STAFF REPORTS

NO. 929
JUNE 2020

REVISED
SEPTEMBER 2023

Credit and Income Inequality

Manthos Delis | Fulvia Fringuellotti | Steven Ongena

FEDERAL RESERVE BANK of NEW YORK

Credit and Income Inequality
Manthos Delis, Fulvia Fringuellotti, and Steven Ongena
Federal Reserve Bank of New York Staff Reports, no. 929
June 2020; revised September 2023
JEL classification: G21, D31, E24, O15

Abstract

How does credit access for small business owners affect income inequality? A bank's cutoff rule, employed in the decision to grant loans and based on applicants' credit scores, provides us with the exogenous variation needed to answer this question. Analyzing uniquely detailed loan application data, we find that application acceptance increases recipients' income five years later by more than 10 percent compared to denied applicants. This effect is mostly driven by upward mobility of poor individuals, especially if credit-constrained, thereby reducing income inequality among those who get credit. Looking across various salient groups of applicants, we find that relatively constrained groups—that is, firms from low-income regions, new or high-growth firms, or female-owned firms—display higher responses to credit origination for their relatively poor applicants, while the effects for the rich applicants are at best negligible.

Key words: credit constraints, income inequality, business loans, economic mobility, regression discontinuity design

---

Fringuellotti: Federal Reserve Bank of New York (email: fulvia.fringuellotti@ny.frb.org). Delis: Audencia Business School (email: mdelis@audencia.com). Ongena: Department of Banking and Finance, University of Zurich (email: steven.ongena@bf.uzh.ch). This paper was previously circulated under the titles "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. We 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, we thank Knut Are Aastveit, Christoph Basten, Tobias Berg, Rajashri Chakrabarti, Nicola Cetorelli, Jerry Coakley, 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, Enrico Sette, László Tétényi, Wilbert van der Klaauw, and Diego A. Vera-Cossio.

1. Introduction

Over past decades, the gap between the rich and the poor has risen in most OECD countries (OECD, 2015), posing serious concerns for economic growth and social cohesion (Galor and Zeira, 1993; Alesina and Rodrick, 1994; Galor and Moav, 2004; Persson and Tabellini, 1994 Putnam, 2000; Stiglitz, 2012; Larsen, 2013; Piketty and Saez, 2013). The increase in income inequality has been associated with an increase in intergenerational social immobility in many countries, creating an upward sloping schedule commonly referred as "the Great Gatsby curve" (Corak, 2013; Kearney and Levine, 2016; Chetty et al., 2017). A lively debate ensued on the sources of this phenomenon and the proper measures to contain the problem. The role of finance is at the forefront in shaping economic opportunities of households and businesses.

This study aims to identify and quantify how banks' credit decisions (credit acceptance or rejection) affect applicants' income inequality. We focus on the micro perspective by studying the extent to which applicants that are similar in terms of income and other traits when applying for credit experience significantly different incomes after the credit decision. We show the important implications (real effect of credit) on upward mobility and income inequality.

Theoretically, asymmetric information between lenders and borrowers affects credit availability. Because the enforcement of loan contracts is imperfect, lenders often require borrowers to pledge collateral. Lenders also ration credit based on an expected probability of repayment. In general, when a credit expansion accompanies a relaxation of credit constraints, it leads to more financing opportunities for the full spectrum of potential borrowers (including the poor) and a possible tightening of the income distribution (Banerjee and Newman, 1993; Galor and Zeira, 1993).

However, credit-constrained individuals often have limited wealth, and their exclusion from credit can hinder economic mobility and fuel persistent income inequality. Specifically, financial frictions in the form of informational asymmetry imply an important role for wealth (or capital) endowment in liquidity creation. The endowment represents a fixed cost for credit access. The relatively poor cannot always overcome it, irrespective of the quality of their investment ideas, due to adverse selection and moral hazard in the loan origination process. Thus, returns on capital can lead to high persistence in income growth only for those with substantial wealth (Piketty, 1997; Mookherjee and Ray, 2003; Demirgüç–Kunt and Levine, 2009). Further, returns on investment usually increase with the amount of capital wealthier individuals employ, initiating a second-order effect due to economies of scale in larger projects (e.g., Evans and Jovanovic, 1989; Greenwood and Jovanovic, 1990).

A simple plot between GDP per capita (or the Gini coefficient) and the ratio of private credit to GDP for 150 countries over 1960-2015, shows that income (income inequality) is strongly and positively (negatively) correlated to private credit from banks and other financial institutions over GDP (Figure 1). Of course, this relation cannot be interpreted as causal. It is confounded by reverse causality, meaning that income (income inequality) may actually drive the availability of credit (Kumhof and Rancière, 2010; Rajan, 2010). Omitted-variable bias is an additional concern due to unobserved factors that are difficult to measure (e.g., the availability of new investment ideas), which jointly affect the distribution of income and the degree of financial depth.

[Insert Figure 1 here]

Our study provides the first empirical analysis of how access to credit affects individuals' income distribution in a developed economy, by comparing the future incomes of accepted applicants to those of rejected applicants with approximately the same characteristics (e.g., income and credit quality). We identify this effect using a unique data set of business loan applications to a single large European bank. This is a systemic bank directly supervised by the ECB under the Single Supervisory Mechanism and headquartered in a highly developed northern European country.

Our focus is on loan applications from small and micro enterprises that are majority-owned by individuals. This focus yields two major advantages for investigating our research question. First, the income of such entrepreneurs is highly correlated to the performance of their business. Second, for these applicants, 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 applicants having an exclusive relationship with the bank, meaning that they do not have a lending relationship with another regulated bank even if their loan application(s) at our bank is (are) rejected. This is necessary to ensure that we estimate the effect of credit on income, avoiding potential confounding factors such as other sources of funding beyond this bank.

The uniqueness of our data lies in the available information on the majority owners, which encompasses income, wealth, and the credit scores assigned by the bank, as well as other applicant and firm characteristics.[1] Importantly, the exclusivity of the relationship between the bank and the applicant means that most applicants (accepted and rejected) reapply for loans. This in turn means that the bank maintains information on applicants' income after the original credit decision. We exploit this feature of the data to build a sample of 61,863 loan applications submitted by 15,628

[1] In this regard the bank information we have access to comprises the set in, e.g., Artavanis et al. (2016).

individuals who apply multiple times over the period 2002-2016. For these applicants, the future income with respect to a given loan application corresponds to the historical information on income collected by the bank in the subsequent applications.

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 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).

We show that, on average, a loan origination increases the recipient's income five years onward by 11% compared to denied applicants, regardless of whether we control for application probability. The economic interpretation of this finding is that marginally accepted applicants benefit from an approximately 11% increase in their incomes compared to marginally rejected applicants, thereby significantly affecting the distribution of income in the two groups. This finding is robust to several re-specifications and is not affected by the mix of the control variables. Further, the RDD passes the tests for credit score manipulation, and the control variables are continuous around the cutoff. Overall, our result suggests that bank credit decisions (loan origination or denial) affect individuals' income in a significant way improving upward mobility.

In a series of extensions to our baseline model, we examine the mechanisms behind the effect of credit access on individual income. First, we show that a loan origination entails a stronger increase in the income of applicants owning young firms compared to business owners of old firms. Second, firms of accepted applicants invest more in business operations, are more likely to repay existing bank loans, experience a higher increase in profitability, and grow at a higher rate compared to firms of rejected applicants. Overall, these results reveal that access to credit is pivotal for small firms to exploit good investment opportunities, expand their business, and improve profitability.

We next focus on how the bank's credit decisions affect income inequality within and between the groups of accepted and rejected applicants, hereby providing a micro-foundation to the impact of a credit expansion on income inequality. As a first step, we calculate inequality measures (Gini coefficients and Theil indices) for the loan applicants around the cutoff. We show that the Gini and Theil indices decrease (tighter income distribution) for accepted applicants and increase (wider income distribution) for rejected applicants. These findings are consistent with the theory of a negative nexus between finance and inequality when access to credit is improved (Greenwood and Jovanovic, 1990).

Importantly, we examine the mechanisms behind the observed decrease and increase in income inequality within accepted and rejected applicants, respectively. To this end, we estimate the full distribution of applicants' income responses to credit access by relying on a simultaneous quantile regression approach. We document that the income responses are significantly stronger for poor individuals on the left tail of the income distribution vis-à-vis rich individuals on the right tail. Specifically, a loan origination implies an increase in the income of accepted applicants by 17% compared to denied applicants for the 1st percentile of the income distribution. The magnitude of the effect drops to 11% and 8% for the 10th percentile and the median, respectively. For top income earners, those above the 90th percentile, the effect of access to credit on income becomes negligible (3% increase) and statistically insignificant. We complement this exercise with a probit model examining the likelihood that an applicant moves upward in the income distribution after being granted a loan based on the level of income prior to the credit decision. We show that relatively poor individuals have a much higher probability to experience an upward shift by more than a decile in the income distribution compared to rich individuals. Overall, our findings suggest that access to credit reduces income inequality among accepted applicants by fostering upward mobility of low-income individuals.

We, next, explore potential heterogeneities in the effect of credit origination on income inequality looking at the macroeconomic environment, the geographical location, the age and industry of the firm, and the gender of the business owner. We document that, for relatively poor business owners, a loan origination has a stronger effect on their income in low-income regions (versus high-income regions), during the Great Recession (compared to the pre-crisis period), if the firm is young (vis-à-vis an old firm), if the firm operates in a high-growth industry (versus a low-growth industry), and if the entrepreneur is a female (compared to a male). Broadly speaking, this reveals that credit originations lowers income inequality among the recipients of credit, especially whenever firm owners are credit constrained, have business growth opportunities, or are women.

Lastly, we look at how the effect of credit access on income may vary depending on the private information held by the bank on the loan applicant. Specifically, we decompose the credit score into two components attributable to hard information (e.g., observable characteristics of the applicant) and soft information (e.g., information on the quality of the investment opportunities of the firm), respectively. Then, we estimate the effect of a loan origination on income for the subset of accepted and rejected applicants whose credit score is positively affected by soft information, and the subset of accepted and rejected applicants whose credit score is negatively affected by soft information. We show that the positive impact of credit 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 how efficiently the bank extends credit.

The next section provides a brief review of the literature. Section 3 describes the data set and empirical identification, emphasizing the particular RDD. Section 4 presents the empirical results. Section 5 concludes the paper.

2. Literature

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 (Angelucci et al., 2015; Attanasio et al., 2015; Augsburg et al.; 2015; Banerjee et al., 2015; Banerjee et al., 2015b; Crépon et al., 2015; Tarozzi et al., 2015; Banerjee et al., 2018; Banerjee et al, 2019). These studies show that various microcredit programs did not have a significant impact on individual income in developing countries, except for the subset of well-established entrepreneurs, presumably because recipients on the extensive margin lack good investment projects. 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 and that the impact is stronger for entrepreneurs who are more credit constrained. Importantly, we document that bank credit has a negative impact on income inequality by fostering upward mobility of relatively poor individuals.

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; Chetty and Hendren, 2018a, 2018b; Bell et al., 2019, Bergman et al., 2019; Chetty et al., forthcoming). We contribute to this literature documenting that credit provision to small businesses is pivotal in fostering entrepreneurship and upward mobility.

Our work also relates to the literature that looks broadly at how credit expansions and/or constraints affect income distribution by relying on aggregate (at the country or regional level) measures of inequality (mostly the Gini index) and financial development. This body of literature provides mixed results. Clarke et al. (2006), Beck et al. (2010), Kappel (2010), Hamori and Hashiguchi (2012), Delis et al. (2014), and Naceur and Zhang (2016), for example, document a negative relation between financial development and income inequality, consistent with the idea that a credit expansion corresponds to a relaxing of credit constraints. Denk and Cournède (2015), Jauch and Watzka (2016), and de Haan and Sturm (2017), point instead to a positive relation, suggesting that financial development improves access to credit only for the rich. Kim and Lin (2011), and Brei et al. (2018) identify a non-monotonic relation depending on the degree of financial development and the financial structure of the economy. Minetti et al. (2019) show that local banking structures affect regional income inequality. Our paper also relates to several other studies on finance and income inequality (for a thorough review, see Demirgüç-Kunt and Levine, 2009).[2] We contribute to this literature by proposing a rigorous identification setup to study the effect of credit origination on income inequality at the individual, micro level. We show that access to credit reduces income inequality by lowering the income gap between poor and rich entrepreneurs, and that this phenomenon is more pronounced if business owners are credit constrained, have growth opportunities, or belong to a disadvantaged group.

[2] Our paper also relates to Saez et al. (2012) and Moser et al. (2018), who look at the effect of payroll taxation and credit supply, respectively, on inequality in wages.

Another strand of related recent literature examines how credit constraints affect economic and social outcomes. Looking at the Home Owners Loan Corporation "redlining" maps drawn in the 1930s, Appel and Nickerson (2016) and Aaronson et al. (2019) show that reduced access to credit in certain city neighborhood has negative long-lasting effects on home ownership, house prices, and rents, while increasing racial segregation. Using data on loan applications (such as ours), Berg (2018) documents that credit denial has stronger negative real effects on low-liquidity firms, which need to increase cash holdings and dispose of other assets in response to a loan rejection. In a similar framework, Fracassi et al. (2016) show that access to credit is pivotal for the survival and expansion of startups. Berton et al. (2018) document that a credit crunch has a negative impact on employment. Acabbi et al. (2020) show that a negative credit supply shock reduces employment and productivity, while increasing labor misallocation.

