Modeling Binary Outcomes: Logit & Probit models

Motivating Example: Model Women's labor force participation

$y_i = 1$ if a married women is in the paid labor force
$\quad = 0$ otherwise

$\underset{1 \times k}{x_i'}$ = vector of observed covariates
e.g. # of children
age
education
wage rate (estimated)

Linear model formulation

$$y_i = x_i'\beta + \epsilon_i \qquad i = 1, \ldots, N$$

<u>Note</u>: when $y_i = 1 \Rightarrow \epsilon_i = 1 - x_i'\beta$
$\qquad\qquad y_i = 0 \Rightarrow \epsilon_i = -x_i'\beta$

Interpretation of Regression model

$$E[y_i | x_i] = 1 \cdot \Pr(Y_i = 1 | x_i) + 0 \cdot \Pr(Y_i = 0 | x_i)$$
$$= \Pr(Y_i = 1 | x_i) = x_i'\beta$$

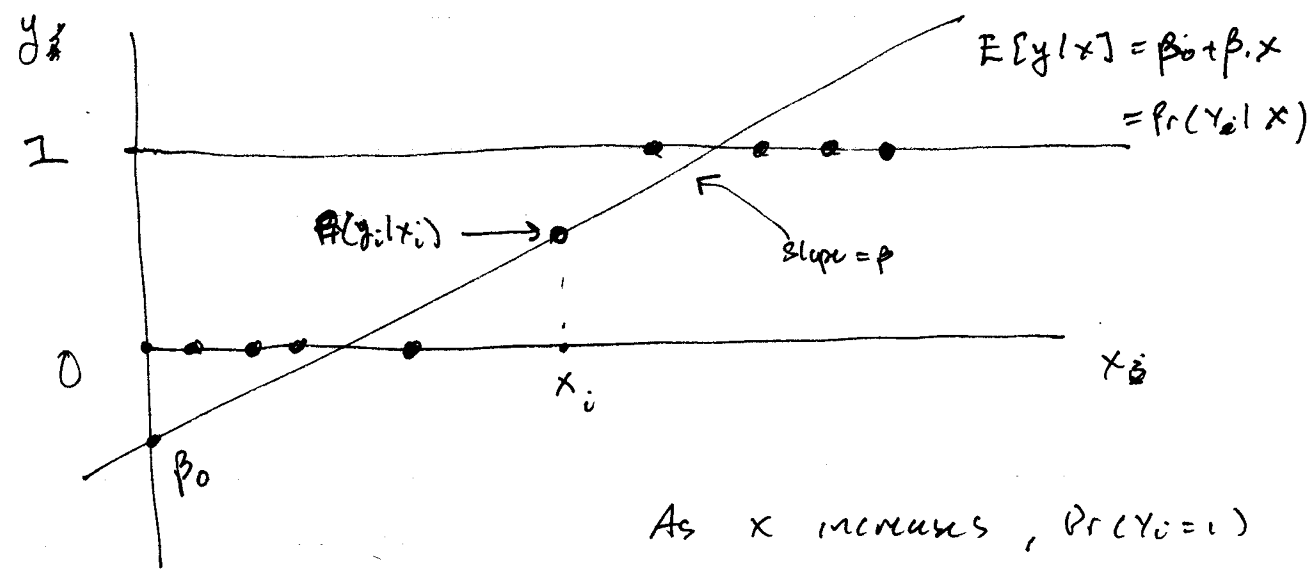
$E[Y_i|X_i] + Y_i - E[Y_i|X_i] = \Pr(Y_i = 1|X_i) + \epsilon_i \qquad \epsilon_i = Y_i - \Pr($

So that

$$\frac{\partial E[y_i | x_i]}{\partial x_i} = \frac{\partial Pr(Y_i = 1 | x_i)}{\partial x_i} = \beta$$

$\Rightarrow \quad \beta$ = change in $Pr(Y_i = 1 | x_i)$ for a unit change in x_i

Graphical illustration for 1 RHS variable



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

<u>Note</u>: Probabilities are not constrained to lie between 0 and 1

Effect of changes in x on $Pr(Y|x)=1$ 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. β_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 $(= 4 * -0.295)$ which is obviously unrealistic!

Problems with the Linear Probability model

(i) $Y_i | X_i$ is heteroskedastic

$$Var(Y_i|X_i) = Pr(Y_i = 1|X_i) \cdot Pr(Y_i = 0|X_i)$$
$$= X_i'\beta(1 - X_i'\beta)$$

(ii) $\varepsilon_i | X_i$ can't be normally distributed.

