Applied Microeconometrics

Home assignment 3

Anastasia Ermolina Maria Lysyuk Artem Smirnov[1]

December 2016

Abstract

In the following paper we attempt to establish a relationship between the interest rate on the mortgage and the facts of having a loan behind in payment. Taking into account that interest rate is discontinuous in a borrower's credit score, we employ a regression discontinuity design. Our key finding is quantitative link between the two variables of interest, and the coefficient is positive.

Introduction

We investigate the possible link between mortgage interest rates and borrowers' delinquencies, namely, delays in payment. In general, this research question is motivated by the increased popularity of securitization, when mortgages are pooled and sold to third-party investors as securities. Since the collapse of the U.S. mortgage market was one of the reasons between 2008 financial crisis, this question has provoked a huge academic discussion. In this paper, we rely on several findings by Keys et al. (2010), who checked whether the conversion of illiquid loans into liquid securities caused financial intermediaries to screen borrowers less intensively. Their results suggest that securitization really lowers the incentives to screen properly. Earlier papers on the issue include Douglas (1984), who has shown that one should incentivize banks to perform screening and monitoring activities, because otherwise the bank's effort may not be efficient. Similar result is provided by Holmstrom and Tirole (1997). We also know that securitized loans are more safe than those that are held by banks (Drucker and Puri (2007)).

One important hypothesis implies a leap in the interest rate at the FICO credit score of 620 (hereinafter referred to as FICO). In other words, this value represents a threshold, which cuts off the set of relatively risky borrowers from the set of relatively reliable customers. Thus, we would like to find (i) a link between the interest rate and a vector of loan and borrower characteristics and (ii) a link between the fact of payment delay and the rate along with the vector of characteristics. The system to identify is therefore:

$$Del_i = \beta rate_i + \gamma X_i + \epsilon_i$$

[1]New Economic School, Moscow, Russia

$$rate_i = \lambda X_i + \nu_i$$

$$corr(\epsilon_i, \nu_i) > 0,$$

where 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.

This specification is not only exposed to bias as the residuals are correlated, but it also complicates the task to find an appropriate instrument. Any factor that affects the interest rate should also influence the propensity to default. Identification strategy is described in the next section.

The paper is organized as follows. First, we present summary statistics in order to better understand the data we are working with, after that we conduct a series of validity checks in order to prove we are able to use the research discontinuity design. Next section presents our estimation strategy and the main assumption we make. Finally, the last section concludes the paper.

Summary statistics

The dataset is composed of level and dummy variables. Key variables are Delinquency, Rate, and FICO. Being a dummy, Delinquency has mean .29 and standard deviation .45, thus the empirical coefficient of variation is 1.55. FICO is a level variable; it contains values from 570 to 670, i.e. equally distanced from 620. It's mean is 641.23 and standard deviation is 22.19. Finally,Rate varies from 0.01 to 0.19, has mean .07 and standard deviation .03 (see Table 1).

