ECON 613: Applied Econometrics

Methods for Cross-sectional Data

February 5, 2019

Binary Response Model

Multinomial Choices

Introduction

Binary response models are models where the variable to be explained y is a random variable taking on the values zero and one which indicate whether or not a certain event has occured.

- $y = 1$ if a person is employed
- $y = 1$ if a family contributes to a charity during a particular year
- $y = 1$ if a firm has a particular type of pension plan
- $y = 1$ if a worker goes to college
- Regardless of what y stands for, we refer to $y = 1$ as a success and $y = 0$ as a failure.

An OLS regression of y on dependent variables denoted x ignores the discreteness of the dependent variable and does not constrain predicted probabilities to be between zero and one.

Linear Probability Model (1)

$$P(y = 1) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k. \quad (1)$$

- If x_1 is continuous, β_1 is the change in the probability of success given one unit increase in x_1
- If x_1 is discrete, β_1 is the difference in the probability of success when $x_1 = 1$ and $x_1 = 0$, holding other x_j fixed.

Linear Probability Model (2)

Given that y is a random variable (Bernouilli), we also have

$$\begin{aligned} E(y \mid x) &= \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k. && (2) \\ Var(y \mid x) &= x\beta(1 - x\beta) && (3) \end{aligned}$$

Implications:

- OLS regression of y on $x_1, x_2, \ldots, x_k$ produces consistent and unbiaised estimators of the β_j.
- Heteroskedacticity, which can be dealt with using standard heteroskedasticity-robust standard errors.
- Problem: OLS fitted values may not be between zero and one.

Maximum Likelihood Estimation: Logit and Probit Models

For binary outcome data, the dependent variable y takes one of two values. We let

$$y = \begin{cases} 1, & \text{with probability } p \\ 0, & \text{with probability } 1 - p \end{cases} \tag{4}$$

Parametrize conditional probabilities:

$$p_i = F_\epsilon(X\beta) \tag{5}$$

And, Marginal Effects

$$\frac{\partial Pr(y_i = 1 \mid x_i)}{\partial x_{ij}} = F_\epsilon'(X\beta)\beta_j \tag{6}$$

Probit Model (1)

The probit model corresponds to the case where $F(x)$ is the cumulative standard normal distribution function

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(\tfrac{1}{2}X^2) dX \tag{7}$$

Where $F(X\beta) = \Phi(X\beta)$.

Probit Model (2)

- Consider the latent approach

$$y^\star = X\beta + \epsilon \tag{8}$$

where $\epsilon \sim N(0,1)$. Think of $y^\star$ as the net utility associated with some action. If the action yields positive net utility, it is undertaken otherwise it is not.

- We would care only about the sign of $y^\star$

$$y = \begin{cases} 1, & \text{if } p \\ 0, & \text{with probability } 1-p \end{cases} \tag{9}$$

- Probabilities

$$\begin{aligned} Pr(y=1) &= Pr(y^\star > 0) = Pr(X\beta + \epsilon \geq 0) \\ &= Pr(\epsilon \geq -X\beta) = Pr(\epsilon \leq X\beta) = \Phi(X\beta) \end{aligned}$$

Logit Model

The logit model specifies the cdf function $F(x)$ is now the logistic function

$$\Lambda(x) = \frac{1}{1 + \exp(-x)} = \frac{exp(x)}{1 + \exp(x)} \tag{10}$$

The logit model is most easily derived by assuming that

$$\log\left(\frac{P}{1-P}\right) = X\beta \tag{11}$$

The logarithm of the odds (ratio of two probabilities) is equal to $X\beta$

Maximum Likelihood Estimation

- Likelihood can not be defined as a joint density function.
- Outcome of a Bernouilli trial

$$f(y_i \mid x_i) = p_i^{y_i}(1-p_i)^{1-y_i} \tag{12}$$

- Given the independence of individuals, the likelihood can be written

$$\mathcal{L}(\beta) = \prod_{i=1}^{n} F(x_i\beta)^{y_i}(1-F(x_i\beta))^{1-y_i} \tag{13}$$

