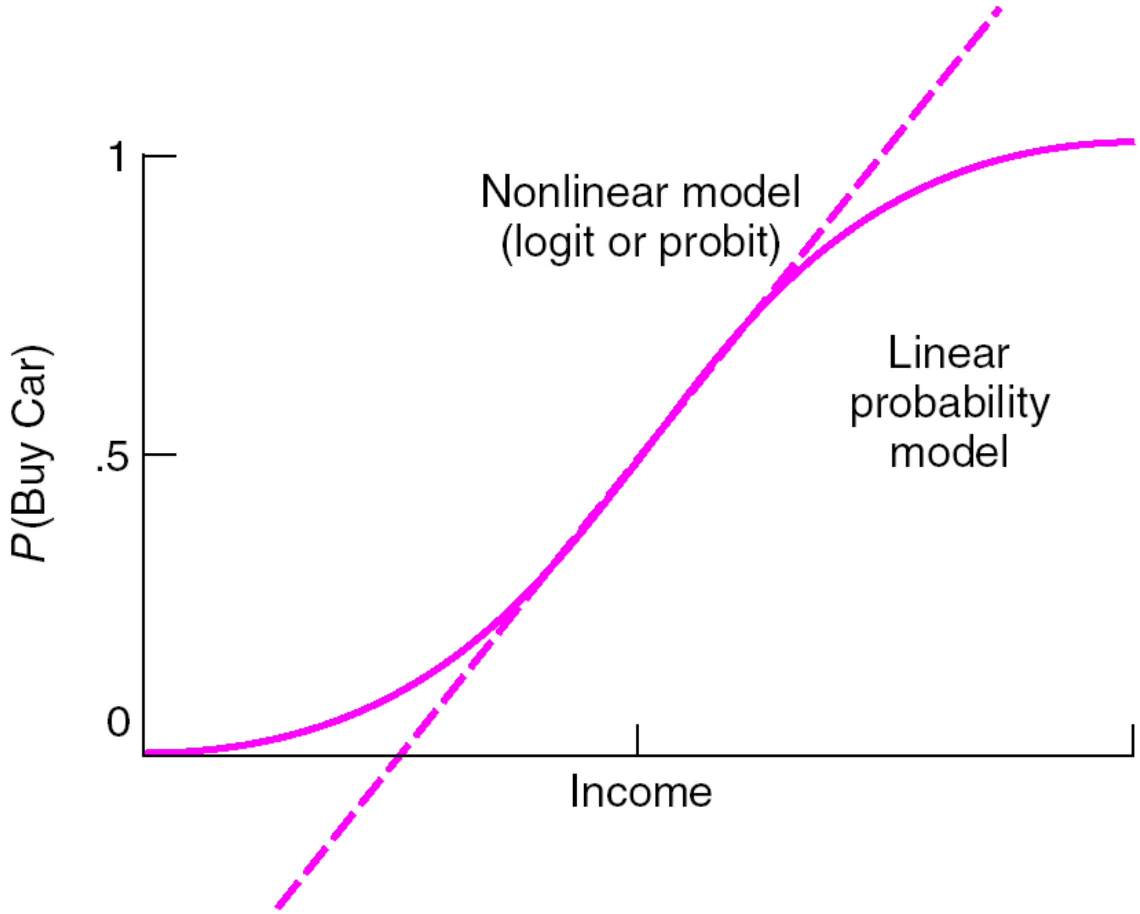
Probit, Logit and Tobit Models

S.Madheswaran
Institute for Social and Economic Change
Bangalore

Logit and Probit Models

- Another criticism of the linear probability model is that the model assumes that the probability that $Y_i = 1$ is linearly related to the explanatory variables
  - However, the relation may be nonlinear
    - For example, increasing the income of the very poor or the very rich will probably have little effect on whether they buy an automobile, but it could have a nonzero effect on other income groups
- Logit and probit models are nonlinear and provide predicted probabilities between 0 and 1

Logit and Probit Models



Logit and Probit Models

- Suppose our underlying dummy dependent variable depends on an unobserved utility index, Y^*
- If Y is discrete—taking on the values 0 or 1 if someone buys a car, for instance
  - Can imagine a continuous variable Y^* that reflects a person's desire to buy the car
    - Y^* would vary continuously with some explanatory variable like income