(iii) Predicted probabilities can be less than zero or greater than 1

(iv) $\dfrac{\partial \Pr(Y_i = 1 \mid x_i)}{\partial x_i} = \beta$ is unrealistic in many situations

e.g. expect each additional child to have a diminishing marginal effect on $\Pr(Y_i = 1 \mid X_i)$

$\Rightarrow$ 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 based on the economic theory underlying labor-leisure trade off

Assume

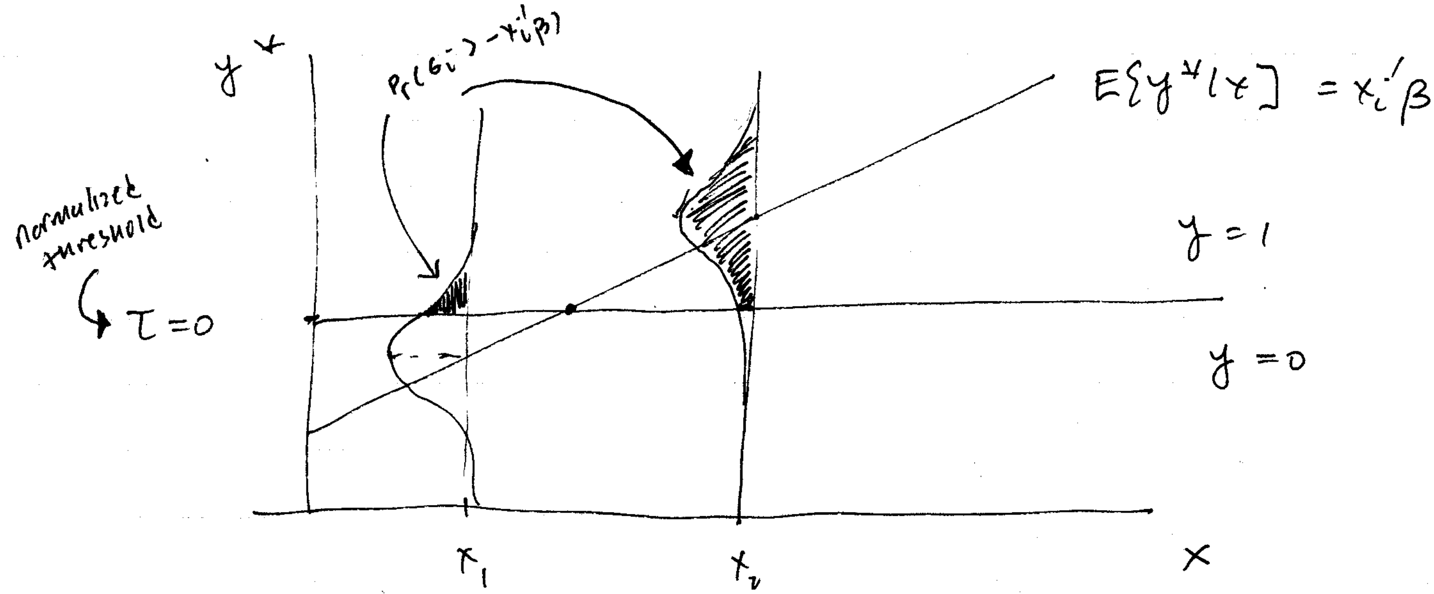
$$y_i^* = x_i'\beta + \epsilon_i$$

and

$$y_i = 1 \quad \text{if} \quad y_i^* > \tau \quad (\tau = 0 = \text{Normalized threshold})$$

$$= 0 \qquad y_i^* \leq \tau$$

So when y_i^* cross the threshold (actual wage > reservation wage) then observe $y_i = 1$



