Econometrics II

Lecture 4 - Binary Choice Models

Lasha Chochua

2025

ISET

Introduction

- This and the next two lectures focus on **limited dependent variables**:
  - Variables that have restricted support (a subset of the real line)
  - This restriction affects econometric modeling
- In this lecture we study the **binary case**, where Y takes two values:
  - Typically 0 and 1, so Y has support $\{0, 1\}$
  - These models are called **binary choice models**

Binary Dependent Variables

- Examples include:
  - Purchase of a single item
  - Market entry
  - Participation
  - Approval of an application/patent/loan
- The dependent variable may appear as:
  - Yes/No, True/False, or $1/-1$
  - But can always be written as $1/0$

Objective of Binary Choice Analysis

- Estimate the **conditional or response probability**:

$$\mathbb{P}(Y = 1 \mid X)$$

- Interest may lie in:
  - The probability itself
  - A transformation like its derivative — the **marginal effect**

Estimation Approaches

- Traditional modeling is **parametric**, using **maximum likelihood**
- There is also literature on **semiparametric estimation**
- In recent years, applied work often uses:
    - **Linear probability models** estimated by **least squares**

Binary Choice Models

- Let (Y, X) be random with $Y \in \{0, 1\}$ and $X \in \mathbb{R}^k$
- The **response probability** of Y given X is:

$$P(x) = \mathbb{P}(Y = 1 \mid X = x) = \mathbb{E}[Y \mid X = x]$$

Marginal Effect

- The **marginal effect** is the derivative of the response probability:

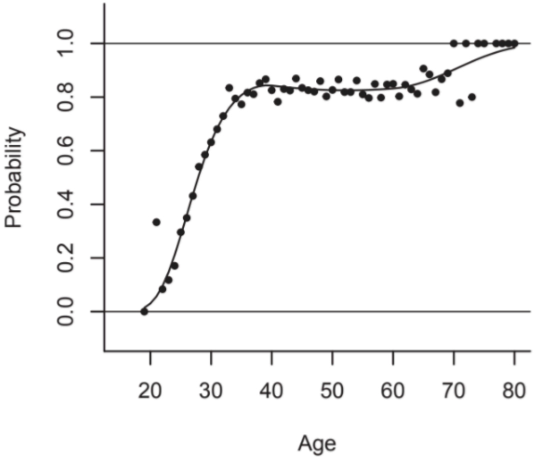
$$\frac{\partial}{\partial x}P(x) = \frac{\partial}{\partial x}\mathbb{P}(Y = 1 \mid X = x) = \frac{\partial}{\partial x}\mathbb{E}[Y \mid X = x]$$

- This is also called the **regression derivative**
- Many economic applications focus on the marginal effect

Illustration: Marriage and Age

- Subset: men with a college degree ($n = 6441$)
- Definition of Y:
  - $Y = 1$ if individual is married or widowed (not separated/divorced)
  - $Y = 0$ otherwise
- Regressor: **age**, taking values in $[19, 80]$

Probability of Marriage



Regression Framework

- The variables satisfy the regression framework:

$$Y = P(X) + e$$

$$\mathbb{E}[e \mid X] = 0$$

Error Term: Two Cases

- Since $Y \in \{0, 1\}$, the error e has **two possible values**:
- Case 1: $Y = 1$
  - Then $e = 1 - P(X)$
  - This happens with probability $P(X)$
- Case 2: $Y = 0$
  - Then $e = 0 - P(X) = -P(X)$
  - This happens with probability $1 - P(X)$
- So

$$e = \begin{cases} 1 - P(X), & \text{with probability } P(X) \\ -P(X), & \text{with probability } 1 - P(X) \end{cases} \tag{1}$$

Error Term: Variance

- Since $\mathbb{E}[e \mid X] = 0$, the conditional variance is:

$$\text{Var}(e \mid X) = \mathbb{E}[e^2 \mid X]$$

- Compute:

$$\text{Var}(e \mid X) = (1 - P(X))^2 \cdot P(X) + (P(X))^2 \cdot (1 - P(X))$$

- Simplify:

$$\text{Var}(e \mid X) = P(X)(1 - P(X)) \tag{2}$$

- The variance depends on $X \rightarrow$ **heteroskedasticity**

Models for the Response Probability

- We now describe the most common models for the response probability $P(x)$.

