Limited Dependent Variables

- $P(y = 1|\boldsymbol{x}) = G(\beta_0 + \boldsymbol{x\beta})$
- $y^* = \beta_0 + \boldsymbol{x\beta} + u,\ y = max(0,y^*)$

Binary Dependent Variables

- So far continuous dependent variables y.

- What if y=0 or 1, i.e., binary variable:

  y=entry to the labor market, “yes”(=1) or “no”(=0)

or

  y= application is denied(=0) or accepted(=1)

Linear Probability Model

- Consider the linear regression model for this case:

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + u$$

- What is the interpretation of the regression line ?

If y=continuous variable, then

$$E(y|x_1 \ldots x_k) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k$$

Linear Probability Model

- Recall that if y=0, 1, then

  $E(y) = 1\text{x } P(y=1)+0 \text{ x } P(y=0) = P(y=1)$

- So, if y=binary

  $E(y|x_1 \ldots x_k) = P(y=1|x_1 \ldots x_k)$

Linear Probability Model

- $P(y = 1|x) = E(y|x)$, when y is a binary variable, so we can write our model as
- $P(y = 1|x) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k$
- So, the interpretation of β_j is the change in the probability of success when x_j changes
- The predicted *y* is the predicted probability of success
- Potential problem that can be outside [0,1]

Linear Probability Model (cont)

- Even without predictions outside of [0,1], we may estimate effects that imply a change in *x* changes the probability by more than +1 or –1, so best to use changes near mean
- This model will violate assumption of homoskedasticity, so will affect inference
- Despite drawbacks, it's usually a good place to start when *y* is binary

Binary Dependent Variables

- Recall the linear probability model, which can be written as $P(y = 1|\mathbf{x}) = \beta_0 + \mathbf{x}\boldsymbol{\beta}$
- A drawback to the linear probability model is that predicted values are not constrained to be between 0 and 1
- An alternative is to model the probability as a function, $G(\beta_0 + \mathbf{x}\boldsymbol{\beta})$, where $0<G(z)<1$

The Probit Model

- One choice for $G(z)$ is the standard normal cumulative distribution function (cdf)
- $G(z) = \Phi(z) \equiv \int \phi(v)dv$, where $\phi(z)$ is the standard normal, so $\phi(z) = (2\pi)^{-1/2}\exp(-z^2/2)$
- This case is referred to as a probit model
- Since it is a nonlinear model, it cannot be estimated by our usual methods
- Use maximum likelihood estimation

The Logit Model

- Another common choice for G(z) is the logistic function, which is the cdf for a standard logistic random variable
- $G(z) = \exp(z)/[1 + \exp(z)] = \Lambda(z)$
- This case is referred to as a logit model, or sometimes as a logistic regression
- Both functions have similar shapes – they are increasing in *z*, most quickly around 0

Probits and Logits

- Both the probit and logit are nonlinear and require maximum likelihood estimation
- No real reason to prefer one over the other
- Traditionally saw more of the logit, mainly because the logistic function leads to a more easily computed model
- Today, probit is easy to compute with standard packages, so more popular

Interpretation of Probits and Logits (in particular vs LPM)

- In general we care about the effect of x on P(y = 1|$\boldsymbol{x}$), that is, we care about $\partial p/\partial x$
- For the linear case, this is easily computed as the coefficient on x
- For the nonlinear probit and logit models, it's more complicated:
- $\partial p/\partial x_j = g(\beta_0 + \boldsymbol{x\beta})\beta_j$, where $g(z)$ is dG/dz

Interpretation (continued)

- Clear that it's incorrect to just compare the coefficients across the three models
- Can compare sign and significance (based on a standard *t* test) of coefficients, though
- To compare the magnitude of effects, need to calculate the derivatives, say at the means
- Stata will do this for you in the probit case

The Likelihood Ratio Test

- Unlike the LPM, where we can compute *F* statistics or *LM* statistics to test exclusion restrictions, we need a new type of test
- Maximum likelihood estimation (MLE), will always produce a log-likelihood, L
- Just as in an F test, you estimate the restricted and unrestricted model, then form
- $LR = 2(L_{ur} – L_r) \sim \chi^2_q$

Goodness of Fit

- Unlike the LPM, where we can compute an R^2 to judge goodness of fit, we need new measures of goodness of fit
- One possibility is a pseudo R^2 based on the log likelihood and defined as $1 – L_{ur}/L_r$
- Can also look at the percent correctly predicted – if predict a probability >.5 then that matches y = 1 and vice versa

Latent Variables

- Sometimes binary dependent variable models are motivated through a latent variables model
- The idea is that there is an underlying variable y*, that can be modeled as
- $y^* = \beta_0 + \boldsymbol{x\beta} + e$, but we only observe
- $y = 1$, if $y^* > 0$, and $y = 0$ if $y^* \leq 0$,

The Tobit Model

- Can also have latent variable models that don't involve binary dependent variables
- Say $y^* = \boldsymbol{x\beta} + u$, $u|\boldsymbol{x} \sim \text{Normal}(0,\sigma^2)$
- But we only observe $y = \max(0, y^*)$
- The Tobit model uses MLE to estimate both $\boldsymbol{\beta}$ and σ for this model
- Important to realize that $\boldsymbol{\beta}$ estimates the effect of $\boldsymbol{x}$ on y*, the latent variable, not y

Interpretation of the Tobit Model

- Unless the latent variable y^* is what's of interest, can't just interpret the coefficient
- $E(y|\mathbf{x}) = \Phi(\mathbf{x}\boldsymbol{\beta}/\sigma)\mathbf{x}\boldsymbol{\beta} + \sigma\phi(\mathbf{x}\boldsymbol{\beta}/\sigma)$, so
- $\partial E(y|\mathbf{x})/\partial x_j = \beta_j \Phi(\mathbf{x}\boldsymbol{\beta}/\sigma)$
- If normality or homoskedasticity fail to hold, the Tobit model may be meaningless
- If the effect of x on $P(y>0)$ and $E(y)$ are of opposite signs, the Tobit is inappropriate

Censored Regression Models & Truncated Regression Models

- More general latent variable models can also be estimated, say
- $y = \boldsymbol{x\beta} + u$, $u|\boldsymbol{x},c \sim \text{Normal}(0,\sigma^2)$, but we only observe $w = \min(y,c)$ if right censored, or $w = \max(y,c)$ if left censored
- Truncated regression occurs when rather than being censored, the data is missing beyond a censoring point

Sample Selection Corrections

- If a sample is truncated in a nonrandom way, then OLS suffers from selection bias
- Can think of as being like omitted variable bias, where what's omitted is how were selected into the sample, so
- $E(y|z, s = 1) = \boldsymbol{x\beta} + \rho\lambda(\boldsymbol{z\gamma})$, where
- $\lambda(c)$ is the inverse Mills ratio: $\phi(c)/\Phi(c)$

Selection Correction (continued)

- We need an estimate of λ, so estimate a probit of *s* (whether *y* is observed) on ***z***
- These estimates of $\boldsymbol{\gamma}$ can then be used along with ***z*** to form the inverse Mills ratio
- Then you can just regress *y* on ***x*** and the estimated λ to get consistent estimates of $\boldsymbol{\beta}$

∀ Important that ***x*** be a subset of ***z***, otherwise will only be identified by functional form