

Advantageous Selection in Fintech Loans

Marco Pelosi*

November 2021

JOB MARKET PAPER

Abstract

Using data from the largest peer-to-peer (P2P) lender in the United States, I document advantageous selection in loan amount. By exploiting a natural experiment within the platform, I show that borrowers who select larger loans are less likely to default. This selection is driven by households who live in states with bankruptcy-friendly laws, where borrowers' default costs are lower. Standard models where borrowers maximize their utility cannot rationalize my results and make the opposite prediction. In a simple model of household borrowing, I show that my results can be explained by the fact that borrowers facing higher loan prices search more intensively for cheaper loans. This effect is stronger for the safest borrowers, as they enjoy the greatest benefits from switching.

1 Introduction

The aftermath of the 2007-08 financial crisis saw the widespread birth and fast rise of peer-to-peer (P2P) and marketplace lenders. As of 2018, fintech loans in the United States accounted for 38% of outstanding balance of unsecured personal loans¹. These platforms allow demand and supply of loans to meet directly, cutting off traditional banks from their role as intermediaries.

Whether and when online marketplaces will disrupt conventional lenders' business model has been an open question since the emergence of credit marketplaces in the early 2000s. Both anecdotal evidence and reports argue that these lenders are attractive as

* London School of Economics and Bank of Italy, m.pelosi@lse.ac.uk. I am grateful to Dirk Jenter and Daniel Paravisini for invaluable advice and support. I would like to thank Mike Burkart, Fabrizio Core, Vicente Cuñat, David De-Meza, Daniel Ferreira, Zhongchen Hu, Peter Kondor, Francesco Nicolai, Martin Oehmke, Alberto Pellicoli, Stefano Pietrosanti, Nunzia Saggiomo, Enrico Sette, Lorenzo Trimarchi, and seminar participants at LSE for comments and suggestions. The views expressed in this paper are mine and do not necessarily reflect the views of the Bank of Italy or of the Eurosystem. All errors are my own.

¹ <https://newsroom.transunion.com/fintechs-continue-to-drive-personal-loans-to-record-levels/>

they charge lower interest rates than traditional banks ([Adams, 2018](#)). Such discount is possibly due to the near absence of physical and monitoring costs.

Indeed, cheaper loans are probably one of the main drivers of the surging popularity of these platforms. For example, about 85% of loans issued through Lending Club (henceforth, LC), the largest online marketplace lender in the United States, are used to refinance other debt. Despite this popularity, we still know little about individuals who switch from traditional credit markets and borrow through P2P platforms. In particular, understanding the riskiness of people turning to online lenders is crucial to learn how these compete with conventional banks.

Adverse selection is often assumed to be pervasive in credit markets, as shown theoretically by [Jaffee and Russell \(1976\)](#) and [Stiglitz and Weiss \(1981\)](#) and documented by a large body of empirical work². Indeed, there is evidence that banks set cyclical credit limits to reduce excessive borrowing from high-risk individuals ([Jambulapati and Stavins, 2014](#)). This phenomenon is likely to be exacerbated in markets with a debtor-friendly bankruptcy code, where it is harder for lenders to seize borrowers' assets. As a result, lenders in such markets will have to charge higher interest rates to break even, and high interest rates discourage borrowing by the safest borrowers.

In this paper, I show that advantageous selection in loan amount may occur when a low-cost borrowing competitor enters a market characterized by adverse selection. More specifically, I document that safe borrowers take large loans in markets with better borrowers' assets protection by bankruptcy law. As this finding contrasts with the above theoretical predictions, I present a model where advantageous selection arises because of the existing credit equilibrium conditions. Consumers that pay higher interest rates on their existing loans put more effort into finding a cheaper lender. Yet, this is optimal only for safer borrowers, as those who are more likely to default do not want to bear the search and switching costs.

Measuring selection on loan size is empirically challenging. Using the universe of LC loans, I exploit the roll-out of larger loans by the platform as a quasi-experiment to disentangle such selection. Analyzing the repayment of borrowers who choose loans of different size might seem sufficient to measure selection on loan amount. Yet, [Karlan](#)

² For a literature review in the context of household finance see [Zinman \(2015\)](#)

and Zinman (2009) and Hertzberg et al. (2018) argue that to isolate empirically selection from the causal effect of contract terms, the lender (or the econometrician) must compare the loan performance of different groups of borrowers who choose an identical contract. Therefore, this especially requires comparing the behavior of borrowers choosing loans of the same size. LC loan menus consist of bundles of price, maturity, and amount. Thus, keeping the first two fixed, comparing the behavior of borrowers who sort into the same size bucket at different times allows me to isolate the effect of amount selection from that of other contract characteristics.

I apply this approach using the loan limit increase announced by LC in March 2016, from \$35,000 to \$40,000. I measure selection by comparing the default probability of borrowers choosing \$35,000 before and after the availability of \$40,000 loans (i.e., before and after the borrowers were selected on loan amount). Taking the difference between borrowers who select identical contracts allows me to rule out that different contractual terms (e.g., installments) cause differences in repayment behavior. If borrowers' composition was not time-varying, I would be able to isolate borrowers selection from this simple difference. Yet, it is hard to argue that this is the case, for example loans take-up could be seasonal.

Therefore, to account for changes over time in the composition of LC borrowers I estimate a difference-in-differences exploiting the increase in the loan limit from \$35,000 to \$40,000 in March 2016, using loan amounts between \$20,000 and just below \$35,000 as counterfactual. In a nutshell, this test compares the default probability of borrowers choosing \$35,000 before and after new and larger loans become available, relative to the change in default rate for loans just below \$35,000. I find that borrowers choosing \$35,000 after the limit extension are 1.9 percentage points more likely to default than smaller loans. This difference implies that borrowers choosing loans larger than \$35,000 after March 2016 are safer. It is worth mentioning that since my sample does not include such loans - as I do not have a counterfactual for larger loans - my estimate sets a lower bound for the selection effect.

To study selection heterogeneity with respect to the prevailing interest rate in traditional credit markets, I then focus on homestead exemption laws. These legislations represent a form of insurance for borrowers who file for bankruptcy under Chapter 7,

and protect the equity in their primary residence up to the level of exemption (Auclert et al., 2019; Indarte, 2019). Because these limits are highly heterogeneous across states (e.g., the protection is unlimited in Florida and none in Pennsylvania) and levels hardly change over time (Hynes et al., 2004), they provide a credible proxy for interest rates charged by banks. To this end, Severino and Brown (2017) show that since banks bear higher costs when a borrower defaults in states with greater protection, they also charge higher prices. However, this is not the case for LC, as I show that the platform does not price the state of residency.

Thus, I can analyze selection heterogeneity by running the same regression outlined above in samples with increasing levels of exemptions. I do not find any evidence of selection in states with exemption level below the median. Instead, the whole differential change in default probability comes from borrowers who live in states with higher than median exemptions, where traditional banks are likely to charge higher interest rates. In these states selected borrowers are 3 percentage points more likely to default than the control group. These results are robust to the sample considered. In another test I divide states into four groups, and I still observe advantageous selection in the two groups with higher protection.

This evidence is counter-intuitive. Standard models of households borrowing predict that when the default cost is low, as in states with high (or even unlimited) homestead exemptions, selection on loan amount is adverse (de Meza et al., 2019). Intuitively, if borrowers do not have to pay anything when they default (for example, because their homes are not seized), the riskiest type finds it optimal to borrow as much as possible. Yet, the above findings contrast this apparently simple prediction, as risky borrowers in LC tend to ask for smaller loans precisely where default costs are low.

To rationalize these results, in the last part of the paper I present a simple model where households search for cheaper loans, and borrow to repay external debt and minimize future cash flows. First, I show that borrowers switch to online lenders as long as their interest rate is lower than traditional banks' one, accounting for switching fees. Then, I prove that the heterogeneity across states arises because of search and switching costs, and credit conditions in conventional lending markets. In fact, paying higher interest rates on outstanding loans leads borrowers living in states with higher homestead

exemptions to exert more effort in searching for cheaper lenders. On top of that, those who have the greatest incentives to look for lower prices are the safest individuals, as they enjoy the greatest benefits from switching.

I present many tests that corroborate the predictions of the model. For example, I show that the existing debt of LC borrowers is significantly lower in states with higher than median exemption, all else equal. While I cannot argue that this difference is causal, it hints that the alternative interest rate these households face is higher. Most importantly, I show that households' private information lies in their default probability. LC data allows me to observe the time series of the FICO score for borrowers with active loans. I show that, *ceteris paribus*, borrowers choosing \$35,000 loans in high exemption states after March 2016 face a bigger drop in their credit score than loans in the control group, hinting that worse income shocks hit them.

To date, there is little evidence on selection in online credit marketplaces. In contrast with my results, [Hertzberg et al. \(2018\)](#) find adverse selection in longer-term contracts using a similar menu extension³. This finding may lead to the conclusion that these markets ultimately do not constitute creative destruction. Instead, they could be another threat to financial stability. The evidence I present sheds new light on the mechanism underlying the move from traditional credit markets of safe and reliable borrowers. My results are also relevant from an investor's perspective, as they shed light on a potential source of alpha in consumer loan portfolio investments.

The critical identification assumption to isolating selection from the menu extension is that any change in loan demand for borrowers with the same characteristics does not differentially influence households borrowing \$35,000 and other amounts at the time of introduction of new menus (for example, due to shocks to economic conditions or to credit supply of other lenders). If this was the case, any difference in default probability would be due to shifts in the composition of borrowers caused by reasons other than selection.

I back this assumption by showing that the switch from \$35,000 occurs exactly at the time of the introduction of new loans. Moreover, there is no evidence of changes in loan demand around March 2016 that could specifically affect borrowers at \$35,000 compared

³ They use the introduction of 60-month maturity for \$10,000-16,000 loans

to others. Finally, I include a rich set of borrowers' characteristics and contract fixed effects in all regressions to compare observationally equivalent loans. These controls include the 4-points FICO score range, debt-to-income ratio, state, loan amount bucket, and maturity. In most specifications, I also control for borrowers' default probability. To predict it, I estimate a random forest algorithm that uses all relevant information on borrowers and contracts.

Related Literature

This paper mainly contributes to two strands of literatures. First, it is related to the growing literature on online unsecured consumer credit. [Tang \(2019\)](#) also uses LC data to pose the question of whether P2P lenders are substitutes or complements to traditional banks. With the same data, [Hertzberg et al. \(2018\)](#) study how lenders can screen borrowers using maturity. [Di Maggio and Yao \(2020\)](#) use both online and credit bureau data to show that Fintech borrowers gain market share by lending to risky borrowers first, and then turns to safer borrowers. [Cespedes \(2019\)](#) analyzes interest rate sensitivity of LC borrowers and suboptimal behaviors. From investors' perspective, [Vallee and Zeng \(2019\)](#) argue that P2P platforms maximize loan volume by trying to limit information asymmetries between sophisticated and unsophisticated investors. Marketplace lenders have also attracted attention from policy makers, as the work by [Adams et al. \(2017\)](#) on US consumer awareness indicates. This work brings novel evidence on the intersection between traditional credit markets and online platforms. While many of the previous studies focus on selection on observable characteristics, I analyze borrowers' loan selection based on their unobserved default risk. Moreover, I am the first to observe that the existing equilibrium in traditional credit markets affects the pool of borrowers who switch to newer (and cheaper) forms of credit. Finally, I also show that online lenders are able to attract the best borrowers where banks' ex-ante losses given default are higher. To this extent, it is possible to conclude that new platforms constitute a complement to traditional banks.

Finally, this paper also contributes to a large literature on credit rationing and selection. In their pioneering study, [Jaffee and Russell \(1976\)](#) and [Stiglitz and Weiss \(1981\)](#) are the first to show how adverse selection leads to credit rationing. Since then, a large literature has documented the impact of asymmetric information on credit markets. Among

these, [De Meza and Webb \(1987\)](#) show that asymmetric information can instead lead to overinvestment. Moreover, their work on advantageous selection in insurance market ([De Meza and Webb, 2001](#)) is seminal, as they show that selection can reverse once risk aversion is introduced. [Einav and Finkelstein \(2011\)](#) provide a great simple framework to think intuitively about both adverse and advantageous selection in insurance market. While there a rich body of empirical work on adverse selection ([Ausubel, 1991](#); [Dobbie and Skiba, 2013](#); [Edelberg, 2003, 2004](#); [Dobbie and Skiba, 2013](#); [Stroebel, 2016](#)), very few papers detect advantageous selection in credit markets. With a structural model of loan demand [Einav et al. \(2012\)](#) find advantageous selection into large loans in the auto loans market. Yet, their findings are specific to secured loans. While this literature is rich, my results not only constitute the first evidence of advantageous selection in unsecured credit markets. They also show it is possible that the selection created by incentives to look for better credit conditions prevails that induced by moral hazard (for example, because of lower default costs).

The rest of the paper proceeds as follows. Section 2 describes the LC platform and data, and homestead exemptions. Section 3 first delineates the main empirical strategy to measure selection, then it shows main empirical results. Section 4 studies selection heterogeneity based on homestead exemptions. Section 5 presents the theoretical model and tests its predictions. Section 6 concludes.

2 Setting

2.1 Lending Club

Lending Club is the largest peer-to-peer lender in the United States. Since its registration with the Security and Exchange Commission in 2008, it has funded over 4 million loans for a total value of \$53.7 billion. As a comparison, the second biggest P2P lender (Prosper) helped just below 1 million people borrowing \$16 billion. Households in all States (except Iowa) can apply for a loan, while all US residents (except for Ohio's) can invest either in the primary or the secondary market.

To apply for a loan, a prospective borrower has to insert the wished amount and the loan purpose in LC's homepage, as shown in Figure 1. She then gives sufficient personal

information (including the Social Security Number) to allow LC to pull credit information from credit score agencies. LC applies a few eligibility criteria to publish the loan on the platform⁴, including a FICO score larger than 660, satisfactory debt-to-income ratio, a credit history of at least 36 months, and a limited number of inquiries in the most recent months.

Once a person is eligible for a loan, LC assigns her one of the 25 possible initial scores (from A1 to E5) according to a proprietary algorithm. After that, she observes a menu of possible loan amounts with two different maturities: 36 and 60 months. The amount-maturity bundle chosen by the borrower puts her in one of the 35 final rating scores (A1 to G5). To each of these final scores corresponds a unique interest rate. Except for the smallest loans, the final price is a non-decreasing function of amount and maturity. Finally, the loan is published on the platform, where both peer households or big investors can participate in the loan funding. LC claims that over 99% of loans are funded. Indeed, in my sample, all but two loans are fully funded.