Linear Probability Model (LPM)

- Definition:

$$P(x) = x'\beta$$

- β is a coefficient vector
- **Properties**:
  - $P(x)$ is a linear function of x
  - Coefficients β represent **marginal effects**
  - Estimation is simple via least squares
- **Limitations**:
  - Violates the $[0, 1]$ probability bound
  - Fitted values can be negative or exceed 1
  - Poor fit for binary outcomes, especially in tail regions

Index Models

- General form:

$$P(x) = G(x'\beta)$$

- $G(\cdot)$ is a **link function** (e.g. normal CDF, logistic CDF)
- Also called **single index models**:
  - $x'\beta$ is the **linear index**
  - $G(u)$ maps the index to $[0, 1]$
- Common choices for $G(u)$:
  - Normal CDF $\Rightarrow$ **Probit model**
  - Logistic CDF $\Rightarrow$ **Logit model**
- Marginal effect:

$$\frac{\partial}{\partial x} P(x) = \beta \cdot g(x'\beta) \quad \text{where} \quad g(u) = \frac{\partial}{\partial u} G(u)$$

Probit and Logit Models

- **Probit model**:

$$P(x) = \Phi(x'\beta)$$

  - $\Phi(\cdot)$ is the **standard normal CDF**
  - Based on normal distribution
  - Easy to interpret, commonly used

- **Logit model**:

$$P(x) = \Lambda(x'\beta) = \frac{1}{1 + \exp(-x'\beta)}$$

  - $\Lambda(\cdot)$ is the **logistic CDF**
  - Produces similar results to probit
  - Closed-form expressions simplify computation

Linear Series Model

- Expands the linear index using basis functions:

$$P(x) = x_K' \beta_K$$

  - $x_K = x_K(x)$ is a vector of transformations of x
  - Can approximate any continuous function
- **Pros**:
  - Simple to estimate with linear methods
- **Cons**:
  - May not respect $[0, 1]$ bounds

Example: Marriage Probability vs. Age

- Suppose we want to model:

$$P(x) = \mathbb{P}(Y = 1 \mid \text{age} = x)$$

- A **linear model** assumes:

$$P(x) = \beta_0 + \beta_1 x$$

  - Too restrictive: assumes a straight-line relationship

What the Series Model Does

- Instead of using just x, we expand it:

$$x_K = \begin{bmatrix} x \\ x^2 \\ x^3 \end{bmatrix}, \quad P(x) = x_K' \beta_K = \beta_1 x + \beta_2 x^2 + \beta_3 x^3$$

- Now $P(x)$ can:
    - Increase quickly
    - Flatten out
    - Increase again

Visual Intuition

- Linear model $\rightarrow$ straight line
- Series model $\rightarrow$ smooth, flexible curve that better fits real data

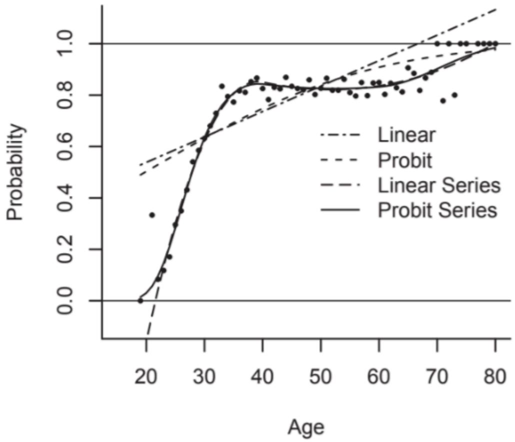
Index Series Model

- Combines a series expansion and a link function:

$$P(x) = G(x'_K \beta_K)$$

- $G(\cdot)$ ensures $P(x) \in [0, 1]$
- Can approximate any continuous **transformed** response probability
- **Pros**:
  - Flexible
  - Boundary-respecting

