

Models with Binary or Categorical Dependent Variables

Erick F Molina

Summer, 2024

ITAM

Linear Probability Model

Linear Probability Model

To model the relationship between formality and education, we start with a linear regression model with a single regressor:

1. What does $\hat{Y}$ mean when Y is binary?
2. What does the fitted line $\hat{Y}$ mean when Y is binary?
3. What does the predicted value $\hat{Y}$ mean when Y is binary?

Interpreting the linear probability model

In the linear probability model:

- The predicted value $\hat{Y}$ is interpreted as the predicted probability that $Y = 1$.
- β_1 is the change in the predicted probability associated with a one-unit change in X .

Thus, when Y is binary, the linear regression model:

- is called the linear probability model because:
 - the predicted value $\hat{Y}$ is the predicted probability that $Y = 1$, given X ;
 - β_1 is the difference in the probability that $Y = 1$ associated with a one-unit difference in X .

The linear probability model specifies $P(Y = 1|X)$ as a linear function of X .

- The model is simple to estimate and interpret.
- Hypothesis testing is identical to multiple regression (note that you should use heteroskedasticity-robust standard errors).
- The model can be estimated in a panel-data framework: for example, fixed-effects estimation can be used.

Limitations of the linear probability model

Two limitations of the linear probability model:

1. The change in the predicted probability for a given change in X is the same for all values of X .
2. Predicted probabilities from a linear probability model can be less than 0 or greater than 1.

We can address both issues using a nonlinear probability model.

The Probit Model

- The name comes from **Probability** and **Unit**.
- It uses a cumulative distribution function to approximate $P(Y_i = 1|X_i)$.
- Predictions always lie in the range $[0, 1]$.
- It is estimated by maximum likelihood.

Suppose we start from the following model:

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

- We are interested in estimating the parameters β .
- We replace Y_i with a *latent* variable, denoted Y_i^* . We define $Y_i = 1\{Y_i^* > 0\}$, an indicator function such that Y_i equals one if Y_i^* is positive.
- We now assume the errors follow a standard normal distribution.

Our model becomes:

$$Y_i^* = X_i' \beta + U_i$$

Therefore, the probability is computed as:

$$P(Y_i = 1|X_i) = P(U_i > -X_i' \beta) = P(U_i < X_i' \beta) = \Phi(X_i' \beta)$$

To estimate by maximum likelihood, we use a Bernoulli distribution (since Y_i is dichotomous). Thus, the probability mass function is $f(y) = p^y(1 - p)^{1-y}$.

Given this distribution, and applying the log transformation, we obtain the *log-likelihood* function:

$$L(\beta) = \sum_{i=1}^N [Y_i \log(\Phi(X_i' \beta)) + (1 - Y_i) \log(1 - \Phi(X_i' \beta))]$$

- The coefficients do not have an intuitive interpretation on their own.
- They reflect changes in the latent variable associated with marginal changes in the regressors.
- To obtain an interpretation similar to the linear probability model, we can use a transformation based on the PDF:

$$\frac{\partial \Pr(Y_i = 1|X_i)}{\partial X_{ki}} = \frac{\partial \Phi(X_i' \beta)}{\partial X_{ki}} = \phi(X_i' \beta) \beta_k$$

- We can then average this effect across all observations (typically reported in percentage points). In that case, the interpretation of β_k is: (*ceteris paribus*) a marginal change in X_k implies a change in the **odds** of $Y_i = 1$ relative to $Y_i = 0$ by a factor related to the marginal effect.¹

¹If you want an odds-ratio interpretation directly, that is more standard in the Logit model.

The Logit Model

- The Logit model is very similar to the Probit model.
- The Logit model assumes the errors follow a logistic distribution:

$$P(Y_i = 1|X_i) = \Lambda(X_i'\beta) = \frac{\exp(X_i'\beta)}{1 + \exp(X_i'\beta)}$$

- In the Logit case, coefficients have a natural interpretation in terms of log-odds.

