Car demand estimation using Random coefficient model by BLP (1995)

Hung-Thuy Nguyen

January 16, 2019

1 Preliminary

Gas price is used to conduct dpm (dollar per mile) variable. The regressions are run using adjusted values of car price and gasoline price, with the reference corresponding to price index of 100. Other explanatory variables include:

- door3, door4, door5: dummies for the number of doors
- at: dummy for having an automatic transmission
- ps: dummy for power steering
- air: dummy for air conditioning
- drv: dummy for front wheel drive
- wt: car weight, in pounds
- hp2wt: ratio of horse power to weight
- hp: horse power
- euro: dummy for european cars
- japan: dummy for japanese cars
- wb: wheelbase
- size: size of the car
- price: adjusted price, **in thousands**

2 Logit Model

From utility function for consumers:

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

the estimation questions are as follow (assuming ϵ_{ij} follows Extreme Value Type I distribution):

$$ln(s_j) - ln(s_0) = X_j\beta - \alpha p_j + \xi_j$$

with s_0 representing the share of outside option. The outside option is not buying a car and is normalized to case in which every coefficient is zero, except the constant.

2.1 Estimation

For this problem, prices are related to observed and unobserved car characteristics, giving rise to endogeneity problem. Proceeding similar to S. Berry, Levinsohn, and Pakes 1995, I have constructed 2 sets of instruments for price.

- Sum of characteristics of other cars by the same firm, except the car in consideration (IV1).
- Sum of characteristics of cars by all other competing firms (IV2).

Moreover, for certain specifications, we want to allow for brand or firm fixed-effects. Hence, a set of dummy variables is needed. Here, there are 19 different firms and the reference is firm with id "1". Table 1 below presents the raw regression results for 4 specifications, using all instruments and brand dummies:

- (1) OLS without IVs and without brand fixed-effects
- (2) OLS without IVs but with brand fixed-effects
- (3) OLS with IVs but without brand fixed-effects
- (4) OLS with both IVs and brand fixed-effects

We can see that the results in column (4) of table 1 do not make much sense. Matlab also gives warning about the possible problem of nearly perfect collinearity. Therefore, it is worth looking closer at the brand dummies and the instrument variables. Regarding brand dummies, the regression reveals that dummy coefficients of firm 14 and 24 are zeros (position 12 and 18 in the list of firm ids). Therefore, I exclude these 2 firms, assuming that firm fixed-effects of these 2 firms are the same as firm 1. This removal does not have any effects on regression coefficients and standard errors in the specification (2), but necessary to have a more precise estimation when combining both brand dummies and IVs.

I have the same issue of nearly perfect collinearity when using 2 sets of instruments at the same time. I suspect that for some instrumental variables, the variations between the 2 instrument sets are not significant. To improve the estimation, I use the correlation between the 2 sets, pair by pair, to leave out some instruments.

The correlation between instruments is presented in table 2 above. From this, I remove instruments from the set IV1 which have correlation coefficient (in absolute value) greater or equal to 0.8 with their counterparts in the set IV2: door4, wt, hp2wt, hp, wb and size. Nevertheless, the coefficients and standard deviations for specification (3), with IVs and without brand dummies, do not change. Therefore, in the end, I decide to keep both sets of instruments.

The regression results using the corrected set of dummies and instruments are reported in table 3. In addition to 4 columns as in table 1, columns (5) and (6) are results of specification (4) when using IV1 and IV2 only, respectively.

Several comments on the results:

Table 1: Logit model using all instruments and brand dummies

	OLS (1)	OLS (2)	OLS (3)	OLS (4)
const	-5.2384	-8.1460	-5.3909	410.1100
dpm	-22.0430	-19.0420	-22.4480	-14.8950
door3	-0.6236	-0.5271	-0.6244	-0.5389
door4	-0.0636	-0.0190	-0.0686	0.0265
door5	0.3424	0.4548	0.3282	0.4958
at	-0.2868	-0.2917	-0.3238	-0.2766
ps	0.2057	0.2052	0.2069	0.1873
air	0.1226	-0.1486	0.0453	0.3096
drv	0.1029	0.1621	0.1142	0.2549
wt	-0.0022	-0.0011	-0.0022	-0.0012
hp2wt	-9.8210	-5.9342	-9.9392	-7.8018
hp	0.0381	0.0228	0.0373	0.0352
euro	-1.7716	-1.4978	-1.8994	540.8900
japan	-0.1290	0.6086	-0.1543	-414.5400
wb	0.0224	0.0431	0.0255	0.0254
size	2.7101	0.6433	2.7410	0.7269
price	-0.0396	-0.0644	-0.0255	-0.1533