A broader body of literature documents how financial constraints affect the transmission of a credit shock due to changes in monetary policy (Gertler and Gilchrist, 1994; Kashyap and Stein, 2000; Jiménez et al. 2012), bank conditions (Klein et al, 2002; Gan, 2007; Duchin et al., 2010; Cingano et al., 2013; Chodorow-Reich, 2014; Balduzzi et al., 2017; Bentolila et al., 2017; Choudhary and Jain, 2017; Acharya et al., forthcoming; Popov and Rocholl, forthcoming), or regulation (Duflo and Banerjee, 2014). We contribute to this literature showing that the effect of credit origination on the income of small business owners is stronger for relatively poor entrepreneurs, and this is especially true at the growth stage of a firm, in low-income regions, and in the crisis period, all instances where entrepreneurs are more credit constrained.

From a methodological perspective, we use unique granular data from a single bank as in Iyer and Puri (2012), Fracassi et al. (2016), Berg (2018), and Delis et al. (2020). We show that our bank is similar across different attributes to 32 other systematically important European banks (identified based on the EBA's guidelines). Importantly, the detailed information on loan applications that we exploit ensures that we rigorously assess the effect of credit decisions on individual income and inequality at the micro level.

3. 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.[3] The bank provides credit to a wide array of small and large firms, as well as 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).[4] 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

[3] 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-.

[4] 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.

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). The applicant characteristics include income (total income reported by the individual, including wages, "dividends" from the firm, and any other source of income), assets (wealth), 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.[5] 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

[5] 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.

otherwise. 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. We explicitly define the credit score along with all the variables used in our empirical analysis in Table 1 and provide summary statistics in Table 2.

[Insert Tables 1 & 2 about here]

Our original data set includes 97,659 loan applications over the time period 2002-2016.[6] For two reasons, we restrict our sample to loan applications from individuals who have exclusive 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.[7] 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 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).[8] 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

[6] 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.

[7] 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.

[8] Non-banks do not benefit from deposit insurance and implicit government guarantees.

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 at al. (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 is a sample of 61,863 loan applications submitted by 15,628 applicants (firms) over 2002-2016.[9] For each applicant, we know his/her income not only at the time of the loan application, but also in the preceding and subsequent years within our time window. This sample also includes information for the rest of the applicant and firm characteristics defined in Table 1.

All individuals reapply for loans within a four-year period and the average time between two consecutive applications is 2.9 years. Applicants apply on average around four times during our sample period and are either always accepted (11,956 applicants), or sometimes accepted and sometimes rejected (3,672 applicants); 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 suggests 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.

We report summary statistics for the variables used in our empirical analysis in Table 2. The mean future 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

---

[9] 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.

approximately 0.1. The average loan size is 34.8 thousand euro, whereas the average loan duration is roughly three years. Summary statistics for our control variables show that the mean applicant has tertiary education and total wealth of €187,200 (see Table 2). The mean firm size (total assets) is €369,500, and mean firm leverage is 20.7%, which is comparable to European averages (e.g., Carvalho, 2017). Overall, the summary statistics show that our data set is consistent with the mean value of our variables at the European level.

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 fully representative across these dimensions.

3.2. Empirical Identification

This study aims at shedding light on the impact of access to credit on 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 and within accepted and rejected applicants. 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 cannot obtain credit from another regulated commercial bank if his/her application is rejected),[10] 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.

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.[11] This implies a sharp discontinuity in treatment as a function of credit score.[12] 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).

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

[10] 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.
[11] This is after dropping the 72 exceptions due to data entry errors.
[12] Berg (2018) exploits a similar type of discontinuity to investigate how loan rejection affects firms' cash holdings.

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

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 allow for a differential effect 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. The main advantage of this approach is the assignment of higher weights as we move closer to the cutoff using a kernel smoother. We determine the optimal bandwidth following Calonico et al. (2014), and for efficient estimation we mainly base our inference on the local-quadratic bias-correction in Calonico et al. (2018).

The distribution of applicant's income depicted in Figure 2 exhibits a regular shape. 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 precise manipulation 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).

[Insert Figure 2 about here]

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. We demonstrate, through a specific statistical test, that credit score manipulation either by applicants or loan officers is most likely absent in our setup. Specifically, we test for manipulation of the assignment variable around the cutoff. Self-selection or nonrandom sorting of applicants would entail a discontinuous change in the distribution of the credit score. Figure 5 shows that the probability density of the credit score does not jump around the cutoff. 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 (see Table 3 and Figure 3). 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 3 about here]

The RDD models of equations (1) and (2) estimate the average effect of access to credit on applicants' income. We build on these two models to identify the impact of a bank's credit decision on income inequality i) between accepted and rejected applicants, and ii) within the groups of "accepted" and "rejected" applicants.

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 4 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. This indicates that the treatment (loan origination) entails a sharp discontinuity in the outcome variable (income), corroborating our methodological approach.

[Insert Figure 4 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.[13] We find a positive and statistically significant coefficient on *Granted* in all specifications.

[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).[14] 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.

We conduct a very large battery of sensitivity tests, which show that our RDD and the associated average treatment effect is robust. We place all results in Appendix B and document that our results are robust to different bandwidth selection methods and to the inclusion of the applicant's initial wealth among the set of controls. In addition, we show that the interest rate charged on newly granted loans does not play a role in driving the effect of loan acceptance on applicants' income. We also conduct robustness tests to (i) firms' ability to get credit outside banks by looking at leverage ratios of firms before and after loan origination, (ii) income increasing with credit quality irrespective of whether individuals obtaining credit from the bank, (iii) the use of

[13] 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 A4 of Appendix B). This means that our baseline model in equation (2) is well specified, and using the controls will not significantly affect our main result.
[14] The average treatment effect here is the counterpart of the coefficient of the acceptance dummy in equation (1).

Heckman sample selection methods that exclude the possibility that our bank selects specific firms from a wider sample of firms in the country, and (iv) placebo tests on the RDD using invalid cutoff points. These tests are important because we base all subsequent analyses on estimations of this baseline RDD.

Overall, our analysis shows that credit decisions have real effects on 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 increase in income experienced after loan origination 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 individual income (Angelucci et al., 2015; Attanasio et al., 2015; Augsburg et al.; 2015; Banerjee et al., 2015; Banerjee et al., 2015b; Crépon et al., 2015; Tarozzi et al., 2015; Banerjee et al., 2018). 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 in Europe. Therefore, our findings reveal that access to credit has a positive effect on individual income when lending decisions are taken efficiently. Also, the magnitude of this effect is substantial, suggesting that credit provision to small businesses impacts significantly the firm owner's economic opportunities and upward mobility.

The large increase in income experienced by accepted applicants vis-à-vis rejected applicants with similar attributes might show that the bank overlooks good investment opportunities. As mentioned before, the percent of denied applications of this bank is in line with the European averages reported in the Survey on access to finance for enterprises (SAFE) published by the European Commission and the ECB. This suggests that the bank may limit its lending capacity as a result of an optimization process. However, further looking into that optimization process is beyond the scope of this paper and we leave it for further research.

4.2. Economic Channels

In our baseline models, we estimate the effect of loan origination on individual income controlling for the size of the loan. Intuitively, the treatment intensity should be stronger the higher is the loan amount. In the first two specifications of Panel A of Table 5, we replicate the regression of column 6 of Table 4, splitting our full sample into small loans and large loans based on the median loan amount. As expected, we find that the effect of credit access on individual income is 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).

Next, we more broadly examine the mechanism behind the observed effect of a positive credit decision on the income of small business owners. In principle, an accepted applicant may use the borrowed funds to invest and expand the business or to smooth consumption over time. To test these economic channels, we rely on a wide set of econometric models where we consider different subsamples and various firm outcomes as dependent variable.

We start by replicating our baseline regression of column 6 of Table 4, separating our sample into new firms and old firms, which are identified as the 25^{th} and the 75^{th} percentile of the distribution of firm age respectively. The last two specifications of Panel A of Table 5 show that, five years after a bank's credit decision, accepted applicants owning a new firm experience an increase in income of 16.7% (column 3), which is more than double the increase in income observed for those who own old firms (column 4). The nonparametric setup adopted does not allow to test if the difference in the treatment effect estimators of the two models is statistically different from zero.[15] However, the large difference of 10.5 percentage points in the effect of loan origination on individual income for business owners of young versus old firms is economically very meaningful. This suggests that access to credit is crucial at the early stage of a business to allow firm investments that foster growth and expansion.

This conjecture is confirmed from the results presented in Panel B of Table 5, where we explicitly look at the evolution of various firm outcomes in response to a credit origination. From a methodological perspective, we use a similar econometric model to that of column 6 of Table 4, the difference being the dependent variable, which consists in the following firm outcomes: the amount of credit borrowed for working capital, which is used to finance everyday business (column 1); a dummy equal to one if the firm is repaying previous loan obligations with the bank and zero otherwise (column 2); firm profitability as captured by the return on assets (column 3); and the growth rate of firm assets (column 4). We find that, five years after the credit decision, firms of accepted applicants invest more in short-term business operations, are more likely to repay existing bank loans, experience a higher increase in profitability, and grow at higher rate compared to firms of rejected applicants. 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. The impact of a positive credit decision on asset growth is about half of what is estimated in Berg (2018). Since Berg (2018) uses a dataset in which the average firm size (about €5 million) is much higher than that of our sample (€369,500), this points to a certain convexity in

[15] As of now, there is no statistical procedure available to perform this kind of test in the nonparametric RDD framework. While, in principle, we could estimate the corresponding parametric models and assess if the difference between the coefficients of the acceptance dummy *Granted* are statistically different from zero, this would not be ideal as it would require departing from the nonparametric RDD setup, which is the most econometrically efficient approach to estimate the treatment effect.

the effect of access to credit on asset growth depending on firm size. Overall, our findings suggest that access to credit is crucial for small firms to undertake investments, expand their business, and be more profitable. This, in turn, has positive repercussions on the future income of the business majority owner. More generally, our findings reveal that credit provision to small businesses (having good investment opportunities) is pivotal to fostering entrepreneurship and economic mobility.

[Insert Table 5 about here]

4.3. Effect on Income Inequality

A natural implication of our key findings is that the income distribution of applicants changes in response to a bank's credit decisions. In this section, we study how credit origination or denial affect income inequality within and between groups of individuals who have similar characteristics (applicants around the cutoff) but receive different credit decisions (accept vs. reject). This analysis helps us to provide a micro-foundation to the study of how large credit expansions affect income inequality in the economy. We start by constructing inequality measures for individuals' income at the time of loan application (t) and five years ahead (t+5). We focus on the sample around the cutoff by using individuals with credit scores less than the absolute value of 0.1.[16]

Panel A of Table 6 reports the results for the Gini coefficient and the Theil index. Both indices increase from time t to time t+5, reflecting higher income inequality. The effect is economically large and equivalent to that identified in Table 4. Specifically, the Gini coefficient increases by approximately 9% and the Theil index increases by approximately 10%, indicating

[16] Alternatively, we use the effective observations left and right of the cutoff produced by the local linear regression in column 6 of Table 4 Panel B. The results are very similar.

considerably higher income inequality after the bank credit decisions for the sample of individuals close to the cutoff.

In Panel B of Table 6, we construct equivalent Gini and Theil indices for accepted and rejected applicants. The indices show that for accepted applicants, the Gini and Theil indices are significantly lower, whereas for the rejected applicants they are higher. This is consistent with the premise that positive credit decisions allow individuals close to the cutoff to increase their incomes, thereby tightening the income distribution among accepted individuals. In contrast, negative credit decisions are consistent with widening income distribution among rejected individuals, who are the relatively poor.

[Insert Table 6 about here]

We explore closely the mechanisms behind the observed decrease and increase in income inequality within accepted and rejected applicants, respectively. One way would be to investigate if the average effect of credit origination on income is heterogeneous across different levels of applicants' income prior to the credit decision. Naturally, we cannot examine this heterogenous effect because rich individuals are always granted loans. Thus, we adopt a different approach. Rather than quantifying the average treatment effect on different subsets of individuals, we estimate the whole distribution of applicant's income responses to a loan origination by converting equation (1) into a simultaneous quantile regression (as described in Angrist and Pischke, 2008). In other words, we estimate a model that allows us to predict different quantiles of applicant's income five years after the loan decision based on whether the firm is given credit or not. This way we can investigate how the entire distribution of income changes as a function of a loan origination.

Figure 5 reports the estimates of the simultaneous quantile regression in graphical form. The income responses are widely different across quantiles. The effect of *Granted* is approximately 17% for the first quantile of the income distribution, dropping to 11% and 9% for the 10^{th} percentile and 25^{th} percentile, respectively. At the median, the effect is approximately 8%. For the top incomes (90^{th} percentile and higher), the treatment effect is below 3% and becomes statistically insignificant. The coefficient equals almost 0 for the 99^{th} percentile of the income distribution. This evidence provides relevant insights on how the bank's credit decisions affect income inequality. A loan origination has a positive impact predominantly on the poor, it has a gradually declining effect on the middle class, and does not affect the top incomes.

