Automobile Demand Estimation

Zhang Donglei, Liu Chang, Deng Sijing

April 28, 2017

Abstract

This paper adopts BLP to estimate the demand parameters of the automobile markets of UK and Belgium.

1 Introduction

We focus on all automobile models marketed in UK and Belgium from 1970 to 1999. The data is aggregated at the automobile model level. We'll estimate the demand parameters for either country separately and then compare the difference.

Automobile is a typical industry with differentiated products, which are designed with various characteristics. This section provides the first hint for the UK and Belgium markets respectively.

During the 30-year time window, automobile models enter and exit the market. The total number of automobile model is 2301 in UK and 2673 in Belgium. In UK, the number of model more than doubles at the peak of 89 in 1986 compared with the minimum of 40 in 1970. In Belgium, the least number of model, 65, appears in 1970 and the largest number of model, 102, appears in 1984 and 1999. In both countries, the variety of auto models increases steadily in the 70s.

The average annual sales doesn't show any monotonic trend. The UK market sales slumps in the late 70s and then retrieve in the late 80s with slight shock in early 90s. The low-tide periods of Belgium market happens at the early 80s and late 90s. The population of both UK and Belgium gradually increase from 1970 to 1999. We'll use the household number as proxy of total market size.

The discrepancy of the annual sales quantity among different auto models is huge. In UK, the lowest annual sales model is AlfaRomeo in 1994, 296, while the largest sales quantity, 193,784, comes from Ford in 1979. In Belgium, Fiat in 1974 has lowest annual sales, 51 and Rover contribute the highest annual model sales, 62,410, in 1979. The largest annual model sales, either Ford or Rover, surpass the average annual total market sales.

The price variable we choose is a standardized ratio, price relative to per capita income. Price has two depressions at late 80s and late 90s in UK and late 70s and 90s in Belgium.

Both markets behave consistently in the perspectives of liter per kilometer and size. The efficiency, reflecting on the liter per kilometer increases relatively steadily with a slight reverse in the last three years, possibly because of the popularize of larger size vehicle.

2 Dataset and Variables

We use two data files in our program: `autoblp.mat` and `draw.mat`.
`autoblp.mat` contains all necessary characteristics of automobile we used:

- **cdid** - the first column, market index;
- **const** - the second column, a vector of ones;
- **cy** - the third column, cylinder volume or displacement;
- **firm** - the fourth column, firm code;
- **home** - the fifth column, domestic car dummy;
- **hpwt** - the sixth column, horse power per weight;
- **model** - the seventh column, model code;
- **price** - the eighth column, price relative to per capita income (used as price variable);
- **s_jt** - the ninth column, market share;
- **space** - the tenth column, the approximate volume of the automobile, = length*weight*height;

`draw.mat` contains all variables we obtain by random drawing:

- **v** - random draws given for the estimation. For each market 50 iid normal draws are provided. They correspond to 50 "individuals", where for each individual there is a different draw for each column of **x2**.
- **demogr** - draws of demographic variables from the Office for National Statistics of UK government for 50 individuals in each market. The first 50 columns give the income, the next 50 columns the income squared, columns 101 through 150 household size and the last 50 columns the number of children per family.

Table 1 provides summary statistics for the variables enter in the demand model.

Table 1: Summary Statistics

	cy	home	hpwt	price	share	space
Mean	1478.727	0.104	0.057	0.873	0.000373	9685062
Std. Dev.	487.249	0.305	0.013	0.427	0.000485	1575907
Min	425	0	0.018	0.247	5.07e-06	5194530
Max	4520	1	0.118	6.473	0.006343	1.36e+07
Observation	4974	4974	4974	4974	4974	4974

3 Results and Analysis

3.1 Product characteristics and Instruments

The product characteristics we used are cylinder volume or displacement, domestic car dummy, horse power per weight, and the approximate volume of the automobile.

