Applied Econometrics II

Hanh Dao

December 23, 2025

Models

Structural model

Definition 1. Implicit model: each observation satisfies:

$$g(y_i, z_i, u_i|\theta) = 0$$

- $g(\cdot)$: a known function implied by economic theory.
- y_i: endogenous variables, determined within the model.
- z_i: exogenous variables, taken as given from outside the model.
- u_i: random disturbances (shocks) capturing unobserved factors.
- θ: structural parameters governing preferences, technology, or institutions.

This form emphasizes economic restrictions.

Formula 1 (Reduced form)**.** Assume that for every (z_i, u_i) there exists a unique solution for y_i. Then equation (2.1) can be solved for y_i as

$$y_i = f(z_i, u_i \mid \pi), \tag{2.2}$$

which is called the *reduced form.*

Theory $\Rightarrow$ $g(y, z, u \mid \theta) = 0$ Solve $\Rightarrow$ y=f(z,u $|\pi$) Predict $\Rightarrow$ E[y |z] Infer $\Rightarrow$ θ (if the structure is identified).

Binary outcome model

We study binary outcomes:

$$y_i \in \{0, 1\},$$

Logit model

$$\text{odds} := \frac{p_i}{1 - p_i}$$

$$p_i = \frac{\exp(\beta_1 + \beta_2 x_i)}{1 + \exp(\beta_1 + \beta_2 x_i)}.$$

- $0 < p_i < 1$, smooth S-shape.
- If $\beta_2 < 0$, then p_i decreases as x_i increases.
- Marginal effect:

$$\frac{dp_i}{dx_i} = \frac{\exp(\beta_1 + \beta_2 x_i)}{\left(1 + \exp(\beta_1 + \beta_2 x_i)\right)^2}\,\beta_2.$$

The effect depends on the value of x_i and is largest near the middle of the distribution.

Coefficient interpretation

- When $\beta_2 = 0$, X and Y are independent
- When $\beta_2 > 0$, the probability that $Y = 1$ increases with X.
- When $\beta_2 < 0$, the probability that $Y = 1$ decreases with X.

Changes in probability is NOT the same as changes in odds.

Intercept interpretation

You can turn a starting point into a probability. You cannot turn a slope alone into a probability.

$$p_i = \frac{\exp(\beta_1 + \beta_2(0))}{1 + \exp(\beta_1 + \beta_2(0))} = \frac{\exp(\beta_1)}{1 + \exp(\beta_1)}$$

Probit model

$$p_i = \Phi(\beta_1 + \beta_2 x_i),$$

- Marginal effect:

$$\frac{dp_i}{dx_i} = \phi(\beta_1 + \beta_2 x_i)\,\beta_2,$$

- $\Phi(\cdot)$ is the standard normal CDF, $\phi(\cdot)$ is the standard normal density.
  Logit uses logistic CDF
  Probit uses normal CDF

Latent variables

Y^* represents hidden driver of the binary outcomes: an internal, continuous index (utility, net benefit, propensity) that cannot be directly measured in real data.

$$Y^* = x'\beta + \varepsilon, \qquad Y = \begin{cases} 1, & \text{if } Y^* > 0, \\ 0, & \text{if } Y^* \le 0. \end{cases}$$

$$\Pr(Y = 1 \mid x) = Pr(\epsilon > -x'\beta) = Pr(\epsilon \le x'\beta) = F(x'\beta).$$

1. Probit assumption:

$$\epsilon \sim N(0,1) \qquad F = \Phi(x'\beta) = Pr(Y = 1|x) \quad \text{(standard normal CDF)}$$

2. Logit assumption:

$$\varepsilon \sim \text{Logistic}(0,1). \qquad F = \Lambda(x'\beta) = \frac{1}{1 + e^{-x'\beta}} \quad \text{(logit, logistic CDF)}$$

Marginal effect

$$x_i'\beta = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_j x_{ij}.$$

- Each x_{ij} is one regressor of observation i
- Each β_j is the weight that the model assigns to that regressor

Marginal probability

Effect of regressor x_{ij} on probability:

$$\frac{\partial \Pr(y_i = 1 \mid x_i)}{\partial x_{ij}} = F'(x_i'\beta)\,\beta_j.$$

Model-specific $F'(\cdot)$
1. Logit: $F'(z) = \Lambda(z)(1 - \Lambda(z))$
2. Probit: $F'(z) = \phi(z)$

