STAFF REPORTS

NO. 929
JUNE 2020

REVISED
JUNE 2025

Credit and Entrepreneurs' Income

Manthos Delis | Fulvia Fringuellotti | Maria Iosifidi | Steven Ongena

FEDERAL RESERVE BANK of NEW YORK

Credit and Entrepreneurs' Income

Manthos Delis, Fulvia Fringuellotti, Maria Iosifidi, and Steven Ongena

Federal Reserve Bank of New York Staff Reports, no. 929
June 2020; revised June 2025
JEL classification: G21, D31, E24, O15

Abstract

Small business entrepreneurs facing credit constraints may experience significantly different future income trajectories compared to their unconstrained counterparts. We quantify this difference using uniquely detailed loan application data and a regression discontinuity design based on a bank's credit score cutoff rule employed in the loan approval process. Our findings indicate that loan acceptance increases recipients' real income by eleven percent five years later compared to rejected applicants. This effect persists across a wide range of robustness tests and is primarily driven by the utilization of borrowed funds for profitable investments, as captured by the bank's ex-ante soft information and the ex-post firm performance. Additionally, within the cohort of accepted applicants, future income is higher for those who were easily accepted compared to marginally accepted borrowers with similar creditworthiness, highlighting the important efficiency effects of loan usage.

Key words: credit constraints, entrepreneurs' income, business loans, regression discontinuity design, efficient use of loans

____

Fringuellotti: Federal Reserve Bank of New York (email: fulvia.fringuellotti@ny.frb.org). Delis: Audencia Business School (email: mdelis@audencia.com). Iosifidi: Montpellier Business School (email: m.iosifidi@montpellier-bs.com). Ongena: Department of Banking and Finance, University of Zurich (email: steven.ongena@df.uzh.ch). This paper was previously circulated under the titles "Credit, Income, and Income Inequality," "Credit and Income Inequality," "Credit, Income, and Inequality," and "Credit and Income." The authors are grateful to participants of the IMF 5th Annual Macro-Financial Research Conference, ASSA 2021 Virtual Annual Meeting, the European Finance Association 2020 Annual Meeting (Helsinki), the 35th Annual Congress of the European Economic Association (Rotterdam), the 2021 Federal Reserve "Week After" Conference on Financial Markets and Institutions, the CEPR Endless Summer Conference on Financial Intermediation and Corporate Finance (Larnaca), the 9th MoFiR Workshop on Banking, the CompNet/NBS Conference on "Sustainable Development, Firm Performance and Competitiveness Policies in Small Open Economies," and the 18th Paris Finance Meeting for many comments. They also thank seminar participants in the Essex Business School (University of Essex), Adam Smith Business School (University of Glasgow), Durham Business School, Federal Reserve Bank of New York, Montpellier Business School, and the Athens University of Economics and Business. For comments, they thank an anonymous referee, an associate editor, Knut Are Aastveit, Christoph Basten, Tobias Berg, Rajashri Chakrabarti, Nicola Cetorelli, Jerry Coakley, Murillo Campello (editor), Francisco Covas, Hans Degryse, Yota Deli, Giovanni Dell'Ariccia, Filippo Di Mauro, Robin Döttling, René Garcia, Miguel Ferreira, Pedro Gete, Mariassunta Giannetti, Linda Goldberg, Jens Hagendorff, Iftekhar Hasan, Zhiguo He, Rajkamal Iyer, Martina Jasova, Olivier De Jonghe, Sotirios Kokas, Alexandros Kontonikas, Michael Lamla, David Lucca, John Mondragon, Georgios Panos, Nicola Pavanini, Diane Pierret, Maxim Pinkovskiy, Tomasz Piskorski, Matthew Plosser, Andrea Presbitero, Simon Price, Manju Puri, Enrichetta Ravina, Jean Charles Rochet, Farzad Saidi, Asani Sarkar, Robert M. Sauer, Enrico Sette, László Tétényi, Wilbert van der Klaauw, and Diego A. Vera-Cossio.

1. Introduction

What is the effect of a bank's decision to approve or reject a business loan application on entrepreneurs' future income? Can we quantify this effect in developed economies, and what are the implications for entrepreneurs' economic mobility? The answers to these questions have important implications for the income of business owners. Assume we observe two groups of entrepreneurs with similar income levels and other observable characteristics prior to a bank's lending decisions. If credit is granted to one group and denied to the other, these two groups may experience significantly different outcomes in their future income. Entrepreneurs' ex-ante characteristics, such as the quality of their investment opportunities, can influence their future income. Equally important, the efficiency in ex-post utilization of loan proceeds by entrepreneurs can further affect their future income. Our study aims to explicitly quantify these different outcomes in future income and explain the main economic channels driving this effect.

Theoretically, asymmetric information between lenders and borrowers affects credit availability. Because the enforcement of loan contracts is imperfect, lenders ration credit and often require borrowers to pledge collateral. A relaxation of credit constraints may lead to more financing opportunities for the full spectrum of potential borrowers and a possible tightening of the income distribution (Banerjee and Newman, 1993; Galor and Zeira, 1993). However, credit-constrained individuals often have limited wealth. Wealth (or capital) endowment plays a critical role in the loan origination process acting as a fixed cost for credit access. The relatively poor or those with other inferior prospects cannot always overcome it, irrespective of the quality of their investment ideas. As a consequence, their exclusion from credit can hinder economic mobility and fuel persistent income inequality (Piketty, 1997; Mookherjee and Ray, 2003; Demirgüç–Kunt and Levine, 2009).

While several studies examine how credit availability affects individuals' income using aggregate measures of credit supply and economic outcomes or income inequality, understanding and quantifying the extent to which access to credit impacts individuals' income from a micro perspective is paramount. The existing literature on microfinance in developing economies has shed some light on the role of microcredit programs to spur entrepreneurship and help people escaping from poverty traps (e.g., Brei et al., 2018; Minetti, et al., 2019).

We instead study the extent to which entrepreneurs in a developed economy, who are similar in terms of income and other traits when applying for credit, experience significantly different incomes after the credit decision. This analysis involves two channels. The first is how the loan decision of the bank (approval or rejection) affects future income. The second is how the use of credit (efficient or not) affects future income. We show the important implications (real effects of credit) of these two channels on firm growth, upward mobility, and income inequality among entrepreneurs. Our analysis is based on a unique data set of business loan applications to a single large European bank. The uniqueness of our data lies in the available information on business owners, which encompasses income, wealth, and the credit scores assigned by the bank, as well as other applicant and firm characteristics throughout the sample period.

Our focus is on loan applications from small and micro enterprises that are majority-owned by individuals-entrepreneurs. This focus yields two major advantages for investigating our research questions. First, the income of such entrepreneurs is highly correlated to the performance of their business. Second, the bank has information on the business owners' incomes and decides whether to grant the loans based on a credit score cutoff rule. Specifically, each applicant receives a credit score at the time of the loan application. The credit score is an internal rating constructed by the bank and it is not affected by the applicant. Then, credit is granted to applicants whose credit scores are above the cutoff and denied otherwise.

We further restrict our sample to a balanced panel of bank customers i) who apply multiple times to this bank and ii) have an exclusive relationship with the bank. This ensures that we track applicants' income before and after the credit decision, and that we estimate the effect of credit on income avoiding potential confounding factors such as other sources of funding beyond this bank. We closely track that our sample is representative of European averages and note that our results are robust to the use of more general samples, as well as methods to overcome sample-selection bias. Our baseline sample covers 61,863 loan applications submitted by 15,628 individuals over the period 2002-2016.

The availability of credit score and future applicants' income is crucial for our identification strategy because it allows exploiting the cutoff rule as a source of exogenous variation in the credit decision. Our approach builds on the idea that individuals whose credit scores are around the cutoff are virtually the same in terms of credit quality, yielding a sharp regression discontinuity design (RDD). This implies identification from comparing changes in the income of accepted and denied applicants, who prior to the bank's credit decision have similar credit scores (including similar incomes).

Our key finding is that, on average, a loan origination increases the recipient's real income five years onward by 11% compared to rejected applicants. This finding is robust to several re-specifications and is not affected by the mix of the control variables. Further, the RDD passes a battery of tests looking at credit score manipulation, uniqueness of the cutoff point and falsification tests, continuity of applicants' attributes (control variables) around the cutoff, sample representativeness and selection bias.

Overall, this result indicates that a bank's credit decisions (loan origination or rejection) have significant real effects on entrepreneurs' income. Consider two applicants: the first has a credit score slightly above the cutoff; the second has a credit score slightly below the cutoff. At the time of the loan application, the credit quality of these two individuals is virtually the same. However, the cutoff rule implies that credit is granted only to the former. The 11% increase in real income experienced by the accepted applicants vs. the rejected applicants documents a causal link between access to credit and income. This link is not obvious. As documented in various studies on microfinance in developing countries, access to credit may have no impact on the income of small business owners (e.g., Banerjee et al., 2015; Tarozzi et al., 2015). Intuitively, a loan origination improves individual income only if credit is granted to applicants having good investment opportunities. This is likely to be the case for our bank, which is a major financial institution operating in a developed economy in Europe. Therefore, our finding reveals that access to credit has a positive effect on applicants' income when lending decisions are taken efficiently. Also, the magnitude of this effect, which we are the first to identify, is substantial and further allows pinpointing the impact of credit access on firms' economic opportunities and entrepreneurs' upward mobility.

We, thus, explore the economic channels behind the response of applicants' income to a loan origination, as well as the implications of credit access for small business owners. First, we look at the role of applicants' ex-ante characteristics that are observable only to the bank, such as the quality of their investment projects. We show that the positive impact of a positive credit decision on income is more pronounced when the soft information held by the bank enters positively in the calculation of the credit score. This confirms that the effect of a loan origination on income is far from obvious, as it depends on whether credit is granted to applicants having good investment opportunities.

Next, we examine if the positive impact of a loan origination on future income is driven by the ex-post efficient use of the borrowed funds. We show that firms of accepted applicants allocate a larger amount of funds to finance investments and business operations, experience a higher increase in profitability, grow at a higher rate compared to firms of rejected applicants and are more likely to repay previous loan obligations. In addition, we document that easily accepted applicants (those further away from the cutoff point) experience higher future income compared to marginally accepted applicants with similar credit scores, consistent with the idea that the income response is more pronounced for entrepreneurs who are able to use the borrowed funds more efficiently.

Our work relates to the broad literature investigating the effect of bank credit on income (see Berger et al., 2020, for a broad overview). From an empirical viewpoint, our study is close to the strand of literature on microfinance in developing countries (Augsburg et al., 2015; Kaboski and Townsend, 2005; Banerjee et al, 2019). These studies show that various microcredit programs in developing countries did not have a pivotal impact on individual income in developing countries. We show that, in the context of a developed economy and for a leading financial institution, a loan origination has generally a positive and large effect on applicants' income. On the same line, a substantial body of related literature examines how various social and economic conditions (including race, gender, education, parents' socioeconomic class, local neighborhood, income inequality etc.) affect individual opportunities and, hence, economic mobility (Chetty et al., 2014; Bell et al., 2019). We contribute to this literature documenting that credit provision to small businesses is pivotal in fostering entrepreneurship and upward income mobility.

Our work also relates to the literature that looks broadly at how credit expansions and/or constraints affect the income distribution by relying on aggregate (at the country or regional level) measures of inequality (mostly the Gini index) and financial development (Delis et al., 2014; Naceur and Zhang, 2016; de Haan and Sturm, 2017). In addition, our paper is close to several other studies on finance and income or wages (see Demirgüç-Kunt and Levine, 2009; Buera et al., 2011; Moser et al., 2018; Shin, 2018). We contribute to this literature by proposing a rigorous identification setup to study the effect of credit origination on income and the income distribution at the micro level.

A broader body of literature documents how credit constraints affect economic and social outcomes (e.g., Herkenhoff et al., 2012; Balduzzi, et al., 2017; Berg, 2018; Berton et al., 2018) or the transmission of shocks (Kashyap and Stein, 2000; Klein et al, 2002; Duflo and Banerjee, 2014; Popov and Rocholl, 2018). We contribute to this literature showing that the effect of credit origination on the income of small business owners is stronger at the growth stage of a firm, in low-income regions, and all instances where entrepreneurs are disadvantaged or more credit constrained.

2. Data and Empirical Identification

3.1. Loan Applications

We use a unique sample of loan applications to a single large European bank directly supervised by the ECB under the Single Supervisory Mechanism and headquartered in a rich northern European country.[1] The bank provides credit to a wide array of small and large firms, as well as

[1] The bank is considered a systematically important financial institution based on the criteria defined by the European Banking Authority (EBA), see https://eba.europa.eu/risk-analysis-and-data/global-systemically-important-institutions and https://eba.europa.eu/risk-analysis-and-data/other-systemically-important-institutions-o-siis-.

to consumers, households, and the public sector both domestically and abroad. Our sample is limited to loan applications from individuals, firms and administrations that are located in the country where the bank is headquartered.

We consider all types of commercial credit, including working capital loans, mortgages, lines of credit, venture loans for startups, etc. Importantly, we use only loan applications from small and micro enterprises (total assets less than €10 million as per the European Commission's definition) that are majority-owned by specific individuals (i.e., holding more than 50% of equity).[2] The reason why we restrict the sample to this subcategory of applicants is twofold: first, the evolution of income of such entrepreneurs is almost uniquely tied to the performance of their business; second, for these applicants, the bank has information that is essential to address our research question. Specifically, we have information on whether the loan is originated or denied, as well as loan characteristics, firm characteristics, and applicant characteristics. Loan characteristics include the requested amount and maturity, as well as other features such as collateral, covenants, and performance-pricing provisions if the loan is originated. Firm characteristics encompass several accounting variables, such as assets and sales, profits, leverage, as well as the firm's region and industry.

What makes this data unique is information on the applicant (the firm's majority owner). Table 1 defines the variables we use in our analysis and Table 2 reports summary statistics. The applicant characteristics include income (total deflated income reported by the individual, including wages, "dividends" from the firm, and any other source of income),[3] assets (wealth),

[2] Using the European Commission's definition, a small enterprise has total assets less than €10 million; a micro enterprise less than €2 million in assets.

[3] We are unaware of situations in which an entrepreneur obtains a personal loan to invest in real estate, with the firm renting the property or similar cases. Generally, all loans we include in our analysis are obtained by the firm, not the entrepreneur. Even in such situations, the entrepreneur should report this loan as "personal income," and the firm would include it as an expense. Controlling for an expenses ratio does not affect our inferences.

gender, education, relationship with the bank (an exclusive relationship or not), and the credit score assigned by the bank. We identify applicants having an exclusive relationship with the bank as those who do not have a lending relationship with another regulated commercial bank, even if their application(s) to our bank is (are) rejected.[4] The exclusivity of the relationship consists in an objective fact and does not stem from any legal agreement between the firm and the bank.

From a methodological perspective, a crucial piece of information that allows us to investigate our research question is the credit score assigned by the bank. Each applicant is given a credit score at the time of the application, and this score is the decisive factor in loan origination. The credit score consists in a private rating constructed by the bank, which is not accessible to anyone including the applicant. The bank generates the credit score based on both hard information (observable applicant and firm characteristics) and soft information (e.g., the bank's perception of the applicant, the quality of the investment opportunities of the firm, the strength of the firm-bank relationship). For comparative purposes, we normalize the credit score to be around the cutoff value of 0. The bank originates the loan if the credit score is higher than 0 and denies the loan otherwise (Berg, 2018). For very few applications (72 cases), this criterion does not hold. These exceptions are possibly due to data-entry mistakes and thus we disregard them in our analysis. The cutoff rule adopted by the bank does not change over our sample period.

[Insert Tables 1 & 2 about here]

Our original data set includes 97,659 loan applications over the time period 2002-2016.[5] For two reasons, we restrict our sample to loan applications from individuals who have exclusive

[4] Our bank has information on any credit relationship in place between a firm and another supervised bank (by the EBA or the country's regulatory and supervisor authority) from both the firm and the national credit register, irrespective of whether the loan application to our bank is accepted or rejected.
[5] This data set is generated starting from a broader panel at the firm-year level that includes all the information collected by the bank on each applicant. Specifically, applicants (firms) are the cross-sectional unit of the panel and the years from 2002 to 2016 are the time unit.

relationships with the bank (as per our definition) and apply multiple times during the sample period. First, the bank has income information for these applicants for several years before and after the loan decision.[6] Second, these applicants are generally unable to obtain credit from another bank, especially if their application is denied; moreover, they cannot access capital markets due to the firm's small size. This ensures that we can estimate the effect of access to credit on income avoiding potential confounding factors due to other sources of funding beyond this bank.

In principle, a rejected applicant may seek credit in the shadow-banking sector which is largely unregulated. However, ceteris paribus, non-banks are likely to charge higher interest rates and, generally, apply worse credit terms than banks given their higher cost of capital (Chen et al., 2017). In addition, a number of reports by Deutsche Bank (2014), OECD (2014), and BIS and FSB (2017) suggest that, in Europe during our sample period, SMEs had very limited access to credit outside of the banking system. Consistently with that, in the subsection presenting our empirical findings, we show evidence that denied applicants (having an exclusive relationship with our bank) do not get credit elsewhere after a rejection. In general, it is fairly common for small and micro firms to have an exclusive relationship with a bank. In our full sample this is the case for 65% of firms, which is close to the value of 71% documented by Berger and Schaeck (2011) for SMEs in three large European countries (i.e., Germany, Italy and UK). Overall, these characteristics of our sample allow us to identify the effect of the bank's credit decision on applicants' income.

