The Role of Credit Reports in Digital Lending: a Case Study from Mexico[§]

Alfredo Burlando[¶] Michael A. Kuhn[‖] Silvia Prina[**]

May 18, 2023

Abstract

Digital credit that uses non-traditional scoring techniques has already expanded credit access to many new populations. A point of policy debate is whether digital lenders should be fully integrated into information sharing systems, such as those offered by credit reference bureaus (CRBs). We study an example of a digital lender in Mexico adopting credit bureau scores into their screening process. Using unique administrative data, we estimate a regression discontinuity in time around the lender's integration of credit bureau scores. We find that the acquisition of credit scores reduces defaults, with the likelihood of borrowers' repayment increasing by 10-13%.

[§]The authors thank the Digital Credit Observatory for generous research support. We thank our digital lender partner for providing us with the data and answering our many questions. We benefited from several conversations with policymakers and digital lenders in Mexico and sub-Saharan Africa who wish to remain anonymous, and from seminar participants at the University of Oregon. Any errors are our own. The dataset analyzed in the current study is not publicly available because it is proprietary to our lender partner, but is available from the corresponding author on reasonable request. All code is available on request.

[¶]University of Oregon and CEGA. Department of Economics, 1285 University of Oregon, Eugene, OR 97403. E-mail: burlando@uoregon.edu

[‖]University of Oregon and CEGA. Department of Economics, 1285 University of Oregon, Eugene, OR 97403. E-mail: mkuhn@uoregon.edu

[**]Northeastern University, Department of Economics, 310A Lake Hall, 360 Huntington Avenue, Boston, MA 02115, United States. E-mail: s.prina@northeastern.edu

1 Introduction

Credit Reference Bureaus (CRBs) play an important role in financial markets. By collecting information about individual borrowers and making it available to credit providers, they lower lenders' costs of screening and acquiring new clients. For borrowers, they can be either signals of creditworthiness that lead to expanded access and improved terms, or signals of risk that lead to the opposite.

In low- and middle-income countries (LMICs), credit markets are evolving very rapidly due to rapid advancements in information technology, and the proliferation of financial technology companies (hereafter Fintechs). In particular, new forms of remote loan distribution (broadly defined as "digital credit") are challenging the primacy of formal credit markets, and rapidly expanding the reach of formal lenders into populations that previously lacked access. Digital lenders are often not integrated into traditional borrower information sharing schemes, and there is substantial variation across countries in whether and how this is done.

In this paper, we start by discussing some of the complications that arise in the integration of digital credit into existing CRBs, drawing examples from three markets at the forefront of innovations in the digital credit realm: Kenya, Uganda and Mexico. First, Fintechs are often regulated differently than traditional banks, including in their CRB reporting requirements. Second, the regulatory environment is constantly shifting; regulators have been introducing new rules about information sharing while contending with the rapidly changing landscape. Third, due to the unique features of digital credit, including a large fraction of customers that are not currently included in the formal financial system, the value of CRB information to Fintechs may differ from that of traditional lenders.

In the second part of the paper, we focus on the Mexican market, where regulators do require digital lenders to share information through CRBs. We provide evidence suggesting that digital lenders may benefit from gathering information from CRBs, although this may come at the cost of the financial inclusion of low-income households. Specifically, we analyze unique administrative data on loans from a digital lender containing outcomes for over 11,000 loans issued over a year-long period. In the middle of the sample period, the digital lender began incorporating credit reports from a credit bureau. We use a regression discontinuity in time strategy to compare repayments of loans issued immediately before the introduction of credit reports to those issued immediately after. Our findings show that the acquisition of credit scores reduces defaults: the likelihood of borrowers' repayment increases by 10-13% immediately after the acquisition of credit scores, a result that is statistically significant and economically meaningful. We also consider how the characteristics of approved borrowers changes at the same time. We suspect CRB scores at least partially supplant observables in the screening process, and this is unlikely to be a balanced effect across characteristics. Most notably, we find that approved borrower income and requested loan amount increase when CRB scores are adopted.

Our empirical results on the importance of credit scoring are consistent with models in which information sharing increases repayments by reducing asymmetric information (Stiglitz and Weiss, 1981; Pagano and Jappelli, 1993), improves borrower incentives to repay (Padilla and Pagano, 1997; Padilla and Pagano, 2000; Flatnes, 2021), and increases loan sizes in equilibrium (McIntosh and Wydick, 2009). They are also in line with an existing empirical literature from LMICs (e.g., Behr and Sonnekalb, 2012, De Janvry et al., 2010) which finds that the introduction of credit reporting increases repayment. We extend this literature to the context of digital credit. We also show that the increase in loan sizes following the introduction of the CRB (also found by De Janvry et al., 2010) is attributable to the lender choosing borrowers who asked for larger loans.

Our study is also connected to the literature on the impact of digital credit on households' ability to cope with unexpected shocks and reduce liquidity constraints for investments (e.g., Karlan and Zinman, 2010; Morse, 2011). Bharadwaj, Jack, and Suri (2019) find that digital credit in Kenya has improved household resilience to negative shocks, Björkegren et al. (2022) find that it increases subjective well-being, and Brailovskaya, Dupas, and Robinson (2021) find that increases perceived financial well-being. Existing research on digital credit has mostly focused on the welfare effects of expanding credit markets for populations excluded from the formal financial system (Francis et al. (2017); Björkegren and Grissen (2018)). Burlando et al. (2022) focuses on repayment, and shows that reducing the speed of delivery of digital loans increases the likelihood that loans are repaid.

The paper proceeds as follows. Section 2 reviews the existing literature on information sharing and discusses the current policy landscape. Section 3 illustrates our Mexico case study. Section 4 discusses policy implications and concludes.

2 Background

2.1 Digital credit and financial inclusion

Sources of consumer credit in LMICs are limited, and digital credit plays an important role in expanding access to finance. For example, according to the information collected by Findex 2023, the fraction of borrowers of formal loans from formal financial service providers including mobile services in East Africa is much higher than the fraction of borrowers from providers excluding mobile services. In Kenya, only 15% of adults in the bottom 40% of the income distribution borrowed from commercial banks; this increases to a third once mobile sources are included. In Uganda, the proportions are 15% and 24%, respectively. In Mexico, while rates of financial inclusion are higher, only 31% has access to some type of credit, which leaves a significant percentage of the population outside the financial system (Finnovista, 2021).

