TA session2# 4

Jun Sakamoto

October 25,2018

Contents

1 Nested Logit Model

A discrete choice model that does not possess the independence of irrelevant alternative (IIA) property is the nested logit model.

IIA property

$$\frac{P(y=j|B,s)}{P(y=i|B,s)} = \frac{P(y=j|\{i,j\},s)}{P(y=i|\{i,j\},s)} \tag{1}$$

where B is the choice set and s is the attribute. This assumption means the odds between i and j does not depend on the third alternatives.
By the assumption, the probability of i choice is

$$1 = \sum_{j=0}^{J} P(y=j|B,s) = (\sum_{j=0}^{J} \frac{P(y=j|\{i,j\},s)}{P(y=j|\{i,j\},s)})P(y=i|B,s) \tag{2}$$

Suppose the odds ratio be

$$\frac{P(y=j|\{i,j\},s)}{P(y=j|\{i,j\},s)} = \exp(X\beta_j - X\beta_i) \tag{3}$$

Then

$$P(y = j|B, s) = \frac{\exp(X\beta_j)}{1 + \exp(X\beta_i)} \tag{4}$$

When the IIA assumption is considered inappropriate, the alternative model is proposed such as a nested logit model.
For this model, the set of possible choices is decomposed into subset. See an example.

"Alternative models of demand for automobiles" Charlotte Wojcik (2000) Economic letters 68,113-118
We consider deciding to buy a car. In this situation, we decides by three steps. 1st, we choose the class of car (small, standard, luxury / sports, or the outside alternative of not buying a new car), 2nd the country of origin (domestic, European, or Japanese /Korean), and finally the specific model.
The following is the variable list.
$SHARE_{jt}$: the market share (taking account of the outside alternative) of model j;

$SHARE_{0t}$: the market share of the outside alternative in year t;
$PRICE_{jt}$: the retail list price of the base model, in thousands of 1983 dollars;
$INCOME_t$: the average household income plus the simulated deviation in year t;
HP/WT_{jt}: horsepower per 10 lbs of vehicle weight;
AIR_{jt}: a dummy for whether air conditioning is standard;
$MPDOL_{jt}$: MPG (in tens) divided by the retail price of unleaded gas;
$LN3WD_{jt}$: length 3 width, in units of 10,000 square inches;

Result is as bellow.

Table 1
Nested logit and BLP demand results[a]

Variable	Nested logit	BLP
ln(PRICE)	−0.336	
	(0.073)	
ln(INCOME −PRICE)		32.890
		(1.890)
Constant	−4.369	−6.696
	(0.167)	(3.643)
HP/WT	0.286	2.651
	(0.080)	(0.585)
AIR	0.065	1.070
	(0.028)	(0.612)
MPDOL	0.125	−0.115
	(0.010)	(0.764)
LN×WD	0.632	3.349
	(0.081)	(3.381)
$\ln(s_{jt\|oc})$	0.822	
	(0.028)	
$\ln(s_{ot\|c})$	0.742	
	(0.052)	
Preferences× constant		2.670
		(1.875)
Preferences× HP/WT		2.911
		(5.689)
Preferences× AIR		2.182
		(1.628)
Preferences× MPDOL		0.424
		(1.510)
Preferences× LN×WD		1.335
		(1.357)

[a] Dependent variables: nested logit demand: ln(SHARE*jt*)−ln(SHARE0*t*); BLP demand: SHARE*jt*.

Also, calculation result of share probability is

Table 4
Prediction error by country of origin and market class (% difference between actual and predicted shares)

	Actual share	Nested logit	BLP	Nested logit prediction error	BLP prediction error
Origin					
Domestic	79.16	79.16	60.07	0.03	24.07
European	4.90	5.21	23.74	−7.65	−409.02
Japanese	15.42	15.13	15.93	1.83	−22.63
Class					
Small	46.22	45.76	45.63	1.10	0.59
Medium/large	39.42	39.90	26.94	−1.15	31.69
Luxury/sports	14.36	14.33	27.43	−0.16	−92.89

2 Truncated Regression Model

Suppose that y_i and x_i is satisfying following linear regression model.

$$y_i = x_i'\beta + u_i, u_i \sim N(0, \sigma^2) \tag{5}$$

where y_i is iid dependent variable and $x_i \in R^k$ is explanatory variable. We will consider follow ing simple truncation rule.

$$(y_i, x_i) = \left\{ \begin{array}{l} observable\ if\ \ y_i > c \\ unobservable\ if\ \ y_i < c \end{array} \right\} \tag{6}$$

