Chapter on Discrete Choice Models

Linear Probability Model: Sample observations on \((I_i, x_i)\) where \(I_i = 0, 1\) is a dichotomous indicator,
\[
I_i = x_i \beta + \epsilon_i, \quad i = 1, \cdots, n
\]
where \(E(\epsilon_i|x) = 0\). This is a linear regression specification.

As \(E(\epsilon_i|x) = 0\), it implies that \(E(I_i|x) = x_i \beta\). Because \(E(I_i|x) = P(I_i = 1|x)\) by the expectation formula of a discrete r.v., \(x_i \beta\) can be interpreted as the choice probability of \(I_i = 1\) given \(x_i\).

The limitation of the linear probability model is that there is no restriction on the possible value of \(x \beta\) between 0 and 1. So, for certain values of \(x\) and \(\beta\), the predicted values of \(x \beta\) can be less than 0 and greater than 1.

Furthermore, the variance of \(\epsilon_i\) is not a constant. The variances of \(\epsilon s\) are heterokedastic. As \(\text{var}(\epsilon|x) = \text{var}(I|x) = P(I = 1|x)[1 - P(I = 1|x)] = x \beta(1 - x \beta)\), the variance of \(\epsilon s\) depends on \(x\) and \(\beta\).

Probit and Logit models: To impose probability structure, specify \(P(I_1 = 1|x) = F(x \beta)\), where \(F\) is a distribution function. A Probit model specifies that \(F(x) = \Phi(x)\) where \(\Phi(x)\) is a standard normal distribution. A Logit model specifies that \(F(x) = e^x/(1 + e^x)\). The linear probability corresponds to \(F(x) = x - \text{a uniform distribution (if } x\text{ is restricted to lie between 0 and 1).}

The logistic distribution has a zero mean and a variance = \(\pi^2/3\). The standard logistic distribution \(e^{\lambda x}/(1 + e^{\lambda x})\) with \(\lambda = \pi/\sqrt{3}\) has slightly heavier tails than the standard normal distribution.

These models can be estimated by the method of maximum likelihood. With an independent sample for \(I's\), the log likelihood function is
\[
\ln L = \sum_{i=1}^{n} [I_i \ln F(x_i \beta) + (1 - I_i) \ln (1 - F(x_i \beta))].
\]

The log likelihood functions of the logit and probit models are concave in parameters.

The Conditional Logit Model

The classical consumer demand model assumes that consumers are rational in the sense that they make choices maximizing their perceived utility subject to constraints on expenditure. Under this classical framework, McFadden (1974) provides microeconomic justification of the specification of popular probability choice models.

Suppose there are \(m\) choice alternatives for a consumer. To allow unobservable attributes, unobservable characteristics or imperfect perception, the utility function is assumed to be a random function from the econometrician’s point of view:
\[
y^*_j = V(x_j, z, \theta) + \epsilon, \quad j = 1, \ldots, m.
\]

Let \(I_j\) be a dichotomous indicator that the \(jth\) alternative is chosen. Under the utility maximization hypothesis, \(I_j = 1\) if \(y^*_j = \max(y^*_1, \cdots, y^*_m)\) and \(I_j = 0\), otherwise. In practice, one may assume that \(V(x_j, z, \theta) = x_j \beta + z \alpha_j\). With stochastic distributions for the disturbances \(\epsilon\), choice probabilities can be derived. The choice probabilities may not have, in general, closed expressions. But for some specific parametric distributions, choice probabilities can be derived analytically with closed form expressions.

Suppose that \(\epsilon_j, j = 1, \ldots, m\) are i.i.d. with the Gumbel type I extreme-value distribution:
\[
F(\epsilon_j < \epsilon) = \exp(-e^{-\epsilon}).
\]
The corresponding density function is
\[
f(\epsilon) = \exp(-e^{-\epsilon})e^{-\epsilon}.
\]
Under this distribution, the choice probability function is a logistic probability (logit),
\[
\text{Prob}(I_j = 1|x) = \frac{e^{V_j}}{\sum_{k=1}^{m} e^{V_k}}.
\]
where $V_j$ denotes $V(x_j, z)$ for simplicity. This can be shown as follows. Since $y_j^* = \max(y_1^*, \ldots, y_m^*)$ is equivalent to $\epsilon_k < \epsilon_j + V_j - V_k$ for all $k, j$, 

$\text{Prob}(I_j = 1|x) = \text{Prob}(\epsilon_k < \epsilon_j + V_j - V_k, \text{for all } k, j) \quad \text{for all } j$. 

$\int_{-\infty}^{\infty} \prod_{k \neq j} F(\epsilon + V_j - V_k) f(\epsilon) \, d\epsilon 
= \int_{-\infty}^{\infty} \exp(-\epsilon - e^{-\epsilon} \sum_{k=1}^{m} e^{V_k - V_j}) \, d\epsilon 
= \left( \sum_{k=1}^{m} e^{V_k - V_j} \right)^{-1} \int_{-\infty}^{\infty} \exp(-e^\epsilon - e^{-\epsilon}) \, d\epsilon 
= \frac{e^{V_j}}{\sum_{k=1}^{m} e^{V_k}},$ 

by using the transformation $e^\epsilon = \epsilon - \ln(\sum_{k=1}^{m} e^{V_k - V_j})$ and $\int_{-\infty}^{\infty} \exp(-e^\epsilon - e^{-\epsilon}) \, d\epsilon = 1.$

