Microeconometrics

Binary and discrete outcomes

Alex Armand

alex.armand@novasbe.pt

Can we use linear models when y is not continuous?

1. y is **discrete**: numeric variable with countable number of values.
   - **Nominal variables**: gender, religious or political affiliation, etc.
   - **Ordinal variables**: income levels, school grades, etc.
   - **Discrete interval variables with few values**: number of times married, number of children, etc.
   - **Continuous variables grouped in categories**: income grouped into subsets, blood pressure levels, etc.
2. y is **binary**: special case with only 2 values
   - Most common case are dummy variables
   - Examples: employed/unemployed, high income/low income

When y is binary

It follows some properties of **Bernoulli (zero-one) random variables**

$$\begin{aligned} E(y|\mathrm{x}) &= P(y=1|\mathrm{x}) = p(\mathrm{x}) \\ Var(y|\mathrm{x}) &= p(\mathrm{x})[1-p(\mathrm{x})] \end{aligned}$$

- We are therefore interested in modelling **conditional probabilities**
  - Correspondent of modelling conditional means (see OLS)
  - $p(\mathrm{x})$ is often called **response probability**
- A binary variable has natural **heteroskedasticity**
  - Conditional variance of y depends on x
  - Except in the special case where $p(\mathrm{x})$ does not depend on x.

Binary variables and partial effects

Similar to the linear case, we are interested in how does x **causes a change in the response probability**?

1. For **continuous** x_j:

$$\frac{\partial p(\mathrm{x})}{\partial x_j}$$

2. For **discrete** x_j: look at changes in the response probability holding other variables fixed

   - Example: if x_K is a training indicator and y is an employment indicator

$$p(x_1, ..., x_{K-1}, 1) - p(x_1, ..., x_{K-1}, 0)$$

   is the effect of training on the employment probability, at given values for the other covariates.

Binary variables and partial effects

1. **Average partial effect (APE)**
   - Weighted average of the partial effects at each observation
     $$E_x\left[\frac{\partial p(x)}{\partial x_j}\right]$$
   - More appeal since it averages partial effects for actual units
2. **Partial effect at the average (PEA)**
   - Partial effect when xs equal their sample average
     $$E_x\left[\frac{\partial p(x)}{\partial x_j}\right]$$

In nonlinear models, the APE and PEA can be very different!

How can we estimate response probabilities?

1. **Linear probability model (LPM)**
   - This is OLS applied to binary outcome variables
   - **Ease of interpretation**: the estimated coefficients give direct estimates of the effects of each x_j on the response probability.
2. **Linear index models**

Linear probability model

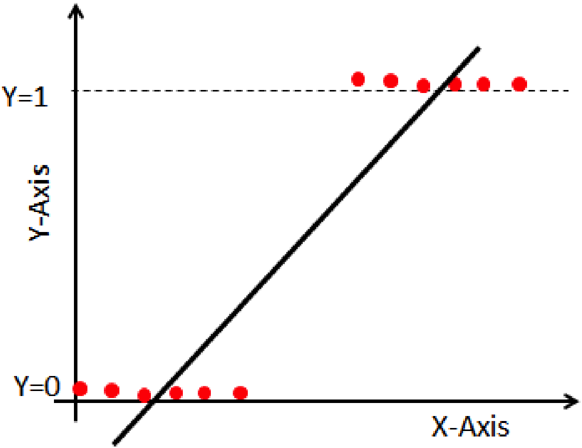
When y is binary, LPM models the response probability as a **function linear in parameters**

$$\begin{aligned} y &= \beta_1 + \beta_2 x_2 + ... + \beta_K x_K + u \equiv \mathrm{x}\beta + u \\ E(y|\mathrm{x}) &= P(y = 1|\mathrm{x}) = \mathrm{x}\beta \end{aligned}$$

- Because $P(y = 1|\mathrm{x}) = E(y|\mathrm{x})$
  - If the conditional mean is truly $\mathrm{x}\beta$, **OLS consistently estimates** β
  - What assumptions required for identification?
- Because $Var(y|\mathrm{x}) = \mathrm{x}\beta(1 - \mathrm{x}\beta)$
  - Inference for OLS should be made robust to **heteroskedasticity**

