Econometrics 2

Regression with a Binary Dependent Variable

Lasha Chochua

2026

Motivation and Setup

Motivation – Detecting Discrimination

- Suppose two identical people, differing only in race, apply for a mortgage.
- By law, banks must treat them equally.
- The question: do they actually receive equal treatment in practice?

- Regulators worry about possible racial discrimination.
- A single denial doesn't prove discrimination – need **statistical evidence**.
- Must compare outcomes **holding applicant characteristics constant**.

Binary Dependent Variables

- Many economic questions involve **binary outcomes**:
  - Loan approved or denied
  - Student goes to college or not
  - Country receives aid or not
- Binary outcome means the dependent variable takes values 0 or 1.
- These are examples of **limited dependent variables**.

Example – Mortgage Applications

- Data: Boston mortgage applications (1990), HMDA dataset.
- Question: Is **race** a factor in loan denial?
- Define variables:

- deny $= 1$ if denied, 0 if accepted
- P/I ratio $= \frac{\text{monthly payment}}{\text{monthly income}}$

- Higher P/I ratio $\rightarrow$ more likely to be denied.

Visualizing the Relationship

- Figure 1: Scatterplot of deny vs. P/I ratio
- Key patterns:
  - P/I ratio $< 0.3 \rightarrow$ few denials
  - P/I ratio $> 0.4 \rightarrow$ many denials

- OLS regression line shows increasing probability of denial as P/I ratio increases.

Visualizing the Relationship (Cont.)

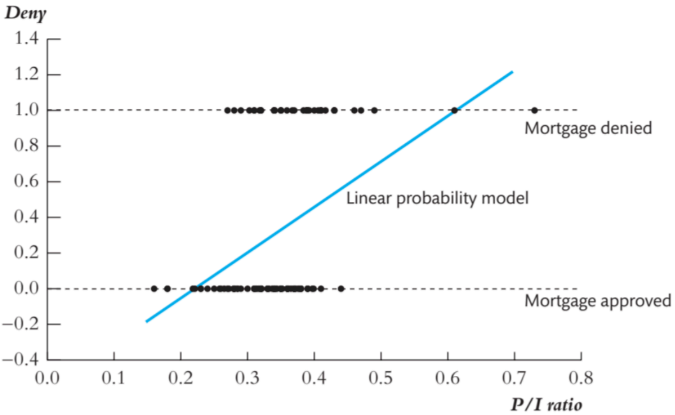

Figure 1: Mortgage Application Denial and the Payment-to-Income Ratio

Regression Framework

- We use **multiple regression** to estimate the effect of race on denial.
- The dependent variable is binary.
- Linear model predicts:

$$\mathbb{E}(Y \mid X_1, \dots, X_k) = \Pr(Y = 1 \mid X_1, \dots, X_k)$$

- In our case: predicts **probability of denial**.

Application and Broader Context

- Use regression to test for **racial bias** in mortgage denials.
- Control for other factors (e.g., income, loan size).
- Binary regression models help identify **statistical discrimination**.

- Broader class: **limited dependent variable models**.
- Other models handle outcomes with more than two categories.

The Linear Probability Model

The Linear Probability Model

- The **linear probability model (LPM)** is a linear regression model with a binary dependent variable.
- It estimates the probability that $Y = 1$ given X.
- Interpretation: $\widehat{\beta}_1$ is the **change in probability** of $Y = 1$ for a **unit change in** X.
- The predicted value $\hat{Y}_i$ is interpreted as a probability.

The LPM – Definition

The Linear Probability Model

The **linear probability model** is the linear multiple regression model,

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i,$$

applied to a binary dependent variable Y_i. Because Y is binary,

$$\mathbb{E}(Y \mid X_1, X_2, \ldots, X_k) = \Pr(Y = 1 \mid X_1, X_2, \ldots, X_k).$$

So,

$$\Pr(Y = 1 \mid X_1, X_2, \ldots, X_k) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k.$$

The coefficient β_1 is the difference in the probability that $Y = 1$ associated with a one-unit change in X_1, holding other regressors fixed. Coefficients can be estimated by OLS, and standard errors should be heteroskedasticity-robust.

Application – Boston HMDA Data

- OLS regression of the binary variable $deny$ on P/I ratio using 2,380 observations:

$$\widehat{deny} = -0.080 + 0.604 \cdot P/I \ ratio$$
$$(0.032) \quad (0.098)$$

- Coefficient on P/I ratio is statistically significant at the 1% level ($t = 6.13$).
- Interpretation: applicants with higher payment-to-income ratios are more likely to be denied.