[Insert Figure 5 about here]

As a second exercise, we examine the probability that an applicant moves in a different income bracket after the bank's loan decision. To this end, we separately estimate the following probit models:

$$P\ upward_{i,t+5} = b_0 + b_1 D_{it} + b_2(x_{it} - \bar{x}) + b_3 D_{it}(x_{it} - \bar{x}) + u_{it}. \quad (3)$$

$$P\ downward_{i,t+5} = b_0 + b_1 D_{it}^{-} + b_2(x_{it} - \bar{x}) + b_3 D_{it}^{-}(x_{it} - \bar{x}) + u_{it}. \quad (4)$$

In equation (3), *P upward* is a binary variable equal to 1 if applicant *i* moves at least one decile up in the income distribution and zero otherwise. Complementary to (3), equation (4) examines whether a negative credit decision (D^- equals 1 if the credit score is below the cutoff and zero otherwise) moves the applicant at least one decile down the income distribution. The coefficient of interest is b_1, which captures the treatment effect.

We report the estimates of the marginal effects at means in Table 7, with those of equation (3) being in column 1 and those from equation (4) being in column 2. The first specification shows that a loan origination leads to a 7.7% increase in the probability that an applicant moves at least one decile upward in the income distribution. The corresponding effect in model 2 is considerably smaller, with a credit denial increasing the probability that a rejected applicant moves downward the income distribution by only 2% (and barely statistically significant).

[Insert Table 7 about here]

Consistent with the evidence in Figure 5, we expect the estimates of b_1 in equations (3) and (4) to be more pronounced for individuals at the lower end of the income distribution, at least with regard to the probability of accepted applicants moving upward. To test this hypothesis, we estimate the marginal effect of *Granted* in equation (3) at different levels of applicants' income at the time of the loan application. We plot the estimates for the income deciles in Figure 6.[17] Consistent with our prior, the marginal effects are considerably higher for applicants in the left tail of the income distribution, starting with a probability of an upward shift of approximately 18% for those within the 1st percentile and declining to approximately 4% for the median applicant.

[Insert Figure 6 about here]

Overall, these results indicate that the observed reduced income inequality among accepted applicants is driven by a strong effect of a loan origination on the income of poor individuals, which becomes progressively weaker once we move towards the right tail of the income distribution and almost vanishes for top income-earners. One important caveat of our analysis, however, is that, albeit we focus on a large systemic bank, credit originated by this lender is small relative to the size of the economy. Thus, from an external validity viewpoint, we cannot ascertain if a larger scale credit expansion would generate the same effects on income inequality.

4.4. Heterogeneity in the Effect on Income Inequality

[17] We include the marginal effect for the individuals below or equal the 1st percentile of the income distribution as zero on the horizontal axis.

A natural extension of our inequality analysis is to investigate if there is any relevant heterogeneity in the effect of credit origination on income inequality. To this end, we conduct a suite of tests that explore different dimensions, including geographical location, macroeconomic environment, firm age, firm industry, and gender of the business owner. We report the results from these regressions for the 1st, 10th, 25th, median, 75th, 90th, and 99th percentiles in Figure 7 (graphical form). Each panel includes two figures, separating high-income regions from low-income regions, the pre-crisis period (2002-2008) from the crisis period (2009-2014), new firms from old firms, high-growth industries from low-growth industries, and male business owners from female owners.

[Insert Figure 7 about here]

In virtually all graphs, the future incomes of the relatively constrained groups (e.g., low-income regions, crisis period, new firms, etc.) reflect higher responses to credit origination for the relatively poor applicants (left tail of the income distribution). By order of significance, this is more evident in the high-growth industries, new firms, and firms with female owners. Again, for the richer applicants, the effects shown in all graphs are economically negligible.

Specifically, Panel A of Figure 7 concerns the role of applicant location based on regional income, distinguishing between low-income regions and high-income regions.[18] Looking at the first percentile of the income distribution, we find that, five years after a bank's credit decision, accepted applicants have 17% higher incomes than rejected applicants in low-income regions. The equivalent effect in the high-income regions is 15%, 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). The wedge between the treatment effect in low-income regions and

[18] This analysis is in the same spirit of Agarwal et al. (2018), who document an income-based geographical heterogeneity in the effect of a micro-credit program on financial access in Rwanda.

high-income regions shrinks to almost disappear once we move to higher quantiles of the income distribution.

As a second exercise, we consider the role of the Great Recession. During this period, Europe experienced sharp losses in household wealth and aggregate demand, substantial contraction of credit, and increased unemployment (e.g., IMF, 2009; ECB, 2016). In such context, entrepreneurs face riskier investment opportunities and lower profits. We split the sample into the 2000-2008 and the 2009-2016 periods. We leave 2008 in the pre-crisis period because credit from banks in European countries was still rising that year. Similarly, we include the full period after the crisis because credit from banks to the private sector over GDP decreased in 2009-2016.[19] Panel B of Figure 10 reports the results from the two samples. We do find a larger response for the particularly poor (1st and 10th centiles) but overall se do not notice large differences in the distribution of our estimates.

Next, we focus on firm attributes, starting with firm age. The evidence presented in specifications 3 and 4 of Table 5 reveals that the average impact of access to credit on the income of small business owners is considerably stronger for young firms compared to old firms. Panel C of Figure 7 reports the results of the simultaneous quantile regressions. We confirm that the difference in the treatment effect between new firms and old firms is particularly pronounced for the relatively poor (3 pp at the first percentile and 2 pp at the 25th percentile), it declines for the middle class (1 pp at the median), and it disappears for individuals with top incomes (0 pp at the 99th percentile). These findings confirm the importance of bank credit for startups to foster expansions and highlights how it is particularly crucial for relatively poor entrepreneurs.

[19] See https://data.worldbank.org/indicator/FS.AST.PRVT.GD.ZS?locations=XC.

We also distinguish between firms operating in high-growth sectors and firms operating in low-growth sectors (again separating them using the median). This analysis (Panel D of Figure 7) provides the largest heterogeneity in the results. The positive effect of a loan origination on the income of a small business owner is 8 pp larger for firms in high-growth industries compared to firms in low-growth industries at the first percentile of the income distribution. Again, the wedge shrinks for the middle-income group (1 pp at the median) and vanishes for the rich (0 pp at the 99th percentile).

Lastly, we look at male versus female applicants. The quantile regression estimates in Panel E of Figure 7 reveal that low-income female applicants gain more than male applicants in the same income buckets from a positive credit decision (a 2.5 pp higher increase in income at the first percentile). Moreover, the coefficient of *Granted* for female entrepreneurs stays above that for males and is statistically significant across most levels of the income distribution below the 80th percentile, with the only exception of the interval that goes approximately from the 5th quantile to the 30th percentile. This finding highlights the importance of credit for female entrepreneurship and how it can contribute to narrow the gender-pay gap.[20]

Interestingly, we also observe some important differences of the effects across the income distribution. For all graphs, the relatively constrained groups (e.g., low-income regions, crisis, new firms, female owners) have higher responses almost up to the 75th centile of the distribution, whereas the equivalent effects for the relatively unconstrained groups fade after the 25th centile or the median. Overall, these findings highlight important heterogeneity in the effect of credit

[20] Additional tests run on i) geographical areas with a different percentage of non-western immigrants, ii) different industry types (by standard industry classification codes), and iii) individuals' age groups (decades) did not show significant differences in the estimates reported for these groups.

origination across income groups with different characteristics and pinpoint important mechanisms via which credit origination/denial affects the future income distribution.

4.5. Hard Information and Soft Information

In this section, we expand the RDD setup to explore the role played by hard and 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 Table 8 show that there is no statistical evidence of an artificial manipulation of credit scores from loan officers.

[Insert Table 8 about here]

As a second step, we replicate the nonparametric regressions in columns 4-6 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). Table 9 reports the results. In specifications 1-3, 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 specifications 4-6, 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 3), compared to 7% when soft information contributes negatively (column 6).[21] This finding suggests that the marginal benefit of getting a loan is stronger when a loan acceptance is favored by a 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

[21] As mentioned earlier, no statistical procedure is available to test if the difference in the treatment effect estimators, obtained on different subsamples in the nonparametric RDD framework, is statistically significantly different from zero. However, the difference of 6.5 percentage points between the estimators of specifications 3 and 4 is economically very meaningful.

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.

[Insert Table 10 about here]

5. Conclusions

Credit constraints potentially hinder income growth opportunities, especially for those with low incomes and a lack of collateral. Using data from business loan applications to a single large European bank, we study and quantify how a bank's credit decisions (acceptance or rejection) affect the distribution of applicants' income and inequality.

We look at a unique sample of loan applications from small and micro enterprises for which we have detailed information on past and future income, the credit score assigned by the bank, and the exclusivity of relationship lending with that bank (among many other applicant and firm characteristics). Our identification strategy comprises a regression discontinuity design, exploiting exogenous variation in the credit decision from the cutoff rule based on the credit score. Essentially, with this strategy we compare individuals with credit scores (and thus very similar characteristics guiding the credit decision) around the cutoff.

We first show that access to credit has a sizeable positive effect on individual income three to five years after the loan application. This finding is robust to several re-specifications and robustness tests. We also show that firms of accepted applicants use the borrowed funds to make investments and expand their business, ultimately experiencing higher profitability and growth rates compared to firms of rejected applicants.

Overall, these results suggest that (an efficient) credit provision to small businesses has a positive impact on individual upward mobility. To this end, we closely investigate the implications of such effects in terms of income inequality. Using the Gini and Theil indices, we show that the distribution of income is tighter among accepted applicants and wider among rejected applicants. The decline in income inequality within recipients of credit is driven by an upward shift of poor business owners in the distribution of income, meaning that low-income entrepreneurs become closer to top-income entrepreneurs once they get a loan. In other words, access to credit fosters upward mobility and reduced income inequality.

Looking across interesting groups of applicants, we document that our results are more pronounced in low-income regions (vs. high-income regions), for young firms (vs. old firms), for firms operating in high-growth industries (vs. low-growth industries), and for female entrepreneurs vs. male entrepreneurs). This means that a credit expansion has a negative impact on income inequality especially if business owners are credit constrained, have growth opportunities, or belong to disadvantaged groups, hereby reinforcing the negative relation between credit availability and income inequality. Lastly, we show that the effect of credit origination on income and its distribution is more pronounced when a loan acceptance is favored by soft information held by the bank (for example on the quality of the investment opportunities of the firm).

Our findings have two key and interrelated economic implications. First, efficient credit decisions strongly affect applicants' future income and its subsequent dynamics, altering lifetime income expectations and potentially applicants' economic decisions. Second, credit decisions exert substantial effects on income inequality among individuals who prior to the credit decision have similar credit scores.

In general, the evidence that efficient credit decisions affect positively economic mobility and reduce inequality provides support to policy interventions aimed at increasing credit access to loan applicants rejected by the banking system due to lack of credit history or collateral. Relevant actions are those of the European Bank for Reconstruction and Development (EBRD) and the European Investment Bank (EIB), which selectively target credit-constrained individuals with good investment ideas, and of the Small Business Administration, which guarantees loans to small firms lacking access to credit but having good business financials. We leave the thorough examination of the effects of these policies to future research.

References

Aaronson, D., Hartley, D., Mazumder, B., Forthcoming. The effects of the 1930s HOLC "Redlining" Maps. *American Economic Journal: Economic Policy*.

Acabbi, E. M., Panetti, E., Sforza, A., 2020. The financial channels of labor rigidities: Evidence from Portugal. Working Paper.

Acharya, V., Eisert, T., Eufinger, C., Hirsch, C., Forthcoming. Real effects of the sovereign debt crisis in Europe: Evidence from syndicated loans. *Review of Financial Studies*.

Adams, R. B., Ferreira, D., 2009. Women in the boardroom and their impact on governance and performance. *Journal of Financial Economics* 94, 291-309.

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

Agarwal, S., Kigabo, T., Minoiu, C., Presbitero, A., Silva, A., Forthcoming. Serving the underserved: Microcredit as a pathway to commercial banks. *Review of Economics and Statistics*.

Alesina, A., Rodrik, D. 1994. Distributive politics and economic growth. *Quarterly Journal of Economics* 109 (2), 465-490.

Angelucci, M., Karlan, D., Zinman, J., 2015. Microcredit impacts: Evidence from a randomized microcredit program placement experiment by Compartamos Banco. *American Economic Journal: Applied Economics* 7 (1), 151-182.

Angrist, J. D., Pischke, J., 2008. Mostly harmless econometrics. Princeton University Press.

Appel, I., Nickerson, J., 2016. Pockets of poverty: The long-term effects of redlining. Working paper.

Artavanis, N., Morse, A., Tsoutsoura, M., 2016. Measuring income tax evasion using bank credit: Evidence from Greece. *Quarterly Journal of Economics* 131 (2), 739-798.

Attanasio, O., Augsburg, B., De Haas, R., Fitzsimons, E., Harmgart, H., 2015. The impacts of microfinance: Evidence from joint-liability lending in Mongolia. *American Economic Journal: Applied Economics* 7 (1), 90-122.

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., Karlan, D., Zinman, J., 2015. Six randomized evaluations of microcredit: Introduction and further steps. *American Economic Journal: Applied Economics* 7 (1), 1-21.

Banerjee, A., Duflo, E., Hornbeck, R., 2018. How much do existing borrowers value microfinance? Evidence from an experiment on bundling microcredit and insurance. *Economica* 85, 671-700.

