Economics 671
Problem Set #2

(1) The data set on the website, **carsdata.txt**, contains information on a sample of 43 different compact and midsized automobiles sold in the United States. This data ist taken from *Consumer Reports* January 2018 reliability ratings of different cars.

The following variables are available in the data set, with description of each below. These follow the ordering of the columns in *carsdata* (we won't use all of them in the analysis below):

- $Price$ = MSRP (Manufacturer's Suggested Retail Price) of automobile. (NOTE: Divide this by $1,000 in the regressions to follow to aid / ease coefficient interpretation).
- $Reliability$ = A categorical measure of expected car reliability, ranging from 1 (most reliable) to 5 (least reliable).
- $OwnerSat$ = A categorical measure of average owner satisfaction, ranging from 1 (most satisfied) to 5 (least satisfied).
- $RoadScore$ = A numerical summary of driving performance based on objective tests (e.g., braking) and subjective experiences. The variable ranges from 0-100, with 100 being the highest score.
- MPG = Overall Miles Per Gallon
- $ZeroTo60$ = Time (in seconds) to accelerate from a stop to 60 mph.
- HP =Horsepower (higher numbers indicate more power).

(1a) Write a MATLAB program to estimate the following regression model:

$$Price_i = \beta_0 + \beta_1 NotReliable_i + \beta_2 RoadScore_i + \beta_3 MPG_i + \beta_4 Zeroto60_i + u_i,$$

where $NotReliable$ is a binary variable that you construct, which equals one if *Reliablity* is equal to 4 or 5 (the two worst categories). Within MATLAB, you can create this variable using syntax like the following:

```
NotReliable = (Reliability==4 | Reliability ==5)
```

Your regression model should include: (a) an estimate of the coefficient vector $\hat{\beta}$, (b) standard errors associated with your estimate, (c), and estimate of the variance parameter, $\hat{\sigma}^2$

and (d) R^2. I have attached STATA output for your reference. Print out and submit the m-file you used to perform these calculations. Briefly comment on your coefficient estimates and how to interpret their magnitudes.

(1b) You will notice that the coefficient estimate on *mpg* is negative. Would you agree, then, with the conclusion that US automobile consumers, all else equal, prefer cars that are less fuel efficient?

(1c) Consider the coefficient $\hat{\beta}_4$. I want you to reproduce this coefficient in two different ways.

- First, by directly coding in MATLAB the "short regression" estimate of this coefficient (i.e., page 13 of the "Regression 2 ‘' slides)
- Second, by following the two-step procedure (page 15 of that same set of notes) that helps to provide an interpretation of what multiple regression coefficients do and how they are interpreted.

(1d) The 40th observation in the sample is an Audi A6, the sweet ride driven by our very own Dean. (He sighs when typing this, thinking of his own cracked-windshield Hyundai). Based on the observed characteristics of the A6, would you expect the price to be more or less than the actual MSRP? By how much?

(1e) You find yourself reading a car magazine and are drawn to a new car, the Toyota Vector (ha ha). An MSRP has not yet been released for this vehicle, but you are in the market for a new car and have a $35,000 budget. You see that the Vector is predicted to be very reliable, receives a driving score of 75, gets 27 mpg and has a zero-to-60 time of 9 seconds. Do you predict that the Vector will be within your budget?

(1f) Using your regression in (1a), calculate the fitted value vector $\hat{y}$ using MATLAB. Now, regress *Price* on those fitted values. What coefficient estimate (approximately) do you get? Finally, prove (using linear algebra and what we have reviewed in class) why this should be the case.