Econometrics

Qualitative and Limited Dependent Variable Models

Anna Donina

Lecture 8

Introduction

So far, the dependent variable (Y) was continuous:

- Average wage
- Number of children
- Money growth rate

But what if it is a binary variable?

$Y = 1$, if person has college degree, 0 otherwise;

$Y = 1$, if person smokes, 0 otherwise;

The linear probability model (LPM)

Non-linear probability model

- Probit
- Logit

Limited Dependent Variable Models

- **Limited dependent variables (LDV)**
  - LDV are variables whose range is substantively restricted
    - Binary variables, e.g. employed/not employed
    - Nonnegative variables, e.g. wages, prices, interest rates
    - Nonnegative variables with excess zeros, e.g. labor supply
    - Count variables, e.g. the number of arrests in a year
    - Censored variables, e.g. unemployment durations

A Binary Dependent Variable: The Linear Probability Model

- **Linear regression when the dependent variable is binary**

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

$$\Rightarrow \quad E(y|\mathbf{x}) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k$$

$$E(y|\mathbf{x}) = 1 \cdot P(y=1|\mathbf{x}) + 0 \cdot P(y=0|\mathbf{x})$$

If the dependent variable only takes on the values 1 and 0

$$\Rightarrow \quad \boxed{P(y=1|\mathbf{x}) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k}$$

<u>Linear probability model (LPM)</u>

$$\Rightarrow \quad \beta_j = \partial P(y=1|\mathbf{x}) / \partial x_j$$

In the linear probability model, the coefficients describe the effect of the explanatory variables <u>on the probability that y=1 (the probability of „success“)</u>

A Binary Dependent Variable: The Linear Probability Model

- **Example: Labor force participation of married women**

=1 if in labor force, =0 otherwise

Non-wife income (in thousand dollars per year)

$$\widehat{inlf} = \underset{(.154)}{.586} - \underset{(.0014)}{.0034}\ nwifeinc + \underset{(.007)}{.038}\ educ + \underset{(.006)}{.039}\ exper$$

$$- \underset{(.00018)}{.00060}\ exper^2 - \underset{(.002)}{.016}\ age - \underset{(.034)}{.262}\ kidslt6$$

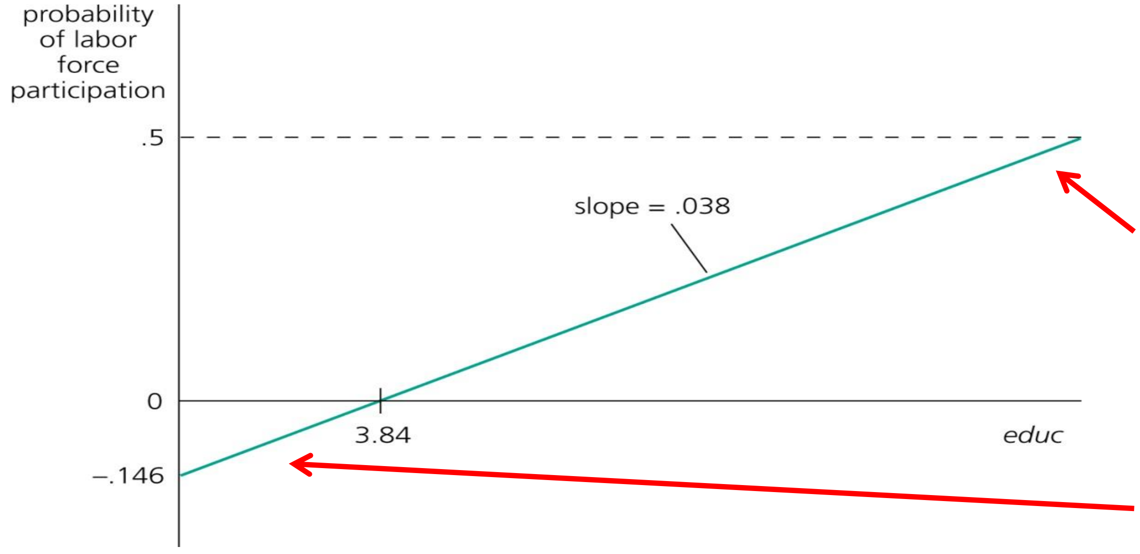
$$+ \underset{(.0132)}{.0130}\ kidsge6,\ n = 753, R^2 = .264$$

