Dynamic Incentives in Credit Markets: An Exploration of Repayment Decisions on Digital Credit in Africa*

Stacy Carlson†

January 8th, 2018

Abstract

Many lenders operating in unsecured credit markets utilize dynamic incentives, whereby incentives to repay are generated by promising access to future loans. In this project, I explore the impact of dynamic incentive schemes on borrower behavior in the digital credit market. To do so, I analyze unique data from a digital lender in Africa who relies heavily on dynamic incentives to encourage repayment. I use a series of quasi-experiments induced by policy nonlinearities to estimate the effect of progressive lending policies on borrower repayment decisions. I find that new borrowers who receive a larger initial loan are more likely to default on that loan, consistent with positive moral hazard and repayment burden effects. By contrast, repeat borrowers who receive a larger loan (relative to their previous loan) are actually less likely to default. I provide evidence that this reflects a strategic repayment motive, whereby borrowers repay in order to get access to larger loans in the future. I then write down and estimate a dynamic structural model consistent with my empirical results. I use the estimation results to simulate the profit-maximizing dynamic lending scheme for the lender in this setting.

*I am indebted to my advisors, Robert Townsend, Tavneet Suri, and Abhijit Banerjee. I also benefited from conversations with Frank Schilbach, Nikhil Agarwal, Tite Yokossi, and many other members of the MIT Department of Economics. Numerous seminar participants at the Development Lunch, Development Seminar, and Applied Microeconomics Summer Lunch provided helpful feedback. I am also grateful to the African digital lender for providing the data and making the research possible.

†Massachusetts Institute of Technology, carlson1@mit.edu

1 Introduction

Many lending relationships are dynamic in nature, involving one-period contracts but repeated interactions over time. In such settings, lenders often try to improve repayment performance through the use of dynamic incentives, defined as incentives to repay generated by increased access to future loans. Dynamic incentives encompass two distinct policies, usually used in tandem as a carrot-and-stick strategy: lenders threaten to exclude defaulting borrowers from access to additional credit, and start borrowers at small initial loan sizes but promise progressively larger loans conditional on repayment (called progressive lending). The purpose of dynamic incentives is to mitigate information asymmetries by increasing the opportunity cost of default and enabling screening on early, small loans. At the same time, however, dynamic incentives may have perverse effects by encouraging strategic borrowing behavior, whereby borrowers repay a series of loans until they have worked their way up to a large loan size and then "cash out" by defaulting. Despite the theoretical ambiguity, there is limited empirical work on the impact of dynamic incentives on borrower behavior.[1] More practically, a paramount concern for many lenders remains the determination of an optimal dynamic incentive scheme.

This paper provides novel empirical insight into how borrowers respond to dynamic incentives. I explore this issue using a unique proprietary dataset from a digital credit provider. As new financial technology ("fintech") products such as digital credit have only recently begun to transform consumer finance across the world, this setting is both important and under-studied. Digital credit refers to unsecured loans where credit decisions are made instantaneously based on mobile data and loans are requested, delivered, and repaid electronically. It has proliferated particularly rapidly in developing countries, made possible by mobile technology, improved data analytics, and new delivery channels. Millions of low-income individuals can now access small loans with just a few clicks on a mobile phone. However, digital lenders face unique challenges in encouraging repayment. The small size and remote nature of borrower-lender interactions negate most typical repayment enforcement mechanisms[2] and prevent any in-person screening or the collection of "soft" information. Additionally, credit information sharing is often nascent in the settings in which these lenders operate; even when credit reporting bureaus (CRBs) are present, most digital credit borrowers don't have access to loans from the formal sector, so the threat of a negative report is less impactful. As a result, digital lenders primary recourse for discouraging default is the use of dynamic incentives, making digital credit an ideal setting for studying the design and effectiveness of such schemes.

To do so, I first measure the impact of dynamic incentive policies on borrower repayment decisions using a series of nonlinearities, either across time or across borrowers, in the lender's policies. Taken together, the results imply that borrowers behave strategically when faced with dynamic incentives, and that this behavior depends critically on the stage of the relationship between the borrower and the lender. Next, I write down a simple model of borrower behavior to capture these empirical results. In this model, borrowers vary in the value they place on future loans and are more likely to repay their current loan when they expect to move

[1] Zinman (2014) argues that a major shortcoming in the existing literature "is the lack of empirical evidence on the extent to which dynamic contracting mediates the effects of asymmetric information".

[2] Examples include collateralization, the threat of legal action, group liability, and social sanctions.

up to a larger loan size in the future. The model captures a tradeoff in loan size growth in the loan ladder schedule: if growth is too gradual, the opportunity cost of defaulting today is low, so even those types that most value future loans choose to walk away; conversely, if it is too rapid, then borrowers have an incentive to wait longer to default and then do so at much larger loan sizes. Finally, I use the structure of the model to estimate the distribution of the value placed by borrowers on access to future loans from a given lender. I then use this estimated distribution to determine some features of an optimal dynamic lending scheme for the lender in this setting.

While I focus on dynamic incentives as utilized by digital credit providers, my analysis is more broadly applicable to many other unsecured lending markets. Use of dynamic incentives is particularly common among lenders in developing countries, where most credit is unsecured, credit information sharing is scarce, and legal recourse is more limited. For instance, dynamic incentives are a standard feature of the canonical microfinance contract. It is well documented that most microfinance lenders permanently exclude any defaulters and use progressive lending to reward repayers.[3] However, while recent work has quantified the impact of the other enforcement mechanisms used by microfinance institutions, such as group lending, weekly meetings, and high-frequency repayment, there has been limited empirical focus on the dynamic incentives piece of the puzzle.[4]

This paper also provides novel insight into the fintech sector, and specifically how fintech companies can disrupt the consumer lending industry. The number and financial resources of fintech start-ups has grown exponentially in recent years, by one count increasing from 800 companies to over 2000 over the course of 2015 alone (McKinsey & Co., 2015) and receiving more than 50 billion dollars in total funding between 2012 and 2016 (McKinsey & Co., 2017). The disruptive power of digital lenders stems from their ability to provide loans at scale to a much broader set of borrowers. They achieve this by using data science and machine learning to generate two main advantages relative to traditional lenders: dramatic cost reductions in loan provision and improved risk scoring of borrowers. By reducing overhead costs and focusing on small, short term loans, the amount of capital needed to acquire even a large number of borrowers is relatively low. Machine-learning based risk-scoring technologies utilizing alternative data allow lenders to vastly expand the feasible set of borrowers, to include those with no formal financial history or that face significant geographical barriers in accessing to traditional lenders. Consistent with findings by Björkegren and Grissen (2015), my empirical results document that these risk scores work, in the sense that they are highly predictive of default; this allows digital lenders to maintain sound risk statistics even while expanding access.

I utilize a unique dataset that is particularly rich in both breadth and depth with regards to individual borrowing behavior over time. It contains detailed information on borrowing and repayment decisions and

[3] For instance, Armendáriz and Morduch (2005) show that Grameen Bank provides borrowers in Bangladesh with a continuous sequence of loans that increase quickly in size.

[4] Field and Pande (2008) find no significant difference between repayment rates amoung weekly and monthly repayers. Feigenberg et al. (2013) find that a higher frequency of group meetings leads to stronger social cohesion and group risk-sharing, and thus lower default rates. Field et al. (2013) show that relaxing the requirement that borrowers start repaying their loan immediately leads to higher profits but also higher default rates. Giné and Karlan (2014) find no difference in repayment rates between group liability and individual liability loans. See Banerjee (2013) for additional references.

loan characteristics for 400 thousand borrowers and two million loans. For each borrower, I am able to observe the risk score assigned by the lender's data-based algorithms, the sequence and timing of loans taken out, subsequent repayment decisions, and various demographic characteristics. On the lender's side, I have full information on the structure of the loan ladder and and all relevant credit policies.

During the sample period that I focus on in this paper,[5] the lender relies primarily on both forms of dynamic incentives to enforce repayment. All borrowers start out at a small initial loan size of $10 or less. Borrowers who repay a sufficient number of installments of their current loan on time proceed up the loan ladder to a larger loan, to a maximum possible loan size of $500. Borrowers who default are blocked permanently from any future credit from this lender. Critically for my empirical strategy, the lender frequently experiments with its progressive lending policies - both the structure of the loan ladder and how borrowers progress along it.

My empirical analysis relies on a series of natural experiments induced by certain aspects of the lender's policies, as well as temporary, unannounced policy changes made by the lender. I use this quasi-experimental variation to estimate the causal impact of changing the two key components of the progressive lending scheme: the initial loan size and the loan ladder slope. Three key results emerge. First, I find that giving a larger initial loan to new borrowers leads to a higher default rate on that loan. This is consistent with positive moral hazard and repayment burden effects, as documented elsewhere in the literature.[6] By contrast, I find that giving a larger loan to repeat borrowers - which amounts to making the loan ladder steeper - results in a *lower* default rate on that loan. I provide evidence that the explanatory mechanism is a strategic repayment motive: borrowers are more likely to repay when they expect to receive a larger loan next time, and a larger current loan increases their expectations of future loan growth and thus increases the value of repaying their current loan. It also suggests that the stage of dynamic relationship matters when considering the impact of loan size on repayment. Finally, a steeper loan ladder, while reducing default rates on early loans, actually results in higher borrower-level default rates. This suggests that progressive lending doesn't mitigate overall default risk but does cause borrowers to shift the timing of when they default.

I next write down a dynamic model of my setting that captures these empirical results. The model has three key elements. First, the lender relies on progressive lending and full threat of exclusion of defaulters to incentivize repayment. Second, borrowers vary in the value they place on access to future loans from the lender, and thus how responsive to dynamic incentives they are. Third, borrowers vary in credit needs. Each borrower has a maximum number of loans that they need (or equivalently, that they can afford to repay), but do not know this cap until they receive a signal that they have reached it. This generates an additional reason for borrowers to default. Neither of these borrower characteristics is directly observable by the lender.

This model generates several key predictions. First, it can be shown that an optimal loan ladder exists. This reflects a tradeoff in loan size growth, or equivalently, the slope of the loan ladder. If loan growth is too

[5] The sample period is defined in Section 3.3.

[6] Examples include Karlan and Zinman (2009) in the South African consumer credit market and Adams et al. (2009) in the US sub-prime auto loan market.

gradual (the loan ladder is relatively flat), then the opportunity cost of defaulting today is relatively low, so even borrowers that most value future loans may default. In the most extreme case, everyone defaults early on, so the lender makes losses and has no customers remaining. Conversely, if loan growth is too quick (the loan ladder is relatively steep), borrowers wait a long time to default, and then to do so at much larger loan sizes. Second, the model generates complementary explanations for the different observed default outcomes when borrowers receive a larger initial loan versus a larger repeat loan (which is equivalent to facing a steeper loan ladder). The first-order cause relates to selection, such that the set of new and repeat borrowers differ. A second-order explanation is that the information set of borrowers differs between the first and subsequent loans. New borrowers have no direct information yet about the future path of loan sizes, whereas repeat borrowers in the model use past loan growth to form expectations about future loan growth.

With a few additional assumptions, I am able to estimate the model structurally in order to simulate the profit-maximizing dynamic incentive scheme from the lender's perspective. Specifically, I estimate the distribution of borrowers' net outside options and the probability of reaching their maximum loan cycle in each period. I then show that the lender minimizes losses on the first loan by offering a smaller loan size, but maximizes overall profits with a combination of a larger initial loan and a moderate loan ladder slope.

The rest of the paper is organized as follows. Section 2 summarizes the relevant related literature. Section 3 introduces the setting and discusses the data used in the empirical analysis. Sections 4 and 5 discuss the empirical analyses of the impact of changing the initial loan size and loan ladder progression policies, respectively. Section 6 presents a dynamic discrete choice model of borrower behavior in this setting. Section 7 outlines the structural estimation exercise, and Section 8 uses the results of this exercise to frame the empirical results and simulate an optimal progressive lending scheme for the lender in this setting. Section 9 concludes.

2 Related Literature

This paper most directly contributes to the predominantly theoretical literature on dynamic incentives. The notion that unsecured debt can be self-enforced in the case of repeated interactions was first formalized by Eaton and Gersovitz (1981) in the context of sovereign debt markets, sparking a robust body of research that further developed the idea. Besley (1995) and Morduch (1999) highlight the importance of dynamic incentives in credit markets in developing countries. Hulme and Mosley (1996) and Armendáriz and Morduch (2005) model dynamic incentives in the microfinance context. Tedeschi (2006) derives conditions under which the optimal period of exclusion after strategic default by a microfinance client is less than infinite. Another line of research, including Karaivanov and Townsend (2014) and Albuquerque and Hopenhayn (2004), consider dynamic lending relationships more broadly, testing for different types of dynamic frictions and deriving optimal contracts in such settings.

Focusing on sovereign debt, Bulow and Rogoff (1989) highlight an important limitation of dynamic incentives. They show that the threat of exclusion itself is not sufficient to prevent default, because a borrower can save the money she would have repaid to the lender and instead use it in place of obtaining another loan. For dynamic incentives to work, it must be the case that the rate of loan growth exceeds the interest rate on savings, or put another way, the borrower is not able to scale up their own resources faster than the lender can. Bond and Krishnamurthy (2004) explore in more detail the assumptions required to give dynamic incentives teeth.

The two most closely related theoretical papers are Ghosh and Ray (2016) and Shapiro (2015). Both focus on dynamic incentives in a competitive lending market; in Ghosh and Ray, the pool of credit providers is composed of informal moneylenders, whereas in Shapiro they are microfinance providers. Each starts with exclusion of defaulters as a primitive, and shows that the efficient equilibrium involves progressive lending. Shapiro then discusses the impact of borrowing from two lenders simultaneously, while Ghosh and Ray show the implications of enabling information sharing between lenders. Unlike this paper, Shapiro assumes lenders correctly anticipate a borrower's probability of repayment in equilibrium and set loan terms accordingly, whereas I assume heterogeneity in types across borrowers that is never observable by the lender; Ghosh and Ray only differentiate between first time loans and repeat loans, and thus don't have anything to say about the optimal design of a more realistic progressive lending scheme. Additionally, both papers focus only on the theory and do not attempt any empirical exercises.

The main contribution of this paper is bridging the gap between theory and empirics by estimating how borrowers respond to dynamic incentives. I am able to quantify directly for the first time the impact of variation in progressive lending schemes on borrower behavior, and in doing so show that how borrowers respond to such incentive schemes varies in interesting ways depending on the stage of their relationship with the lender. Empirical work on dynamic incentives remains thin, although there are some recent applications in various development contexts. Most closely related to my setting, Karlan and Zinman (2009) find in an experiment with repeat borrowers in South Africa that the promise of a subsidized future loan improves repayment on a borrower's current loan, which they characterize as evidence of positive moral hazard. Giné et al. (2010) show that the threat of credit denial reduces both risk-taking and default in a game setting. Giné et al. (2012) use a fingerprinting intervention to randomize a lender's ability to implement dynamic incentives among farmers in Malawi, and find that the threat of exclusion improves repayment only among the riskiest types. Breza (2013) demonstrates that concerns over microfinance lender viability (and thus over whether future loans will be accessible from that lender) in India leads to higher default.

By analyzing the effect of plausibly exogenous increases in loan size on repayment, this paper also contributes to a broader empirical literature on moral hazard in consumer credit. My results add to this literature by explicitly testing for the effect of a larger loan size across borrowers in different stages in their relationship with the same lender, thus allowing me to tease out possible heterogeneity in the direction of moral hazard that was previously unexplored. The vast majority of existing research finds evidence of positive moral hazard, similar to my findings for first-time borrowers in Section 4. In the same study mentioned above, Karlan and Zinman (2009) find that offering a 100 basis point discount on the next loan in the South African medium-term consumer credit market leads to a 13 to 21 percent decrease in default. Adams et al. (2009) study the sub-prime car loan market in the United States, and find that a $1,000 increase in loan size increases the probability of default by 16 percent. De Giorgi et al. (2015) focus on the Mexican credit card market and find that receiving an additional credit card, which translates into an additional $1,000 in available credit, leads to a 34% increase in default. Consistent with my findings of a negative moral hazard effect among repeat borrowers in Section 5, Dobbie and Skiba (2013) show that a $50 increase in loan size lowers the probability that a payday borrower defaults by 17 to 33 percent. While they do not explicitly differentiate between new and repeat borrowers, most payday borrowers take out a sequence of loans similar to the borrowers in my sample. One hypothesis that they put forward to explain their results, but do not explicitly test, is similar to the story I focus on here: specifically, that borrowers repay larger loans in order to maintain a larger credit line in the future.

3 Setting and Data

3.1 Empirical Setting

I study dynamic incentives as utilized by a digital credit provider in Africa. Digital credit refers to unsecured loans where credit decisions are made instantaneously based on mobile data and loans are requested, delivered, and repaid electronically. This type of credit is relatively new but growing rapidly, enabled by the rapid spread of mobile phones and improved big data processing tools. In emerging markets, there are over 20 active digital lenders reaching an estimated total of 24 million users, six of which have over one million users (CGAP, 2017). While digital credit is now available in many regions, it is particularly prevalent in Sub-Saharan Africa due to the region's high mobile money penetration.

Conditions are in place for digital credit to continue to expand rapidly over the years to come, making this study increasingly relevant. A key input for digital credit is the existence of a robust and widely-utilized mobile money network. As documented by Suri (2017), mobile money has spread like wildfire across the developing world. By the end of 2015, 271 mobile money services were available in 93 different countries, with an additional 110 in the works. At this date, there were 411 million registered mobile money accounts globally, of which 134 million had been utilized within the past 90 days.

Digital credit is characterized by the following features. First, loan applicant eligibility depends on owning a phone (in some cases, a smart phone), but applicants don't need a formal bank account, any credit history, or any formal proof of income. Loan decisions are automated and depend on non-traditional digital data that are scraped automatically from the phone. Loans tend to be smaller, shorter-term, and costlier than traditional consumer loans. Finally, all interactions between borrowers and lenders are done remotely. Borrowers don't need to live near or travel to a bank to apply, receive a loan, make repayments, or engage with customer service representatives.

Reflecting these features, digital credit is typically used for different purposes than formal bank loans or microcredit. Small, instantly-delivered loans are ideal for filling short-term household liquidity gaps or meeting working capital needs for businesses, rather than financing large-scale asset purchases or starting a business. A recent survey in Kenya, which has the highest penetration of digital credit, confirmed this, finding that mobile loans are the primary source of credit used for both day-to-day needs and for emergencies (CGAP, 2017b).

Digital lenders rely heavily on dynamic incentives. This reflects a few considerations. First, average loan sizes are generally too small for it to make financial sense to pursue legal action against defaulters. Second, because of the remote nature of interactions, lenders don't have any sort of personal relationship with borrowers and often aren't located anywhere near them, making many other punishment techniques infeasible. Finally, they predominantly operate in settings with only nascent credit reporting bureaus (CRBs) and little credit information sharing among lenders, so threats to report defaulters in this sense are less effective. And even in countries where CRBs are more robust, most digital credit borrowers don't also have access to formal credit, thus reducing the power of the reporting threat.

3.2 Lender

The data for this study comes from a digital lender active in Africa since 2015. Like other digital lenders, the lender uses data science and machine learning algorithms to make credit decisions. This lender uses a smart phone application ("app") platform as an interface, thus restricting the sample of potential borrowers to those with access to a smart phone. However, given recent trends, this does not represent a significant restriction.[7] The country's primary mobile money system provides the rails for all loan disbursements and repayments.

To apply for a loan, individuals must first download the app, fill in some basic identification and KYC[8] information, read and agree to confidentiality and data-sharing disclosures, and select a loan product for which to apply. At this point, additional data is gathered from the phone automatically, and an internal risk score is generated for - but not shared with - the applicant. Alternative digital data sources include call data records (CDRs), mobile money transactions, information from social media accounts, social network connections, and various types of text notifications. These data provide a richer picture of the individual than is available to traditional lenders. If the applicant's risk score is above the relevant threshold and the applicant passes a number of fraud checks, she is notified of approval and the loan amount is disbursed immediately into her mobile money account. She can then repay the loan by transferring mobile money funds back to the lender. She receives text message reminders prior to each repayment date, as well as several late payment notifications in the case of a missed repayment. She can also check the app to view the remaining balance and additional loan details at any time.

[7] According to the Consumer Barometer survey by Google, smart phone penetration in the relevant country was close to 50% by 2016, relative to 90% total mobile penetration.

[8] "Know Your Customer" information required by regulation.

The lender relies primarily on both forms of dynamic incentives to encourage repayment. During the sample period that I focus on in my empirical analysis, the dynamic incentive scheme took the following form. All borrowers start out at a small initial loan size. Borrowers who repay a sufficient number of installments of their current loan on time proceed up the loan ladder to a larger loan. Borrowers who repay their current loan but with an insufficient on-time payment record either remain on the same rung or move down (depending on the degree of lateness). Borrowers who default are blocked from the system permanently. The lender provides no specific information about the loan ladder or its credit policies to borrowers, beyond general statements that on-time repayment is necessary to qualify for larger loans in the future. Critically for my empirical strategy, the lender frequently experiments with its progressive lending policies - both the structure of the loan ladder and how borrowers progress along it - over the sample period.

3.3 Data

My dataset contains detailed information on application, borrowing, and repayment decisions as well as loan characteristics for all loans made between March 2015 and September 2017. I focus my analysis on the subset of loans issued through July 2016, which equates to 85 thousand borrowers and 340 thousand loans.[9] All summary statistics discussed refer only to this subsample of borrowers and loans. For each borrower, I observe various demographic, financial, locational, and network characteristics obtained when they download and register for the mobile app. After they apply for a loan, I additionally observe the risk score assigned by the lender's data-based algorithms, as well as the risk score inputs. For successful applicants, I see the sequence and timing of loans taken out and all subsequent repayment decisions. On the lender's side, I have full information on loan parameters, the structure of the loan ladder, and and all relevant credit policies. Key loan parameters include loan size, repayment frequency, loan term, and the interest rate.

3.4 Descriptive Statistics

3.4.1 Loan Parameters

The mobile loans offered by the lender are small, short term, and relatively expensive. Table 1 displays summary statistics on key contract terms, split by initial versus repeat loans. The median loan to a first-time borrower is \$10, has an 8% flat fee attached, and is meant to be repaid in three weekly installments. The median loan to a repeat borrower is \$20, comes with an 11% fee attached, and is repaid in four weekly installments. Overall, loan sizes range from \$2.50 to \$400.[10] Most loans have a 28-day maturity. Interest is charged as a flat fee on top of the loan principal, which ranges between 5% and 17%. There are no other fees associated with the loans; most notably, there are no late fees and interest is not compounded on loans

[9] This decision is predicated on three considerations. First, there was a technology upgrade made to the lender's app in August 2016 that complicates the comparison of outcomes before and after this change and also makes the earlier period preferable for my purposes. Second, the lender did not want statistics on its full portfolio shared. Third, this means I have at least a year of time to observe borrowing patterns for these borrowers, which is sufficient to get data on the longer-term outcomes I consider in my empirical analyses.

[10] No borrower in the sample has reached the maximum offered loan size of \$500.

that are repaid after the final due date. The high fees and short maturities translate into high APRs, which range from 19% to 334%.

3.4.2 Borrowers

The lender does not require borrowers to provide much self-reported information loan application stage, instead preferring to rely on hard data scraped from the mobile phone. However, the lender has conducted several third-party surveys in order to learn more about the composition of its customer base. They do collect information on the age and gender of all individuals who download the app; the prototypical borrower is relatively young, at around 30 years of age, and male, with men comprising over two-thirds of the borrowing pool. Table 2 displays additional information gathered from surveys on borrower location, education, income, income security, occupation, business ownership, other credit sources, and reported loan use. 45% of borrowers live in urban settings. Most borrowers are very well-educated: over 50% have graduated from university, 95% have graduated from secondary school or higher, and nearly all have at least graduated from primary school. About a third of surveyed borrowers are still in school, either full- or part-time. Average self-reported monthly income is around $420, implying annual income of $5,400, which is much higher than average income per capita in the relevant country. Only about 24% report having sufficient savings such that they would be able to last at least six months if they were to lose their main source of income. Half of surveyed borrowers report having a salaried job, with slightly more reporting that they own a business.[11] Many borrowers report owning more than one business.

Nearly all surveyed borrowers report having access to other credit sources; three-quarters utilize other forms of digital credit, while only 23% report having access to formal bank loans. Most borrowers (over 70%) report using their loan for consumption smoothing purposes, such as for emergencies and unexpected expenses. Slightly less than half report using their loan for business purposes. Of those who do use the loan to invest in a business, only about one-quarter state that it was to start a new business, while the rest indicate using it for an existing business.

3.4.3 Borrowing Patterns

Most borrowers take out a sequence of loans in relatively quick succession. Tables 3 and 4 display various statistics about how borrowers utilize this lender at the borrower and loan level, respectively. Three-quarters of borrowers have taken out more than one loan. The median borrower has taken out five loans total over the course of 5.5 months. The median number of loans per month is 1.1. The distribution of the number of loans per borrower has a long right tail, with 23% of borrowers having taken out ten or more loans and three borrowers having taken out more than 50 loans. Borrowers usually repay loans early: 82% of loans that are repaid are done so before the final scheduled repayment date, and the median loan is repaid two full weeks early. Borrowers also tend to take out another loan very quickly after repaying their last loan:

[11] Note that these are not mutually exclusive - many salaried workers report also running a business on the side.

the median time between loans is zero; in other words, the median borrower took out another loan the same day as repaying her previous loan.

Most borrowers continue taking out loans until defaulting. About 14% of the sample of borrowers has an outstanding loan with the lender at the end of the sample period. Of borrowers without an outstanding loan currently, around half have defaulted, with the remaining evenly split between those who chose not to reapply for another loan and those who reapplied and were rejected.

Figure 1 depicts some key moments related to borrowing behavior delineated by loan cycle (i.e., how many loans a borrower has taken out thus far). As shown in Panel A, 75% of borrowers have taken out at least two loans, 50% have taken out at least five loans, and 23% have taken out ten or more loans. Panel B depicts that, conditional on continuing to borrow, it takes the average borrower four weeks to reach their second loan, 15 weeks to reach their fifth loan, and 35 weeks to reach their tenth loan. Panel C shows that average loan sizes climb from under $10 on the first loan to $25 on the fifth and $45 by the tenth.

3.4.4 Default Patterns

Default rates are relatively low at the loan level, but many borrowers end up defaulting and thus being excluded from any future loans from this lender. About 9% of loans have ended in default, while 50% of borrowers during the sample period default eventually. Of loans that end in default, 76% have no partial repayments made. This suggests that many defaulters decide to walk away with the entire loan principal, rather than that they default because they struggle to keep up with the repayment schedule.

Average default rates are nonlinear in loan cycle position, as shown in Figure 1, Panel D. Default is much more likely on the first loan: first-time borrowers during the sample period default about 14% of the time, while the repeat borrower default rate is only about 6%. The default rate drops significantly on the second loan relative to the first loan, then rises through the next several loan cycle positions before falling steadily thereafter.

One key predictor of a borrower's eventual default outcome during this period is how quickly she takes out and repays loans. The borrower-level default rate for borrowers who, on average, take out their next loan within one day of repaying their prior loan is 72%, relative to only 34% for borrowers who take loans out more slowly. Figure 2 depicts this strong negative relationship between the probability of default and the speed with which a borrower moves through loans from this lender. I use this fact in attempting to disentangle mechanisms in my empirical analysis in Section 5 of repeat borrower responsiveness to dynamic incentives.