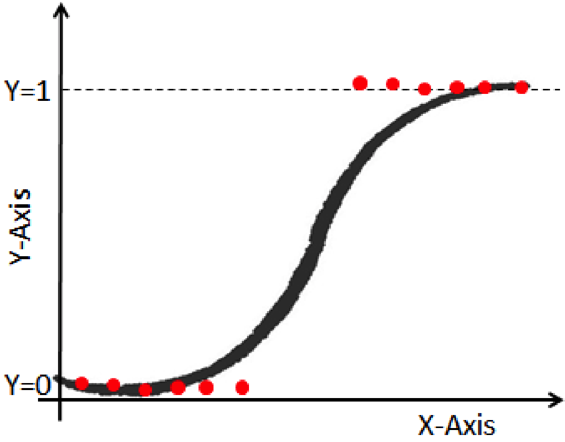
Pros and cons of LPM

1. Nothing guarantees fitted values $\hat{y}_i = \mathrm{x}_i\hat{\beta}$ are in the unit interval.
2. Difficult to impose diminishing effects of the x_j on the $p(\mathrm{x})$



Pros and cons of LPM

1. Nothing guarantees fitted values $\hat{y}_i = \mathrm{x}_i\hat{\beta}$ are in the unit interval.
2. Difficult to impose diminishing effects of the x_j on the $p(\mathrm{x})$



How can we estimate response probabilities?

1. **Linear probability model (LPM)**
   - This is OLS applied to binary outcome variables
   - **Ease of interpretation**: the estimated coefficients give direct estimates of the effects of each x_j on the response probability.
2. **Linear index models**
   - Non-linear models designed to model conditional probabilities
   - Probit / logit models
   - Notice that these models have this name because they use a *linear index* into a *non-linear function*

Linear index models

The response probability has the form

$$P(y = 1|\mathrm{x}) = G(\mathrm{x}\beta)$$

for some $G : \mathbb{R} \to (0, 1)$

1. $\mathrm{x}\beta$ is the **linear index**
2. $G(\cdot)$ is the **non-linear function** transforming the linear index into a real number bounded between 0 and 1
    - In most cases, $G(\cdot)$ is a cumulative distribution function for a continuous random variable with density $g(\cdot)$

Estimating partial effects in linear index models

- If x_j be continuous, the partial effect equals

$$\frac{\partial p(\mathrm{x})}{\partial x_j} = \beta_j g(\mathrm{x}\beta)$$

- Because $g(\mathrm{x}\beta) > 0$, β_j **gives the direction of the partial effect**
- For two continous covariates, the ratio of the coefficients give the ratio of the partial effects, independent of x

$$\frac{\partial p(\mathrm{x})/\partial x_j}{\partial p(\mathrm{x})/\partial x_h} = \frac{\beta_j g(\mathrm{x}\beta)}{\beta_h g(\mathrm{x}\beta)} = \beta_j/\beta_h$$

- No sense to compare magnitudes of coefficients across probit / logit and LPM because their interpretation is totally different

Estimating partial effects in linear index models

1. **Average partial effect**
   - For a continuous x_j: $\widehat{APE}_j = \hat{\beta}_j \left[N^{-1} \sum_{i=1}^{N} g(\mathrm{x}_i \hat{\beta}) \right]$
   - For a discrete x_j:
     $\widehat{APE}_K = N^{-1} \sum_{i=1}^{N} [G(\mathrm{x}_{i(K)} \hat{\beta}_{(K)} + \hat{\beta}_K) - G(\mathrm{x}_{i(K)} \hat{\beta}_{(K)})]$
2. **Partial effect at the average**
   - For a continuous x_j: $\widehat{PEA}_j = \hat{\beta}_j g(\bar{x}\hat{\beta})$
   - For a discrete x_j: $\widehat{PEA}_K = G(\bar{x}_{(K)} \hat{\beta}_{(K)} + \hat{\beta}_K) - G(\bar{x}_{(K)} \hat{\beta}_{(K)})$