Most of LC earnings come from two sources. First, an origination fee is paid by borrowers upfront. This means that once the loan is funded, applicants receive the desired amount net of the fee. According to the platform's website, such fee ranges between 1-5% of the loan amount, depending on the above final rating. Second, a service fee paid by investors who receive monthly payments from borrowers, of approximately 1%. On top of these two standard fees, investors pay between 30-40% of any amount collected when a borrower stops repaying and the loan becomes *charged off*. In this case, LC "makes a reasonable effort to recover the money owed to investors"⁵. In my sample, LC raises on average 11% of the residual debt.

2.2 The Increase in Maximum Loan

Before March 2016, it was possible to borrow any amount between \$1,000 and \$35,000. In March 2016, without any previous announcement, LC increased the maximum amount lent to \$40,000 and published a blog post on its website. Borrowers can ask for any amount in \$25 intervals. It is clear from Figure 2 that the share of \$35,000 loans issued

⁴ <http://www.snl.com/Cache/c33047201.html>

⁵ <https://www.lendingclub.com/investing/investor-education/interest-rates-and-fees>

goes down in response to the increase in the loan limit. While the share peaks at 16% of issued loans in February 2016, it drops to just above 11% by the end of the sample period.

2.3 Bankruptcy in the United States

Bankruptcy helps borrowers discharge or make a plan to repay their debts. Individuals in the United States can choose to file under three chapters of the Bankruptcy Code: Chapter 7, Chapter 13, and Chapter 11. The first is the most bankruptcy-friendly type, as it allows filers to keep some exempt property. Exemptions vary state by state and can include real estate, various sources of income (e.g., wages, pensions), and insurance. Households who file for bankruptcy under Chapter 13 must repay their lenders at least the amount they would reimburse under Chapter 7. To do so, they must propose a repayment plan that can last from three to five years. Compared to Chapter 7, this plan allows borrowers to stop foreclosure proceedings and to keep their properties. Since in bankruptcy borrowers tend to have a low level of non-exempt properties, 61 percent of filers opted for Chapter 7 in 2016. In 2016 alone, 458 thousand households filed for bankruptcy under Chapter 7, discharging a total of \$138 billion ⁶.

The Homestead Exemption

Of all debt relief tools, the homestead exemption is the most important, as it allows borrowers to keep some of their equity in their primary residence. Also, it induces the greatest cross-state variation in potential debt relief [Auclert et al. \(2019\)](#). Such exemption is non-existent in New Jersey and Pennsylvania and unlimited in eight states, including Texas and Florida. Table 1 lists homestead exemptions for all States. When a house owner files for bankruptcy under Chapter 7, her seizable assets can be determined according to the following rule:

$$Seizable\ Assets = \text{Max}\{Home\ Equity - Homestead\ Exemption, 0\}$$

[Goodman \(1993\)](#) and [Skeel \(2001\)](#) argue that states set the initial levels of exemptions in the nineteenth century for reasons that are likely to be unrelated to their current state

⁶ <https://www.uscourts.gov/statistics-reports/bapcpa-report-2016>

of the credit market and personal bankruptcy. To support this view, Figure 3 maps the different levels of exemptions across all States. There is neither a geographical nor an economic pattern in the variation of debt relief generosity. Also, [Auclert et al. \(2019\)](#) and [Hynes et al. \(2004\)](#) note that the state variation has been stable over time and that cross-state differences in exemption levels are persistent. States mainly update their exemption levels based on inflation.

2.4 Data

The two main datasets I use come from LC's website. The first contains information on newly issued loans. Crucially for this study, each borrower LC publishes the loan repayment status, terms - such as the loan amount, price, and maturity - and some personal information. Such information includes the location (at state and three-digit ZIP code levels), the employment status and tenure, the annual salary, homeownership, plus a rich set of variables from the credit pull. Not only can I observe the FICO score range, the debt-to-income ratio, but also the debt outstanding on various lines of credit (e.g. revolving loans, installment debt), the number of credit inquiries, and the number of delinquencies in the previous two years.

The second dataset consists of the whole history of payments for each loan. Because of the focus on charged-off loans, this dataset allows me to observe the month of the default, the outstanding principal at the time of default, and the post-charge off recovery.

I conduct the analysis using data downloaded in April 2019. I restrict the sample to loans issued between October 2015 and July 2016. This time interval contains the date at which LC increased the maximum loan amount (March 2016), but also limits the length of the sample period to address the concerns raised by [Bertrand et al. \(2004\)](#). Also, I only consider loans larger or equal to \$20,000 to make a meaningful comparison between loans of similar size. The final sample is made of 119,747 loans.

Figure 4 plots the distributions of loan amounts before and after the increase in the loan limit. There are two interesting patterns. First, borrowers tend to choose loans in round numbers, despite [Cesp  des \(2019\)](#) shows that is a sub-optimal behavior because of the excessive interest paid. Second, people bunch strongly at every multiple of \$5,000. This effect is particularly evident at \$35,000, where LC set the initial loan limit. While

such bunching persists after March 2016, it is clear that there is a significant pool of borrowers applying for larger loans.

Table 2 presents the summary statistics of loan and borrowers characteristics of the 63,330 loans issued before March 2016. Panel A shows the characteristics of loans. When borrowers have only loans up to \$35,000 available, the average loan amount is \$26,143. The average APR is 12.96%, with a corresponding installment of \$740. At the take-up stage, borrowers self-report the purpose of the loan. Table 3 shows that in my sample, 86% of borrowers use their funds to refinance another debt, either in a credit card or other lines of credit. Only a smaller portion takes a loan to make home improvements or to finance small businesses. Half of the sample borrows at 60 months, which is the longest maturity available. Finally, roughly 18% of loans in this period are in default - defined as being delinquent for more than 120 days.

In Panel B, I present statistics about borrowers' characteristics. LC only admits prime borrowers to its platform. Indeed, the average FICO score is 700, and the minimum score in the sample is 660. A share 19.45% of borrowers' monthly income is spent on repaying current debt. Their outstanding revolving balance is \$29,413, which represents 56.43% of the available line of credit. Almost three-quarters are either on a mortgage or own a house.

3 Measuring Advantageous Selection

3.1 Empirical Strategy

To document advantageous selection in loan amount, I exploit the increase in the maximum loan size that occurred in March 2016. Figure 5 helps visualizing the research design. Given the bunching at the previous limit of \$35,000, two groups of people borrowed such amount before the extension. A first group actually wanted to borrow \$35,000, and therefore was not rationed to start with. Together with these, some borrowers would have preferred to ask for a higher amount but were capped by the loan limit.

The empirical strategy compares the default rate of observationally equivalent households who borrow \$35,000 before and after the increase in the loan limit (that is, before and after they are selected on loan amount). After March 2016, borrowers who would

have been rationed without the limit extension can take up larger loans. Because of that, I expect that those who still decide to borrow \$35,000 would have asked for the same amount under the old cap. Then, the difference in the default probability around the extension captures the shift in the unobserved quality of borrowers who choose \$35,000 loans.

Given the very simple structure of Lending Club loans, the natural experiment is surprisingly clean. Three possible outcomes can arise. Suppose risky borrowers now choose larger loans and safe individuals keep asking for \$35,000. In that case, I will observe a decrease in the default rate at this threshold, therefore concluding that the selection is adverse. Instead, the default rate will increase if risky borrowers stay at the old limit and safe borrowers ask for larger loans, thus finding advantageous selection. Finally, the no selection case arises if the default probability does not change, as risky and safe borrowers pick large loans once they are available.

Crucially, nothing else moves when introducing new menus, neither the credit model LC uses nor the interest rate for any rating. Also, given that nearly every loan posted on the platform is funded I am able to rule out any supply effect, and conclude that the demand for loans drives all the variation in the quality of borrowers. I consider a loan as in default if LC categorizes it as in *Default* or *Charged Off*. According to LC's website, a loan is in *Default* if a payment has been past due for more than 120 days. It becomes *Charged Off* when investors should stop expecting additional payments.

However, differences in the default probability of loans of the same size could also be driven by time-varying unobserved factors. The composition of borrowers could change over time due to external factors. If that was the case, disentangling the effect of amount choice from other drivers would be infeasible. Therefore, I use a difference-in-differences strategy that compares the default probabilities of other loans on top of the above variation. The rationale for doing this is two-fold. First, this variation captures the change in the quality of loans of similar size. Second, it computes the difference in the default probability across borrowers who do not apply for the largest possible loan. Therefore, these consumers picked their menus without any constraint both before and after the extension.

Table 4 summarizes observable characteristics of borrowers asking for \$35,000 before

and after the limit increase. Panel A already hints that the ex-post default rate increases by two percentage points, suggesting that the selection may be advantageous. Borrowers characteristics in Panel B are similar in all respects, except for the revolving balance in other accounts. Thus, it is possible to conclude that some borrowers who asked for \$35,000 before March 2016 would have asked more to repay outside loans had the limit been higher.

The underlying identification assumption is that unobserved trends in loan demand (e.g. economic conditions, meaningful events in LC competitors) does not affect differentially borrowers who choose \$35,000 and other bundles in March 2016. Within this assumption, comparing the ex-post performance of borrowers in the treated and control amounts allows me to disentangle the effect of the selection on the amount that followed the increase in the loan limit.

A few factors are supporting this assumption. First, there is no reason to believe that any shift in loan demand should affect differently borrowers choosing \$35,000 and bundles just below the threshold in March 2016. Also, none of LC's competitors made meaningful moves that shifted the composition of borrowers from one platform to another (in line with LC, Prosper had the loan limit at \$35,000 in March 2016 and increased it to \$40,000 only in May 2018). LC did not change any of the eligibility criteria, prices charged to borrowers, or its credit scoring algorithm in March 2016. They made the news about newly available loans via a blog post at the time of the introduction. It is impossible to find hints that the platform was about to increase the limit through web searches, nor that users were demanding such change.

Most importantly, I run an event study to provide more convincing evidence that the shift in borrowers' composition occurred precisely when LC introduced new loans. Finally, using a rich set of time-varying fixed effects - including location, FICO score among others - allows me to account for unobserved trends common to borrowers facing the same economic conditions and with equal creditworthiness at the time of take-up.

3.2 Take-up of Larger Loans

First, Figure 2 shows a substantial decline in the take up of loans at the old loan limit. While before March 2016 the average frequency was between 15% and 16% of all loans

in my sample, once larger amounts become available, it settles around 12%. The same trend can be seen in Figure 4. Red bars represent the distribution of loans before the limit increase. Borrowers prefer round numbers, and in particular in multiples of \$5,000. Most importantly, the loan limit at \$35,000 leads to a strong bunching effect, as it includes people who would like to borrow more. This effect is more clear once the platform introduces larger loans. Blue bars depict the distribution of loans after March 2016. While the first observation remains true, many borrowers choose newly available loans instead of those at the old limit, thus making the bunching effect weaker.

Nevertheless, while the plots above suggest a weaker take-up of \$35,000 loans, take-up at other size bundles changes as well. Therefore, it is useful to verify that the increase in loan limit corresponds a drop at the old cap, all else equal. To do that, I have to check whether similar borrowers are less likely to take \$35,000 loans once \$40,000 loans are also offered in the menu. Therefore, I run the following regression:

$$D35_{ilt} = \alpha + \beta POST_t + \gamma Rate_l + \Theta_{il} + \varepsilon_{ilt} \quad (1)$$

where $D35_{ilt}$ is a dummy that equals one if the loan amount is \$35,000, $POST_t$ is a time-dummy that turns on after March 2016, $Rate_l$ is the interest rate on the loan. Finally, Θ_{il} is a rich set of fixed effects that includes maturity, state, 4-digit range FICO score, employment status, home ownership, income bin, number of delinquencies, debt-to-income ratio, and default probability fixed effects.

To construct the latter, I perform a random forest algorithm that produces a default probability for each loan. Starting from October 2015, I estimate the probability of default using as training set all loans issued in the previous 12 months (i.e., starting in October 2014). Thereafter, and until the end of the sample (July 2016), I also include all loans issued in the previous month. Finally, I categorize loans in 20 buckets based on their default probability. The advantage of using machine learning is dual. First of all, it allows me to control for non-linear determinants of the default probability and to select the main variables that influence it. Second, it is the best attempt to match what LC does to assign borrowers' risk grades.

Results of the above regression are shown in Table 5. The only difference across the

three columns is the sample used. In the regression in the first column, I include all loans between \$20,000 and \$35,000. Individuals taking loans after March 2016 are 3.69 percentage points less likely to ask for \$35,000 compared to equivalent borrowers in months before. To better understand the effect, since 15% of people in my sample borrow \$35,000, it represents a drop of roughly 25%. In the second and third columns I report the same regression results using different samples, \$25,000-\$35,000 and \$30,000-\$35,000, respectively. They show similar patterns, with a drop of loans at the old limit always larger than 17%. These effects survive the inclusion of a rich set of fixed effects, including the default probability.

These conclusions allow strengthening the interpretation of the above graphs. Some borrowers were rationed when the old limit was in place and would have borrowed more in the presence of larger loans available. Now it remains to establish whether safe or risky individuals prefer larger loans.

3.3 Evidence of Advantageous Selection

The most basic question of this paper is: does riskiness matter when choosing the amount of a loan? If it does, which borrowers choose large loans? To document such an effect, I run a difference-in-differences model to understand the unobserved quality of borrowers who ask for \$40,000. I define a loan as *Treated* if its amount is exactly \$35,000, then I estimate the following regression:

$$y_{lit} = \alpha + \beta DID_{lt} + \Theta_{lit} + \varepsilon_{lit} \quad (2)$$

where y_{lit} is the outcome variable, DID_{lt} is a dummy equal to one if loan l is *Treated* after March 2016. The main coefficient of interest is β , the difference-in-differences estimator. In the main specification, the outcome variable is a default dummy. Therefore, β measures the change in the default rate of treated loans relative to loans in the control group. I include a granular set of fixed effects to compare borrowers who can choose similar menus, have the same credit risk, and face similar business cycles. As in the above regression, Θ_{lit} includes includes amount bin, default probability bin, maturity, state, employment status, home ownership, 4-digits range FICO score, income bin, debt-

to-income ratio bin, and number of delinquencies (all except amount bin interacted with month) fixed effects. I remove the loan interest rate as that results from borrowers' choice, thus probably endogenous to their privately known default probability.