Probability of Marriage



Latent Variable Interpretation

- Index models can be interpreted as **latent variable models**.
- Let:

$$Y^* = X'\beta + e, \quad e \sim G(e)$$

- Y^* is unobserved (latent), linear in X, plus a random error e
- The observed variable Y is:

$$Y = \mathbb{1}\{Y^* > 0\} = \begin{cases} 1 & \text{if } Y^* > 0 \\ 0 & \text{otherwise} \end{cases}$$

Response Probability

- The event $Y = 1$ is equivalent to:

$$X'\beta + e > 0 \quad \Leftrightarrow \quad e > -X'\beta$$

- Thus:

$$P(x) = \mathbb{P}(Y = 1 \mid X = x) = \mathbb{P}(e > -x'\beta)$$

- If G is **symmetric around zero**, then:

$$P(x) = 1 - G(-x'\beta) = G(x'\beta)$$

Interpretation and Scale

- This model corresponds to a **structural choice** model:
  - Y^* represents utility (or profit) difference
  - Choose $Y = 1$ if utility exceeds zero
- Model type depends on G:
  - **Probit**: e is standard normal
  - **Logit**: e is standard logistic
- **Standard** means the distribution has:
  - Mean 0
  - **Unit variance**

Why Unit Variance?

- Suppose instead:

$$e = \sigma\varepsilon, \quad \varepsilon \sim G(u) \text{ with unit variance}$$

- Then:

$$P(x) = \mathbb{P}(\sigma\varepsilon > -x'\beta) = G\left(\frac{x'\beta}{\sigma}\right)$$

- Define scaled coefficients:

$$\beta^* = \frac{\beta}{\sigma}$$

- Then:

$$P(x) = G(x'\beta^*)$$

What Is Identified?

1. **Scaled coefficients**: $\beta^* = \beta/\sigma$
2. **Ratios of coefficients**: $\beta_1/\beta_2 = \beta_1^*/\beta_2^*$
3. **Marginal effects**:
$$\frac{\partial}{\partial x} P(x) = \beta^* \cdot g(x'\beta^*)$$

- These depend only on $\beta^* \rightarrow$ they **are identified**

Nonparametric Perspective

- Generalize the model:

$$Y^* = m(X) + e$$

- Then:

$$P(x) = 1 - G(-m(x)) \tag{4}$$

- If $G(e)$ and $m(x)$ are both nonparametric:
  - They cannot be separately identified
  - Only the composite $1 - G(-m(x))$ is identified

Likelihood

- Probit and logit models are typically estimated by **maximum likelihood (MLE)**.

Bernoulli Likelihood

- If $Y \sim \text{Bernoulli}(p)$:

$$\mathbb{P}(Y = y) = p^y(1-p)^{1-y}, \quad y \in \{0, 1\}$$

- In index models, $p = G(X'\beta)$:

$$\pi(Y \mid X) = G(X'\beta)^Y(1 - G(X'\beta))^{1-Y}$$

- Equivalently:

$$\pi(Y \mid X) = G(Z'\beta)$$

where:

$$Z = \begin{cases} X & \text{if } Y = 1 \\ -X & \text{if } Y = 0 \end{cases}$$

Log-Likelihood

- Taking logs and summing over observations:

$$\ell_n(\beta) = \sum_{i=1}^{n} \log G(Z_i'\beta)$$

- For **probit** and **logit**:

$$\ell_n^{\text{probit}}(\beta) = \sum_{i=1}^{n} \log \Phi(Z_i'\beta)$$

$$\ell_n^{\text{logit}}(\beta) = \sum_{i=1}^{n} \log \Lambda(Z_i'\beta)$$

First and Second Derivatives

- Let $h(x) = \frac{d}{dx} \log G(x)$ and $H(x) = -\frac{d^2}{dx^2} \log G(x)$
- For **logit**:

$$h_{\text{logit}}(x) = 1 - \Lambda(x)$$

$$H_{\text{logit}}(x) = \Lambda(x)(1 - \Lambda(x))$$

- For **probit**:

$$h_{\text{probit}}(x) = \frac{\phi(x)}{\Phi(x)} = \lambda(x) \quad \text{(Inverse Mills ratio)}$$

$$H_{\text{probit}}(x) = \lambda(x)(x + \lambda(x))$$

Logit Model Derivatives

We derive:

