Tutorial 4: Logit & Nested-logit demand estimation with a simple counterfactual analysis

Yiran Hao

Nov.19. 2018

Outline:

1. *Manually generate market shares & within group market shares*
2. *Run logit & nested-logit estimation*
3. *Manually calculate marginal cost & price-cost margins under Bertrand competition*
4. *Automatically generate marginal cost & price-cost margins*
5. *A simple counterfactual example of a hypothetical merger of two firms*

Step 0: Loading the dataset ***verboven_cars.dta***

Step 1: Preparation for demand estimation: manually generate market shares

1) Generate new variables***gen logq = ln(qu)***

   gen logp = ln(eurpr)
   gen logpop = ln(pop)
   gen loggdp = ln(ngdp)
   gen msize = pop/4

2) Construct market share s_j
   gen share = qu/msize
   egen sum_share = sum(share), by(ma ye)

3) Construct outside good's market share s_0

   gen share0 = 1 - sum_share

   sum share share0

4) Generate log odd ratio (all above have done in tutorial 3)

   gen lsj_ls0 = ln(share/share0)

5) Manually generate market shares & within group market shares

Groups are defined as whether the car is a domestic or foreign car, which is denoted as variable "home" in the dataset. This is a simpler version of Nested -Logit of Bjornerstedt&Verboven Stata J.pdf.

sum home

```
sum home
```

Variable	Obs	Mean	Std. Dev.	Min	Max
home	11,549	.1878085	.3905761	0	1

If home=1, it means the car is domestic; home=0 represents a foreign car.

The following code is generating the sum of all domestic products' market shares in a given year and country and the sum of all foreign products' market shares in a given year and country.

bys ma ye home: egen denom= total(share)

For a given product, the percentage it accounts for among all domestic/foreign cars is the within-group market share:

gen s_within = share/denom

Then we generate its logarithm form.

gen ln_s_within=ln(s_within)

we generate the denominator of within-group market shares: for a given country and year, the summation of all products' market shares should contain at most 2 values: one for domestic group; the other for foreign group. We can check by using the following commands:

bys ye ma: tab denom

-> ye = 70, ma = Belgium

denom	Freq.	Percent	Cum.
.1079466	65	100.00	100.00
Total	65	100.00	

-> ye = 70, ma = France

denom	Freq.	Percent	Cum.
.0162458	36	66.67	66.67
.0804139	18	33.33	100.00
Total	54	100.00	

-> ye = 70, ma = Germany

denom	Freq.	Percent	Cum.
.0620833	37	66.07	66.07
.0646751	19	33.93	100.00
Total	56	100.00	

Step 2: Logit & Nested-logit estimation

1) Logit estimation including model attributes:

reghdfe lsj_ls0 logp hp li wi cy le he logpop loggdp, vce(robust) a(ma ye brd)

```
HDFE Linear regression                            Number of obs   =     11,549
Absorbing 3 HDFE groups                           F(   9,  11467) =     341.86
Statistics robust to heteroskedasticity           Prob > F        =     0.0000
                                                  R-squared       =     0.4018
                                                  Adj R-squared   =     0.3975
                                                  Within R-sq.    =     0.2203
                                                  Root MSE        =     1.1656
```

lsj_ls0	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
logp	-1.164701	.1038432	-11.22	0.000	-1.368251	-.9611507
hp	-.0137051	.0017671	-7.76	0.000	-.0171689	-.0102413
li	-.0415553	.0136634	-3.04	0.002	-.0683379	-.0147726
wi	.0638831	.0033199	19.24	0.000	.0573755	.0703907
cy	-.0006896	.000084	-8.21	0.000	-.0008542	-.000525
le	-.0000936	.0007787	-0.12	0.904	-.00162	.0014327
he	-.017626	.0030341	-5.81	0.000	-.0235734	-.0116785
logpop	.349325	.2300971	1.52	0.129	-.1017048	.8003547
loggdp	.2826532	.062891	4.49	0.000	.1593761	.4059302

