Determinants of Sports Car Price

Zach Cohen
ECON125 Econometrics
Linus Yamane
02/24/2021

Introduction

This research seeks to assess what factors are significant in determining the price of a sports car; Does top speed matter more than 0-60? Are there preferences for different aspiration types? The data covers 62 different sports cars and all relevant explanatory variables for each.

Review of Previous Literature

Before beginning this project, I examined two external sources, each of which ran regressions on car price. *Linear Regression on Car Price Prediction* by Sinha et al. uses a regression to estimate the selling price of cars for a fictional car manufacturer. Sinha et al. use a similar set of explanatory variables to mine; however, they included more variables pertaining to the specific engine specifications and the dimensions of the car. Notably, the researchers restricted their regression to only significant variables, which included aspiration, engine location, width, weight, height, engine type, cylinder number, engine size, stroke, and peak rpm

I had a hard time finding linear regression models that addressed the price of new cars, specifically new performance cars such as mine. Regression analyses has been used more often with used cars. In *Building a Linear Regression Model for Predicting the Price of a Used Car,* Chang examines the prices of used cars and employs specification techniques which I will go on to use in my research. Chang takes the natural log of price to account for variability in skewed data, which I will also do. Chang's regression uses variables that will not be relevant to the sale of new cars, such as mileage, low mileage, city mpg, highway mpg, and years old. However, he does include a dummy variable for luxury cars similar to a variable I will include called LIMIT, which is a dummy variable for whether a car is limited in production or not.

Regression specification

LNPRICE = β_1 + β_2HP + β_3sSEAT + β_4LIMIT+ β_5TOP + β_6RWD + β_7YEAR + β_8ACC + β_9NA + β_{10}TT + β_{11}TTT+ β_{12}HYB + β_{13}SPEED + β_{14}WEIGHT + β_{15}rRATIO + β_{16}HP2 + u_i

The preferred specification uses the natural logarithmic form to account for the large spread of prices included in the data.

Dependent Variable:

Natural log of price (LNPRICE): is the price in us dollars ($) and recorded as the manufacturer suggested retail price (MSRP) from the most recent year the car was produced. The average price was $492,000, the minimum price was $59,895, and the maximum price was $3,500,000. Due to the large variance in prices, the natural log of price was used to smooth out the data.

Independent Variables:

Horsepower (HP): horsepower is the amount of power an engine produces. Generally speaking, a car with more power tends to be more expensive, so this variable should have a positive effect on MSRP

Horsepower ^{2} (HP2): HP2 accounts for the fact that horsepower has a nonlinear relationship with price. As the amount of horsepower increases, each extra horsepower has a gradually diminishing effect on price.

Number of seats (SEAT): for sports cars, whether or not the car has four seats as opposed to two is generally a good performance indicator. A four-seater car is usually intended for more practical use, whereas a two-seater car is usually intended for performance use. Therefore, the number of seats may play some role in affecting price.

Limited production (LIMIT): is a dummy variable that assesses whether or not the production of the car is capped at a certain number. 1 indicates that the car is limited, 0 means the car does not have a restricted production. Limited production scale tends to positively affect the MSRP of a car given that the supply is fewer than other mass-produced cars. A car being a limited model can have a large effect on its resell value, given that it becomes rare.

Convertible (TOP): is a dummy variable that measures whether or not a car is a convertible. 1 indicates the car is convertible, 0 indicates otherwise. Convertible models are generally more expensive than their non-convertible counterparts, given the additional materials and mechanics required to build the drop top mechanism. TOP should have a positive effect on price.

Rear-wheel drive (RWD): is a dummy variable that says whether a car is rear wheel drive or otherwise. Rear wheel drive sports cars are known by many to have better handling characteristics than all-wheel drive cars or 4 wheel drive cars. However, all-wheel drive cars generate more traction during acceleration, allowing them to accelerate faster. Both of these phenomena are likely to affect price.

Year (YEAR): The year the subject car was produced. All cars were documented in the most recent production year for this research.

0-60 acceleration (ACC): this is how fast, in seconds, a car can go from 0-60 miles an hour. 0-60 times are a benchmark statistic for all performance cars. A faster 0-60 time should generally have a positive effect on MSRP. However, some cars are not particularly fast on acceleration but have good handling characteristics and aerodynamic complexity, which makes them fast around a track. ACC will likely have a positive effect on price.

Aspiration type, natural aspiration, twin-turbo, turbo, supercharged, tri-turbo, turbo-hybrid (NA, TT, T, SUP, TTT, HYB): these dummy variables measure whether or not the car is naturally aspirated or has a forced induction system, as well as which kind of forced induction a car has. Aspiration is how a car moves air into the engine; an engine either gets air from an intake and feeds it to the cylinders. This process is assisted by a turbo or supercharger that pressurizes the air before it reaches the cylinders. Forced induction engines are more powerful than naturally aspirated engines of the same caliber. Naturally aspirated engines can still be powerful and are also known to sound better than forced induction engines which could affect price because people who buy sports cars tend to prefer loud engines.

Top speed (SPEED): measures the manufacturer's quoted top speed of a car in MPH. Speed is a desired trait and should therefore have a positive effect on price.

Weight (WEIGHT): measured in KG, weight is the curb weight of a car. Heavier cars are generally slower therefore, weight must have a negative effect on price.

Power to weight ratio (RATIO): measures a car's power to weight ratio. Horsepower matters, and weight matters, but they also matter relative to each other. A car could be relatively low horsepower but still be very fast if its light and vice versa.

