Understanding and Applying the Probit Model

Prepared & presented by: Prof. GUECHARI Yasmina

1.Probit Model

1) What is the Probit Model?

- **The Probit model** is a type of regression used when the dependent variable is binary (e.g., 0 or 1, "yes" or "no"). It estimates the probability of an event occurring as a function of independent variables, assuming the error terms follow a standard normal distribution.

- **Purpose:**

➢ To model the relationship between a binary dependent variable and one or more independent variables.

➢ Similar to the Logit model, but assumes a normal cumulative distribution function (CDF).

1. Probit Model

- **The Probit and Logit models** are often used interchangeably because they both model binary outcomes. The main difference is in the assumed distribution of the error term:
- Logit model assumes a logistic (S-shaped) distribution of errors.
- Probit model assumes a normal distribution of errors.
- If you believe the normality assumption of errors is more realistic, use Probit.

1. Probit Model

- **Formula:** The Probit model can be expressed as:

$$P(Y = 1/X) = \Phi(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots\cdots + \beta_k X_k)$$

Where Φ is the CFD of standard normal distribution

P(Y=1/X): The probability of the event occurring.

β_0 is the intercept.

$\beta_1,\beta_2,\ldots,\beta_k$ are the coefficients of the independent variables $X_1,X_2,\ldots,X_k$.

$$\Phi(x) = P(Z \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

- Since there is no closed-form solution for this equation, values of Φ(x) are typically found using statistical tables, For example: **Φ(1) ≈ 0.8413**

1. Probit Model

- The Probit model is widely used in situations where the outcome is binary. Common examples include:

1.Economics and Finance:

✓Predicting the probability of loan default based on credit score, income, etc.

✓Modeling whether an investor chooses to invest or not.

2.Marketing:

✓Modeling whether a customer will renew a subscription.

1. Probit Model

○ What is CFD?

- The **Cumulative Distribution Function (CDF)** tells us the probability that a random variable Z is less than or equal to a particular value z: $\Phi(z)=P(Z\leq z)$
- This is calculated as the area under the standard normal curve from −.
- Example: If $z=0$, $\Phi(0)=0.5$ (exactly half the area of the curve is to the left of 0).
- If $z=1.96$, $\Phi(1.96)\approx 0.975$
- If $z=-1.96$ $\Phi(-1.96)\approx 0.025$ or 2.5%.
- In practice, the CDF converts the value of z into a **probability between 0 and 1**.
- https://www.rit.edu/academicsuccesscenter/sites/rit.edu.academicsuccesscenter/files/documents/math-handouts/Standard%20Normal%20Distribution%20Table.pdf

1. Probit Model

○ How Does the Standard Normal CDF Apply to the Probit Model?

- In the Probit model, the **Standard Normal CDF (Φ) transforms a linear combination of independent variables (X) into a probability that falls between 0 and 1.**
- **The Probit model equation is:** $P(Y = 1/X) = \Phi(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots\cdots + \beta_k X_k)$
- $\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots\cdots + \beta_k X_k$ is the **linear combination** of predictors.
- This value can range from $-\infty$ to $+\infty$, depending on the values of X and the coefficients (β).The **Standard Normal CDF (Φ) maps this range to a probability between 0 and 1**

1. Probit Model

- If the linear combination is very large (e.g., $z=5$), $\Phi(z) \approx 1$, meaning almost certain that $Y=1$.
- If the linear combination is very small (e.g., $z=-5$), $\Phi(z) \approx 0$, meaning almost certain that $Y=0$.
- **Standard Normal CDF:**
- The CDF is the cumulative area under the PDF, producing a curve like this:
- At $z=-\infty$: Probability = 0. At $z=0$: Probability = 0.5. At $z=+\infty$: Probability = 1.

1. Probit Model

- **Probit Model Mapping:**
- If the linear combination of predictors is, for example:
- $z = -1$: $\Phi(-1) = 0.1587$ (15.87% probability).
- $z = 0$: $\Phi(0) = 0.5$ (50% probability).
- $z = 2$: $\Phi(2)=0.9772$ (97.72% probability).

1. Probit Model

- **Example1 :** Suppose we want to predict whether someone will default (Y=1) based on income, income-to-debt ratio and credit score:
- $Default\ probability = \Phi(-4{,}5 + 0{,}12DTI - 0{,}008Credit\ score + 0{,}09Income)$
- If a customer has an income of $50,000 DTI = 30% and Credit score=650, then,
- z = −4.5 +0.12(0,30) − 0.008(650) +0.09(50) = -5.16.
- The computed Z-score is **-5.16**, which is extremely low. Using the standard normal CDF, the default probability **is effectively zero.**
- This means that, given the provided debt-to-income ratio, credit score, and income, the model predicts an **almost negligible probability** of default for this individual.

1. Probit Model

Example 2: The probit model estimates the probability of loan approval based on the **standard normal distribution**.

- **Probit Equation:**
- $P(Y=1) = \Phi(z)$, $z=\beta_0+\beta_1X_1+\beta_2X_2$
- Using the coefficients from E-views output: $\beta_0=-3$, $\beta_1=0.02$, $\beta_2=0.01$
- For an applicant with an income of \$50,000 = 50\$ and a credit score of 700: $z=-3+(0.02)(50)+(0.01)(700)$, $z=-3+1+7=5$ $P(Y=1) = \Phi(5)=0.9997$
- **Interpretation:**
- The probability of loan approval is nearly **100%** for this applicant.

