

NES Master of Arts in Economics
Applied Microeconometrics
Group Assignment #3

Assignment date: December 1st, 2016

Due date: December 11th, 2016 23:59

Delivery: One electronic copy per group (text and codes zipped and uploaded to my.nes)

For this problem you may find it helpful to take a look at Keys et al. (2010) and Jiang et al. (2014) (both available on the reading list).

You are given a dataset that contains artificially generated data on mortgage loans. The variables in the dataset are the following:

Age – borrower age

Area – indicator for the area of the property

Balance – original balance on the loan

Broker – loan origination dummy (1 if the loan is originated at the broker level, 0 if at the bank level)

Del – delinquency dummy variable (1 if the loan is at least 60 days behind in payment, 0 otherwise)

Fico – borrower credit score

Gender – gender dummy variable (1 if female, 0 if male)

Hard – prepayment penalty dummy variable (1 if there is a hard prepayment penalty in the loan contract, 0 otherwise)

Income – monthly income of the borrower, in some units

LTV – the ratio of loan to value of the property

Rate – initial interest rate on the mortgage

Refinance – refinance dummy variable (1 if the mortgage purpose is for refinance, 0 if initial purchase)

White – race dummy variable (1 if white, 0 otherwise)

The main goal of this assignment is to practice the full procedure of regression discontinuity design.

You are interested in identifying the following system:

$$\text{Del}_i^* = \beta \text{rate}_i + \gamma X_i + \varepsilon_i$$

$$\text{rate}_i = \lambda X_i + v_i$$

$$\text{Del}_i = 1 \text{ if } \text{Del}_i^* > 0 \text{ and } 0 \text{ otherwise}$$

$$\text{corr}(\varepsilon_i, v_i) > 0$$

In the above equations, Del is a dummy variable for loan delinquency; rate is the interest rate on the loan; and X is a vector of loan and borrower characteristics. The difficulty of identification lies in the fact that residuals from the two equations are correlated: higher propensity to default (conditional on observables) should be priced if the lender uses any (soft) information that is not in the dataset. Moreover, there is no obvious instrument (since anything that affects default should also affect the interest rate).

You hypothesize that there is a jump in the interest rate at FICO = 620 (please verify). Apply the regression discontinuity design to identify the system, most importantly, the coefficient β . Properties in the same area tend to be highly correlated, and therefore area should be the appropriate level for clustering standard errors. You may also wish to take logs of some right-skewed variables.

Do not forget to do various checks to support validity of your findings. If there are jumps in covariates, interpret them, do they make your results stronger or weaker? You may also wish to consult Jiang et al (2014) paper as additional evidence on correlations between covariates and delinquency.

What do your results suggest about the effects of securitization other than those suggested by Keys et al (2010)? Do you see any differences with respect to loan origination channel?

Please write up your results in a mini-paper format (abstract, introduction, results, conclusion, references), no more than 10 pages length (including tables and graphs).

If you would like to do this assignment in Stata, then you may want to install additional packages (such as rd) or use the IV procedures. Please consult the TA if there are any questions with respect to this software.