Our final data set includes 15,628 applicants (firms) and 61,863 loan applications over 2002-2016. We conduct an extensive set of tests to show that the 61,863 loan applications used in our analysis (out of the total 97,659) do not introduce any selection bias (see Section 4.2). This is

[6] To understand how we exploit this feature to build our dataset, consider the case of a business owner who lodges two loan applications during our sample period. Then, the future income of the entrepreneur with respect to the time of the first application corresponds to the information on his/her past income collected by the bank at the time of the second application.

a balanced panel of entrepreneurs, who (i) apply multiple times to this bank, ii) do not have a credit relationship with another bank at the time of any of their loan application, and iii) provide information to the bank throughout our sample period (2002-2016), irrespective of whether they apply for a loan in a given year. Entrepreneurs who borrow from another bank or cease to exist anytime between 2002 and 2016 are excluded from the sample. For each applicant, we know the income and the other characteristics defined in Table 1 during the entire sample period.

The number of loan applications in a given year ranges between 3,500 and 4,750, with historical peaks in 2006 and 2016 and a marked drop during the financial crisis (see Figure 1). All individuals reapply for loans within a four-year period and the average time between two consecutive applications is 2.9 years. Business owners who apply from 3 to 5 times account for 70% of applicants (see Figure 1). Individuals apply on average around four times during our sample period and are either always accepted (11,956 applicants, or 77%), or sometimes accepted and sometimes rejected (3,672 applicants, or 23%); no business owner is always rejected. The bank accepts 87% of loan applications and rejects 13%. Applicants that experience at least one loan denial make on average 4.4 loan applications and are accepted 52% of the time. This is a first piece of evidence suggesting that accepted and rejected applicants are similar enough and this is especially true for individuals whose credit score is in the neighborhood of the cutoff.

Applicant's income in natural logarithm ranges between 9.85 (min) and 12,29 (max), corresponding to €18,996 and €217,510, respectively. The average income (11.01 in natural logarithm, or €60,475) is above GDP per capita of the Eurozone, which ranges between 23,000 euro and 32,000 euro during our sample period (2002-2016).[7] This is not surprising as we focus on small business owners and the bank operates in a developed northern European country. The

[7] See Eurostat PPS and Eurostat GDP Per Capita.

mean future real income (respectively, in one year, three years, and five years) tends to rise over time for loan applicants. After its transformation, the mean credit score is positive and equal to approximately 0.1.

[Insert Figure 1 about here]

The bank provides credit to firms in all industries, according to the Statistical Classification of Economic Activities in the European Community (commonly known as NACE codes). Our sample includes firms in all industries except from loans to firms in the “Public Administration and Defence; Compulsory Social Security” and “Activities of Extraterritorial Organisations and Bodies” industries. Most loan applicants are firms in wholesale and retail trade, but all industries are fairly well represented in our sample. In Figure 2, we report a chart with the share of firms by industry. The wholesale/retail and market services industries are by far the most widely represented, followed by manufacturing and construction/real estate.

[Insert Figure 2 about here]

Using data from a single entity is not an unusual practice when the research question is detailed (Adams et al., 2009; Iyer and Puri, 2012; Fracassi et al., 2016; Berg, 2018; Delis et al., 2020). In our case, we take advantage of granular application-level data for one bank to document how the decision to grant or deny credit affects individuals’ income. Also, the bank that we look at is a major financial institution operating on a national scale. This ensures that the bank is representative enough for the banking system, so that we can reasonably generalize the results of our study.

We, nonetheless, perform three formal checks to verify that the bank and firms in our sample (i.e., small and micro businesses that have an exclusive relationship with the bank and apply at multiple times in our sample period) exhibit similar characteristics to other systemic

European banks and other small European firms, respectively. These tests include a comparison of (i) the bank's characteristics with averages of other European banks, (ii) access to credit by our firms vis-à-vis other similar European firms, and (iii) characteristics of firms in our sample with European averages. As shown in Appendix A, our sample is representative across these dimensions.

3.2. Empirical Identification

This study aims at shedding light on the impact of access to credit on income and income inequality from a micro-perspective. A natural way to identify this effect is to assess how a bank's credit decision (credit origination or denial) affects the distribution of income across accepted and rejected applicants, and how this effect varies depending on loan, applicant and macroeconomic conditions. Three important features of our data set making this a viable approach are the availability of information about (i) originated and denied loans, (ii) the exclusivity of the relationship between loan applicants and banks (the applicant does not obtain credit from another regulated commercial bank if his/her application is rejected),[8] and (iii) applicants' income before and after the loan application. Based on these features, a standard identification method would compare the incomes of approved applicants (the treated group) with the incomes of rejected applicants (the control group) before and after the loan decision. Unfortunately, the treatment here is endogenous to several factors behind the bank's decision to grant the loan, making a difference-in-differences exercise far from optimal.

[8] The bank has this information from the applicants, meaning that no other bank is able/willing to finance the same project. This feature of our sample implies that the loan applicants do not leave the sample; therefore, we do not have such attrition bias.

The fourth and most important feature of our data set for identification purposes is the availability of information on credit scores and the perfect correlation of the scores above the cutoff with loan origination. This implies a sharp discontinuity in treatment as a function of credit score. Therefore, we rely on a sharp RDD using credit score as the assignment (also referred to as "the running" or "the forcing") variable (Imbens and Lemieux, 2008; Lee and Lemieux, 2010). Moreover, the credit score controls for the ex-ante perceptions of the bank regarding the expected profitability of the firms' projects. This limits the reverse causality due to future expected profits.

Assuming that the relation between access to credit and income is linear, a simple form of the RDD is:

$$y_{i,t+n} = a_0 + a_1 D_{it} + a_2(x_{it} - \bar{x}) + a_3 D_{it}(x_{it} - \bar{x}) + u_{it}. \tag{1}$$

In equation (1), y is applicant's i income in the n^{th} year ahead of the loan application, which takes place in year t. D is a binary variable that equals 1 if the credit score x is above the cutoff $\bar{x}$ and zero otherwise, which determines whether the loan is granted. Thus, a_1 captures the average treatment effect. Also, $x_{it} - \bar{x}$ is the distance between the cutoff and applicant i's credit score given at the time of the loan application. Finally, the interaction $D_{it}(x_{it} - \bar{x})$ is included to capture non-linearities in the relationship between applicant's income and the credit score (i.e., a differential slope of this relationship on the two sides of the cutoff).

While the linear model of equation (1) is intuitive, it presents an important limitation, namely it identifies the treatment effect placing equal weight on all the information available in the sample. This may lead to a potential bias, as observations far from the cutoff are treated in the same way as observations close to the cutoff. To overcome this issue, we also consider a local linear regression model (for a general description, see Imbens and Lemieux, 2008; Calonico et al., 2014). According to this model, the average treatment effect is nonparametrically identifiable as:

$$\tau_{RDD} = \lim_{\varepsilon \to 0^+} \mathbb{E}\left[y_{i,t+n} | x_{it} = \bar{x} + \varepsilon\right] - \lim_{\varepsilon \to 0^-} \mathbb{E}\left[y_{i,t+n} | x_{it} = \bar{x} + \varepsilon\right], \quad (2)$$

where the two conditional expectations are estimated by fitting linear regression functions to the observations on either side of the cutoff in a neighborhood of it. The advantage of the nonparametric model is twofold. First, it requires using a data-driven optimal bandwidth to identify observations that are close enough to the cutoff. Note that the credit score encompasses all the applicants' characteristics observable to us, as well as attributes that are observable solely by the bank (e.g., soft information). This implies that applicants with similar credit scores look alike across several dimensions. Thus, the optimal bandwidth ensures one considers a neighborhood of the credit score around the cutoff that is sufficiently narrow to include applicants that are virtually identical except for the treatment (i.e., the loan outcome). Second, this approach allows using a kernel smoother to assign higher weights as we move closer to the cutoff. Following Calonico et al. (2014) and Calonico et al. (2018), we use the mean squared error optimal bandwidth and a triangular kernel in our nonparametric estimation. In addition, we mainly base our inference on the local-quadratic bias-correction of Calonico et al. (2018) for efficient estimation.

The main assumption for the validity of the linear model of equation (1) and the nonparametric model of equation (2), similar to any other RDD, is that applicants cannot precisely manipulate their credit scores and loan officers do not artificially adjust the credit scores to move applicants on either side of the cutoff. If applicants, even while having some influence, are unable to manipulate their credit scores precisely and loan officers do not perform ad hoc adjustments of the credit scores, the variation in treatment around the cutoff provides a randomized experiment. The lack of non-random sorting and self-selection is the most compelling requirement of the RDD vis-à-vis other identification methods, such as differences-in-differences or instrumental variables (Lee and Lemieux, 2010).

Theoretically, precise manipulation by applicants is unlikely, as loan officers' prudent behavior should prevent applicants from having exact information on their credit scores. Although credit underwriting has increasingly become an automated process in the past decades thanks to digitalization (Straka, 2000; Frame et al., 2001; Evans and Schmalensee, 2005), we cannot fully rule out that loan officers manipulate the credit score of their applicants fostering an approval or a rejection. In our setup, self-selection or non-random sorting of applicants would entail a discontinuous change in the distribution of the credit score around the cutoff. A simple and immediate approach to verify if this condition is met is to check if there is any discontinuity in the empirical density of the assignment variable (credit score) and the outcome variable (business owners' income after a loan application) in our sample. Figure 3 shows that the probability density of the credit score does not jump around the cutoff and the distribution of applicant's income exhibits a regular shape, hereby validating our regression discontinuity design.

[Insert Figure 3 about here]

We, next, rely on specific statistical test to formally demonstrate that credit score manipulation either by applicants or loan officers is absent in our setup. Specifically, we test for manipulation of the assignment variable around the cutoff. This test consists in a data-driven statistical technique which relies on local polynomial to construct the density of the running variable. We present the outcome of the test in Table 3 and its graphical representation in Figure 4. Consistent with the approach adopted in our baseline nonparametric model of equation 2, we estimate the density of the credit score relying on a local quadratic estimator with cubic bias-correction and triangular kernel. The solid black line in Figure 4 represents the estimated density of the credit score, whereas the shaded grey indicates its 95% confidence interval. The test statistics and p-values presented in Table 3, and, similarly, the large overlapping of the 95% confidence interval on the two sides of the cutoff in Figure 4, indicate that the null hypothesis of no manipulation (i.e., no discontinuity around the cutoff) cannot be rejected. Thus, in line with the graphical evidence, the formal test of Cattaneo et al. (2018) confirms there is no statistical evidence of manipulation of the forcing variable. As we show later in our empirical results, we do not find evidence of manipulation of the credit scores even when we focus on the subsample of applicants for which the soft information held by the bank enters positively (or negatively) in the calculation of the credit score. This further corroborates that loan officers do not artificially adjust the credit scores of applicants around the cutoff.

[Insert Table 3 & Figure 4 about here]

As discussed above, a crucial aspect of a RDD is to estimate the treatment effect by comparing treated and control units that are sufficiently similar to each other. The credit score encompasses all the applicants' characteristics observable to us, as well as attributes that are observable solely by the bank (e.g., soft information). Thus, considering a subset of applicants with a credit score around the cutoff, as in the nonparametric model, ensures that one estimates the treatment effect by comparing applicants that are virtually the same along different dimensions. Nevertheless, we still may want to explicitly check that indeed the treated and control groups are sufficiently similar to each other. In Section 3.1 we presented a series of statistics suggesting that accepted and rejected applicants in our balanced panel share a similar borrowing behavior. In addition, we verify that the relation between observable loan and applicant characteristics (loan amount, loan maturity, initial income, initial wealth, education, firm size, firm leverage) and the credit score is smooth around the cutoff. The graphical evidence presented in Figure 5 reveals that applicants above and below a close neighborhood of the cutoff are comparable along all these dimensions.

[Insert Figure 5 about here]

The RDD models of equations (1) and (2) estimate the average effect of a bank's credit decision on applicants' income. We build on these two models to identify the differential impact of access to credit on income based on loan, applicant and macroeconomic conditions.

4. Empirical Results

4.1. Average Treatment Effect

We begin our RDD analysis with a graphical inspection of the relation between access to credit and income. Figure 6 shows applicants' income five years after the loan decision against the credit score. There is a clear upward shift in applicants' income around the cutoff point. This indicates that the treatment (loan origination) entails a sharp discontinuity in the outcome variable (income), corroborating our methodological approach. Moreover, the cutoff is clearly unique (no multiple cutoff points), pointing to a sharp RDD.

[Insert Figure 6 about here]

Also, the plot shows a linear relation between applicants' income and the credit score on both sides of the cutoff. The relation looks slightly increasing below the cutoff and almost flat above. This evidence suggests that the econometric analysis should focus on a linear regression model or a local linear regression model, as we do. More importantly, the upward discontinuity in applicants' income at the cutoff, as well as the flat relationship between income and credit score above the cutoff, reveal that access to credit plays a preeminent role in shaping the future income path of small business owners.

The starting point of our formal empirical analysis is to identify the average effect of credit origination on applicants' income (estimation of equations 1 and 2). Table 4 reports the results, with Panel A reporting the parametric OLS results and Panel B the nonparametric results. Specifications 1-3 use as a dependent variable the applicants' income one year ahead, three years ahead, and five years ahead of the loan application. Specifications 4-6 replicate the results by additionally using control variables.[9] We find a positive and statistically significant coefficient on *Granted* in all specifications. The magnitude of the effect does not exhibit a marked difference depending on whether we include or not the set of controls, likely because these covariates capturing hard information are largely accounted for by the credit score.

[Insert Table 4 about here]

For economic inferences, we rely on the nonparametric results, which place more weight on individuals around the cutoff (as per our discussion of equation 2). Each specification is estimated on a subset of our balanced panel pertaining to business owners having a credit score within the range of the optimal bandwidth, as indicated by the effective number of observations used above and below the cutoff. For each specification, we report the conventional RD estimates with conventional variance estimator (*Conventional*), the bias-corrected RD estimates with conventional variance estimator (*Bias-corrected*), and the bias-corrected RD estimates with robust variance estimator (*Robust*). We find an income increase of approximately 6% among approved applicants one year or three years after the loan origination, and an increase of approximately 11% five years ahead. This effect translates to approximately 7,000 euro higher real income five years ahead. While the positive impact of credit access on income increases from 1 to 5 years after the loan application, it decays from 6 years onwards. The evidence that the treatment effect peaks

[9] On the use of control variables, a key assumption of the RDD is that the expectation of the outcome variable conditional on the assignment variable is continuous. This requires that the relation between the covariates and the credit score is smooth around the cutoff. A graphical inspection confirms that this condition is fulfilled (Figure 5). This means that our baseline model in equation (2) is well specified, and using the controls will not significantly affect our main result.

around 5 years suggests that a loan origination affects the income of small business owners mostly because it allows to undertake investments and expand the business, rather than to smooth temporary shocks. It takes time for investments to fully have their effect, and it appears this occurs over the medium term. This conjecture is further confirmed in Section 4.4 where we investigate the economic channels behind the impact of credit on income more closely.

4.2. Robustness of the Average Treatment Effect

We conduct a very large battery of sensitivity tests, which show that our RDD and the associated average treatment effect is robust. First, we conduct robustness tests on the parametric model. In Table A1 of Appendix B, we document that our parametric results in Table 4 are robust to (i) the inclusion of industry, loan type, and year fixed effects. In Table A2, we show that the parametric results are robust to controlling for the entrepreneurs' initial wealth. Most important, in Table A3 we show that our parametric results come closer to the nonparametric ones when restricting our sample around the cutoff point.

We next turn to the robustness tests of our baseline nonparametric model, on which we base our main inference. We first assess in a more explicit way if business owners are able to obtain credit outside of the banking system after a loan rejection from our bank. Table A4 reports values for total firm debt, before and after the loan application, measured relative to total assets in the year prior to the loan application for the subsets of accepted and rejected applicants considered in the 17,917 "effective observations" around the cutoff where we estimate the nonparametric RDD of Table 4. While firms of accepted applicants, especially those in the tail of the distribution of leverage, experience an increase in total debt right after a loan origination, debt financing of firms or rejected applicants remains almost unchanged after a loan denial and, if anything, slightly declines.

A second concern is that changes in income within and across groups of individuals may be influenced by reasons that are independent from the bank's credit decision. We, thus, conduct two validation tests of our RDD approach to rule out this hypothesis. The first includes falsification tests on different (invalid) cutoff points for the credit score. Specifically, we estimate a placebo version of specification (6) of Panel B of Table 4 by arbitrarily setting the cutoff at the credit score values -1.5, -1, -0.5, 0.5, 1, 1.5. We report the RD estimates from these six regressions in Table A5. All estimates are statistically insignificant, showing no effect of a positive credit decision on future income at these falsified cutoffs.

