Binary Response Models

(GB: Chapter 15.1–15.3; UGB: Chapter 17.1)

Isabel Casas

SYDDANSK UNIVERSITET

Binary Response Models

- The Linear Probability Model
  - Properties
  - Estimation
  - Problems
- Index Models:
  - Two popular versions: Logit/Probit, and their interpretations
  - Motivations of these models

Binary models

In many cases, the response variable, $\mathbf{y}$, has only a finite and discrete number of outcomes.

Examples:

- Accept/reject a job offer
- Choice of transportation mode (train, bus, car/ bicycle)
- Introduce/not-introduce new software system
- Being unemployed, having a job/ out of the labour force
- Married/not married
- Eat out/eat at home
- Attend this class/don't attend

Binary response

We shall focus (for now) on the case with two outcomes $\Rightarrow$

A binary dependent variable:

$$\begin{aligned} y &= 1 \text{ (called the ''successful'' outcome)} \\ y &= 0 \text{ (the ''failure'')} \end{aligned}$$

Examples:

- $y = 1$ if the person accepts a job; $y = 0$ if she doesn't
- $y = 1$ if eating out; $y = 0$ if eating at home.

Thus, we define in each case what a ''success'' is.

The Linear Probability Model

$$\mathbf{y} = G(\mathbf{X}\boldsymbol{\beta}) + \boldsymbol{\epsilon} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

$\mathbf{y}$ is normal and the link function is the identity.

We are estimating the conditional mean of $\mathbf{y}$ given $\mathbf{X}$, i.e.

$$E(\mathbf{y}|\mathbf{X}) = P\left(y = 1\,|\mathbf{X}\right) = \mathbf{X}\boldsymbol{\beta}$$

which is a probability, then:

$\boldsymbol{\beta}$ must satisfy $0 \leq \mathbf{x}_j\boldsymbol{\beta} \leq 1$ for all $j = 1, \ldots, n$.

Partial effects of $\mathbf{x}_i$ in the LPM

How do we interpret β_i?

- It is not interpret as the ratio by which $E(\mathbf{y}|\mathbf{X})$ changes when $\mathbf{x}_i$ changes by one unit (holding other factors fixed).
- Because $\mathbf{y}$ is 0 or 1, then $E(\mathbf{y}|\mathbf{X}) \in [0, 1]$
- β_i measures the change in the probability of success when $\mathbf{x}_i$ changes by one unit (holding other factors fixed).
- For example, the probability of a woman as part of the labour force increases by 0.038 for each year of education added.

Partial effects of $\mathbf{x}_i$ in the LPM

1. If $\mathbf{x}_i$ is continuous:

$$\frac{\partial P\left(y=1\,|\mathbf{X}\right)}{\partial \mathbf{x}_i}=\frac{\partial \mathbf{X}\boldsymbol{\beta}}{\partial \mathbf{x}_i}$$

2. If $\mathbf{x}_i$ is a dummy $(x_{ij}=0,1)$ or a discrete variable $(x_{ij}=0,1,2,3,\ldots)$:

$$P\left(y=1\,|\mathbf{x}_1,\ldots,\mathbf{x}_{i-1},1,\mathbf{x}_{i+1},\ldots,\mathbf{x}_k\right)-$$
$$P\left(y=1\,|\mathbf{x}_1,\ldots,\mathbf{x}_{i-1},0,\mathbf{x}_{i+1},\ldots,\mathbf{x}_k\right)$$

Questions: The effect of $\mathbf{x}_i$ is constant. Is that reasonable? If $\mathbf{x}_i=age$ and age^2 is also in the model, what is the partial effect of changing age?

Disadvantages of LPM

The observed value of $\mathbf{y}$ is then:

$$\mathbf{y} = P\left(y = 1 \,|\mathbf{X}\right) + \boldsymbol{\epsilon} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

It is linear, so use OLS!

The error term $\boldsymbol{\epsilon}$ is given by:

$$\begin{aligned} \epsilon_j &= 1 - \mathbf{x}_j\boldsymbol{\beta} \qquad \text{with probability } \mathbf{x}_j\boldsymbol{\beta} \\ \epsilon_j &= -\mathbf{x}_j\boldsymbol{\beta} \qquad \text{with probability } 1 - \mathbf{x}_j\boldsymbol{\beta} \end{aligned}$$

- The errors are not normally distributed.
- Their distribution depends on $\mathbf{X}$ and they are heteroskedastic.
- The OLS estimator is unbiased because $E(\boldsymbol{\epsilon}) = 0$ but it is not efficient.