4 Empirical Results: Initial Loan Size

I use quasi-experimental variation in the lender's progressive lending policies - holding constant their policy of full exclusion of defaulters - to estimate the causal impact of dynamic incentives on borrower behavior. I focus on two key parameters of the lender's loan ladder: the starting loan size for new borrowers, and the rate of growth in loan size as borrowers progress up the ladder. In both cases, I focus on how exogenous changes to each parameter impact outcomes.

The initial loan size is an important piece of the progressive lending scheme from the lender's perspective for three main reasons. First, the initial loan can be viewed as a customer acquisition tool. A larger first-time loan will allow it to attract more borrowers in a competitive setting. Second, to the extent that it influences expectations of future loan sizes, as posited in the model in Section 6, a higher starting base also strengthens dynamic repayment incentives for both the first loan and future loans. Finally, the initial loan serves as an additional, cheap screening device, whereby the lender can sort out unobservably bad types before they reach larger loan sizes. A smaller first-time loan will reduce the cost of this screening strategy. Taken together, these considerations generate a tradeoff for the lender in deciding on an initial loan size. Ultimately, how borrowers adjust repayment behavior in response to the initial loan size determines how the lender manages this tradeoff. In the analysis presented below, I find that a larger initial loan leads to higher default rates on the current loan, but to no change in overall (borrower-level) default rates. Therefore, consistent with the results in Section 5, progressive lending policies appear to influence the timing of default, but don't reduce the overall probability of default.

4.1 Empirical Strategy

In order to estimate the impact of the initial loan size on default, I use a regression discontinuity (RD) approach that exploits discontinuities in the relationship between risk score and loan size. In late 2015 and early 2016, first-time borrowers were sorted into initial loan sizes by their internal risk score. This process worked as follows. All potential new borrowers first applied for an initial loan of \$10. Applicants with risk scores above a fixed threshold were approved, while applicants with risk scores below the threshold were rejected but could reapply for a smaller loan. This sorting of new borrowers into different initial loan products thus created discontinuous jumps in loan size across individuals with very similar risk scores around the \$10 threshold.

I use this new borrower sorting policy to estimate the impact of a larger initial loan size using a fuzzy RD approach with the risk score as the running variable.[12] The intuition behind my identification strategy is the following. A borrower's risk score is continuously (negatively) related to her probability of default. Being above the risk score threshold for the \$10 loan increases the probability of receiving the larger loan relative to the smaller loan. Any jump in the default rate at this threshold is thus plausibly due to the change in loan size.

Formally, consider the following model of the causal relationship between default on the first loan D_1 and initial loan size L_1 for individual i:

[12]I only need to use fuzzy RD because there are a small number of cases in which the sorting algorithm malfunctioned, such that some borrowers do not receive the correct loan size. Specifically, 0.4% of borrowers mistakenly received a \$10 loan despite not having a sufficient credit score, and 0.5% of borrowers who should have received a \$10 loan instead got a \$5 loan.

$$D_{1,i} = \alpha + \beta \cdot L_{1,i} + \varepsilon_i \tag{1}$$

I focus on default on the initial loan, but also present results on longer term credit outcomes. The parameter of interest is β, which measures the causal impact of loan size on default. However, loan size is determined by the risk score, which is highly correlated by design with the probability of default: larger loans were given to borrowers who were predicted to be lower risk. Therefore, direct estimation of β by OLS will be biased. Borrowers who received the larger loan should be less likely to default than those who received a smaller loan in the event that they had both received a loan of the same size: $E[\varepsilon_i|L_{1,i}] < 0$.

The key identifying assumption behind the fuzzy RD approach is that the risk score is continuously related to default probabilities, and in particular does not jump discontinuously at the threshold for determining eligibility for the larger loan. This requires that all other unobserved determinants of default are continuously related to the risk score, or more formally, that the distribution of unobserved determinants of default conditional on risk score, $E[\varepsilon_i|score_i]$, trends smoothly through the loan size-eligibility threshold used by the lender:

$$E[\varepsilon_i|score_i - cutoff_i = \triangle]_{\triangle \to 0^+} = E[\varepsilon_i|score_i - cutoff_i = \triangle]_{\triangle \to 0^-} \tag{2}$$

where $score_i$ is individual i's risk score and $cutoff_i$ is the loan-eligibility threshold for the larger loan at the time of individual i's initial application. For ease of presentation, I consider the score relative to the cutoff. Because loan size is a discontinuous function of the risk score at the threshold whereas the conditional distribution of all unobservable determinants of default, ε_i, is continuous at the threshold, I can now identify β. Any change in the default rate right at the threshold can be attributed to the receipt of a larger loan.

I use two complementary strategies to estimate β, the causal impact of initial loan size on default. First, I use an instrumental variables (IV) approach with global risk score controls. I instrument for loan size using a binary indicator of having a risk score greater than the cutoff. This implies the following semi-parametric first-stage and reduced-form equations:

$$L_{1,i} = f(score_i - cutoff_i) + \delta \cdot \mathbf{1}(score_i > cutoff_i) + u_i \tag{3}$$

$$D_{1,i} = g(score_i - cutoff_i) + \lambda \cdot \mathbf{1}(score_i > cutoff_i) + \eta_i \tag{4}$$

Here, δ and λ represent the impact on loan size and default, respectively, of having a risk score that clears the \$10 loan eligibility threshold. I present results using both a third-order polynomial functional form for $f(\cdot)$ and $g(\cdot)$, as well as a linear specification. I include interaction terms between each power of the score (relative to the cutoff) and the indicator of being above the threshold, thus allowing the slope to vary on either side of the threshold. I also include weekly loan cohort fixed effects to control for any time trends.

Second, as suggested by Gelman and Imbens (2014) and Calonico et al. (2014), among others, I use a non-parametric local polynomial approach. This involves zooming in on data in a neighborhood around the cutoff, which reduces the dependency of the estimate on the correct functional form assumption. In practice, this procedure computes weighted averages of $D_{1,i}$ and $L_{1,i}$ on either side of the risk score threshold, where points closer to the threshold receive larger weights, and points farther from the threshold receive lower weights. Two key decisions in implementing this approach are the order of the polynomial and the bandwidth to use. Following Imbens and Kalyanaraman (2012), I use a local linear specification and the mean squared error (MSE) optimal bandwidth, where I allow the bandwidth to be different on either side of the cutoff. For implementation, I use the tools developed by Calonico et al. (2014) and expanded by Calonico et al. (2017).

There are a few main threats to the causal interpretation of my results. The first concerns the possibility of selection. Adverse (or advantageous) selection, whereby worse (better) borrowers systematically choose larger loans, is not technically possible in my setting. Everyone chose to apply for the same initial loan of \$10, and thus was willing to borrow ("selected in") at the same initial terms. Additionally, borrowers were not able to choose any other loan size besides the one they were sorted into according to their risk score. However, it is possible for applicants to opt out of borrowing upon discovering they are not eligible for the larger loan. Such selective borrowing would be problematic in my setting, as in any regression discontinuity analysis, because it could create discontinuous differences in unobservable borrower characteristics at the risk score threshold that could also be correlated with the probability of default. I address this in Section A.1 in two ways. First, I compute the percentage of applicants who chose not to reapply despite being eligible for the smaller loan, and find it to be too small to explain my results. Second, I show that the distribution of risk scores for those who chose to reapply for the smaller loan versus those who didn't are very similar.

Another potential issue is the possibility of manipulation of the risk score by borrowers. If individuals could influence the risk score in such a way as to ensure they were just above the threshold and thus eligible for the larger loan, this would again create a discontinuity in unobservable borrower characteristics at the threshold. Borrowers with sufficient knowledge or ability to influence the score would be more likely to end up just above the threshold, while borrowers with less information or capacity to impact the score would be more likely to fall below it. As a result, the two groups would not provide proper counterfactuals for one another, undermining the identification strategy. This type of manipulation is highly unlikely in my setting, as the risk score is made up of many components selected via a machine-learning algorithm, and neither the components nor the risk score itself are disclosed to the applicants. Nevertheless, I provide both graphical and test-based evidence in Section A.1 to assuage any concerns.

A final potential problem is that my regression discontinuity design is misspecified. To ensure that my results identify real discontinuities caused by the change in loan size receipt, I replicate my empirical analysis using data from two alternative periods. First, I analyze a later period during which the lender eliminated the initial loan size sorting policy. During this period, nearly all borrowers took out \$10 loans, so there is no longer a discontinuity in loan size at the risk score threshold. Reassuringly, I also find that the discontinuity in default disappears. Second, I replicate my analysis using data from another later period during which the lender reintroduced the initial loan size sorting policy, and obtain very similar estimates to my baseline analysis.

4.2 Impact of Loan Eligibility on Initial Loan Size

The first-stage analysis provides evidence that the lender closely followed its policy of sorting borrowers into initial loan sizes according to their risk score. Each plot in Figure 3 shows average loan sizes across risk score bins with a width of .01 units.[13] Panel A uses a third-order polynomial to control for the risk score, allowing the shape to vary on either side of the threshold:

$$L_{1,i} = \gamma_0 + \sum_{p=1}^{3} \gamma_{1p} \cdot (score_i - cutoff_i)^p + \sum_{p=1}^{3} \gamma_{2p} \cdot \mathbf{1}(score_i > cutoff_i) \cdot (score_i - cutoff_i)^p + \delta \cdot \mathbf{1}(score_i > cutoff_i) + u_i \quad (5)$$

Panel B instead includes fitted values from a first-stage regression with a linear trend, interacted with an indicator of being above the cutoff:

$$L_{1,i} = \gamma_0 + \gamma_1 \cdot (score_i - cutoff_i) + \gamma_2 \cdot \mathbf{1}(score_i > cutoff_i) \cdot (score_i - cutoff_i) + \delta \cdot \mathbf{1}(score_i > cutoff_i) + u_i \quad (6)$$

In each specification, δ represents the impact on loan size of being above the risk score threshold. Panel C shows the local linear approximation within the optimal data-determined bandwidths above and below the cutoff. Figure 4 replaces average loan size with the probability of receiving the larger loan.

These charts provide strong visual evidence for the discontinuity in loan size at the risk score threshold. The average loan size jumps from approximately \$5 to \$10 at the threshold, and the probability of getting the larger loan shifts from close to 0% to nearly 100%. This is unsurprising given the lender's sorting policy and the automated nature of most lending decisions by this lender.

Table 5 presents the formal first-stage results. Columns 1-3 use loan size as the dependent variable, while Columns 4-6 use a dummy for obtaining the larger loan. Columns 1 and 4 include a third-order polynomial as controls for the risk score fully interacted with an eligibility indicator, corresponding to Panel A in Figures 3 and 4. Columns 2 and 5 replace the polynomial with linear controls, corresponding to Panel B in Figures 3 and 4. Columns 3 and 6 use local linear estimation, corresponding to Panel C in Figures 3 and 4.

Consistent with the visual evidence, eligibility for the larger loan is highly predictive of receipt of that loan. Results are very similar across both specifications. Borrowers with risk scores above the threshold receive loans that are around \$4.67 - \$4.73 larger, or put another way, have a 93%-95% higher probability of obtaining a \$10 loan relative to a \$5 loan. These results imply that the RD design is quite close to qualifying

[13] The full distribution of new borrower risk scores during this period had a range of .53, so each bin represents about 2% of this range.

as "sharp", and that the degree of scaling in my IV estimation of the effect of initial loan size on default will be minimal.

4.3 Impact of Initial Loan Size on Default

Given the strong evidence for a jump in loan size precisely at the risk score threshold, I next exploit this discontinuity to estimate the causal impact of receiving a larger initial loan on the probability of default on that loan. Figure 5 depicts the reduced-form relationship between the default rate and the risk score relative to the $10 eligibility threshold. Each plot shows the average default rate across risk score bins. Consistent with the approach followed in the first-stage analysis, Panels A and B contain fitted values from the reduced-form regression including global polynomial and linear controls fully interacted with an indicator of being above the cutoff, while Panel C shows the local linear approximation within the optimal data-determined bandwidths above and below the cutoff.

There is clear visual evidence of a jump upward in default rates precisely at the risk score threshold. This implies that increasing the initial loan size to otherwise similar borrowers leads to a higher probability of default on the initial loan. The figure also illustrates, as expected, a negative relationship between the risk score and default rates. This relationship is steeper and tighter for borrowers above the threshold. This reflects the lower density of borrowers below the threshold, which reduces the precision of the estimation results, as well as related fact that the risk score model tends to be less predictive for lower quality borrowers because fewer of them are let into the system in the first place. A formal test, as reported in Table 6, confirms that the slope change is indeed significant.

Table 7, Columns 1-3 presents formal two-stage least squares estimates of the impact of doubling the initial loan size on the probability of default by new borrowers. I instrument for initial loan size using the maximum eligible loan, which is $10 and $5 for borrowers above and below the threshold, respectively. I also include loan cohort week fixed effects. Column 1 corresponds to the global polynomial risk control specification, Column 2 utilizes global linear controls, and Column 3 features local linear estimation. I find that doubling the initial loan size leads to a large, statistically significant, and economically meaningful increase in default among new borrowers. With polynomial controls, borrowers who get the larger loan are 7.7ppts more likely to default, relative to an average new borrower default rate during this period of 17.7%. This thus represents an economically meaningful 44% increase in the probability that a new borrower defaults on her first loan. The global linear specification yields a slightly higher estimate of a 8.2ppt positive impact, corresponding to a 46% increase in the default rate, while the local linear estimate is slightly lower at 6.1ppts (a 34% increase in default).

These results can be interpreted as an exogenous increase in initial loan size, net of any selection effects, leading to a higher likelihood of default by new borrowers on their first loan. In other words, starting borrowers out at a higher loan size reduces the probability of repayment, implying the lender both loses the principal on these loans and loses the defaulting borrowers as customers. This is consistent with some combination of positive moral hazard and repayment burden effects, although I don't have sufficient variation to identify each separately. Moral hazard captures the idea that individuals who receive larger loans have weaker repayment incentives, consistent with a number of underlying mechanisms. The repayment burden effect refers to the fact that larger loans mechanically require more to be paid back, thus making repayment more difficult. It is worth pointing out that much of the literature on the impact of loan size (or interest rates) on default rates does not distinguish between these two effects, and instead refers to any positive impact on default rates from an exogenous increase in loan size or interest rates as moral hazard.[14]

4.4 Impact of Initial Loan Size on Longer-Term Credit Outcomes

I also consider how varying the initial loan size impacts longer-term credit outcomes at the borrower level, beyond the probability of default on the initial loan. The identification strategy, and thus the first-stage portion of the analysis, remains the same.

Figure 6, Panel A shows the graphical results for the reduced-form impact of a larger initial loan on borrower-level default rates. There are no clear discontinuities in the overall probability that a borrower defaults. In other words, receiving a larger initial loan does not make a borrower more likely to default eventually, even though that borrower is more likely to have defaulted on her first loan. Putting these two pieces together, this implies that giving a larger initial loan shifts the timing of default forward. Table 7, Columns 4-6 confirm these results.

Figure 6 also gives visual evidence for the reduced-form effect of a larger initial loan on the total number of loans and the last loan size (or equivalently, how far up the loan ladder the borrower progressed). Table 8 presents the corresponding formal estimates. Receiving a larger initial loan results in a borrower taking out 0.9-1.4 fewer loans. However, because these borrowers also started at a higher ladder rung, it does not lead to a significant change in how far up the ladder a borrower makes it.[15]

5 Empirical Results: Loan Ladder Progression

I next analyze the impact of an exogenous change in the loan ladder progression policies of the lender. This is the core piece of the lender's progressive lending structure. Faster progression has the upside of creating stronger incentives for repayment, but also comes at the risk of increasing the ultimate probability of borrower default and the cost to the lender in the event of default once borrowers reach much larger loan sizes. My empirical design allows me to analyze this tradeoff in my particular setting. In doing so, I am able to provide novel evidence on the impact of dynamic contracting, and in particular how the stage of the dynamic relationship between borrowers and lenders influences repayment outcomes.

[14]Examples include Adams et al. (2009) and Dobbie and Skiba (2013).

[15]For all of these results, the predictive content is stronger for individuals with risk scores above the threshold: the relationship between the risk score and the variable of interest is tighter, more steeply sloped above the threshold in most cases. As discussed previously, this likely reflects the lower density of borrowers below the threshold and the fact that the risk score model used by the lender does less well for lower quality borrowers because it has less data on these types off which to train.

5.1 Empirical Strategy

I estimate the impact of loan ladder progression on borrower behavior by exploiting a series of experiments in how repeat borrowers proceeded up the loan ladder conducted by the lender. These experiments were unannounced to borrowers and thus provide a source of exogenous variation in the rate at which borrowers move up the loan ladder. Specifically, the lender introduced a policy whereby repeat borrowers were only upgraded to a larger loan if their on-time payment (OTP) percentage on their previous loan was above a certain threshold, and then experimented with the level at which this threshold was set. The OTP percentage refers to the fraction of loan payments that were completed before the scheduled repayment due dates. Borrowers below this threshold either remained on the same loan ladder level, or were downgraded if their OTP percentage was sufficiently low. Prior to the introduction of the OTP upgrade threshold, all borrowers who repaid their previous loan were upgraded to a larger loan. The lender introduced the threshold in mid-2015 and set it at a particular level, which I will denote as $P\%$. They then experimented with the level of the threshold over the next few months: lowering it by 25ppts to $(P-25)\%$ in November 2015, increasing it back to $P\%$ in December 2015, lowering it back to $(P-25)\%$ in March 2016, and then increasing it back to $P\%$ a month later. Figure 7 summarizes these changes. This variation in the OTP threshold generates "upgrading shocks": periods when borrowers who normally would have had to take out a next loan of the same size instead got upgraded to a larger next loan. The objective of this analysis is to determine how this variation in loan ladder progression impacts repayment outcomes.

I estimate the impact of getting upgraded to a larger loan on default using a differences-in-differences (DD) approach, comparing default outcomes for borrowers whose upgrade status changed depending on whether they reapplied before or after the policy change to those who always got upgraded.[16] The treatment group is the set of borrowers whose upgrade status depended on the timing of their application for their next loan. It is composed of the set of borrowers whose OTP percentage on their previous loan was in the range of $(P-25)\%$ to $P\%$. These borrowers were only upgraded to a larger subsequent loan during the periods in which the upgrade threshold was temporarily lowered to $(P-25)\%$. The control group is formed by borrowers who were always upgraded to a larger loan because their OTP percentage on their previous loan was between $P\%$ and 100%. I restrict my sample to second-time borrowers to limit the impact of possible selection issues caused by borrowers who default (and thus are ineligible for future loans from this lender) dropping out of the sample. I further restrict my sample to those second-time borrowers whose first loan was \$10, to prevent any compositional issues induced by the results presented in Section 4.

I use the timing of the policy changes to delineate the temporal dimension of the DD analysis. Due to very low volume of less than 50 loans a week, I exclude the period prior to November 2015 from my main analysis, although I do incorporate the earlier period as a robustness check in Section A.2. The "off" period consists of when the upgrading shock was not in effect, defined as when the OTP threshold was set at $P\%$.

[16] While an RD strategy is in theory possible in this setting as well, with the OTP percentage as the running variable, there is not sufficient variation in OTP percentages for this strategy to be implementable. This reflects the fact that most loans in the sample have either three or four weekly payments due, thus constraining the possible values of the OTP percentage significantly.

While in theory this is true both from December 2015 through Feb 2016 and from April 2016 onward, in practice a few other upgrading restrictions were added in April 2016 that make this period not a suitable comparison. Naturally, the "on" period is when the upgrading shock was in effect, which consists of two time segments: when the OTP upgrade threshold was temporarily lowered to $(P-25)\%$ in November 2015 and March 2016.[17]

Figure 8 depicts the identification strategy. The control group is always upgraded to a larger second loan relative to their first loan. Given that their first loan was $10 and the second loan on the loan ladder during this period was $20 in size, members of the control group face an effective loan ladder slope[18] of two. During the "off" period, from mid-December 2015 through early March 2016, the treatment group is not upgraded to a larger second loan. Instead, they receive a $10 loan again, so the loan ladder slope they observe is one. Conversely, during the two "on" periods, the treatment group is also upgraded, so their loan ladder slope increases to two. Therefore, the treatment group gets a loan that is twice as large if they reapply during the "on" period instead of the "off" period, or put another way, face an upward-sloping loan ladder instead of a flat one.

Formally, I estimate DD regression equations that take the following form:

$$D_i = \alpha + \delta \cdot on_t + \gamma \cdot treated_i + \beta \cdot (on \cdot treated)_{it} + \eta \cdot X_{it} + \varepsilon_{it} \tag{7}$$

where i denotes the borrower, t denotes the week the loan is disbursed, and X_{it} is a vector of covariates. The coefficient of interest is β, which represents the impact of the upgrading shock on default. δ captures the effect on default of being in the "on" period when the upgrading shock was in effect for the borrowers who would have been upgraded regardless, and γ gives the effect of being in the treatment group when the upgrading shock was not in effect.

The key identifying assumption behind my DD approach is that, in the absence of the change in the loan ladder progression policy, default rates would depend on group and period but not on their interaction. Put another way, the average change in default rates between the "off" and "on" periods for borrowers who were always upgraded represents a valid counterfactual change for the treatment group if the change in loan ladder progression was never instituted. While this assumption is inherently untestable, I provide some corroborating evidence that it holds in Section 5.2. I show that default rates trend similarly for both groups during the period when both are upgraded (termed the "on" period), although there is a level difference induced by the fact that the control group is composed of observably better borrowers (given that they repaid more of the installments on their previous loan on time than did the treatment group). In Section A.2, I also run the DD for an extended sample period with an additional "off" segment, and find similar results.

[17] Note that "off" and "on" are analgous to "pre" and "post" in traditional DD parlance. I define them in this way so that the analysis yields an estimate the impact of larger repeat loan on default rates and thus can be compared to the results from Section 4 on the impact on repayment of a larger initial loan.

[18] I define the loan ladder slope in proportional terms; it is equal to the ratio of the current loan size to the previous loan size.

Another potential issue for the causal interpretation of my results is selection. As was the case in the first set of empirical results on initial loan size, there is no possibility of differential loan size selection by risk type in this setting. Everyone borrowed ("selected in" at) the same initial loan, and borrowers could not select a different second loan size besides the one into which they were sorted according to their OTP percentage on their first loan. In practice, after repaying their first loan, borrowers could only view and apply for the one loan product that they qualified for next, as determined by the OTP threshold rules. This means the OTP threshold changes can be viewed as exogenous variation in whether a subset of borrowers were upgraded to a larger loan (relative to their first loan) or not. However, selective borrowing is possible and potentially problematic: borrowers may have chosen not to take out another loan if they were not upgraded. This could create differences between treatment and control groups that vary over time but reflect a selection effect rather than the causal effect of the upgrading shock. I show that there is little evidence that this occurs in my setting in Section A.2.

5.2 Impact of Loan Ladder Progression on Default

Figure 9 depicts the estimated difference in default rates between the treatment and control group by loan cohort week, along with the corresponding 95 percent confidence intervals. It provides a few different pieces of information. First, it provides graphical support for the parallel trends assumption. Default rates are slightly higher for the treatment group relative to control during the "on" period when both groups are upgraded, but this difference is not statistically significant and is relatively constant across time. Next, it suggests a strong effect of loan ladder progression on default rates. Treatment default rates rise substantially relative to control when the treatment group is *not* upgraded to a larger second loan. While there is heterogeneity by cohort week, the positive difference in default rates between treatment and control is both statistically significant and economically meaningful in nine weeks out of 12.

Table 9, Column 1 provides the formal DD regression estimates of Equation 7. It contains the standard DD specification, incorporating controls for borrower risk score, the speed at which the second loan is taken out, and borrower cohort fixed effects. The estimation results show that an exogenous doubling in loan size (relative to the previous loan) to a second-time borrower, or an "upgrading shock", leads to a statistically significant decline in default by 2.9ppts. This is relative to an average default rate of 5.0 percent for this sample during this period. Therefore, it implies that a doubling in the second loan size relative to no loan size improvement causes the default rate to fall by 58%.

This is a striking result. It can be interpreted as an exogenous increase in the rate of loan ladder progression, net of any selection effects, leading to a much lower likelihood of default on that loan by repeat borrowers. In other words, giving repeat borrowers more money on their second loan relative to their first loan actually makes them less likely to default. Under the assumption that any repayment burden effect is always positive (a larger loan is harder to repay because the borrower has to come up with a bigger sum of money), this implies a negative moral hazard effect. Very few others have found evidence of negative moral hazard in credit markets, with one prominent exception being Dobbie and Skiba (2013) in the payday lending market in the United States, as discussed in Section 2.

It also contrasts sharply with the results for first time borrowers in the same setting, presented in Section 4. New borrowers who receive an exogenously larger loan are *more* likely to default, whereas repeat borrowers who receive an exogenously larger loan (relative to their first loan) are *less* likely to default. This suggests that there are important differences between borrowers and how they make default decisions depending on which stage of the dynamic relationship with the lender they are currently in. I investigate some possible mechanisms to explain these contrasting results in Section 5.4 and in the model presented in Section 6.

5.3 Impact of Loan Ladder Progression on Longer-Term Credit Outcomes

Given that faster loan ladder progression leads to lower default on the current loan, a natural next question is how it impacts borrower-level default rates. Does the loan ladder structure impact overall default probabilities, or just shift the timing of default? Figure 10, Panel A displays the difference in the rates of eventual default between borrowers in the treatment and control groups. Borrower-level default rates are relatively constant and not statistically different in the "on" period, when both treatment and control borrowers were upgraded on their second loan. However, the rate at which borrowers from the treatment group default eventually actually drops substantially during the "off" period, during which they were forced to stay on the same loan ladder level on their second loan.

Table 9, Column 2 presents formal estimates of this effect. An exogenous doubling in loan size on the second loan increases the probability of eventual default 8.6ppts, relative to an average borrower-level default rate among this sample of 50.9%. This result is robust to including various controls, including the internally-calculated risk score, the number of loans taken out before default, the total amount of time between loans, and loan cohort fixed effects. Therefore, this implies that the steeper loan ladder early on increases eventual default by 16.9%, which is substantial. Taken together with the results on default on the second loan, this suggests that faster loan ladder progression encourages borrowers to postpone defaulting until they move further up the ladder and increases the overall probability of default.

Figure 10, Panels B and C depict the impact of faster loan ladder progression on two other borrower-level outcomes, the total number of loans taken out and the final loan ladder rung reached. Treatment borrowers, on average, take out fewer loans and make it significantly less far up the loan ladder. Both results are consistent with the fact that the treatment group are observably somewhat worse borrowers, given their worse OTP record on their first loan and their slightly lower average risk score. Treatment borrowers don't take out more loans overall when they face a steeper loan ladder during the "on" periods, but do end up making it farther up the ladder relative to during the "off" period; in other words, while their last loan is still significantly smaller than that of control borrowers on average, the gap is much smaller when they were upgraded on their second loan. Table 9, Columns 3-4 provides formal DD estimates to corroborate the visual evidence.

5.4 Mechanisms

The results above suggest that some borrowers respond strategically to progressive lending. As is fleshed out in Section 6, this is consistent with a model with two key elements: first, underlying heterogeneity across borrowers in the value placed on larger future loans, and second, borrowers are forward-looking and form expectations about future loan growth based off of their recent experience the lender. Some borrowers place greater value in reaching a larger loan size and base repayment decisions off of this preference; or in other words, they act "strategically" to try to move up the loan ladder to larger loan size. If these borrowers observe loan growth on the second loan relative to the first, they are more likely to repay this loan because they believe they will again move up the ladder on their next loan. If they observe no loan growth on the second loan, then they are more likely to default strategically now. It is important to note that this is not necessarily nefarious behavior; borrowers may be trying to game the system and reach the largest loan size they can before defaulting, or they might just need a much larger loan for their business or consumption needs and don't view the smaller loans as very useful.

