Privacy and Financial Inclusion in Credit Markets: A Lenders' Perspective*

Sumit Agarwal† Pulak Ghosh‡ Shohini Kundu§ Nishant Vats¶

June 24, 2025

Abstract

How do privacy regulations impact financial inclusion? Discussions on privacy regulations globally have primarily focused on concerns such as price discrimination, data abuse, unethical advertising, and financial fraud. However, there is a lack of attention to the unequal effects these regulations may have on financial inclusion. This paper aims to address this gap by examining the consequences of a specific Google policy change that prohibited a FinTech lender in India from accessing contact details in the borrowers' phones. Utilizing a regression discontinuity setup, we examine the impact of this privacy regulation on financial inclusion. Before the Google policy change, the lender used the phone numbers of all contacts in the borrowers' phones as social collateral, i.e., the lender could contact anyone in the borrowers' contact list in case of default. This threat helped reduce the incidence of default, enabling the lender to adopt a somewhat more relaxed approach to screening. However, the implementation of the privacy regulation removed this threat. Using detailed application-level data, we find that the policy resulted in a significant increase in the rejection rate and a small increase in the number of applications. However, while the positive effect on demand is limited, the negative effect on acceptance is large. Specifically, our estimates indicate a 25.14% decline in application acceptance, coupled with a 3.5% increase in the number of loan applications. These results suggest that while privacy regulations may be successful in boosting demand for loans by addressing privacy concerns, they have large negative effects on acceptance rates. Notably, there was no observable change in the default rate, indicating that lenders, in response to regulations limiting social collateral, tightened their selection process. From the lender's perspective, their total revenues declined as well. Furthermore, our analysis revealed that this tightened selection disproportionately impacted individuals with low income, younger individuals, those new to credit, and individuals belonging to lower social status. Overall, the results suggest that while privacy regulations may increase loan demand, such regulations are likely to have a large negative effect on financial inclusion and lender's profits.

*

†National University of Singapore, Email: ushakri@yahoo.com
‡Indian Institute of Management Bangalore, Email: pulak.ghosh@iimb.ac.in
§UCLA Anderson School of Management, Email: shohini.kundu@anderson.ucla.edu
¶Olin Business School, Washington University at St Louis, Email: vats@wustl.edu

1 Introduction

The rise of digital financial services has the potential to revolutionize financial inclusion, particularly for underbanked populations. However, global privacy regulations, while crucial for safeguarding user data, can have unintended consequences. While discussions surrounding these regulations often address concerns like price discrimination, data abuse, and financial fraud, their potential to disproportionately impact financial inclusion has received less scrutiny.

In this paper, we examine the impact of privacy regulation on financial inclusion. We analyze the consequences of a specific Google privacy regulation in India that restricted FinTech lenders' access to call detail records (CDRs). Previously, CDRs served as an auxiliary data source for assessing creditworthiness. However, this practice raised significant ethical concerns. The ability to contact an individual's social network in case of default posed a risk of undue pressure and potential harassment, compromising borrower privacy and potentially hindering their ability to access financial services.

Thus, the relation between privacy regulation and financial inclusion is empirically unclear. On the one hand, privacy regulations are essential to safeguard user data and prevent its misuse. They empower individuals with control over their personal information and protect them from exploitation. This fosters trust in the financial system and encourages responsible data handling practices. On the other hand, stringent privacy regulations can inadvertently hinder financial inclusion. Limiting access to certain data points traditionally used for credit assessment might make it challenging for lenders, particularly FinTech institutions, to effectively evaluate the creditworthiness of potential borrowers. This could lead to stricter selection criteria, potentially excluding individuals with limited credit history or those belonging to underbanked populations who rely heavily on alternative financial services.

Identifying the impact of privacy regulations on financial inclusion is a major empirical challenge. Several factors, beyond just the regulation itself, can influence both the implementation of the regulation and the observed changes in financial inclusion. To address this challenge, we utilize a Regression Discontinuity Design (RDD). We leverage a specific Google privacy regulation that created a sharp discontinuity in a FinTech lender's CDR. Google, adhering to its privacy regulations, issued a directive to a FinTech lender requiring them to cease collecting call detail records from their customers' phones. Non-compliance with this directive would result in the app's removal from the Google Play Store. Consequently, the FinTech halted the collection of call log data from January 1, 2019. This event serves as a natural experiment for the RDD analysis.

The RDD methodology leverages the Google privacy regulation as a natural experiment to isolate the causal effect of the privacy regulation on loan application outcomes. The RDD analysis unveils significant consequences of the Google privacy regulation on various aspects, including loan application rejection rates, demand for loans, default rates, and lender profitability. We

observe a significant increase in loan application rejections, with a 25.14% decline in approvals, indicating a stricter selection process in response to the regulation's limitations on data access. In addition, we find an uptick of 3.5% in the number of applications, suggesting that the privacy regulation potentially assuaged user concerns and may have even encouraged a minor rise in loan applications. Importantly, there is no observable change in default rates, signaling that the more rigorous selection process did not adversely affect borrower behavior. However, the FinTech lender's total revenue declines due to the significant decrease in approved loans.

Further, our analysis reveals that the tightened selection criteria disproportionately impacts specific demographics. Assessing creditworthiness for new borrowers, particularly those lacking extensive financial data, becomes more challenging without access to alternative data points like CDRs. This limitation is further exacerbated by privacy regulation. We find that the privacy regulation adversely affects individuals with lower income, younger applicants, those with limited credit history, and individuals belonging to a lower social strata. Individuals belonging to these demographics have limited traditional credit footprints, making it difficult for lenders to assess their creditworthiness solely based on the remaining data. Thus, while privacy regulation may address privacy concerns and potentially fosters a slight increase in loan demand, there is a substantial negative impact on access to financial services as evinced by the substantial decline in loan approvals. Our findings suggest that privacy regulation can hinder access to financial services, particularly for vulnerable populations.

Overall, this study underscores the critical need for a balanced approach that prioritizes robust data privacy while simultaneously fostering financial inclusion.

2 Institutional Setting

Our analysis focuses on a prominent FinTech lender operating in India's digital lending ecosystem. The company was established in 2016 and represents a new generation of alternative lending platforms that emerged to serve India's underbanked millennial population. The institutional setting is particularly relevant for understanding privacy regulation impacts on financial inclusion, as this lender pioneered the use of alternative data sources for credit assessment.

2.1 Business Model and Target Market

The FinTech operates primarily through a mobile application platform, providing unsecured short-term personal loans. The company specifically targets millennials and individuals from marginalized backgrounds who are often excluded from traditional banking services due to limited or no credit history. Loan amounts range from approximately ₹10,000 to ₹300,000 (roughly $120 to $3,600), with repayment tenures spanning from 15 days to 18 months. The average loan amount is approximately ₹35,000-40,000 ($420-480). SK: Check these numbers using our dataset. These numbers are based on articles. we can make a summary stats table and refer to it here.

The lender's primary value proposition centers on rapid loan processing and disbursement, with the ability to approve and disburse loans within minutes of application submission. This speed advantage is crucial in the competitive Indian FinTech market, where traditional banks often require weeks for loan processing. The company processes thousands of loan applications daily and has maintained high customer retention rates, with approximately 70% of borrowers returning for additional loans.

2.2 Alternative Credit Scoring Methodology

The most distinctive feature of this institutional setting is the lender's proprietary alternative credit scoring system, which we refer to as the "Alternative Credit Quotient" (ACQ) to maintain anonymity. This system represents a significant departure from traditional credit assessment methods that rely primarily on credit bureau data and financial transaction history. The ACQ algorithm incorporates multiple non-traditional data sources to evaluate creditworthiness, including:

- **Mobile device behavioral data:** Usage patterns, device characteristics, and application usage
- **Social media footprint:** Social network connections, activity levels, and engagement patterns
- **Contact list analysis:** The nature and composition of the borrower's contact network
- **SMS and communication patterns:** Salary credit confirmations, bill payment receipts, and communication behavior

- **Educational and professional profile:** Employment status, educational background, and career trajectory
- **Geographic and demographic indicators:** Location stability and demographic characteristics

The algorithm generates scores ranging from 0 to approximately 1000, with higher scores indicating better creditworthiness. This scoring system is designed to be dynamic and forward-looking, updating in real-time as new behavioral data becomes available. Unlike traditional credit scores that primarily reflect historical financial behavior, the ACQ attempts to predict future repayment behavior based on current behavioral patterns and social signals.

2.3 Changees to Privacy Policy Chang

Prior to January 1, 2019, the lender's data collection practices were comprehensive, requiring borrowers to grant extensive permissions for accessing various smartphone data streams. Central to their credit assessment process was access to call detail records (CDRs), which encompassed contact lists, SMS records, call logs, and communication metadata. These CDRs served as a critical auxiliary data source for creditworthiness evaluation, forming an integral component of their alternative credit scoring algorithm.

The use of CDRs enabled the lender to implement what could be characterized as "social collateral" - the implicit threat that the borrower's social network could be contacted in cases of payment default. While this mechanism potentially reduced default rates by creating reputational stakes, it raised significant ethical and privacy concerns. The ability to access and potentially contact individuals in a borrower's social network created risks of undue pressure, social embarrassment, and harassment, thereby compromising borrower privacy and autonomy.

In January 2019, Google implemented a privacy policy directive that fundamentally altered this data collection landscape. The directive specifically prohibited the lender from continuing to collect CDRs from their customers' mobile devices. This policy change was enforced through Google's platform control mechanism – non-compliance would result in the immediate removal of the lender's application from the Google Play Store, effectively eliminating their primary customer acquisition and service delivery channel.[1] More broadly, in May 2025, the Reserve Bank of India (RBI) issued Reserve Bank of India (Digital Lending) Directions requiring digital lending apps to desist from accessing mobile phone resources such as file and media, contact list, call logs, telephony functions etc.