In this case, this rule is often called a truncation from below. Next, the distribution after the truncation $y > c$ is defined over the interval (c, ∞) and given by this

$$f(y|y > c) = \frac{f(y)}{P(y > c)} \tag{7}$$

where $f(y|g)$ means the conditional distribution of y given x. Suppose that $y \sim N(\mu, \sigma^2)$ the mean and variance of the truncated distribution are

$$E[y|y > c] = \mu + \sigma\lambda(v) \tag{8}$$
$$V[y|y > c] = \sigma^2\{1 - \lambda(v)[\lambda(v) - v]\} \tag{9}$$
$$v = \frac{c - \mu}{\sigma} \tag{10}$$
$$\lambda(v) = \frac{\phi(v)}{1 - \Phi(v)} \tag{11}$$

where ϕ and Φ mean the pdf and the cdf of the standard normal distribution function respectively. We sometimes call $\lambda(v)$ inverse Mill's ratio and hazard function. Clearly above equation shows that the mean of the truncated distribution is not μ but $\mu + \sigma\lambda(v)$. That is it has sample selection bias. By using above results, we can get following results for the linear regression model.

$$E[y_i|x_i] = x_i'\beta + \sigma\lambda(\frac{c - x_i'\beta}{\sigma}) \tag{12}$$
$$V[y|x_i] = \sigma^2\{1 - \lambda(\frac{c - x_i'\beta}{\sigma})[\lambda(\frac{c - x_i'\beta}{\sigma}) - \frac{c - x_i'\beta}{\sigma}]\} \tag{13}$$

for if and only if $y_i > c$. In this case, the OLSE has bias. Because

$$E[\hat{\beta}] = E[(X'X)^{-1}X'y] \tag{14}$$
$$= (\sum_i x_i x_i')^{-1} \sum_i x_i E[y_i] \tag{15}$$
$$= (\sum_i x_i x_i')^{-1} \sum_i x_i[\mu + \sigma\lambda(v_i)] \tag{16}$$
$$= (\sum_i x_i x_i')^{-1} \sum_i x_i[x_i'\beta + \sigma\lambda(\frac{c - x_i'\beta}{\sigma})] \tag{17}$$
$$= \beta + \sigma(\sum_i x_i x_i')^{-1} \sum_i x_i\lambda(v_i) \tag{18}$$

holds. Therefore the OLSE has bias. So we use MLE.

3 Tobit Model

Suppose that y_i and x_i is satisfying following linear regression model.

$$y_i^* = x_i'\beta + u_i, u_i \sim N(0, \sigma^2) \tag{19}$$

where y_i is iid dependent variable and $x_i \in R^k$ is explanatory variable. We will consider the following censoring.

$$(y_i, x_i) = \left\{ \begin{array}{l} y_i^* \ if \ y_i^* > c \\ c \ if \ y_i^* \leq c \end{array} \right\} \tag{20}$$

that is

$$y_i = max\{x_i'\beta + u_i, c\} \tag{21}$$

This model called Tobit model. Unlike in truncated model, there is no truncation here. The feature that distinguishes the cenored regression model from usual regression model is that the dependent variable is censored. For the observations that satisfy $y_i^* > c$, its density is given by

$$\sigma^{-1}\phi(\frac{y_i - x_i'\beta}{\sigma}) \tag{22}$$

Because there is no truncation, this density differs from the density for the truncated model in this case. For the observations which satisfy $y_i^* \leq c$, the probability of which is given by

$$P(y_i^* \leq c|x_i) = P(\frac{y_i^* - x_i'\beta}{\sigma} \leq \frac{c - x_i'\beta}{\sigma}|x_i) \tag{23}$$

$$= \Phi(\frac{c - x_i'\beta}{\sigma}) \tag{24}$$

Combining these two results. We can get

$$[\sigma^{-1}\phi(\frac{y_i - x_i'\beta}{\sigma})]^{I_i} \times [\Phi(\frac{c - x_i'\beta}{\sigma})]^{1-I_i} \tag{25}$$

$$I_i = \left\{ \begin{array}{l} 1 \ if \ y_i^* > c \\ 0 \ if \ y_i^* \leq c \end{array} \right\} \tag{26}$$

Taking logs, we obtain the log conditional likelihood for the observation $i = 1, 2, ...$

$$l(\beta, \sigma^2) = \sum_i I_i log[\sigma^{-1}\phi(\frac{y_i - x_i'\beta}{\sigma})] + \sum_i (1 - I_i) log\Phi(\frac{c - x_i'\beta}{\sigma}) \tag{27}$$