Disadvantages of LPM

- The estimates of the probability of success $(X\hat{\boldsymbol{\beta}})$ can lie outside $[0, 1]$. How to interpret that?
- This occurs when $\mathbf{x}_j\boldsymbol{\beta}$ are not in [0,1] for all j
- Heteroscedasticity:

$$Var\left(\boldsymbol{\epsilon}\,|\mathbf{X}\right) = Var\left(\mathbf{y}\,|\mathbf{X}\right) = \mathbf{X}\boldsymbol{\beta}\left(1 - \mathbf{X}\boldsymbol{\beta}\right)$$

i.e. OLS.3 is violated. What does this mean?

LPM and heteroscedasticity: first solution

Obtain the OLS estimator but use heteroscedasticity-robust standard errors for the hypothesis testing:

- To test individual effects
  - t–test with robust standard errors
  - In R, $coeftest(model, vcov = vcovHC(model, type = "HC0"))$
- To test the joint effect of a set of variables
  - Wald and LM tests (robust versions) can then be used
  - In R, $wald.test(model, model_{res}, vcov = vcovHC(model, type = "HC0"))$
  - The LM test has to be programmed by yourself
  - For the particular case: $H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$ we can use the classical F-test.

Syddansk Universitet

LPM and heteroscedasticity: second solution

- Use Weighted Least Squares (WLS)
- We know the form of the heteroscedasticity:
  $Var\left(\boldsymbol{\epsilon}\,|\mathbf{X}\right) = \mathbf{X}\boldsymbol{\beta}\left(1 - \mathbf{X}\boldsymbol{\beta}\right).$
- If we divide all variables by this $\Rightarrow$ transformed model is homokedastic:

$$\frac{Y}{\sqrt{Var\left(\boldsymbol{\epsilon}\,|\mathbf{X}\right)}} = \frac{X}{\sqrt{Var\left(\boldsymbol{\epsilon}\,|\mathbf{X}\right)}}\boldsymbol{\beta} + \frac{\boldsymbol{\epsilon}}{\sqrt{Var\left(\boldsymbol{\epsilon}\,|\mathbf{X}\right)}}$$

$$\Downarrow$$

$$\tilde{Y} = \tilde{X}\boldsymbol{\beta} + \tilde{\boldsymbol{\epsilon}}$$

Syddansk Universitet

LPM and heteroscedasticity: second solution

- The transformed model is homokedastic

$$
\begin{aligned}
Var(\tilde{\boldsymbol{\epsilon}}\,|\mathbf{X}) &= Var\left(\left. \frac{\boldsymbol{\epsilon}}{\sqrt{Var\left(\boldsymbol{\epsilon}|X\right)}} \right| X \right) \\
&= \frac{1}{Var\left(\boldsymbol{\epsilon}\,|\mathbf{X}\right)} Var\left(\boldsymbol{\epsilon}|X\right) = 1
\end{aligned}
$$

How does WLS it work in practice?

WLS in practice

1. Run OLS on original model: $\mathbf{y}$ on $\mathbf{X}$ and obtain the fitted values $\hat{\mathbf{y}}$
2. Construct estimate of conditional variance of $\mathbf{y}$ (and $\boldsymbol{\epsilon}$) for each jth observation:

$$\hat{\sigma}_j^2 = \mathbf{x}_j \hat{\boldsymbol{\beta}} \left(1 - \mathbf{x}_j \hat{\boldsymbol{\beta}}\right) = \hat{y}_j \left(1 - \hat{y}_j\right)$$

3. Transform variables:

$$\tilde{y}_j = y_j / \hat{\sigma}_j \qquad \text{and} \qquad \tilde{\mathbf{x}}_j = \mathbf{x}_j / \hat{\sigma}_j$$

4. Regress $\tilde{y}_j$ on $\tilde{\mathbf{x}}_j$ using OLS.

Problem: What if $\tilde{y}_j > 1$ or $\tilde{y}_j < 0$? I.e. predicted probability is outside the unit interval.

Example 15.1 Wooldridge (15 minutes)

Consider the model for woman's labor market participation (mroz.dat and mroz.des):

$$P(\text{inlf} = 1\,|\mathbf{X}\,) = \beta_1 + \beta_2 \text{ nwifeinc} + \beta_3 \text{ educ} + \beta_4 \text{ exper} + \beta_5 \text{ exper}^2 + \beta_6 \text{ age} + \beta_7 \text{ kidslt6} + \beta_8 \text{ kidsge6}$$

- Estimate by LPM
- LPM with heteroskedasticity-robust se
- WLS
- Interpret the results.
- How should we interpret the parameter β_2? Is the effect of $nwifeinc$ important?
- What about $kidslt6$?
- Are there any fitted probabilities outside [0,1]?
- Do the robust se are any different?