If $\tau = 0$ then ~~If $\tau \neq 0$ then~~ $y_i = 1$ if $y_i^* > 0$ and

$$Pr(y_i = 1|x_i) = Pr(y_i^* > 0|x_i)$$

$$= Pr(x_i'\beta + \epsilon_i > 0|x_i)$$

$$= Pr(\epsilon_i > -x_i'\beta \mid x_i)$$

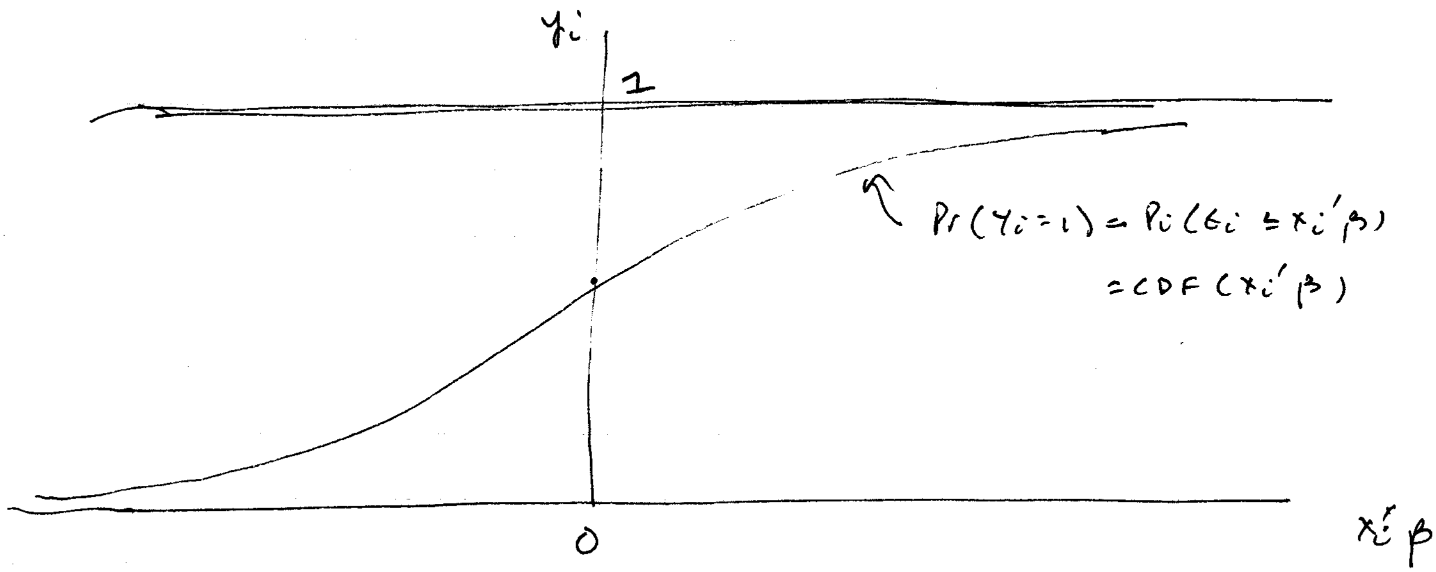
$$= Pr(\epsilon_i \leq x_i'\beta) \quad \text{for symmetric distn.}$$

y_i

$Pr(Y_i = 1 | x_i) = Pr(\epsilon_i \le x_i'\beta)$ for symmetric distn

$= $ CDF of ϵ_i evaluated at $x_i'\beta$

Graphically



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 & 1 because it is based on a CDF for ϵ_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 $\Pr(Y_i = 1 | x_i)$ depends on the distribution of ϵ_i in the latent variable formulation.

Two Popular choices for distn of ϵ_i:

(i) $\epsilon_i \sim N(0,1)$: Probit model

$$\Pr(Y_i = 1 | x_i) = \Pr(\epsilon_i \le x_i'\beta)$$
$$= \Phi(x_i'\beta)$$

where

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

$-\infty < z < \infty$

(ii) $\epsilon_i \sim$ Logistic : Logit model

$$\Pr(Y_i = 1 | x_i) = \Pr(\epsilon_i \le x_i'\beta)$$
$$= \Lambda(x_i'\beta)$$

where

$$\Lambda(z) = \frac{e^z}{1 + e^z} = \frac{1}{e^{-z} + 1}$$

$-\infty < z <$

Note: $\Lambda(z) = \int_{-\infty}^{z} \lambda(x)\,dx$ where

$$\underbrace{\lambda(z)}_{\text{logit pdf}} = \frac{d}{dz}\underbrace{\Lambda(z)}_{\text{logit CDF.}} = \Lambda(z)(1-\Lambda(z))$$

