Econometrics 6027
Lecture 3
Binary Choice Models and
Multiple Discrete Choice Models (Continued)

So far, we have addressed the following binary and multiple discrete choice models:

- Probit
- Logit

In this lecture, we will address these additional binary and multiple discrete choice models:

- Ordered Probit
- Ordered Logit
- Sequential Probit/Logit
- Multinomial Probit/Logit
- Simultaneous Bivariate Probit
- Bivariate with selection

This lecture will use indicator notation, and latent variables, quite extensively, so it may be useful for you to review those now. Note also that:

- $f(.)$ represents a PDF
- $F(.)$ represents a CDF
- $\phi(.)$ is the normal PDF
- $\Phi(.)$ is the normal CDF

1 Ordered Probit

1.1 Theoretical framework

Say y_i doesn't take on values of either 0 or 1, but can rather take on a range of values, say, $y_i = 0, 1, 2, ...K$. The underlying, latent value of y_i, though, is continuous:

$$\begin{aligned} \alpha_0 &< Y_i^* \leq \alpha_1 \\ \alpha_1 &< Y_i^* \leq \alpha_2 \\ &\cdots \\ \alpha_K &< Y_i^* \leq \alpha_{K+1} \end{aligned}$$

For example, the true income is $y_i^* = x_i'\beta + u_i$. I observe only y_i and x_i. In this situation, $\alpha_0 < \alpha_1 < \alpha_2 < .... < \alpha_K < \alpha_{K+1}$, that is, the bounds have a natural ordering. This ordinal ranking has meaning: for example, which income box range is ticked off in a survey.

Let y_i have K possible outcomes $y_i = k$, with $k = 1, ..., K$. Natural ordering means that $y_i = k+1$ is in some sense better than $y_i = k$; for example, whether the top degree a prospective employee has is high school, undergrad, masters, or PhD.

We can write the probability of each outcome, where y_i is the observed data (ie. which income range box is ticked off) and y_i^* is the latent index (ie. actual income) as:

$$
\begin{aligned}
(y_i = 0) &= Pr(\alpha_0 < y_i^* \leq \alpha_1) \\
(y_i = 1) &= Pr(\alpha_1 < y_i^* \leq \alpha_2) \\
&..... \\
(y_i = K) &= Pr(\alpha_K < y_i^* \leq \alpha_{K+1})
\end{aligned}
$$

More generally, the conditional probability of observing $y_i = k$ for $k = 1, ..., K$ is

$$
\begin{aligned}
Pr\,(y_i = k|x_i) &= \Pr(\alpha_{k-1} \leq y_i^* \leq \alpha_k) = \Pr(\alpha_{k-1} \leq x_i^{'}\beta + u_i \leq \alpha_k) \\
&= \Pr(\alpha_{k-1} - x_i^{'}\beta \leq u_i \leq \alpha_k - x_i^{'}\beta) \\
&= \Pr(u_i \leq \alpha_k - x_i^{'}\beta) - \Pr(u_i \leq \alpha_{k-1} - x_i'\beta)
\end{aligned}
$$

If u_i is normal, then this equals

$$= F(\alpha_1 - x_i'\beta) - F(\alpha_0 - x_i'\beta)$$

1.2 Estimation of the Ordered Probit Model

The likelihood for ordered discrete models is (where $I(y_i = 0)$ is the indicator function):
for an individual: $L_i = Pr(y_i = 0)^{I(y_i=0)} Pr(y_i = 1)^{I(y_i=1)} ... Pr(y_i = K)^{I(y_i=K)}$
for the whole population: $L^T = \prod_{i=1}^{n} Pr(y_i = 0)^{I(y_i=0)} Pr(y_i = 1)^{I(y_i=1)} .... Pr(y_i = K)^{I(y_i=K)}$

To evaluate the conditional probability one needs to make assumptions about the distribution for $u_{i.}$ Thus,
if u_i is standard normal $\Rightarrow$ Ordered Probit
if u_i is logistic $\Rightarrow$ Ordered Logit