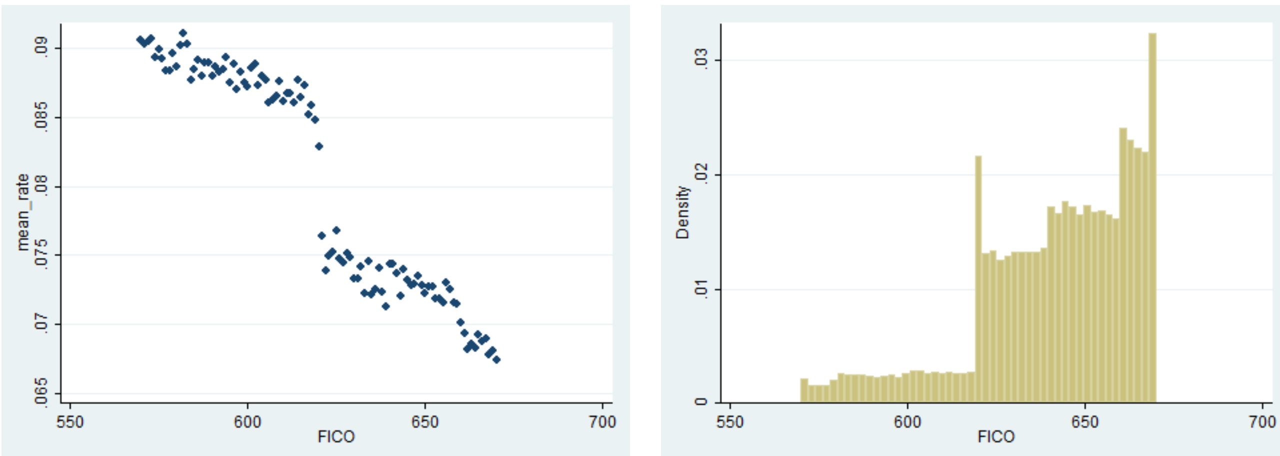
Variable	Obs	Mean	Std. Dev.	Min	Max
Age	79885	46.94	8.9	22	89
Area	85805	509.53	263.53	1	900
Balance	81125	7198.77	4441.99	76.48	147144.8
Broker	79885	.38	.49	0	1
Delinquency	85805	.29	.45	0	1
FICO	85805	641.23	22.19	570	670
Gender	79885	.38	.49	0	1
Hard	79885	.34	.47	0	1
Income	77581	304.47	825.57	0	206723
LTV	82276	70.96	23.18	4.61	107.39
Rate	85805	.07	.03	.01	.19
Refinance	79885	.4	.49	0	1
White	85805	.24	.43	0	1

Table 1: Summary statistics

Validity checks

In order our results to be credible we should conduct a number of the validity checks.

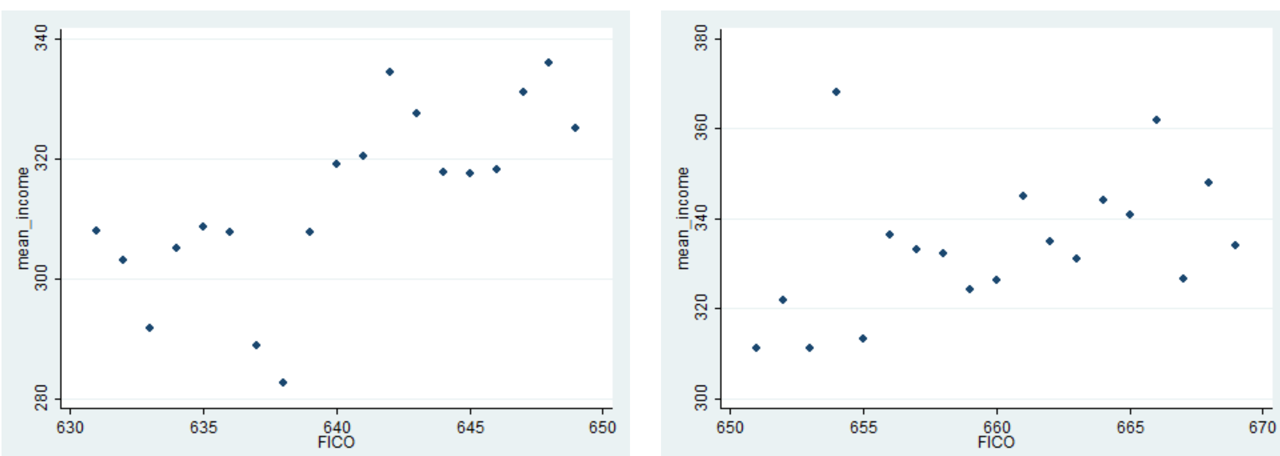
Validity check 1: First, let's check that no other threshold drives the result. From Figure 1 you can indeed ensure that there is a jump in the interest rate at $FICO = 620$. Figure 2 reflects the histogram of FICO from which we can retrieve 2 other possible candidates for threshold: they are $FICO = 640$ and $FICO = 660$. Figure 3 represents the scatter plot of Rate in the neighbourhood of $FICO = 640$ and Figure 4 represents the scatter plot of Rate in the neighbourhood of $FICO = 660$. As it can be seen from both Figures 3 and 4 there is no jump in the Rate near other candidates for threshold, so that there are no other unexplained jumps in Rate.



