2018

# Essays in industrial organization of Peer-to-Peer online credit markets

BOSTON UNIVERSITY

GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

# ESSAYS IN INDUSTRIAL ORGANIZATION OF PEER-TO-PEER ONLINE CREDIT MARKETS

by

**FNU TALAL UR RAHIM**

M.Sc., Lahore School of Economics, 2009
M.Sc., London School of Economics & Political Science, 2011

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2018

Approved by

First Reader _____

Marc Rysman, PhD
Professor of Economics

Second Reader _____

Berardino Palazzo, PhD
Assistant Professor of Finance

Third Reader _____

Hiroaki Kaido, PhD
Assistant Professor of Economics

## Acknowledgments

First of all, I owe a big thank you to my family who provided me all kinds of support one can ever dream of. My mom, Saeeda Yasmine Khan, and my dad, Nasrullah Khan, raised me with immense love and care, and that has made me the man I am today. Furthermore, my siblings always took care of me as the youngest member of the family and advised me at several stages of my life.

At Boston University, I would like to express my gratitude to my advisors Marc Rysman, Berardino Palazzo and Hiroaki Kaido who provided invaluable advice and support during my graduate studies. In particular, I am forever indebted to Marc Rysman who encouraged me to pursue my research in the field of Industrial Organization. His guidance and advice in the past five years were of prime importance to my work and without it I would not have achieved my goals from graduate school. Berardino Palazzo provided an additional source of insight as to how credit markets work and how I should shape my research work. Discussions with him allowed me to improve my work significantly. Lastly, Hiroaki Kaido taught me the details of Econometric research and the intricacies that underlie great research, for which I am extremely grateful. I am truly honored to have these three great researchers to guide me during my dissertation stage.

I would also like to give a big thank you to many professors in the department of Economics who provided useful suggestions at various stages of my dissertation. In particular, I would like to thank Kehinde Ajayi, Randall Ellis, and Stephen Terry. Furthermore, many enthusiastic participants, at several reading groups and seminars at BU, provided great feedback throughout my dissertation stage.

My friends and colleagues at Boston University gave me a lot of encouragement and useful advice during my entire time here. Peter Wang, Svet Semov, Matt John-

son, Calvin Luscombe, Yuan Tian, Steven Bhardwaj, Jonathan Hercsh, Samarth Gupta, Hsyn Seren and Wenjia Zhu were instrumental in helping me survive graduate school. I owe them, and many more, a hearty thank you.

Members of the Boston University Economics Department also provided logistical support and occasionally moral support. In particular, Andrew Campolieto and Deborah Kasabian were very helpful.

# ESSAYS IN INDUSTRIAL ORGANIZATION OF PEER-TO-PEER ONLINE CREDIT MARKETS

**FNU TALAL UR RAHIM**

Boston University, Graduate School of Arts and Sciences, 2018

Major Professors: Marc Rysman, PhD, Professor of Economics

Berardino Palazzo, PhD, Assistant Professor of Finance

ABSTRACT

This dissertation consists of three separate essays on Peer-to-Peer (P2P) online credit markets. The first essay presents new empirical evidence of a decrease in loan demand and repayment when prices in the market are determined by competing lenders in an auction as compared to the case in which a platform directly controls all prices. The paper develops an econometric model of loan demand and repayment which is then used to predict borrower choices when they are offered prices set by lenders in a market. I find that when lenders set prices, borrowers are more likely to pick loans of shorter maturity and smaller sizes, and repay less. Aggregated at the market level, demand and repayment of credit fall by 10% and 2% respectively.

In the second paper, I quantify the effects of implementation of finer credit-scoring on credit demand, defaults and repayment in the context of a large P2P online credit platform. I exploit an exogenous change in the platform's credit scoring policy where the centralized price setting rules ensure that the one-to-one relationship between credit scores and prices remains intact unlike in a traditional credit market

where it is broken. The results show that a 1% increase in interest rate due to the implementation of finer credit scoring results in an average decrease of 0.29% in the requested loan amount, an average increase of 0.01 in the fraction of borrowers who default and an average increase of 0.02 in the fraction of loan repaid. These findings contribute to a better understanding of how a reduction in information asymmetry affects borrower choices in a credit market.

The third paper explores the main drivers behind the geographic expansion in demand for credit from P2P online platforms. It uses data from the two largest platforms in the United States to conduct an empirical analysis. By exploiting heterogeneity in local credit markets before the entry of P2P online platforms, the paper estimates the effect of local credit market conditions on demand for credit from P2P platforms. The paper uses a spatial autoregressive model for the main specification. We find that P2P consumer credit expanded more in counties with poor branch networks, lower concentration of banks, and lower leverage ratios.