A third concern is that our framework considers applicants with an exclusive relationship with the bank. While working on such balanced panel limits concerns of attrition bias and allows us to estimate the treatment effect focusing on individuals for which we have comprehensive information, there is a downside related to the potential introduction of a selection bias. This is because we overlook one-time applicants who may drop out of the sample because they turn to another lender or decide to stop operating their business (for example after a denied application). We also discard firm owners who have credit relationships with multiple banks. If these applicants differ in a substantial way from individuals who have an exclusive relationship with the bank and apply multiple times, we may either underestimate or overestimate the effect of credit of income.

As a first exercise on this note, we compare applicants in our sample of 61,863 loan applications (i.e., those who have an exclusive relationship with the bank and apply multiple times during our sample period) to those in the discarded sample of 35,796 loan applications based on a set of observables. Summary statistics reported in Table A6 suggest that the two groups are comparable across all attributes. In addition, the analysis presented in Appendix A shows that small firms in our sample are on average comparable to other small firms operating in the euro area. While this limits concerns of a potential selection bias in our sample, we need to address the issue in a formal way.

To this end, we begin with a parametric two-stage selection model as in, e.g., Heckman (1976), Dass and Massa (2011), and Jiménez et al. (2014). In the first stage, we estimate the probability that a loan application is submitted in a specific year by a bank customer who has an exclusive relationship with the bank and applies multiple times in our sample period (probit model). We run this regression on our broad data set at the firm-year level including all the information on applicants collected by the bank and spanning the time window 2002-2016. This consists in an unbalanced panel of all applicants, irrespective of whether they have an exclusive relationship with the bank or not and apply a single or multiple times. The right-hand side variables in the first stage encompass the applicant's attributes of columns 4-6 of Tables 4, excluding the credit score (which is unknown to the applicant) and including *Gender*. In the second stage, we run a similar regression to the one implied in equation (2), in which we use the predicted instantaneous probability of applying for a loan (Mills ratio) from the first stage as an additional control variable.[10]

Concerning the exclusion restriction, we find that *Gender* is significantly and positively correlated with the probability of a loan application by an individual with a long-lasting relationship with the bank but does not explain future income in the baseline specifications. In other words, males are more likely to apply for credit than females, as documented also in Delis

[10] Given that the sample of our baseline RDD is a balanced panel of bank customers with an exclusive credit relationship and these customers appear in the panel irrespective of whether they apply for a loan in a given year, we can also model the probability of receiving a loan application in the baseline setup. The results of this exercise are similar to those here and are available upon request.

et al. (2020), but any effect on the future income of male and female entrepreneurs is transmitted via this higher probability of male entrepreneurs to apply for credit, once having accounted for other individual and firm characteristics. Importantly, we also document that the bank's credit decision is not driven by gender (i.e., we find no evidence that the bank discriminates between male and female applicants, *ceteris paribus*). For these reasons, we argue that *Gender* satisfies the exclusion restriction, and we include this variable only in the first stage regression.

Table 5 reports the estimation results. The first-stage results show that income, wealth and education positively and strongly affect the probability of a loan application by an individual with a long-lasting exclusive relationship with the bank. The same holds for owners of more leveraged firms. Interestingly, we also find that male applicants are 0.8% more likely to apply for credit than female applicants. The second-stage results are fully in line with Table 4, with the Mills ratio having a positive but insignificant coefficient (which is indication of limited endogeneity in the OLS model). This suggests that the selection effect is very low and the estimation of the treatment effect using a balanced panel of individuals having a long-lasting exclusive relationship with the bank delivers reliable results.

To account for selection of loan applicants, we prefer to use the conventional parametric model because it is standard in the applied economics/finance literature, whereas the nonparametric models are quite rare in this respect. However, we do an experiment with a semiparametric model, where we save the parametric first-stage prediction and include it in the nonparametric second stage. Again, the results, reported in Table A7, are consistent with those of Table 4.

[Insert Table 5 about here]

We conduct several additional tests in Appendix B. Specifically, we (i) hold the number of effective observations across the different time horizons in columns 4 to 6 of Panel B of Table 4 constant (Table A8), (ii) include the applicant's initial wealth in the nonparametric model (Table A9), (iii) control for total credit received by each firm from the bank to deal with the history of the specific bank-firm relationship (Table A10), and (iv) use alternative bandwidth selection methods (Table A11). Also, Figure 7 shows that the significance of the conventional estimate in model (3) of Panel B of Table 4 is robust to a continuum of different windows around the cutoff where (small-sample) inference is conducted.

[Insert Figure 7 about here]

Two additional tests highlight that our baseline results are robust to reverse causality due to higher profitability of loan recipients that is not captured by the credit score. The first is to use information on whether the firm co-finances the projects(s) for which it applies for credit; generally, this implies more favorable prospects. We include the dummy variable *Co-financing* that equals 1 if this is the case (equals 0 if there is no co-financing). The results in specification 1 of Table A12 show that our inferences remain robust.

The second test is to include ex-post information on the credit score, given that this is regularly available to us and allows capturing if expectations on future firm performance at the time of the loan application have de facto materialized. Specifically, we use the difference between the first or the third lead of the credit score and the current credit score as a control variable and report the results in specifications 2 and 3 of Table A12. Again, our results remain robust, indicating that the credit score assigned by the bank at the time of the loan application does, on average, a good job in capturing the firms' prospects.

Last, we examine if the impact of credit decisions on applicants' income varies based on macro conditions across geographical regions and over the business cycle. The results in Table A13 show that, five years after a loan origination, accepted applicants have 12% higher incomes than rejected applicants in low-income regions. The equivalent effect in the high-income regions is 9%, indicating that the incomes of individuals in high-income regions are less affected by credit decisions compared to low-income regions (where credit constraints are higher). Considering the role of the business cycle, we use three subsamples: (i) loan applications submitted in the pre-global financial crisis period (2002-2007), when GDP of the Eurozone exhibits a positive growth; (ii) loan applications submitted during the double-dip recession period (2008-2013), when the euro area experienced negative GDP growth, substantial contraction of credit, and increased unemployment (e.g., IMF, 2009; ECB, 2016); (iii) loan applications submitted during the post-crisis recovery period (2014-2016), when GDP of the Euro area was growing but had not achieved its pre-crisis level and credit to the private sector over GDP was still decreasing. The results presented in columns 3-6 of Table A13 indicate that the positive impact of a loan origination on income is somewhat stronger in periods of economic growth (2002-2007 and 2014-2016) compared to periods of recession (2008-2013).

4.3. The Role of Soft Information

In this section, we explore the role played by soft information held by the bank in driving the real effect of credit decisions on individuals' income. Hard information consists in the observable characteristics listed in Table 1. Soft information includes any other relevant feature of the applicant and the firm that is unobservable, such as the quality of the investment opportunities of the firm, the bank's perception of the loan applicant, the strength of the firm-bank relationship, etc. While both hard information and soft information contribute to the bank's credit decision, what leads the effect of credit on income is far from clear.

To decompose the credit score into hard information and soft information, we regress the credit score on the set of observables capturing hard information (income, wealth, education, firm size, firm leverage, loan amount, maturity, availability of collateral, and use of loan covenants). We then interpret the residuals as the component of the credit score ascribable to soft information. We find that 77% of the credit score is explained by hard information, with the remaining ascribable to soft information. A natural question is whether loan officers make ad hoc adjustments to the credit scores, which depart from an unbiased assessment of the applicant, to influence an acceptance or a rejection. Such adjustments would be embedded in the component of the credit score represented by soft information and would imply a discontinuity in the probability density function of the credit scores in a neighborhood of the cutoff. As discussed in Section 3, we do not detect any form of manipulation when we look at the entire distribution of the credit scores in our sample. As a complementary more granular exercise, we replicate the statistical test of Cattaneo et al. (2018) also on the subsamples of observations where soft information enters positively and negatively in the calculation of the credit score (i.e., the subsamples where the residuals are positive and negative, respectively). The results in Panel A of Table 6 show that there is no statistical evidence of an artificial manipulation of credit scores from loan officers.

[Insert Table 6 about here]

As a second step, we replicate the nonparametric regression in column 6 of Table 4, splitting the data in two subsamples based on the sign of the residuals (positive residuals in the first subsample and negative or equal to zero in the second). Panel B of Table 6 reports the results. In specification 1, we compare the future income of accepted and rejected applicants for which the private assessment of the loan officer affects positively the credit score; in specification 2, instead, we compare the future income of accepted and rejected applicants for which the soft information held by the loan officer negatively affects the credit score. Even though soft information explains only 23% of the credit score, the effect of credit origination on individuals' income is stronger when soft information makes a loan acceptance more likely. In particular, five years after a bank's credit decision, accepted applicants experience an increase in income of 13.5% when soft information enters positively into the credit score (column 1), compared to 7% when soft information contributes negatively (column 2).

This finding suggests that the marginal benefit of getting a loan is stronger when a loan acceptance is favored by an ex-ante positive assessment of the bank on unobservable characteristics of the applicant. To see this more clearly, let us consider a simple example of two entrepreneurs, A and B, who have an exclusive credit relationship with the bank. Entrepreneur A falls on the right side of the cutoff, whereas B is on the left side. Our estimates suggest that the difference in income between A and B after the credit decision is more pronounced if A and B have good investment opportunities (positive soft information) than bad investment opportunities (negative soft information). This further confirms that the effect of loan origination on income is far from trivial, as it depends on the level of efficiency of the bank in granting credit.

4.4. Ex-post Use of Proceeds

Our results in Table 4 essentially answer the question of how the bank's credit decisions affect entrepreneurs' future income and wealth. However, these findings provide limited information on how differences in productivity and efficiency related to loan usage impact entrepreneurs' future income and wealth. This limitation arises from a fundamental observation on the RDD: while we do not observe the counterfactual (i.e., how rejected borrowers would have used the loans), the RDD setup ensures that rejected applicants near the cutoff would likely have utilized the loan proceeds with similar efficiency levels as the approved applicants near the cutoff, given their similar traits. In other words, productivity and efficiency differences across firms (which are typically time-invariant) are largely accounted for in the RDD. This is also the essence of the results in Figure 5.

Thus, examining how the efficient use of loan amount after the bank's credit decision affects entrepreneurial outcomes requires a different approach. Comparing accepted applicants near the cutoff point with those further awthe ay from it can provide insights into this question. To this end, we re-estimate equation 2, replacing the treated group with accepted applicants whose credit score is far from the cutoff point (i.e., "easily" accepted applicants) and the control group with accepted applicants whose credit score is close to the cutoff point, respectively. The definitions of "close" and "far" are derived from the analysis in Table 4 and Table A3. Thus, accepted applicants near the cutoff are those within the bandwidth of the nonparametric models of columns (4) to (6) of panel B in Table 4. Accepted applicants far from the cutoff are the remaining accepted loan applicants. We use the same covariates as in Table 4, with loan characteristics being especially important to control for scale effects.

The results in Table 7 show considerable differences in the efficiency of loan use between marginally accepted and easily accepted loan applicants, particularly over time. In the first year after loan origination, there is a 2.5% higher income for easily accepted entrepreneurs compared to marginally accepted entrepreneurs. This income gap increases to 7.2% three years after loan origination and to 11.6% five years after loan origination. Given our controls for loan-level characteristics, especially loan amount, these findings can be attributed to the more efficient use of financing by easily accepted applicants.

In Table 8, we expand our analysis of borrowed funds efficiency by examining the loan purpose, loan type, and firms' financial outcomes after the lending decision. From a methodological perspective, we re-estimate the benchmark nonparametric model of equation 2 by either i) substituting the dependent variable or ii) splitting the sample into different subsamples and test if the average treatment effect is statistically different across these subsamples. To perform this heterogeneity analysis, we rely on a Z-test as follows:

$$H_0\colon \tau_{RDD1} = \tau_{RDD2}$$

$$H_1\colon \tau_{RDD1} \neq \tau_{RDD2}$$

$$Z = \frac{abs(\tau_{RDD1} - \tau_{RDD2})}{\sqrt{SE(\tau_{RDD1})^2 + SE(\tau_{RDD2})^2}}$$

where τ_{RDD1} and τ_{RDD2} denote the treatment effects estimated on the two subsamples according to the nonparametric model of equation (2), and $SE(\tau_{RDD1})$ and $SE(\tau_{RDD2})$ are the standard errors calculated based on the robust variance estimator of Calonico et al. (2018). In a sharp regression discontinuity setup characterized by a large proportion of compliers (i.e., accepted applicants) in each subsample, this approach delivers reliable inference (Hsu and Shen, 2019).[11]

We start by distinguishing between credit requested to finance investment projects versus credit requested to smooth short-term liquidity shocks and gauge which of these two economic channels is the key driver of the positive impact of credit on the income of small business owners.

[11] The heterogeneity tests are run by splitting the sample into subgroups, where each subsample has several thousand observations, from which we select an optimal bandwidth around the cutoff following Calonico et al. (2018).

While we cannot observe the specific purpose of the loan requested, we can nevertheless shed some light on the use of the borrowed funds by exploring two alternative outcome variables. In the first two specifications of Table 8, we use i) the natural logarithm of the amount of credit used for corporate purposes (e.g., expansion projects, investments, working capital needs, inventory purchases, equipment acquisition, or other operational expenses) and ii) an indicator for whether the borrower has repaid previous loan obligations.

We find that, five years after a loan origination, the amount of funds used to finance investments and business operations increases by 13% for accepted applicants compared to rejected ones (column 1). At the same time, the likelihood that accepted applicants repay previous loan obligations with the bank increases by 5% relative to rejected applicants (column 2). This suggests that the positive impact of a loan origination on applicants' income likely stems from the use of the borrowed funds to make investments and support operational expenses.

[Insert Tables 7, 8 & 9 about here]

In specifications 3 to 6 of Table 8, we use *Firm growth*$_{t+5}$, *ROA*$_{t+5}$, *Income growth*$_{(t+5)}$, and the ratio of total revenue to total leverage (*Debt turnover*$_{t+5}$) to analyze future firm performance. In the debt turnover specifications, we exclude *Firm leverage* from the controls. We expect that results using these measures move in tandem with those from entrepreneurs' income, and our findings support these expectations. Specifically, forward ROA is 0.048 units higher in accepted applicants compared to rejected ones (or 51% higher for the mean applicant), *Income growth*$_{(t+5)}$ is 14% higher, and *Debt turnover* is 6.7% higher. These findings further highlight the superior productivity and performance of accepted firms following the bank's credit decision. These results are largely consistent with those of Berg (2018), who shows, also in a RDD setup, that loan origination has a positive effect on firm growth, investments and employment.

In specifications 7 and 8 of Table 8, we split the sample into term loans and credit lines, excluding a few other loans that do not clearly fall into these categories, such as invoice financing and equipment financing. The treatment effect estimator is somewhat larger for term loans that are more likely to finance long-term investment projects, but it is also significant for credit lines. Moreover, in specifications 9 and 10, we split the sample into secured and unsecured loans. The difference in the estimates between the two specifications is statistically and economically insignificant. Lastly, in specifications 11 and 12 of Table 8, we differentiate between limited liability and unlimited liability firms. Although most firms in our sample are limited liability firms, the results for both groups are consistent with our baseline.

We conclude our study by investigating if, conditional on applicants' characteristics, the size, cost and sequence of credit originations affects the impact of a new loan on entrepreneurs' future income. Borrowers obtaining larger and cheaper loans, or benefiting from a long-lasting relationship with the bank through a series of loans, may find it easier to break even and profit from their investments compared to other borrowers with a similar credit quality. We start by looking at the role of loan size. In the first two specifications of Table 9, we replicate the nonparametric regression of column 6 of Panel B of Table 4, splitting our full sample into small loans and large loans, which are identified as those below the 25th and above the 75th percentile of the distribution of *Loan amount*, respectively. We find that the effect of credit access on individual income is somewhat stronger for larger loan amounts. In particular, the income of approved applicants rises by 11.8% five years ahead of the loan origination for large loans (column 2), versus 10.5% for small loans (column 1).

In specification 3, we zoom into the continuous intensity of treatment driven by loan size to shed light on the full-scale effects of the loan amount on future income. This approach requires departing from the RDD setup to focus solely on accepted applicants. To this end we estimate a modified version of the parametric model of column 6 of Panel A of Table 4 where we use only approved loans and *Loan amount* instead of *Granted* as the main explanatory variable. A limitation of this approach is that the identification of causal effects is not as robust because we cannot utilize the sharp discontinuity. However, controlling for the credit score (i.e., overall creditworthiness) and the component of the credit score ascribable to soft information (i.e., quality of investment opportunities, bank-firm lending relationship etc.), as we do, can yield valuable inference. The results indicate that a 1% increase in the loan amount is associated with a 0.5% increase in the entrepreneur's future income. Thus, we observe scale effects of larger loan amounts even within the sample of accepted loan applicants. Note that the coefficient of *Soft information* is positive and highly significant, underscoring the influence of unobservable factors such as the quality of firm's projects, the bank-firm lending relationship, and other relevant soft conditions on future income.

In specification 4, we analyze cost of credit effects for approved borrowers by using the *Loan spread* as our main explanatory variable. We find a negative coefficient, suggesting that borrowers paying higher rates experience significantly lower increases in their future income and, hence, require more profitable projects to break even. The estimates indicate that a 1% increase in the loan spread is associated with a 0.33% decrease in income five years after loan origination.

