# LECTURE 15
# BINARY RESPONSE MODELS

In this lecture, we discuss models in which the dependent can take on only two values, usually zero or one. For example, $Y_i$ can measure whether or not individual $i$ participates in labor force: $Y_i = 1$ (yes) or $Y_i = 0$ (not). As we will see later, in such cases, the linear regression model might not be an appropriate tool for the analysis.

## Bernoulli trials

The econometrician observes $\{Y_i : i = 1, \ldots, n\}$ where $Y_i$'s are iid random variables. The distribution of $Y_i$ is given by

$$Y_i = \begin{cases} 1 & \text{with probability } \theta, \\ 0 & \text{with probability } 1-\theta. \end{cases}$$

Such a distribution is called Bernoulli($\theta$). Note that

$$\begin{aligned} EY_i &= \theta \\ &= P(Y_i = 1). \end{aligned} \tag{1}$$

We can write the PMF of $Y_i$ as

$$p(y_i, \theta) = \theta^{y_i} (1-\theta)^{1-y_i} \text{ for } y_i = 0, 1.$$

Thus, the log-likelihood function is given by

$$\begin{aligned} \log L_n(\theta) &= \left( n^{-1} \sum_{i=1}^n Y_i \right) \log \theta + \left( 1 - n^{-1} \sum_{i=1}^n Y_i \right) \log(1-\theta) \\ &= \overline{Y}_n \log \theta + \left(1 - \overline{Y}_n\right) \log(1-\theta), \end{aligned}$$

where $\overline{Y}_n = n^{-1} \sum_{i=1}^n Y_i$. It follows that the ML estimator of $\theta$ is the average value of $Y_i'$s:

$$\widehat{\theta}_n^{ML} = \overline{Y}_n.$$

Next,

$$\frac{d^2 \log p(Y_i, \theta)}{d\theta^2} = -\frac{Y_i}{\theta^2} - \frac{1-Y_i}{(1-\theta)^2},$$

Equation (1) implies that

$$\begin{aligned} E\frac{d^2 \log p(Y_i, \theta)}{d\theta^2} &= -\frac{1}{\theta} - \frac{1}{1-\theta} \\ &= -\frac{1}{\theta(1-\theta)}. \end{aligned}$$

Thus, in this model, the information (scalar) is given by

$$I(\theta) = \frac{1}{\theta(1-\theta)},$$

and the results in Lecture 14 imply that

$$\begin{aligned} \overline{Y}_n &\to_p \theta, \\ n^{1/2}\left(\overline{Y}_n - \theta\right) &\to_d N(0, \theta(1-\theta)). \end{aligned}$$

We can estimate the asymptotic variance consistently by $\overline{Y}_n \left(1 - \overline{Y}_n\right)$. The $1 - \alpha$ asymptotic confidence interval for $\theta$ can be constructed as follows:

$$\left[\overline{Y}_n \pm z_{1-\alpha/2}\sqrt{\frac{\overline{Y}_n \left(1 - \overline{Y}_n\right)}{n}}\right].$$

The Bernoulli trials is a univariate model. Next, we extend it to the case where the probability of $Y_i$ taking on 1 is a function of some exogenous explanatory variables.

# Linear probability model

Suppose that the econometrician observes $\{(Y_i, X_i) : i = 1, \ldots, n\}$, where $Y_i$'s are binary $\{0, 1\}$ variables, and $X_i$'s are $k$-vectors of exogenous variables that explain $P(Y_i = 1|X_i)$. Let's assume that

$$P(Y_i = 1|X_i) = F(X_i'\beta),$$

for some function $F$. Contrary to the Bernoulli trials model, the probability of $Y_i$ taking on the value one or zero is not constant and depends on some observable characteristics $X_i$.

If we assume that $Y_i$ and $X_i$ are related by the liner regression model, i.e.

$$E(Y_i|X_i) = X_i'\beta,$$

then we obtain that

$$F(X_i'\beta) = X_i'\beta.$$

In other words, the probability of $Y_i$ taking on 1 is a linear function of $X_i$. Such an assumption is a good starting point for the analysis, however, it suffers from a serious flaw. If some of the $X$'s are continuous, and since $X_i'\beta$ cannot be restricted to the zero-one interval, the model will produce probabilities that are negative or greater than one. Thus, in order to avoid nonsense predictions, one has to consider functions $F$ restricted to zero-one interval, and therefore nonlinear. A natural choice for $F$ is any CDF.