Despite the rapid expansion of digital credit, it cannot be considered a perfect substitute for commercial credit: digital credit is expensive, and loan amounts are typically small. For example, Mswhari loans in Kenya start at 1,000 Ksh ($7.35) and top out at 50,000 KSh ($370)[1] while in Uganda, MoKash limits loans to UGX 1 million ($270).[2] In Mexico, Vivus charges 50% interest and fees for a month-long loan to a first-time borrower, with a limit of MXN 8,000 ($450 USD).[3] Thus, information sharing through a CRB could benefit marginalized borrowers who are initially reliant on digital credit. By allowing consumers to build a credit history with digital lenders, good borrowers may be able to gain access to the larger, cheaper loans offered by traditional lenders, which they would not otherwise have been able to obtain. This is, of course, provided that digital lenders report successfully repaid loans to the CRB, in addition to defaults, an issue we discuss later.

[1]SafariCom Kenya website, accessed on May 11, 2023.

[2]MTN Uganda website, accessed on May 11, 2023.

[3]Vivus Mexico website, accessed on May 11, 2023.

2.2 Role of CRBs

The classic model of information sharing in credit markets is based on the adverse selection model by Pagano and Jappelli (1993), in which heterogeneous agents borrow in a market with imperfect competition, and where information sharing allows lenders to price discriminate. Padilla and Pagano (1997) extends the model to two periods and incorporates moral hazard. In the first period, the lender gains information about the outcome of the borrower's first loan. In the second period, borrowers who repay their loans are rewarded by receiving another loan. If information sharing among lenders via a credit bureau is missing, borrowers pay a high interest rate and the lender earns information rents. With information sharing, second-period rents are eliminated. Nevertheless, information sharing raises lender's profits in the first period: borrowers strategically increase their effort levels in the first period, in order to obtain lower interest rates in the second period. Padilla and Pagano (1997) show that a CRB is required to sustain this outcome by preventing profitable deviations, such as lenders choosing not to share information on high-quality borrowers to keep them in their portfolio.[4]

There are theoretical implications of CRBs for financial inclusion as well. In particular, theory predicts that reporting reduces interest rates and therefore increases the number of high quality borrowers willing to borrow at the lower rate; thus, credit increases (even if it is never rationed). In Bennardo et al. (2015), entrepreneurs borrow from multiple sources, leading to large loans and suboptimal risk-taking when information sharing is lacking and lenders cannot internalize the lending choices of competitors. Under certain conditions, this leads to credit rationing. In this scenario, the presence of CRBs alleviates the distortion: lenders learn the true levels of borrowers' debt and offer loan contracts conditional on that debt. In equilibrium, interest rates and credit rationing decline. A similar mechanism in McIntosh and Wydick (2009) shows that CRBs lead to higher loans sizes conditional on lending, as

[4] A similar result can be obtained under perfect competition, provided that information sharing is limited to negative repayment outcomes (Padilla and Pagano, 2000)). On the other hand, Flatnes (2021) finds that information sharing could increase the interest rate on loans because lenders reduce the amount required as collateral. This provides an alternative pathway through which the expansion of information due to digital credit increases access to traditional finance to borrowers with little collateral.

lenders are able to optimize loan sizes conditional on current levels of indebtedness.

The application of these models to digital credit is not straightforward. In these models, lenders either know the borrowers' type (as in Pagano and Jappelli (1993)) or learn it during the lending process. Fintechs, instead, use costly alternative sources of information to screen new un-banked borrowers. Information sharing between Fintechs and CRBs would impact investments in information acquisition, influencing the financial inclusion of the un-banked in ways that depends on the availability and costs of different sources of repayment-relevant information. Furthermore, collecting borrower information *across* Fintechs would likely decrease heterogeneity in credit scoring across firms, leading to a richer market for borrowers with a good CRB signal, but reducing the likelihood that any given borrower has access to credit from *at least one* source.

Empirically, several studies find that information sharing has positive effects on repayment and mixed results in terms of credit availability. While much of this literature is based on the U.S. (Doblas-Madrid and Minetti, 2013; Einav et al., 2013), in LMICs Behr and Sonnekalb (2012) shows that the introduction of mandatory reporting to a credit bureau in Albania reduces the probability that loans are overdue. De Janvry et al. (2010) study the case of a microfinance institution in Guatemala that introduced CRB reports to its branches in a staggered fashion. They find a decrease in loan delinquency and a change in the demographic characteristics of new borrowers, attributable to a reduction in adverse selection. They also find that, following the introduction of the credit bureau, the total number of loans and their value increased. The authors then carried out a randomized intervention informing clients of the existence of *reporting to* a CRB. They found that repayments improved further, consistent with a decrease in moral hazard among borrowers.[5]

2.3 Integration of digital credit into CRBs

The regulatory environment for digital lenders is still evolving in many countries. We provide here an overview of some of these challenges, following conversations with some representa-

[5]A related macro literature, which relies in across country variation in CRB density, shows findings consistent with the empirical micro literature (Galindo and Miller (2001), Jappelli and Pagano (2002), Love and Mylenko (2003), Djankov et al. (2007)).

tives of the digital credit industry and policymakers in Mexico and East Africa. We center our discussion around the experiences of Kenya, Uganda and Mexico, as these countries have a thriving digital credit market and are at the forefront of both innovation and regulation.

Regulatory environment In many countries, digital finance remains unregulated and there is no credit information sharing for digital lenders. In Kenya, the financial technology sector was unregulated until the 2021 Digital Credit Providers Regulations, which established the Bank of Kenya (the central bank) as regulator. In Uganda, the Tier 4 Microfinance Institutions Act and Money Lenders Act of 2016 has been regulating digital lenders and some other institutions. Under the act, the regulator is the Uganda Microfinance Regulatory Authority (UMRA). While the Act requires UMRA to "establish a mechanism of reporting by Tier 4 microfinance institutions to the Credit Reference Bureau," to date digital lenders are not required to report to a CRB.[6] In contrast, in Mexico, the Law to Regulate Financial Technology Institutions (known as "Ley Fintech"), introduced in 2018 and amended in 2021, sets the Central Bank as regulator. While the law allows innovative Fintechs to operate outside of the regulations for at most two years, the regulations dictate that digital lenders report credit outcomes to a CRB.

Borrower tracking An important hurdle for the widespread adoption of CRBs for digital credit is the absence of national ID numbers that can be used to track borrowers across institutions. While both Mexico and Kenya have a robust national ID system that is used by local CRBs, in Uganda it is entirely absent. In its place, the financial system requires borrowers to obtain a "financial card" which serves as an alternative ID and is used by local CRBs to track borrowers. To get a financial card, a borrower must visit a commercial bank,

[6] An important caveat exists for digital credit offered through mobile money operators in many sub-Saharan Africa countries. Often (but not always) the digital loan product offered originate from a commercial bank. For instance, in Uganda, MTN's Mokash offers NCBA loans and Airtel's Quick Loans offers Housing Finance Corporation loans; in Kenya, Safaricom's M-shwari service offer loans from the Kenya Commercial Bank. In these instances, lenders must follow pre-existing banking regulations and not Fintech regulations. Thus, in Uganda, mobile money lenders follow the Financial Institutions Act, the Financial Institutions (Credit Reference Bureau) Act of 2022, and other related laws; and the regulator is the central bank, and not UMRA.

pay a fee, and get a scan of all finger digits and a photo taken. The system covers a limited number of people, as only 5.3% of Ugandans in 2013 were covered by CRB (i.e., had financial cards) (BOU, 2017). It is also not easily transferable to digital lenders, raising the issue of how to effectively track information across lending institutions.