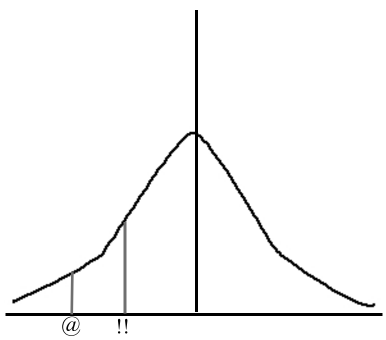
For ordered Probit, $u_i \sim N(0, 1)$, we have:

$$Pr\,(y_i = k|x_i) = F(\alpha_k - x_i'\beta) - F(\alpha_{k-1} - x_i'\beta)$$

More specifically,

$$\begin{aligned}
\Pr(y_i &= 1|x_i) = \Phi(\alpha_1 - x_i'\beta) - \Phi(-x_i'\beta) \\
&\vdots \\
\Pr(y_i &= k|x_i) = \Phi(\alpha_k - x_i'\beta) - \Phi(\alpha_{b-1} - x_i'\beta) \\
&\vdots \\
\Pr(y_i &= K|x_i) = 1 - \Phi(\alpha_{K-1} - x_i'\beta).
\end{aligned}$$

Graphically,



where the first threshold, marked by "@", is $\alpha_{k-1} - x_i\beta$ and the second threshold "!!" is $\alpha_k - x_i\beta$. The probability is the space between the two thresholds; the total probability, the CDF, is the space between the top and the bottom thresholds.

If we assume that the CDF runs from $-\infty$ to ∞, then we assume thresholds $\alpha_0 = -\infty$ and $\alpha_{K+1} = \infty$, then the total likelihood is:

$$L^T = \frac{n}{u}(F(\alpha - x_i\beta)^{I(y=0)}(F(\alpha_2 - x_i\beta)) - (F(\alpha_1 - x_i\beta))^{I(y_i=1)} * ...$$
$$... * (F(\alpha_K - x_i\beta) - F(\alpha_{K-1} - x_i\beta))^{I(y_i=K-1)}(1 - F(\alpha_K - x_i\beta))^{I(y_i=K)})$$

In alternate notation, the likelihood is:

$$L_i = \prod_{k=0}^{K} Pr(y_i = k)^{I(y_i=k)}$$

$$L^T = \prod_{i=1}^{N}\prod_{k=0}^{K} Pr(y_i = k)^{I(y_i=k)}$$

For the purposes of this course, you will be given a threshold and you will have to plug in. If $U \sim N(0, \sigma^2)$, need to divide by σ.

Maximum likelihood estimates solve

$$\left\{\widehat{\alpha}_{ML}, \widehat{\beta}_{ML}\right\} = \arg\max_{\alpha,\beta} (\ln L)$$

2 Ordered Logit

When u_i ~logistic $\left(f(u_i) = \Lambda(u_i)(1-\Lambda(u_i)),\ \text{with } \Lambda(u_i) = \frac{1}{1+e^{-u_i}}\right)$, then

$$\begin{aligned}
\Pr(y_i &= 1|x_i) = \Lambda(\alpha_1 - x_i'\beta) - \Lambda(-x_i'\beta) \\
&\vdots \\
\Pr(y_i &= k|x_i) = \Lambda(\alpha_k - x_i'\beta) - \Lambda(\alpha_{k-1} - x_i'\beta) \\
&\vdots \\
\Pr(y_i &= K|x_i) = 1 - \Lambda(\alpha_{K-1} - x_i'\beta).
\end{aligned}$$

2.1 Estimation of the Ordered Logit Model

Since this model is once again non-linear, we estimate by maximum likelihood methods.

$$\begin{aligned}
\hat{\beta}_L = \arg\max \ln L^T = \arg\max \sum_{i=1}^{N} (I(y_i = 0) \ln \Lambda(\alpha_n - x_i\beta) + ... \\
+ ...I(y_i = 1) \ln(\Lambda(\alpha_2 - x\beta) - \Lambda(\alpha_1 - x\beta)) + ... \\
+ ...I(y_i = K) \ln(\Lambda - \Lambda(\alpha_K - x_i\beta)))
\end{aligned}$$

Analyzing the Probit and Logit

Note that we've assumed here that the probability of one category – say that defined from α_0 to α_1 – is the same as another category – say α_1 to α_2. But that is not necessarily the case. Categories can be different sizes; α_0 to α_1 can be smaller or larger than α_1 to α_2.