To test this hypothesis, I split the sample by a variable that is plausibly correlated with strategic borrowing behavior: the speed at which a borrower takes out loans. Specifically, because the analysis above focuses on borrowers' second loans, I sort borrowers based on the number of days between their first and second loans. This measure is highly correlated with default, as shown in Figure 11. I then split the sample into two groups based on whether they are above ("fast turnover") or below ("slow turnover") the median in this sample of one day, and repeat the DD analysis separately for each group. About 72% of the control group and 55% of the treatment group take out their second loan within one day of repaying their last. The results presented here are robust to using different variables in the same spirit to split the data, such as the number of days repaid early.

Figure 12, Panels A and B depict the differences in default rates on the second loan for the fast and slow turnover borrowers, respectively. There is a clear difference in patterns between the two groups, despite some volatility across loan cohort weeks. Only the fast turnover borrowers in the treatment group see a sharp rise in their average default rate relative to the control group when faced with a flat loan ladder during the "off" period. This provides visual support for the notion that the full sample results are largely driven by the set of borrowers who are taking loans out most quickly, and thus are more likely to repay strategically to try to get to a higher loan size.

The DD regression estimates in Table 10 confirm this. Columns 1 and 2 estimate the model separately for the fast and slow turnover groups, respectively, including controls for the risk score and the week in which a borrower took out her first loan. The DD estimate is large, negative, and statistically significant for the former group, and small and indistinguishable from zero for the latter. Column 3 includes a full set of interactions with a binary indicator for having taken out a second loan at least as fast as the median borrower, again including the additional controls. In this regression, the coefficient of interest is the one on the triple interaction term, and it is again statistically significant and economically meaningful. These results suggest that, among borrowers who take out loans quickly, getting upgraded on the second loan relative to the first instead of having to take out the same loan size twice causes default rates to fall by 4.8% to 5.4%. Again, this is relative to an average default rate of 5.0%, implying it causes default to fall by around 100%.

6 Model

6.1 Introduction

I next introduce a simple model that attempts to capture my empirical evidence. To recap, there are three empirical results that the model should help to explain. First, an exogenously larger initial loan to new borrowers leads to higher default rates. Second, an exogenously larger loan relative to the previous loan to repeat borrowers, or equivalently faster loan ladder progression, results in lower default rates on the current loan, but doesn't reduce overall borrower-level default rates. And finally, the latter result is driven by borrowers who appear to be acting strategically, as captured by the speed at which they take out loans.

To explain these results, I write down a dynamic discrete choice model of individual borrowing behavior, in which borrowers repay their current loan in order to remain eligible for future loans. I then use the model to derive conditions for when and why different borrowers choose to default. In Sections 7 and 8, respectively, I also use the structure of the model to estimate the underlying distribution of borrower types and then to simulate lender profits under various progressive lending schemes.

The logic of the model is as follows. Borrowers vary in the value they place in future access to loans. They don't have ex-ante information about loan ladder; instead, they form expectations using their own loan history. A steeper loan ladder creates stronger incentives for borrowers to repay current loan, because they expect larger loans in the future. However, if the loan ladder is too steep, it can lead to borrowers postponing default until much farther up the loan ladder, which can ultimately be more costly for the lender.

6.2 Setup

Consider a setting with one lender and a continuum of borrowers. Each borrower can interact with the lender a potentially infinite number of times, but both sides can only commit to one-period contracts.

The lender is assumed to be a profit-maximizing entity operating in a competitive environment, with cost of funds given by r. The lender has a limited ability to screen out first-time applicants. Conditional on repayment of the first loan, the lender doesn't reject repeat borrowers. The lender cannot collect collateral and has no ability to monitor, so it must rely exclusively on dynamic incentives to encourage repayment. It does this using a two-pronged approach. First, any borrowers who default are permanently excluded from borrowing in the future from the lender. Second, it uses a progressive lending structure, such that loan sizes increase as borrowers successfully repay loans and take out subsequent loans. The lender determines loan ladder progression based on beliefs about the borrower's probability of repayment. However, I assume that loan sizes and loan ladder progression are taken as exogenous by the borrower. This assumption is motivated by the reality that the process is opaque from borrower's perspective. The lender also has a maximum loan size that they are able to offer, $\bar{L}$.

Each borrower starts out with an initial capital stock K_0, which is assumed to be constant across all borrowers. Loans from this lender are assumed to *augment* the existing capital stock, rather than providing the sole funds available for investment in each period. This reflects the small average loan size and the fact that most borrowers report using the loans for consumption-smoothing purposes or to purchase working capital for an existing business, rather than to start new businesses. Borrowers have access to a production function $F(\cdot)$, which is increasing and concave and satisfies the usual Inada conditions ($F(0) = 0$, $F(\infty) = \infty$, $F'(0) = \infty$, $F'(\infty) = 0$). Borrowers don't have access to any savings devices, but they do potentially have access to loans from other lenders.

Borrowers differ in three dimensions. First, borrowers differ in the extent to which they value access to future loans from the lender. This is captured by heterogeneity across borrowers in a net outside option parameter, ω_i, which they receive if positive or pay if negative in the event of default. ω_i can alternatively be interpreted as the inverse of the default cost. Borrowers with a higher net outside option (or equivalently, a lower default cost) value future loans from this lender less. I assume that ω_i is exogenous and known to the borrower, but not observable to the lender. Second, borrowers vary in credit needs. Each borrower has a maximum number of loans $\bar{T}_i$ that they need (or equivalently, that they can afford to repay). Once a borrower reaches her maximum loan cycle, she ends her relationship with the lender. I assume that borrowers do not know their maximum loan cycle in advance, but instead receive a signal each period indicating whether or not they have reached it. Finally, borrowers are heterogeneous in how quickly they expect loan sizes to increase upon taking out their first loan. This is represented as variation in $\tilde{\lambda}^e_{2,i}$, the expected proportional increase in the second loan relative to a borrower's first loan. For notational conciseness, I drop individual subscripts on all variables except for these sources of heterogeneity.

Time is discrete. Each period t corresponds to a loan cycle. For instance, a borrower in period $t = 2$ is on their second loan with the lender. In period t, assuming no previous defaults, a borrower starts with her initial capital K_0 and funds from previous loans $\sum_{s=1}^{t-1} L_s$. I denote capital at the beginning of period t (prior to loan receipt) as:

$$K_{t-1} = \begin{cases} K_0 & \text{if } t = 1 \\ K_0 + \sum_{s=1}^{t-1} L_s & \text{if } t > 1 \end{cases} \tag{8}$$

The borrower receives loan (L_t, R_t) from the lender, where L_t is the loan size and R_t is the gross interest rate. She invests her current loan and gets certain net return $F(K_{t-1} + L_t) - F(K_{t-1})$, and owes the lender $R_t L_t$. She then receives a signal S_t indicating whether she has reached her maximum loan cycle, where $S_t = 0$ if $t < \bar{T}_i$ and $S_t = 1$ if $t = \bar{T}_i$. I assume this signal occurs with the same probability p to each borrower in every period, but borrowers do not know the value of p and cannot infer it. She then decides whether to repay her current loan or default. This decision is based off of a comparison of the net benefit of repayment versus the net benefit of default. If she repays, she gets access to another, possibly larger, loan in the next period. Note that if she has reached her loan cycle cap, this consideration is no longer relevant. If she defaults, she walks away with the full loan amount L_t and gets her net outside option ω_i if positive, or pays it if negative. Borrowers thus trade off having more money today if they default on their current loan versus the value of obtaining an additional loan next period if they repay. If she repays and is not at the loan cap, she proceeds to period $t+1$ and takes out another loan.[19] If she defaults, she is no longer eligible for credit from this lender and the relationship ends.

The borrower's value function in period t thus takes the following form:

$$\begin{aligned} V_i^t(K_t, L_t, R_t) \; &= \max\{(F(K_{t-1}+L_t) - F(K_{t-1}) - R_t L_t + \delta E[V_i^{t+1}(K_{t+1}, L_{t+1,} R_{t+1})|K_t, L_t, R_t], \\ &F(K_{t-1}+L_t) - F(K_{t-1}) + \omega_i\} \end{aligned} \tag{9}$$

where $E[V_i^{t+1}(K_{t+1}, L_{t+1,} R_{t+1})|K_t, L_t, R_t] = 0$ if the borrower receives a signal in period t that $t = \bar{T}_i$, which occurs in each period with probability p. However, the signal comes as a surprise and the borrower does not have any information about p, so it does not enter directly into the value function in period t.

Borrowers have no knowledge about the structure of the lender's loan ladder. I assume that borrowers form expectations about future loan sizes based on their own loan history. Specifically, repeat borrowers ($t > 1$) expect their next loan to increase by the same proportion as their current loan:

$$E[L_{t+1}] = L^e_{t+1} = \lambda^e_{t+1} L_t \tag{10}$$

where λ^e_{t+1}, the expected proportional change in a borrower's next loan relative to her current loan, is:

$$\lambda^e_{t+1} = \frac{L_t}{L_{t-1}} \tag{11}$$

New borrowers ($t = 1$) don't have any loan history to extrapolate forward. I assume that there is heterogeneity across borrowers in loan growth expectations, captured by $\tilde{\lambda}^e_{2,i}$:

$$E[L_2] = L^e_2 = \tilde{\lambda}^e_{2,i} L_1 \tag{12}$$

Given my focus on loan size rather than price, I assume that borrowers expect the interest rate to remain constant over all loans[20]:

[19] The reapplication rate (after the repayment of the prior loan) averages close to 90%, so I ignore the third option of repaying and deciding to end the relationship with lender.

[20] This is consistent with the data, in the sense that the variation in loan sizes is much more significant than the variation in interest rates.

$$E[R_{t+1}] = R^e_{t+1} = R_t \tag{13}$$

Given the lender's loan ladder policy $\{\lambda_t\}$ and initial loan terms $\{L_1, R_1\}$, which determine the subsequent loan terms $\{L_t, R_t\}$, we can use this to define a function $T(\omega_i)$ which represents the optimal number of loans the consumer takes out before defaulting, conditional on not hitting their maximum number of loans:

$$T(\omega_i) = \arg\max_T \left\{ \sum_{t=1}^{T-1} \delta^t (F(K_{t-1} + L_t) - F(K_{t-1}) - R_t L_t) + \delta^T (F(K_{T-1} + L_T) - F(K_{T-1})) + \delta^T \omega_i \right\} \tag{14}$$

$$\text{s.t.} \quad \begin{aligned} K_{t-1} &= \begin{cases} K_0 & \text{if } t = 1 \\ K_0 + \sum_{s=1}^{t-1} L_s & \text{if } t > 1 \end{cases} \\ L_t &= \lambda_t L_{t-1} \\ R_t &= R_{t-1} \\ \lambda_t &= \frac{L_{t-1}}{L_{t-2}} \\ t &\leq \bar{T} \\ L &\leq \bar{L} \\ K_0, L_1, R_1 & \quad \textit{given} \end{aligned}$$

where $0 < T(\omega_i) < \infty$. Because the borrower has no information about her maximum loan cycle, they cannot take it into account when determining their optimal default time. Also, note that loan sizes cannot grow above the lender's maximum offered loan size $\bar{L}$.

In each period, the borrower receives a signal about her maximum loan cycle before making her default decision. I consider each case separately.

Case 1: $S_t = 0$ $(t < \bar{T}_i)$

Let ω_t be the net outside option for a borrower who is just indifferent between defaulting in period t and period $t+1$, conditional on not receiving the loan cap shock in the current period:

$$F(K_{t-1} + L_t) - F(K_{t-1}) - R_t L_t + \delta \left(F(K_t + L^e_{t+1}) - F(K_t) + \omega_t \right) = F(K_{t-1} + L_t) - F(K_{t-1}) + \omega_t \tag{15}$$

Equation 15 is similar in spirit to the Euler equation. Recall that the Euler equation says that the individual is indifferent between reallocating an infinitesimal amount of consumption between periods t and $t+1$. Here, the choice variable is the binary default decision, and indifference is between defaulting in periods

t and $t+1$.

Solving for ω_t yields the following equation:

$$\omega_t = \frac{\delta}{1-\delta}\left(F(K_t + \lambda^e_{t+1}L_t) - F(K_t)\right) - \frac{1}{1-\delta}\cdot R_t L_t \tag{16}$$

We can alternatively write this condition in terms of the marginal product of capital (MPK_i) by taking a first order approximation of the production function. I allow the marginal product of capital to vary across individuals, but not over loan cycles.[21] The net outside option of the indifferent borrower is now given by:

$$\omega_t = \frac{1}{1-\delta}L_t\left(\delta \cdot MPK_i \cdot \lambda^e_{t+1} - R_t\right) \tag{17}$$

ω_t can be interpreted as the net value of repaying loan t. Borrowers with $\omega_i > \omega_t$ prefer default at t to default at $t+1$, because the net outside option they receive if they default is greater than the net value of repaying the loan. Borrowers with $\omega_i < \omega_t$ prefer to postpone default to $t+1$ because they gain more from repaying this loan and taking out another than from defaulting today. For a given distribution of ω_i across borrowers, anything that causes ω_t to increase will lead to lower default on loan t, as fewer borrowers will have a sufficiently high net outside option to make default worthwhile. Likewise, if ω_t falls, more borrowers will have a net outside option that exceeds the net value of repayment, and thus default will increase.

Equation 17 is similar to the Bulow and Rogoff (1989) condition that the rate of loan growth must exceed the interest rate in order for dynamic incentives to have any teeth. It generates a lower bound on the desirable rate of growth: if expected loan growth is very slow (λ^e_{t+1} is low), ω_t will fall, inducing more borrowers to default. In the extreme, if ω_t is sufficiently low, then all borrowers default.

It can be shown that ω_t is unambiguously increasing in the discount factor δ and the marginal product of capital MPK_i. More patient and higher-return borrowers stand to gain more from taking out additional loans in the future, and thus are more likely to repay today. Clearly, ω_t is decreasing in the gross interest rate R_t: borrowers are less likely to repay expensive loans.

The impact of loan size on ω_t varies depending on whether an individual is a new or repeat borrower. For new borrowers, a larger initial loan has two countervailing effects on $\omega_{t=1}$:

$$\frac{\partial \omega_{t=1}}{\partial L_1} = \delta \cdot MPK_i \cdot \tilde{\lambda}^e_{2,i} - R_1 \tag{18}$$

First, it increases the expected second loan size by raising the starting point of the loan ladder, thus increasing the marginal benefit of repayment. Second, it increases the repayment burden, thus raising the marginal cost of repayment. If the latter effect dominates, then $\omega_{t=1}$ will be decreasing in L_1. This means that a larger initial loan leads to higher default rates, as demonstrated empirically in Section 4.

For repeat borrowers, a larger current loan has an additional positive effect on the marginal benefit of repayment:

[21] This reflects the fact that loans are very small, thus making the first order approximation reasonable.

$$\frac{\partial \omega_t}{\partial L_t} = \delta \cdot MPK_i \cdot \lambda^e_{t+1} + \delta \cdot MPK_i \cdot \frac{\partial \lambda^e_{t+1}}{\partial L_t} L_t - R_t \tag{19}$$

Increasing L_t not only has a direct effect on the base loan size, but now also has an indirect effect on the expected loan ladder slope. All else equal, this makes it more likely that the marginal benefit of repayment will outweigh the marginal cost, such that ω_t is increasing in L_t. This means that a larger repeat loan leads to lower default on that loan, consistent with the empirical results in Section 5.

We can also determine how the borrower responds if she hits the maximum loan size on the lender's loan ladder. When this occurs, she can continue borrowing but loan sizes will no longer grow. Therefore, if she reaches $\bar{L}$ in period $t-1$, the indifference condition between defaulting and repaying will take the following form from period t forward:

$$\omega_t^{\bar{L}} = \frac{1}{1-\delta} \bar{L} \left(\delta \cdot MPK_i - R_t \right)$$

Clearly, reaching the maximum loan size offered by the lender reduces repayment incentives. This puts an upper bound on desired loan growth, as borrowers are more likely to default once they reach the maximum loan size.

Case 2: $S_t = 1$ $(t = \bar{T}_i)$

Next, consider the case when a borrower on loan cycle t receives a signal that she has reached her maximum loan cycle. I assume that the borrower ends the relationship with the lender at this point, either by repaying her last loan and not taking out another, or by defaulting and thus being blocked from any future credit by the lender. Let $\omega_t^{\bar{T}}$ be the net outside option for a borrower who is just indifferent between these two options:

$$F(K_{t-1} + L_t) - F(K_{t-1}) - R_t L_t = F(K_{t-1} + L_t) - F(K_{t-1}) + \omega_t^{\bar{T}} \tag{20}$$

This leads to a very simple equation for $\omega_t^{\bar{T}}$:

$$\omega_t^{\bar{T}} = -R_t L_t \tag{21}$$

Borrowers who find out they are at their maximum loan cycle will default if $\omega_i > \omega_t^{\bar{T}}$, or equivalently when their net outside option is greater than the total loan cost. $\omega_t^{\bar{T}}$ is unambiguously decreasing in both $L_{\bar{T}}$ and $R_{\bar{T}}$, meaning borrowers are more likely to default on larger and more expensive loans when they reach their maximum required loan cycle. Like the maximum loan size offered by the lender, this also puts upper bound on desired loan growth. Borrowers are more likely to default in that event as $L_{\bar{T}}$ increases.

6.3 Predictions

The model generates a number of useful predictions. First, it suggests that an optimal progressive lending scheme exists. Specifically, the lender can choose the loan ladder parameters to balance the tradeoff for lender profitability in loan size growth. If loan improvement is too gradual, this reduces the opportunity cost of defaulting, so even those types that most value future loans may default. Mechanically, a low λ reduces the net value of repayment ω_t and thus increases default on loan t. In the extreme case, everyone defaults early on, so the lender makes large losses and has no customers remaining. Conversely, if loan sizes grow too quickly, borrowers will postpone default and thus reach higher loan sizes before receiving a positive cap signal. This reflects the fact that, in the model, high λ increases the net value of repayment ω_t and thus reduces default by borrowers who have not reached their loan cycle cap. Borrowers are then more likely to default at these higher loan sizes, captured by the fact that $\omega_t^{\bar{T}}$ is decreasing in L_t. At the extreme, borrowers will postpone default until reaching the maximum loan size offered by the lender $\bar{L}$, at which point repayment incentivizes weaken significantly and default becomes much more likely for even borrowers with a low net outside option. Section 7 explores this further by trying to estimate structurally what the optimal average rate of loan growth is in this particular setting.

Next, the model captures how progressive lending generates endogenous screening. Borrowers who value access to future loans less (higher ω_i) default earlier on, at smaller loan sizes. In effect, this allows the lender to "test" borrowers with the smaller early loans and sort out unobservably worse borrowers relatively cheaply.

Finally, it provides a useful lens for interpreting the empirical results. There are two broad reasons in the model for the different observed impacts on default from increasing the first loan size versus increasing later loan sizes (i.e., increasing the loan ladder slope). The first is selection, which creates differences in types between the pools of first time and repeat borrowers. There is selection on the value placed in future loans, such that borrowers with the highest net outside options will default on the first loan. Selection also operates via borrowers' initial expectations of the rate of loan size growth. Borrowers who expect the lowest rate of loan growth after the first loan are more likely to default on the first loan, because the perceived benefit of repayment is lower. The second reason stems from the difference between initial and repeat loans in how expectations are formed about subsequent loan size growth. Intuitively, first time borrowers face a fundamentally different problem, because they have no experience yet with the lender and thus less information to utilize when thinking about the future path of loan sizes. In the model, a larger initial loan doesn't influence the expected growth rate of future loans, whereas larger repeat loans do. As a result, larger repeat loans have a compounded effect on the expected path of future loan sizes, and thus also on the marginal cost of default. Mechanically, this is reflected in the additional positive term in Equation 19 relative to Equation 18, which makes it more likely that the net value of repayment ω_t is increasing in L_t and thus that the probability of default is decreasing in L_t.

7 Structural Estimation

I next use the structure of the model to estimate two key features governing borrower behavior: the distribution of the net outside option, ω_i, and the probability of hitting the maximum loan size in each period, p_t.

7.1 Distribution of Borrower Net Outside Options

The model implies a series of bounds on borrower net outside options, ω_i. I can use these bounds to estimate the distribution of ω_i in my sample.[22]

First, consider a borrower who repays her last loan in the sample period, loan T. There are two potential reasons in the model why she would do so: either her net outside option is no greater than the net value of repayment on loan T ($\omega_i \leq \omega_T$), or she receives a signal that she has reached her maximum loan size but her net outside option is below the repayment burden on loan t ($\omega_i \leq \omega_T^{\bar{T}} = -R_T L_T$). This implies the following upper bound on the net outside option of non-defaulting borrowers:

$$\omega_i \leq \max\{\omega_T, \omega_T^{\bar{T}}\} = \omega_{i,upper\ bound}^{ND} \tag{22}$$

In most cases, $\omega_T^{\bar{T}} < \omega_T$, so the upper bound is given by ω_T.

Next, consider a borrower who defaults at loan T, where $T > 1$. Again, the model gives two possible explanations: either her net outside option exceeds the net value of repayment on loan t ($\omega_i > \omega_T$), or she receives a signal that she has reached her maximum loan cycle and her net outside option is greater than the repayment burden on loan T ($\omega_i > \omega_T^{\bar{T}} = -R_T L_T$). Given that the borrower made it to loan T, she successfully repaid loan $T-1$. This means her net outside option must be no greater than the net value of repaying loan $T-1$ ($\omega_i < \omega_{T-1}$). Therefore, the net outside option of a borrower who defaults on loan T must fall within the following bounds:

$$\omega_{i,lower\ bound}^{D,RB} = \min\{\omega_T, \omega_T^{\bar{T}}\} \leq \omega_i < \omega_{T-1} = \omega_{i,upper\ bound}^{RB,D} \tag{23}$$

In most cases, $\omega_T^{\bar{T}} < \omega_T$, so the lower bound is given by $\omega_T^{\bar{T}}$.

Finally, consider a first-time borrower who defaults ($T = 1$). While she may do so for the same two reasons as repeat borrowers, we don't have any information on previous loans to translate into an upper bound. Therefore, we can only derive a lower bound for defaulters on first-time loans:

$$\omega_{i,lower\ bound}^{D,FTB} = min\{\omega_1, \omega_1^{\bar{T}}\} \leq \omega_i \tag{24}$$

In most cases, $\omega_1^{\bar{T}} < \omega_1$, so the lower bound is given by $\omega_1^{\bar{T}}$.

[22] Note that no borrowers have reached the lender's maximum loan size offered, so I do not use any bounds related to this constraint in this section.

To summarize, I can calculate an upper bound on the net outside option of non-defaulters using loan terms on their last observed loan during the sample period. For defaulters on the initial loan, I calculate a lower bound using the loan terms on their first loan. For defaulters on repeat loans, I calculate an upper (lower) bound using the loan terms on their second-to-last (last) observed loan during the sample period.

To do so, I first need to compute ω_t and $\omega_t^{\bar{T}}$ for each borrower over the relevant loan cycles. Recall that $\omega_t = \frac{1}{1-\delta} L_t \left(\delta \cdot MPK_i \cdot \lambda_{t+1}^e - R_t\right)$ and $\omega_t^{\bar{T}} = -R_t L_t$. While loan size L_t and the interest rate R_t are available in the data, I need to make a few additional assumptions. I assume a monthly discount factor of $\delta = .9$. I simulate $n = 100$ draws of monthly marginal returns to capital for each individual, where MPK_i is drawn from a uniform distribution between 5% and 25%.[23] I assume repeat borrowers form expectations about future loan growth adaptively, as described in the model, such that $\lambda_{t+1}^e = \frac{L_t}{L_{t-1}}$. I need to use an alternative assumption for new borrowers, however, given that they have no loan history off which to extrapolate. I assume rational expectations, meaning that new borrowers correctly anticipate the lender's loan ladder progression policies. Borrowers don't know there is heterogeneity in loan ladder progression, so this means the loan ladder slope between the first and second loan is a random variable with a fixed probability distribution given by the observed probabilities:

$$E[\lambda_2] = \lambda_2^e = \sum_{j=1}^{N} P(\lambda_{j,2}) \cdot \lambda_{j,2} \tag{25}$$

where $\lambda_2 = \frac{L_2}{L_1}$ is the proportional loan growth in period $t = 1$, $\lambda_{j,2}$ is one of j different possible loan growth rates between periods $t = 1$ and $t = 2$, and $P(\lambda_{j,2})$ are the empirically observed proportion of borrowers who are allocated to each $\lambda_{j,2}$.

I first present results on the variation in the net value of repayment ω_t and inverse repayment burden $\omega_t^{\bar{T}}$ for $t > 1$. I exclude the first loan cycle because of the difference in assumptions about λ_{t+1}^e. Figure 13, Panels A and B depict how ω_t varies by loan cycle unconditionally and conditional on the loan cycle at which a borrower eventually defaults, respectively. Panel A can be interpreted as the average upper bound on the net outside option of people who repay loan t and the average lower bound on the net outside option of people who don't, conditional on not reaching their maximum loan cycle. Panel B controls for selection effects, and gives a visual depiction of the strategy for identifying bounds for repeat borrowers who default: the upper bound is given by ω_T (the last point on each line) and the lower bound for those who default strategically (not because they hit their maximum loan cycle) is given by ω_{T-1}. Figure 14 shows how the inverse of the repayment burden, $\omega_t^{\bar{T}}$, evolves by loan cycle. It provides visual confirmation of the assumption in the model that the repayment burden is increasing in t. Figure 15, Panels A and B displays how average default rates vary across ω_t and $\omega_t^{\bar{T}}$, respectively. As would be expected, ω_t is inversely related to default: the higher the net value of repayment, the fewer borrowers default. The relationship between $\omega_t^{\bar{T}}$ and default is much noisier.

[23]This is consistent with the range of values in recent studies, including De Mel et al. (2008) and Hussam et al. (2017).

Next, I compute upper and lower bounds using Equations 22, 23, and 24. Figure 16, Panels A and B show the distribution of upper and lower bounds, respectively, by borrower type. The upper bound estimates, which I have for both defaulters and non-defaulters and which primarily reflect variation in ω_t, appear to be approximately normally distributed, an observation which underlies my distribution estimation strategy.

Finally, I use these bounds to estimate the distribution of net outside options across all borrowers. To do this, I use a censored regression model to estimate the mean and variance of net outside options within MPK_i buckets. I assume that, conditional on the MPK_i draw, the net outside option is normally distributed: $w_i|MPK_i \sim N(\mu, \sigma^2)$.

I have an upper bound ($\omega_{i,upper}$) on ω_i for non-defaulters and defaulters on repeat loans and a lower bound ($\omega_{i,lower}$) on ω_i for all defaulters. This implies that non-defaulters (ND) are left censored (only an upper bound), defaulters on initial loans (D_{FTB}) are right censored (only a lower bound), and defaulters on repeat loans (D_{RB}) are on an interval (both lower and upper bounds). The likelihood contribution for each borrower is the probability of them being on the interval in which they are observed:

$$\begin{aligned} \ln \mathscr{L} = & \sum_{i \in ND} \ln \Phi\left(\frac{\omega_{i,upper} - \mu}{\sigma}\right) \\ & + \sum_{i \in D_{FTB}} \ln\left\{1 - \Phi\left(\frac{\omega_{i,lower} - \mu}{\sigma}\right)\right\} \\ & + \sum_{i \in D_{RB}} \ln\left\{\Phi\left(\frac{\omega_{i,upper} - \mu}{\sigma}\right) - \Phi\left(\frac{\omega_{i,lower} - \mu}{\sigma}\right)\right\} \end{aligned} \tag{26}$$

I then estimate the censored regression model via maximum likelihood separately for each decile of MPK_i draws. The distributional parameters of the regression model with no covariates included are given in Table 11, Panel A. The mean of the distribution is generally negative but close to zero, and increasing in the MPK_i range. The distribution also grows tighter for higher values of MPK_i . Figure 17 plots the estimated distribution for a range of values of MPK_i . These results imply that there is significant heterogeneity in the value placed in future loans, with the average borrower actually facing a cost to defaulting (the equivalent of a negative net outside option). Note that, because I conservatively use $\omega_T^{\bar{T}}$ as the lower bound for defaulters in most cases, even though many of these borrowers are likely defaulting despite not having reached their loan size cap, my results generate a lower bound on the distribution estimates.