3.4 Empirical Results

I report the main results of regression (2) using default as outcome variable in Table 6. The first column reports a difference-in-differences estimator of 2.33 percentage points. That is, selected borrowers are 2.33 percentage points relatively more likely to default than those who pick smaller loans. The mechanism is described above and depicted in Figure 5. Once new loans are available, borrowers who take up loans larger than the old limit - but would have borrowed \$35,000 had new amounts not been offered - are less likely to default. This finding is evidence of advantageous selection. The point estimate is statistically significant at the 1% level, and it is robust across samples including different amount limits. Such effect survives the inclusion of a granular set of fixed effects. These absorb any variation in default common to all loans with the same maturity, of borrowers living in the same state, with the same employment status, etc. In other words, one should interpret this result keeping in mind that I am comparing loans with the same maturity of observably equivalent borrowers.

The second column exhibits the same regression results after I include twenty time-varying default probability bins in the set of controls. When added to the aforementioned set of fixed effects, they allow me to study any difference in default probability not captured by a random forest algorithm. While slightly smaller, the advantageous selection remains sizeable and strongly statistically significant. The DID estimator suggests that new loans at the old loan limit are 1.9 percentage points more likely to default when compared with other loans in the sample.

These estimates are economically sizeable. Taking into account that the mass of people who choose larger loans equals 25% of borrowers, the DID estimator implies that individuals selecting large loans are 7.6 ($1.9/25$) percentage points less likely to default than the average \$35,000 borrower in my sample, who has an ex-post default probability of 19%. Thus, if the difference in default probability comes only from borrowers choosing large loans, it follows that these are 40% less likely to default.

Lending Club data allows me to analyze these results more deeply. In fact, not only can I study the effect of the extension on the extensive margin (i.e., default probability), but also on the intensive margin. It could be the case that even though the default probability increases, eventually the principal repaid does not change. I test this hypothesis by using the share of initial loan eventually repaid as the outcome variable of regression (2). Results are displayed in Table 7. The first column shows a DID estimator of -0.0139, meaning that selected borrowers pay 1.39 percentage points less of their loan than other borrowers. This effect is again strongly statistically significant. When I add default probability bin fixed effects, the point estimate drops (in absolute value) to -0.0122, but it keeps its statistical significance.

The above estimates imply that the share repaid by borrowers choosing large loans is 4.9 (1.22/25) percentage points larger than those selecting \$35,000 loans. Since the average \$35,000 borrower in my sample pays back 78% of the initial loan, it means that larger loans pay 6.2% dollars more of their initial debt. These calculations suggest that LC made a win-win choice. Not only it introduced larger loans that led to more fees for the platforms, it also improved the pool of borrowers sorting into these loans, arguably improving its reputation and increasing ex-post returns for investors.

Supporting Identification Assumption

The above analysis can measure the selection on loan amount as long as other determinants of borrowers' creditworthiness do not affect differentially borrowers taking \$35,000 and those taking other amounts in March 2016. For example, it could be the case that economic conditions (loan demand, meaningful changes at LC alternatives) drive the choice of borrowers in the same way that the menu expansion does. If this was true, the DID estimator might be capturing effects not due to the selection of borrowers into new loans. In order to rule out this possibility, I run the regression in Equation (1) breaking *POST* into month dummies:

$$D35_{ilt} = \alpha + \sum_{t=-5}^5 \beta_t EXP_t + \gamma Rate_l + \Theta_{il} + \varepsilon_{ilt}$$

In the alternative story, conditional on a rich set of fixed effects, I should observe a

change in the probability of \$35,000 take up before March 2016, and the above selection would merely pick up a change in the composition of borrowers due to external phenomena. Instead, under the assumption that the effects I show in the previous section are due to the increase in the loan limit, I should observe a shift from the \$35,000 bundle precisely in March 2016. Each dot in Figure 6 represents the point estimate of the dummy of the correspondent month in the above regression, together with its 95% confidence interval. I drop the February 2016 dummy to make the comparison more meaningful.

Once I control for characteristics that lead borrowers to make their optimal choice, there is no pre-trend in the unobserved probability of picking \$35,000 loans. Instead, there is a sizeable drop just after new loans become available, in March 2016. This plot supports the assumption that the increase in the maximum loan limit induced some borrowers to apply for new loans, not other unobserved events.

Alternative Explanations

A possible explanation for the mechanism described above is that the selection of borrowers changes across the whole distribution of loan amounts. Since there is significant bunching at multiples of \$5,000, it could be the case that once new loans become available the composition of borrowers changes at those bundles. For example, borrowers might refrain from asking the maximum to avoid signaling that they need as much as possible, perhaps as they fear some stigma. Under this theory, once \$35,000 is not the loan limit anymore in March 2016, people who did not ask as much only because that was the maximum may now pick it. If such shift is particular only to a specific type of borrowers (i.e., either risky or safe), the change in composition I detect could be due to these borrowers moving to \$35,000.

In order to rule out this possibility, I run the regression in Equation (2) in the sample of loans strictly smaller than \$35,000, using \$30,000 as the treated group. If the above mechanism holds, the DID estimator should be significant, as the unobserved type of borrowers who choose that threshold after March 2016 changes relative to smaller loans. I present evidence for this test in Table 8. None of the two specifications shows a significant DID estimator for the \$30,000 loan bundle, meaning that the ex-post performance of loans strictly below \$35,000 does not differentially change. In turn, the selection effect

I highlight is specific to loans at the old loan limit and due to new borrowers choosing new loans.

Finally, it is worth stressing that the change in loan performance found above is due to equivalent borrowers who would have most likely chosen \$35,000, absent any change in the platform rules. If the better performance of new loans was only driven by the *extensive margin* - that is, borrowers who join the platform only once larger loans become available - then I would not observe any differential change in the default rate of already existing menus, as the average type who choose \$35,000 would not differ. Instead, comparing borrowers who choose always-available bundles before and after the extension allows me to isolate the effect of selection into new bundles of borrowers who would have been on the platform even without the limit increase.

4 Understanding the Source of Selection

It is well known that debt relief generosity affects credit supply ([Severino and Brown, 2017](#)) and bankruptcy behavior ([Dobbie and Song, 2015](#); [Indarte, 2019](#)), but their influence over the selection of borrowers has received limited attention from scholars. Such selection is driven by the interest rate charged by traditional lenders. In particular, [Severino and Brown \(2017\)](#) show that banks set higher interest rates where homestead exemptions are larger, as they bear most of the default cost.

My analysis aims at studying whether (and how) the pool of borrowers switching to LC is affected by the price they face in conventional credit markets. I subdivide states into groups of equal size based on their homestead exemption level to shed light on this issue. For example, when I form four groups, the least bankruptcy-friendly group includes states with exemptions lower than \$20,000. Instead, the thirteen most generous states have an exemption at least as large as \$250,000 (and eight do not have any limit). As there is significant heterogeneity across groups, estimating the regression in Equation (2) in these different samples allows me to see how debt relief generosity leads borrowers to make different choices. As already pointed out, homestead exemption levels were set in the nineteenth century (possibly not randomly) for reasons unlikely related to nowadays credit markets.

First, Table 9 reports some relevant information about borrowers in two groups of states defined as *Low Exemption* and *High Exemption*. The first comprises the 24 states with exemptions lower than the median \$75,000, while all the remaining 27 are in the second (the two groups are not equally sized as there are five states with median exemption). Before March 2016, individuals in both groups borrowed just above \$26,000 at a rate roughly equal to 13%. Consumers in *High Exemption* states prefer shorter loans, which translates into a higher monthly installment. Most importantly, the default rate is 18% in both states. The average borrower in both groups has a 699 FICO score and a debt-to-income ratio just below 20%. Residents of high exemption states earn \$5,000 larger incomes and are less indebted by \$3,000. Also, they are as likely to have been delinquent in the previous two years, and more likely to rent their home rather than being on a mortgage.

First of all, I check whether borrowers in the two groups of states differ in their probability of choosing newly available loans. To test this, I estimate the following difference-in-differences regression:

$$D35_{ilst} = \alpha + \beta DID_{st} + \gamma Rate_l + \Theta_{lit} + \varepsilon_{ilst} \quad (3)$$

where $D35_{ilst}$ is a dummy that equals one for \$35,000 loans, while the dummy DID_{st} turns on for loans in *High Exemption* states after March 2016. Remaining variables are the usual controls: $Rate_l$ is the price of the loan, and Θ_{lit} is the above-defined set of fixed-effect. The coefficient of interest is β , as it measures the differential take-up of \$35,000 loans in the two groups of states after March 2016.

Results are reported in Table 10. As in previous cases, the only difference between the two columns is the inclusion of the default probability bin fixed effects. The difference-in-differences estimator is not statistically significant in both columns. There is no differential take-up of larger loans after March 2016 between the two groups of states. If this was the case, any heterogeneous selection across states could have been related to different preferences for loan amount. Instead, any selection in the analysis that follows has to be linked to the unobserved qualities of borrowers.

As anticipated, the main test of this section consists in estimating the regression in

Equation (2) in different samples. First, I only consider two groups called *Low Exemption* and *High Exemption* states. Second, I further split every sample in two, forming four groups. Main results are reported in Table 11 and Table 12.

In the analysis reported in Table 11, I split the sample into two parts. States whose homestead exemption is below and above the median \$75,000, respectively. The first column summarizes results in the *Low Exemption* group. Interestingly, the point estimate is very close to zero and not statistically significant. This result means that in states more hostile to bankruptcy filers the relative quality of selected borrowers does not change. Thus, the average quality of borrowers choosing \$35,000 and \$40,000 after March 2016 is the same.

The second column reports the opposite result. Here, the relative increase in default probability amounts to 3 percentage points, suggesting that selected borrowers are of poorer quality, as safer individuals prefer larger loans. The coefficient is strongly statistically significant despite the richest set of fixed effects (which comprises default probability bins, interacted with months). Taking the two columns together, it is clear that the effect seen above in Table 6 is entirely driven by the selection in more generous states.

Using the same back of the envelope calculations I used in the previous paragraph, these estimates imply that, in states with high exemptions, borrowers choosing large loans are 12.5 (3/25) percentage points less likely to default, compared to \$35,000 loans (whose average default probability is 19%). That is, in these states switching borrowers who choose large loans are 65% less likely to default.

In order to deliver more convincing evidence, I split the above two samples further. Table 12 reports the results of the same regression, this time with four samples with increasing levels of debt relief generosity. Column one reports the results in states with the lowest level of exemptions, while the last sample considers the most generous states. Conclusions do not change. Borrowers' quality in states more hostile to bankruptcy filers does not show any differential change. Indeed, difference-in-differences estimators in the first two columns are both not statistically significant. Instead, estimators in the third and fourth columns are both strongly significant and sizeable, as they both show a 3 percentage points increase in bankruptcy probability.

The last test I run is possibly the more convincing. As I have established that there is

no selection in states whose exemption is below the median, I can add states to the sample one level of exemption at a time and observe any pattern in the DID point estimate. I plot the observed point estimates, together with their 95% confidence intervals, in Figure 7.

The first coefficient on the left is the same shown in the first column of Table 11. Indeed, it is very close to zero and not statistically significant. As I add states to the sample, one level at a time, the DID estimator starts increasing while still being not statistically significant. It shows any relevance when I add states with homestead exemption at least equal to \$165,500. And as I enrich the sample, it keeps being significant and with a higher magnitude. This trend is the second interesting fact. Advantageous selection is almost monotonic in the debt relief generosity. That is, the vast majority of safe borrowers who switch to LC live in bankruptcy-friendly states and are, therefore, likely to face higher banks' interest rates.

5 A Simple Model of Debt Refinancing

In my model, borrowers make an effort to search for lenders charging lower interest rates. Once they find it, they decide how much to borrow to minimize future cash flows. This assumption is motivated by the fact that almost 90% of LC borrowers report using new loans to refinance their existing debt. Anecdotal evidence suggests that the main advantage of online lenders lies in the cheaper interest rates they can charge borrowers. It follows that borrowers and personal finance experts alike perceive this as the main comparative advantage relative to conventional lenders (Lux and Chorzempa, 2017). It is worth noting that the average LC borrower has more than \$70,000 in outstanding debt, excluding mortgages. Therefore, even in the event complete refinancing is optimal, it would not be feasible.

Each household lives for one period. At the beginning of the period, she faces uncertainty over her disposable income (w). For simplicity, I assume this is either positive (w) or equal to zero. Borrowers' heterogeneity lies in their exogenous probability of default (p_i). She has an outstanding debt (k) she wishes to refinance. Such debt expires in the next period and carries an interest rate r_B . Each borrower chooses how much effort (t) to exert to find a new lender. Given the effort, she finds a cheaper lender with proba-

bility $\alpha(t)$, an increasing and concave function of the effort. Yet, searching costs $c(t)$, an increasing and convex function of the effort exerted.

If she finds a lender (for example, LC), she chooses how much to borrow (d). This debt has the same maturity as external debt, but she pays an interest rate r_{LC} . After a loan is funded, she pays an origination fee (δ) upfront. That is, this is deducted from the money received from the lender. Therefore, each borrower actually gets $d(1 - \delta)$. In the bad state of the world, she defaults on their LC and external debt. The default cost is fixed (Θ), representing stigma, the impossibility of future borrowing, etc. If she defaults, she only repays a share θ of the outstanding debt with the new lender⁷. Finally, households own a house of value H and, depending on the states a household resides in, she is subject to a homestead exemption (e).

Household i maximizes her expected utility in the payment period. Denoting with U^{LC} and U^B households' utilities if they switch to LC or stick with traditional bank, respectively, they are defined as follows:

$$\begin{aligned}\mathbb{E}[U^{LC}] &= [(1 - p_i)u(c_N^{LC}) + p_i(u(c_D^{LC}) - \Theta)] \\ \mathbb{E}[U^B] &= [(1 - p_i)u(c_N^B) + p_i(u(c_D^B) - \Theta)]\end{aligned}\tag{4}$$

where $u(c)$ is an increasing and concave function of consumption, and subscripts N and D denote consumption in the non-default and default states, respectively. Therefore, the final maximization problem looks as follows:

$$\max_{d,t} \{ \alpha(t) \mathbb{E}[U^{LC}] + (1 - \alpha(t)) \mathbb{E}[U^B] - c(t) \}\tag{5}$$

subject to the following budget constraints:

⁷ This is motivated by the fact that in case of default LC "makes a reasonable effort to recover the money owed to investors". While it is true that traditional lenders could do the same, the predictions of the model are not affected by the settlement borrowers negotiate with them.

$$\begin{aligned}
c_N^{LC} &= w - d(1 + r_{LC}) - (k - d(1 - \delta))(1 + r_B) + H \\
&= w + d[(1 - \delta)(1 + r_B) - (1 + r_{LC})] - k(1 + r_B) + H \\
c_D^{LC} &= \min\{e, H\} - \theta d \\
c_N^B &= w - k(1 + r_B) + H \\
c_D^B &= \min\{e, H\}
\end{aligned} \tag{6}$$

When they do not default, borrowers consume their income w after repaying all their debts to LC and traditional banks. Also, they consume their housing wealth. In default, they enjoy their housing wealth only up to the exemption limit.