- $h_{\text{logit}}(x) = \frac{d}{dx} \log \Lambda(x)$
- $H_{\text{logit}}(x) = -\frac{d^2}{dx^2} \log \Lambda(x)$

Derivative of $\log \Lambda(x)$

- Start with:

$$\Lambda(x) = \frac{1}{1 + e^{-x}}$$

- Then:

$$h_{\text{logit}}(x) = \frac{d}{dx} \log \Lambda(x) = \frac{\Lambda'(x)}{\Lambda(x)}$$

- First, compute $\Lambda'(x)$:
- Let $f(x) = 1 + e^{-x}$, then:

$$\Lambda(x) = \frac{1}{f(x)} \quad \Rightarrow \quad \Lambda'(x) = -\frac{f'(x)}{f(x)^2}$$

- But $f'(x) = -e^{-x}$, so: $\Lambda'(x) = \frac{e^{-x}}{(1+e^{-x})^2}$

Express $\Lambda'(x)$ in terms of $\Lambda(x)$

- Recall:

$$\Lambda(x) = \frac{1}{1+e^{-x}}, \quad 1 - \Lambda(x) = \frac{e^{-x}}{1+e^{-x}}$$

- Then:

$$\Lambda(x)(1-\Lambda(x)) = \left(\frac{1}{1+e^{-x}}\right)\left(\frac{e^{-x}}{1+e^{-x}}\right) = \frac{e^{-x}}{(1+e^{-x})^2}$$

- So:

$$\Lambda'(x) = \Lambda(x)(1-\Lambda(x))$$

- Then:

$$h_{\text{logit}}(x) = \frac{\Lambda'(x)}{\Lambda(x)} = \frac{\Lambda(x)(1-\Lambda(x))}{\Lambda(x)} = 1 - \Lambda(x)$$

Second Derivative

- We compute:

$$H_{\text{logit}}(x) = -\frac{d}{dx}h_{\text{logit}}(x) = -\frac{d}{dx}(1 - \Lambda(x)) = \Lambda'(x)$$

- From earlier:

$$H_{\text{logit}}(x) = \Lambda(x)(1 - \Lambda(x))$$

Log-Concavity and Global Maximization

- Both $H_{\text{logit}}(x)$ and $H_{\text{probit}}(x) > 0$
- This implies **global concavity** of the log-likelihood:

$$\ell_n(\beta) \text{ is globally concave in } \beta$$

Score and Hessian

- **Score function**:

$$S_n(\beta) = \sum_{i=1}^{n} Z_i h(Z_i'\beta) \tag{6}$$

- **Hessian matrix**:

$$\mathcal{H}_n(\beta) = \sum_{i=1}^{n} X_i X_i' H(Z_i'\beta) \tag{7}$$

- Concavity ensures that MLE is:

$$\widehat{\beta}^{\text{probit}} = \arg\max_{\beta} \ell_n^{\text{probit}}(\beta), \quad \widehat{\beta}^{\text{logit}} = \arg\max_{\beta} \ell_n^{\text{logit}}(\beta)$$

Score Function: First Derivative of Log-Likelihood

- The score is the gradient of the log-likelihood:

$$S_n(\beta) = \frac{\partial \ell_n(\beta)}{\partial \beta} = \sum_{i=1}^{n} \frac{d}{d\beta} \log G(Z_i'\beta) = \sum_{i=1}^{n} Z_i \cdot h(Z_i'\beta)$$

- Where:

$$h(x) = \frac{d}{dx} \log G(x)$$

- This comes directly from the chain rule:

$$\frac{d}{d\beta} \log G(Z_i'\beta) = \frac{d}{d(Z_i'\beta)} \log G(Z_i'\beta) \cdot \frac{d(Z_i'\beta)}{d\beta} = h(Z_i'\beta) \cdot Z_i$$