In order to explore heterogeneity across borrowers, I also estimate Equation 26 with current loan terms included as covariates, again conditional on MPK_i decile. This allows me to generate predicted net outside options at the borrower level. Figure 18 displays the predicted net outside option distributions for non-defaulters and defaulters across all MPK_i draws. Consistent with the predictions of the model, borrowers who do not default have net outside options that are lower on average than defaulting borrowers. Table 11, Panel B contains the mean and standard deviation of the distribution for defaulters and non-defaulters by MPK_i decile. The mean predicted net outside option of defaulting borrowers is positive for MPK_i values above 15%, but not until 19% for non-defaulters. The distribution for defaulters is also tighter within MPK_i

bands, likely reflecting the fact I can use both upper and lower bounds in the censored regression model. Table 11, Panel C further differentiates between first-time and repeat borrowers.

7.2 Probability of Hitting Maximum Loan Cycle

I next estimate the probability of hitting an individual's maximum number of loans in each loan cycle. To do so, I need to differentiate between borrowers in my sample who default because their net outside option is greater than the net value of repaying the loan versus those who default because they receive the loan cap signal. I am only able to do this convincingly for repeat borrowers. Repeat borrowers who default for the first reason must satisfy the condition that $\omega_{T-1} > \omega_T$, or else they would have defaulted on their previous loan. Borrowers who default for the latter reason do not need this condition to hold. Therefore, in the context of the model, repeat borrowers who default on loan T when $\omega_{T-1} < \omega_T$ must be defaulting because they hit the loan cap. The proportion of repeat borrowers who satisfy this property thus provides a lower bound estimate of p. Using this approach, I estimate $p_{lower\ bound} = 13.6\%$, with a standard error of .01%. This implies that at least 50% of borrowers will have reached their loan cap by their fifth loan and 75% will have done so by the 10th loan.

8 Structural Estimation Applications

8.1 Results for Repeat Borrower Quasi-Experiment Sample

I can use the results from Section 7 to interpret the empirical analyses of changes in progressive lending policies to first-time borrowers and repeat borrowers in Sections 4 and 5, respectively.

Starting with the initial loan size quasi-experiment, recall that borrowers with a risk score just above the threshold received a loan double the size of the loan received by borrowers just below the threshold. This resulted in an increase in default on the first loan. This result is predicted by the model if the following condition holds:

$$\frac{\partial \omega_{t=1}}{\partial L_1} = \delta \cdot MPK_i \cdot \tilde{\lambda}_2 - R_1 < 0 \tag{27}$$

Rearranging, I obtain the following condition on expected loan growth:

$$\tilde{\lambda}_2 < \frac{R_1}{\delta \cdot MPK_i}$$

This condition does not hold using the median value of MPK used in the structural estimation ($MPK_i =$ .15) and the empirically observed average of λ_2 (i.e., $\tilde{\lambda}_2 = 1.52$), combined with my assumption of δ =.9. It would require the monthly interest rate to exceed 20%, which is not true in the data. With the lower observed interest rates in the data, ω_t increases from the \$5 loan to the \$10 loan, which would lead to fewer borrowers choosing to default on larger loan. This implies that either expected loan size growth is lower than implied by rational expectations, or that productivity among first time borrowers is on the lower end of the range we consider in the structural estimation. Alternatively, it is consistent with borrowers not understanding or misinterpreting interest rates.

We can also interpret the repeat borrower experiment through the lends of the model and structural estimation exercise. Recall that, in Section 5, the treatment group gets an "upgrading shock" in the post period such that they are upgraded to a larger loan relative to their first loan, while the control group is always upgraded. In the context of the model, this will lead to an increase in the value of repaying the loan (a higher ω_t) if the following condition holds:

$$\frac{\partial \omega_t}{\partial L_t} = \delta \cdot MPK_i \cdot \lambda^e_{t+1} + \delta \cdot MPK_i \cdot \frac{\partial \lambda^e_{t+1}}{\partial L_t} L_t - R_t > 0 \tag{28}$$

Figure 19 shows ω_t for the treatment and control group in each period, assuming the median value of the MPK_i range considered in the estimation exercise.[24] Consistent with Equation 28, ω_t increases significantly between the "off" period, when treatment borrowers receive a second loan of the same size relative to their first loan, and the "on" period, when they are upgraded to a larger loan. The model then implies that, because the net value of repaying the loan is greater for the treatment group in the "on" period, fewer borrowers will have net outside options that are sufficiently high that they choose to default. This is consistent with the empirical finding that the upgrading shock leads to a decline in the default rate of the treatment group relative to the control group on their current loan. In addition, the model implies that borrowers will be more likely to default eventually, because the faster loan ladder progression induced by the upgrading shock means treatment borrowers will on average be on a higher loan cycle t when they are hit with the maximum loan signal, and default probabilities are increasing in t.

8.2 Lender Simulation

Next, I consider how the lender can optimize the loan ladder. I assume the lender has the choice over two variables: the initial loan size, L_1, and the average rate of proportional loan growth, λ. To simplify the analysis, I assume λ is constant over loan cycles. If this assumption were relaxed, the lender would want to condition λ explicitly on borrower quality, such that better borrowers are able to progress faster up the loan ladder to higher loan sizes whereas worse borrowers with a higher probability of default are forced to remain at low loan sizes. Note that there are two possible interpretations for λ. First, all borrowers receive a loan that is λ times the size of their previous loan, conditional on repaying their previous loan. Conversely, the lender chooses some proportion of borrowers to upgrade to a loan that is λ_{UG} times their previous loan (with the rest remaining on the same level), and λ is the weighted average of λ_{UG} and 1. This latter formulation is more consistent with the empirical analysis in Section 5, when the relevant experiment was not a change

[24] Results are qualitatively similar using any other value of MPK_i or averaging over all values of MPK_i; all that is impacted is the scale.

in slope per se but a change in the percentage of borrowers who could proceed up the loan ladder.

8.3 Derivation

Because of my focus on loan size rather than price, I assume the lender sets a fixed interest rate across all loans. Specifically, the lender sets the interest rate at some fixed markup m over its cost of funds:

$$R_t = R = r + m \tag{29}$$

This implies that the profit on a repaid loan is a linear function of the loan size:

$$\pi^R(L_t) = RL_t - rL_t = mL_t \tag{30}$$

Given the binary default decision assumed in the model, profit on a defaulted loan is given by:

$$\pi^D(L_t) = -rL_t \tag{31}$$

The lender seeks to maximize profits per borrower. I assume that borrowers take out one loan per month (the modal duration offered by the lender), and the lender discounts future profits with monthly discount factor $\delta = .9$. Therefore, the profit for borrower i who takes out T loans before defaulting is given by:

$$\Pi = \sum_{t=1}^{T-1} \beta^{t-1}\pi^R(L_t) + \beta^{T-1}\pi^D(L_T) \tag{32}$$

In the first period, the lender's choice of initial loan size determines the proportion of the distribution that defaults strategically. The initial loan size determines ω_1:

$$\omega_1 = \frac{1}{1-\delta} L_1 \left(\delta \cdot MPK_i \cdot \tilde{\lambda}_2^e - R\right) \tag{33}$$

Borrowers with $\omega_i > \omega_1$ default strategically, and borrowers with $\omega_i < \omega_1$ do not. In each period, including the first, there is a probability p that a borrower hits her loan cap. In that case, borrowers default if $\omega_i > \omega_t^{\bar{T}} = -R_t L_t$ and repay (but do not take another loan) otherwise. With a weakly increasing loan ladder, it is always the case that $\omega_t > \omega_t^{\bar{T}}$. Therefore, on the first loan cycle, expected profits per borrower for the lender are given by:

$$E[\Pi_1] = p\int_{-\infty}^{\omega_t^{\bar{T}}} \pi^R(L_1)dF(\omega_i) + p\int_{\omega_t^{\bar{T}}}^{\infty} \pi^D(L_1)dF(\omega_i) + (1-p)\int_{-\infty}^{\omega_1} \pi^R(L_1)dF(\omega_i) + (1-p)\int_{\omega_1}^{\infty} \pi^D(L_1)dF(\omega_i) \tag{34}$$

Plugging in and rearranging, we obtain:

$$E[\Pi_1] = L_1 \left\{(1-p)RF(\omega_1) + pRF(\omega_1^{\bar{T}}) - r\right\} \tag{35}$$

Next, I derive expected profits per borrower on repeat loans and combine this with equation 35 to compute total expected profits per borrower. In order to provide the strongest incentives for repayment, the loan ladder must be upward sloping. This is due to the fact that, with an upward sloping loan ladder and a flat interest schedule, ω_t is weakly increasing in t. Therefore, borrowers either default on the first loan or continue to repay until hitting their loan cap. This implies that the expected value of profits per borrower on loan t, conditional on being below the maximum loan size offered by the lender, is given by:

$$E[\Pi_{t|t>1}|L_t < \bar{L}] = (1-p)^{t-1}\left\{(1-p)\int_{-\infty}^{\omega_1} \pi^R(L_t)dF(\omega_i) + p\int_{-\infty}^{\omega_t^{\bar{T}}} \pi^R(L_t)dF(\omega_i) + p\int_{\omega_t^{\bar{T}}}^{\omega_1} \pi^D(L_t)dF(\omega_i)\right\} \tag{36}$$

Plugging in:

$$E[\Pi_{t|t>1}|L_t < \bar{L}] = (1-p)^{t-1}\lambda^{t-1}L_1\left\{mF(\omega_1) + pR(F(\omega_t^{\bar{T}}) - F(\omega_1))\right\} \tag{37}$$

Conversely, if a borrower reaches the maximum loan size, expected profits drop for a few different reasons. First, some portion of borrowers will now default strategically, because $\omega^{\bar{L}} < \omega^{\bar{L}-1}$, reflecting the drop in the loan ladder slope from some positive value to flat. Second, because loan sizes are no longer growing, this no longer counteracts the negative pressure from borrowers deciding to default because they satisfy their (randomly allocated) borrowing needs. For all loans after the cap, expected profits per borrower are given by:

$$\begin{aligned} E[\Pi_{t|t>1}|L_t = \bar{L}] =& (1-p)^{t-1}\left\{(1-p)\int_{-\infty}^{\omega^{\bar{L}}} \pi^R(L_t)dF(\omega_i) + (1-p)\int_{\omega^{\bar{L}}}^{\omega^{\bar{L}-1}} \pi^D(L_t)dF(\omega_i)\right\} \\ &+ (1-p)^{t-1}\left\{p\int_{-\infty}^{\omega_t^{\bar{T}}} \pi^R(L_t)dF(\omega_i) + p\int_{\omega_t^{\bar{T}}}^{\omega^{\bar{L}-1}} \pi^D(L_t)dF(\omega_i)\right\} \\ =& (1-p)^{t-1}\bar{L}\left\{(1-p)RF(\omega^{\bar{L}}) - rF(\omega^{\bar{L}-1}) + pRF(\omega_t^{\bar{T}})\right\} \end{aligned} \tag{38}$$

Let $t^{\bar{L}}$ indicate the period a borrower reaches the lender's maximum loan size. Then, total expected profits on all repeat loans are:

$$\sum_{t=2}^{\infty} E[\Pi_{t|t>1}] = \sum_{t=2}^{t^{\bar{L}-1}} E[\Pi_{t|t>1}|L_t < \bar{L}] + \sum_{t=\bar{L}}^{\infty} E[\Pi_{t|t>1}|L_t = \bar{L}] \tag{39}$$

$$= L_1 \sum_{t=2}^{t^{\bar{L}-1}} (1-p)^{t-1}\lambda^{t-1} \left\{ mF(\omega_1) + pR(F(\omega_t^{\bar{T}}) - F(\omega_1)) \right\}$$

$$+ \bar{L} \sum_{t=\bar{L}}^{\infty} (1-p)^{t-1} \left\{ (1-p)RF(\omega^{\bar{L}}) - rF(\omega^{\bar{L}-1}) + pRF(\omega_t^{\bar{T}}) \right\}$$

Total expected profits per borrower are given by the sum of Equations 35 and 39:

$$E[\Pi] = E[\Pi_1] + \sum_{t=2}^{\infty} \beta^{t-1} E[\Pi_{t|t>1}]$$

$$= (1-p)^{t-1}\lambda^{t-1} L_1 \left\{ mF(\omega_1) + pR(F(\omega_t^{\bar{T}}) - F(\omega_1)) \right\} \tag{40}$$

$$+ L_1 \sum_{t=2}^{t^{\bar{L}-1}} \beta^{t-1}(1-p)^{t-1}\lambda^{t-1} \left\{ mF(\omega_1) + pR(F(\omega_t^{\bar{T}}) - F(\omega_1)) \right\}$$

$$+ \bar{L} \sum_{t=\bar{L}}^{\infty} \beta^{t-1}(1-p)^{t-1} \left\{ (1-p)RF(\omega^{\bar{L}}) - rF(\omega^{\bar{L}-1}) + pRF(\omega_t^{\bar{T}}) \right\}$$

8.4 Results

For the simulation exercise, I have to make a number of assumptions about parameter values. Specifically, I impose the maximum loan size observed in the data and offered initial loan sizes, as well as the current average monthly interest rate. I consider profits per borrower over a finite number of loans, ranging from 10 to 30. I assume a cost of funds of 10%. I use the estimated lower bound of 13.6% for p, the probability of a borrower hitting her maximum loan cycle in each period, as my baseline, but also consider additional values ranging from 20%-40%. I then simulate Equation 40 to determine the profit-maximizing λ.

My baseline simulation results (with $p = 13.6\%$) for an upward-sloping loan ladder are shown in Figure 20. Table 12, Panel A contains numerical estimates of the optimal loan ladder slope (λ) , by initial loan size, as well as the corresponding level of profits on the initial loan and in total. Note that while profits vary depending on the total number of loans considered for the simulation, the optimal λ does not.

A few key observations are apparent. First, profits are increasing in initial loan size. This is despite the fact that profits on the initial loan are actually decreasing in initial loan size, consistent with the empirical analysis in Section 4. Second, the optimal loan ladder slope is decreasing in initial loan size. Taken together, these reflect the tradeoff in advancing borrowers quickly up the loan ladder. Quick advancement provides strong incentives for repayment (in this case, any positive λ generates sufficiently strong incentives because of the simplifying assumption of a constant interest rate) but also means borrowers will be more likely to default upon the event they reach the limit of their credit needs or hit the lender's loan ladder cap. A higher initial loan size is equivalent to starting farther up the ladder, so this is balanced by a more gradual optimal rate of loan ladder progression.

I also compute results for alternative assumptions for p, the probability of a borrower hitting her maximum loan cycle in each period. Specifically, I consider $p = \{20\%,\ 30\%,\ 40\%\}$. The motivation for exploring various higher values of p is two-fold. First of all, I am only able to estimate a rough lower bound for p from the data, and so estimates based on $p = 13.6\%$ are likely to be biased. Second, my simplifying assumption that p is constant across loan cycles is likely unrealistic, but necessary for estimation purposes. As borrowers reach higher loan cycles, they are more likely to hit ability-to-pay constraints, which would be captured by an increase in p. Figure 21, Panels A - C and Table 12, Panels B - D contain results for $p = 20\%$, $p = 30\%$, and $p = 40\%$, respectively. The optimal loan ladder slope corresponding to each initial loan size increases slightly between $p = 13.6\%$ and $p = 20\%$, but is then decreasing thereafter. Initial profits and overall profits are both strictly decreasing in p. This reflects that a higher p reduces the lender's incentive to advance borrowers up the loan ladder, because they are now more likely to receive a signal that they no longer need to borrow any longer, and are more likely to default in this event if they are higher up the loan ladder. The level of profits for higher values of p is more consistent with actual profits observed by the lender, suggesting these higher p values may do a better job of capturing actual borrower behavior than the lower bound estimate of p used in the baseline results.

These results imply that, for the average borrower, the lender should choose its largest feasible initial loan size ($10 in this setting) combined with a moderate loan ladder slope. Interestingly, the optimal progressive lending policy for a one-fit-all-setting implied by these results is very close to what the lender follows during the sample period. In the baseline simulations, the optimal policy is an initial loan size of $10 and a loan ladder slope of 1.49. The lender does indeed start most borrowers out with an initial loan size of $10, and the average loan ladder slope is around 1.5 (although it varies over the loan ladder).

9 Conclusion

In this paper, I empirically analyze the impact of the design of dynamic incentive schemes on borrower repayment behavior. I use quasi-experimental variation in an African digital lender's progressive lending policies to isolate the impact of changes in the initial loan size and loan ladder progression on loan-level and borrower-level default rates. I find that giving an exogenously larger loan to new borrowers leads to a higher rate of default on the first loan, consistent with positive moral hazard and repayment burden effects, but no change in a borrower's overall probability of default. By contrast, I show that increasing the size of the second loan relative to the first, or in other words, making the loan ladder steeper, leads to a lower loan-level default rate but a higher probability of eventual default. These results suggest that progressive lending policies cause borrowers to shift their timing of default in meaningful ways. I provide suggestive evidence that the underlying mechanism explaining these results is a strategic repayment motive, whereby some subset of borrowers repays loans in order to obtain larger loans in the future. I then write down and estimate a structural model consistent with this story. I find that there is a large estimated degree of heterogeneity in borrower valuation of access to future loans, a fact that lenders can utilize to mitigate risk if they are able to effectively identify and sort borrowers according to this metric. Finally, I use these results to simulate a profit-maximizing progressive lending scheme for the lender in this setting.

My results have important implications for lenders and regulators. As fintech startups continue to upend consumer lending markets, pushing them to become more geographically decentralized and based on digital data rather than "soft" relationship-based information, lenders are relying increasingly on dynamic incentives and other such tools to encourage repayment. This is particularly true in developing countries, where access to formal sector loans is generally lower. My results suggest that the stage of the relationship matters significantly for the impact of various policies on outcomes, something lenders should take into account when designing their lending schemes. My results also demonstrate that progressive lending can have potentially perverse effects on borrower behavior, encouraging certain borrowers to try to take advantage of the system. Lenders need to implement carefully-designed, risk-based dynamic schemes that encourage strategic types to exit the system at low loan sizes by restricting loan size growth for these borrowers, while allowing better borrowers to continue to progress up to larger, more profitable loan sizes. From a regulatory perspective, more robust credit information sharing systems, which would increase the cost to borrowers of such a strategy by fully excluding defaulters from any credit - rather than just credit from a particular lender - would also help to ameliorate this issue. Even in settings where credit reporting bureaus are present, it is important for all lenders to be required to contribute both positive and negative reports and borrowers to be made aware of the importance of remaining in good standing.

References

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

Albuquerque, R. & Hopenhayn, H. A. (2004), 'Optimal lending contracts and firm dynamics', *The Review of Economic Studies* **71**(2), 285–315.

Alexander Tedeschi, G. (2006), 'Here today, gone tomorrow: Can dynamic incentives make microfinance more flexible?', *Journal of Development Economics* **80**(1), 84–105.

Armendáriz, B. & Morduch, J. (2010), *The Economics of Microfinance*, MIT press.

Banerjee, A. V. (2013), 'Microcredit under the microscope: What have we learned in the past two decades, and what do we need to know?', *Annual Review of Economics* **5**(1), 487–519.

Besley, T. (1995), 'Savings, credit and insurance', *Handbook of Development Economics* **3**, 2123–2207.

Bjorkegren, D. & Grissen, D. (2015), 'Behavior revealed in mobile phone usage predicts loan repayment'.

Bond, P. & Krishnamurthy, A. (2004), 'Regulating exclusion from financial markets', *The Review of Economic Studies* **71**(3), 681–707.

Breza, E. (2013), 'Peer effects and loan repayment: Evidence from the krishna default crisis'.

Bulow, J. & Rogoff, K. (1989), 'Sovereign debt: Is to forgive to forget?', *The American Economic Review* **79**(1), 43.

Calonico, S., Cattaneo, M., Farrell, M. & Titiunik, R. (2017), 'Rdrobust: Software for regression discontinuity designs', *Stata Journal, Forthcoming* .

Calonico, S., Cattaneo, M., Titiunik, R. et al. (2014), 'Robust data-driven inference in the regression-discontinuity design', *Stata Journal* **14**(4), 909–946.

Cattaneo, M., Jansson, M. & Ma, X. (2017), 'Simple local polynomial density estimators'.

CGAP (2017*a*), 'Consumer protection in digital credit'.

CGAP (2017*b*), 'Digital credit's evolving landscape: 3 things you need to know'.

De Giorgi, G., Drenik, A. & Seira, E. (2015), 'Moral hazard in the credit market'.

de Mel, S., McKenzie, D. & Woodruff, C. (2008), 'Returns to capital in microenterprises: Evidence from a field experiment', *The Quarterly Journal of Economics* **123**(4), 1329–1372.

Dobbie, W. & Skiba, P. M. (2013), 'Information asymmetries in consumer credit markets: Evidence from payday lending', *American Economic Journal: Applied Economics* **5**(4), 256–282.

Eaton, J., Gersovitz, M. et al. (1981), 'Debt with potential repudiation: Theoretical and empirical analysis', *Review of Economic Studies* **48**(2), 289–309.

Feigenberg, B., Field, E. & Pande, R. (2013), 'The economic returns to social interaction: Experimental evidence from microfinance', *Review of Economic Studies* **80**(4), 1459–1483.

Field, E. & Pande, R. (2008), 'Repayment frequency and default in microfinance: Evidence from india', *Journal of the European Economic Association* **6**(2-3), 501–509.

Gelman, A. & Imbens, G. (2014), Why high-order polynomials should not be used in regression discontinuity designs, Technical report, National Bureau of Economic Research, Inc.

Ghosh, P. & Ray, D. (2016), 'Information and enforcement in informal credit markets', *Economica* **83**(329), 59–90.

Giné, X., Goldberg, J. & Yang, D. (2012), 'Credit market consequences of improved personal identification: Field experimental evidence from malawi', *The American Economic Review* **102**(6), 2923–2954.

Giné, X., Jakiela, P., Karlan, D. & Morduch, J. (2010), 'Microfinance games', *American Economic Journal: Applied Economics* **2**(3), 60–95.

Giné, X. & Karlan, D. S. (2014), 'Group versus individual liability: Short and long term evidence from philippine microcredit lending groups', *Journal of Development Economics* **107**, 65–83.

Hulme, D. & Mosley, P. (1996), *Finance Against Poverty*, Vol. 2, Psychology Press.

Hussam, R., Rigol, N. & Roth, B. (2017), 'Targeting high ability entrepreneurs using community information: Mechanism design in the field'.

Imbens, G. & Kalyanaraman, K. (2012), 'Optimal bandwidth choice for the regression discontinuity estimator', *The Review of Economic Studies* **79**(3), 933–959.

Karaivanov, A. & Townsend, R. M. (2014), 'Dynamic financial constraints: Distinguishing mechanism design from exogenously incomplete regimes', *Econometrica* **82**(3), 887–959.

Karlan, D. & Zinman, J. (2009), 'Observing unobservables: Identifying information asymmetries with a consumer credit field experiment', *Econometrica* **77**(6), 1993–2008.

McCrary, J. (2008), 'Manipulation of the running variable in the regression discontinuity design: A density test', *Journal of Econometrics* **142**(2), 698–714.

McKinsey (2015), 'Cutting through the fintech noise: Markers of success, imperatives for banks'.

McKinsey (2017), 'Fintech: Opportunities and challenges for banks and regulators'.

Morduch, J. et al. (1999), 'The microfinance promise', *Journal of Economic Literature* **37**(4), 1569–1614.

Shapiro, D. (2015), 'Microfinance and dynamic incentives', *Journal of Development Economics* **115**, 73–84.

Suri, T. (2017), 'Mobile money', *Annual Review of Economics* **9**, 497–520.

Zinman, J. (2014), 'Consumer credit: Too much or too little (or just right)?', *The Journal of Legal Studies* **43**(S2), S209–S237.

Appendix

A Specification Checks

A.1 Initial Loan Size

In this section, I present a series of specification checks to address potential threats to the causal interpretation of my results from Section 4. I first check for any evidence of selective borrowing based on loan size eligibility in my sample. This would invalidate my regression discontinuity results by creating additional discontinuities in borrower characteristics around the loan size threshold. For instance, if unobservably worse applicants were more likely not to reapply for the smaller loan, this could explain why I find lower default rates just below the threshold relative to just above. To explore this issue, I first compute the percentage of applicants rejected for the larger loan who chose not to reapply for the smaller loan despite being eligible. This amounts to only about 5% of applicants. This suggests that, even if there were large differences in borrower qualities between those who reapplied and those who didn't, the magnitude of dropouts is too small to drive my results. Second, I compare the distribution of risk scores based on reapplication choice. As shown in Figure 22, the two distributions are very similar. While I cannot directly test for any differences in unobservable risk characteristics between the two groups, the fact that they are very similar on observable risk measures is reassuring.

My second set of specification checks concerns the possibility of manipulation of the risk score by borrowers. If applicants could influence the risk score and were differentially able to do so, this again could create discontinuities in unobservable borrower characteristics around the loan eligibility threshold. In my setting, this type of manipulation is very unlikely: applicants aren't even told that a risk score is computed, and have no knowledge of how it is computed. In addition, many of the variables that enter into the risk score are inherently unable to be manipulated. Figure 23 provides visual evidence confirming the lack of bunching above the threshold. I also conduct a formal manipulation test based on density discontinuities, as first proposed by McCrary (2008), using a nonparametric density estimator proposed in Cattaneo et al. (2017). Formally, this procedure tests the null hypothesis that the density of the running variable (the risk score in this setting) is continuous at the threshold. Reassuringly, the test produces a T-statistic of 1.45 with an associated p-value of .15, such that I am not able to reject the null.

Next, I check that my results identify real discontinuities caused by the change in initial loan size receipt by replicating the analysis using a sample of borrowers from a later time period, when the lender had eliminated its new borrower sorting policy. Figure 24, Panel A depicts the first-stage relationship between the risk score (relative to the $10 threshold) and initial loan size during my sample period and the later period. During the later period, no borrowers are required to take out a smaller loan because of an insufficiently high risk score, so nearly all borrowers end up with the larger $10 loan. This means there is no longer a policy-induced discontinuity in loan size at the original risk score threshold. Table 13 presents a formal estimation. While the effect is significant under the linear specification, the magnitude is not economically meaningful. Figure 24, Panel B shows the reduced-form relationship between risk score and default. It provides graphical evidence that the sharp discontinuity in default rates at the threshold during my sample period disappears in the later, no-sorting period. It also confirms the non-linear relationship between risk score and default, and in particular the higher predictive power of the risk score above the threshold relative to below. Because of the lack of a meaningful first stage, I do not present formal IV estimates.