The first order condition with respect to d is then:

$$(1 - p_i)u'(c_N^{LC})[(1 - \delta)(1 + r_B) - (1 + r_{LC})] = p_i u'(c_D^{LC})\theta \tag{7}$$

The interpretation of this first order condition is very intuitive. Conditional on finding LC and refinancing, borrowers have to equate the marginal benefit (the savings on interests paid on their outstanding loans) with the marginal cost (the marginal settlement in case of default). Moreover, the necessary condition for an internal solution is that the interest rate savings must be positive, accounting for entry fees. That is:

$$(1 - \delta)(1 + r_B) - (1 + r_{LC}) > 0$$

The first order condition with respect to t is:

$$\alpha'(t)(\mathbb{E}[U^{LC}] - \mathbb{E}[U^B]) = c'(t) \tag{8}$$

That is, the marginal gains from exerting effort (gain in expected utility) must equal marginal costs. Moreover, the gain in expected utility over the *status quo* is a necessary condition for exerting any effort.

It is this last FOC that allows me to interpret my empirical findings. As explained above, without searching for new lenders, borrowers pay r_B on their bank loans. In order to understand why the advantageous selection comes from states with higher exemption

levels, it is necessary to analyze the equilibrium in traditional credit markets.

From a bank's perspective, the expected loss in case of a borrower's default is higher when such borrower can keep a portion (or sometimes the entire) house of residence. Therefore, it is optimal to set a higher price $r_B(e)$. [Severino and Brown \(2017\)](#) note that risk-neutral banks' break-even condition for a loan of size k is:

$$(1 + r_f)k = \tilde{\mathbb{E}}[\min \{(1 + r_B(e))k, \max\{H - e, 0\}\}] \quad (9)$$

With a constant left-hand side, as e increases banks must raise $r_B(e)$ to be remunerated for risk-bearing. They also provide empirical evidence for this prediction. They show that interest rates on unsecured credit tend to be higher in states with higher debt relief generosity. They explain these findings with a high demand for unsecured debt. Therefore, it is likely that LC borrowers face different outside options. In particular, LC borrowers living in states with high exemptions are likely to face higher $r_B(e)$.

The assumption that LC's interest rate r_{LC} does not depend on the state of residency seems strong, but the data back it. In the next section, I show that, while I cannot access LC's current credit scoring algorithm, I can exclude that the variation in the interest rate charged is affected by the state of residence. Moreover, looking at old versions of LC's website through the Internet Archive's Wayback Machine, it is clear that when the algorithm was published, none of the criteria was about the location of loan applicants⁸.

From the above logic, all else equal, the pool of individuals who are willing to borrow from traditional banks will be different across states. It will be worse in states with higher exemption because of the higher price. However, conditional on borrowing (as every LC borrower has also external debt), it follows that the safest borrower in high exemption states pays a higher premium over her symmetric information interest rate relative to an identical borrower in states with low exemptions. Therefore, it is intuitive to conclude that safe borrowers in high states are more likely to look for better credit conditions elsewhere.

I can formalize this prediction by studying the sensitivity of the optimal effort exerted with respect to the interest paid on borrowers' current debt r_B . By applying the Implicit

⁸ <https://web.archive.org/web/20121031160219/http://www.lendingclub.com/public/how-we-set-interest-rates.action>

Function Theorem to the first order condition in Equation (8):

$$\frac{\partial t^*}{\partial r_B} = -\frac{\alpha'(t)(1-p_i) [u'(c_N^B)k - u'(c_N^{LC})(k - d(1-\delta))]}{\alpha''(t)(\mathbb{E}[U^{LC}] - \mathbb{E}[U^B]) - c''(t)} \quad (10)$$

If consumers' consumption is higher when they switch to LC, which is always true under the FOC in Equation (7), the numerator is positive. Moreover, from the FOC in Equation (8), it must be the case that

$$\mathbb{E}[U^{LC}] - \mathbb{E}[U^B] > 0$$

Therefore, for a concave $\alpha(t)$ and a convex $c(t)$, the above sensitivity is positive. Accordingly, all else equal, borrowers that face higher interest rates are more likely to exert more effort to find cheaper loans.

The positive selection arises from the fact that borrowers have to pay a search and a switching cost to save on their interest payments. Intuitively, for someone that is sure to default there is no reason to search for cheaper loans. Formally, taking the sensitivity of the optimal effort with respect to default probability yields the following:

$$\frac{\partial t^*}{\partial p_i} = -\frac{\alpha'(t) [u(c_N^B) - u(c_N^{LC})]}{\alpha''(t)(\mathbb{E}[U^{LC}] - \mathbb{E}[U^B]) - c''(t)} \quad (11)$$

As in the previous case, if consumption in the non-default state is higher when switching to LC, this derivative is negative. That is, keeping everything constant, a riskier borrower exerts a lower effort to find a cheaper lender. Thus, they are more likely to find themselves at the stage of choosing the loan amount, conditional on finding LC.

5.1 Testing Model Predictions

The data and the setting largely justify the assumptions I make in this model. First, I assume LC's interest rate r_{LC} does not depend on the state of residence. This pricing model is arguably suboptimal from LC's perspective. By charging higher prices, they could be able to make up for higher expected losses given default. In Figure 8 I plot the R-squared of eight regressions of the same dependent variable (interest rate), on a richer and richer set of controls. Once I add states to previous controls (in *Model 7*), there is no

gain in the explained variance of r_{LC} , thus hinting that borrowers do not pay different rates based on where they live.

It could also be that LC is simply pricing loans wrongly. In this case, borrowers in states with a high exemption are implicitly subsidized by those in more bankruptcy-hostile states. Yet, if this were true, I would also observe different post charge-off recovery, namely lower settlements in more generous states.

First, in Table 14 I rule out mispricing by LC. In the first column, I report the result of a regression of the post charge-off recovery on a *High Exemption Dummy* and a rich set of controls. Similar borrowers do not settle for a different payment in the two groups of states, leading to the conclusion that the typical LC's borrower should not expect to be shielded against the platform's attempt to be repaid depending on her residency. The same conclusion follows from the second column. Here I use the percentage of the outstanding debt recovered, and it does not seem advantageous to live in states with higher exemption.

While I cannot directly observe the outside options that LC borrowers face (i.e., the interest rate on their outstanding loans), I can test whether they have different debt stocks before applying for a loan on LC. By assuming an inverse relation between loan demand and price, observing differential indebtedness would imply that interest rates faced are also different. Absent any difference in the rate charged by banks and in the exemption level, identical borrowers should not have different loan demand. If one allows the exemption to vary, keeping rate fixed, it should be optimal to borrow more in more generous states. Then, observing lower debt stock in *High Exemption* states could be evidence of the fact that banks mitigate excess borrowing by charging higher prices.

I test this hypothesis by running a regression of several debt classifications - the total outstanding debt (excluding mortgages), the installment loans balance, and the revolving balance - on a *High Exemption Dummy*, controlling for a large set of observable characteristics. I only use loans issued before March 2016 to avoid overlap with new menus offered, but results are robust when using an extended sample.

Results are reported in Table 15. The dummy of interest is always negative and strongly significant, except the third column on revolving debt. In particular, the first column shows that households applying to LC living in generous states have \$1,947 lower

total debt stock than identical borrowers in other states. While I cannot perfectly disentangle the price effect from other confounding factors, such a differential take-up is evidence of a different interest rate faced.

Similarly, the second column shows results of the identical regression using the balance in installment loans (e.g., auto loans) as the dependent variable. As in the previous case, the coefficient on the exemption dummy is negative and strongly statistically significant. Borrowers in states with high exemption have \$1,565 lower debt to be paid in installments, suggesting that also in this case households may face differential banks' rates. Only in the last column, using revolving debt balance as dependent variable, such dummy is not significant. All in all, given the above discussion about exemption level, bank rates, and loan demand, this evidence is suggestive of different prices faced by borrowers depending on the exemption they are subject to.

In the model, the nature of private information is the probability of realization of the bad state of the world. More realistically, borrowers' private information stems in their future ability to repay their debt. One measure of such ability is the FICO score. LC distributes a *Payments* dataset that contains the time series of all payments made, together with the FICO score at the time of each payment date. I construct a new variable:

$$\Delta FICO = FICO_{issue} - FICO_{last}$$

to study how borrowers' credit score evolves over the life of the loan. Then, I run the regression in Equation (2) using the above variable as the outcome. If it is the case that borrowers have private information about their future income shocks, these are likely to appear in worse FICO scores. Results are in Table 16 and Table 17.

Table 16 shows the results using the entire sample. The only difference between the two columns is the inclusion of the default probability bin fixed effects, but the results are consistent across the two specifications. As the model predicts, selected borrowers have a FICO score decline at least 2.8 points higher than other borrowers. Both these point estimates are statistically significant at the 1% level. Interestingly, the average LC borrower sees her FICO score decreasing over the relationship with the platform. Yet, this evidence confirms that borrowers at the old limit deal with worse income shocks and thus

are more likely to default. All these results are both statistically and economically strong.

Table 17 shows the estimates of the identical regression, but this time dividing states above and below the median exemption. This evidence corroborates the fact that it is only in *High Exemption* states that LC members display private information about their future income shocks. The DID estimator in the first column, which only includes less generous states, does not show any relative difference in the FICO score of borrowers treated and in the control group. Instead, when I only look at individuals in states above the median exemption, the DID estimator shows a relative decline of 3.3 scores for \$35,000 loans, compared to other size bundles. Therefore, it is exactly in these states that private information leads borrowers to choose different loan size choices.

To provide further evidence on the private information underlying size choice, it is useful to look at the performance of \$40,000 loans. If it is the case that borrowers choosing the new maximum are less likely to be exposed to negative income and creditworthiness shocks, then in turns, their ex-post default probability should be lower than borrowers choosing \$35,000. It turns out that the default probability of \$40,000 loans issued after March 2016 is roughly 11%, against 20.5% at \$35,000. Yet, characteristics of borrowers choosing these two bundles could be different.

Thus, it is more convincing to test this hypothesis using the regression in Equation (2), this time including newly available loans in the sample. It is important to point out that I am not perfectly able to measure selection for new loans, as I do not have a proper counterfactual for these borrowers. Therefore, it is possible that the results are affected by individuals who applied for a loan only because \$40,000 was available, and would not have signed up before March 2016. Although this is not likely to have happened, as they had the alternative to ask for a smaller loan, I cannot test this assumption. With this potential caveat in mind, I report evidence for this prediction in Table 18 and Table 19.

Estimates in Table 18 use the entire sample, and should be interpreted in contrast with Table 6. The first column displays results without the inclusion of default probability bin fixed effects. The DID estimator shows a relative increase in default probability of selected borrowers by 2.79 percentage points compared to other loan sizes. It is statistically significant at the 1% level.

In order to understand how the default probability of borrowers choosing new loans

compares to the treated group, it is useful to compare this point estimate with its equivalent in Table 6. In both cases, the regression compares identical borrowers choosing different loan amounts before and after the maximum amount increase. Table 18 only adds loans larger than \$35,000 in the sample. Thus, the only reason why the last point estimate is 0.46 percentage points larger than its parallel in Table 6 is that the relative performance of selected loans is even worse when compared to larger ones. Identical reasoning applies to the second column. This estimate confirms that the above arguments are robust to the use of non-linear default probability bins. As in the previous case, the DID estimator is strongly significant and higher than its correspondent in Table 6, as the model predicts.

Table 19 displays the results of the identical regression in states with low and high exemptions. I do this to verify that safer borrowers choose large loans especially in states where the interest rate differential is higher. If this is the case, I should not observe any relative difference with the results in Table 11 in states with lower than the median exemption. It is only in more generous states that borrowers choose loan size based on their private information. Therefore, my model implies that large loans are less likely to default than smaller loans only in the second group of states.

Results confirm this thesis. The first column, which focuses on less generous states, does not show any differential performance of selected loans. This is the same result I show in the first column of Table 11. On the contrary, the DID estimator in the second column is significant at the 1% level and suggests a relative increase in default probability of selected borrowers of 3.5 percentage points. To complement this analysis, this point estimate is 0.5 percentage points higher than its equivalent in Table 11. This difference implies that once new loans are added to the sample the performance of borrowers choosing the old limit is even worse, indicating that large borrowers have a lower default probability.

Finally, my model does not say anything about borrowers applying for a loan for purposes other than refinancing. For these individuals, incentives are less obvious. On the one hand, default costs should prevent them from asking for high amounts. On the other hand, someone might prefer immediate consumption against the expected future bankruptcy costs. It follows that claims about the role of the interest rate differential in determining the selection of borrowers are at least mitigated when they care less about

it. I can test this implication with the regression in Equation (2) and by only including borrowers applying for other purposes. Estimates are reported in Table 20. As predicted, in none of the two columns the DID estimator is significant, suggesting that the above reasoning on the offsetting incentives might apply.

6 Conclusion

The growing literature on fintech lending still lacks evidence on how borrowers choose marketplace lenders based on their private information. In this paper, I document advantageous selection in loan amount in the context of the online unsecured credit market. In particular, I show that safer borrowers choose large loans in bankruptcy-friendly states. In the theoretical model, I argue that this is because the main gain from switching to online markets to refinance debt is represented by interest rate savings. Under asymmetric information, traditional lenders set higher prices where exemptions are more generous, which lead to adverse selection. Therefore, when a low-cost competitor enters the market, the safest borrowers have the highest incentives to refinance their loans. Instead, risky borrowers are not willing to bear the costs and stay with their current lenders.

This work is among the first to document empirically advantageous selection in unsecured loans. It is relevant as it shows how fintech lenders can beat conventional banks in attracting safe borrowers by offering simple contracts and sizeable price savings. It is important for policymakers to investigate which borrowers switch to less regulated online lenders, and why. In fact, as the market share of these players grows, so does the share of borrowers that are not subject to standard regulation. Understanding these incentives is crucial to limit (and possibly govern) their impact on financial stability.