Table 2: Pair-wise correlation between IVs in IV1 and IV2

dpm	-0.477
door3	0.177
door4	-0.960
door5	-0.110
at	-0.655
ps	-0.526
air	-0.082
drv	-0.182
wt	-0.924
hp2wt	-0.809
hp	-0.929
euro	-0.337
japan	-0.330
wb	-0.858
size	-0.898

Table 3: Logit model using selected instruments and brand dummies

	(1)	(2)	(3)	(4)	(5)	(6)
const	-5.2384*** (1.4214)	-8.1460*** (1.3756)	-5.3909*** (1.4311)	-6.5265*** (1.6057)	-6.4693*** (1.6905)	-6.2098*** (1.6633)
dpm	-22.0430*** (3.4604)	-19.0420*** (3.1515)	-22.4480*** (3.4880)	-14.8950*** (3.7716)	-14.7490*** (4.0053)	-14.0850*** (3.9201)
door3	-0.6236*** (0.1642)	-0.5271*** (0.1503)	-0.6244*** (0.1644)	-0.5389*** (0.1565)	-0.5393*** (0.1569)	-0.5412*** (0.1591)
door4	-0.0636 (0.1029)	-0.0190 (0.0933)	-0.0686 (0.1031)	0.0265 (0.0992)	0.0281 (0.1006)	0.0354 (0.1013)
door5	0.3424* (0.1977)	0.4548** (0.1771)	0.3282* (0.1984)	0.4958*** (0.1852)	0.4972*** (0.1862)	0.5038*** (0.1885)
at	-0.2868* (0.1615)	-0.2917* (0.1606)	-0.3238* (0.1660)	-0.2766* (0.1672)	-0.2761* (0.1678)	-0.2737*** (0.1701)
ps	0.2057 (0.1496)	0.2052 (0.1487)	0.2069 (0.1498)	0.1873 (0.1550)	0.1867 (0.1555)	0.1838 (0.1576)
air	0.1226 (0.1592)	-0.1486 (0.1600)	0.0453 (0.1779)	0.3096 (0.2648)	0.3258 (0.3028)	0.3993 (0.2838)
drv	0.1029 (0.1104)	0.1621 (0.1111)	0.1142 (0.1111)	0.2549** (0.1229)	0.2582** (0.1267)	0.2731** (0.1263)
wt	-0.0022*** (0.0004)	-0.0011*** (0.0004)	-0.0022*** (0.0004)	-0.0012*** (0.0004)	-0.0012*** (0.0004)	-0.0012*** (0.0004)
hp2wt	-9.8210*** (2.3918)	-5.9342*** (2.2820)	-9.9392*** (2.3968)	-7.8018*** (2.5185)	-7.8677*** (2.5944)	-8.1670*** (2.5869)
hp	0.0381*** (0.0084)	0.0228*** (0.0080)	0.0373*** (0.0084)	0.0352*** (0.0100)	0.0357*** (0.0108)	0.0377*** (0.0105)
euro	-1.7716*** (0.1451)	-1.4978*** (0.3742)	-1.8994*** (0.1954)	-1.2799*** (0.4016)	-1.2722*** (0.4086)	-1.2373*** (0.4105)
japan	-0.1290 (0.1493)	0.6086** (0.2672)	-0.1543 (0.1516)	0.7100** (0.2818)	0.7136** (0.2844)	0.7299** (0.2872)
wb	0.0224 (0.0169)	0.0431** (0.0174)	0.0255 (0.0172)	0.0254 (0.0197)	0.0248 (0.0206)	0.0219 (0.0204)
size	2.7101*** (0.6952)	0.6433 (0.7208)	2.7410*** (0.6965)	0.7269 (0.7511)	0.7299 (0.7537)	0.7433 (0.7639)
price	-0.0396*** (0.0095)	-0.0644*** (0.0140)	-0.0255 (0.0173)	-0.1533*** (0.0425)	-0.1565*** (0.0512)	-0.1707*** (0.0466)

Significance code: *** 1%; ** 5%; * 10%

- Without instrumental variables, the coefficient for price is underestimated, almost 2.5 times.
- The use of different instruments and their combination (the last 3 columns), at least for this case, does not alter very much the estimation results.
- The coefficient for price in specification (3) is not significant, which is unexpected and hard to interpret.

2.2 Testing for weak instruments and over-identification

To test for relevance and validity of instruments, I use the 'ivreg' Matlab routine which provides results for F-statistics, and Sargan statistics and p-value. The results are shown in table 4. For all 3 base cases, the Sargan test suggests that we reject the null hypothesis, meaning that IVs in general are not valid. In this section, I tried several judgments on what characteristics to include as instruments to avoid the invalidity.