Example

- Suppose the **response probability** is equal to

$$p(\mathrm{z}) = G[\beta_0 + \beta_1 z_1 + \beta_2 z_1^2 + \beta_3 \log(z_2) + \beta_4 z_3] \equiv G(\mathrm{x}\beta)$$

- Some examples of **partial effects**

$$\begin{aligned}
\frac{\partial p(\mathrm{z})}{\partial z_1} &= (\beta_1 + 2\beta_2 z_1) g(\mathrm{x}\beta) \\
\frac{\partial p(\mathrm{z})}{\partial z_2} &= \frac{\beta_3}{z_2} g(\mathrm{x}\beta) \\
\frac{\partial \log p(\mathrm{z})}{\partial \log z_2} &= \beta_3 \frac{g(\mathrm{x}\beta)}{G(\mathrm{x}\beta)}
\end{aligned}$$

How to model discrete variables: probit model

Suppose y_i is generated by a **linear latent variable model**:

$$\begin{aligned} y_i^* &= \mathrm{x}_i\theta + e_i \\ & \quad e_i|\mathrm{x}_i \sim \mathcal{N}(0,1) \\ y_i &= 1 \text{ if } y_i^* > 0 \\ &= 0 \text{ if } y_i^* \leq 0 \end{aligned}$$

- The latent variable y_i^* is <u>not observed by the researcher</u>
- **The assumption of standard-normality of e is fundamental**
  - Used to derive the response probability
  - Remember that MLE applications always require distributional assumptions

Probit model and response probability

Derive the **response probability** from the latent model assumptions

$$\begin{aligned} P(y_i = 1|\mathrm{x}_i) &= P(y_i^* > 0|\mathrm{x}_i) \\ &= P(\mathrm{x}_i\theta + e_i > 0|\mathrm{x}_i) \\ &= P(e_i > -\mathrm{x}_i\theta|\mathrm{x}_i) \\ &= 1 - \Phi(-\mathrm{x}_i\theta) \\ &= \Phi\left(\mathrm{x}_i\theta\right) \end{aligned}$$

- $\Phi(z) = \int_{-\infty}^{z} \phi(v)dv$ is the **standard normal c.d.f.**
- $\phi(v) = (2\pi)^{-1/2} \exp(-v^2/2)$ is the **standard normal p.d.f.**

Probit model and response probability

Completely characterized the conditional distribution of y using a well-known c.d.f. function

1. Write the **conditional density of** y

$$
\begin{aligned}
P[y = 1|\mathrm{x}] &= \Phi(\mathrm{x}\theta) \\
P[y = 0|\mathrm{x}] &= 1 - \Phi(\mathrm{x}\theta)
\end{aligned}
$$

$$
\begin{aligned}
f(y|\mathrm{x};\theta) &= [1 - \Phi(\mathrm{x}\theta)]^{(1-y)}\Phi(\mathrm{x}\theta)^{y} \text{ if } y \in \{0,1\} \\
&= 0 \text{ if } y \notin \{0,1\}
\end{aligned}
$$

2. Compute **log-likelihood function**

$$
\ell_i(\theta) = (1 - y_i)\log[1 - \Phi(\mathrm{x}\theta)] + y_i \log[\Phi(\mathrm{x}\theta)]
$$

3. Apply **MLE**   Derivation of parameters   Derivation of asymptotic variance

Probit versus logit

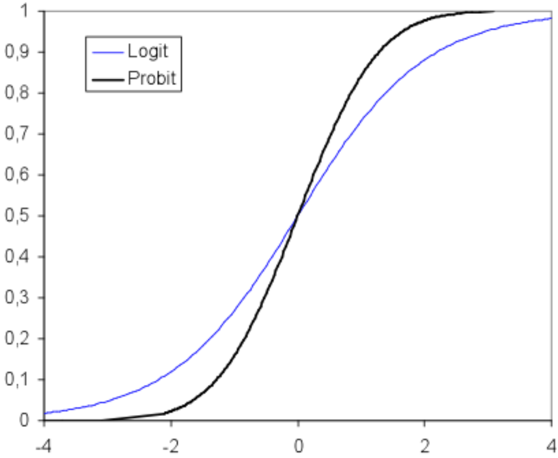
Used for modelling of the tails of the distribution