Finally, I confirm that my results are not unique to the specific time period and set of borrowers I use in my main analysis. To do so, I replicate the analysis using data from a later period in which the lender reinstituted its policy of sorting borrowers into initial loan sizes according to risk scores. One small difference during this period is that borrowers who qualified for larger loans had the option to choose a smaller loan if they so desired, although in practice very few do so. Panels A and B of Figure 25 replicate the first-stage and reduced-form figures, respectively, using this alternative sorting period. As in the earlier period, there is a sharp jump in both average loan size and the probability of default exactly at the risk score threshold. Tables 14 and 15, which provide formal regression estimates of the first stage and IV, confirm this conclusion.[25] Borrowers with a risk score above the $10 threshold receive a loan that is about $4.10 larger on average, or equivalently are about 80% more likely to take out the $10 initial loan product; as a result, they are 8.2 to 9.6 ppts more likely to default, relative to an average default rate during this period of 16.9%. Therefore, receiving a loan that is two times larger leads to a 49 to 57 percent increase in default, identical in direction and similar in magnitude to my main results.

A.2 Loan Ladder Progression

In this section, I present some checks on the interpretation of my results from Section 5. The first concern I address is selective borrowing, which would be problematic if borrowers differentially decide to take out a second loan depending on whether they qualify for a larger loan relative to their first loan or not. Note that, in order to explain the results on repeat borrowers' loan ladder progression, selective borrowing would

[25] Note that I do not present local linear estimates for this exercise because of the small range of observations below the threshold.

actually need to work in the opposite direction as in the analysis of initial loan size in Section 4. Among new borrowers, unobservably riskier applicants would need to be less likely to apply for a smaller loan relative to a larger loan to explain the lower relative default rate on smaller loans around the loan eligibility threshold. Among repeat borrowers, unobservably riskier applicants would need to be *more* likely to reapply for the smaller loan that they qualified for during the "off" period relative to the larger loan during the "on" periods in order to explain the observed pattern of default. This type of selective borrowing seems highly unlikely. And even if it were plausible, very few borrowers end their relationship with the lender after successfully repaying their first loan. Figure 26 shows the probability of reapplying and taking out another loan conditional on being upgraded or remaining on the same level relative to the previous loan, by loan cycle. In the primary analysis, I focus only on second-time borrowers. For these borrowers, the probabilities of reapplying and successfully getting another loan, respectively, are 98.4% and 98.0% if upgraded and 95.8% and 94.3% if they stay on the same level. While borrowers are slightly less likely to reapply if they aren't upgraded, overall re-application and re-approval rates are too high for this to pose any empirical issues. By contrast, reapplication rates are much lower if borrowers are downgraded. This is a primary reason for not studying the impact of downgrading on borrower behavior, as selective borrowing would present a much larger issue for inference.

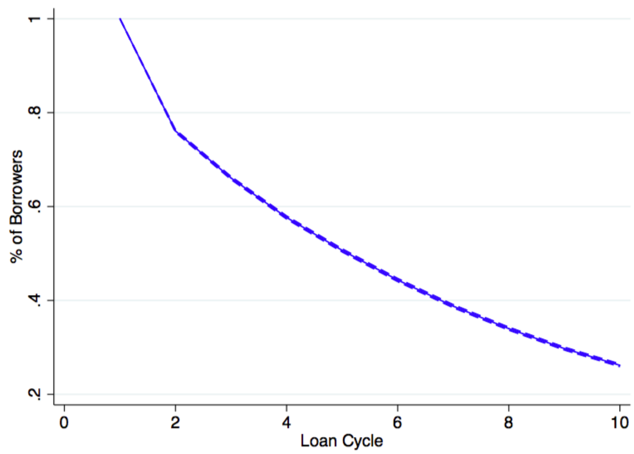
Next, I present some additional suggestive evidence that corroborates my results. As mentioned previously, the OTP upgrade threshold was set at $P\%$ prior to November 2015. This means there is an additional "off" period in which borrowers in the $(P - 25)\%$ to $P\%$ were not upgraded, as shown in Panel A of Figure 27. Unfortunately for my purposes, low loan volume means the data are quite noisy, particularly when I split the sample, so I exclude them from my core results. However, my regression results are robust to the inclusion of weeks in this period with at least 50 loans in both the treatment and control groups, as shown in Table 16. Graphical evidence in Panel B of Figure 27 provides suggestive evidence of the same pattern observed in the later data.

Unfortunately, because the lender has always utilized the OTP threshold system to regulate movement up and down the loan ladder, it is not possible to use an alternative time period when the policy was not in effect (and thus when the DD estimate should be null) as a robustness check. Likewise, it is not possible to construct an alternative control group as an additional robustness check. The only other logical alternative control would be a set of borrowers who face a flat loan ladder regardless of period. This would then identify the upgrading shock that the treatment group experiences in the "on" periods relative to a control group that is never upgraded or downgraded, rather than a control group who is always upgraded. However, there is no OTP percentage range for whom this holds.

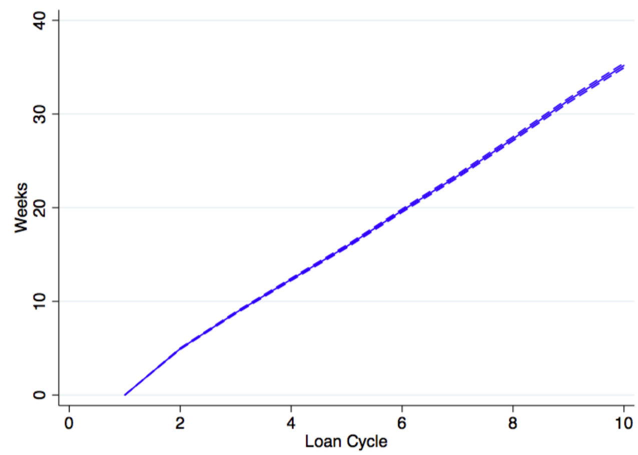
Figures

Figure 1: Key Moments by Loan Cycle

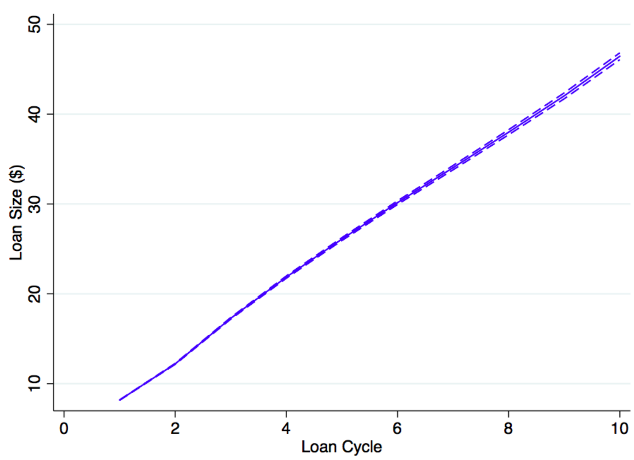
(a) Fraction Still Borrowing



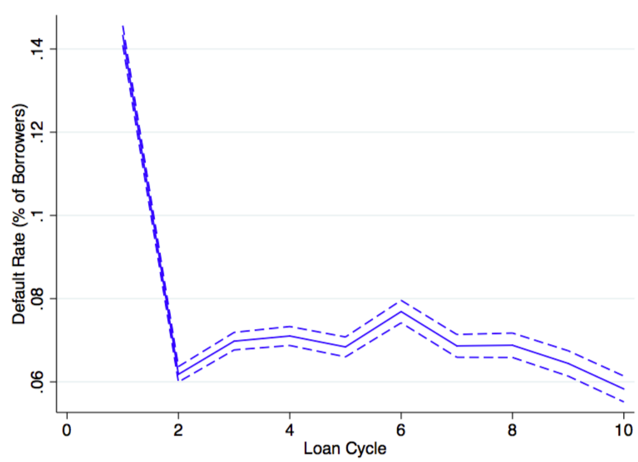
(b) Weeks Since First Loan Disbursement



(c) Loan Size

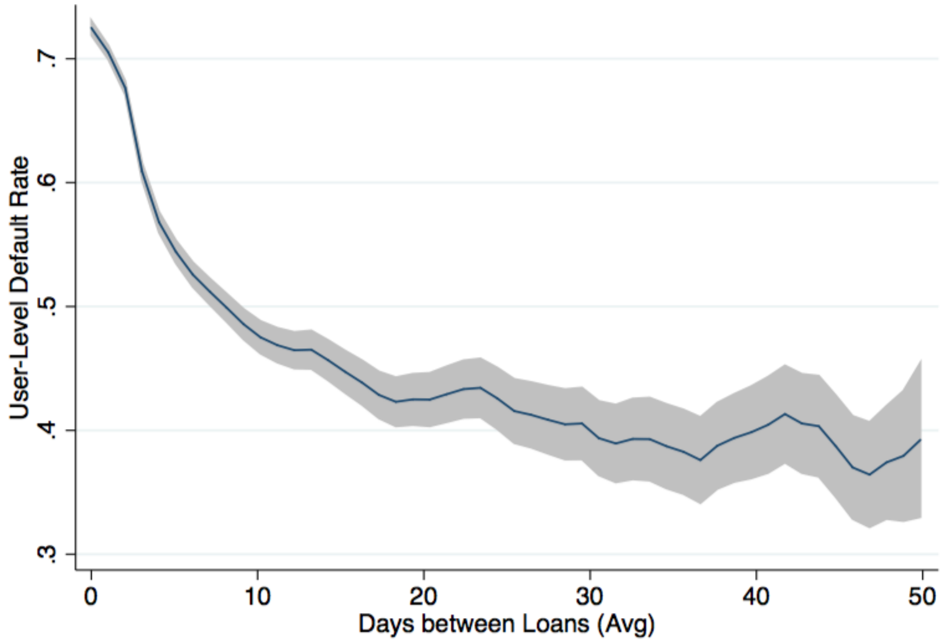


(d) Default Rate



This figure presents some key moments related to borrowing behavior delineated by loan cycle. Solid lines indicate means, while dashed lines indicate 95th percent confidence intervals. Loan cycle refers to the number of loans a borrower has taken out thus far. Fraction still borrowing refers to the percentage of borrowers who are still active at a given loan cycle. Weeks since first loan disbursement is calculated as the number of weeks between the disbursement of the first loan and the disbursement of the n-th loan, where n refers to the loan cycle indicated. Loan size is the amount of principal disbursed. The default rate is measured as the percentage of borrowers at that loan cycle with loans that reach 90 days past due.

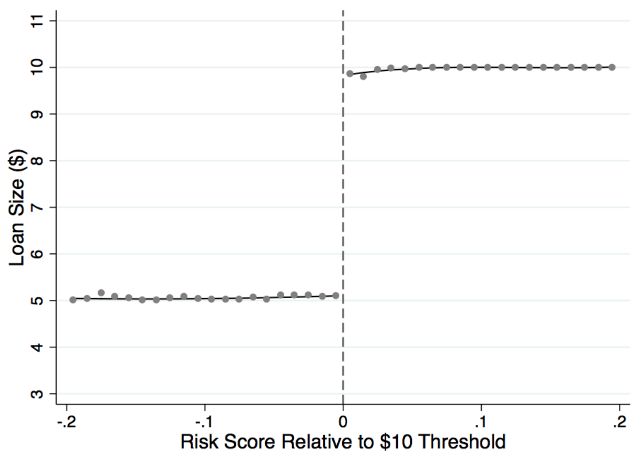
Figure 2: Default Rates by Loan Turnover Rate



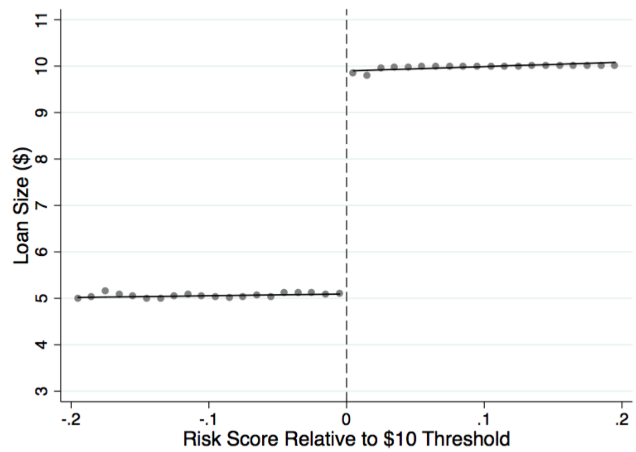
This figure displays a kernal-weighted local polynomial regression of the average borrower-level default rate on the average number of days a borrower waits between loans. The solid line indicates the smoothed values, and the shaded gray area indicates the 95th percent confidence interval. Borrower-level default rates are calculated as the percentage of borrowers who ever default. A borrower is deemed to have defaulted if her current loan reaches 90 days past due. Because of the lender's policy of full exclusion of defaulters, this is analagous to borrowers who default on their last loan with the lender. The number of days between loans is calculated as the number of days between the last payment made on a borrower's previous loan and the disbursement of her next loan.

Figure 3: Initial Loan Size: First Stage, Loan Size

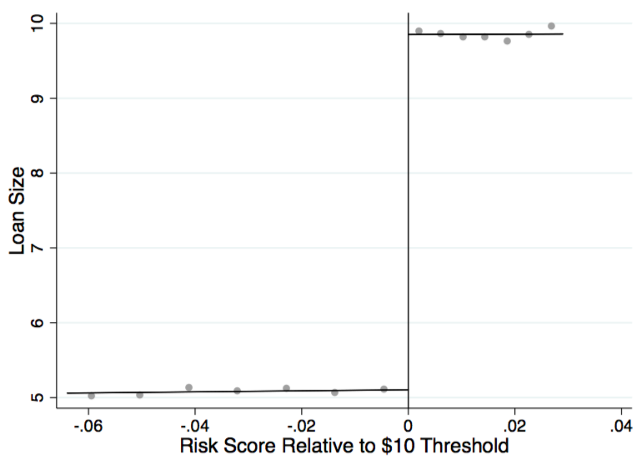
(a) Global Polynomial



(b) Global Linear

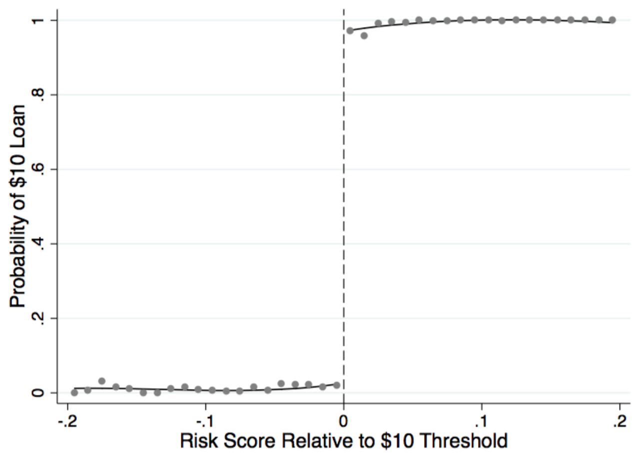


(c) Local Linear

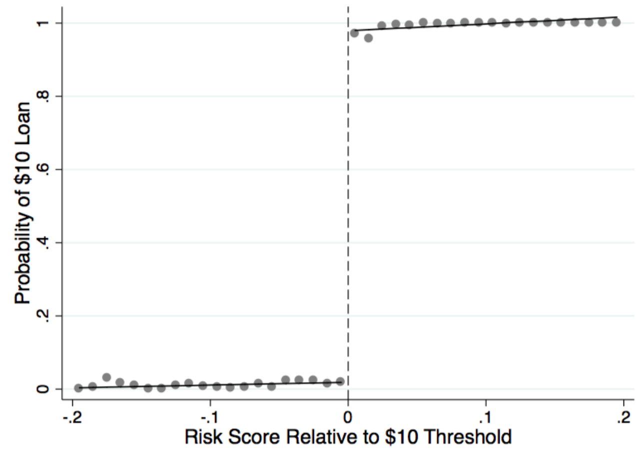


This figure depicts the first-stage relationship between the risk score threshold and the average initial loan size received by new borrowers during the sample period. The risk score is an internally-calculated measure of borrower credit-worthiness. There were only two initial loan sizes during this period, equal approximately to $5 or $10. Each panel plots average loan sizes across risk score bins, where the risk score is measured relative to the threshold. It is presented in this way because the actual value of the risk score did not have any meaning during this period; it was only relevant in terms of how it compared to the threshold. Panel A plots average loan sizes across risk score bins with a width of .01 units (about 2% of the total range of risk scores during this period), as well as fitted values from a first-stage regression with a third-order polynomial to control for the risk score, where the shape is allowed to vary on either side of the threshold. Panel B instead uses a linear trend, interacted with an indicator of being above the cutoff. Panel C shows the local linear approximation within the optimal data-determined bandwidths above and below the cutoff.

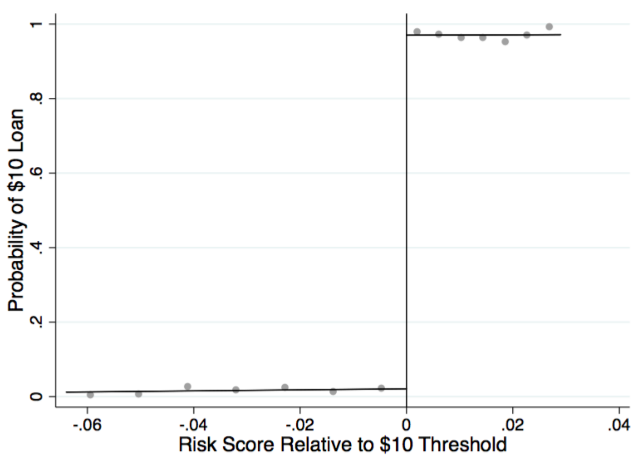
Figure 4: Initial Loan Size: First Stage, $10 Indicator



(a) Global Polynomial



(b) Global Linear

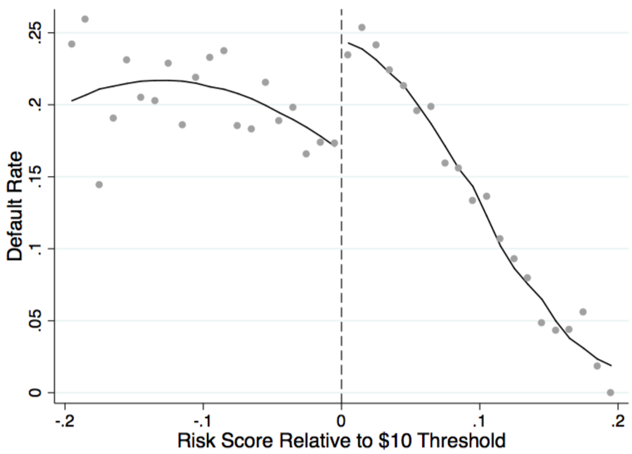


(c) Local Linear

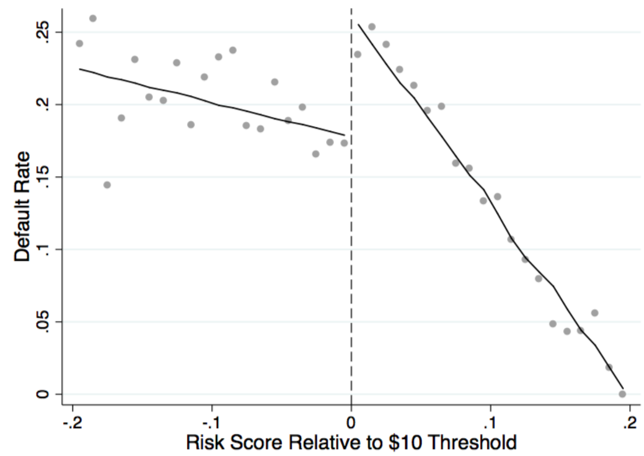
This figure depicts the first-stage relationship between the risk score threshold and the probability that a new borrower receives the larger initial loan of $10 during the sample period. The risk score is an internally-calculated measure of borrower credit-worthiness. There were only two initial loan sizes during this period, equal approximately to $5 or $10. Each panel plots average probabilities across risk score bins, where the risk score is measured relative to the threshold. It is presented in this way because the actual value of the risk score did not have any meaning during this period; it was only relevant in terms of how it compared to the threshold. Panel A plots probabilities of getting the larger loan across risk score bins with a width of .01 units (about 2% of the total range of risk scores during this period), as well as fitted values from a first-stage regression with a third-order polynomial to control for the risk score, where the shape is allowed to vary on either side of the threshold. Panel B instead uses a linear trend, interacted with an indicator of being above the cutoff. Panel C shows the local linear approximation within the optimal data-determined bandwidths above and below the cutoff.

Figure 5: Initial Loan Size: Reduced Form, Loan-Level Default Rate

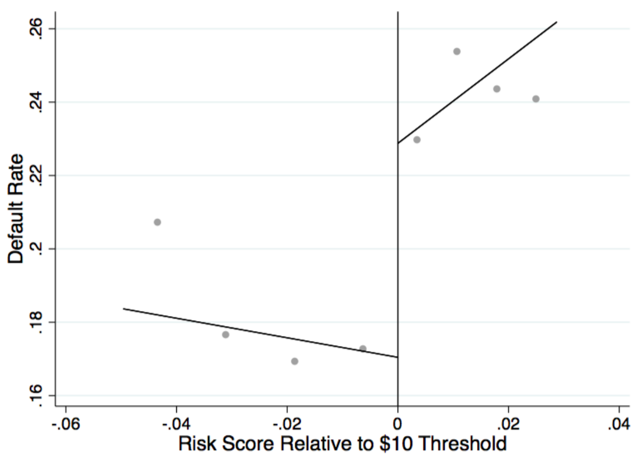
(a) Global Polynomial



(b) Global Linear



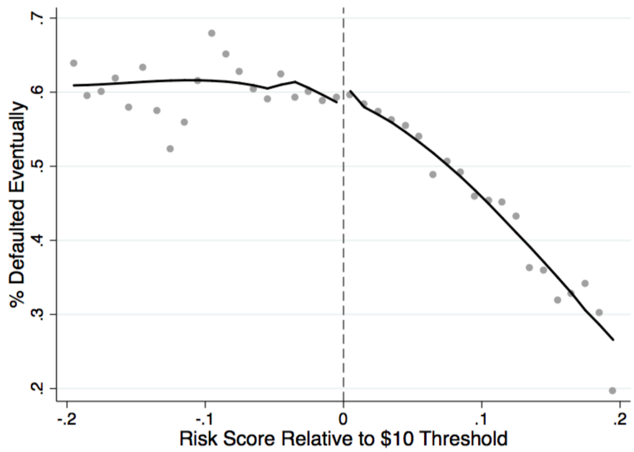
(c) Local Linear



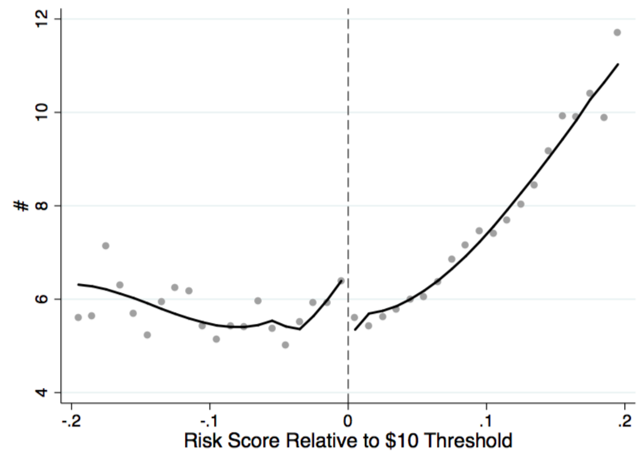
This figure depicts the reduced-form relationship between the risk score threshold and average default rates on loans to new borrowers during the sample period. The risk score is an internally-calculated measure of borrower credit-worthiness. The average default rate is calculated as the percentage of loans that reach 90 days past due. Each panel plots average default rates across risk score bins, where the risk score is measured relative to the threshold. It is presented in this way because the actual value of the risk score did not have any meaning during this period; it was only relevant in terms of how it compared to the threshold. Panel A plots average default rates across risk score bins with a width of .01 units (about 2% of the total range of risk scores during this period), as well as fitted values from a reduced-form regression with a third-order polynomial to control for the risk score, where the shape is allowed to vary on either side of the threshold. Panel B instead uses a linear trend, interacted with an indicator of being above the cutoff. Panel C shows the local linear approximation within the optimal data-determined bandwidths above and below the cutoff.

Figure 6: Initial Loan Size: Reduced Form, Longer-Term Credit Outcomes

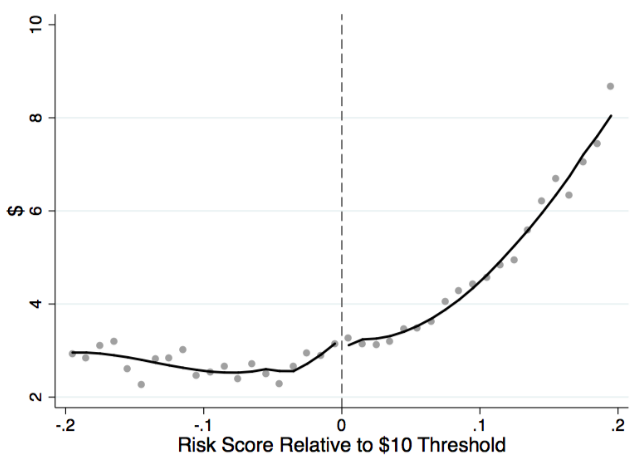
(a) Borrower-Level Default Rate



(b) Number of Loans

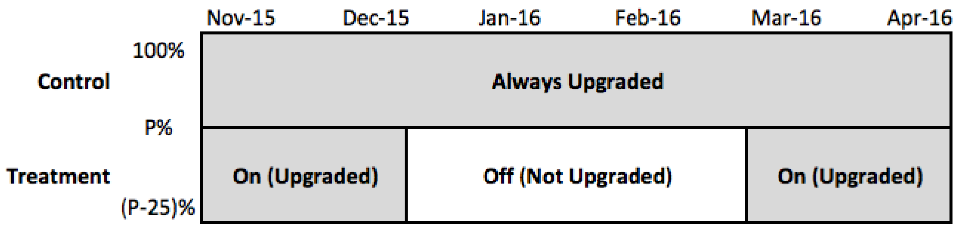


(c) Last Loan Size



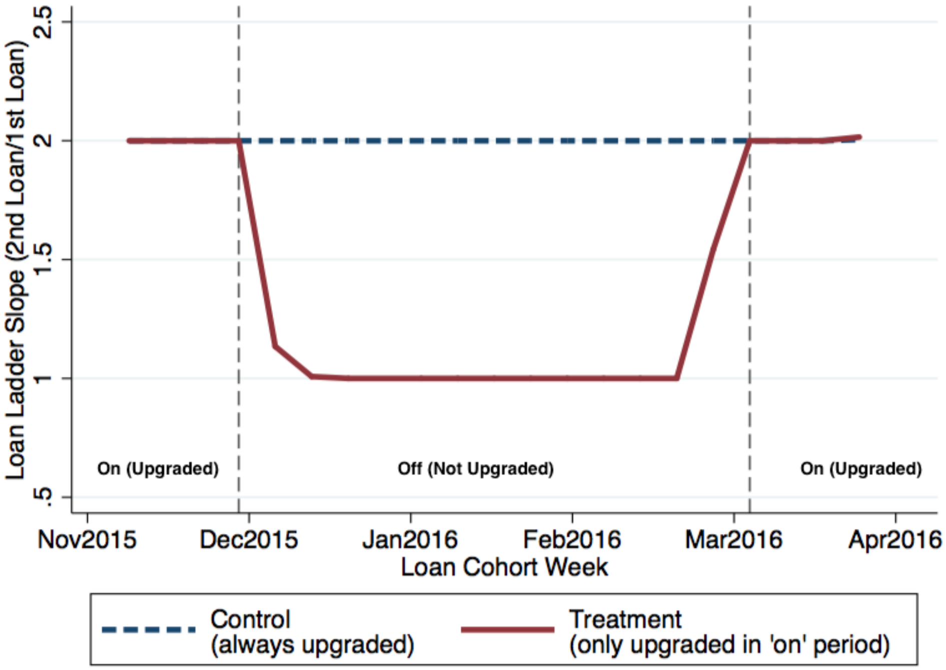
This figure depicts the reduced-form relationship between the risk score threshold for new borrowers and longer-term credit outcomes. Each outcome is measured as of September 2017, representing an average of 85 weeks since a borrower's first loan was disbursed. Each panel plots average values of the outcome variable across risk score bins with a width of .01 units (about 2% of the total range of risk scores during this period), where the risk score is measured relative to the threshold. Each panel also includes fitted values from a reduced-form regression with a third-order polynomial to control for the risk score, where the shape is allowed to vary on either side of the threshold. The risk score is an internally-calculated measure of borrower credit-worthiness. Borrower-level default rates are calculated as the percentage of borrowers who ever default, where a borrower is deemed to have defaulted if her current loan reaches 90 days past due. The number of loans refers to the total number of loans taken out by a borrower from this lender. The last loan size is the amount disbursed on the final loan recorded for each borrower.

