

Prices of Hybrid Vehicles

Shuangjie Zhang & Yuqing Zhang

UCSC

12/4/2019

Find the most valuable car by 2013

- Find relationship behind suggested retail price and hybrid car features.
- <https://shuangjiezhang.shinyapps.io/STA204-Final-Project/>

Overview

Data Review

Method

$$Y = X\beta + \epsilon, \epsilon \sim N(0, \sigma^2 I)$$

Variable Selection & Model Comparison

Outlier Test

Conclusion

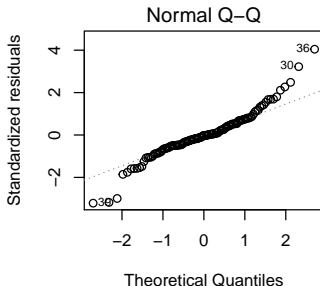
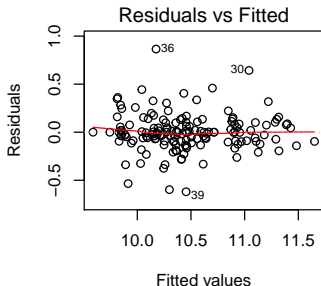
Data Review

- Transform msrp, mpg, mpgmpge into log form
- $\log(\text{msrp}) \sim \text{year} + \text{accelrate} + \log(\text{mpg}) + \log(\text{mpgmpge}) + \text{carclass}$
- Additional variables: brand, company, country

	Estimate S	td. Error	t value	Pr(> t)
(Intercept)	14.7014041	17.0213437	0.8637041	0.3892070
year	-0.0019610	0.0084967	-0.2307977	0.8178042
accelrate	0.0864023	0.0119159	7.2510202	0.0000000
mpg	-0.5482133	0.1750384	-3.1319610	0.0021091
mpgmpge	0.1636797	0.1175983	1.3918541	0.1661428
carclassL	0.3903007	0.1465221	2.6637669	0.0086204
carclassM	-0.0797795	0.0769701	-1.0365009	0.3017304
carclassMV	0.2985002	0.1673562	1.7836217	0.0766211
carclassPT	-0.0809727	0.1756483	-0.4609933	0.6455087
carclassSUV	0.0632425	0.0999264	0.6328907	0.5278229
carclassTS	-0.2553617	0.1305925	-1.9554088	0.0524983

Variable Selection

- All parameters
- $\log(\text{msrp}) \sim \text{year} + \text{accelrate} + \log(\text{mpg}) + \log(\text{mpgmpge}) + \text{carclass} + \text{brand} + \text{company} + \text{country}$



Variable Selection

Step function with AIC/BIC

Use stepwise method & AIC/BIC criterion to select variables.

- AIC result:
year + accelrate + log(mpg) + log(mpgmpge) + carclass + brand
-> Prefer more variables
- BIC result: accelrate + log(mpg)
-> Prefer fewer variables
- $\log(153) = 5.030438$: the penalty term k for BIC is big.

Lasso

- Lasso is able to shrink coefficients to 0.
- `gamlr` function

Variable Selection

Lasso Result 1

42 x 1 sparse Matrix of class "dgCMatrix"

	s0		
year	.		
accelrate	0.06425803		
mpg	-0.20643582		
mpgmpge	.		
carclassL	0.14683584		
carclassM	-0.05010318		
carclassMV	0.06430422		
carclassPT	.		
carclassSUV	.		
carclassTS	.		
brandAudi	.	brandLincoln	.
brandBesturn	-0.12544422	brandMazda	-0.07453768
brandBMW	0.21584200	brandMercedes-Benz	0.32936708
brandBuick	.	brandMercury	.
brandBYD	.	brandNissan	.
brandCadillac	0.51753996	brandPeugeot	0.01894611
brandChevrolet	.	brandPorsche	0.18281847
brandChrysler	.	brandSaturn	-0.17052238
brandDodge	.	brandToyota	.
brandFord	.	brandVauxhall	.
brandGMC	0.13154439	brandVolkswagen	.
brandHonda	-0.26970231	country_combinedJapan	.
brandHyundai	-0.02326671	country_combinedKorea	-0.02415707
brandInfiniti	.	country_combinedOthers	.
brandJeep	-0.34807474	country_combinedUSA	.
brandKia	.		
brandLexus	0.12021810		

