

What Is Your Car Really Worth?

Jonathan Shing
Econometrics, Fall 2002

Executive Summary

Searching for the best deal when purchasing a vehicle can often be an arduous task. My goal is to simplify it. I have investigated the effect of different variables on used car pricing as well as depreciation and formed three different models.

The first model, Model One, accounts for car prices using the National Automotive Dealers Association's guides that it releases quarterly. It revealed that 94% of used car pricing is explained by the N.A.D.A. valuations as well as if the car is American or foreign. Model One is by far the best model however it does not explain the real factors behind the price.

Model Two attempts to explain what Model One left a mystery, that is, a pricing model based purely on objective attributes of a vehicle. Running a regression, I have found that both age and mileage cause the value of a car to deteriorate, while original price, horsepower and the global origin of the vehicle contribute positively. This model combined with Model One provide two methods of measurement, to ensure that the an ideal transaction will be made.

The third and final model, Model Three, studies the effect of the objective variables on percent depreciation. Depreciation is important for not only car-pricing but also for tax effects as well as asset valuation. Current methods for measuring depreciation make many assumptions about salvage value and lifespan that Model Three clarifies. More specifically, Age, Mileage, Number of Doors, and Domestic or Foreign-Made are the variables that had a significant effect on percent depreciation.

I have also conducted a test of the model using out of sample data to check the accuracy of the model. It turns out, that there are two ways to interpret the output that the models yield. Assuming the models are indeed the true predictors, then the difference between the actual and predicted price will tell whether a car is undervalued or overvalued. On the other hand, a discrepancy in price could also signal that the models are either flawed or there is another variable not considered. The best method of verification is to look at the raw data and check if the car is priced above what the N.A.D.A. guide recommends.

Only a couple true downfalls exist in these three models. Subjective factors such as the upholstery condition, the timeliness of maintenance and on-site bargaining are difficult to measure. Given these concerns, the models still yield a fair estimate of what is a "good" or "bad" deal when purchasing a vehicle.

Introduction

The automobile market is both vast and vague. If a person walks into a car dealership or flips through the classified advertisements, how does he know if he is really receiving a good deal? Pricing seems arbitrary, and the “bargaining factor” extends this matter further. I have investigated the key elements that I believe contribute the price of automobiles. More specifically, I will explain the effect of each variable on used car prices.

Hypothesis

As previously alluded to, car prices vary. Searching for a car and the respective “best” price often leaves people frustrated and unsure of their purchase. My goal is to create a model that approximates the amount that someone should pay for a vehicle. I am using the following factors construct this model: age, origin of manufacturer, number of doors, horsepower, mileage, original price, National Automobile Dealers Association (N.A.D.A.) average trade-in value, and N.A.D.A. average retail value. I believe that each of these is something that people take into consideration, directly or indirectly, when pricing a car.

Using the same variables, I will also look at depreciation of vehicles. Since car prices can range from less than \$1000 to over \$100000, looking at percent depreciation will uncover a clearer more objective model. For the car-buyer, these models will signal if a car is overpriced or under-priced and on the other side of things, these models will calculate the optimal price for a seller to charge.

Data

I have extracted data from a number of sources including: Seattle Times Classified Advertisements, “Auto Trader” magazine of automotive advertisements, Standard Catalog of American Cars 1976-1999 by James M. Flamming & Ron Kowalke, Standard Catalog of Imported Cars 1946-2002 by Mike Covello, National Automobile Dealers Association Official Used Car Guide Fall 2002 (N.A.D.A.) and local car dealerships. I should note that the price of the vehicle is in today’s terms and is actually the asking price of a private seller. The original price represents the average price of the car when it was new, with no inflation adjustments made.