Next, we examine the role of path dependencies (positive serial correlation in loan approvals). To this end, we rely on the approach by Kirschenmann (2016), who investigates the effect of each additional loan on credit rationing, measured by the ratio of the granted loan amount to the requested loan amount. In our setting, given that all entrepreneurs apply at least twice, we use dummy variables for loan numbers 3 to 6 (i.e., equal to 1 if a firm obtains a third/fourth/fifth/sixth loan during our sample period and 0 otherwise) as the main explanatory variables. Thus, we analyze the effect of obtaining each additional loan, from loan 3 to loan 6, on future income.[12]

The results in specification 5 show that each additional loan positively and significantly contributes to entrepreneurs' future income, with the economic significance of the estimates being largely comparable. Therefore, our results imply that each additional loan has a similar positive effect on income five years after its origination.

This remains true in specification 6, where we remove *Soft information* from the set of controls, to avoid accounting for unobservable factors that may affect the importance of each additional loan, such as the evolution of firm's investment projects. While the loan dummies become more significant, we still find that, at least from an economic perspective, the differences in the coefficients between *Loan 3* and *Loan 6* are not substantial to indicate strong path dependence.

5. Conclusions

Credit constraints can hinder income growth opportunities, particularly for individuals with low incomes and insufficient collateral. Using unique data from business loan applications to a large European bank, we quantify how the bank's credit decisions (acceptance or rejection) affect applicants' future real income. Our identification strategy is based on a regression discontinuity design, which exploits exogenous variation in credit decisions derived from the cutoff rule based on credit scores. This approach allows us to compare individuals with credit scores near the cutoff, ensuring that the characteristics influencing the credit decision are comparable.

[12] Additional tests exploring the interaction between *Loan 3* to *Loan 6* indicators with the total number of loans for each firm or the total loan amount received by each firm during our sample period (proxies for relationship lending) are available upon request.

We demonstrate that access to credit has a significant positive effect on individual income three to five years after the loan application, with a key channel being the soft information held by the bank. Additionally, we find that firms of accepted applicants utilize the borrowed funds to make investments and expand their businesses, ultimately achieving higher profitability and growth rates compared to firms of rejected applicants. Lastly, we show that the income response is stronger for easily accepted applicants than for marginally accepted ones with a similar credit quality, indicating larger income effects for entrepreneurs who are able to use the borrowed funds more efficiently.

References

Adams, W., Einav, L., Levin, J., 2009. Liquidity constraints and imperfect information in subprime lending. *American Economic Review* 99 (1), 49-84.

Augsburg, B., De Haas, R., Harmgart, H., Meghir, C., 2015. The impacts of microcredit: Evidence from Bosnia and Herzegovina. *American Economic Journal: Applied Economics* 7 (1), 183-203.

Balduzzi, P., Brancati, E., Schiantarelli, F., 2017. Financial markets, banks' cost of funding, and firms' decisions: Lessons from two crises. *Journal of Financial Intermediation* 36, 1-15.

Banerjee, A., Newman, A. F., 1993. Occupational choice and the process of development. *Journal of Political Economy* 101 (2), 274-298.

Banerjee, A., Duflo, E., Glennerster, R., Kinnan, C., 2015. The miracle of microfinance? Evidence from a randomized evaluation. *American Economic Journal: Applied Economics* 7 (1), 22-53.

Banerjee, A., Breza, E., Duflo, E., Kinnan, C., 2019. Can microfinance unlock a poverty trap for some entrepreneurs? NBER Working Paper no. 27346.

Bell, A., Chetty, R., Jaravel, X., Petkova, N., Van Reenen, J., 2019. Who becomes an inventor in America? The importance of exposure to innovation. *Quarterly Journal of Economics*, 134 (2), 647-713.

Berg, T., 2018. Got rejected? Real effects of not getting a loan. *Review of Financial Studies* 31 (12), 4912-4957.

Berger, A. N., Schaeck, K., 2011. Small and medium-sized enterprises, bank relationship strength, and the use of venture capital. *Journal of Money, Credit and Banking* 43, 461-490.

Berger, A. N., Molyneux, P., Wilson, J. O. S., 2020. Banks and the real economy: An assessment of the research. *Journal of Corporate Finance* 62, 101513.

Berton, F., Mocetti, S., Presbitero, A. F., Richiardi, M., 2018. Banks, firms and jobs. *Review of Financial studies* 31(6), 2113-2156.

Buera, F. J., Kaboski, J. P., Shin, Y., 2011. Finance and Development: A Tale of Two Sectors. *American Economic Review* 101, 1964–2002.

BIS and FSB, 2017. FinTech credit: Market structure, business models and financial stability implications.

Brei, M, Ferri, G., Gambacorta, L., 2018. Financial structure and income inequality. BIS Working Papers No. 756.

Calonico, S., Cattaneo, M. D., Titiunik, R., 2014. Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica* 82 (6), 2295-2326.

Calonico, S., Cattaneo, M. D., Farrell, M. H., 2018. On the effect of bias estimation on coverage accuracy in nonparametric inference. *Journal of the American Statistical Association* 113(522), 767-779.

Carvalho, I., 2017. What is the impact of the access to external finance on the capital structure of SMEs in Europe? PhD Dissertation, Católica Lisbon School of Business & Economics.

Cattaneo, M. D., Titiunik, R., Vazquez-Bare, G., 2016. Inference in regression discontinuity design under local randomization. *Stata Journal* 16 (2), 331-367.

Cattaneo, M. D., Jansson, M., Ma, X., 2018. Manipulation testing based on density discontinuity. *Stata Journal* 18 (1), 234-261.

Chen, B. S., Hanson, S. G., Stein, J. C., 2017. The decline of big-bank lending to small business: Dynamic impacts on local credit and labor markets. NBER Working Paper No. 23843.

Chetty, R., Hendren, N., Kline, P., Saez, E., 2014. Where is the land of opportunities? The geography of intergenerational mobility in the United States. *Quarterly Journal of Economics* 129 (3), 1553-1623.

Dass, N., Massa, M., 2011. The impact of a strong bank-firm relationship on the borrowing firm. *Review of Financial Studies* 24, 1204-1260.

De Haan, J., Sturm, J. E., 2017. Finance and income inequality: A review and new evidence. *European Journal of Political Economy* 50, 171-195.

Delis, M. D., Hasan, I., Iosifidi, M., Ongena, S., 2020. Gender, credit, and firm outcomes. *Journal of Financial and Quantitative Analysis*, 1-56.

Delis, M. D., Hasan, I., Kazakis, P., 2014. Bank regulations and income inequality: Empirical evidence. *Review of Finance* 18, 1811-1846.

Deutsche Bank Research, 2014. SME financing in the euro area: New solutions to an old problem. EU Monitor, Global Financial Markets.

Duflo, E., Banerjee, A., 2014. Do firms want to borrow more? Testing credit constraints using a direct lending program. *Review of Economic Studies* 81, 572-607.

ECB, 2016. The household finance and consumption survey: Results from the second wave. ECB Statistics Paper No 18. December 2016. Household Finance and Consumption Network.

Evans, D. S., Schmalensee, R., 2005. Paying with plastic. The digital revolution in buying and borrowing. Second Edition, MIT Press, Cambridge, Massachusetts, London, England.

Frame, W. S., Srinivasan, A., Woosley, L., 2001. The effect of credit scoring on small-business lending. *Journal of Money Credit and Banking* 33 (3), 813-825.

Fracassi, C., Garmaise, M.J., Kogan, S., Natividad, G., 2016. Business microloans for US subprime borrowers. *Journal of Financial and Quantitative Analysis*, 51 (1), 55-83.

Galor, O., Zeira, J., 1993. Income distribution and macroeconomics. *Review of Economic Studies* 60 (1), 35-52.

Greenwood, J., Jovanovic, B., 1990. Financial development, growth, and the distribution of income. *Journal of Political Economy* 98 (5), 1076-1107.

Heckman, J. J., 1976. The common structure of statistical models of truncation, sample selection, and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement* 5, 475–492.

Herkenhoff, K., Phillips, G. M., Cohen-Cole, E., 2012. The impact of consumer credit access on self-employment and entrepreneurship. Journal of Financial Economics 141(1), 345-371.

Hsu, Y., Shen. S., 2019. Testing treatment effect heterogeneity in regression discontinuity designs. *Journal of Econometrics* 208, 468-486.

Imbens, G. W., Lemieux, T., 2008. Regression discontinuity designs: A guide to practice. *Journal of Econometrics* 142, 615-635.

IMF, 2009. World economic outlook. Crisis and recovery. April 2009. World Economic and Financial Surveys.

Iyer, R., Puri, M., 2012. Understanding bank runs: The importance of depositor-bank relationships and networks. *American Economic Review* 102 (4), 1414-1445.

Jiménez, G., Ongena, S., Peydró, J. L., Saurina, J., 2014. Hazardous times for monetary policy: What do twenty-three million bank loans say about the effects of monetary policy on credit risk-taking? *Econometrica* 82 (2), 463-505.

Kaboski, J. P., Townsend, R. M., 2005. Policies and impact: An analysis of village-level microfinance institutions. *Journal of the European Economic Association* 3(1), 1–50.

Kashyap, A., Stein., J., 2000. What do a million observations on banks say about the transmission of monetary policy. *American Economic Review* 90, 407–28.

Klein, M. W., Peek, J., Rosengren, E. S., 2002. Troubled banks, impaired foreign direct investment: The role of relative access to credit. *American Economic Review* 93 (3), 664-682.

Lee, D. S., Lemieux, T., 2010. Regression discontinuity designs in economics. *Journal of Economic Literature* 48, 281-355.

Minetti, R., Murro, P., Peruzzi, V., 2019. One size does not fit all. Cooperative banking and income inequality. CASMEF Working Paper No. 2, September 2019.

Mookherjee, D., Ray, D., 2003. Persistent inequality. *Review of Economic Studies* 70 (2), 369-393.

Moser, C., Saidi, F., Wirth, B, 2018. The effects of credit supply on wage inequality between and within firms. Working paper.

Naceur, S. B., Zhang, R., 2016. Financial development, inequality and poverty: Some international evidence. IMF Working Paper 16/32.

OECD, 2014. Non-bank debt financing for SMEs: The role of securitization, private placements and bonds. OECD Journal: Financial Markets Trends 2014/1, 139-159.

Piketty, T., 1997. The dynamics of wealth distribution and the interest rate with credit rationing. *Review of Economic Studies* 64 (2), 173-189.

Popov, A., Rocholl., J., 2018. Do credit shocks affect labor demand? Evidence for employment and wages during the financial crisis. *Journal of Financial Intermediation* 36, 16-27.

Shin, Y., 2018. Finance and Economic Development in the Very Long Run: A Review Essay. *Journal of Economic Literature* 56(4), 1577–1586.

Straka, J. W., 2000. A shift in the mortgage landscape: The 1990s move to automated credit evaluations. *Journal of Housing Research* 11 (2).

Tarozzi, A., Desai, J., Johnson, K., 2015. The impacts of microcredit: Evidence from Ethiopia. *American Economic Journal: Applied Economics* 7 (1), 54-89.

Table 1
Data and variable definitions

Variable	Description
A. Dimension of the data	
Individuals	Loan applicants who have an exclusive relationship with the bank and are majority owners (own more than 50%) of a firm. These borrowers apply to the bank for one or more business loans during the period 2002-2016 and the loan is either originated or denied. Due to the exclusive relationship, the bank holds information on the individuals' income and wealth even outside the year of loan application.
Year	The years covering the period 2002-2016.
B. Dependent variables	
Income	The natural logarithm of the euro amount of individuals' real total annual income (deflated by the Harmonized Index of Consumer Prices), as reported by the loan applicant. This variable reflects all taxable income, including salary, dividends from the firm, and other sources, as reported by the entrepreneur in the loan application.
Corporate purpose	The natural logarithm of the amount of a loan used for corporate purposes (e.g., expansion projects, investments, working capital needs, inventory purchases, equipment acquisition, or other operational expenses).
Debt repay	A dummy variable equal to 1 if the borrower is repaying previous loan obligations and 0 otherwise.
ROA	The ratio of firm's net income to total assets.
Debt turnover	The ratio of firm's total revenue to total leverage.
Firm growth	The annual growth of firm assets, calculated as the difference between the current assets minus previous year's assets and that difference over the previous year's assets.
C. Explanatory Variables: Running variable and cutoff	
Credit score	The credit score of the applicant, as calculated by the bank. We normalize this variable to take values around the cutoff of 0. The bank originates the loan if the credit score is higher than 0 and denies the loan otherwise.
Granted	A dummy variable equal to 1 if the loan is originated (Credit score>640) and 0 otherwise (Credit score<640).
D. Other covariates	
Education	An ordinal variable ranging between 0 and 5 if the individual completed the following education. 0: No secondary; 1: Secondary; 2: Post-secondary, non-tertiary; 3: Tertiary; 4: MSc, PhD or MBA.
Firm size	The natural logarithm of the total firm assets.
Firm leverage	The ratio of firm total debt to total assets.
Firm age	The firm's age in years.
Gender	A dummy variable equal to 1 if the applicant is a male and 0 otherwise.
Loan amount	The natural logarithm of the requested loan amount in thousands of euros.
Loan spread	The difference between the loan rate and the LIBOR (in natural logarithm basis points).
Maturity	Requested loan maturity in months.
Collateral	A dummy variable equal to 1 if the requested loan is secured by collateral and 0 otherwise.
Covenant	A dummy variable equal to 1 if there is one or more covenants associated with the requested loan and 0 otherwise.

Wealth	The natural logarithm of the euro amount of individuals' total wealth, as estimated by the bank and reported by the loan applicant. This includes only the portion of wealth that is pledgeable by the bank. The variable is deflated by the Harmonized Index of Consumer Prices.
Initial wealth	Individuals' wealth in the first year before the loan application in which this information is available (one to five years before). The variable is deflated by the Harmonized Index of Consumer Prices.

Table 2
Summary statistics

The table reports summary statistics (number of observations, mean, standard deviation, minimum, and maximum) for the variables used in the empirical analysis. The variables are defined in Table 1. For the variable Income, *t-1*, *t+1*, *t+3* and *t+5* stand for 1 year before, 1 year after, 3 years after, and 5 years after the loan application occurring at time *t*, respectively.

	Obs.	Mean	St. dev.	Min.	Max.
Income	61,863	11.01	0.376	9.852	12.29
Income t-1	57,682	10.58	0.406	9.804	12.62
Income t+1	57,766	11.10	0.388	9.866	12.58
Income t+3	49,514	11.14	0.373	9.987	12.57
Income t+5	41,391	11.16	0.363	10.04	12.62
Granted	61,863	0.867	0.498	0	1
Credit score	61,863	0.103	1.205	-2.921	2.100
Gender	61,863	0.811	0.387	0	1
Education	61,863	2.975	1.018	0	5
Firm size	61,863	12.821	0.806	2.500	16.12
Firm leverage	61,863	0.207	0.0249	0.143	0.917
Firm age	61,863	14.20	14.87	0	182
Loan amount	61,863	3.551	1.948	0.679	10.960
Loan spread	61,863	6.014	4.021	3.500	6.893
Maturity	61,863	34.35	10.14	7	103
Wealth	61,863	12.14	0.556	8.564	14.05
Initial wealth	40,953	12.09	0.406	7.952	14.20
Corporate purpose	61,863	1.925	0.714	0.679	5.825
ROA	61,863	0.094	0.160	-0.711	0.836
Debt turnover	61,863	3.140	1.406	0.162	12.28
Firm growth	61,863	0.193	0.386	-1.938	6.484

Percentile	10%	25%	75%	90%	Median
Credit score	0.047	0.068	1.185	1.637	0.116
Loan amount	2.953	3.201	5.042	5.931	3.620
Firm size	10.954	11.290	14.104	15.403	12.913
Firm age	1	7	30	59	15
Firm leverage	0.161	0.189	0.317	0.577	0.234

Table 3

Manipulation test

The table reports results from the manipulation testing procedure using the local polynomial density estimator proposed by Cattaneo et al. (2018). To perform this test, we rely on the local quadratic estimator with cubic bias-correction and triangular kernel. We report the conventional test statistic (Conventional) and the robust bias-corrected statistic (Robust) along with the corresponding p-value. The null hypothesis consists in no manipulation.