Interpreting the Coefficient

- A 0.1 increase in P/I ratio implies a 6 percentage point increase in denial probability:

$$0.604 \times 0.1 \approx 0.060$$

- For example, if $P/I\ ratio = 0.3$, then:

$$\widehat{deny} = -0.080 + 0.604 \cdot 0.3 = 0.101$$

- Meaning: 10.1% probability of denial for an applicant whose P/I ratio is 0.3.

Adding Race to the Model

- To test for racial bias, add a dummy variable $black$:
  - $black = 1$ if applicant is Black
  - $black = 0$ if applicant is white

$$\widehat{deny} = \underset{(0.029)}{-0.091} + \underset{(0.089)}{0.559} \cdot P/I\ ratio + \underset{(0.025)}{0.177} \cdot black$$

- Interpretation: holding P/I ratio constant, Black applicants have a 17.7% higher probability of denial.

Interpretation and Omitted Variable Bias

- Estimate suggests racial bias may exist.
- But be **cautious**: omitted variables (e.g., credit score, income potential) may bias results.
- $black$ may be correlated with unobserved factors, causing omitted variable bias.
- We delay conclusions until more robust analysis.

Shortcomings of the LPM

- LPM is easy to estimate, but has major flaws:
  - **Predicted probabilities** can be less than 0 or greater than 1.
  - Effects are assumed **constant** across X, but real-world effects can be nonlinear.
- For example:
  - Increasing P/I ratio from 0.3 to 0.4 may strongly affect denial probability.
  - Increasing P/I ratio from 0.8 to 0.9 may have no effect if already likely to be denied.
- Solution: switch to **nonlinear** models like **probit** and **logit**.

Probit and Logit Regression

Probit and Logit Regression

- **Probit** and **logit** regression are **nonlinear models** designed for binary dependent variables.
- These models estimate the probability that $Y = 1$.
- Unlike the LPM, they **force predicted probabilities to lie in** $[0, 1]$.

- The idea: use a **cumulative distribution function (c.d.f.)** to model probabilities.
  - Probit regression: uses the **standard normal c.d.f.**
  - Logit regression (or **logistic regression**): uses the **logistic c.d.f.**

- These models are preferred when:
  - You want valid predicted probabilities.
  - You expect **nonlinear** relationships between regressors and the outcome.

Probit Regression – Single Regressor

- **Model:**

$$\Pr(Y = 1 \mid X) = \Phi(\beta_0 + \beta_1 X)$$

- $\Phi(\cdot)$ is the standard normal cumulative distribution function.
- β_1 affects the **z-value**, not directly the **probability**.

Example – Interpreting Probit Coefficients

- Suppose:
  - Y = mortgage denial
  - X = payment-to-income ratio
  - $\beta_0 = -2$, $\beta_1 = 3$
- Then:

$$\Pr(Y = 1 \mid P/I\ ratio = 0.4) = \Phi(-2 + 3 \cdot 0.4) = \Phi(-0.8) \approx 21.2\%$$

- Compute $z = -2 + 3 \cdot 0.4 = -0.8$, then look up $\Phi(-0.8)$.
- Interpretation: applicant has a 21.2% chance of being denied.

Probit Probability Curve

- The function has an **S-shape**:
  - Flat near 0 and 1
  - Steep in the middle range
- Examples from the model:
  - $P/I = 0.2 \Rightarrow \Pr(\text{deny}) = 2.1\%$
  - $P/I = 0.3 \Rightarrow \Pr(\text{deny}) = 16.1\%$
  - $P/I = 0.4 \Rightarrow \Pr(\text{deny}) = 51.9\%$
  - $P/I = 0.6 \Rightarrow \Pr(\text{deny}) = 98.3\%$

Probit Probability Curve (Cont.)