$$= \left(\frac{e^z}{1+e^z}\right)\left(1-\frac{e^z}{1+e^z}\right)$$

$$= \left(\frac{e^z}{1+e^z}\right)\left(\frac{1}{1+e^z}\right)$$

$$= \frac{e^z}{(1+e^z)^2}$$

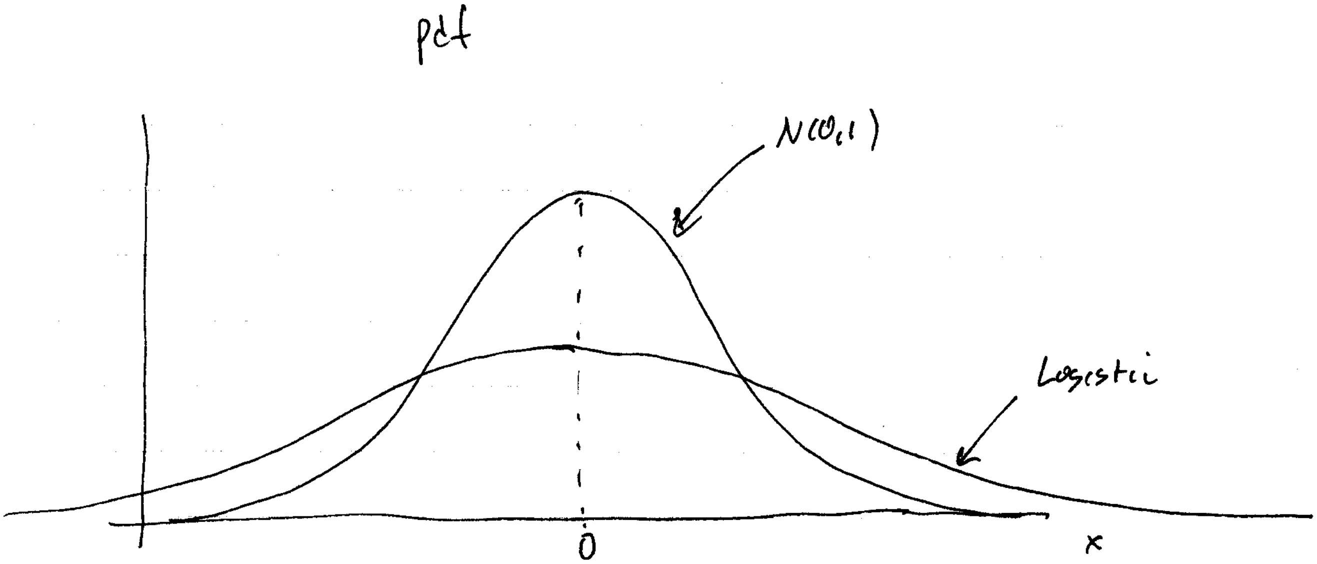
Remarks

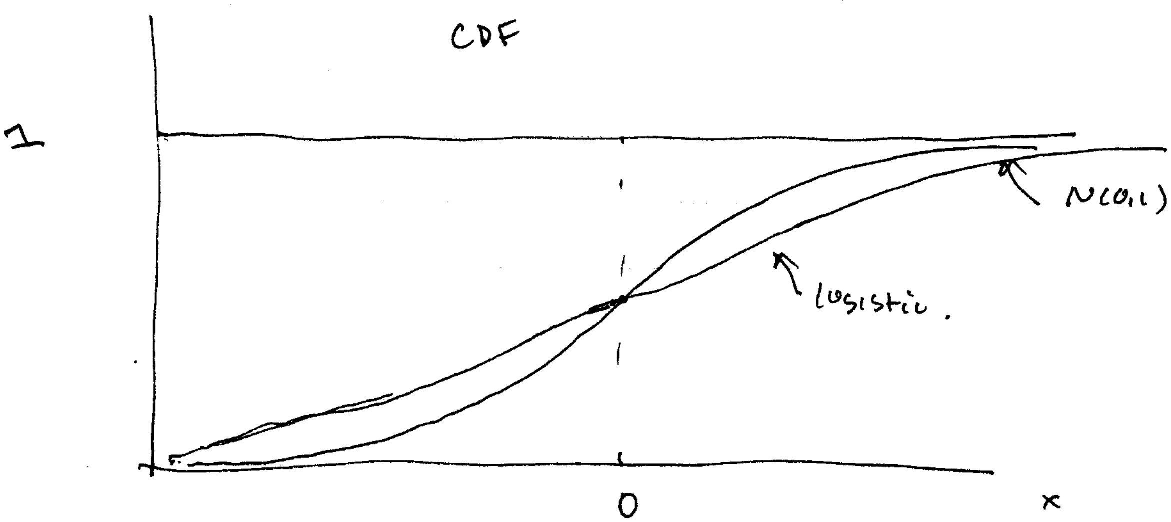
If $\epsilon_i \sim$ logistic then

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

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

$\Rightarrow$ logistic distribution is similar to Student-t with 7 d.f.