Hurdles to CRB adoption by digital lenders Another challenge is that the appetite for digital lenders for accessing CRBs might be limited, due to a number of reasons. First, acquisition costs of CRB information are high, relative to the (typically small) loan amounts disbursed. Second, a large share of digital lenders' customers is unbanked, and therefore lack long credit histories. Third, digital lenders rely on alternative credit scoring mechanisms built using alternative sources of non-credit information. For example, mobile lenders typically generate credit scores using mobile money and airtime use. Many Fintechs specialize in lending to farmers, and leverage real-time data generated by the agricultural value chain. Other still lend to mobile money agents and have business models that are fine tuned to the clients' needs in terms of managing float.[7] Hence, CRB credit scores may be strategic substitutes to internal scores. In fact, digital lenders may openly resist a system that requires them to share hard-won information on borrower quality with traditional lenders with little value in exchange.

3 The value of credit bureau reports: a Mexican case study

3.1 Setting and data

Next, we test empirically whether acquiring information from a credit bureau about potential customers is advantageous for a digital lender already operating. We do so in the context of Mexico. We have administrative data on all approved loans from June 2018 to May 2019 from an online digital lender. The loan amounts range from 1,500 to 3,000 Mexican pesos

[7] For example, In Uganda Emata provides a dashboard that dairy cooperatives use to manage the flow of information from farmers, and uses that information to create credit scores. In Kenya and Uganda, Flow lends exclusively to mobile money agents.

(approximately USD 75 to 150),[8] and the loan terms vary from seven to 30 days. The cost of a loan (and thus repayment amount) is fixed at the time it is taken; early payment is allowed at any time, but does not save the borrower any interest. Costs consist of interest, taxes, and fees, with the implied APRs reaching up to 478.8%. There is a one-time penalty added to the amount owed for delinquent loans. The characteristics of this loan product are comparable to those of other digital lenders in the market at the time of the study. Potential borrowers interact with the lender using a browser on a smartphone or computer. The lender's home page prominently reports the interest rate and other costs, including taxes and fees, at the bottom of the window. Potential borrowers are advised that they can get a loan in "minutes."

Users select their preferred amount and term of the loan. Applicants need to satisfy the following requirements to obtain a loan: proof of citizenship (a photo of the national identification card); age between 20-65 years; a photo taken from a phone or computer camera; regular income; cellphone number and e-mail address; and a bank account. Starting on November 1st, 2018, five months into our data, the digital lender began pulling the applicant's credit history from a credit bureau for all first-time applicants.[9]

3.1.1 Sample

There are 11,961 approved loans in our sample.[10] For each loan, we observe the timestamps of application submission and loan disbursement; the repayment status and date of final repayment for each loan; the borrower's age, sex, marital status, number of dependents, and reported personal income. Furthermore, we have information on both the requested and approved loan amounts and terms.

Table 1 provides summary statistics for the entire sample and for the sub-sample of borrowers within 60 days of the credit scores acquisition. Borrowers are low-income: the median monthly income reported is 917 pesos (roughly $46). After winsorizing extreme positive outliers, the mean is 1,816 pesos, which is just about one-third of the national

[8]The exchange rate during the study period was approximately USD 1 = MXP 20.

[9]The administrative data also contain information on repeat borrowers. As the lender only uses their prior repayment behavior, rejecting applicants who failed to fully pay back previous loans, we do not include them in our analysis.

[10]See Burlando et al. (2022) for additional details on sample construction.

Table 1: **Summary statistics**

Sample:	All loans	Within 60 days of credit score adoption		
		Before	After	Difference (3) - (2)
	(1)	(2)	(3)	(4)
Monthly income	1,815.717	1,719.996	2,112.899	392.903
(pesos, winsorized)	[9007.519]	[8067.431]	[10,597.97]	(277.757)
# of dependents	1.252	1.327	1.215	-0.112***
	[1.143]	[1.135]	[1.141]	(0.034)
Married	0.511	0.569	0.478	-0.091***
	[0.500]	[0.495]	[0.500]	(0.015)
Female	0.434	0.440	0.449	0.008
	[0.496]	[0.497]	[0.497]	(0.015)
Age	37.443	38.442	37.820	-0.622**
	[9.625]	[9.771]	[9.611]	(0.291)
Term of loan	22.033	21.742	21.467	-0.275
(days)	[7.410]	[8.315]	[7.495]	(0.239)
Requested term of loan	22.638	21.620	22.131	0.511**
(days)	[7.509]	[8.364]	[7.567]	(0.241)
Loan amount	1,793.515	1,727.716	1,731.129	3.953
(pesos)	[363.738]	[338.034]	[333.994]	(10.093)
Requested loan amount	2,263.490	2,348.588	2,274.212	-74.377***
(pesos)	[638.270]	[592.397]	[645.802]	(18.495)
Repaid	0.636	0.588	0.684	0.096***
	[0.481]	[0.492]	[0.465]	(0.014)
Observations	11,961	2,550	1,996	4,516

Notes: $^{***} \Rightarrow p < 0.01, ^{**} \Rightarrow p < 0.05, ^{*} \Rightarrow p < 0.10$, given by t-tests of means. Standard deviations of the statistics are shown in brackets, and standard errors of the differences are shown in parentheses. Monthly income is winsorized at the 99.9th percentile of the distribution.

average in 2018.[11] The average loan is equivalent to a month of income, is due three weeks after disbursement, and is repaid 64% of the time.[12]

The summary statistics from Table 1 suggest that credit scores at least partially supplant borrower characteristics as a source of repayment-relevant information. In fact, approved borrowers after the acquisition of credit scores look different than approved borrowers prior:

[11] Source: CEIC

[12] We cannot observe partial repayments, only when a loan has been entirely repaid.

on average they have fewer dependents, are less likely to be married, and are younger. They also request longer, smaller loans, though the size and term of the disbursed loans do not change. Moreover, repayment is 9.6pp higher in the 60 days after the adoption of credit bureau scores than in the 60 days before. In Section 3.3.2 we use our regression discontinuity model to formally estimate changes in borrower characteristics at the credit score adoption, and find that some of the naïve differences shown in Table 1 are misleading; at the discontinuity, we identify a shift towards higher-income, single, (perhaps) male borrowers requesting (and receiving) *larger* loans.