References

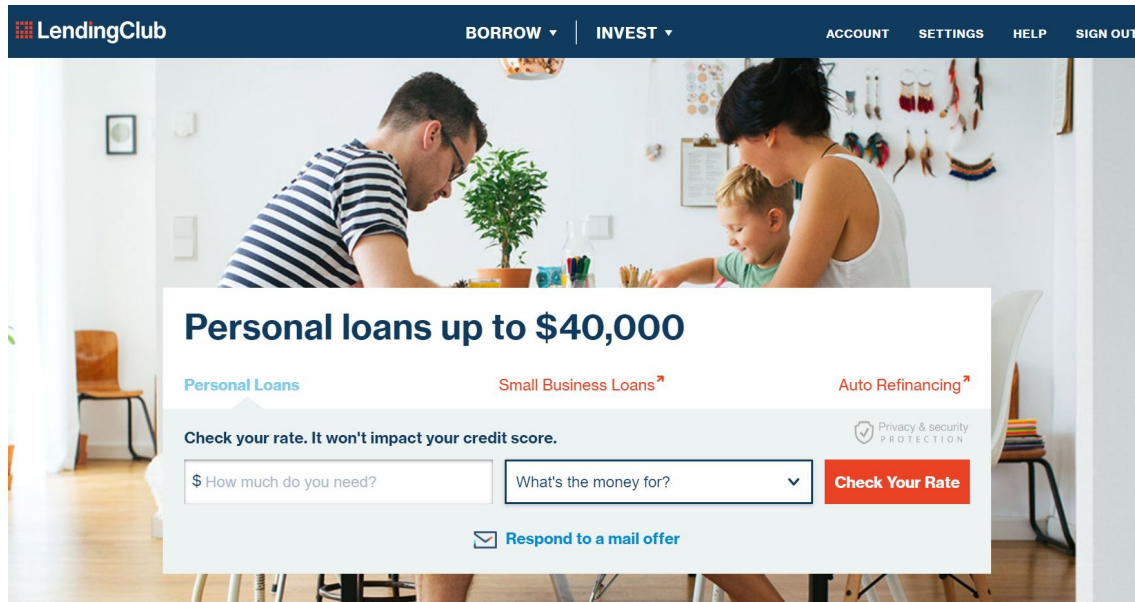
- R. Adams, T. Dore, C. Greene, T. Mach, and J. Premo. U.s. consumers' awareness and use of marketplace lending. *Working Paper*, 2017.
- R. M. Adams. Do marketplace lending platforms offer lower rates to consumers? *FED Notes*, 2018.
- A. Auclert, W. S. Dobbie, and P. Goldsmith-Pinkham. Macroeconomic effects of debt relief: Consumer bankruptcy protections in the great recession. *Working paper*, 2019.
- L. M. Ausubel. The failure of competition in the credit card market. *The American Economic Review*, pages 50–81, 1991.
- M. Bertrand, E. Duflo, and S. Mullainathan. How much should we trust differences-in-differences estimates? *The Quarterly journal of economics*, 119(1):249–275, 2004.
- J. Cespedes. Heterogeneous sensitivities to interest rate changes: Evidence from consumer loans. *Working paper*, 2019.
- D. De Meza and D. C. Webb. Too much investment: a problem of asymmetric information. *The Quarterly Journal of Economics*, 102(2):281–292, 1987.
- D. De Meza and D. C. Webb. Advantageous selection in insurance markets. *RAND Journal of Economics*, pages 249–262, 2001.
- D. de Meza, F. Reito, et al. Too little lending: A problem of symmetric information. Technical report, Working Paper, 2019.
- M. Di Maggio and V. Yao. Fintech Borrowers: Lax Screening or Cream-Skimming? *The Review of Financial Studies*, 2020.
- W. Dobbie and P. M. Skiba. Information asymmetries in consumer credit markets: Evidence from payday lending. *American Economic Journal: Applied Economics*, 5(4):256–82, 2013.
- W. Dobbie and J. Song. Debt relief and debtor outcomes: Measuring the effects of consumer bankruptcy protection. *American Economic Review*, 105(3):1272–1311, 2015.

- W. Edelberg. Risk-based pricing of interest rates in household loan markets. *Available at SSRN 484522*, 2003.
- W. Edelberg. Testing for adverse selection and moral hazard in consumer loan markets. *Available at SSRN 515903*, 2004.
- L. Einav and A. Finkelstein. Selection in insurance markets: Theory and empirics in pictures. *Journal of Economic Perspectives*, 25(1):115–38, 2011.
- L. Einav, M. Jenkins, and J. Levin. Contract pricing in consumer credit markets. *Econometrica*, 80(4):1387–1432, 2012.
- P. Goodman. The emergence of homestead exemption in the united states: Accommodation and resistance to the market revolution, 1840-1880. *The Journal of American History*, 80(2):470–498, 1993.
- A. Hertzberg, A. Liberman, and D. Paravisini. Screening on loan terms: Evidence from maturity choice in consumer credit. *The Review of Financial Studies*, 31(9):3532–3567, 2018.
- R. M. Hynes, A. Malani, and E. A. Posner. The political economy of property exemption laws. *The Journal of Law and Economics*, 47(1):19–43, 2004.
- S. Indarte. The impact of debt relief generosity and liquid wealth on household bankruptcy. *Working paper*, 2019.
- D. M. Jaffee and T. Russell. Imperfect information, uncertainty, and credit rationing. *The Quarterly Journal of Economics*, 90(4):651–666, 1976.
- V. Jambulapati and J. Stavins. Credit card act of 2009: What did banks do? *Journal of Banking & Finance*, 46:21–30, 2014.
- D. Karlan and J. Zinman. Observing unobservables: Identifying information asymmetries with a consumer credit field experiment. *Econometrica*, 77(6):1993–2008, 2009.
- M. Lux and M. Chorzempa. When markets quake: Online banks and their past, present and future. *Working Paper*, 2017.

- F. Severino and M. Brown. Personal bankruptcy protection and household debt. *Working Paper*, 2017.
- D. A. Skeel. *Debt's Dominion*. Princeton University Press, 2001.
- J. E. Stiglitz and A. Weiss. Credit rationing in markets with imperfect information. *The American Economic Review*, 71(3):393–410, 1981.
- J. Stroebe. Asymmetric information about collateral values. *The Journal of Finance*, 71(3): 1071–1112, 2016.
- H. Tang. Peer-to-peer lenders versus banks: Substitutes or complements? *The Review of Financial Studies*, 32(5):1900–1938, 2019.
- B. Vallee and Y. Zeng. Marketplace lending: a new banking paradigm? *The Review of Financial Studies*, 32(5):1939–1982, 2019.
- J. Zinman. Household debt: Facts, puzzles, theories, and policies. *Annual Review of Economics*, 7(1):251–276, 2015.

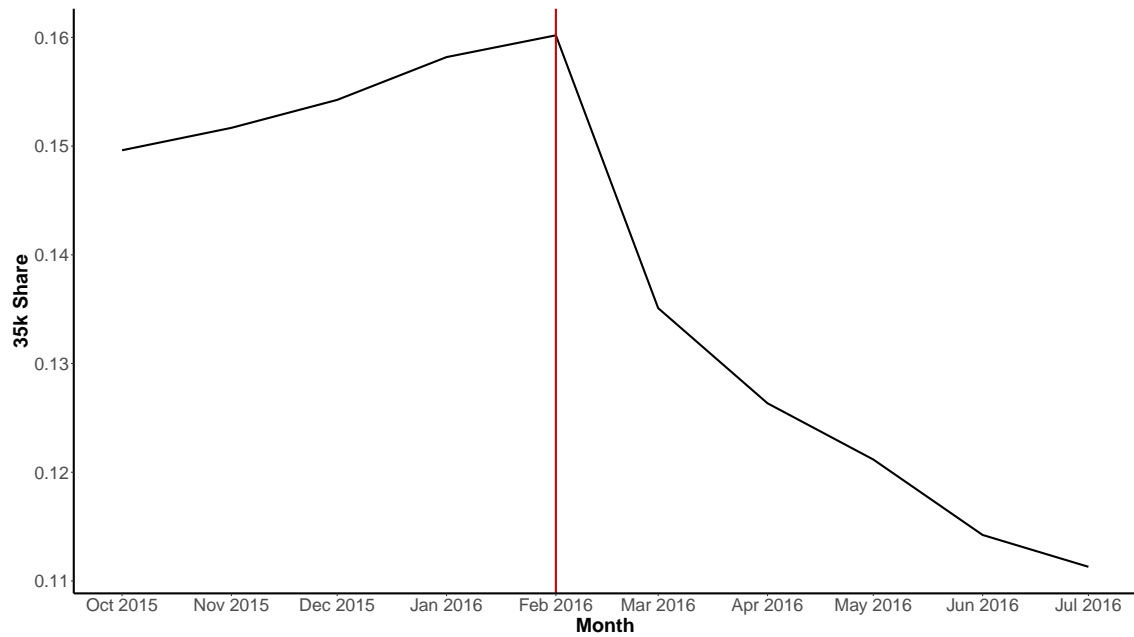
Figures

Figure 1: Lending Club Homepage



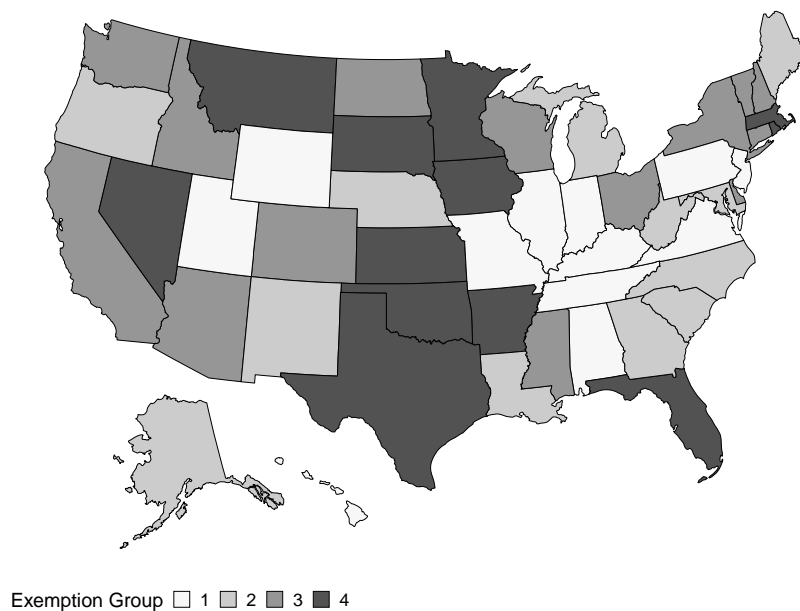
This figure is a screenshot of the Lending Club's website. On the homepage the platform only asks for the amount needed and the purpose. Following steps require more information, such as complete name, address, and Social Security Number.

Figure 2: Share of 35k loans



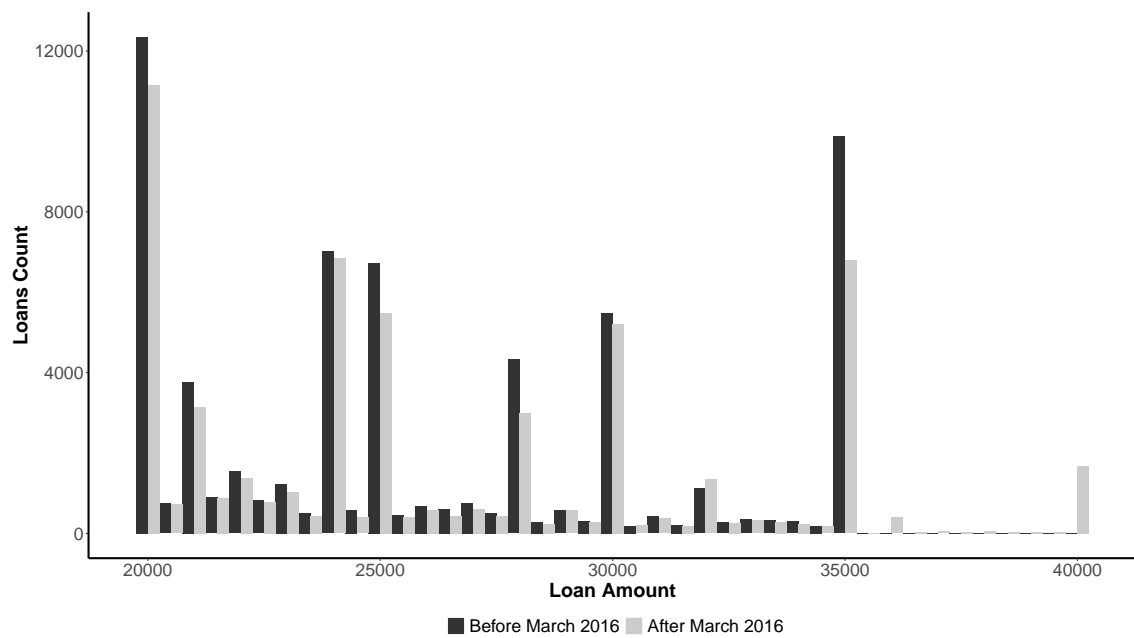
This figure presents the time-series of the share of \$35,000 loans issued by Lending Club between October 2015 and July 2016.

Figure 3: Homestead Exemption Map



This figure represents the different level of Homestead Exemption in the United States. States in the first group are those with lower exemption levels (such as New Jersey and Pennsylvania), while states in the last group offer the highest level of debt relief (e.g. Texas and Florida).