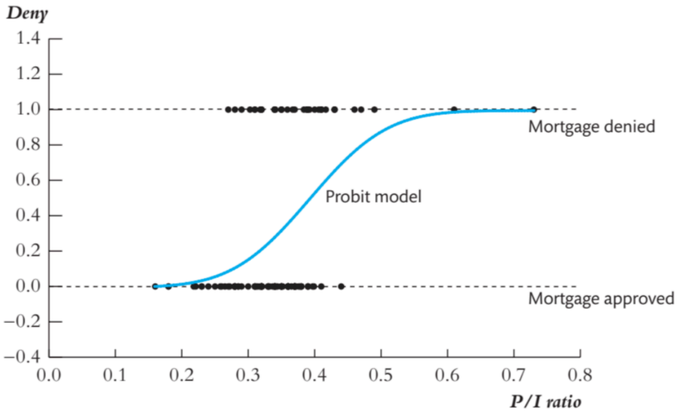

Figure 2: Probit Model of the Probability of Denial Given P/I Ratio

Probit Regression – Multiple Regressors

- **Model:**

$$\Pr(Y = 1 \mid X_1, X_2) = \Phi(\beta_0 + \beta_1 X_1 + \beta_2 X_2)$$

- Example: $\beta_0 = -1.6$, $\beta_1 = 2$, $\beta_2 = 0.5$; $X_1 = 0.4$, $X_2 = 1$
- Then:

$$z = -1.6 + 2 \cdot 0.4 + 0.5 \cdot 1 = -0.3 \Rightarrow \Pr(Y = 1) = \Phi(-0.3) \approx 38\%$$

- This generalizes the probit model to handle multiple covariates.

Interpreting Probit Coefficients

- Coefficients affect the **z-value**, not directly the probability.
- A positive β_1 increases z, thus increases $\Pr(Y = 1)$.
- A negative β_1 decreases the probability.
- Effect on probability is **nonlinear**, even though z is linear in X.

The Probit Model – Predicted Probabilities and Effects

The Probit Model

The population probit model with multiple regressors is

$$\Pr(Y = 1 \mid X_1, X_2, \ldots, X_k) = \Phi(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k),$$

where Φ is the cumulative standard normal distribution function.

- To compute the predicted probability, calculate $z = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k$ and look up $\Phi(z)$.
- The coefficient β_1 is the difference in the z-value from a one-unit change in X_1, holding all other variables constant.
- The change in predicted probability: compute initial probability, compute new probability after changing a regressor, take the difference.

Application – Probit Model Specification

- We fit a probit model to 2,380 observations on mortgage denial.
- The estimated probit model is:

$$\text{Pr}(\text{deny} = 1 \mid P/I \text{ ratio}) = \Phi(-2.19 + 2.97 \cdot P/I \text{ ratio})$$

- Standard errors: (0.16) and (0.47).

- The estimated coefficients (-2.19 and 2.97) are difficult to interpret directly.
- They affect the probability of denial via the z-value.
- The payment-to-income ratio is positively related to denial probability.
- Relationship is statistically significant: t-statistic $= 2.97/0.47 = 6.32$.

Predicted Probability Change – P/I Ratio

- What is the change in predicted denial probability when P/I ratio increases from 0.3 to 0.4?

- For P/I ratio $= 0.3$:

$$\Phi(-2.19 + 2.97 \times 0.3) = \Phi(-1.30) = 0.097$$

- The probability of denial is 9.7%.

- For P/I ratio $= 0.4$:

$$\Phi(-2.19 + 2.97 \times 0.4) = \Phi(-1.00) = 0.159$$

- The probability of denial is 15.9%.

- The estimated change in denial probability is $0.159 - 0.097 = 0.062$, i.e. a **6.2 percentage point increase**.

Nonlinearity of the Probit Function

- The probit regression function is nonlinear.
- The effect of a change in X depends on the starting value of X.

- For P/I ratio $= 0.5$:

$$\Phi(-2.19 + 2.97 \times 0.5) = \Phi(-0.71) = 0.239$$

- Change when P/I ratio increases from 0.4 to 0.5: $0.239 - 0.159 = 0.080$, i.e. **8.0 percentage points**.
- This is larger than the 6.2 percentage points from 0.3 to 0.4 – illustrating nonlinearity.

Effect of Race on Mortgage Denial

- To estimate the effect of race ($black$) while holding P/I ratio constant, we add $black$ as a regressor.
- The estimated probit model is:

$$\text{Pr}(\text{deny} = 1 \mid P/I \text{ ratio}, \; black) = \Phi(-2.26 + 2.74 \cdot P/I \text{ ratio} + 0.71 \cdot black)$$

- Standard errors: (0.16), (0.44), and (0.083).