3.2 Empirical strategy

To study the impact of credit score on repayment, we take advantage of the fact that the digital lender began using credit bureau some time after starting its business. We compare repayment rates of borrowers just before and after the acquisition of credit scores using a regression-discontinuity approach. In particular, for each first-time loan j, we run the following regression:

$$\begin{aligned} LoanPaid_j = \alpha + \beta_1 PostCS_j + \beta_2 DaysAfter_j + \\ \beta_3 PostCS_j \times DaysAfter_j + \delta X_j + \epsilon_j \end{aligned} \tag{1}$$

where $PostCS$ is an indicator equal to one for any loan whose application arrived after the credit score adoption and $DaysAfter$ is the number of days relative to credit score adoption on $DaysAfter = 0$. Our main outcome variable, $LoanPaid$, is an indicator of whether the loan was repaid. The coefficient β_1 identifies the effect of the introduction of credit scores over the time window considered.

We consider three different specifications: without controls; controlling for the term and amount of the loan; and controlling for borrower characteristics (gender, log income, age and age squared, marital status, number of dependents, and the *requested* loan term and amount) and the timing of the loan application (application submission day-of-week, hour-of-day and day-of-month fixed effects), in addition to loan term and amount. Standard errors are clustered by day. For each of these specifications, we take both a fixed 60-day bandwidth and an asymmetric optimal bandwidth approach. In general, the optimal bandwidths are smaller than 60-days, ranging from 28-44 days. We maintain a linear model of the running variable as our main specification of equation 1 throughout the analysis. As mentioned in Section 3.3.1, the discontinuity in repayment is robust to higher-order specifications, but that these specifications over-fit the data.

3.3 Results

Figure 1 shows the discontinuity in the repayment rate when credit bureau scores are introduced into the lender's process using a 60-day bandwidth. The fit lines are estimated without control variables and a triangular kernel over the fixed 60-day window before and after the cutoff. We find a 7.8 p.p. increase in repayment. This corresponds to a 13.3% increase in repayment from a base of 58.8%. The regression estimate in Table 2, column 1, shows that the discontinuity is statistically significant with $p = 0.003$. When adding loan amount and term as controls (column 2), and borrower and application controls (column 3), results do not change. Columns (4)-(6) repeat these three specifications with an asymmetric optimal bandwidth approach. Again, results are unchanged. Overall, our estimates of the impact of the introduction of credit bureau scores range from a 7.8 p.p. (13.3%) on the high end to 5.8 p.p. (9.8%) on the low end. All estimates are statistically significant at the 5% level, although the p-values that account for bias in bandwidth selection (which correspond to discontinuities estimated over a larger-than-optimal bandwidth) range from 0.062 to 0.107, suggesting some sensitivity in precision with respect to bandwidth choice.

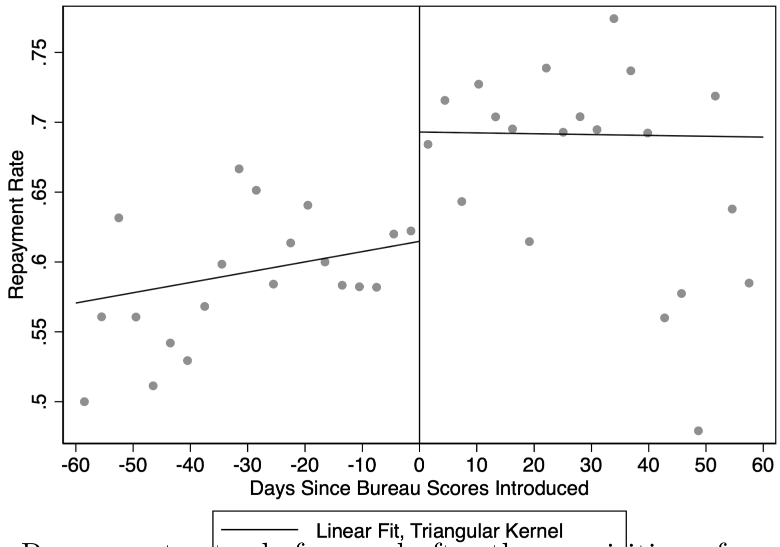

Figure 1: Repayment rates before and after the acquisition of credit scores

Table 2: **Impact of credit score acquisition on loan repayment**

Bandwidth type:	60-day fixed			Optimal		
	(1)	(2)	(3)	(4)	(5)	(6)
PostCS	0.078***	0.067***	0.064***	0.073**	0.059**	0.058**
	(0.026)	(0.025)	(0.024)	(0.030)	(0.028)	(0.028)
% of pre-CS mean	13.3%	11.4%	10.9%	12.4%	10.1%	9.8%
Est. p-value	0.003	0.008	0.009	0.015	0.034	0.041
BW bias-corr. p-value				0.062	0.101	0.107
Bandwidth (days)	[60,60]	[60,60]	[60,60]	[28,40]	[30,44]	[32,40]
Obs. in bandwidth	4,502	4,502	4,502	2,757	3,014	3,036
Loan controls	N	Y	Y	Y	Y	Y
Borrower & app. controls	N	N	Y	Y	Y	Y

Notes: $^{***} \Rightarrow p < 0.01,^{**} \Rightarrow p < 0.05,^{*} \Rightarrow p < 0.10$, according to the estimate p-values. Columns (1)-(3) report specifications with a fixed bandwidth 60 days. Columns (4)-(6) report specifications using optimal bandwidth selection allowing for asymmetric bandwidths around the date the lender stared using credit bureau scores. Standard errors of the estimates are shown in parentheses below the estimates and are clustered by day. All estimates use a triangular kernel. Below each estimate, we report: the estimate as a percentage of the 60-day pre-treatment mean, the p-value of the estimate, the bandwidth (rounded to the nearest integer in the case of the optimal bandwidths), and the number of observations within the bandwidth. For the optimal bandwidth models we also report the bandwidth selection bias-corrected p-values. Columns (1) and (4) have no control variables, columns (2) and (5) add controls for the actual amount and term of the loan, and columns (3) and (6) also feature borrower controls (age, age squared, sex, marital status, number of dependents and log income) and fixed effects for application day-of-month, day-of-week, and hour-of-day.

