Qualitative and Limited Dependent Variable Models

Ketevani Kapanadze

Brno, 2020

A Single Dummy Independent Variable

- **Qualitative Information**

  - Examples: gender, race, industry, region, rating grade, …
  - A way to incorporate qualitative information is to use dummy variables
  - They may appear as the dependent or as independent variables

Dummy Variables

- Dummy variable - takes on the values of 0 or 1, depending on a qualitative attribute;
- Examples of dummy variables are:

$$Male = \begin{cases} 1 & \text{if the person is male} \\ 0 & \text{if the person is female} \end{cases}$$

$$Weekend = \begin{cases} 1 & \text{if the day is on weekend} \\ 0 & \text{if the day is a work day} \end{cases}$$

$$NewStadium = \begin{cases} 1 & \text{if the team plays on new stadium} \\ 0 & \text{if the team plays on old stadium} \end{cases}$$

Intercept Dummy

- Dummy variable included in a regression alone (not interacted with other variables) is an intercept dummy;
- It changes the intercept for the subset of data defined by a dummy variable condition:

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 X_\mathrm{i} + u_i$$

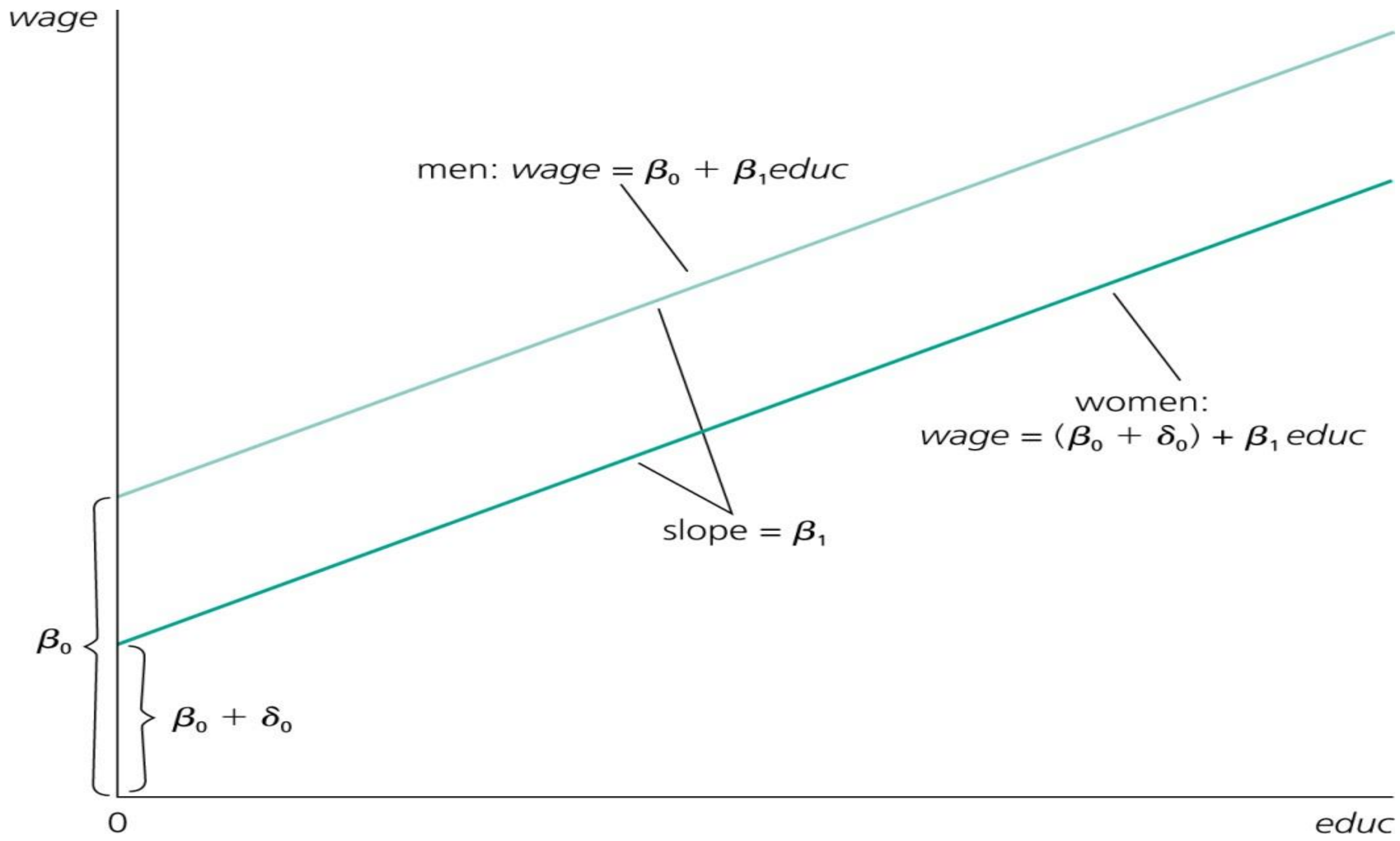
where

$$D_i = \begin{cases} 1 & \text{if the } i\text{-th observation meets a particular condition} \\ 0 & \text{otherwise} \end{cases}$$

- We have: (on the board)

Intercept Dummy

- **Graphical Illustration**



Example

- Estimating the determinant of wages:

$$wage_i = -3.89 + \underset{(0.270)}{2.156}\ M_i + \underset{(0.051)}{0.603}\ educ_{\mathrm{i}} + \underset{(0.064)}{0.010}\ exper_{\mathrm{i}}$$

- Interpretation of the dummy variable M: men earn on average $2.156 per hour more than women, ceteris paribus

A Single Dummy Independent Variable

- **Estimated wage equation with intercept shift**

$$\widehat{wage} = \underset{(.72)}{-\ 1.57} \ \underset{(.26)}{-\ 1.81}\ female + \underset{(.049)}{.572}\ educ$$

$$+ \underset{(.012)}{.025}\ exper + \underset{(.021)}{.141}\ tenure$$

Holding education, experience, and tenure fixed, women earn 1.81$ less per hour than men