2) Nested-logit estimation including model attributes:

By simply adding one more explanatory variable: the logarithm of within-group market shares into the regression, the nested-logit model is estimated.

reghdfe lsj_ls0 logp ln_s_within hp li wi cy le he logpop loggdp, vce(robust) a(ma ye brd)

```
HDFE Linear regression                              Number of obs   =     11,549
Absorbing 3 HDFE groups                             F(  10,   11466) =   17778.91
Statistics robust to heteroskedasticity             Prob > F        =     0.0000
                                                    R-squared       =     0.9524
                                                    Adj R-squared   =     0.9521
                                                    Within R-sq.    =     0.9380
                                                    Root MSE        =     0.3287
```

lsj_ls0	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
logp	.0347379	.028136	1.23	0.217	-.0204135	.0898893
ln_s_within	1.059716	.0030231	350.54	0.000	1.05379	1.065641
hp	.0022768	.0004699	4.85	0.000	.0013557	.0031979
li	.0005364	.0042338	0.13	0.899	-.0077626	.0088354
wi	-.0042111	.0009483	-4.44	0.000	-.0060699	-.0023523
cy	7.53e-06	.0000222	0.34	0.734	-.0000359	.000051
le	.0002348	.0002179	1.08	0.281	-.0001922	.0006619
he	.0009742	.0009389	1.04	0.299	-.0008661	.0028145
logpop	-.162994	.0666984	-2.44	0.015	-.2937343	-.0322537
loggdp	.4837022	.0182389	26.52	0.000	.4479509	.5194535

Comments: Nested-logit is not valid here since the coefficient of within-group market shares is larger than 1. We could reconsider the nests by setting groups as cars' sizes: compact, suv, mpv, etc. then setting subgroups as domestic & foreign cars. This type of nests corresponds to the paper: Bjornerstedt&Verboven Stata J.pdf.

Step 3: Manually calculate marginal cost & price-cost margins under Bertrand competition

1) Rerun the Logit model

Since Nested logit does not perform well here, we use standard logit to calculate marginal cost and price-cost margins.Before calculating the price-cost margin, we need to rerun the standard logit estimation to get the coefficients in Stata's memory.

reghdfe lsj_ls0 logp hp li wi cy le he logpop loggdp, vce(robust) a(ma ye brd)

2) Generating margins

The coefficient of price is saved in Stata's memory as _b[logp]. We can calculate the price-cost margin as follows:

gen margin=-eurpr/(_b[logp]*(1-share))

Here the margin is measured in monetary units, in euros.

sum margin

```
sum margin
```

Variable	Obs	Mean	Std. Dev.	Min	Max
margin	11,549	7181.547	4760.113	732.3913	42572.36

3) Generating marginal revenue

Now, we can calculate the marginal revenue by subtracting the margin from price.

gen mr= eurpr-margin
gen log_mr= ln(mr)

```
sum mr
```

Variable	Obs	Mean	Std. Dev.	Min	Max
mr	11,549	1170.972	781.0226	98.27631	7008.289

4) Estimating marginal costs

By running an OLS regression on marginal revenue using the quantities and product's attributes as explanatory variables, we could estimate marginal cost as follows:

reghdfe log_mr logq hp li wi cy le he logpop loggdp, vce(robust) a(ma ye brd)

```
HDFE Linear regression                          Number of obs   =     11,549
Absorbing 3 HDFE groups                         F(   9,  11467) =    8552.44
Statistics robust to heteroskedasticity         Prob > F        =     0.0000
                                                R-squared       =     0.9724
                                                Adj R-squared   =     0.9723
                                                Within R-sq.    =     0.8860
                                                Root MSE        =     0.1100
```

