PROBIT REGRESSIONS

ECON 398B
A. JOSEPH GUSE

1. Binary Choice. Why not OLS?
   (a) A Linear Probability Model.
   (b) The Unboundedness Problem. See Studemund, p.455, for a nice diagram related to the unboundedness problem.
   (c) When is Linear Probability Good or Pretty OK? According to Hugo, Linear Probability Models have certain advantages over the probit/logit approach despite the unboundedness problem. For example, probit estimates are inconsistent unless the error term in the latent variable equation is truly normal. OLS may be more robust to specification errors. Also, says Hugo, a linear probability model may be an especially good choice when the right-hand side variables are mostly or all dummies since it will not suffer from the unboundedness problem. Really? See if you can prove this.

2. Solutions to Unboundedness Problem.
   (a) Properties of the Normal cdf.
   (b) Properties of the logistic function.

3. Setting up the Probit Model (Follows Wooldridge (2002))
   (a) The latent variable. Suppose that the true value of an observation is given by an unobserved latent variable $z_i$

   
   $$z_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

   Instead of directly observing this value we only see a binary choice $y_i$ which is equal to 1 if $z_i$ is positive and 0 if $z_i$ is negative. In other words when $z_i$ is high enough taking some action is “worth it” and all we see is wether the agent took the action or not. So in many cases we can think of the latent variable $z_i$ as the benefits minus the costs for individual $i$ of doing something. (Taking the bus, buying a camera, getting married, etc.)

   $$y_i = 1 (z_i \geq 0)$$

4. Why Probit? If we good reason to believe that residual $\varepsilon_i$ in the specification of the latent variable follows a normal distribution then the probability that $z_i$ will be positive is given by

   $$Prob(z_i > 0) = Prob(\beta_0 + \beta_1 x_i + \varepsilon_i > 0)$$

   $$= Prob(\varepsilon_i > -\beta_0 - \beta_1 x_i)$$

   $$= \Phi(\beta_0 + \beta_1 x_i)$$

   On the other hand the probability that the true $z_i$ is negative given $x_i$ is

   $$Prob(z_i < 0) = 1 - \Phi(\beta_0 + \beta_1 x_i)$$
(5) **The Likelihood Function** Given an estimate of $\beta$, what are the chances that we would observe the choices we observe? This is the question behind the likelihood function.

(a) **The odds of a single observation.** $y_i$ is either 0 or 1. You either took the bus (or got married or got arrested or...) or you didn’t. What are the chances of a single observation $y_i$ given our estimate of $\beta$?

$$Prob(y_i|x_i, \beta) = \Phi(\beta x_i)^y(1 - \Phi(\beta x_i)^{1-y})$$

Here $\beta x_i$ is shorthand for $\beta_0 + \beta_1 x_i$. What is going on here. Remember $y$ is either zero or one. If it is 0 then this expression just reduces to $(1 - \Phi(\beta x))$ which is what we said was the probability that $z_i < 0$. If $y = 1$ then the expression reduces to $\Phi(\beta x)$ which we said was the probability that $z_i > 0$.

(b) **The odds of two observations.** Now suppose you observe $(y_i, x_i)$ and $(y_j, x_j)$. Given your estimate for $\beta$ what are the chances of those two observations?

$$Prob(y_i \cap y_j|x_i, x_j, \beta) = [\Phi(\beta x_i)^y(1 - \Phi(\beta x_i)^{1-y})] \times [\Phi(\beta x_j)^y(1 - \Phi(\beta x_j)^{1-y})]$$

(c) **How about all the observations?**

$$Prob(\bigcap_i y_i|\{x_i\}, \beta) = \prod_i [\Phi(\beta x_i)^y(1 - \Phi(\beta x_i)^{1-y})]$$

(6) **Estimation by Maximum Likelihood**

(7) **Interpreting the Results** (Follows Wooldridge (2003), p.556) Suppose we estimate that $\beta_1 = .5$. What the hell does that mean? What we can guess is that a positive value for $\beta_1$ means that increasing $x$ make the action more likely ($y = 1$). But how much more likely? Let’s differentiate the probability of taking the action with respect to $x$.

