Analysis of the Hedging Rate of Used Cars in Beijing by Linear Regression Model

Shuaiming Jing[1, *]

[1] Department of Mathematics, University of California, Santa Barbara, US

*Corresponding author: Shuaiming@ucsb.edu

Abstract. The hedging rate for used cars is a very important index that measures the ratio of how much a car can be sold after driving multiple years out of the price when it was bought. It measures the popularity of the cars in used-car market. The higher the hedging rate for a used car always makes the car more popular in the market. As a result, it is crucial to make an accurate prediction of the hedging rate. This paper selects 12 basic indices for 1500 used cars in Beijing based on a popular used-car website in China to predict which factors have an influence on the hedging rate. By using linear regression model and comparing the p-value of each index with 0.05, 5 indices that have less significance on the hedging rate are eliminated, which are fuel consumption, maximum capacity, environmental protection standard, luggage capacity, and number of transfers. The result shows that the hedging rate has a positive relationship with swept volume, seats, and maximum speed, while it has a negative relationship with registration time, mileage, maximum horsepower, and doors.

Keywords: Hedging rate; used cars; multiple linear regression model.

1. Introduction

As technology develops and the living standards of people become better and better, cars become major transportation for people, and they make people's life much more convenient. However, it is still very difficult for many families to buy a new car, which makes used cars popular in society. In this context, the market for used cars become more and more popular, and the websites of used cars are the most common places where people can buy suitable cars.

Prices are always the first thing people will consider when choosing used cars, which makes the hedging rate a hotly-debated issue. The hedging rate is the current prices or the prices people can sell right now divided by the prices when buying the cars. A higher hedging rate means people can sell their cars at higher prices even though driving for multiple years, which means the cars are popular in the market for used cars. Having a better understanding of this index is good for both car buyers and sellers. Ajupov, et al. stated that due to the more competitive automotive industry in Russia, the hedging rate became more and more and more important for customers and buyers to optimize the prices so that they could make the best decision. [1] As a result, many researchers focus on the factors which would influence the hedging rate of used cars to make an accurate prediction. Hu used a value evaluation model to predict the hedging rate of household used cars based on Hedonic Price Theory [2]. She used 22 indices about the basic configurations, car conditions, and market characteristics. Zhang predicted the hedging rate of used cars in Beijing by using a multiple linear regression model and random forest model regression model [3]. Sameerchand used the combination of multiple linear regression, K-nearest neighbor, decision trees, and Naive Bayes to predict the price of used cars [4]. Winand and George investigated the behavior of both buyers and sellers in the used-car market by using the lemon model based on the assumption that owners of used cars have an information advantage compared to the potential buyers [5]. Kazuyuki and Shigeru explored the used car market in Japan to understand how hybrid vehicles are used on the street [6]. Xiu stated that Japanese cars had the highest hedging rate, which is 59.3%, which was far greater than Germany cars, which had the second highest hedging rate. French cars had the lowest hedging rate, which was 42.4%.[7] Yun used Evaluation Model and Gradient Boosting Decision Tree (GBDT) to predict the hedging rate of used electric vehicles in China. The R^2 of GBDT is model is 97.7%, showing that this model gave the most accurate prediction on the hedging rate among all models. [8] Bao and Jia used multiple linear regression model to predict the hedging rate of used cars on Guaji, which is a popular used car

website in China. [9] Yanqing Chen contended that the hedging rate of new energy vehicle was generally low, but nowadays the hedging rate would improve a lot due to the increasing cost of power battery. [10]

This article mainly uses a multiple regression model to make a prediction of the hedging rate of used cars in Beijing based on 12 basic indices of used cars, and then uses the hypothesis test and p-value to check whether each index has a significant influence on the hedging rate. After excluding some indices which has less significance on the hedging rate from the model, the linear regression model is run again by using R to show the positive or negative relationship between each predictor and the hedging rate.

2. Methodology

2.1 Section Headings

A Linear regression model is a very useful approach to predict the relationship between dependent variables and independent variables, and when there is more than one predictor, a multiple linear regression model would be a good method for the prediction because it can directly accommodate multiple predictors. Instead of fitting a separate simple linear regression model for each predictor, a multiple linear regression model gives each predictor a separate slope coefficient in a single model, which makes the prediction model the most concise.