Log Likelihood

- The log likelihood

$$\log \mathcal{L}(\beta) = \sum_{i=1}^{n} y_i lnF(x_i\beta) + (1 - y_i)ln(1 - F(x_i\beta)) \tag{14}$$

- First order conditions

$$\sum_{i=1}^{n} \frac{y_i - F(x_i\beta)}{F(x_i\beta)(1 - F(x_i\beta))} F^{'}(x_i\beta)x_i = 0 \tag{15}$$

Empirical considerations

- Probit and logit yield same outcomes. Only difference is how parameters are scaled.
- The natural metric to compare models is the fitted log-likelihood provided that the models have the same number of parameters.
- Although estimated parameters are different, marginal effects are quite similar.

Pseudo R2

$$R^2_{\text{Binary}} = 1 - \frac{\mathcal{L}(\hat{\beta})}{N[\bar{y} ln \bar{y} + (1-\bar{y}) ln(1-\bar{y})]} \qquad (16)$$

Predicted Outcomes

- The criterion $\sum_i (y_i - \hat{y}_i)^2$ gives the number of wrong predictions.
  - average rule: let $\hat{y} = 1$ when $\hat{p} = F(X\beta) > 0.5$
  - Receiver Operating Characteristics (ROC) curve plots the fractions of $y = 1$ correctly classified against the fractions of $y = 0$ incorrectly specified as the cutoffs $\hat{p} = F(X\beta) > c$ varies.

Example: Describing the data

		No Affair	Affair
Gender	female	0.54	0.48
	male	0.46	0.52
Age	17.5	0.01	0.02
	22	0.22	0.11
	27	0.26	0.24
	32	0.17	0.25
	37	0.14	0.15
	42	0.08	0.12
	47	0.04	0.05
	52	0.03	0.04
	57	0.04	0.02
Years Married	0.125	0.02	0.01
	0.417	0.02	0.01
	0.75	0.06	0.02
	1.5	0.17	0.08
	4	0.17	0.18
	7	0.13	0.15
	10	0.11	0.14
	15	0.31	0.41
Share		0.75	0.25

Example: Affairs

		No Affair	Affair
Children	no	0.319	0.18
	yes	0.681	0.82
Religiousness	1	0.062	0.133
	2	0.273	0.273
	3	0.191	0.287
	4	0.348	0.22
	5	0.126	0.087
Education	9	0.011	0.013
	12	0.069	0.087
	14	0.264	0.233
	16	0.211	0.133
	17	0.137	0.18
	18	0.175	0.22
	20	0.133	0.133

Probit

	Model 1	Model 2	Model 3	Model 4
(Intercept)	-0.74^{***}	-0.02	-0.18	-0.02
	(0.08)	(0.51)	(0.52)	(0.81)
as.factor(gender)male	0.14	0.13	0.14	0.21
	(0.11)	(0.12)	(0.12)	(0.12)
as.factor(age)22		-1.12^{*}	-1.07^{*}	-1.45^{*}
		(0.53)	(0.54)	(0.61)
as.factor(age)27		-0.76	-0.82	-1.44^{*}
		(0.53)	(0.53)	(0.62)
as.factor(age)32		-0.48	-0.60	-1.36^{*}
		(0.53)	(0.53)	(0.63)
as.factor(age)37		-0.70	-0.84	-1.79^{**}
		(0.53)	(0.54)	(0.66)
as.factor(age)42		-0.50	-0.65	-1.61^{*}
		(0.54)	(0.55)	(0.67)
as.factor(age)47		-0.56	-0.68	-1.72^{*}
		(0.58)	(0.59)	(0.71)
as.factor(age)52		-0.62	-0.76	-1.75^{*}
		(0.59)	(0.60)	(0.71)
as.factor(age)57		-1.17	-1.31^{*}	-2.36^{**}
		(0.62)	(0.62)	(0.74)
as.factor(children)yes			0.31^{*}	0.11
			(0.15)	(0.17)
as.factor(yearsmarried)0.417				0.01
				(0.77)
as.factor(yearsmarried)0.75				-0.37
				(0.67)
as.factor(yearsmarried)1.5				0.22
				(0.56)
as.factor(yearsmarried)4				0.62
				(0.57)
Log Likelihood	-336.91	-327.84	-325.74	-319.69

