Univariate Probit Model

Unobservable utility index (y^*) or latent var.

$$y^* = x\beta + \varepsilon, \quad y = 1 \quad if \quad y^* > 0, \; 0 \quad otherwise$$

Then,

$$P_i = \Pr(y = 1 | x) = P(y^* > 0) = F(x\beta) = \Phi(x\beta)$$

Assume $\Phi(.)$ is cumulative standardized normal distribution function.

Bivariate Probit Models

Two binary choice dependent variables.

The models assume standardized bivariate normal distribution.

$$y_1^* = x_1\beta_1 + \varepsilon_1, \qquad y_1 = 1 \quad if \quad y_1^* > 0, \quad 0 \quad otherwise$$
$$y_2^* = x_2\beta_2 + \varepsilon_2, \qquad y_2 = 1 \quad if \quad y_2^* > 0, \quad 0 \quad otherwise$$

$$\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)$$

ρ is Correlation between ε_1 and ε_2

Bivariate Probit Models

The probability function of the model can be stated as:

$$P(y_1 = 1, y_2 = 1) = P(y_1^* > 0, y_2^* > 0)$$

$$= P(\varepsilon_1 < x_1\beta_1, \varepsilon_2 < x_2\beta_2)$$

$$= \int_{-\infty}^{x_1\beta_1} \int_{-\infty}^{x_2\beta_2} \phi(z_1, z_2, \rho)\, dz_1 dz_2$$

$$= \Phi(x_1\beta_1, x_2\beta_2, \rho)$$

where: $\Phi(.)$ is Cumulative Standardized Bivariate Normal Distribution Function.

Bivariate Probit Models

Maximum Likelihood Estimation method is employed to estimate the model.

However, ρ cannot be directly estimated.

Inverse hyperbolic tangent $(\operatorname{atanh} \rho)$ is estimated instead.

$$\operatorname{atanh} \rho = \frac{1}{2} \ln\left(\frac{1+\rho}{1-\rho}\right)$$

Bivariate Probit Models

<u>Test of Bivariate Probit</u>

If $\rho = 0$, the log-likelihood of Bivariate Probit $(\log L_{BVP})$ is equal to sum of log-likelihood of two separate Univariate Probit models $(\log L_P)$.

Likelihood Ratio (LR) test compared between Bivariate Probit and two separate Univariated Probit can be performed to test the appropriateness of Bivariate Probit Models.

$$LR=2\left(\log L_{BVP} - \left[\log L_{P_1} + \log L_{P_2}\right]\right) \sim \chi_1^2$$

Multivariate Probit Models

More than two binary choice dependent variables. The model assumes multivariate normal distribution.

$$y_1^* = x_1\beta_1 + \varepsilon_1, \qquad y_1 = 1 \;\; if \;\; y_1^* > 0, \;\; 0 \;\; otherwise$$

$$y_2^* = x_2\beta_2 + \varepsilon_2, \qquad y_2 = 1 \;\; if \;\; y_2^* > 0, \;\; 0 \;\; otherwise$$

$$\vdots$$

$$y_k^* = x_k\beta_k + \varepsilon_k, \qquad y_k = 1 \;\; if \;\; y_k^* > 0, \;\; 0 \;\; otherwise$$

$$\begin{matrix} \varepsilon_1 & & 0 & 1 & & & \\ \varepsilon_2 & & 0 & \rho_{21} & 1 & & \end{matrix}$$

[redacted]

$$\begin{matrix} \varepsilon_k & & 0 & \rho_{k1} & \rho_{k2} & \cdots & 1 \end{matrix}$$

ρ_{ij} is Tetrachoric Correlation between ε_i and ε_j

Multivariate Probit Models

<u>Example</u> - 4 binary choice dependent variables:

$$
\begin{aligned}
y_1^* &= x_1\beta_1 + \varepsilon_1, \qquad & y_1 &= 1 \quad if \quad y_1^* > 0, \ \ 0 \ \ otherwise \\
y_2^* &= x_2\beta_2 + \varepsilon_2, \qquad & y_2 &= 1 \quad if \quad y_2^* > 0, \ \ 0 \ \ otherwise \\
y_3^* &= x_3\beta_3 + \varepsilon_3, \qquad & y_3 &= 1 \quad if \quad y_3^* > 0, \ \ 0 \ \ otherwise \\
y_4^* &= x_4\beta_4 + \varepsilon_4, \qquad & y_4 &= 1 \quad if \quad y_4^* > 0, \ \ 0 \ \ otherwise
\end{aligned}
$$

$$
\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & & & \\ \rho_{21} & 1 & & \\ \rho_{31} & \rho_{32} & \ddots & \\ \rho_{41} & \rho_{42} & \rho_{43} & 1 \end{bmatrix} \right)
$$

ρ_{ij} is Tetrachoric Correlation between ε_i and ε_j

Multivariate Probit Models

The probability function of the model can be stated as:

$$P(y_1 = 1, y_2 = 1, y_3 = 1, y_4 = 1) = P(y_1^* > 0, y_2^* > 0\ y_3^* > 0\ y_4^* > 0)$$

$$= P(\varepsilon_1 < x_1\beta_1, \varepsilon_2 < x_2\beta_2, \varepsilon_3 < x_3\beta_3, \varepsilon_4 < x_4\beta_4)$$

$$= \int_{-\infty}^{x_1\beta_1}\int_{-\infty}^{x_2\beta_2}\int_{-\infty}^{x_3\beta_3}\int_{-\infty}^{x_4\beta_4} \phi\left(z_1, z_2, z_3, z_4, \rho_{21}, \rho_{31}, \rho_{41}, \rho_{32}, \rho_{42}, \rho_{43}\right) dz_1 dz_2 dz_3 dz_4$$

$$= \Phi\left(x_1\beta_1, x_2\beta_2, x_3\beta_3, x_4\beta_4, \rho_{21}, \rho_{31}, \rho_{41}, \rho_{32}, \rho_{42}, \rho_{43}\right)$$

where: $\Phi(.)$ is Cumulative Standardized Multivariate Normal Distribution Function.

Multivariate Probit Models

Since the likelihood function of the models has no close-form solution, the models are estimated by Maximum Simulated Likelihood Estimation method.

However, ρ_{ij} cannot be directly estimated.

$\text{atanh}\,\rho_{ij}$ are estimated instead.

$$\text{atanh}\,\rho_{ij} = \frac{1}{2}\ln\left(\frac{1+\rho_{ij}}{1-\rho_{ij}}\right)$$

Multivariate Probit Models

Test of Multivariate Probit

If all $\rho_{ij} = 0$, the log-likelihood of Multivariate Probit is equal to sum of log-likelihood of all separate Univariate Probit models.

Likelihood Ratio (LR) test compared between Multivariate Probit and all separate Univariated Probit models can be performed to test the appropriateness of Multivariate Probit Models.