Data Description

```
. summm
```

Variable	Obs	Mean	Std. dev.	Min	Max
price	61	492000.7	741149.9	59895	3500000
hp	61	667.3279	185.6631	405	1603
seat	61	2.590164	.9197291	2	4
limit	61	.3114754	.4669398	0	1
top	61	.2622951	.4435328	0	1
rwd	61	.6393443	.4841758	0	1
year	61	2019.754	2.668681	2010	2022
acc	61	3.022951	.4356579	2.2	4.2
na	61	.295082	.4598646	0	1
tt	61	.6393443	.4841758	0	1
t	61	.0163934	.1280369	0	1
sup	61	0	0	0	0
ttt	61	.0327869	.1795562	0	1
hyb	61	.0491803	.2180389	0	1
speed	61	202.5738	25.64726	130	300
weight	61	1581.361	187.6855	1198	2032
ratio	61	.4322951	.1563905	.22	1.21
hp2	61	477146.9	338498.1	164025	2569609
lnacc	61	1.096542	.1387964	.7884573	1.435085
lnprice	61	12.55178	.9365382	11.00035	15.06827

Fig 1. Summary statistics for all explanatory x variables showing the mean, standard deviation, minimum and maximum value

The data for each car was gathered primarily from Car and Driver reports and using a variety of internet tools that report the specifications of sports cars. The curb weight for each car was used, which is the car's weight at the curb with fuel. Additionally, the manufacturer quoted top speed was used, as well as for 0-60 times. There is room for discrepancies since some third parties may test a car and get different numbers.

One thing to note is that the difference between the minimum and maximum values for some variables is quite significant. A few cars were included that fall into the "hyper-car" category, accounting for this variance. Hyper-cars are cars that exceed the 1000 horsepower mark, and their prices usually fall somewhere around or above $1,000,000. A possible issue with the data could be that only several hyperarcs are included while the rest of the cars fall below that threshold.

Results

lnprice	Coefficient	Robust std. err.
hp	.0093541	.0019673
seat	.0689429	.0615635
limit	.2722681	.106141
top	.1233727	.0805018
rwd	.0057979	.095875
year	.0115939	.0114687
acc	.6920651	.2706087
na	-.0725237	.3242173
tt	-.3402328	.3044439
ttt	-.5130256	.3145513
hyb	.4353462	.2027381
speed	.0162779	.004018
weight	-.0011392	.000453
hp2	-3.87e-06	9.01e-07
_cons	-18.90648	24.02959

Number of obs	=	61
F(14, 46)	=	70.71
Prob > F	=	0.0000
R-squared	=	0.9337
Root MSE	=	.27531

Fig 2. Regression output with all explanatory variables included (sup and turb omitted due to collinearity problem) Statistically significant:

The regression output on all of the explanatory variables with SUP and TURB omitted due to collinearity issues has confirmed that some of the chosen explanatory variables are significant. RATIO, which measures the power to weight ratio, was omitted because it measures the same effect that HP and WEIGHT measure independently. Several variables returned with statistical significance; HP (4.75), LIMIT(2.57), TOP(1.53), ACC(2.56), HYB(2.15), SPEED(4.05), WEIGHT(-2.51), and HP2(-4.29). The general theory surrounding sports cars would agree with this assessment that horsepower, limited production, convertible, acceleration, hybrid power, top speed, and weight are the key determinants of price.

Horsepower is one of the first things a buyer might look at when assessing a sports car because it will determine how fast the car will be and how it will feel on the gas. Therefore, it makes sense that horsepower positively affects price and that each additional horsepower contributes to a 0.93% increase in price. It also makes sense that there is a negative relationship between horsepower2 and price. Above a certain threshold, each extra horsepower may become increasingly less desirable. Traction could be one of the main limitations here; the tires can only withstand a certain amount of power before they lose all grip, so having a car with too much power may be less desirable.

After horsepower, 0-60 acceleration can be one of the defining selling characteristics of a sports car. Given that these cars are designed to be driven on the road, most of the performance will be felt on acceleration to 60mph. Consequentially ACC has a positive relationship with price; each extra second has a 69% increase in price.

I was surprised to see that top speed came back as significant because I believed it to be a more superficial performance measure. Unlike 0-60 acceleration, which can be performed regularly and safely on the road, drivers will very rarely be able to achieve a car's actual top speed. Nonetheless, each additional mph of top speed causes a 1% increase in price. A high top speed may be representative of other desirable characteristics, such as high horsepower or a light car, or a combination of the two, and it thus has a positive effect on price.

Limited production means that there will be less of a car on the market, and the production will be capped at a small number, whereas cars usually are mass-produced. A smaller supply means that the retailer will have to charge more for the cars and be able to do so because they will likely appreciate after the sale due to their limited nature. When a car is limited in production, the price will be 27% higher than when a car is not limited.

When a car is convertible, it takes more materials to assemble, meaning that the manufacturer will take on more costs in production. There might also be some added desire and allure associated with the idea of driving a convertible. The extra costs will be reflected in the MSRP as convertibles cost 12% more than hardtops.

Amongst the dummy variables for engine aspiration, hybrid power was the only one that came back as significant. This makes sense because hybrid power is very different from the other variables, such as turbocharged or supercharged. Hybrid power is the most technologically demanding form of forced induction and is a newly developed green engine technology that can be used to lower emissions and boost power. According to the regression, sports cars with hybrid assisted engines will cost 43% more. There may be some bias at play here as well; the HYB numbers may be inflated because only 3 of 61 cars in the sample were hybrids and those three cars are all hypercars.

Discussion

Variable	*ß**
LIMIT	0.1357
TOP	0.0583
ACC	0.3277
HYB	0.1013
SPEED	0.4436
WEIGHT	0.2204
HP + HP2	0.4512

Fig 3. ß value table*

To test and compare each variables economic significance I used ß* = $\left|\frac{bOLS * SD_x}{SD_y}\right|$. From this I found that the effect of horsepower, measured with HP and HP2, has the most influence on the variation of price (45%). I also found that top speed (SPEED) explains 44% of the variation in price. Additionally, 0-60 acceleration (ACC) explains 32% of the variation in price.