If the number of kids under six years increases by one, the proprobability that the woman works falls by 26.2%

Not significant

A Binary Dependent Variable: The Linear Probability Model

- **Example: Female labor participation of married women (cont.)**



Graph for nwifeinc=50, exper=5, age=30, kindslt6=1, kidsge6=0

The maximum level of education in the sample is educ=17. For the given case, this leads to a predicted probability to be in the labor force of about 50%.

Negative predicted probability but no problem because no woman in the sample has educ < 5.

The Linear Probability Model: Heteroskedasticity

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki} + u_i$$

The variance of a Bernoulli random variable:

$$Var(Y) = Pr(Y = 1) \times (1 - Pr(Y = 1))$$

We can use this to find the conditional variance of the error term

$$\begin{aligned}
Var(u_i|X_{1i}, \cdots, X_{ki}) &= Var(Y_i - (\beta_0 + \beta_1 X_{1i} + \cdots \beta_k X_{ki})|\, X_{1i}, \cdots, X_{ki}) \\
&= Var(Y_i|\, X_{1i}, \cdots, X_{ki}) \\
&= Pr(Y_i = 1|\, X_{1i}, \cdots, X_{ki}) \times (1 - Pr(Y_i = 1|\, X_{1i}, \cdots, X_{ki})) \\
&= (\beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki}) \times (1 - \beta_0 - \beta_1 X_{1i} - \cdots - \beta_k X_{ki}) \\
&\neq \sigma_u^2
\end{aligned}$$

Solution: always use heteroskedasticity robust standard errors when estimating a LPM

A Binary Dependent Variable: The Linear Probability Model

- **Disadvantages of the linear probability model**
  - Predicted probabilities may be larger than one or smaller than zero
  - Marginal probability effects sometimes logically impossible
  - <u>The linear probability model is necessarily heteroskedastic</u>

$$Var(y|\mathrm{x}) = P(y=1|\mathrm{x})\left[1 - P(y=1|\mathrm{x})\right]$$

Variance of Bernoulli variable

  - Heterosceasticity consistent standard errors need to be computed
- **Advantanges of the linear probability model**
  - Easy estimation and interpretation
  - Estimated effects and predictions often reasonably good in practice

Logit and Probit Models for Binary Response

- **Disadvantages of the LPM for binary dependent variables**
  - Predictions sometimes outside the unit interval
  - Partial effects of explanatory variables are constant
- **Nonlinear models for binary response**
  - Response probability is a nonlinear function of explanat. variables

$$P(y=1|\mathbf{x}) = G(\beta_0+\beta_1 x_1+\cdots+\beta_k x_k) = G(\mathbf{x}\boldsymbol{\beta})$$

Probability of a „success" given explanatory variables

A cumulative distribution function $0 < G(z) < 1$. The response probability is thus a function of the explanatory variables x.

Shorthand vector notation: the vector of explanatory variables x also contains the constant of the model.

Logit and Probit Models for Binary Response

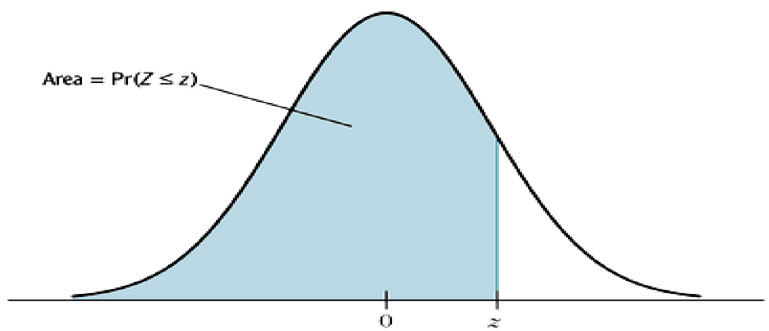
- **Choices for the link function**

Probit: $G(z) = \Phi(z) = \int_{-\infty}^{z} \phi(v)dv$ (standard normal distribution)

