. Therefore, linear regression is in fact a statistical method used for forecasting and modelling, mainly for examining how two or more variables relate to one another. The basic principle is to fit a straight line to the data points to determine whether the input and output variables have a linear relationship.

2.1 One-Dimensional Linear Regression

Linear regression is separated into multiple linear regression and one-dimensional linear regression. The distinction is also simple: Multiple linear regression has several independent variables, whereas one-dimensional linear regression simply has one. The general form of a univariate linear regression model is [6]

$$\hat{y}=\hat{b}x+\hat{a} \tag{1}$$

where the random variable y is called the explained variable (dependent variable) and x is called the explanatory variable (independent variable). Also, a is the vector of true residuals $n\times 1$, and one should estimate the parameter b in the regression equation using the least squares method, i.e.,

$$\hat{b}=\frac{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)}{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2} \tag{2}$$

This method is susceptible to outliers, which, if present, can produce misleading results because outliers in the data can drag the least squares fit towards itself. However, robust regression methods can detect outliers even in complex data and give efficient results and give efficient results.

2.2 Multiple Linear Regression

2.2.1 Formulas

One-dimensional linear regression only considers a single variable easily ignores the possible correlation between individual variables, while multiple linear regression is extended by one-dimensional linear regression, which investigates the interdependence of multiple random variables, and can reflect the factual situation more objectively and comprehensively. Let the general variable be and the random variable y's linear regression model be [7]

$$y=\beta_0+\beta_1x_1+\beta_2x_2+\cdots+\beta_nx_n+\epsilon \tag{3}$$

where $\beta_0,\beta_1\cdots\beta_n$ are $n+1$ unknown functions, β_0 is the regression constant, $\beta_0\cdots\beta_n$ is the regression coefficient, y is called the dependent variable, while $x_1,\cdots,x_n$ is a general variable that is precisely controllable and measurable, and is called the independent variable, ϵ is random error. In general, the following steps are involved in developing and accessing a linear regression model: establishing indicator variables in accordance with the study's objectives, gathering and compiling statistical data, figuring out the theoretical regression model's mathematical form, estimating the model's parameters, testing and adjusting the model, and putting the regression model into practice.

2.2.2 Error Function

The regression model is actually finding a line to fit these sample points so that the error between them is as small as possible. Different lines correspond to different errors. From this, an error function is introduced to measure the value of the error and find a line fit with the smallest error. The commonly used error functions are mean square error (MSE), root mean square (RMSE), mean average error (MAE), and R-squared. The percentage of variation in the dependent variable that can be explained by the independent variable is represented by the R-square, which is a relative indicator of how well the regression line fits the observed sample. The formula of R-squared is

$$R^2 = \frac{SSR}{SST} = \frac{\sum_{i=1}^{n}\left(\hat{y}_i - \bar{y}\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} \tag{4}$$

The sum of squares formula can be decomposed as

$$\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 = \sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 + \sum_{i=1}^{n}\left(\hat{y}_i - \bar{y}\right)^2 \tag{5}$$

The sum of squares of a regression is $SSR = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$, and it reflects the magnitude of fluctuations in the n estimates of $\hat{y}_1, \hat{y}_2, \cdots, \hat{y}_n$ [8]. This is because Y and the independent variables $x_1, x_2 \cdots x_n$ do, in fact, have a linear relationship, which may be induced by altering the independent variables $x_1, x_2 \cdots x_n$.

The sum of squares of the residuals $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$, is caused by everything other than the linear relationship of $x_1, x_2 \cdots x_n$ to Y. The total sum of squared deviation (SST) $\sum_{i=1}^{n}(y_i - \bar{y})^2$, which reflects the observed value of Y, with $y_1, y_2 \cdots y_n$ being size of the total fluctuation. Base on the theory of the regression sum of squares and the residual sum of squares, linear regression is more effective when it accounts for a higher proportion of the regression sum of squares in the total deviation of squares, i.e., the regression straight line is a better fit to the sample observations. Whereas if the residual sum of squares accounts for a larger proportion of the total deviation sum of squares, the regression straight line is not a good fit to the sample observations. Therefore, If the coefficient of R-squared is close to 1, the regression equation has a good fit and can explain the majority of the variability. However, if knowledge is limited, the regression equation may need to be modified.