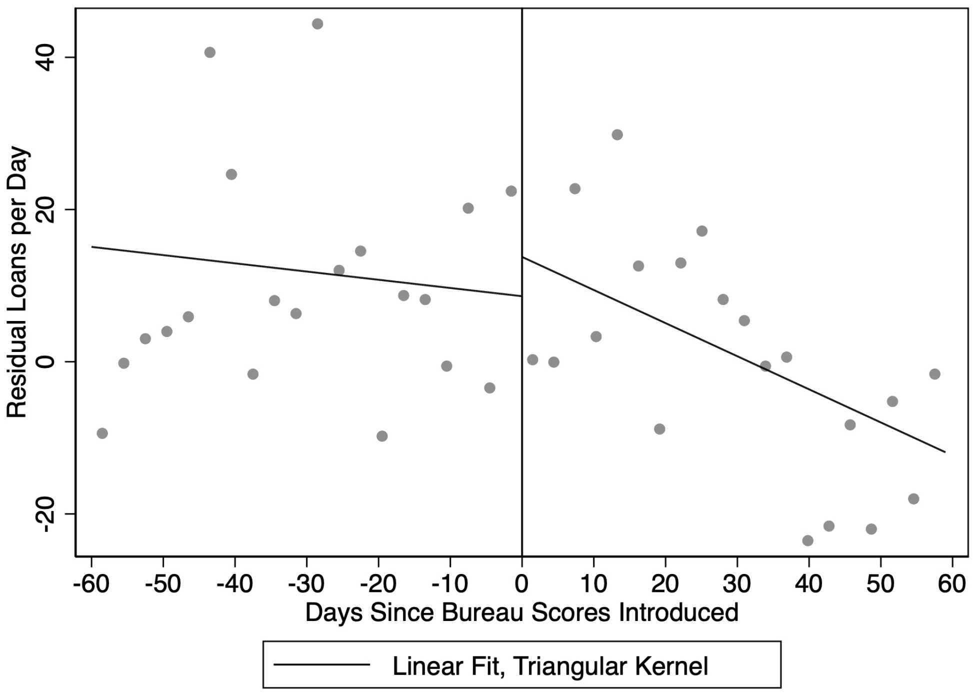

Figure 2: Loan disbursement volume by day

Next, we report on a smoothness test. Figure 2 shows that approved loan volume (after controlling for calendar time, application day-of-month, and application day-of-week) is smooth through the introduction of credit bureau scores. The p-value associated with the 3.9 loan-per-day increase is 0.586.[13,14]

Finally, we note that our results can be the result of the lender taking substantially more risk in the run up to the policy change. Perhaps, anticipating tighter future credit decision making (internally), this was a last opportunity to learn about repayment predictors among risky borrowers. In that case, our results would lead us to falsely conclude that the increase in repayment rate was a result of score adoption (and/or any other coincident changes in business practices) and not due to increased risk-taking. Figure 3 provides some evidence

[13] Estimate comes from an optimal bandwidth regression discontinuity without control variables (beyond those already residualized). The optimal bandwidth for this model is [39,54], and the bandwidth-bias-robust p-value of the estimate is 0.701.

[14] A true test of validity would require checking for discontinuities in the the volume of applications received immediately after score adoption. While we cannot observe application volume due to constraints in our data, we have no reason to expect a discontinuous shift in applications given that product pricing remained constant.

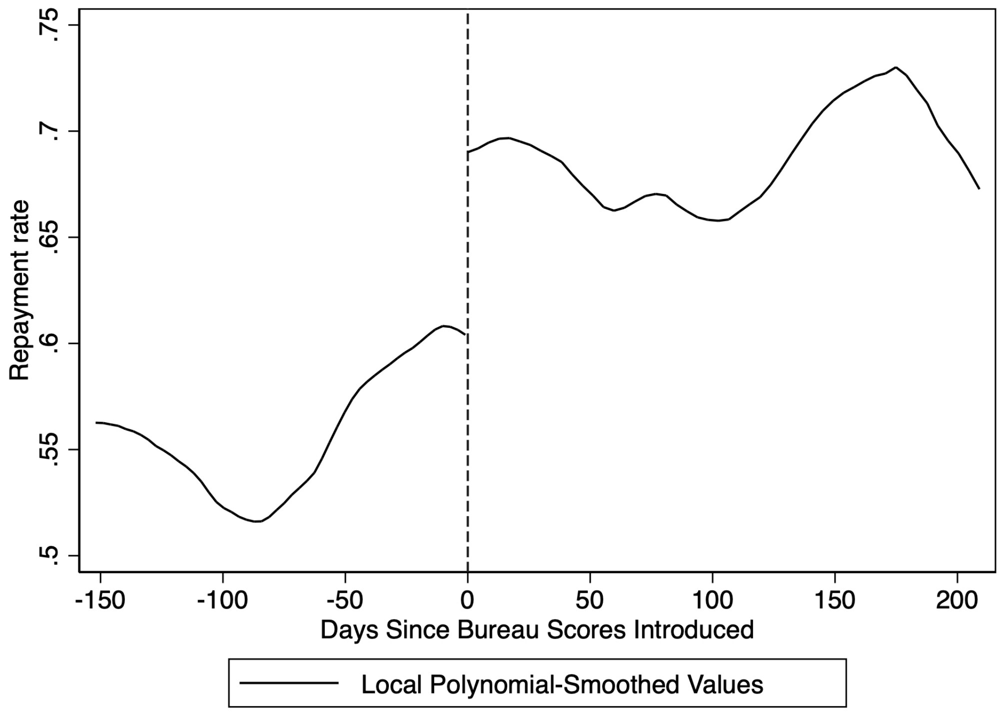

Figure 3: Time path of repayment rate, estimated separately before and after cutoff

that this is unlikely to be a concern. We plot the time path of the repayment rate, estimated using local polynomial regressions.[15] The repayment rate is on an upward trend beginning about 90 days prior to the adoption of the credit scores. While that trend attenuates right before score adoption, this is not to a degree that appear to explain the discontinuity.

3.3.1 Robustness

Figure 4 shows that the point estimate from column (1) of Table 2 is stable with respect to the fixed bandwidth used for estimation. The point estimates range from 6.6 p.p. to 8.6 p.p. as the bandwidth grows from 20 to 80 days, and the 95% confidence interval excludes zero for all bandwidths above 23 days. The optimal-bandwidth estimate of 7.3 p.p. in column (4) of Table 2 is also within this range.

Figure 5 shows the robustness of the estimate to higher-order polynomial specifications of the running variable. In order to match Figure 1, we use the 60-day bandwidth, but

[15]We select a 14-day bandwidth to ensure each day-of-week effect is accounted for on either side of any day in the sample.