- IV4: all but door dummies and dummies on the origin of the car from IV1
- IV5: include only dpm, drv, wt, hp, size from IV1
- IV6: only dummies on the origin from IV1
- IV7: only dummies on the origin from IV2

Table 4: Instrument Tests

	F-statistcs	Sargan statistics	Sargan p-value
IV1	9.0080	75.1288	2.2411e-10
IV2	5.8417	72.5726	6.5773e-10
IV1 & IV2	6.5655	93.5697	1.0175e-08
IV4	12.7977	75.6694	1.1655e-12
IV5	6.7902	32.7291	1.3571e-06
IV6	15.2097	1.3196	0.2507
IV7	9.7740	1.0572	0.3039

From the table, we can see that IV6 and IV7 seem to be good candidates because of high F-statistics and that the null hypothesis of Sargan test is not rejected. In essence, IV6 (IV7) is the number of cars that a firm (rival firms) offers in a given year minus 1. It is interesting to see that solely the number of products offered by a firm and competitors can explain the set price that well. However, using IV6 gives a positive coefficient for price which is not expected. Using IV7 is better and in fact, the coefficient for price in specifications (3) and (4) is not too different, even though the coefficients for other explanatory variables do change substantially.

3 Nested Logit Model

With Nested Logit Model, products are grouped into mutually exclusive nests based on several characteristics. For this exercise, the dimension based on which the grouping is done is car size. There are 3 nests, namely compact, midsize and large. The utility for this kind of model is given as:

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

Following S. T. Berry 1994 ans steps in class lecture notes, we can derive the estimation equations:

$$ln(s_j) - ln(s_0) = X_j\beta - \alpha p_j + \sigma ln(\bar{s}_{j/g}) + \xi_j$$

with $\bar{s}_{j/g}$ is the share of product j within the nest g. It is still assumed that ϵ_{ij} follows Extreme Value Type I distribution.

3.1 Estimation

Here, we use the same set of instruments for the price as in logit model. The within-group shares become another endogenous variable because it is also correlated with unobserved characteristics of a product. By grouping products into nests, we allow for the substitution pattern between products of the same group to be different from that of a product with products in other groups. Therefore, the within-group shares can be instrumented by characteristics of other products in the same nest. This new set of instruments is called IV3. For brand fixed-effects, the same corrected dummies are used, meaning that dummies of firm 14 and 24 are left out.
The regression results are presented in table 5 below. Note that this model is also run for 4 specifications, as in previous section.

Table 5: Nested Logit model using selected instruments and brand dummies

	(1)	(2)	(3)	(4)
const	–4.7122***	-5.8547***	-4.6970***	-6.8127***
	(0.6013)	(0.6346)	(0.7435)	(1.6407)
dpm	-8.1369***	-9.1900***	-12.1310***	-15.9100***
	(1.4931)	(1.4676)	(1.9513)	(4.0002)
door3	-0.0134	0.0662	-0.2091**	-0.5798***
	(0.0707)	(0.0705)	(0.0909)	(0.1910)
door4	-0.0311	0.0031	-0.0355	0.0215
	(0.0435)	(0.0429)	(0.0536)	(0.1032)
door5	0.1019	0.1758**	0.1967*	0.5123**
	(0.0838)	(0.0817)	(0.103)7	(0.1996)
at	0.0367	-0.0741	-0.0228	-0.2932
	(0.0687)	(0.0740)	(0.0877)	(0.1785)
ps	-0.1059*	-0.0123	-0.0068	0.2040
	(0.0636)	(0.0685)	(0.0794)	(0.1658)
air	0.0162	-0.0004	0.1442	0.2639
	(0.0674)	(0.0736)	(0.0903)	(0.2464)
drv	0.1260***	0.0432	0.1049*	0.2562**
	(0.0467)	(0.0511)	(0.0577)	90.1288)
wt	-0.0003*	-0.0001	-0.0009***	-0.0013***
	(0.0002)	(0.0002)	(0.0002)	(0.0004)
hp2wt	-1.9554*	-1.3074	-4.3492***	-7.9845***
	(1.0253)	(1.0547)	(1.3099)	(2.7364)
hp	0.0104***	0.0066*	0.0203***	0.0354***
	(0.0036)	(0.0037)	(0.0046)	(0.0105)
euro	-0.1817***	0.3266*	-0.5395***	-1.4254***
	(0.0700)	(0.1775)	(0.1304)	(0.5035)
japan	0.0878	0.3118**	0.0485	0.7232**
	(0.0633)	(0.1230)	(0.0795)	(0.2991)
wb	0.0003	0.0107	0.0036	0.0290
	(0.0072)	(0.0080)	(0.0090)	(0.0205)
size	1.6909***	1.0592***	1.9821***	0.6912
	(0.2948)	(0.3315)	(0.3658)	(0.7887)
price	-0.0165***	-0.0142**	-0.0411***	-0.1500***
	(0.0041)	(0.0066)	(0.0083)	(0.0399)
nest share	0.8896***	0.8759***	0.6027***	-0.0618
	(0.0189)	(0.0210)	(0.0452)	(0.1437)

Significance code: *** 1%; ** 5%; * 10%

3.2 Testing for weak instruments and over-identification

The test for weak instruments is the same as previous logit model, based on F-statistics from the first stage regression. For over-identification test, as now there are 2 endogenous variables, the Cragg-Donald test should be used. However, I do not know how to implement this test simultaneously for 2 endogenous variables. Therefore, I have done it separately for price and within-nest shares. The results are reported in table 6 below. Similar to IV6 and IV7, IV8 includes 2 dummies for origin extracted from IV3 set.

