

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/370523018>

data analysis project used car sale price

Research · May 2023

CITATIONS

0

READS

2,059

1 author:



Priyadarshini Angadi
Purdue University West Lafayette

2 PUBLICATIONS 0 CITATIONS

SEE PROFILE

The Sale Price of Used Cars

Priyadarshini Angadi

Abstract

The market for used cars is a complex and constantly evolving landscape. Buyers and sellers alike are constantly seeking ways to gain an edge in this market, and one key factor is understanding the value of a particular vehicle. In this study, we analyze a dataset consisting of 18,647 used vehicles listed for sale in 2023 within 25KM proximity of downtown Toronto, Ontario, Canada. By using econometric tools, the aim is to identify the factors that have the most significant impact on the price of a used vehicle. The factors analyzed are year, mileage, and controlling variables such as make, body type, engine, exterior colors, passengers, and fuel type. By analyzing these factors and their respective relationships with the price of a used vehicle, I can help buyers and sellers make more informed decisions, promoting efficiency in the used car market. Overall, researching used car prices is an important economic activity that can lead to better outcomes for consumers, sellers, and the economy.

1. Introduction

The market for used cars is a significant segment of the automotive industry, with millions of used cars sold each year around the world. This market includes a variety of vehicles, ranging from economy cars to luxury vehicles, and is driven by a wide range of consumer preferences and needs. The global used car market size was valued at \$1.41 trillion in 2020 and is projected to reach \$2.48 trillion by 2028, growing at a CAGR of 7.9% during the forecast period.

Research on the factors influencing used car prices can provide valuable insights into the market dynamics affecting supply and demand. A stable and efficient used car market can provide affordable transportation options for consumers, promote competition and innovation among businesses, and contribute to the economy's overall health. Understanding the relationship between supply and demand in the used car market is also important. Factors such as the availability of financing, the level of competition among dealerships, and the age and condition of the vehicles can all impact the prices that consumers are willing to pay for used cars. By studying these and other factors, policymakers and businesses can develop strategies to promote a healthy and efficient used car market. This can include initiatives to promote transparency and fair competition, efforts to reduce barriers to entry for new players in the market, and investments in technologies and infrastructure that can improve market efficiency and accessibility. In this paper, I analyze the effect of two of these factors, year and mileage on used car sales prices while controlling other variables like fuel type, engine type, make and passenger seats.

2. Data

The dataset for used car sales was sourced from Kaggle [1]. The used cars data includes details of 18,647 used vehicles listed for sale in 2023 within 25KM proximity of downtown Toronto, Ontario, Canada and includes year, make, model, kilometers, body type, engine, transmission, drivetrain, exterior color, interior color, passengers, doors, fuel type, city fuel economy, highway fuel economy, price, and various other numeric features. The summary statistics for factors relevant to this study, namely year and mileage (referred to as kilometers in the rest of this paper), are shown in Fig 2.1 below.

Variable	Obs	Mean	Std. dev.	Min	Max
price	18,647	47450.54	53371.45	2000	1699998
kilometres	18,647	65777.06	63845.7	0	480000
year	18,647	2018.713	4.003967	1958	2023

Fig. 2.1 Summary statistics for Price (in thousands), milage (kilometer), and year.

Fig 2.2 and Fig. 2.3 show Price vs kilometers and Price vs Year respectively. The variables in the dataset are difficult to read intuitively, so throughout this paper, I will be using $\ln(\text{Price})$ so that the coefficient estimates provide much cleaner interpretations similar to the method described in [2]. I will also be using $\ln(\text{Kilometers})$ for the same reason.

