A Forecast of CO_2 Emissions Based on the Fuel Consumption Rating of a Particular Vehicle

Zikun Chen *

Department of Journalism and Communication, Beijing Sport University, Beijing, China

* Corresponding author: chenzikun2022012157@bsu.edu.cn

Abstract. The accumulation of carbon dioxide will not only cause global warming but also cause a series of catastrophic consequences and cause irreversible damage to the earth's ecology. The prediction of carbon dioxide emissions from light-duty vehicles can provide decision-making support for consumers when purchasing cars and is beneficial to consumers in purchasing environmentally friendly cars. This paper constructs a vehicle carbon dioxide emission model based on a multiple linear regression model, using the four indicators of vehicle engine size, vehicle fuel consumption, Hwy (L/100km), and Comb (L/100km) as output variables. The quantity is the predictor variable. The correlation coefficient of this study is 0.9675. The regression equation of this linear regression model is Y=22.2344+2.7157X1+32.9843X2+26.6259X3-39.1429X4. The order of impact of the four indicators studied on carbon emissions is Comb (L/100km), vehicle fuel consumption, Hwy (L/100km), and vehicle engine size. The determination of carbon emissions of specific models of new light vehicles can refer to the multiple linear regression model proposed in this study.

Keywords: Carbon emissions, multiple linear regression analysis, automobile road testing.

1. Introduction

A series of human activities such as the extensive use of fossil fuels, large-scale deforestation, and over-exploitation of land have led to drastic changes in the global climate. The emissions of greenhouse gases such as CO2 have increased significantly. The accumulation of greenhouse gases in the atmosphere has prevented the heat from the earth's surface from dissipating into space, causing the earth's temperature to rise and causing a series of widespread and far-reaching impacts, including rising sea levels, The frequent occurrence of extreme weather events, destruction of ecosystems, changes in agricultural output, etc., have had a major impact on human society and the natural environment [1]. With the accelerated development of industrialization and the expansion of the scale of emerging economies, China has since become the globe's largest manufacturer and consumer of energy, and it also emits the most CO2 emissions all over the globe. [2]. Currently, modeling analysis is the mainstream of current research for the prediction of CO2 emissions. Utilizing the carbon emission benchmark data from 1980 to 2009, Liu Guangwei recruited the technique of the discrete second-order difference prediction method (DDEPM) to estimate China's carbon emission intensity in 2020 [3]. Qu Shenning et al. compiled panel data from 30 provinces and cities in China and used the STIRPAT model to predict China's future carbon emission peak [4]. Implementing carbon emission data from 1980 to 2009, Zhao Xi et al. established an isolated the second-order difference equation forecasting framework to estimate China's carbon emissions in 2020 [5]. Zhu Yongbin, Wang Zheng, et al. obtained the carbon emissions under the future stable economic growth path by improving the Moon-Sonn model. The peak of carbon emissions will be reached in 2040 [6]. Xiao Zhihong et al. used the combined ARIMA-BP model to predict China's carbon emissions based on the characteristics of carbon emissions [7]. Zhang Fafa et al. combined system clustering and BP neural network fusion model to predict world carbon emissions [8]. Ji Guangyue came across that the gray correlation assessment principle can be utilized when implementing the BP neural network model to the gauge of China's carbon emissions [9].

As an indispensable means of transportation for human beings, cars are one of the important chains of human carbon emissions. Accurate estimation and prediction of the carbon emissions of light vehicles can effectively help consumers measure and compare during the car purchase process, which will help consumers choose environmentally friendly light vehicles and reduce carbon emissions. This article uses multiple linear regression analysis, using the four indicators of vehicle engine size, vehicle fuel consumption, Hwy (L/100km), and Comb (L/100km) as the benchmark, and passed the highway test of Canadian light vehicles in 2015 and 2017. The resulting data can be used to predict and analyze the carbon emissions of light vehicles, helping consumers choose environmentally friendly cars and reduce carbon dioxide emissions during travel.