These ß* results support the existing knowledge that horsepower, speed and acceleration are the three key performance markers of a sportscar.

```
. estat vif

    Variable |       VIF       1/VIF
-------------+----------------------
          na |     11.20    0.089254
          tt |     10.59    0.094443
       speed |      6.46    0.154882
          hp |      5.47    0.182898
        seat |      3.65    0.274280
         acc |      2.94    0.340446
      weight |      2.74    0.364937
         ttt |      2.61    0.383279
         rwd |      1.97    0.506854
       limit |      1.94    0.516308
         hyb |      1.94    0.516525
        year |      1.56    0.641134
         top |      1.10    0.906031
-------------+----------------------
    Mean VIF |      4.17
```

Fig 4. estat vif results.

The *estat vif* or a variance inflation factor test examines the presence of multicollinearity issues. The test results indicate that we likely do not have a collinearity issue. A mean VIF below 10 is generally regarded as acceptable, and such is the case here (4.17). Additionally, multiple variables from the regression model came back with significance, so it is likely that we do not have a collinearity issue. NA(11.20) and TT(10.59) may have returned as insignificant due to a collinearity problem.

```
. estat ovtest

Ramsey RESET test for omitted variables
Omitted: Powers of fitted values of lnprice

H0: Model has no omitted variables

F(3, 44) =  16.88
Prob > F = 0.0000
```

In figure 3 we see that the ovtest has returned a p-value that is smaller than 0.05 (0.0000), meaning that we will move forward to reject the null hypothesis that the model has no omitted variables. This means we might have a problem; if we have omitted a relevant explanatory variable, the model will be meaningless, and the coefficients will be biased. Ive experimented with adding various other functional forms but have not been able to pass the test.

Fig 5. estat ovtest is a regression specification error test which asses whether or not there is an omitted variable. P < .05 rejects H0: Model has no omitted variables. P > 0.05 can follow H0: Model has no omitted variables.

Correlation Analysis

The dependent variable here is LNPRICE, it should be correlated with all of the explanatory variables, but the explanatory vairables should not be correlated with each other. There are some red flags in this sense: HP2 and HP seem to be correlated with each other (.9689), and HP2 and ratio seem to be correlated with each other (.9296). Both correlations make sense; HP2 could be correlated with hp because they measure the same thing. Additionally, horsepower is included in RATIO (hp/weight), so it makes sense that HP2 and ratio are correlated. On account of these correlation issues, a new regression specification was generated, which omitted RATIO.

Heteroskedasticity

```
. predict yhat
(option xb assumed; fitted values)

. predict res, residual

. gen res2 = res*res

. twoway scatter res yhat
```

In order to examine our results to see if we have a heteroskedasticity issue, we must first define some variables. We want to plot two variables, the residuals (*res*) and the fitted y values (*yhat*). To do so, we must tell Stata that we want to predict yhat as well as predict the residuals. We also want to be able to look at our residuals2, so we must define *res2.* We can then go on to graph *res* against *yhat*.

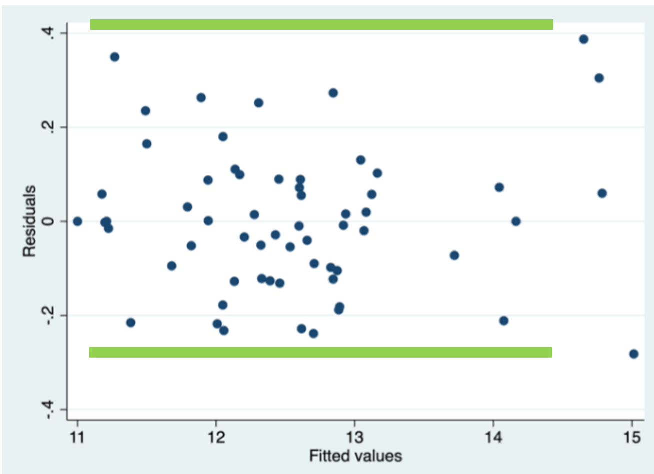

Fig 6. Plot of the residuals on the fitted values

From figure 6., a plot of the residuals against *yhat*, we can conclude that there is homoskedasticity which means that there is seemingly no correlation between the fitted values and the residuals. It seems that the residuals are scattered randomly, and there is no systematic change in the spread over time. This trend has been illustrated by two straight green lines between which all of the data are somewhat evenly spread. This indicates that we do not have heteroskedasticity, and our OLS estimates will be unbiased, consistent, and efficient.

```
. estat hettest

Breusch–Pagan/Cook–Weisberg test for heteroskedasticity
Assumption: Normal error terms
Variable: Fitted values of lnprice

H0: Constant variance

    chi2(1) =  17.46
Prob > chi2 = 0.0000
```

Fig 7. Breusch-Pagan/Cook-Weisberg test for heteroskedasticity

Even though the graph of the fitted values on the residual's visual looks homoskedastic, a Breusch-Pagan/Cook-Weisberg test revealed that there is some correlation of the error terms indicating a heteroskedasticity problem. Because the chi2 statistic is less than 0.05 we reject the null hypothesis that there is constant variance. To account for this problem, we run a VCE Robust standard error regression which will be valid even with the presence of a heteroskedasticity issue.

```
. imtest, white

White's test
H0: Homoskedasticity
Ha: Unrestricted heteroskedasticity

   chi2(60) =  61.00
Prob > chi2 = 0.4397

Cameron & Trivedi's decomposition of IM-test
```

Source	chi2	df	p
Heteroskedasticity	61.00	60	0.4397
Skewness	30.66	14	0.0062
Kurtosis	2.19	1	0.1391
Total	93.84	75	0.0695

While the Breusch-Pagan/Cook-Weisberg test revealed that we might have a heteroskedasticity issue, the White's test results say otherwise. Under this test, we cannot reject the null hypothesis that there is homoskedasticity, and therefore according to this test, we have no issue. However, we must keep in mind that the Breusch-Pagan test revealed a heteroskedasticity issue. For this reason, we still decide to run a VCE Robust regression to account for the possible presence of heteroskedasticity. After running the VCE Robust regression, several variables come in with significance.