Faced with this ultimatum, the lender had no viable alternative but to comply immediately. Consequently, the collection of call log data and related CDR information ceased entirely on

[1]The enforcement mechanism was particularly significant given that the vast majority of the lender's user base accessed services through Android devices via the Google Play Store.

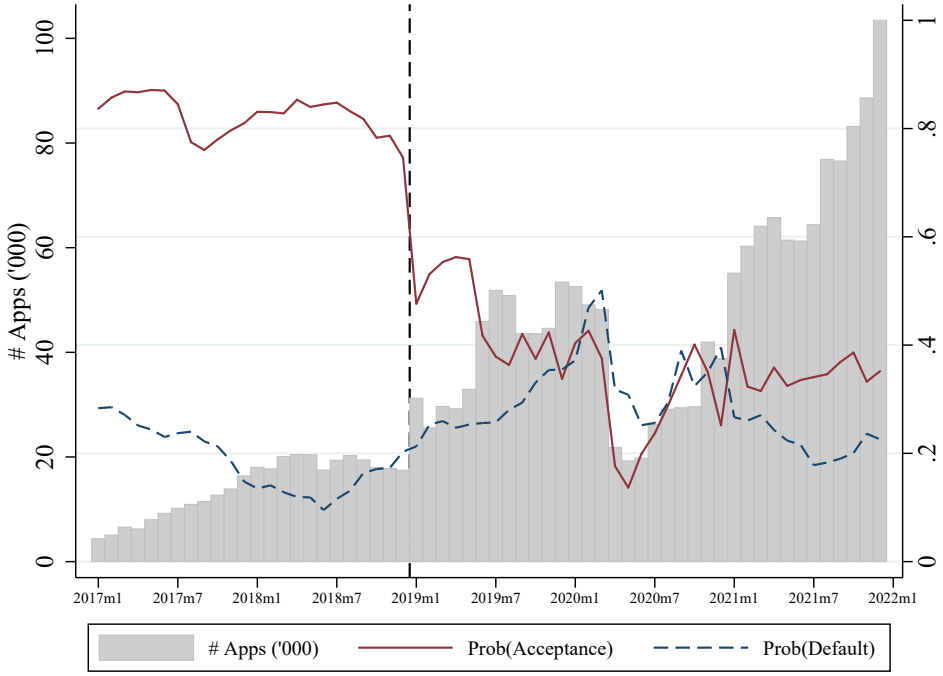
January 1, 2019. This policy implementation was executed without advance notice to users or extensive media coverage, creating a clean regulatory discontinuity that was largely external to the lender's control and independent of other market or business factors.

The abrupt nature of this privacy policy change created several important conditions for empirical analysis. First, the timing was exogenous to both borrower behavior and the lender's business strategy, as it was imposed by an external platform provider rather than emerging from internal business decisions or broader regulatory changes. Second, the binary nature of the change – from full CDR access to complete prohibition – creates a sharp discontinuity in data availability that affected all loan applications processed after the implementation date. Third, the lender's core business model, target market, and operational procedures remained otherwise unchanged, ensuring that the policy change represents the primary source of variation in lending outcomes during the study period. Overall, the research design benefits from the clean identification provided by this regulatory discontinuity, which allows for robust causal inference regarding the relationship between privacy protection measures and financial inclusion outcomes.

This institutional context provides a clean natural experiment for examining the impact of privacy regulation on financial inclusion. The lender's substantial reliance on alternative data sources, particularly CDRs, combined with their focus on underbanked populations with limited traditional credit histories, creates a setting where privacy restrictions can generate measurable and significant effects on both lending decisions and borrower access to credit. The availability of detailed application-level data spanning the policy change, combined with well-defined customer segments and consistent data collection practices, enables precise measurement of how privacy regulations differentially affect various demographic groups and lending outcomes.

3 Aggregate Effects

Figure 1: Number of Applications, Probability of Acceptance and Default over time



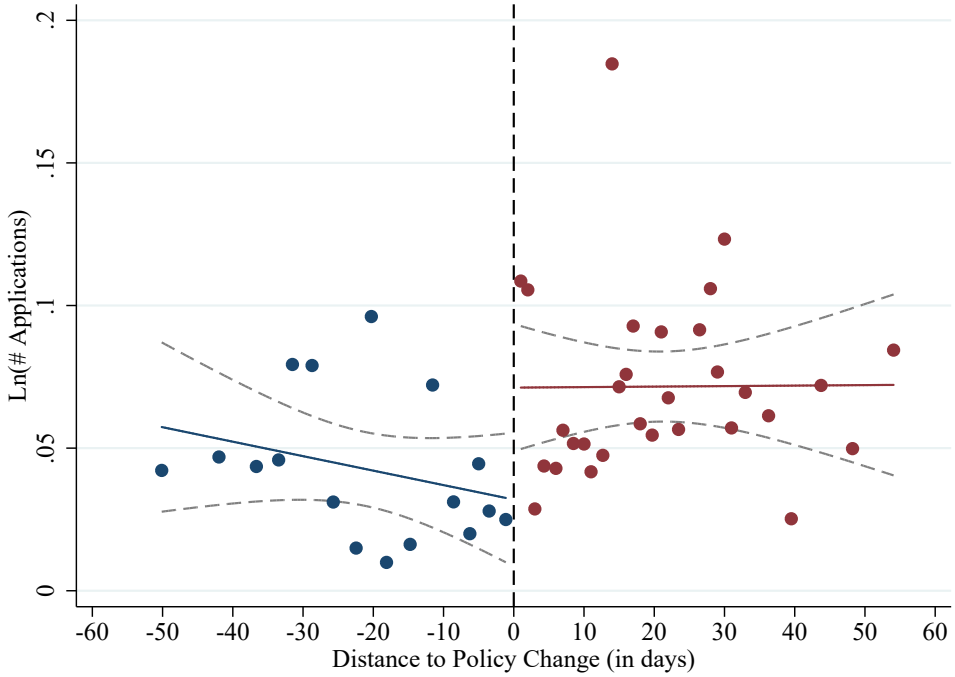
This figure illustrates the number of monthly applications, the monthly average probability of acceptance, the monthly average probability of default from January 2017 until December 2021.

Table 1: Summary of Key Results

	Measured Using	Result	Magnitude	Comments
Demand	# of Daily Applications	Increase	3.50%	Distribution is unchanged
Supply	Application Acceptance	Decrease	25.14%	Implications for inclusion
Ex-Post Risk	Default	No Effect	-	Monitoring is substituted by Screening

4 Effect on Demand or Number of Applications

Figure 2: Effect on Number of Applications: RD Estimate



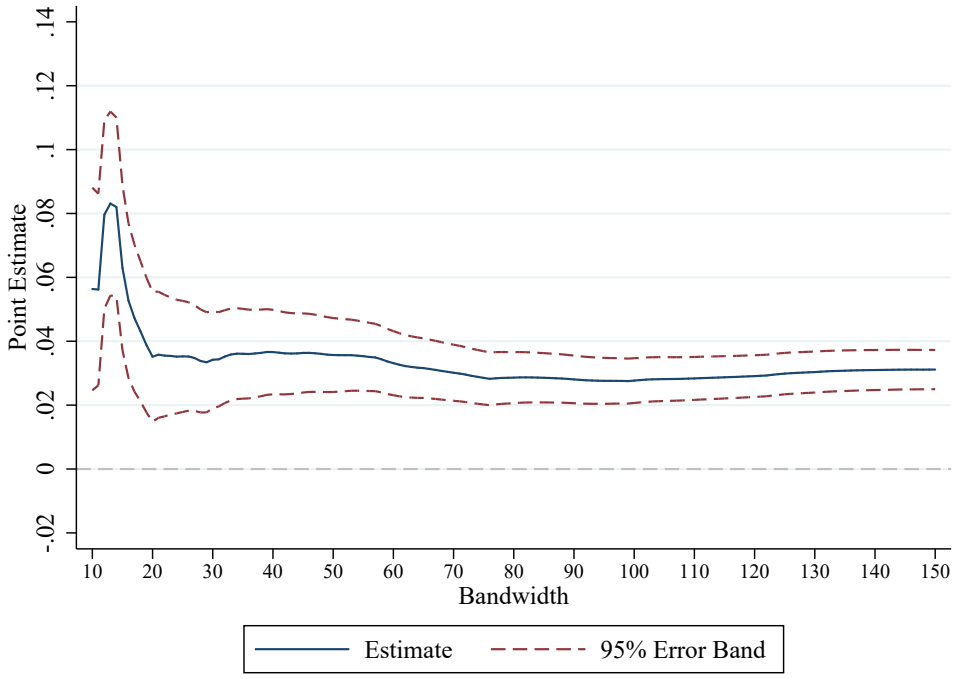
The figure presents RD plots for natural logarithm of the number of applications during the day at different bins along the running variable (distance to policy change) as well as with a local linear trend estimated separately, along with 95% confidence intervals, on each side of the discontinuity. The figure is estimated after including ZIPcode × loan reason and the day of the week fixed effects. The unit of observation is a day. The red dots indicate days after the call log collection ban policy and the blue dots indicate days before the call log collection ban policy was implemented.