log_mr	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
logq	-.0188848	.0009599	-19.67	0.000	-.0207663	-.0170032
hp	.0084214	.0001798	46.84	0.000	.008069	.0087738
li	.0001027	.0013618	0.08	0.940	-.0025666	.002772
wi	.0034008	.00034	10.00	0.000	.0027343	.0040674
cy	.0000968	8.75e-06	11.06	0.000	.0000796	.0001139
le	.0022009	.0000757	29.08	0.000	.0020526	.0023493
he	-.0033937	.0002729	-12.44	0.000	-.0039286	-.0028587
logpop	.4230247	.0200337	21.12	0.000	.3837551	.4622942
loggdp	.1295281	.0050102	25.85	0.000	.1197073	.139349

There is an endogeneity problem, i.e., logq is negatively correlated with the unobserved component of the cost. As usual, we need to find IV for logq. However, to make our life easier, we could impose the assumption of constant marginal costs as follows:

reghdfe log_mr logq hp li wi cy le he logpop loggdp, vce(robust) a(M=ma Y=ye B=brd)

```
HDFE Linear regression                          Number of obs   =     11,549
Absorbing 3 HDFE groups                         F(   8,  11468) =    8855.84
Statistics robust to heteroskedasticity         Prob > F        =     0.0000
                                                R-squared       =     0.9713
                                                Adj R-squared   =     0.9711
                                                Within R-sq.    =     0.8814
                                                Root MSE        =     0.1122
```

log_mr	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
hp	.0088716	.0001872	47.38	0.000	.0085045	.0092386
li	.0008985	.0014203	0.63	0.527	-.0018855	.0036824
wi	.0022548	.0003321	6.79	0.000	.0016039	.0029058
cy	.0001122	9.38e-06	11.96	0.000	.0000938	.0001306
le	.0022507	.0000774	29.09	0.000	.0020991	.0024024
he	-.003132	.0002753	-11.38	0.000	-.0036715	-.0025924
logpop	.406027	.02019	20.11	0.000	.3664511	.4456028
loggdp	.1272735	.0051288	24.82	0.000	.1172202	.1373267

predict log_mc_hat,xbd

gen mc= exp(log_mc_hat)

M,Y, and B save these three types of FEs into dataset. We predict marginal cost by including all FEs using “xbd”. Then marginal cost is generated by taking the exponential form of the prediction.

```
. sum mc
```

Variable	Obs	Mean	Std. Dev.	Min	Max
mc	11,549	1161.811	751.4056	126.116	7634.299

5) Taking the average marginal cost for each firm cross years

bys frm : egen mc_frm=mean(mc)

or we could take the average marginal cost for each firm in a given year:

bys frm ye: egen mc_frm=mean(mc)

Step 4: Automatically generate marginal cost & price-cost margins

Note: firstly, here, step 1 (1) is the only necessary part for generating marginal cost & price-cost margins automatically. Secondly, which is also most important, using command “mergersim” gives a different estimate of margins, mr, and mc is that the command takes into account that these firms have multiproduct, and therefore the expression for the f.o.c is not same as the previous mc manual calculation. The previous estimation, by contrast, is assuming each firm has only one product. Additionally, “mergersim” sets price in levels as an explanatory variable rather than the logarithm form that we use in the previous manual estimation.

1) Set the dataset as 2 dimensions of the panel data: the car model and the market (year &

country)

egen yect=group(ye ma),label
xtset co yect

2) Initializing the merger settings

If you want to use nested-logit, do the following. The code will generate a group share variable "M_lsjg" into the dataset if setting the nests. The market shares are also generated automatically , which is denoted as M_ls. There is no need to manually generate market shares.

mergersim init,nests(home) price(pr) quantity(qu) marketsize(msize) firm(frm)

```
MERGERSIM: Merger Simulation Program
Version 1.0, Revision: 218
------------------------------------------------------------------------
Unit demand one-level nested logit

                  Depvar              Price               Group shares
                  ------------------------------------------------------
                  M_ls                eurpr               M_lsjg
------------------------------------------------------------------------
Variables generated: M_ls M_lsjg
```

