ECON 121, Applied Econometrics and Data Analysis **Prof. Tom Vogl**

Fall 2023 **University of California San Diego**

LECTURE NOTE 7: BINARY DEPENDENT VARIABLES

1 Introduction

In this lecture note, we take a fresh look at the estimation of models with binary dependent variables. We will first discuss the linear probability model, exploring estimation details that you may not have discussed in your introductory econometrics course. We will then move on to the estimation of probit and logit models by maximum likelihood.

2 Linear Probability Model

The simplest approach to estimate an equation with a binary dependent variable is the linear probability model. Suppose D_i takes on values 0 or 1. We write the model:

$$
\begin{aligned}
D_i &= \beta_0 + \beta_1 X_{1i} + \cdots + \beta_K X_{Ki} + \varepsilon_i \\
&= X_i' \beta + \varepsilon_i
\end{aligned}
$$

where the error term ε_i has expectation zero conditional on $X_{1i}, \cdots, X_{Ki}$. We interpret $\hat{D}_i = X_i'\beta = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_K X_{Ki}$ as the expected probability that $D_i = 1$ given $X_{1i}, \cdots, X_{Ki}$. Thus, we interpret $\beta_k = \frac{\partial Pr[D_i=1]}{\partial X_{ki}}$ as the derivative of the expected probability with respect to X_{ki}. If X_{ki} is a binary variable, then $\beta_k = Pr[D_i = 1|X_{ki} = 1, X_{ji} = x_{ji} \forall j \neq k] - Pr[D_i = 1|X_{ki} = 0, X_{ji} = x_{ji} \forall j \neq k]$. (We hold all other X_{jt} constant at some value x_{jt}, although the value does not matter.) We can then use our coefficient estimates to generate a predicted probability: $\widehat{Pr}[D_i = 1|X_i] = X_i'\hat{\beta} = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \cdots + \hat{\beta}_K X_{Ki}$.

Two problems plague the linear probability model. The first, probably better known, problem is that $\widehat{Pr}[D_i = 1]$ may lie outside the $[0, 1]$ interval. The second is that a applying a linear model to a binary outcome creates heteroskedasticity, which makes the estimator inefficient.

3 Probit and Logit Models

Lecture Note 4 discussed maximum likelihood estimation of a Bernoulli random variable. If D_i is a binary variable that equals 1 with probability p and equals 0 with probability $1-p$, then the likelihood is:

$$L = \prod_{i=1}^{N} p^{D_i} (1-p)^{(1-D_i)}$$

and the log-likelihood is:

$$\ln L = \sum_{i=1}^{N} \{D_i \ln(p) + (1-D_i) \ln(1-p)\}$$

In Lecture Note 4, we were okay with having a single p for the overall sample, Now, however, we are interested in allowing p_i to be individual-specific in a way that depends on X_i. To do so, let us define a function G and parameters β such that $p_i = G(X_i'\beta)$. Then the likelihood becomes:

$$L = \prod_{i=1}^{N} G(X_i'\beta)^{D_i} [1 - G(X_i'\beta)]^{(1-D_i)}$$

and the log-likelihood becomes:

$$\ln L = \sum_{i=1}^{N} \{D_i \ln [G(X_i'\beta)] + (1-D_i) \ln [1 - G(X_i'\beta)]\}$$

Now we only need to choose an appropriate function G.

Because p_i (and therefore also $G(X_i'\beta)$) is a probability, we want the function G to satisfy a few properties. First, its range should be the interval $[0,1]$. Second, as $X_i'\beta$ approaches infinity, $G(X_i'\beta)$ should approach 1. Third (and analogously), as $X_i'\beta$ approaches negative infinity, $G(X_i'\beta)$ should approach 0. Note that these properties are central features of cumulative distribution functions (CDFs). Indeed, the two most common choices for G are the CDFs for the normal distribution and the logistic distribution. We write $\Phi[X_i'\beta]$ for the standard normal distribution, and $\Lambda[X_i'\beta]$ for the logistic distribution. The method using the standard normal distribution is called *probit regression*, and the method using the logistic distribution is called *logit* or *logistic regression*. The next two subsections describe the likelihood functions for both methods. In neither case does the maximum likelihood problem have a closed-form solution, so statistical software uses numerical optimization methods to find the maximum of the likelihood function.