Logit

	Model 1	Model 2	Model 3	Model 4
(Intercept)	-1.22^{***}	-0.03	-0.31	-0.07
	(0.13)	(0.82)	(0.84)	(1.48)
as.factor(gender)male	0.24	0.21	0.23	0.34
	(0.19)	(0.20)	(0.20)	(0.21)
as.factor(age)22		-1.88^{*}	-1.78^{*}	-2.44^{*}
		(0.86)	(0.87)	(1.05)
as.factor(age)27		-1.24	-1.35	-2.41^{*}
		(0.84)	(0.85)	(1.06)
as.factor(age)32		-0.77	-0.99	-2.29^{*}
		(0.84)	(0.86)	(1.08)
as.factor(age)37		-1.14	-1.38	-2.99^{**}
		(0.86)	(0.88)	(1.13)
as.factor(age)42		-0.82	-1.07	-2.70^{*}
		(0.87)	(0.89)	(1.14)
as.factor(age)47		-0.90	-1.11	-2.87^{*}
		(0.94)	(0.95)	(1.21)
as.factor(age)52		-1.00	-1.26	-2.95^{*}
		(0.95)	(0.97)	(1.21)
as.factor(age)57		-1.97	-2.22^{*}	-3.99^{**}
		(1.03)	(1.05)	(1.29)
as.factor(children)yes			0.54^{*}	0.16
			(0.27)	(0.30)
as.factor(yearsmarried)0.417				0.09
				(1.49)
as.factor(yearsmarried)0.75				-0.50
				(1.31)
as.factor(yearsmarried)1.5				0.42
				(1.10)
as.factor(yearsmarried)4				1.11
				(1.10)
Log Likelihood	-336.91	-327.90	-325.85	-320.24

Probit VS Logit

	Probit	Logit
(Intercept)	-0.02	-0.07
	(0.81)	(1.48)
as.factor(gender)male	0.21	0.34
	(0.12)	(0.21)
as.factor(age)22	-1.45^{*}	-2.44^{*}
	(0.61)	(1.05)
as.factor(age)27	-1.44^{*}	-2.41^{*}
	(0.62)	(1.06)
as.factor(age)32	-1.36^{*}	-2.29^{*}
	(0.63)	(1.08)
as.factor(age)37	-1.79^{**}	-2.99^{**}
	(0.66)	(1.13)
as.factor(age)42	-1.61^{*}	-2.70^{*}
	(0.67)	(1.14)
as.factor(age)47	-1.72^{*}	-2.87^{*}
	(0.71)	(1.21)
as.factor(age)52	-1.75^{*}	-2.95^{*}
	(0.71)	(1.21)
as.factor(age)57	-2.36^{**}	-3.99^{**}
	(0.74)	(1.29)
as.factor(children)yes	0.11	0.16
	(0.17)	(0.30)
as.factor(yearsmarried)0.417	0.01	0.09
	(0.77)	(1.49)
AIC	675.38	676.48
BIC	754.55	755.66
Log Likelihood	-319.69	-320.24

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Identification considerations

- β is identified up to a scale.
- We observe only whether $X\beta + \epsilon > 0$.
- Implication for the interpretation of the coefficients.

