Demand for Differentiated Products

Problem Set 1
Empirical Industrial Organization

Zixuan, Anoosheh, Shuo

November 2, 2024

1 Data

The dataset is a panel of aggregate (model level) car sales along with the characteristics of each model. The data spans 1977-1981 in the U.S., giving rise to 5 markets in total. We divide the quantity sold `q` by the market size (the number of households `hh` in that year) to get the market share for each model. We also adjust the car price `p` by the consumer price index `cpi` to get the real price.

$$p_{adj} = \frac{100 \times p}{cpi}$$

We rescale the price by $p/1000$ to make the coefficient on price more readable. An overview of the mean and standard deviation of the variables is shown in Table 1.

	mean	sd
dpm	0.06	0.02
door3	0.08	0.27
door4	0.52	0.50
door5	0.05	0.22
at	0.28	0.45
ps	0.38	0.49
air	0.14	0.35
drv	0.21	0.40
wt	2887.27	702.02
hp2wt	0.36	0.08
hp	104.19	34.42
euro	0.27	0.44
japan	0.14	0.34
size	1.32	0.24
wb	104.47	9.14
p	10.89	7.82

Table 1: Descriptive statistics

2 Logit Model

2.1 Utility function

The standard utility function takes the following linear in characteristics form:

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

where i represents individual and j represents the product. X_j includes the following variables: `dpm, door3, door4, door5, at, ps, air, drv, wt, hp2wt, hp, euro, japna, size, wb,`
We assume that ϵ_{ij} is distributed type I extreme value and is independent across all products j. Thus, the market share for each product is given by (by normalizing the utility of the outside good to zero)

$$s_j = \frac{e^{u_{ij}}}{1 + \sum_{k=1}^{J} e^{u_{ik}}}$$

Thus,

$$\log s_j - \log s_0 = \beta_0 + \beta_1 x_{j1} + \cdots + p_j \alpha + \xi_j \tag{2}$$

where ξ_j is the fixed effect/unobserved heterogeneity of each product j. For the remaining section, we assume that all characteristics except price are exogenous.

2.2 Estimation

The OLS estimation is unbiased and consistent if we assume that ξ_j is uncorrelated with p. We perform two OLS estimations based on equation (2) with and without brand fixed effect. The results are shown in Table 2. However, it is generally not the case. Price is generally correlated with unobserved characteristics. We construct instruments based on Berry et al. (1995). That is, for each product characteristic x^k, we compute

- $z_j^k = \sum_{k \in \mathcal{F}_j, j' \neq j} x_{j'}^k$, the sum of characteristics of all other products belonging to the same firm.
- $z_j^k = \sum_{k \notin \mathcal{F}_j} x_{j'}^k$, the sum of characteristics of all other products belonging to rival firms.

This is a measure of the competitive environment surrounding firms, capturing the substitutability of other available cars to any particular model.
Since we have many x^k, we have a large number of instruments. Following Berry et al. (1995), we selected the following to instrument price p: `const_*, dpm_*, hp2wt_*, size_*, air_*`.
The IV estimation results are shown in Table 2.

2.3 Results

We observe that the coefficient on price, α, is larger (in absolute terms) but noisily estimated using IV compared to OLS. This makes sense as the simple OLS regression yields downward-biased estimates (in terms of magnitude) because failing to account for endogeneity underestimates the effect of price on market share.

