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

$$y^* = \beta_0 + x\beta_1 + u, \ u|x \sim N(0, \sigma^2)$$

but we only observe

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

and

$$y = 0 \text{ if } y^* \leq 0$$
This formulation is equivalent to a Probit model. To see this note that

\[
P(y = 1|x) = P(y^* > 0|x) \\
= P(\beta_0 + x\beta_1 > -u|x) \\
= P(- (\beta_0 + x\beta_1)/\sigma < u/\sigma|x) \\
= 1 - \Phi(- (\beta_0 + x\beta_1)/\sigma) \\
= \Phi((\beta_0 + x\beta_1)/\sigma)
\]

- We note form the last line that we can not separately estimate the scale of \(\beta_0, \beta_1\) and \(\sigma\).
The Tobit Model

- We can also have latent variable models that don’t involve binary dependent variables.

- The Tobit model can be used to handle datasets where outcome variables have natural limitations of definition. For example, quantities consumed of a particular good can not be negative.

- If a significant fraction of a population decides to consume zero amounts of a particular good, then a linear regression model may not be well suited to describe the outcome variable. In particular, as with the linear probability model, a linear model potentially produces negative predictors for the outcome variable.
• An example of such a good is the amount spent on alcohol in a given month.

• A second example is hours worked in the sample of married women. Of the 753 women in the sample, 428 worked and had annual working hours ranging from 12 to 4950 while 325 women did not work and had annual working hours equal to zero. This sample is thus well suited for the Tobit model with a substantial fraction of the sample being at the corner solution of zero and the remainder of the sample having a larger number of different values of working hours.

• We are looking for a model which does not imply negative predictions for some choices of the $x$ variables. This can be achieved by specifying a latent variable.
• Define the latent variable $y^*$ as a function of observed covariates $x$

$$y^* = \beta_0 + x\beta_1 + u, \quad u|x \sim \text{Normal}(0, \sigma^2)$$

• But we only observe

$$y = \max(0, y^*).$$

• Because, conditional on $x$, $y^*$ has continuous distribution with a normal density function, it follows that $y$ has the same continuous distribution when $y^* > 0$.

• Furthermore, when $y^* \leq 0$, it follows that $y$ has point mass at $y = 0$. In
other words,

\[ P(y = 0|x) = P(y^* \leq 0|x) \]
\[ = P(u \leq -(\beta_0 + x\beta_1)) \]
\[ = P(u/\sigma \leq -(\beta_0 + x\beta_1)/\sigma) \]
\[ = \Phi(-(\beta_0 + x\beta_1)/\sigma) \]
\[ = 1 - \Phi((\beta_0 + x\beta_1)/\sigma) \]

- The Tobit model uses MLE to estimate both \( \beta \) and \( \sigma \) for this model.

- The log-likelihood function is given for each observation pair \((y_i, x_i)\) as

\[
l_i(\beta, \sigma) = 1(y_i = 0) \log (1 - \Phi ((\beta_0 + x\beta_1)/\sigma)) + 1(y_i > 0) \log (\sigma^{-1} \phi ((y_i - \beta_0 + x_i\beta_1)/\sigma))
\]
where $\phi$ is the density of the standard normal distribution

$$
\phi (z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}
$$

and $\Phi (z)$ is the cumulative distribution function of the standard normal defined as

$$
\Phi (z) = \int_{-\infty}^{z} \phi (x) \, dx = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \, dx.
$$

- The log-likelihood function of the sample then is obtained by summing over the individual log lik functions:

$$
l(\beta, \sigma) = \sum_{i=1}^{n} l_i (\beta, \sigma)
$$

and the parameter estimates $\hat{\beta}$ and $\hat{\sigma}$ are obtained from maximizing the log-likelihood:

$$
\hat{\beta}, \hat{\sigma} = \arg \max_{\beta,\sigma} l(\beta, \sigma)
$$
• Statistical software packages such as Stata can compute the parameters and associated standard errors and t-statistics. They can also carry out joint hypothesis tests.

• A possible measure for the goodness of fit is the squared correlation between \( y \) and \( \hat{y} \) where

\[
\hat{y} = E(y|x).
\]

• Tobit models (as any nonlinear model) are sensitive to the assumptions used to derive them. In particular, we assumed that the innovations are Gaussian and that they are homoskedastic. If these two assumptions do not hold then the Tobit model is not valid and will likely give misleading results. This is quite different from linear regression which is less sensitive to misspecification of this form.
• We can check if a probit model of the participation decision leads to different parameter estimates (compare $\hat{\beta}/\hat{\sigma}$ to probit parameters). If specification is a possible problem the parameter estimates are likely different. In the latter case it is possible to estimate models (called two part models) that allow for different parameters in the selection and outcome part.
Interpretation of the Tobit Model

- It is important to realize that $\beta$ estimates the effect of $x$ on $y^*$, the latent variable, not on $y$.

