

Case 1: Explaining prices of U.S. cars

Deadline: Monday February 28, 2022, 17.00 AMS (upload)

This assignment should be done in teams of exactly 2 students.

Requirements

- Put your **group number, names, VU ID numbers, and email addresses** of both group members on the front at a separate title page.
- You have to **answer each (sub) question separately**. Hence do **NOT** write a report!!!
- Reporting style as explained in class. Title, text layout, clarity of tables and figures, no screen dumps, tables no vertical lines, tables and figures in Journal of Finance layout (style + explanatory note, to be stand-alone), etc. Failure to meet the guidelines results in loss of points.
- If you have to make a Figure, delete the blue background color created by Stata.
- If you are asked to **test** something, always provide H_0/H_A , the test-statistic (formula and actual outcome) and your **economic** conclusion.
- Only one of you has to upload the final version before the deadline. Do ***not*** upload two versions of the paper / files, one by each one of you.
- You **upload your report in PDF** format via Canvas. Name of the file: student ID 1 nr _ student ID 2 nr plus "case1_QRM3.pdf", e.g., 2056791_111222_case1_QRM3.pdf.
Be aware that you should provide your **annotated / commented DO file** in the appendix of your report for me to run. The do file should be such that I can run it to get precisely the tables and figures you have in your report. Also make sure your do file has a "`cd your-map-name`" command at the start, such that all file names in the rest of the file are relative to this map [see the lab sessions].
- Clarity of descriptions, replicability, clarity of argumentation, interpretation and conclusions do earn you points.

Description and Data

In this case you will work with prices and other characteristics of cars in the United States. The first goal is to explain the car price by characteristics such as miles per gallon (mpg), weight, length, trunk space etc. Here you will make use of the *linear regression model*.

Second, we transform the car price into a binary variable that equals 1 if the car is expensive and 0 otherwise. Then we would like to model the probability that a car is expensive given the aforementioned car characteristics.

Each group has its own dataset, which you can find via **canvas-> files > cases > Case 1 > Data**. In summary, you have the following data:

Variable	Description
car nr	variable that runs from 1 until N, the number of cars in the sample
price	Price (dollars)
mpg	Mileage (miles per gallon)
headroom	Headroom (inches)
trunk	Trunk space (cubic foot)
weight	Weight (pounds, lbs)
length	Length (inches)
turn	Turn Circle (foot)
displacement	Displacement (cubic foot)
gear-ratio	Gear ratio
foreign	Car type: 1 if foreign, 0 if domestic (US).
Ind_buick	1 if car make is a Buick, 0 else
Ind_olds	1 if car make is an Olds

Table A: Overview of car data. In parentheses, you find how the variable is expressed.

Part I: Explaining car prices

1a) Create a table with summary statistics of your variables used in your assignment. Put at least the mean, standard deviation, min and max in **Table 1**. Also create a variable **ln(price)**, the (natural) logarithm of the price, into your table.

Remark 1: Be sure that all your tables and figures are according to the Journal of Finance (JF) style.

Example: See Table 1 of this article: <https://onlinelibrary.wiley.com/doi/full/10.1111/jofi.12505>

1b) Estimate a linear regression model with **price** as dependent variable and at least the following variables as independent variables: **headroom**, **weight**, **length** and **turn**.

Put the estimated coefficients and standard errors, as well as the F test-statistic on the joint significance of all variables and the R-squared into **Table 2**, making use of `outreg2`.

Remark 2: You are free to include other variables as well next to the four variables mentioned here, but do **NOT** include any dummy variable into your model yet.

Remark 3: In question 1d, 1f, and 1i you need to extend **Table 2**. Therefore, show in your report just **once** Table 2, containing all the joint estimation results from 1b, 1d, 1f and 1i.

Remark 4: See Stata lab 0 for more information about `outreg2`.

Remark 5: Sometimes `outreg2` does not provide you all the results you need. You are allowed to change your table manually.

1c) Given the model of 1b, what is the exact interpretation of

1c1: the estimated constant?

1c2: the estimated **length** coefficient?

1d) Again estimate the model of 1b, but now also include the dummy variables **ind_buick**, **ind_olds** and **foreign** into the model.

Extend Table 2 with a new column containing estimated coefficients, standard errors, the F-stat and the R-squared of your new model.

Provide the *exact* theoretical model equation corresponding to your new model.

1e) Consider the estimated coefficients of the model of **1d**. What is the interpretation of the estimated **ind_buick** coefficient?

1f) You hypothesize that, compared to the other car makes, the effect of **weight** and **length** on the car price is not only different for **buicks**, but also for **olds**.

Estimate a new linear regression model by extending the model of **1b** such that you are able to test your hypothesis.

Remark 6: you don't have to **test** your hypothesis yet; you will do this in question **1g**.

Extend Table 2 with a new column containing estimated coefficients, standard errors, the F-stat and the R-squared of your new model.

Provide the *exact* theoretical model equation corresponding to your new model.

Interpret **all** coefficients related to the variable **weight**.

1g) Test your hypothesis of **1f** by performing an adequate statistical test. What is your (economic) conclusion?

1h) Given the model of **1d**, you are worried about heteroskedasticity. Test on this phenomena manually (i.e. do **not** use a build-in Stata command that runs a test on heteroskedasticity). Report at least the possible used test-equation(s), the test-statistic and your conclusion.

1i) Estimate again the regression model of **1d**, but now with **ln(price)** as dependent variable.

Extend Table 2 with a new column containing estimated coefficients, standard errors, the F-stat and the R-squared of your new model.

Interpret the estimated **foreign** coefficient.

1j) Given the model of **1i**, you are again worried about heteroskedasticity.

Would you expect that heteroskedasticity is **less** or **more** likely apparent in the model of **1i** compared to the model of **1d**? Explain your answer!

NOTE: See next page for Part II of this Case

Part II: Explaining expensive cars

1k) Create a variable **expensive** that equals 1 if the car price exceeds 8000 dollar and 0 otherwise. What is the percentage of non-expensive cars in your sample?

1l) Estimate a logit model with **expensive** as the binary variable and **headroom**, **weight**, **turn** and **foreign** as independent variables.

Put your estimation results plus the pseudo R-squared into **Table 3** using `outreg2`.

Provide the exact theoretical model equation of your logit model.

What can you say about the estimated **foreign** coefficient?

1m) Compute the marginal effects implied by your estimated logit model of all four variables according to the following two methods:

Method I: the average marginal effect

Method II: the marginal effect at the mean

Put both types of marginal effects in **Table 4**.

Interpret the estimated marginal effect of **turn** according to method I and II. Comment on possible differences. Can you explain this?

Note: Do not forget to include an annotated DO-file into your final PDF at the very end!