	T-stat	P-value
Conventional	1.5861	0.1127
Robust	1.2064	0.2277

Table 4
RDD results

The table reports coefficients and standard errors (in parentheses). The dependent variable is given in the first row of the table and all variables are defined in Table 1. *t-1*, *t+1*, *t+3* and *t+5* stand for 1 year before, 1 year after, 3 years after, and 5 years after the loan application occurring at time *t*, respectively. Estimation method in Panel A is OLS on the RDD model of equation (1). In Panel B, the estimation method is the local linear regression with triangular kernel on the RDD model of equation (2). For each specification in panel B, we report the conventional RD estimates with conventional variance estimator, the bias-corrected RD estimates with conventional variance estimator, and the bias-corrected RD estimates with robust variance estimator. Specifications (1) to (3) do not include any covariate besides the assignment variable (Credit score). More covariates are included in specifications (4) to (6). The *, **, and *** marks denote statistical significance at the 10%, 5%, and 1% level, respectively. In panel B, Effective obs. are the effective number of observations (determined by the bandwidth) left and right of the cutoff. BW estimate is the estimate of the bandwidth and BW bias is the associated bias. Following, Calonico et al. (2014), we use the mean squared error optimal bandwidth. The bias-corrected RD estimator and the robust variance estimator are obtained according to Calonico et al. (2018) and Calonico at el. (2014), respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	Income t+1	Income t+3	Income t+5	Income t+1	Income t+3	Income t+5
Panel A: Parametric results						
Granted	0.0512***	0.0730***	0.0699***	0.0536***	0.0754***	0.0718***
	(0.0062)	(0.0064)	(0.0069)	(0.0063)	(0.0066)	(0.0072)
Credit score	-0.0015	0.0060	0.0120***	-0.0056	0.0027	0.0084*
	(0.0038)	(0.0039)	(0.0042)	(0.0039)	(0.0041)	(0.0044)
Granted x Credit score	-0.0013	-0.0122**	-0.0216***	0.0026	-0.0087	-0.0168***
	(0.0052)	(0.0053)	(0.0057)	(0.0053)	(0.0056)	(0.0060)
Income t-1				0.0958***	0.0653***	0.0452***
				(0.0041)	(0.0043)	(0.0045)
Education				0.0023	-0.0017	0.0004
				(0.0016)	(0.0017)	(0.0019)
Firm size				-0.0004	0.0030	-0.0015
				(0.0021)	(0.0022)	(0.0024)
Firm leverage				0.1872***	0.2877***	0.2435***
				(0.0672)	(0.0745)	(0.0778)
Loan amount				-0.0008	-0.0023	-0.0014
				(0.0020)	(0.0021)	(0.0023)
Maturity				0.0004**	0.0001	0.0002
				(0.0002)	(0.0002)	(0.0002)
Observations	57,766	49,514	41,391	53,585	45,333	37,210
Adjusted R-squared	0.004	0.010	0.010	0.015	0.015	0.013
Clustering	Individual	Individual	Individual	Individual	Individual	Individual
Panel B: Nonparametric results						
Conventional	0.0599***	0.0605***	0.107***	0.0623***	0.0605***	0.105***
	(0.0127)	(0.0134)	(0.0166)	(0.0126)	(0.0146)	(0.0170)
Bias-corrected	0.0632***	0.0572***	0.113***	0.0649***	0.0564***	0.112***
	(0.0127)	(0.0134)	(0.0166)	(0.0126)	(0.0146)	(0.0170)
Robust	0.0632***	0.0572***	0.113***	0.0649***	0.0564***	0.112***
	(0.0150)	(0.0159)	(0.0188)	(0.0150)	(0.0172)	(0.0194)
Controls	No	No	No	Yes	Yes	Yes
Eff. obs. left of cutoff	8,731	7,510	4,487	8,274	6,171	4,061
Eff. obs. right of cutoff	9,186	7,855	4,686	8,670	6,398	4,232
BW estimate	61.37	61.30	44.03	62.61	54.76	44.08
BW bias	98.59	97.00	79.73	97.82	88.67	79.28

Table 5
Controlling for sample selection in the parametric RDD

The table reports coefficients and standard errors (in parentheses) from a two-stage Heckman model. The first stage models the probability that a loan application is submitted in a given year by individuals who have an exclusive relationship with the bank and apply multiple times during our sample period (probit model). The first stage is estimated on a dataset including all the information on loan applicants collected by the bank and spanning the time period 2002-2016. This is an unbalanced panel including all applicants, irrespective of whether they have an exclusive relationship with the bank or not and apply a single or multiple times. The second stage is equivalent to the estimation of equation (1) as in columns 4-6 of Panel A of Table 4, but including the fitted value of the *Mills ratio* (i.e., the instantaneous probability of loan application) obtained in the first stage. The dependent variable is given in the first row of the table and all variables are defined in Table 1. *t+1*, *t+3* and *t+5* stand for 1 year, 3 years, and 5 years after the loan application occurring at time *t*, respectively. The *, **, and *** marks denote statistical significance at the 10%, 5%, and 1% level, respectively.

	Second-stage results		
	(1)	(2)	(3)
	Income t+1	Income t+3	Income t+5
Granted	0.0533***	0.0761***	0.0795***
	(0.0179)	(0.0185)	(0.0188)
Credit score	-0.0021	-0.0011	-0.0051
	(0.0311)	(0.0350)	(0.0205)
Granted x Credit score	0.0184	0.0038	0.0087
	(0.0367)	(0.0401)	(0.0233)
Mills ratio	0.9150	0.9683	0.6129
	(1.3962)	(1.3121)	(0.8163)
Observations	53,585	45,333	37,210
Controls	Yes	Yes	Yes
Clustering	Individual	Individual	Individual
	First-stage results		
	Pr. application t	Pr. application t	Pr. application t
Income	0.0739***	0.0767***	0.0781***
	(0.0083)	(0.0083)	(0.0108)
Wealth	0.0580**	0.0625**	0.0642**
	(0.0270)	(0.0305)	(0.0316)
Education	0.0245***	0.0220***	0.0237**
	(0.0072)	(0.0079)	(0.0094)
Firm size	0.0014	0.0026*	0.0034**
	(0.0024)	(0.0015)	(0.0014)
Firm leverage	0.2870***	0.3022**	0.3147**
	(0.0331)	(0.0610)	(0.1103)
Gender	0.0081***	0.0081***	0.0074***
	(0.0023)	(0.0028)	(0.0031)
Observations	228,507	228,507	228,507
Clustering	Individual	Individual	Individual

Table 6
Hard and soft information

Panel A reports the results from the manipulation testing procedure using the local polynomial density estimator proposed by Cattaneo et al. (2018) performed on the subsample where the residuals of the linear regression of the credit score on a set of observables (income, wealth, education, firm size, leverage, loan amount, maturity, and two dummies reflecting the use of collateral and covenants) capturing hard information are positive and the subsample where the residuals are negative or zero. To perform this test, we rely on the local quadratic estimator with cubic bias-correction and triangular kernel. We report the conventional test statistic (Conventional) and the robust bias-corrected statistic (Robust) along with the corresponding p-value. The null hypothesis consists in no manipulation. Panel B replicates the analysis of column 6 of Panel B of Table 4 on different subsamples depending on the residuals of the linear regression of the credit score on a set of observables (income, wealth, education, firm size, leverage, loan amount, maturity, and two dummies reflecting the use of collateral and covenants) capturing hard information. The residuals of these regressions are interpreted as soft information held by the bank. Specification 1 is estimated on the subsample where the residuals are positive and specification 2 where the residual are negative or zero. The table reports coefficients and standard errors (in parentheses). The dependent variable is given in the first row of the table and all variables are defined in Table 1. *t+5* stands for 5 years after the loan application occurring at time *t*. The estimation method is the local linear regression with triangular kernel. For each specification, we report the bias-corrected RD estimates with robust variance estimator. All specifications include the control variables as in specifications 4-6 in Table 4. The *, **, and *** marks denote statistical significance at the 10%, 5%, and 1% level, respectively.

Panel A. Manipulation test

	Residuals>0		Residuals≤0	
	T-stat	P-value	T-stat	P-value
Conventional	0.3129	0.7543	1.2656	0.2057
Robust	0.2732	0.7847	0.4447	0.6566

Panel B. Heterogeneity analysis

	Residuals>0	Residuals≤0
	(1)	(2)
Dependent variable	Income t+5	Income t+5
Robust	0.135***	0.0695*
	(0.0293)	(0.0378)
Test difference in coefficients	0.000***	
Controls	Yes	Yes
Eff. obs. left of cutoff	2,549	2,373
Eff. obs. right of cutoff	2,720	2,556
BW estimate	47.11	41.28
BW bias	79.26	76.64

Table 7
Efficient use of loans

The table reports coefficients and standard errors (in parentheses). The dependent variable is given in the first row of the table and all variables are defined in Table 1. *t+1*, *t+3* and *t+5* stand for 1 year, 3 years, and 5 years after the loan application occurring at time *t*, respectively. The estimation method is the local linear regression with triangular kernel on the RDD model of equation (2). For each specification, we report the bias-corrected RD estimates with robust variance estimator. The table replicates columns (4) to (6) of Panel B of Table 4 by replacing the treated group with "easily accepted applicants" (i.e., accepted applicants whose credit score is far from the cutoff point) and the control group with "not easily accepted applicants" (i.e., accepted applicants whose credit score is close to the cutoff point), respectively. The former are accepted applicants within the bandwidth of the nonparametric models of columns (4) to (6) of panel B in Table 4; the latter are the remaining accepted loan applicants. The *, **, and *** marks denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1)	(2)	(3)
Dependent variable	Income t+1	Income t+3	Income t+5
Robust	0.025**	0.072***	0.116***
	(0.012)	(0.020)	(0.023)
Controls	Yes	Yes	Yes
Eff. obs. left of cutoff	7,329	5,508	3,649
Eff. obs. right of cutoff	7,816	5,905	3,852
BW estimate	67.11	63.13	58.42
BW bias	71.21	69.13	68.85

Table 8
Additional tests and firm performance

The table reports coefficients and standard errors (in parentheses). The dependent variable is given in the first row of the table and all variables are defined in Table 1. *t+5* stands for 5 years after the loan application occurring at time *t*. The estimation method is the local linear regression with triangular kernel on the RDD model of equation (2). For each specification, we report the bias-corrected RD estimates with robust variance estimator. Each specification replicates model 6 of Panel B of Table 4 with a different outcome variable or on a different subsample. The outcome variables of specifications 1-6 consist in the following items: i) the natural logarithm of the amount of credit used for corporate purposes (e.g., expansion projects, investments, working capital needs, inventory purchases, equipment acquisition, or other operational expenses) in column 1; ii) a dummy variable equal to one if the firm is repaying previous loan obligations and zero otherwise in column 2; iii) the growth rate of the firm in column 3; iv) the return on asset of the firm in column 4; v) the income growth of the applicant in column 5; vi) the ratio of total revenue to total leverage of the firm in column 6. Specification 7 and specification 8 are estimated on the subsamples of term loans and credit lines, respectively. Specification 9 and specification 10 are estimated on the subsamples of secured loans and unsecured loans, respectively. Specification 11 and specification 12 are estimated on the subsamples of limited liability firms and firms with unlimited liability, respectively. *Firm leverage* is excluded by the set of controls in specification 6. The *, **, and *** marks denote statistical significance at the 10%, 5%, and 1% level, respectively.

Dependent variable	(1) Corporate purpose t+5	(2) Debt repay t+5	(3) Firm growth t+5	(4) ROA t+5	(5) Income growth t+5	(6) Debt turnover t+5
Robust	0.131***	0.048**	0.035***	0.048***	0.014***	0.211**
	(0.019)	(0.022)	(0.0118)	(0.021)	(0.004)	(0.091)
Test difference in coefficients	0.000***					
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Eff. obs. left of cutoff	2,969	1,024	3,716	3,804	3,755	3,860
Eff. obs. right of cutoff	3,177	1,180	3,908	3,986	3,962	4,046
BW estimate	20.6	13.24	67.91	61.27	48.89	43.11
BW bias	22.46	15.72	107.18	95.16	88.61	76.19

Dependent variable	(7) Income t+5	(8) Income t+5	(9) Income t+5	(10) Income t+5	(11) Income t+5	(12) Income t+5
Robust	0.132***	0.098***	0.108***	0.103***	0.107***	0.099***
	(0.019)	(0.016)	(0.017)	(0.016)	(0.017)	(0.015)
Test difference in coefficients	0.000***		0.107		0.086*	
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Eff. obs. left of cutoff	1,470	1,720	2,917	1,258	3,376	643
Eff. obs. right of cutoff	1,588	1,842	3,062	1,343	3,490	699
BW estimate	34.19	36.20	35.22	31.80	44.01	33.11
BW bias	68.11	70.03	63.18	49.74	78.18	49.97

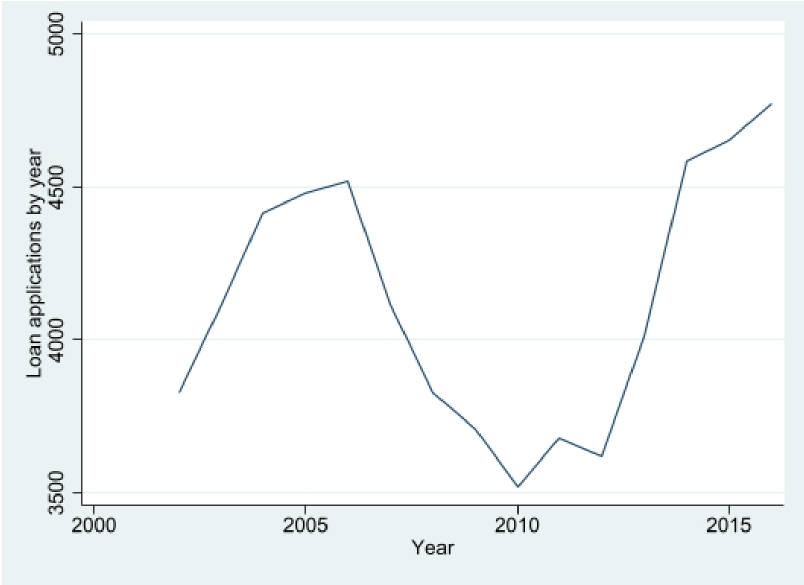
Table 9
Loan amount, loan spread, and path dependency in approvals

The table reports the coefficients and standard errors (in parentheses). The dependent variable is given in the first row of the table and all variables are defined in Table 1. *t+5* stands for 5 years after the loan application occurring at time *t*. Specifications 1-2 replicate model 6 of Panel B of Table 4 but estimated on the subsamples of small and large loans, which are identified as those below the 25th and above the 75th percentile of the distribution of *Loan amount*, respectively. The estimation method is the local linear regression with triangular kernel on the RDD model of equation (2). Specifications 3-6 report the parametric estimates of a modified version of model (6) of Panel A of Table 4 estimated on the subset of accepted applicants and by using i) Loan amount (column 3), ii) Loan spread (column 4), and iii) dummy variables equal to one if the firm obtains a 3rd, 4th, 5th or 6th loan, respectively, at time *t* (columns 5-6) as the main explanatory variables. All specifications include the control variables of specifications 4-6 in Table 4. Specifications 3-6 include *Credit score* as an additional control variable. Specifications 3-5 include the soft information component of the credit score, as calculated in Section 4.3, *Soft information*, as an additional control variable. The *, **, and *** marks denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	Income t+5	Income t+5	Income t+5	Income t+5	Income t+5	Income t+5
Robust	0.105***	0.118***				
	(0.017)	(0.022)				
Loan amount			0.502**			
			(0.217)			
Loan spread				-0.325***		
				(0.131)		
Loan 3					0.048**	0.057**
					(0.021)	(0.023)
Loan 4					0.053**	0.059***
					(0.022)	(0.022)
Loan 5					0.040**	0.045**
					(0.020)	(0.020)
Loan 6					0.045*	0.050**
					(0.024)	(0.025)
Soft information			0.082***	0.084***	0.079***	
			(0.007)	(0.007)	(0.007)	
Test difference in coefficients	0.039**					
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Eff. obs. left of cutoff	1,499	403				
Eff. obs. right of cutoff	2,022	416				
BW estimate	14.69	8.67				
BW bias	16.52	10.11				
Observations			37,210	37,210	37,210	37,210
Clustering			Individual	Individual	Individual	Individual

Figure 1
Statistics on loan applications

The first graph depicts the number of loan applications by year over 2002-2016. The second graph depicts the number of applicants for each number of loan applications (minimum 2, maximum 23).



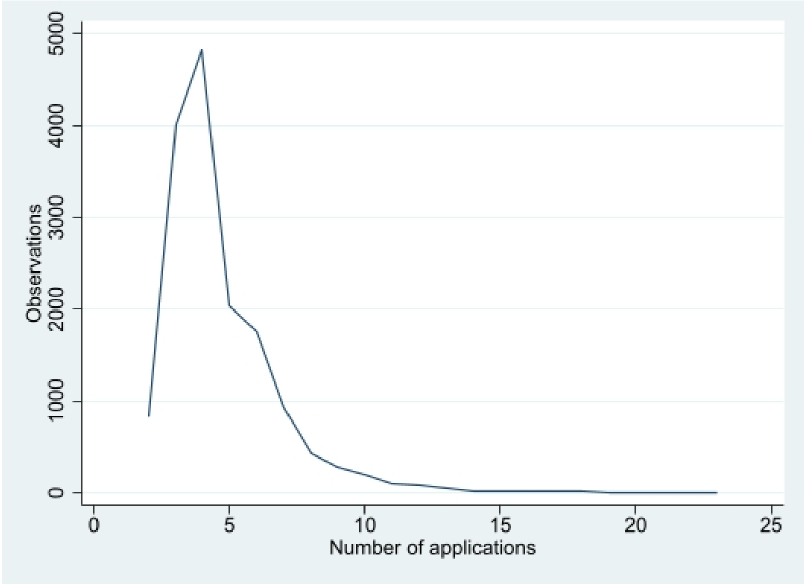


Figure 2
Applicants by industry

The figures report the share of loan applicants by industry.