Introduction

- Generalised Linear Models
- Index Models

Generalised Linear Models (GLM)

Generalised linear models are an extension of the MLR allowing for:

- non–normal distribution of the response $\mathbf{y}$
- non–linear relationship between the $E(\mathbf{y}|\mathbf{X})$ and $\mathbf{X}$

The GLM unify the MLR, Logit and Probit models... and more.

Generalised Linear Models (GLM)

The basic elements of the GLMs are:

$$\mathbf{y} = G(\mathbf{X}\boldsymbol{\beta}) + \boldsymbol{\epsilon}, \quad E(\boldsymbol{\epsilon}|\mathbf{X}) = 0$$

1. $\mathbf{y}$ follows an exponential distribution
   - Normal, Exponential, Gamma, Poisson and Binomial
2. The mean value $E(y_j|\mathbf{X}) = G(\mathbf{x}_j\boldsymbol{\beta}) = \mu_j$ is a function of the *linear predictor* η_j which is:

$$\eta_j = \beta_0 + \beta_1 x_{1j} + \ldots + \beta_k x_{kj}$$

3. The relationship between μ_j and η_j is defined by the *link function*, g: $\eta_j = g(\mu_j)$ or $\mu_j = g^{-1}(\eta_j) = G(\eta_j)$
   - The link function is smooth and invertible
4. We are interested in estimating $G(\mathbf{X}\boldsymbol{\beta})$ (the conditional mean of $\mathbf{y}$)

Example

- Given a MLR: $y_j = \beta_0 + \beta_1 x_{1j} + \ldots + \beta_k x_{kj} + \epsilon_j$
- If $\epsilon_j \sim N(0, \sigma^2) \Rightarrow y_j \sim N(\mu_j, \sigma^2)$ - normal response
- The conditional mean is: $\mu_j = \beta_0 + \beta_1 x_{1j} + \ldots + \beta_k x_{kj}$
- Clearly the link function $g(u) = u$ is the identity function, $g^{-1}(u) = G(u) = u$

$$\Downarrow$$

The mean of MLR

$$E(y_j|X) = \mu_j = g^{-1}(\eta_j) = G(\eta_j) = \eta_j = \beta_0 + \beta_1 x_{1j} + \ldots + \beta_k x_{kj}$$

$$\Downarrow$$

The MLR is a type of GLM with a normal response and the identity as link function

SYDDANSK UNIVERSITET

Links for responses in GLM

- The combination of the distribution of the response $\mathbf{y}$ and a link function is called the *family* of the GLM.
- In the example, the combination of normal with the identity function is the MLR
- Each response distribution can be combined with a variety of links
- Only certain combinations are natural or canonical

Links for responses in GLM

Response/error distribution	Canonical link	Mean
Normal/Normal	identity: $g(\mu) = \mu$	$G(\eta) = \eta$
Binomial/Normal	probit: $g(\mu) = \Phi^{-1}(\mu)$	$G(\eta) = \Phi(\eta)$
Binomial/logistic	logit: $g(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$	$G(\eta) = \frac{e^{\eta}}{1+e^{\eta}}$
Poisson	log: $g(\mu) = \log(\mu)$	
Gamma	inverse: $g(\mu) = \frac{1}{\mu}$	

Exercise (3 minutes)

- Assume y_j is a discrete binary variable that takes values 0 and 1 only. What is the expression of $\mu_j = E(y_j|\mathbf{X})$?
- Assume the *logit* function as the link function:

$$g(u) = \log\left(\frac{u}{1-u}\right)$$

- Is this function g smooth? and invertible?
- Find the inverse, $\mu_j = g^{-1}(\eta_j)$. Hint: write the function as $\eta_j = \ldots$
- The inverse is the logistic distribution

SYDDANSK UNIVERSITET

How do we define a model for $\mathbf{y}$

We are modelling the conditional mean:

$$E(\mathbf{y}\,|\mathbf{X}) = E(\mathbf{y}\,|\mathbf{x}_1, \ldots, \mathbf{x}_k) = G(\mathbf{X}\boldsymbol{\beta})$$

which for a binary $\mathbf{y}$ is:

$$E\,(\mathbf{y}\,|\mathbf{X}) = 1 \cdot P\,(y = 1\,|\mathbf{X}) + 0 \cdot P\,(y = 0\,|\mathbf{X}) = P\,(y = 1\,|\mathbf{X})$$

Therefore, we are estimating the probability of success:

$$P(y = 1\,|\mathbf{X}) = P(y = 1\,|\mathbf{x}_1, \ldots, \mathbf{x}_k)$$

and the effect of $\mathbf{x}_i$ on this probability.