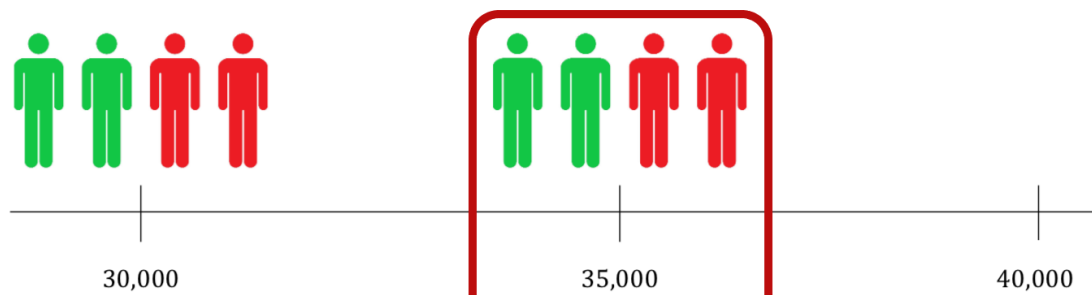
Figure 4: Loan amount distributions



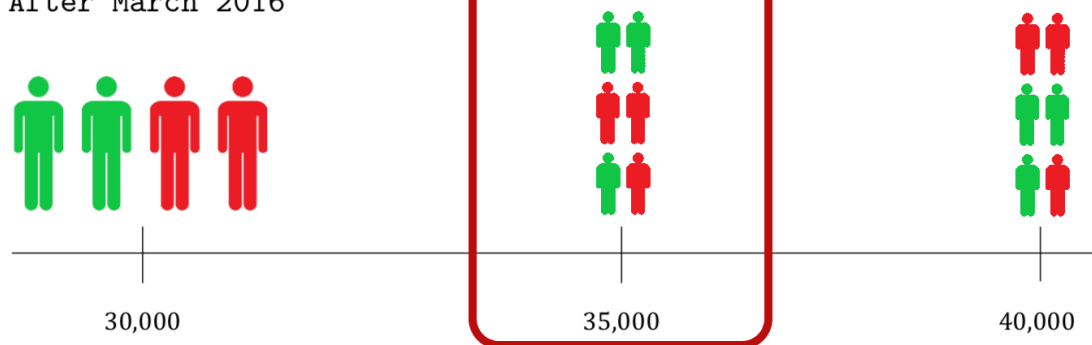
This histogram represents the number of loans issued by Lending Club in \$500 bands. The sample includes loans issued between October 2015 and July 2016.

Figure 5: Research Design

Before March 2016



After March 2016



This figure depicts the empirical design. In my difference-in-differences specification I compare the change in the default probability of \$35,000 loans before and after March 2016 (the squared group) with the same change for smaller loans (to the left). Without loss of generality, I fill safe borrowers in green and risky borrowers in red.

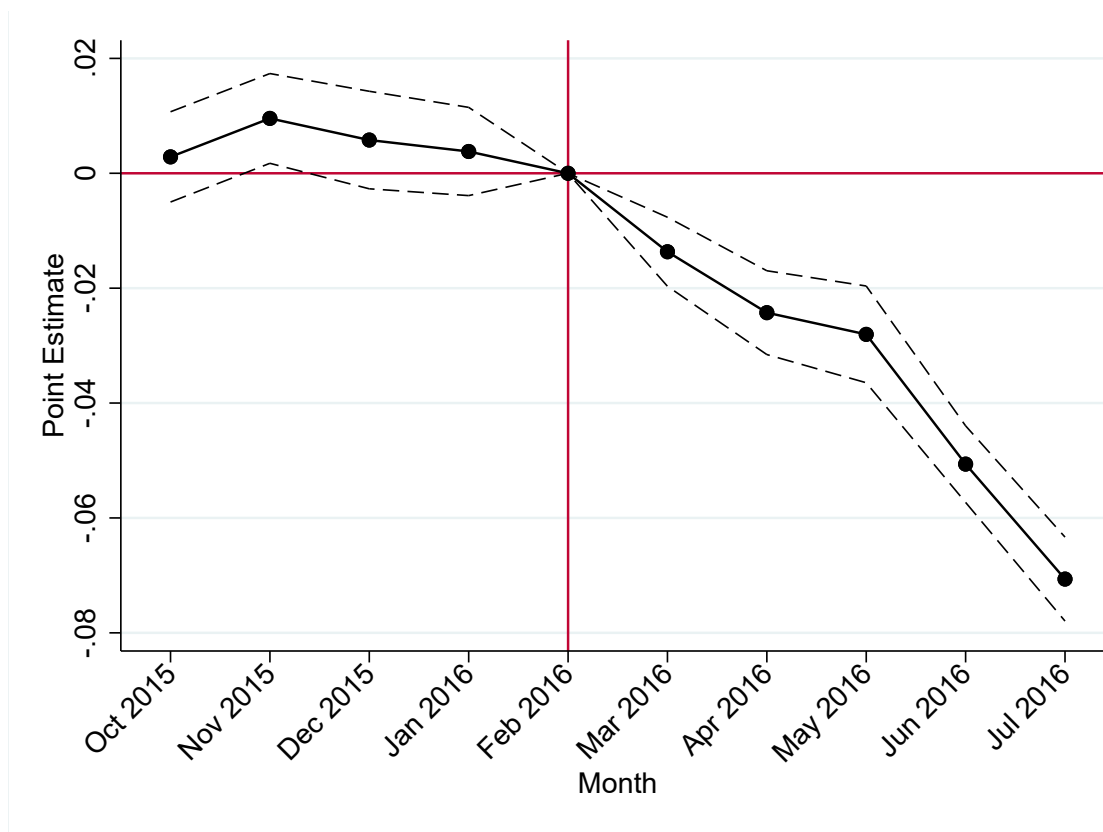


Figure 6: Pre-trend of \$35,000 Origination

This figure presents point estimates and 95% confidence intervals of the regression in Equation 1, where the dummy POST has been replaced by monthly dummies (excluding February 2016). The unit of observation is a loan. The dependent variable is a dummy that equals one if the loan amount equals \$35,000, and zero otherwise. I compute 20 default probability bins to form Default Bin FE. The regression contains default bin, maturity, state, employment length, home ownership, 4-digit range FICO score, \$10,000 income bin, delinquencies in the last 2 years, and 10% debt-to-income ratio fixed effects. Loans in the sample are issued between October 2015 and July 2016 for amounts between \$20,000 and \$35,000. Standard errors are robust to heteroskedasticity.

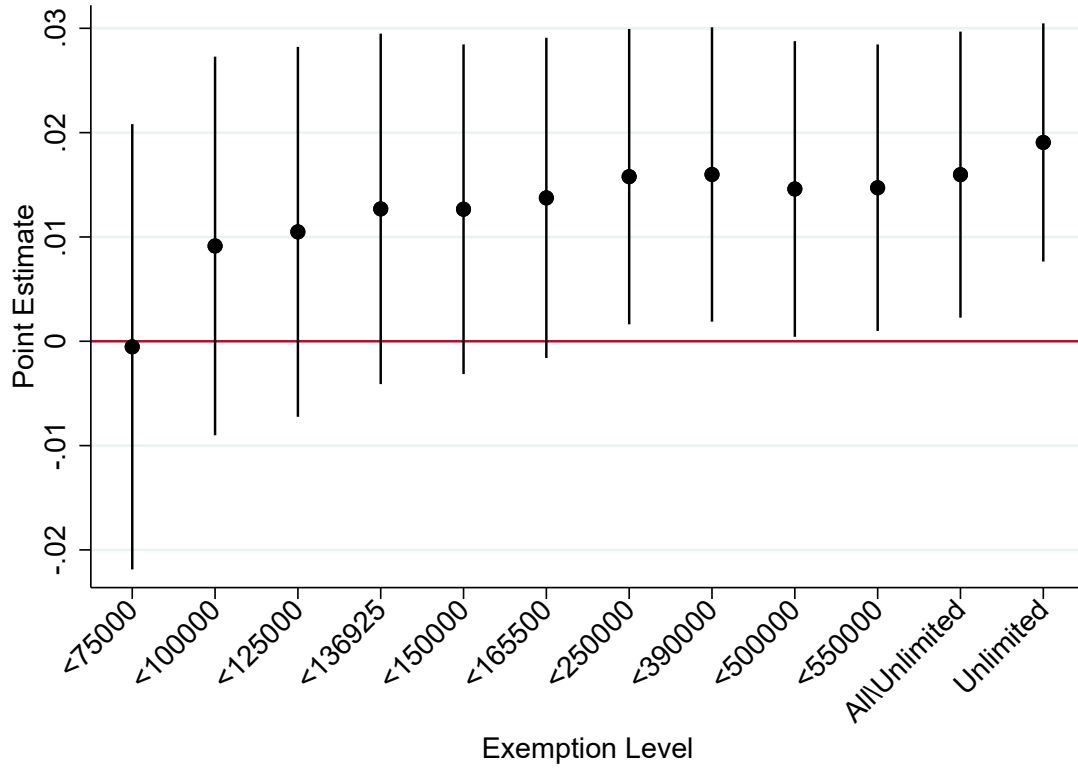
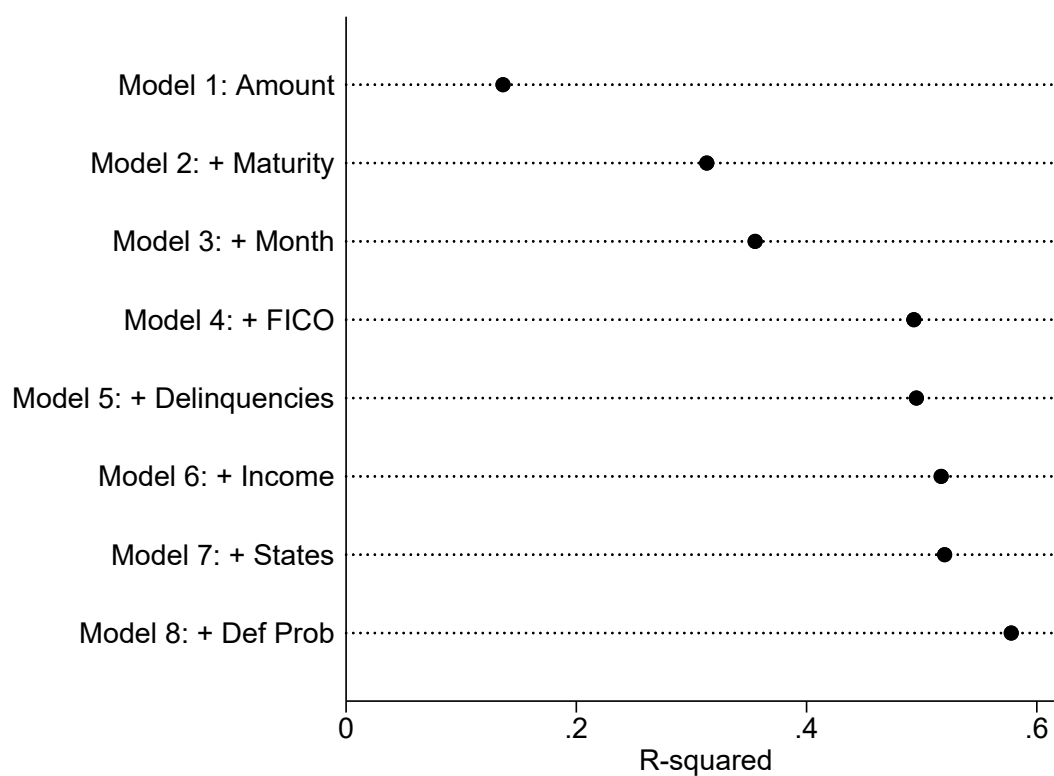


Figure 7: DID point estimate in different samples

This figure presents point estimates and 95% confidence intervals of the main regression in Equation 2, in sample with different Homestead Exemptions (on the x-axis). The unit of observation is a loan. In all regressions the dependent variable is a dummy that equals one if the loan is in default, and zero otherwise. The independent variable DID is a dummy that equals one if the loan is issued after March 2016 and its amount is \$35,000, and zero otherwise. I compute 20 default probability bins to form Default Bin FE. All regressions contain \$5,000 loan amount bin, default bin, maturity, state, employment length, home ownership, 4-digit range FICO score, \$10,000 income bin, delinquencies in the last 2 years, and 10% debt-to-income ratio (all interacted with month except loan amount bin) fixed effects. All regressions include loans issued between October 2015 and July 2016 for amounts between \$20,000 and \$35,000. Standard errors are clustered at state level.

Figure 8: R-squared of LC Interest Rate controlling for different observables



This figure depicts the R-squared of eight regressions of the same dependent variable (interest rate), on a richer and richer set of controls. Model 1 contains only loan size fixed effects. Model 2 adds maturity fixed effects. Model 3 adds month FE. Model 4 adds 4-digit FICO score range fixed effects. Model 5 adds delinquencies in the last two years fixed effects. Model 6 adds \$10,000 income bin fixed effects. Model 7 adds state fixed effects. Finally, Model 8 adds default probability bin fixed effects.

Tables

Table 1: Homestead exemptions by State

State	Homestead Exemption	State	Homestead Exemption
Alabama	15,000	Montana	250,000
Alaska	72,900	Nebraska	60,000
Arizona	150,000	Nevada	550,000
Arkansas	Unlimited	New Hampshire	100,000
California	75,000	New Jersey	0
Colorado	75,000	New Mexico	60,000
Connecticut	75,000	New York	165,550
Delaware	125,000	North Carolina	35,000
District of Columbia	Unlimited	North Dakota	100,000
Florida	Unlimited	Ohio	136,925
Georgia	21,500	Oklahoma	Unlimited
Hawaii	20,000	Oregon	40,000
Idaho	100,000	Pennsylvania	0
Illinois	15,000	Rhode Island	500,000
Indiana	19,300	South Carolina	58,255
Iowa	Unlimited	South Dakota	Unlimited
Kansas	Unlimited	Tennessee	5,000
Kentucky	5,000	Texas	Unlimited
Louisiana	35,000	Utah	20,000
Maine	47,500	Vermont	125,000
Maryland	22,975	Virginia	5,000
Massachusetts	500,000	Washington	125,000
Michigan	30,000	West Virginia	25,000
Minnesota	390,000	Wisconsin	75,000
Mississippi	75,000	Wyoming	20,000
Missouri	15,000		

This table shows the homestead exemption level in all the United States.

Table 2: Summary Statistics

	Mean	Median	SD
Panel A: Loan Characteristics			
Loan amount	26,143.09	25,000	5,208.29
Interest rate (%)	12.96	12.59	4.65
Installment	739.76	704.10	199.36
Long Maturity	0.50		
Default	0.183		
Panel B: Borrowers Characteristics			
FICO	699.71	695	31.22
Debt-to-Income (%)	19.45	8.63	18.94
Annual Income (\$)	109,602.40	92,000	90,750.80
Revolving balance	29,412.97	21,981	35,520.93
Revolving utilization (%)	56.43	57.7	23.72
Delinquencies (2yrs)	0.21		
Home owner	0.10		
Mortgage	0.61		
Rent	0.29		
N		63,300	

This table shows summary statistics of the main sample of Lending Club borrowers in pre-expansion months, which includes all loans whose listing date is between October 2015 and February 2016, for an amount between \$20,000 and \$35,000.