(a) Rate

(b) FICO histogram

Figure 1: Validity check 1: thresholds of FICO



(a) FICO about 640

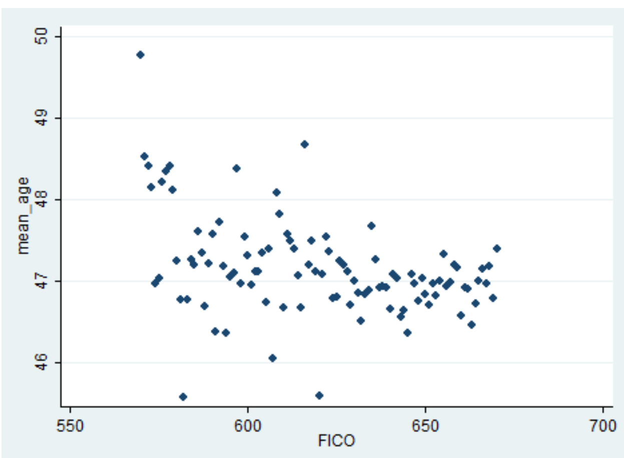
(b) FICO about 660

Figure 2: Validity check 1: thresholds of FICO

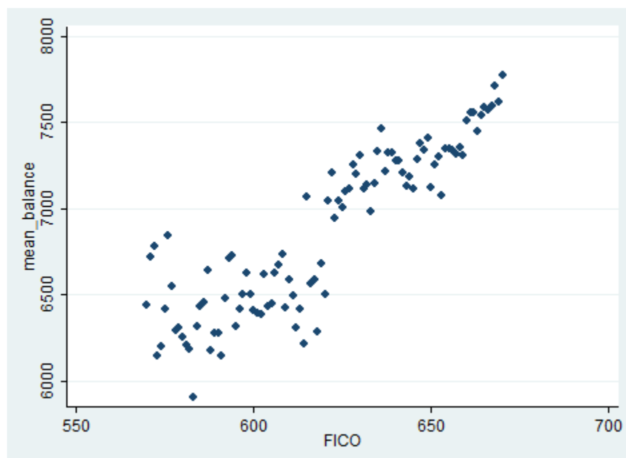
Validity check 2: Second, we have to check that there are no jumps in covariates at $FICO = 620$. Near we provide evidence for some of the variables. For example, for variables *Age* and

Balance we do not see any jump in covariates, however for covariates *Hard* and *Broker* we do see it. Intuition tells us that if local randomization works, then average covariates should be the same on both sides of the threshold. However, since we see that some of the covariates jump at $FICO = 620$, then there may be a separate reason why both *Rate*, *Broker* and *Hard* jump at the threshold, that is different from forcing variable crossing the threshold. Actually, if we can exclude these covariates from the analysis, our results will not be weakened. Unfortunately, there is some evidence which makes take into consideration these covariates. Thus, Jiang et al. (2014) found that loans issued by brokers have delinquency probabilities that are 3.7 percentage points higher than those issued by correspondents; in turn, delinquency probabilities for correspondent-issued loans are 10.6 percentage points higher than those issued by the bank. So, broker do matters. Beyond this they claim that prepayment penalty is a significant factor at the broker level.

Moreover, looking at the scatter plots for *Income* and *LTV* (see Figure 5) it's necessary to mention that there are obvious outliers in the sample which could bias our results, so that they will be eliminated for future analysis.