P-value is an important index that needs to be considered for the multiple linear regression model. It helps to decide whether a hypothesis is significant or not. In fact, is a standard level for the hypothesis. If p-value is less than 0.05, it means that there is enough evidence against the hypothesis test. On the other hand, if the p-value is greater than 0.05, it means that we have enough evidence to explain the hypothesis test. In this model, the p-value is mainly used to exclude some insignificant predictors.

F statistic is also used to test the hypothesis. It is given by:

$$F=\frac{(TSS-RSS)/p}{RSS/(n-p-1)}\text{ , where TSS}=\sum(y_i-\overline{y})^2\text{, RSS}=\sum(y_i-\hat{y})^2 \quad (1)$$

If F-statistic has a value close to 1, it means there is no relationship between the hedging rate and the predictors. On the other hand, F is greater than 1 means at least one predictor has a relationship with the hedging rate.

In this paper, 12 pieces of basic indices of used cars are collected from the website as predictors to find out which factors may influence the hedging rate by using the hypothesis test and multiple linear regression model. The formula of the model is:

$$Y=\beta_0+\beta_1X_1+\cdots+\beta_{12}X_{12}+\epsilon \quad (2)$$

where X_j (1≤j≤12) is each predictor, and β_j is the corresponding coefficient. Y represents the hedging rate, β_0 is a constant term and ϵ is an error term. To find out whether there exists a relationship between each predictor and the hedging rate, we need to check whether the coefficient of each predictor is 0. (i.e. whether $\beta_1=\beta_2=\cdots=\beta_{12}=0$). By using the hypothesis test, we test the null hypothesis:

H_o: $\beta_2=\cdots=\beta_{12}=0$ versus the alternative hypothesis:

H_a: at least one β_j is non-zero.

Then, we compare the p-value of each predictor with 0.05. If it is greater than 0.05, it means there is no significant relationship between the predictor and the hedging rate. Otherwise, if it is smaller than 0.05, it means this predictor has a significant influence on the hedging rate, and the smaller of the p-value means the bigger the influence.

3. Results and Discussion

To explore the market of used cars in China, this paper selects the basic information of 1500 used cars in Beijing, which is the capital of China. The data is collected from You Xin website, which is a popular website for used cars in China.

3.1 Introduction of Some Indices

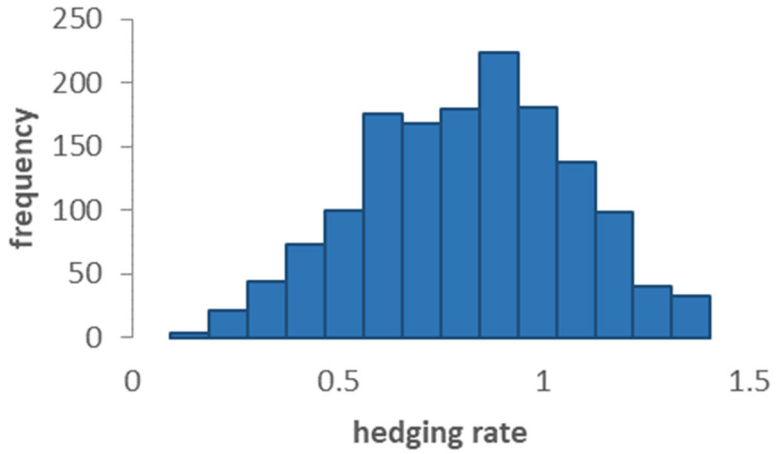

Fig. 1 Histogram of the hedging rate

The hedging rate is a very important index people need to consider when buying new cars or used cars. The hedging rate is a value between 0 and 1. It is given by the current price of a car after using a period time divided by the price of the new car. Car buyers can choose to buy cars at higher hedging rates so that they can sell their cars at relatively higher prices after driving for a long time. According to the histogram of the hedging rate, the value of the hedging rate mostly ranges from 0.3 to 0.8.

