Binary Choice Models Jerome Dumortier

Numerical Methods

Concepts
Probit Model

organic Comparison LPM vs. Logit

Coefficicent Estimates

Estimation with R

Base Results

Marginal Effects

Marginal Effects with mfx package

Probabilities
Fitted Values in a

Binary Choice Model
Probit Model

Additional

Binary Choice Models

Jerome Dumortier

04 March 2025

Numerical Methods

Concepts

Example using organic

vs. Logit
Coefficieen

Estimation with Base Results

Results from mf:

Effects

Marginal Effects w mfx package

Predicted Probabilities

Fitted Values in a Binary Choice Mod Probit Model

Additional Example

Binary choice models and possible research questions

- Did you vote during the last election?
- Does an individual get arrested after being released from prison?
- Does an individual participate in the labor market?

Dependent variable y takes on one of two values: 0 or 1

Binary Choice Mc

Additional Example Consider the following equations

$$y = x^2$$
$$y = x^2 + \sqrt{x}$$

Questions

- What is the value of y if x = 4 or x = 9?
- What is the value of x if y = 81 or x = 14?

Need for numerical methods to determine the answer to the last question more generally for the second equation

> Jerome Dumortier

Numerical Methods

Theoretical Concepts

Probit Model

example using organic Comparison LP vs. Logit

Coefficieer Estimates

Estimation with

Base Results

Results from

Margina Effects

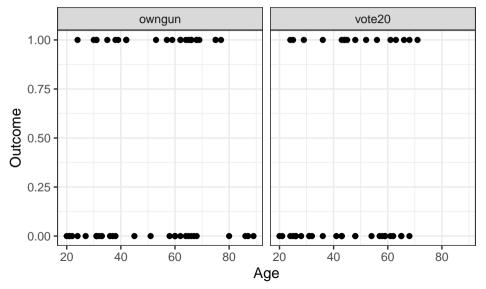
Marginal Effects wit

Predicted Probabilitie

Fitted Values in a Binary Choice Mc Probit Model

Additional Example

Sample Binary Choice Data from the GSS



Dumortier

Numerical Methods

Concepts
Probit Model
Example using

Coefficieent

Estimation with Base Results

Base Results

Margina

Marginal Effects wi

Predicted Probabilities

Fitted Values in Binary Choice M Probit Model

Additional Example

Linear Probability Model (LPM)

Most rudimentary model for binary choice

• Use of linear regression model, i.e., $y_i = \beta_0 + \beta_1 \cdot x_i + \epsilon$

Problems

- Possibility of $E(y_i|x_i) > 1$ or $E(y_i|x_i) < 0$
- Error terms are neither normally distributed nor homoscedastic

Alternative: Logit and Probit

• Calculation of the probability that the outcome equals 1 given exogenous variables, i.e, Pr(y=1|x)

Jerome Dumortier

Binary Choice Models

Common Setup for Logit and Probit Models

Theoretical

Concepts

General assumption about some function $G(\cdot)$ for all values of z

$$0 \leq G(z) \leq 1$$

 $z = \beta_0 + \beta_1 \cdot x_1 + \cdots + \beta_k \cdot x_k$

vs. Logit

Let

Then, we have

$$P(y=1|x) = G(\beta_0 + \beta_1 \cdot x_1 + \cdots + \beta_k \cdot x_k)$$

Logit Model

Methods

Theoretical Concepts Probit Model

organic Comparison LPM vs. Logit

vs. Logit

Coefficicent Estimates

Estimation wit

Results from mi

Marginal

Marginal Effects w

Predicted Probabilities

Fitted Values in a Binary Choice Model Probit Model

Additional

Remember the Bernoulli distribution from statistics:

$$Pr(y = 1) = p$$

 $Pr(y = 0) = 1 - p$

with E(y) = p. We have the following for the logit model

$$Pr(y=1) = G(z) = \frac{e^z}{1+e^z} = \frac{1}{1+e^{-z}}$$

where $z = \beta_0 + \beta_1 \cdot x_1 + \cdots + \beta_k \cdot x_k$

Models Jerome Dumortier

Binary Choice

Probit Model

Probit Model

vs. Logit

Instead of using the cumulative logistic distribution, the probit model uses the cumulative normal distribution:

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

Both models lead to similar results (not similar coefficients!).