Model Selection

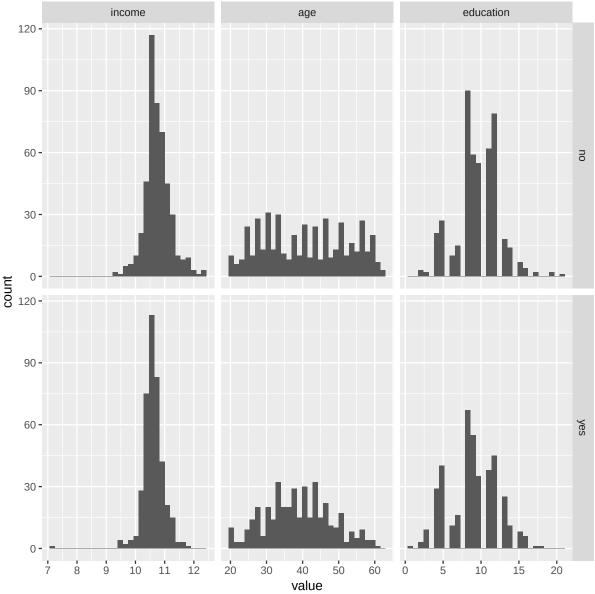
- $AIC = 2k - 2\log\mathcal{L}$
- $BIC = \log(n)k - 2\log\mathcal{L}$

Determinants of Female Labor Supply

Cross-section data originating from the health survey SOMIPOPS for Switzerland in 1981.

- participation Factor: Did the individual participate in the labor force?
- income: Logarithm of nonlabor income.
- age: Age in decades (years divided by 10).
- education: Years of formal education.
- youngkids: Number of young children (under 7 years of age).
- oldkids:Number of older children (over 7 years of age).
- foreign Factor: Is the individual a foreigner (i.e., not Swiss)?

Describing the data (1)



Describing the data (2)

		Participation	
		No	Yes
Number of young kids	0	0.6921	0.8454
	1	0.2123	0.1172
	2	0.0892	0.0324
	3	0.0064	0.005
Number of old kids	0	0.4904	0.404
	1	0.2166	0.2394
	2	0.2166	0.2643
	3	0.0573	0.0698
	4	0.0149	0.0175
	5	0.0042	0
	6	0	0.005

Effect of income

	Model 1	Model 2	Model 3
(Intercept)	5.97^{***}	-16.60	116.89
	(1.19)	(10.14)	(67.96)
income	-0.57^{***}	3.69	-37.41
	(0.11)	(1.91)	(20.52)
$I(income^2)$		-0.20^{*}	3.97
		(0.09)	(2.06)
$I(income^3)$			-0.14^{*}
			(0.07)
Log Likelihood	-587.91	-585.54	-583.06
Num. obs.	872	872	872

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Effect of age

	Model 1	Model 2	Model 3
(Intercept)	0.34^{*}	-3.52^{***}	-1.10
	(0.17)	(0.63)	(2.20)
age	-0.01^{**}	0.19^{***}	-0.00
	(0.00)	(0.03)	(0.18)
$I(age^2)$		-0.00^{***}	0.00
		(0.00)	(0.00)
$I(age^3)$			-0.00
			(0.00)
Log Likelihood	-597.85	-576.39	-575.71
Num. obs.	872	872	872

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Specification

	Model 1	Model 2	Model 3
(Intercept)	7.66^{***}	4.65^{***}	-7.42
	(1.28)	(1.40)	(10.73)
income	-0.56^{***}	-0.74^{***}	1.55
	(0.12)	(0.13)	(2.03)
age	-0.03^{***}	0.23^{***}	0.23^{***}
	(0.01)	(0.04)	(0.04)
education	-0.03	-0.02	-0.02
	(0.02)	(0.02)	(0.02)
youngkids	-0.71^{***}	-0.64^{***}	-0.64^{***}
	(0.10)	(0.10)	(0.10)
oldkids	-0.01	-0.16^{**}	-0.16^{**}
	(0.04)	(0.05)	(0.05)
$I(age^2)$		-0.00^{***}	-0.00^{***}
		(0.00)	(0.00)
$I(income^2)$			-0.11
			(0.10)
Log Likelihood	-550.08	-526.38	-525.86
Num. obs.	872	872	872