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

Beck, T., Levine, T., Levkov, A., 2010. Big bad banks? The winners and losers from bank deregulation in the United States. *Journal of Finance* 65 (5), 1637-1667.

Becker, S. O., Egger, P. H., Von Ehrlich, M., 2013. Absorptive capacity and the growth and investment effects of regional transfers: A regression discontinuity design with heterogeneous treatment effects. *American Economic Journal: Economic Policy* 5 (4), 29-77.

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.

Bentolila, S., Jansen, M., Jiminéz, G., Ruano, S., 2017. When credit dries up: Job losses in the great recession. *Journal of the European Economic Association* 16 (3), 650-695.

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.

Bergman, P., Chetty, R., DeLuca, S., Hendren, S., Katz, L. F., Palmer, C., 2019. Creating moves to opportunity: Experimental evidence on barriers to neighborhood choice. NBER Working Paper No. 26164.

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

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.

Charness, G., Gneezy, U., 2012. Strong evidence of gender differences in risk taking. *Journal of Economic Behavior and Organization* 83, 50-58.

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., Grusky, D., Hell, M., Hendren, N., Manduca, R., Narang, J., 2017. The fading American dream: Trends in absolute mobility since 1940. *Science* 356 (6336), 398-406.

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.

Chetty, R., Hendren, N., 2018a. The effects of neighborhoods on intergenerational mobility I: Childhood exposure effects. *Quarterly Journal of Economics* 133 (3), 1107-1162.

Chetty, R., Hendren, N., 2018b. The effects of neighborhoods on intergenerational mobility II: County level estimates. *Quarterly Journal of Economics* 133 (3), 1663-1228.

Chetty, R., Hendren, N., Jones, M. R., Porter, S. R., 2020. Race and economic opportunity in the United States: An intergenerational perspective. *Quarterly Journal of Economics* 135 (2), 711-783.

Chetty, R., Hendren, N., Katz, L., 2016. The effects of exposure to better neighborhoods on children: New evidence from the moving opportunity experiment. *American Economic Review* 106 (4), 855-902.

Chodorow-Reich, G., 2014. The employment effects of credit market disruptions: Firm-level evidence from the 2008-2009 financial crisis. *Quarterly Journal of Economics* 129, 1–59.

Choudhary, M. A., Jain, A., 2017. Finance and inequality: The distributional impacts of bank credit rationing. International Finance Discussion Papers No. 1211.

Cingano, F., Manaresi, F., Sette., E., 2013. Does credit crunch investments down? *Review of Financial Studies* 29, 2737–73.

Clarke, G. R. G., Xu, L. C., Zou, H. F., 2006. Finance and income inequality: What do data tell us? *Southern Economic Journal* 72 (3), 578-596.

Corak, M., 2013. Income inequality, equality of opportunity, and intergenerational mobility. *Journal of Economic Perspective* 27 (3), 79-102.

Crépon, B, Devoto, F., Duflo, E., Parienté, W., 2015. Estimating the impact of microcredit on those who take it up: Evidence from a randomized experiment in Morocco. *American Economic Journal: Applied Economics* 7 (1), 123-150.

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.

Demigrüç-Kunt, A., Levine, R., 2009. Finance and inequality: Theory and evidence. *Annual Review of Financial Economics* 1, 287-318.

Denk, O., Cournède, B., 2015. Finance and income inequality in OECD countries. OECD Economic Department Working Papers No. 1224.

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

Duchin, R., Ozbas, O., Sensoy, B. A., 2010. Costly external finance, corporate investment, and the subprime mortgage crisis. *Journal of Financial Economics* 97, 418–35.

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.

Eckel, C. C., Grossman, P. J., 2002. Sex differences and statistical stereotyping in attitudes toward financial risk. *Evolution and Human Behavior* 23 (4), 281-295.

Eckel, C. C., Grossman, P. J., 2008. Men, women and risk aversion: Experimental evidence. Handbook of Experimental Economics Results, Volume I, Chapter 113, Elsevier.

Evans, D. S., Jovanovic, B., 1989. An estimated model of entrepreneurial choice under liquidity constraints. *Journal of Political Economy*, 97 (4), 808-827.

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., Moav, O. 2004. From physical to human capital accumulation: Inequality and the process of development. *Review of Economic Studies* 71 (4), 1001-1026.

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

Gan, J., 2007. The real effects of asset market bubbles: Loan- and firm-level evidence of a lending channel. *Review of Financial Studies* 20, 1941–73.

Gertler, M., Gilchrist., S., 1994. Monetary policy, business cycles, and the behavior of small manufacturing firms. *Quarterly Journal of Economics* 109, 309–40.

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

Hamori, S., Haschiguchi, Y., 2012. The effect of financial deepening on inequality: Some international evidence. *Journal of Asian Economics* 23, 353-359.

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.

Holt, C. A., Laury, S. K., 2002. Risk aversion and incentive effects. *American Economic Review* 92 (5), 1644-1655.

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., Peydró, J. L., da-Rocha-Lopes, S., Schoar, A., 2014. Interbank liquidity crunch and the firm credit crunch: evidence from the 2007-2009 crisis. *Review of Financial Studies* 27 (1), 347-372.

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

Jauch, S., Watzka, S., 2016. Financial development and income inequality: A panel data approach. *Empirical Economics* 51, 291-314.

Jianakoplos, N. A., Bernasek, A., 1998. Are women more risk averse? *Economic Inquiry* 36, 620-630.

Jiménez, G., Ongena, S., Peydró, J. L., Saurina, J., 2012. Credit supply and monetary policy: Identifying the bank balance-sheet channel with loan applications. *American Economic Review* 102, 2301–6.

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.

Kappel, V., (2010). The effects of financial development on income inequality and poverty. CER-ETH Working Paper 10/127.

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.

Kearney, M. S., Levine, P. B., 2016. Income Inequality, social mobility, and the decision to drop out of high school. *Brookings Papers on Economic Activity*, Spring 2016, 333-380.

Kim, D. H., Lin, S. C., 2011. Nonlinearity in the financial development-income inequality nexus. *Journal of Comparative Economics* 39 (3), 310-325.

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.

Kruse, J. B., Thompson, M. A., 2003. Valuing low probability risk: Survey and experimental evidence. *Journal of Economic Behavior and Organization* 50, 495-505.

Kumhof, M., Rancière, R., 2010. Inequality, leverage and crises. IMF Working Paper 10/268.

Larsen, A. C., 2013. The rise and fall of social cohesion: The construction and de-construction of social trust in the US, UK, Sweden and Denmark. Oxford University Press.

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

Li, H., Squire, L., Zou, H. F., 1998. Explaining international and intertemporal variations in income inequality. *Economic Journal* 108, 26-43.

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.

Moore, E., Eckel, C. C., 2003. Measuring ambiguity aversion. Working Paper.

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.

OECD, 2015a. Financing SMEs and Entrepreneurs 2015: An OECD scoreboard.

OECD, 2015b. In it together: Why less inequality benefits all. Overview of inequality trends, key findings and policy directions. OECD Publishing.

Persson, T., Tabellini, G., 1994. Is inequality harmful for growth? *American Economic Review* 84 (3), 600-621.

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

Piketty, T., Saez, E., 2013. Top incomes and the great recession: Recent evolutions and policy implications. *IMF Economic Review* 61 (3), 456-478.

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.

Putnam, R., 2000. Bowling alone: The collapse and revival of American community. Simon & Schuster Paperbacks.

Rajan, R. G., 1992. Insiders and outsiders: The choice between informed and arm's-length debt. *Journal of Finance* 47 (4), 1367-1400.

Rajan, R. G., 2010. Fault lines. How hidden fractures still threaten the world economy. Princeton University Press.

Saez, E., Matsaganis, M., Tsakloglou, P., 2012. Earnings determination and taxes: Evidence from a cohort-based payroll tax reform in Greece. *Quarterly Journal of Economics* 127, 493-533.

Schubert, R., Brown, M., Gysler, M., Brachinger, H. W., 1999. Gender and economic transactions. *AEA Papers and Proceedings* 89 (2), 381-385.

Sharpe, S., 1990. Asymmetric information, bank lending and implicit contracts: A stylized model of consumer relationships. *Journal of Finance* 45 (4), 1069-1087.

Solt, Frederick. 2016. "The Standardized World Income Inequality Database". *Social Science Quarterly* 97. SWIID Version 6.2, March 2018.

Stiglitz, J., 2012. The price of inequality: How today's divided society endangers our future. W. W. Norton & Company.

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 euro amount of individuals' total annual income (in log).
Working capital loan	Log of the amount of a working capital facility.
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.
Firm growth	The annual growth rate of firm 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	Total firm's assets (in log).
Firm leverage	The ratio of firm's 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	Log of the requested loan amount in thousands of euros.
Maturity	Requested loan duration 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 euro amount of individuals' total wealth, as estimated by the bank (in log).
Initial wealth	Individuals' wealth in the first year before the loan application in which this information is available (one to five years before).

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.

	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
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
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
Working capital loan	61,863	1.925	0.714	0.679	5.825
ROA	61,863	0.094	0.160	-0.711	0.836
Firm growth	61,863	0.193	0.386	-1.938	6.484

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. 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, Obs. is the original number of observations. 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. The bandwidth selection procedure is the one proposed by Calonico et al. (2014). 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)
Observations	57,766	49,514	41,391	53,585	45,333	37,210
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
Economic channels

The table reports coefficients and standard errors (in parentheses). The dependent variable is given in the first row of each panel and all variables are defined in Table 1. 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 first two specifications of panel B distinguish between small and large loans, which are identified as the 25th and the 75th percentile of the distribution of *loan amount*, respectively. The last two specifications of panel B distinguish between new and old firms, which are identified as the 25th and the 75th percentile of the distribution of *firm age*, respectively. The dependent variables in panel B consist in various firm outcomes, including the amount of a working capital loan taken to finance short-term operations, a dummy equal to one if the firm is repaying previous loan obligations, the return on asset and the growth rate of the firm. The *, **, and *** marks denote statistical significance at the 10%, 5%, and 1% level, respectively. Obs. is the original number of observations. 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. The bandwidth selection procedure is the one proposed by Calonico et al. (2014). The robust variance estimator is obtained according to Calonico at el. (2014).

Panel A. Small vs large loans, new vs old firms

	Small loans	Large loans	New firms	Old firms
	(1)	(2)	(3)	(4)
Dependent variable	Income t+5	Income t+5	Income t+5	Income t+5
Robust	0.105***	0.118***	0.167***	0.0623***
	(0.0171)	(0.0216)	(0.0386)	(0.0162)
Observations	8,226	3,507	2,727	13,245
Eff. obs. left of cutoff	1,499	403	662	2,015
Eff. obs. right of cutoff	2,022	416	679	2,026
BW estimate	14.69	8.67	10.07	14.55
BW bias	16.52	10.11	12.81	17.39

Panel B. Firm outcomes

	(1)	(2)	(3)	(4)
Dependent variable	Corporate purpose t+5	Debt repay t+5	ROA t+5	Firm growth t+5
Robust	0.131***	0.048**	0.048**	0.035***
	(0.019)	(0.022)	(0.0207)	(0.0118)
Observations	27,628	7,311	41,391	41,391
Eff. obs. left of cutoff	5,211	1,361	4,815	4,927
Eff. obs. right of cutoff	5,440	1,407	5,003	5,093
BW estimate	20.6	13.24	61.27	67.91
BW bias	22.46	15.72	95.16	107.18

Table 6
Inequality measures

Panel A reports the Gini coefficient and the Theil index for individuals' income at time t and time t+5 around the cutoff (credit score < |0.1|). Panel B compares the equivalent Gini coefficients and Theil indices for the samples of granted and non-granted loans.