Interpreting Coefficients with Race

- Coefficient values are difficult to interpret directly; sign and significance matter.
- The coefficient on $black$ is positive (0.71): African American applicants face a higher probability of denial, holding the payment-to-income ratio constant.
- The coefficient on $black$ is statistically significant: t-statistic $= 8.55$ (significant at the 1% level).

Predicted Probability – White vs. Black Applicant

- For a **white applicant** $(black = 0)$ with P/I ratio $= 0.3$:
  - Predicted denial probability is **7.5%**.
- For a **Black applicant** $(black = 1)$ with P/I ratio $= 0.3$:
  - Predicted denial probability is **23.3%**.
- The difference in denial probabilities is **15.8 percentage points**.

Estimation of Probit Coefficients

- The probit coefficients are estimated using the **method of maximum likelihood**.
- This method produces **efficient (minimum variance) estimators**.
- It is suitable for regression with a binary dependent variable.

- The maximum likelihood estimator is **consistent** and **normally distributed in large samples**.
- This allows for constructing t-statistics and confidence intervals:

$$\text{estimated coefficient} \pm 1.96 \times \text{standard error}$$

Logit Regression Model

- The logit regression model is similar to the probit model.
- The standard normal c.d.f. (Φ) is replaced by the **cumulative standard logistic distribution function**, denoted F.

The Logit Model

The population logit model of the binary dependent variable Y with multiple regressors is

$$\Pr(Y = 1 \mid X_1, X_2, \ldots, X_k) = F(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k)$$

$$= \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k)}}$$

Logit regression is similar to probit regression except that the cumulative distribution function is different.

Logit Coefficients and Estimation

- The logistic c.d.f. is defined in terms of the exponential function.
- Logit coefficients are best interpreted by computing predicted probabilities and differences in predicted probabilities.

- Coefficients of the logit model are estimated by **maximum likelihood**.
- The maximum likelihood estimator is consistent and normally distributed in large samples.
- So, t-statistics and confidence intervals can be constructed.

Comparing Logit and Probit

- The logit and probit regression functions are similar, as illustrated in Figure 3.
- Figure 3 graphs both functions for the dependent variable $deny$ using the single regressor P/I ratio.

- Historically, logit regression was motivated by computational speed.
- With modern computers, this distinction is no longer important.

Probit and Logit Curves

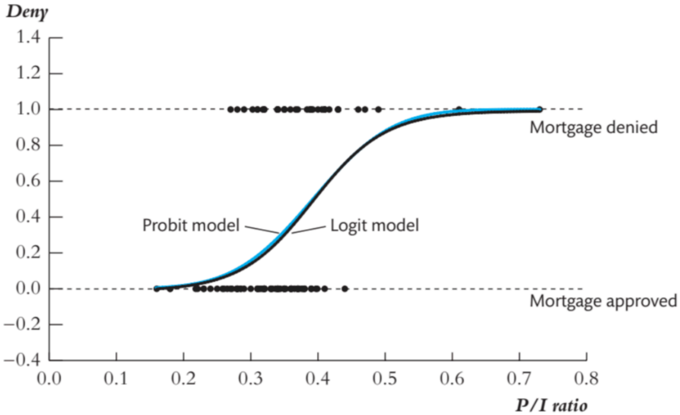

Figure 3: Probit and Logit Models of the Probability of Denial Given P/I Ratio

Application – Boston HMDA Data (Logit)

- We apply logit regression to the Boston HMDA data.
- Dependent variable: mortgage denial ($deny$); regressors: P/I ratio and $black$.

- The estimated regression function is:

$$\text{Pr}(\text{deny} = 1 \mid P/I \text{ ratio}, \ black) = F(-4.13 + 5.37 \cdot P/I \text{ ratio} + 1.27 \cdot black)$$

- Standard errors: (0.35), (0.96), and (0.15) respectively.

Interpreting the $black$ Coefficient

- The coefficient on $black$ is positive (1.27): higher probability of denial for Black applicants, consistent with earlier probit findings.
- The coefficient on $black$ is statistically significant: t-statistic $= 8.47$ (significant at the 1% level).

Predicted Probability – White vs. Black Applicant (Logit)

- For a **white applicant** ($black = 0$) with P/I ratio $= 0.3$:

$$1/[1 + e^{-(-4.13+5.37\times0.3+1.27\times0)}] = 1/[1 + e^{2.52}]$$

  - Predicted denial probability: **7.4%**.