Results

Model One

In the first model, I studied the effect on car price from age, region of production, number of doors, horsepower, miles, original price, N.A.D.A. average retail value and N.A.D.A. average trade-in value. By going through systematically and eliminating the least significant variable, only three out of the original eight are significant. These are age, American-made, N.A.D.A. retail value and N.A.D.A. trade-in value. They explain over 94% of the prices, so these three variables are in fact good predictors of used car prices. The best model is given as:

$$\text{Current Selling Price} = b_1 * \text{American-Made} + b_2 * \text{Retail Price} + b_3 * \text{Trade-in Price}$$

$$\text{Current Selling Price} = 1052 * \text{American} + -0.212916 * \text{Retail Price} + 1.385827 * \text{Trade-in Price}$$

T-Statistics: 2.14 (American Made), -2.54 (Retail), 14.24 (Trade-in Price)

(*Note that American is a dummy variable with a value of 0 or 1)

In terms of the origin of the vehicle, cars built in America are significantly higher in price relative to foreign cars. The average American used car costs \$1052 higher than its foreign competitors at a comparable price level. This does not mean the American cars cost more in general, just that they cost more relative to the average trade-in and retail values. The remainder model ends up relying completely on the retail value and the trade-in values given by N.A.D.A.

The N.A.D.A. trade-in value is the average amount a dealer will give when a car is traded in, and the N.A.D.A. retail value is the average price that the dealer will sell the car for. According to the regression, private sellers price used cars between the N.A.D.A. trade-in value and the N.A.D.A. retail value. In more focus, sellers price cars on average at 12% above N.A.D.A. average trade-in value or at 93% of N.A.D.A. average retail value. The widespread use of the N.A.D.A. guide for car valuation explains this occurrence. The symbiotic nature between buyer and seller allows the seller to receive more than if he traded the car in and the buyer to pay less than if he bought it from a dealer. The N.A.D.A. values already account age, mileage, and many other things into vehicle's price, making many of the other variables redundant and thus insignificant. Although this model is conclusive in the fact that it yielded a useful tool, it does not tell anything about the individual pieces behind the N.A.D.A. values.

Model Two

In this model, I have excluded the N.A.D.A. variables, in order to find out the factors behind the prices. Although this model only explains 74% of the pricing, it will still work good enough in practical application to reach a “ballpark” figure. I will first examine the factors that work against a car’s value and then those which increase it. Using the same elimination method explained in model one, the resulting model is as follows:

$$\text{Current Selling Price} = b_1 * \text{Original Price} + b_2 * \text{Age} + b_3 * \text{Mileage} + b_4 * \text{Horsepower} + b_5 * \text{American} + b_6 * \text{Asian} + b_7 * \text{European}$$

$$\text{Current Selling Price} = 0.264 * \text{Original Price} + -348 * \text{Age} + -.15 * \text{Miles} + 51 * \text{Horsepower} + 4989 * \text{American} + 8544 * \text{Asian} + 10740 * \text{European}$$

T-Statistics: 3.52 (Org. Price), -2.13 (Age), -6.75 (Mileage), 3.22 (Horsepower), 1.93 (American), 3.62 (Asian), 3.99 (European)

(*Note again that American, Asian, and European are dummy variables 0 or 1)

Age and miles both diminish the value of a vehicle, which most people assume and know to be true anyways. According to the model, each year reduces a car’s value by \$348. I expected a more substantial effect from age, however mileage has a greater effect, causing a \$0.15 decrease in price per mile driven.

The basis for the value of a car is from its original cost. People who paid \$50000 for a luxury automobile should not be selling it for the same price as someone who only paid \$10000 for an economy vehicle, unless there is something severely wrong with the former. For that reason, I included the original price term in the model. Model Two puts an initial value at 26.4% of the original price of the car, and the other variables increase or decrease it, arriving at the current selling price.

Where a car manufacturer bases itself factors in as well at 90% confidence. American cars receive an additional \$4989, the lowest of the three measured regions. Cars produced in Asia are priced with an \$8544 markup, while those from Europe are the most expensive at \$10740 on top. The model prices European cars higher because they are simply more expensive to start out. In my sample, the average original price of a European automobile was \$43048 and its respective average selling price is \$24426. In comparison, American cars’ average original price was only \$22616 and its average selling price at \$11959.

In addition, people value power, more specifically about \$51 per unit of horsepower. In the US, horsepower is a well-marketed feature of a car, namely the more the better. Thus it is sensible that horsepower contributes positively to the value of a vehicle.

Model Two illustrates the individual components that comprise a car's price, which Model One did not. Given the right information, a person can use this tool to find a general price of what a car should cost without using the N.A.D.A. guide. I personally recommend using both resources to see comparability. As mentioned above, Model Two yields a good ballpark figure to work from before bargaining and negotiations begin.