Table 2: Effect of the Policy Change on Number of Applications

Dep Var: Ln(# Applications)	(1)	(2)	(3)	(4)	(5)
Post	0.0564***	0.0630***	0.0352***	0.0353***	0.0342***
	(0.0162)	(0.0132)	(0.0104)	(0.0089)	(0.0077)
ZIP Code X Loan Purpose FE	Yes	Yes	Yes	Yes	Yes
Day of the Week FE	Yes	Yes	Yes	Yes	Yes
Bandwidth	10	15	20	25	30
Kernel	Triangular	Triangular	Triangular	Triangular	Triangular
# Obs	6,805	12,280	17,130	23,230	30,494
R^2	0.4162	0.3449	0.304	0.2831	0.2689

This table presents the RD estimates using the natural logarithm of the number of applications during the day as the dependent variable. We use different bandwidths of 10, 15, 20, 25, and 30 days on either side of the policy change date (January 1, 2019), in columns (1), (2), (3), (4), and (5) respectively. All regressions include ZIPcode × loan purpose and the day of the week fixed effects, a local linear specification estimated separately on each side of the policy change date, and use a triangular kernel. Standard errors reported in parentheses are clustered at the ZIPcode level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure 3: Number of Application: RD Estimates for various Bandwidth choices



This figure presents RD estimates and 95% confidence interval for the RD specification using the natural logarithm of number of applications during the day as the dependent variable. We estimate the RD specification for several bandwidths between 2 days and 150 days in increments of 1 day. The Y-axis reports the RD estimate, and the X-axis reports the bandwidth used to estimate the RD coefficient. The unit of observation is a day. All regressions include ZIPcode × loan reason and the day of the week fixed effects, a local linear specification estimated separately on each side of the policy change date, and use a triangular kernel. The solid red line reports the RD estimate, and the dashed grey lines indicate the 95% confidence bands calculated based on standard errors clustered at the block level.

Table 3: Robustness to Covariates: Effect of the Policy Change on Application Acceptances

Dep Var: LN(# Applications)	(1)	(2)	(3)	(4)	(5)	(6)
Post	0.0286***	0.0347***	0.0324***	0.0387***	0.0482***	0.0342***
	(0.0044)	(0.0056)	(0.0046)	(0.0057)	(0.0082)	(0.0077)
ZIP Code FE		Yes		Yes		
Loan Purpose FE			Yes	Yes		
ZIP Code X Loan Purpose FE					Yes	Yes
Day of the Week FE						Yes
Bandwidth	30	30	30	30	30	30
Kernel	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular
# Obs	39,534	36,928	39,534	36,928	30,494	30,494
R^2	0.0021	0.161	0.0091	0.1705	0.2614	0.2689

This table presents the RD estimates using the natural logarithm of the number of applications during the day as the dependent variable. We use a bandwidth of 30 days on either side of the policy change date (January 1, 2019). All regressions include different set of fixed effects, a local linear specification estimated separately on each side of the policy change date, and use a triangular kernel. 95% confidence bands calculated based on standard errors clustered at the ZIPcode level. Standard errors reported in parentheses are clustered at the ZIPcode level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

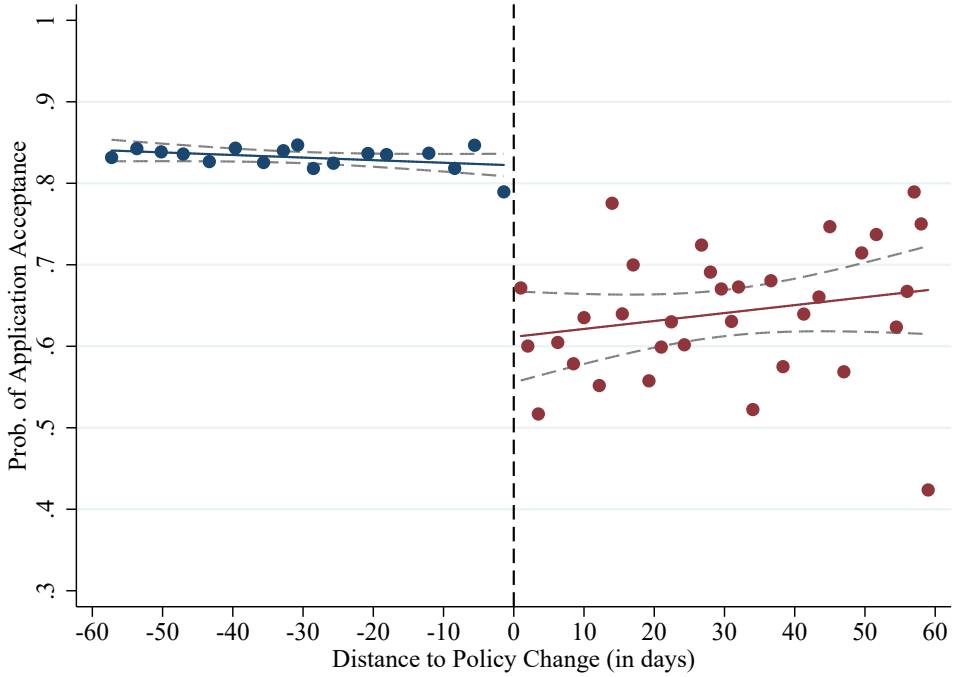
Table 4: Effect of the Policy Change on Composition of Applications

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	LN(Amount Asked)	LN(Income)	LN(DTI)	Credit Score	Social Score	LN(Age)	Female
Post	-0.0129	-0.0089	-0.004	0.6544	0.0085	-0.0173***	0.014
	(0.0356)	(0.0208)	(0.0269)	(2.0006)	(0.0061)	(0.0055)	(0.0103)
ZIP Code X Loan Purpose FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day of the Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bandwidth	30	30	30	30	30	30	30
Kernel	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular
# Obs	30,625	30,625	30,625	30,025	30,618	30,566	30,625
R^2	0.348	0.3606	0.3128	0.311	0.2984	0.3443	0.3016

This table presents the RD estimates indicating the effect of the policy change on the composiiton of the applications. Column 1 uses the natural logarithm of the application loan amount as the dependent variable. Column 2 uses the natural logarithm of the monthly customer income as the dependent variable. Column 3 uses the natural logarithm of the ratio of loan amount and monthly income (DTI, debt to income ratio) as the dependent variable. Column 4 uses the applicants CIBIL credit score as the dependent variable. Column 5 uses the proprietary social score as the dependent variable. Column 6 uses the natural logarithm of the applicants age as the dependent variable. Column 7 uses a binary variable as the dependent variable that takes a value of one for female applicants and zero otherwise. We use a bandwidth of 30 days on either side of the policy change date (January 1, 2019). All regressions include different set of fixed effects, a local linear specification estimated separately on each side of the policy change date, and use a triangular kernel. 95% confidence bands calculated based on standard errors clustered at the ZIPcode level. Standard errors reported in parentheses are clustered at the ZIPcode level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

5 Effect on Application Acceptances

Figure 4: Effect on Acceptance: RD Estimate



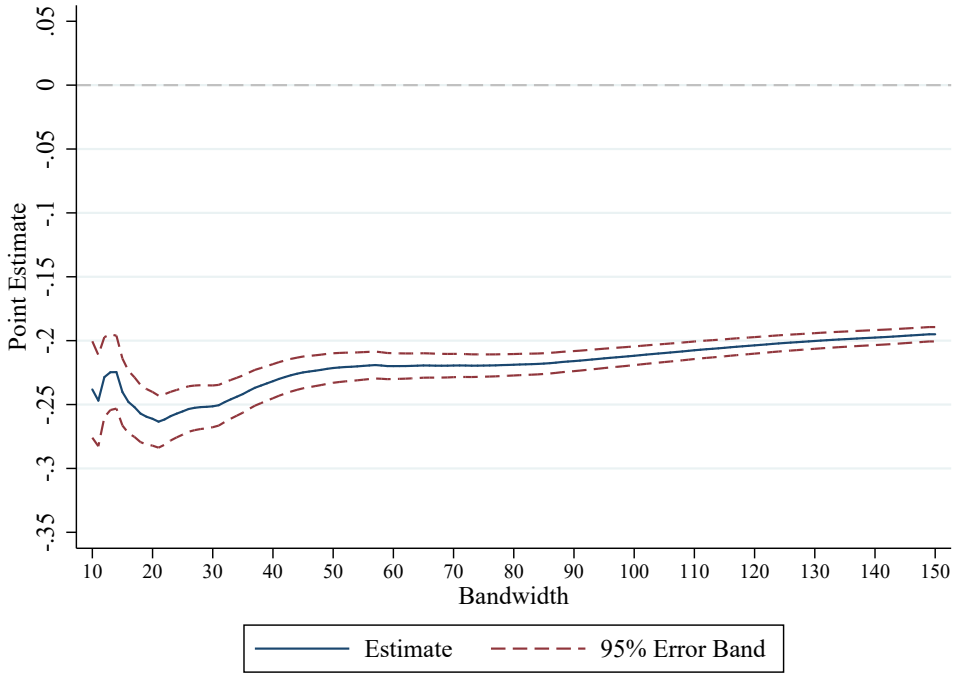
The figure presents RD plots for the probability of acceptance at different bins along the running variable (distance to policy change) as well as with a local linear trend estimated separately, along with 95% confidence intervals, on each side of the discontinuity. The figure is estimated after including ZIPcode × loan reason and the day of the week fixed effects. The unit of observation is a day. The red dots indicate days after the call log collection ban policy and the blue dots indicate days before the call log collection ban policy was implemented.