$$n = 526, R^2 = .364$$

- **Does that mean that women are discriminated against?**
  - Not necessarily. Being female may be correlated with other productivity characteristics that have not been controlled for.

A Single Dummy Independent Variable

- **Comparing means of subpopulations described by dummies**

$$\widehat{wage} = \underset{(.21)}{7.10} - \underset{(.26)}{2.51}\, female$$

Not holding other factors constant, women earn 2.51$ per hour less than men, i.e. the difference between the mean wage of men and that of women is 2.51$.

$$n = 526,\, R^2 = .116$$

- **Discussion**
  - It can easily be tested whether difference in means is significant
  - The wage difference between men and women is larger if no other things are controlled for; i.e. part of the difference is due to differences in education, experience and tenure between men and women

Slope Dummy

- If a dummy variable is interacted with another variable (x), it is a slope dummy;
- It changes the relationship between x and y for a subset of data defined by a dummy variable condition:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 (X_i * D_i) + u_i$$

where

$$D_i = \begin{cases} 1 & \text{if the } i\text{-th observation meets a particular condition} \\ 0 & \text{otherwise} \end{cases}$$

- We have: (on the board)

Example

- Estimating the determinant of wages:

$$wage_i = -2.620 + 0.450\ educ_{\mathrm{i}} + 0.17\ M_i * educ_{\mathrm{i}} + 0.010\ exper_{\mathrm{i}}$$
$$\qquad\qquad(0.054)\qquad\quad(0.021)\qquad\qquad\qquad(0.065)$$

- Interpretation: men gain on average 17 cents per hour more than women for each additional year of education, ceteris paribus

Multiple categories

- What if a variable defines three or more qualitative attributes?
- Example: level of education - elementary school, high school, and college;
- Define and use a set of dummy variables:

$$H = \begin{cases} 1 & \text{if high school} \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad C = \begin{cases} 1 & \text{if college} \\ 0 & \text{otherwise} \end{cases}$$

