Applied Economics

Probability Models

1. We are studying the factors behind the likelihood that a mortgage is denied to an individual. We use data from Stock and Watson, `hmda.gdt`, with information about mortgage applications in the Boston area. In the first part we will use the following variables: the variable *deny*, that takes the value one if the mortgage is denied and zero otherwise, *pi_rat* that indicates the debt payment to income ratio, and *black*, a dummy variable that takes a value of one if the applicant is black, and zero otherwise.

   i) What is the fraction of individuals who are denied a mortgage? Summarize variables for blacks.

   ii) Using a linear specification, estimate a model that has as dependent variable the variable *deny* and as unique control the variable PI_rat. Give an interpretation of the estimated coefficient. Should we use robust standard error? Explain.

   iii) Include to the previous specification the *black*. How should we interpret the estimated coefficient for this last variable?

   iv) Calculate the predicted value for a black individual whose payment to income ratio is 0.15. What is the interpretation for this predicted value? What is the predicted value for another individual but classified as non-black? Have you detected any problem? Explain.

   v) Estimate the model in iii) using a Logit and Probit model. What is the impact of being black for an individual with PI_rat equal to the mean in the sample? How does this value compare with the one obtained with the linear specification.

   vi) Does the impact of race change when we control for the education (*hischl*), marital status (*single*, *married*), self-employed status ($selfemp$), probability of unemployment (*probunmp*) and whether or not the individual has public records associated with credit problems (*pubrec*).

2. Using US Census data for the year 1980, `us1980.dta,` we are studying the determinants of attending private school. The variable private is a dummy variable that indicate whether or not a student is enrolled in a private school and zero, otherwise. The variable *higrade_mom* indicates mother's years of education and $n1child$ the number of siblings.

i) Estimate using a linear regression specification, a model that has as dependent variable, *private* and as controls mother's education and number of siblings.

ii) Do you detect any problems with the fitted values of the model. Explain.

iii) What is the average difference in the probability of attending a private school between two children, one with a mother with 12 years of education and another one with a mother with 16 years of education. Does the result depend on the number of siblings?

iv) Repeat the previous parts using a probit and a logit model.

v) Using the models estimated in previous questions, what is the effect of additional sibling for a child with a mother with high school completed and who lives in household with two children? Does it make sense?