	Income t	Income t+5
Panel A. Inequality measures around the cutoff		
Gini coefficient	0.207	0.226
Theil index	0.067	0.074
Panel B. Inequality measures for accepted vs. denied applicants		
Credit is granted		
Gini coefficient	0.224	0.200
Theil index	0.080	0.065
Credit is denied		
Gini coefficient	0.193	0.214
Theil index	0.058	0.073

Table 7
Probability of applicants moving to a different decile of the income distribution

The table reports estimates of the marginal effects and standard errors (in parentheses) from the probit regressions of equation (3) in column 1 and equation (4) in column 2. The dependent variable is given in the first row of the table and all variables are defined in Table 1. The *, **, and *** marks denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1)	(2)
Dependent variable	P upward	P downward
Granted	0.077***	
	(0.013)	
Rejected		0.020*
		(0.011)
Credit score	0.232***	-0.176***
	(0.040)	(0.038)
Obs.	53,585	53,585
Controls as in Table 4	Yes	Yes
Clustering	Individual	Individual

Table 8
Manipulation test for soft information

The table reports 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 of Table 11 are positive (panel A) and the subsample where the residual are negative or zero (panel B). 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 A. Residuals>0		Panel B. 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

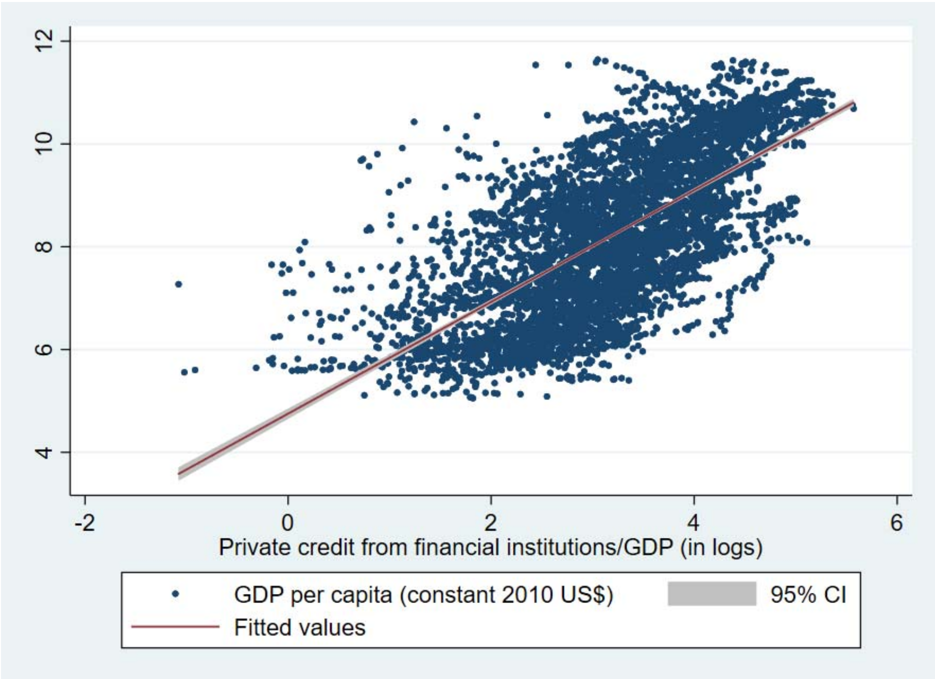
Table 9
The role of soft information

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. 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 table replicates the analysis of columns 4-5 of Table 5 on different subsamples depending on the residuals of a 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. Specifications 1 to 3 are estimated on the subsample where the residuals are positive and specifications 4-6 where the residual are negative or zero. The *, **, and *** marks denote statistical significance at the 10%, 5%, and 1% level, respectively. Obs. is the original number of observations. 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. The bandwidth selection procedure is the one proposed by Calonico et al. (2014). The bias-corrected RD estimator and the robust variance estimator are obtained according to Calonico et al. (2018) and Calonico at el. (2014), respectively.

	Residuals>0			Residuals≤0		
	(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
Robust	0.0764***	0.0595**	0.135***	0.0856***	0.0391	0.0695*
	(0.0244)	(0.0234)	(0.0293)	(0.0319)	(0.0318)	(0.0378)
Observations	26,817	22,835	18,955	26,768	22,498	18,255
Eff. obs. left of cutoff	4,649	3,927	2,549	3,748	3,375	2,373
Eff. obs. right of cutoff	4,937	4,118	2,720	4,549	3,373	2,556
BW estimate	56.13	54.27	47.11	54.20	52.29	41.28
BW bias	94.29	93.18	79.26	92.16	90.25	76.64

Figure 1
Income and income inequality against credit

The first graph depicts GDP per capita (in constant 2010 US$) against the ratio of private credit to GDP (x-axis). The second graph depicts the Gini index against the ratio of private credit to GDP (x-axis). We report individual values, as well as fitted values using a linear regression model. The estimated slopes of the linear regressions are 1.087 and -0.077, respectively, and are statistically significant at the 1% level. Data on the Gini index are from the Standardized World Income Inequality Database (SWIID); data on credit and GDP per capita are from the World Development Indicators.



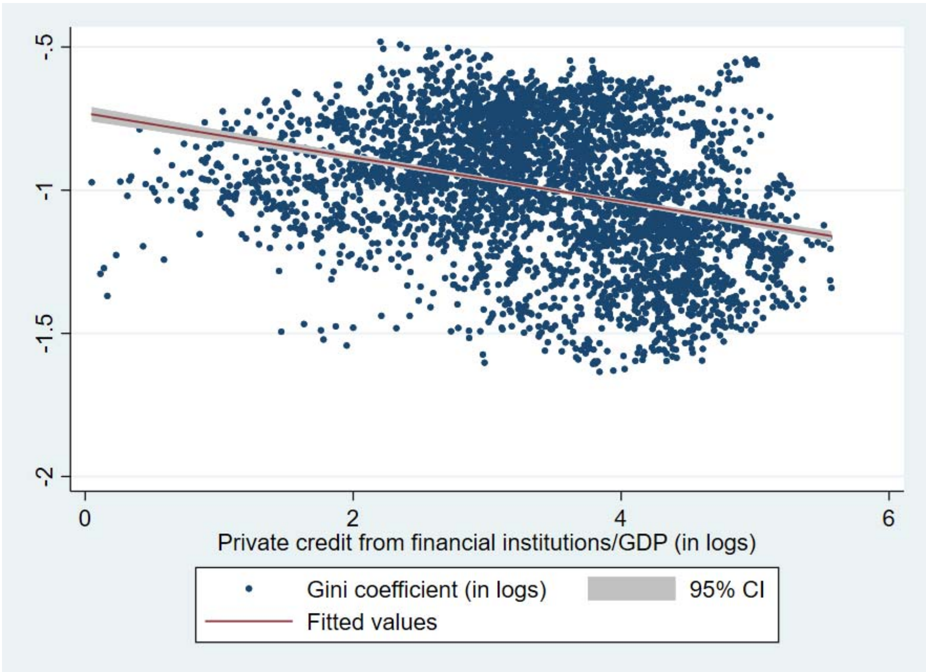


Figure 2
Densities of outcome and assignment variables

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

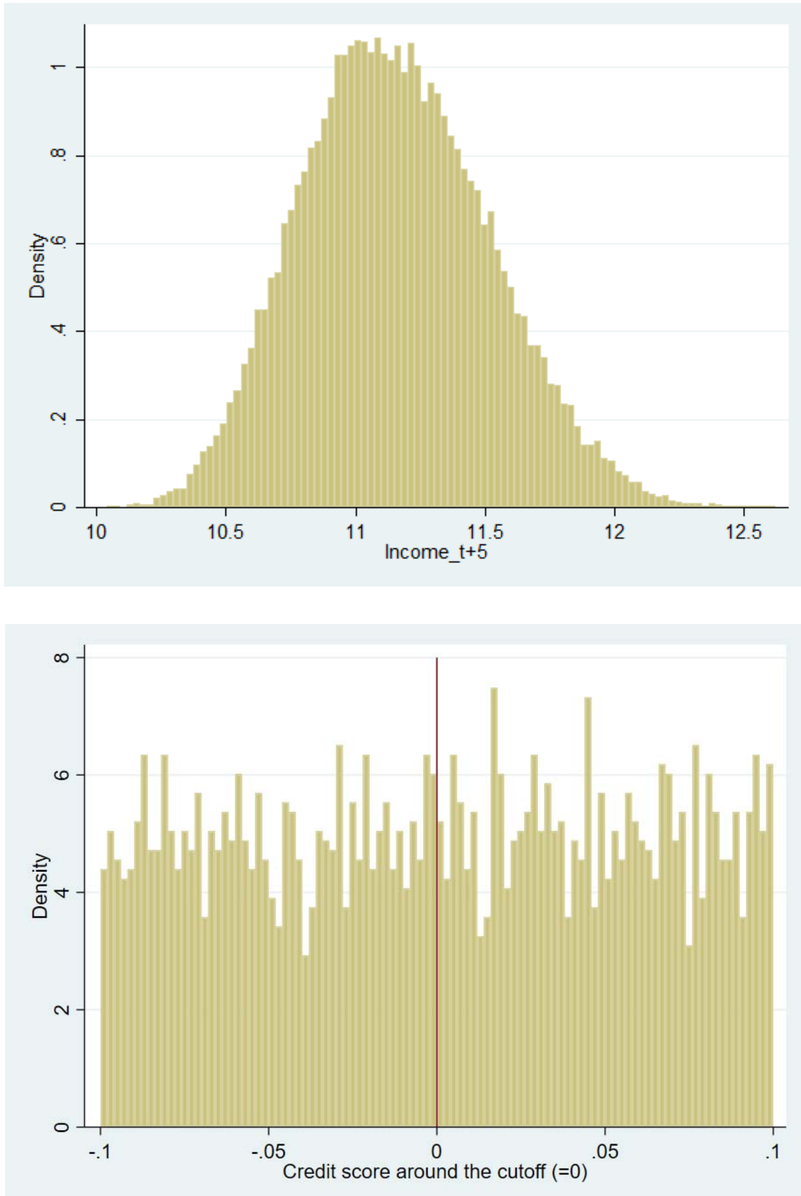


Figure 3
Manipulation test

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.

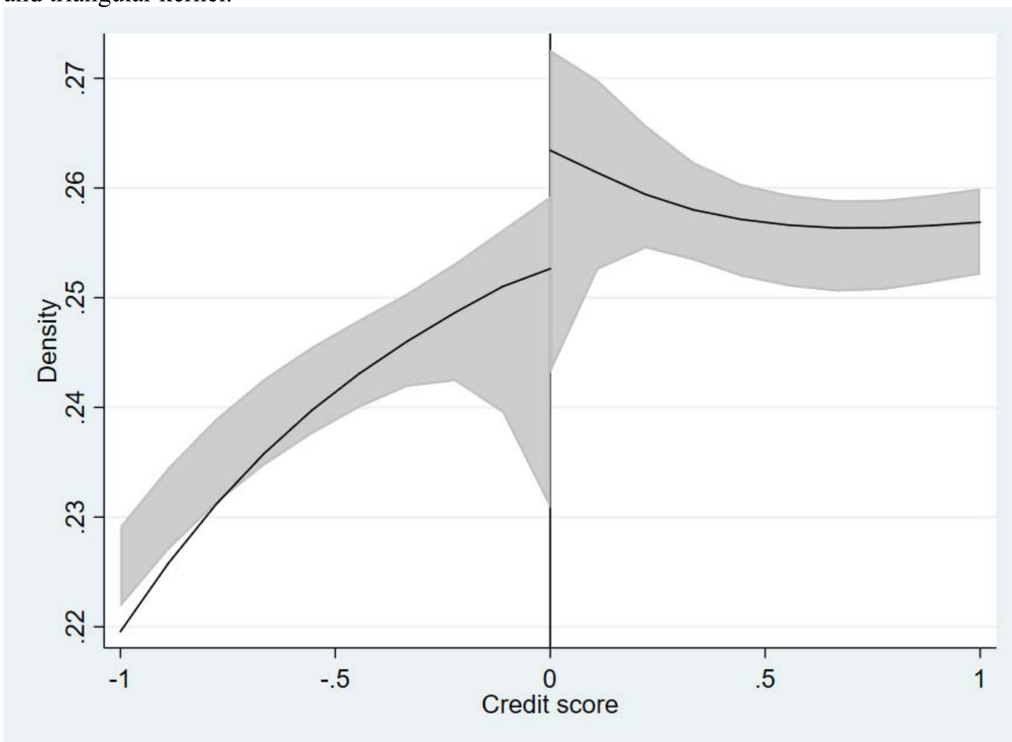


Figure 4
Applicants' income around the cutoff

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.

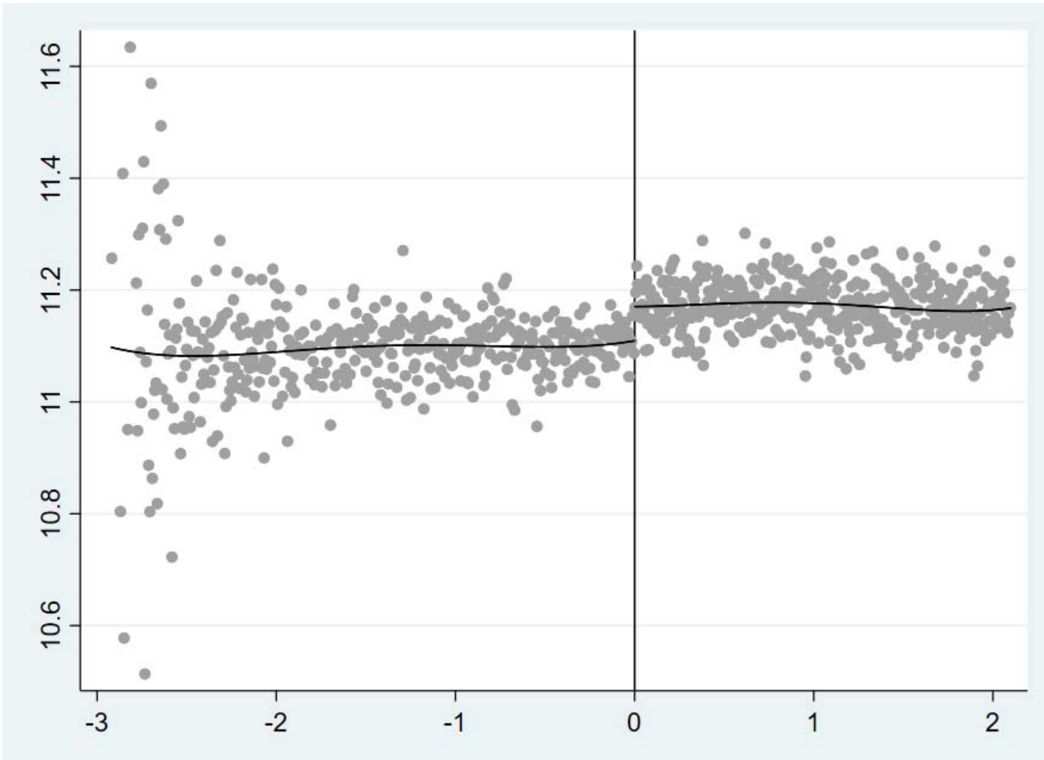


Figure 5. Estimates from simultaneous quantile regressions

The figure shows coefficient estimates from the simultaneous quantile regressions of equation (1) along with their 95% confidence interval. The estimates are for the 1%, 5%, 10%, 25%, 50% (median), 75%, 90%, 95%, and 99% of the distribution of income five years after the credit decision.