3. Applications of Linear Regression

3.1 Empirical Analysis of China‘s GDP Growth

3.1.1 Background and Data Selection

Gross domestic product (GDP) is the main indicator used in national economic accounting to assess the economic situation and level of development of a country or region. GDP is the measurement of a country's or region's overall economic output over a given time period. This paragraph analyses China's GDP growth by constructing a regression model and makes relevant recommendations [9].

Because GDP (Y) equal to consumptions (R) plus investment (I) plus government expenditures (G) plus net export (N). Base on this principle and in the context of a practical search for relevant information. Four important variables were selected to measure GDP growth, which are Dependent variable: GDP(Y), and independent variable: consumptions (R), investment (I), and net export (NE). This research uses data from the National Statistical Yearbook from 1978 to 2019. Since the economic growth and the variables are in the form of exponential growth, the data are logarithmically processed to transform the original exponential relationship into a linear one. Because of the exponential growth between economic growth and variables, the data can be logarithmically processed, so that the original exponential relationship can be transformed into a linear relationship, and the model of the study is also simplified.

3.1.2 Modeling Of Multiple Linear Regression

Processing the data and making scatter plots of each variable against GDP reveals that they are all linearly related to each other. Make Correlation matrix, find out the correlation coefficients between the explanatory variables are very high, so the first guess is that there is a multivariate linear relationship between the variables. The author shall construct a preliminary theoretical model of multiple linear regression:

$$lny = \beta + \alpha_1 lnI + \alpha_2 lnG + \alpha_3 lnCR + \alpha_4 lnNX \tag{6}$$

The data of the variables after taking the logarithm were imported into the software SPSS 22.0, and this model uses the theory of least squares to seek to maintain the minimum sum of squares of error between the data obtained and the actual data. In this regard, the following results were obtained through regression analysis based on the relevant data.

First of all, from the model summary it is easy to find out that the R-square before and after the adjustment is above 90 percent, indicating that the regression model at this stage is well fitted and realistic. Second, according to the

ISSN 2959-6157

ANOVA results. The sig. value of the ANOVA is significantly less than 0.05, indicating that there is a significant linear relationship between lnY and lnI, lnR, lnG, and $lnNX$.

Table 1. Results of multiple regression analysis.

Model	Unstandardized Coefficients		Standardized Coefficients	Sig.	t
	Std. Error	B	Beta		
(constant)	0.242	1.246		0.000	5.147
lnI	0.079	0.201	0.252	0.017	2.529
lnG	0.056	−0.127	−0.134	0.051	−1.338
lnR	0.096	0.857	0.827	0.056	1.534
lnNX	0.010	0.039	0.059	0.000	4.082

Table 1 is the list of coefficients of multiple linear regression. then the multiple linear regression equation based on the model is:

$$lnY = 1.246 + 0.201lnI - 0.127lnG + 0.857lnR + 0.039lnNX \quad (7)$$

Here, only the sig. value of lnI and $lnNX$ are less than 0.05, so only these two parameters passed the test of significance, while lnG and lnR did not pass the test of significance of the coefficients and the coefficients of lnG are negative, which is not in line with the reality of the economic implications. assessed the level of multicollinearity in multiple linear regression models using the variance inflation factor (VIF). Combining the results of the model coefficient test and the coefficient of variance inflation, it can be determined that there is multicollinearity between the variables So in this paper, stepwise regression is chosen to eliminate multicollinearity. Use ANOVA again and find that the ANOVA results of each model, it can be seen that the observed value of F-statistic of the model is 11332.118 and the sig. value is much less than 0.05, then at the significance level of 0.05 it can be assumed that: there is a linear relationship between lnR, $lnNX$ and lnY.

According to the coefficients for the multiple linear regression, the final model that eliminates the coefficients is

$$lnY = 0.896 + 0.969lnR + 0.047lnNX \quad (8)$$

By making a model fit of the stepwise regression results, the Durbin-Watson test statistic is 0.837, which is a large departure from 2, and there is no definitive proof to show that there is no autocorrelation between the errors [10]. The standardized residual series for the final model, and it can be seen that there is a period of time in the middle where 13 residuals in a row are positive, followed immediately by the last 8 negative ones, it is plausible to believe that there is first-order autocorrelation in the final model error. The validity of the ordinary least squares' estimator cannot be guaranteed when the model is autocorrelated, but this does not affect the unbiasedness and consistency of the estimator, so the ordinary least squares estimator is still used and its corresponding variance is corrected by the serial correlation robust standard error method. According to Eq. (8), gross consumer goods and net exports are the main factors affecting the growth of GDP in the country. Gross consumer goods, net exports, and GDP are positively correlated. Each unit change in total consumer goods has an impact of 0.969 on GDP; an increase in net exports leads to an average increase in GDP, which is in line with China's actual situation.