The coefficient β is the same for all categories, but some of the smaller categories may not significantly affect the β. The Brant test is a means to test the thresholds we've chosen to define the categories to see if the probability of moving between categories is the same. Essentially, the test groups some of the categories together, keeping the order intact, to see if the estimates for the parameters are the same. If with the grouped categories the estimate of β doesn't change significantly, then the categories should be grouped. The Brant test thus collapses some of the groups that don't significantly change the β.

3 The Sequential Probit and Logit Models

These models can be used when the dependent variable can be separated into a sequence of binary choices.

For the simplest sequential model, assume u_i independent. Examples:
Labour force status
Question: Do you work? Yes/No. If "Yes" part-time/full-time?
Transport model
Question: What type of transportation do you use? Public/Private. If "Public" do you use bus/train?

Consider a sample of data $\{y_{0i}, y_{1i}, x_i, z_i\}$, where
- y_{0i} represent a binary indicator variable for some discrete choice
- y_{1i} represent a second discrete choice, observed only when $y_{0i} = 1$.
- let the k_0 explanatory variables x_i influence the first choice
- let the k_1 explanatory variables z_i influence the conditional choice

(the two sets of explanatory variables can overlap partially or completely)

We model the first stage in this sequence of choices using standard binary choice techniques. Assume:

$$y_{0i}^* = x_i'\beta + u_{0i}$$

where $u_{0i} \sim NID(0, 1)$ for a probit latent relationship.

Observe

$$y_{0i} = 0 \text{ if } y_{0i}^* \leq 0$$
$$y_{0i} = 1 \text{ if } y_{0i}^* > 0$$

Thus,

$$\Pr(y_{0i} = 1 | x_i) = \Phi(x_i'\beta)$$

Estimation by standard Probit MLE on the full sample yields ML estimates $\widehat{\beta}$.

If $y_{0i} = 1$, we have a second stage where we select the n_1 observations for which $y_{0i} = 1$ and define:

$$y_{1i}^* = x_i'\delta + u_{1i}$$

where $u_{1i} \sim NID(0, 1)$ and independent of u_{0i}. We have to assume that u_{0i} is independent of u_{1i} for the model to work. This assumption allows us in the second stage to run a probit model on that part of the sample for which $y_{0i} = 1$.

If we observe $y_{0i} = 1$ in the first stage, the second stage observability criterion is:

$$y_{1i} = 0 \text{ if } y_{1i}^* \leq 0$$
$$y_{1i} = 1 \text{ if } y_{1i}^* > 0$$

Thus,

$$\Pr(y_{1i} = 1|x_i) = \Phi(x_i'\delta).$$

The second stage probability $\Pr(y_{0i} = 1, y_{1i} = 1)$ can be decomposed into a product of the marginal probability and the conditional probability,

$$\Pr(y_{0i} = 1, y_{1i} = 1) = \Pr(y_{0i} = 1)\Pr(y_{1i} = 1|y_{0i} = 1).$$

To estimate the second conditional probability we use the same estimating fashion as for the first stage. Estimate $\widehat{\beta}_1$ by standard Probit MLE on the selected sample.

There are 3 possible outcomes. The overall probabilities of these outcomes are:

$$\begin{aligned}
\Pr(y_{0i} &= 0|x_i) = 1 - \Phi(x_i'\beta_0) \\
\Pr(y_{0i} &= 1, y_{1i} = 0|x_i, z_i) = \Phi(x_i'\beta_0)[1 - \Phi(z_i'\beta_1)] \\
\Pr(y_{0i} &= 1, y_{1i} = 1|x_i, z_i) = \Phi(x_i'\beta_0)\Phi(z_i'\beta_1),
\end{aligned}$$

where

$$\Pr(y_{0i} = 0|x_i) + \Pr(y_{0i} = 1, y_{1i} = 0|x_i, z_i) + \Pr(y_{0i} = 1, y_{1i} = 1|x_i, z_i) = 1.$$

To estimate the full sequential probit, we estimate the first binary choice by univariate Probit ML on the full sample to obtain $\hat{\beta}$ and the second by Probit ML on the selected sample n_i for which $y_{0i} = 1$ to get $\hat{\delta}$. We can only use this trick if the errors are independent.