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Marginal Effects

- Recall that marginal effects are given by

$$\frac{\partial Pr(y_i = 1 \mid x_i)}{\partial x_{ij}} = F'_{\epsilon}(X\beta)\beta_j \tag{17}$$

- Different definitions
  - Average marginal effects in the sample.
  - Marginal effect evaluated at the mean.

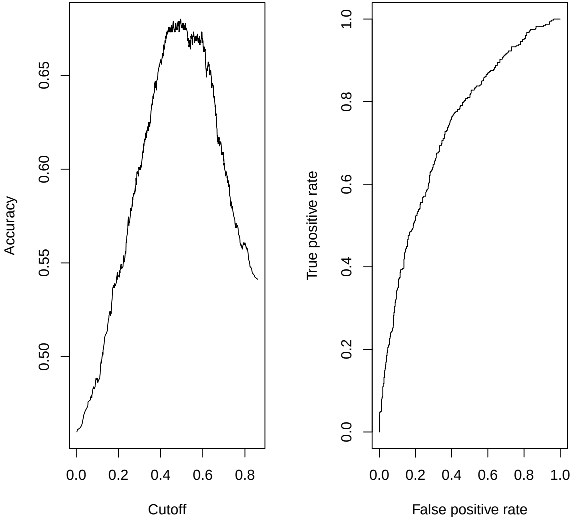
Marginal Effects

income	0.53
age	0.08
education	-0.01
youngkids	-0.22
oldkids	-0.05
I(age^2)	-0.00
I($income^2$)	-0.04

Prediction

	Model	
Data	0	1
0	325	146
1	137	264

ROC



Accuracy
0.50
0.55
0.60
0.65
0.0
0.2
0.4
0.6
0.8
Cutoff
True positive rate
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0
False positive rate

By citizenship

	Native	Immigrants
(Intercept)	-6.85	-67.57
	(11.24)	(48.18)
income	1.17	13.31
	(2.11)	(9.16)
age	0.24^{***}	0.14
	(0.05)	(0.09)
education	0.05^{*}	-0.03
	(0.02)	(0.04)
youngkids	-0.76^{***}	-0.75^{***}
	(0.13)	(0.18)
oldkids	-0.16^{**}	-0.18
	(0.06)	(0.12)
$\mathrm{I}(age^2)$	-0.00^{***}	-0.00
	(0.00)	(0.00)
$\mathrm{I}(income^2)$	-0.09	-0.66
	(0.10)	(0.43)
Log Likelihood	-384.24	-117.82
Num. obs.	656	216

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Murder Rates

Cross-section data on states in 1950.

- rate: Murder rate per 100,000 (FBI estimate, 1950).
- convictions: Number of convictions divided by number of murders in 1950.
- executions: Average number of executions during 1946–1950 divided by convictions in 1950.
- time: Median time served (in months) of convicted murderers released in 1951.
- income: Median family income in 1949 (in 1,000 USD).
- lfp: Labor force participation rate in 1950 (in percent).
- noncauc: Proportion of population that is non-Caucasian in 1950.
- southern: Factor indicating region.

Warnings

- Binary model of the determinants of having an execution
- This is very bad economics
- An example to illustrate some technical problems

```
glm(formula = I(executions > 0) ~ time + income + noncauc +
    southern, family = binomial, data = MurderRates)

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  10.99326   20.77336   0.529   0.5967
time          0.01943    0.01040   1.868   0.0617 .
income       10.61013    5.65409   1.877   0.0606 .
noncauc      70.98785   36.41181   1.950   0.0512 .
lfp          -0.66763    0.47668  -1.401   0.1613
southernyes  17.33126 2872.17069   0.006   0.9952
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

- Diagnostic
- Point estimate and Std.

