Empirical IO PS1

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., give 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.

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

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 **??**.

	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

2 Question 1:Logit

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

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

Thus,

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

where ξ_j 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 estimation based on equation **??** with and without brand fixed effect. The results are shown in Table **??**. However, it is generally not the case. Price is generally correlated with unobserved characteristics. We construct instruments based on (**?**). That is, for each product characteristics 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 belong to the same firm.
- $z_j^k = \sum_{k \notin \mathcal{F}_j} x_{j'}^k$, the sum of characteristics of all other products belong to the rival firms.

Since we have many x^k, we have a large number of instruments. Following (**?**), we selected the following to instrument price p:
`const_*, dpm_*, hp2wt_*,size_*, air_*` The iv estimation results are shown in Table **??**.

2.3 Results

3 Question 2:Nested logit

3.1 Utility function

We relax the assumption that ε_{ij} are 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

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

Now 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 now $\log(\bar{s}_{j|g})$ is also endogenous. We follow (**?**) 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. The estimation results are shown in Table **??**.

3.3 Results

The F-test of the first stage is commonly named weak instrument test. The null hypothesis is that the instruments are weak, which is rejected by the test. 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.

$\hat{\alpha}$ We see that in both cases, the IV estimation gives a larger estimate of the price coefficient, which indicates that price is positively correlated with the unobserved characteristics. The OLS estimation is upward biased.

$\hat{\sigma}$ The estimate of σ is smaller in IV that in OLS, implying that the group share $\log(\bar{s}_{j|g})$ positively correlates with ξ_j. The OLS estimation is also upward biased.

4 Question 3

4.1 Utility function

We now consider a random coefficient model. The utility function is given by

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

where $eta_i \sim N(0, \sigma^2)$.

4.2 Estimation

Unlike before, the relationship between $log(s_j)$ and δ_j is no longer as explicit as before ($\delta_j = \log(s_j) - \log(s_0)$). We need to find a way to recover δ_j from the observed market share s_j. We follow the method provided by (**?**).

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

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

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 (**?**) 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, disp_local, wb_local, hp_local, hp2wt_local`

4.3 Results

We plot the GMM objective function w.r.t. σ in Figure **??**. The estimation results by regressing δ_j on X_j and p_j are shown in Table **??**.