Probit

$$G(z) = \Phi(z)$$

Logit

$$G(z) = \frac{\exp(z)}{[1 + \exp(z)]}$$



Goodness of fit

Define a **binary predictor** for each i

$$\begin{aligned}\tilde{y}_i &= 1 \text{ if } G(\mathrm{x}_i\hat{\beta}) \geq .5 \\ &= 0 \text{ if } G(\mathrm{x}_i\hat{\beta}) < .5\end{aligned}$$

- Define a **correct prediction** if $y_i = \tilde{y}_i$
  - N_0 (N_1) is the number of observations with $y_i = 0$ ($y_i = 1$)
  - N_{00} (N_{11}) is the number of observations with $y_i = 0$ and $\tilde{y}_i = 0$ ($\tilde{y}_i = 1$ and $y_i = 1$)
- **Proportions correctly predicted** are a measure of goodness of fit

$$q_0 = \frac{N_{00}}{N_0} \qquad q_1 = \frac{N_{11}}{N_1}$$

APPLICATION: Married labor force participation (Mroz 1987)

THE SENSITIVITY OF AN EMPIRICAL MODEL OF MARRIED WOMEN'S HOURS OF WORK TO ECONOMIC AND STATISTICAL ASSUMPTIONS

BY THOMAS A. MROZ[1]

This study undertakes a systematic analysis of several theoretic and statistical assumptions used in many empirical models of female labor supply. Using a single data set (PSID 1975 labor supply data) we are able to replicate most of the range of estimated income and substitution effects found in previous studies in this field. We undertake extensive specification tests and find that most of this range should be rejected due to statistical and model misspecifications. The two most important assumptions appear to be (i) the Tobit assumption used to control for self-selection into the labor force and (ii) exogeneity assumptions on the wife's wage rate and her labor market experience. The Tobit models exaggerate both the income and wage effects. The exogeneity assumptions induce an upwards bias in the estimated wage effect; the bias due to the exogeneity assumption on the wife's labor market experience, however, substantially diminishes when one controls for self-selection into the labor force through the use of unrestricted generalized Tobit procedures. An examination of the maintained assumptions in previous studies further supports these results. These inferences suggest that the small responses to variations in wage rates and nonwife income found here provide a more accurate description of the behavioral responses of working married women than those found in most previous studies.

Dataset

- Focus on the dummy variable for labour force participation
  - What do you notice about the distribution?
  - Can you learn something about probit or logit use?

```
. use mroz

. tab inlf

   =1 if in |
  lab frce, |
       1975 |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |        325       43.16       43.16
          1 |        428       56.84      100.00
------------+-----------------------------------
      Total |        753      100.00
```

Dataset

```
. des nwifeinc educ exper expersq age kidslt6 kidsge6

              storage  display     value
variable name   type   format      label      variable label
------------------------------------------------------------------------
nwifeinc        float  %9.0g                  (faminc - wage*hours)/1000
educ            byte   %9.0g                  years of schooling
exper           byte   %9.0g                  actual labor mkt exper
expersq         int    %9.0g                  exper^2
age             byte   %9.0g                  woman's age in yrs
kidslt6         byte   %9.0g                  # kids < 6 years
kidsge6         byte   %9.0g                  # kids 6-18

. sum nwifeinc educ exper expersq age kidslt6 kidsge6

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
    nwifeinc |       753    20.12896     11.6348  -.0290575         96
        educ |       753    12.28685    2.280246          5         17
       exper |       753    10.63081     8.06913          0         45
     expersq |       753    178.0385    249.6308          0       2025
         age |       753    42.53785    8.072574         30         60
-------------+--------------------------------------------------------
     kidslt6 |       753    .2377158     .523959          0          3
     kidsge6 |       753    1.353254    1.319874          0          8
```