Binary Choice Models Jerome

Dumortier

Numerical Methods

Concepts

Probit Model

Example using

organic Comparison LPI

vs. Logit

Coefficicen Estimates

Estimation wit

base Results

Margina

Marginal Effects wi

Predicted Probabilities

Fitted Values in a Binary Choice Me Probit Model

Additional Example

Example using organic

Data description

- *income* of the respondent in \$ 1,000
- buying of organic food: yes (1) or no (0)

Results of interest for the binary choice model (for other models as well)

- Coefficient estimates
- Marginal effects
- Predicted probabilities

Notes

- Estimation through Maximum Likelihood
- Difficulty interpreting the values of coefficient

Binary Choice Models Jerome

Dumortier

Numerical Methods

Concepts
Probit Model

Example using organic

Comparison LPM vs. Logit

Coefficieer

Estimation wit

Results from mfy

Marginal

Marginal Effects with mfx package

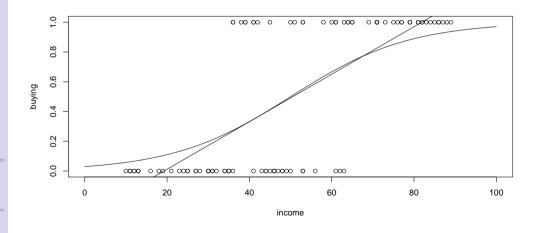
Predicted Probabilities

Fitted Values in a Binary Choice Mod Probit Model

Frobit Model

Additiona Example

Comparison LPM vs. Logit



lerome Dumortier

Binary Choice Models

vs. Logit

Estimation with R

Estimation with R

Coefficient estimates using the built-in R command:

bhatglm = glm(buying~income, family=binomial(link="logit"), data=organic)

library(mfx)

bhatmfx = logitmfx(buying~income,data=organic)

Obtaining summary from bhatmfx

summary(bhatmfx\$fit)

```
Binary Choice
Models
```

Jerome Dumortier

Numerical

Probit Mode

Example using organic
Comparison LPN

Coefficicer Estimates

Latimation w

Base Results

Marginal Effects wit

Predicted Probabilities

Fitted Values in a Binary Choice Mo

Probit Mode

Additiona Example

Base Results

```
##
## Call:
## glm(formula = buving ~ income, family = binomial(link = "logit").
      data = organic)
##
## Coefficients:
              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -5.87557
                          1.13842
                                   -5.161 2.45e-07 ***
## income
               0.11709
                          0.02247
                                    5.211 1.87e-07 ***
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
##
  (Dispersion parameter for binomial family taken to be 1)
##
       Null deviance: 138.469 on 99 degrees of freedom
## Residual deviance: 70.931 on 98 degrees of freedom
## ATC: 74.931
## Number of Fisher Scoring iterations: 6
```

Jerome Dumortier

Numerical Methods

Concepts

Example using organic
Comparison LPN vs. Logit

Coefficieer Estimates

Estimation with Base Results

Results from mfy

Marginal

Marginal Effects wit

Predicted Probabilities

Fitted Values in a Binary Choice Mo

Additiona

Results from mfx

```
##
## Call:
## glm(formula = formula, family = binomial(link = "logit"), data = data,
       start = start, control = control, x = T)
##
## Coefficients:
              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -5.87557
                          1.13842
                                   -5.161 2.45e-07 ***
## income
               0.11709
                          0.02247
                                    5.211 1.87e-07 ***
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
##
  (Dispersion parameter for binomial family taken to be 1)
##
       Null deviance: 138,469 on 99 degrees of freedom
## Residual deviance: 70.931 on 98 degrees of freedom
## ATC: 74.931
## Number of Fisher Scoring iterations: 6
```

Binary Choice Models Jerome Dumortier

Marginal Effects with mfx package

Numerica Methods

Concepts

Example using organic
Comparison LPM

Coefficieen

Estimation with

Base Results

Marginal Effects

Marginal Effects with mfx package

Predicted Probabilitie

Fitted Values in a Binary Choice Mo Probit Model

Additional Example Advantage of mfx package: Estimation of marginal effects

bhatmfx\$mfxest

```
## dF/dx Std. Err. z P>|z| ## income 0.02919553 0.005634262 5.181785 2.197728e-07
```

Important note:

Marginal effects are estimated at the mean of the independent variable(s)!

Models Jerome Dumortier

Binary Choice

Bamora

Theoretic

Probit Model
Example using organic

vs. Logit
Coefficieent

Estimation with F

Base Results

Marginal Effects

Marginal Effects w mfx package

Predicted Probabiliti

Fitted Values in a Binary Choice Model Probit Model

Additional Example

Fitted Values in a Binary Choice Model

Example:

• What are the predicted probabilities of a person purchasing organic given their annual income (in \$ 1,000) of 25, 50, and 75?