(a) Age



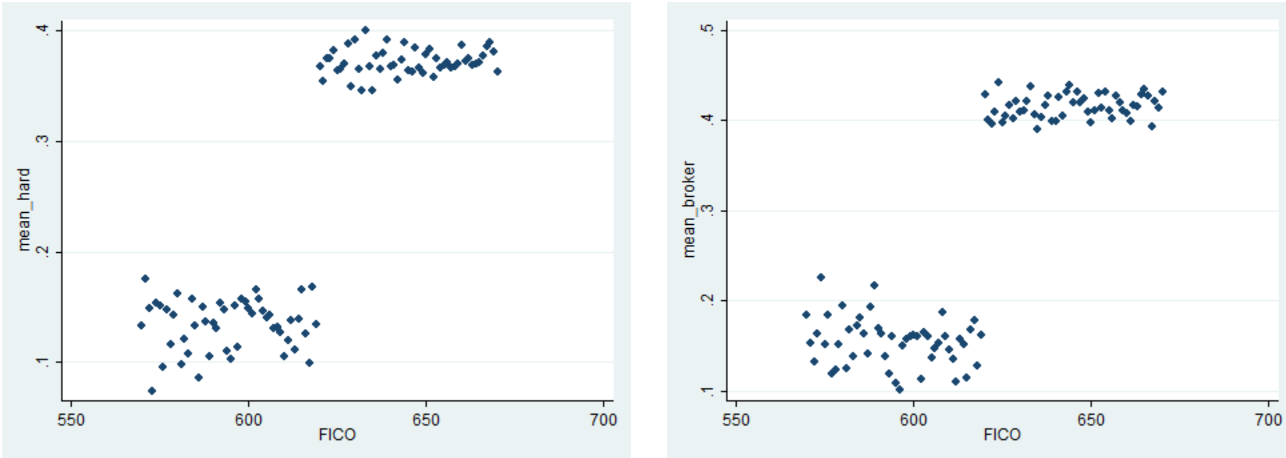
(b) Balance

Figure 3: Scatter plots of covariates

Validity check 3: Third, let's make the placebo test: test the non-eligible (high-income) group. As it can be seen from Figure 6 there is no jump in *Rate* at $FICO = 620$, so the analysis survives this robustness check.

Validity check 4: Finally, there should be no discontinuity in the density of FICO at the threshold. Since it is not the case (see Figure 6), we conclude the agents do not manipulate the threshold, so that our analysis is valid.

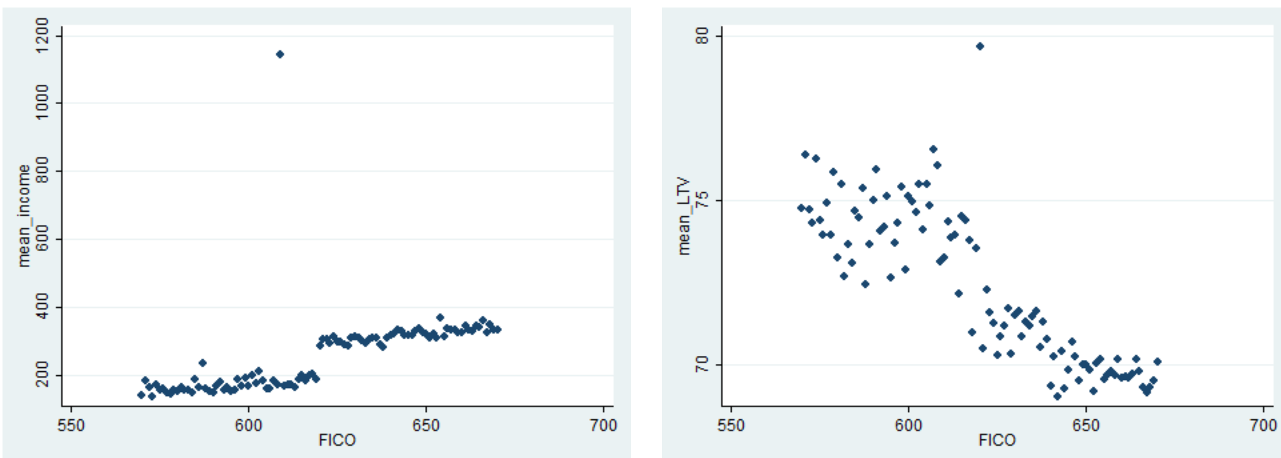
Check specification. Moreover, let's look at the empirical distributions. Most of the variables are close to have normal distribution or left-skewed (see Figures 7, 8), so it will be a good idea to