Logit: $G(z) = \Lambda(z) = \exp(z)/\left[1 + \exp(z)\right]$ (logistic function)

Logit and Probit Models for Binary Response

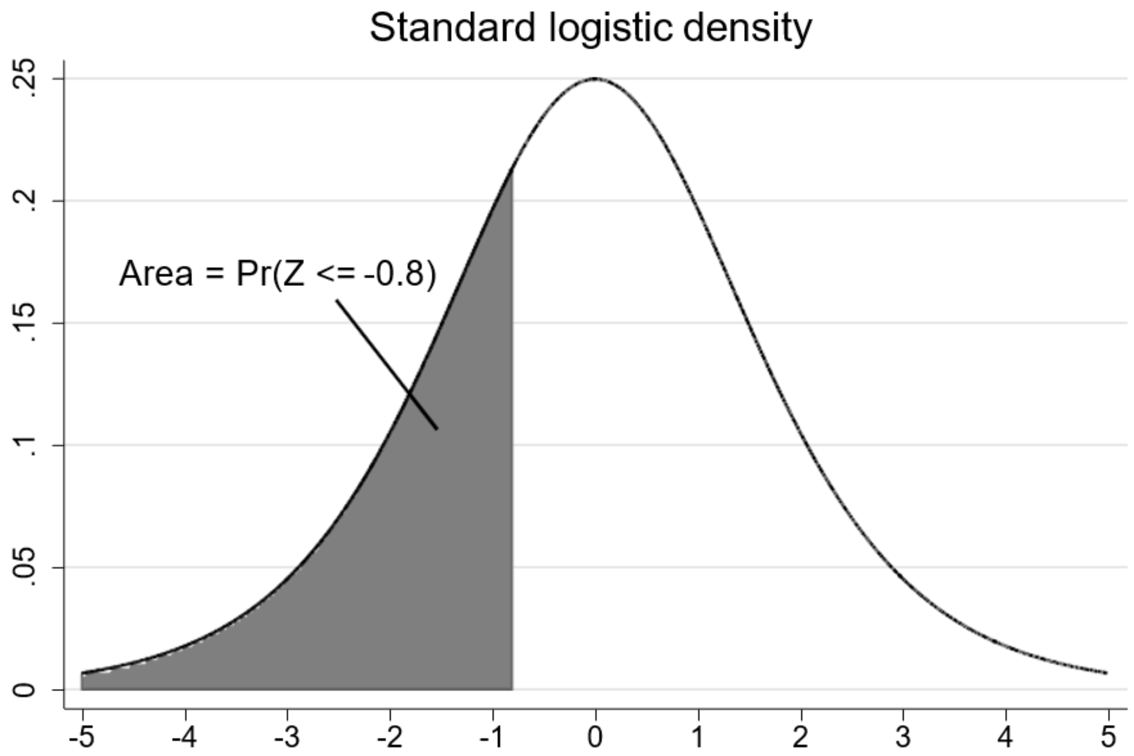
TABLE 1 The Cumulative Standard Normal Distribution Function, $\Phi(z) = Pr(Z$ ″ $z)$



z	Second Decimal Value of z									
	0	1	2	3	4	5	6	7	8	9
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
−0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121