- Should we include also a third dummy in the regression, which is equal to 1 for people with elementary education?
  - No, unless we exclude the intercept!
  - Using full set of dummies leads to perfect multicollinearity (dummy variable trap)

A Single Dummy Independent Variable

- **Dummy variable trap**

This model cannot be estimated (perfect collinearity)

$$wage = \beta_0 + \gamma_0 \boxed{male} + \delta_0 \boxed{female} + \beta_1 educ + u$$

<u>When using dummy variables, one category always has to be omitted:</u>

$$wage = \beta_0 + \delta_0 \boxed{female} + \beta_1 educ + u$$

The base category are men

$$wage = \beta_0 + \gamma_0 \boxed{male} + \beta_1 educ + u$$

The base category are women

<u>Alternatively, one could omit the intercept:</u>

<u>Disadvantages:</u>
1) More difficult to test for differences between the parameters
2) R-squared formula only valid if regression contains intercept

$$wage = \gamma_0 \boxed{male} + \delta_0 \boxed{female} + \beta_1 educ + u$$

Interactions Involving Dummy Variables

- **Allowing for different slopes**

Interaction term

$$\log(wage) = \beta_0 + \delta_0 female + \beta_1 educ + \delta_1 \boxed{female \cdot educ} + u$$

β_0 = <u>intercept men</u> β_1 = <u>slope men</u>

$\beta_0 + \delta_0$ = <u>intercept women</u> $\beta_1 + \delta_1$ = <u>slope women</u>

- **Interesting hypotheses**

$\boxed{H_0 : \delta_1 = 0}$ $\boxed{H_0 : \delta_0 = 0, \delta_1 = 0}$

The <u>return to education</u> is the same for men and women

The <u>whole wage equation</u> is the same for men and women

A Binary Dependent Variable:
The Linear Probability Model

A Binary Dependent Variable: The Linear Probability Model

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

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

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

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

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

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

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

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

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)

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

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

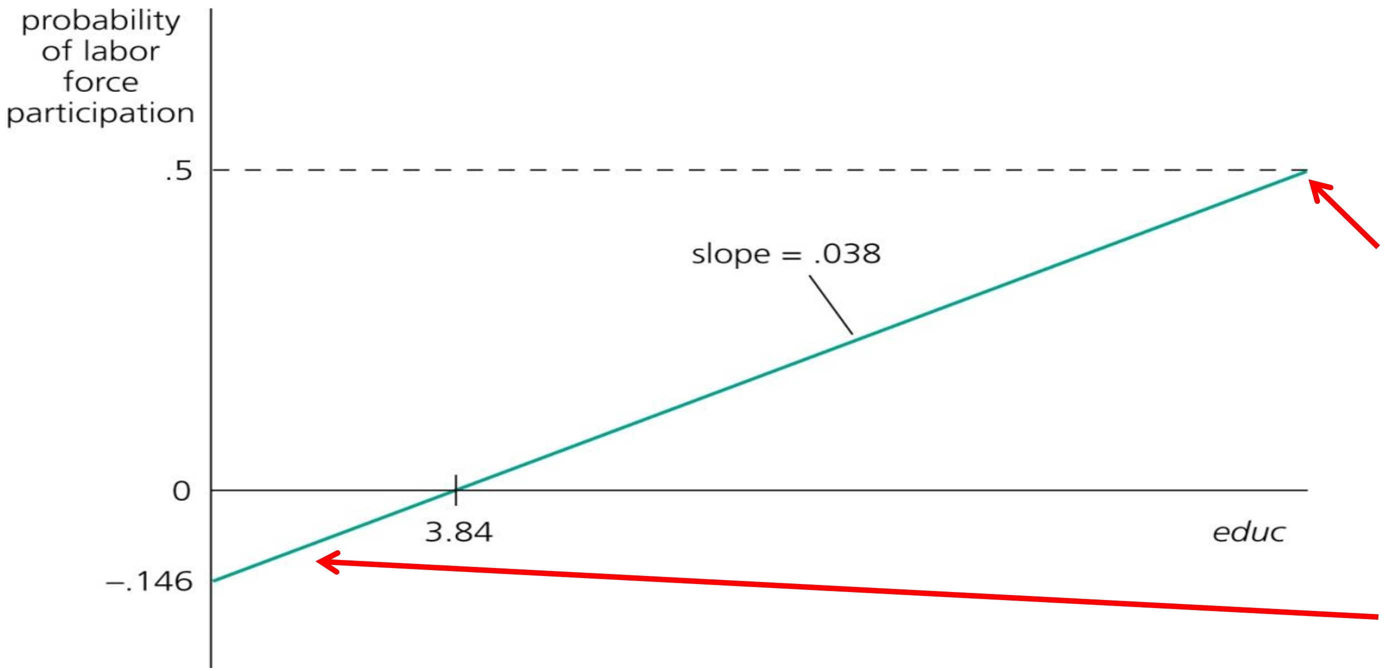
$$+ \underset{(.0132)}{.0130}\ kidsge6,\ n = 753, R^2 = .264$$