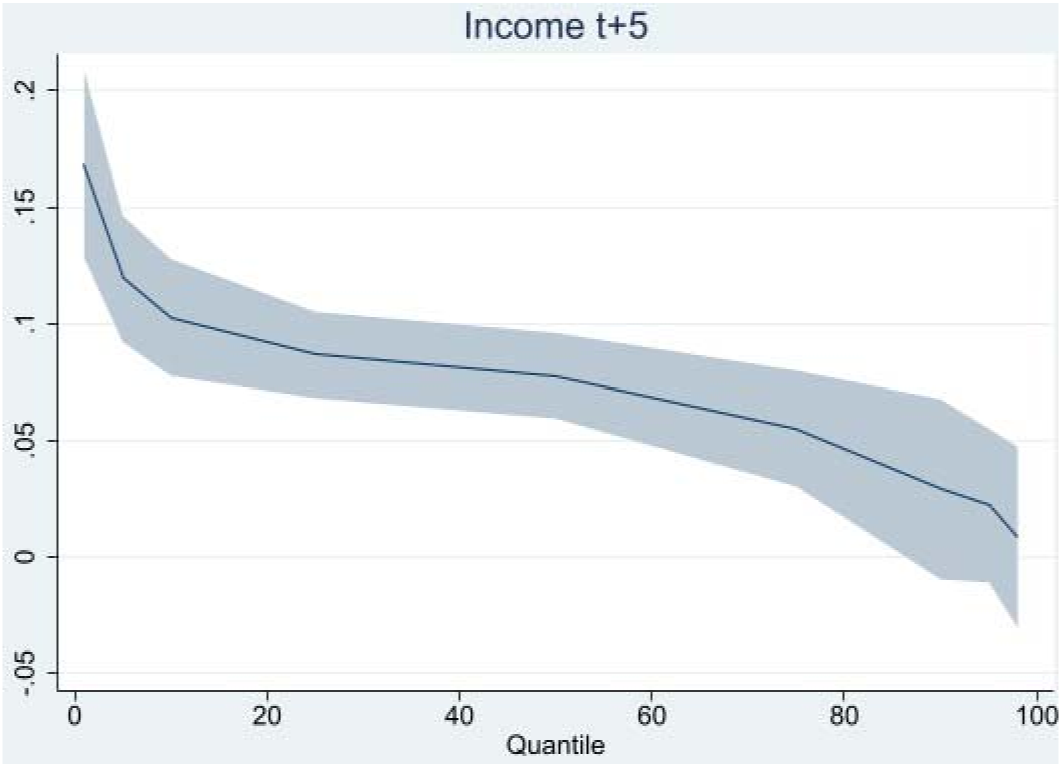


Figure 6. Probability of accepted applicants moving upward in the income distribution for different levels of initial income

The figure shows the predictive marginal effects of *Granted* from the estimation of equation (3) along with their 95% confidence intervals. Marginal effects are estimated for each decile of the distribution of individual income at the time of the loan application (including the 1[st] percentile as zero on the horizontal axis).

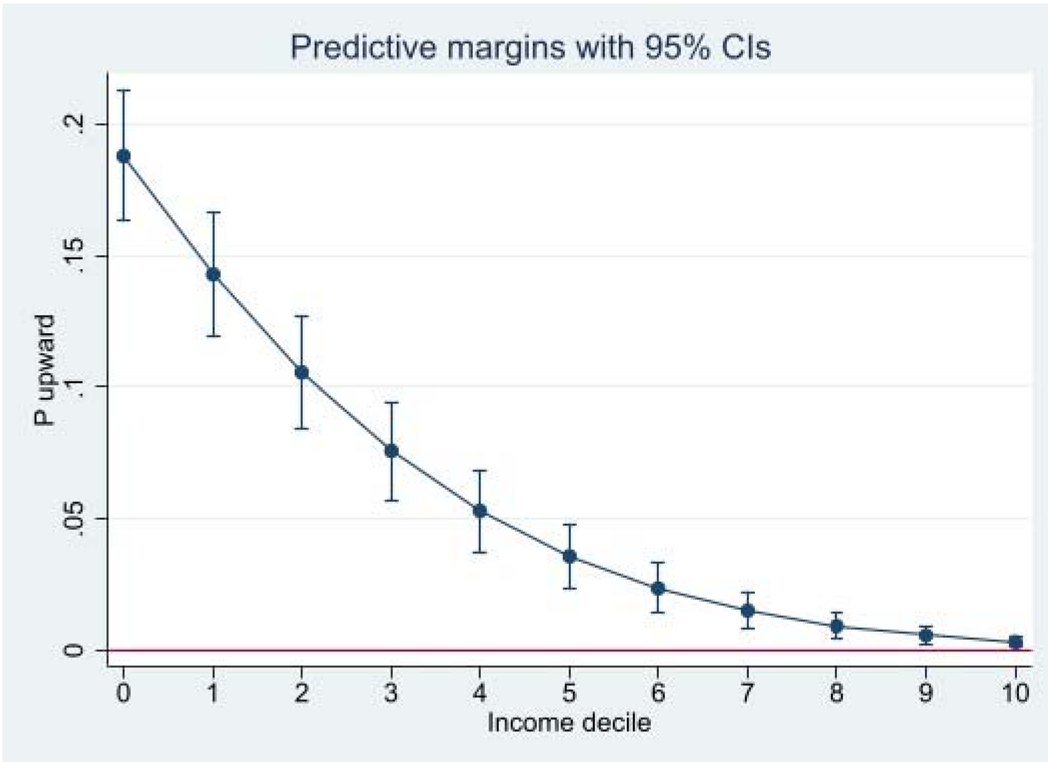
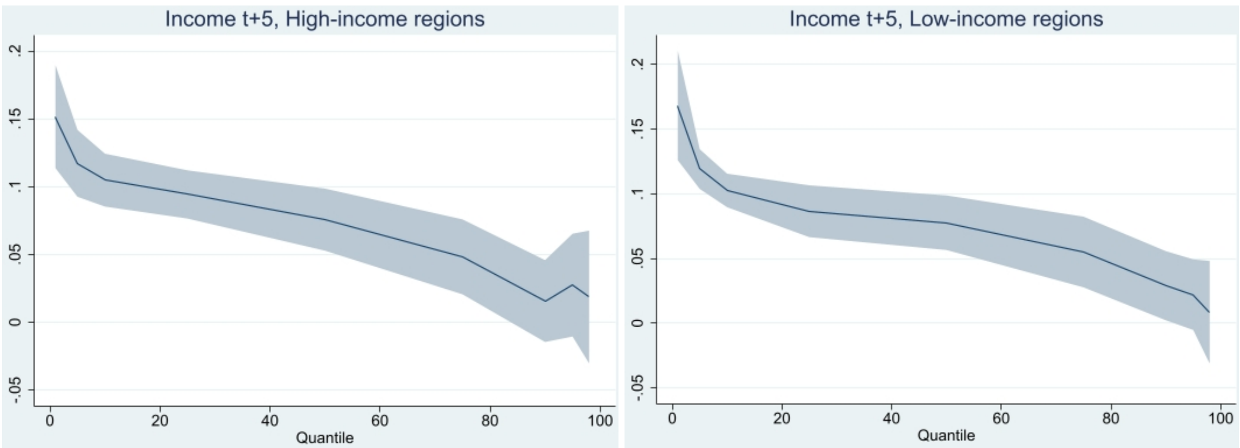


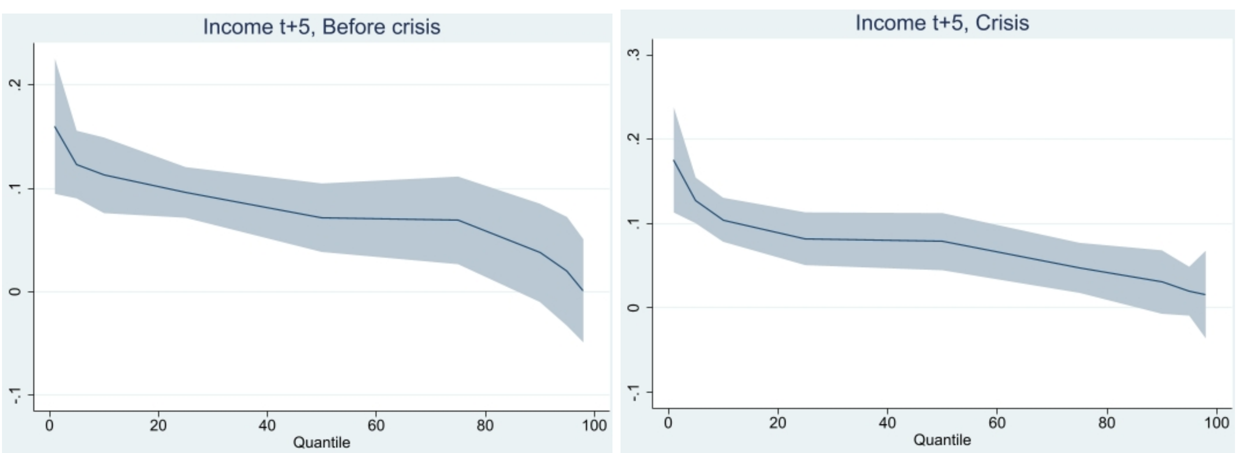
Figure 7. Heterogeneity analysis using simultaneous quantile regressions

Figure show coefficient estimates from the simultaneous quantile regressions of equation (1) along with their 95% confidence interval obtained from: i) the subsamples of low income and high income regions (panel A); ii) the subsamples of the pre-crisis period, 2000-2008, and crisis and post-crisis period, 2009-2016 (panel B); iii) the subsamples of new firms and old firms (panel C); iv) the subsamples of high-growth industries and low-growth industries (panel D); v) the subsamples of male entrepreneurs and female entrepreneurs (panel F). The estimates are for the 1%, 5%, 10%, 25%, 50% (median), 75%, 90%, 95%, and 99% of the distribution of income five years after the credit decision.

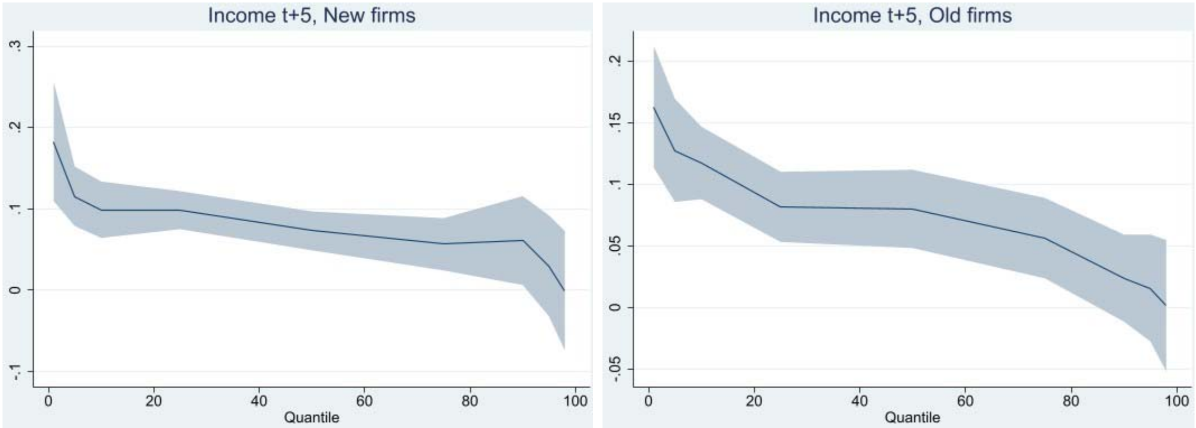
Panel A. Low-income regions vs. high-income regions



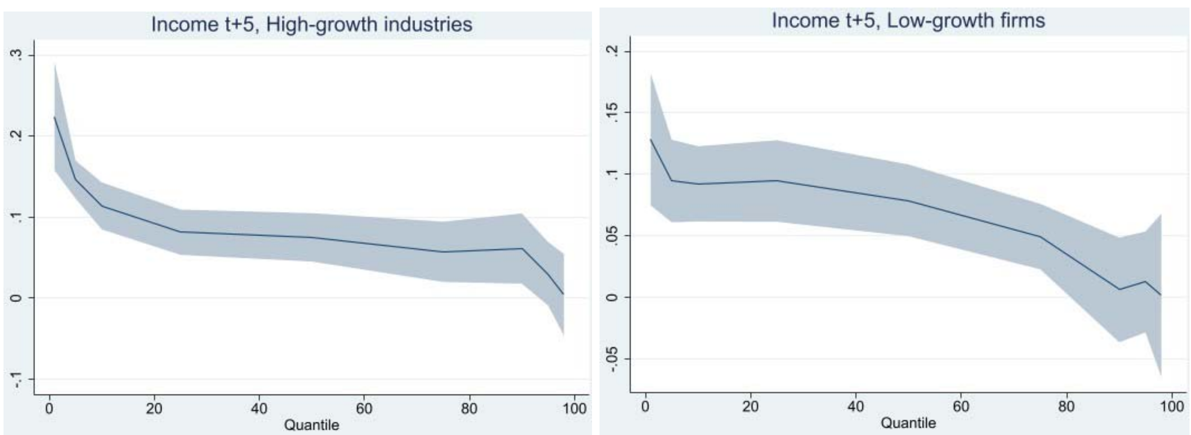
Panel B. Pre-crisis period vs. crisis and post-crisis period



