Faculty of Economics & Politics

Part 2 Paper 3: Microeconometrics and Panel Data

Class 2: Discrete Choice

All datasets and accompanying variable description files, can be found on my web page and in `J:\Students\mw217`

1. The binary variable *approve* is equal to one if a mortgage loan to an individual is approved. The key explanatory variable is *white*, a dummy variable equal to one if applicant was white. The applicants in the data set are black and Hispanic. To test for discrimination in the mortgage loan market, a linear probability model can be used:

$$approve = \beta_0 + \beta_1 white + other\ factors$$

   (i) If there is discrimination against minorities, and the appropriate factors have been controlled for, what is the sign of β_1?

   (ii) Regress *approve* on *white* and report the results in the usual form. Interpret the coefficient on *white*. Is it statistically significant? Is it practically large?

   (iii) As controls, add the variables *hrat, obrat, loanprc, unem, male, married, dep, sch, cosign, chist, pubrec, mortlat1, mortlat2,* and *vr*. What happens to the coefficient on *white?* Is there still evidence of discrimination against non-whites?

   (iv) Now allow the effect of race to interact with the variable measuring other obligations as a percent of income *(obrat)*. Is the interaction term significant?

   (v) Using the model from part (iv) what is the effect of being white on the probability of approval when *obrat=32*, which is roughly the mean value in the sample? Obtain a 95% confidence interval for this effect. Use the data in LOANAPP.XLS for this exercise.

2. Let *grad* be a dummy variable for whether a student athlete at a large university graduates in five years. Let *hsGPA* and *SAT* be high school grade point average and *SAT* score. Let *study* be the number of hours spent per week in organised study hall. Suppose that, using data on 420 students-athletes, the following logit model is obtained:

$$\hat{P(grad} = 1|hsGPA, SAT, study) = \Lambda(-1.17 + .24\ hsGPA + .00058\ SAT + .073\ study)$$

   where $\Lambda(z) = \exp(z)/[1 + \exp(z)]$ is the logit function. Holding *hsGPA* fixed at 3.0 and *SAT* fixed at 1.200, compute the estimated difference in the graduation probability for someone who spent 10 hours per week in the study hall and someone who spent five hours per week.

3. Use the data in PNTSPRD.XLS for this exercise.

   (i) The variable *favwin* is a binary variable if the team *favoured* by the Las Vegas point spread wins[1]. A linear probability model to estimate the probability that the *favoured* team wins is

   $$P(favwin = 1|spread) = \beta_0 + \beta_1 spread$$

   Explain why, if the spread incorporates all relevant information, we expect $\beta_0 = .5$.

   (ii) Estimate the model from part (i) by OLS. Test $H_0 : \beta_0 = .5$ against a two-sided alternative. Use both the usual and heteroskedasticity-robust standard errors.

   (iii) Is *spread* statistically significant? What is the estimated probability that the favoured team wins when $spread = 10$?

   (iv) Now estimate a probit model for $P(favwin = 1|spread)$. Interpret and test the null hypothesis that the intercept is zero. *[Hint*: Remember that $\Phi(0) = .5$].

   (v) Use the probit model to estimate the probability that the favoured team wins when $spread = 10$. Compare this with the LPM estimate from part (iii).

   (vi) Add the variables *favhome, fav25,* and *und25* to the probit model and test joint significance of these variables using the likelihood ratio test. (How many *df* are in the chi-square distribution?) Interpret this result, focussing on the question of whether the spread incorporates all observable information prior to a game.

   Note: the Likelihood Ratio test statistic, for the restricted (R) and unrestricted model (UN) is given by

   $$LR = 2 * (LogL^{UN} - LogL^{R}) \sim \chi^2_{(q)}$$

   See Wooldridge, Chapter 17.

4. We now return to the problem of testing for discrimination in the mortgage loan market.

   (i) Estimate a probit model of *approve* on *white*. Find the estimated probability of loan approval for both whites and nonwhites. How do these compare with the linear probability estimates?

[1]For team A and B the spread, say $S_A - S_B$, represents the difference in the points awarded to each team based upon recent form, injury, previous meetings of the two teams.

(ii) Now, add the variables *hrat, obrat, loanprc, unem, male, married, dep, sch, cosign, chist, pubrec, mortlat1, mortlat2,* and *vr* to the probit model. Is there statistically significant evidence of discrimination against nonwhites?

(iii) Estimate the model from part (ii) by logit. Compare the coefficient on *white* to the probit estimate.

(iv) How would you compare the size of the discrimination effect between probit and logit.

Melvyn Weeks
Lent 2008