Table 5: Effect of the Policy Change on Application Acceptances

Dep Var: Application Accepted	(1)	(2)	(3)	(4)	(5)
Post	-0.2382***	-0.2402***	-0.2611***	-0.2554***	-0.2514***
	(0.0192)	(0.0134)	(0.0107)	(0.0093)	(0.0083)
ZIP Code X Loan Purpose FE	Yes	Yes	Yes	Yes	Yes
Day of the Week FE	Yes	Yes	Yes	Yes	Yes
Bandwidth	10	15	20	25	30
Kernel	Triangular	Triangular	Triangular	Triangular	Triangular
# Obs	7,089	12,698	17,236	23,231	30,625
R^2	0.7652	0.7555	0.7438	0.741	0.7392

This table presents the RD estimates using the application acceptance status as the dependent variable. The dependent variable takes a value of one if the application was accepted and zero otherwise. We use different bandwidths of 10, 15, 20, 25, and 30 days on either side of the policy change date (January 1, 2019), in columns (1), (2), (3), (4), and (5) respectively. All regressions include ZIPcode × loan reason and the day of the week fixed effects, a local linear specification estimated separately on each side of the policy change date, and use a triangular kernel. Standard errors reported in parentheses are clustered at the ZIPcode level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure 5: Application Acceptance: RD Estimates for various Bandwidth choices



This figure presents RD estimates and 95% confidence interval for the RD specification using application acceptance status as the dependent variable. We estimate the RD specification for several bandwidths between 2 days and 150 days in increments of 1 day. The Y-axis reports the RD estimate, and the X-axis reports the bandwidth used to estimate the RD coefficient. The unit of observation is a day. All regressions include ZIPcode × loan reason and the day of the week fixed effects, a local linear specification estimated separately on each side of the policy change date, and use a triangular kernel. The solid red line reports the RD estimate, and the dashed grey lines indicate the 95% confidence bands calculated based on standard errors clustered at the block level.

Table 6: Robustness to Covariates: Effect of the Policy Change on Application Acceptances

Dep Var: Application Accepted	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Post	-0.4565***	-0.4530***	-0.2450***	-0.2514***	-0.2511***	-0.2509***	-0.2488***	-0.2488***	-0.2404***	-0.2428***	-0.2430***
	(0.0076)	(0.0087)	(0.0068)	(0.0083)	(0.0083)	(0.0084)	(0.0083)	(0.0084)	(0.0080)	(0.0080)	(0.0079)
Female (=1)						-0.0178**	-0.0094	-0.0118	-0.0048	-0.0059	-0.0037
						(0.0080)	(0.0079)	(0.0081)	(0.0079)	(0.0078)	(0.0078)
ZIP Code FE		Yes	Yes								
Loan Purpose FE			Yes								
ZIP Code X Loan Purpose FE				Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day of the Week FE				Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Education FE					Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age Decile FE							Yes	Yes	Yes	Yes	Yes
Credit Score Decile FE								Yes	Yes	Yes	Yes
Social Score Decile FE									Yes	Yes	Yes
Loan Maturity FE										Yes	Yes
Income Decile FE											Yes
Bandwidth	30	30	30	30	30	30	30	30	30	30	30
Kernel	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular
# Obs	33,584	32,733	32,733	30,625	30,625	30,625	30,566	29,972	29,970	29,970	29,970
R^2	0.202	0.3263	0.6606	0.7392	0.7402	0.7403	0.7424	0.7444	0.7588	0.761	0.764

This table presents the RD estimates using the application acceptance status as the dependent variable. The dependent variable takes a value of one if the application was accepted and zero otherwise. We use a bandwidth of 30 days on either side of the policy change date (January 1, 2019). All regressions include different set of fixed effects, a local linear specification estimated separately on each side of the policy change date, and use a triangular kernel. 95% confidence bands calculated based on standard errors clustered at the ZIPcode level. Standard errors reported in parentheses are clustered at the ZIPcode level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

5.1 Heterogeneity in Application Acceptances

Table 7: Heterogeniety & Effect of the Policy Change on Application Acceptances

Dep Var: Application Accepted	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post	-0.2514***	-0.2393***	-0.1657***	-0.1818***	-0.2147***	-0.2079***	-0.2472***	-0.0107
	(0.0083)	(0.0128)	(0.0096)	(0.0116)	(0.0100)	(0.0106)	(0.0088)	-0.0179
Sub-Prime Credit Score X Post		-0.0207						-0.0212
		(0.0160)						(0.0152)
Sub-Prime Credit Score		0.0025						0.0033
		(0.0090)						(0.0084)
Low Social Score X Post			-0.1745***					-0.1525***
			(0.0132)					(0.0134)
Low Social Score			0.0072					0.0076
			(0.0083)					(0.0083)
New Customer X Post				-0.1220***				-0.1229***
				(0.0144)				(0.0147)
New Customer				0.0022				0.0036
				(0.0088)				(0.0089)
Low Income X Post					-0.0817***			-0.0944***
					(0.0150)			(0.0150)
Low Income					-0.0046			-0.0034
					(0.0086)			(0.0087)
Young X Post						-0.0829***		-0.0746***
						(0.0148)		(0.0147)
Young						-0.0055		-0.0026
						(0.0089)		(0.0087)
Female X Post							-0.0374	-0.0254
							(0.0250)	(0.0239)
Female							0.007	0.0097
							(0.0144)	(0.0144)
ZIP Code X Loan Purpose FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day of the Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bandwidth	30	30	30	30	30	30	30	30
Kernel	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular
# Obs	30,625	30,025	30,618	30,625	30,625	30,566	30,625	29,970
R^2	0.7392	0.7406	0.7535	0.7464	0.7425	0.7416	0.7393	0.7667

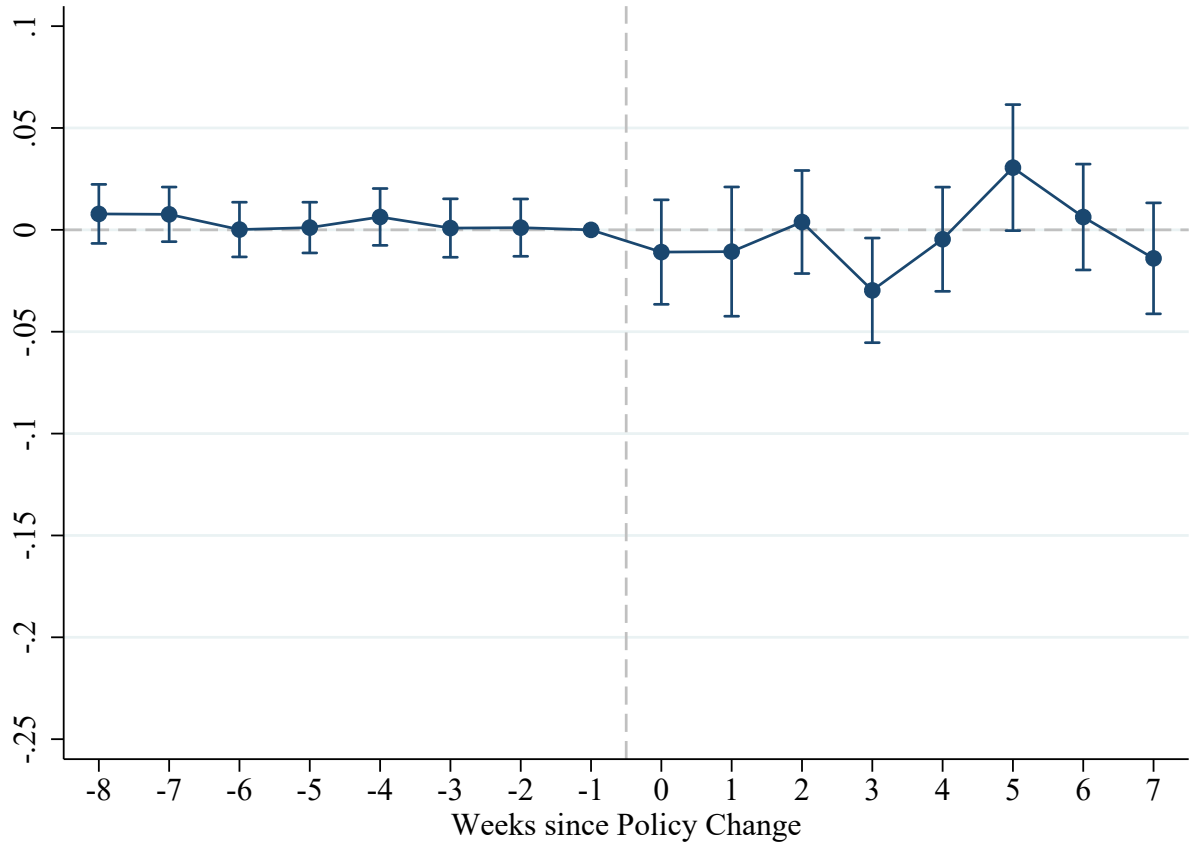
This table presents the RD estimates using the application acceptance status as the dependent variable. The dependent variable takes a value of one if the application was accepted and zero otherwise. All regressions use a bandwidth of 30 days on either side of the policy change date (January 1, 2019). All regressions include ZIPcode × loan reason and the day of the week fixed effects, a local linear specification estimated separately on each side of the policy change date, and use a triangular kernel. Standard errors reported in parentheses are clustered at the ZIPcode level. Sub-prime credit score is a binary variable that takes a value of one for customers with CIBIL credit score less than 680, and zero otherwise. Low social score is a binary variable that takes a value of one for customers with proprietary social score less than the sample median value, and zero otherwise. New customer is a binary variable for customers who are first time applicants. Low income is a binary variable that takes a value of one for customers with monthly income less than the sample median value, and zero otherwise. young is a binary variable that takes a value of one for customers with age less than the sample median value, and zero otherwise. Female is a binary variable that takes a value of one for females, and zero otherwise. $^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$.