# Probit and logit models

It is convenient to introduce a latent (unobservable) variable $Y_i^*$ for which the usual linear regression model holds:

$$Y_i^* = X_i'\beta + U_i.$$

We assume further that $\{(Y_i^*, X_i) : i = 1, \ldots, n\}$ are iid, and that

$$Y_i = \begin{cases} 1 & \text{if } Y_i^* > 0, \\ 0 & \text{if } Y_i^* \leq 0. \end{cases}$$

Let $F$ be the conditional CDF of $U_i$ given $X_i$:

$$F(u|X_i) = P(U_i \leq u|X_i)$$

Then,

$$\begin{aligned} P(Y_i = 1|X_i) &= P(Y_i^* > 0|X_i) \\ &= P(X_i'\beta + U_i > 0|X_i) \\ &= P(U_i > -X_i'\beta|X_i) \\ &= 1 - F(-X_i'\beta). \end{aligned}$$

A usual assumption in this framework is that $F$ is symmetric around zero, i.e.

$$F(u) = 1 - F(-u).$$

Under this assumption, we obtain that

$$P(Y_i = 1|X_i) = F(X_i'\beta).$$

The common choices for $F$ are

- Probit model: $F$ is the standard normal CDF, denoted by $\Phi$.
- Logit model: $F$ is logistic CDF, denoted by $\Lambda$:

$$\begin{aligned}\Lambda(X_i'\beta) &= \frac{\exp(X_i'\beta)}{\exp(X_i'\beta)+1} \\ &= \frac{1}{1+\exp(-X_i'\beta)}.\end{aligned}$$

For either choice of $F$, the model is estimated by the ML method. The PMF of $Y_i$ conditional on $X_i = x_i$ is similar to the PMF of $Y_i$ in the Bernoulli trials model, however with $\theta$ replaced by $F(x_i'\beta)$:

$$p(y_i, \beta|x_i) = F(x_i'\beta)^{y_i}(1 - F(x_i'\beta))^{1-y_i} \text{ for } y_i = 0, 1.$$

Thus, the log-likelihood is given by

$$\log L_n(\beta) = n^{-1}\sum_{i=1}^{n} Y_i \log F(X_i'\beta) + n^{-1}\sum_{i=1}^{n}(1-Y_i)\log(1-F(X_i'\beta)).$$

The first-order condition for $\widehat{\beta}_n^{ML}$ is

$$\begin{aligned}0 &= \frac{\partial \log L_n\left(\widehat{\beta}_n^{ML}\right)}{\partial\beta} \\ &= n^{-1}\sum_{i=1}^{n} Y_i \frac{\partial F\left(X_i'\widehat{\beta}_n^{ML}\right)/\partial\beta}{F\left(X_i'\widehat{\beta}_n^{ML}\right)} - n^{-1}\sum_{i=1}^{n}(1-Y_i)\frac{\partial F\left(X_i'\widehat{\beta}_n^{ML}\right)/\partial\beta}{1-F\left(X_i'\widehat{\beta}_n^{ML}\right)} \\ &= n^{-1}\sum_{i=1}^{n} Y_i \frac{f\left(X_i'\widehat{\beta}_n^{ML}\right)}{F\left(X_i'\widehat{\beta}_n^{ML}\right)} X_i - n^{-1}\sum_{i=1}^{n}(1-Y_i)\frac{f\left(X_i'\widehat{\beta}_n^{ML}\right)}{1-F\left(X_i'\widehat{\beta}_n^{ML}\right)} X_i,\end{aligned}$$

where $f$ is the PDF of $F$:

$$f(u) = \frac{dF(u)}{du}.$$

There is no closed form expression for $\widehat{\beta}_n^{ML}$ and it must be computed numerically. Statistical software packages such as Eviews, Stata, SAS can produce the ML estimates and their asymptotic standard errors for probit and logit models given the data on $Y_i$ and $X_i$.

In the case of logit, the PDF of $\Lambda$ satisfies

$$\frac{d\Lambda(u)}{du} = \Lambda(u)(1-\Lambda(u)). \tag{2}$$

Therefore, for the logit model,

$$\begin{aligned} \frac{f\left(X_i'\beta\right)}{F\left(X_i'\beta\right)} &= 1-\Lambda\left(X_i'\beta\right), \\ \frac{f\left(X_i'\beta\right)}{1-F\left(X_i'\beta\right)} &= \Lambda\left(X_i'\beta\right). \end{aligned}$$

The first-order condition simplifies to

$$\begin{aligned} \frac{\partial \log L_n\left(\widehat{\beta}_n^{ML}\right)}{\partial \beta} &= n^{-1}\sum_{i=1}^{n}\left(Y_i-\Lambda\left(X_i'\widehat{\beta}_n^{ML}\right)\right)X_i \\ &= 0. \end{aligned}$$