Binary Model

As we have seen the probability of success depends on the values of $\mathbf{X}$.

$$P(y = 1\,|\mathbf{X}\,) = E\,(\mathbf{y}\,|\mathbf{X}\,) = G(\mathbf{X}\boldsymbol{\beta})$$

Show:

$$Var\,(\mathbf{y}\,|\mathbf{X}\,) = G(\mathbf{X}\boldsymbol{\beta})(1 - G(\mathbf{X}\boldsymbol{\beta}))$$

Hint: $Var\,(\mathbf{y}\,|\mathbf{X}\,) = E\left(Y^2\,|\mathbf{X}\right) - [E\,(\mathbf{y}\,|\mathbf{X}\,)]^2$

How to estimate $P\left(y=1\,|\mathbf{X}\right)=E\left(\mathbf{y}\,|\mathbf{X}\right)$?

1. Use OLS on the model with a binary response.
   - the Linear Probability Model (LPM)
   - Simple model, but it has a number of problems: the most important is that fitted values might be <0 or >1.
2. Alternative: The "index models"
   - Logit model: $G(\mathbf{X}\boldsymbol{\beta})=\dfrac{e^{\mathbf{X}\boldsymbol{\beta}}}{1+e^{\mathbf{X}\boldsymbol{\beta}}}$
   - Probit model: $G(\mathbf{X}\boldsymbol{\beta})=\Phi(\mathbf{X}\boldsymbol{\beta})$

Index Models

Main problem of the LPM:

- $P(y = 1 \,|\mathbf{X}\,) = \mathbf{X}\boldsymbol{\beta}$ can take values outside unit interval
- Hard to interpret
- Creates problems for WLS

Solution: Use an index model:

$$P(y = 1 \,|\mathbf{X}\,) = G(\mathbf{X}\boldsymbol{\beta})$$

where G is function of $\mathbf{X}\boldsymbol{\beta}$ that takes values between 0 and 1, for example a c.d.f.

Index Models

Two popular choices of G:

- $G(\mathbf{X}\boldsymbol{\beta}) = \Lambda(\mathbf{X}\boldsymbol{\beta}) = \frac{\exp(\mathbf{X}\boldsymbol{\beta})}{1+\exp(\mathbf{X}\boldsymbol{\beta})} \Rightarrow$ logit model
- $G(\mathbf{X}\boldsymbol{\beta}) = \Phi(\mathbf{X}\boldsymbol{\beta}) = \int_{-\infty}^{\mathbf{X}\boldsymbol{\beta}} \phi\left(z\right) dz \Rightarrow$ probit model

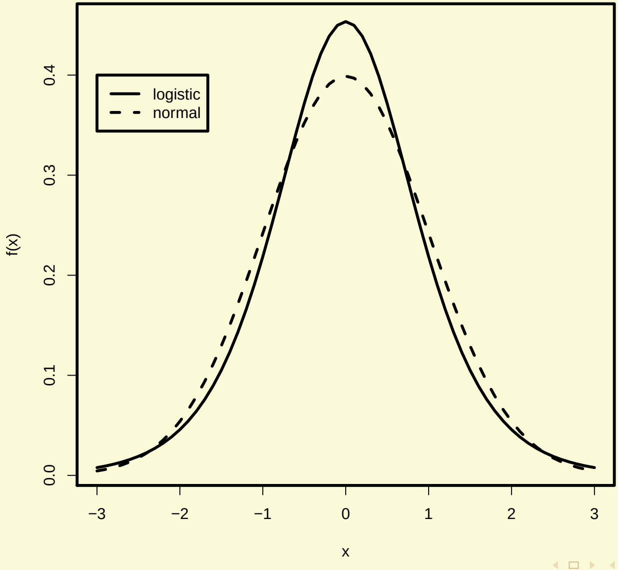
Instead of $G(\mathbf{X}\boldsymbol{\beta}) = \mathbf{X}\boldsymbol{\beta} \Rightarrow$ PLM

SYDDANSK UNIVERSITET

Exercise(3 minutes)

- If $\mathbf{X}\boldsymbol{\beta} = \beta_0 + \beta_1\mathbf{x}_1 + \beta_2\mathbf{x}_2$
- Write the probit and logit models $P(y = 1|\mathbf{X}) =$?
- If the estimate $\hat{\boldsymbol{\beta}}^{logit} = (0.2, 0.5, 1)'$ What is the estimated probability of success for the individual with $(\mathbf{x}_1 = 2, \mathbf{x}_2 = 3)$?
- If the estimate $\hat{\boldsymbol{\beta}}^{probit} = (0.125, 0.3, 0.6)'$ What is the estimated probability of success for $(\mathbf{x}_1 = 2, \mathbf{x}_2 = 3)$?

