Modeling

Binary Outcomes : Logit & Probit models

Motivating Example : Model women’s labor force participation

\[ y_i = 1 \text{ if a woman is in the paid labor force} \]
\[ = 0 \text{ otherwise} \]

\[ \epsilon_i \text{ is vector of observed covariates} \]
\[ \epsilon \text{ e.g. # of children} \]
\[ \text{age} \]
\[ \text{education} \]
\[ \text{wage rate (estimated)} \]

Linear model formulation

\[ y_i = \epsilon_i \beta + \epsilon_i \]
\[ i = 1, \ldots, N \]

Note: When \[ y_i = 1 \] \[ \implies \epsilon_i = 1 - \epsilon_i \beta \]
\[ y_i = 0 \] \[ \implies \epsilon_i = -\epsilon_i \beta \]

Interpretation of Regression model

\[ E[y_i | x_i] = 1 \cdot \Pr (y_i = 1 | x_i) + 0 \cdot \Pr (y_i = 0 | x_i) \]
\[ = \Pr (y_i = 1 | x_i) = \epsilon_i \beta \]
So that

\[
\frac{d \text{E}[y_i | x_i]}{d x_i} = \frac{2 \Pr(Y_i = 1 | x_i)}{2 x_i} = \beta
\]

\[
\Rightarrow \beta = \text{change in } \Pr(Y_i = 1 | x_i) \text{ for a unit change in } x_i
\]

**Graphical illustration for I RHS variance**

\[
E(y_i | x_i) = \beta_0 + \beta_1 x_i = \Pr(Y_i = 1 | x_i)
\]

As \( x \) increases, \( \Pr(Y_i = 1) \) increases

Note: Probabilities are constrained to lie between 0 and 1.

Effect of changes in \( x \) on \( \Pr(Y_i = 1) \) is the same for all values of \( x \) — may not be realistic.
Comments on Empirical Analysis of linear Probability model for Women's labor Force Participation

(i) Effect of a variable is the same regardless of the values of the other variables
   i.e. $p_i$ is not affected by $x_j$

(ii) The effect of a unit change for a variable is the same regardless of the current value of that variable

   e.g. If a woman has 4 young children compared to no children, her predicted probability of employment decreases by 1.18 ($= 4 \times -0.295$) which is obviously unrealistic!

Problems with the linear probability model

(i) $Y_i$ is heterogeneous

   $\text{Var}(Y_i) = \Pr(Y_i=1|x_i) \cdot \Pr(Y_i=0|x_i)$
   $= \xi \phi (1 - \xi \beta)$

(ii) $E_i$ cannot be normally distributed.

(iii) Predicted probabilities can be less than zero or greater than 1
\( (iv) \quad \frac{d P(Y_i = 1 | x_i)}{d x_i} = \beta \) is unrealistic in many situations

e.g. expect each additional child to have a diminishing effect on \( P(Y_i = 1 | x_i) \) maximal

\( \Rightarrow \) Non linear effects are more realistic.

"Latent Variable (Index Model) Formulation"

\( y_i = 1, 0 \) : observed variable

\( y_i^* \in \mathbb{R}^+ \) is an unobserved latent or index variable

Idea: large values of \( y_i^* \) generate \( y_i = 1 \)
small values of \( y_i^* \) "\( y_i = 0 \"

Example: \( y_i = 1 \) if woman in LF
\( = 0 \) otherwise

\( y_i^* = \) underlying propensity to work for a particular woman \( \beta \) is the economic theory underlying labor-leisure trade off
Assume

\[ y_i^x = x_i' \beta + \varepsilon_i \]

and

\[ y_i^* = \begin{cases} T & \text{if } y_i^x > T \text{ (threshold)} \\ 0 & \text{if } y_i^x \leq T \end{cases} \]

So when \( y_i^* \) cross the threshold (actual wave > reservation wave), then observe \( y_i = 1 \)

\[ \Pr (y_i = 1 | y_i^* > T) = \Pr (y_i^* > 0 | y_i) \]

\[ = \Pr (x_i' \beta + \varepsilon_i > 0 | x_i') \]

\[ = \Pr (\varepsilon_i > -x_i' \beta | x_i') \]

\[ = \Pr (E_i \leq x_i' \beta) \text{ for symmetric dist.} \]
\[ \Pr(Y_i = 1|x_i) = \Pr(t_i \leq x_i'\beta) \text{ for some } \beta \]

Graphically:

\[ \Pr(Y_i = 1) = \Pr(t_i \leq x_i'\beta) = CDF(x_i'\beta) \]

Notice that the latent variable formulation produces a "non-linear" probability model for \( \Pr(Y_i = 1|x_i) \). By construction, \( \Pr(Y_i = 1|x_i) \) lies between 0 and 1 because it is based on a CDF for \( t_i \).

Notice that

\[
\lim_{x_i'\beta \to \infty} \Pr(Y_i = 1|x_i) = 1 \\
\lim_{x_i'\beta \to -\infty} \Pr(Y_i = 1|x_i) = 0.
\]
So $\Pr(Y_i = 1 | \xi_i)$ depends on the distribution of $\xi_i$ in the latent variable formulation.