Compute the odds of $Y_i = 1$ relative to $Y_i = 0$:

$$\frac{P(Y_i = 1|X_i)}{P(Y_i = 0|X_i)} = \frac{P(Y_i = 1|X_i)}{1 - P(Y_i = 1|X_i)} = \exp(X_i'\beta)$$

Taking logs:

$$\log \left(\frac{Pr(Y_i = 1|X_i)}{Pr(Y_i = 0|X_i)} \right) = X_i'\beta$$

It is common to report a transformation of the coefficients known as the *odds ratio*, rather than reporting the raw coefficients. Define:

$$\frac{Pr(Y_i = 1|X_i)}{Pr(Y_i = 0|X_i)} = \left(\frac{p}{1-p} \middle| X_i \right)$$

If we want the effect of increasing one explanatory variable by one unit, we have:

$$\log \left(\frac{p}{1-p} \middle| X_{1i} = 50 \right) = \beta_0 + \beta_1(50) + \beta_2 X_{2i} + \cdots + \beta_K X_{Ki}$$

$$\log \left(\frac{p}{1-p} \middle| X_{1i} = 49 \right) = \beta_0 + \beta_1(49) + \beta_2 X_{2i} + \cdots + \beta_K X_{Ki}$$

Now, consider the difference between these two values and then exponentiate:

$$\begin{aligned} & \exp \left[\log \left(\frac{p}{1-p} \middle| X_{1i} = 50 \right) - \log \left(\frac{p}{1-p} \middle| X_{1i} = 49 \right) \right] \\ &= \frac{\exp \left[\log \left(\frac{p}{1-p} \middle| X_{1i} = 50 \right) \right]}{\exp \left[\log \left(\frac{p}{1-p} \middle| X_{1i} = 49 \right) \right]} \end{aligned}$$

Hence:

$$\text{Odds Ratio} = \exp(\beta_1)$$

Multinomial Choice Models

Suppose now that Y is not a dummy variable but a categorical variable. For example: which party to vote for (PRI/PAN/PRD/Morena), or which console to buy (PlayStation/XBox/Nintendo Switch).

- $Y \in \{0, 1, \dots, J\}$ with no natural ordering across categories.
- Each value represents an alternative that can be chosen.

Now we want to generate appropriate predictions for the alternatives in the categorical outcome.

As in the binary case, $P(Y_i = j|X) = P_j(X_i\beta)$.

Multinomial Logit

Using these probabilities, derive the following maximum likelihood function:

$$\mathcal{L} = \sum_{i=1}^N \sum_{j=0}^J \mathbf{1}\{Y_i = j\} \log P_j(X_{i,j}, \beta)$$

This is a simple model that assumes we only have individual-level covariates.

Extending the Logit model, we obtain:

$$P(Y_i = j|X_i) = \frac{\exp(X_i' \beta_j)}{1 + \sum_{l=0}^J \exp(X_i' \beta_l)}$$

$$P(Y_i = 0|X_i) = \frac{1}{1 + \sum_{l=0}^J \exp(X_i' \beta_l)}$$

Multinomial Logit

$$\log \left(\frac{P(Y_i = j|X_i)}{P(Y_i = 0|X_i)} \right) = X_i' \beta_j$$

which allows us to interpret the β 's.

This model can be seen as a special case of the conditional Logit model, which includes explanatory variables that vary across alternatives.

$$P(Y_i = j|X_{i,0}, X_{i,1}, \dots, X_{i,J}) = \frac{\exp(X_{i,j}'\beta_j)}{\sum_{l=0}^J \exp(X_{i,l}'\beta_l)}$$

Link to utility maximization:

$$U_{i,j} = X_{i,j}'\beta + \epsilon_{i,j}$$

We assume an individual chooses j if it yields the highest utility:

$$Y_i = j \quad \text{if} \quad U_{i,j} \geq U_{i,l} \quad \forall l \neq j$$

Multinomial Logit

Problem: Independence of Irrelevant Alternatives (IIA)

- $Y = 0$ denotes the unemployed
- $Y = 1$ denotes those with part-time employment
- $Y = 2$ denotes those working full-time

Example

- Suppose the government introduces a health insurance subsidy for part-time employees.
- This will induce some individuals to switch from $Y = 2$ to $Y = 1$ (part-time).
- However, $Y = 0$ should not be affected by this change:

$P(Y = 0) \rightarrow$ should remain unchanged

But...

- In practice, introducing this subsidy *does* affect the probabilities of all available options.

Solution: relax the IIA assumption. We can use a Probit model.

Multinomial Probit Model

- Let $U_i = X_i\beta + \epsilon_i$ denote the utility level associated with choice i .
- Parameters $\beta = (\beta_0, \beta_1, \dots, \beta_k)$.
- $P(Y_i = j|X) = P(U_{ij} > U_{ik} \text{ for all } k \neq j)$.
- There is no closed-form solution, but given a distribution for the errors ϵ , the choice probabilities can be estimated.