Comparison of probit and logit G



Logit

The variable $\mathbf{y}$ has values 0 and 1, then $E(\mathbf{y}|\mathbf{X}) = G(\mathbf{X}\boldsymbol{\beta}) \in [0,1]$.

Choose G as a cumulative distribution function, then this is satisfied. For example:

$$G(\mathbf{X}\boldsymbol{\beta}) = \frac{e^{\mathbf{X}\boldsymbol{\beta}}}{1 + e^{\mathbf{X}\boldsymbol{\beta}}},$$

the logistic cumulative distribution function.

In R:

$$my.model = glm(y \sim \mathbf{x}_1 + \mathbf{x}_2 + .., data = mydata, family = binomial)$$

Probit

A popular choice is

$$G(\mathbf{X}\boldsymbol{\beta}) = \Phi(\mathbf{X}\boldsymbol{\beta})$$

because the normal cumulative density function is from the exponential family and behaves well.

$$\Downarrow$$

Probit Model

In R:

$$my.model = glm(y \sim \mathbf{x}_1 + \mathbf{x}_2 + .., data = mydata, \\ family = binomial("probit"))$$

SYDDANSK UNIVERSITET

Marginal effects of $\mathbf{x}_i$ on probabilities

$\mathbf{x}_j\boldsymbol{\beta}$ for $j = 1, \ldots n$ is the strength of the stimulus for the outcome $y_j = 1$ because

$$P(y_j = 1) = G(\mathbf{x}_j\boldsymbol{\beta}) \to 1 \quad \text{if} \quad \mathbf{x}_j\boldsymbol{\beta} \to \infty$$
$$P(y_j = 1) = G(\mathbf{x}_j\boldsymbol{\beta}) \to 0 \quad \text{if} \quad \mathbf{x}_j\boldsymbol{\beta} \to -\infty$$

- All the values are in the [0,1] interval
- Large $\mathbf{X}$ will imply a high probability of success
- Small $\mathbf{X}$ will imply a low probabily of success

Marginal effects of $\mathbf{x}_i$ on probabilities

The marginal effect of a linear continuous $\mathbf{x}_i$ is

$$\frac{\partial P(y=1|\mathbf{X})}{\partial \mathbf{x}_i} = G'(\mathbf{X}\boldsymbol{\beta})\frac{\partial \mathbf{X}\boldsymbol{\beta}}{\partial \mathbf{x}_i}$$

- The effect depends on our sample $\mathbf{X}$, non-constant.
- For the logit and probit, $G'(u) > 0$ for all u, therefore the sign of the parameter estimate is the sign of the effect.
- This means that the largest effect of $\mathbf{x}_i$ occurs for observations with $\mathbf{X}\boldsymbol{\beta}$ around the mean of the distribution G (usually zero).
- The relative effect is the same than the OLS relative effect, independent of the sample,

$$\frac{\partial P(y=1|\mathbf{X})/\partial \mathbf{x}_i}{\partial P(y=1|\mathbf{X})/\partial \mathbf{x}_k} = \frac{\beta_i}{\beta_k}$$

Marginal effects of $\mathbf{x}_i$ on probabilities

Remember how to calculate the derivative of G. This is useful for the nonlinear continuous variables:

$$\frac{\partial(G(\mathbf{X}\boldsymbol{\beta}))}{\partial \mathbf{x}_i} = G'(\mathbf{X}\boldsymbol{\beta}) \cdot \frac{\partial(\mathbf{X}\boldsymbol{\beta}))}{\partial \mathbf{x}_i}$$

For example, if we have the fitted model:

$$P(y=1|\mathbf{X}) = G(\beta_0 + \beta_1 age + \beta_2 age^2 + \beta_3 educ)$$

What is $\frac{\partial(G(\mathbf{X}\boldsymbol{\beta}))}{\partial age}$?

Q: Think about the marginal effect of interaction variables

Marginal effects of $\mathbf{x}_i$ on probabilities

The marginal effect of a binary $\mathbf{x}_i$ is

$$G(\beta_0 + \beta_1 \mathbf{x}_1 + \beta_{i-1}\mathbf{x}_{i-1} + \beta_i + \beta_{i+1}\mathbf{x}_{i+1} + \ldots + \beta_k \mathbf{x}_k)$$
$$- G(\beta_0 + \beta_1 \mathbf{x}_1 + \beta_{i-1}\mathbf{x}_{i-1} + 0 + \beta_{i+1}\mathbf{x}_{i+1} + \ldots + \beta_k \mathbf{x}_k)$$