Table 8: Strict Identification: Heterogeniety & Effect of the Policy Change on Application Acceptances

Dep Var: Application Accepted	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Sub-Prime Credit Score X Post	-0.0099						-0.0068
	(0.0165)						(0.0160)
Sub-Prime Credit Score	-0.0046						-0.0051
	(0.0077)						(0.0073)
Low Social Score X Post		-0.1636***					-0.1419***
		(0.0138)					(0.0141)
Low Social Score		0.0059					0.0040
		(0.0064)					(0.0066)
New Customer X Post			-0.1167***				-0.1159***
			(0.0152)				(0.0154)
New Customer			0.0025				0.0005
			(0.0076)				(0.0075)
Low Income X Post				-0.0718***			-0.0840***
				(0.0164)			(0.0164)
Low Income				-0.0078			-0.0078
				(0.0072)			(0.0074)
Young X Post					-0.0813***		-0.0779***
					(0.0152)		(0.0151)
Young					-0.0026		0.0038
					(0.0075)		(0.0077)
Female X Post						-0.0179	0.0000
						(0.0248)	(0.0239)
Female						-0.0121	-0.0130
						(0.0108)	(0.0116)
ZIP Code X Loan Purpose FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day of the Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
ZIP Code X Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Loan Purpose X Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bandwidth	30	30	30	30	30	30	30
Kernel	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular
# Obs	30,025	30,618	30,625	30,625	30,566	30,625	29,970
R^2	0.7406	0.7535	0.7464	0.7425	0.7416	0.7393	0.7667

This table presents the RD estimates using the application acceptance status as the dependent variable. The dependent variable takes a value of one if the application was accepted and zero otherwise. All regressions use a bandwidth of 30 days on either side of the policy change date (January 1, 2019). All regressions include ZIPcode × loan reason and the day of the week fixed effects, a local linear specification estimated separately on each side of the policy change date, and use a triangular kernel. Standard errors reported in parentheses are clustered at the ZIPcode level. Sub-prime credit score is a binary variable that takes a value of one for customers with CIBIL credit score less than 680, and zero otherwise. Low social score is a binary variable that takes a value of one for customers with proprietary social score less than the sample median value, and zero otherwise. New customer is a binary variable for customers who are first time applicants, and zero otherwise. Low income is a binary variable that takes a value of one for customers with monthly income less than the sample median value, and zero otherwise. Young is a binary variable that takes a value of one for customers with age less than the sample median value, and zero otherwise. Female is a binary variable that takes a value of one for females, and zero otherwise. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure 6: DID: Application Acceptance & Credit Scores

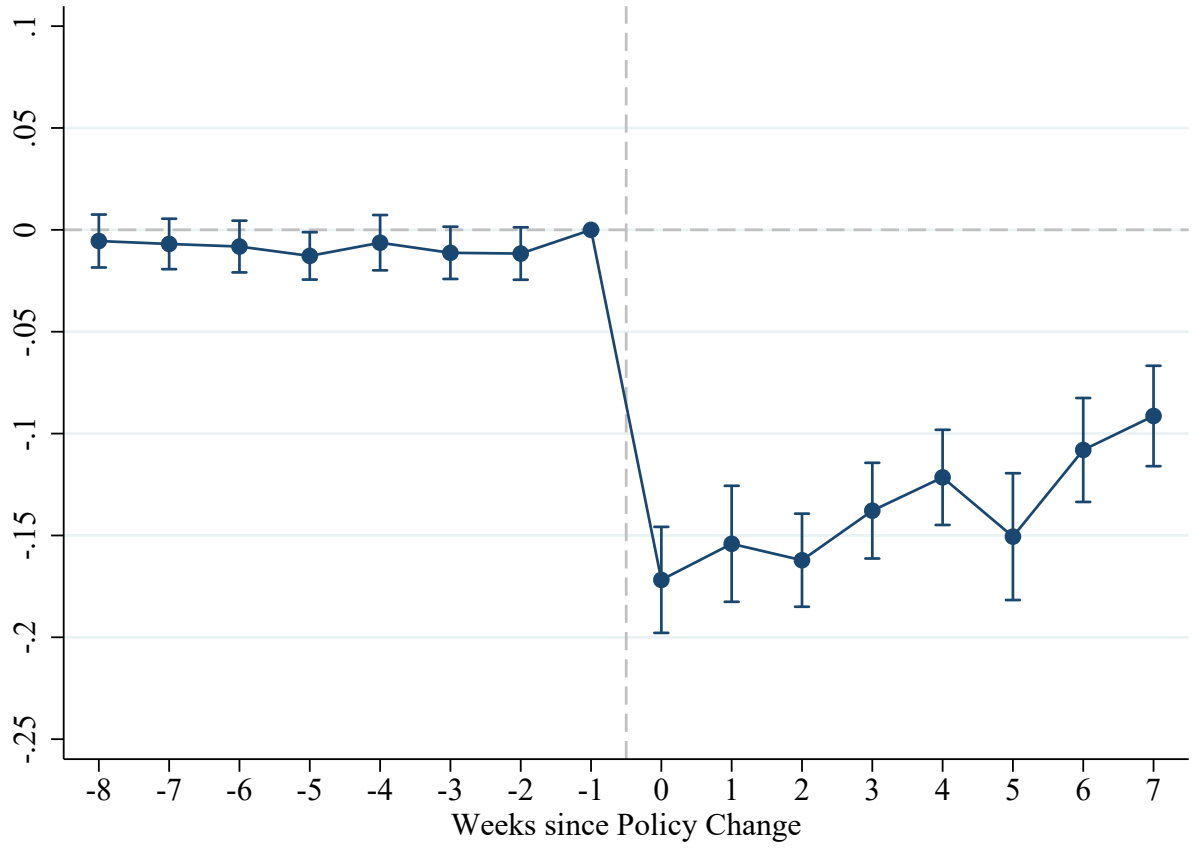


This figure presents the DID estimates and 95% confidence interval for the following specification using application acceptance status as the dependent variable.

$$y_{i,t} = \sum_{k=-8, k\neq -1}^{k=+7} \beta_k \cdot \text{Sub-prime Credit Score}_i \times \mathbb{I}(t = k) + \theta_{z,p} + \theta_{z,t} + \theta_{p,t} + \theta_w + \varepsilon_{i,t}$$

All regressions include ZIPcode × loan purpose ($\theta_{z,p}$), ZIPcode × week ($\theta_{z,t}$), loan purpose × week ($\theta_{p,t}$), and the day of the week (θ_w) fixed effects. Sub-prime credit score is a binary variable that takes a value of one for customers with CIBIL credit score less than 680, and zero otherwise. 95% confidence bands calculated based on standard errors clustered at the ZIPcode level.

Figure 7: DID: Application Acceptance & Social Scores

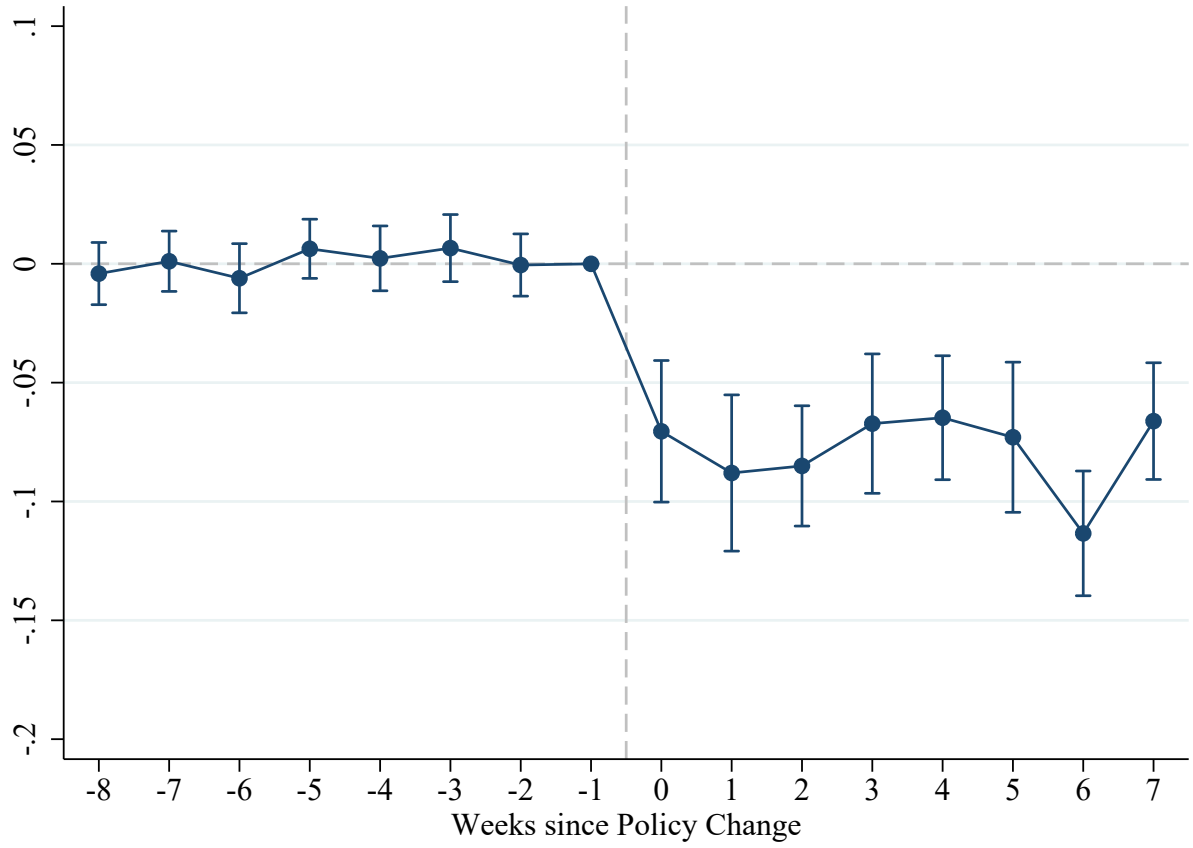


This figure presents the DID estimates and 95% confidence interval for the following specification using application acceptance status as the dependent variable.

$$y_{i,t} = \sum_{k=-8,k\neq-1}^{k=+7} \beta_k \cdot \text{Low Social Score}_i \times \mathbb{I}(t = k) + \theta_{z,p} + \theta_{z,t} + \theta_{p,t} + \theta_w + \varepsilon_{i,t}$$

All regressions include ZIPcode × loan purpose ($\theta_{z,p}$), ZIPcode × week ($\theta_{z,t}$), loan purpose × week ($\theta_{p,t}$), and the day of the week (θ_w) fixed effects. Low social score is a binary variable that takes a value of one for customers with proprietary social score less than the sample median value, and zero otherwise. 95% confidence bands calculated based on standard errors clustered at the ZIPcode level.