- For a **Black applicant** ($black = 1$) with P/I ratio $= 0.3$:

$$1/[1 + e^{-(-4.13+5.37\times0.3+1.27\times1)}] = 1/[1 + e^{1.25}]$$

  - Predicted denial probability: **22.2%**.

- The **difference** between the two probabilities is **14.8 percentage points**.

Marginal Effects in Probit and Logit

- The raw coefficients affect the **z-value**, not directly the probability – so how do we report a single interpretable number?
- Answer: compute the **marginal effect** of x_j:

$$\delta_j(x) = \frac{\partial}{\partial x_j} \Pr(Y = 1 \mid X = x) = \beta_j \cdot g(x'\beta)$$

where $g(u) = G'(u)$ is the PDF corresponding to the link function G.

- For **probit**: $\delta_j(x) = \beta_j \cdot \phi(x'\beta)$, where ϕ is the standard normal PDF.
- For **logit**: $\beta_j \cdot F(x'\beta)(1 - F(x'\beta))$.
- The marginal effect **varies with** x – it is not a single number.

Average Marginal Effect (AME)

- To summarize across the sample, compute the **Average Marginal Effect**:

Average Marginal Effect (AME)

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

Sample estimator:

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

- Compute the marginal effect at **each observation's actual covariate values**, then average.
- Preferred over evaluating at sample means, which may not represent a typical observation.
- **Logit and probit AMEs are typically nearly identical** even though their raw coefficients differ in scale – model choice is usually not critical.

Measures of Fit

Measures of Fit

- R^2 is a poor measure of fit for binary dependent variable models.
- Two main measures exist:
  - **Fraction correctly predicted**
  - **Pseudo-R^2**

Fraction Correctly Predicted

- Uses the following prediction rule:
  - If $Y_i = 1$ and predicted probability exceeds 50%, prediction is correct.
  - If $Y_i = 0$ and predicted probability is below 50%, prediction is correct.
  - Otherwise, Y_i is incorrectly predicted.

- The measure is the fraction of n observations correctly predicted.

- **Advantage**: easy to understand.
- **Disadvantage**: does not reflect prediction quality – a predicted probability of 51% or 90% are treated the same when $Y_i = 1$.

Pseudo-R^2

- The pseudo-R^2 measures fit using the likelihood function.
- Adding a regressor to a probit or logit model increases the maximized likelihood – analogous to adding a regressor reducing the sum of squared residuals in OLS.

- The pseudo-R^2 compares the maximized likelihood of the full model to the likelihood of a model with no regressors:

$$\text{pseudo-}R^2 = 1 - \frac{\ln(f^{\text{max}}_{\text{probit}})}{\ln(f^{\text{max}}_{\text{Bernoulli}})}$$

- $f^{\text{max}}_{\text{probit}}$: maximized probit likelihood (includes the X's).
- $f^{\text{max}}_{\text{Bernoulli}}$: maximized Bernoulli likelihood (no X's included).

Back to the Boston HMDA Data

Variable	Definition	Sample Average
Financial Variables		
P/I ratio	Ratio of total monthly debt payments to total monthly income	0.331
housing expense-to-income ratio	Ratio of monthly housing expenses to total monthly income	0.255
loan-to-value ratio	Ratio of size of loan to assessed value of property	0.738
consumer credit score	1 if no "slow" payments or delinquencies 2 if one or two slow payments or delinquencies 3 if more than two slow payments 4 if insufficient credit history for determination 5 if delinquent credit history with payments 60 days overdue 6 if delinquent credit history with payments 90 days overdue	2.1
mortgage credit score	1 if no late mortgage payments 2 if no mortgage payment history 3 if one or two late mortgage payments 4 if more than two late mortgage payments	1.7
public bad credit record	1 if any public record of credit problems (bankruptcy, charge-offs, collection actions) 0 otherwise	0.074
Additional Applicant Characteristics		
denied mortgage insurance	1 if applicant applied for mortgage insurance and was denied, 0 otherwise	0.020
self-employed	1 if self-employed, 0 otherwise	0.116
single	1 if applicant reported being single, 0 otherwise	0.393
high school diploma	1 if applicant graduated from high school, 0 otherwise	0.984
unemployment rate	1989 Massachusetts unemployment rate in the applicant's industry	3.8
condominium	1 if unit is a condominium, 0 otherwise	0.288
black	1 if applicant is black, 0 if white	0.142
deny	1 if mortgage application denied, 0 otherwise	0.120

Back to the Boston HMDA Data (Cont.)