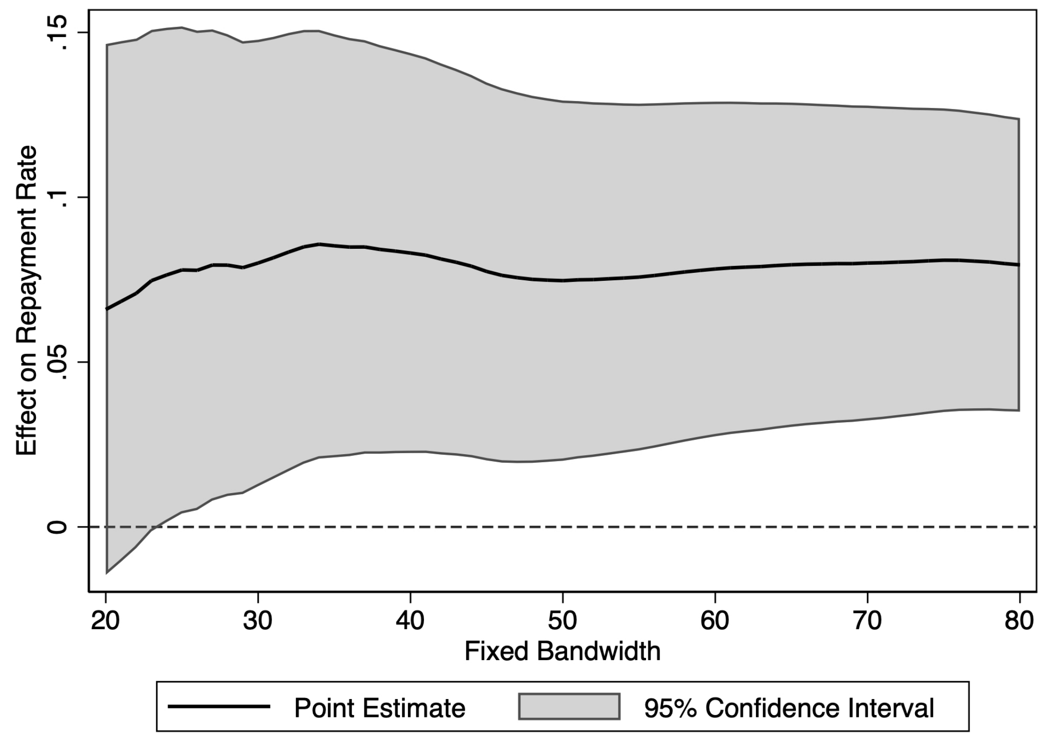

Figure 4: Robustness of the repayment rate discontinuity to fixed bandwidths

we include data in a 120-day bandwidth to assess the out-of-sample fit of each polynomial. The estimated discontinuity is robust up to a quartic specification (although the effect size remains consistent with the range identified in Table 2). We estimate a 7.7 p.p. (13.1%) increase ($p = 0.049$) for the quadratic model, a 8.8 p.p. (14.9%) increase ($p = 0.059$) for the cubic model, and a 5.0 p.p. (8.5%) increase ($p = 0.314$) for the quartic model. The out-of-sample fit for the non-linear models appears poor due to over-fitting, most often predicting large declines in repayment moving away from the cutoff in both directions. The linear model does a better job predicting a smooth increase in repayment leading up to the acquisition of credit bureau scores, and a stable repayment rate afterwards.

While our application controls for weekly patterns, there remains substantial day-to-day volatility in loan application arrival. Since weekly volume is more stable, as a robustness check we also collapse repayments by week and re-run our analysis at the weekly level where the running variable is measuring the number of weeks relative to the introduction of credit bureau scores. Figure 6 shows two different weekly discontinuity estimates: one estimated using the optimal bandwidth of 4 weeks (Panel A), and one estimated using the entire sample

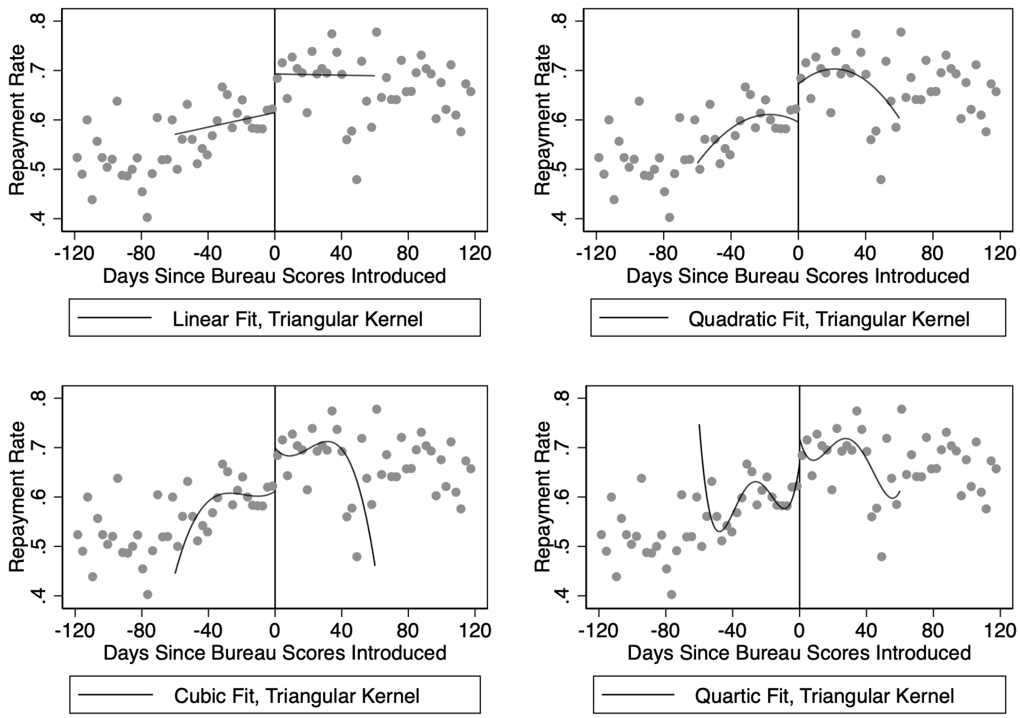

Figure 5: Robustness of repayment rate discontinuity to higher-order polynomial fits

(Panel B). The estimates are similar both to one another and to the daily estimates. The optimal bandwidth estimate represents an increase of 7.7 p.p. (13.2%, $p = 0.009$) and the full sample estimate is an increase of 6.6 p.p. (11.2%, $p = 0.002$). Notably, the bandwidth-bias p-values in these specifications are smaller than the daily specification ($p = 0.039$ and $p = 0.016$, respectively), reflecting the stability of the weekly estimate across bandwidths.