```
table(I(MurderRates$executions > 0), MurderRates$southern)

       no yes
  FALSE  9   0
  TRUE  20  15
```

Quasi Separation

- We have here β^0 such that

$$\begin{aligned} y_i = 0 \quad & \text{whenever} \quad x_i'\beta^0 \leq 0 \\ y_i = 1 \quad & \text{whenever} \quad x_i'\beta^0 \geq 0 \end{aligned}$$

- The maximum likelihood estimate does not exist.

Binary Response Model

Multinomial Choices

Multinomial Models

Dependent variable has several possible outcomes, that are mutually exclusive

- Commute to work (car, bus, bike, walking)
- Employment status (full time, part time, unemployed)
- Occupation choice, field of study, product choice
- Ordered choices eg. education choices
- Unordered choices eg. fishing mode

Ordered Discrete Response

Suppose that $y^{\star}(= x_i\beta + \epsilon_i)$ is continuously distributed with standard deviation σ but the observed response y_i is an ordered discrete choice (ODR) taking 0,1,. . . ,R-1 determined by fixed threesholds γ_r. Formally, we have

$$y = \begin{cases} 0, & \text{if } x\beta + \epsilon < \gamma_1 \\ 1, & \text{if } \gamma_1 \leq x\beta + \epsilon < \gamma_2 \\ 2, & \text{if } \gamma_2 \leq x\beta + \epsilon < \gamma_3 \\ \dots & \\ R-1, & \text{if } \gamma_{R-1} \leq x\beta + \epsilon \end{cases} \tag{18}$$

Identification

- Not all parameters are identified as in the binary response model.
- Consider $\gamma_r \leq x\beta + \epsilon < \gamma_{r+1}$. Then, we have $\frac{\gamma_r - \gamma_1}{\sigma} \leq \frac{x\beta + \epsilon - \gamma}{\sigma} < \frac{\gamma_{r+1} - \gamma_1}{\sigma}$
- Then, the identified parameters are

$$\alpha = \left(\frac{\beta_1 - \gamma_1}{\sigma}, \frac{\beta_2}{\sigma}, \ldots, \frac{\beta_k}{\sigma} \right) \tag{19}$$

$$\rho_r = \frac{\gamma_r - \gamma_1}{\sigma} \tag{20}$$

for $r = 2, \ldots, R - 1$.

Toward the Likelihood

$y = r$

$$
\begin{aligned}
\gamma_r &\leq x\beta + \epsilon &&< \gamma_{r+1} \\
\gamma_r - x\beta &\leq \epsilon &&< \gamma_{r+1} - x\beta \\
\frac{\gamma_r - \gamma_1}{\sigma} + \frac{\gamma_1 - x\beta}{\sigma} &\leq \frac{\epsilon}{\sigma} &&< \frac{\gamma_{r+1} - \gamma_1}{\sigma} + \frac{\gamma_1 - x\beta}{\sigma} \\
\rho_r - x\alpha &\leq \frac{\epsilon}{\sigma} &&< \rho_{r+1} - x\alpha
\end{aligned}
$$

Likelihood

- Choice probabilities

$$P(y = r \mid x) = P(\rho_r - x\alpha \leq \frac{\epsilon}{\sigma} < \rho_{r+1} - x\alpha)$$
$$= F(\rho_{r+1} - x\alpha) - F(\rho_r - x\alpha)$$

- Likelihood

$$\mathcal{L}(a, t) = \prod_{i=1}^{N} \prod_{r=0}^{R-1} [P(y_i = r \mid x)]^{y_{ir}} \tag{21}$$