Logit and Probit Models

- Written formally as

$$Y_i^* = \beta_0 + \beta_1 X_{1i} + \varepsilon_i$$

  - If the utility index is “high enough,” a person will buy a car

$$Y_i = 1 \text{ if } Y_i^* \geq 0$$

  - If the utility index is not “high enough,” a person will not buy a car

$$Y_i = 0 \text{ if } Y_i^* < 0$$

Logit and Probit Models

$$
\begin{aligned}
P_i &= \text{Prob}(Y_i = 1) \\
&= \text{Prob}(Y_i^* \geq 0) \\
&= \text{Prob}(\beta_0 + \beta_1 X_{1i} + \varepsilon_i \geq 0) \\
&= \text{Prob}(\varepsilon_i \geq -\beta_0 - \beta_1 X_{1i}) \\
&= 1 - F(-\beta_0 - \beta_1 X_{1i}) \text{ where } F \text{ is the c.d.f. for } \varepsilon \\
&= F(\beta_0 + \beta_1 X_{1i}) \text{ if } F \text{ is symmetric}
\end{aligned}
$$

- The basic problem is selecting *F*—the cumulative density function for the error term
  - This is where where the two models differ

Logit and Probit Models

- Interested in estimating the β's in the model
  - Typically done using a maximum likelihood estimator (MLE)
- Each outcome Y_i has the density function $f(Y_i) = P_i^{Y_i}(1 - P_i)^{1 - Y_i}$
  - Each Y_i takes on either the value of 0 or 1 with probability $f(0) = (1 - P_i)$ and $f(1) = P_i$

Logit and Probit Models

- The likelihood function is

$$\begin{aligned}\ell &= f(Y_1, Y_2, \ldots Y_n) \\ &= f(Y_1) f(Y_2) \ldots f(Y_n) \\ &= P_1^{Y_1}(1 - P_1)^{1-Y_1} P_2^{Y_2}(1 - P_2)^{1-Y_2} \ldots P_n^{Y_n}(1 - P_n)^{1-Y_n} \\ &= \prod_{i=1}^{n} P_i^{Y_i}(1 - P_i)^{1-Y_i}\end{aligned}$$

and

$$\ln \ell = \sum_{i=1}^{n} Y_i \ln P_i + (1 - Y_i) \ln(1 - P_i)$$

which, given $P_i = F(\beta_0 + \beta_1 X_{1i})$, becomes

$$\ln \ell = \sum_{i=1}^{n} Y_i \ln F(\beta_0 + \beta_1 X_{1i}) + (1 - Y_i) \ln(1 - F(\beta_0 + \beta_1 X_{1i}))$$

Logit Model

- For the logit model we specify

$$\text{Prob}(Y_i = 1) = F(\beta_0 + \beta_1 X_{1i}) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_{1i})}}$$

- $\text{Prob}(Y_i = 1) \to 0$ as $\beta_0 + \beta_1 X_{1i} \to -\infty$
- $\text{Prob}(Y_i = 1) \to 1$ as $\beta_0 + \beta_1 X_{1i} \to \infty$
    - Thus, probabilities from the logit model will be between 0 and 1

Logit Model

- A complication arises in interpreting the estimated β's
  - With a linear probability model, a β estimate measures the *ceteris paribus* effect of a change in the explanatory variable on the probability *Y* equals 1
- In the logit model

$$\frac{\partial \text{Prob}(Y_i = 1)}{\partial X_1} = \frac{\partial F(\hat{\beta}_0 + \hat{\beta}_1 X_{1i})}{\partial X_1}\hat{\beta}_1$$

$$= \frac{\hat{\beta}_1 e^{-(\beta_0+\beta_1 X_{1i})}}{[1 + e^{-(\beta_0+\beta_1 X_{1i})}]^2}$$

The derivative is nonlinear and depends on the value of *X.*

Probit Model

- In the probit model, we assume the error in the utility index model is normally distributed
  - $\varepsilon_i \sim N(0,\sigma^2)$

$$\text{Prob}(Y_i = 1) = F\left(\frac{\beta_0 + \beta_1 X_{1i}}{\sigma}\right)$$

  - Where F is the standard normal cumulative density function (c.d.f.)

$$\text{Prob}(Y_i = 1) = F\left(\frac{\beta_0 + \beta_1 X_{1i}}{\sigma}\right) = \int_{-\infty}^{\frac{\beta_0 + \beta_1 X_{1i}}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

Probit Model

- The c.d.f. of the logit and the probit look quite similar
- Calculating the derivative is moderately complicated

$$\frac{\partial \text{Prob}(Y_i = 1)}{\partial X_1} = \frac{\partial F\left(\frac{\beta_0 + \beta_1 X_{1i}}{\sigma}\right)}{\partial X_1} = f\left(\frac{\beta_0 + \beta_1 X_{1i}}{\sigma}\right)\beta_1$$

  - Where f is the density function of the normal distribution

Probit Model

- The derivative is nonlinear
  - Often evaluated at the mean of the explanatory variables
- Common to estimate the derivative as the probability $Y = 1$ when the dummy variable is 1 minus the probability $Y = 1$ when the dummy variable is 0
  - Calculate how the predicted probability changes when the dummy variable switches from 0 to 1

Which is Better? Logit or Probit?