Hessian Matrix: Second Derivative of Log-Likelihood

- The Hessian is the matrix of second derivatives:

$$\mathscr{H}_n(\beta) = \frac{\partial^2 \ell_n(\beta)}{\partial\beta\,\partial\beta'} = \sum_{i=1}^{n} \frac{d^2}{d\beta\,d\beta'} \log G(Z_i'\beta)$$

- Using the product rule:

$$\frac{d^2}{d\beta\,d\beta'} \log G(Z_i'\beta) = Z_i Z_i' \cdot H(Z_i'\beta)$$

Summary

- Logit and probit likelihoods are smooth, concave, and easy to optimize
- MLE has a **unique global maximum**
- Optimization relies on computing:
  - $\ell_n(\beta)$ (log-likelihood)
  - $S_n(\beta)$ (score)
  - $\mathcal{H}_n(\beta)$ (Hessian)

Marginal Effects in Index Models

- Consider the index model:

$$\mathbb{P}(Y = 1 \mid X = x) = G(x'\beta)$$

- The marginal effect of x is:

$$\delta(x) = \frac{\partial}{\partial x} P(x) = \beta g(x'\beta)$$

- This varies with x

Average Marginal Effect (AME)

- To summarize marginal effects over the sample:

$$\text{AME} = \mathbb{E}[\delta(X)] = \beta\, \mathbb{E}[g(X'\beta)]$$

- Estimator of $\delta(x)$:

$$\hat{\delta}(x) = \hat{\beta}\, g(x'\hat{\beta})$$

- Estimator of the AME:

$$\widehat{\text{AME}} = \frac{1}{n}\sum_{i=1}^{n} \hat{\delta}(X_i) = \hat{\beta}\left(\frac{1}{n}\sum_{i=1}^{n} g(X_i'\hat{\beta})\right)$$

Nonlinear Transformations in X

- For model:

$$\mathbb{P}(Y = 1 \mid X = x) = G(\beta_0 + \beta_1 x + \cdots + \beta_p x^p)$$

- The marginal effect is:

$$\delta(x) = (\beta_1 + \cdots + p\beta_p x^{p-1}) \cdot g(\beta_0 + \beta_1 x + \cdots + \beta_p x^p)$$

- Estimator of $\delta(x)$:

$$\hat{\delta}(x) = (\hat{\beta}_1 + \cdots + p\hat{\beta}_p x^{p-1}) \cdot g(\hat{\beta}_0 + \hat{\beta}_1 x + \cdots + \hat{\beta}_p x^p)$$

- Estimator of AME:

$$\widehat{\text{AME}} = \frac{1}{n} \sum_{i=1}^{n} \hat{\delta}(X_i)$$

Application: Marriage Probability and Marginal Effects

- Dataset: `cps09mar`, men aged 19–35 ($n = 9137$)
- Regressors: age, education, race (Black, Asian, Hispanic), and region (3 dummies)
- Models: linear **logit** and **probit**
- Estimates: coefficients, **average marginal effects (AME)**, and **robust SEs**

Binary Choice regressions for Marriage

	Logit		Probit	
	Coefficient	AME	Coefficient	AME
age	0.217	0.044	0.132	0.045
	(0.006)	(0.001)	(0.003)	(0.001)
education	0.014	0.003	0.009	0.003
	(0.010)	(0.002)	(0.006)	(0.002)
Black	−0.767	−0.156	−0.454	−0.153
	(0.092)	(0.018)	(0.054)	(0.018)
Asian	0.033	0.007	0.025	0.008
	(0.103)	(0.021)	(0.063)	(0.021)
Hispanic	−0.084	−0.017	−0.048	−0.017
	(0.063)	(0.013)	(0.038)	(0.013)
MidWest	0.272	0.056	0.165	0.056
	(0.074)	(0.011)	(0.045)	(0.015)
South	0.338	0.069	0.203	0.069
	(0.070)	(0.014)	(0.043)	(0.014)
West	0.383	0.078	0.228	0.077
	(0.072)	(0.015)	(0.044)	(0.015)
Intercept	−6.45		−3.93	
	(0.21)		(0.12)	