If using standard logit, we could use following code:

mergersim init, price(eurpr) quantity(qu) marketsize(msize) firm(frm)

```
. mergersim init, price(eurpr) quantity(qu) marketsize(msize) firm(frm)
------------------------------------------------------------------------
MERGERSIM: Merger Simulation Program
Version 1.0, Revision: 218
------------------------------------------------------------------------
Unit demand unnested logit

                  Depvar              Price               Group shares
                  ------------------------------------------------------
                  M_ls                eurpr
------------------------------------------------------------------------
Variables generated: M_ls
```

3) Set the dependent variable in demand estimation as M_ls.

Stata needs to know which regression you would like to use so you need to rerun the regression and the dependent variable should be same as the dependent variable that is shown in the above table. Note: here price doesn't take the logarithm form since it should be consistent with the setting of part (2)

reghdfe M_ls eur pr hp li wi cy le he logpop loggdp, vce(robust) a(ma ye brd)

```
HDFE Linear regression                          Number of obs   =     11,549
Absorbing 3 HDFE groups                         F(   9,  11467) =     343.09
Statistics robust to heteroskedasticity         Prob > F        =     0.0000
                                                R-squared       =     0.3975
                                                Adj R-squared   =     0.3932
                                                Within R-sq.    =     0.2148
                                                Root MSE        =     1.1697
```

M_ls	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
eurpr	-.0000432	5.55e-06	-7.79	0.000	-.0000541	-.0000323
hp	-.0162624	.0019841	-8.20	0.000	-.0201516	-.0123731
li	-.0506472	.0138631	-3.65	0.000	-.0778213	-.0234732
wi	.0572194	.0033083	17.30	0.000	.0507346	.0637042
cy	-.0008931	.0000869	-10.28	0.000	-.0010633	-.0007228
le	-.0015961	.0007632	-2.09	0.037	-.0030922	-.0001001
he	-.0138003	.0030159	-4.58	0.000	-.0197121	-.0078885
logpop	.0189387	.226602	0.08	0.933	-.42524	.4631174
loggdp	.1801754	.0621139	2.90	0.004	.0584215	.3019292

4) Show firm's marginal cost & pre-merger lerner

If you want to look at average mc for each firm cross years:

mergersim market

If you want to see the average mc for each firm in a given year, for example, 1983:

mergersim market if ye == 83

Pre-merger Market Conditions
Unweighted averages by frm

firm code	eurpr	Marginal costs	Pre-merger Lerner
AlfaRomeo	8385.140	-1.48e+04	2.957
BMW	8608.927	-1.46e+04	2.815
Toyota	3771.713	-1.94e+04	6.256
Fiat	5392.861	-1.83e+04	5.000
Ford	6221.833	-1.73e+04	4.030
Honda	6140.053	-1.70e+04	3.957
Hyundai	3610.902	-1.95e+04	6.410
DeTomaso	4275.521	-1.89e+04	5.418
Mazda	6419.721	-1.68e+04	3.861

Step 5: A simple counterfactual example of a hypothetical merger of two firms

1) check market & firm's label

d

variable name	storage type	display format	value label	variable label
ye	byte	%9.0g		year (=first dimension of panel)
ma	byte	%9.0g	market	market (=second dimension of panel)
co	int	%9.0g		model code (=third dimension of panel)
zcode	int	%9.0g		alternative model code (predecessors and successors get same number)
brd	byte	%21.0g	brand	brand code
type	str40	%40s		name of brand and model
brand	str11	%11s		name of brand
model	str11	%11s		name of model
org	byte	%13.0g	origin	origin code (demand side, country with which consumers associate model)
loc	byte	%13.0g	location	location code (production side, country where producer produce model)
cla	byte	%12.0g	class	class or segment code
home	byte	%9.0g		domestic car dummy (appropriate interaction of org and ma)
frm	byte	%25.0g	firm	firm code

