NATIONAL RESEARCH UNIVERSITY

ADVANCED ECONOMETRICS

Lecture 09.12.15

Binary choice models

Olga Demidova
E-mail:demidova@hse.ru
Department of Applied Economics

BINARY CHOICE MODELS

- Why do some people go to college while others do not?
- Why do some women enter the labor force while others do not?
- Why do some people buy houses while others rent?
- Why do some people migrate while others stay put?

Economists are often interested in the factors behind the decision-making of individuals or enterprises, examples being shown above.

The models that have been developed for this purpose are known as qualitative response or binary choice models, with the outcome, which we will denote Y, being assigned a value of 1 if the event occurs and 0 otherwise.

BINARY CHOICE MODELS

$$Y = \beta_1 + \beta_2 X_2 + \ldots + \beta_k X_k + u$$

$$p_i = P(Y_i = 1), \quad i = 1, \ldots, n$$

$$P(Y_i = 0) = 1 - p_i$$

$$\hat{Y} = \hat{\beta}_1 + \hat{\beta}_2 X_2 + \ldots + \hat{\beta}_k X_k,$$

$$\hat{Y} - ?$$

BINARY CHOICE MODELS

$$E(Y_i) = p(Y_i = 1)\cdot 1 + p(Y_i = 0)\cdot 0 = p_i$$

$$E(Y_i) = E(\beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + u_i) =$$

$$= \beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + E(u_i) =$$

$$= \beta_1 + \beta_2 X_{1i} + \ldots + \beta_k X_{ki}$$

LINEAR PROBABILITY MODEL

$$p(Y=1)=\beta_1+\beta_2 X_2+\ldots+\beta_k X_k$$

Shortcoming of the linear probability model:

1) **it may predict probabilities of more than 1. It may also predict probabilities less than 0.**
2) **Distribution of u is not normal**

$$Y_i=1 \quad \Rightarrow \quad u_i=1-\beta_1-\beta_2 X_{2i}-\ldots-\beta_k X_{ki}$$

$$Y_i=0 \quad \Rightarrow \quad u_i=-\beta_1-\beta_2 X_{2i}-\ldots-\beta_k X_{ki}$$

Since u does not have a normal distribution, the standard errors and test statistics are invalid. Its distribution is not even continuous.

LINEAR PROBABILITY MODEL

Shortcoming of the linear probability model:

3) Heteroscedasticity

$$\sigma_{u_i}^2 = (\beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki})(1 - \beta_1 - \beta_2 X_i - \ldots - \beta_k X_{ki})$$

BINARY CHOICE MODELS

$$P(Y_i = 1) = F(Z_i), \quad Z_i = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki}, i = 1, \ldots, n.$$

The usual way of avoiding this problem is to hypothesize that the probability is a sigmoid (S-shaped) function of Z, F(Z), where Z is a function of the explanatory variables.

BINARY CHOICE MODELS

$$Y_i^* = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + u_i,$$

$$E(u_i) = 0, \operatorname{var}(u_i) = \sigma_u^2, \quad i = 1, \ldots, n,$$

$$F - cumulative \quad function \quad of \quad u / \sigma_u,$$

$$f = F' \; is \quad a \quad symmetric \quad function.$$

$$\begin{cases} Y_i = 1, \; if \;\; Y_i^* \geq 0 \\ Y_i = 0, \; if \;\; Y_i^* < 0 \end{cases}$$

BINARY CHOICE MODELS

$$P(Y_i=1)=P(Y_i^*\geq 0)=P(\beta_1+\beta_2X_{2i}+\ldots+\beta_kX_{ki}+u_i\geq 0)=$$

$$=P(u_i\geq -\beta_1-\beta_2X_{2i}-\ldots-\beta_kX_{ki})=P(u_i\leq \beta_1+\beta_2X_{2i}+\ldots+\beta_kX_{ki})=$$

$$=F\left(\frac{\beta_1}{\sigma_u}+\frac{\beta_2}{\sigma_u}X_{2i}+\ldots+\frac{\beta_k}{\sigma_u}X_{ki}\right)$$

Most popular F(Z)