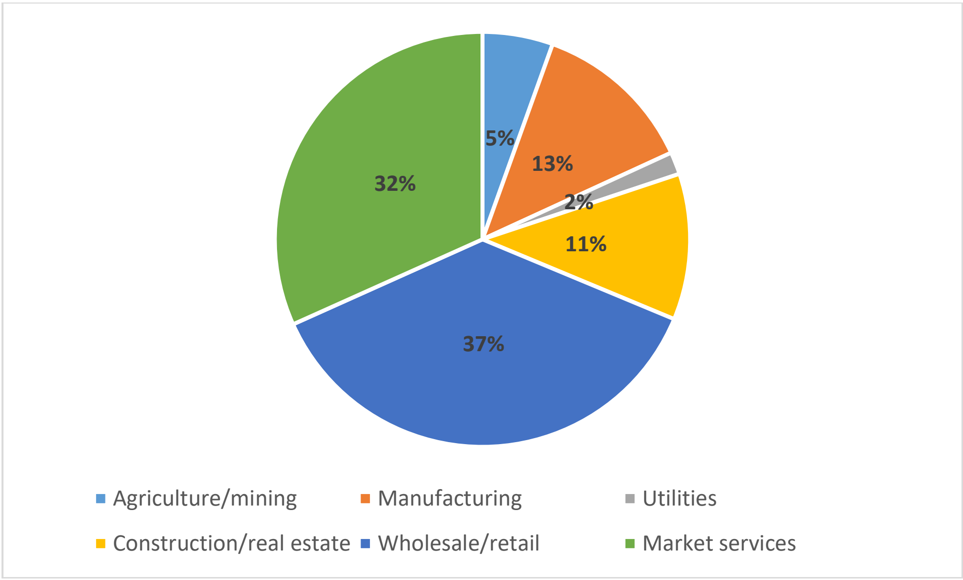
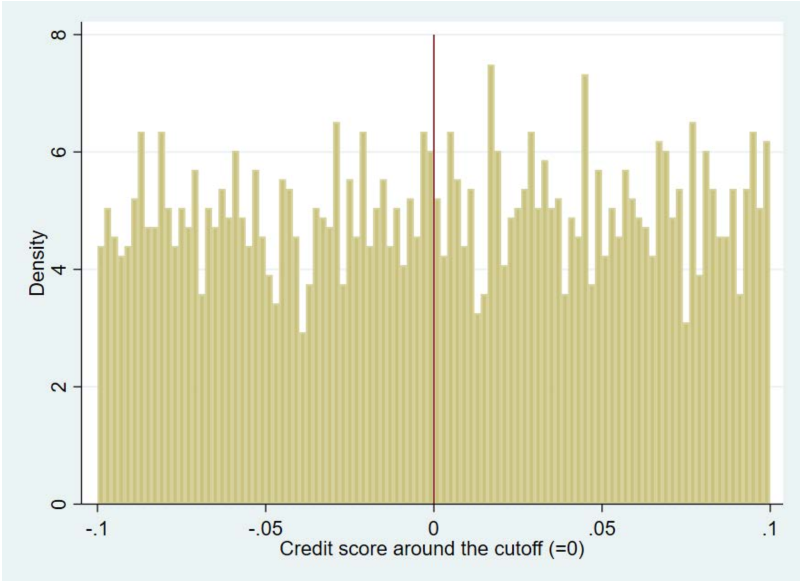


Figure 3
Densities of assignment and outcome variables

The figures report the probability densities for the assignment variable Credit score (top) and the outcome variable Income t+5 (bottom).



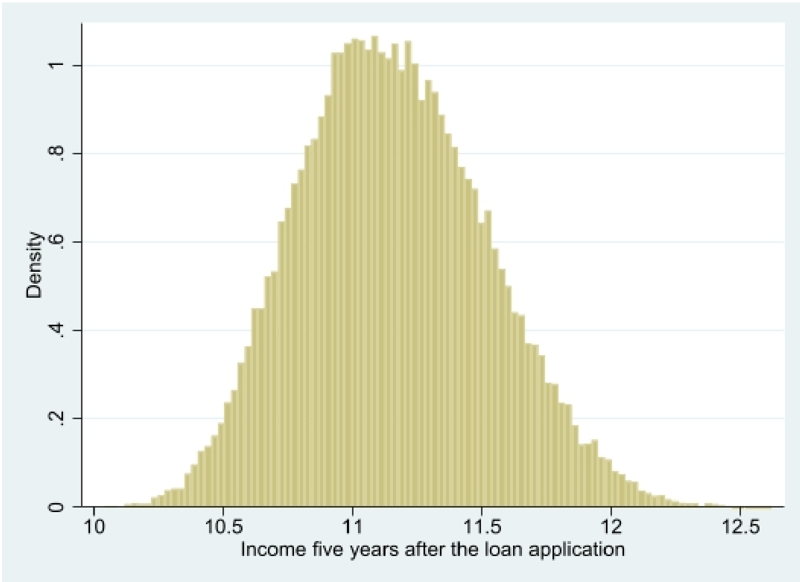


Figure 4
Manipulation test

The figure reports results from the manipulation testing procedure using the local polynomial density estimator proposed by Cattaneo et al. (2018). To perform this test, we rely on the local quadratic estimator with cubic bias-correction and triangular kernel.

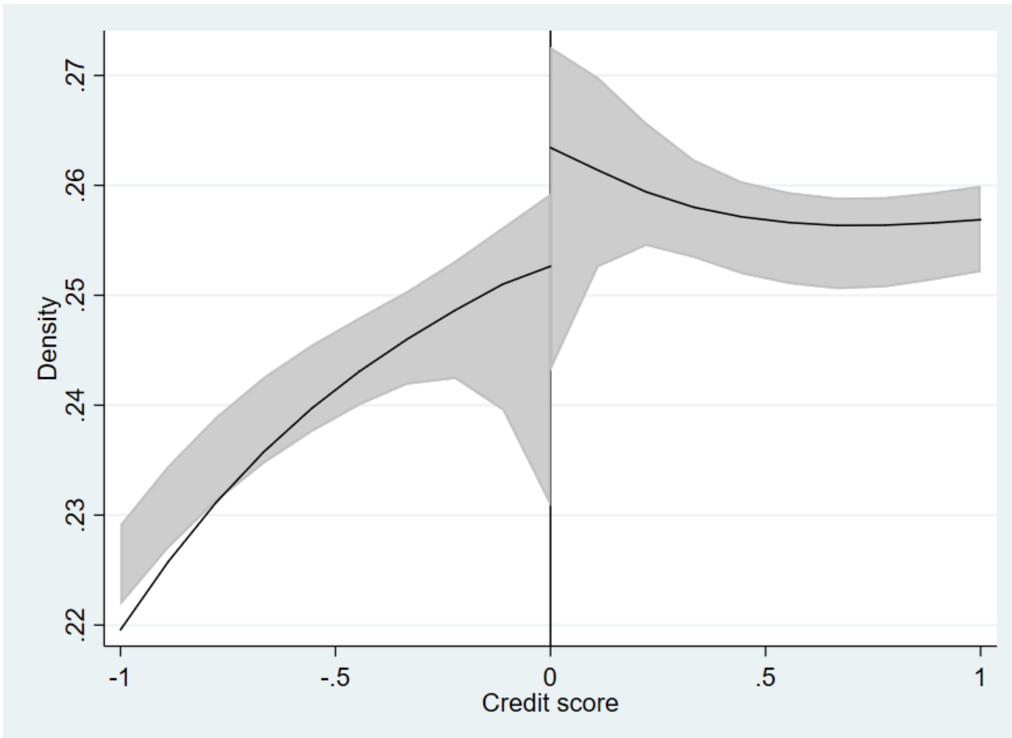
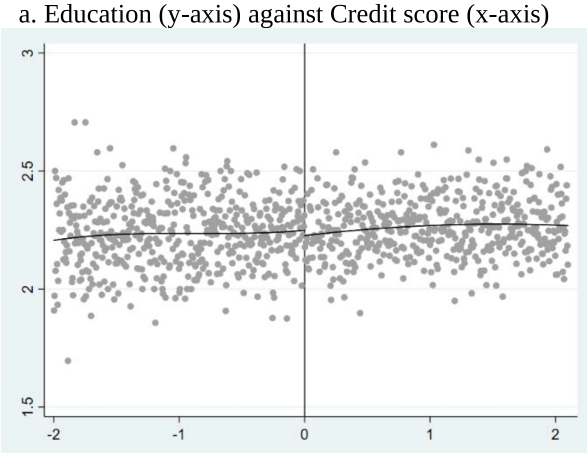


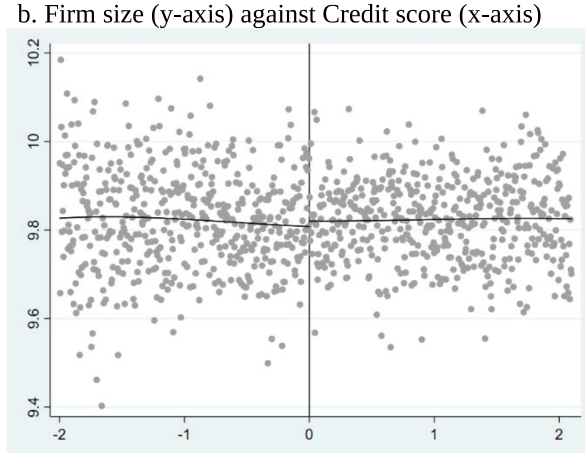
Figure 5
Covariates around the cutoff

The figure reports a plot for set of covariates against the Credit score. The covariates include Education, Firm size, Firm leverage, Loan amount, Maturity and Wealth (first instance of wealth before the loan application). The continuous line represents a fourth order polynomial fit used to approximate the conditional mean of each covariate below and above the cutoff.

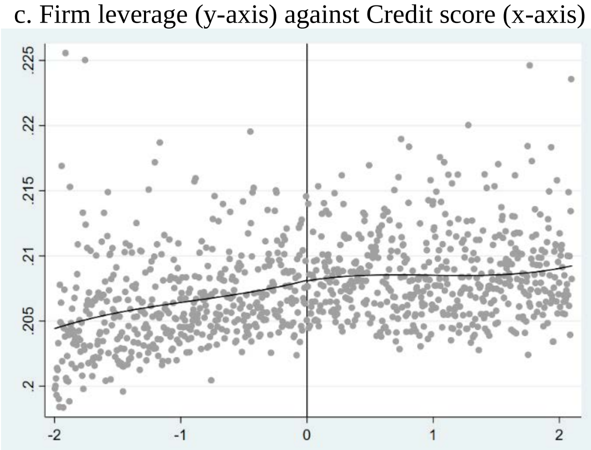
a. Education (y-axis) against Credit score (x-axis)



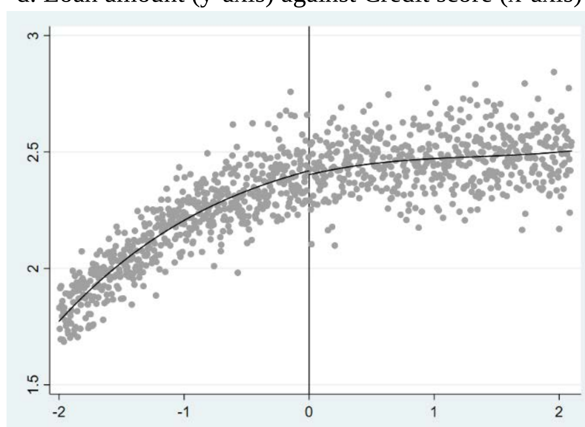
b. Firm size (y-axis) against Credit score (x-axis)



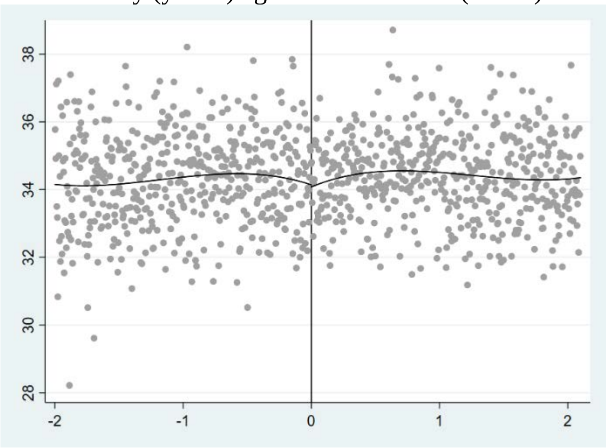
c. Firm leverage (y-axis) against Credit score (x-axis)



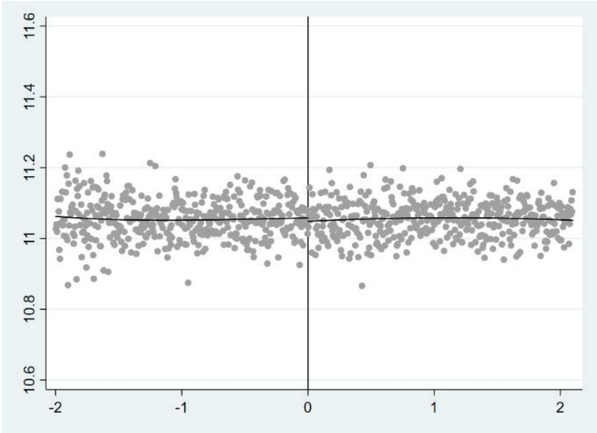
d. Loan amount (y-axis) against Credit score (x-axis)



e. Maturity (y-axis) against Credit score (x-axis)



f. Income t-1 (y-axis) against Credit score (x-axis)



g. Wealth t-5 (y-axis) against Credit score (x-axis)

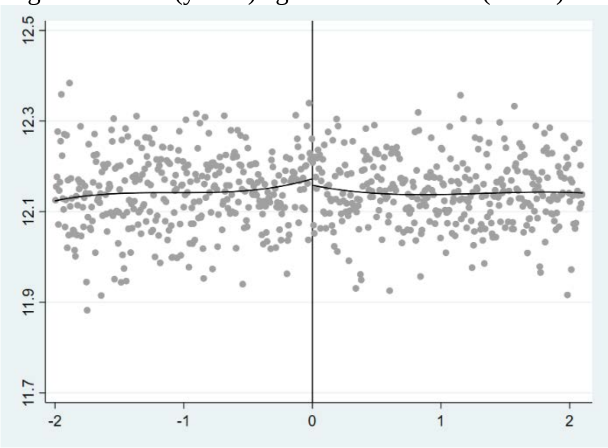


Figure 6
Applicants' income around the cutoff

The figure depicts applicants' Income five years after the loan decision (y-axis) against the Credit score (x-axis). The points represent local sample means of the applicant's income for a set of disjoint bins of control and treatment units spanning the full sample. We select evenly spaced bins that mimic the underlying variability of the data using spacings estimators. The continuous line represents a fourth order polynomial fit used to approximate the conditional mean of applicants' income below and above the cutoff.

