Problem Set #3 Econometrics IV Lizal

Limited Dependent Variable - Binary Choice

This problem set is designed to give you experience with limited dependent variables. The setting is a small local referendum in which the voters should decide whether to support additional taxes to finance additional schooling expenditures (vote YES) or to leave the taxes unchanged (vote NO). As you may know, in the USA schooling expenditures are financed by local taxes, the parents thus support only school in their own district (i.e. they finance education of their own child{ren}). And since they expect better education for higher taxes we cannot expect all of them to vote NO. Remember, that this problem set is somehow open ended, as usually.

1) As usual, start by familiarizing with the data. Use plots if appropriate.

2) Specify a reasonable modem of vote determination. Justify **each** variable that you include in the model by explaining why and how the characteristic would influence the way a person might vote. You should have a prior on the sign of the effect (i.e. coefficient should be positive / negative). You should have prior why the coefficient of variable you do not include should be insignificant.

3) Estimate your model using three different functional forms: (**a**) linear probability, (**b**) probit, and (**c**) logit. In addition to reported estimates of coefficients and standard errors (or t-stats), report the marginal probability effects at the points of means (similarly as it is done in Greene's book in the Table 21.2).

4) Interpret the results from part 3). Can you improve your model?

5) Calculate the appropriate measures of fit. How can you use them to decide which of your models is the best? Can you use these measures to decide whether the logit or probit model is preferred?

6) Use your models to predict votes for each of the three functional forms (linear, logit, probit). Can you use the results to judge the performance of your model(s)? Do you get predictions of probability outside the <0,1> interval in the case of the linear probability model?

7) Try to group observations into 8 groups according crossproducts of the number of children in public school (4 groups) and indicator whether the person has lived in the community at least for 8 years (2 groups, i.e., 4*2=8 groups). For each group compute the proportion of YES votes and run a regression on log of income and log of property taxes paid. Report the results and compare to the previous results above.

DATA DESCRIPTION:

PUB1_2	1=one or two children in public school
PUB3_4	1=three or four children in public school
PUB5	1=five or more children in public school
PRIV	1=person has at least one child in private school
YEARS	number of years living in the community
SCHOOL	person is working as a teacher (anywhere)
logINC	logarithm of family income per year
PTCON	log of tax property paid last year
YESVM	1=person voted for tax increase, 0 = against