If the deterministic utility component $V$ is linear in parameters, the log likelihood function of the logit model for an independent sample is concave in parameters. Without loss of generality, suppose that $V_j = x_j \beta$. Let $m_i$ be the number of alternatives available for individual $i$. The choice probability for alternative $j$ for individual $i$ is

$P_{ij} = \frac{\exp(x_{ij} \beta)}{\sum_{k=1}^{m_i} \exp(x_{ik} \beta)}.$

The log likelihood function for an independent sample of size $n$ is

$\ln L = \sum_{i=1}^{n} \sum_{j=1}^{m_i} I_{ij} \ln P_{ij}.$

It follows that

$\frac{\partial \ln L}{\partial \beta} = \sum_{i=1}^{n} \sum_{j=1}^{m_i} I_{ij} \frac{\partial P_{ij}}{\partial \beta}$

$= \sum_{i=1}^{n} \sum_{j=1}^{m_i} I_{ij}(x_{ij} - \bar{x}_i(\beta))',$

where $\bar{x}_i(\beta) = \sum_{k=1}^{m_i} P_{ik} x_{ik}$. This is as follows. As $\ln P_{ij} = x_{ij} \beta - \ln \sum_{k=1}^{m_i} e^{x_{ik} \beta}$, it follows that $\frac{\partial \ln P_{ij}}{\partial \beta} = x_{ij} - \frac{1}{\sum_{j=1}^{m} e^{x_{ij} \beta}} \sum_{k=1}^{m_i} e^{x_{ik} \beta} x_{ik} = (x_{ij} - \sum_{k=1}^{m_i} P_{ik} x_{ik})'$, i.e., $\frac{\partial P_{ij}}{\partial \beta} = P_{ij}(x_i - \sum_{k=1}^{m_i} P_{ik} x_{ik})'$. For the second order derivatives, 

$\frac{\partial^2 P_{ij}}{\partial \beta \partial \beta'} = (x_{ij} - \bar{x}_i(\beta))' \frac{\partial P_{ij}}{\partial \beta} - P_{ij} \sum_{k=1}^{m_i} x_{ik}' \frac{\partial P_{ik}}{\partial \beta'}$

$= (x_{ij} - \bar{x}_i(\beta))' P_{ij}(x_{ij} - \bar{x}_i(\beta)) - P_{ij} \sum_{k=1}^{m_i} x_{ik}' P_{ik}(x_{ik} - \sum_{l=1}^{m_i} P_{il} x_{il})$

$= (x_{ij} - \bar{x}_i(\beta))' P_{ij}(x_{ij} - \bar{x}_i(\beta)) - P_{ij} \sum_{k=1}^{m_i} (x_{ik} - \sum_{l=1}^{m_i} P_{il} x_{il})' P_{ik}(x_{ik} - \sum_{l=1}^{m_i} P_{il} x_{il}),$

and, hence,

$\frac{\partial^2 \ln L}{\partial \beta \partial \beta'} = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \frac{I_{ij}}{P_{ij}} \left[ \frac{\partial^2 P_{ij}}{\partial \beta \partial \beta'} - \frac{1}{P_{ij}} \frac{\partial P_{ij}}{\partial \beta} \frac{\partial P_{ij}}{\partial \beta'} \right]$

$= -\sum_{i=1}^{n} \sum_{j=1}^{m_i} \frac{I_{ij}}{P_{ij}} \sum_{k=1}^{m_i} (x_{ik} - \bar{x}_i(\beta))' P_{ik}(x_{ik} - \bar{x}_i(\beta))'$

$= -\sum_{i=1}^{n} \sum_{k=1}^{m_i} (x_{ik} - \bar{x}_i(\beta))' P_{ik}(x_{ik} - \bar{x}_i(\beta))'.$
The Hessian matrix is non-positive definite and the log likelihood function is concave.

The conditional logit model has a restrictive property known as the independence of irrelevant alternative property (IIA): The probability odds ratio for the $j$th and the $k$th choices is the same irrespective of the total number $m$ of choices. This is so, because

$$\frac{\text{Prob}(I_j = 1|\{1, 2, \ldots, m\})}{\text{Prob}(I_k = 1|\{1, 2, \ldots, m\})} = \frac{e^{V_j}}{e^{V_k}} = \frac{\text{Prob}(I_j = 1|\{j, k\})}{\text{Prob}(I_k = 1|\{j, k\})}.$$  

The IIA property is inappropriate for some situations. An example is the so-called blue bus and red bus paradox. Suppose that each consumer has three alternatives in choosing the red bus, the blue bus or his/her own automobile to work. Let $x_1$ (red bus), $x_2$ (blue bus) and $x_3$ (car) be the attributes of the three transportation modes. Suppose that consumers treat the two buses indifferent and are also indifferent between the automobile mode and the bus mode. In such situation, $\text{Prob}(I_1 = 1|x_1, x_2) = \text{Prob}(I_1 = 1|x_1, x_3) = \text{Prob}(I_2 = 1|x_2, x_3) = 1/2$ and $\text{Prob}(I_1 = 1|x_1, x_2, x_3) = \text{Prob}(I_2 = 1|x_1, x_2, x_3) = 1/4$. It follows that $\frac{\text{Prob}(I_1=1|x_1,x_2,x_3)}{\text{Prob}(I_2=1|x_1,x_2,x_3)} = 1/2$ and $\frac{\text{Prob}(I_1=1|x_1,x_3)}{\text{Prob}(I_3=1|x_1,x_3)} = 1$. The IIA property is not satisfied in this situation. The inconsistency occurs because the two bus modes are perceived as similar alternatives rather than independent alternatives by the individual.