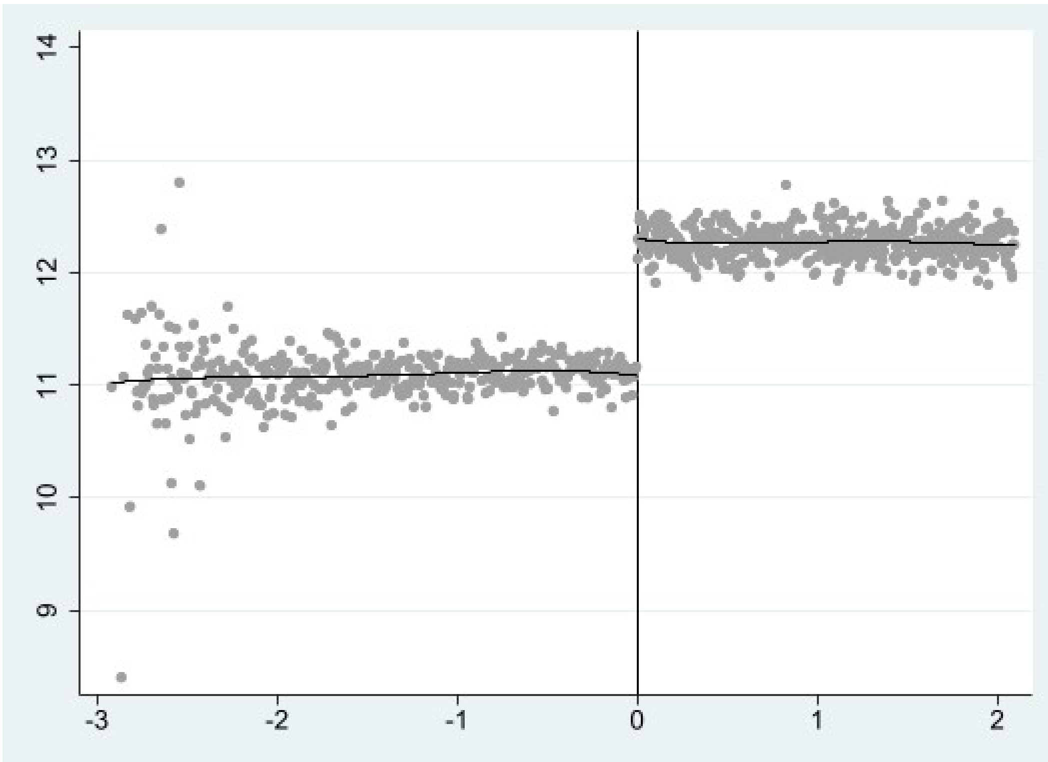
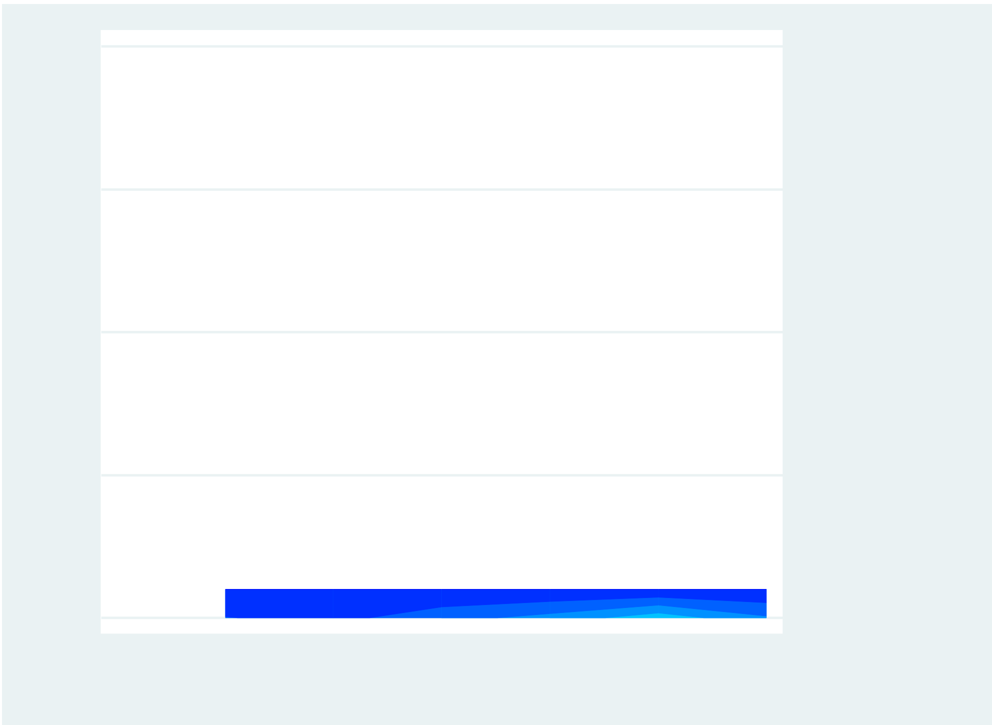


Figure 7
Sensitivity analysis for the RDD

The figure reports results from a sensitivity analysis under local randomization (see Cattaneo et al., 2016). We perform a sequence of hypotheses tests for different windows around the cutoff. Specifically, we show the test statistic of the null hypothesis of no treatment effect (x-axis) against the window length (y-axis). The p-values are calculated using randomization inference methods.



Online Appendix

This online appendix includes information on our sample's representativeness (Appendix A) and further robustness tests.

Appendix A. More on Sample Representativeness

We start by comparing annual averages of key attributes of 32 systematically important European banks (identified as per EBA's guidelines) with the corresponding characteristics of our bank. To this end, we collect the data on banks' balance sheets from Compustat. We focus on three metrics: the liquidity ratio (i.e., the ratio of cash plus short-term securities to total assets), the market-to-book ratio, and the (before tax) returns on assets (ROA). In Figures A1a to A1c, we show scatterplots and a linear fit of our bank's annual values (y-axes) against the corresponding averages for the set of systemic banks (x-axes). The coefficients of the three linear regressions are all positive and highly statistically significant, suggesting that liquidity, market value and profitability conditions of our bank are similar to the average counterparts of other European systemic banks.

[Insert Figure A1 about here]

We next use data from the Survey on Access to Finance of Enterprises (SAFE) to compare access to credit of small and micro firms operating in the euro area with that of firms in our sample.[13] Figure A2 shows the time series of the average rejection rate in the euro area along with the rejection rate in our sample of 61,863 applications during 2002-2016. The two series follow a similar path over time, with the rejection rate of our bank being somewhat higher than the euro area's average in 2010-2014 and slightly lower from 2015 onward.

[Insert Figure A2 about here]

As a last exercise, we present a comparative analysis of leverage and profitability of the 15,628 firms in our sample versus small and micro firms located in six representative European countries (i.e., Austria, Belgium, Denmark, France, Germany, and the Netherlands). We collect

[13] Both groups of firms comply with the requirements set by European Commission to define a firm as a small or micro business.

balance sheet data on small businesses operating in these countries from Bureau van Dijk Orbis. Figures A3a and A3b show that the average leverage ratio and profitability of the two groups are closely aligned during the whole sample period, although firms in our sample exhibit a slightly lower leverage and higher ROA.[14] Such small differences are probably explained by the fact that our sample country is characterized by a high per-capita income and was less affected from the economic downturn of 2010-2014 compared to other European countries. We conclude that small firms in our sample are comparable, across different dimensions, to small firms located in representative European countries.

[Insert Figure A3 about here]

[14] Additional plots comparing other firm characteristics are available upon request.

Figure A1
Our bank versus other systemic European banks

Figure A1a shows a scatter plot and a linear fit of the annual liquidity ratio of our bank against the annual average of liquidity ratios of 32 European systemic banks over the period 1985-2018. Figures A1b and A1c show similar scatter plots and regressions for the market-to-book value ratio and ROA. The coefficient estimates of all three lines are statistically significant at the 1% level and correlation coefficients are 0.34, 0.43, and 0.35, respectively.

Figure A1a

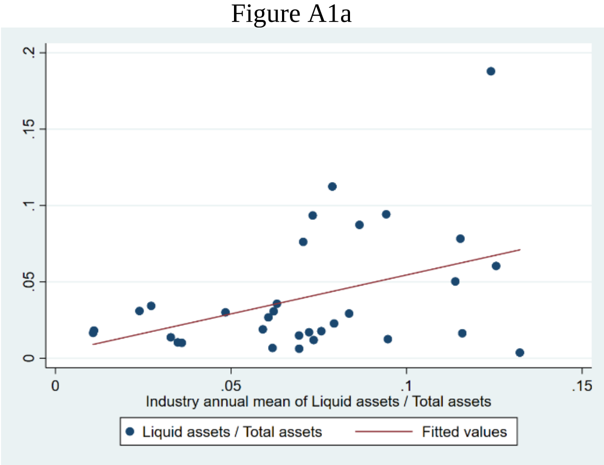


Figure A1b

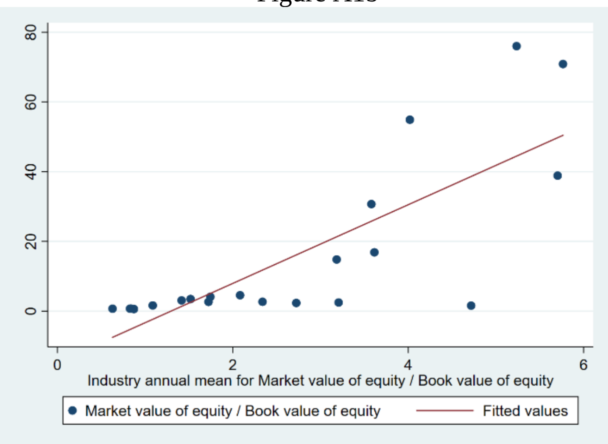


Figure A1c

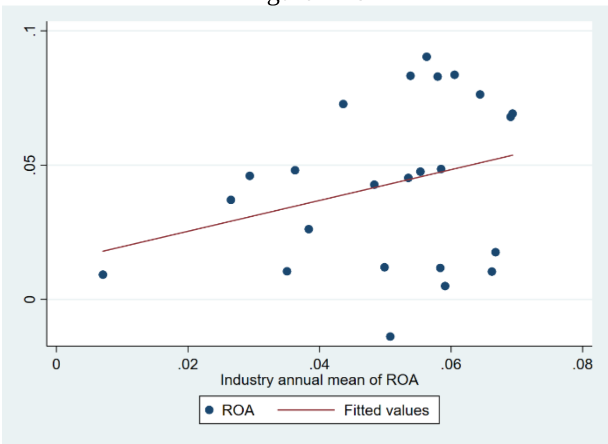


Figure A2

Percent of rejected loans to small and micro firms in the euro area and by our bank

The figure plots the annual average (in percent) of rejected loan applications to small and micro firms in the euro area, obtained from the (SAFE), and the rejection rate (in percent) for the 61,863 loan applications in our sample.

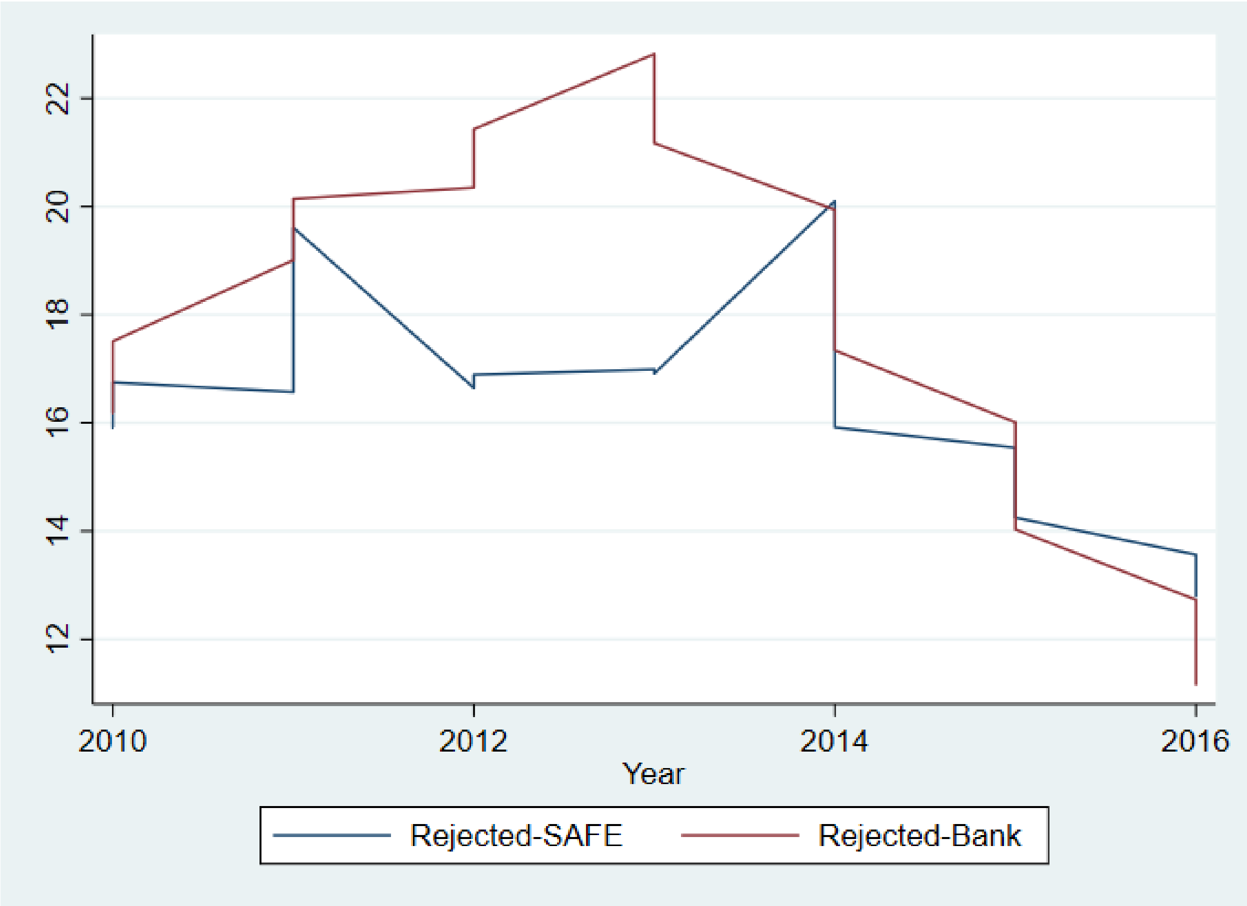


Figure A3

Leverage and ROA of North European small firms versus small firms in our sample

The figure plots the annual average of leverage (Figure A3a) and ROA (Figure A3b) of small and micro firms in Austria, Belgium, Denmark, France, Germany, and the Netherlands (blue lines) and the equivalent for the 15,628 firms in our sample (red lines).

Figure A3a

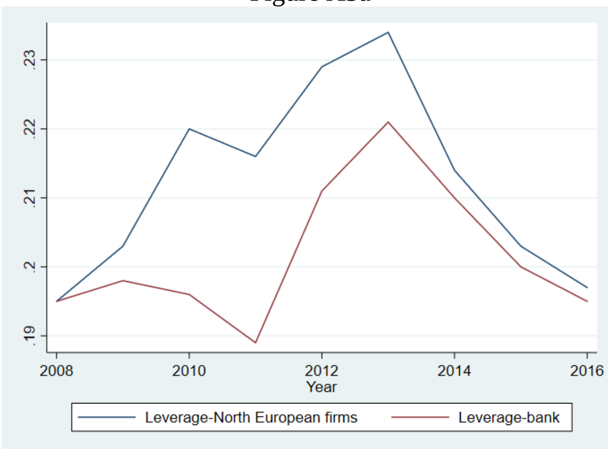
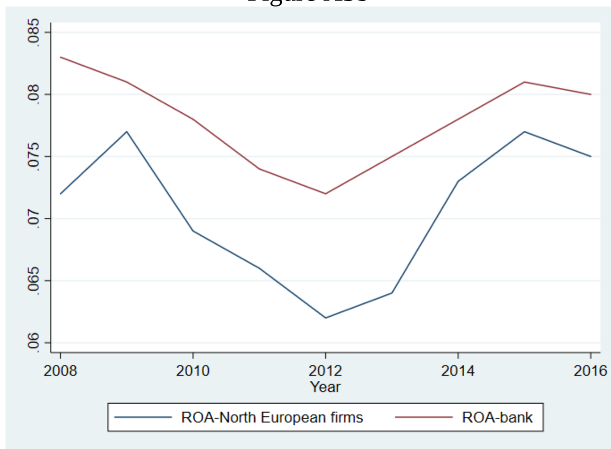


Figure A3b



Appendix B. Robustness of the RDD

In this appendix, we report the results of several robustness tests on the validity of our RDD, as described in Section 4.2, along with additional details on some of these tests.

The analysis presented in Table A8 relates to the different horizons considered in Table 4. The number of observations reported in Table 4 declines from column 1 to column 3 and from column 4 to column 6 due to a "truncation" affecting the right-hand side of the sample, i.e. business owners whose last loan application occurs less than five years prior to 2016. To ensure that the estimates across different horizons are fully comparable, Table A8 reports the estimates of the baseline nonparametric models of columns 4 and 5 of Panel B of Table 4 (those pertaining to the 1-year and 3-year horizon) run on the subsample of applicants considered in specification 6 of Panel B of Table 4. Results are virtually the same to those of Table 4.

In Table A9 we examine the role of initial wealth. In principle, wealthier individuals should be able to maintain higher incomes over time through higher investment. Accordingly, part of the macro inequality literature highlights the role of initial GDP per capita and suggests controlling for some sort of historical (or initial) wealth conditions when estimating models of inequality (e.g., Li et al., 1998). To this end, we use individual wealth in the first year before the loan application in which this information is available (Initial wealth; see Table 1). As with the rest of the control variables, we show in Figure 5 that *Initial wealth* is continuous around the cutoff. Of course, adding this variable to our covariates entails a substantial drop in the number of observations in the sample. This is the reason we leave this exercise as a robustness test. The nonparametric results in Table A9 show that including initial wealth does not yield significantly different results. If anything, the treatment effect is slightly stronger, with the only exception of the three-year horizon from the loan decision. We obtain similar patterns when using the parametric RDD (Table A2).

Table A10 deals with the history of the credit relationship between the applicant and the bank. As discussed in Section 3.1, applicants included in our balance panel apply multiple times throughout the sample period. Some applications may be approved, others may be denied. Thus, to account for credit obtained by the bank within the horizon considered after a loan application, we re-estimate models (4)-(6) of Panel B of Table 4 by including the total credit received by the firm from the bank in the period *t* to *t*+5 as a control variable. The estimates presented in Table A10 are virtually the same as those of table 4.

Table A11 focuses on bandwidth selection. Despite the advantage of focusing on observations close to the cutoff, the nonparametric approach does not necessarily represent the ideal functional form of the RDD. In light of that, Lee and Lemieux (2010) suggest relying on different bandwidth-selection methods to test if the results are stable across different specifications. Table A11 shows that the results presented in Panel B of Table 4 remain unchanged when using two different MSE-optimal bandwidth selectors below and above the cutoff, and one common coverage error (CER)-optimal bandwidth selector.