1) $Logit \quad model$

$$F(z) = \Lambda(z) = \frac{1}{1+e^{-z}},$$

$$f(z) = \frac{e^{-z}}{(1+e^{-z})^2}.$$

2) $Probit \quad model$

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2},$$

$$F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2} dt.$$

Estimation

$$L(Y,X|\beta)=\prod_{Y_i=1}F(\beta_1+\beta_2X_{2i}+\ldots+\beta_kX_{ki})\prod_{Y_i=0}(1-F(\beta_1+\beta_2X_{2i}+\ldots+\beta_kX_{ki}))=$$

$$=\prod_{i=1}[F(\beta_0+\beta_1X_{1i}+\ldots+\beta_kX_{ki})]^{Y_i}[1-F(\beta_0+\beta_1X_{1i}+\ldots+\beta_kX_{ki})]^{1-Y_i}$$

$$l(X,Y|\beta)=\ln L(X,Y|\beta)=\sum_{i=1}^{n}[Y_i\ln F(\beta_1+\beta_2X_{2i}+\ldots+\beta_kX_{ki})+$$

$$+(1-Y_i)\ln(1-F(\beta_1+\beta_2X_{2i}+\ldots+\beta_kX_{ki}))]\to\max$$

Conditions for an optimum

$$\frac{\partial l(X,Y|\beta)}{\partial \beta_j} =$$

$$= \sum_{i=1}^{n} X_{ji} \left[Y_i \frac{f(\beta_1 + \beta_2 X_{1i} + \ldots + \beta_k X_{ki})}{F(\beta_1 + \beta_2 X_{1i} + \ldots + \beta_k X_{ki})} + (1 - Y_i) \frac{(-f(\beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki}))}{1 - F(\beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki})} \right] = 0,$$

$$j = 1, \ldots, k.$$

$$For \quad logit \quad model:$$

$$\sum_{i=1}^{n} [Y_i - \Lambda(\beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki})] X_{ji} = 0,$$

$$H = \frac{\partial^2 l}{\partial \beta \partial \beta'} = -\sum_i \Lambda_i (1 - \Lambda_i) x_i x_i'$$

$$is \quad negatively \quad definite \Rightarrow \max$$

Conditions of max

For probit model:

$$l = \sum_{Y_i = 0} \ln[1 - \Phi(\beta_1 + \beta_2 X_{1i} + \ldots + \beta_k X_{ki})] +$$

$$+ \sum_{Y_i = 1} \ln \Phi(\beta_1 + \beta_2 X_{1i} + \ldots + \beta_k X_{ki}) = 0,$$

$$\frac{\partial l}{\partial \beta} = \sum_{Y_i = 0} \frac{-\phi_i}{1 - \Phi_i} x_i + \sum_{Y_i = 0} \frac{\phi_i}{\Phi_i} x_i = 0,$$

$$H = \frac{\partial^2 l}{\partial \beta \partial \beta'}$$

is also negatively definite

(Green, ed 7, p.691 – 692) $\Rightarrow$ max

Covariance matrix

$$I(\theta) = -E\left[\frac{\partial^2 l}{\partial\theta\partial\theta'}\right]$$

$$\hat{\theta}_{ML} \overset{a}{\sim} N(\theta, I^{-1}(\theta))$$

$$H_0 : \beta_j \neq 0$$

$$z = \frac{\hat{\beta}_j^{ML}}{\hat{\sigma}_{\hat{\beta}_j}}$$

Marginal effects

For continuous variables:

$$\frac{\partial \hat{p}_i}{\partial X_j}(X_1,\ldots,X_k)=\frac{\partial F}{\partial Z}\frac{\partial Z}{\partial X_j}(X_1,\ldots,X_k)=f(X_1,\ldots,X_k)\hat{\beta}_j, i=1,\ldots,n, \quad j=1,\ldots,k$$

$$(X_1,\ldots,X_k)-?$$

$$By \quad default \quad (\bar{X}_1,\ldots,\bar{X}_k).$$

$$\frac{\partial \hat{p}_i}{\partial X_j}(X_1,\ldots,X_k)=\hat{p}(X_1,\ldots,X_{j-1},D=1,X_{j+1},\ldots,X_k)-\hat{p}(X_1,\ldots,X_{j-1},D=0,X_{j+1},\ldots,X_k),$$