3.1 Probit Likelihood

The probit model uses the standard normal CDF:

$$Pr[D_i = 1|X_i] = \Phi[X_i'\beta] = \int_{-\infty}^{X_i'\beta} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(x'\beta\right)^2\right\} dx$$

Thus, the likelihood for the probit is:

$$L = \prod_{i=1}^{N} \Phi\left[X_i'\beta\right]^{D_i} \left[1 - \Phi[X_i'\beta]\right]^{(1-D_i)}$$

and the log-likelihood is:

$$\begin{aligned}
\ln L &= \sum_{i=1}^{N} \left\{D_i \ln\left[\Phi[X_i'\beta]\right] + (1 - D_i)\ln\left[1 - \Phi[X_i'\beta]\right]\right\} \\
&= \sum_{i=1}^{N} \left\{D_i \ln\left[\Phi[X_i'\beta]\right] + (1 - D_i)\ln\left[\Phi[-X_i'\beta]\right]\right\}
\end{aligned}$$

where the second line follows because the normal distribution is symmetric. Implementation in Stata uses the `probit` command. Implementation in R uses the `glm()` function, with `family` set to `binomial(link="probit")`.

3.2 Logit Likelihood

The logit model uses the logistic CDF:

$$Pr[D_i = 1|X_i] = \Lambda[X_i^{'}\beta] = \frac{\exp\left\{X_i'\beta\right\}}{1 + \exp\left\{X_i'\beta\right\}} = \frac{1}{1 + \exp\left\{-X_i'\beta\right\}}$$

where $\exp\left\{X_i'\beta\right\}$ is alternative notation for $e^{X_i'\beta}$. The likelihood for the logit is:

$$L = \prod_{i=1}^{N} \Lambda\left[X_i'\beta\right]^{D_i} \left[1 - \Lambda[X_i'\beta]\right]^{(1-D_i)}$$

and the log-likelihood is:

$$\begin{aligned}
\ln L &= \sum_{i=1}^{N} \left\{D_i \ln\left[\Lambda[X_i'\beta]\right] + (1 - D_i)\ln\left[1 - \Lambda[X_i'\beta]\right]\right\} \\
&= \sum_{i=1}^{N} \left\{D_i \ln\left[\Lambda[X_i'\beta]\right] + (1 - D_i)\ln\left[\Lambda[-X_i'\beta]\right]\right\}
\end{aligned}$$

where the second line follows because the logistic distribution is symmetric. Implementation in Stata uses the `logit` command. Implementation in R uses the `glm()` function, with `family` set to `binomial(link="logit")`.

4 Latent Variables Representation

As an alternative way to conceptualize the problem, we can think of the binary dependent variable as a discretization of an underlying (unobserved) continuous random variable. We suppose that a continuous variable Y_i exists, but we cannot observe it. Y_i is determined as follows:

$$Y_i = X_i'\beta + \varepsilon_i$$

where ε_i has either a logistic or a normal distribution. However, we can only observe the binary variable D_i:

$$D_i = \begin{cases} 1 & \text{if } Y_i > 0 \\ 0 & \text{if } Y_i \leq 0 \end{cases}$$

In this setup, D_i is just a coarse version of the underlying latent variable Y_i.

We can derive the probability that $D_i = 1$:

$$\begin{aligned} Pr[D_i = 1|X_i] &= Pr[Y_i > 0|X_i] \\ &= Pr[X_i'\beta + \varepsilon_i > 0|X_i] \\ &= Pr[\varepsilon_i > -X_i'\beta|X_i] \\ &= 1 - F[-X_i'\beta] \\ &= F[X_i'\beta] \end{aligned}$$

where $F[\cdot]$ is either the standard normal CDF or the logistic CDF. The last step follows from the fact that both distributions have mean zero and are symmetric. This derivation shows that the latent variables representations are equivalent to the probit and logit models we originally specified.

5 Interpreting the Results of Probit and Logit Models

Despite their limitations, linear probability models have the attractive property that we can interpret their coefficients as changes in the absolute probability of the event that $D_i = 1$. In contrast, probit and logit results are more difficult to interpret. In those models, the coefficients depend on the variance we choose for the latent variable Y_i, which is arbitrary. As a result, the signs and relative sizes of the coefficients within a regression are meaningful, but one generally cannot compare the sizes of coefficients across regressions. Furthermore, even within a single regression, the changes in probability implied by the coefficients are not always obvious. Nonetheless, we can use the probit and logit coefficients to calculate more interpretable quantities. I outline two approaches below. In both cases, we will take interest in assessing the *magnitude* of the results. For significance tests, we can rely on the original probit or logit coefficients and standard errors.

5.1 Marginal Effects

In the economics literature, by far the most common approach is to compute what are called the "marginal effect" of each X_{ki}. The marginal effect corresponds to the derivative $\frac{\partial Pr[D_i=1]}{\partial X_{ki}}$, which is exactly the same as the estimand in the linear probability model. We write:

$$\frac{\partial Pr[D_i = 1]}{\partial X_{ki}} = \frac{\partial F[X_i'\beta]}{\partial X_{ki}} = F'[X_i'\beta]\beta_k = f(X_i'\beta)\beta_k$$

where $F[\cdot]$ and $f(\cdot)$ are the CDF and PDF for either the normal or the logistic distribution. You can see clearly here that the marginal effect of X_{ki} depends on the values of the other X_{ji}'s. This dependency is due to the non-linearity of the model; it does not arise in OLS. We usually choose either to set each X_{ji} equal to its mean or to compute $\frac{\partial Pr[D_i=1]}{\partial X_{ki}}$ for every observation and then average them. In Stata, you can type `mfx compute` (for marginal effects at the average X_{ji}) or `margins,dydx(*)` (for average marginal effects) after running a probit or logit regression. Stata also has a command `dprobit` that automatically reports marginal effects at the average X_{ji}. In R, you can use `logitmfx()` or `probitmfx()` from the `mfx` package. In practice, logits and probits almost always produce nearly identical marginal effects, and the marginal effect at the average is almost always nearly identical to the average marginal effect.