2. A Forecast of CO_2 Emissions based on the Fuel Consumption Rating of a Particular Vehicle2

2.1. Data Introduction and Variable Setting

This article selects four indicators that are closely related to carbon dioxide emissions. The specific indicators are car engine size, car fuel consumption, Hwy (L/100km), and Comb (L/100km), which are expressed as X1, X2, X3, and X4, and set these four indicators as independent variables, and set carbon emissions as the dependent variable. The data used for the indicator come from two on-road test results in 2015 and 2017 for specific models of new light vehicles for retail sale in Canada in 2023.

2.2. Model Introduction and Research Purpose

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon \tag{1}$$

This article utilizes a model that utilizes multiple linear regression as its data-driven model. The straight-line correlation between a dependent variable (Y) and one or more independent variables (X) is described by linear regression analysis. By fitting the connection between the dependent and independent variables linearly, the model parameters are found. To obtain the regression equation, use the regression equation to predict the changing trend of the dependent variable, and use the regression analysis method to establish a mathematical model that reflects the specific quantitative relationship, that is, the regression model. A phenomenon in real life typically has several contributing components. It is more efficient and practical to forecast or estimate the dependent variable using the best possible combination of several independent factors than it is to use just one independent variable. The number of independent variables is mainly determined by actual influencing factors [10]. The equation of this study is as shown above.

This article aims to explore the linear relationship between the four indicators of automobile engine size, automobile fuel consumption, Hwy (L/100km), Comb (L/100km), and carbon emissions through a multiple linear regression model, as well as the relationship among these four indicators. Which indicator has the greatest impact on carbon emissions, carbon emissions are predicted through regression analysis of the four indicators.

2.3. Data Results and Data Analysis

Table 1. Regression statistics

Multiple R	0.9675
R Square	0.9360
Adjusted R	0.9357
Standard Error	16.4379
Observed Value	833

Table 2. Variance analysis

	df	SS	MS	F	Significance F
regression analysis	4	3271847.8552	817961.9638	3027.2058	0
residual	828	223728.5937	270.2036		
total	832	3495576.4490			

Table 3. Functional relation

	Coefficients	standard error	t Stat	P-value	Lower 95%	Upper 95%	Lower limit 95.0%	Upper limit 95.0%
Intercept	22.2344	2.4344	9.1333	4.9954	17.4560	27.0127	17.4560	27.0127
X1 Variable	2.7157	0.7558	3.5931	0.0003	1.2321	4.1992	1.2322	4.1992
X2 Variable	32.9842	1.2098	27.2641	2.5981	30.6096	35.3589	30.6096	35.3589
X3 Variable	26.6259	1.3451	19.7948	1.0571	23.9857	29.2660	23.9857	29.2660
X4 Variable	-39.1429	2.3395	-16.7316	2.3639	-43.7349	-34.5509	-43.7349	-34.5509

As shown in table 2 and table 3, according to the regression statistics table of multiple linear analysis, the regression equation of this linear regression model is formula 1:

$$Y=22.2344+2.7157X1+32.9843X2+26.6259X3-39.1429X4 \tag{2}$$

As shown in table 1, the correlation coefficient of this study is 0.9675 and is greater than 0.8, so the study's dependent variables and independent variables endure a statistically significant positive correlation. The goodness-of-fit value of this study is 0.935997 greater than 0.8, so the regression model of this study has a good fit. According to the analysis of the variance table, the significance test result of this study is 0. Subsequently, it is extremely noteworthy that each of the dependent and independent variables in this study suffers from a linear relationship. The correlation between each of the dependent variables and its independent variable is competently portrayed through a linear model. The critical value of the significance level is set to 0.01. The third table highlights the study's T-test recommendations. For the sake of the independent factor X1, the P value is 0.0003. This value is far lower than the critical value of the significance level, so the independent variable X1 has strong linear significance for the dependent variable, and its prediction results have extremely significant statistical significance. The parametric correlation between the effect of the independent variables X2, X3, and X4 on the factor that is dependent is not strong given that their P values are all in excess of the critical value of the significance level.