LPM with heteroskedasticity

- Coefficients give direct estimates of effects on response probability.
- These are best interpreted as average partial effects.

```
. reg inlf nwifeinc educ exper expersq age kidslt6 kidsge6, robust

Linear regression                                      Number of obs =     753
                                                       F(  7,   745) =   62.48
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.2642
                                                       Root MSE      =  .42713

------------------------------------------------------------------------------
             |               Robust
        inlf |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nwifeinc |  -.0034052   .0015249    -2.23   0.026    -.0063988   -.0004115
        educ |   .0379953    .007266     5.23   0.000      .023731    .0522596
       exper |   .0394924     .00581     6.80   0.000     .0280864    .0508983
     expersq |  -.0005963     .00019    -3.14   0.002    -.0009693   -.0002233
         age |  -.0160908    .002399    -6.71   0.000    -.0208004   -.0113812
     kidslt6 |  -.2618105   .0317832    -8.24   0.000    -.3242058   -.1994152
     kidsge6 |   .0130122   .0135329     0.96   0.337     -.013555    .0395795
       _cons |   .5855192   .1522599     3.85   0.000     .2866098    .8844287
------------------------------------------------------------------------------
```

Probit estimates

- Probit estimates are much larger in magnitude
- But **probit parameters are not partial effects!**

```
. probit inlf nwifeinc educ exper expersq age kidslt6 kidsge6

Probit regression                         Number of obs   =        753
                                          LR chi2(7)      =     227.14
                                          Prob > chi2     =     0.0000
Log likelihood = -401.30219               Pseudo R2       =     0.2206

------------------------------------------------------------------------------
        inlf |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nwifeinc |  -.0120237   .0048398    -2.48   0.013    -.0215096   -.0025378
        educ |   .1309047   .0252542     5.18   0.000     .0814074     .180402
       exper |   .1233476   .0187164     6.59   0.000     .0866641    .1600311
     expersq |  -.0018871      .0006    -3.15   0.002     -.003063   -.0007111
         age |  -.0528527   .0084772    -6.23   0.000    -.0694678   -.0362376
     kidslt6 |  -.8683285   .1185223    -7.33   0.000    -1.100628    -.636029
     kidsge6 |    .036005   .0434768     0.83   0.408     -.049208    .1212179
       _cons |   .2700768    .508593     0.53   0.595    -.7267473    1.266901
------------------------------------------------------------------------------
```

Partial effects at the average after probit

- To interpret coefficient, we need to focus on partial effects

```
. mfx

Marginal effects after probit
      y  = Pr(inlf) (predict)
         =  .58154201
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
nwifeinc |  -.0046962      .00189   -2.48   0.013  -.008401 -.000991   20.129
    educ |   .0511287      .00986    5.19   0.000   .031805  .070452  12.2869
   exper |   .0481771      .00733    6.57   0.000   .033815  .062539  10.6308
 expersq |  -.0007371      .00023   -3.14   0.002  -.001197 -.000277  178.039
     age |  -.0206432      .00331   -6.24   0.000  -.027127  -.01416  42.5378
 kidslt6 |  -.3391514      .04636   -7.32   0.000  -.430012 -.248291  .237716
 kidsge6 |   .0140628      .01699    0.83   0.408  -.019228  .047353  1.35325
------------------------------------------------------------------------------
```

Average partial effects after probit

- Small differences with partial effects at the average – why?

```
. margeff

Average partial effects after probit
      y  = Pr(inlf)

------------------------------------------------------------------------------
    variable |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nwifeinc |  -.0036162   .0014414    -2.51   0.012    -.0064413   -.0007911
        educ |   .0393088   .0071877     5.47   0.000     .0252212    .0533964
       exper |    .037046    .005131     7.22   0.000     .0269893    .0471026
     expersq |  -.0005675   .0001771    -3.20   0.001    -.0009146   -.0002204
         age |  -.0158917   .0023569    -6.74   0.000     -.020511   -.0112723
     kidslt6 |  -.2441788   .0258995    -9.43   0.000    -.2949409   -.1934167
     kidsge6 |   .0108274   .0130538     0.83   0.407    -.0147576    .0364124
------------------------------------------------------------------------------
```