Figure 8: DID: Application Acceptance & Monthly Customer Income

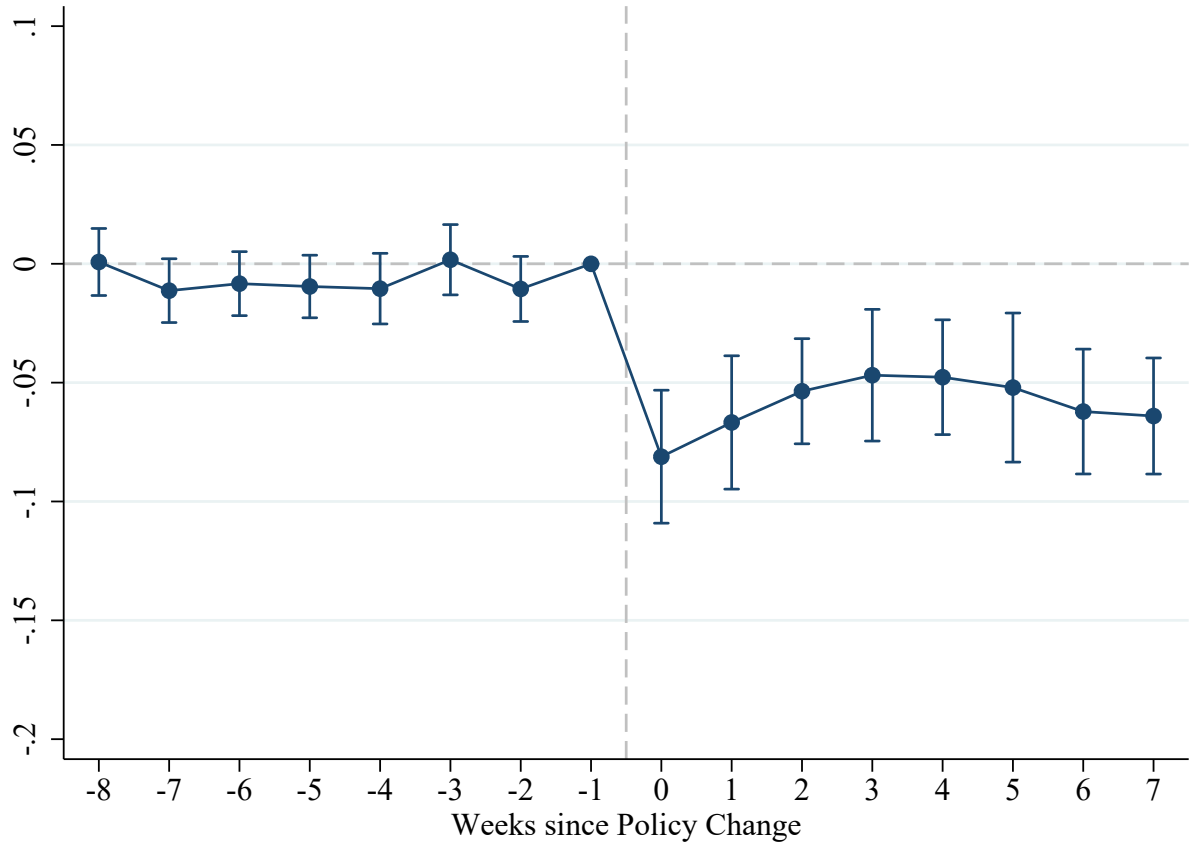


This figure presents the DID estimates and 95% confidence interval for the following specification using application acceptance status as the dependent variable.

$$y_{i,t} = \sum_{k=-8, k\neq -1}^{k=+7} \beta_k \cdot \text{Low Income}_i \times \mathbb{I}(t = k) + \theta_{z,p} + \theta_{z,t} + \theta_{p,t} + \theta_w + \varepsilon_{i,t}$$

All regressions include ZIPcode × loan purpose ($\theta_{z,p}$), ZIPcode × week ($\theta_{z,t}$), loan purpose × week ($\theta_{p,t}$), and the day of the week (θ_w) fixed effects. Low income is a binary variable that takes a value of one for customers with monthly income less than the sample median value, and zero otherwise. 95% confidence bands calculated based on standard errors clustered at the ZIPcode level.

Figure 9: DID: Application Acceptance & Age

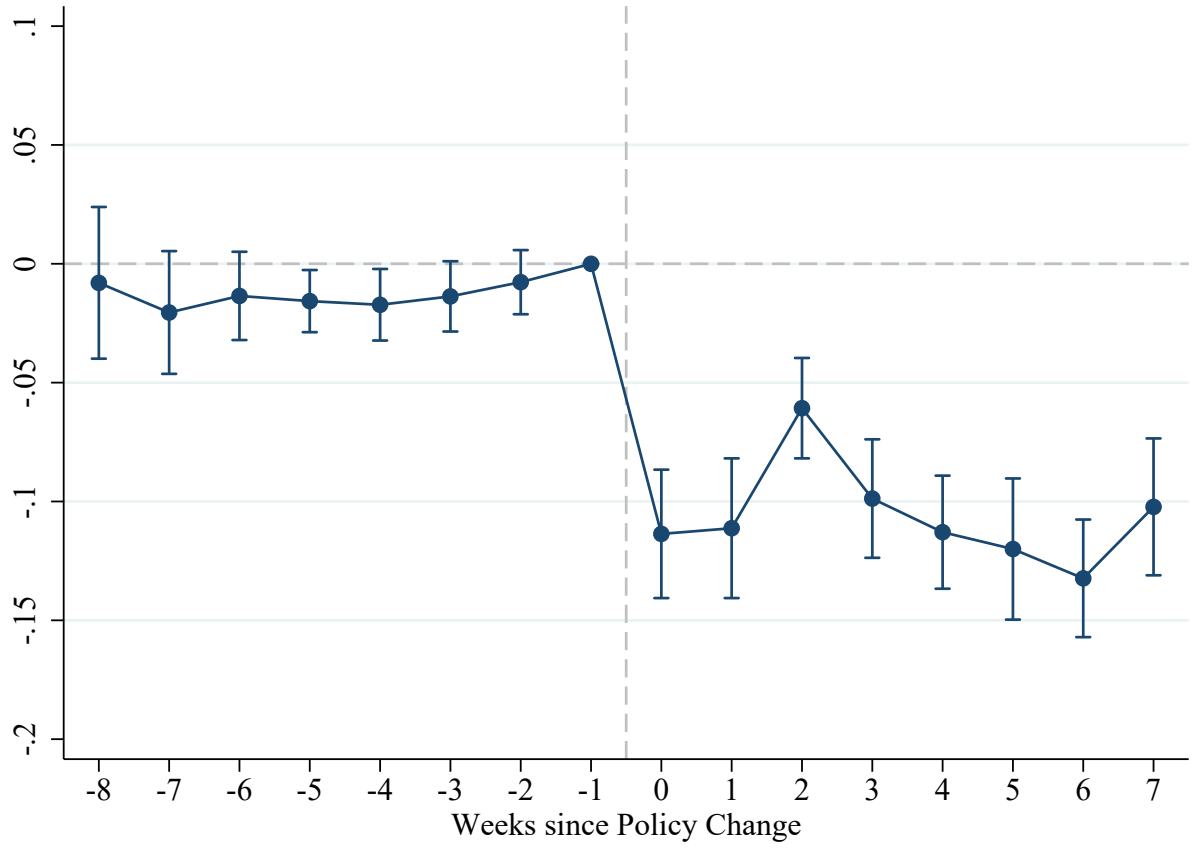


This figure presents the DID estimates and 95% confidence interval for the following specification using application acceptance status as the dependent variable.

$$y_{i,t} = \sum_{k=-8, k\neq -1}^{k=+7} \beta_k \cdot \text{Young}_i \times \mathbb{I}(t = k) + \theta_{z,p} + \theta_{z,t} + \theta_{p,t} + \theta_w + \varepsilon_{i,t}$$

All regressions include ZIPcode × loan purpose ($\theta_{z,p}$), ZIPcode × week ($\theta_{z,t}$), loan purpose × week ($\theta_{p,t}$), and the day of the week (θ_w) fixed effects. Young is a binary variable that takes a value of one for customers with age less than the sample median value, and zero otherwise. 95% confidence bands calculated based on standard errors clustered at the ZIPcode level.