Dependent Variable:	log(mktshr)-log(shr_0)			
Estimator	OLS		IV	
Model:	(1)	(2)	(3)	(4)
Variables				
Constant	-5.238***		-4.616**	
	(1.956)		(2.312)	
dpm	-22.04***	-19.04***	-20.39***	-13.19**
	(4.601)	(3.929)	(5.260)	(5.922)
door3TRUE	-0.6235***	-0.5271***	-0.6199***	-0.5421***
	(0.2038)	(0.1876)	(0.2089)	(0.1964)
door4TRUE	-0.0636	-0.0190	-0.0431	0.0388
	(0.1493)	(0.1344)	(0.1528)	(0.1418)
door5TRUE	0.3424	0.4548**	0.4005	0.5102**
	(0.3062)	(0.2010)	(0.3106)	(0.2059)
at	-0.2868	-0.2917	-0.1360	-0.2736
	(0.2102)	(0.1902)	(0.2353)	(0.2042)
ps	0.2057	0.2052	0.2012	0.1861
	(0.1738)	(0.1684)	(0.1782)	(0.1880)
air	0.1226	-0.1486	0.4378	0.4374
	(0.1797)	(0.1689)	(0.3030)	(0.3578)
drv	0.1029	0.1621	0.0569	0.2813*
	(0.1467)	(0.1363)	(0.1646)	(0.1500)
wt	-0.0022***	-0.0011**	-0.0019***	-0.0012**
	(0.0005)	(0.0004)	(0.0006)	(0.0006)
hp2wt	-9.821***	-5.934**	-9.339**	-8.330**
	(3.242)	(2.356)	(4.171)	(3.388)
hp	0.0381***	0.0228**	0.0415***	0.0385***
	(0.0112)	(0.0088)	(0.0149)	(0.0139)
euro	-1.772***		-1.250***	
	(0.2136)		(0.4288)	
japan	-0.1290		-0.0259	
	(0.2246)		(0.2441)	
size	2.710***	0.6433	2.584***	0.7710
	(0.9020)	(0.9060)	(0.9831)	(1.142)
wb	0.0224	0.0431*	0.0095	0.0203
	(0.0234)	(0.0254)	(0.0277)	(0.0311)
p	-0.0396***	-0.0644***	-0.0974**	-0.1770***
	(0.0126)	(0.0179)	(0.0448)	(0.0605)
Fixed-effects				
firmids		Yes		Yes
Fit statistics				
Observations	501	501	501	501
Adjusted R^2	0.63354	0.72408	0.60557	0.68596
Wald (1st stage), p			3.5472	5.8321
Wald (1st stage), p-value, p			0.00015	2.69×10^{-8}
Sargan			63.992	5.9720
Sargan, p-value			2.27×10^{-10}	0.74272

Clustered (modelid) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 2: Logit estimation results

3 Nested Logit Model

3.1 Utility function

We relax the assumption that ε_{ij} is independent across j. Instead, we group the products into nests based on their size. The 3 nests are `compact, midsize, large` plus one outside option. Now the utility function is given by

$$\begin{aligned} u_{ij} &= X_j\beta + \alpha p_j + \xi_j + \eta_{ig} + \epsilon_{ij} \\ &= \delta_j + \eta_{ig} + (1-\sigma)\epsilon_{ij} \end{aligned} \tag{3}$$

Following the derivation in Berry (1994), the estimation equation is

$$\log(s_j) - \log(s_0) = x_j\beta + p_j\alpha + \sigma\log(\bar{s}_{j|g}) + \xi_j \tag{4}$$

3.2 Estimation

In addition to p, $\log(\bar{s}_{j|g})$ is also endogenous. We follow Berry (1994) to construct additional instruments for $\log(\bar{s}_{j|g})$, which is the sum of product characteristics of the rival firms in the same nest g. We exploit the competition posed by rivals with the added feature of this 'competition' coming from within the same group (i.e. size category). The results from demand estimation are presented in Table 3.

3.3 Results

Price α Once again, the OLS estimates of α are biased towards zero. The OLS estimation is upward biased.

Within nest substitutability $\hat{\sigma}$ The estimate of σ is smaller in IV than in OLS, suggesting a lower degree of intra-group substitutability once we have accounted for the endogeneity of prices and within-group shares.

Instrument test The F-test of the first stage is a weak instrument test. The null hypothesis is that the instruments are weak, which is rejected by the test, implying that our instruments are not too weak. The Sargan-Hansen test is a test of over-identification. The null hypothesis is that all instruments are valid, which is rejected by the test, implying that some instruments are not suitable.
From the results of the test of overidentifying restrictions, we are able to reject the null that *all* instruments are valid. Our instruments also appear to be weak as indicated by the conditional likelihood ratio (CLR) test for weak IV-robustness (for an over-identified system). We cannot reject the joint null of $H_0 : \alpha = 0,\ \sigma = 0$.
This could point to a possible failure of the exogeneity of the instruments we use.

