ECON 4151
Lab Session 4: Logit/Probit Model

Wonjin Lee

Washington University in St. Louis

October 9, 2020

Outline

1. Overview: LDV Models

2. Review of Probit/Logit Models (Wooldridge Ch. 17)

Limited Dependent Variable Models

- A limited dependent variable is a dependent variable whose range of values is substantively restricted.
- LDV Models:
  1. Binary dependent variable $\Longrightarrow$ logit and probit models
  2. Corner solution dependent variable $\Longrightarrow$ Tobit model
     - Lumpiness in an outcome variable (Ex. zero values for a nontrivial fraction of the population)
  3. Count dependent variable (nonnegative integer values) $\Longrightarrow$ Poisson regression models
  4. Data censoring (issue of data observability) $\Longrightarrow$ selectivity model

Binary response models

- Binary response models can be derived from an underlying latent variable model:

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

  where u is independent of x.
- Observation rule for y: $y_i = \mathbb{1}[y_i^* > 0]$.
- Assume that u either has the standard normal distribution or the standard logistic distribution.

Binary response models

- We can derive the response probability for y:

$$\begin{aligned}
\Pr\left[y_i = 1 \mid x_i\right] &= \Pr\left[y_i^* \geq 0 \mid x_i\right] \\
&= \Pr\left[u_i \geq -x_i'\beta\right] \\
&= 1 - \Pr\left[u_i \leq -x_i'\beta\right] \\
&= 1 - G\left(-x_i'\beta\right) \\
&= G\left(x_i'\beta\right) \quad (\because \text{symmetric distribution})
\end{aligned}$$

  where G is the cdf of u.
- In the probit model, G is the standard normal cdf.
- In the logit model, G is the logistic function.

Binary response models

- Partial effect (marginal effect):
  - For continuous explanatory variables,

$$\frac{\partial \Pr[y=1|x]}{\partial x_j} = G'(x'\beta)\beta_j = g(x'\beta)\beta_j$$

  - For discrete explanatory variables, the effect of x_k going from c_k to c_k+1 is

$$G\left[\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k\left(c_k+1\right)\right] \\ -G\left(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k c_k\right)$$

- For logit and probit, the direction of the effect of x_j on the response probability and the latent variable is always the same.
- The magnitudes of each β_j are not useful because the latent variable rarely has a well-defined unit of measurement. Ex) the difference in utility levels from two different actions.

Binary response models

- Partial effect at the average:

$$\underbrace{g(\overline{x}'\hat{\beta})}\hat{\beta}_j$$

- Average partial effect:

$$\underbrace{\frac{1}{N}\sum_{i=1}^{N} g(x_i'\hat{\beta})}\hat{\beta}_j$$

Maximum likelihood estimation of binary response models

- In the linear probability model, the OLS estimator is unbiased but inefficient due to heteroskedasticity.
- MLE is based on the distribution of y given x, the heteroskedasticity in $\text{Var}(y|x)$ is automatically accounted for.

Maximum likelihood estimation of binary response models

- The density of y_i given x_i is

$$f\left(y_i \mid x_i'\beta\right) = G\left(x_i'\beta\right)^{y_i}\left(1 - G\left(x_i'\beta\right)\right)^{1-y_i}, \quad y_i = 0, 1$$

- The joint distribution of the particular sequence we observe in a sample of N observations is simply:

$$f\left(y_1, y_2, \ldots, y_N\right) = \prod_{i=1}^{N} G\left(x_i'\beta\right)^{y_i}\left(1 - G\left(x_i'\beta\right)\right)^{1-y_i}$$

- We want to find $\hat{\beta}$ which maximizes the joint distribution; the joint distribution is a function of β

Maximum likelihood estimation of binary response models

- By taking the log of the joint distribution, we get the log-likelihood function

$$\log \mathcal{L}\left(\beta; y_1, y_2, \ldots, y_N\right) = \sum_{i=1}^{N} \Big[y_i \log G\left(x_i'\beta\right) + \left(1 - y_i\right) \log\left(1 - G\left(x_i'\beta\right)\right) \Big]$$

- Find $\hat{\beta}$ s.t. $\frac{\partial \log \mathcal{L}}{\partial \beta} = 0$.
- Since $\frac{\partial \log \mathcal{L}}{\partial \beta}$ is nonlinear in β, we usually do not have explicit solution.
- As long as the model is correctly specified, MLE $\hat{\beta}$ is consistent, efficient, and and asymptotically normal.
- $\hat{\beta}$ comes with an (asymptotic) standard error which can be used to construct (asymtotic) t tests and confidence intervals just as with OLS.