Regression discontinuity in time estimates are subject to concerns regarding the volatility in the outcome variable over time. Hence, we conduct a placebo test to determine how likely it would be to obtain our estimate when applying our model to a randomly-selected day in the sample. We estimate every possible 60-day bandwidth regression discontinuity in our sample, starting 92 days before the introduction of credit scores, through 150 days after. Figure 7 shows the estimated kernel density of all the estimates (as percentages of the relevant 60-day pre-cutoff means). We add the 95th and 97.5th percentiles of the distribution (corresponding to the one- and two-sided 95% density regions), as well as the points representing the true cutoff, and the days before and after. Our main estimate (corresponding to column (1) in Table 2) is in the 98th percentile of the distribution, as is the estimate from the day

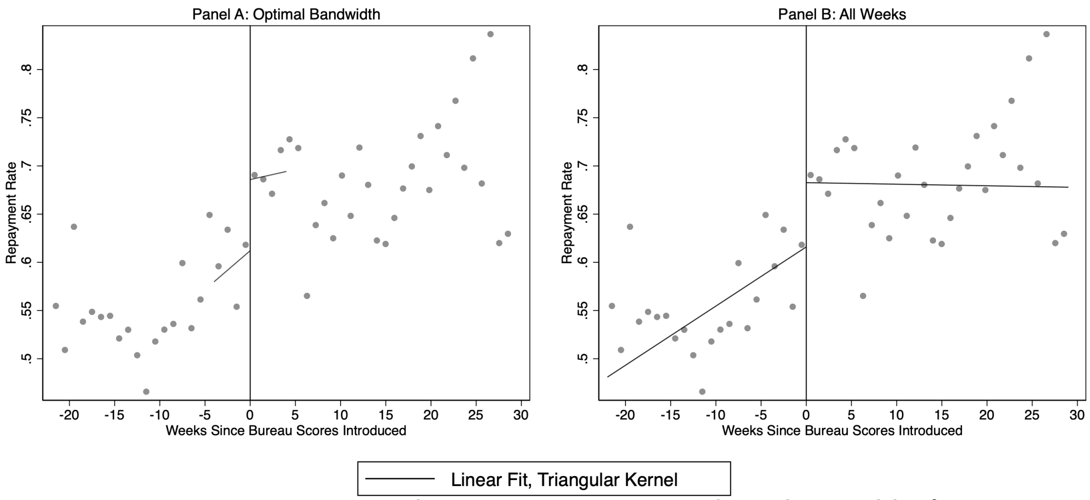

Figure 6: Repayment rate discontinuity estimated at the weekly frequency

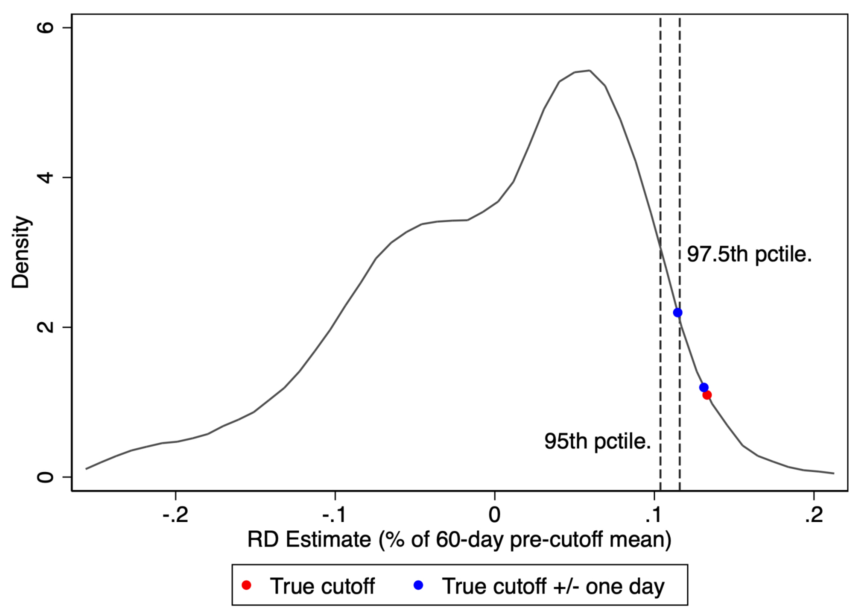

Figure 7: Kernel density of all possible 60-day bandwidth RD estimates

immediately prior. The estimate from the day immediately after is in the 96.5th percentile. We also consider the z-statistics associated with each estimate as a joint measure of effect size and precision. Our main estimate using the true cutoff features the second-largest z-statistic of any of the 243 estimates.

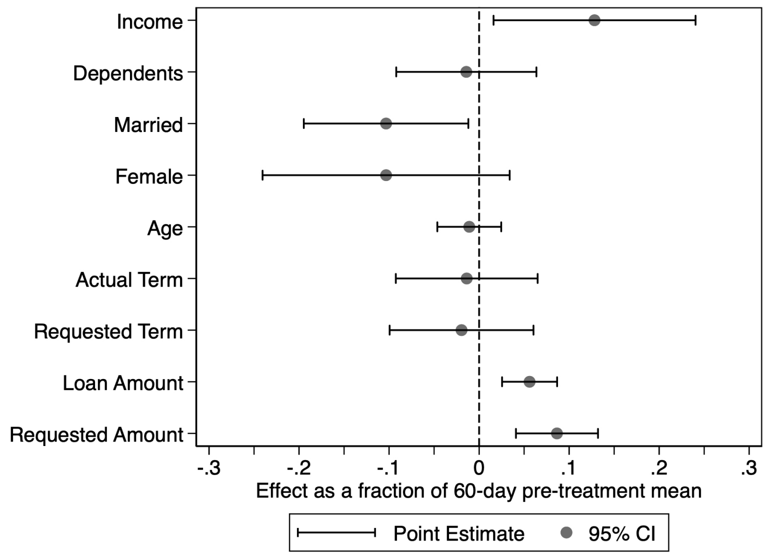

Figure 8: Changes in loan & borrower observables when credit scores are acquired

3.3.2 Borrower characteristics

We next use the same regression discontinuity model to estimate changes in borrower and loan characteristics when credit bureau scores are acquired. The estimates come from an optimal bandwidth model without control variables (corresponding to column (4) in Table 2) and are plotted in Figure 8 as percentages of the 60-day pre-treatment mean. We find that after acquiring credit scores, approved borrowers have higher incomes, asked for larger loans, and received them. The increase in loan size is in line with the findings from De Janvry et al. (2010). They attribute this findings to a reduction of credit rationing (McIntosh and Wydick (2009)), while in our setting, this is due to selection of higher income individuals with larger demands for credit. This further indicates that the lender did not use access to credit bureau scores to identify missed opportunities among the previously rejected low-income applicants for small loans. Another effect of credit score acquisition appears to be that single applicants (reduction of 5.9 p.p., estimate and bandwidth bias-corrected estimate of $p = 0.027$ and $p = 0.081$, respectively) and perhaps women (reduction of 4.5 p.p., estimate and bandwidth bias-corrected estimate of $p = 0.140$ and $p = 0.091$, respectively) are a smaller fraction of the approved loans after credit bureau scores are introduced.