- Thus, the parameters of the Tobit model cannot be directly interpreted in the same way we interpret parameters in a linear regression model.

- In Tobit models two types of expectations are of interest:

  \[ E(y|y > 0, x) \]

  is the expectation of $y$ given that $y > 0$ and $x$.

- The second expectation of interest is

  \[ E(y|x) \]
We note that

\[ y = y \cdot 1(y > 0) + y \cdot 1(y = 0) \]

\[ = y \cdot 1(y > 0) \]

such that for \( y|x \) with conditional density \( f(y|x) \)

\[ E(y|x) = \int_{y>0} y f(y|x) \, dy \]

\[ = P(y > 0|x) \int_{y>0} y \frac{f(y|x)}{P(y > 0|x)} \, dy \]

\[ = P(y > 0|x) E(y|y > 0, x) \]

- To evaluate \( E(y|y > 0, x) \) we need a result for normally distributed random variables:
• If \( z \sim N(0, 1) \) then

\[
E(z|z > c) = \int_{z>c} z \frac{\phi(z)}{1 - \Phi(c)} dz
\]

\[
= \frac{\phi(c)}{1 - \Phi(c)}
\]

because

\[
de^{-\frac{1}{2}z^2} = -ze^{-\frac{1}{2}z^2} = -z\phi(z)
\]

such that

\[
\int_{a}^{b} z\phi(z) dz = \phi(b) - \phi(a).
\]

• It then follows from this that

\[
E(y|y > 0, x) = E(\beta_0 + x\beta_1 + u|y > 0, x)
\]
because when \( y > 0 \), it follows that \( y = \beta_0 + x\beta_1 \). Then,

\[
E(\beta_0 + x\beta_1 + u|y > 0, x) = \beta_0 + x\beta_1 + E(u|y > 0, x)
\]

\[
= \beta_0 + x\beta_1 + E(u|u > -(\beta_0 + x\beta_1))
\]

\[
= \beta_0 + x\beta_1 + \sigma E\left(\frac{u|u > -\beta_0 + x\beta_1}{\sigma}\right)
\]

But now it follows that \( u/\sigma \) is a standard normal random variable. Thus, we can apply the previous result and deduce that for \( z = u/\sigma \) and \( c = -\frac{\beta_0 + x\beta_1}{\sigma} \) it follows that

\[
E\left(\frac{u|u > -\beta_0 + x\beta_1}{\sigma}\right) = \frac{\phi\left(-\frac{\beta_0 + x\beta_1}{\sigma}\right)}{1 - \Phi\left(-\frac{\beta_0 + x\beta_1}{\sigma}\right)}
\]

\[
= \frac{\phi\left(\frac{\beta_0 + x\beta_1}{\sigma}\right)}{\Phi\left(\frac{\beta_0 + x\beta_1}{\sigma}\right)}
\]

because

\[
\phi(c) = \phi(-c)
\]
and

\[ \Phi (c) = 1 - \Phi (-c) \]

- It is useful to define the so called inverse Mills ratio as

\[ \lambda (c) = \frac{\phi (c)}{\Phi (c)}. \]

- We can then write that

\[ E(y|y > 0, x) = \beta_0 + x\beta_1 + \sigma \lambda \left( \frac{\beta_0 + x\beta_1}{\sigma} \right) \]

- This equation is important. It shows that running OLS only on the observations where \( y_i > 0 \), estimating the equation

\[ y_i = \beta_0 + x_i\beta_1 + u_i \]
for the observations where $y_i > 0$ will lead to biased estimates for $\beta_0 \beta_1$ because the inverse Mills ratio is essentially an omitted variable which is correlated with $x$ and thus leads to biased OLS estimators.

- We can now return to $E(y|x)$. Combining previous results we find

\[
E(y|x) = P(y > 0|x) E(y|y > 0, x) \\
= \phi \left( \frac{\beta_0 + x\beta_1}{\sigma} \right) \left( \beta_0 + x\beta_1 + \sigma \lambda \left( \frac{\beta_0 + x\beta_1}{\sigma} \right) \right) \\
= \phi \left( \frac{\beta_0 + x\beta_1}{\sigma} \right) (\beta_0 + x\beta_1) + \sigma \phi \left( \frac{\beta_0 + x\beta_1}{\sigma} \right).
\]

- It can be shown that $E(y|x) > 0$ for all values of $x$ and $\beta$. It also follows immediately that $E(y|x)$ is not a linear function of $x$ and thus that a usual linear regression model would be misspecified.

- The Tobit model depends on the correctness of the normality assumption.
• The interpretation of the parameters $\beta$ becomes more difficult than in the linear model. We need to compute partial effects of changing $x$ as we have done for the Logit and Probit model. These partial effects depend not only on $\beta$ but also on $x$ and $\sigma$.