Fig 8. Plot of the residuals on the fitted values

Reverse Causality?

Reverse causality is another issue that may be present in this data; we must ask whether or not the dependent variable price has any influence on the independent variables. To some extent, yes, cars are not just cars to most manufacturers; instead, cars can be made to achieve some financial goal. The car is a mode by which a manufacturer can meet financial goals, so it is possible that before a car is developed, there is a specific price that will meet a projected sales threshold that a manufacturer is aiming for. Thus, the price may be determined before the actual make-up of the car, meaning that it would be impossible for the independent variables to affect the price, given that the price was already determined.

Conclusions

I will take tentative conclusions from the results, given that we cannot fully trust the regression due to the failure of the Ramsey RESET test for omitted variables. I found that the effects of horsepower, top speed, and 0-60 acceleration are the three key determinants of price based on their beta star coefficients. Consequentially, when buying a car, buyers should consider these three factors as well as whether or not the car is convertible, the weight, and whether or not the car has a hybrid assisted motor.

Upon examining the residuals of price, seen at the end of the appendix, some buyers seem to have misunderstood the relationship between the explanatory variables and price and have subsequently overpaid. For example, it seems as if the Pagani Huyra, which costs $3,500,000, is overvalued by 30%. Additionally, the BMW M5 Competition seems to be overvalued by 34%. The implications of this regression have cost-saving potential for consumers of higher-end sports cars who may be suffering from a lack of information about the valuation of cars based on their specifications.

Bibliography

New car reviews, buying advice and news. Car and Driver. (n.d.). Retrieved April 6, 2022, from https://www.caranddriver.com/

Sinha, Sayan, Azim, Riazul, Das, Sourav. *Linear Regression on Car Price Prediction*. University of Kalyani. Published May 4, 2020, from file:///Users/zachcohen/Downloads/car-price-prediction.pdf

Chang, Hamilton. *Building a Linear Regression Model for Predicting the Price of a Used Car.* Towards Data Science. Published March 6, 2020, from https://towardsdatascience.com/building-a-linear-regression-model-for-predicting-the-price-of-a-used-car-now-with-logs-54a478438d1.

Appendix

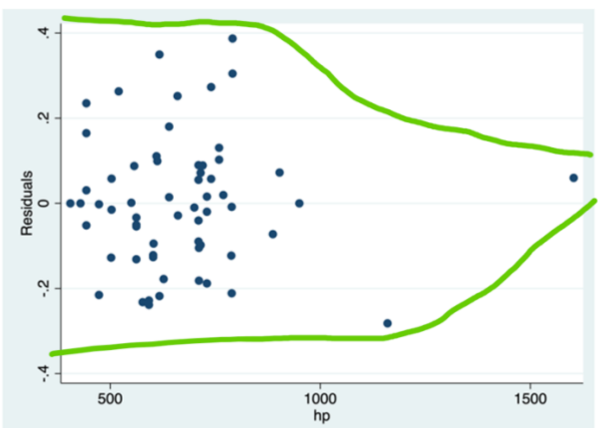

Fig 6. Scatter plot of the residuals on horsepower (hp) showing a possible heteroskedasticity problem

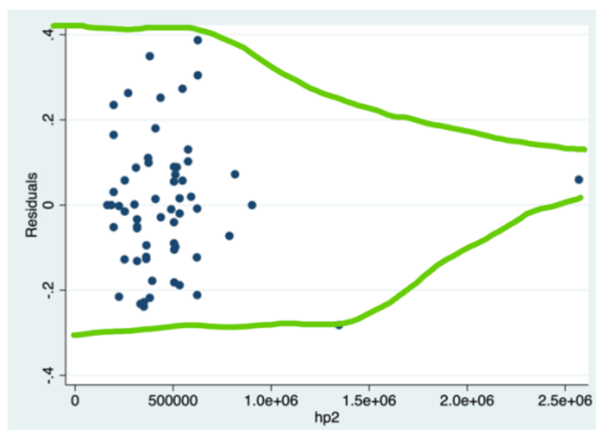

Fig 7. Scatter plot of the residuals on horsepower^2 (hp2) showing a possible heteroskedasticity problem

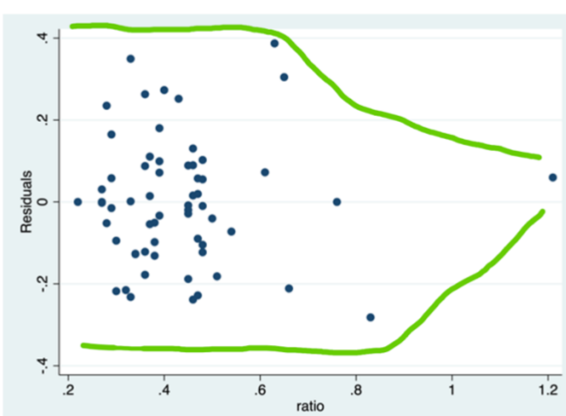

Fig 8. Scatter plot of the residuals on the power to weight ratio (ratio) showing a possible heteroskedasticity problem

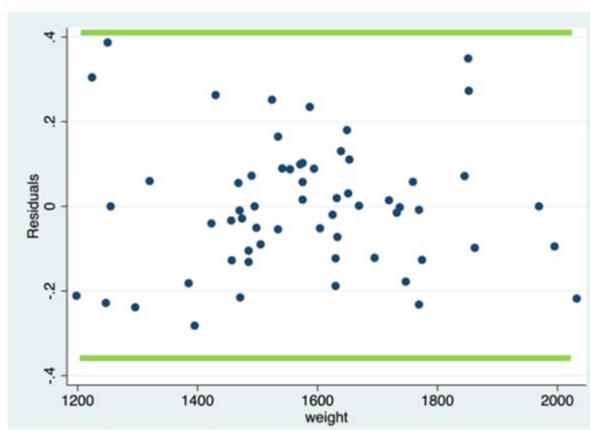

Fig 9. Scatter plot of the residuals on weight showing homoskedasticity)