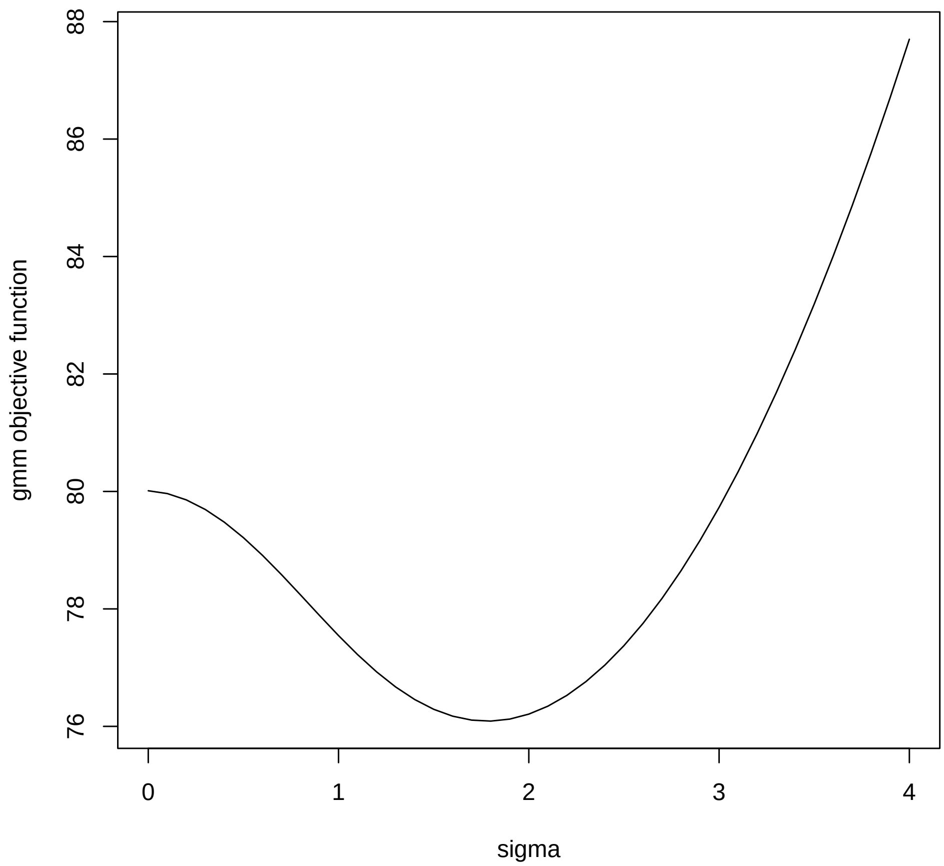

Figure 1: GMM objective function w.r.t. σ

5 Question 4+5 : Markups, 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 the product j.

The optimal price is given by taking the first order derivative wrt 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 the matrix form gives,

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

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.

5.2 Estimation

We take $\hat{\alpha} = -0.0785$. We estimate marginal cost (negative), markup and elasticiy for each year. We simulate a merger between firm17 and firm7, by modifying the ownership matrix. The mean results are shown in Table **??**. Table **??** shows the results. 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.

We calculate consumer surplus pre and post merger in monetary terms,

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

Notice that

$$s_0 = \frac{1}{1 + \sum_j \exp \delta_j + \mu_{ij}}$$

Thus

$$CS_i = \frac{-\ln(s_0)}{\alpha_i}$$

In total,

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

The result is shown in Table **??**

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 1: Logit estimation results

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), $\log(\text{mktshr}_g)$			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, $\log(\text{mktshr}_g)$			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 2: Nested Logit estimation results

Dependent Variable:	y
Model:	(1)
Variables	
Constant	-2.136
	(1.927)
p	-0.0812*
	(0.0420)
dpm	-20.06***
	(5.345)
door3TRUE	-0.6427***
	(0.2076)
door4TRUE	-0.0069
	(0.1524)
door5TRUE	0.4570
	(0.3128)
at	-0.3181
	(0.2241)
ps	0.2280
	(0.1719)
air	0.5108*
	(0.2803)
drv	0.0804
	(0.1527)
wt	-0.0022***
	(0.0006)
hp2wt	-9.957***
	(3.669)
hp	0.0411***
	(0.0131)
euro	-1.293***
	(0.4126)
japan	0.0120
	(0.2291)
wb	-0.0138
	(0.0222)
Fit statistics	
Observations	501
Adjusted R^2	0.68233
F-test (1st stage), p	4.6944
F-test (1st stage), p-value, p	2.99×10^{-7}
Sargan	73.662
Sargan, p-value	2.45×10^{-11}

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

Table 3: Random coefficient on size

	year	p	mc	mk	elas	p_new
1	1977	9.89	-2.73	12.62	-124.46	9.89
2	1978	10.67	-1.95	12.61	-134.01	10.67
3	1979	10.50	-2.06	12.56	-131.42	10.50
4	1980	10.71	-1.82	12.53	-133.78	10.71
5	1981	12.42	-0.06	12.48	-154.55	12.42

	year	nb_hh	shr_0	shr_0_new	cs_pre	cs_post	diff
1	1977	74142	0.88	0.92	121396.89	79581.15	-41815.75
2	1978	76030	0.88	0.91	117721.84	90084.00	-27637.85
3	1979	77330	0.89	0.92	106467.24	76718.77	-29748.47
4	1980	80776	0.91	0.94	94937.42	63350.37	-31587.04
5	1981	82368	0.91	0.93	92085.25	69479.11	-22606.13