Dependent variable: *deny* = 1 if mortgage application is denied, = 0 if accepted; 2380 observations.

Regression Model	LPM	Logit	Probit	Probit	Probit	Probit
Regressor	(1)	(2)	(3)	(4)	(5)	(6)
black	0.084** (0.023)	0.688** (0.182)	0.389** (0.098)	0.371** (0.099)	0.363** (0.100)	0.246 (0.448)
P/I ratio	0.449** (0.114)	4.76** (1.33)	2.44** (0.61)	2.46** (0.60)	2.62** (0.61)	2.57** (0.66)
housing expense-to-income ratio	−0.048 (0.110)	−0.11 (1.29)	−0.18 (0.68)	−0.30 (0.68)	−0.50 (0.70)	−0.54 (0.74)
medium loan-to-value ratio (0.80 ≤ *loan-value ratio* ≤ 0.95)	0.031* (0.013)	0.46** (0.16)	0.21** (0.08)	0.22** (0.08)	0.22** (0.08)	0.22** (0.08)
high loan-to-value ratio (*loan-value ratio* > 0.95)	0.189** (0.050)	1.49** (0.32)	0.79** (0.18)	0.79** (0.18)	0.84** (0.18)	0.79** (0.18)
consumer credit score	0.031** (0.005)	0.29** (0.04)	0.15** (0.02)	0.16** (0.02)	0.34** (0.11)	0.16** (0.02)
mortgage credit score	0.021 (0.011)	0.28* (0.14)	0.15* (0.07)	0.11 (0.08)	0.16 (0.10)	0.11 (0.08)
public bad credit record	0.197** (0.035)	1.23** (0.20)	0.70** (0.12)	0.70** (0.12)	0.72** (0.12)	0.70** (0.12)
denied mortgage insurance	0.702** (0.045)	4.55** (0.57)	2.56** (0.30)	2.59** (0.29)	2.59** (0.30)	2.59** (0.29)
self-employed	0.060** (0.021)	0.67** (0.21)	0.36** (0.11)	0.35** (0.11)	0.34** (0.11)	0.35** (0.11)
single				0.23** (0.08)	0.23** (0.08)	0.23** (0.08)
high school diploma				−0.61** (0.23)	−0.60* (0.24)	−0.62** (0.23)
unemployment rate				0.03 (0.02)	0.03 (0.02)	0.03 (0.02)
condominium					−0.05 (0.09)	
black × *P/I ratio*						−0.58 (1.47)
black × *housing expense-to-income ratio*						1.23 (1.69)
additional credit rating indicator variables	no	no	no	no	yes	no
constant	−0.183** (0.028)	−5.71** (0.48)	−3.04** (0.23)	−2.57** (0.34)	−2.90** (0.39)	−2.54** (0.35)

Maximum Likelihood Estimation

Introduction to MLE

- We now briefly introduce MLE for binary response models.
- We derive the MLE of the success probability (p) for i.i.d. Bernoulli random variables.
- We then extend this to probit and logit models.
- Discussion includes pseudo-R^2 and standard errors for predicted probabilities.

What Is MLE Doing? – Intuition

- Consider a coin with unknown probability p of heads.
- You flip it 10 times and observe: H, H, T, H, T, T, H, H, H, T – i.e. 6 heads, 4 tails.
- **Question**: what value of p makes this outcome most plausible?

- The probability of observing exactly this sequence is:

$$p^6(1-p)^4$$

- MLE answers: find the p that **maximizes** this expression.

- The answer is $\hat{p} = 6/10 = 0.6$ – the sample proportion.
- Intuitively: MLE picks the parameter value under which the data we actually saw is most likely.

- The same logic applies in probit/logit: find $(\beta_0, \beta_1, \dots, \beta_k)$ that make the observed pattern of 0s and 1s most probable.

