Introductory Econometrics
Binary dependent variable

by Hieu Nguyen

Fall 2024

1.

Use the data in `loanapp_b.gdt` for this exercise. The binary variable to be explained is `approve`, which is equal to one if a mortgage loan to an individual was approved. The key explanatory variable is `whiteskin`, a dummy variable equal to one if the applicant has light skin. The other applicants in the data set are `darkskin` and `Hispanic`. To test for discrimination in the mortgage loan market, a LPM can be used:

$$\text{approve} = \beta_0 + \beta_1 \text{whiteskin} + \text{other factors}.$$

(a) Regress `approve` on `whiteskin` and report the results in the usual form. Interpret the estimated coefficient on `whiteskin`. Is it significant? Is it practically large?

(b) As controls, add the variables `hrat`, `obrat`, `loanprc`, `unem`, `male`, `married`, `dep`, `sch`, `cosign`, `chist`, `pubrec`, `mortlat1`, `mortlat2`, and `vr`. What happens to the estimated coefficient on `whiteskin`? Is there still statistically significant evidence of discrimination against non-white skin individuals?

(c) Estimate the equation in part (b) computing the White heteroskedasticity-consistent robust standard errors. Compare the 95% confidence interval on $\beta_{\text{whiteskin}}$ with the non-robust confidence interval.

(d) Obtain the fitted values from the regression in part (c). Are any of them less than zero? Are any of them greater than one?

(e) Estimate a Probit model of `approve` on `whiteskin`. Check the direction of the effect and the statistical significance of `whiteskin`. Find the estimated probability of loan approval for both `whiteskin` and non-white skin individuals. How do these compare with the LPM estimates?

(f) Now, add the variables `hrat`, `obrat`, `loanprc`, `unem`, `male`, `married`, `dep`, `sch`, `cosign`, `chist`, `pubrec`, `mortlat1`, `mortlat2`, and `vr`. Is there still statistically significant evidence of discrimination against non-white skin people? Interpret also other information from the Gretl output.

(g) Estimate the model from part (f) by Logit. Compare the estimated coefficient on `whiteskin` to the Probit model.

(h) Estimate the sizes of the discrimination effects for Probit and Logit.