(a) Hard

(b) Broker

Figure 4: Scatter plots of covariates



(a) Income

(b) LTV

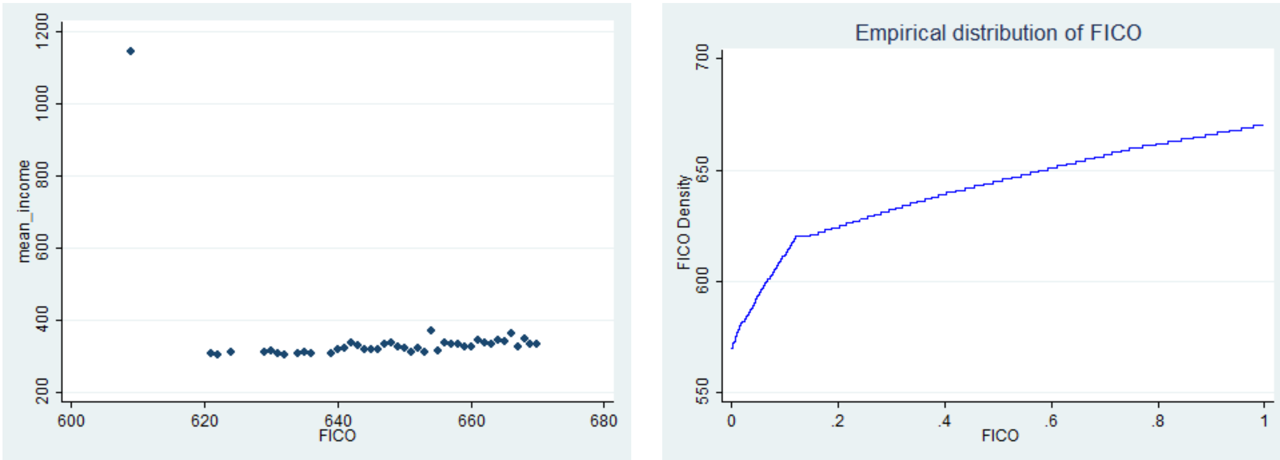
Figure 5: Scatter plots of covariates

smooth those which are right-skewed. Hence, for further analysis let's take the log of LTV[2].

Parametric regressions

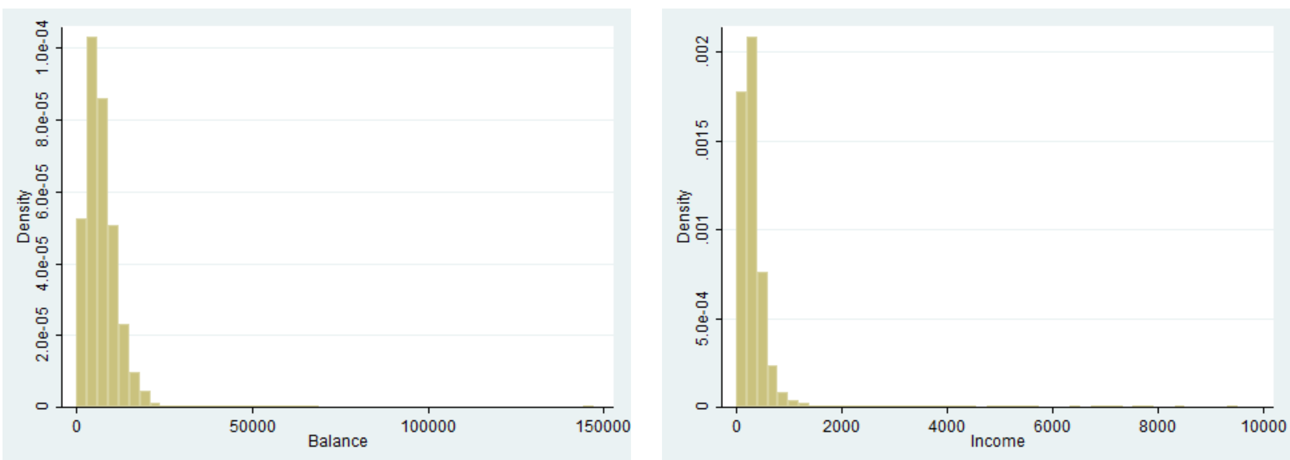
The main problem in our setup is that rate and delinquency both depend on the soft information, while delinquency in turn depends on rate. Probably, the best idea is to find the valid instrument but due to the case we have artificially constructed data it is impossible. Consequently, we have to make the simplifying assumption. The main assumption which we make is the following