Logit estimates

```
. logit inlf nwifeinc educ exper expersq age kidslt6 kidsge6

Logistic regression                               Number of obs   =        753
                                                  LR chi2(7)      =     226.22
                                                  Prob > chi2     =     0.0000
Log likelihood = -401.76515                       Pseudo R2       =     0.2197

------------------------------------------------------------------------------
        inlf |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nwifeinc |  -.0213452   .0084214    -2.53   0.011    -.0378509   -.0048394
        educ |   .2211704   .0434396     5.09   0.000     .1360303    .3063105
       exper |   .2058695   .0320569     6.42   0.000     .1430391    .2686999
     expersq |  -.0031541   .0010161    -3.10   0.002    -.0051456   -.0011626
         age |  -.0880244    .014573    -6.04   0.000     -.116587   -.0594618
     kidslt6 |  -1.443354   .2035849    -7.09   0.000    -1.842373   -1.044335
     kidsge6 |   .0601122   .0747897     0.80   0.422     -.086473    .2066974
       _cons |   .4254524   .8603696     0.49   0.621    -1.260841    2.111746
------------------------------------------------------------------------------
```

Average partial effects after logit

- Very similar to the ones after probit – why?

```
. margeff

Average partial effects after logit
      y  = Pr(inlf)

------------------------------------------------------------------------------
    variable |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    nwifeinc |  -.0038118   .0014824    -2.57   0.010    -.0067172   -.0009064
        educ |   .0394323   .0072593     5.43   0.000     .0252044    .0536602
       exper |   .0367123   .0051289     7.16   0.000     .0266598    .0467648
     expersq |  -.0005633   .0001774    -3.18   0.001    -.0009109   -.0002156
         age |  -.0157153   .0023789    -6.61   0.000    -.0203779   -.0110527
     kidslt6 |   -.240805   .0259425    -9.28   0.000    -.2916515   -.1899585
     kidsge6 |   .0107335   .0133282     0.81   0.421    -.0153893    .0368564
------------------------------------------------------------------------------
```

Specification issues in linear index models

There is much confusion about specification issues in probit and logit because **inappropriate parallels are made with linear models**

- Let's look at some cases using the probit model
  - **Omitted variable independent of covariates**
  - **Heteroskedasticity in the latent variable model**
  - **Non-normality in the latent variable model**

Specification issues in linear index models

There is much confusion about specification issues in probit and logit because **inappropriate parallels are made with linear models**

- Let's look at some cases using the probit model
  - **Omitted variable independent of covariates**
  - **Heteroskedasticity in the latent variable model**
  - **Non-normality in the latent variable model**

Omitted variable independent of covariates

- c is an omitted variable independent of x

$$c|\mathrm{x} \sim \mathcal{N}(0, \sigma_c^2)$$

- The **underlying latent variable model** is

$$y^* = \mathrm{x}\beta + c + e, \;\; D(e|\mathrm{x}, c) = \mathcal{N}(0, 1)$$

- The **probit model we estimate** is

$$y^* = \mathrm{x}\beta + v, \;\; D(v|\mathrm{x}) = \mathcal{N}(0, 1)$$

- However notice that assumption we make is wrong

$$D(v|\mathrm{x}) = D(c + e|\mathrm{x}) = Normal(0, \sigma_c^2 + 1)$$

Omitted variable independent of covariates

The correct formulation of the response probability is

$$P(y = 1|\mathrm{x}) = \Phi[\mathrm{x}\beta/(1+\sigma_c^2)^{1/2}]$$

- Using a scaled parameter vector we can write

$$\begin{aligned} \beta_c &\equiv \beta/(1+\sigma_c^2)^{1/2} \\ P(y = 1|\mathrm{x}) &= \Phi(\mathrm{x}\beta_c) \end{aligned}$$

- Probit of y_i on x_i consistently estimates β_c, not β
  - β_c is attenuated toward zero (**attenuation bias**)
  - This would not happen in a linear model

$$E(y|\mathrm{x}) = \mathrm{x}\beta + E(c|\mathrm{x}) = \mathrm{x}\beta$$

Specification issues in linear index models

There is much confusion about specification issues in probit and logit because **inappropriate parallels are made with linear models**