Study case 1: Example: Predicting Internet Subscription Renewal Using a Probit Model

- we use a **Probit model** to predict whether a customer will **renew their internet subscription** based on their characteristics. We have the following variables:
- **Renewal (Y)**: 1 if the customer renews their subscription, 0 otherwise (binary dependent variable).
- **Monthly Bill (X1)**: The amount the customer pays per month (in $).
- **Contract Length (X2)**: The duration of the customer's contract (in months).
- **Customer Support Calls (X3)**: The number of times the customer called customer support in the past 6 months.

Studycase 1: Example: Predicting Internet Subscription Renewal Using a Probit Model

ID	Renewal	Monthly Bill	Contract Length	Customer Support Calls
1	0	50	6	1
2	0	70	12	5
3	1	40	24	0
4	1	80	36	7
5	0	55	6	2
6	1	45	12	1
7	1	90	24	8
8	0	60	36	3
9	1	85	6	6
10	0	50	24	0

Probit E-views output

Dependent Variable: RENEWAL
Method: ML - Binary Probit (Newton-Raphson / Marquardt steps)
Date: 02/27/25 Time: 14:19
Sample: 1 10
Included observations: 10
Convergence achieved after 3 iterations
Coefficient covariance computed using observed Hessian

Variable	Coefficient	Std. Error	z-Statistic	Prob.
MONTHLY_BILL	-0.096004	0.128387	-0.747772	0.4546
CUSTOMER_SUPPORT_CAL	0.762832	0.806399	0.945973	0.3442
CONTRACT_LENGTH	0.002873	0.041498	0.069238	0.9448
C	3.431338	5.704726	0.601491	0.5475

McFadden R-squared	0.166051	Mean dependent var	0.500000
S.D. dependent var	0.527046	S.E. of regression	0.584595
Akaike info criterion	1.956099	Sum squared resid	2.050509
Schwarz criterion	2.077133	Log likelihood	-5.780494
Hannan-Quinn criter.	1.823325	Deviance	11.56099
Restr. deviance	13.86294	Restr. log likelihood	-6.931472
LR statistic	2.301956	Avg. log likelihood	-0.578049
Prob(LR statistic)	0.512146		

Obs with Dep=0	5	Total obs	10
Obs with Dep=1	5		

Interpretation

- The **Probit model** estimates the probability of renewal using the cumulative standard normal distribution function:
- $P(Y = 1) = \Phi(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3)$
- $P(Y = 1) = \Phi(\beta_0 + \beta_1 Bill + \beta_2 Lenght + \beta_3 call.sup)$
- $P(Y = 1) = \Phi(3{,}43 - 0{,}096 Bill + 0{,}003 Lenght + 0{,}763 call.sup)$
- **Steps for Interpretation……Examine Coefficients:**
  - A positive coefficient means the variable increases the probability of adoption.
  - A negative coefficient means the variable decreases the probability.

Interpretation

- **Intercept (3.43)** means the probability of renewal is 0.9997 or 99.97% when all explanatory variables are zero.
- **Bill (-0.096), Negative sign**: As the monthly bill increases, the probability of renewal decreases.it’s effect is medium.
- **Length (+0.003) Positive sign**: A longer contract length slightly increases the probability of renewal. The effect is small because 0.003 is close to 0.
- **Call Support (+0.763),** Positive sign: Customers who get a call support have a higher probability of renewal. Since 0.763 is relatively large, it has a stronger impact than the other variables.
- because satisfied customers who received help tend to stay.

Prediction

- Let’s compute a few cases:
- **Customer A:**
- **Bill = 50, Length = 12, Call.Sup = 1**
- **Latent Score**: $3.43 - (0.096 \times 50) + (0.003 \times 12) + (0.763 \times 1) = -0.571$
  Probability: $P(Y=1) = \Phi(-0.571) \approx 0.284$
- **Interpretation**: This customer has a **28.4% probability** of renewing.
- **Customer B:**
- **Bill = 20, Length = 24, Call.Sup = 1**
- **Latent Score**: $3.43 - (0.096 \times 20) + (0.003 \times 24) + (0.763 \times 1) = 2.345$
- **Probability**: $P(Y=1) = \Phi(2.345) \approx 0.99$
- **Interpretation**: This customer has a **99% probability** of renewing.

Interpretation

- **Higher Bills → Lower Probability of Renewal**
- **Longer Contracts → Slightly Higher Probability of Renewal**
- **Contacting Support → Stronger Influence on Renewal**

Studycase 2: Adoption of Renewable Energy Technology

- What factors influence a household's decision to adopt renewable energy technology (e.g., solar panels)?
- **Dependent Variable (Binary):**
- **Adoption** (Y):
  - Y=1: The household has adopted renewable energy technology.
  - Y=0: The household has not adopted renewable energy technology.
- **Independent Variables:**

1. **Income** (X_1): Annual household income (in $1,000s).
2. **Education Level** (X_2): Number of years of education of the household head.
3. **Environmental Awareness** (X_3): A score (1 to 10) measuring the household's awareness of environmental issues.
4. **Electricity Costs** (X_4): Monthly electricity expenditure (in $100s).

Example Study: Adoption of Renewable Energy Technology

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-2.5	0.7	-3.57	0.0004
Income	0.05	0.02	2.5	0.012
Education	0.10	0.03	3.33	0.001
Awareness	0.20	0.05	4.00	0.000
Costs	0.15	0.04	-3.75	0.0002

Example Study: Adoption of Renewable Energy Technology

- **Interpretation:**

- **Income**: A $1,000 increase in income increases the probability of adoption (significant at 5% level).
- **Education**: Higher education significantly increases the probability.
- **Awareness**: Greater environmental awareness strongly drives adoption.
- **Costs**: Higher electricity costs increase the likelihood of adoption.