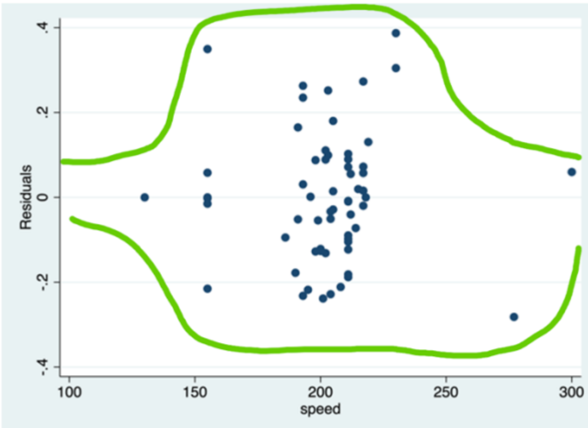

Fig 10. Scatter plot of the residuals on speed showing a possible heteroskedasticity issue.

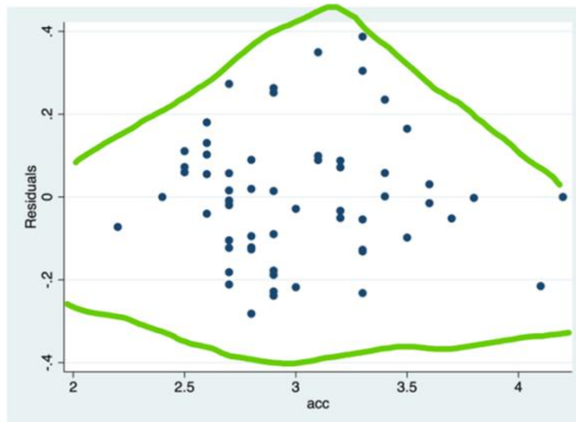

Fig 11. Scatter plot of the residuals on 0-60 acceleration (acc) showing a possible heteroskedasticity issue

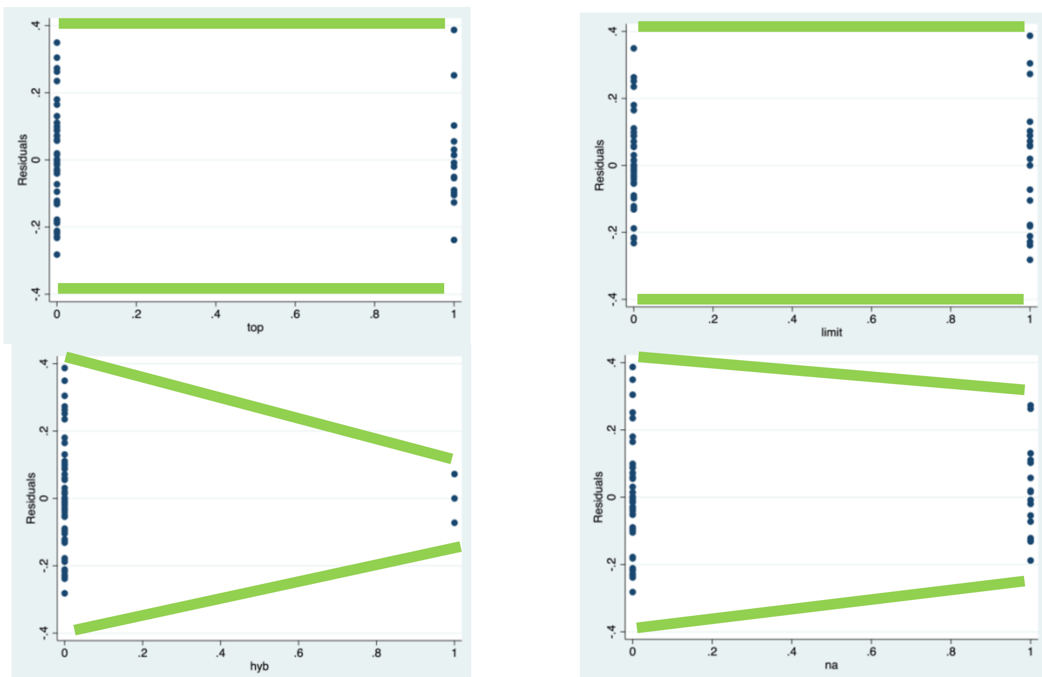

Fig 12. Residual graph for all dummy variables or variables with two distinct category's, hybrid (hyb), convertible (top), limited production (limit), and naturally aspirated (na). Top and limit seem to have homoskedasticity but na and hyb seem to have a heteroskedasticity issue.

Correlation Matrix

```
. correlate lnprice hp hp2 ratio weight seat limit top rwd year acc na tt t ttt hyb speed
(obs=61)
```