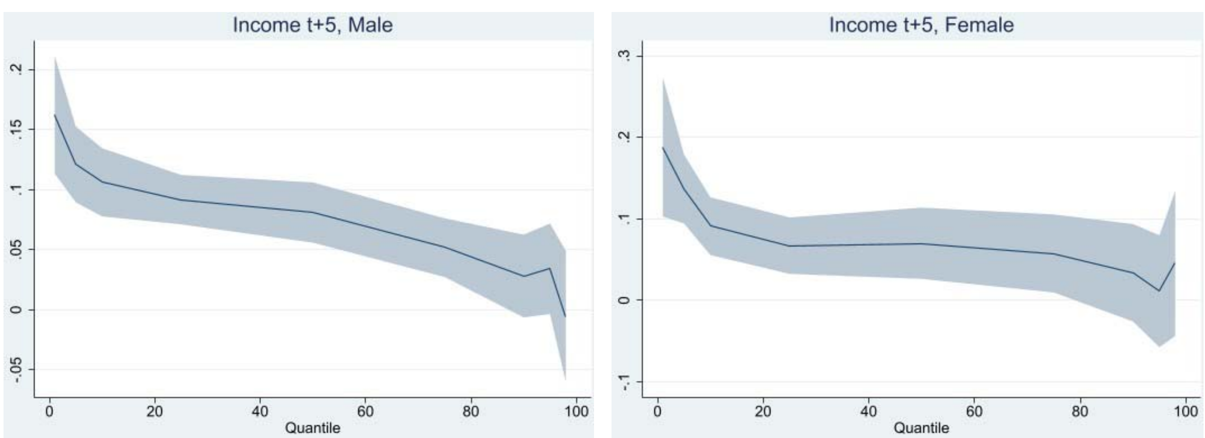
Panel C. New firms vs. old firms



Panel D. High-growth industries vs. Low-growth industries



Panel E. Male entrepreneur vs. female entrepreneur



Online Appendix

This online appendix includes information on our sample's representativeness (Appendix A), and robustness tests on the validity of the RDD (Appendix B).

Appendix A. 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.[22] 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 balance sheet data on small businesses operating in these countries from Bureau van Dijk Orbis.

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

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.[23] 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 very similar, across different dimensions, to small firms located in representative European countries.

[Insert Figure A3 about here]

[23] 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

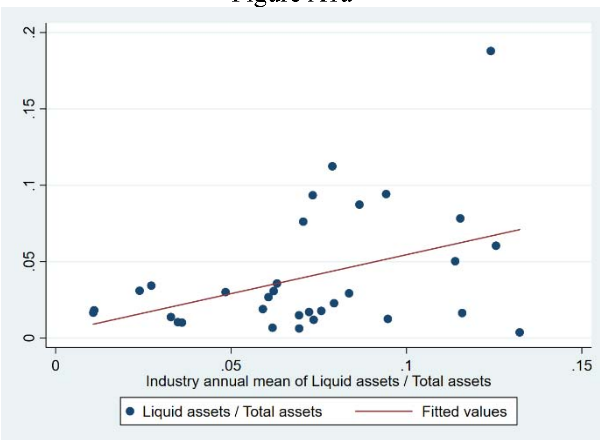


Figure A1b

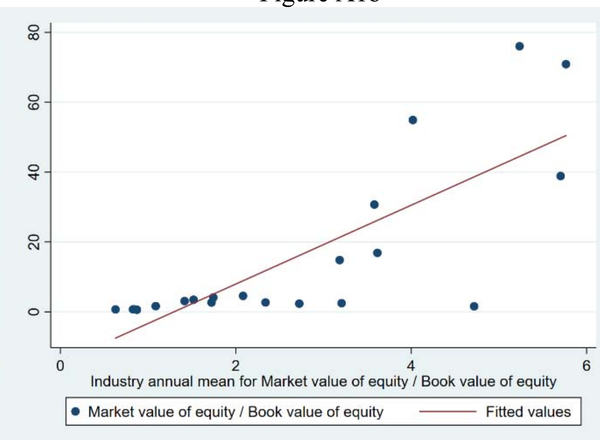


Figure A1c

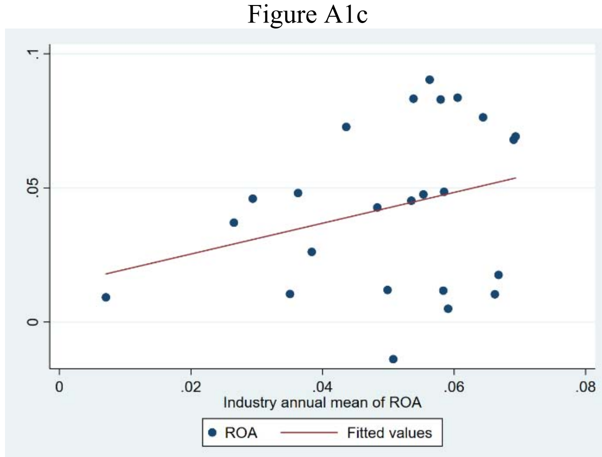


Figure A2
Percent of rejected loans to small and micro firms in the euro area and by our bank

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.

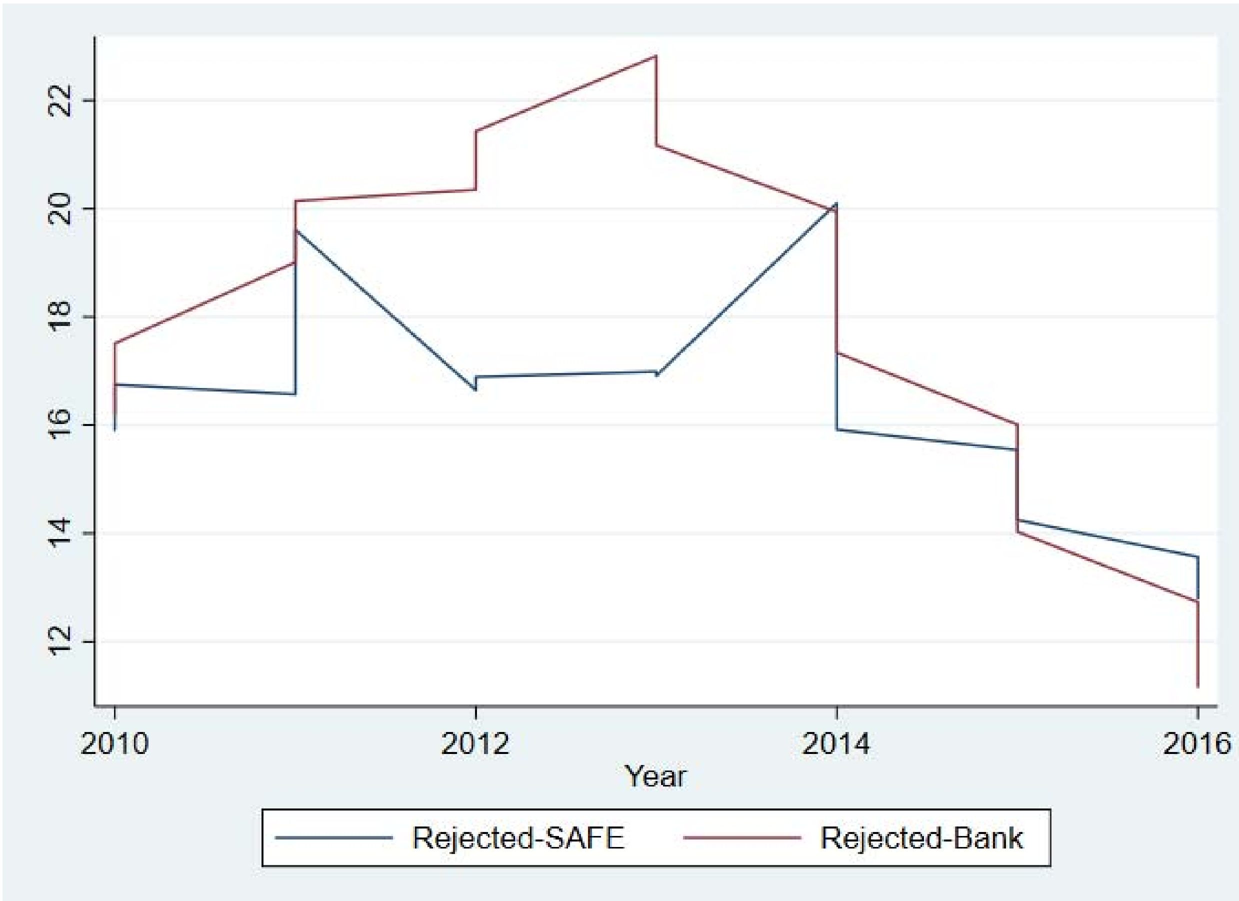


Figure A3
Leverage and ROA of North European small firms versus small firms in our sample

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

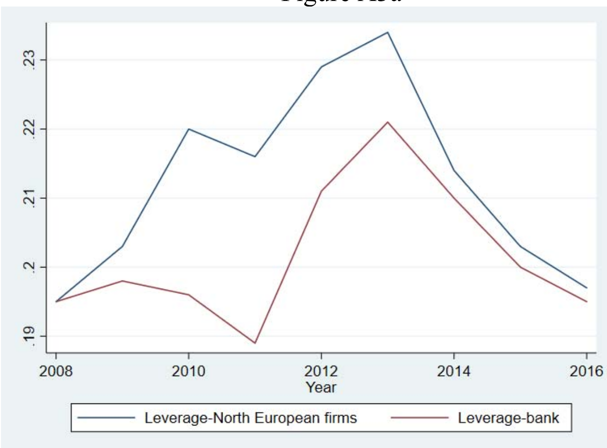
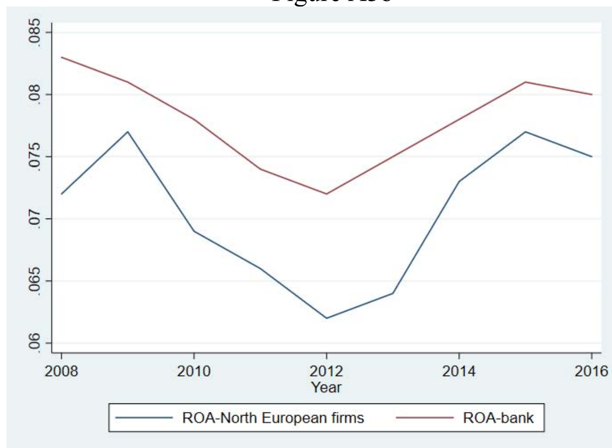


Figure A3b



Appendix B. Robustness of the RDD

In this appendix, we report the results of several robustness tests on the validity of our RDD. First we verify that the relation between the covariates (education, firm size, firm leverage, loan amount, loan maturity, initial income and initial wealth) and the credit score is smooth around the cutoff. Results are in Figure A4.

We next present some robustness tests for the parametric model. The estimates presented in Table A1 and Table A2 show that the results of the parametric RDD of Table 4 are robust to the inclusion in the econometric specification of i) firm industry, loan type and year fixed effects and ii) initial wealth of the entrepreneur, respectively.

Most important, we present robustness tests for the nonparametric RDD model of equation (2). The first test relates to 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 A3 shows that the results presented in Table 4 remain unchanged when using the mean-squared error (MSE) or the common coverage error (CER) bandwidth selector. Also, Figure A5 shows that the significance of Conventional in model (3) of Table 4 is robust to different windows around the cutoff where (small-sample) inference is conducted.[24]

The second test examines 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

[24] Inference in Table 5 is based, instead, on large-sample approximations (Calonico et al., 2014).

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 A4 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 A4 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).

The third test focuses on the lending rate. In nonparametric specifications 4-6 of Table 4 we estimate the effect of credit on income controlling for a wide set of loan, firm, and applicant characteristics, including the requested loan amount and maturity. The lending rate applied on a new loan determines the future stream of payments and, hence, may affect the recipient's future income. Specifically, we would expect that the higher is the credit score of a borrower, the lower is the interest rate applied. Figure A6 shows that the income of accepted applicants considered in the nonparametric RDD one year after the loan decision is a flat function of the lending rate. This means that the interest rate charged on newly granted loans does not influence the effect of loan acceptance on individual income.

While relying on this nonparametric model allows us to restrict our attention to accepted and rejected applicants who are virtually the same in terms of credit quality (as captured by the credit score), we may still wonder if these two groups are perfectly comparable. Specifically, we know that applicants who are rejected are not getting credit elsewhere in the banking system, but we cannot exclude that they may turn to non-bank financing. If this is the case, the estimated treatment effect would carry a bias, as the control group (rejected applicants) would not be a proper counterfactual for the treatment group (accepted applicants). If anything, the bias would be against our results, i.e., leading to an underestimation of the effect of credit on income.