Advantages: the model is easy to estimate.
Disadvantages: we ignore the potential correlation between the unobservables (u_{1i} is assumed independent of u_{0i}). In addition, there will be fewer degrees of freedom in the estimation of $\hat{\delta}$.

4 The Multinomial Logit

Consider a model with multiple alternatives and unordered outcomes (eg. the decision to use a form of public transportation: bus, train, taxi, subway). There is no obvious ordering or sequence of these alternatives and ordered/sequential and bivariate models are inapplicable.

We can assign probabilities in this case by comparing everything to a benchmark outcome. The Multinomial Logit Model (MLM) expresses the probability of outcome b by comparing it to the probability of benchmark outcome B using the following probability function expression (CDF):

$$F(x'\beta_b) = \frac{P_b}{P_b + P_B}$$

With B discrete alternatives to a discrete choice $b = 1, ..., B$, each with an associated probability $P_{ib} = \Pr(y_i = b)$ for individuals $i = 1, ..., N$. This is a probability function (CDF) because (1) all the individual probabilities sum to 1:

$$P_1 + P_2 + ... + P_B = 1$$

and since $P_b \in (0, 1)$ (2):

$$\begin{aligned} \frac{P_b}{P_b + P_B} &\rightarrow 0 \text{ as } P_b \rightarrow 0 \\ \frac{P_b}{P_b + P_B} &\rightarrow 1 \text{ as } P_b \rightarrow 1. \end{aligned}$$

and because $F(.)$ is a monotone increasing function of its argument (3).

$$\begin{aligned} F(x'\beta_b) &\rightarrow 0 \text{ as } x'\beta_b \rightarrow -\infty \\ F(x'\beta_b) &\rightarrow 1 \text{ as } x'\beta_b \rightarrow \infty \end{aligned}$$

The expression for the CDF implies that

$$\frac{P_b}{P_B} = \frac{F(x'\beta_b)}{1 - F(x'\beta_b)} = \lambda(x'\beta_b)$$

because

$$P_bF + P_BF = P_b \rightarrow P_b - P_bF = P_BF \rightarrow P_b(1-F) = P_BF \rightarrow P_b = P_B\frac{F}{1-F} \rightarrow \frac{P_b}{P_B} = \frac{F}{1-F}$$

If have P_B as a function of λ, then I can have any P_b as a function of λ.

$$\begin{aligned} P_b &= P_B\lambda(x_i'\beta_B) \\ P_1 &= P_B\lambda(x_i'\beta_1) \\ P_2 &= P_B\lambda(x_i'\beta_2) \\ &... \\ P_{B-1} &= P_B\lambda(x_i'\beta_{B-1}) \\ P_B &= P_B \end{aligned}$$

since the left side must sum to 1, so must the right side:

$$P_B\lambda(x_i'\beta_1) + P_B\lambda(x_i'\beta_2) + ... + P_B\lambda(x_i'\beta_{B-1}) + P_B = 1$$

Factoring out the P_B:

$$P_B[\lambda(x_i'\beta_1) + \lambda(x_i'\beta_2) + ... + \lambda(x_i'\beta_{B-1}) + 1] = 1$$

$$P_B = \frac{1}{1 + \sum_{j=1} \lambda(X_i'\beta_\delta)} = \left[1 + \sum_{j=1}^{B-1} \frac{P_j}{P_B}\right]^{-1}$$

and the remaining $B-1$ probabilities:

$$P_b = P_B \lambda(x_i' \beta_B) = \frac{\lambda(x_i'\beta_B)}{1+\sum_{j=1} \lambda(X_i'\beta_\delta)}$$

for all $b = 1, ..., B-1$.