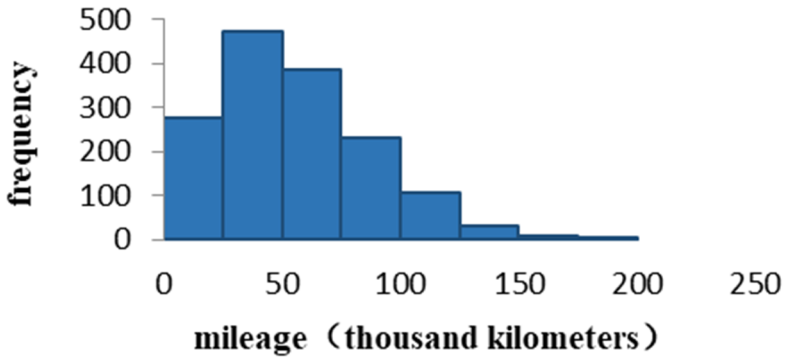

Fig. 2 Histogram of the mileage

Mileage is an index that measures the degree of usage of a car. It refers to the distance the cars run by the previous owner. Buyers are more willing to buy a car with less mileage since it means the car is newer. According to the histogram of the mileage, most used cars have mileage between 0 to 100 thousand kilometers.

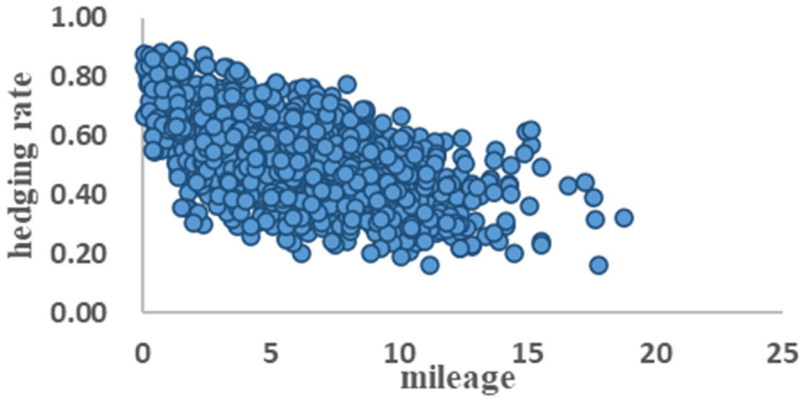

Fig. 3 Scatter plot between mileage and hedging rate

Mileage is also an important index that needs to consider in the multiple linear regression model since it directly influences the hedging rate. According to the scatter plot of mileage and hedging rate, it is easy to see that the hedging rate and mileage have a negative relationship. A used car that has more mileage always tends to have a lower hedging rate, which also means that the coefficient of mileage is negative in the model.

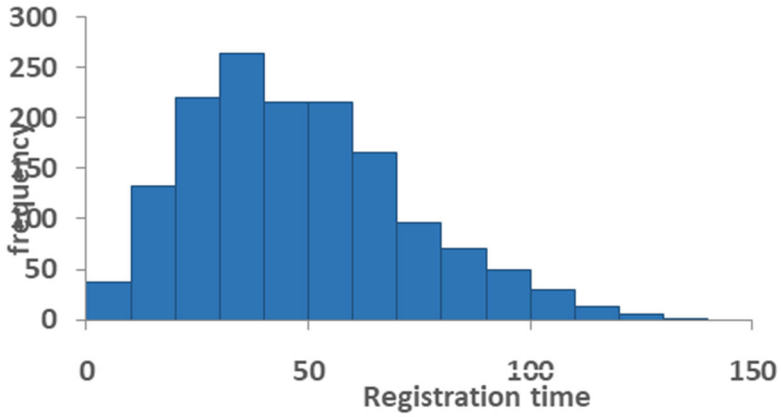

Fig. 4 Histogram of registration time

Registration time is the time that a new car first gets its license plate. It is the age of a car. In this article, registration refers to the months from the first time a car gets its license plate to August 2022. According to the histogram of registration time, most car owners decide to sell their cars within 100 months, which is about 8 years. In addition, few people decide to sell their cars within 10 months or after 100 months. In other words, few people are willing to sell the cars if they are fairly new or very old.