To identify the price and other nonlinear parameters, we employ three types of instruments. First, we make use of the own characteristics of the product. Second, the sum of the characteristics of other products in the same firm in the market. Third, the sum of the characteristics of the products in the rival firm in the market.

3.2 Result

We use different kinds of optimization algorithms when doing estimation, including Nelder-Mead Simplex, Quasi-Newton, Mesh Adaptive Direct Search (MADS), Generalized Pattern Search (GPS) etc. GPS algorithm outperforms others on our dataset. Table 2 shows our estimation result.

Table 2: Estimation Result

Parameters	Variables	Parameter Estimate#	Standard Error
Means	price	-2.517**	1.053
	const	-1.346*	0.750
	hpwt	-12.082***	4.232
	cy	-7.991***	1.285
	space	1.833*	0.981
	home	-4.827***	0.595
Std. Dev.	price	2.164***	0.599
	const	0.698	0.477
	hpwt	-5.366	7.022
	cy	-2.464***	0.808
	space	-2.194***	0.706

Parameter estimates are reported in the tables with *, **, and *** indicating significance at the 10%, 5%, and 1% levels, respectively.

Our estimation results contains two panels, the first panel shows the estimation of the parameters corresponding to means, and the second panel provides the parameter estimates of standard deviations of the taste distribution. For the mean part, all of the parameters are statistically significant. The parameters on `price`, `const`, `hpwt`, `cy`, and `home` are negative, while the parameter on `space` is positive. The negative parameter on price indicates that on average people prefer lower price automobiles while holding other variables constant.

On the other hand, the estimate of the standard deviations of the distribution of marginal utilities for `price`, `cy` and `space` are substantial and estimated precisely enough to be considered significant at 99% significant levels. Thus, each of the included variables is estimated to have a significantly effect on the mean of the distribution of utilities, while `price`, `cy` and `space` is estimated to have significantly effect on both mean and standard deviation of the distribution of utilities.

4 Appendix: Code Description

4.1 Executable Codes

The executable codes are:

- **main.m**: The main program. Including Initializing global data, defining paths, and providing a complete frame for optimization. In the optimization loop, variables that enter the linear and non-linear part of the estimation are declared and initialized, as well as market shares and demographic variables. Several different algorithms are specified for optimization purpose.

- **gmmobj.m, gmmobj2.m, gmmobj3.m**: Constructing the GMM objective function. The subtle differences between them are some specific modifies for each optimization routine.

- **mktsh.m**: The predicted market shares. $\frac{1}{nx}\sum_{i=1}^{ns} s_{jti}$ in (A-1).

- **ind_sh.m**: Individual probabilities of choosing each brand. s_{jti} in (A-1).

- **mufunc.m**: $\sum_{k=1}^{K} x_{jt}^{k}(\sigma_k v_i^k + \pi_{k1} D_{i1} + ... + \pi_{kd} D_{id})$ in (A-1).

- **gradobj.m**: Compute the gradient of the objective function.

- **meanval.m**: Calculate the $J \times T - dimentional$ vector of mean valuations, δ_{jt}.

- **jacob.m**: Compute the Jacobian matrix of δ.

- **var_cov.m**: Compute the variance-covariance matrix.

- **printm.m**: Standard format for printing logs.

- **optim_results_summary.m**. It can generate an xls-workbook concludes almost all estimation results such as the optimal values of objective function and θ.

- **drawrnd.m**: Take random draw of demographic data and v.

- **gen_matlab_data.m**: Generate .mat data from .dta data.

4.2 Results

The results are saved in folder "`Optimization logs`" and "`Optimization results`". To obtain the similar result,

- run `datagen.do` in `Data.rar` (line 1-29) to obtain the `.csv` data file;
- run `gen_matlab_data.m` in `BLP_code.rar` to obtain the `.mat` data file;
- run `main.m` (edit : routine 9, perturbs = 50 , mymaxfunevals = 500);
- run `optim_results_summary.m` to obtain the excel file (edit: routine 9, and you can find the excel file in "~\`Optimization results`\..."