- From an empirical standpoint logits and probits typically yield similar estimates of the relevant derivatives
  - Because the cumulative distribution functions for the two models differ slightly only in the tails of their respective distributions
- The derivatives are different only if there are enough observations in the tail of the distribution
- While the derivatives are usually similar, the parameter estimates associated with the two models are not
  - Multiplying the logit estimates by 0.625 makes the logit estimates comparable to the probit estimates

Censored Regression Model

- Often the dependent variable is constrained (or censored)
  - Takes on a positive value for some observations and zero for other observations
  - Represents non-continuous data as there is a large cluster of observations at zero
- Using OLS leads to biased estimates of the parameters

Censored Regression Model

- Examples include data sets containing information on
  - The number of hours people worked last week along with their age
    - Some people will have worked a positive number of hours
    - Others (such as retirees) will not have worked at all and will report working zero hours
  - Families' expenditures on new automobile purchases during a particular year

Censored Regression Model

- For the probit and logit we defined a latent variable $Y^*_i = \beta X_i + u_i$ with

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

- If Y_i is not a binary variable but rather is observed as Y^*_i if $Y^*_i > 0$ and is not observed for $Y^*_i \leq 0$, then

$$Y_i = Y_i^* = \beta X_i + u_i \text{ if } Y_i^* > 0$$
$$= 0 \text{ otherwise}$$

u is assumed to follow the normal distribution with mean 0 and variance σ^2.

Censored Regression Model

- Called the Tobit model or the censored regression model
- To estimate this model, specify the likelihood function for this problem and generate the maximum likelihood estimator
- The (log) likelihood for the Tobit model is

$$\log L = \sum_{Y_i > 0} -\frac{1}{2}\left[\log(2\pi) + \log \sigma^2 + \frac{(Y_i - \beta X_i)^2}{\sigma^2}\right] + \sum_{Y_i = 0} \log\left[1 - F\left(\frac{\beta X_i}{\sigma}\right)\right]$$

Heckman Two-Step Estimator

- As an alternative to estimation of the Tobit model using maximum likelihood methods, James Heckman has developed a two-step estimation procedure
  - Yields consistent estimates of the parameters
- Suppose the model takes the form

$$\begin{aligned} Y_i &= Y_i^* = \beta X_i + u_i \text{ if } Y_i^* > 0 \\ &= 0 \text{ otherwise} \end{aligned}$$

Heckman Two-Step Estimator

- The mean value of Y (if it is greater than zero) may be written as

$$E[Y_i / Y_i > 0] = \beta X_i + E[u_i / u_i > 0] = BX_i + E[u_i / u_i > -Bx_i]$$

- It can be shown that

$$E[u_i / u_i > \; - \; \beta x_i] = \frac{\sigma f(\beta X_i)}{F(\beta X_i)} \equiv \sigma\lambda$$

  - Where

$$\lambda = \frac{f(\beta X_i)}{F(\beta X_i)}$$

Called the inverse Mills ratio or the hazard rate.

Heckman Two-Step Estimator

- Regressing the positive values of Y_i on X_i would lead to omitted variable bias
- If we could get an estimate of λ we could run ordinary least squares on X and λ

Heckman Two-Step Estimator

- Heckman proposes
  - Defining *I as* a dummy variable taking on the value 1 for the positive values of *Y* and 0 otherwise
    - $I_i = 1$ if $Y_i > 0$; 0 otherwise
  - Estimate λ by estimating a probit model of I_i on *X*
    - Since the probit model specifies Prob($Y = 1$) = $F(\beta X_i)$, we can get estimates of β by estimating the probit model
    - Can use these estimates to form

$$\hat{\lambda} = \frac{f(\hat{\beta} X_i)}{F(\hat{\beta} X_i)}$$

  - Using the positive values of *Y*, run OLS on *X* and the estimated λ—will yield consistent estimates of β