Two popular choices for $\xi_i$ are:

(i) $\xi_i \sim N(0, 1)$ \& Probit model

$$
\Pr(Y_i = 1 | \xi_i) = \Pr(\xi_i \leq \xi_i' \beta) = \Phi(\xi_i' \beta)
$$

where

$$
\Phi(z) = \int_{-\infty}^{z} \phi(x) \, dx, \quad \phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} x^2}
$$

(ii) $\xi_i \sim \text{Logistic} : \text{Logit model}$

$$
\Pr(Y_i = 1 | \xi_i) = \Pr(\xi_i \leq \xi_i' \beta) = \Lambda(\xi_i' \beta)
$$

where

$$
\Lambda(z) = \frac{e^{z}}{1 + e^{z}} = \frac{1}{e^{-z} + 1}
$$
Note: \[ \Lambda (t) = \int_{-\infty}^{t} \lambda (x) \, dx \]

where

\[ \lambda (t) = \frac{d}{d\logit} \Lambda (t) = \Lambda (t) (1 - \Lambda (t)) \]

\[ \logit \quad \text{pdf} \quad \logit \quad \text{CDF.} \]

\[
= \left( \frac{e^t}{1 + e^t} \right) \left( 1 - \frac{e^t}{1 + e^t} \right) \\
= \left( \frac{e^t}{1 + e^t} \right) \left( \frac{1}{1 + e^t} \right) \\
= \frac{e^t}{(1 + e^t)^2}
\]

Remarks

If \( \epsilon_i \sim \text{logistic} \) then

(i) \( E \{ \epsilon_i \} = 0 \)

(ii) \( \text{var} \{ \epsilon_i \} = \frac{\pi^2}{3} \)

\[ \Rightarrow \] logistic distribution is similar to Student-t with 7 d.f.
Probabilities differ mostly in the tails of the distribution.
Identifying in the OPM

In the logit and probit there are important identifying assumptions

(i) \( E[\epsilon_i | x_i] = 0 \)

(ii) \( \text{var}(\epsilon_i | x_i) \) is constant

(iii) \( Z = 0 \) if identifies the constant or intercept.

These assumptions allow the identification of \( \beta \).
Suppose $\varepsilon \sim N(0, \sigma^2)$ and $y_i^* = \beta_0 + \beta_1 x_i + \varepsilon_i$

Then $\Pr(y_i = 1) = \Pr(y_i^* > 0)$

\[
= \Phi \left( \frac{\beta_0 + \beta_1 x_i + \varepsilon_i}{\sigma} \right)
\]

\[
= \Pr \left( \frac{\varepsilon_i}{\sigma} \leq \frac{\beta_0 + \beta_1 x_i}{\sigma} \right)
\]

\[
= \Pr \left( \frac{\varepsilon_i}{\sigma} \leq \frac{\beta_0}{\sigma} + \frac{\beta_1}{\sigma} x_i \right)
\]

For different values of $\sigma$ we get different values of $\beta_0^* + \beta_1^*$, by setting $\sigma = 1$ a priori we can identify the $\beta$'s.

Can only estimate $\beta_0^* = \frac{\beta_0}{\sigma}$ and $\beta_1^* = \frac{\beta_1}{\sigma}$

and not $\beta_0$, $\beta_1$, and $\sigma$ individually.
For the logit we have

$$\frac{d \Pr(Y_i = 1 | x_i)}{d x_{ki}} = \Lambda(x_i' \beta)(1 - \Lambda(x_i' \beta)) \cdot \beta_k$$

and for the Probit

$$\frac{d \Phi(Y_i = 1 | x_i)}{d x_{ki}} = \varphi(x_i' \beta) \cdot \beta_k$$

Give more examples.
Maximum likelihood estimation of Probit and Logit models

\[ Y_i^* = x_i' \beta + \epsilon_i \] , \( \epsilon_i \) has CDF \( F \) and p.d.f. \( f \)

\[ y_i = 1 \text{ if } Y_i^* > 0 \]
\[ = 0 \text{ if } Y_i^* \leq 0 \]

\( \epsilon_i \) iid \( \mathcal{N}(0,1) \) and \( F(z) = \Phi(z) \): Probit

\( \epsilon_i \) iid logistic and \( F(z) = \Lambda(z) \): Logit

Since \( y_i \) only takes values 0 and 1 we can treat \( y_i \) as a Bernoulli r.v. with

\[ \pi_i = \Pr(Y_i = 1 \mid x_i) = \Pr(\epsilon_i \leq x_i' \beta) = F(x_i' \beta) \]

and

\[ 1 - \pi_i = \Pr(Y_i = 0 \mid x_i) = 1 - F(x_i' \beta) \]

The likelihood function for \( \beta \) is based on an iid sample \((y_1, x_1), \ldots, (y_n, x_n)\) is

\[ L(\beta; y, x) = \prod_{i=1}^{n} \pi_i^{y_i} (1-\pi_i)^{1-y_i} \]
$$L(\beta | y, x) = \prod_{i=1}^{n} F(x_i; \beta)^{y_i} (1 - F(x_i; \beta))^{1 - y_i}$$

The log-likelihood is

$$\ln L(\beta | y, x) = \sum_{i=1}^{n} \left\{ y_i \cdot \ln F(x_i; \beta) + (1 - y_i) \ln (1 - F(x_i; \beta)) \right\}$$

The F.O.C.'s are

$$\frac{\partial \ln L(\beta)}{\partial \beta} = S(\beta)$$

$$= \sum_{i=1}^{n} \left\{ y_i \cdot \frac{\partial \ln F}{\partial \beta} - (1 - y_i) \frac{\partial \ln (1 - F)}{\partial \beta} \right\}$$

$$= \sum_{i=1}^{n} \left\{ y_i \cdot \frac{f(x_i; \beta)}{F(x_i; \beta)} x_i + (1 - y_i) \frac{f(x_i; \beta)}{1 - F(x_i; \beta)} x_i \right\}$$

$$= \sum_{i=1}^{n} \left\{ y_i \cdot \frac{f(x_i; \beta)}{F(x_i; \beta)} + (1 - y_i) \frac{f(x_i; \beta)}{1 - F(x_i; \beta)} \right\} x_i$$

= System of k nonlinear equations in $\beta$.