Model Three

Unlike the first two models that study the car price, Model Three studies the factors of depreciation. I believe this to be useful because it looks at the percent depreciation, calculated by: [(price-original price)/original price]. The arbitrary pricing does not directly affect the model, because it measures percent change, not actual dollars. I have also included an American term, to look at depreciation between American and foreign vehicles. Model Three resulted as:

$$\% \text{Depreciation} = b_1 * \text{Age} + b_2 * \text{Miles} + b_3 * \text{Number of Doors} + b_4 * \text{American Made}$$

$$\% \text{Depreciation} = -0.029737 * \text{Age} + -0.003384 * \text{Miles} + -0.032732 * \text{Number of Doors} + -0.050475 * \text{American Made}$$

T-Statistics: -5.40 (Age), -2.01 (American), -4.62 (Doors), -6.42 (MilesK)

(*Note that American Made is a dummy variable with value of 0 or 1)

The resulting model yields some fascinating aspects. First, cars produced by American automakers are worth 5% less on average than their foreign brothers and sisters. There exists a common perception that American cars depreciate faster than those from Europe or Asia do. I cannot differentiate if American cars actually depreciate quicker, or just base themselves on private perceptions of depreciation. Because of the discrepancy in depreciation between foreign and American vehicles, I isolated them into their individual regions. I regressed an interaction term age*origin for America, Asia and Europe. This focus measures the percent depreciation per year for each region. It turns out, that cars from all regions depreciate at about 8% per year. The standard errors overlapped, thus the origin of the vehicle does not in fact have a true effect on depreciation.

Parallel to model two, miles contribute to depreciation as well. All else held constant, a car depreciates at 0.33% per thousand miles driven. Age and miles affect depreciation the most because they are variable, unlike American-made and number of doors, which are essentially dummy variables. The more miles a person racks up, or the older a car gets, forever decreases the worth of a car. The exception to this is cars when become collectors' items and appreciate over a long period of time.

Model Three, like the other two models, is useful in its own respect. Depreciation is another way to measure the worth of a vehicle. Whether it is for pricing purposes, tax adjustments or just asset valuation, this model is capable of measuring all three.

Forecast

Now that we have these models, we want to find out if they work. I have randomly selected a car from the classified advertisements to test the findings. I selected a 1996 Honda Civic DX listed at \$6500 and 62000 miles. The models predict the prices and depreciation as follows:

Model 1: \$5300

Model 2: \$5540

Model 3: 45% depreciation (Price of \$6100 based on \$11100 original cost)

Based on the models I have constructed, there are two ways to look at these results. Either all three models undervalue the car, or it is overpriced. In this instance, the owner has priced the car above the N.A.D.A retail value for such a vehicle, so rationally I believe it is indeed overpriced.

Conclusion & Notes

These three models are each useful in their own right, however they are only estimates. Keep in mind that I constructed these models from a sample of about 128 automobiles. Evidence may be stronger with a larger sample. In addition, auto dealerships and classified advertisements often overstate their prices. They do so in order to cushion themselves when bargaining with a buyer -- hence under-pricing these models is fair game. There will always be variation based on the condition of a car, a subjective variable that these models cannot account for. Other than these issues, the factors in each model make sense logically, similar to the thinking of someone who is actually selling a vehicle. The difference is that the models now associate a coefficient to each factor and can now be an excellent tool when car shopping.

Appendix

Regression output for Model One

Final Regression for Model One

Dependent Variable: PRICE
Method: Least Squares
Date: 11/12/02 Time: 13:29
Sample(adjusted): 36 128 IF USED=1 AND AGE<15
Included observations: 90 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AMERICAN	1052.113	490.0315	2.147032	0.0346
RETAIL	-0.212916	0.083560	-2.548078	0.0126
TRADE	1.385827	0.097304	14.24223	0.0000
R-squared	0.944742	Mean dependent var		16658.21
Adjusted R-squared	0.943471	S.D. dependent var		11020.14