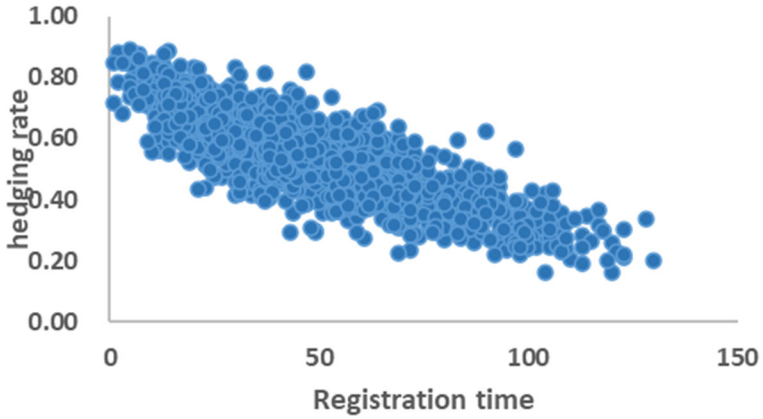

Fig. 5 Scatter plot between registration time and hedging rate

Registration time is also a very important index in the model which directly influences the hedging rate. According to the scatter plot, a used car that has a longer registration time tends to have less hedging rate. In other words, the older the car is, the lower the hedging rate it has. This also means the coefficient of registration time is negative in the model.

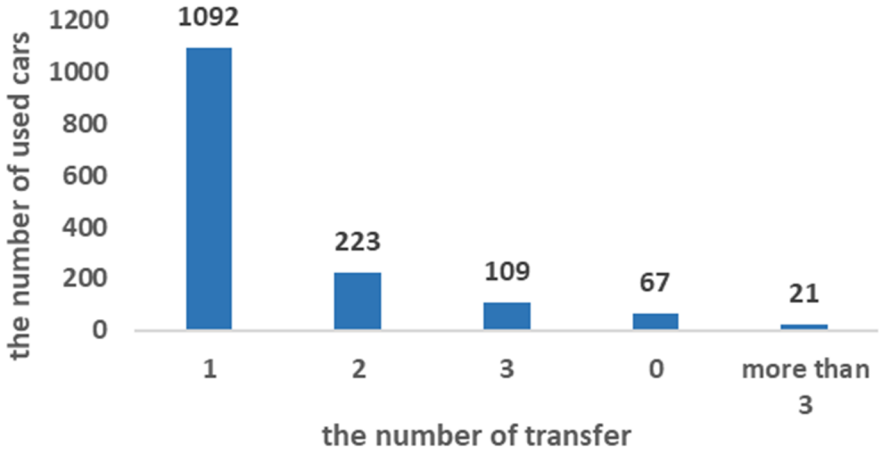

Fig. 6 Column chart of the number of transfer

The number of transfers lets buyers know how many previous owners a used car has. Generally, a car has a much number of transfer means it not satisfy the basic needs of its previous owners. According to the column chart, the number of transfers is 1 for most used cars, and just 21 used cars have more than 3 for the number of transfers, which means the vehicle condition may not be very good.