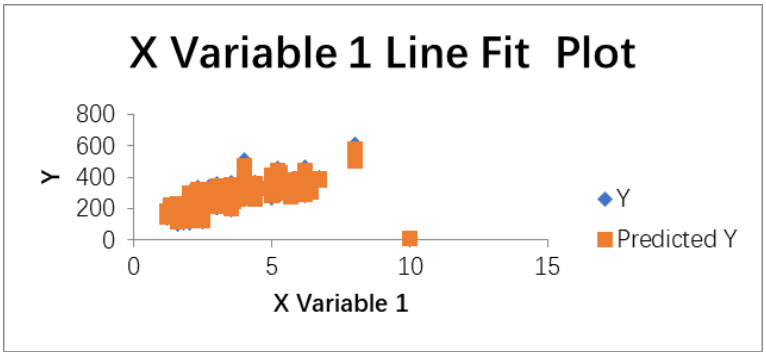

Figure 1. X Variable 1 Line Plot

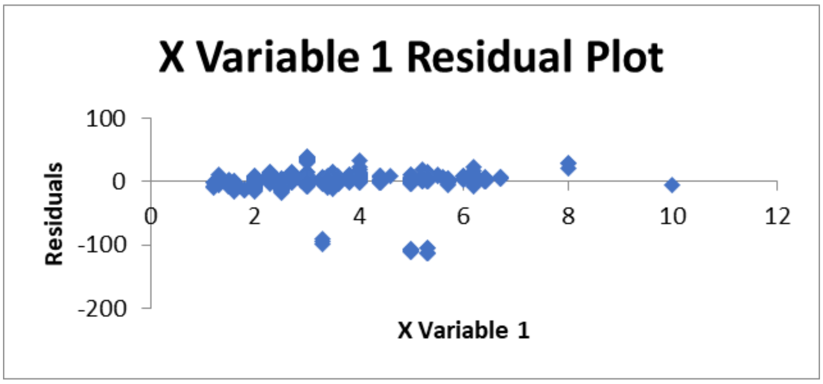

Figure 2. X Variable 1 Residual Plot

From the Figure 1. X Variable 1 Line Plot, it can be seen that the Y value of this experiment and the Y value predicted by the linear regression equation are basically consistent in distribution and numerical size. Combining Figure 1. X Variable 1 Line Plot and the Table 1. Regression statistics illustrate the tremendous relevance of the linear causal connection between the uncorrelated variable (X1) and the variable that serves as the dependent (Y), the linear regression equation has a good fitting effect, and the predicted value of Y basically conforms to the linear regression between the relation. It can be seen from Figure 2. X Variable 2 Residual Plot Most of the scattered points composed of (X1, σ^2) fall within the horizontal banded interval of (-50, +50), but there are some abnormal points in the sample data. Each Xi of the standard residual diagram has a corresponding σ^2, and the entire standard residual diagram shows increasing characteristics, indicating that the random error term has heteroskedasticity. Considering Figure 1. X Variable 1 Line Plot and Figure 2. X Variable 2 Residual Plot, the sample regression model in this study fits the data well.