Table A1

Including industry, loan type, and year fixed effects in the parametric RDD

The table reports coefficients and standard errors (in parentheses). The dependent variable is given in the first row of the table and all variables are defined in Table 1. *t-1*, *t+1*, *t+3* and *t+5* stand for 1 year before, 1 year after, 3 years after, and 5 years after the loan application occurring at time *t*, respectively. Estimation method is OLS on the RDD model of equation (1). Specifications (1) to (3) do not include any covariate besides the treatment and assignment variables. More covariates are included in specifications (4) to (6). All specifications include industry, loan type, and year fixed effects. The *, **, and *** marks denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	Income t+1	Income t+3	Income t+5	Income t+1	Income t+3	Income t+5
Granted	0.0534***	0.0751***	0.0713***	0.0536***	0.0754***	0.0718***
	(0.0063)	(0.0066)	(0.0072)	(0.0063)	(0.0066)	(0.0072)
Credit score	-0.0051	0.0029	0.0089**	-0.0056	0.0027	0.0084*
	(0.0038)	(0.0040)	(0.0044)	(0.0039)	(0.0041)	(0.0044)
Granted x Credit score	0.0021	-0.0089	-0.0172***	0.0025	-0.0087	-0.0168***
	(0.0052)	(0.0055)	(0.0059)	(0.0053)	(0.0056)	(0.0060)
Income t-1				0.0975***	0.0657***	0.0447***
				(0.0053)	(0.0056)	(0.0058)
Education				0.0023	-0.0017	0.0004
				(0.0016)	(0.0017)	(0.0019)
Firm size				-0.0004	0.0030	-0.0015
				(0.0021)	(0.0022)	(0.0024)
Firm leverage				0.1872***	0.2877***	0.2435***
				(0.0672)	(0.0745)	(0.0778)
Loan amount				-0.0008	-0.0023	-0.0014
				(0.0020)	(0.0021)	(0.0023)
Maturity				0.0004**	0.0001	0.0002
				(0.0002)	(0.0002)	(0.0002)
Constant	0.0429***	0.0297***	0.0209***	-0.0020	-0.0004	0.0005
	(0.0029)	(0.0030)	(0.0032)	(0.0038)	(0.0039)	(0.0041)
Observations	53,585	45,333	37,210	53,585	45,333	37,210
Clustering	Individual	Individual	Individual	Individual	Individual	Individual

Table A2
Controlling for "initial" wealth in the parametric model

The table reports coefficients and standard errors (in parentheses). The dependent variable is given in the first row of the table and all variables are defined in Table 1. *t-1*, *t+1*, *t+3* and *t+5* stand for 1 year before, 1 year after, 3 years after, and 5 years after the loan application occurring at time *t*, respectively. The estimation method is OLS on the RDD model of equation (1). The table essentially replicates columns (3) to (6) of Panel A of Table 4, the difference being the inclusion of Wealth t-5 as a control variable. The *, **, and *** marks denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1)	(2)	(3)
Dependent variable	Income t+1	Income t+3	Income t+5
Granted	0.0514***	0.0726***	0.0814***
	(0.0072)	(0.0080)	(0.0094)
Credit score	-0.0071	-0.0023	0.0003
	(0.0044)	(0.0050)	(0.0059)
Granted x Credit score	0.0028	-0.0020	-0.0083
	(0.0060)	(0.0068)	(0.0079)
Income t-1	0.0816***	0.0600***	0.0450***
	(0.0051)	(0.0056)	(0.0064)
Education	0.0032*	-0.0027	0.0013
	(0.0018)	(0.0021)	(0.0024)
Firm size	-0.0001	0.0024	-0.0007
	(0.0024)	(0.0027)	(0.0031)
Firm leverage	0.1898**	0.1764**	0.2908***
	(0.0765)	(0.0850)	(0.1051)
Loan amount	0.0001	0.0014	0.0006
	(0.0023)	(0.0026)	(0.0030)
Maturity	0.0004*	-0.0000	0.0001
	(0.0002)	(0.0002)	(0.0003)
Wealth t-5	0.0215***	0.0148***	0.0046
	(0.0032)	(0.0035)	(0.0040)
Constant	9.9057***	10.2427***	10.5395***
	(0.0736)	(0.0803)	(0.0929)
Observations	36,856	28,604	20,481
Clustering	Individual	Individual	Individual

Table A3
Using a restricted subsample around the cutoff of the parametric model

The table reports coefficients and standard errors (in parentheses). The dependent variable is given in the first row of the table and all variables are defined in Table 1. *t+1*, *t+3* and *t+5* stand for 1 year, 3 years, and 5 years after the loan application occurring at time *t*, respectively. The estimation method is OLS on the RDD model of equation (1). The table essentially replicates columns (4) to (6) of Panel A of Table 4 by restricting the sample to the observations around the cutoff that are used in the nonparametric models of column (4)-(6) of Panel B of Table 4. The *, **, and *** marks denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1)	(2)	(3)
Dependent variable	Income t+1	Income t+3	Income t+5
Granted	0.0702***	0.0766***	0.123***
	(0.019)	(0.018)	(0.020)
Credit score	-0.0042	0.0019	0.0091*
	(0.0043)	(0.0048)	(0.0050)
Granted x Credit score	-0.0040	-0.0017	-0.0136**
	(0.0061)	(0.0065)	(0.067)
Controls	Yes	Yes	Yes
Observations	16,944	12,569	8,293
Clustering	Individual	Individual	Individual

Table A4
Firm debt before and after the loan application

The table reports summary statistics of firm leverage one year before the loan application, debt(t-1)/assets(t-1), and the ratio of total debt one year after the loan application to total assets in the year preceding the loan application, debt (t+1)/assets(t-1), for firms of accepted and rejected applicants belonging to the restricted sample of 17,917 "effective observations" around the cutoff where we estimate the nonparametric RDD model of specification 1 of Panel B of Table 4.

	Accepted		Rejected	
	debt(t-1)/assets(t-1)	debt(t+1)/assets(t-1)	debt(t-1)/assets(t-1)	debt(t+1)/assets(t-1)
min	0.130	0.143	0.149	0.147
25^{th} percentile	0.199	0.201	0.196	0.190
median	0.205	0.207	0.203	0.199
mean	0.208	0.209	0.208	0.205
75^{th} percentile	0.212	0.222	0.210	0.207
max	0.916	0.921	0.917	0.916

Table A5
Falsification tests on the RDD: Setting invalid cutoff points

The table reports coefficients and standard errors (in parentheses). The dependent variable is Income t+5 and all variables are defined in Table 1. *t+1*, *t+3* and *t+5* stand for 1 year, 3 years, and 5 years after the loan application occurring at time *t*, respectively. Each specification reports the estimate of the average treatment effect by replicating specification 6 of Panel B of Table 4 using -1.5, -1, -0.5, 0.5, 1, 1.5 as the cutoff values, respectively. The estimation method is the local linear regression with triangular kernel. For each specification, we report the bias-corrected RD estimates with robust variance estimator. BW estimate is the estimate of the bandwidth and BW bias is the associated bias. Following, Calonico et al. (2014), we use the mean squared error optimal bandwidth. The bias-corrected RD estimator and the robust variance estimator are obtained according to Calonico et al. (2018) and Calonico at el. (2014), respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
	Cutoff = -1.5	Cutoff = -1	Cutoff = -0.5	Cutoff = 0.5	Cutoff = 1	Cutoff = 1.5
Dependent variable	Income t+5	Income t+5	Income t+5	Income t+5	Income t+5	Income t+5
Robust	0.002	0.004	0.007	0.007	0.005	-0.000
	(0.019)	(0.020)	(0.020)	(0.019)	(0.019)	(0.022)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Eff. obs. left of cutoff	7,141	6,244	4,509	6,988	5,873	4,011
Eff. obs. right of cutoff	7,855	6,593	4,630	7,098	6,022	4,247
BW estimate	63.59	60.11	46.16	64.90	57.22	47.02
BW bias	79.22	78.72	80.90	80.82	78.67	79.16

Table A6
Equality of means of variables in the full sample and the used sample

The table compares the means of observables between the 35,796 loan applications that we do not use (one-time applicants, lack of information on forward income) and the 61,863 loan applications used in our sample.

	Discarded sample	Used sample	Equality test (p-value)
Equality of means			
Credit score	0.105	0.103	0.009
Income	10.99	11.01	0.000
Wealth	12.12	12.14	0.000
Education	2.897	2.975	0.091
Gender	0.801	0.802	0.002
Marital status	0.580	0.589	0.040
Dependents	1.890	1.895	0.021
Firm size	12.826	12.821	0.002
Firm leverage	0.206	0.207	0.000
Firm ROA	0.096	0.094	0.032
Firm age	14.227	14.203	0.042
Observations	35,796	61,863	

Table A7
Controlling for sample selection in the nonparametric RDD

The table reports coefficients and standard errors (in parentheses) from a quasi-two-stage Heckman model. The table essentially replicates the analysis of columns 4-6 of Panel B of Table 4, the difference being the inclusion of the *Mills Ratio* obtained in the first stage regressions of Table 5 as a control variable in the nonparametric RDD estimation. The dependent variable is given in the first row of the table and all variables are defined in Table 1. *t+1*, *t+3* and *t+5* stand for 1 year, 3 years, and 5 years after the loan application occurring at time *t*, respectively. The estimation method is the local linear regression with triangular kernel. For each specification, we report the bias-corrected RD estimates with robust variance estimator. The *, **, and *** marks denote statistical significance at the 10%, 5%, and 1% level, respectively.

	Second-stage results		
	(1)	(2)	(3)
Dependent variable	Income t+1	Income t+3	Income t+5
Robust	0.0601***	0.0613***	0.106***
	(0.014)	(0.0163)	(0.0182)
Controls	Yes	Yes	Yes
Eff. obs. left of cutoff	8,203	6,049	4,080
Eff. obs. right of cutoff	8,480	6,261	4,197
BW estimate	62.4	56.13	45.09
BW bias	96.25	87.24	79.11

Table A8

Holding the number of observations constant across the different time horizons

The table reports coefficients and standard errors (in parentheses). The dependent variable is given in the first row of the table and all variables are defined in Table 1. *t+1*, *t+3* and *t+5* stand for 1 year, 3 years, and 5 years after the loan application occurring at time *t*, respectively. The estimation method is the local linear regression with triangular kernel. For each specification, we report the bias-corrected RD estimates with robust variance estimator. The table replicates columns (4) to (6) of Panel B of Table 4 run on the subsample of applicants considered in specification 6 of Panel B of Table 4. The *, **, and *** marks denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1)	(2)	(3)
Dependent variable	Income t+1	Income t+3	Income t+5
Robust	0.0614***	0.0625***	0.105***
	(0.013)	(0.015)	(0.017)
Controls	Yes	Yes	Yes
Eff. obs. left of cutoff	4,061	4,061	4,061
Eff. obs. right of cutoff	4,232	4,232	4,232

Table A9

Controlling for “initial” wealth: Nonparametric model

The table reports coefficients and standard errors (in parentheses). The dependent variable is given in the first row of the table and all variables are defined in Table 1. *t+1*, *t+3* and *t+5* stand for 1 year, 3 years, and 5 years after the loan application occurring at time *t*, respectively. The table essentially replicates columns (4) to (6) of Panel B of Table 4, the difference being the inclusion of Wealth t-5 as a control variable. The estimation method is the local linear regression with triangular kernel. For each specification, we report the conventional RD estimates with conventional variance estimator, the bias-corrected RD estimates with conventional variance estimator, and the bias-corrected RD estimates with robust variance estimator. The *, **, and *** marks denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1)	(2)	(3)
Dependent variable	Income t+1	Income t+3	Income t+5
Conventional	0.0646***	0.0491***	0.112***
	(0.0148)	(0.0171)	(0.0227)
Bias-corrected	0.0681***	0.0450***	0.121***
	(0.0148)	(0.0171)	(0.0227)
Robust	0.0681***	0.0450**	0.121***
	(0.0175)	(0.0202)	(0.0260)
Controls	Yes	Yes	Yes
Eff. obs. left of cutoff	5,312	4,238	2,207
Eff. obs. right of cutoff	5,572	4,386	2,295
BW estimate	57.92	58.91	42.43
BW bias	91.65	94.75	74.35

Table A10
Control for total credit received by the bank

The table reports coefficients and standard errors (in parentheses). The dependent variable is given in the first row of the table and all variables are defined in Table 1. *t+1*, *t+3* and *t+5* stand for 1 year, 3 years, and 5 years after the loan application occurring at time *t*, respectively. The table essentially replicates columns (4) to (6) of Panel B of Table 4, the difference being the inclusion of total credit received by the firm from the bank in the period *t* to *t*+5 and include it as a control variable. The estimation method is the local linear regression with triangular kernel on the RDD model of equation (2) For each specification, we report the bias-corrected RD estimates with robust variance estimator. The *, **, and *** marks denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1)	(2)	(3)
Dependent variable	Income t+1	Income t+3	Income t+5
Robust	0.0649***	0.0570***	0.114***
	(0.0152)	(0.0171)	(0.0192)
Controls	Yes	Yes	Yes
Eff. obs. left of cutoff	8,284	6,180	4,070
Eff. obs. right of cutoff	8,681	6,409	4,240
BW estimate	62.73	54.89	44.21
BW bias	97.89	88.76	79.40

Table A11
Alternative bandwidth selection methods

The table reports coefficients and standard errors (in parentheses). The dependent variable is given in the first row of the table and all variables are defined in Table 1. *t+1*, *t+3* and *t+5* stand for 1 year, 3 years, and 5 years after the loan application occurring at time *t*, respectively. The estimation method is the local linear regression with triangular kernel. For each specification, we report the conventional RD estimates with conventional variance estimator. The specifications replicate models 1-3 of Panel B of Table 4 using different bandwidth selection methods. Each specification does not include any covariate besides the assignment variable (credit score). Specifications 1, 3, and 5 use the two mean squared error (MSE)-optimal bandwidth selectors (below and above the cutoff) for the RD treatment effect. Specifications 2, 4, and 6 use one common coverage error (CER)-optimal bandwidth selector for the RD treatment effect. The *, **, and *** marks denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	Income t+1	Income t+1	Income t+3	Income t+3	Income t+5	Income t+5
Robust	0.0611***	0.0716***	0.0610***	0.0645***	0.103***	0.0956***
	(0.0127)	(0.0167)	(0.0131)	(0.0178)	(0.0159)	(0.0215)
Eff. obs. left of cutoff	7,743	5,053	8,260	4,373	5,180	2,599
Eff. obs. right of cutoff	10,530	5,284	7,802	4,536	4,831	2,738

Table A12
Co-financing of projects and future credit score

The table reports the bias-corrected coefficients and standard errors (in parentheses). The dependent variable is given in the first row of the table and all variables are defined in Table 1. *t+5* stands for 5 years after the loan application occurring at time *t*. The estimation method is the local linear regression with triangular kernel on the RDD model of equation (2). Each of the first three specifications replicate the last specification of Panel B of Table 4, but with additional controls. The first specification controls for *Co-financing*$_t$, the second for *Credit score*$_{t+1}$ and the third for *Credit score*$_{t+3}$. The *, **, and *** marks denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1)	(2)	(3)
Dependent variable	Income t+5	Income t+5	Income t+5
Robust	0.107***	0.101***	0.113***
	(0.017)	(0.017)	(0.020)
Controls	Yes	Yes	Yes
Eff. obs. left of cutoff	3,711	3,750	3,819
Eff. obs. right of cutoff	3,850	3,883	3,956
BW estimate	44.29	44.81	44.85
BW bias	79.78	79.15	79.60

Table A13
Macroeconomic conditions

The table reports coefficients and standard errors (in parentheses). The dependent variable is given in the first row of the table and all variables are defined in Table 1. *t+5* stands for 5 years after the loan application occurring at time *t*. Estimation method is the local linear regression with triangular kernel on the RDD model of equation (2). For each specification, we report the bias-corrected RD estimates with robust variance estimator. All specifications include the control variables as in specifications 4-6 of Table 4. The first two specifications distinguish between firms located in low income and high income regions based on the median regional income. Specifications 3-5 distinguish between loan applications submitted during the pre-global financial crisis period (2002-2007), the double-dip recession period (2008-2013), and the post-crisis recovery period (2014-2016). The *, **, and *** marks denote statistical significance at the 10%, 5%, and 1% level, respectively.

	Low income	High income
	(1)	(2)
Dependent variable	Income t+5	Income t+5
Robust	0.1203***	0.0926***
	(0.0380)	(0.0263)
Test difference in coefficients	0.000***	
Controls	Yes	Yes
Eff. obs. left of cutoff	2,311	2,290
Eff. obs. right of cutoff	2,384	2,297
BW estimate	43.28	41.18
BW bias	75.61	72.16

	2002-2007	2008-2013	2014-2016
	(3)	(4)	(5)
Dependent variable	Income t+5	Income t+5	Income t+5
Robust	0.103***	0.093***	0.116***
	(0.029)	(0.015)	(0.017)
Test difference in coefficients	0.001***		0.000***
Controls	Yes	Yes	Yes
Eff. obs. left of cutoff	1,445	1,310	832
Eff. obs. right of cutoff	1,619	1,495	984
BW estimate	16.60	15.36	11.89
BW bias	22.47	20.18	15.11