	lnprice	hp	hp2	ratio	weight	seat	limit	top	rwd	year	acc	na	tt	t	ttt	hyb	speed
lnprice	1.0000																
hp	0.8297	1.0000															
hp2	0.7333	0.9689	1.0000														
ratio	0.8677	0.9436	0.9296	1.0000													
weight	-0.5292	-0.3122	-0.3057	-0.5895	1.0000												
seat	-0.5540	-0.4115	-0.3235	-0.4948	0.5887	1.0000											
limit	0.6758	0.5604	0.4921	0.6125	-0.4286	-0.3575	1.0000										
top	0.0861	-0.0523	-0.0783	-0.0493	-0.0326	-0.0589	-0.0792	1.0000									
rwd	-0.1465	-0.2779	-0.2722	-0.1431	-0.2949	-0.1877	-0.2320	-0.0178	1.0000								
year	-0.2236	-0.1149	-0.0644	-0.1216	0.1485	0.3046	-0.1916	0.0272	-0.0698	1.0000							
acc	-0.5760	-0.6313	-0.4962	-0.5492	0.1870	0.5147	-0.4208	-0.0058	0.3718	0.1727	1.0000						
na	0.1382	0.0721	0.0099	-0.0372	0.1349	-0.4186	0.0305	0.0228	-0.1877	-0.3609	-0.3505	1.0000					
tt	-0.1433	-0.0925	-0.0276	-0.0041	-0.1179	0.4111	-0.0109	-0.0178	0.1469	0.3430	0.3164	-0.8614	1.0000				
t	-0.1864	-0.1671	-0.1139	-0.1775	0.2689	0.1995	-0.0868	-0.0770	0.0970	0.0608	0.3517	-0.0835	-0.1719	1.0000			
ttt	0.0057	0.0427	0.0137	0.0383	-0.0577	-0.1191	-0.1238	0.0995	0.1383	0.0867	-0.0737	-0.1191	-0.2451	-0.0238	1.0000		
hyb	0.3486	0.3038	0.2418	0.2997	-0.1491	-0.1471	0.3381	-0.1356	-0.1449	-0.4085	-0.3454	0.0191	-0.1449	-0.0294	-0.0419	1.0000	
speed	0.8416	0.8444	0.7963	0.8350	-0.4450	-0.5332	0.4775	0.1008	-0.2367	-0.1204	-0.6489	0.1522	-0.0904	-0.3683	0.0610	0.1230	1.0000

correlation matrix between the dependent variable (lnprice) and all of the explanatory x variables. Correlation coefficients close to 1 indicate that the two relevant variables are corelated

Beta Star Values

Variable	*ß**
LIMIT	0.1357
TOP	0.0583
ACC	0.3277
HYB	0.1013
SPEED	0.4436
WEIGHT	0.2204
HP + HP2	0.4512

Beta Star Calculations Table

$$\text{ß* HP} = \left|\frac{.00935 * 185.66}{.9365}\right| = \mathbf{1.85}$$

$$\text{ß* HP2} = \left|\frac{-3.87e{-}06 * 338498}{.9365}\right| = \mathbf{1.3988}$$

$$\text{ß* HP + HP2}\ \frac{-3.87e{-}06 * 338498}{.9365} + \frac{.00935 * 185.66}{.9365} = \mathbf{0.452}$$

$$\text{ß* LIMIT} = \left|\frac{.2722 * .4669}{.9365}\right| = \mathbf{0.1357}$$

$$\text{ß* TOP} = \left|\frac{.1233 * .4435}{.9365}\right| = \mathbf{0.0583}$$

$$\text{ß* ACC} = \left|\frac{.6920 * .4356}{.9365}\right| = \mathbf{0.3277}$$

$$\text{ß* HYB} = \left|\frac{.4353 * .2180}{.9365}\right| = \mathbf{0.1013}$$

$$\text{ß* SPEED} = \left|\frac{.0162 * 25.647}{.9365}\right| = \mathbf{0.4436}$$

$$\text{ß* WEIGHT} = \left|\frac{-.0011 * 187.68}{.9365}\right| = \mathbf{0.2204}$$

ß calculations for all statistically significant explanatory variables*

Incorrect Original Regression Specification

```
. reg lnprice hp seat limit top rwd year acc na tt ttt hyb speed weight ratio hp2
```

Source	SS	df	MS			
				Number of obs	=	61
				F(15, 45)	=	75.51
Model	50.6151943	15	3.37434629	Prob > F	=	0.0000
Residual	2.01103706	45	.044689712	R-squared	=	0.9618
				Adj R-squared	=	0.9490
Total	52.6262314	60	.877103856	Root MSE	=	.2114

lnprice	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
hp	.0011149	.0017087	0.65	0.517	-.0023266	.0045564
seat	.0159975	.0576013	0.28	0.782	-.1000176	.1320125
limit	.2274228	.0865236	2.63	0.012	.0531554	.4016902
top	.14287	.0653149	2.19	0.034	.0113191	.2744209
rwd	.0178968	.081687	0.22	0.828	-.1466292	.1824228
year	.0122074	.012786	0.95	0.345	-.013545	.0379598
acc	.6897965	.1248022	5.53	0.000	.4384319	.941161
na	.7333222	.2519003	2.91	0.006	.2259689	1.240675
tt	.3319499	.2259341	1.47	0.149	-.1231049	.7870046
ttt	.2896781	.2902631	1.00	0.324	-.2949418	.874298
hyb	.6877422	.1791332	3.84	0.000	.3269493	1.048535
speed	.0175169	.0027097	6.46	0.000	.0120594	.0229745
weight	.0027054	.0007217	3.75	0.001	.0012517	.0041591
ratio	13.66987	2.443786	5.59	0.000	8.747829	18.5919
hp2	-4.81e-06	4.86e-07	-9.88	0.000	-5.79e-06	-3.83e-06
_cons	-27.00904	25.97477	-1.04	0.304	-79.32491	25.30682

Fig 10. Original specification which includes RATIO and HP2

lnPrice = -27.009 + 0.0011 hp + 0.0159 seat + 0.2274 limt + 0.1428 top + 0.0178 rwd
(25.97) (.0017) (.0576) (.0865) (.0653) (.0816)

+ 0.0122 year + 0.6897 acc + 0.7333 na + 0.3319 tt + 0.2896 ttt + 0.6877 hyb + 0.0175 speed
(.0122) (.1248) (.2519) (.2259) (.2902) (.1791) (.0027)

+ 0.0027 weight + 13.669 ratio + -4.81e-06 hp2
(.0007) (2.4437) (4.86e-07)