• We start by computing the effect of a change of $x$ on $E(y|y > 0, x)$. We have

$$\frac{\partial E(y|y > 0, x)}{\partial x} = \frac{\partial \left( \beta_0 + x\beta_1 + \sigma \lambda \left( \frac{\beta_0 + x\beta_1}{\sigma} \right) \right)}{\partial x}$$

$$= \beta_1 + \frac{\sigma \partial \lambda \left( \frac{\beta_0 + x\beta_1}{\sigma} \right)}{\partial x}$$
where
\[
\frac{\partial \lambda (c)}{\partial c} = \frac{\partial \phi (c) / \partial c}{\Phi (c)} - \frac{\phi (c)}{\Phi (c)} \frac{\partial \Phi (c)}{\partial c} \\
= -\frac{\phi (c) c}{\Phi (c)} - \frac{\phi (c)^2}{\Phi (c)^2} \\
= -\lambda (c) (c + \lambda (c))
\]

- It then follows that
\[
\frac{\partial E(y|y > 0, x)}{\partial x} = \beta_1 \left( 1 - \lambda \left( \frac{\beta_0 + x \beta_1}{\sigma} \right) \left( \frac{\beta_0 + x \beta_1}{\sigma} + \lambda \left( \frac{\beta_0 + x \beta_1}{\sigma} \right) \right) \right)
\]
Which shows that the partial effect does not just depend on \( \beta_1 \) but also on an adjustment factor
\[
\left( 1 - \lambda \left( \frac{\beta_0 + x \beta_1}{\sigma} \right) \left( \frac{\beta_0 + x \beta_1}{\sigma} + \lambda \left( \frac{\beta_0 + x \beta_1}{\sigma} \right) \right) \right)
\]
which can be shown to be strictly between 0 and 1.
• This equation shows that the partial effect also depends on $\sigma$. Thus, $\sigma$ needs to be estimated in order to compute the partial effects.

• As in the case of Logit and Probit partial effects are computed by evaluating

$$\frac{\partial E (y|y > 0, x)}{\partial x}$$

at the estimated parameter values $\hat{\beta}_0, \hat{\beta}_1$ and $\hat{\sigma}$ and some specific value for $x_i$. One popular choice is to use the sample average $\bar{x}$.

• If $x$ is binary then the effect of interest is the difference

$$E (y|y > 0, x = 1) - E (y|y > 0, x = 0)$$

rather than the derivative. Similarly, discrete changes are computed for discrete variables such as the number of children.
• Stata can carry out these calculations automatically.

• We can also compute the partial effect on $E(y|x)$. This partial effect takes into account that people starting out at $y = 0$ might decide the switch to $y > 0$.

• We have

$$\frac{\partial E(y|x)}{\partial x} = \frac{\partial (P(y > 0|x) E(y|y > 0, x))}{\partial x}$$

and by the product rule of differentiation

$$\frac{\partial E(y|x)}{\partial x} = \frac{\partial P(y > 0|x)}{\partial x} E(y|y > 0, x) + P(y > 0|x) \frac{\partial E(y|y > 0, x)}{\partial x}.$$ 

• Because

$$P(y > 0|x) = \Phi \left( \frac{\beta_0 + x\beta_1}{\sigma} \right)$$
it follows that

$$\frac{\partial P(\mathbf{y} > 0|\mathbf{x})}{\partial x} = \left(\frac{\beta_1}{\sigma}\right) \phi \left(\frac{\beta_0 + x\beta_1}{\sigma}\right).$$

- Using the fact that

$$\Phi(c) \lambda(c) = \phi(c)$$

and substituting into the right hand side expression for $\frac{\partial E(\mathbf{y}|\mathbf{x})}{\partial x}$ we obtain

$$\frac{\partial E(\mathbf{y}|\mathbf{x})}{\partial x} = \beta_1 \Phi \left(\frac{\beta_0 + x\beta_1}{\sigma}\right)$$

- This last equation gives us a way to compare the OLS and Tobit estimates. Remember that for a linear regression model we have

$$\frac{\partial E(\mathbf{y}|\mathbf{x})}{\partial x} = \beta_1$$
so that

$$\Phi \left( \frac{\beta_0 + x\beta_1}{\sigma} \right)$$

is an adjustment factor stemming from the Tobit regression.

- Also note that

$$P(y > 0|x) = \Phi \left( \frac{\beta_0 + x\beta_1}{\sigma} \right)$$

which means that if

$$P(y > 0|x) = 1$$

then the Tobit and OLS coefficients are the same. This makes sense because

$$P(y > 0|x_i) = 1$$

for all $$x_i$$ means that there are no cases in the sample where $$y_i = 0$$. 

Example: Hours worked

- In the Mzor.dta dataset we compute

\[ n^{-1} \sum_{i=1}^{n} \Phi \left( \frac{\hat{\beta}_0 + x\hat{\beta}_1}{\hat{\sigma}} \right) = .589 \]

- This means that the marginal effects are approximately equal to the parameter estimate times .589.

- Comparing these results with conventional OLS estimates, we find that the effects estimated by Tobit are generally larger.