Econ 4811 - Problem Set #6
Due Wednesday, April 18th before class

---

1. Discrimination Research

Marianne Bertrand (Chicago Booth) and Sendhil Mullainathan (Harvard) recently conducted an experiment in which they created a series of fictitious resumes and randomly assigned a name that was very common amongst the African-American community and uncommon amongst the Caucasian community to 50% of the resumes (Tyrone and Ebony, for example) and a name that was common amongst the Caucasian community but not common amongst the African-American community to the other 50% (Brad and Emily, for example). Resumes were sent in response to advertisements posted by employers in Boston and in Chicago. The investigators kept track of the resumes that received a call-back. Data from the experiment, in **ps6q1.dta**, will be used to test whether there is evidence of discrimination in the labor market. (Note: A student research assistant put this data together so be VERY careful assuming all of the variables are entered/coded the way you need/want).

a. What is the advantage of running this experiment versus using data on the call back rate of actual real life resumes?

b. What is the difference in the call-back rate for resumes with African-American (afram is an indicator variable for African American sounding name on the resume) and Caucasian sounding names? Test the hypothesis that the call-back rate is the same for resumes with African-American and Caucasian sounding names using OLS.

c. Now estimate the model controlling gender, quality of resume, education and city. Does estimating this model change your answers in (b)? Why would you expect it to change the answer or not?

d. Using a logit model, estimate the marginal effect of having an African-American name on the call-back rate, controlling gender, quality of resume, education and city. How does estimating this model change your answers in (c)? You may find it helpful to look at the marginal effects when comparing your answer to (c).

An enterprising student followed up this experiment with another experiment in which actors were sent to interviews in response to the call-backs. The actors were given scripts to follow so that there were no differences in the qualifications and suitability for the job amongst African-American and Caucasian interviewees and, at the interviews, African-Americans and Caucasians were dressed in similar ways. The student recorded the number of job offers.

e. Using these data, test two the hypotheses. 1) that the job offer rate is the same for African-Americans and Caucasian actors who were called-back for an interview and 2) that, conditional on sending a resume, the offer rate is the same for African-Americans and Caucasians.

f. What can you conclude about discrimination in the labor market? What questions about discrimination in the labor market are still unanswered?

2. Discriminatory Lending

The Home Mortgage Disclosure Act was passed in 1975 to monitor minority and low-income access to the mortgage market. The original data collected for this purpose showed that minorities are more than twice as likely to be denied a mortgage as whites. However, variables correlated with both race and creditworthiness were omitted from the first set of lending data, making any conclusion about race's role impossible. In an effort to further examine whether discriminatory lending occurs, the Federal Reserve Bank of Boston began collecting additional variables important to the mortgage lending decision during the mid-1990s from all banks in the metro Boston area. A random subset of this data is included in **ps6q2.dta**.

a) Begin by testing whether or not a minority borrower is more likely to be denied a loan by estimating the simple linear probability model below. Remember to use the appropriate standard errors.

$$deny_i = \alpha_0 + \alpha_1 minority_i + u_i$$

b) Based on your results, what fraction of non-minority borrowers is denied?

c) What fraction of minority borrowers is denied? Is the difference in denial rates for minorities statistically significant at the 5% level?

These results are of course just a correlation and loan officers said there is definitely not discrimination in their judgments. They claim this simple analysis only looks like there is discrimination because it suffers from omitted variable bias, and if the right variables were included there would be no evidence of discrimination.

d) What characteristics must the omitted variables have for the loan officers to be right?

When the results in a) first became public, loan officers claimed that they really only cared about the payment to income ratio (P/I ratio) and absolutely nothing about race. The P/I ratio is the fraction of a borrower's monthly income that would go to their mortgage payment if the loan is approved.

e) Prove whether this factor has the characteristics necessary to cause the type of omitted variable bias the loan officers claim.

f) Estimate the following model to analyze the loan officers claim:

$$deny_i = \beta_0 + \beta_1 minority_i + \beta_2 pi\_rat_i + u_i$$

g) What is the interpretation of your estimate for $\hat{\beta}_2$?

h) Does your regression support the loan officers' claims? Why or why not?

i) Using the regression from part f), for each racial group (i.e. minority and non-minority), what percentage of the predicted values lie outside of the [0,1] interval?

Hint: You can do this by first generating the predicted vales for each racial group (e.g. predict yhat_min if minority==1 and predict yhat_non if minority==0) then generating a variable for each racial group that is equal to 1 if the prediction is between 0 and 1 and 0 otherwise. Then use the "tab" command on your newly generated variables for each racial group separately using an "if" statement.

The primary purpose of the Boston Fed's data collection was to be able to control for additional covariates in models predicting loan denial. One thought was that the minority variable was picking up features of loan applications that would lead to denial regardless of race on top of P/I ratio.

j) Estimate and show the results of a regression of loan denial against minority status, P/I ratio, household expense-to-income ratio (hse_inc), loan to value ratio (measured in 3 indicator variables), consumer credit score (ccred), public bad credit record (pubbad), and whether or not the individual is self employed (selfemp).

k) What is your estimate on the minority coefficient after controlling for the additional variables? Is it statistically significant at the 5% level? How does it compare with the raw mean you calculated in part (a)? What does this tell you about lending practices and race in Boston in the mid-1990s?

3. Binary dependent variable

This is problem C17.8 from Wooldridge

The dataset **ps6q3.dta** contains data on a job training experiment for a group of men. Men could enter the program starting in January 1976 through about mid-1977. The program ended in December 1977. The idea is to test whether participation in the job training program had an effect on unemployment probabilities and earnings in 1978.

i) The variable *train* is the job training indicator. How many men in the sample participated in the job training program? What was the highest number of months a man actually participated in the program?

ii) Run a linear probability model 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?

iii) Estimate each coefficient (excluding the constant) of the regression estimated in (ii), and state whether the variable significantly predicts the probability of job training.

iv) Estimate a probit version of the linear model in part (ii). How does estimating this model change your answers in (ii)? You may find it helpful to look at the marginal effects when comparing your answer to (ii).

v) Mr. Garrido is a single hispanic, 20 years old, unemployed both in 1974 and 1975, and with 11 years of education. Using the probit results from (iv), calculate the probability of getting job training for Mr. Garrido. Show your estimations.

vi) Interpret the effect of education on the probability of getting job training for Mr. Garrido. What would the LPM have predicted the effect of education on the predicted probability of job training for Mr. Garrido?

vii) Based on your answers to (ii) through (vi), do the probit and linear probability model results differ? If they do, which results make more sense?

viii) Based on your answers to part (ii) and (iii), does it appear that participation in job training can be treated as exogenous for explaining 1978 unemployment status? Explain.

ix) Run a LPM model of *unem78* on *train* and report the results. What is the estimated effect of participating in the job training program on the probability of being unemployed in 1978? Is it statistically significant?

x) Run a probit if *unem78* on *train*. Does it makes sense to compare the coefficient on *train* with the coefficient obtained from the LPM model in (ix)?

4. Using the file
https://docs.google.com/spreadsheets/d/1LJ2As5XiZEAOJ-9Q1YZyeL__ZQ7EYoIkjMhdnho_zhg/edit?usp=sharing

posted on Canvas. Write down all the finals you have scheduled for Wednesday, December 9th in addition to ECON 4811. Please include the class number (e.g. ENGL XXXX), time, and professor's name so that I can contact them if needed.