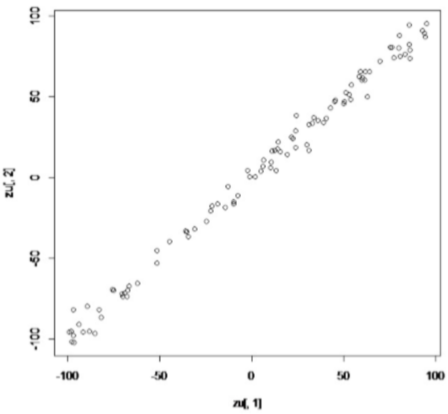
which implies

$$l(\beta, \sigma^2) = n^{-1} \sum_i log[\sigma^{-1} \phi(\frac{y_i - x_i' \beta}{\sigma})] + n^{-1} \sum_i log \Phi(\frac{c - x_i' \beta}{\sigma}) \tag{28}$$

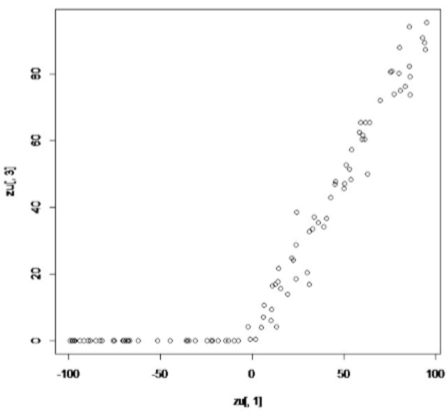
Empirical example

Check estimate of OLS and Tobit regression.
First we check scatter plot of $\{y_i, x_i\}$ and $\{y_i^*, x_i\}$.

$\{y_i^*, x_i\}$

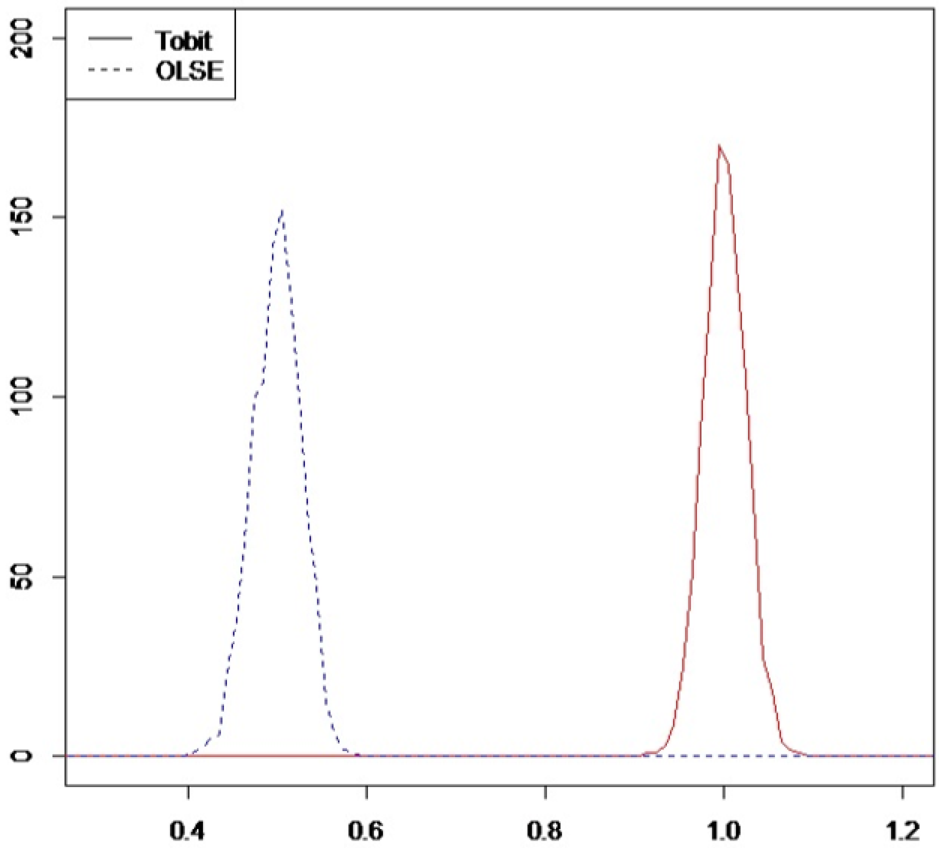


$\{y_i, x_i\}$



1.We generate samples $X \sim U[-100, 100]$ and $u \sim N(0, 5)$.(t=50)
2.y is made by X and u from step1.(We assume true beta is 1)
3.We estimate $\hat{\beta}_{OLS}$ and $\hat{\beta}_{TR}$.
4.$\hat{\beta}$ is obtained by repeating above steps 1000 times.

Distribution of there estimate are


Tobit
OLSE
0
50
100
150
200
0.4
0.6
0.8
1.0
1.2