$$i=1,\ldots,n, \quad j=1,\ldots,k$$

Marginal effects

For dummy variables:

$$\frac{\partial \hat{p}_i}{\partial X_j}(X_1,\ldots,X_k)=\hat{p}(X_1,\ldots,X_{j-1},D=1,X_{j+1},\ldots,X_k)-\hat{p}(X_1,\ldots,X_{j-1},D=0,X_{j+1},\ldots,X_k),$$

$$i=1,\ldots,n,\quad j=1,\ldots,k$$

Measuring goodness of fit

1) *McFadden's likelihood ratio index:*

$$R^2_{MF} = 1 - \frac{\hat{l}}{l_0},$$

$\hat{l}$ – *the maximum likelihood value of the model,*

l_0 – *the maximum value of the loglikelihood function when all parameters except the intercept, are set to zero,*

2) $$pseudoR^2 = 1 - \frac{1}{1 + 2(\hat{l} - l_0)/n} \quad (Amemiya)$$

Cross-tabulation of actual and predicted outcomes

		$\hat{Y}_i$		
		0	1	Total
Y_i	0	n00	n01	N0
	1	n10	n11	N1
	Total	n0	n1	N

Measuring goodness of fit

The implied predictions are

$$\hat{Y}_i = 1 \quad if \quad F(x_i'\beta) > 1/2 \Leftrightarrow x_i'\beta > 0,$$

$$\hat{Y}_i = 0 \quad if \quad F(x_i'\beta) \le 1/2 \Leftrightarrow x_i'\beta \le 0.$$

The proportion of incorrect predictions:

$$wr_1 = \frac{n_{01} + n_{10}}{N},$$

which can be compared with the proportion of incorrect predictions based on the model with an intercept term only.

Measuring goodness of fit

This simplest model will predict 1 for all observations if $\hat{p} > N_1 / N > 1/2$ *and 0 otherwise. The proportion of incorrect predictions:*

$$wr_0 = \begin{cases} 1-\hat{p} & if \quad \hat{p} > 0.5, \\ \hat{p} & if \quad \hat{p} \le 0.5 \end{cases},$$

$$R_p^2 = 1 - \frac{wr_1}{wr_0}.$$

Hypothesis tests

$$P(Y_i = 1) = F(Z_i), \quad Z_i = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki}, i = 1, \ldots, n.$$

$$H_0 : R\beta = q,$$

$$Wald \quad test:$$

$$W = (R\hat{\beta} - q)'(R\hat{Var}(\hat{\beta})R')^{-1}(R\hat{\beta} - q),$$

$$Likelihood \quad ratio \quad test:$$

$$LR = -2[\ln \hat{L}_R - \ln \hat{L}_U]$$

Example of logit model

```
. logit GRAD ASVABC SM SF MALE

Iteration 0:   log likelihood = -118.67769
Iteration 1:   log likelihood = -104.73493
Iteration 2:   log likelihood = -97.080528
Iteration 3:   log likelihood = -96.806623
Iteration 4:   log likelihood = -96.804845
Iteration 5:   log likelihood = -96.804844

Logit estimates                                   Number of obs   =        540
                                                  LR chi2(4)      =      43.75
                                                  Prob > chi2     =     0.0000
Log likelihood = -96.804844                       Pseudo R2       =     0.1843

------------------------------------------------------------------------------
        GRAD |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      ASVABC |   .1329127   .0245718     5.41   0.000     .0847528    .1810726
          SM |   -.023178   .0868122    -0.27   0.789    -.1933267    .1469708
          SF |   .0122663   .0718876     0.17   0.865    -.1286307    .1531634
        MALE |   .1279654   .3989345     0.32   0.748    -.6539318    .9098627
       _cons |  -3.252373   1.065524    -3.05   0.002    -5.340761   -1.163985
------------------------------------------------------------------------------
```