Probabilities differ mostly in the tails of the distribution.

Identification in the BRM

In the logit & Probit there are ~~[illegible]~~ 3 important identifying assumptions

$$\left.\begin{array}{l} \text{(i) } E\{\epsilon_i | x_i\} = 0 \\ \text{(ii) } var(\epsilon_i | x_i) \text{ is constant} \end{array}\right\}$$ necessary b/c the latent variable y_i^* is not observable

(iii) $\tau = 0$ } identifies the constant or intercept.

These assumptions allow the identification of β

Do later other likelihood function.

Ex: Probit:

Suppose $\epsilon_i \sim N(0, \sigma^2)$ and $y_i^* = \beta_0 + \beta_1 \cdot x_i + \epsilon$

$\tau = 0$

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

$= \cancel{Pr(y_i^*)} \; Pr(\beta_0 + \beta_1 x_i + \epsilon_i > 0)$

$= \cancel{Pr(\beta_0 + \beta_1}$

$Pr(\epsilon_i \leq \beta_0 + \beta_1 x_i)$

$= Pr\left(\frac{\epsilon_i}{\sigma} \leq \frac{\beta_0}{\sigma} + \frac{\beta_1}{\sigma} x_i \right)$

$= Pr(z_i \leq \beta_0^* + \beta_1^* x_i)$

For different values of σ we get different values of β_0^* & β_1^*. By setting $\sigma = 1$ a priori we can identify the β's.

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

and not β_0, β_1 and σ individually.

For the logit we have

$$\frac{\partial \Pr(Y_i = 1 \mid x_i)}{\partial x_{Ki}} = \Lambda(x_i'\beta)\left(1 - \Lambda(x_i'\beta)\right) \cdot \beta_K$$

and for the Probit

$$\frac{\partial \Pr(Y_i = 1 \mid x_i)}{\partial x_{Ki}} = \varphi(x_i'\beta) \cdot \beta_K$$

Give mini example

Maximum Likelihood Estimation of Probit e Logit models

$$y_i^* = x_i'\beta + \epsilon_i \quad , \quad \epsilon_i \text{ has CDF } F \text{ and p.d.f. } f$$

$$y_i = 1 \text{ if } y_i^* > 0$$
$$= 0 \quad y_i^* \leq 0$$

$\epsilon_i \sim$ iid $N(0,1)$ and $F(z) = \Phi(z)$: Probit

$\epsilon_i \sim$ 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

$$\pi_i = Pr(Y_i = 1 | x_i) = Pr(\epsilon_i \leq x_i'\beta)$$
$$= F(x_i'\beta)$$

and

$$1 - \pi_i = Pr(Y_i = 0 | X_i) = 1 - F(x_i'\beta)$$

The likelihood function for β ~~is~~ based on an iid sample $(y_1, x_1) \ldots (y_N, x_N)$ is

$$L(\beta | \underset{\sim}{y}, \underset{\sim}{x}) = \prod_{i=1}^{N} \pi_i^{y_i} (1-\pi_i)^{1-y_i}$$

$$L(\beta \mid \underset{\sim}{y}, x) = \prod_{i=1}^{N} F(x_i'\beta)^{y_i} \left(1 - F(x_i'\beta)\right)^{1-y_i}$$

The log-likelihood is

$$\ln L(\beta \mid \underset{\sim}{y}, x) = \sum_{i=1}^{N} \left\{ y_i \cdot \ln F(x_i'\beta) + (1-y_i) \ln\left(1 - F(x_i'\beta)\right) \right\}$$

The F.O.C's are

$$\underset{k \times 1}{\frac{\partial \ln L(\beta)}{\partial \beta}} = S(\beta)$$

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

$$= \sum_{i=1}^{N} \left\{ y_i \frac{f(x_i'\beta)}{F(x_i'\beta)} x_i + \frac{(1-y_i) * \left(-f(x_i'\beta)\right)}{1-F(x_i'\beta)} x_i \right.$$

$$= \sum_{i=1}^{N} \left\{ y_i * \frac{f(x_i'\beta)}{F(x_i'\beta)} - \frac{(1-y_i) f(x_i'\beta)}{1-F(x_i'\beta)} \right\} x_i$$

= System of k nonlinear equations in β!