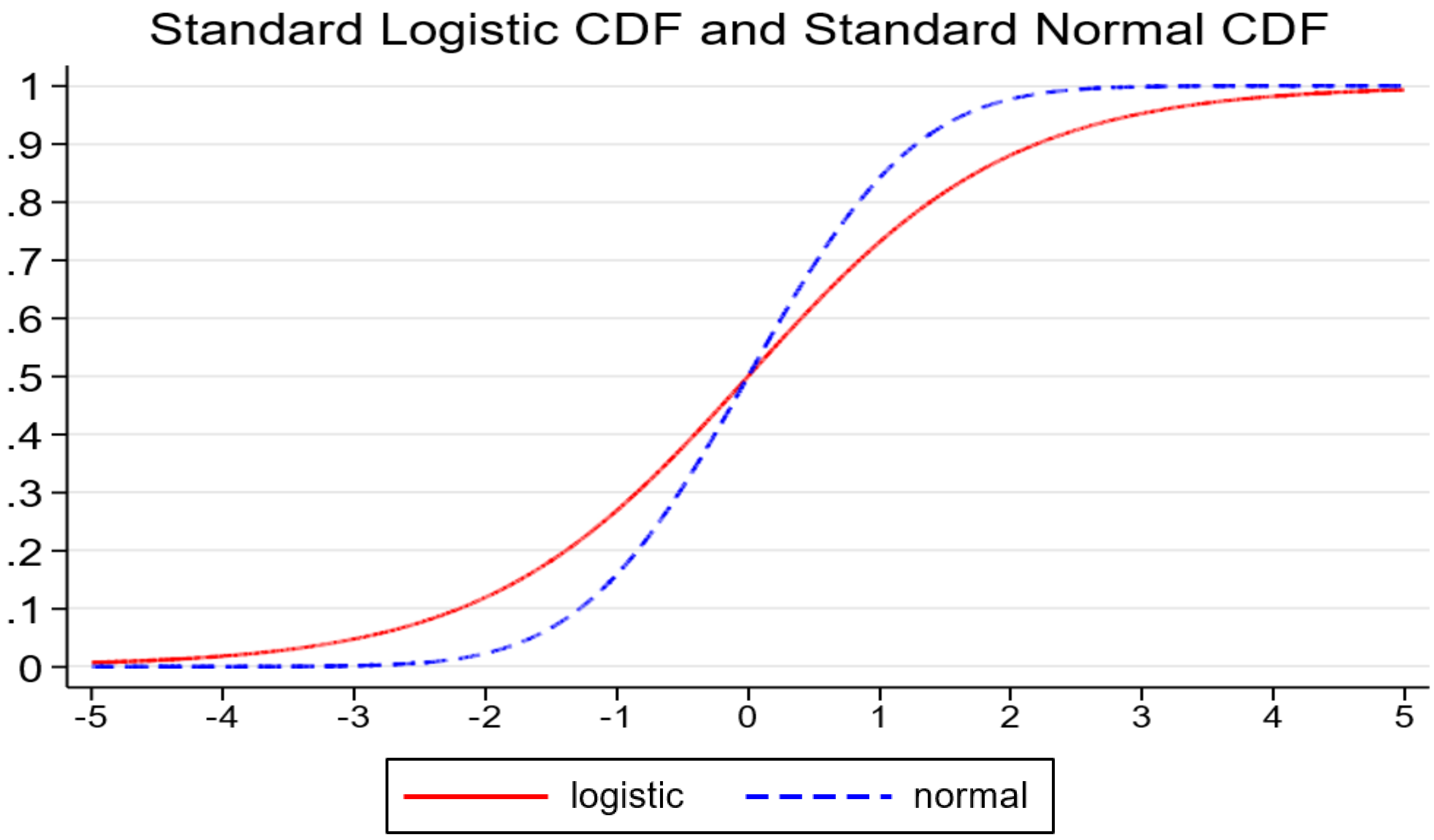
$$Pr(Y = 1) = Pr(Z \leq -0.8) = \Phi(-0.8) = 0.2119$$

Logit and Probit Models for Binary Response



$$\Pr(Y = 1) = \Pr(Z \leq -0.8) = \frac{1}{1 + e^{0.8}} = 0.31$$

Logit and Probit Models for Binary Response



Logit and Probit Models for Binary Response

- **<u>Interpretation of coefficients in Logit and Probit models</u>**

$$\boxed{\frac{\partial P(y=1|\mathbf{x})}{\partial x_j}} = g(\mathbf{x}\boldsymbol{\beta})\beta_j \quad \underline{\text{where}} \quad g(z) = \partial G(z)/\partial z > 0$$

How does the probability for y=1 change if explanatory variable x_j changes by one unit?

<u>Discrete explanatory variables:</u>

$$G\left[\beta_0 + \beta_1 x_1 + \cdots + \beta_k \boxed{(c_k + 1)}\right] - G\left[\beta_0 + \beta_1 x_1 + \cdots + \beta_k \boxed{c_k}\right]$$

For example, explanatory variable x_k increases by one unit.