Regression to find typical percent of N.A.D.A. average retail value

Dependent Variable: PRICE
Method: Least Squares
Date: 11/10/02 Time: 17:54
Sample(adjusted): 36 128 IF USED=1
Included observations: 93 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	766.4471	918.6952	0.834278	0.4063
RETAIL	0.936341	0.046671	20.06250	0.0000
R-squared	0.815604	Mean dependent var		16317.62
Adjusted R-squared	0.813578	S.D. dependent var		11013.95
S.E. of regression	4755.450	Akaike info criterion		19.79324
Sum squared resid	2.06E+09	Schwarz criterion		19.84771
Log likelihood	-918.3857	F-statistic		402.5041
Durbin-Watson stat	1.758800	Prob(F-statistic)		0.000000

Regression to find typical percent of N.A.D.A. average trade-in value

Dependent Variable: PRICE
Method: Least Squares
Date: 11/10/02 Time: 17:54
Sample(adjusted): 36 128 IF USED=1
Included observations: 93 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	659.1667	522.3901	1.261828	0.2102
TRADE	1.122188	0.031009	36.18915	0.0000
R-squared	0.935030	Mean dependent var		16317.62
Adjusted R-squared	0.934316	S.D. dependent var		11013.95
S.E. of regression	2822.744	Akaike info criterion		18.75008
Sum squared resid	7.25E+08	Schwarz criterion		18.80454
Log likelihood	-869.8786	F-statistic		1309.654
Durbin-Watson stat	1.704063	Prob(F-statistic)		0.000000

Regression Output for Model Two

Final Regression Output for Model Two

Dependent Variable: PRICE

Method: Least Squares

Date: 11/10/02 Time: 20:08

Sample(adjusted): 36 128 IF USED=1

Included observations: 93 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AGE	-348.4685	163.5850	-2.130199	0.0360
AMERICAN	4989.695	2576.017	1.936981	0.0560
ASIAN	8544.420	2356.734	3.625535	0.0005
EURO	10740.24	2690.373	3.992099	0.0001
HORSEPOWER	51.48463	15.95724	3.226412	0.0018
MILES	-0.150975	0.022338	-6.758579	0.0000
ORGRPRICE	0.264754	0.075171	3.522033	0.0007
R-squared	0.740596	Mean dependent var		16317.62
Adjusted R-squared	0.722498	S.D. dependent var		11013.95
S.E. of regression	5801.984	Akaike info criterion		20.24207
Sum squared resid	2.90E+09	Schwarz criterion		20.43270
Log likelihood	-934.2564	Durbin-Watson stat		1.573012

Regression Output for Model Three

Final Regression Output for Model Three

Dependent Variable: DEPPERCENT
Method: Least Squares
Date: 11/10/02 Time: 21:02
Sample(adjusted): 36 128 IF USED=1 AND AGE<15
Included observations: 90 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AGE	-0.029237	0.005414	-5.400003	0.0000
AMERICAN	-0.050475	0.025083	-2.012340	0.0473
DOORS	-0.032732	0.007071	-4.628976	0.0000
MILESK	-0.003384	0.000527	-6.423614	0.0000
R-squared	0.742897	Mean dependent var		-0.419284
Adjusted R-squared	0.733929	S.D. dependent var		0.224571
S.E. of regression	0.115838	Akaike info criterion		-1.429814
Sum squared resid	1.153995	Schwarz criterion		-1.318711
Log likelihood	68.34163	Durbin-Watson stat		1.674771

Regression with age and origin interaction terms to find per year depreciation for each individual region.

Dependent Variable: DEPPERCENT
Method: Least Squares
Date: 11/11/02 Time: 10:29
Sample(adjusted): 36 128 IF DEP>-40000 AND USED=1 AND AGE<15
Included observations: 86 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AGE*ASIAN	-0.081881	0.006704	-12.21461	0.0000
AGE*AMERICAN	-0.080942	0.005711	-14.17325	0.0000
AGE*EURO	-0.077949	0.007999	-9.744433	0.0000
R-squared	0.281508	Mean dependent var		-0.401805
Adjusted R-squared	0.264195	S.D. dependent var		0.213878
S.E. of regression	0.183462	Akaike info criterion		-0.519355
Sum squared resid	2.793648	Schwarz criterion		-0.433738
Log likelihood	25.33224	Durbin-Watson stat		1.285078

Forecast Information

1996 Honda Civic DX

Price: \$6500

Miles: 62000

2 Doors

106 Horsepower

Original Cost: \$11100

N.A.D.A. Average Trade-in Value: \$4800

N.A.D.A. Average Retail Value: \$6225

Made in Japan