Dependent Variable:	log(mktshr)-log(shr_0)			
Estimator	OLS		IV	
Model:	(1)	(2)	(3)	(4)
Variables				
Constant	-4.712***		-4.467***	
	(0.8660)		(1.277)	
dpm	-8.137***	-9.190***	-11.71***	-17.32***
	(1.616)	(1.561)	(2.923)	(6.254)
door3TRUE	-0.0134	0.0662	-0.2170*	-0.6752**
	(0.0658)	(0.0602)	(0.1128)	(0.2799)
door4TRUE	-0.0311	0.0031	-0.0281	0.0197
	(0.0569)	(0.0511)	(0.0778)	(0.1636)
door5TRUE	0.1019	0.1758**	0.2226	0.5586**
	(0.0988)	(0.0717)	(0.1591)	(0.2699)
at	0.0367	-0.0741	0.0300	-0.3274
	(0.0969)	(0.1086)	(0.1258)	(0.2398)
ps	-0.1059	-0.0123	-0.0038	0.2381
	(0.0983)	(0.1127)	(0.1124)	(0.2113)
air	0.0162	-0.0004	0.2665*	0.2584
	(0.0748)	(0.0836)	(0.1496)	(0.3645)
drv	0.1261**	0.0432	0.0870	0.2789
	(0.0498)	(0.0515)	(0.0834)	(0.1836)
wt	-0.0003	-5.32×10^{-5}	-0.0008*	-0.0014*
	(0.0003)	(0.0002)	(0.0005)	(0.0008)
hp2wt	-1.955	-1.307	-4.284	-8.799**
	(1.468)	(1.202)	(2.767)	(4.100)
hp	0.0104**	0.0066	0.0220**	0.0385**
	(0.0052)	(0.0046)	(0.0093)	(0.0164)
euro	-0.1817*		-0.3641	
	(0.0977)		(0.2763)	
japan	0.0878		0.0847	
	(0.0836)		(0.1283)	
size	1.691***	1.059**	1.949***	0.6279
	(0.4538)	(0.4710)	(0.6055)	(1.195)
wb	0.0003	0.0107	-0.0010	0.0335
	(0.0111)	(0.0112)	(0.0151)	(0.0345)
p	-0.0165***	-0.0142	-0.0636***	-0.1616**
	(0.0058)	(0.0095)	(0.0230)	(0.0686)
log(mktshr_g)	0.8896***	0.8759***	0.5891***	-0.2019
	(0.0256)	(0.0285)	(0.0856)	(0.2678)
Fixed-effects				
firmids		Yes		Yes
Fit statistics				
Observations	501	501	501	501
Adjusted R^2	0.93445	0.94171	0.88708	0.58920
F-test (1st stage), p			4.8290	6.9594
F-test (1st stage), log(mktshr_g)			9.1389	4.3604
F-test (1st stage), p-value, p			1.64×10^{-7}	1.12×10^{-11}
F-test (1st stage), p-value, log(mktshr_g)			6.51×10^{-16}	1.31×10^{-6}
Sargan			146.28	16.304
Sargan, p-value			2.17×10^{-26}	0.09124

Clustered (modelid) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 3: Nested Logit estimation results

4 Random Coefficients Model

4.1 Utility function

Next, we turn to our final specification of a random-coefficients model à la Berry et al. (1995) where indirect utility is defined as

$$u_{ij} = X_j\beta + \alpha p_j + \xi_j + \eta_i \text{size}_j + \epsilon_{ij} = \delta_j + \eta_i \text{size}_j + \epsilon_{ij} \tag{5}$$

where $\eta_i \sim N(0, \sigma^2)$. The difference here is that the coefficient η_i is allowed to vary across households i. After refactoring the indirect utility, we arrive at the following expression, where the last line arises since we have learned that $\bar{\eta}$ is zero. So that, even if the ϵ_{ij}'s are still i.i.d. TIEV across j, the composite error is not.
Now, the aggregate market share becomes

$$s_j = \int \frac{\exp\left(\delta_j + \eta_i \text{Size}_j\right)}{1 + \sum_{j'}^{J} \exp\left(\delta_{j'} + \eta_i \text{Size}_{j'}\right)} dF(\eta_i)$$

4.2 Estimation

Unlike before, the relationship between $\log(s_j)$ and δ_j is no longer as explicit. We need to find a way to recover δ_j from the observed market share s_j. We follow the method provided by Berry et al. (1995). The procedure involves several steps, including drawing values from a normal distribution, contraction mapping, and GMM estimation.
The procedure is the following:

1. We draw 500 values from a normal distribution $N(0, 1)$.

2. We pick an initial value of σ.

3. Given a value of σ, we calculate the predicted market share based on σ by approximate the following integral

$$s_j(\delta_1, \ldots, \delta_j) = \int \frac{e^{\delta_j + \eta_i \text{size}_j}}{1 + \sum_{k=1}^{J} e^{\delta_k + \eta_i \text{size}_k}} dF(\eta_i)$$

4. We use contraction mapping to find the value of δ vector that makes the predicted market share equal to the observed market share.