- Let's look at some cases using the probit model
  - **Omitted variable independent of covariates**
  - **Heteroskedasticity in the latent variable model**
  - **Non-normality in the latent variable model**

Heteroskedasticity

- Allow **heteroskedasticity in the error** e by assuming, for example, that the variance of e is a function of the regressor x_1

$$y^* = \mathrm{x}\beta + e,\ \ D(e|\mathrm{x}) = \mathcal{N}(0, \exp(2x_1\delta))$$

- Heteroskedasticity changes the response probability

$$\begin{aligned} P(y &= 1|\mathrm{x}) = P(e > -\mathrm{x}\beta|\mathrm{x}) \\ &= P[\exp(-\mathrm{x}_1\delta)e > -\exp(-\mathrm{x}_1\delta)\mathrm{x}\beta|\mathrm{x}] \\ &= 1 - \Phi[-\exp(-\mathrm{x}_1\delta)\mathrm{x}\beta] = \Phi[\exp(-\mathrm{x}_1\delta)\mathrm{x}\beta] \end{aligned}$$

- Probit of y_i on x_i does not estimate consistently β
  - β and δ can be estimated using the correct response probability
  - $H_0 : \delta = 0$ is a test for homoskedasticity – **why?**

Heteroskedasticity

Notice that partial effects are dependent on heteroskedasticity

- The derivatives and changes in $P(y = 1|\mathrm{x})$ are much more complicated, and need not have the same sign as the relevant coefficient
- For example, if x_K is in x_1

$$
\begin{aligned}
\frac{\partial P(y = 1|\mathrm{x})}{\partial x_K} &= \phi[\exp(-\mathrm{x}_1\delta)\mathrm{x}\beta] \cdot \\
&\quad \{\beta_K \exp(-\mathrm{x}_1\delta) - \delta_K \exp(-\mathrm{x}_1\delta)\mathrm{x}\beta\} \\
&= \phi[\exp(-\mathrm{x}_1\delta)\mathrm{x}\beta] \exp(-\mathrm{x}_1\delta)\{\beta_K - \delta_K \mathrm{x}\beta\}
\end{aligned}
$$

Specification issues in linear index models

There is much confusion about specification issues in probit and logit because **inappropriate parallels are made with linear models**

- Let's look at some cases using the probit model
  - **Omitted variable independent of covariates**
  - **Heteroskedasticity in the latent variable model**
  - **Non-normality in the latent variable model**

Non-normality

Consider the standard latent variable model

$$y^* = \mathrm{x}\beta + e,\ \ D(e|\mathrm{x}) = \mathcal{F}(\cdot)$$

- $F(\cdot) \neq \mathcal{N}(0,1)$ is the cfd of e
- The **response probability** is therefore equal to

$$P(y=1|\mathrm{x}) = 1 - F(-\mathrm{x}\beta) \neq \Phi(\mathrm{x}\beta)$$

- **Probit MLE is not consistent for β if $F(\cdot) \neq \Phi(\cdot)$**
    - Again we need to compute the correct response probability
    - Is normality a strong assumption?

Extention to discrete variables

We can easily extend the results to **multivariate probit or logit models**

- Cases with two or more binary responses to model
  - Call them y_g, where $g = 1, ..., G$ are individual binary response variables for each category.
  - Any combination of zeros and ones is possible
- **Example**: $G = 2$
  - y_1 indicates when a worker has employer-sponsored health insurance
  - y_2 indicates having an employer-sponsored pension plan

