Department of Economics
Columbia University

UN3412
Fall 2020

Problem Set 6
Introduction to Econometrics
for both sections
Michael C. Best and Seyhan Erden

Instructions: Please aim for clarity and do not write more than a few sentences, make sure to give the answer to the question and nothing else. This problem set will help you to get into the habit of getting your point across efficiently.

1. (50p) What is the effect of children on the labor force participation of mothers? In the data set MROZ.dta, 428 of the 753 women in the sample report being in the labor force at some point during the year. Variables in the data file MROZ.dta are:

inlf	"in the labor force" =1 if women reports working for a wage outside the home at some point during the year, and zero otherwise
educ	Years of education
exper	Past years of labor market experience
expersq	Experience squared
age	Age in years
kidslt6	Number of kids less than 6 years old
kidsge6	Number of kids between 6 and 18 years of age
nwifeinc	Other sources of income including husband's earnings (in $1000s)

(a) (6p) Estimate a linear probability model by regressing *inlf* on *nwifeinc, educ, exper, expersq, age, kidslt6 and kidsge6.*

(b) (6p) Estimate a probit model using the same control variables in part (a).

(c) (6p) Estimate a logit model using the same control variables in part (a).

(d) (8p) Report your results from 3(a), 3(b) and 3(c) on Table 1.

(e) (6p) Test the hypothesis that the coefficient on *kidslt6* is zero in the population version of the regression in 3(a), against the alternative that it is nonzero, at the 5% significance level.

(f) (6p) Test the hypothesis that the probability of being in the labor force does not depend on the amount of work experience in the regression in 3(a). In words, describe the estimated relationship between working and experience (holding the other regressors constant).

(g) (6p) Test the hypothesis that the coefficient on *kidslt6* is zero in the population version of the regression in 3(b), against the alternative that it is nonzero, at the 5% significance level.

(h) (6p) Test the hypothesis that the probability of being in the labor force does not depend on the amount of work experience in the regression in 3(b). In words, describe the estimated relationship

between working and experience (holding the other regressors constant).

Table 2
LPM, Probit and Logit results of ps6 question 1

Dependent Variable: *inlf*			
Independent Variables:	LPM (OLS)	Probit (MLE)	Logit (MLE)
nwifeinc	( )	( )	( )
educ	( )	( )	( )
exper	( )	( )	( )
exper2	( )	( )	( )
age	( )	( )	( )
kidslt6	( )	( )	( )
kidsge6	( )	( )	( )
constant	( )	( )	( )
Percent Correctly Predicted *Log-Likelihood Value* *Pseudo R-Squared*	----		

2. (50p) In this problem we will study how effective campaign expenditures are in congressional elections. To do this, we will use the dataset "vote1.dta".[1] This is a cross-sectional dataset of different election campaigns and how much party "A" and party "B" spent in them. The main variables we will use are the share of votes won by party A voteA, expenditure by party A (in thousands of dollars) expendA, expenditure by party B (thousands of dollars) expendB, the share of voters in the district who voted for party A in the presidential election prtystrA, and a dummy indicating whether party A is the Democratic party democA.

(a) (5p) Generate a dummy variable indicating whether party A won, call it winA. Generate a variable for the difference between expenditure by party A and party B called dSpend. Start by running a linear probability model of the win dummy on expenditures by party B, dSpend, prtystrA and democA.

(i) (5p) What is the interpretation of the coefficient on dSpend? How does the predicted probability of party A winning change when dSpend goes from -250 to -249? How about when dSpend goes from 0 to 1? Does this make sense?

(ii) (5p) Draw a scatterplot of the actual wins (winA, on the vertical axis) against the predicted winning probabilities.

(iii) (5p) What fraction of predictions are "correct" (i.e. either have a predicted probability > 0.5 and party A won or a predicted probability <= 0.5 and party A lost)?

(b) (5p) Now let's try a probit model. Run the same model but by probit.

(i) (5p) What shortcomings of the linear probability model does the probit model overcome?

(ii) (5p) What is the interpretation of the coefficient on dSpend? How does the predicted probability of party A winning change when dSpend goes from -250 to -249? How about when dSpend goes from 0 to 1? How about when dSpend goes from 250 to 251? Does this make sense? You will find the margins command useful for this. For example, if I type "margins, dydx(prtystrA) at(prtystrA = (25 50 75))" that will calculate the change in the predicted probability A wins evaluated at prtystrA = 25, 50, and 75.

(iii) (5p) Draw a scatterplot of the actual wins (winA, on the vertical axis) against the predicted winning probabilities.

---

[1] Incidentally, you can access this, and all the other datasets used in the Stock & Watson (https://fmwww.bc.edu/ec-p/data/stockwatson/datasets.list.html) and Wooldridge (http://fmwww.bc.edu/ec-p/data/wooldridge/datasets.list.html) textbooks through Boston College. They even set up a nice stata command that lets you read them straight into stata called `bcuse`. To install it type "`ssc install bcuse`" into stata. Then you can load this dataset by typing "`bcuse vote1, clear`"

(iv) (5p) What fraction of predictions are "correct" (i.e. either have a predicted probability > 0.5 and party A won or a predicted probability <= 0.5 and party A lost)?

(v) (5p) Why might you still prefer the linear probability model to the probit model?

Following questions will not be graded, they are for you to practice and will be discussed at the recitation:

3. [ungraded] SW Empirical Exercise 11.1

4. [ungraded] SW Empirical Exercise 11.2