Figure 10: DID: Application Acceptance & Prior Relationship

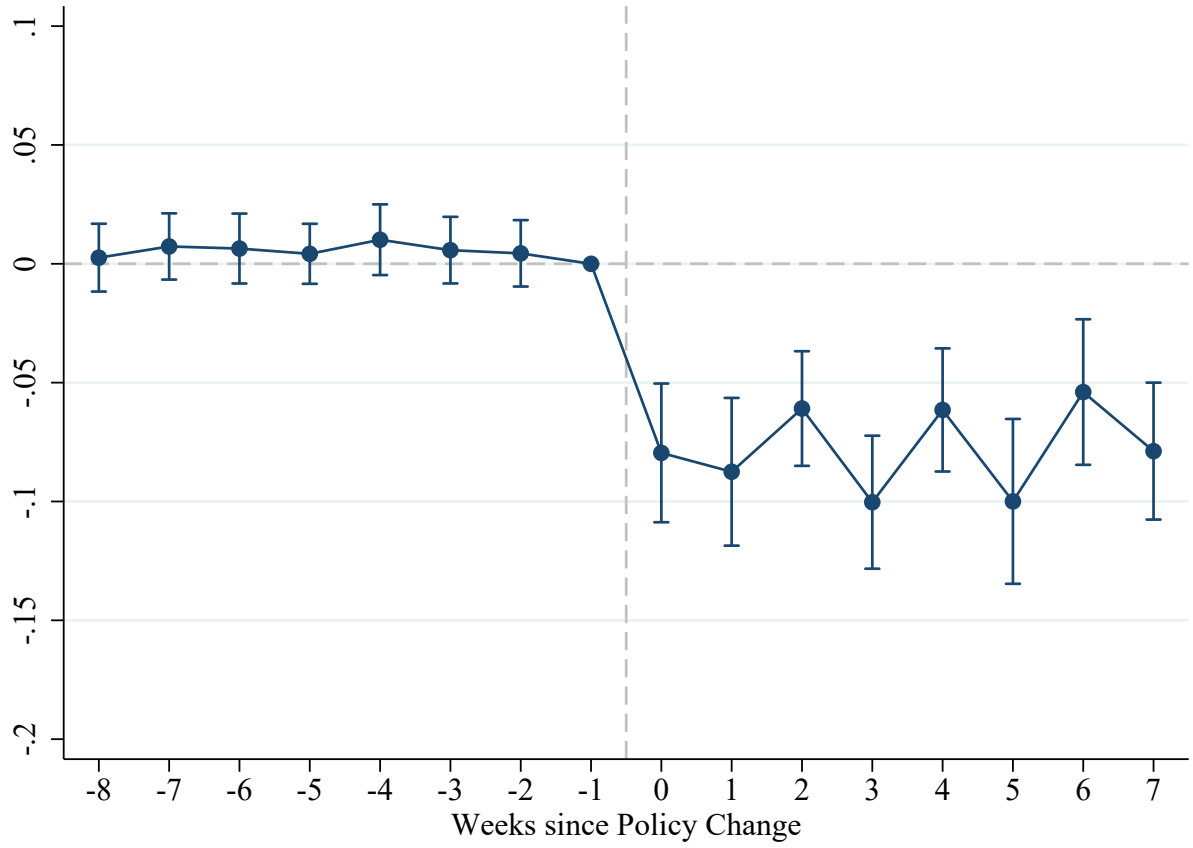


This figure presents the DID estimates and 95% confidence interval for the following specification using application acceptance status as the dependent variable.

$$y_{i,t} = \sum_{k=-8, k\neq -1}^{k=+7} \beta_k \cdot \text{New Customer}_i \times \mathbb{I}(t = k) + \theta_{z,p} + \theta_{z,t} + \theta_{p,t} + \theta_w + \varepsilon_{i,t}$$

All regressions include ZIPcode × loan purpose ($\theta_{z,p}$), ZIPcode × week ($\theta_{z,t}$), loan purpose × week ($\theta_{p,t}$), and the day of the week (θ_w) fixed effects. New customer is a binary variable for customers who are first time applicants, and zero otherwise. 95% confidence bands calculated based on standard errors clustered at the ZIPcode level.

Figure 11: DID: Application Acceptance & Caste

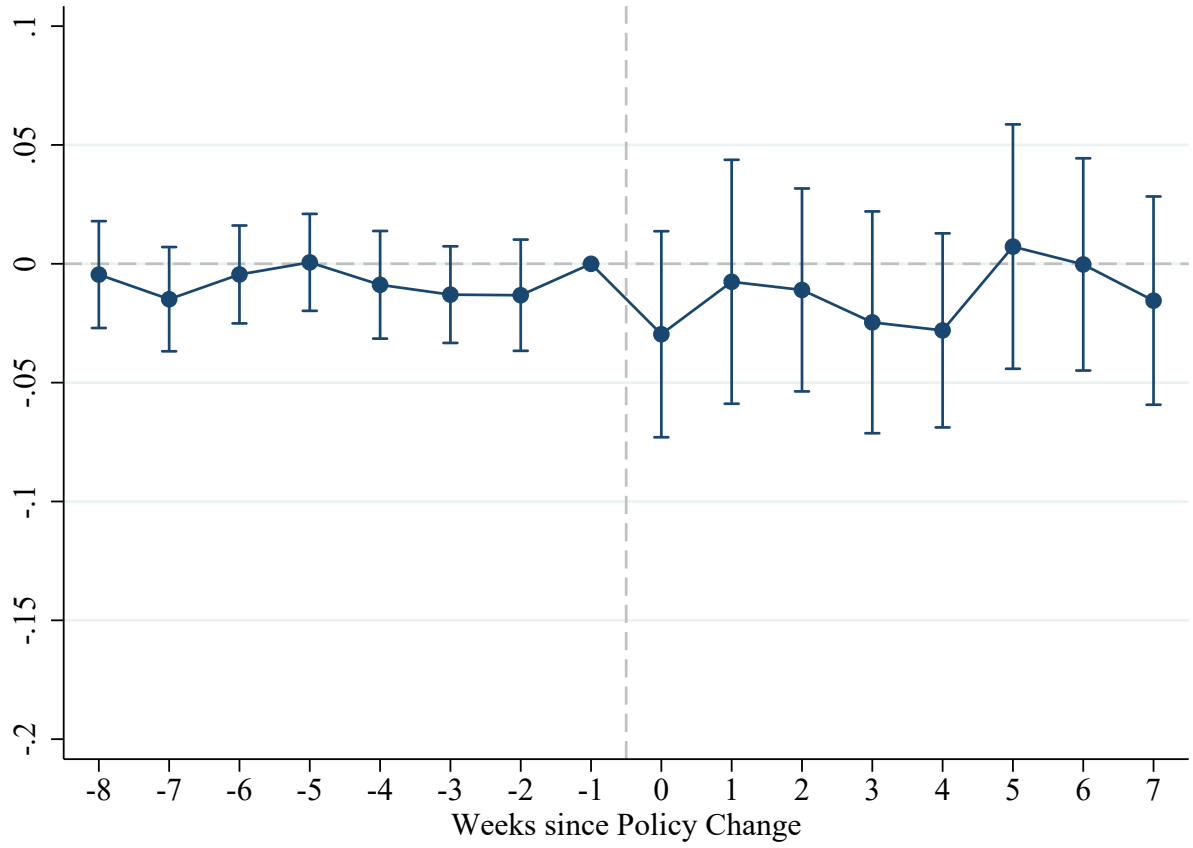


This figure presents the DID estimates and 95% confidence interval for the following specification using application acceptance status as the dependent variable.

$$y_{i,t} = \sum_{k=-8, k\neq -1}^{k=+7} \beta_k \cdot \text{Low Caste}_i \times \mathbb{I}(t = k) + \theta_{z,p} + \theta_{z,t} + \theta_{p,t} + \theta_w + \varepsilon_{i,t}$$

All regressions include ZIPcode × loan purpose ($\theta_{z,p}$), ZIPcode × week ($\theta_{z,t}$), loan purpose × week ($\theta_{p,t}$), and the day of the week (θ_w) fixed effects. Low Caste is a binary variable that takes a value of one for applicants belonging to the Scheduled caste, scheduled tribe and other backward class, and zero otherwise. 95% confidence bands calculated based on standard errors clustered at the ZIPcode level.

