Other Binary Choice Models

Logit

$$\Pr(y=1) = \frac{\exp\{\underset{\sim}{x}^T \underset{\sim}{\beta}\}}{1+\exp\{\underset{\sim}{x}^T \underset{\sim}{\beta}\}} \equiv \Lambda(\underset{\sim}{x}^T \underset{\sim}{\beta})$$

Λ is the logistic CDF

The logit p.d.f.

$$\frac{d\,\Lambda(x^T\beta)}{\partial(\underset{\sim}{x}^T\beta)} = \frac{\exp\{x^T\beta\}}{\left[1+\exp(\underset{\sim}{x}^T\underset{\sim}{\beta})\right]^2}$$

$$= \Lambda(x^T\beta)\left[1-\Lambda(\underset{\sim}{x}^T\beta)\right]$$

$\therefore$ The marginal effects

$$\frac{\partial E(y \mid x)}{\partial \underset{\sim}{x}} = \Lambda(x^T\beta)\left(1-\Lambda(x^T\beta)\right)\underset{\sim}{\beta}$$

Hessian: $$\frac{\partial^2 \ln L}{\partial\beta\,\partial\beta^T} = -\sum \Lambda_i(1-\Lambda_i)\,\underset{\sim}{x}_i \underset{\sim}{x}_i^T$$

$$\Lambda_i \equiv \Lambda(\underset{\sim}{x}_i^T \underset{\sim}{\beta})$$

The tails of a logit dist are heavier than those of the normal.

$\sim t_7$

If the index $x_i^T\beta$ $i=1,\dots,n$ lie between ± 1.2 then the results will be very similar.

The biggest differences occur when either $x_i^T\beta$ is very large or small.

Predictions will differ especially if (1) there are very few 0's or 1's and (2) if there is very wide variation in an x..

Logit & Probit are both symmetric.

<u>Nonsymmetric</u>

Gumbel

$$\text{Prob}(y=1 \mid \underset{\sim}{x}) = \exp\{-\exp(-x_T^T\beta)\}$$

Complementary Log-Log

$$\text{Prob}(y=1 \mid \underset{\sim}{x}) = 1-\exp\{-\exp\{x^T\beta\}\}$$

Skewed logit

$$\text{Prob}(y=1) = 1 - \frac{1}{[1+\exp\{x_i^T\beta\}]^{\alpha}}$$

STATA: The main reason to use this model is that the maximum ~~prob~~ effect of the regressors on the prob of success is not constrained to be when $p=.5$

Instead, its used when that max occurs ~~at~~ between .3 and .6

* How you code 1 & 0 matters
* you could use "Power Logit" if $p > .63$
* you could reverse the signs and talk about the marginal effect of x on failure

Model with interactions

$$\text{Prob}(\text{DocVis} = 0) = \Lambda(\beta_1 + \beta_2 \text{Age} + \beta_3 \text{Inc.}$$
$$+ \beta_4 \text{Kids} + \beta_5 \text{Educ} + \beta_6 \text{MARR} + \beta_7 \text{Age} * \text{Educ})$$

$$\frac{\partial \text{Prob}(\text{DocVis} = 0)}{\partial \text{Age}} = \Lambda(x_i'\beta)(1 - \Lambda(x'\beta))(\beta_2 + \beta_7 \text{Educ})$$

$$\frac{\partial \text{Prob}(\text{DocVis} = 0)}{\partial \text{Educ}} = \Lambda(x\beta)[1 - \Lambda(x\beta)](\beta_5 + \beta_7 \text{Age})$$

So far so good

But,

$$\frac{\partial^2 \text{Prob}(\text{DocVis})}{\partial \text{Age}\, \partial \text{Educ}} = \Lambda(x\beta)[1 - \Lambda(x\beta)]\beta_7 +$$

$$\Lambda(1-\Lambda)(\beta_5 + \beta_7 A)(\beta_2 + \beta_7 E)$$
$$-2\Lambda(\Lambda(1-\Lambda))(\beta_5 + \beta_7 A)$$
$$\times(\beta_2 + \beta_7 E)$$

$$= \Lambda(1-\Lambda)\beta_7 + \Lambda(1-\Lambda)[1-2\Lambda]$$
$$\times(\beta_2 + \beta_7 \text{Educ})(\beta_5 + \beta_7 \text{Age})$$

Fit

pseudo $- R^2 \quad 1 - \dfrac{\ln L}{\ln L_0}$

$\ln L$ is log likelihood overall

$\ln L_0$ is log likelihood with all "slopes" set to 0, i.e.,

$$\beta_2 = \beta_3 = \cdots = \beta_k = 0$$

This stat $= 0$ if $\beta_2 = \cdots = \beta_k = 0$

But can never quite $= 1$ (unless there is something seriously wrong with your model i.e., $x\beta \to \infty$

Another

$$R^2_{BL} = \frac{1}{n} \sum_{i=1}^{n} y_i \hat{F}_i + (1 - y_i)(1 - \hat{F}_i)$$

Avg prob of correct prediction

However, in badly balanced samples the less freq outcome is seldom predicted.

Cramer.

$$\lambda = (\text{Average } \hat{F} \mid y_i = 1) - (\text{Average } \hat{F} \mid y_i = 0)$$

This heavily penalizes incorrect predictions.

<u>TABLE</u>

Might want to Δ $p = .5$ rule.