By looking at the descriptions of variables, we know the label of firm is “firm” and label of market is just “market”. Then we would like to see the label list for “firm” & “label”.

label list firm
label list market

```
firm:
           1 AlfaRomeo
           2 BMW
           3 Toyota
           4 Fiat
           5 Ford
           6 Honda
           7 Hyundai
           8 DeTomaso
           9 Kia
          10 Lada
          11 Mazda
          12 Mercedes
          13 Mitsubishi
          14 Nissan
          15 GeneralMotors
          16 Peugeot
          17 Porsche
          18 Renault
          19 Rover
          20 Saab
          21 Seat
          22 AZNP
          23 FujiHI
          24 Suzuki
          25 Toyota
          26 VW
```

```
market:
           1 Belgium
           2 France
           3 Germany
           4 Italy
           5 UK
```

2) The example below considers a merger where General Motors (GM) (.#15) sells its operations to VW (#26) for German in 1998

mergersim simulate if ye == 98 & ma == 3, seller(15) buyer(26) detail

The following tables show the price change after merger between GM & VW and pre & post merger's market shares.

Prices

Unweighted averages by frm

firm code	Pre-merger	Post-merger	Relative change
BMW	14369.057	14369.688	0.000
Toyota	7739.311	7739.334	0.000
Fiat	12281.086	12281.448	0.000
Ford	10483.486	10484.302	0.000
Honda	12633.703	12633.801	0.000
Hyundai	10338.669	10338.720	0.000
Kia	9028.601	9028.627	0.000
Mazda	11393.104	11393.317	0.000
Mercedes	19168.445	19169.281	0.000
Mitsubishi	12676.833	12676.963	0.000
Nissan	12091.164	12091.379	0.000
GeneralMotors	15950.850	17063.985	0.082
Peugeot	13129.019	13129.310	0.000
Renault	12244.376	12244.807	0.000
Suzuki	7386.362	7386.413	0.000
Toyota	13109.080	13109.283	0.000
VW	13757.575	14362.484	0.054
Volvo	17734.826	17734.930	0.000
Daewoo	10796.207	10796.260	0.000
Daimler	6917.631	6917.648	0.000

Variables generated: **M_price2 M_quantity2 M_price_ch (Other M_ variables dropped)**

Market shares by quantity
Unweighted averages by frm

firm code	Pre-merger	Post-merger	Difference
BMW	0.074	0.075	0.001
Toyota	0.003	0.003	0.000
Fiat	0.043	0.043	0.001
Ford	0.095	0.097	0.002
Honda	0.012	0.012	0.000
Hyundai	0.006	0.006	0.000
Kia	0.003	0.003	0.000
Mazda	0.025	0.026	0.000
Mercedes	0.098	0.099	0.002
Mitsubishi	0.015	0.016	0.000
Nissan	0.025	0.026	0.000
GeneralMotors	0.166	0.161	-0.005
Peugeot	0.034	0.035	0.001
Renault	0.051	0.051	0.001
Suzuki	0.006	0.006	0.000
Toyota	0.024	0.024	0.000
VW	0.300	0.297	-0.003
Volvo	0.012	0.013	0.000
Daewoo	0.006	0.006	0.000
Daimler	0.002	0.002	0.000

	Pre-merger	Post-merger
HHS:	1495	2424
C4:	65.87	72.83
C8:	86.01	88.36

	Change
Consumer surplus:	-1,130,545,700
Producer surplus:	114,552,512

3) Plot the graph of price change

gen perc_price_ch=M_price_ch*100

graph bar (mean) perc_price_ch if ma==3 & ye==98, ///
over(frm, sort(perc_price_ch) descending label(angle(vertical))) ///

ytitle(Percentage) title(Average percentage price increase per firm)