Figure 12: DID: Application Acceptance & Gender



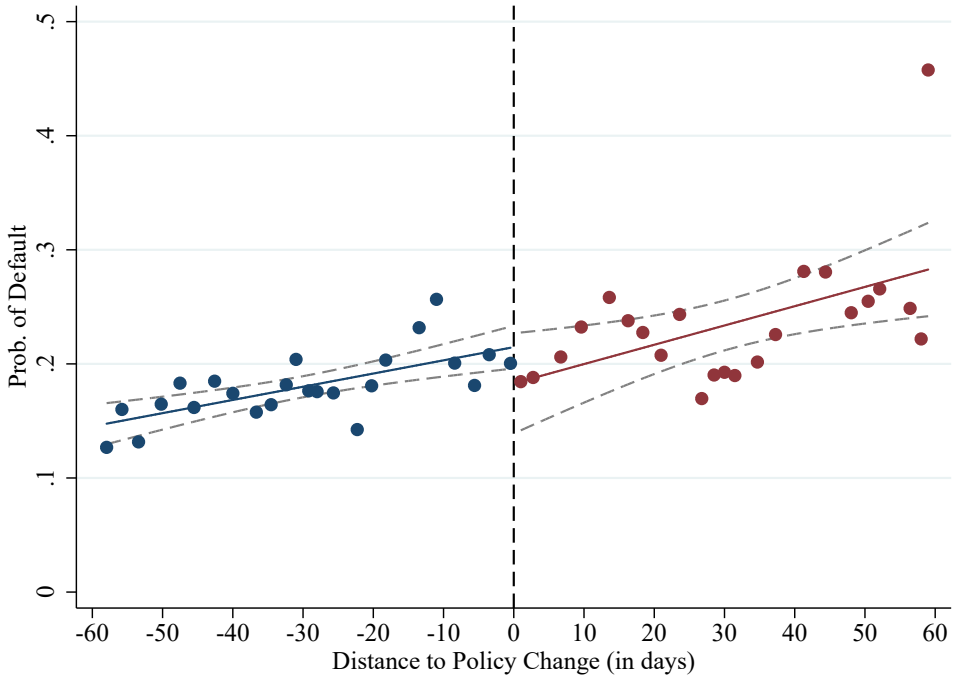
This figure presents the DID estimates and 95% confidence interval for the following specification using application acceptance status as the dependent variable.

$$y_{i,t} = \sum_{k=-8,k\neq-1}^{k=+7} \beta_k \cdot \text{Female}_i \times \mathbb{I}(t = k) + \theta_{z,p} + \theta_{z,t} + \theta_{p,t} + \theta_w + \varepsilon_{i,t}$$

All regressions include ZIPcode × loan purpose ($\theta_{z,p}$), ZIPcode × week ($\theta_{z,t}$), loan purpose × week ($\theta_{p,t}$), and the day of the week (θ_w) fixed effects. Female is a binary variable that takes a value of one for females, and zero otherwise. 95% confidence bands calculated based on standard errors clustered at the ZIPcode level.

6 Effect on Default

Figure 13: Effect on Default: RD Estimate



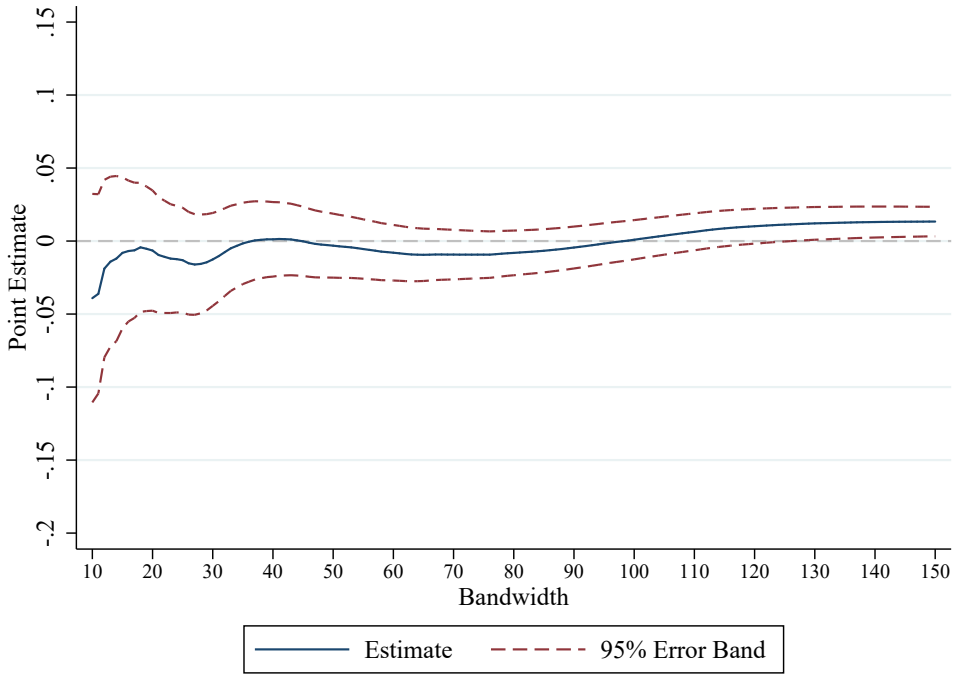
The figure presents RD plots for the probability of default at different bins along the running variable (distance to policy change) as well as with a local linear trend estimated separately, along with 95% confidence intervals, on each side of the discontinuity. The unit of observation is a day. The red dots indicate days after the call log collection ban policy and the blue dots indicate days before the call log collection ban policy was implemented.

Table 9: Effect of the Policy Change on Default

Dep Var: Loan Default	(1)	(2)	(3)	(4)	(5)
Post	-0.0391	-0.0081	-0.0065	-0.0131	-0.0126
	(0.0364)	(0.0264)	(0.0210)	(0.0183)	(0.0162)
ZIP Code X Loan Purpose FE	Yes	Yes	Yes	Yes	Yes
Day of the Week FE	Yes	Yes	Yes	Yes	Yes
Bandwidth	10	15	20	25	30
Kernel	Triangular	Triangular	Triangular	Triangular	Triangular
# Obs	4,293	8,341	11,068	14,893	20,568
R^2	0.3427	0.3097	0.2726	0.2543	0.2565

This table presents the RD estimates using the default status of a loan as the dependent variable. The dependent variable takes a value of one if the loan is in default and zero otherwise. We use different bandwidths of 10, 15, 20, 25, 30, 40, 50, and 60 days on either side of the policy change date (January 1, 2019), in columns (1), (2), (3), (4), (5), (6), (7) and (4), respectively. All regressions include ZIPcode × loan reason and the day of the week fixed effects, a local linear specification estimated separately on each side of the policy change date, and use a triangular kernel. Standard errors reported in parentheses are clustered at the ZIPcode level. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Figure 14: Default: RD Estimates for various Bandwidth choices



This figure presents RD estimates and 95% confidence interval for the RD specification using loan default status as the dependent variable. We estimate the RD specification for several bandwidths between 10 days and 150 days in increments of 1 day. The Y-axis reports the RD estimate, and the X-axis reports the bandwidth used to estimate the RD coefficient. The unit of observation is a day. All regressions include ZIPcode × loan reason and the day of the week fixed effects, a local linear specification estimated separately on each side of the policy change date, and use a triangular kernel. The solid red line reports the RD estimate, and the dashed grey lines indicate the 95% confidence bands calculated based on standard errors clustered at the block level.

Table 10: Robustness to Covariates: Effect of the Policy Change on Default

Dep Var: Loan Default	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Post	-0.0132	-0.0205	-0.0197	-0.0126	-0.0124	-0.0122	-0.0113	-0.013	-0.0011	-0.0033	0.0037
	(0.0116)	(0.0132)	(0.0132)	(0.0162)	(0.0162)	(0.0162)	(0.0162)	(0.0158)	(0.0156)	(0.0147)	(0.0148)
Female (=1)						-0.0129	-0.0135	-0.0098	-0.0115	-0.0182	-0.0225*
						(0.0131)	(0.0133)	(0.0135)	(0.0135)	(0.0131)	(0.0130)
ZIP Code FE		Yes	Yes								
Loan Purpose FE			Yes								
ZIP Code X Loan Purpose FE				Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day of the Week FE				Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Education FE					Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age Decile FE							Yes	Yes	Yes	Yes	Yes
Credit Score Decile FE								Yes	Yes	Yes	Yes
Social Score Decile FE									Yes	Yes	Yes
Loan Maturity FE										Yes	Yes
Income Decile FE											Yes
Bandwidth	30	30	30	30	30	30	30	30	30	30	30
Kernel	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular	Triangular
# Obs	23,184	22,226	22,226	20,568	20,568	20,568	20,531	20,004	20,004	20,004	20,004
R^2	0.0007	0.1256	0.1291	0.2565	0.2566	0.2567	0.2575	0.2713	0.2761	0.3317	0.3449

This table presents the RD estimates using loan default status as the dependent variable. The dependent variable takes a value of one if the loan is in default and zero otherwise. We use a bandwidth of 30 days on either side of the policy change date (January 1, 2019). All regressions include ZIPcode × loan purpose and the day of the week fixed effects, a local linear specification estimated separately on each side of the policy change date, and use a triangular kernel. 95% confidence bands calculated based on standard errors clustered at the ZIPcode level. Standard errors reported in parentheses are clustered at the ZIPcode level.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

References

Berg, Tobias, Valentin Burg, Ana Gombović, and Manju Puri. 2020. "On the rise of fintechs: Credit scoring using digital footprints." *The Review of Financial Studies*, 33(7): 2845–2897.

Cai, Jing, Muhammad Meki, Simon Quinn, Erica Field, Cynthia Kinnan, Jonathan Morduch, Jonathan de Quidt, and Farah Said. 2021. "Microfinance." *VoxDevLit*, 3(1): 26.

Czura, Kristina, Anett John, and Lisa Spantig. 2020. "Flexible Microcredit: Effects on Loan Repayment and Social Pressure."

Diep-Nguyen, Ha, and Huong Dang. 2020. "Social collateral." mimeo.

Jack, William, Michael Kremer, Joost De Laat, and Tavneet Suri. 2023. "Credit Access, Selection, and Incentives in a Market for Asset-Collateralized Loans: Evidence From Kenya." *Review of Economic Studies*, rdad026.

Karlan, Dean, Markus Mobius, Tanya Rosenblat, and Adam Szeidl. 2009. "Trust and social collateral." *The Quarterly Journal of Economics*, 124(3): 1307–1361.

Lin, Mingfeng, Nagpurnanand R Prabhala, and Siva Viswanathan. 2013. "Judging borrowers by the company they keep: Friendship networks and information asymmetry in online peer-to-peer lending." *Management science*, 59(1): 17–35.

Thakor, Anjan V. 2020. "Fintech and banking: What do we know?" *Journal of Financial Intermediation*, 41: 100833.