- Change on the probability of success when we go from one event to other
- The sign of the parameter will be the sign of the effect
- Ex. if $\mathbf{x}_i = 1$ for male and the dependent variable is an employment indicator, the marginal effect will measure the change

Marginal effects of $\mathbf{x}_i$ on probabilities

The marginal effect of a discrete $\mathbf{x}_i$ (for example number of kids) is:

$$G(\beta_0 + \beta_1\mathbf{x}_1 + \beta_{i-1}\mathbf{x}_{i-1} + \beta_i m + \beta_{i+1}\mathbf{x}_{i+1} + \ldots + \beta_k\mathbf{x}_k)$$
$$- G(\beta_0 + \beta_1\mathbf{x}_1 + \beta_{i-1}\mathbf{x}_{i-1} + n + \beta_{i+1}\mathbf{x}_{i+1} + \ldots + \beta_k\mathbf{x}_k)$$

- Change on the probability of success when going from n to m kids

Average Marginal Effects of $\mathbf{x}_i$

The marginal effects has n values, how do we report this?

- Partial effect at average (PEA). Take $\bar{\mathbf{X}}$ and report $G'(\bar{\mathbf{X}}\beta) \cdot \frac{\partial(\bar{\mathbf{X}}\boldsymbol{\beta}))}{\partial \mathbf{x}_i}$
- Problems if any value if $\mathbf{X}$ is nonlinear such as $\log(income)$ or age^2 because then we are using $\log(in\bar{c}ome)$ and $a\bar{g}e^2$ instead of $\log(inc\bar{o}me)$ and $a\bar{g}e^2$
- Problems also using the average of dummy variables
- Instead use APE, median, quantiles or a particular type of person

Average Marginal Effects of $\mathbf{x}_i$

What is the APE?

$$\text{marginal\_effect}_{ij} = G'(\mathbf{x}_j\boldsymbol{\beta})\frac{\partial(\mathbf{X}\boldsymbol{\beta}))}{\partial\mathbf{x}_i}$$

This called Average Partial Effect of $\mathbf{x}_i$ or Mean Marginal Effect:

$$\frac{1}{n}\sum_{j=1}^{n}\text{marginal\_effect}_{ij}$$

We can report the marginal effect of the average value of $\mathbf{x}_i$

Marginal effects of $\mathbf{x}_i$

- The logit and probit have both symmetric density functions (G') but as we saw the logit has fatter tails and larger mean.
- If $\mathbf{x}_j\boldsymbol{\beta} = 0$ for a given observation j:
  - Logit marginal effect of $\mathbf{x}_i$= 0.25 *β_i^{logit}
  - Probit marginal effect of $\mathbf{x}_i$= 0.40*β_i^{probit}
- $\beta_i^{logit} \approx 1.6\beta_i^{probit}$ will result in the same marginal effects around $\mathbf{X}\boldsymbol{\beta}$.
- This is a good check of your results which should hold if you run both models

Comparison of probit and logit models

- Both the logistic and standard normal densities have mean zero, both are unimodal and symmetric.
- The standard deviation is constant, 1 for the normal and $\pi/\sqrt{3} \approx 1.8$ for the logistic. (graphics)
- Compared to the probit model, the logit model has marginal effects that are somewhat larger around the mean and in the tails
- But smaller marginal effects in between the mean and the tails
- What model to choose? No strong reasons for one of the other
- In the logit model, the cumulative distribution has an analytical closed-form

Exercise (15 minutes)

Using the same data set than before, go through Example 15.1 (Married Women's Labor Force Participation) with the probit and logit... compare with the OLS results.

- Do model.probit and model.logit
- Compare the signs and statistical significance of coefficients
- Divide the coefficients of the logit by 4 and the probit by 2.5 to compare them with the LPM coefficients
- $fitted(model)$
- How many estimates $fitted(y) > 0.5$
- Create a prediction variable $y_2 = 1$ if fitted is greater than 0.5

Application of Index Models in Economics

- Latent variable problem
- Random utility model

Latent variable model

Assume the following model for an underlying (unobserved) variable:

$$\mathbf{y}^* = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

Note: The latent variable is linear and possibly continuous

$$\begin{aligned} y &= 1 \quad \text{if} \quad y^* > 0 \\ y &= 0 \quad \text{if} \quad y^* < 0 \end{aligned}$$

We have the binary model

$$\mathbf{y} = G(\mathbf{X}\boldsymbol{\beta}) + \boldsymbol{\epsilon}$$

SYDDANSK UNIVERSITET

Example

$\mathbf{y}^*$ is the Net Present Value (NPV) of an investment.

NPV is an indicator of how much value an investment or project adds to the firm. We would like to decide whether a certain investment should be carried out.