- Log Likelihood of the probit model

$$\log \mathcal{L}(a, t) = \sum_{i=1}^{N} \sum_{r=0}^{R-1} y_{ir} \log(\Phi(t_{r+1} - xa) - \Phi(t_r - xa)) \tag{22}$$

where a and t are respectively the parameters to be estimated and the cutoffs.

Marginal Effects

- The marginal effects of x on each choice probabilities can be derived as:

$$\frac{\partial P(y = r \mid x)}{\partial x} = -\alpha(\phi(\rho_{r+1} - x\alpha) - \phi(\rho_r - x\alpha)) \qquad (23)$$

- Caution

$$\sum_{r=0}^{R} P(y = r \mid x) = 1 \qquad (24)$$

Then

$$\sum_{r=0}^{R} \frac{\partial P(y = r \mid x)}{\partial x} = 0 \qquad (25)$$

An increase in some choice probability necessarily entails a decrease in some other choice probabilities.

Multinomial Logit

The models differ according to whether or not regressors vary across alternatives.

- Conditional logit model

$$p_{ij} = \frac{\exp(x_{ij}\beta)}{\sum_{l=1}^{m} \exp(x_{il}\beta)} \qquad j = 1, \ldots, m. \tag{26}$$

- Multinomial logit model

$$p_{ij} = \frac{\exp(x_i\beta_j)}{\sum_{l=1}^{m} \exp(x_i\beta_l)} \qquad j = 1, \ldots, m. \tag{27}$$

- Mixed logit model

$$p_{ij} = \frac{\exp(x_{ij}\beta + w_i\gamma_j)}{\sum_{l=1}^{m} \exp(x_{il}\beta + w_i\gamma_l)} \qquad j = 1, \ldots, m. \tag{28}$$

Marginal Effects

- Conditional logit

$$\frac{\partial p_{ij}}{\partial x_{ik}} = p_{ij}(\delta_{ijk} - p_{ik})\beta \tag{29}$$

  where δ_{ijk} is an indicator variable equal to 1 if $j = k$ and equal to 0 otherwise.

- Multinomial logit

$$\frac{\partial p_{ij}}{\partial x_i} = p_{ij}(\beta_j - \bar{\beta}_i) \tag{30}$$

  where $\bar{\beta}_i = \sum_l p_{il}\beta_l$

Independence of Irrelevant Alternatives

- A limitation of the conditional logit and multinomial logit is that discrimination among the m alternatives reduces to a series of pairwise comparisons that are unaffected by the characteristics of alternatives other than the pair under consideration.
- The choice probabilities must be unaffected by the removal of one alternative.

That is because

$$Pr(y = j \mid y = k) = \frac{p_j}{p_j + p_k} \tag{31}$$

Testing for IIA

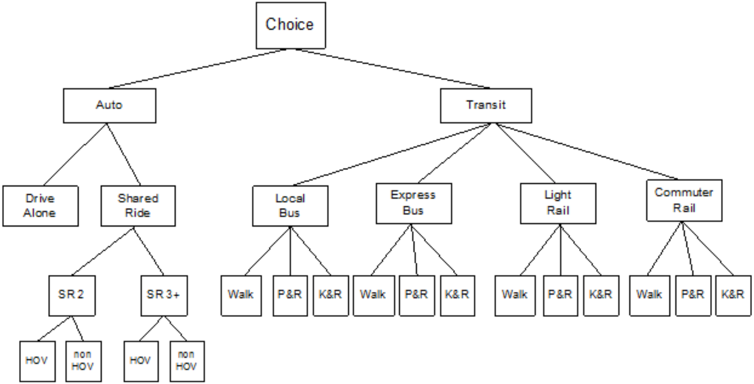
- Estimate the model twice
  - On the full set of alternatives and obtain θ_{full}
  - On a subset of alternatives and obtain θ_{subset}
- Compare $\mathcal{L}_{subset}(\theta_{full})$ and $\mathcal{L}_{subset}(\theta_{subset})$. If there is a significant different, then IIA is violated.

It is very (very) rare that IIA is not violated.

Alternatives

- Generalized Extreme Value Model
- Nested Logit Model
- Random Parameters Logit