Solution in R:

```
datablock = data.frame(income=c(25,50,75))
test = predict(bhatglm,newdata=datablock,type="response")
```

```
Models

Jerome
Dumortier

Numerical
Methods
```

Very similar results compared to Logit:

bhatmfx = probitmfx(buying~income,data=organic)

bhatmfx\$mfxest

dF/dx Std. Err. z P>|z| ## income 0.02771441 0.004753676 5.830101 5.539374e-09

Income 0.02771441 0.004703070 3.030101 3.333374e 03

Probabilities
Fitted Values in a

Binary Choice

Probit Model
Additiona

vs. Logit

Additional Example

lerome Dumortier

F 10 1

Food Purchases fpdata

Food purchases data:

- strawberries_org: Frequency of strawberry purchases per month
- tomatoes org: Frequency of strawberry purchases per month
- age: Age of the respondent
- kidsunder12: Presence of kids under the age of 12
- rootsurban: Urban (as opposed to rural) upbringing of respondent
 - education: Education level
- income: Income

```
Binary Choice
Models
```

Jerome Dumortier

Numerical Methods

Theoretica

Concepts

Probit Model Example using organic

Comparison LPM vs. Logit

Coefficicent Estimates

Estimation with

Results from m

Marginal

Marginal Effects with

Predicted Probabilities

Fitted Values in a Binary Choice Mode

Additional Example

Data Preparation and Estimation

```
fpdata$strawberriesorg
                          ifelse(fpdata$strawberriesorg==0,0,1)
fpdata$tomatoesorg
                          ifelse(fpdata$tomatoesorg==0,0,1)
bhats = glm(strawberriesorg~age+kidsunder12+rootsurban+
            education+income.
            family=binomial(link="logit"),
            data=fpdata)
bhatt = glm(tomatoesorg~age+kidsunder12+rootsurban+
            education+income.
            family=binomial(link="logit"),
            data=fpdata)
```

Jerome Dumortier

lumerical

Concepts

Probit Mode

Example using organic

vs. Logit
Coefficieer

Estimation with F

Base Results

Results from wf

Marginal

Marginal Effects wit

Predicted Probabilities

Binary Choice M

Additional

Results Strawberries

```
##
## Call:
## glm(formula = strawberriesorg ~ age + kidsunder12 + rootsurban +
       education + income, family = binomial(link = "logit"), data = fpdata)
##
## Coefficients:
                Estimate Std. Error z value Pr(>|z|)
## (Intercept)
               6.961e-02 6.909e-01
                                      0.101
                                             0.91974
               -8.478e-03
                          1.121e-02
                                     -0.756
                                             0.44947
## kidsunder12
               8.526e-02
                          3.709e-01
                                      0.230
                                             0.81820
## rootsurban
               3 507e-01
                          3.312e-01
                                      1.059
                                             0.28972
              -1.203e-01
                          1.329e-01
                                     -0.905
                                             0.36528
## education
               1.524e-05 5.597e-06
                                      2.722
                                             0.00649 **
## income
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
  (Dispersion parameter for binomial family taken to be 1)
##
       Null deviance: 232.45 on 171 degrees of freedom
## Residual deviance: 222.77 on 166 degrees of freedom
    (4 observations deleted due to missingness)
## ATC: 234.77
##
## Number of Fisher Scoring iterations: 4
```

Jerome Dumortier

vs. Logit

Results Tomatoes

```
##
## Call:
## glm(formula = tomatoesorg ~ age + kidsunder12 + rootsurban +
       education + income, family = binomial(link = "logit"), data = fpdata)
##
## Coefficients:
                Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.329e-01 7.010e-01
                                     -0.190
                                             0.84967
               -5.728e-03
                          1.138e-02
                                     -0.503
                                             0.61466
                          3.770e-01
## kidsunder12 -1.104e-01
                                     -0.293
                                             0.76956
## rootsurban
               3 603e-01
                          3.364e-01
                                      1.071
                                             0.28417
              -4.158e-02
                          1.338e-01
                                     -0.311
                                             0.75606
## education
               1.708e-05 5.892e-06
                                      2.899
                                             0.00374 **
## income
## Signif. codes: 0 '***' 0.001 '**' 0.05 '.' 0.1 ' ' 1
  (Dispersion parameter for binomial family taken to be 1)
##
       Null deviance: 227.06 on 171 degrees of freedom
## Residual deviance: 216.18 on 166 degrees of freedom
    (4 observations deleted due to missingness)
## ATC: 228.18
##
## Number of Fisher Scoring iterations: 4
```

Binary Choice Models Jerome Dumortier

Additional Questions

Methods

Probit Mod

Example using organic
Comparison LPN vs. Logit

Coefficieen

Estimation with Base Results

Results from mf

.

Marginal Effects v

Predicted Probabilities

Fitted Values in a Binary Choice Mode Probit Model

Additional Example

For the strawberries and tomatoes regression, do the following:

- Calculate the marginal effects of all independent variables
- Calculate the predicted probability for each observation