Extention to discrete variables: multinomial probit

- The marginal distributions are assumed to follow probits:

$$P(y_g = 1|\mathrm{x}) = \Phi(\mathrm{x}_g \beta_g), g = 1, ..., G.$$

- Multivariate probit can be obtained from

$$\begin{aligned} y_{i1}^* &= \mathrm{x}_{i1}\beta_1 + e_{i1} \\ y_{i2}^* &= \mathrm{x}_{i2}\beta_2 + e_{i2} \\ &\vdots \\ y_{iG}^* &= \mathrm{x}_{iG}\beta_G + e_{iG} \end{aligned}$$

  with $\mathrm{e}_i|\mathrm{x}_i \sim Normal(0, \mathrm{W})$ with unit variances

- Estimation is performed with MLE

EXTRA: probit estimates Back to probit

1. Compute the **gradient** of $\ell_i(\theta)$

$$\begin{aligned}
\nabla_\theta \ell_i(\theta) &= -(1-y_i)\mathrm{x}_i\phi(\mathrm{x}_i\theta)/[1-\Phi(\mathrm{x}_i\theta)] + y_i\mathrm{x}_i\phi(\mathrm{x}_i\theta)/\Phi(\mathrm{x}_i\theta) \\
&= \phi(\mathrm{x}_i\theta)\mathrm{x}_i\frac{-(1-y_i)\Phi(\mathrm{x}_i\theta) + y_i[1-\Phi(\mathrm{x}_i\theta)]}{\Phi(\mathrm{x}_i\theta)[1-\Phi(\mathrm{x}_i\theta)]} \\
&= \phi(\mathrm{x}_i\theta)\mathrm{x}_i\frac{[y_i-\Phi(\mathrm{x}_i\theta)]}{\Phi(\mathrm{x}_i\theta)[1-\Phi(\mathrm{x}_i\theta)]}
\end{aligned}$$

2. Compute the **score** from the gradient

$$\mathrm{s}_i(\theta) = \phi(\mathrm{x}_i\theta)\mathrm{x}_i'\frac{[y_i-\Phi(\mathrm{x}_i\theta)]}{\Phi(\mathrm{x}_i\theta)[1-\Phi(\mathrm{x}_i\theta)]}$$

3. Set **optimality conditions**

$$E[\mathrm{s}_i(\theta_o)|\mathrm{x}_i] = \phi(\mathrm{x}_i\theta_o)\mathrm{x}_i'\frac{[E(y_i|\mathrm{x}_i)-\Phi(\mathrm{x}_i\theta_o)]}{\Phi(\mathrm{x}_i\theta_o)[1-\Phi(\mathrm{x}_i\theta_o)]} = 0$$

EXTRA: probit asymptotic variance

Back to probit

1. Rewrite optimality conditions as

$$E[\mathrm{s}_i(\theta_o)|\mathrm{x}_i] = [E(y_i|\mathrm{x}_i) - \Phi(\mathrm{x}_i\theta_o)]\frac{\phi(\mathrm{x}_i\theta_o)\mathrm{x}_i'}{\Phi(\mathrm{x}_i\theta_o)[1-\Phi(\mathrm{x}_i\theta_o)]}$$

2. Compute the **Hessian**

$$\mathrm{H}_i(\theta) = \nabla_\theta \mathrm{s}_i(\theta) = \frac{-[\phi(\mathrm{x}_i\theta)]^2\mathrm{x}_i'\mathrm{x}_i}{\Phi(\mathrm{x}_i\theta)[1-\Phi(\mathrm{x}_i\theta)]} + \mathrm{L}(\mathrm{x}_i,\theta)[y_i - \Phi(\mathrm{x}_i\theta)]$$

   where $\mathrm{L}(\mathrm{x}_i,\theta)$ is the Jacobian of $\frac{\phi(\mathrm{x}_i\theta)\mathrm{x}_i'}{\Phi(\mathrm{x}_i\theta)[1-\Phi(\mathrm{x}_i\theta)]}$.

3. Compute the asymptotic variance estimator

$$(-E[\mathrm{H}_i(\theta_o)|\mathrm{x}_i])^{-1} = \left(\sum_{i=1}^{N}\frac{[\phi(\mathrm{x}_i\hat{\theta})]^2\mathrm{x}_i'\mathrm{x}_i}{\Phi(\mathrm{x}_i\hat{\theta})[1-\Phi(\mathrm{x}_i\hat{\theta})]}\right)^{-1} \xrightarrow{p} \mathrm{A}_0^{-1}$$