5. Once we have the δ vector which is linear in X.

$$\delta_j = X_j\beta + \alpha p_j + \xi_j$$

   We can estimate the β and α with some instruments, thus getting the residual ξ_j.

6. We compute the empirical moment condition is $Q(\sigma) = Z'\xi$ where Z is the instruments and ξ is the residual, which is of dimension #instr $\times$ 1. The GMM objective function is thus $Q(\sigma)'WQ(\sigma)$.
$$\min_{\sigma} Q(\sigma)'WQ(\sigma)$$

Note that by creating nests, we create a new endogenous variable $\log(\bar{s}_{j|g})$ for which we create new instruments. Similarly, by creating a random coefficient term η_i on size, we create extra endogeneity. (How so?) We follow Gandhi and Houde (2019) to construct extra differentiation instrument.

1. Quadratic: $\sum_{j' \neq j} (d^k_{j'j})^2$ where $d^k_{j'j}$ is the distance between product j and j' in characteristic k.

2. Local: $\sum_{j' \neq j} 1\{d^k_{j'j} < sd^k\}$ where sd^k is the standard deviation of x^k acorss all markets.

We select the following instruments in addition to the previous nested logit instruments: `size_local, wb_local, hp_local, hp2wt_local`.

4.3 Results

We plot the GMM objective function w.r.t. σ in Figure 1.
The sigma that minimizes the GMM objective function is 2.49. Given the $\hat{\sigma}$, we recover the mean utility δ_j for each product in every market. The estimation results by regressing δ_j on X_j and p_j are shown in Table 4.

5 Markups and Merger Simulation

5.1 Price setting equation

The profit of a multiproduct firm is

$$\Pi_f = \sum_{j \in \mathcal{J}_f} (p_j - c_j) s_j M$$

since $s_j M$ represents the sales of product j. The optimal price is given by taking the first order derivative with respect to p:

$$s_j + \sum_{k \in \mathcal{J}_f} \frac{\partial s_k}{\partial p_j} (p_k - c_k) = 0 \quad \forall j \in \mathcal{J}_f$$

The first term is the direct impact of changing p_j. The second term is the impact on the profits of other products of the firm. For firm f, the pricing equation written in matrix form gives:

$$S^f + \Omega^f (P^f - C^f) = 0$$

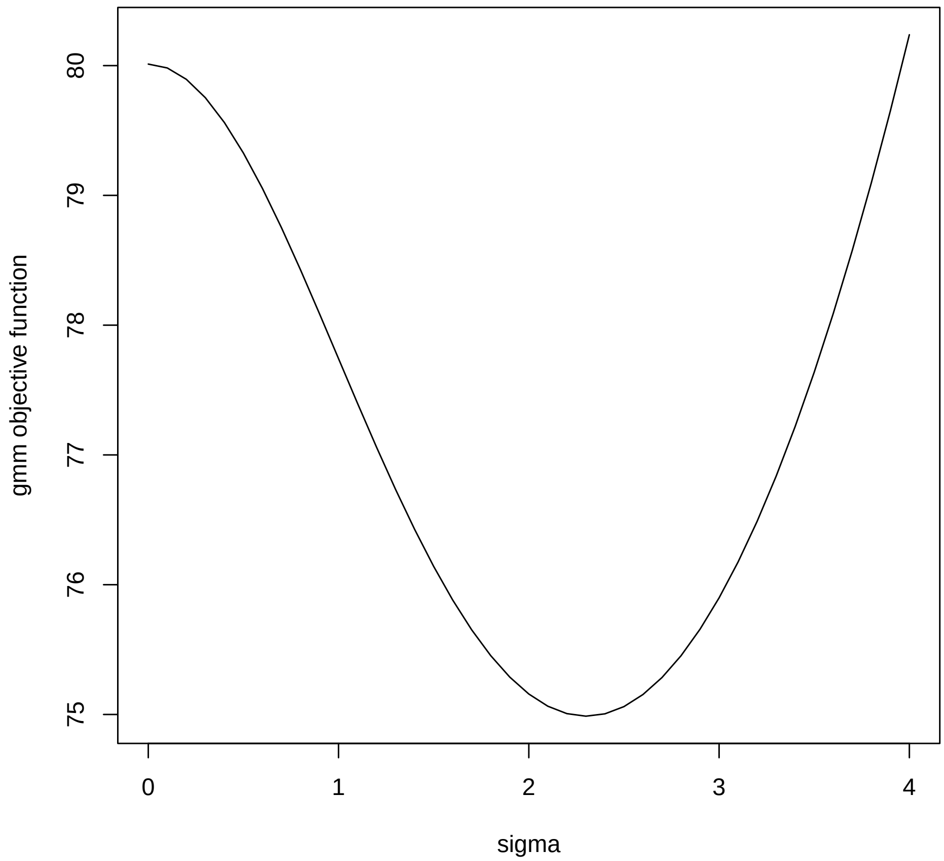

Figure 1: GMM objective function w.r.t. σ

where Ω^f is the matrix of cross-price elasticities. For example, if firm f has two products, the matrix is

$$\Omega^f = \begin{bmatrix} \frac{\partial s_1}{\partial p_1} & \frac{\partial s_1}{\partial p_2} & 0 \\ \frac{\partial s_2}{\partial p_1} & \frac{\partial s_2}{\partial p_2} & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Therefore, aggregate the pricing equation for all firms, we have

$$S + \Omega \odot O(P - C) = 0$$

where O is the ownership matrix. We take $\hat{\alpha} = -0.098$. We estimate marginal cost (negative), markup, and elasticity for each year. We simulate a merger between firm 17 and firm 7 by modifying the ownership matrix. The mean results are shown in Table 5.
The anomaly is that we get negative marginal cost, this is due to the small $\hat{\alpha}$ from the estimation. Also, the effect of average price seems to be very insignificant.

Dependent Variable:	y
Model:	(1)
Variables	
Constant	-1.741
	(1.938)
p	-0.0870**
	(0.0428)
dpm	-20.46***
	(5.398)
door3TRUE	-0.6440***
	(0.2082)
door4TRUE	-0.0094
	(0.1522)
door5TRUE	0.4822
	(0.3175)
at	-0.2956
	(0.2242)
ps	0.2206
	(0.1728)
air	0.5580*
	(0.2874)
drv	0.0961
	(0.1538)
wt	-0.0023***
	(0.0006)
hp2wt	-10.12***
	(3.726)
hp	0.0421***
	(0.0134)
euro	-1.212***
	(0.4153)
japan	0.0573
	(0.2280)
wb	-0.0207
	(0.0225)
Fit statistics	
Observations	501
Adjusted R^2	0.69806
F-test (1st stage), p	4.6944
F-test (1st stage), p-value, p	2.99×10^{-7}
Sargan	72.714
Sargan, p-value	3.71×10^{-11}

Clustered (modelid) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table 4: Random coefficient on size

	year	p	mc	mk	elas	p_new
1	1977	9.89	-0.55	10.44	-102.98	9.89
2	1978	10.67	0.23	10.44	-110.88	10.67
3	1979	10.50	0.11	10.39	-108.74	10.50
4	1980	10.71	0.34	10.36	-110.69	10.71
5	1981	12.42	2.09	10.33	-127.88	12.42

Table 5: Markup

5.2 Estimation

We calculate consumer surplus pre- and post-merger in monetary terms:

$$CS_i = E(\max_j u_{ij}) = \frac{\ln(1 + \sum_j \exp(\delta_j + \mu_{ij}))}{\alpha_i}$$

Thus in total:

$$CS = M \int CS_i(\theta) dF(\theta)$$

The result is shown in Table 6.

	year	nb_hh	shr_0	shr_0_new	cs_pre	cs_post	diff
1	1977	74142	0.88	0.91	100444.81	68026.70	-32418.12
2	1978	76030	0.88	0.91	97404.05	76085.12	-21318.92
3	1979	77330	0.89	0.92	88091.89	64588.86	-23503.03
4	1980	80776	0.91	0.94	78552.02	52631.03	-25920.99
5	1981	82368	0.91	0.93	76192.11	58267.48	-17924.63

Table 6: Consumer surplus

References

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, 63(4):841–890.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, pages 242–262.

Gandhi, A. and Houde, J.-F. (2019). Measuring substitution patterns in differentiated-products industries. *NBER Working paper*, (w26375).