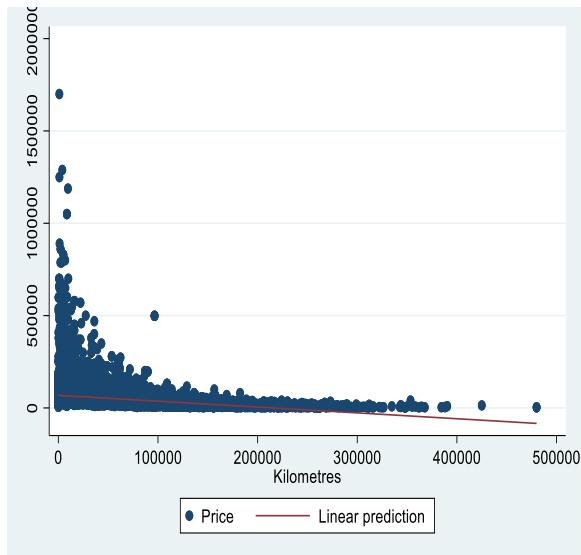


Fig 2.2 Price v/s kilometers

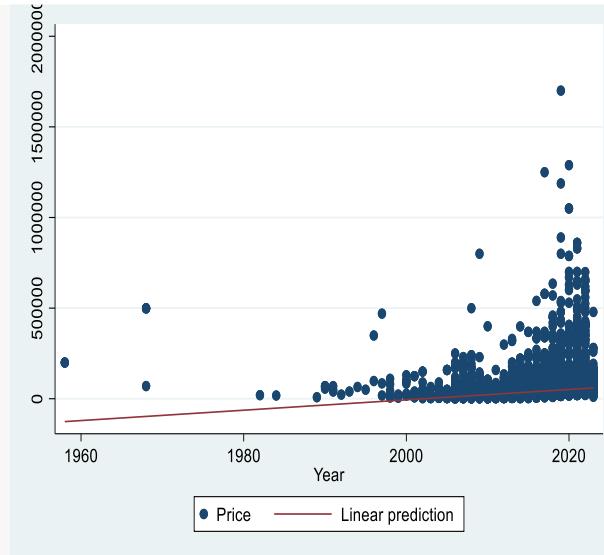


Fig 2.3 Price v/s Year

3. Methodology

This paper will focus on comparing the effects of year and $\ln(\text{kilometer})$ on the sale price for used cars, so it is useful to establish a first impression of those relationships. The sale price of used cars is expected to be negatively correlated with mileage but positively correlated with age. The graphs in Fig 3.1 and Fig. 3.2 demonstrate the simplest relationship between each explanatory variable on the outcome variable. The graphs presented reflect this expected correlation.

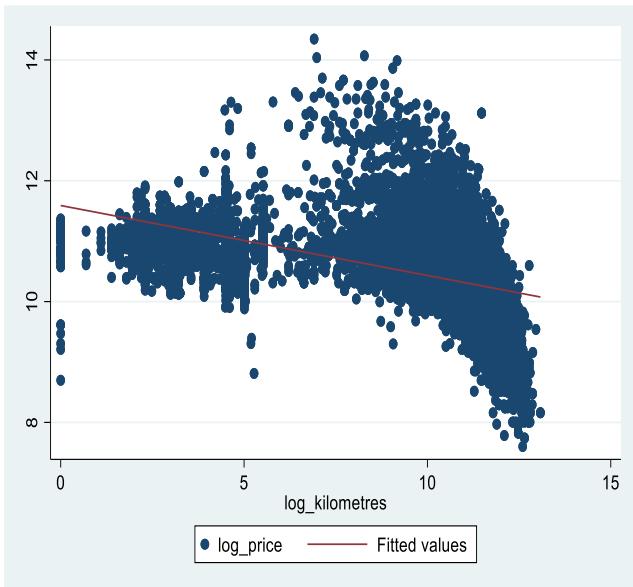


Fig 3.1 $\ln(\text{Price})$ v/s $\ln(\text{kilometers})$

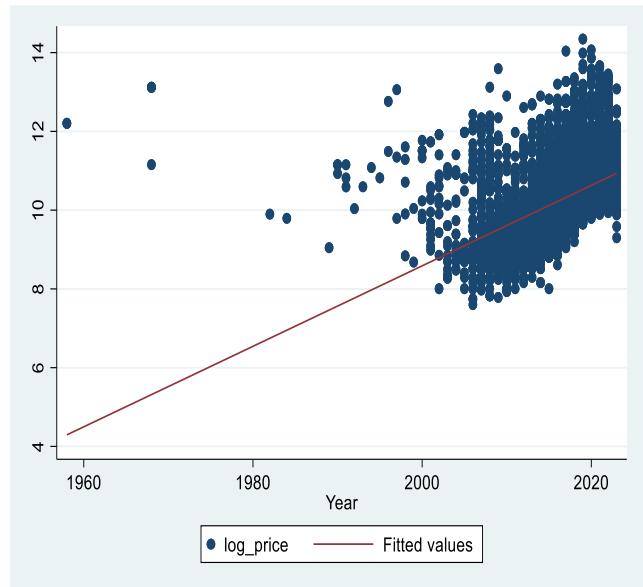


Fig 3.2 $\ln(\text{Price})$ v/s Year

The magnitudes of the slopes of the two variables to determine which variable has a stronger impact on the sale price of used cars are then compared. The slope of age(year) is expected to be more pronounced than the slope of $\ln(\text{kilometer})$, suggesting that age(year) has a stronger impact on the sale price of used cars. However, other factors that may affect the sale price of used cars, such as the make, fuel type, engine type of the car need to be controlled to refine our understanding of the relationship between mileage, age, and sale price.