$\mathbf{y}^*$	$\mathbf{y}$
NPV > 0	1
NPV ≤ 0	0

We want to estimate the probability of success $y = 1$ after knowing certain properties of the investment which are expressed in $\mathbf{X}$

Latent variable model

$$\mathbf{y} = G(\mathbf{X}\boldsymbol{\beta}) + \boldsymbol{\nu}$$

where:

- $E(\boldsymbol{\epsilon}|\mathbf{X}) = 0$
- $\mathbf{X}$ is exogenous, independent of $\boldsymbol{\epsilon}$
- The c.d.f G is from an exponential family and symmetric around zero $\Rightarrow$ $G(\mathbf{X}\boldsymbol{\beta}) = 1 - G(-\mathbf{X}\boldsymbol{\beta})$.

Then:

$$\begin{aligned} P\left(y = 1 \,|\mathbf{X}\right) =& P\left(y^* > 0 \,|X\right) = P\left(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} > 0 \,|\mathbf{X}\right) \\ =& P\left(\boldsymbol{\epsilon} > -\mathbf{X}\boldsymbol{\beta} \,|\mathbf{X}\right) = P\left(\boldsymbol{\epsilon} < \mathbf{X}\boldsymbol{\beta} \,|\mathbf{X}\right) \\ =& G\left(\mathbf{X}\boldsymbol{\beta}\right) \end{aligned}$$

SYDDANSK UNIVERSITET

Latent variable model

Case 1: ϵ has a standard logistic distribution, then:

$$P\left(y=1\left|\mathbf{X}\right.\right)=P\left(\boldsymbol{\epsilon}<\mathbf{X}\beta\left|\mathbf{X}\right.\right)=G\left(\mathbf{X}\boldsymbol{\beta}\right)=\Lambda\left(\mathbf{X}\boldsymbol{\beta}\right)=\frac{\exp\left(\mathbf{X}\boldsymbol{\beta}\right)}{1+\exp\left(\mathbf{X}\boldsymbol{\beta}\right)}$$

and we have the logit model.

Case 2: If $\epsilon \sim N\left(0,1\right)$, then:

$$P\left(y=1\left|\mathbf{X}\right.\right)=P\left(\boldsymbol{\epsilon}<\mathbf{X}\beta\left|\mathbf{X}\right.\right)=G\left(\mathbf{X}\boldsymbol{\beta}\right)=\Phi\left(\mathbf{X}\boldsymbol{\beta}\right)=\int_{-\infty}^{\mathbf{X}\boldsymbol{\beta}}\phi\left(\boldsymbol{\epsilon}\right)d\boldsymbol{\epsilon}$$

and we have the probit model.

Latent variable model

Subcase: If $\epsilon \sim N\left(0, \sigma^2\right)$ where $\sigma \neq 1 \Rightarrow \epsilon/\sigma$ is standard normal, and:

$$\begin{aligned} P\left(y = 1 \left| \mathbf{X} \right.\right) &= P\left(\boldsymbol{\epsilon} < \mathbf{X}\boldsymbol{\beta} \left| X \right.\right) \\ &= P\left(\frac{\boldsymbol{\epsilon}}{\sigma} < X\frac{\boldsymbol{\beta}}{\sigma}\right| X\left.\right) \\ &= \Phi\left(X\frac{\boldsymbol{\beta}}{\sigma}\right) = \Phi\left(X\tilde{\boldsymbol{\beta}}\right) \end{aligned}$$

Again, we have the probit, but this time with parameter $\tilde{\boldsymbol{\beta}} = \boldsymbol{\beta}/\sigma$. Hence, in this case, only $\boldsymbol{\beta}/\sigma$ is identified – not $\boldsymbol{\beta}$

Latent variable model

Intuition:

Consider the following two latent models (where $\boldsymbol{\epsilon} \sim N(0, \sigma^2)$):

$$\begin{aligned} \mathbf{y}^* &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \\ \frac{\mathbf{y}^*}{\sigma} &= \mathbf{X}\frac{\boldsymbol{\beta}}{\sigma} + \frac{\boldsymbol{\epsilon}}{\sigma} \end{aligned}$$

- $y^* > 0 \Leftrightarrow y^*/\sigma > 0$.
- The two models are observationally equivalent (same value $\mathbf{y}$) $\Rightarrow$ We can never know which model is "true".
- The first has parameters $\boldsymbol{\beta}$ and σ. The second has parameters $\boldsymbol{\beta}/\sigma$ and 1 $\Rightarrow$ all we can infer from observing $\mathbf{y}$ is $\boldsymbol{\beta}/\sigma$. To recover $\boldsymbol{\beta}$, we must assume a value for σ.
- We get estimates of $\boldsymbol{\beta}/\sigma$.