Key Results

- **Logit and probit** give:
  - Same signs and qualitative results
  - Logit coefficients are larger in magnitude
  - **AMEs are nearly identical**
- **Interpretation** of marginal effects:
  - Age: +4.5% increase in marriage probability per year
  - Education: +0.3% per year
  - Black: –15% vs. omitted group
  - Asian: +1%; Hispanic: –2%
  - Midwest, South, West: +6–8% vs. NorthEast
  - NorthEast: –7% relative to other regions

Takeaways

- Use **AMEs coefficients** for interpretation
- **Logit vs. Probit**: model choice not critical if both approximate $P(Y = 1|X)$ well
- **Model specification matters**:
    - Sample restricted to 19–35: marginal effect of age $=$ 4.5%
    - Full sample: marginal effect of age would drop to 1% $\rightarrow$ **misspecification bias**
- Always verify that your **functional form** and **sample** match the phenomenon

Optional

Pseudo-True Values

- We define the **expected log mass function**:

$$\ell(\beta) = \mathbb{E}[\log G(Z'\beta)]$$

Correct vs Misspecified Models

- The model is **correctly specified** if:

$$\mathbb{P}(Y = 1 \mid X) = G(X'\beta_0)$$

  for some β_0.

- Then β_0 maximizes $\ell(\beta)$.

$$\beta_0 = \arg\max_{\beta} \ell(\beta) \tag{8}$$

Global Concavity and Uniqueness

- When $G(x)$ is **log-concave** (as in logit and probit), $\ell(\beta)$ is globally concave.
- Define:

$$Q(\beta) = -\frac{\partial^2}{\partial\beta\partial\beta'}\ell(\beta) = \mathbb{E}\left[XX'H(Z'\beta)\right]$$

- Log-concavity implies $H(x) > 0$, so:
  - $Q(\beta) \succeq 0$
  - $\ell(\beta)$ is **concave**
  - Uniqueness if $Q(\beta_0)$ is positive definite:

$$\mathbb{E}\left[XX'H(X'\beta_0)\right] > 0 \tag{9}$$

Probit and Logit Pseudo-True Values

- Probit:

$$\ell^{\text{probit}}(\beta) = \mathbb{E}[\log \Phi(Z'\beta)], \quad \beta_0^{\text{probit}} = \arg\max_{\beta} \ell^{\text{probit}}(\beta)$$

- Logit:

$$\ell^{\text{logit}}(\beta) = \mathbb{E}[\log \Lambda(Z'\beta)], \quad \beta_0^{\text{logit}} = \arg\max_{\beta} \ell^{\text{logit}}(\beta)$$

Full Rank Conditions

- For **logit**, since $H_{\text{logit}}(x) = \Lambda(x)(1 - \Lambda(x))$, we require:

$$Q_{\text{logit}} \equiv \mathbb{E}\left[XX'\Lambda(X'\beta)(1 - \Lambda(X'\beta))\right] > 0 \tag{10}$$

- For **probit**, using $H_{\text{probit}}(x) = \lambda(x)(x + \lambda(x))$, we require:

$$Q_{\text{probit}} \equiv \mathbb{E}\left[XX'H_{\text{probit}}(Z'\beta)\right] > 0 \tag{11}$$

- These are the **full rank conditions** needed to ensure that the pseudo-true values β_0^{probit} and β_0^{logit} are **unique**.