4. Results

Effect of Milage and Year on the Price of the Used Car

Variables	In(price)	In(price)	In(price)	In(price)	In(Price)
log_kilometres	-.0551003 (.0017498)	-.0554952 (.0017394)	-.0399138 (.0015298)	-.0331623 (.0009309)	-.0307414 (.0009251)
year	.0749574 (.0013501)	.0729558 (.0012917)	.0848854 (.0011815)	.0924146 (.0007054)	.0917296 (.0006989)
fuel type	N/A	YES	YES	YES	YES
engine type	N/a	N/A	YES	YES	YES
make	N/A	N/A	N/A	YES	YES
passenger seats	N/A	N/A	N/A	N/A	.0041812 (.0049463)
Observations	18,647	18,647	18,647	18,647	18,167
R-squared	0.3608	0.3747	0.5252	0.8769	0.8819

Fig. 3.3 Coefficient estimates on mileage and year's effects on Price. Fuel type, engine type, make, and passenger controls are included.

The above statistical analysis in Fig 3.3 presents a multiple linear regression model to predict the sale price of a used car based on various predictor variables. The model uses a sample of 18,647 observations and has a high significance level, with a p-value of 0.0000, indicating that the predictor variables strongly influence the sale price.

The results indicate that the logarithm of the car's mileage has a negative coefficient of -.0307414%, which means that as every 100 unit increases in the log kilometer(117kil), the sale price decreases by 3%. On the other hand, the year of the car (year) has a positive coefficient of .0917296%, indicating that newer cars tend to fetch higher sale prices holding all other variables constant.

The presence of control variables in the model such as fuel type, engine type and make helped us to understand the marginal effect of mileage and year on price better. The number of passengers has a positive coefficient, indicating that as the number of passengers a car can accommodate increases, the sale price tends to increase. Controlling fuel type, engine type and make variables has removed the bias from the milage(kilometer) and the year coefficient.

The model could achieve an overall 88.19% accuracy for the given data. Overall, this multiple linear regression model provides insights into the various factors that influence the sale price of a used car, which can be useful for both buyers and sellers in the used car market.

5. Conclusion

The regression model above was used to find significant influences on the sale price of used cars based on several variables. The model had a high degree of fit, with an R-squared value of 0.8819, indicating that approximately 88.19% of the variation in the sale price of used cars was explained by the predictor variables. Of these variables, the year had the most significant effect on the sale price, with a coefficient of .0917296. This means that for a year increase in the age of a car significantly fetches a higher price about .09%, provided all other variables remain the same. The sale price of the used car decreased by 0.03074% for a unit increase in the $\ln(\text{kilometer})$, provided all other variables remain the same. Overall, the model provides valuable insights into the factors that influence the sale price of used cars and can be useful for making predictions about the sale prices of cars in the future.

One of the main reasons why people buy used cars is affordability. Used cars generally have a lower price than new cars, making them a popular option for buyers who want to save money. Additionally, used cars often have a history of maintenance and repairs that can give buyers a better understanding of the car's condition.

However, buying a used car can also come with some risks. The car may have hidden issues or questionable histories, such as accidents or flood damage, which could impact its value and safety. To mitigate these risks, buyers often conduct a thorough inspection and request vehicle history reports before making a purchase. Factors that influence the price of a used car include the car's age, mileage, fuel type, engine type, make, and model. This study shows that out of these factors, year and mileage are the most economically significant factors in deciding the price of a used car.

6. References

- [1] <https://www.kaggle.com/datasets/farhanhossein/used-vehicles-for-sale?resource=download> used car sales web scrapped data from autotarder.com
- [2] Price Anomalies in the Used Car Market: Marco Haan and Peter Koo reman [\(PDF\) Price Anomalies in the Used Car Market \(researchgate.net\)](#).