Table 3: Distribution of Loan Purposes

Purpose	Frequency	Percent
Car	422	0.35
Credit Card	29,220	24.40
Debt Consolidation	74,131	61.91
Home Improvement	7,979	6.66
House	491	0.41
Major Purchase	1,963	1.64
Medical	463	0.39
Moving	211	0.18
Other	3,352	2.80
Renewable Energy	36	0.03
Small Business	1,354	1.13
Vacation	125	0.10
N	119,747	100

This table shows the distribution of self-reported loan purposes of the main sample of Lending Club borrowers, which includes all loans whose listing date is between October 2015 and July 2016, for an amount between \$20,000 and \$35,000.

Table 4: Summary Statistics of USD 35,000 loans

	35,000 Pre	35,000 Post
Panel A: Loan Characteristics		
Interest rate (%)	14.33	15.82
Installment	1,011.41	1,078.82
Long Maturity	0.51	0.40
Default	0.19	0.21
Panel B: Borrowers Characteristics		
FICO	700.1	698.21
Debt-to-Income (%)	18.82	18.84
Annual Income (\$)	148,471	147,173.5
Revolving balance (\$)	43,428	40,244.24
Revolving utilization (%)	59.58	58.56
Delinquencies (2yrs)	0.22	0.24
Home owner	0.11	0.12
Mortgage	0.65	0.64
Rent	0.24	0.23
N	9,791	6,712

This table shows summary statistics of Lending Club borrowers choosing \$35,000 loans in pre-expansion months, which includes loans whose listing date is between October 2015 and February 2016.

Table 5: Event study of takeup

	(1)	(2)	(3)
	35k Loan	35k Loan	35k Loan
POST	-0.0369*** (0.00179)	-0.0554*** (0.00333)	-0.0941*** (0.00539)
Interest Rate	1.860*** (0.0412)	2.230*** (0.0542)	2.208*** (0.0567)
Default Bin FE	Yes	Yes	Yes
Maturity FE	Yes	Yes	Yes
State FE	Yes	Yes	Yes
Employment FE	Yes	Yes	Yes
Home Ownership FE	Yes	Yes	Yes
Fico FE	Yes	Yes	Yes
Income Bin FE	Yes	Yes	Yes
Delinquencies FE	Yes	Yes	Yes
DTI Bin FE	Yes	Yes	Yes
Cluster	State	State	State
Sample	20-35	25-35	30-35
Mean	.15	.29	.53
N	117329	60691	33637

This table shows the result of regression in Equation (1) and measures the extent of the selection into new loans after March 2016. The unit of observation is a loan. In all columns the dependent variable is a dummy that equals one if the loan amount equals \$35,000, and zero otherwise. The independent variable POST is a dummy that equals one if the loan is issued after March 2016, and zero otherwise. I compute 20 default probability bins to form Default Bin FE. All regressions contain default bin, maturity, state, employment length, home ownership, 4-digit range FICO score, \$10,000 income bin, delinquencies in the last 2 years, and 10% debt-to-income ratio fixed effects. Loans in all columns are issued between October 2015 and July 2016. The first column contains loans between \$20,000 and \$35,000. The second column contains loans between \$25,000 and \$35,000. The last column contains loans between \$30,000 and \$35,000. Standard errors are robust to heteroskedasticity. *, ** and *** represent significance at the 10%, 5%, and 1% respectively.

Table 6: Regression of default on selected loans

	(1)	(2)
	Default	Default
DID	0.0233*** (0.00532)	0.0189*** (0.00563)
Amount Bin FE	Yes	Yes
Default Bin-Month FE	No	Yes
Maturity-Month FE	Yes	Yes
State-Month FE	Yes	Yes
Homeowner-Month FE	Yes	Yes
Employment-Month FE	Yes	Yes
FICO-Month FE	Yes	Yes
Income Bin-Month FE	Yes	Yes
Delinquencies-Month FE	Yes	Yes
DTI Bin-Month FE	Yes	Yes
Cluster	State	State
Sample	20-35	20-35
Mean	.18	.18
N	117135	117135

This table shows the result of regression in Equation (2) and measures the relative change in default rate of \$35,000 loans after March 2016, relative to smaller loans. The unit of observation is a loan. In all columns the dependent variable is a dummy that equals one if the loan is in default, and zero otherwise. The independent variable DID is a dummy that equals one if the loan is issued after March 2016 and its amount is \$35,000, and zero otherwise. I compute 20 default probability bins to form Default Bin FE. The regression in the first column contains \$5,000 loan amount bin, maturity, state, employment length, home ownership, 4-digit range FICO score, \$10,000 income bin, delinquencies in the last 2 years, and 10% debt-to-income ratio (all interacted with month except loan amount bin) fixed effects. The regression in the second column adds default bin-month fixed effects. Loans in all columns are issued between October 2015 and July 2016 for amounts between \$20,000 and \$35,000. Standard errors are clustered at state level. *, ** and *** represent significance at the 10%, 5%, and 1% respectively.

Table 7: Regression of share repaid on selected loans

	(1)	(2)
	% Repaid	% Repaid
DID	-0.0139*** (0.00472)	-0.0122** (0.00486)
Amount Bin FE	Yes	Yes
Default Bin-Month FE	No	Yes
Maturity-Month FE	Yes	Yes
State-Month FE	Yes	Yes
Homeowner-Month FE	Yes	Yes
Employment-Month FE	Yes	Yes
FICO-Month FE	Yes	Yes
Income Bin-Month FE	Yes	Yes
Delinquencies-Month FE	Yes	Yes
DTI Bin-Month FE	Yes	Yes
Cluster	State	State
Sample	20-35	20-35
Mean	.78	.78
N	117135	117135

This table shows the result of regression in Equation (2) and measures the relative change in shared repaid by \$35,000 loans after March 2016, relative to smaller loans. The unit of observation is a loan. In all columns the dependent variable is the share of the initial loan repaid, computed as outstanding debt at the end of the loan divided by initial debt. The independent variable DID is a dummy that equals one if the loan is issued after March 2016 and its amount is \$35,000, and zero otherwise. I compute 20 default probability bins to form Default Bin FE. The regression in the first column contains \$5,000 loan amount bin, maturity, state, employment length, home ownership, 4-digit range FICO score, \$10,000 income bin, delinquencies in the last 2 years, and 10% debt-to-income ratio (all interacted with month except loan amount bin) fixed effects. The regression in the second column adds default bin-month fixed effects. Loans in all columns are issued between October 2015 and July 2016 for amounts between \$20,000 and \$35,000. Standard errors are clustered at state level. *, ** and *** represent significance at the 10%, 5%, and 1% respectively.

Table 8: Regression of default on 30,000 loans

	(1)	(2)
	Default	Default
DID 30k	0.00466 (0.00432)	0.00332 (0.00464)
Amount Bin FE	Yes	Yes
Default Bin-Month FE	No	Yes
Maturity-Month FE	Yes	Yes
State-Month FE	Yes	Yes
Homeowner-Month FE	Yes	Yes
Employment-Month FE	Yes	Yes
FICO-Month FE	Yes	Yes
Income Bin-Month FE	Yes	Yes
Delinquencies-Month FE	Yes	Yes
DTI Bin-Month FE	Yes	Yes
Cluster	State	State
Sample	20-34,975	20-34,975
Mean	.18	.18
N	Yes	Yes

This table shows the result of a robustness test. Using the same regression in Equation (2) I test whether there is a relative change in default rate of \$30,000 loans after March 2016, relative to other loans (but excluding \$35,000 loans). The unit of observation is a loan. In all columns the dependent variable is a dummy that equals one if the loan is in default, and zero otherwise. The independent variable DID 30k is a dummy that equals one if the loan is issued after March 2016 and its amount is \$30,000, and zero otherwise. I compute 20 default probability bins to form Default Bin FE. The regression in the first column contains \$5,000 loan amount bin, maturity, state, employment length, home ownership, 4-digit range FICO score, \$10,000 income bin, delinquencies in the last 2 years, and 10% debt-to-income ratio (all interacted with month except loan amount bin) fixed effects. The regression in the second column adds default bin-month fixed effects. Loans in all columns are issued between October 2015 and July 2016 for amounts between \$20,000 and \$34,975. Standard errors are clustered at state level. *, ** and *** represent significance at the 10%, 5%, and 1% respectively.

Table 9: Summary Statistics of loans and borrowers in low and high exemption states before March 2016

	Low Exemption	High Exemption
Panel A: Loan Characteristics		
Loan amount	26,201.02	26,105
Interest rate (%)	13.03	12.91
Installment	734.28	743.32
Long Maturity	0.53	0.48
Default	0.18	0.18
Panel B: Borrowers Characteristics		
FICO	699.90	699.58
Debt-to-Income (%)	19.99	19.11
Annual Income (\$)	106,635.8	111,527.7
Outstanding Debt (no mortgage) (\$)	74073.17	71270.66
Delinquencies (2yrs)	0.34	0.33
Home owner	0.12	0.11
Mortgage	0.66	0.56
Rent	0.22	0.33
N	24,924	38,406

This table shows summary statistics of Lending Club borrowers in pre-expansion months, which includes all loans whose listing date is between October 2015 and February 2016, for an amount between \$20,000 and \$35,000. The first column shows statistics for borrowers living in states with exemption lower or equal than median (\$75,000). The second column shows statistics for borrowers living in states with higher than median exemption.

Table 10: Take up of 35,000 loans on exemption level

	(1)	(2)
	35k Loan	35k Loan
DID HSE	0.000891 (0.00373)	0.000545 (0.00374)
Interest Rate	2.038*** (0.0457)	1.918*** (0.0406)
Amount Bin FE	Yes	Yes
Default Bin-Month FE	No	Yes
Maturity-Month FE	Yes	Yes
State FE	Yes	Yes
Homeowner-Month FE	Yes	Yes
Employment-Month FE	Yes	Yes
FICO-Month FE	Yes	Yes
Income Bin-Month FE	Yes	Yes
Delinquencies-Month FE	Yes	Yes
DTI Bin-Month FE	State	State
Cluster	20-35	20-35
Sample	.14	.14
Mean	117135	117135

This table shows the result of regression in Equation (3) and measures whether the extent of the selection into new loans after March 2016 is different in states with higher exemption, relative to states with lower exemptions. The unit of observation is a loan. In all columns the dependent variable is a dummy that equals one if the loan amount equals \$35,000, and zero otherwise. The independent variable DID HSE is a dummy that equals one if the loan is issued after March 2016 in a state with higher than median exemption, and zero otherwise. I compute 20 default probability bins to form Default Bin FE. The regression in the first column contains \$5,000 loan amount bin, maturity, state, employment length, home ownership, 4-digit range FICO score, \$10,000 income bin, delinquencies in the last 2 years, and 10% debt-to-income ratio (all interacted with month except loan amount bin) fixed effects. The regression in the second column adds default bin-month fixed effects. Loans in all columns are issued between October 2015 and July 2016 for amounts between \$20,000 and \$35,000. Standard errors are clustered at state level. *, ** and *** represent significance at the 10%, 5%, and 1% respectively.

Table 11: Regression of default on selected loans in different HSE samples

	(1)	(2)
	Default	Default
DID	-0.000528 (0.0103)	0.0299*** (0.00514)
Amount Bin FE	Yes	Yes
Default Bin-Month FE	Yes	Yes
Maturity-Month FE	Yes	Yes
State-Month FE	Yes	Yes
Homeowner-Month FE	Yes	Yes
Employment-Month FE	Yes	Yes
FICO-Month FE	Yes	Yes
Income Bin-Month FE	Yes	Yes
Delinquencies-Month FE	Yes	Yes
DTI Bin-Month FE	Yes	Yes
Cluster	State	State
Sample	20-35	20-35
Exemption	Low	High
Mean	.18	.18
N	45968	70988

This table shows the result of regression in Equation (2) and measures the relative change in default rate of \$35,000 loans after March 2016, relative to smaller loans, in samples with different exemptions levels. The unit of observation is a loan. In all columns the dependent variable is a dummy that equals one if the loan is in default, and zero otherwise. The independent variable DID is a dummy that equals one if the loan is issued after March 2016 and its amount is \$35,000, and zero otherwise. I compute 20 default probability bins to form Default Bin FE. All regressions contain \$5,000 loan amount bin, default bin, maturity, state, employment length, home ownership, 4-digit range FICO score, \$10,000 income bin, delinquencies in the last 2 years, and 10% debt-to-income ratio (all interacted with month except loan amount bin) fixed effects. The sample in the first column contains loans issued in states with exemption lower or equal than median (\$75,000). The sample in the second columns contains loans issued in states with higher than median exemption. Loans in all columns are issued between October 2015 and July 2016 for amounts between \$20,000 and \$35,000. Standard errors are clustered at state level. *, ** and *** represent significance at the 10%, 5%, and 1% respectively.

Table 12: Regression of default on selected loans in different HSE samples

	(1)	(2)	(3)	(4)
	Default	Default	Default	Default
DID	0.00517 (0.0122)	-0.0133 (0.0219)	0.0308*** (0.00722)	0.0289** (0.0103)
Amount Bin FE	Yes	Yes	Yes	Yes
Default Bin-Month FE	Yes	Yes	Yes	Yes
Maturity-Month FE	Yes	Yes	Yes	Yes
State-Month FE	Yes	Yes	Yes	Yes
Homeowner-Month FE	Yes	Yes	Yes	Yes
Employment-Month FE	Yes	Yes	Yes	Yes
FICO-Month FE	Yes	Yes	Yes	Yes
Income Bin-Month FE	Yes	Yes	Yes	Yes
Delinquencies-Month FE	Yes	Yes	Yes	Yes
DTI Bin-Month FE	Yes	Yes	Yes	Yes
Cluster	State	State	State	State
Sample	20-35	20-35	20-35	20-35
HSE Quartile	First	Second	Third	Fourth
Mean	.18	.18	.18	.19
N	26769	19033	42653	28185

This table shows the result of regression in Equation (2) and measures the relative change in default rate of \$35,000 loans after March 2016, relative to smaller loans, in samples with different exemptions levels. The unit of observation is a loan. In all columns the dependent variable is a dummy that equals one if the loan is in default, and zero otherwise. The independent variable DID is a dummy that equals one if the loan is issued after March 2016 and its amount is \$35,000, and zero otherwise. I compute 20 default probability bins to form Default Bin FE. All regressions contain \$5,000 loan amount bin, default bin, maturity, state, employment length, home ownership, 4-digit range FICO score, \$10,000 income bin, delinquencies in the last 2 years, and 10% debt-to-income ratio (all interacted with month except loan amount bin) fixed effects. The sample in the first column contains loans issued in states with exemption lower than the first quartile (\$21,500). The sample in the second column contains loans issued in states whose exemption is between the first and the second quartile (\$75,000). The sample in the third column contains loans issued in states whose exemption is between the second and the third quartile (\$250,000). Finally, the sample in the last column contains loans issued in states whose exemption is above or equal than the third quartile. Loans in all columns are issued between October 2015 and July 2016 for amounts between \$20,000 and \$35,000. Standard errors are clustered at state level. *, ** and *** represent significance at the 10%, 5%, and 1% respectively.