Example of probit model

```
. probit GRAD ASVABC SM SF MALE

Iteration 0:   log likelihood = -118.67769
Iteration 1:   log likelihood = -98.195303
Iteration 2:   log likelihood = -96.666096
Iteration 3:   log likelihood = -96.624979
Iteration 4:   log likelihood = -96.624926

Probit estimates                                  Number of obs   =        540
                                                  LR chi2(4)      =      44.11
                                                  Prob > chi2     =     0.0000
Log likelihood = -96.624926                       Pseudo R2       =     0.1858

------------------------------------------------------------------------------
        GRAD |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      ASVABC |   .0648442   .0120378     5.39   0.000     .0412505    .0884379
          SM |  -.0081163   .0440399    -0.18   0.854     -.094433    .0782004
          SF |   .0056041   .0359557     0.16   0.876    -.0648677    .0760759
        MALE |   .0630588   .1988279     0.32   0.751    -.3266368    .4527544
       _cons |  -1.450787   .5470608    -2.65   0.008    -2.523006   -.3785673
------------------------------------------------------------------------------
```

Odd ratio for the logit models

$$OR = \frac{\Pr(Y = 1)}{\Pr(Y = 0)}$$

$$\ln(OR) = \beta_1 + \beta_2 X_2 + \ldots \beta_k X_k$$

Ordered response models

$$Y_i = 1, 2, \dots, m$$

$$Y_i^* = x_i^{'}\beta + \varepsilon_i,$$

$$Y_i = j \quad if \quad c_{j-1} < Y_i^* < c_j, \quad j = 1, \dots, m,$$

$$c_0 = -\infty, \quad \dots, \quad c_m = \infty,$$

$$P(Y_i = j) = F(c_j - x_i^{'}\beta) - F(c_{j-1} - x_i^{'}\beta),$$

$$L = \prod_{j=1}^{m} \prod_{i:Y_i = j} (F(c_j - x_i^{'}\beta) - F(c_{j-1} - x_i^{'}\beta)) \to \max_{\beta, c_j}$$

Ordered response models

$$P(Y_i = j) = F(c_j - x_i'\beta) - F(c_{j-1} - x_i'\beta),$$

$$\Rightarrow P(Y_t \le k \mid X) = F(c_k - x_i'\beta),\ k = 1, \ldots, m$$

(parallel regression assumption)

Brant test

Odd ratios for the ordered logit models

$$\frac{P(Y_t \le k \mid X)}{P(Y_t > k \mid X)} = \exp(c_k - x_i^{'}\beta),\ k = 1, \ldots, m$$

$$\Rightarrow \frac{P(Y_t \le k \mid X)(X, x_k + 1)}{P(Y_t > k \mid X)(X, x_k)} = \exp(-\beta_k)$$

Marginal effects for the ordered logit models

$$\frac{\partial P(Y_i = 1)}{\partial X_k} = -\beta_k f(c_1 - (X\beta)),$$

$$\frac{\partial P(Y = j)}{\partial X_k} = -\beta_k [f(c_j - (X\beta)) -$$

$$- f(c_{j-1} - (X\beta))], \quad j = 1,\ldots, m-1$$

$$\Rightarrow \frac{\partial P(Y = j)}{\partial X_k} = \beta_k f(c_{m-1} - (X\beta))$$

Multinominal logit models

$$P(Y_t = 1) = \frac{1}{1 + \exp(X_t\beta_2) + \ldots + \exp(X_t\beta_m)},$$

$$P(Y_t = j) = \frac{\exp(X_t\beta_j)}{1 + \exp(X_t\beta_2 + \ldots + \exp(X_t\beta_m)}$$

$$\Rightarrow \frac{P(Y_t = j)}{P(Y_t = k)} = \frac{\exp(X_t\beta_j)}{\exp(X_t\beta_k)} = \exp(X_t(\beta_j - \beta_k))$$

Multinominal logit models

IIA – independence from irrelevant alternatives
Test Small-Hsiao

NATIONAL RESEARCH UNIVERSITY

Thank you for your attention!

20, Myasnitskaya str., Moscow, Russia, 101000
Tel.: +7 (495) 628-8829, Fax: +7 (495) 628-7931
www.hse.ru