Similarly, we have that

$$\frac{\partial \log p\left(y_i,\beta|x_i\right)}{\partial \beta} = \left(y_i-\Lambda\left(x_i'\beta\right)\right)x_i,$$

and

$$\frac{\partial^2 \log p\left(y_i,\beta|x_i\right)}{\partial \beta \partial \beta'} = -\Lambda\left(x_i'\beta\right)\left(1-\Lambda\left(x_i'\beta\right)\right)x_i x_i'.$$

Thus, the information matrix is given by

$$\begin{aligned} I\left(\beta\right) &= -E\frac{\partial^2 \log p\left(Y_i,\beta|X_i\right)}{\partial \beta \partial \beta'} \\ &= E\Lambda\left(X_i'\beta\right)\left(1-\Lambda\left(X_i'\beta\right)\right)X_i X_i'. \end{aligned}$$

Hence, we have that, in the case of the logit model $\widehat{\beta}_n^{ML} \to_p \beta$, and

$$n^{1/2}\left(\widehat{\beta}_n^{ML}-\beta\right) \to_d N\left(0,\left(E\Lambda\left(X_i'\beta\right)\left(1-\Lambda\left(X_i'\beta\right)\right)X_i X_i'\right)^{-1}\right).$$

The asymptotic variance-covariance matrix of $\widehat{\beta}_n^{ML}$ can be estimated consistently by

$$\left(n^{-1}\sum_{i=1}^{n}\Lambda\left(X_i'\widehat{\beta}_n^{ML}\right)\left(1-\Lambda\left(X_i'\widehat{\beta}_n^{ML}\right)\right)X_i X_i'\right)^{-1}.$$

# Marginal effects

In the linear regression model, the marginal effect of $X_i$ on $E\left(Y_i|X_i\right)$ is given by the slope coefficients $\beta$. In the case of nonlinear binary choice models like probit or logit,

$$\begin{aligned} \frac{\partial E\left(Y_i|X_i\right)}{\partial X_i} &= \frac{\partial P\left(Y_i=1|X_i\right)}{\partial X_i} \\ &= \frac{\partial F\left(X_i'\beta\right)}{\partial X_i} \\ &= f\left(X_i'\beta\right)\beta, \end{aligned}$$

and

$$\begin{aligned} \frac{\partial P\left(Y_i=0|X_i\right)}{\partial X_i} &= \frac{\partial\left(1-P\left(Y_i=1|X_i\right)\right)}{\partial X_i} \\ &= -f\left(X_i'\beta\right)\beta. \end{aligned}$$

Thus, in the case of the probit model, the marginal effect of $X_i$ on $P(Y_i = 1|X_i)$ is given by

$$\phi(X_i'\beta)\,\beta,$$

where $\phi(u) = d\Phi(u)/du$ is the standard normal density. In the case of logit, equation (2) implies that $\partial P(Y_i = 1|X_i)/\partial X_i$ is given by

$$\Lambda(X_i'\beta)\,(1 - \Lambda(X_i'\beta))\,\beta.$$

The marginal effects can be estimated by replacing the unknown coefficients $\beta$ with their ML estimators. Fix $X_i = x$, then

$$\frac{\partial \widehat{P(Y_i = 1|X_i = x)}}{\partial X_i} = f\left(x'\widehat{\beta}_n^{ML}\right)\widehat{\beta}_n^{ML}.$$

The Slutsky's Theorem and consistency of the ML estimator of $\beta$ imply that

$$\begin{aligned}\frac{\partial \widehat{P(Y_i = 1|X_i)}}{\partial X_i} &= f\left(x'\widehat{\beta}_n^{ML}\right)\widehat{\beta}_n^{ML} \\ &\to_p f(x'\beta)\,\beta \\ &= \frac{\partial P(Y_i = 1|x)}{\partial X_i},\end{aligned}$$

provided that $f$ is continuous.

The asymptotic distribution of $f\left(x'\widehat{\beta}_n^{ML}\right)\widehat{\beta}_n^{ML}$ can be obtained using the delta method:

$$\begin{aligned}n^{1/2}\left(f\left(x'\widehat{\beta}_n^{ML}\right)\widehat{\beta}_n^{ML} - f(x'\beta)\,\beta\right) &\to_d N\left(0, \frac{\partial\left(f(x'\beta)\,\beta\right)}{\partial\beta'} I^{-1}(\beta)\,\frac{\partial\left(f(x'\beta)\,\beta'\right)}{\partial\beta}\right) \\ &= N\left(0, \left(\frac{\partial f(x'\beta)}{\partial u}\beta x' + f(x'\beta)\,I_k\right) I^{-1}(\beta)\left(\frac{\partial f(x'\beta)}{\partial u}\beta x' + f(x'\beta)\,I_k\right)'\right).\end{aligned}$$