[2]test for normality (sktest in Stata) rejected the hypothesis of normality



(a) placebo: high income (b) FICO_emp

Figure 6: Placebo test and empirical distribution



(a) hist_Balance (b) hist_Income (<10000)

Figure 7: Histograms of covariates

one:

Assumption: *The main channel through which the rate and unobserved effects which influence delinquency are connected is the soft information.*

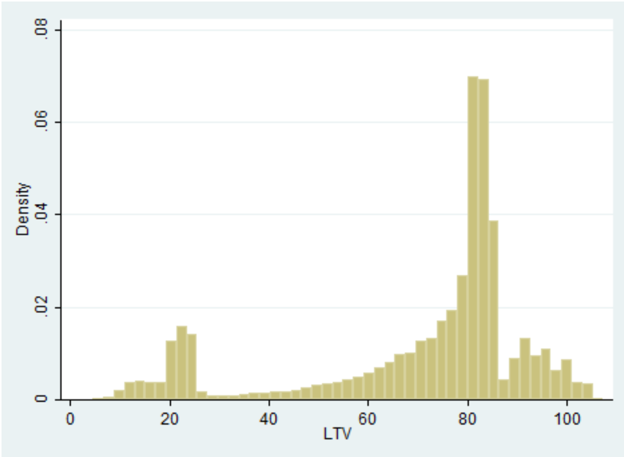
The assumption is based on the evidence that at $FICO = 620$ the number of defaults increases by about 20% because lenders require not full information about the borrower, so that the soft information is mostly not used (Keys et al., 2008).

Hence, the estimation is as follows:

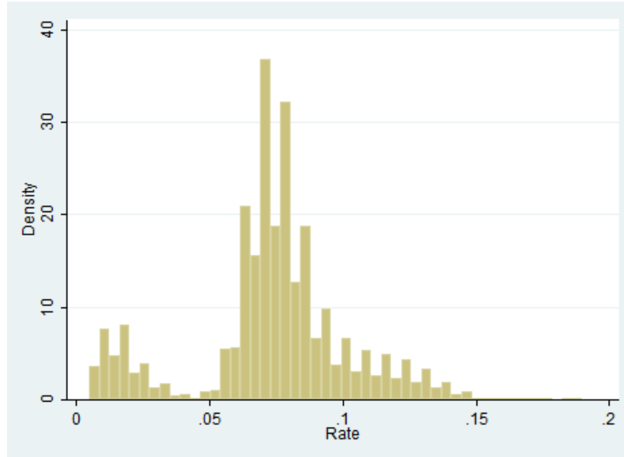
1 step. Estimate by the regression discontinuity design $rate_i = \lambda X_i + v_i$.

2 step. Save the estimates of the residuals $\hat{v}_i$.