MLE for i.i.d. Bernoulli Random Variables

- For n i.i.d. observations $Y_1, \ldots, Y_n$ of a Bernoulli random variable, the joint probability distribution is:

$$\Pr(Y_1 = y_1, \ldots, Y_n = y_n) = p^{(y_1 + \cdots + y_n)}(1-p)^{n-(y_1+\cdots+y_n)}$$

- The **likelihood function** treats this as a function of the unknown parameter p. Let $S = \sum_{i=1}^{n} Y_i$:

$$f_{\text{Bernoulli}}(p; Y_1, \ldots, Y_n) = p^S(1-p)^{n-S}$$

Maximizing the Likelihood Function

- The MLE of p is the value that maximizes the likelihood.
- It is convenient to maximize the **log-likelihood** (the logarithm is strictly increasing, so the maximizer is the same):

$$S \ln(p) + (n - S) \ln(1 - p)$$

- The derivative with respect to p:

$$\frac{d}{dp} \ln \left[f_{\text{Bernoulli}}(p; Y_1, \dots, Y_n) \right] = \frac{S}{p} - \frac{n - S}{1 - p}$$

- Setting this to zero and solving yields $\hat{p} = S/n = \bar{Y}$.

MLE for the Probit Model

- For the probit model, the probability that $Y_i = 1$ is:

$$p_i = \Phi(\beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki})$$

- Assuming $(X_{1i}, \dots, X_{ki}, Y_i)$ are i.i.d., the conditional probability for observation i is $p_i^{y_i}(1 - p_i)^{1-y_i}$.
- The **log-likelihood function** for the probit model is:

$$\sum_{i=1}^{n} Y_i \ln[\Phi(\beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki})] + \sum_{i=1}^{n}(1 - Y_i)\ln[1 - \Phi(\beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki})]$$

Maximizing the Probit Likelihood

- The MLE for the probit model maximizes the log-likelihood function above.
- There is no simple, closed-form formula for the probit MLE.
- The likelihood function must be maximized using a **numerical algorithm** on the computer.

- Under general conditions, the MLE is **consistent** and has a **normal sampling distribution in large samples**.
- These properties allow for standard statistical inference (t-statistics, confidence intervals).

MLE for the Logit Model

- The likelihood for the logit model is derived in the same way as for the probit model.
- Replace $\Phi(\beta_0 + \beta_1 X_{1i} + \cdots + \beta_k X_{ki})$ with $[1 + e^{-(\beta_0+\beta_1 X_{1i}+\cdots+\beta_k X_{ki})}]^{-1}$.

- As with the probit model, there is no closed-form formula for the MLE of the logit coefficients.
- The log-likelihood must be maximized **numerically**.

Standard Errors for Predicted Probabilities – Setup

- For simplicity, consider a single regressor in the probit model.
- The predicted probability at a fixed value x is:

$$\hat{p}(x) = \Phi(\hat{\beta}_0^{MLE} + \hat{\beta}_1^{MLE} x)$$

- Because $\hat{p}(x)$ depends on the MLEs, it also has a **sampling distribution**.

- To compute the variance of $\hat{p}(x)$, approximate the nonlinear function $\Phi(\hat{\beta}_0^{MLE} + \hat{\beta}_1^{MLE} x)$ by a linear function using a first-order Taylor series expansion.

Standard Errors for Predicted Probabilities – Approximation

- First-order Taylor expansion:

$$\hat{p}(x) \approx c + a_0(\hat{\beta}_0^{MLE} - \beta_0) + a_1(\hat{\beta}_1^{MLE} - \beta_1)$$

- where $c = \Phi(\beta_0 + \beta_1 x)$, and:

$$a_0 = \left.\frac{\partial \Phi(\beta_0 + \beta_1 x)}{\partial \beta_0}\right|_{\hat{\beta}_0^{MLE}, \hat{\beta}_1^{MLE}}, \quad a_1 = \left.\frac{\partial \Phi(\beta_0 + \beta_1 x)}{\partial \beta_1}\right|_{\hat{\beta}_0^{MLE}, \hat{\beta}_1^{MLE}}$$

Standard Errors for Predicted Probabilities – Variance

- Using the approximation and the formula for the variance of a sum of two random variables:

$$\mathsf{var}[\widehat{p}(x)] \approx a_0^2\,\mathsf{var}(\widehat{\beta}_0^{MLE}) + a_1^2\,\mathsf{var}(\widehat{\beta}_1^{MLE}) + 2a_0a_1\,\mathsf{cov}(\widehat{\beta}_0^{MLE}, \widehat{\beta}_1^{MLE})$$

- The standard error of $\widehat{p}(x)$ is computed using estimates of the variances and covariance of the MLEs.

Required Reading

- **Stock and Watson (2020)** Chapter 11, Sections 11.1–11.5; Appendix 11.2