Problem Set 1: Demand Estimation for Differentiated Products
Empirical Industrial Organization

In this assignment you have to estimate a simple version of a random-coefficients model for car demand. The objective of the assignment is to familiarize with these type of discrete choice models, to learn to estimate non-linear models via GMM, and measure markups in differentiated products markets.

You can work in groups of 3 people, each group has to send a pdf report as well as the code files used to generate the results.

When reporting results, present them in tables that are easy to read and discuss your findings. Do not simply attach a printout of the computer program you used.

Deadline: Please send your problem set before **November 4th**.

1 Data

The main dataset is a sample of market shares and car characteristics sold in the U.S. between 1977 and 1981. This is included in the data file `carpanel.csv`.

Construct the market shares using the aggregate quantities. For the market size, use the number of households per year that in included in `UShouseholds.csv`.

In addition, `gasprice.csv` and `USCPI.csv` contain average gasoline prices and the consumer price index for those years. Use these, to construct measures of dollar per mile. The prices of cars and the price for gas need to be corrected by CPI to make different years comparable:

$$P_{adj} = P * \frac{100}{CPI}$$

2 Logit Model

Consider the following specification:

$$u_{ij} = X_j\beta + \alpha p_j + \xi_j + \epsilon_{ij} \tag{1}$$

where X_j includes the following variables:

`dpm, door3, door4, door5, at, ps, air, drv, wt, hp2wt, hp, euro, japan, size, and wb`

They are respectively, dollars per mile, dummies for having 3, 4, or 5 doors, dummies for having automatic transmission, power steering, air conditioning, and front wheel drive; the car weight, horse power, horse power to weight ratio, dummies for being european or japanese, size, and wheelbase.

Construct a linear estimating equation and estimate α and β using OLS with and without brand fixed-effects. The brand is given by the variable firmids.

Construct a set of instruments for price and estimate α and β using IV with and without brand fixed-effects. Compare your results across the four specifications.

3 Nested Logit Model

Consider now a nested logit, where the nests g are defined using three size categories: compact, midsize, and large.

$$u_{ij} = X_j\beta + \alpha p_j + \xi_j + \eta_{ig} + (1-\sigma)\epsilon_{ij}$$

Use the same X_{jt} variables as in equation (1).

Follow the derivation in Berry(1994) to construct a linear estimating equation for the parameters α, β, and σ.

Now your estimating equation involves two endogenous variables, so in addition to the instruments used in the previous section you have to construct new sets of instruments.

Estimate the parameters using OLS and IV, with and without brand fixed-effects. For each specification report the results of joint weak IV tests and over-identification tests.

4 Random-Coefficients Logit Model

Consider the following random-coefficients model:

$$u_{ij} = X_j\beta + \alpha p_j + \xi_j + \eta_i Size_j + \epsilon_{ij} \tag{2}$$

Assume that the marginal utility for car size, $\eta_i \sim N(0, \sigma_\eta^2)$ and use the same X_{jt} variables as in equation 1.

In addition to the IVs you created in the previous section, compute a new set of instruments that take into account the fact that you now have a continuous measure of differentiation. Estimate the model by GMM using the nested logit IVs and the new IVs you created (without model fixed-effects).

Hint: The minimization problem involves only one parameter (σ_η), since β can be concentrated out of the minimization problem.

5 Markups and Elasticities

Calculate the implied marginal-costs, markups, and demand elasticities. For the supply side, follow the pricing model of Berry, Levinsohn, and Pakes (1995). Assume that prices are set as a Bertrand game with multi-product firms.

6 Merger Simulation

Simulate and analyze a merger where Fiat (firm=17) sells its operations to VW (firm=7) in 1981.

Appendix: Variables in the Dataset

US Households Data:

nb_hh: number of households in the US in thousands.

Panel Car Data:

name: model.name
id: model id
yr: model year abbreviated
cy: cylinders
dr: number of doors
at: automatic transmission
ps: power steering
air: air conditioning
drv: front-wheel drive
p: price
wt: weight
dom: domestic brand
disp: engine displacement
hp: horse power
lng: length
wdt: width
wb: wheelbase
mpg: miles per gallon
q: quantity
firmids: car brand
euro: European brand
reli: reliability index
dfi: direct fuel injection
hp2wt: horsepower to weight ratio
size: car size
japan: Japanese brand
year: model year
cat: size category