Average marginal effect (AME):

$$\text{AME}_j := \frac{1}{N}\sum_{i=1}^{N} F'(x_i'\beta)\,\beta_j.$$

$$\text{At mean regressors: } F'(\bar{x}'\beta)\,\beta_j,$$
$$\text{At mean probability } \bar{y} : \quad F'\big(F^{-1}(\bar{y})\big)\,\beta_j.$$

For logit specifically,

$$\text{ME at } \bar{y} = \bar{y}(1 - \bar{y})\,\beta_j.$$

- Sign of β_j = sign of marginal effect
- Ratios: β_j/β_k = ratio of marginal effects (single-index property)

Remark. The marginal effects differ with the point of evaluation x_i, as for any nonlinear model, and differ with different choices of $F(\,\cdot\,)$.

Multinomial models

Family of models

Multinominal models are a family of models used when the outcome has more than two choices.

Utility

$$U_{ij} = \underbrace{X_i^\top \beta}_{\text{things about the person}} + \underbrace{Z_j^\top \gamma}_{\text{things about option } j} + \varepsilon_{ij}$$

Conditional Logit Model

We study choice among alternatives. "Why does a person choose this alternative instead of the others?"

Regressors

2 kinds of information (regressors)
1. Alternative-invariant regressors
2. Alternative-specific regressors

Alternative-specific regressors

$$x_{ik} \neq x_{ij} \qquad \text{for } k \neq j$$

Choice of Fishing Model: 4 mutually exclusive alternatives:
1. Beach
2. Pier
3. Private boat
4. Charter boat

$$X_{ij} = \begin{pmatrix} P_{ij} \\ C_{ij} \end{pmatrix}, \qquad \beta = \begin{pmatrix} \beta_P \\ \beta_C \end{pmatrix}$$

$$U_{ij} = \beta^\top X_{ij} + \varepsilon_{ij} \qquad \text{(Utility score by option } j)$$

$$p_{ij} = \frac{\exp(\beta_P P_{ij} + \beta_C C_{ij})}{\sum_{k=1}^{4} \exp(\beta_P P_{ik} + \beta_C C_{ik})}.$$

Marginal effect

A regressor change affects all probabilities, not just one.

$$\frac{\partial p_{ij}}{\partial x_{ikr}}$$

$$\frac{\partial p_{ij}}{\partial x_{ikr}} = \begin{cases} p_{ij}\big(1 - p_{ij}\big)\beta_r, & \text{if } j = k \quad \text{(own effect)}, \\ -\,p_{ij} p_{ik} \beta_r, & \text{if } j \neq k \quad \text{(cross effect)}. \end{cases}$$

Average Marginal effect

$$\frac{1}{N} \sum_{i=1}^{N} \frac{\partial p_{ij}}{\partial x_{ikr}}$$

Suppose individual i faces the following four alternatives:

Option k	Mode	P_{ik}	C_{ik}
1	Beach	10	3
2	Pier	12	4
3	Charter	30	8
4	Private	20	5

Consider the regressor which is the price of the charter boat.

$$x_{i,3,P} = 30,$$

An increase in $x_{i,3,P}$ affects the choice probabilities as follows:

$$\frac{\partial p_{i3}}{\partial x_{i,3,P}} \quad \text{(own effect)},$$

$$\frac{\partial p_{i1}}{\partial x_{i,3,P}}, \quad \frac{\partial p_{i2}}{\partial x_{i,3,P}}, \quad \frac{\partial p_{i4}}{\partial x_{i,3,P}} \quad \text{(cross effects)}.$$

Multinominal Logit

Any variable that is identical across all alternatives for an individual (such as income) cancels out of the choice probabilities.
1. MNL Utility (different observations have different incomes):

$$U_{ij} = \alpha_j + \beta_{Ij} I_i + \varepsilon_{ij}, \qquad I_i \neq I_{i'} \text{ for } i \neq i'.$$

2. Choice probability:

$$p_{ij} = \frac{\exp(\alpha_j + \beta_{Ij} I_i)}{\sum_{k=1}^{4} \exp(\alpha_k + \beta_{Ik} I_i)}.$$

Marginal effect

$$\frac{1}{N} \sum_{i=1}^{N} \frac{\partial p_{ij}}{\partial I_i}$$

$$\frac{1}{f} N \sum_{i=1}^{N} \hat{p}_{ij} \left(\hat{\beta}_j - \hat{\bar{\beta}}_i \right)$$

$$\hat{\bar{\beta}}_i = \sum_{l=1}^{m} p_{il} \hat{\beta}_l$$