Table 13: Regression of share repaid on selected loans in different HSE samples

	(1)	(2)
	% Repaid	% Repaid
DID	-0.00183 (0.00986)	-0.0178*** (0.00421)
Amount Bin FE	Yes	Yes
Default Bin-Month FE	Yes	Yes
Maturity-Month FE	Yes	Yes
State-Month FE	Yes	Yes
Homeowner-Month FE	Yes	Yes
Employment-Month FE	Yes	Yes
FICO-Month FE	Yes	Yes
Income Bin-Month FE	Yes	Yes
Delinquencies-Month FE	Yes	Yes
DTI Bin-Month FE	Yes	Yes
Cluster	State	State
Sample	20-35	20-35
Exemption	Low	High
Mean	.77	.79
N	45968	70988

This table shows the result of regression in Equation (2) and measures the relative change in shared repaid by \$35,000 loans after March 2016, relative to smaller loans, in samples with different exemptions levels. The unit of observation is a loan. In all columns the dependent variable is the share of the initial loan repaid, computed as outstanding debt at the end of the loan divided by initial debt. The independent variable DID is a dummy that equals one if the loan is issued after March 2016 and its amount is \$35,000, and zero otherwise. I compute 20 default probability bins to form Default Bin FE. All regressions contain \$5,000 loan amount bin, default bin, maturity, state, employment length, home ownership, 4-digit range FICO score, \$10,000 income bin, delinquencies in the last 2 years, and 10% debt-to-income ratio (all interacted with month except loan amount bin) fixed effects. The sample in the first column contains loans issued in states with exemption lower or equal than median (\$75,000). The sample in the second columns contains loans issued in states with higher than median exemption. Loans in all columns are issued between October 2015 and July 2016 for amounts between \$20,000 and \$35,000. Standard errors are clustered at state level. *, ** and *** represent significance at the 10%, 5%, and 1% respectively.

Table 14: Regression of Recovery on High Exemption Dummy

	(1)	(2)
	Recovery	% Recovered
High HSE Dummy	17.68 (11.42)	0.000797 (0.000596)
Default Bin-Month FE	Yes	Yes
Amount Bin FE	Yes	Yes
Maturity-Month FE	Yes	Yes
Homeowner-Month FE	Yes	Yes
Employment-Month FE	Yes	Yes
FICO-Month FE	Yes	Yes
Income Bin-Month FE	Yes	Yes
Delinquencies-Month FE	Yes	Yes
DTI Bin-Month FE	Yes	Yes
SE	Robust	Robust
Sample	Pre March 2016	Pre March 2016
Mean	403.22	.02
N	63207	62361

This table tests whether post charge-off recoveries are different for states below and above the median exemption. In the first column the dependent variable is the post charge off recovery (in dollars). The dependent variable in the second column is the ratio between the amount recovered and the debt outstanding at the time of default. In both columns High HSE Dummy is equal to one if the loan is issued in a state with higher or equal than median exemption (\$75,000). I compute 20 default probability bins to form Default Bin FE. All regressions contain \$5,000 loan amount bin, default bin, maturity, state, employment length, home ownership, 4-digit range FICO score, \$10,000 income bin, delinquencies in the last 2 years, and 10% debt-to-income ratio (all interacted with month except loan amount bin) fixed effects. Loans in all columns are issued in the pre-expansion period, between October 2015 and February 2016 and for amounts between \$20,000 and \$35,000. Standard errors are clustered at state level. *, ** and *** represent significance at the 10%, 5%, and 1% respectively.

Table 15: Regression of outstanding existing debt on High Exemption Dummy

	(1)	(2)	(3)
	Balance (No Mort)	Install Loans	Revolving
High HSE Dummy	-1946.8*** (372.3)	-1565.8*** (498.6)	-229.0 (237.9)
Amount Bin FE	Yes	Yes	Yes
Default Bin-Month FE	Yes	Yes	Yes
Homeowner-Month FE	Yes	Yes	Yes
Employment-Month FE	Yes	Yes	Yes
FICO-Month FE	Yes	Yes	Yes
Income Bin-Month FE	Yes	Yes	Yes
Delinquencies-Month FE	Yes	Yes	Yes
DTI Bin-Month FE	Yes	Yes	Yes
SE	Robust	Robust	Robust
Sample	Pre March 2016	Pre March 2016	Pre March 2016
Mean	72148.71	43426.33	29303.12
N	63207	30305	63207

This table tests whether borrowers in states with higher exemptions have different levels of indebtedness outside LC. In the first column the dependent variable is total debt (in dollars, excluding mortgages). The dependent variable in the second column is debt in installment loans (in dollars). The dependent variable in the third column is revolving debt balance (in dollars). In all columns High HSE Dummy is equal to one if the loan is issued in a state with higher or equal than median exemption (\$75,000). I compute 20 default probability bins to form Default Bin FE. All regressions contain \$5,000 loan amount bin, default bin, maturity, state, employment length, home ownership, 4-digit range FICO score, \$10,000 income bin, delinquencies in the last 2 years, and 10% debt-to-income ratio (all interacted with month except loan amount bin) fixed effects. Loans in all columns are issued in the pre-expansion period, between October 2015 and February 2016 and for amounts between \$20,000 and \$35,000. Standard errors are clustered at state level. *, ** and *** represent significance at the 10%, 5%, and 1% respectively.

Table 16: Regression of FICO(end)-FICO(start) on selected loans

	(1)	(2)
	$\Delta(\text{FICO})$	$\Delta(\text{FICO})$
DID	-3.761***	-2.784**
	(1.073)	(1.119)
Amount Bin FE	Yes	Yes
Default Bin-Month FE	No	Yes
Maturity-Month FE	Yes	Yes
State-Month FE	Yes	Yes
Homeowner-Month FE	Yes	Yes
Employment-Month FE	Yes	Yes
FICO-Month FE	Yes	Yes
Income Bin-Month FE	Yes	Yes
Delinquencies-Month FE	Yes	Yes
DTI Bin-Month FE	Yes	Yes
Cluster	State	State
Sample	20-35	20-35
Mean	-14.2	-14.2
N	114213	114213

This table shows the result of regression in Equation (2) and measures the relative change in the 4-digit range FICO score by \$35,000 loans after March 2016, relative to smaller loans. The unit of observation is a loan. In both columns the dependent variable is difference between the last and the first recorded 4-digit FICO score band. The independent variable DID is a dummy that equals one if the loan is issued after March 2016 and its amount is \$35,000, and zero otherwise. I compute 20 default probability bins to form Default Bin FE. The regression in the first column contains \$5,000 loan amount bin, maturity, state, employment length, home ownership, 4-digit range FICO score, \$10,000 income bin, delinquencies in the last 2 years, and 10% debt-to-income ratio (all interacted with month except loan amount bin) fixed effects. The regression in the second column adds default bin-month fixed effects. Loans in all columns are issued between October 2015 and July 2016 for amounts between \$20,000 and \$35,000. Standard errors are clustered at state level. *, ** and *** represent significance at the 10%, 5%, and 1% respectively.

Table 17: Regression of FICO(end)-FICO(start) on selected loans

	(1)	(2)
	$\Delta(\text{FICO})$	$\Delta(\text{FICO})$
DID	-1.516	-3.332**
	(1.582)	(1.400)
Amount Bin FE	Yes	Yes
Default Bin-Month FE	Yes	Yes
Maturity-Month FE	Yes	Yes
State-Month FE	Yes	Yes
Homeowner-Month FE	Yes	Yes
Employment-Month FE	Yes	Yes
FICO-Month FE	Yes	Yes
Income Bin-Month FE	Yes	Yes
Delinquencies-Month FE	Yes	Yes
DTI Bin-Month FE	Yes	Yes
Cluster	State	State
Sample	20-35	20-35
Exemption	Low	High
Mean	-14.47	-13.99
N	44827	69211

This table shows the result of regression in Equation (2) and measures the relative change in the 4-digit range FICO score by \$35,000 loans after March 2016, relative to smaller loans, in samples with different exemptions levels. The unit of observation is a loan. In both columns the dependent variable is difference between the last and the first recorded 4-digit FICO score band. The independent variable DID is a dummy that equals one if the loan is issued after March 2016 and its amount is \$35,000, and zero otherwise. I compute 20 default probability bins to form Default Bin FE. All regressions contain \$5,000 loan amount bin, default bin, maturity, state, employment length, home ownership, 4-digit range FICO score, \$10,000 income bin, delinquencies in the last 2 years, and 10% debt-to-income ratio (all interacted with month except loan amount bin) fixed effects. The sample in the first column contains loans issued in states with exemption lower or equal than median (\$75,000). The sample in the second columns contains loans issued in states with higher than median exemption. Loans in all columns are issued between October 2015 and July 2016 for amounts between \$20,000 and \$35,000. Standard errors are clustered at state level. *, ** and *** represent significance at the 10%, 5%, and 1% respectively.

Table 18: Regression of default on selected loans (up to 40,000)

	(1)	(2)
	Default	Default
DID	0.0279*** (0.00507)	0.0231*** (0.00537)
Amount Bin FE	Yes	Yes
Default Bin-Month FE	No	Yes
Maturity-Month FE	Yes	Yes
State-Month FE	Yes	Yes
Homeowner-Month FE	Yes	Yes
Employment-Month FE	Yes	Yes
FICO-Month FE	Yes	Yes
Income Bin-Month FE	Yes	Yes
Delinquencies-Month FE	Yes	Yes
DTI Bin-Month FE	Yes	Yes
Cluster	State	State
Sample	20-40	20-40
Mean	.18	.18
N	119506	119506

This table shows the result of regression in Equation (2) and measures the relative change in default rate of \$35,000 loans after March 2016, relative to other loans (including newly available \$40,000). The unit of observation is a loan. In all columns the dependent variable is a dummy that equals one if the loan is in default, and zero otherwise. The independent variable DID is a dummy that equals one if the loan is issued after March 2016 and its amount is \$35,000, and zero otherwise. I compute 20 default probability bins to form Default Bin FE. The regression in the first column contains \$5,000 loan amount bin, maturity, state, employment length, home ownership, 4-digit range FICO score, \$10,000 income bin, delinquencies in the last 2 years, and 10% debt-to-income ratio (all interacted with month except loan amount bin) fixed effects. The regression in the second column adds default bin-month fixed effects. Loans in all columns are issued between October 2015 and July 2016 for amounts between \$20,000 and \$40,000. Standard errors are clustered at state level. *, ** and *** represent significance at the 10%, 5%, and 1% respectively.

Table 19: Regression of default on selected loans (up to 40,000)

	(1)	(2)
	Default	Default
DID	0.00188 (0.00941)	0.0349*** (0.00515)
Amount Bin FE	Yes	Yes
Default Bin-Month FE	Yes	Yes
Maturity-Month FE	Yes	Yes
State-Month FE	Yes	Yes
Homeowner-Month FE	Yes	Yes
Employment-Month FE	Yes	Yes
FICO-Month FE	Yes	Yes
Income Bin-Month FE	Yes	Yes
Delinquencies-Month FE	Yes	Yes
DTI Bin-Month FE	Yes	Yes
Cluster	State	State
Sample	20-40	20-40
Exemption	Low	High
Mean	-14.47	-13.98
N	46935	72395

This table shows the result of regression in Equation (2) and measures the relative change in default rate of \$35,000 loans after March 2016, relative to other loans (including newly available \$40,000), in samples with different exemptions levels. The unit of observation is a loan. In all columns the dependent variable is a dummy that equals one if the loan is in default, and zero otherwise. The independent variable DID is a dummy that equals one if the loan is issued after March 2016 and its amount is \$35,000, and zero otherwise. I compute 20 default probability bins to form Default Bin FE. All regressions contain \$5,000 loan amount bin, default bin, maturity, state, employment length, home ownership, 4-digit range FICO score, \$10,000 income bin, delinquencies in the last 2 years, and 10% debt-to-income ratio (all interacted with month except loan amount bin) fixed effects. The sample in the first column contains loans issued in states with exemption lower or equal than median (\$75,000). The sample in the second columns contains loans issued in states with higher than median exemption. Loans in all columns are issued between October 2015 and July 2016 for amounts between \$20,000 and \$35,000. Standard errors are clustered at state level. *, ** and *** represent significance at the 10%, 5%, and 1% respectively.

Table 20: Regression of default on selected loans for other purposes

	(1)	(2)
	Default	Default
DID	-0.00472 (0.0254)	-0.00240 (0.0242)
Amount Bin FE	Yes	Yes
Default Bin-Month FE	No	Yes
Maturity-Month FE	Yes	Yes
State-Month FE	Yes	Yes
Homeowner-Month FE	Yes	Yes
Employment-Month FE	Yes	Yes
FICO-Month FE	Yes	Yes
Income Bin-Month FE	Yes	Yes
Delinquencies-Month FE	Yes	Yes
DTI Bin-Month FE	Yes	Yes
Cluster	State	State
Sample	20-35	20-35
Mean	.2	.2
N	15737	15737

This table shows the result of regression in Equation (2) and measures the relative change in default rate of \$35,000 loans after March 2016, relative to smaller loans, in the sample of loan with purposes other than refinancing. The unit of observation is a loan. In all columns the dependent variable is a dummy that equals one if the loan is in default, and zero otherwise. The independent variable DID is a dummy that equals one if the loan is issued after March 2016 and its amount is \$35,000, and zero otherwise. I compute 20 default probability bins to form Default Bin FE. The regression in the first column contains \$5,000 loan amount bin, maturity, state, employment length, home ownership, 4-digit range FICO score, \$10,000 income bin, delinquencies in the last 2 years, and 10% debt-to-income ratio (all interacted with month except loan amount bin) fixed effects. The regression in the second column adds default bin-month fixed effects. Loans in all columns are issued between October 2015 and July 2016 for amounts between \$20,000 and \$35,000 and purpose different than "Credit Card" and "Debt Consolidation". Standard errors are clustered at state level. *, ** and *** represent significance at the 10%, 5%, and 1% respectively.