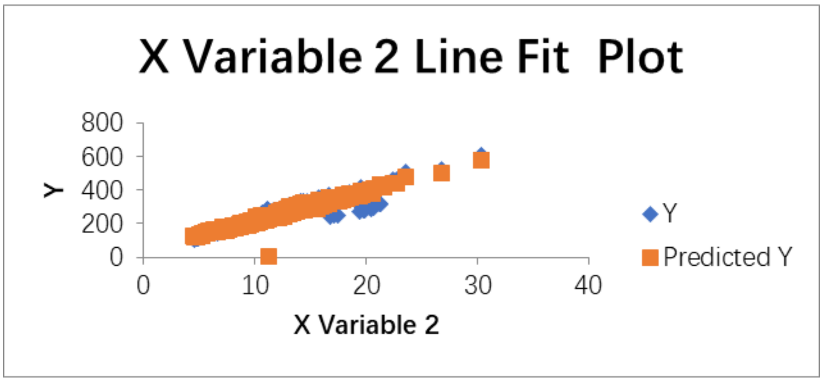

Figure 3. X Variable 2 Line Plot

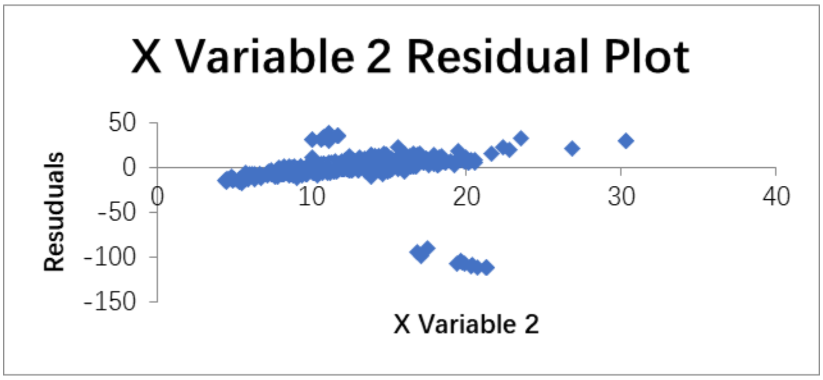

Figure 4. X Variable 2 Residual Plot

From the Figure 3. X Variable 2 Line Plot, the Y values in this experiment are roughly consistent with the Y values predicted by the linear regression equation in terms of distribution and numerical size. There are some Y values that are inconsistent with the predicted values. Combining Figure 3. X

Variable 2 Line Plot and the Table 1. In regression statistics, the Y values in this experiment are roughly consistent with the Y values predicted by the linear regression equation in terms of distribution and numerical size. There are some Y values that are inconsistent with the predicted values. Combining Figure 3. X Variable 2 Line Plot and Table 1. The formula for linear regression has a good fitting effect, the anticipated result of the linear regression association between the uncorrelated variable X2 and the factor that is dependent Y is not substantial, and these findings are readily apparent in the regression statistics. It can be seen from the Figure 4. X Variable 2 Residual Plot Each Xi of the standard residual diagram has a corresponding σ^2, and the entire standard residual diagram shows increasing characteristics, indicating that the random error term has heteroskedasticity. Considering the linear fitting plot and the standard contrast plot, the sample regression model in this study fits the data well.