In other words, each probability can be expressed in terms of the explanatory variables and an unknown set of parameters $\{\beta_b, \text{ with } b = 1, ..., B-1\}$. Note however that we can only identify up to $B-1$ β's.

4.1 Estimation of Multinomial Logit Model

Our definition of $\lambda(X_i'\beta)$ was rather arbitrary. Recall that we defined it as:

$$\lambda(x'\beta_b) = \frac{P_b}{P_B} = \frac{F(x'\beta_b)}{1-F(x'\beta_b)}$$

If instead we choose the following functional form:

$$\lambda(X_i'\beta) = e^{X_i'\beta}$$

then the probabilities follow a logistic distribution:

$$\begin{aligned} P_b &= \frac{\lambda(x_i'\beta_b)}{1+\sum_{i=1}^{b1} \lambda(x_i'\beta_j)} \\ &= \frac{e^{x_i'\beta}}{1+\sum_{j=1}^{K-1} e^{x_i'\beta_j}} \end{aligned}$$

where $b = 1, ..., B+1$.

Estimation is done using maximum likelihood estimates. The individual likelihood is:

$$L_i = \prod_{B=1}^{B} P_1^{I(y_i=1)} P_2^{I(y_i=2)} ....P_B^{I(y_i=B)}$$

The total likelihood is:

$$L^T = \prod_{i=1}^{N} \prod_{b=1}^{B} \left(\frac{\lambda(X_i'\beta_1)}{1+\sum_{b=1}^{B-1} \lambda(x_i'\beta_B)}\right)^{I(y_i=1)} \left(\frac{\lambda(x_i'\beta_2)}{1+\sum_{b=1}^{B-1} \lambda(x_i'\beta_B)}\right)^{I(y_i=2)} * ....$$

$$... * \left(\frac{\lambda(x_i'\beta_{B-1})}{1+\sum_{b=1}^{B-1} \lambda(x_i'\beta_B)}\right)^{I(y_i=B-1)} \left(\frac{1}{1+\sum_{b=1}^{B-1} \lambda(x_i'\beta_B)}\right)^{I(y_i=B)}$$

if we assume $\lambda(X_i'\beta) = e^{X_i'\beta}$, this is

$$L^T(\beta_1, ..., \beta_N | y, x) = \prod_{i=1}^{N} \prod_{b=1}^{B} \frac{exp(X_i'\beta)}{1+\sum_{b=1}^{B} exp(x_i'\beta_b)}$$

Advantages of MLM :

- the simplicity of the estimation: we don't need to integrate across multivariate distributions;
- MLM models have "nice" computational properties;
- you can use any outcome as your benchmark: it's your choice. The upper or lower bound, or the midpoint, will do, so long as $P_B \neq 0$;
- if $X_i'\beta$ goes to $-\infty$, then the probability at $P_b = 0$;
- the specification of L_i ensures that if $y_i = 1$ then all others are zero.

Disadvantages of MLM:

- if one probability is zero, then the others can't be zero. If this was the case, then $\frac{F}{1-F}$ would go to infinity if $F = 1$, which would be a problem;
- the "independence of irrelevant alternatives" critique. Recall the formulae for the probabilities in the MLM :

  $$P_b = \frac{\exp(x'\beta_b)}{1 + \sum_{j=1}^{B-1} \exp(x'\beta_j)}, \text{ for all } b = 1, 2, .., B - 1.$$

  However, if we take the ratio of two probabilities P_j and P_k, $\left(\frac{P_j}{P_k}\right)$, we have that

  $$\frac{P_j}{P_k} = \frac{\exp(x'\beta_j)}{\exp(x'\beta_k)}.$$

  In other words the ratio of probabilities of any 2 outcomes is independent of the probability of any other outcome. Adding an extra outcome to the range of choices therefore leaves this ratio of probabilities unchanged. This unreasonable characteristic is known as "The Independence of Irrelevant Alternatives" and forms the major criticism of the MLM.