Figure 7: Loan Ladder Progression: Identification Strategy



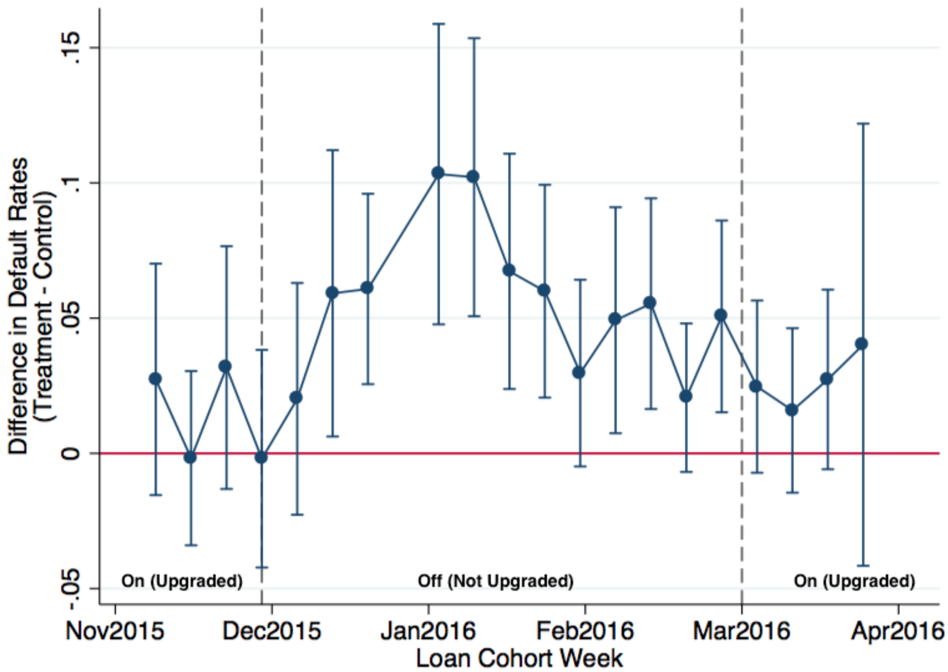
This figure depicts the identification strategy for the loan ladder progression difference-in-differences analysis. The control group is composed of borrowers with an on-time payment percentage on their first loan of greater than $P\%$. These borrowers were always upgraded to a larger loan size for their second loan. The treatment group is made up of borrowers with an on-time payment percentage on their first loan of between $(P-25)\%$ and $P\%$. These borrowers were upgraded to a larger loan in November 2015 and March 2016, but were forced to take out a second loan of the same size as their first loan if they reapplied between December 2015 and February 2016.

Figure 8: Loan Ladder Progression: Loan Ladder Slope (Ratio of Second Loan Size to First Loan Size)



This figure depicts the ratio of the size of a borrower's second loan to the size of her first loan, split by treatment and control groups in the loan ladder progression quasi-experiment. The control group is composed of second-time borrowers who took out a \$10 loan initially and repaid a sufficient number of installments of their first loan such that they were always upgraded to a second loan of \$20. The treatment group is composed of second-time borrowers who also took out a \$10 loan initially, but whose on-time payment record on their first loan was in a lower range. For this group, whether they were upgraded to a larger second loan of \$20 or were required to take a second loan that was the same size as their first loan depended on when they applied for their second loan, due to variation over time in the lender's upgrading policies. The "off" period corresponds to when treatment borrowers were not upgraded to a larger second loan, and thus received a second loan that was the same size as their first; the "on" periods correspond to when treatment borrowers were upgraded to a larger second loan.

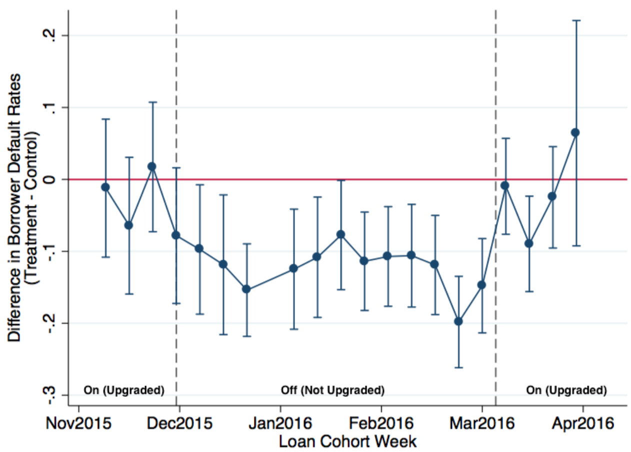
Figure 9: Loan Ladder Progression: Difference in Default Rates (Treatment - Control)



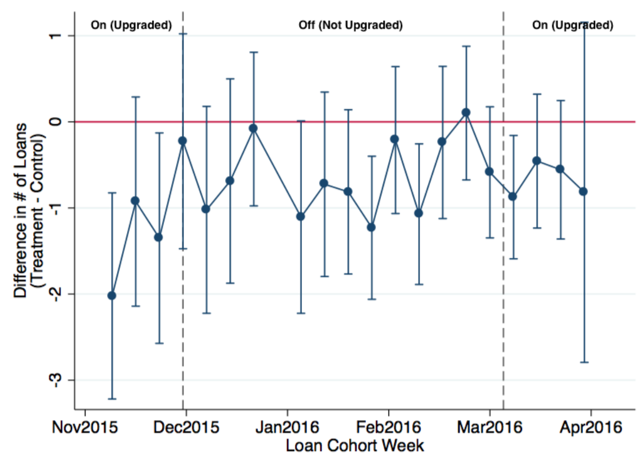
This figure depicts the estimated difference in second-loan default rates between the treatment and control groups by loan cohort week, along with the corresponding 95 percent confidence intervals. The control group is composed of second-time borrowers who took out a $10 loan initially and repaid a sufficient number of installments of their first loan such that they were always upgraded to a second loan of $20. The treatment group is composed of second-time borrowers who also took out a $10 loan initially, but whose on-time payment record on their first loan was in a lower range. For this group, whether they were upgraded to a larger second loan of $20 or were required to take a second loan that was the same size as their first loan depended on when they applied for their second loan, due to variation over time in the lender's upgrading policies. The "off" period corresponds to when treatment borrowers were not upgraded to a larger second loan, and thus received a second loan that was the same size as their first; the "on" periods correspond to when treatment borrowers were upgraded to a larger second loan. A loan is marked as defaulted when it is 90 days past due.

Figure 10: Loan Ladder Progression: Other Longer-Term Credit Outcomes

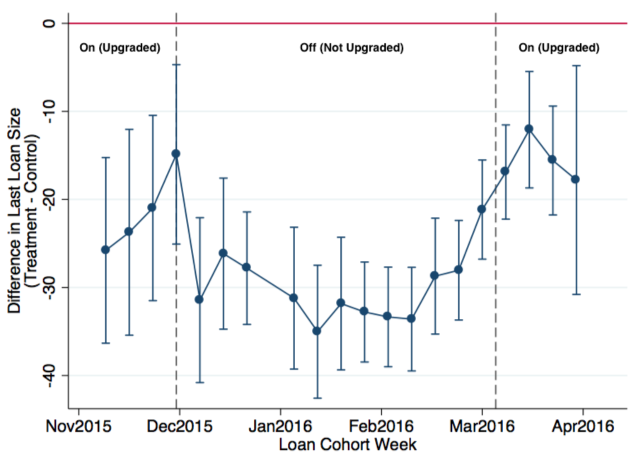
(a) Borrower-Level Default Rate



(b) Number of Loans

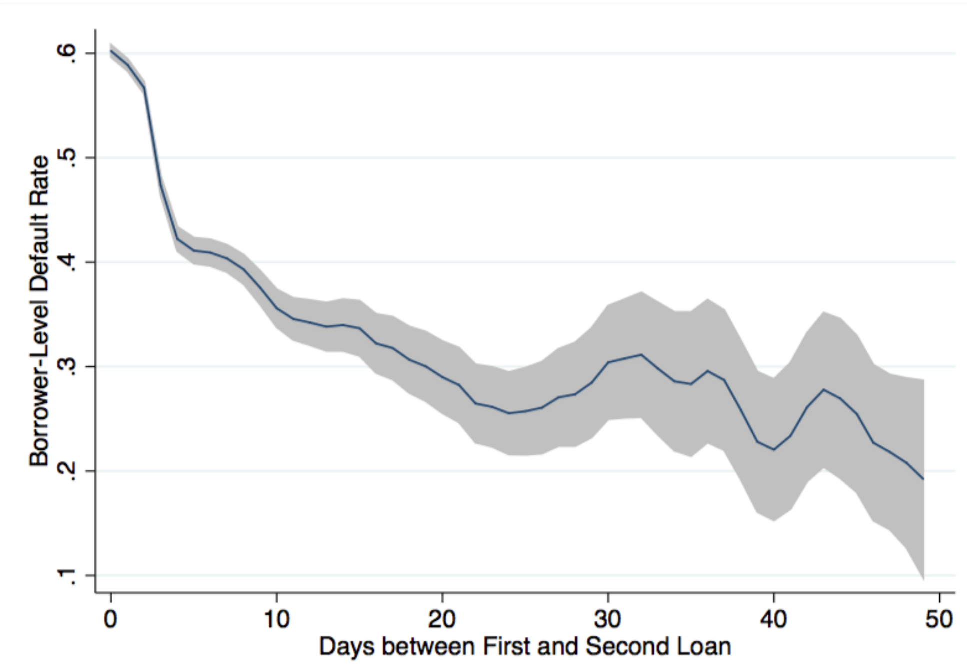


(c) Last Loan Size



This figure depicts the estimated difference in longer-term credit outcomes between the treatment and control groups by loan cohort week, along with the corresponding 95 percent confidence intervals. Each outcome is measured as of September 2017, representing an average of 88 weeks since a borrower's first loan was disbursed. The control group is composed of second-time borrowers who took out a \$10 loan initially and repaid a sufficient number of installments of their first loan such that they were always upgraded to a second loan of \$20. The treatment group is composed of second-time borrowers who also took out a \$10 loan initially, but whose on-time payment record on their first loan was in a lower range. For this group, whether they were upgraded to a larger second loan of \$20 or were required to take a second loan that was the same size as their first loan depended on when they applied for their second loan, due to variation over time in the lender's upgrading policies. The "off" period corresponds to when treatment borrowers are not upgraded to a larger second loan, and thus receive a second loan that is the same size as their first; the "on" periods correspond to when treatment borrowers are upgraded to a larger second loan. Borrower-level default rates are calculated as the percentage of borrowers who ever default, where a borrower is deemed to have defaulted if her current loan reaches 90 days past due. The number of loans refers to the total number of loans taken out by a borrower from this lender. The last loan size is the amount disbursed on the final loan recorded for each borrower.

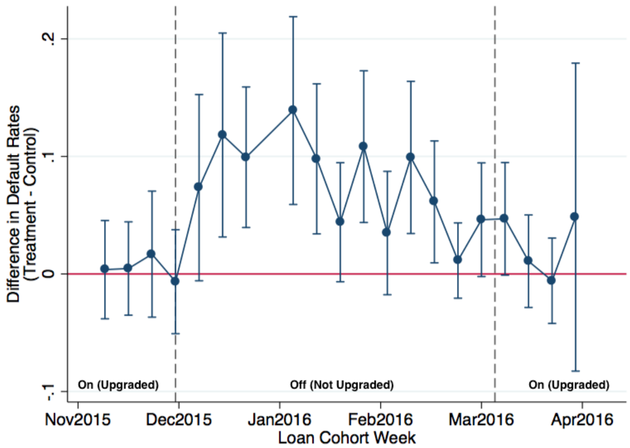
Figure 11: Loan Ladder Progression: Default Rates by Loan Turnover Rate between First and Second Loans



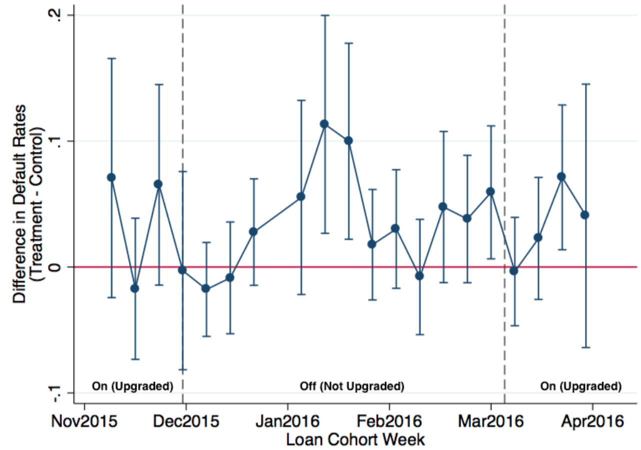
This figure displays a kernal-weighted local polynomial regression of the average borrower-level default rate on the average number of days a borrower waits between the first and second loan. The solid line indicates the smoothed values, and the shaded gray area indicates the 95th percent confidence interval. Borrower-level default rates are calculated as the percentage of borrowers who ever default. A borrower is deemed to have defaulted if her current loan reaches 90 days past due. Because of the lender's policy of full exclusion of defaulters, this is analagous to borrowers who default on their last loan with the lender. The number of days between the first and second loan is calculated as the number of days between the last payment made on a borrower's first loan and the disbursement of her second loan.

Figure 12: Loan Ladder Progression: Default Rates, Split by Loan Turnover Rate

(a) Days between Loans <= Median



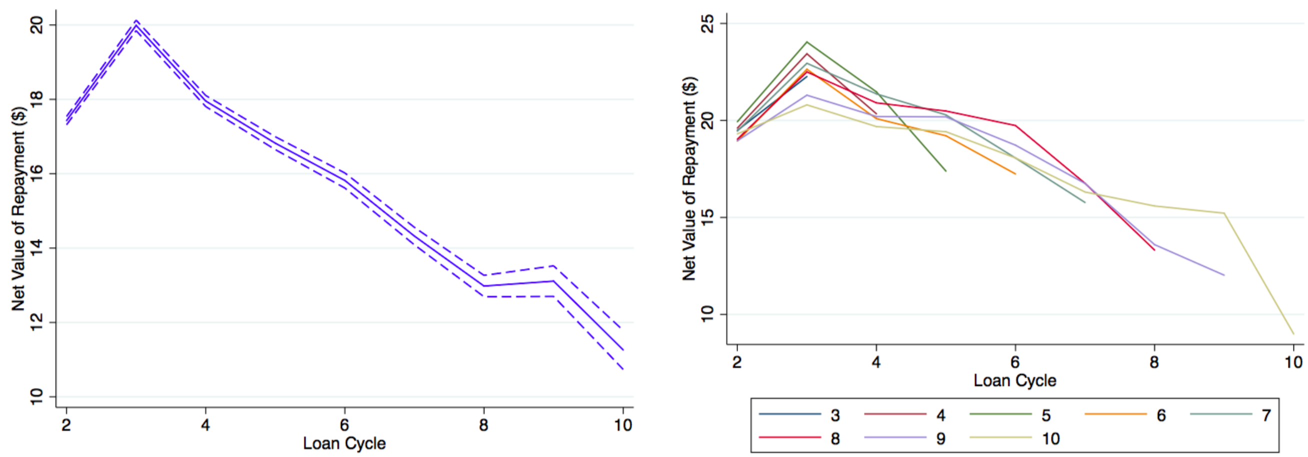
(b) Days between Loans > Median



This figure depicts the estimated difference in second-loan default rates between the treatment and control groups by loan cohort week, split by the loan turnover rate, along with the corresponding 95 percent confidence intervals. Panel A shows the difference in default rates for "fast turnover" borrowers, defined as those who took out their second loan within the median time of one day after repaying their first loan. Panel B replicates the analysis for the "slow turnover" borrowers, corresponding to those who waited longer than one day after repaying their first loan to take out their second. A loan is marked as defaulted when it is 90 days past due. The control group is composed of second-time borrowers who took out a $10 loan initially and repaid a sufficient number of installments of their first loan such that they were always upgraded to a second loan of $20. The treatment group is composed of second-time borrowers who also took out a $10 loan initially, but whose on-time payment record on their first loan was in a lower range. For this group, whether they were upgraded to a larger second loan of $20 or were required to take a second loan that was the same size as their first loan depended on when they applied for their second loan, due to variation over time in the lender's upgrading policies. The "off" period corresponds to when treatment borrowers are not upgraded to a larger second loan, and thus receive a second loan that is the same size as their first; the "on" periods correspond to when treatment borrowers are upgraded to a larger second loan.

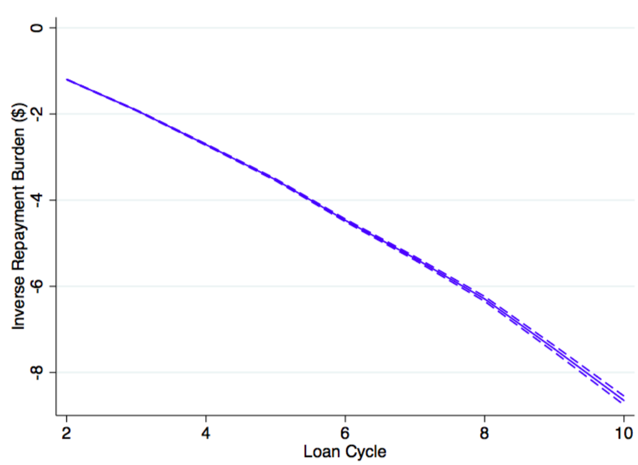
Figure 13: Structural Estimation: Unconditional Net Value of Repayment (ω_t)

(a) Unconstrained Net Value of Repayment, by Loan Cycle (b) Unconstrained Net Value of Repayment, by Defaulted Loan Cycle



This figure depicts how the unconditional net value of repayment (ω_t) varies by loan cycle. ω_t is defined as the net outside option for a borrower who is just indifferent between defaulting in period t and period $t+1$, conditional on not reaching the maximum required loan cycle. As derived in Section 6 and shown in Equation 17, ω_t is a function of loan size L_t, the interest rate R_t, expected loan size growth λ^e_{t+1}, the discount factor δ, and the marginal product of capital draw MPK_i. It can therefore be calculated for each loan, conditional on assumptions about λ^e_{t+1}, δ, and MPK_i. I assume $\lambda^e_{t+1} = \frac{L_t}{L_{t-1}}$, $\delta = .9$, and $MPK_i = 15\%$. Panel A shows the average value of ω_t for loans taken out at a given loan cycle. Panel B shows average values of ω_t across loan cycles, conditional on the loan cycle at which a borrower eventually defaults. For both figures, I exclude the first loan cycle because of the difference in assumptions about expected loan growth λ^e_{t+1}. Units for ω_t are in terms of dollars.

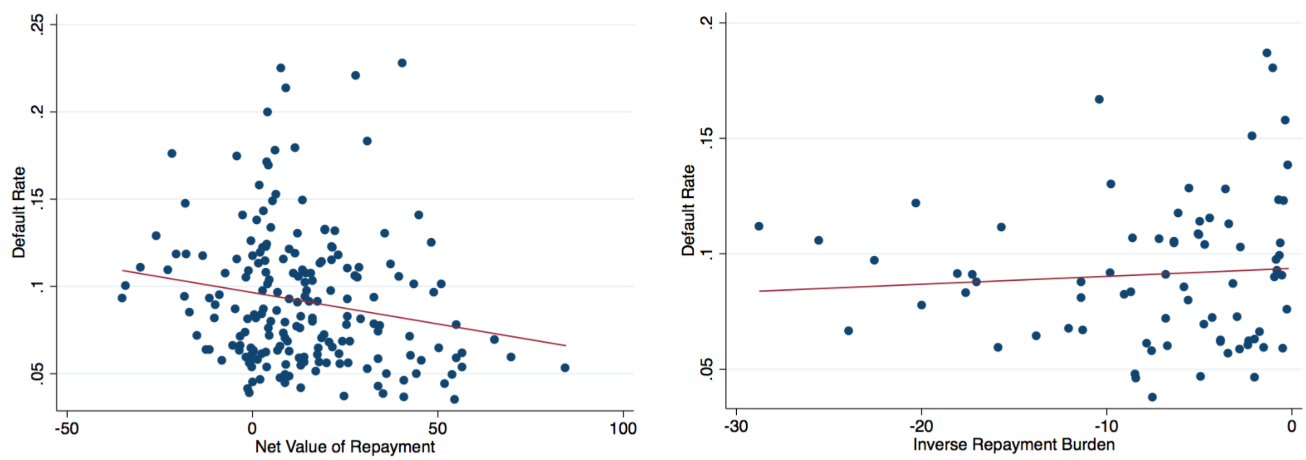
Figure 14: Structural Estimation: Inverse Repayment Burden ($\omega_t^{\bar{T}}$)



This figure shows how the inverse of the repayment burden, $\omega_t^{\bar{T}}$, evolves by loan cycle. In each period, each borrower has some (unknown) probability of receiving a signal that she has reached her maximum loan cycle. Borrowers who find out they are at their maximum loan cycle will default if their net outside option is greater than $\omega_t^{\bar{T}}$, or equivalently when their net outside option is greater than the total cost of repaying the loan. As derived in Section 6 and shown in equation 21, $\omega_t^{\bar{T}}$ is a function of loan size L_t and the interest rate R_t, and thus can be calculated for each loan. Units for $\omega_t^{\bar{T}}$ are in terms of dollars.

Figure 15: Structural Estimation: Net Value of Repayment (ω_t) and Inverse Repayment Burden ($\omega_t^{\bar{T}}$) versus Average Default Rates

(a) Net Value of Repayment (ω_t) (b) Inverse Repayment Burden ($\omega_t^{\bar{L}}$)

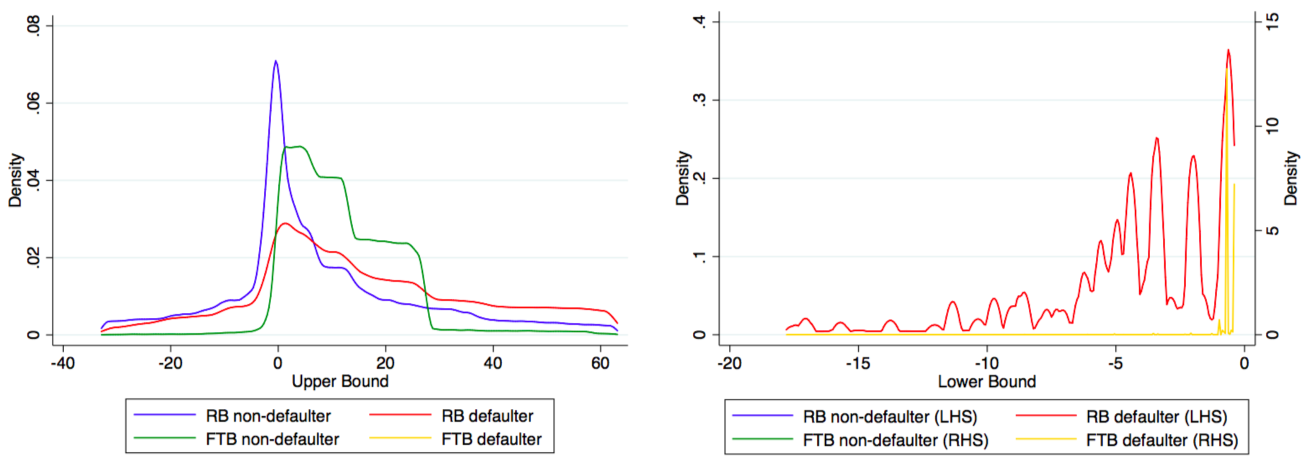


This figure displays how average default rates vary across the net value of repayment (ω_t) and the inverse repayment burden ($\omega_t^{\bar{T}}$). In the model, borrowers default once the net value of repayment falls below their net outside option if they don't reach their maximum required loan cycle, and if their net outside option is above the repayment burden once they do reach their last required loan. ω_t is a function of loan size L_t, the interest rate R_t, expected loan size growth λ_{t+1}^e, the discount factor δ, and the marginal product of capital draw MPK_i. It can therefore be calculated for each loan, conditional on assumptions about λ_{t+1}^e, δ, and MPK_i. I assume $\lambda_{t+1}^e = \frac{L_t}{L_{t-1}}$, $\delta = .9$, and $MPK_i = 15\%$. Likewise, $\omega_t^{\bar{T}}$ is only a function of loan size L_t and the interest rate R_t, and thus can also be calculated for each loan. Average default rates are calculated as the percentage of loans marked as defaulted, which occurs when it reaches 90 days past due. Units for ω_t and $\omega_t^{\bar{T}}$ are in terms of dollars.