Correct Regression Specification

```
. reg lnprice hp seat limit top rwd year acc na tt ttt hyb speed weight hp2, vce(robust)
```

```
Linear regression                               Number of obs     =         61
                                                F(14, 46)         =      70.71
                                                Prob > F          =     0.0000
                                                R-squared         =     0.9337
                                                Root MSE          =     .27531
```

lnprice	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
hp	.0093541	.0019673	4.75	0.000	.0053942	.013314
seat	.0689429	.0615635	1.12	0.269	-.0549781	.1928638
limit	.2722681	.106141	2.57	0.014	.0586174	.4859188
top	.1233727	.0805018	1.53	0.132	-.038669	.2854144
rwd	.0057979	.095875	0.06	0.952	-.1871885	.1987844
year	.0115939	.0114687	1.01	0.317	-.0114914	.0346792
acc	.6920651	.2706087	2.56	0.014	.147358	1.236772
na	-.0725237	.3242173	-0.22	0.824	-.7251393	.5800919
tt	-.3402328	.3044439	-1.12	0.270	-.9530466	.272581
ttt	-.5130256	.3145513	-1.63	0.110	-1.146185	.1201333
hyb	.4353462	.2027381	2.15	0.037	.0272556	.8434369
speed	.0162779	.004018	4.05	0.000	.0081901	.0243657
weight	-.0011392	.000453	-2.51	0.015	-.002051	-.0002274
hp2	-3.87e-06	9.01e-07	-4.29	0.000	-5.68e-06	-2.06e-06
_cons	-18.90648	24.02959	-0.79	0.435	-67.27554	29.46257

HP(4.75) LIMIT(2.57) TOP(1.53) ACC(2.56) HYB(2.15) SPEED(4.05) WEIGHT(-2.51) HP2(-4.29)

Data

	price	hp	seat	limit	top	rwd	year	acc	na	tt	t	sup	ttt	hyb	speed	weight	ratio	hp2	lnacc	lnprice
1	59895	405	2	0	0	1	2021	4.2	0	1	0	0	0	0	155	1495	.27	164025	1.43508	11.0003
2	73900	429	4	0	0	1	2021	4.2	0	0	1	0	0	0	130	1969	.22	184041	1.43508	11.2105
3	70895	473	4	0	0	1	2021	4.1	0	1	0	0	0	0	155	1471	.32	223729	1.41099	11.169
4	72795	473	4	0	0	1	2021	3.8	0	1	0	0	0	0	155	1737	.27	223729	1.335	11.1954
5	129250	443	4	0	1	1	2021	3.7	0	1	0	0	0	0	191	1604	.28	196249	1.30833	11.7695
6	73795	503	4	0	0	1	2021	3.6	0	1	0	0	0	0	155	1732	.29	253009	1.28093	11.209
7	136550	443	4	0	1	0	2021	3.6	0	1	0	0	0	0	193	1651	.27	196249	1.28093	11.8244
8	337525	715	4	0	1	1	2021	3.5	0	1	0	0	0	0	211	1862	.38	511225	1.25276	12.7294
9	116450	443	4	0	0	1	2021	3.5	0	1	0	0	0	0	191	1534	.29	196249	1.25276	11.6652
10	75695	503	4	0	0	1	2021	3.4	0	1	0	0	0	0	155	1759	.29	253009	1.22378	11.2345
11	154000	550	2	0	0	1	2021	3.4	0	1	0	0	0	0	196	1669	.33	302500	1.22378	11.9447
12	123750	443	4	0	0	0	2021	3.4	0	1	0	0	0	0	193	1587	.28	196249	1.22378	11.726
13	226000	562	2	0	0	1	2010	3.3	1	0	0	0	0	0	202	1485	.38	315844	1.19392	12.3283
14	263000	562	2	0	1	1	2010	3.3	1	0	0	0	0	0	199	1534	.37	315844	1.19392	12.4799
15	136500	577	4	0	0	0	2019	3.3	0	1	0	0	0	0	193	1769	.33	332929	1.19392	11.8241
16	3.5e+06	791	2	1	0	1	2021	3.3	0	1	0	0	0	0	230	1224	.65	625681	1.19392	15.0683
17	3.4e+06	791	2	1	1	1	2020	3.3	0	1	0	0	0	0	230	1250	.63	625681	1.19392	15.0393
18	163500	502	2	0	0	1	2022	3.3	1	0	0	0	0	0	198	1457	.34	252004	1.19392	12.0046
19	319125	715	4	0	0	1	2021	3.2	0	1	0	0	0	0	211	1845	.39	511225	1.16315	12.6733
20	193000	562	2	0	0	1	2021	3.2	0	1	0	0	0	0	204	1456	.39	315844	1.16315	12.1704
21	213800	562	2	0	1	1	2022	3.2	0	1	0	0	0	0	204	1498	.38	315844	1.16315	12.2728
22	167650	557	2	0	0	1	2021	3.2	0	1	0	0	0	0	198	1554	.36	310249	1.16315	12.0296
23	111095	617	4	0	0	0	2021	3.1	0	1	0	0	0	0	155	1851	.33	380689	1.1314	11.6181
24	213195	612	2	0	0	1	2021	3.1	0	1	0	0	0	0	203	1571	.39	374544	1.1314	12.27
25	327000	720	2	1	0	1	2021	3.1	0	1	0	0	0	0	202	1594	.45	518400	1.1314	12.6977
26	131995	617	4	0	0	0	2022	3	0	1	0	0	0	0	195	2032	.3	380689	1.09861	11.7905
27	242737	661	2	0	0	1	2016	3	0	1	0	0	0	0	205	1474	.45	436921	1.09861	12.3997
28	142995	627	4	1	0	0	2022	2.9	0	1	0	0	0	0	190	1747	.36	393129	1.06471	11.8706
29	284700	660	2	0	1	1	2016	2.9	0	1	0	0	0	0	203	1524	.43	435600	1.06471	12.5592
30	326745	730	2	0	0	1	2016	2.9	1	0	0	0	0	0	211	1630	.45	532900	1.06471	12.6969
31	302000	710	2	0	1	1	2021	2.9	0	0	0	0	1	0	211	1505	.47	504100	1.06471	12.6182
32	240000	592	2	1	0	1	2019	2.9	0	1	0	0	0	0	204	1247	.47	350464	1.06471	12.3884
33	259000	592	2	1	1	1	2020	2.9	0	1	0	0	0	0	201	1296	.46	350464	1.06471	12.4646
34	217650	640	4	0	1	0	2021	2.9	0	1	0	0	0	0	205	1719	.37	409600	1.06471	12.2906
35	190050	520	2	0	0	1	2019	2.9	1	0	0	0	0	0	193	1430	.36	270400	1.06471	12.155
36	200295	602	2	0	0	0	2021	2.8	1	0	0	0	0	0	200	1695	.36	362404	1.02962	12.2075
37	211695	602	2	0	1	0	2021	2.8	1	0	0	0	0	0	200	1774	.34	362404	1.02962	12.2629
38	490000	769	2	1	0	1	2016	2.8	1	0	0	0	0	0	215	1632	.47	591361	1.02962	13.1022
39	280000	710	2	0	0	1	2021	2.8	0	0	0	0	1	0	211	1541	.46	504100	1.02962	12.5425
40	2.5e+06	1160	2	1	0	0	2021	2.8	0	1	0	0	0	0	277	1395	.83	1.3e+06	1.02962	14.7318
41	107500	603	4	0	0	0	2021	2.8	0	1	0	0	0	0	186	1995	.3	363609	1.02962	11.5852
42	331000	711	2	1	0	1	2020	2.7	0	1	0	0	0	0	211	1385	.51	505521	.993252	12.7099
43	351300	711	2	1	1	0	2020	2.7	0	1	0	0	0	0	211	1485	.48	505521	.993252	12.7694
44	404494	789	2	0	1	1	2021	2.7	1	0	0	0	0	0	211	1769	.45	622521	.993252	12.9104
45	335000	788	2	0	0	1	2021	2.7	1	0	0	0	0	0	211	1630	.48	620944	.993252	12.7219
46	421321	730	2	0	0	0	2021	2.7	1	0	0	0	0	0	217	1575	.46	532900	.993252	12.9511
47	463917	730	2	0	1	0	2021	2.7	1	0	0	0	0	0	217	1625	.45	532900	.993252	13.0475
48	497895	740	2	1	0	0	2016	2.7	1	0	0	0	0	0	217	1852	.4	547600	.993252	13.1181
49	530000	740	2	1	0	0	2016	2.7	1	0	0	0	0	0	217	1575	.47	547600	.993252	13.1806
50	1.1e+06	789	2	1	0	1	2021	2.7	0	1	0	0	0	0	208	1198	.66	622521	.993252	13.8643
51	293200	700	2	0	0	1	2021	2.7	0	1	0	0	0	0	211	1470	.48	490000	.993252	12.5886
52	526443	759	2	1	0	0	2021	2.6	1	0	0	0	0	0	219	1639	.46	576081	.955511	13.1739
53	577461	759	2	1	1	0	2021	2.6	1	0	0	0	0	0	211	1575	.48	576081	.955511	13.2664
54	301500	710	2	0	0	1	2021	2.6	0	1	0	0	0	0	212	1423	.5	504100	.955511	12.6165
55	318000	710	2	0	1	1	2021	2.6	0	1	0	0	0	0	212	1468	.48	504100	.955511	12.6698
56	204850	640	4	0	0	0	2021	2.6	0	1	0	0	0	0	205	1649	.39	409600	.955511	12.23
57	2.8e+06	1603	2	1	0	0	2021	2.5	0	1	0	0	0	0	300	1320	1.21	2.6e+06	.916291	14.8451
58	208571	610	2	0	0	1	2021	2.5	1	0	0	0	0	0	202	1653	.37	372100	.916291	12.248
59	1.4e+06	903	2	1	0	1	2014	2.5	0	1	0	0	0	1	217	1490	.61	815409	.916291	14.1156
60	1.4e+06	950	2	1	0	0	2016	2.4	0	0	0	0	0	1	218	1255	.76	902500	.875469	14.1636
61	845000	887	2	1	0	0	2015	2.2	1	0	0	0	0	1	214	1633	.54	786769	.788457	13.6471

