

ECON3360 Causal Inference for Microeconometrics

Tutorial 11: Limited dependent variables

Instructor: Julie Moschion

Problem I: racial discrimination in the mortgage market

Background Mortgage applications in 1990 in the Boston area reveal that 28% of minority applicants were denied mortgages but only 9% of white applicants were denied mortgages. This suggests that there may be some discrimination. But this difference in rates of applications denied could also be due to other differences between minority and white applicants (observed or unobserved).

To study whether there is discrimination on the mortgage market, we can analyse the relationship between the outcome of a mortgage application and whether the applicant is from a minority group using regressions. The outcome is a binary variable for *approve*, which is equal to one if a mortgage loan is approved, and zero otherwise. The key explanatory variable is *white*, a dummy variable equal to one if the applicant is white, and zero otherwise (i.e. if the applicant is black or Hispanic).

The relationship between the outcome and explanatory variables can be specified using a linear probability model (LPM):

$$approve = \beta_0 + \beta_1 white + controls + u$$

Use the data in loanapp.dta for the following questions.

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

(2) Estimate the LPM by OLS without including any controls. How do you interpret β_1 ? Is it statistically significant? Is it practically large?

(3) Re-run the OLS regression adding the following controls: *hrat, obrat, loanprc, unem, male, married, dep, sch, cosign, chist, pubrec, mortlat1, mortlat2, and vr*. What happens to the coefficient on *white*? Is there evidence of potential discrimination against minorities?

(4) Estimate a probit model of *approve* on *white*. How does it compare with the linear probability estimate? Calculate the estimated probability of loan approval for both white and minority applicants (hint: check out the *predict* command). How does this predicted probability differ between the two groups?

(5) 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 evidence of potential discrimination against minorities? Calculate the marginal effect of *white*.

(6) Estimate the same regression without controls by *logit*. Compare the coefficient on *white* to the probit estimate. Using the *margins* command, compute the marginal effect of *white* on *approve*: how does it compare to the LPM and probit marginal effects?

Problem II: job training and unemployment

Background A sample of men were **quasi-randomly** allocated to a job training program from Jan 1976 to Dec 1977. Allocation to the program was mostly independent of men's characteristics, except that men who participated were less likely to be high school dropouts. Participants could start the program from Jan 1976 to mid-1977. The program ended in Dec 1977. **We want to evaluate if participation in the program decreased the probability of unemployment in 1978.** Use the data in jtrain2.dta for the following questions.

- (1) The variable *train* is the job training indicator. How many men in the sample participated in the job training program? What was the maximum number of months of participation in the program?
- (2) Run a linear regression of *train* on several demographic and pre-training variables: *unem74*, *unem75*, *age*, *educ*, *black*, *hisp*, and *married*. Are these variables jointly significant at the 5% level?
- (3) Run the same regression as above using a **probit**. Compare these estimates to those of the **LPM**.
- (4) Based on the previous questions, does participation in the program appear to be exogenous?
- (5) Run an OLS regression of *unemp78* on *train*. Interpret the coefficient estimate on *train*. Is it statistically significant?
- (6) Run a **probit** of *unemp78* on *train*. Does it make sense to compare the probit coefficient on *train* with the coefficient obtained from the linear probability model in (5)? **What is the marginal effect of training on the probability of being unemployed in 1978?**
- (7) Calculate the predicted probabilities from (5) and (6). Are they identical?
- (8) Run a probit where you include the control variables from question 2 and estimate the marginal effect of *train*. Run the LPM model with the same controls. Compare the estimated effects.

What do we expect after adding controls?