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.

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>

$$Var(y|\mathbf{x}) = P(y=1|\mathbf{x})\,[1 - P(y=1|\mathbf{x})]$$

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

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?

- **Marginal effects are nonlinear and depend on the level of X !**

Marginal effects for the logit model

$$\partial p/\partial \mathrm{x}_j = \Lambda(\mathbf{x}'\beta)[1 - \Lambda(\mathbf{x}'\beta)]\beta_j = \frac{e^{\mathbf{x}'\beta}}{\left(1 + e^{\mathbf{x}'\beta}\right)^2}\beta_j$$

Marginal effects for the probit model

$$\partial p/\partial \mathrm{x}_j = \phi(\mathbf{x}'\beta)\beta_j$$

Estimating marginal effects

Marginal effects at the mean

- The marginal effects are estimated for the average person in the sample $\bar{\mathbf{x}}$.

$$\partial p/\partial \mathrm{x}_j = \mathrm{F}'(\bar{\mathbf{x}}'\beta)\beta_j$$

- Most papers report marginal effects at the mean.
- A problem is that there may not be such a person in the sample.

Average marginal effects

- The marginal effects are estimated as the average of the individual marginal effects.

$$\partial p/\partial \mathrm{x}_j = \frac{\sum \mathrm{F}'(\mathbf{x}'\beta)}{\mathrm{n}}\,\beta_j$$

- This is a better approach of estimating marginal effects, but papers still use the previous approach.
- In practice, the two ways to estimate marginal effects produce almost identical results most of the time.

Partial effects for discrete variables

- Predict the probabilities for the two discrete values of a variable and take the difference:

$$F(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2(k+1)) - F(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2(k))$$

Interpretation of marginal effects

- An increase in x increases (decreases) the probability that y=1 by the marginal effect expressed as a percent.
- For dummy independent variables, the marginal effect is expressed in comparison to the base category (x=0).
- For continuous independent variables, the marginal effect is expressed for a one-unit change in x.
- We interpret both the sign and the magnitude of the marginal effects.
- The probit and logit models produce almost identical marginal effects.

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

$$\text{R-squared} = 1 - L_{ur}/L_r$$

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

Discussion about binary outcome models

Choice between the logit and probit model

- The choice depends on the data generating process, which is unknown.
- The models produce almost identical results (different coefficients but similar marginal effects).
- The choice is up to you.

Coding of the dependent variable

If we reverse the categories 0 and 1, the signs of the coefficients are reversed (positive become negative and vice versa) but the magnitudes are the same.

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 $g_{Logit}(0)/g_{Probit}(0) \approx 1/1.6$

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

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

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

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

(Larger decrease in probability for the first child)

Next Class – 10.04
In the Zoom at 1pm

Regression Analysis with Time Series Data