Additive random utility model

A consumer chooses between the alternatives 0 and 1 according to which has the higher satisfaction or utility (e.g. choice between car=0 and train=1)

$$\begin{aligned}\mathbf{y}_0^* &= \mathbf{X}\boldsymbol{\alpha}_0 + \boldsymbol{\epsilon}_0 \\ \mathbf{y}_1^* &= \mathbf{X}\boldsymbol{\alpha}_1 + \boldsymbol{\epsilon}_1\end{aligned}$$

- $\mathbf{y}_k^* =$ "utility" of alternative k
- $\mathbf{X} =$ observable individual characteristics, e.g. income, closeness to railway, etc.
- $\boldsymbol{\epsilon}$'s: random components of the utility
- The alternative with higher utility is chosen.

Additive random utility model

Decision rule:

$$\begin{aligned} y_1^* \quad > \quad y_0^* \;\Leftrightarrow\; y_1^* - y_0^* > 0 \;\Rightarrow\; \text{take the train} \;\; (y = 1) \\ y_1^* \quad < \quad y_0^* \;\Leftrightarrow\; y_1^* - y_0^* < 0 \;\Rightarrow\; \text{take the car} \;\; (y = 0) \end{aligned}$$

then:

$$\begin{aligned} y_1^* - y_0^* &= X\alpha_1 + \boldsymbol{\epsilon}_1 - (X\alpha_0 + \boldsymbol{\epsilon}_0) \\ &= X(\alpha_1 - \alpha_0) + (\boldsymbol{\epsilon}_1 - \boldsymbol{\epsilon}_0) \\ &\Downarrow \\ \mathbf{y}^* &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \end{aligned}$$

where:

$$\mathbf{y}^* = \mathbf{y}_1^* - \mathbf{y}_0^*, \quad \boldsymbol{\beta} = \boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_0, \quad \boldsymbol{\epsilon} = \boldsymbol{\epsilon}_1 - \boldsymbol{\epsilon}_0$$

Additive random utility model

Decision rule:

$$
\begin{aligned}
y^* &> 0 \Rightarrow \text{take the train} \Rightarrow y = 1 \\
y^* &< 0 \Rightarrow \text{take the car} \Rightarrow y = 0
\end{aligned}
$$

where:

$$\mathbf{y}^* = (\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_0)\,\mathbf{X} + (\boldsymbol{\epsilon}_1 - \boldsymbol{\epsilon}_0) = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

If $\boldsymbol{\epsilon}_1$ and $\boldsymbol{\epsilon}_0$ are independently normally distributed with mean 0 and variances σ_0^2 and $\sigma_1^2 \Rightarrow \boldsymbol{\epsilon}$ is normal with variance $\sigma^2 = \sigma_0^2 + \sigma_1^2 \Rightarrow \sigma = \sqrt{\sigma_0^2 + \sigma_1^2}$

Additive random utility model

$$
\begin{aligned}
P\left(y=1\,|\mathbf{X}\right)=&P\left(y^{*}>0\,|X\right)=P\left(\boldsymbol{\epsilon}<\mathbf{X}\boldsymbol{\beta}\,|\mathbf{X}\right)\\
=&P\left(\frac{\boldsymbol{\epsilon}}{\sigma}<X\frac{\boldsymbol{\beta}}{\sigma}\middle|X\right)=\Phi\left(X\frac{\boldsymbol{\beta}}{\sigma}\right)=\Phi\left(X\tilde{\boldsymbol{\beta}}\right)
\end{aligned}
$$

where $\tilde{\boldsymbol{\beta}}=\boldsymbol{\beta}/\sigma=(\alpha_1-\alpha_0)/\sqrt{\sigma_1^2+\sigma_0^2}$. We have a probit with parameter vector $\tilde{\boldsymbol{\beta}}$

Summary

A binary dependent variable : $y = 1$ or $y = 0$ which we model by

$$P\left(y = 1 \,|\mathbf{X}\right)$$

using:

A) LPM: $P\left(y = 1 \,|\mathbf{X}\right) = \mathbf{X}\boldsymbol{\beta}$

- Easy to estimate, but hard to interpret
- Problems with heteroscedasticity

B) An index model: $P\left(y = 1 \,|\mathbf{X}\right) = G\left(\mathbf{X}\boldsymbol{\beta}\right) \in (0, 1)$

- Models are non-linear $\Rightarrow$ more complicated partial effects
- Two popular version: Probit $(G = \Phi)$ and Logit $(G = \Lambda)$
- Can be motivated by, e.g, a latent variable model or a random utility model

How do we estimate the index models?