Variable Selection

Lasso Result 2

14 x 1 sparse Matrix of class "dgCMatrix"
s0

year	.
accelrate	0.06355273
mpg	-0.26222492
mpgmpge	.
carclassL	0.08420204
carclassM	.
carclassMV	.
carclassPT	.
carclassSUV	.
carclassTS	.
country_combinedJapan	.
country_combinedKorea	.
country_combinedOthers	.
country_combinedUSA	.

Variable Selection

Lasso Result 3

38 x 1 sparse Matrix of class "dgCMatrix"

s0	
year	.
accelrate	0.06451812
mpg	-0.21044835
mpgmpge	.
carclassL	0.14889670
carclassM	-0.04364799
carclassMV	0.02936425
carclassPT	.
carclassSUV	.
carclassTS	.
brandAudi	.
brandBesturn	-0.07000165
brandBMW	0.18955555
brandBuick	.
brandBYD	.
brandCadillac	0.47480093
brandChevrolet	.
brandChrysler	.
brandDodge	.
brandFord	.
brandGMC	0.10404587
brandHonda	-0.25229957
brandHyundai	-0.02464128
brandInfiniti	.
brandJeep	-0.28771098
brandKia	.
brandLexus	0.10221202
brandLincoln	.
brandMazda	-0.03957247
brandMercedes-Benz	0.30217752
brandMercury	.
brandNissan	.

Variable Selection

Summary of first stage selection

- Surprisingly, year and country are not significant.
- Left variables: `accelrate`, `log(mpg)`, `carclass`, `brand`
- Remark: lasso treats each level separately as different covariates.
- Next, we considered interaction terms and combine some `carclass/brand` levels.

Model Comparison

Model LM1 - LM4

- LM1:
 $\log(\text{msrp}) = \text{accelrate} + \log(\text{mpg}) + \text{carclass} + \text{brand}$
 - Significant:
 $\text{accelrate}, \text{carclassL}, \text{carclassMV}, \text{brandCadillac}, \text{brandJeep}$
- LM2: $\log(\text{msrp}) = \text{accelrate} * \log(\text{mpg})$
 - All coefficients are significant at 95% confidence level.
 - Adj R-squared is only 0.571.
- LM3: $\log(\text{msrp}) = \text{accelrate} * \log(\text{mpg}) + \text{carclass}$
 - Significant:
 $\text{accelrate}, \text{carclassL}, \text{carclassTS}, \text{accelrate}:\log(\text{mpg})$
- LM4: $\log(\text{msrp}) = \text{accelrate} * \log(\text{mpg}) + \text{brand}$
 - Significant: $\text{accelrate}, \log(\text{mpg}), \text{brandCadillac}, \text{brandJeep}, \text{brandMercedesBenz}, \text{accelrate}:\log(\text{mpg})$

Model Comparison

Table 2: Comparison of LM1 through LM4

	AIC	BIC
LM1	-413.3292	-301.2030
LM2	-346.8987	-334.7770
LM3	-353.3303	-323.0259
LM4	-427.6842	-330.7102

Model Comparison

Model LM5 - LM7

- Identify representative brands and carclasses
- LM5: $\log(\text{msrp}) = \text{accelrate} * \log(\text{mpg}) + \text{brandCadillac} + \text{brandJeep} + \text{brandBenz}$
 - We encode brand into three significant brands and others with 0.
 - All coefficients are significant, adj R-squared=0.6513.
- LM6: $\log(\text{msrp}) = \text{accelrate} * \log(\text{mpg}) + \text{brandCadillac} + \text{brandJeep} + \text{brandBenz} + \text{classL} + \text{classMV} + \text{classTS}$
 - Similarly, add in three significant carclasses.
 - Only TS (Two-Seater) is significant among the three carclasses.
- LM7: $\log(\text{msrp}) = \text{accelrate} * \log(\text{mpg}) + \text{brandCadillac} + \text{brandJeep} + \text{brandBenz} + \text{classTS}$
 - All coefficients are significant, adj R-squared=0.6686.