3.2 Exploring the Influencing Factors of Foreign Exchange Earnings

3.2.1 backgrounds

Tourism has a significant impact on economic development, employment opportunities, and improving people's quality of life. As China's international status continues to improve, inbound tourism has had a remarkable performance in recent years. The scale of the inbound overnight market and the inbound market for foreigners has gradually expanded. The foreign exchange earnings of international tourism is an important economic indicator to measure the development speed and scale of China's inbound tourism. Foreign exchange earnings from international tourism refers to all the expenses incurred by inbound tourists for transportation, sightseeing, lodging, catering, shopping, entertainment and so on during their travels and excursions in China (mainland). This chapter examines the elements impacting international tourism's foreign exchange revenues, aiming to improve foreign exchange earnings, promote economic and social development, and

help China better go global [11].

3.2.2 Research Methods, Data Processing

This paper collects panel data to produce a regional distribution table of international tourism foreign exchange earnings to qualitatively analyze the spatial variability of this economic indicator, and then establishes a multivariate linear regression model to quantitatively analyze the correlation between international tourism foreign exchange earnings, the dependent variable, and the three independent variables.

According to data from the National Bureau of Statistics (NBS), each of the eight economic regions earned millions of dollars in foreign exchange from travel abroad in 2016, 2017, and 2018 respectively, were: the Northeast and Northwest have been less than 5,000, the northern coast is higher than 10,000 and is on the rise, the eastern coast is close to 15,000 and is on a downward trend, and the southern coast is in the shape of a mountain peak. Hovering around 25,000, the middle reaches of the Yellow River are close to 5,000, the middle reaches of the Yangtze River are more than 5,000, the southwest region is more than 10,000, and all are on a downward trend. The calculation shows that the average value of foreign exchange earnings from international tourism in the eight economic regions of China in 2016, 2017 and 2018 were 9280.20, 10120.95 and 9750.74 respectively

Firstly, based on theoretical analysis and experience, three variables are identified and relevant data are collected. The data are collected, respectively, international tourism foreign exchange income ($ billion) gross domestic product ($ billion), consumer price index (previous year = 100), and the number of international routes of civil aviation (airlines) during the period of 2010-2019, as shown in the Table 2.

Table 2. Graph of determining and determined coefficients.

particular year	Year No.	Foreign exchange earnings from international tourism (US$ billion)	GDP (billions)	Consumer price index (previous year =100)	Number of international routes by civil aviation (number)
2010	1	458.14	412119.3	103.3	302
2011	2	484.64	487940.2	105.4	443
2012	3	500.28	538580	102.6	381
2013	4	516.64	592963.2	102.6	427
2014	5	1053.8	643563.1	102	490
2015	6	1136.5	688858.2	101.4	660
2016	7	1200	746395.1	102	739
2017	8	1234.27	832035.9	101.6	803
2018	9	1271.03	919281.1	102.1	849
2019	10	1312.54	990865.1	102.9	953

To proceed further, one can construction of multiple regression equations: $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$, where the dependent variable Y represents foreign exchange earnings from international tourism, the independent variables x_1, x_2, x_3 represent GDP, consumer price index, and the number of international air routes of civil aviation, respectively, and $\beta_0, \beta_1, \beta_2, \beta_3$ are the model parameters.

The results of the multiple regression analysis were the following. The Multiple R is 0.959980459, R Square is 0.921562481, Adjusted R Square is 0.882343722, standard error is 128.3868018, observed value is 10. The final results of the three variables are shown in Table 3. The multiple linear regression equation can be obtained as:

$$Y = 11182.01055 - 0.000596651X_1 - 106.6334693X_2 + 1.791238964X_3 \quad (9)$$

Table 3. Results of the regression analysis.