5.2 Odds Ratios

Finally, the logit model has a convenient property for measuring the proportional effects of a change in X_{ki}. Recall that is an event has probability p, then the odds of the even are $\frac{p}{1-p}$. In the logit model, we have $p = \frac{\exp\left\{X_i'\beta\right\}}{1+\exp\left\{X_i'\beta\right\}}$, so that we can write the logarithm of the odds (or the "log-odds") as:

$$\ln(\text{odds}) = \ln\left(\frac{\frac{\exp\left\{X_i'\beta\right\}}{1+\exp\left\{X_i'\beta\right\}}}{1-\frac{\exp\left\{X_i'\beta\right\}}{1+\exp\left\{X_i'\beta\right\}}}\right) = \ln\left(\exp\left\{X_i'\beta\right\}\right) = X_i'\beta$$

The log-odds are linear in X_i in the logit model. (For this reason, the log-odds of an event are often called the logit of an event.) As a result, β_k measures the derivative of the log-odds with respect to X_{ki}. Note that the equation above implies that we can write the odds in the logit model as:

$$\text{odds} = e^{X_i'\beta} = e^{\beta_0+\beta_1 X_{1i}+\cdots+\beta_K X_{Ki}}$$

Consider the effect of changing X_{1i} from 0 to 1. Above, we represented this effect in terms of the *absolute* change in probability. As an alternative, we can calculate the *proportional* change in the odds. We do so using an *odds ratio*:

$$\begin{aligned}
\text{odds ratio} &= \frac{\frac{Pr[D_i=1|X_{1i}=1,X_{2i},\cdots,X_{Ki}]}{1-Pr[D_i=1|X_{1i}=1,X_{2i},\cdots,X_{Ki}]}}{\frac{Pr[D_i=1|X_{1i}=0,X_{2i},\cdots,X_{Ki}]}{1-Pr[D_i=1|X_{1i}=0,X_{2i},\cdots,X_{Ki}]}} \\
&= \frac{\exp\{\beta_0+\beta_1\cdot 1+\beta_2X_{2i}+\cdots+\beta_KX_{Ki}\}}{\exp\{\beta_0+\beta_1\cdot 0+\beta_2X_{2i}+\cdots+\beta_KX_{Ki}\}} \\
&= \frac{\exp\{\beta_0\}}{\exp\{\beta_0\}}\cdot\frac{\exp\{\beta_1\cdot 1\}}{\exp\{\beta_1\cdot 0\}}\cdot\frac{\exp\{\beta_2X_{2i}+\cdots+\beta_KX_{Ki}\}}{\exp\{\beta_2X_{2i}+\cdots+\beta_KX_{Ki}\}} \\
&= e^{\beta_1}
\end{aligned}$$

So the exponentiated logit coefficient is the odds ratio! Consequently, many researchers report the exponentiated coefficient rather than the actual coefficient or the marginal effect. This representation is especially popular outside of economics, perhaps because economists like thinking about marginal changes, but researchers in many other fields like thinking about discrete changes. For instance, when epidemiologists want to know whether exposure to a toxin is associated with birth defects, they will often run logit regressions and report the exponentiated coefficient on the dummy for toxin exposure. That exponentiated coefficient represents the proportional change in the odds of a birth defect that is associated with exposure to a toxin. If the odds ratio is below 1, then toxins are negatively associated with the odds of a birth defect; if the odds ratio is above 1, the association is positive. Notably, the proportional change in the odds is *not* the same as the proportional change in the probability (i.e., the risk of a birth defect). The proportional change in the probability, $\frac{Pr[D_i=1|X_{1i}=1,X_{2i},\cdots,X_{Ki}]}{Pr[D_i=1|X_{1i}=0,X_{2i},\cdots,X_{Ki}]}$, is called the *relative risk*. The relative risk and the odds ratio always fall on the same side of 1, but the odds ratio is more extreme than the relative risk. That is to say, $|OR-1|\geq|RR-1|$.

In Stata, you can obtain the odds ratio in two ways. First, you can run the original logit command with the odds ratio option: `logit d x,robust or`. Second, you can use the `logistic` command, which automatically reports odds ratios: `logistic d x,robust`. In R, you can use `logitor()` from the `mfx` package.