$$\frac{\partial \Phi(\beta_0 + \beta_1 x)}{\partial x} = \beta_1 \phi(\beta_0 + \beta_1 x)$$

Since $\Phi$ is the normal cdf, $\phi$, the derivative, is the normal pdf. This derivative should tell us the marginal effect of increases in $x$ on the probability of taking the action. But note that it DEPENDS not only on $\beta_1$, it depends on $x$ as well. In other words, the marginal probability effect of changes in $x$ depends on $x$ itself. Well, in a way, it has to. If it were constant, we would have a linear probability model and be back to our unboundedness problem. Recall that

$$\phi(\varepsilon) = \frac{e^{-\varepsilon^2/2}}{\sqrt{2\pi}}$$

The standard normal pdf reaches a maximum value at 0 where it is equal to about .4. $\phi(0) = \frac{1}{\sqrt{2\pi}} = 0.39894$. At $\varepsilon = 1 -$ which would be one SD away from 0, it drops to about .242

(8) **Example: MacKie-Mason’s Model of Financing.** Recall that $y = 1$ means the firm finance with debt and $y = 0$ means they financed with equity. Let’s look at a couple of his estimates
<table>
<thead>
<tr>
<th>RHS Variable</th>
<th>β</th>
<th>Max Marg. Prob. Effect</th>
<th>“Sample Deriv”</th>
</tr>
</thead>
<tbody>
<tr>
<td>(in millions)</td>
<td></td>
<td>(β × .4)</td>
<td>(in percentage points)</td>
</tr>
<tr>
<td>Tax Loss Carryforwards</td>
<td>-1.86</td>
<td>-74</td>
<td>-9.36</td>
</tr>
<tr>
<td>Investment Tax Credit</td>
<td>28.8</td>
<td>11.52</td>
<td>8.54</td>
</tr>
<tr>
<td>ITC * Bankruptcy Predictor</td>
<td>-33.8</td>
<td>-13.52</td>
<td>-10.8</td>
</tr>
</tbody>
</table>

(9) Example: Bernheim et al (2004) Recall authors estimated the effect of estate tax exemption levels on probability of an inter-vivos gift. The key variables of interest were year dummies, bracket (Group I, II and III) dummies and interactions of these dummies. Instead of reporting the “betas” from the probit model, their table 2 converts the numbers to marginal probability effects. The question becomes, what method are they using to convert? Here is what they reported

<table>
<thead>
<tr>
<th>RHS Variable</th>
<th>Marg. Prob Effect</th>
<th>Stdd Err</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCF year =1995</td>
<td>0.027</td>
<td>0.034</td>
</tr>
<tr>
<td>SCF year =1998</td>
<td>0.074</td>
<td>0.035</td>
</tr>
<tr>
<td>SCF year =2001</td>
<td>0.010</td>
<td>0.041</td>
</tr>
<tr>
<td>Group 1</td>
<td>0.093</td>
<td>0.034</td>
</tr>
<tr>
<td>Group 3</td>
<td>0.053</td>
<td>0.033</td>
</tr>
<tr>
<td>Group 2 * SCF year =1995</td>
<td>0.045</td>
<td>0.060</td>
</tr>
<tr>
<td>Group 2 * SCF year =1998</td>
<td>0.108</td>
<td>0.049</td>
</tr>
<tr>
<td>Group 2 * SCF year =2001</td>
<td>0.137</td>
<td>0.051</td>
</tr>
<tr>
<td>Group 3 * SCF year =1995</td>
<td>0.066</td>
<td>0.040</td>
</tr>
<tr>
<td>Group 3 * SCF year =1998</td>
<td>0.015</td>
<td>0.041</td>
</tr>
<tr>
<td>Group 3 * SCF year =2001</td>
<td>0.008</td>
<td>0.041</td>
</tr>
</tbody>
</table>

The stata manual gives us some clue about how we should probably interpret these numbers saying “By default, margins evaluates this derivative for each observation and reports the average of the marginal effects.” (http://www.stata.com/support/faqs/stat/mfx_size.html)

1. References