	Standard error	Coefficients	P-value	t Stat	Upper 95%	Lower 95%	Upper limit 95%	Lower limit 95%
Intercept	4647.90	11182.01	0.05287	2.40582	22555.0	190.995	22555.0	190.995

ISSN 2959-6157

X Var. 1	0.00104	0.00059	0.58575	0.57570	0.00194	0.00313	0.00194	0.00313
X Var. 2	43.9994	106.633	0.05161	2.42352	1.02918	214.296	1.02918	214.296
X Var. 3	0.83693	1.79123	0.07612	2.14023	3.83914	0.25666	3.83914	0.25666

The practical significance of the above regression equation is that for every 100 million yuan increase in GDP, International tourism generates significant foreign exchange revenues. decrease by 0.000596651 billion US dollars. For every one-unit increase in the consumer price index, foreign exchange earnings from international tourism decreased by $10,663,346.93 million; For every additional international airline route, foreign exchange earnings from international tourism increase by $179,123,896.4 million. Meanwhile, the three variables mentioned above fit better with international tourism foreign exchange earnings from the perspective of goodness of fit.

Overall, the model fits well, with a substantial linear relationship between the dependent variable and three independent factors, but it does not get the GDP, consumer price index and the number of international air routes, which have a significant impact on the foreign exchange earnings of international tourism, and from the multiple regression equation, it is observed that the GDP and the consumer price index are negatively correlated with the foreign exchange earnings of international tourism, this is slightly different from the previous knowledge. The number of international air routes is positively related to foreign exchange earnings from international tourism. The higher the number of international air routes, the higher the foreign exchange earnings from international tourism.

4. Conclusion

This paper presents two articles using multiple linear regression models respectively.3.1 For the study, four variables—fixed asset investment, gross social consumer goods, government fiscal expenditure, and net exports—were chosen to gauge GDP growth. The construction of the multiple linear regression model shows that GDP, gross social consumption goods, and net exports are in a positive relationship, while the coefficient of determination is 99.9%, which indicates that 99.9% of the change in China's GDP is due to gross social consumption goods and net exports.3.2, on the other hand, three variables were selected, namely, GDP, consumer price index, and the number of international routes of civil aviation, to measure the growth of The results of multiple regression analysis show that GDP, consumer price index and international tourism foreign exchange income are inversely proportional to each other, and the coefficient of determination is 88.2343722%, which indicates that the above three variables fit well with international tourism foreign exchange income from the perspective of goodness of fit. In general, the results of this paper are accurate and fine with certain reference value, but there are also places that can be improved, there are many other factors that affect the GDP and international tourism foreign exchange income, adding other factors can make the model more accurate.

References

[1] Altman, Naomi and Martin Krzywinski. Simple Linear Regression. Nature Methods, 2015, 12(11): 999–1000.

[2] Wiedermann, Wolfgang, et al. Heteroscedasticity as a Basis of Direction Dependence in Reversible Linear Regression Models. Multivariate Behavioral Research, 2017, 52(2): 222–41.

[3] Guo Yali, Shi Li, Li Zimeng. Empirical analysis of demand forecasting for agricultural product logistics in Chengdu based on multiple linear regression. China Storage and Transportation, 2024, (07): 127-128

[4] Liu Dong. Prediction and Analysis of Water Consumption in Chifeng City Based on Multiple Linear Regression Model. Journal of Chifeng University (Natural Science Edition), 2022, 38 (07): 11-16

[5] Zou Yuping, Sun Weiguo. Application of Multiple Linear Regression in Engineering Quota Design of Landscape Architecture Projects. Chinese Journal of Landscape Architecture, 2021, 37 (06): 87-92.

[6] Li Jun. Prediction method using the multivariate linear regression model based on the factor analysis and its applications on the share price forecast. CAJEPH, 2014, 5(30): 1-40.

[7] Nikolaos Pandis. Multiple linear regression analysis. American journal of orthodontics and dentofacial orthopedics 149 (3), 431-434, 2016.

[8] Weixin Yao, Longhai Li. A New Regression Model: Modal Linear Regression. Scandinavian Journal of Statistics, 2014, 41(3), 656-671.

[9] Jain, Kirti, et al. Applications of Multitarget Regression Models in Healthcare. Machine Learning in Healthcare and Security, CRC Press, 2024.

[10] Zimeras, S., and F. Gerogiakodis. Bayesian Models for Medical Image Biology Using Monte Carlo Markov Chains Techniques. Mathematical and Computer Modelling, 2005, 42(7–8): 759–68.

[11] Schneider, Astrid, et al. Linear Regression Analysis. Deutsches Ärzteblatt International, 2010, 107(44): 776–82.