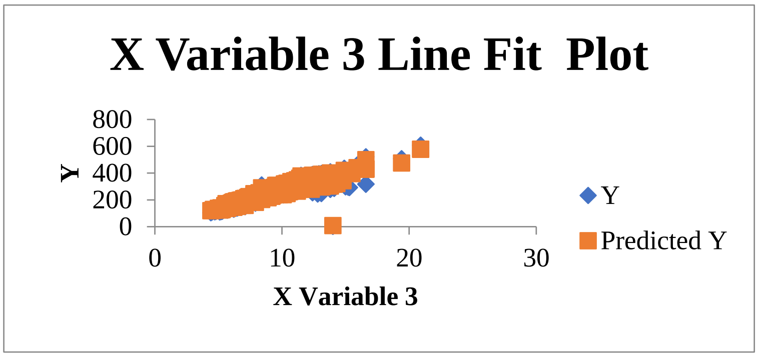

Figure 5. X Variable 3 Line Plot

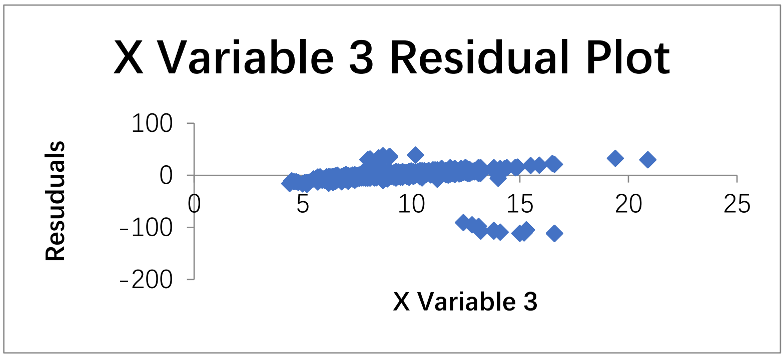

Figure 6. X Variable 3 Residual Plot

From Figure 5. X Variable 3 Line Plot, it can be seen that the Y value of this experiment and the Y value predicted by the linear regression equation are basically consistent in distribution and numerical size. Combining Figure 5. X Variable 3 Line Plot and the Table 1. Regression statistics disclose the existence is a generally statistically significant linear relationship between the control variable X3 and the factor that is dependent Y, and that the formula for linear regression encompasses a good fitting effect. The overwhelming majority of the scattered points rendered up of (X3, σ^2) fall underneath the diagonally banded time frame of (-50, +50), as can be seen from the Figure 6. X Variable 3 Residual Plot. Nevertheless, the sample data incorporates some abnormal points. Each Xi of the standard residual diagram has a corresponding σ^2, and the entire standard residual diagram shows increasing characteristics, indicating that the random error term has heteroskedasticity. Considering Figure 5. X Variable 3 Line Plot and Figure 6. X Variable 3 Residual Plot, the sample regression model in this study fits the data well.

