# Probit and Logit Models

Ani Katchova

# Probit and Logit Models Overview

- Examples of probit and logit models
- Binary dependent variable
- Linear regression model, probit, and logit models functional forms and properties
- Model coefficients and interpretations
- Marginal effects (and odds ratios) and interpretations
- Goodness of fit statistics (percent correctly predicted and pseudo R-squared)
- Choice between probit and logit
- Economic models that lead to use of probit and logit models

# Probit and Logit Models (Binary Outcome Models)

## Binary outcome examples

- Consumer economics: whether a consumer makes a purchase or not.
- Labor economics: whether an individual participates in the labor market or not.
- Agricultural economics: whether or not a farmer adopts or uses organic practices, marketing/production contracts, etc.

## Binary outcome dependent variable

- The decision/choice is whether or not to have, do, use, or adopt.
- The dependent variable is a binary response
- It takes on two values: 0 and 1.

$$y = \begin{cases} 0 \; if \; no \\ 1 \; if \; yes \end{cases}$$

# Binary outcome models

- Binary outcome models are among the most used in applied economics.
- A look at the OLS model: $y = \mathbf{x}'\beta + e$
- Binary outcome models estimate the probability that $y$=1 as a function of the independent variables.

$$p = \mathrm{pr}[y = 1|\mathbf{x}] = \mathrm{F}(\mathbf{x}'\beta)$$

- There are three different models depending on the functional form of $\mathrm{F}(\mathbf{x}'\beta)$.

## Regression model (linear probability model)

- In the linear probability model, $\mathrm{F}(\mathrm{x}'\beta) = \mathrm{x}'\beta$

$$p = \mathrm{pr}[y = 1|\mathrm{x}] = \mathrm{x}'\beta$$

- A problem with the regression model is that the predicted probabilities will not be limited between 0 and 1.
- We do not use the regression model with binary outcome data.

## Logit model

- For the logit model, $\mathrm{F}(\mathrm{x}'\beta)$ is the cdf of the logistic distribution.

$$\mathrm{F}(\mathbf{x}'\beta) = \Lambda(\mathbf{x}'\beta) = \frac{e^{\mathbf{x}'\beta}}{1 + e^{\mathbf{x}'\beta}} = \frac{\exp{(\mathbf{x}'\beta)}}{1 + \exp{(\mathbf{x}'\beta)}}$$

- The predicted probabilities are limited between 0 and 1.

## Probit model

- For the probit model, $\mathrm{F}(\mathbf{x}'\beta)$ is the cdf of the standard normal distribution.

$$\mathrm{F}(\mathbf{x}'\beta) = \Phi(\mathbf{x}'\beta) = \int_{-\infty}^{\mathbf{x}'\beta} \phi(z)dz$$

- The predicted probabilities are limited between 0 and 1.

## Model coefficients

- Probit and logit models are estimated using the maximum likelihood method.

### Interpretation of coefficients

- An increase in x increases/decreases the likelihood that y=1 (makes that outcome more/less likely). In other words, an increase in x makes the outcome of 1 more or less likely.
- We interpret the *sign* of the coefficient but not the *magnitude*. The magnitude cannot be interpreted using the coefficient because different models have different scales of coefficients.

### Comparison of coefficients

- Coefficients differ among models because of the functional form of the *F* function.

$$\beta_{logit} \simeq 4\beta_{OLS}$$

$$\beta_{probit} \simeq 2.5\beta_{OLS}$$

$$\beta_{logit} \simeq 1.6\beta_{probit}$$

- We should not compare the magnitude of the coefficients among different models.

## Marginal effects

- When estimating probit and logit models, it is common to report the marginal effects after reporting the coefficients.
- The marginal effects reflect the change in the probability of $y$=1 given a 1 unit change in an independent variable x.

### Marginal effects for the regression model

- For the OLS regression model, the marginal effects are the coefficients and they do not depend on x.

$$\partial p / \partial \mathrm{x}_j = \beta_j$$

- The index $j$ refers to the $j^{\text{th}}$ independent variable.
- [When we use the index $i$, it refers to the $i^{\text{th}}$ observation.]

**Marginal effects for the binary models (probit and logit)**

- For the logit and probit models, the marginal effects are calculated as:

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

- The marginal effects depend on x, so we need to estimate the marginal effects at a specific value of x (typically the means).
- Coefficients and marginal effects have the same signs because $\mathrm{F}'(\mathbf{x}'\beta) > 0$.

**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.

*Odds ratio/relative risk for the logit model*

- The odds ratio or relative risk is p/(1-p) and measures the probability that y=1 relative to the probability that y=0.

$$p = \frac{\exp(\mathbf{x}'\beta)}{1 + \exp(\mathbf{x}'\beta)}$$

$$\frac{p}{1-p} = \exp(\mathbf{x}'\beta)$$

$$ln \frac{p}{1-p} = \mathbf{x}'\beta$$

- An odds ratio of 2 means that the outcome y=1 is twice as likely as the outcome of y=0.
- Odds ratios are estimated with the logistic model.
- Reporting marginal effects instead of odds ratios is more popular in economics.

## Predicted probabilities and goodness of fit measures

- After estimating the models, we can predict the probability that y=1 for each observation.

$$\hat{p} = \mathrm{pr}[y = 1|\mathbf{x}] = \mathrm{F}\left(\mathbf{x}'\hat{\beta}\right)$$

- For the regression model, the predicted probabilities are *not* limited between 0 and 1.
- For the logit and probit models, the predicted probabilities are limited between 0 and 1.
- The predicted probability indicate the likelihood of y=1. If the predicted probability is greater than 0.5 we can predict that y=1, otherwise y=0.

**Goodness of fit measures**

*Percent correctly predicted values*

- If the predicted probability is greater than 0.5 we can predict that y=1, otherwise y=0.
- We can create the following table:

| | Actual y=1 | Actual y=0 |
|---|---|---|
| Predicted yhat=1 | True | False |
| Predicted yhat=0 | False | True |

- We have four cases of 0/1: two of them are correct predictions and two of them are wrong predictions.
- The percent correctly predicted values are the proportion of true predictions to total predictions.

*Pseudo R-squared (McFadden R-squared)*

- The pseudo R-square is calculated as:

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

- It compares the unrestricted log-likelihood $L_{ur}$ for the model we are estimating and the restricted log-likelihood $L_r$ with only an intercept.
- If the independent variables have no explanatory power, the restricted model will be the same as unrestricted model and R-squared will be 0.

## 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.

*Latent variable models*

- A latent variable is a variable that is incompletely observed $y^*$. Latent variables can be introduced into binary outcome models in two ways: index functions and random utility models.

*Index function models*

- The latent variable is an index of the unobserved propensity for the event to occur.
- Index models are used in two step models, which will be covered later in class.
  - Example: We cannot observe how much people want to work, only if they work or not.

$$y = \begin{cases} 1 \; if \; y* > 0 \\ 0 \; if \; y* \leq 0 \end{cases}$$

*Random utility models*

- The latent variable is the difference in utilities if the event occurs or does not occur.
- They are often a result of individual choice.
  - Example: a consumer chooses one product or another depending on which utility is higher.

$$U_0 = V_0 + e_0$$

$$U_1 = V_1 + e_1$$

$$p(y = 1) = p(U_1 > U_0)$$