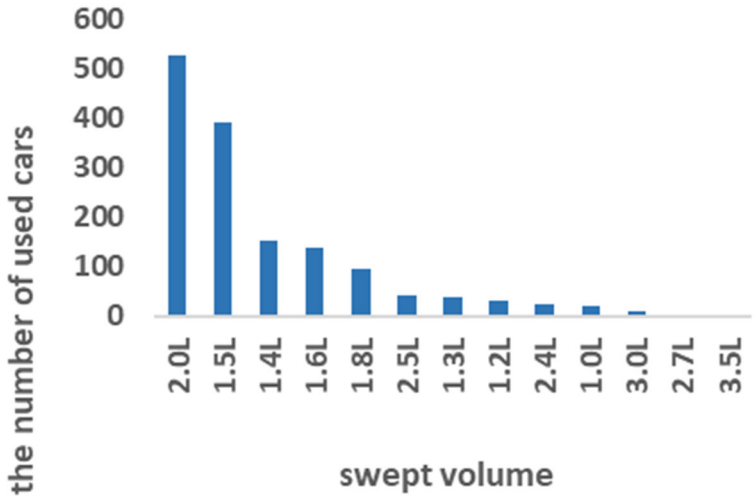

Fig. 7 Column chart of swept volume

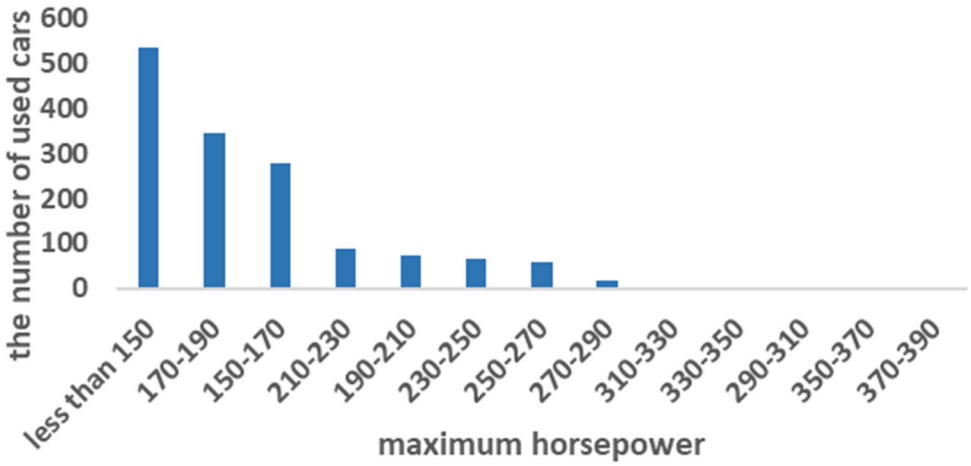

Fig. 8 Column chart of maximum horsepower

The swept volume and maximum horsepower the performance of the engine. The more swept volume and maximum horsepower a car has means it can accelerate more quickly, and many buyers will focus on that when choosing their vehicles. According to the two charts above, most used cars have 2.0L swept volume and 1.5L swept volume, and have a maximum horsepower of less than 150.

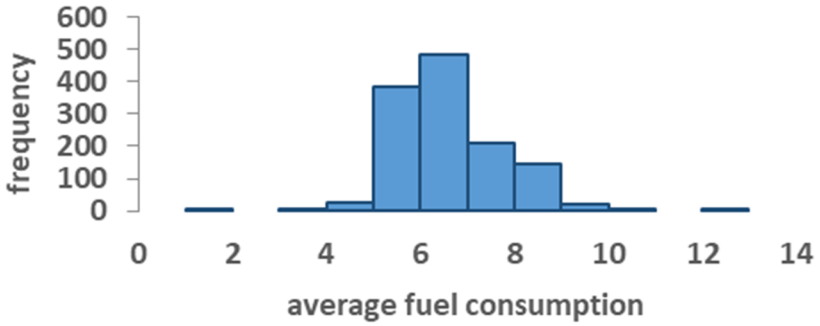

Fig. 9 Histogram of average fuel consumption

Average fuel consumption is an important index for the vehicle. Buyers need to consider it when buying cars because it helps them to predict how much it would cost for the future use of the cars. According to the histogram, the average fuel consumption of most used cars is 6.

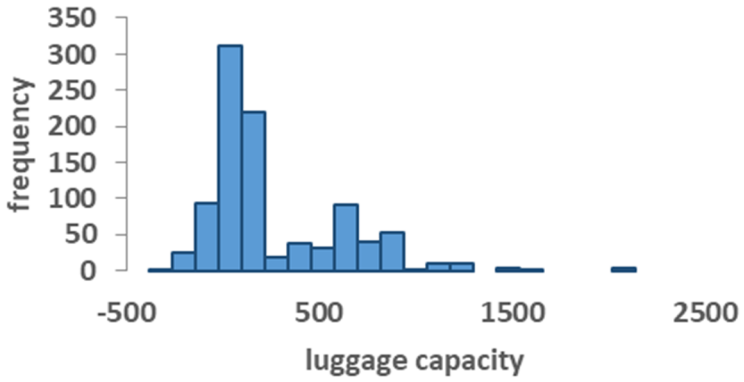

Fig. 10 Histogram of luggage capacity

Luggage capacity is not important for some buyers, but for people who travel a lot or for some commercial use, it is an important index that will be considered because these people always need more luggage capacity to accommodate their luggage.