- Multinomial Probit

Nested Logit



Nested Logit

The nested logit model breaks decision making into groups. The utility for the alternative is given

$$U_{jk} = V_{jk} + \epsilon_{jk} \quad k = 1, 2, \ldots, K_j, \quad j = 1, 2, \ldots, J \tag{32}$$

Utilities are given by:

- $V_{11} + \epsilon_{11}$
- $\ldots$
- $V_{JK_J} + \epsilon_{JK_J}$

Choice Probabilities

- Choice probability

$$p_{jk} = p_j \times p_{k|j}. \tag{33}$$

- This arises from GEV joint distribution

$$F(\epsilon) = exp(-G(e^{-\epsilon_{11}}, \ldots, e^{-\epsilon_{1K_1}}; \ldots; e^{-\epsilon_{J1}}, \ldots, e^{-\epsilon_{JK_J}}))$$

with

$$G(Y) = G(Y_{11}, \ldots, Y_{1K_1}, \ldots, Y_{JK_J}) = \sum_{j=1}^{J} \left(\sum_{k=1}^{K_j} Y_{jk}^{\frac{1}{\rho_j}} \right)^{\rho_j} \tag{34}$$

Model

Consider

$$V_{jk} = z_j\alpha + x_{jk}\beta_j \quad k = 1, \ldots, K_j, \quad j = 1, \ldots, J \tag{35}$$

The probability of the nested logit model

$$p_{jk} = p_j \times p_{k|j} = \frac{\exp(z_j\alpha + \rho_j I_j)}{\sum_{m=1}^{J} \exp(z_m\alpha + \rho_m I_m)} \times \frac{\exp(x_{jk}\beta/\rho_j + \rho_j I_j)}{\sum_{m=1}^{J} \exp(z_m\alpha + \rho_m I_m)}$$

where

$$I_j = ln\left(\sum_{l=1}^{K_j} \exp(x_{jl}\beta_j/\rho_j)\right)$$

is the inclusive value or the log-sum.

Likelihood

For the *ith* observation, we observe $K_1 + \ldots + K_J$ outcomes y_{ijk}, where $y_{ijk} = 1$ if alternative jk is chosen and is zero otherwise. Then the density of one observation y_i can be expressed

$$f(y_i) = \prod_{j=1}^{J} \prod_{k=1}^{K_J} [p_{ij} \times p_{ik|j}]^{y_{ijk}} = \prod_{j=1}^{J} p_{ij}^{y_{ij}} \left(\prod_{k=1}^{K_J} p_{ik|j}^{y_{ijk}} \right)$$

The log likelihood is given by

$$lnL = \sum_{i=1}^{N} \sum_{y=1}^{J} y_{ij} \log(p_{ij}) + \sum_{i=1}^{N} \sum_{y=1}^{J} \sum_{k=1}^{K_J} y_{ijk} \log(p_{ik|j})$$

Discussions and Limitations

- Nested Logit can be estimated in two steps following the construction of the densities
- Not all choices are easy to nest.

Multinomial Probit

- Consider m-choice with

$$U_j = V_j + \epsilon_j \qquad j = 1, 2, \ldots, m. \tag{36}$$

  where $\epsilon \sim \mathbb{N}(0, \Sigma)$.
- Σ is the matrix of variance-covariance, which can be left unrestricted to capture the correlation between choices.
- Choice probabilities.. For 4 choices, we have

$$P(Y = 1) = \int_{-\infty}^{-V_{41}} \int_{-\infty}^{-V_{31}} \int_{-\infty}^{-V_{21}} f(x, y, z) dz dy dx \tag{37}$$

  where $f(x, y, z)$ is the pdf of the trivariate normal.

Need simulation to compute the choice probabilities and the likelihood.

Applications