3 step. Estimate β by Logit(Probit)-regression $Del_i = \beta rate_i + \gamma X_i + \alpha \hat{v}_i * (1 - w_i) + \epsilon_i$



(a) hist_LTV



(b) hist_Rate

Figure 8: Histograms of covariates

The last step allows us to obtain consistent estimates since the correlation between rate and ϵ is now eliminated because the estimate of soft information is incorporated into $\hat{v}_i * (1 - w_i)$ and included as the additional term. By the assumption above it allows us to eliminate endogeneity: the soft information is excluded from error term ϵ_i since we assume that the lower requirements to lenders who have credit rating more than 620 lead to the minor effect for the usage of soft information.

Below are presented two tables which reflect the information about the estimation of β. For the estimation of the first step 3 specifications are used (mostly all are highly significant, we've chosen those with the least standard errors). They are:

$$RD1 : rate_i = \alpha + \lambda_1 FICO_i + \lambda_2 W_i + v_i$$

$$RD2 : rate_i = \alpha + \lambda_1 (FICO_i - 620) + \lambda_2 W_i + \lambda_3 W_i * (FICO_i - 620) + v_i$$

$$RD3 : rate_i = \alpha + \lambda_1 (FICO_i - 620) + \lambda_2 W_i + \lambda_3 W_i * (FICO_i - 620) + \lambda_4 W_i * (FICO_i - 620)^2 + v_i$$

In the table *FICO1* reflects the FICO minus threshold in the corresponding power, while *inter1* and *inter2* reflect the interactions between treatment W_i and $FICO_i$.

Logit regession was chosen as a specification which estimates β, *residuals* reflect $\hat{v}_i$ and *int* the interaction term $\hat{v}_i * (1 - w_i)$, motivation for which is stated above.

All the specifications in Table 2 show highly significant coefficient of β. So that increase in rate by 1% increases the probability of delinquency by 20%.

Comparing the results after adding the control variables (consider Table 3), we can see that in our specifications variables *Broker* and *Hard* are insignificant (surprisingly contrary to the results predicted by theory). However, lucky we are, since we can exclude them and do not bother about the third factors driving the results behind. The more complicated specification as ones above are not good after adding the controls, so we leave the simpliest one. The result stays the same in sign but drops dramatically in the magnitude. β equals 8.97 which means that 1% increase in rate will lead to 9% increase in probability of delinquency in comparison with 20% above.

Comparing our results with previous findings, Keys et al. conclude in their paper that securitization influence screening process of the lenders and make it less precise. Then, weaker screening process increases delinquency. It coincides with the results we get. Our results show that higher interest rate increases probability of delinquency. Higher securitisation can be a result of higher interest rate, because it leads to higher profitability. However, there can be more direct effect. When interest rates go up, it becomes more difficult for borrowers to pay debt. It leads to higher probability of default. In this case there will be a difference in loan origination channel. Here delinquency is associated with difficulty of borrowing at high interest rate. In Keys' paper delinquency is a result of poor screening process and the fact that some borrowers should not have got a loan.

	(1)	(2)	(3)	(4)	(5)	(6)
Variables	RD1	Logit1	RD2	Logit2	RD3	Logit3
FICO	-0.000144***	-0.00859***		-0.00855***		-0.00880***
	(1.16e-05)	(0.00106)		(0.00104)		(0.00106)
Rate		19.73***		19.78***		18.81***
		(4.423)		(4.362)		(4.009)
int		10.04***		10.05***		10.06***
		(1.824)		(1.833)		(1.832)
residuals		-13.53***		-13.58***		-12.61***
		(4.769)		(4.730)		(4.334)
w	-0.00865***		-0.00971***		-0.0112***	
	(0.00104)		(0.00105)		(0.00102)	
FICO1			-9.08e-05***		-9.08e-05***	
			(1.38e-05)		(1.38e-05)	
inter1			-5.94e-05***		0.000100***	
			(1.82e-05)		(2.75e-05)	
inter2					-3.01e-06***	
					(4.97e-07)	
Constant	0.174***	3.108***	0.0859***	3.083***	0.0859***	3.316***
	(0.00733)	(0.968)	(0.000863)	(0.951)	(0.000863)	(0.931)
Observations	83,165	83,165	83,165	83,165	83,165	83,165
R-squared	0.039		0.039		0.039	