3.2 Model construction

This article uses R to construct a multiple linear regression model based on 12 variables: registration time, mileage, swept volume, fuel consumption, maximum capacity, maximum horsepower, environmental protection standard, doors, seats, luggage capacity, maximum speed and number of transfer. The result is following:\

Table 1. The result of the multiple linear regression model

	Estimate	Std.Error	t-value	Pr(>|t|)	Sig.
Intercept	5.951×10^{-1}	5.45×10^{-2}	10.907	<0.001	***
Registration time	-4.178×10^{-3}	1.60×10^{-4}	-25.990	<0.001	***
Mileage	-6.183×10^{-4}	8.31×10^{-5}	-7.438	<0.001	***
Swept volume	8.979×10^{-2}	8.15×10^{-3}	11.012	<0.001	***
Fuel consumption	7.223×10^{-4}	2.57×10^{-3}	0.281	0.778	
Maximum capacity	-1.622×10^{-4}	1.36×10^{-4}	-1.193	0.233	
Maximum horsepower	-3.62×10^{-4}	1.18×10^{-4}	-3.071	0.002	**
Environmental	-4.64×10^{-4}	6.14×10^{-3}	-0.076	0.939	
Doors	-9.22×10^{-3}	3.97×10^{-3}	-2.323	0.020	*
Seats	1.758×10^{-2}	4.56×10^{-3}	3.850	<0.001	***
Luggage capacity	-1.79×10^{-5}	1.74×10^{-5}	-1.028	0.304	
Maximum speed	3.062×10^{-4}	1.37×10^{-4}	2.234	0.025	*
Number of transfer	2.678×10^{-4}	2.72×10^{-3}	0.098	0.921	

The F-statistic is 276.8, which is greater than 1. This means H_a is true. In other words, at least one predictor is related to the hedging rate. The corresponding p-value is far less than 0.05, which means the model passes the test. However, the p-values of some predictors are smaller than 0.05, which means they have little relationship with the hedging rate. The "*" in the table also shows how strong of the relationship. The more "*" a predictor has, the more significant it is. It is obvious that 4 predictors have a p-value greater than 0.05, and they all have no "*", which means they have little influence on the hedging rate. After deleting these predictors, the result is following:

Table 2. The revised linear regression model

	Estimate	Std.Error	t-value	Pr(>|t|)	Signif. codes
Intercept	5.97×10^{-1}	3.69×10^{-2}	16.186	$< 2\times 10^{-16}$	***
Registration time	-4.14×10^{-3}	1.08×10^{-4}	-38.189	$< 2\times 10^{-16}$	***
Mileage	-6.17×10^{-4}	8.24×10^{-5}	-7.479	1.2×10^{-13}	***
Swept volume	8.91×10^{-2}	8.02×10^{-3}	11.115	$< 2\times 10^{-16}$	***
Maximum horsepower	-4.63×10^{-4}	7.18×10^{-5}	-6.444	1.5×10^{-10}	***
Doors	-9.29×10^{-3}	3.90×10^{-3}	-2.379	0.017	*
Seats	1.694×10^{-2}	4.48×10^{-3}	3.774	0.0001	***
Maximum speed	2.625×10^{-4}	4.48×10^{-3}	2.022	0.043	*

The F-statistic is greater than 1 with a corresponding p-value far less than 0.05, which means the adjusted model also passes the test, and now each predictor has a p-value less than 0.05, so they all have a significant influence on the hedging rate. However, doors and maximum speed have comparatively less influence compared to other predictors since the corresponding p-values of these two predictors are not small enough.

Thus, the equation of the multiple linear regression model can be written as:

$$\hat{Y} = -0.004148X_1 - 0.000617X_2 + 0.08919X_3 - 0.0004631X_4 - 0.009293X_5 + 0.01694X_6 + 0.0002625X_7 + 0.5978 \quad (3)$$

where X_1, X_2, X_3, X_4, X_5, X_6, X_7 represent registration time, mileage, swept volume, maximum horsepower, doors, seats, and maximum speed respectively.