Figure 16: Structural Estimation: Upper and Lower Bound Densities for Borrower Net Outside Options (ω_i)

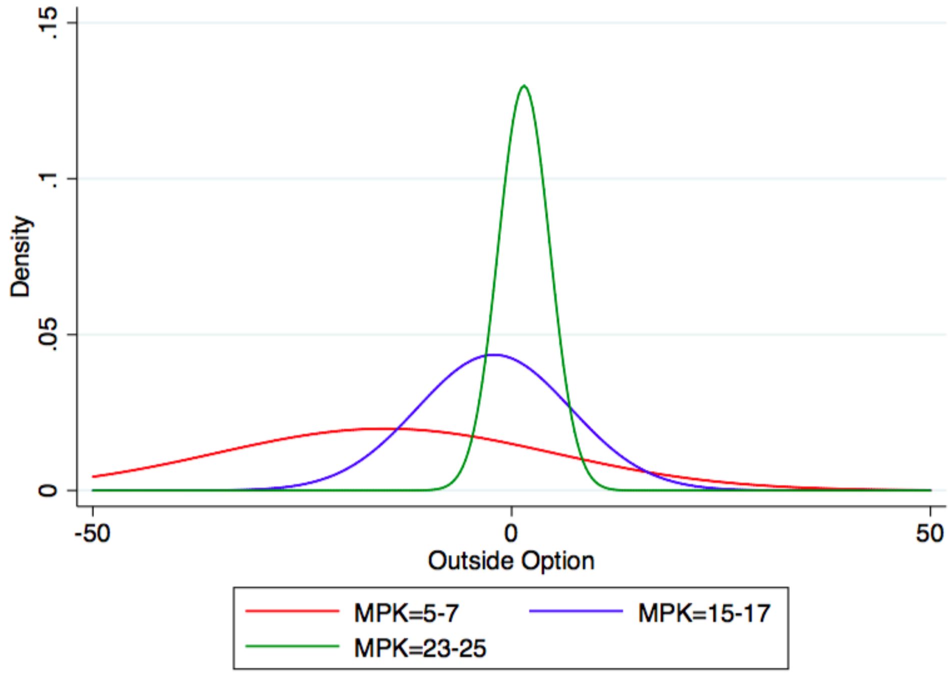
(a) Upper Bounds

(b) Lower Bounds



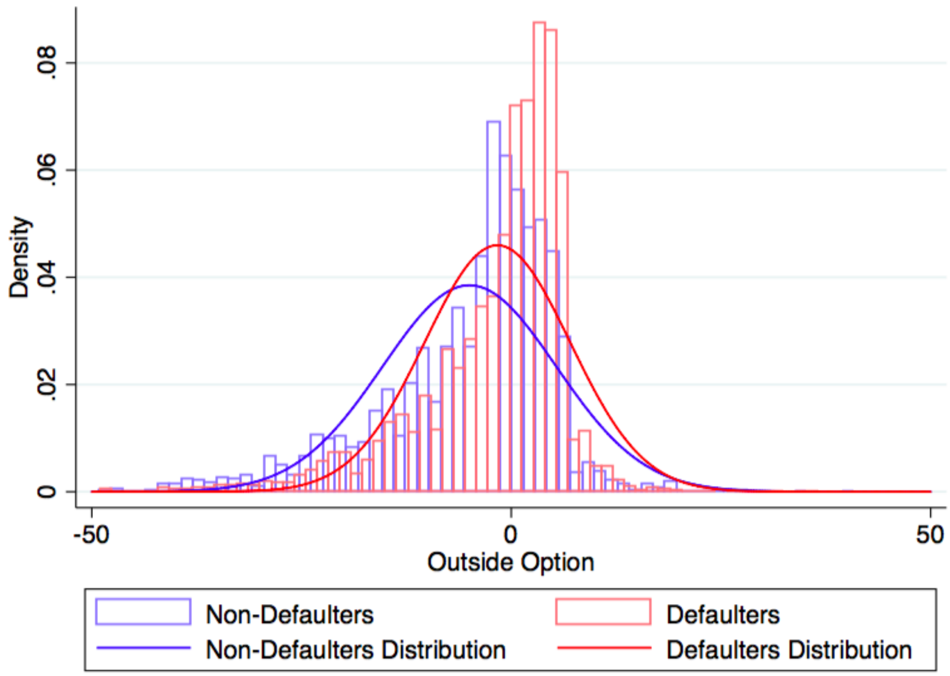
This figure shows the densities of the estimated upper and lower bounds on borrower net outside options, split by first-time borrowers (FTB) versus repeat borrowers (RB) and borrower-level default status. In the model presented in Section 6, borrowers differ in the extent to which they value access to future loans from the lender. This is captured by heterogeneity across borrowers in a net outside option parameter, ω_i, which they receive if positive or pay if negative in the event of default. The model implies a series of bounds on the value of a borrower's net outside option, depending on whether or not they have defaulted and how many loans they have taken out. These bounds can then be calculated as a function of loan size L_t and the interest rate R_t and conditional on assumptions about expected future loan growth λ^e_{t+1}, the discount factor δ, and a borrower's marginal product of capital draw MPK_i. I assume $\lambda^e_{t+1} = \frac{L_t}{L_{t-1}}$, $\delta = .9$, and I simulate $n = 100$ draws of monthly marginal returns to capital for each individual, where MPK_i is drawn from a uniform distribution between 5% and 25%. Units for the bounds are in terms of dollars.

Figure 17: Structural Estimation: Borrower Net Outside Option (ω_i) Distribution Estimates



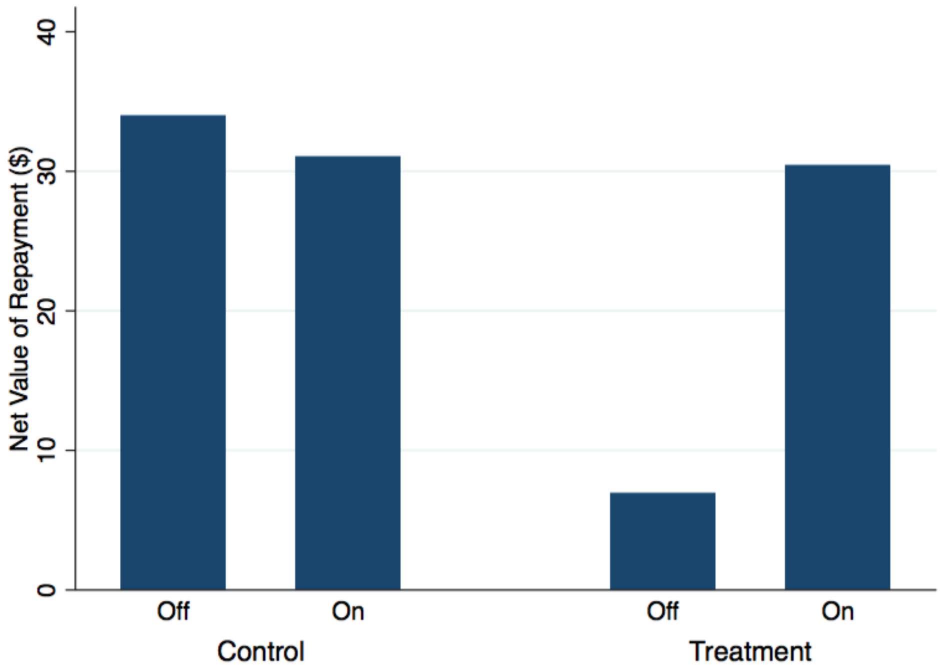
This figure plots the estimated distribution of borrowers' net outside option (ω_i) for a range of values of the marginal product of capital draw, MPK_i. In the model presented in Section 6, borrowers differ in the extent to which they value access to future loans from the lender. This is captured by heterogeneity across borrowers in a net outside option parameter, ω_i, which they receive if positive or pay if negative in the event of default. I use a censored regression model to estimate the mean and variance of net outside options within MPK_i buckets, assuming that, conditional on the MPK_i draw, the net outside option is normally distributed: $w_i|MPK_i \sim N(\mu, \sigma^2)$. Units for ω_i are in terms of dollars.

Figure 18: Structural Estimation: Net Outside Option Distribution Estimates by Default Status



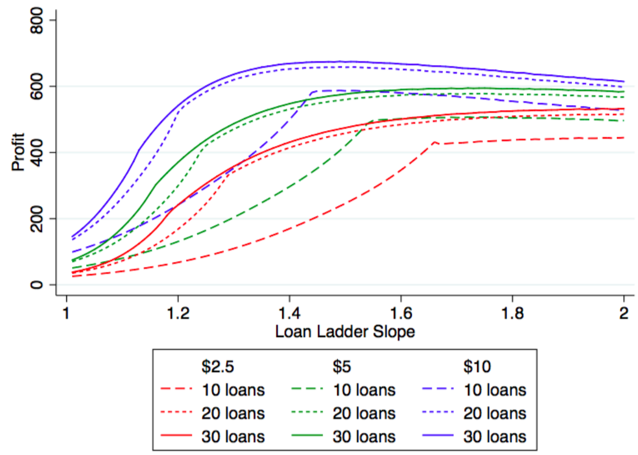
This figure plots the estimated values and distribution of borrowers' net outside option (ω_i), split by borrower-level default status. In the model presented in Section 6, borrowers differ in the extent to which they value access to future loans from the lender. This is captured by heterogeneity across borrowers in a net outside option parameter, ω_i, which they receive if positive or pay if negative in the event of default. I use a censored regression model to estimate the mean and variance of net outside options within MPK_i buckets, assuming that, conditional on the MPK_i draw, the net outside option is normally distributed: $w_i|MPK_i \sim N(\mu, \sigma^2)$. This plot shows average values across all possible MPK_i draws. Defaulters are defined as borrowers whose final loan during the sample period ended in default, meaning it was at least 90 days past due. Units for ω_i are in terms of dollars.

Figure 19: Structural Estimation Applications: Net Value of Repayment (ω_t) in Repeat Borrower Experiment



This figure shows the average net value of repayment (ω_t) for the treatment and control group in each period. ω_t is defined as the net outside option for a borrower who is just indifferent between defaulting in period t and period $t+1$, conditional on not reaching the maximum required loan cycle. A higher ω_t implies that fewer borrowers will choose to default. As derived in Section 6 and shown in Equation 17, ω_t is a function of loan size L_t, the interest rate R_t, expected loan size growth λ^e_{t+1}, the discount factor δ, and the marginal product of capital draw MPK_i. It can therefore be calculated for each loan, conditional on assumptions about λ^e_{t+1}, δ, and MPK_i. I assume $\lambda^e_{t+1} = \frac{L_t}{L_{t-1}}$, $\delta = .9$, and $MPK_i = 15\%$. Units for ω_i are in terms of dollars.

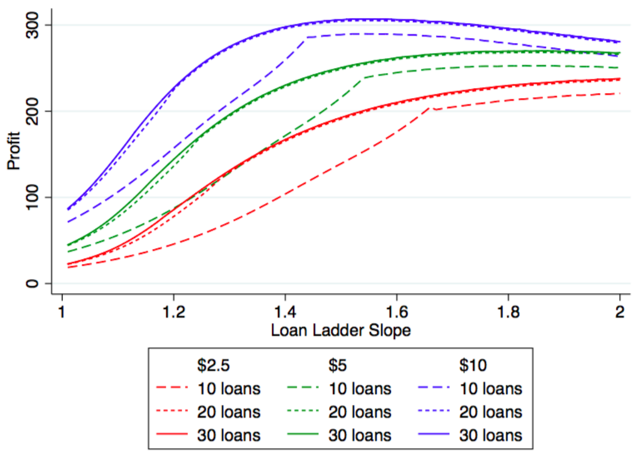
Figure 20: Loan Ladder Optimization Results: Profits across Possible Loan Ladder Slope Values



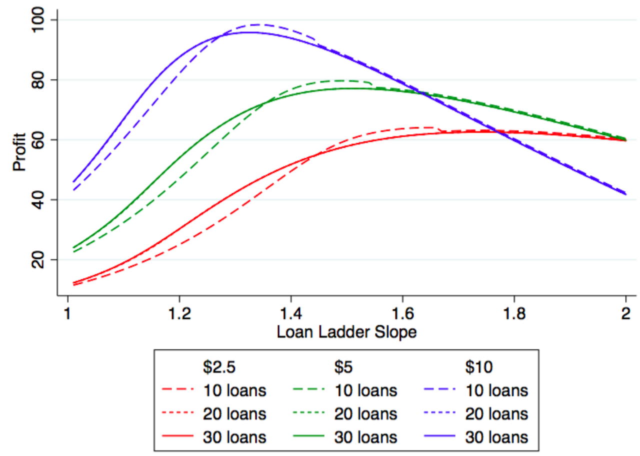
This figure shows simulated profits for the lender across different values of the loan ladder slope, λ. It depicts expected profits for the three initial loan sizes offered by the lender ($2.5, $5, and $10) and over three different loan cycle horizons (10, 20, or 30 loans). It uses the baseline assumption that the probability of a borrower hitting her maximum loan cycle in each period, p, is given by the value estimated in the data (13.6%). Other assumptions about parameter values are given in Section 8.4. The loan ladder slope is calculated as the proportional rate of loan growth. Units for profits are in terms of dollars.

Figure 21: Loan Ladder Optimization Results: Alternative Values for p

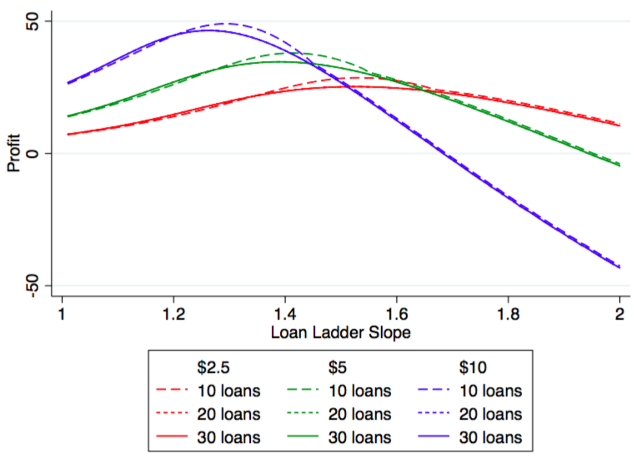
(a) $p = 20\%$



(b) $p = 30\%$

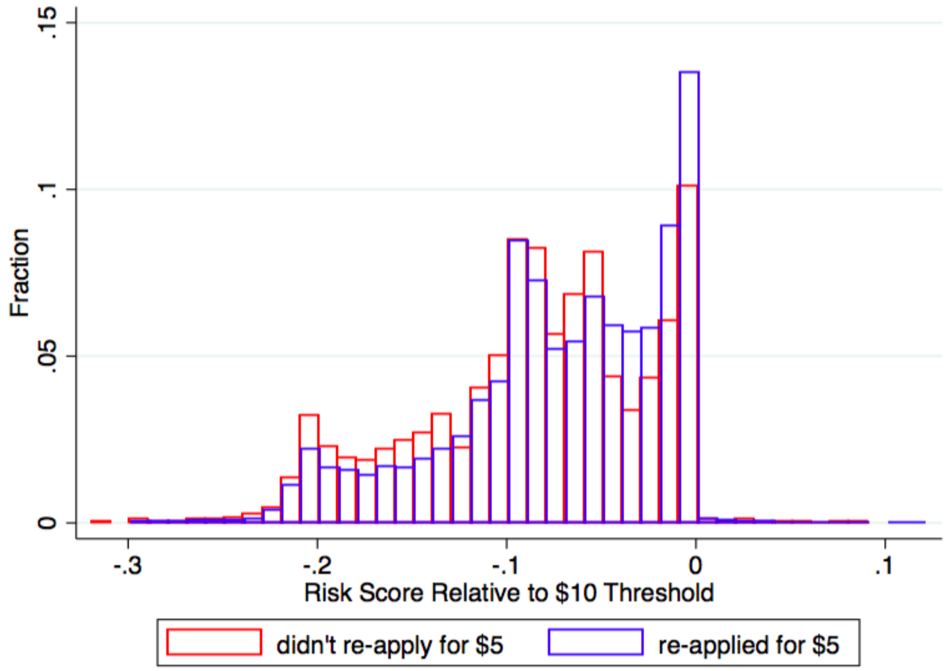


(c) $p = 40\%$



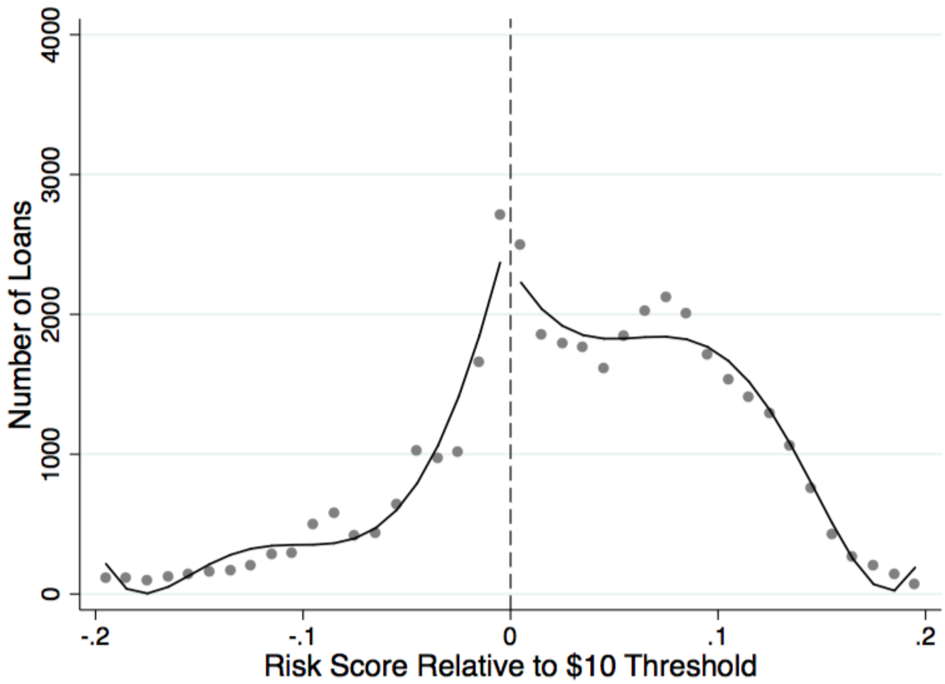
This figure shows simulated profits for the lender across different values of the loan ladder slope, λ. It depicts expected profits for the three initial loan sizes offered by the lender ($2.5, $5, and $10) and over three different loan cycle horizons (10, 20, or 30 loans). Each panel uses an alternative assumption about the probability of a borrower hitting her maximum loan cycle in each period, p. Other assumptions about parameter values are given in Section 8.4. The loan ladder slope is calculated as the proportional rate of loan growth. Units for profits are in terms of dollars.

Figure 22: Initial Loan Size: Specification Checks - Risk Score Distribution by Reapplication Status



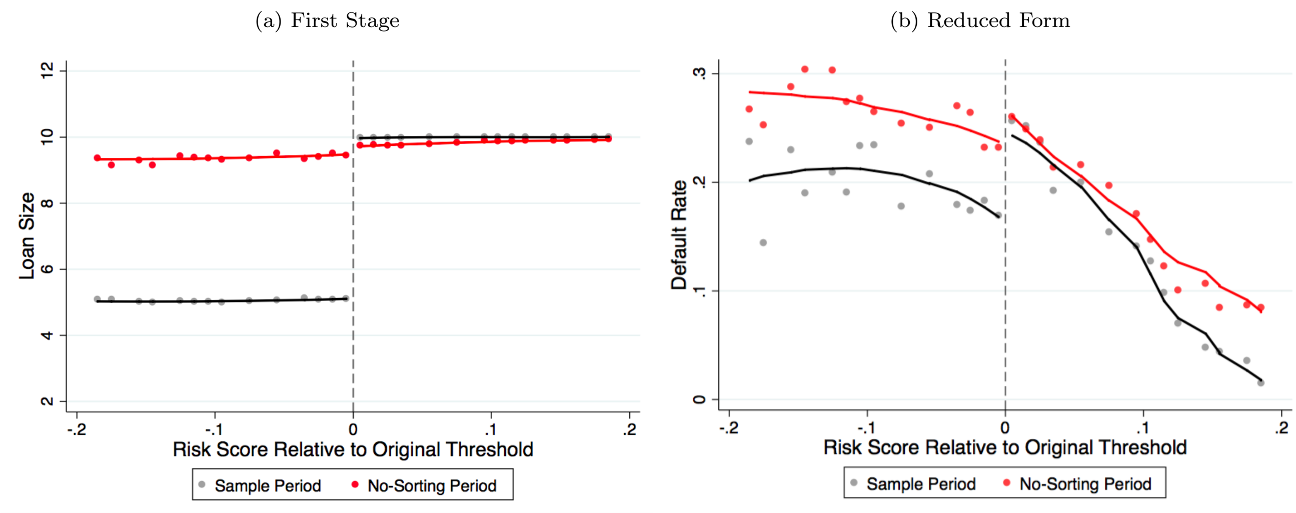
This figure shows the distribution of risk scores relative to the $10 loan threshold for individuals who were rejected for the larger $10 initial loan, split by whether or not they chose to re-apply for the smaller $5 loan. It provides visual evidence for the lack of selective borrowing among borrowers who were ineligible for the larger loan. The risk score is an internally-calculated measure of borrower credit-worthiness. Borrowers with a risk score above the $10 threshold received the larger $10 loan, whereas those below the threshold were instead directed to reapply for the smaller $5 loan.

Figure 23: Initial Loan Size: Specification Checks - Risk Score Manipulation Check



This figure shows the average number of first-time loans across risk score bins with a width of .01 units (about 2% of the total range of risk scores during this period), as well as fitted values from a regression with a third-order polynomial to control for the risk score, where the shape is allowed to vary on either side of the threshold. It provides visual evidence for the lack of manipulation of the risk score (the running variable in the regression discontinuity approach). The risk score is an internally-calculated measure of borrower credit-worthiness. Borrowers with a risk score above the $10 threshold received the larger $10 loan, whereas those below the threshold were instead directed to reapply for the smaller $5 loan.

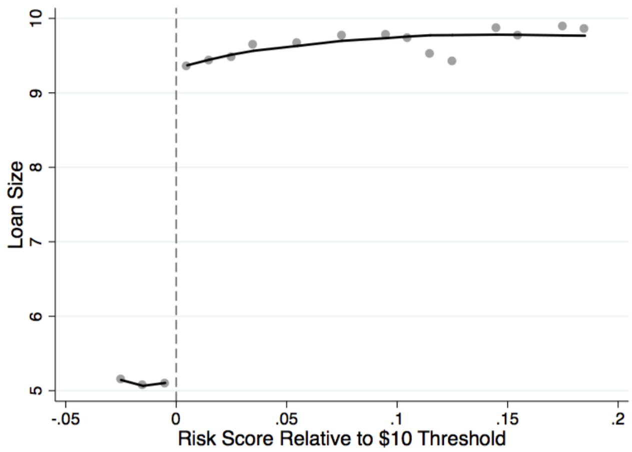
Figure 24: Initial Loan Size: Specification Checks - Comparison between Periods with and without Risk Score Sorting



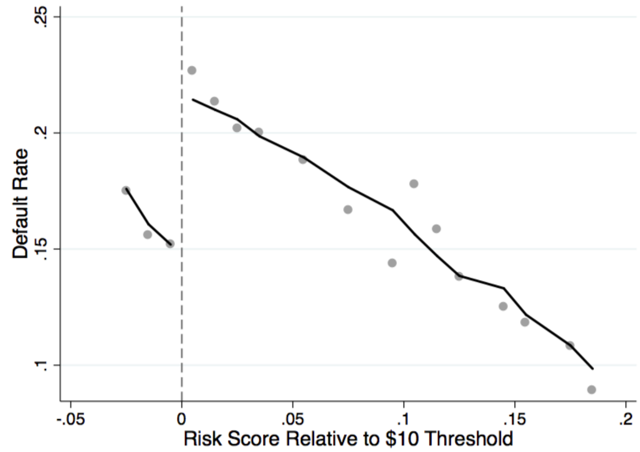
This figure replicates the regression discontinuity analysis in an alternative later period when borrowers were no longer sorted into initial loan sizes by their risk score. Panel A depicts the first-stage relationship between the risk score and initial loan size during my sample period and the later period. Panel B shows the reduced-form relationship between the risk score and default. Each panel plots average values of the outcome variable across risk score bins with a width of .01 units (about 2% of the total range of risk scores during this period), where the risk score is measured relative to the threshold. Each panel also includes fitted values from a regression with a third-order polynomial to control for the risk score, where the shape is allowed to vary on either side of the threshold. The risk score is an internally-calculated measure of borrower credit-worthiness. The average default rate is calculated as the percentage of loans that reach 90 days past due.

Figure 25: Initial Loan Size: Specification Checks - Alternative Time Period with Risk Score Sorting

(a) First Stage

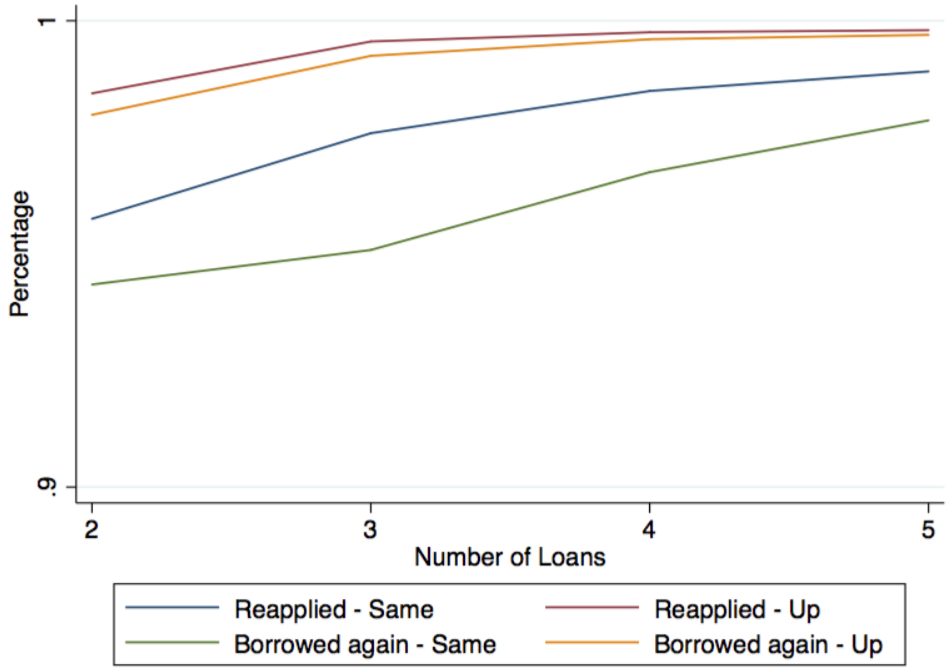


(b) Reduced Form



This figure replicates the regression discontinuity analysis in an alternative later period when borrowers were once again sorted into initial loan sizes by their risk score. Panel A depicts the first-stage relationship between the risk score and initial loan size. Panel B shows the reduced-form relationship between the risk score and default. Each panel plots average values of the outcome variable across risk score bins with a width of .01 units, where the risk score is measured relative to the threshold. Each panel also includes fitted values from a regression with a third-order polynomial to control for the risk score, where the shape is allowed to vary on either side of the threshold. The risk score is an internally-calculated measure of borrower credit-worthiness. The average default rate is calculated as the percentage of loans that reach 90 days past due.

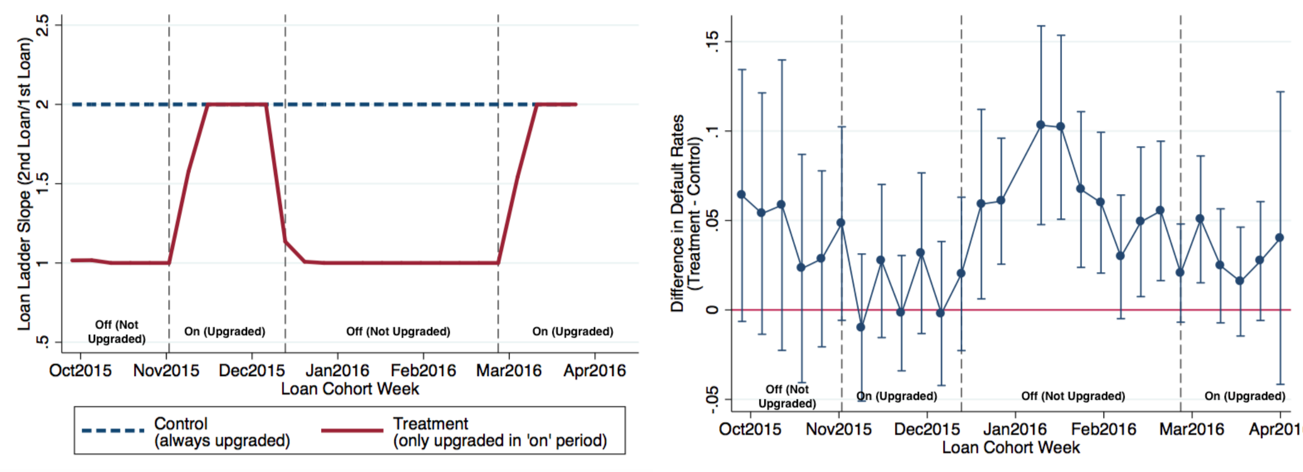
Figure 26: Loan Ladder Progression: Specification Checks - Evidence on (Lack of) Selective Borrowing



This figure shows the probability of reapplying and taking out another loan, conditional on being upgraded or remaining on the same level relative to the previous loan, by loan cycle. “Same” refers to borrowers who stayed on the same loan size relative to their prior loan; “up” refers to borrowers who were upgraded to a larger loan relative to their prior loan. Note that the scale starts at 90% and extends to 100%.