Consistency of Logit Estimation

Theorem 1 *(Consistency of Logit Estimation)*

If (1) (Y_i, X_i) are i.i.d.; (2) $\mathbb{E}\|X\| < \infty$; (3) $Q_{\text{logit}} > 0$; and (4) B is compact, then:

$$\widehat{\beta}^{\text{logit}} \xrightarrow{p} \beta^{\text{logit}} \quad \text{as } n \to \infty.$$

Consistency of Probit Estimation

Theorem 2 *(Consistency of Probit Estimation)*

If (1) (Y_i, X_i) are i.i.d.; (2) $\mathbb{E}\|X\|^2 < \infty$; (3) $Q_{\text{probit}} > 0$; and (4) B is compact, then:

$$\widehat{\beta}^{\text{probit}} \xrightarrow{p} \beta^{\text{probit}} \quad \text{as } n \to \infty.$$

Asymptotic Normality of Logit MLE

Theorem 3 *(Asymptotic Normality of Logit MLE)*

If the conditions of Theorem 25.1 hold, $\mathbb{E}\|X\|^4 < \infty$, and β^{logit} is in the interior of B, then:

$$\sqrt{n}\left(\hat{\beta}^{\text{logit}} - \beta^{\text{logit}}\right) \xrightarrow{d} \mathcal{N}(0, V_{\text{logit}})$$

where:

$$V_{\text{logit}} = Q_{\text{logit}}^{-1} \Omega_{\text{logit}} Q_{\text{logit}}^{-1}$$

Asymptotic Normality of Probit MLE

Theorem 4 *(Asymptotic Normality of Probit MLE)*

If the conditions of Theorem 25.2 hold, $\mathbb{E}\|X\|^4 < \infty$, and β^{probit} is in the interior of B, then:

$$\sqrt{n}\left(\hat{\beta}^{\text{probit}} - \beta^{\text{probit}}\right) \xrightarrow{d} \mathcal{N}(0, V_{\text{probit}})$$

where:

$$V_{\text{probit}} = Q_{\text{probit}}^{-1} \Omega_{\text{probit}} Q_{\text{probit}}^{-1}$$

Covariance Matrix Estimation: Logit Model

- Define fitted values:

$$\hat{\Lambda}_i = \Lambda(X_i'\hat{\beta}^{\text{logit}})$$

- Let:

$$\hat{Q}_{\text{logit}} = \frac{1}{n}\sum_{i=1}^{n} X_i X_i' \hat{\Lambda}_i(1-\hat{\Lambda}_i)$$

$$\hat{\Omega}_{\text{logit}} = \frac{1}{n}\sum_{i=1}^{n} X_i X_i' (Y_i - \hat{\Lambda}_i)^2$$

- Sandwich covariance estimator:

$$\hat{V}_{\text{logit}} = \hat{Q}_{\text{logit}}^{-1}\hat{\Omega}_{\text{logit}}\hat{Q}_{\text{logit}}^{-1}$$

- Under correct specification:

$$\hat{V}_{\text{logit}}^{0} = \hat{Q}_{\text{logit}}^{-1}$$

Covariance Matrix Estimation: Probit Model

- Define:

$$\hat{\mu}_i = Z_i' \hat{\beta}^{\text{probit}}, \quad \hat{\lambda}_i = \lambda(\hat{\mu}_i)$$

- Let:

$$\hat{Q}_{\text{probit}} = \frac{1}{n} \sum_{i=1}^{n} X_i X_i' \hat{\lambda}_i (\hat{\mu}_i + \hat{\lambda}_i)$$

$$\hat{\Omega}_{\text{probit}} = \frac{1}{n} \sum_{i=1}^{n} X_i X_i' \hat{\lambda}_i^2$$

- Sandwich covariance estimator:

$$\hat{V}_{\text{probit}} = \hat{Q}_{\text{probit}}^{-1} \hat{\Omega}_{\text{probit}} \hat{Q}_{\text{probit}}^{-1}$$

Covariance Matrix Estimation: Probit Model (Cont.)

- Under correct specification:

$$\hat{Q}^0_{\text{probit}} = \frac{1}{n}\sum_{i=1}^{n} X_i X_i' \lambda(X_i'\hat{\beta}^{\text{probit}})\lambda(-X_i'\hat{\beta}^{\text{probit}})$$

$$\hat{V}^0_{\text{probit}} = \left(\hat{Q}^0_{\text{probit}}\right)^{-1}$$

Reading

- **Hansen (2022), Econometrics.**
  - Sections: 25.1-25.10