Model Comparison

Table 3: Comparison of LM2, LM4, LM5, LM6, LM7

	AIC	BIC	adj_R_squared
LM2	-346.8987	-334.7770	0.5710
LM4	-427.6842	-330.7102	0.7839
LM5	-375.7193	-354.5062	0.6513
LM6	-382.5683	-352.2639	0.6726
LM7	-382.5769	-358.3334	0.6686

- Check for interaction: $\text{step}(lm(\log_msrp \sim (\text{accelrate} + \log_mpg + \text{brandCadillac} + \text{brandJeep} + \text{brandBenz} + \text{classTS})^2))$
- We also got LM7.

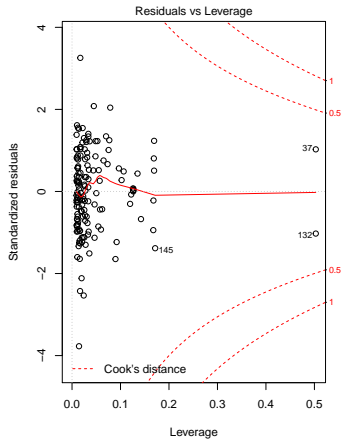
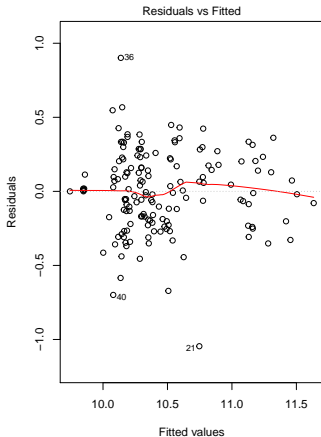
Final Model

Table 4: Final model: LM7

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.5130315	1.1829141	4.660551	0.0000071
accelrate	0.5218653	0.0999225	5.222703	0.0000006
log_mpg	1.1280302	0.3407428	3.310504	0.0011752
brandCadillac	0.8982650	0.2051142	4.379341	0.0000227
brandJeep	-0.6994143	0.2806561	-2.492069	0.0138260
brandBenz	0.4770443	0.1198410	3.980642	0.0001083
classTS	-0.3127326	0.1063798	-2.939775	0.0038235
accelrate:log_mpg	-0.1269567	0.0294695	-4.308073	0.0000302

Outlier Test

```
## Warning: not plotting observations with leverage one:  
##      59
```



Mean shift outlier test

- $Y = X\beta + d_i\phi + \epsilon, \epsilon \sim N(0, \sigma^2 I)$
- d_i is an n -vector with i -th element equal to 1, and all other elements are 0.
- $H_0 : \phi = 0$ and $H_1 : \phi \neq 0$
- Bonferroni p-values for testing each observation in turn to be a mean-shift outlier

```
##          rstudent  unadjusted p-value Bonferroni p
## 21 -3.958069          0.00011826      0.017976
```

- <https://shuangjiezhang.shinyapps.io/STA204-Final-Project/>

Conclusions

- Acceleration rate and mpg both positively affects the price, but their interaction will moderate the effect.
- Three representative brands are Cadillac(+), Jeep(-) and Mercedes-Benz(+).
- TS (Two-Seater) has significantly lower price than other car classes.



'# of people looking at the same car!' is just from random number generation. Don't believe it!

Reference

- ① Lim, D.-J., Jahromi, S. R., Anderson, T. R., Tudorie, A.-A. (2015). Comparing technological advancement of hybrid electric vehicles (HEV) in different market segments. *Technological Forecasting and Social Change*, 97, 140-153. doi: 10.1016/j.techfore.2014.05.008.
- ② University of Toronto R homework:
<https://www.coursehero.com/file/12498270/a3/>
- ③ University of Illinois at Urbana-Champaign:
<https://www.studocu.com/en-us/document/university-of-illinois-at-urbana-champaign/applied-regression-and-design/mandatory-assignments/kanuru-sai-hw-06-homeworks-given-every-lecture/3821289/view>



Thank
you!!