Table 6: Instrument Tests

	F-statistcs	Sargan statistics	Sargan p-value
IVs for price			
IV1	9.0080	75.1288	2.2411e-10
IV2	5.8417	72.5726	6.5773e-10
IV3	2.4530	15.4938	0.3453
IV1 & IV2	6.5655	93.5697	1.0175e-08
IV1 & IV3	8.5703	130.9368	6.5503e-15
IV2 & IV3	8.5703	130.9368	6.5503e-15
IV1, IV2 & IV3	5.5256	130.9368	1.4435e-10
IV6	15.2097	1.3196	0.2507
IV7	9.7740	1.0572	0.3039
IV8	4.2400	3.8579	0.0495
IVs for within-nest shares			
IV1	6.3048	71.0271	1.2569e-09
IV2	6.4312	63.7865	2.5163e-08
IV3	2.8754	501.0000	0.0000
IV1 & IV2	3.9827	92.7496	1.3660e-08
IV1 & IV3	5.4912	115.7403	2.5677e-12
IV2 & IV3	5.4912	115.7403	2.5677e-12
IV1, IV2 & IV3	3.5404	121.3151	3.6998e-09
IV6	2.4827	0.6749	0.4114
IV7	2.0859	1.1157	0.2908
IV8	3.5351	1.7423	0.1868

It is quite clear that most instruments here lead to over-identification issue.

4 Random-Coefficients Model

Random-coefficients utility function:

$$u_{ij} = X_j\beta + \alpha p_j + \xi_j + \eta_i size_j + \epsilon_{ij}$$

The assumption over η_i: $\eta_i \sim N(0, \sigma_\eta^2)$.
The nested logit model allows the taste for size categories to vary. The random-coefficients model

also allows the taste for size to vary, but it treats size as a continuous variable instead of discrete size categories. As a result, we still have to instrument for price using IV1 and IV2.
In this section, I explain the procedure used to estimate to standard deviation of the taste for size, namely σ_η.

- Step 0: Guess a value for σ_η. Draw 100 values for η from $N(0,1)$, then multiply by σ_η
- Step 1: Make an initial guess for the mean utility values $\delta = \delta_j$ for $j = 1...J$. From the given (guessed) σ_η and δ, compute predicted market shares
- Step 2: Keep σ_η constant, run a loop to find values of δ. In this step, we have to use the contraction mapping given by S. Berry, Levinsohn, and Pakes 1995.
- Step 3: Use 2SLS with the values of δ we just found to estimate α, β and the residuals $\hat{\xi_j}$. The residuals are interacted with the instruments to calculate the GMM objective function. Step 1, 2 and 3 are done inside the Matlab objective function (objfunc.m)
- Step 4: Search for σ_η that minimizes the GMM objective function. The method used here is 'Quasi-newton' which is said to depend on the initial guess. However, I encountered no problem regarding that claim.
- Step 5: Check if this value of σ_η is different from our guess. If it is the case, use the average of these two as new guess for σ_η, start again from Step 0 until σ_η converges.

Using this algorithm, I have found σ_η for different cases, as presented in table 7. They are very different, depending on the instruments used.

Table 7: Estimation for σ_η

IV	σ_η
IV1 & IV2	1.3647
IV1	2.4072
IV2	1.7634

Comments:
The basis of comparison is specification (1) and (3) in which brand fixed-effects are not included. Take the results in columns (1) and (3) in table 5, the coefficients for size are 1.690 and 1.9821 respectively. In random-coefficients model, size is still included in vector X_j, as a result, σ_η measures the deviation from these numbers. Given that, we can say that the divergence in taste for size is substantial. With $\sigma_\eta = 1.3647$, there would be people with zero taste for size and some people with the coefficient equal 3.0 for instance within 2 standard deviations from the mean value.

References

Berry, Steven T (1994). "Estimating discrete-choice models of product differentiation". In: *The RAND Journal of Economics*, pp. 242–262.
Berry, Steven, James Levinsohn, and Ariel Pakes (1995). "Automobile prices in market equilibrium". In: *Econometrica: Journal of the Econometric Society*, pp. 841–890.