Figure 27: Loan Ladder Progression: Specification Checks - Extended Sample

(a) Loan Ladder Slope (b) Default Rates



This figure replicates the difference-in-differences analysis for an extended period of time. It includes an additional "off" period, in October 2015, during which borrowers in the treatment group were not upgraded to a larger loan. Panel A depicts the ratio of the size of a borrower's second loan to the size of her first loan, split by treatment and control groups in the loan ladder progression quasi-experiment. Panel B depicts the estimated difference in second-loan default rates between the treatment and control groups by loan cohort week, along with the corresponding 95 percent confidence intervals. A loan is marked as defaulted when it is 90 days past due.

B Tables

Table 1: Loan Characteristics

	Initial Mean	Median	SD	Min	Max	Repeat Mean	Median	SD	Min	Max
Loan Size	8.17	10.00	3.60	2.50	30.00	31.22	20.00	27.85	2.50	400.00
Fee Rate	0.08	0.08	0.02	0.05	0.13	0.11	0.11	0.02	0.05	0.17
Monthly Rate	0.13	0.12	0.05	0.07	0.28	0.12	0.12	0.03	0.02	0.28
APR	1.54	1.39	0.55	0.87	3.34	1.45	1.43	0.30	0.19	3.34
Repayment Freq.	7.00	7.00	0.00	7.00	7.00	7.74	7.00	3.07	7.00	28.00
Loan Term	19.47	21.00	2.89	14.00	28.00	27.85	28.00	8.25	14.00	168.00
Observations	85143					255867				

This table displays summary statistics on key contract terms, split by initial and repeat loans. Loan size is the amount disbursed, in dollars. The fee rate is calculated as the the flat fee charged on top of loan principal, relative to the loan principal. The monthly rate is a conversion of the fee rate, taking into account the loan term. The annual percent rate (APR) is the annualized version of the monthly rate. The repayment frequency refers to the fixed number of days between loan installment due dates. The loan term refers to the number of days between the disbursement date and the final loan installment due date.

Table 2: Borrowers (Survey Sample)

	Mean	Std. Dev.
Age	33.71	8.06
Gender (Female = 1)	0.26	0.44
Urban	0.45	0.50
Edu: At Least Primary	1.00	0.06
Edu: At Least Secondary	0.95	0.22
Edu: University	0.52	0.50
Currently In School	0.34	0.47
Monthly Income	421.27	401.58
Sufficient Savings to Last 6+ Mo. w/o Income	0.24	0.43
Salaried Job	0.50	0.50
Owns a Business	0.54	0.50
Number of Businesses (cond.)	1.60	0.97
Access to Other Credit Sources	0.96	0.21
Number of Other Credit Sources (cond.)	1.82	1.04
Other Credit Sources: Bank	0.23	0.42
Other Credit Sources: Other Digital Credit	0.75	0.43
Reported Loan Use: Business Purposes	0.46	0.50
Reported Loan Use: Smoothing Purposes	0.71	0.46
Observations	4903	

This table displays information on borrowers gathered from several third-party surveys, which were conducted by the lender in order to learn more about its customer base. Age is in years. Gender gives the percentage of the survey sample that is female. Urban is the percentage that report living in a city. The education-related variables give the percentage that report completing at least primary school, at least secondary school, and university. Currently in School refers to the percentage that report currently being enrolled in school either full- or part-time. Monthly income is in dollars. Sufficient Savings refers to the percentage of borrowers who report having enough savings to last at least six months in the event that they lost their primary source of income. Salaried Job gives the perentage that report currently having a salaried job, while Owns a Business is the percetange that report currently operating a self-owned business; note that these are not mutually exclusive. Number of Businesses gives the average number of businesses owned by borrowers who report owning at least one business. Access to Other Credit Sources is the percentage of borrowers who report having access to at least one other credit source. Number of Other Credit Sources is the number of other credit sources utilized by borrowers that have access to at least one other credit source. The Other Credit Sources variables refer to the percentage of borrowers with access to specific other sources. The Reported Loan Use variables give the percentage of borrowers who report using variables for the stated purpose; note that borrowers could select more than one, so the categories are not mutually exclusive.

Table 3: Borrowing Patterns (Borrower Level)

	Mean	Median	Std Dev
Num. Loans	6.79	5.00	6.40
One Loan Only	0.24	0.00	0.43
Ten + Loans	0.23	0.00	0.42
Months b/w First and Last Loan	8.11	5.47	7.43
Loans per Month	3.25	1.11	9.72
Last Loan: Defaulted	0.50	1.00	0.50
Last Loan: Outstanding	0.14	0.00	0.34
Last Loan: Repaid, Reapplied, Rejected	0.20	0.00	0.40
Last Loan: Repaid, Not Reapplied Yet	0.16	0.00	0.37
Observations	85170		

This table displays information at the borrower level on how borrowers utilize this lender. Number of Loans refers to the total number of loans taken out by the borrower. One Loan Only gives the percentage of borrowers who only take out one loan, whereas Ten + Loans refers to the percentage that have taken out at least ten loans. Months b/w First and Last Loan is the total number of months between the disbursement dates of the first and last loans. The Last Loan variables split the borrower pool by last loan status.

Table 4: Borrowing Patterns (Loan Level)

	Mean	Median	Std Dev
Loan Cycle	3.93	3.00	3.11
Initial Loans	0.25	0.00	0.43
Days bw Loans	4.25	0.00	12.17
Days bw Loan Disbursements	19.16	14.00	21.37
Defaulted	0.09	0.00	0.29
Repaid Early (cond. on repaid)	0.82	1.00	0.39
Num. Days Repaid Early (cond. on early)	14.74	14.00	9.93
OTP Percentage	0.63	0.75	0.41
No Repayments Made (cond. on default)	0.76	1.00	0.43
Observations	341011		

This table displays information at the loan level on how borrowers utilize this lender. Loan Cycle refers to the number of loans the borrower has taken out up to this loan (including the current loan). Initial Loans gives the percentage of loans that go to first-time borrowers. Days bw Loans gives the number of days between the last repayment made on a borrower's prior loan and the disbursement date of the current loan. Days bw Loan Disbursements gives the number of days between the disbursement date of a borrower's prior loan and that of the current loan. Defaulted is the percentage of loans that reach 90 days past due. Repaid Early is the percentage of repaid loans that are fully repaid prior to the final installment due date. Number of Days Repaid Early gives the total number of days between the final repayment and the final installment due date, conditional on a loan having been repaid early. The On-Time Payment (OTP) Percentage is calculated as the percentage of loan installments that were successfully repaid on time. No Repayments Made gives the percentage of defautled loans for which no repayments were ever made.

Table 5: Initial Loan Size: First Stage Results

VARIABLES	(1) Initial Size	(2) Initial Size	(3) Initial Size	(4) 10 Ind.	(5) 10 Ind.	(6) 10 Ind.
RS Threshold for 10	4.668***	4.733***	4.677***	0.934***	0.947***	0.935***
	(0.045)	(0.043)	(0.022)	(0.009)	(0.009)	(0.004)
Observations	39,405	39,405	39,405	39,405	39,405	39,405
R-squared	0.954	0.954		0.954	0.954	
Week FE	YES	YES	YES	YES	YES	YES
Global Polynomial	YES			YES		
Global Linear		YES			YES	
Local Linear			YES			YES

This table presents first-stage regressions of a measure of receiving the larger $10 loan on an indicator of being above the risk score threshold for qualifying for the $10 loan. Columns 1-3 use loan size as the dependent variable, while Columns 4-6 use a dummy for obtaining the larger loan. Columns 1 and 4 include a third-order polynomial as controls for the risk score, fully interacted with an eligibility indicator. Columns 2 and 5 replace the polynomial with global linear controls. Columns 3 and 6 use local linear estimation. All regressions also include loan cohort week fixed effects. Robust standard errors are shown in parantheses. $***p < 0.01, **p < 0.05, *p < 0.1$

Table 6: Test of Change in Risk Score Predictive Power

VARIABLES	(1) Below Threshold Default Rate	(2) Above Threshold Default Rate	(3) Combined Default Rate
Risk Score	-0.242***	-1.338***	-0.242***
	(0.078)	(0.069)	(0.078)
Risk Score x Elig for 10			-1.096***
			(0.106)
Elig for 10			0.088***
			(0.008)
Observations	11,947	26,400	38,347
R-squared	0.001	0.026	0.018
Test of Slope Change			
P-value			0
F-stat			107.3

This table presents results from regressions testing for a difference in the slope of the relationship between risk score and default rates above and below the risk score threshold for qualifying for the $10 loan. In Columns 1 and 2, an indicator of loan default is regressed on the risk score (measured relative to the threshold) for the sample of borrowers below and above the $10 risk score threshold, respectively. Column 3 uses the full set of loans and regresses an indicator of loan default on risk score, an indicator of being above the threshold, and their interaction. An F-test is presented to test for a significant difference in the slope on either side of the threshold. Robust standard errors are shown in parantheses. $***p < 0.01, **p < 0.05, *p < 0.1$

Table 7: Initial Loan Size: RD Results, Default Rates

VARIABLES	(1) Loan Default	(2) Loan Default	(3) Loan Default	(4) Borrower Default	(5) Borrower Default	(6) Borrower Default
Initial Loan Size	0.077***	0.082***	0.061***	0.011	0.014	0.009
	(0.013)	(0.011)	(0.013)	(0.011)	(0.010)	(0.015)
Observations	38,347	38,347	38,347	35,930	35,930	35,930
R-squared	0.021	0.020		0.027	0.026	
Week FE	YES	YES	YES	YES	YES	YES
Global Polynomial	YES			YES		
Global Linear		YES			YES	
Local Linear			YES			YES

This table presents two-stage least squares regressions measuring the impact of doubling the initial loan size on the probability of default by new borrowers. I instrument for initial loan size using the maximum eligible loan, which is \$10 and \$5 for borrowers above and below the risk score threshold for qualifying for the \$10 loan, respectively. I also include loan cohort week fixed effects. Columns 1-3 use loan-level default on the first loan as an outcome variable, while Columns 4-6 use borrower-level default instead. Columns 1 and 4 correspond to the global polynomial risk control specification, Column 2 and 5 utilize global linear controls, and Column 3 and 6 feature local linear estimation. A loan is marked as defaulted when it reaches 90 days past due. A borrower is marked as having defaulted eventually if she defaults on any loan; in practice, because the lender practices full exclusion of defaulters, this equates to a borrower who defaults on her last loan with the lender. Robust standard errors are shown in parantheses. $***p < 0.01, **p < 0.05, *p < 0.1$

Table 8: Initial Loan Size: IV Results, Other Longer Term Credit Outcomes

VARIABLES	(1) # of Loans	(2) # of Loans	(3) # of Loans	(4) Last Loan Size	(5) Last Loan Size	(6) Last Loan Size
Initial Loan Size	-1.390***	-0.871***	-0.913***	-2.530	-2.468	1.336
	(0.268)	(0.247)	(0.197)	(1.921)	(1.789)	(1.331)
Observations	38,347	38,347	38,347	35,930	35,930	35,930
R-squared	0.043	0.041		0.056	0.052	
Week FE	YES	YES	YES	YES	YES	YES
Global Polynomial	YES			YES		
Global Linear		YES			YES	
Local Linear			YES			YES

This table presents two-stage least squares regressions measuring the impact of doubling the initial loan size on other longer-term credit outcomes. I instrument for initial loan size using the maximum eligible loan, which is \$10 and \$5 for borrowers above and below the risk score threshold for qualifying for the \$10 loan, respectively. I also include loan cohort week fixed effects. Columns 1-3 use number of loans as an outcome variable, while Columns 4-6 use the last loan size instead. Columns 1 and 4 correspond to the global polynomial risk control specification, Column 2 and 5 utilize global linear controls, and Column 3 and 6 feature local linear estimation. The number of loans refers to the total number of loans taken out by a borrower from this lender. The last loan size is the amount disbursed on the final loan recorded for each borrower. Robust standard errors are shown in parantheses. $***p < 0.01, **p < 0.05, *p < 0.1$

Table 9: Loan Ladder Progression: DD Results

VARIABLES	(1) Default Rate	(2) Borrower Default Rate	(3) Number of Loans	(4) Last Loan Size
Treatment * On	-0.029***	0.086***	-0.307**	13.565***
	(0.009)	(0.019)	(0.142)	(1.508)
Treatment	0.053***	-0.086***	-0.904***	-31.271***
	(0.007)	(0.012)	(0.122)	(1.696)
On	0.010*	-0.053***	0.403**	-0.655
	(0.005)	(0.014)	(0.161)	(1.525)
Risk Score	-0.433***	-0.876***	22.451***	206.377***
	(0.040)	(0.084)	(1.611)	(18.439)
Number of Loans		-0.028***		
		(0.001)		
Weeks Between Loans	-0.001	-0.053***	0.246***	0.117
	(0.001)	(0.002)	(0.023)	(0.272)
Constant	0.015**	1.059***	9.404***	68.527***
	(0.007)	(0.116)	(1.447)	(10.351)
Observations	17,229	17,229	17,229	17,229
R-squared	0.018	0.190	0.055	0.080
Week of First Loan FE	YES	YES	YES	YES

This table presents difference-in-differences regression estimates of the impact of an exogenous doubling in loan size (relative to the previous loan) to a second-time borrower on various outcomes. It uses the standard DD specification. Treatment is an indicator of being in the treatment group, which is composed of borrowers with an on-time payment percentage on their prior loan of between $(P - 25)\%$ and $P\%$, such that whether or not they are upgraded on their second loan depends on when they apply. "On" is an indicator of a borrower applying for her second loan during the "on" periods, when the upgrading "shock" for treatment borrowers is in effect. Column 1 uses an indicator of default on the second loan as an outcome variable, Column 2 uses an indicator of whether a borrower defaults eventually, Column 3 uses the number of loans, and Column 4 uses the last loan size. All columns incorporate additional controls for borrower risk score, the speed at which the second loan is taken out, borrower cohort fixed effects, and the number of loans (where applicable). A loan is marked as defaulted when it reaches 90 days past due. A borrower is marked as having defaulted eventually if she defaults on any loan; in practice, because the lender practices full exclusion of defaulters, this equates to a borrower who defaults on her last loan with the lender. The number of loans refers to the total number of loans taken out by a borrower from this lender. The last loan size is the amount disbursed on the final loan recorded for each borrower. Robust standard errors are shown in parantheses. $***p < 0.01, **p < 0.05, *p < 0.1$

Table 10: Loan Ladder Progression: DD Results, Loan-Level Default Rate (Split by Loan Turnover Rate)

VARIABLES	(1) Below Median	(2) Above Median	(3) Interaction
Treatment * On	-0.048***	-0.004	0.002
	(0.013)	(0.015)	(0.014)
Below Median * Treatment * On			-0.054***
			(0.019)
Treatment	0.068***	0.035***	0.031***
	(0.009)	(0.009)	(0.008)
Below Median * Treatment			0.039***
			(0.012)
On	-0.002	0.014	0.007
	(0.009)	(0.009)	(0.007)
Below Median * On			0.001
			(0.006)
Risk Score	-0.418***	-0.455***	-0.431***
	(0.042)	(0.061)	(0.035)
Constant	0.005	0.005	0.008
	(0.014)	(0.009)	(0.007)
Observations	11,806	5,423	17,229
R-squared	0.021	0.022	0.019
Week of First Loan FE	YES	YES	YES

This table presents difference-in-differences regression estimates of the impact of an exogenous doubling in loan size (relative to the previous loan) to a second-time borrower on the probability of default, where the sample is split according to the loan turnover rate. The loan turnover rate is measured as the number of days between the final repayment on a borrower's first loan and the disbursement date of her second loan. I split borrowers into two groups based on whether they are below ("fast turnover") or above ("slow turnover") the median of one day in this sample. Columns 1 and 2 estimate the model separately for the fast and slow turnover groups, respectively, including controls for the risk score and the week in which a borrower took out her first loan. Column 3 includes a full set of interactions with a binary indicator for having taken out a second loan at least as fast as the median borrower, again including the additional controls. Robust standard errors are shown in parantheses. $***p < 0.01, **p < 0.05, *p < 0.1$

Table 11: Distribution of Net Outside Option (ω_i)

(a) Overall

10 Quantiles of MPK	Mean	SD
5-7	-15.12	20.12
7-9	-11.15	16.11
9-11	-9.25	15.23
11-13	-7.58	14.74
13-15	-4.91	12.55
15-17	-2.12	9.16
17-19	0.03	4.97
19-21	0.76	3.26
21-23	1.21	3.07
23-25	1.48	3.07
Total	-4.66	10.23

(b) By Default Status

10 Quantiles of MPK	Mean		SD	
	Defaulters	Non-Defaulters	Defaulters	Non-Defaulters
5-7	-11.88	-16.67	13.15	15.21
7-9	-7.18	-11.73	9.91	11.68
9-11	-4.32	-9.02	8.73	10.43
11-13	-2.25	-7.01	8.16	9.79
13-15	-0.78	-4.62	7.06	8.46
15-17	0.14	-2.56	5.61	6.67
17-19	0.90	-1.14	3.57	4.31
19-21	2.14	0.17	2.69	3.27
21-23	3.14	1.11	2.94	3.40
23-25	3.65	1.63	3.17	3.61
Total	-1.65	-4.98	8.68	10.36

(c) By Default Status (First-Time vs Repeat Borrowers)

10 Quantiles of MPK	Mean				SD			
	D (RB)	N-D (RB)	D (FTB)	N-D (FTB)	D (RB)	N-D (RB)	D (FTB)	N-D (FTB)
5-7	-18.65	-20.60	-0.25	-1.03	11.94	14.45	3.36	3.87
7-9	-12.33	-14.93	1.71	0.96	8.71	10.78	3.43	3.85
9-11	-8.81	-11.98	3.43	2.61	7.57	9.44	3.76	4.15
11-13	-6.35	-9.78	4.82	3.99	7.10	8.78	3.95	4.31
13-15	-4.40	-7.00	5.43	4.78	6.17	7.67	2.98	3.33
15-17	-2.74	-4.41	5.15	4.70	4.96	6.14	1.96	2.22
17-19	-0.90	-2.35	4.00	3.66	3.12	3.90	1.63	1.79
19-21	1.18	-0.65	3.77	3.43	2.69	3.03	1.74	1.86
21-23	2.66	0.49	3.97	3.58	3.30	3.39	1.92	2.02
23-25	3.36	1.09	4.14	3.74	3.63	3.71	2.07	2.13
Total	-4.70	-7.01	3.61	3.05	9.35	10.53	3.25	3.55

This table reports estimates of the parameters characterizing the distribution of the net outside option ω_i. I use a censored regression model to estimate the mean and variance of net outside options within MPK_i buckets, assuming that, conditional on the MPK_i draw, the net outside option is normally distributed: $w_i|MPK_i \sim N(\mu, \sigma^2)$. Panel A reports overall estimates from a model with no covariates. Panels B and C explore heterogeneity by default status and by both default status and first-time versus repeat borrowers, respectively. Defaulters are defined as borrowers who default eventually during the sample period, whereas non-defaulters have not yet defaulted. A borrower defaults when a loan reaches 90 days past due. First-time borrowers are borrowers who have only taken out one loan, while repeat borrowers have taken out more than one loan. Units are in terms of dollars.

Table 12: Lender Optimization Results

(a) Baseline ($p = 13.6\%$)

Loan Size	Optimal Lambda	Initial Profit	Max Profit (10)	Max Profit (20)	Max Profit (30)
2.5	2.00	-0.02	516.13	516.13	516.13
5	1.72	-0.03	578.11	578.11	578.11
10	1.49	-0.09	658.63	658.63	658.63
Total	1.74	-0.05	584.29	584.29	584.29

(b) $p = 20\%$

Loan Size	Optimal Lambda	Initial Profit	Max Profit (10)	Max Profit (20)	Max Profit (30)
2.5	2.00	-0.02	236.26	236.26	236.26
5	1.83	-0.04	268.40	268.40	268.40
10	1.54	-0.09	305.35	305.35	305.35
Total	1.79	-0.05	270.00	270.00	270.00

(c) $p = 30\%$

Loan Size	Optimal Lambda	Initial Profit	Max Profit (10)	Max Profit (20)	Max Profit (30)
2.5	1.65	-0.02	62.59	62.59	62.59
5	1.49	-0.04	77.14	77.14	77.14
10	1.34	-0.09	95.81	95.81	95.81
Total	1.49	-0.05	78.51	78.51	78.51

(d) $p = 40\%$

Loan Size	Optimal Lambda	Initial Profit	Max Profit (10)	Max Profit (20)	Max Profit (30)
2.5	1.53	-0.02	25.25	25.25	25.25
5	1.41	-0.04	34.63	34.63	34.63
10	1.29	-0.10	46.47	46.47	46.47
Total	1.41	-0.05	35.45	35.45	35.45

This table shows the lender profit simulation results. It gives the optimal loan ladder slope, λ, corresponding to each of the three initial loan sizes offered by the lender ($2.5, $5, and $10). It also reports the corresponding profit on the initial loan and maximized profit over three different potential loan cycle horizons (10, 20, or 30 loans). Panel A uses the baseline assumption that the probability of a borrower hitting her maximum loan cycle in each period, p, is given by the value estimated in the data (13.6%). Panels B-D use alternative values for p. Other assumptions about parameter values are given in Section 8.4. The loan ladder slope is defined as the rate of proportional loan growth. Units for profits are in terms of dollars.

Table 13: Initial Loan Size: Specification Checks - First Stage, No Sorting Period

VARIABLES	(1) Initial Size	(2) Initial Size	(3) Initial Size	(4) 10 Ind.	(5) 10 Ind.	(6) 10 Ind.
RS Threshold for 10	0.274***	0.379***	0.208***	0.044***	0.056***	0.012
	(0.034)	(0.029)	(0.044)	(0.006)	(0.005)	(0.009)
Observations	48,550	48,550	48,566	47,916	47,916	47,916
R-squared	0.028	0.028		0.021	0.021	
Week FE	YES	YES	YES	YES	YES	YES
Global Polynomial	YES			YES		
Global Linear		YES			YES	
Local Linear			YES			YES

This table presents first-stage regression estimates for the relationship between the risk score and initial loan size for an alternative later period when borrowers were no longer sorted into initial loan sizes by their risk score. It gives the impact of a measure of receiving the larger $10 loan on an indicator of being above the original risk score threshold, which was no longer being used to split borrowers into different initial loan sizes. Columns 1-3 use loan size as the dependent variable, while Columns 4-6 use a dummy for obtaining the larger loan. Columns 1 and 4 include a third-order polynomial as controls for the risk score, fully interacted with an eligibility indicator. Columns 2 and 5 replace the polynomial with global linear controls. Columns 3 and 6 use local linear estimation. Regressions also include loan cohort week fixed effects. Robust standard errors are shown in parantheses. $***p < 0.01, **p < 0.05, *p < 0.1$

Table 14: Initial Loan Size: Specification Checks - First Stage, Alternative Sorting Period

VARIABLES	(1) Initial Size	(2) Initial Size	(3) Initial Size	(4) 10 Ind.	(5) 10 Ind.	(6) 10 Ind.
RS Threshold for 10	0.041***	0.042***	0.028***	0.802***	0.832***	0.541***
	(0.003)	(0.001)	(0.001)	(0.046)	(0.022)	(0.022)
Observations	21,959	21,960	21,960	21,960	21,959	21,960
R-squared	0.570	0.568		0.621	0.620	
Week FE	YES	YES	YES	YES	YES	YES
Global Polynomial	YES			YES		
Global Linear		YES			YES	
Local Linear			YES			YES

This table presents first-stage regression estimates for the relationship between the risk score and initial loan size for an alternative later period when borrowers were once again sorted into initial loan sizes by their risk score. It gives the impact of a measure of receiving the larger $10 loan on an indicator of being above the risk score threshold for qualifying for the $10 loan. Columns 1-3 use loan size as the dependent variable, while Columns 4-6 use a dummy for obtaining the larger loan. Columns 1 and 4 include a third-order polynomial as controls for the risk score fully interacted with an eligibility indicator. Columns 2 and 5 replace the polynomial with global linear controls. Columns 3 and 6 use local linear estimation. All regressions also include loan cohort week fixed effects. Robust standard errors are shown in parantheses. $***p < 0.01, **p < 0.05, *p < 0.1$

Table 15: Initial Loan Size: Specification Checks - RD Results, Alternative Sorting Period

VARIABLES	(1) Defaulted	(2) Defaulted	(3) Defaulted
Initial Loan Size	0.087***	0.082***	0.071***
	(0.025)	(0.009)	(0.026)
Observations	21,960	21,960	21,960
R-squared	0.024	0.024	
Week FE	YES	YES	YES
Global Polynomial	YES		
Global Linear		YES	
Local Linear			YES

This table presents two-stage least squares regressions measuring the impact of doubling the initial loan size on the probability of default by new borrowers, for an alternative later period when borrowers were once again sorted into initial loan sizes by their risk score. I instrument for initial loan size using the maximum eligible loan, which is $10 and $5 for borrowers above and below the risk score threshold for qualifying for the $10 loan, respectively. I also include loan cohort week fixed effects. Column 1corresponds to the global polynomial risk control specification, Column 2 utilizes global linear controls, and Column 3 features local linear estimation. A loan is marked as defaulted when it reaches 90 days past due. Robust standard errors are shown in parantheses. $***p < 0.01, **p < 0.05, *p < 0.1$

Table 16: Loan Ladder Progression: Specification Checks - DD Results, Extended Sample

VARIABLES	(1) Default Rate	(2) Borrower Default Rate
Treatment * On	-0.029***	0.095***
	(0.009)	(0.017)
Treatment	0.053***	-0.061***
	(0.007)	(0.011)
On	0.010*	-0.065***
	(0.005)	(0.016)
Risk Score	-0.433***	-1.499***
	(0.040)	(0.083)
Weeks Between Loans	-0.001	-0.060***
	(0.001)	(0.002)
Constant	0.015**	0.798***
	(0.007)	(0.121)
Observations	17,229	17,229
R-squared	0.018	0.077
Week of First Loan FE	YES	YES

This table presents difference-in-differences regression estimates of the impact of an exogenous doubling in loan size (relative to the previous loan) to a second-time borrower on the probability of default, for an extended sample period (which includes an additional "off" time period). It uses the standard DD specification. Treatment is an indicator of being in the treatment group, which is composed of borrowers with an on-time payment percentage on their prior loan of between $(P-25)\%$ and $P\%$, such that whether or not they are upgraded on their second loan depends on when they apply. "On" is an indicator of a borrower applying for her second loan during the "on" periods, when the upgrading "shock" for treatment borrowers is in effect. Column 1 uses an indicator of default on the second loan as an outcome variable and Column 2 uses an indicator of whether a borrower defaults eventually. Both columns incorporate additional controls for borrower risk score, the speed at which the second loan is taken out, and borrower cohort fixed effects. A loan is marked as defaulted when it reaches 90 days past due. A borrower is marked as having defaulted eventually if she defaults on any loan; in practice, because the lender practices full exclusion of defaulters, this equates to a borrower who defaults on her last loan with the lender. Robust standard errors are shown in parantheses. $***p < 0.01, **p < 0.05, *p < 0.1$