Residuals on Price

```
. list res price
```

	res	price
1.	-3.55e-15	59895
2.	3.55e-15	73900
3.	-.2152535	70895
4.	-.0021734	72795
5.	-.0518113	129250
6.	-.0149158	73795
7.	.0306747	136550
8.	-.0979665	337525
9.	.1648396	116450
10.	.0579905	75695
11.	.0013754	154000
12.	.2351022	123750
13.	-.1313546	226000
14.	-.0542369	263000
15.	-.2320411	136500
16.	.3047358	3.5e+06
17.	.3871556	3.4e+06
18.	-.1275015	163500
19.	.0716125	319125
20.	-.0333812	193000
21.	-.0506055	213800
22.	.0877368	167650
23.	.3495924	111095
24.	.0993321	213195
25.	.0890742	327000
26.	-.2178323	131995
27.	-.0285849	242737
28.	-.1778311	142995
29.	.2519877	284700
30.	-.1880538	326745
31.	-.0897081	302000
32.	-.2282003	240000
33.	-.2383931	259000
34.	.0144024	217650
35.	.2630681	190050
36.	-.1216444	200295
37.	-.1264904	211695
38.	.0195349	490000
39.	.0897081	280000
40.	-.2818542	2.5e+06
41.	-.094545	107500
42.	-.1816231	331000
43.	-.1046361	351300
44.	-.0083454	404494
45.	-.1228507	335000
46.	.0157432	421321
47.	-.0197506	463917
48.	.2731367	497895
49.	.0574254	530000
50.	-.2112669	1.1e+06
51.	-.0098518	293200
52.	.1304756	526443
53.	.1025057	577461
54.	-.0403289	301500
55.	.0552179	318000
56.	.1801556	204850
57.	.0597609	2.8e+06
58.	.1106883	208571
59.	.0723497	1.4e+06
60.	-3.55e-15	1.4e+06
61.	-.0723497	845000