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2: Parametric regression without controls

Variables	(1) RD1	(2) RR2	(3) Logit2
FICO	-0.000161***	-0.000161***	-0.0106***
	(6.63e-06)	(6.62e-06)	(0.000644)
w	-0.00745***	-0.00753***	-0.143***
	(0.000565)	(0.000561)	(0.0491)
Age	-0.000331***	-0.000331***	-0.0300***
	(3.93e-05)	(3.93e-05)	(0.00261)
Balance	-1.69e-06***	-1.69e-06***	-5.47e-06
	(1.25e-07)	(1.25e-07)	(5.53e-06)
Broker	-0.000178		
	(0.000156)		
Gender	-0.00331***	-0.00331***	-0.294***
	(0.000372)	(0.000372)	(0.0369)
Hard	-0.000135		
	(0.000159)		
Income	5.24e-06***	5.24e-06***	0.000231***
	(8.19e-07)	(8.19e-07)	(4.87e-05)
LTV	-0.000437***	-0.000437***	0.00326***
	(1.38e-05)	(1.38e-05)	(0.000806)
Refinance	-0.000262*	-0.000258*	-0.0180
	(0.000155)	(0.000156)	(0.0156)
White	-0.00195***	-0.00195***	0.308***
	(0.000368)	(0.000368)	(0.0390)
Rate			8.970***
			(1.073)
int4			9.649***
			(1.901)
Constant	0.244***	0.244***	6.480***
	(0.00469)	(0.00468)	(0.393)
Observations	75,109	75,109	75,109
R-squared	0.335	0.335	

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3: Parametric regressions with controls

Conclusion

The main purpose of this paper was to estimate the effect of interest rate on the propensity to default in presence of vector of observables. The main result we were able to obtain is that the relationship between borrowers' delinquencies and the interest rate is statistically significant and positive. The cornerstone of this research was to avoid endogeneity caused by positive correlation between the errors in two regressions.

The identiication strategy contained the following steps: a) to regress the interest rate on observable using regression discontinuity design (parametric and non-parametric estimation) and get the residuals; b) to plug those residuals as a measure of soft characteristics into the delinquency-on-rate regression, thus “clearing” the rate variable from its influence and, therefore, to make the variable exogenous as a regressor; c) to estimate the effect via binary choice model (logistic regression).

The validity of RD estimation relies on a number of conditions, which we check one by one. First, we have checked that no other threshold drives the result. Second, we have verified that there are no jumps in covariates at FICO = 620. Next, the placebo test involving the high-income group is done. Finally, discontinuity in the density of rate at the threshold is disproven. In addition, several variables tended to have right-skewed distribution; their logs were included in the model. This procedure ensured the feasibility of the correct RD estimation, which in turn helped to eliminate endogeneity in the principal regression.

References

Diamond, Douglas (1984), “Financial Intermediation And Delegated Monitoring”, Review of Economic Studies, 51, 393-414.

Drucker, Steven & Puri, Manju (2007), “On Loan Sales, Loan Contracting, and Lending Relationships”, Working Paper.

Holmstrom, Bengt &Tirole, Jean (1997), “Financial Intermediation, Loanable funds, and The Real Sector”, Quarterly Journal of Economics, 52, 663-692

Keys, B. J., Mukherjee, T. K., Seru, A., & Vig, V. (2008). Did securitization lead to lax screening? Evidence from subprime loans. Evidence from Subprime Loans (December 25, 2008). EFA.

Jiang, W., Nelson, A. A., & Vytlacil, E. (2014). Liar’s loan? Effects of origination channel and information falsification on mortgage delinquency. Review of Economics and Statistics, 96(1), 1-18.