5 The Simultaneous Bivariate Probit Model

When we have a joint event – such as parents deciding whether to work full or part time, and whether to put their child in daycare – then decisions are taken simultaneously. Unlike the last two models, the two decisions are not independent; indeed, the errors are correlated: $\text{corr}(u_0, u_1) = \rho$, where ρ is the correlation parameter.

The bivariate probit model addresses such a set of simultaneous binary variables, which can be understood as a system. Say y_{0i} takes on values of 0 and 1, and Y_{1i} takes on values of 0 and 1. More formally, consider $\{y_{0i}, y_{1i}, x_i, z_i\}$ for $i = 1, ..., N$, where y_{0i} and y_{1i} represent two binary indicator variables.

Assume that the observed outcomes are driven by a two equation system of latent propensities

$$\begin{aligned} y_{0i}^* &= x_i'\beta + u_{0i}; \\ y_{1i}^* &= z_i'\delta + u_{1i}, \end{aligned}$$

and u_{1i} is assumed correlated to u_{0i}:

$$corr\left(u_{1i}, u_{0i}\right) = \rho.$$

We observe the indicator variables according to the defined thresholds in the observability criteria:

$$\begin{aligned} y_{0i} &= 1(y_{0i}^* > 0); \\ y_{1i} &= 1(y_{1i}^* > 0). \end{aligned}$$

Assuming that u_{0i} and u_{1i} are normally distributed, the model has the following bivariate standard normal density:

$$PDF = \phi_2(u_0, u_{1i}; \rho) = \frac{1}{2\pi(1-\rho^2)^{1/2}} e^{(\frac{-1}{2}\frac{-(u_0^2+u_1^2-2\rho u_0 u_1)}{1-\rho^2}}$$

and bivariate probability:

$$CDF = \Phi_2(u_o, u_1; \rho) = \int_{-\infty}^{u_0}\int_{-\infty}^{u_1} \phi_2(u_0, u_1; \rho)\mathrm{d}v_0\mathrm{d}v_1$$

When $\rho = 0, \phi_2\left(u_0, u_1; 0\right) = \phi\left(u_0\right)\phi(u_1)$.

5.1 Estimating a Bivariate Probit Model

For the bivariate model there are 4 possible combinations of observed outcomes. From the observability criteria we can derive probabilities P_{jk} for $j, k = 0, 1$ associated to each of these combinations and for any set of parameters.

1. $y_0 = 0$; $y_1 = 0$

$$\begin{aligned} Pr(y_{0i} = 0, y_{1i} = 0|x, z) &= Pr(y_{0i}^* \leq 0, y_{1i}^* \leq 0|x, z) \\ &= Pr(u_{0i} \leq -x_i\beta, u_{1i} \leq z'\delta|x, z, \delta, \rho) \\ &= \Phi_2(-x_{0i}'\beta, -z_i'\delta|x, z, \beta, \delta, \rho) \\ &= \int_{-\infty}^{-X_i'\beta}\int_{-\infty}^{-Z_i'\delta} \phi(u_0, u_1, \rho)\mathrm{d}v_0\mathrm{d}v_1 \end{aligned}$$

2. $y_0 = 1$; $y_1 = 1$

$$\begin{aligned} Pr(y_0 = 1, y_1 = 1|x_0, z_1) &= Pr(y_{0i}^* > 0, y_{1i}^* > 0|x, z, \beta, \delta, \rho) \\ &= Pr(u_{0i} > -x_i\beta, u_{1i} > -z_i\delta) \end{aligned}$$

since the distribution is symmetric, this area under the distribution to the right of a lower threshold set at $-x_i\beta$ is equal to the area to the left of an upper threshold set at $x_i\beta$. So, by symmetry,

$$\begin{aligned} &= \Phi_2(x_i\beta, z_i\delta|x, z, \beta, \delta, \rho) \\ &= \int_{-\infty}^{x_i'\beta}\int_{-\infty}^{z_i'\delta} \phi(u_0, u_1, \rho)\mathrm{d}v_0\mathrm{d}v_1 \end{aligned}$$