Finally, we use borrower characteristics to test the hypothesis that observable borrower data should be less predictive of loan repayment after the lender obtains credit scores. If married borrowers, for example, are less likely to repay loans prior to the adoption of credit scores, credit scores should inform that probability, and the predictive power of marital status among *approved* borrowers should go down. To evaluate this, we use LASSO model selection to predict repayment using borrower characteristics.[16] We do this separately for the periods 30-60 days prior to the adoption of credit scores, 0-30 days prior, and 0-30 days after. This allows us to evaluate the within-sample predictive success of the models, as well as the one-period ahead predictive success both within the pre-cutoff period, and across the cutoff.

The predictive success of the models is poor. However, as all loans in our sample have been approved, this is not surprising. That said, all of our borrower characteristics –log income, number of dependents, age, marital status, sex, requested term, and requested amount– are retained by the selected 30-60 day pre-cutoff model. Age, sex, and marital status are retained by the selected 0-30 day pre-cutoff model. Only income is retained in the selected 0-30 day post-cutoff model. The within-sample R^2 values of the pre-cutoff models are 0.056 and 0.057. However, the R^2 falls to 0.041 for the post-cutoff model (a decrease of 27-28%). The out-of-sample fits are so poor that the R^2 values are negative –suggesting a naïve would outperform the selected models– but the degree of this failure is much larger when predicting pre-cutoff to post-cutoff, than when predicting pre-cutoff to pre-cutoff: a 300% decline vs. a 143% decline in R^2.

4 Conclusion

In this paper, we study the impact of introducing information from a credit bureau on repayments for a digital lender in Mexico. There are three takeaways from our analysis. First, regression discontinuity estimates show that the acquisition of additional information in the form of credit scores improves repayments. The likelihood of borrowers' repayment increases by 10-13% immediately after the acquisition of credit scores. Second, borrower

[16] We search over linear models that always include the amount and term of the loan, day-of-month and hour-of-day fixed effects. Sample fit exercises are conducted using the post-selection coefficients, not subject to shrinkage. Output is reproducible with a seed of one.

characteristics became less predictive of repayment once scores were adopted. Third, the sample of approved borrowers after the lender acquired the credit scores had higher incomes, asked and received larger loans. Because overall volume did not go up, this suggests that lending may have become less inclusive of low-income households.

Our study highlights the potential benefits that information sharing through participation in CRBs could bring to digital lenders. More work is needed to explore the role of CRBs in digital financial markets for both consumers and lenders, including evidence that highlights the impact of mandatory information sharing on the financial inclusion of those lacking ties to the formal financial sector.

References

Behr, P. and S. Sonnekalb (2012). The effect of information sharing between lenders on access to credit, cost of credit, and loan performance–evidence from a credit registry introduction. *Journal of Banking & Finance 36*(11), 3017–3032.

Bennardo, A., M. Pagano, and S. Piccolo (2015). Multiple bank lending, creditor rights, and information sharing. *Review of finance 19*(2), 519–570.

Bharadwaj, P., W. Jack, and T. Suri (2019). Fintech and household resilience to shocks: Evidence from digital loans in kenya. Technical report, National Bureau of Economic Research.

Björkegren, D., J. E. Blumenstock, O. Folajimi-Senjobi, J. Mauro, and S. R. Nair (2022). Instant loans can lift subjective well-being:a randomized evaluation of digital credit in nigeria. *Working paper*.

Björkegren, D. and D. Grissen (2018). The potential of digital credit to bank the poor. In *AEA Papers and Proceedings*, Volume 108, pp. 68–71.

BOU (2017). National financial inclusion strategy. Retrieved from `https://www.bou.or.ug/bou/bouwebsite/bouwebsitecontent/publications/special_pubs/2017/National-Financial-Inclusion-Strategy.pdf`.

Brailovskaya, V., P. Dupas, and J. Robinson (2021). Is digital credit filling a hole or digging a hole? evidence from malawi. Working paper, National Bureau of Economic Research.

Burlando, A., M. Kuhn, and S. Prina (2022). Too fast, too furious? digital credit delivery speed and repayment rates. *Working paper*.

De Janvry, A., C. McIntosh, and E. Sadoulet (2010). The supply-and demand-side impacts of credit market information. *Journal of development Economics 93*(2), 173–188.

Djankov, S., C. McLiesh, and A. Shleifer (2007). Private credit in 129 countries. *Journal of financial Economics 84*(2), 299–329.

Doblas-Madrid, A. and R. Minetti (2013). Sharing information in the credit market: Contract-level evidence from us firms. *Journal of financial Economics 109*(1), 198–223.

Einav, L., M. Jenkins, and J. Levin (2013). The impact of credit scoring on consumer lending. *The RAND Journal of Economics 44*(2), 249–274.

Finnovista (2021). Radar fintech e imcumbentes 2021. *Working paper*.

Flatnes, J. E. (2021). Information sharing and rationing in credit markets. *American Journal of Agricultural Economics 103*(3), 944–960.

Francis, E., J. Blumenstock, and J. Robinson (2017). Digital credit: A snapshot of the current landscape and open research questions. *CEGA White Paper*.

Galindo, A. and M. Miller (2001). Can credit registries reduce credit constraints? empirical evidence on the role of credit registries in firm investment decisions. In *Annual Meetings of the Inter-American Development Bank, Santiago Chile*.

Jappelli, T. and M. Pagano (2002). Information sharing, lending and defaults: Cross-country evidence. *Journal of Banking & Finance 26*(10), 2017–2045.

Karlan, D. and J. Zinman (2010). Expanding credit access: Using randomized supply decisions to estimate the impacts. *The Review of Financial Studies 23*(1), 433–464.

Love, I. and N. Mylenko (2003). *Credit reporting and financing constraints*. Number 3142. World Bank Publications.

McIntosh, C. and B. Wydick (2009). What do credit bureaus do? understanding screening, incentive, and credit expansion effects.

Morse, A. (2011). Payday lenders: Heroes or villains? *Journal of Financial Economics 102*(1), 28–44.

Padilla, A. J. and M. Pagano (1997). Endogenous communication among lenders and entrepreneurial incentives. *The review of financial studies 10*(1), 205–236.

Padilla, A. J. and M. Pagano (2000). Sharing default information as a borrower discipline device. *European Economic Review 44*(10), 1951–1980.

Pagano, M. and T. Jappelli (1993). Information sharing in credit markets. *The journal of finance 48*(5), 1693–1718.

Stiglitz, J. E. and A. Weiss (1981). Credit rationing in markets with imperfect information. *The American economic review 71*(3), 393–410.