- **Partial effects are nonlinear and depend on the level of x !**

Logit and Probit Models: Estimation

- So far, we used OLS to estimate models
- Logit and Probit models are nonlinear in parameters:
  - Hence, in this case the OLS cannot be used
- The method used to estimate Logit and Probit models is Maximum Likelihood Estimation (MLE)
- The MLE are the values of parameters that best describe the full distribution of the data
  - The **likelihood function** is the joint probability distribution of the data, treated as a function of the unknown coefficients
  - The **MLE** are the values of the coefficients that maximize the likelihood function
  - MLE's are the parameter values "most likely" to have produced the data

Logit and Probit Models for Binary Response

- **Goodness-of-fit measures for Logit and Probit models**
  - Percent correctly predicted

$$\tilde{y}_i = \begin{cases} 1 & \text{if } G(\mathbf{x}_i\hat{\beta}) > .5 \\ 0 & \text{otherwise} \end{cases}$$

Individual i's outcome is predicted as one if the probability for this event is larger than .5, then percentage of correctly predicted y=1 and y=0 is counted

  - Pseudo R-squared

$$\tilde{R}^2 = 1 - \log L_0 / \log L_{ur}$$

Compare maximized log-likelihood of the model with that of a model that only contains a constant (and no explanatory variables)

  - Correlation based measures

$$Corr(y_i, \tilde{y}_i), \quad Corr(y_i, G(\mathbf{x}_i\hat{\beta}))$$

Look at correlation (or squared correlation) between predictions or predicted prob. and true values

Logit and Probit Models for Binary Response

- **Reporting partial effects of explanatory variables**
  - The difficulty is that partial effects are not constant but depend on x
  - Partial effects at the average:

$$\widehat{PEA}_j = g(\bar{\boldsymbol{x}}\hat{\boldsymbol{\beta}})\hat{\beta}_j$$

The partial effect of explanatory variable x_j is considered for an „average individual" (this is problematic in the case of explanatory variables such as gender)

  - Average partial effects:

$$\widehat{APE}_j = n^{-1}\sum_{i=1}^{n} g(\mathbf{x}_i\hat{\boldsymbol{\beta}})\hat{\beta}_j$$

The partial effect of explanatory variable x_j is computed for each individual in the sample and then averaged across all sample members (makes more sense)

Logit and Probit Models for Binary Response

- **Example: Married women's labor force participation**

TABLE 17.1 LPM, Logit, and Probit Estimates of Labor Force Participation			
Dependent Variable: *inlf*			
Independent Variables	**LPM (OLS)**	**Logit (MLE)**	**Probit (MLE)**
nwifeinc	−.0034 (.0015)	−.021 (.008)	−.012 (.005)
educ	.038 (.007)	.221 (.043)	.131 (.025)
exper	.039 (.006)	.206 (.032)	.123 (.019)
exper^2	−.00060 (.00018)	−.0032 (.0010)	−.0019 (.0006)
age	−.016 (.002)	−.088 (.015)	−.053 (.008)
kidslt6	−.262 (.032)	−1.443 (.204)	−.868 (.119)
kidsge6	.013 (.013)	.060 (.075)	.036 (.043)
constant	.586 (.151)	.425 (.860)	.270 (.509)
Percentage correctly predicted	73.4	73.6	73.4
Log-likelihood value	—	−401.77	−401.30
Pseudo *R*-squared	.264	.220	.221

The coefficients are not comparable across models

Often, Logit estimated coefficients are 1.6 times Probit estimated because $g_{Logit}(0)/g_{Probit}(0) \approx 1/1.6$

<u>The biggest difference between the LPM and Logit/Probit is that partial effects are nonconstant in Logit/Probit</u>:

$$\widehat{P}(working|\bar{x}, kidslt6 = 0) = .707$$

$$\widehat{P}(working|\bar{x}, kidslt6 = 1) = .373$$

$$\widehat{P}(working|\bar{x}, kidslt6 = 2) = .117$$

(Larger decrease in probability for the first child)