According to the equation, the coefficients of registration time, mileage, and maximum horsepower is negative, which means these three predictors have negative relationships with the hedging rate. The registration time and mileage measure the age of the cars and the degree of usage for used cars. It is obvious that the bigger he value of these two indices means the older the cars are, which will decrease the price of the cars for sure, leading to a lower hedging rate. As for the maximum horsepower, it measures the performance of the engine for the cars. People are willing to buy used cars with better performance for sure, but the value of the maximum horsepower is bigger also brings more pressure to the engine and more cost for the cars, so the price of the cars will decrease, leading to a lower hedging rate.

The swept volume, doors, seats, and maximum speed have positive relationships with the hedging rate. The bigger the swept volume is, the faster a car can accelerate, which makes the car more attractive for most people, so the hedging rate will increase. Doors and seats are basic needs for passengers. More doors and seats mean there is more space in the car, which provides a more comfortable room. Moreover, most cars have two doors and four doors, but two-door cars are sports cars, which have greater maximum horsepower, so they have a lower hedging rate according to the previous explanation. Thus, doors and seats have a positive relationship with the hedging rate. Buyers also tend to pursue cars with greater maximum speed, which leads to a positive relationship with the hedging rate.

As a result, when buying used cars, people should look for cars that have a short registration time, less mileage, and less maximum horsepower. On the other hand, the cars should have more swept volume, doors, and seats, and the maximum speed should be as better as possible. In this case, the hedging rate will become larger, which means the car buyers can sell their cars at good prices even though driving for multiple years.

4. Conclusion

With the development of technology and the improvement of people's living quality, used cars become more and more popular in our lives. Thus, it is particularly important to make an accurate prediction of the hedging rates of used cars so that buyers can make the best decisions. In this paper, a linear regression model is used to predict the hedging rate of used cars in Beijing based on 12 basic indices of used cars: registration time, mileage, swept, fuel consumption, maximum capacity, maximum horsepower, environmental protection standard, doors, seats, luggage capacity, maximum speed, and the number of transfers. Then, five indices that have less significant influences on the hedging rate are eliminated from the model, which are fuel consumption, maximum capacity, environmental protection standard, luggage capacity, and the number of transfers, according to the p-value. By multiple linear regression model, registration time, mileage, and maximum horsepower have a negative relationship with the hedging rate, while the swept volume, doors, seats, and maximum speed have positive relationship with the hedging rate. As a result, when buying used cars, in order to sell it at good prices after driving for multiple years, consumers should focus on cars which have less registration time, mileage and maximum horsepower. On the other hand, they should look for cars with greater swept volume, doors, seats, and maximum speed. In this case, the cars will have a greater hedging rate, which means the cars will be sold at good prices a few years later. For car sellers, they can make a prediction of the current prices of their cars based on the model in this paper so that they can sell their cars at reasonable prices.

References

[1] Ajupov A.A., Kurilova A.A., and Ivanov D.U. Hedging as an important component of the Financial Mechanism of Enterprise Management in the Automotive Cycles. Mediterranean Journal of Social Science, 2015.

[2] Dan Hu. Research on Value Evaluation Model of Household Used Car Based on Hedonic Price Theory. Chongqing University of Technology, 2022.

[3] Yue Zhang. Analysis on the Influencing Factors of the Preservation Rate of Used Vehicles in Beijing. Tianjing University of Finance and Economics, 2019.

[4] Pudaruth Sameerchand. Predicting the Price of Used Cars using Machine Learning Techniques. International Journal of Information & Computation Technology, 2014, 753-764.

[5] Emons Winand and George Sheldon. The market for used cars: new evidence of the lemons phenomenon. University of California, Santa Barbara, 2009, 2867-2885.

[6] Iwata Kazuyuki and Matsumoto. Use of hybrid vehicles in Japan: An analysis of used car market data. ScienceDirect, 2016, pp. 200-206.

[7] Chen Xiujuan. The hedging rate: Japanese brands dominate the list. China Academic Journal Electronic Publishing House. China Academia Journal Electronic Publishing House, 2021.

[8] Lin Yunzhou. Research on Evaluation Model and Application of Used Electric Vehicle Based on GBDT. Chongqing University of Technology, 2022.

[9] Lin Jiahui and Zhou Baojian. Analysis of the hedging rate of used cars based on multiple linear regression model. China Academia Journal Electronic Publishing House, 2021.

[10] Chen Yanqing. The hedging rate of new energy vehicle. China Academia Journal Electronic Publishing House, 2022.