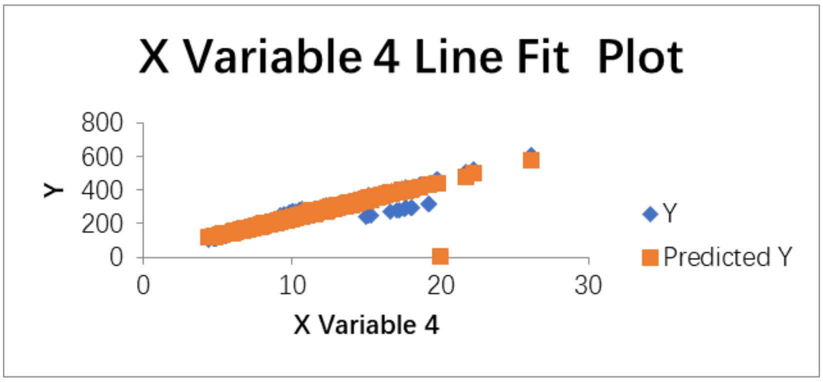

Figure 7. X Variable 4 Line Plot

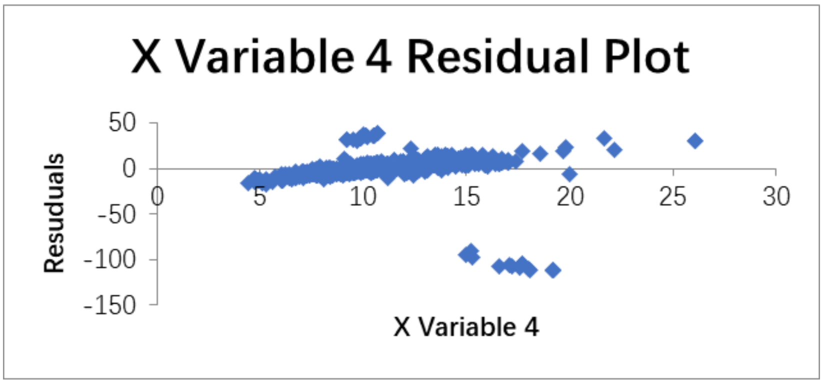

Figure 8. X Variable 4 Residual Plot

From Figure 7. X Variable 4 Line Plot, it can be seen that the Y value of this experiment and the Y value predicted by the linear regression equation are basically consistent in distribution and numerical size. Combining Figure 7. X Variable 4 Line Plot and the Table 1. Regression statistical analysis exhibits the presence is a typically statistically significant linear correlation between the independent variable X4 and the dependent variable Y, and that the linear regression equation encompasses a good blending effect. The bulk of the scattered points executed up of (X4, σ^2) fall within the vertically wrapped interval of (-50, +50), as can be seen in Figure 8. X Variable 4 Residual Plot. Yet, the sample data incorporates some atypical points. Each Xi of the standard residual diagram has a corresponding σ^2, and the entire standard residual diagram shows increasing characteristics, indicating that the random error term has heteroskedasticity. Considering Figure 7. X Variable 4 Line Plot and Figure 8. X Variable 4 Residual Plot, the sample regression model in this study fits the data well.

The standard deviation of the data in this research project is essentially a straight line, as can be recognized in the normal the likelihood diagram. The left tail of the graph formed by the data points is short and curved downward, and the right tail is long and curved upward, showing a right-skewed distribution. The fit between the research data and the research model is good. The structure of the multiple linear regression model aligns with the factors that are independent X1, X2, X3, X4, and the regressive variable Y.

$$Y=22.2344+2.71569X1+32.9843X2+26.6259X3-39.1429X4 \tag{3}$$

3. Conclusion

The regression equation of this linear regression model is: Y=22.2344+2.7157X1+32.9843X2+26.6259X3-39.1429X4. By comparing the regression coefficients of the respective variables in the regression equation, it is unambiguous that the parameter that is dependent Y receives the greatest influence by the uncorrelated variable X4. the independent variable X2 has the second greatest impact on the dependent variable Y, and the independent variable X3 has the third greatest impact on the dependent variable Y, and independent variable X1 has the least influence on dependent variable Y. That is, the impact of the four indicators in this study on carbon emissions is ranked in order: Comb (L/100km), vehicle fuel consumption, Hwy(L/100km), and vehicle engine size. The determination of carbon emissions of specific models of new light vehicles can refer to the multiple linear regression model proposed in this study.

References

[1] Li Shuang, Cao Wenjing, Lu Bin. Research on optimization of my country's energy consumption structure under the constraints of low carbon goals. Journal of Shanxi University (Philosophy and Social Sciences Edition), 2015, 38 (04): 108 - 115.

[2] Chen Liang, Wang Jinhong, He Tao, et al. Research on regional transportation carbon emission prediction based on SVR. Transportation System Engineering and Information, 2018, 18 (02): 13 - 19.

[3] Liu Guangwei, Zhao Tao. China’s carbon emission intensity forecast and tertiary industry proportion test analysis. Economic Management, 2012, 34 (05): 141 - 152.

[4] Qu Shenning, Guo Chaoxian. Research on China’s carbon emission peak prediction based on STIRPAT model. Chinese Population·Resources and Environment, 2010, 20 (12): 10 - 15.

[5] Zhao Xi, Qi Jianmin, Liu Guangwei. China’s carbon emission prediction based on discrete second-order difference algorithm. Arid Area Resources and Environment, 2013, 27 (01): 63 - 69.

[6] Zhu Yongbin, Wang Zheng, Pang Li, et al. Peak prediction of energy consumption and carbon emissions in China based on economic simulation. Acta Geographica Sinica, 2009, 64 (08): 935-944.

[7] Xiao Zhihong, Wang Minghao. Combination model and prediction of China’s carbon emissions. Journal of Chongqing Technology and Business University (Natural Science Edition), 2016, 33 (01): 9 - 15.

[8] Zhang Fafa, Wang Yanxu. Research on world carbon emission prediction model integrating system clustering and BP neural network. Practice and Understanding of Mathematics, 2016, 46 (01): 77 - 84.

[9] Ji Guangyue. Application of BP neural network model based on gray correlation analysis in China's carbon emission prediction. Practice and Understanding of Mathematics, 2014, 44 (14): 243 - 249.

[10] Zhao Jinyuan, Ma Zhen, Tang Hailiang. Comparison of BP neural network and multiple linear regression models in carbon emission prediction. Science and Technology and Industry, 2020, 20 (11): 172 - 176.