**Computer Lab Session 4:**

**Discrete Choice Models in R**

**EF3450 Semester B 2017-18**

The example below is similar to the problem in part 2 of hand-in assignment 4. This handout explains some relevant R commands and output.

## Labor Supply of Married, Working Women

We study the probability that a woman is in the labor force depending on socio-demographic characteristics. We first consider the linear probability model

In the example, the variables are:

* : Labor force participation (1 if yes, 0 otherwise)
* : Years of education
* : Past years of labor market experience
* : Age
* : Number of children less than 6 years old

## Packages

The package "stargazer" is recommended (optional), the packages "lmtest" and "car" are required:

# Install the package "stargazer", which provides the command "stargazer"

install.packages('stargazer')

library(stargazer)

# Install the package "lmtest", which provides the command "coeftest"

install.packages('lmtest')

library(lmtest)

# Install the package "car", which provides the command "hccm"

install.packages('car')

library(car)

## Data

## Set your working directory, load the example data, and display its structure:

## # Set working directory

## setwd('C:\\Users\\EFuser\\Desktop') # Change this path appropriately

## # Read the data set to the data frame "LFdata"

## LFdata <- read.csv('cls4\_data.csv')

## str(LFdata) # Display data structure

## Linear Probability Model

## Estimate the linear probability model:

## # Estimate a linear probability model using OLS

## LPM <- lm(inlf ~ educ+exper+age+kidslt6, data=LFdata)

## Regression Results

## Display the estimation results with heteroskedasticity-robust standard errors:

## # Regression output based on White's heteroskedasticity-robust standard errors

## coeftest(LPM, vcov=hccm(LPM, type="hc0"))

## The output:

## t test of coefficients:

## Estimate Std. Error t value Pr(>|t|)

## (Intercept) 0.7960514 0.1326168 6.0026 3.026e-09 \*\*\*

## educ 0.0339157 0.0067833 4.9999 7.152e-07 \*\*\*

## exper 0.0233905 0.0019924 11.7401 < 2.2e-16 \*\*\*

## age -0.0194387 0.0020961 -9.2736 < 2.2e-16 \*\*\*

## kidslt6 -0.2783017 0.0310574 -8.9609 < 2.2e-16 \*\*\*

## ---

## Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

## Probit Model

## Estimate the corresponding probit model:

## # Estimate probit model using ML

## PM <- glm(inlf ~ educ+exper+age+kidslt6, family = binomial(link=probit), data=LFdata)

## Logit Model

## Estimate the corresponding logit model:

## # Estimate logit model using ML

## LM <- glm(inlf ~ educ+exper+age+kidslt6, family = binomial(link=logit), data=LFdata)

## Probit & Logit Results

## Display the estimation results for the probit and logit models:

## # Display probit & logit estimation results

## stargazer(PM, LM, type = 'text', digits=4)

## The output:

## ==============================================

## Dependent variable:

## ----------------------------

## inlf

## probit logistic

## (1) (2)

## ----------------------------------------------

## educ 0.1117\*\*\* 0.1882\*\*\*

## (0.0234) (0.0400)

## 

## exper 0.0729\*\*\* 0.1240\*\*\*

## (0.0074) (0.0132)

## 

## age -0.0608\*\*\* -0.0998\*\*\*

## (0.0077) (0.0133)

## 

## kidslt6 -0.8782\*\*\* -1.4460\*\*\*

## (0.1155) (0.1984)

## 

## Constant 0.8582\*\* 1.3119\*

## (0.4359) (0.7348)

## 

## ----------------------------------------------

## Observations 753 753

## Log Likelihood -409.4107 -409.4709

## Akaike Inf. Crit. 828.8213 828.9417

## ==============================================

## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

## Predicted Probabilities

## Compute the predicted probability that a woman with *educ* = 17, *exper* = 20, *age* = 39, and *kidslt6* = 0 is in the labor force using the estimated linear probability, probit, and logit models:

## # Compute a certain predicted probability using the linear probability, probit, and logit models

## LPMsummary <- summary(LPM)

## PP\_LPM <- LPMsummary$coefficients[1,1] + LPMsummary$coefficients[2,1]\*17 + LPMsummary$coefficients[3,1]\*20 + LPMsummary$coefficients[4,1]\*39 + LPMsummary$coefficients[5,1]\*0 # linear probability model

## PMsummary <- summary(PM)

## PP\_PM <- pnorm(PMsummary$coefficients[1,1] + PMsummary$coefficients[2,1]\*17 + PMsummary$coefficients[3,1]\*20 + PMsummary$coefficients[4,1]\*39 + PMsummary$coefficients[5,1]\*0) # probit model

## LMsummary <- summary(LM)

## PP\_LM <- 1 / (1 + exp(-(LMsummary$coefficients[1,1] + LMsummary$coefficients[2,1]\*17 + LMsummary$coefficients[3,1]\*20 + LMsummary$coefficients[4,1]\*39 + LMsummary$coefficients[5,1]\*0))) # logit model

## # Display the predicted probabilities

## A <- cbind(PP\_LPM, PP\_PM, PP\_LM)

## colnames(A) <- c("Linear probability model", "Probit model", "Logit model")

## rownames(A) <- c("Predicted probability")

## A

## The output:

## Linear probability model Probit model Logit model

## Predicted probability 1.082321 0.9673826 0.9568543