Excercise

- Linear probability model:

$$\begin{aligned} smoker = &\alpha_0 + \alpha_1 age + \alpha_2 educ \\ &+ \alpha_4 income + \alpha_5 pcigs79 + \epsilon \end{aligned}$$

- Logit/probit model:

$$\begin{aligned} \Pr[smoker = 1|X] = &G(\beta_0 + \beta_1 age + \beta_2 educ \\ &+ \beta_4 income + \beta_5 pcigs79) \end{aligned}$$

Probit Model

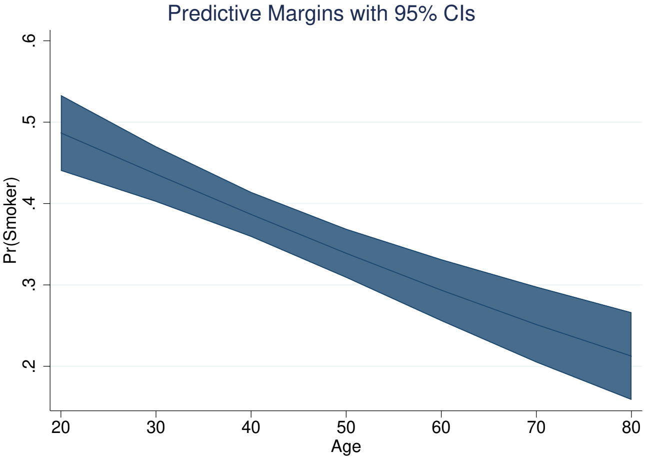

Figure: Predicted Pr[*smoke*] given *age*

Probit Model

- Examples from class:

$$\Pr[inlf = 1|X] = G(\beta_0 + \beta_1 nwifeinc + \beta_2 educ + \beta_3 exper + \beta_5 age + \beta_6 kidslt6 + \beta_7 kidsge6)$$

$$\Pr[inlf = 1|X] = G(\beta_0 + \beta_1 nwifeinc + \beta_2 educ + \beta_3 exper + \beta_4 exper^2 + \beta_5 age + \beta_6 kidslt6 + \beta_7 kidsge6)$$

Probit Model

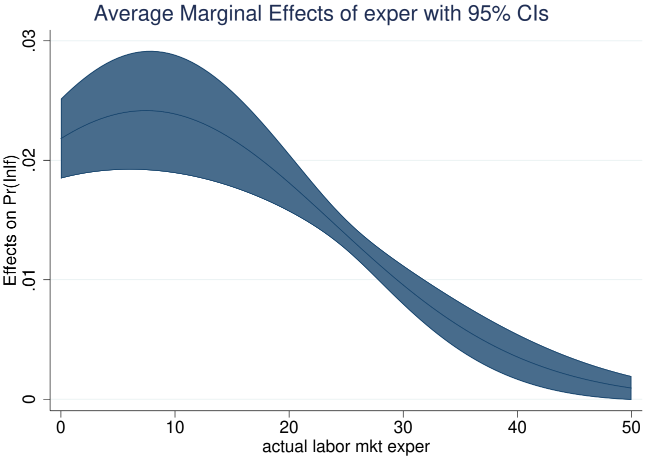

Figure: Average marginal effects of *exper* on Pr[*inlf*]

Probit Model

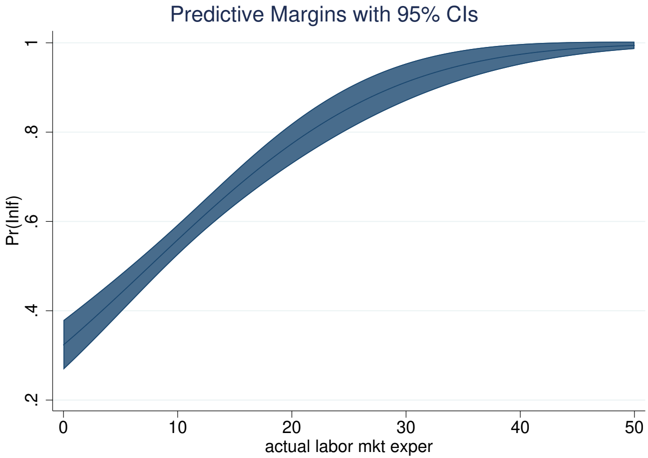

Figure: Predicted $\Pr[inlf]$ given *exper*