Discrete & Limited Dependent Variable Models

In general, if your dependent variable only takes on a few discrete values then the usual Regression model is probably not the one to use.

Likewise if it is continuous, but limited in Range.

Binary Dependent Variable takes on only 1 of 2 possible states. By convention, these are coded 1 or 0.

$$y_t = \begin{cases} 1 & \text{if } A \\ 0 & \text{otherwise.} \end{cases}$$

This is a special case of 'discrete' dependent variable models - here the # of possible outcomes is more than 2, but still countably finite.

- These can follow a natural ordering.

  Rate your level of service

  High
  Moderate
  Low

- unordered

  How did you get to work

  Walk
  Car
  Bike
  Other.

And then, there is count data.

# of traffic tickets.

Sometimes, a dependent variable is continuous but only takes a limited range of values.

Automobile expense -

0

0

275

1250

397

0

.

Here, you may have a number of zeros if sample includes those who do not own a car.

And others

Binary Response

$$y_t = \begin{cases} 1 & \text{if } A \\ 0 & \text{otherwise.} \end{cases}$$

Let P_t denote the probability that $y_t = 1$ conditional on the information set Ω_t.

Ω_t usually includes exogenous and predetermined variables.

Binary response models this conditional probability

$$P_t \equiv \Pr(y_t = 1 \mid \Omega_t)$$

$$= E(y_t \mid \Omega_t)$$

Proof:

$$E(y_t | \Omega_t) = 0 \cdot Pr(y_t = 0 | \Omega_t)$$

$$+ 1 \cdot Pr(y_t = 1 | \Omega_t) = Pr(y_t = 1 | \Omega_t)$$

Suppose X_t' is a Row vector of length k and includes the explanatory variables in your model. X_t' Belongs to Ω_t.

Linear Regression

$E(y_t | \Omega_t)$ is modeled

$E(y_t | X_t\beta)$ as $X_t\beta$.

But fails to impose the restriction

$$0 \leq E(y_t | \Omega_t) \leq 1$$

which must Be true if it is a probability.

One way of imposing this restriction

that $\quad 0 < P_t < 1$

is to specify

$$P_t \equiv E(y_t \mid \Omega_t) = F(x_t'\beta)$$

$x_t'\beta$ is a scalar (index function

and F has the property

$$F(-\infty) = 0 \quad , \quad F(\infty) = 1 \quad \text{and}$$

$$f(x) = \frac{d\,F(x)}{d\,x} > 0$$

These are properties of C.D.F.

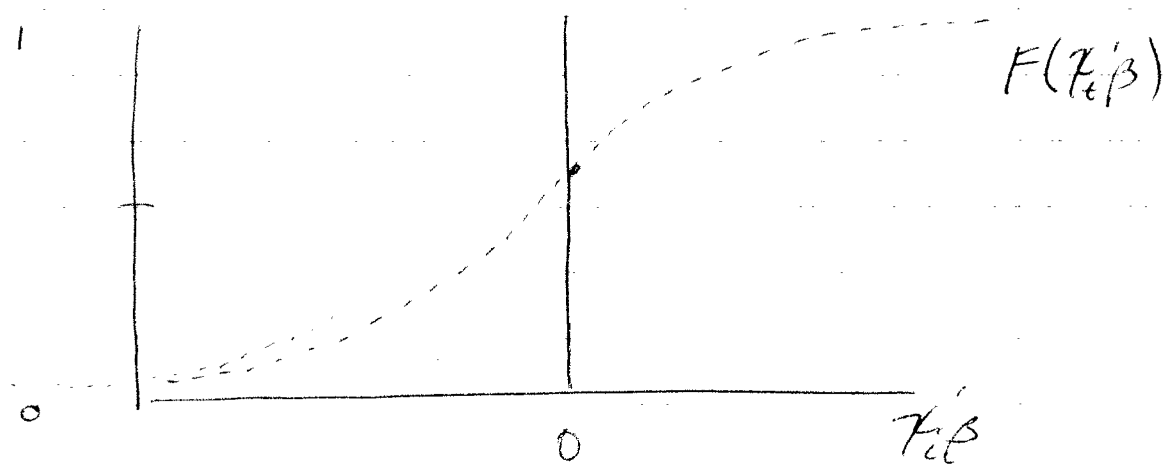
These properties also garantee that F is nonlinear

$$\frac{\partial P_t}{\partial x_{ti}} = \frac{\partial F(x_t'\beta)}{\partial x_{ti}} = f(x_t'\beta)\beta_i$$

The magnitude of the derivative is proportional to $f(x_i'\beta)$.

$\Rightarrow$ Sign of β_i determines sign of marginal effect.

$\Rightarrow$ Usually, marginal effect reaches max at $f(0)$



Choosing $F(\cdot)$

Probit - with Probit

$F(\,)$ is chosen as standard cumulative normal CDF.

$$\Phi(x) \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} -\frac{1}{2} x^2 \, dx$$

Rationale: This can be derived from a linear latent variable model

$$y_t^* = x_t'\beta + u_t \qquad u_t \sim N(0,1)$$

we do not observe y_t^*, only its sign

e.g. Marginal Cost vs Marginal Benefits

If consumer purchase
$MB - MC > 0$
If not, consumer passes.

$$y_t = \begin{cases} 1 & \text{if } y_t^* > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Then, the latent variable is assumed to be a linear function of x_t'

$$\Pr(y_t = 1) = \Pr(y_t^* > 0) = \Pr(x_t'\beta + u_t > 0)$$

$$= \Pr(u_t > -x_t'\beta) = \Pr(u_t < x_t'\beta)$$

due to symmetry.

if $u_t \sim N(0,1)$

$$\Pr(y_t = 1) = \int_{-\infty}^{x_t'\beta} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \, dx.$$

Logit

$$F(x) \equiv \frac{1}{1+e^{-x}} = \frac{e^{x}}{1+e^{x}} \equiv \Lambda(x)$$

Verify.

to get p.d.f. take derivative.

$$\lambda(x) = \frac{\partial F(x)}{\partial x} = e^{-x}/(1+e^{-x})^2$$

$$= \Lambda(x)\,\Lambda(-x)$$

Verify

This latter result suggests it must be symmetric about zero.

$$\Lambda(-x) = 1 - \Lambda(x)$$

verify

Model is easily derived as log odds

$$\log\left(\frac{P_t}{1-P_t}\right) = x_t'\beta$$

Verify

~~log~~

Solve for P_t

$$\frac{P_t}{1-P_t} = e^{x_t'\beta}$$

$$P_t = (1-P_t)\,e^{x_t'\beta}$$

$$P_t(1+e^{x_t'\beta}) = e^{x_t'\beta}$$

$$P_t = \frac{e^{x_t'\beta}}{1+e^{x_t'\beta}} = \Lambda(x)$$