3. $y_0 = 0;\ y_1 = 1$

$$\begin{aligned} Pr(y_0 = 0, y_1 = 1|x_0, z_1) &= Pr(y^*_{0i} \leq 0, y^*_{1i} > 0|x, z, \beta, \delta, \rho) \\ &= Pr(u_{0i} \leq -x_i\beta, u_{1i} > -z_i\delta) \\ \text{By symmetry,} & \\ &= Pr(u_{0i} \leq -x_i\beta, u_{1i} < z_i\delta) \\ &= \Phi_2(-x_i\beta, z_i\delta|x, z, \beta, \delta, \rho) \\ &= \int_{-\infty}^{-x'_i\beta}\int_{-\infty}^{z'_i\delta} \phi(u_0, u_1, \rho)\mathrm{d}v_0\mathrm{d}v_1 \end{aligned}$$

4. $y_0 = 1;\ y_1 = 0$

$$\begin{aligned} Pr(y_0 = 1, y_1 = 0|x_0, z_1) &= Pr(y^*_{0i} > 0, y^*_{1i} \leq 0|x, z, \beta, \delta, \rho) \\ &= Pr(u_{0i} > -x_i\beta, u_{1i} \leq -z_i\delta) \\ \text{By symmetry,} & \\ &= Pr(u_{0i} < x_i\beta, u_{1i} \leq -z_i\delta) \\ &= \Phi_2(x_i\beta, -z_i\delta|xz\beta\delta\rho) \\ &= \int_{-\infty}^{x'_i\beta}\int_{-\infty}^{-z'_i\delta} \phi(u_0, u_1, \rho)\mathrm{d}v_0\mathrm{d}v_1 \end{aligned}$$

Use a maximum likelihood estimator to estimate across all of these possible cases:

$$Pr(y_0 = 1, y_1 = 1) + Pr(y_0 = 1, y_1 = 0) + Pr(y_0 = 0, y_1 = 1) + Pr(y_0 = 0, y_1 = 0) = 1$$

$$L_i = \text{Product of these probabilities}$$

$$= \Phi_2(-x_1\beta, -z_i\delta|x, z; \beta, \delta, \rho)^{(1-y_0)(1-y_1)}\Phi_2(-x_1\beta, -z_i\delta|.)^{(1-y_0)y_1} *$$
$$\Phi_2(-x_1\beta, -z_i\delta|.)^{y_0(1-y_1)}\Phi_2(-x_1\beta, -z_i\delta|.)^{y_0y_1}$$

$$L^T = \prod_{i=1}^{N} L_i$$

$$\ln L = \sum_{i=1}^{N}\{(1 - y_{0i})\,(1 - y_{i1})\ln(y_0 = 0, y_1 = 0) + (1 - y_{0i})\,y_{1i}\ln(y_0 = 0, y_1 = 1)$$
$$+ y_{0i}\,(1 - y_{1i})\ln(y_0 = 1, y_1 = 0) + y_{0i}y_{1i}\ln(y_0 = 1, y_1 = 1)$$

where $(1 - y_0)(1 - y_1)$ is the same as $I(y_0 = 0, y_1 = 0)$. We estimate β, δ, ρ.

Advantages of MLM :

- can express outcomes as a function of threshold ie. in case of unobserved utility;
- can extend the model to the multivariate case.

Disadvantages of MLM :

- tedious calculations.

6 Bivariate with Selection

This model is sequential, but estimated in a bivariate framework.

$$y_{1i} = z_i' + I(y_{i0} = 1|X\beta\gamma)$$

where y_{i0} is an endogenous regressor. If we assume correlation between the u_0s in the regressions of y_{0i} and y_{1i}, then we have to estimate jointly to account for this correlation through a selection regression. The model that does this is called bivariate with selection.

$$\begin{aligned} Y_{0i} &= X_i'\beta + u_{0i} \\ Y_{1i} &= Z_i'\delta + u_{1i} \\ \rho &= corr(u_{1i}, u_{0i}) \end{aligned}$$

Where correlation is present, even though the decisions may be sequential, there is endogeneity so we can't do a probit on each. The bivariate with selection model uses the approach of the bivariate probit to estimate in this context.