A series of facts suggests that rejected applicants are unlikely to seek credit outside of the banking system. First, no applicant in our sample is always rejected, meaning that applicants who experience a loan denial at some point in time get at least another application accepted during the sample period. On average, more than half of credit applications from these applicants are approved in our sample. Second, given the very limited size (average total assets equals is €369,500), firms in our sample are unable to access capital markets. While other forms of non-bank credit might be available to small and micro firms (e.g., fintech lending), Deutsche Bank (2014), OECD (2014), and BIS and FSB (2017) suggest that reliance of SMEs on funding from the shadow banking sector was very limited during our sample period in Europe. Lastly, non-banks are likely to charge higher interest rates than banks, everything else equal, given their higher cost of capital (Chen et al., 2017).

We, nonetheless, 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 A5 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. We conclude that rejected applicants do not obtain non-bank credit after a loan denial from our bank.

A second concern is that changes in income and income inequality within and across groups of individuals may be influenced by reasons that are independent from the bank's credit decision.

For example, income (income inequality) may increase (decrease) among individuals with a high credit quality irrespective of whether they are the recipients of a loan from this bank or not, e.g., because they can invest their own funds in the firm. 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 Table 4 Panel B by arbitrarily setting the cutoff at the credit score values -1.5, -1, -0.5, 0.5, 1, 1.5. We report the coefficient estimates of *Granted* from these six regressions in Table A6. 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. These individuals are firm owners who do not have a lending relationship with another regulated bank at the time of the loan application, and who apply multiple times during the sample period so that we have information on their income for several years before and after the loan decision. 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, 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 A7 suggest that the two groups are very similar across all attributes. In addition, the analysis presented in the paper shows that small firms in our sample are on average very similar 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 use 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.[25]

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 et al. (2020), but any effect on the future income of male and female entrepreneurs is transmitted via this higher

[25] 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.

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 A8 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.[26] 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 A9, are consistent with those of Table 4.

[26] In a two-stage linear Heckman model we also can correctly adjust the standard errors.

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. 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: 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. The estimation method is OLS on the RDD model of equation (1). The table essentially replicates columns (3) to (6) 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
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. 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 do 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. Obs. is the original number of observations. 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.

	(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
	0.0611***	0.0716***	0.0610***	0.0645***	0.103***	0.0956***
	(0.0127)	(0.0167)	(0.0131)	(0.0178)	(0.0159)	(0.0215)
Observations	57,766	57,766	49,514	49,514	41,391	41,391
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 A4
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. The table essentially replicates columns (3) to (6) of Table 5, 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. Obs. is the original number of observations. 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. The bandwidth selection procedure is the one proposed by Calonico et al. (2014). The bias-corrected RD estimator and the robust variance estimator are obtained according to Calonico et al. (2018) and Calonico at el. (2014), respectively.

Dependent variable	(1) Income t+1	(2) Income t+3	(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)
Observations	36,856	28,604	20,481
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 A5
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 models 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 A6
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. Each specification reports the estimate of the average treatment effect by replicating specification 6 of Table 4 Panel B 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. The bandwidth selection procedure is the one proposed by Calonico et al. (2014). 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)
Observations	57,766	49,514	41,391	53,585	45,333	37,210
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 A7
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 A8
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 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. 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)
Obs.	53,585	45,333	37,210
Controls as in Table 4	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)
Obs.	228,507	228,507	228,507
Clustering	Individual	Individual	Individual

Table A9

Controlling for sample selection in the nonparametric RDD

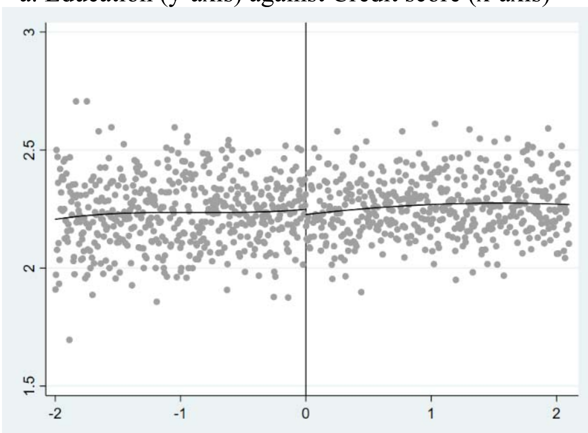
The table reports coefficients and standard errors (in parentheses) from a quasi-two-stage Heckman model. The table essentially replicates the analysis of columns 4-6 of Table 4 Panel B, the difference being the inclusion of the *Mills Ratio* obtained in the first stage regressions of Table A8 as a control 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. 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. Obs. is the original number of observations. 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. The bandwidth selection procedure is the one proposed by Calonico et al. (2014). The bias-corrected RD estimator and the robust variance estimator are obtained according to Calonico et al. (2018) and Calonico at el. (2014), 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)
Observations	53,585	45,333	37,210
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

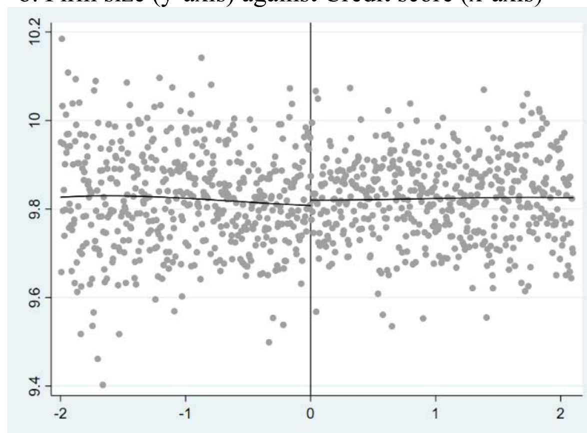
Figure A4
Covariates around the cutoff

The figure reports a plot for each control variable 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.

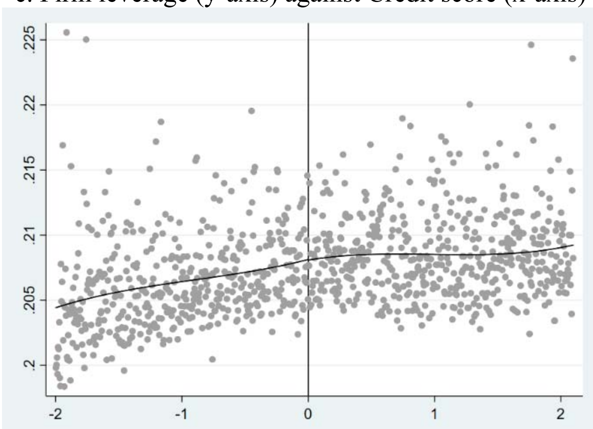
a. Education (y-axis) against Credit score (x-axis)



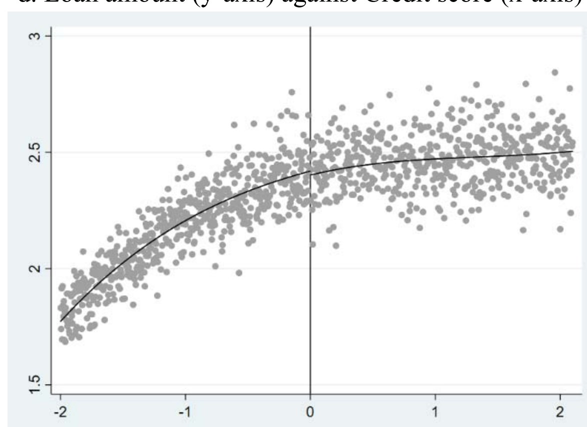
b. Firm size (y-axis) against Credit score (x-axis)



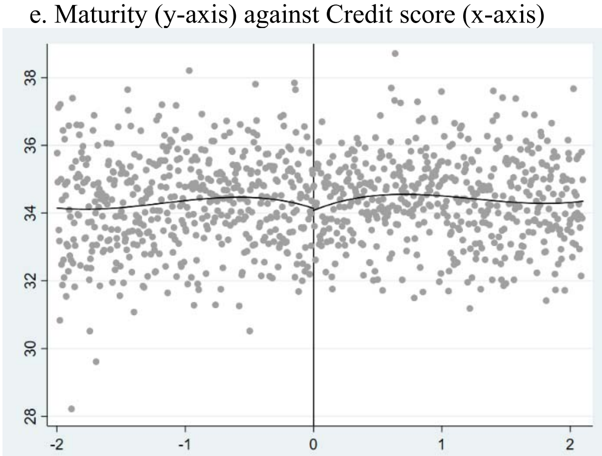
c. Firm leverage (y-axis) against Credit score (x-axis)



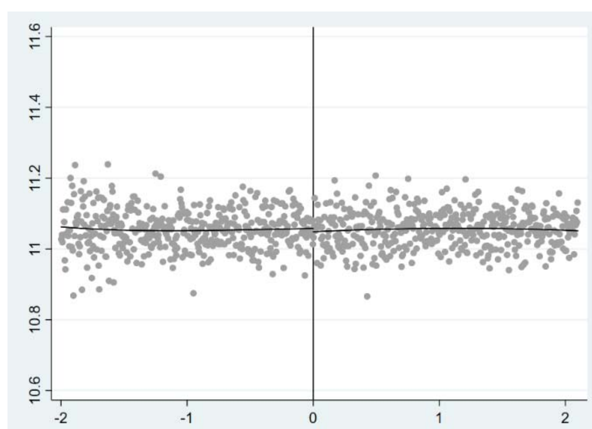
d. Loan amount (y-axis) against Credit score (x-axis)



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)

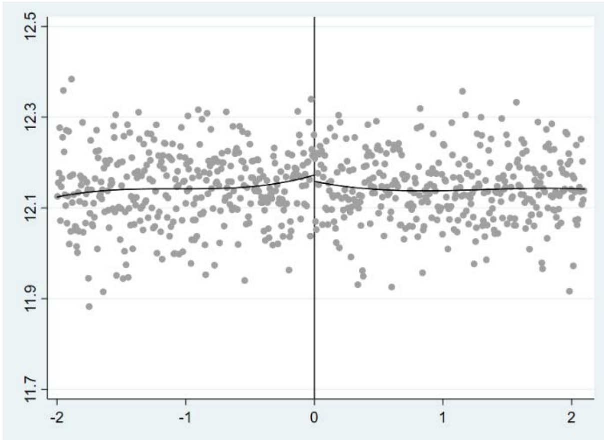

12.5
12.3
12.1
11.9
11.7
-2
-1
0
1
2

Figure A5
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.

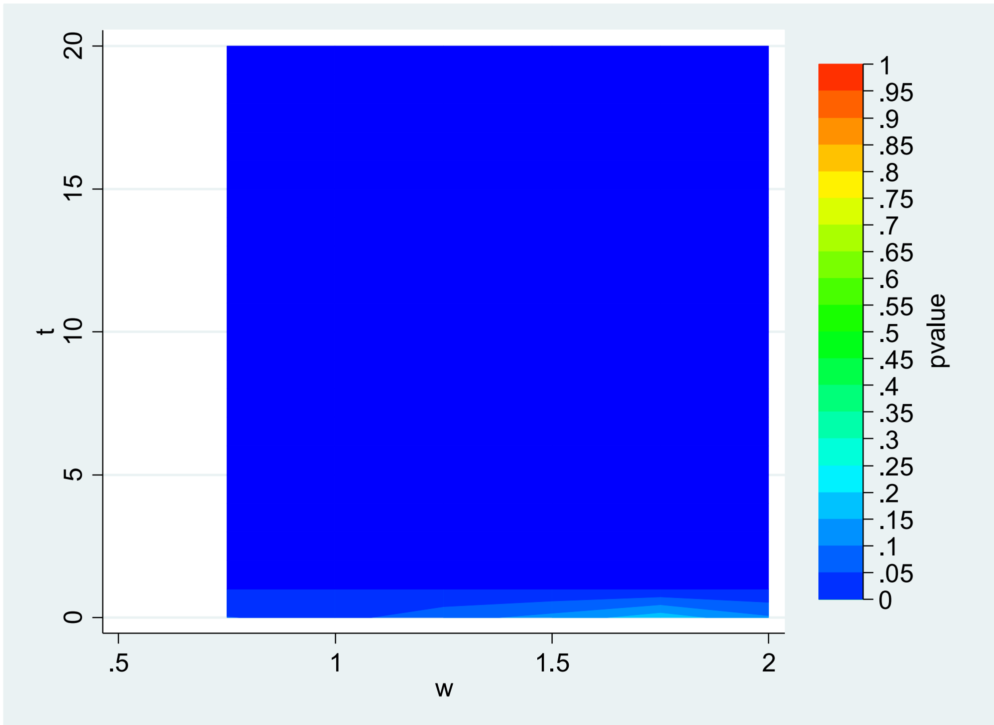


Figure A6
Applicants' income and lending rate around the cutoff

The figure depicts applicants' Income one year after the loan decision (y-axis) against the Lending rate (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 restricted sample where we estimate the nonparametric RDD of Table 5. We select evenly spaced bins that mimic the underlying variability of the data using spacings estimators. The continuous line represents a local polynomial smoother of order zero (i.e. local mean smoother) used to approximate the mean of applicants' income as a function of the lending rate.