# Contents

## List of Tables

## List of Figures

# Chapter 1

# Can Peer-to-Peer Online Platforms Improve Market Outcomes by Controlling Prices?

## 1.1 Introduction

An important function of a market is to allocate resources efficiently by allowing market participants to trade with each other and determine prices of resources in the process. However, this process is often hindered by many different types of frictions and their associated costs which can restrict a market from functioning properly. Recently, several Peer-to-Peer (P2P) online platforms have emerged fundamentally changing how economic agents trade goods and services. The emergence of platforms like Airbnb, Amazon, Ebay, Lending Club, Prosper and Uber have shown how technological advances can help to improve competition, resource allocation and asset utilization by facilitating trade in efficiently designed markets. They provide services like finding the best trader given your needs, providing information about goods and other traders' trustworthiness and trading history, aggregating small and thin local markets into bigger and thicker markets, and facilitating transfer of goods and payments. In doing so, these platforms reduce many frictions and their associated costs found in offline counterparts of such markets.

These platforms often also set the rules of trade which may prevent market failure. One such rule that *some* platforms use is that they directly control the prices within the markets they create.[1] This restricts a fundamental function of a com-

[1]For example, Lending Club, Prosper and Uber directly set prices, whereas Airbnb and Amazon allow sellers to set their own prices. Ebay gives its sellers even more freedom by allowing them to pick their own pricing mechanism which is either an auction or a posted pricing mechanism.

petitive market: its ability to aggregate information which is then reflected in the prices. If prices are not allowed to adjust freely in a market, it can hinder the process of information revelation through price discovery and thus lead to information asymmetry between buyers and sellers. As a consequence of asymmetric information the market may suffer from adverse selection which ultimately prevents the competitive equilibrium in the market from achieving first best allocation Akerlof (1970). This raises the question of whether controlling market prices will result in better or worse outcomes for consumers if a platform uses this market design choice to prevent market failure.

In this paper I use micro data from a large online P2P credit market to show evidence of better market outcomes when prices are set by the platform instead of by competing lenders in an auction. To investigate the channels through which a pricing mechanism can affect outcomes in this credit market, I first study how changes in contract terms, including prices (interest rates), affect borrower demand and repayment of credit. Since these decisions of borrowers are interdependent the effects of changes in contract terms can be nonlinear. Taking this into account, I specify an econometric model of loan demand and repayment with specific emphasis on the role of interdependencies in borrower choices and estimate it using granular data from Prosper.com, which is the second largest P2P online credit platform in the United States by loan volume.

I use the estimated parameters to conduct a counterfactual experiment in which borrowers are offered prices determined in an auction among lenders. To find the set of counterfactual prices I exploit a change in the pricing mechanism implemented by the platform and use machine learning to match borrowers under the two pricing mechanisms based on a rich set of borrower characteristics and market conditions.

Given the inefficiency of simple matching procedures in high dimensions, I turn the problem into a prediction problem: I first use several machine learning techniques to predict the contracted price for each borrower under the auction pricing mechanism using borrower and market characteristics. Here I use sample splitting to select the technique that gives the lowest psuedo out-of-sample prediction error[2]. Next, I use this estimated pricing function to predict the counterfactual prices for borrowers who were issued loans under the platform's posted-price mechanism and plug them back into the estimated loan demand and repayment model. This gives me borrower choices under this counterfactual scenario and then I aggregate them up to determine the new market outcomes.

I find that when lenders set prices using an auction, borrowers are more likely to switch to shorter maturity loan contracts, smaller loan sizes and lower repayment. Aggregated at the market level, demand for credit and repayment of credit owed fall by 10% and 2% respectively. This has important implications for an online platform's ability to improve the allocation of credit by controlling market prices. I discuss these findings stem from the platform's ability to screen borrowers using proprietary credit scoring technology which reduces the average costs of screening since the platform can do it at scale. Here I highlight the differences in the two price distributions. I show that the platform charged, on average, lower prices to borrowers than the market. Furthermore, the difference between the platform price and the market price is increasing in borrower riskiness. The platform charges a lower risk premium to the risky borrower than the market would. This key insight explains many of the distributional effects I find in this paper.

Given the nature of individual level data from a market in which prices are indi-

[2]In section 5 I explain that this approach has several advantages over its alternatives in calculating the counterfactual input prices I need to answer the question motivated above.

vidual specific, I was able to conduct the counterfactual experiment at the individual level. This allowed me to analyze the distributional effects of the in addition to the average effects discussed above. The distribution of the effects of change in pricing mechanism on borrowers' maturity choices show that the effects are positive and increasing in credit scores. This means that borrowers with higher credit scores are more likely to pick a loan of longer maturity than borrowers with low risk. The distribution of the effects on loan amount tell a very interesting story. Over here we have two effects: partial and full effects and both their distributions show positive effects in different ways. The distribution of partial effects on loan amount show that the effects are positive and decreasing in credit score. However, the distribution of full effects on loan amount show that that borrowers with the lowest or the highest credit scores are affected less than the borrowers who have credit scores near the median score. Lastly, the distributions of the partial and full effects on repayment show that the effects are positive and decreasing in credit scores. This means that the increases in repayment from high risk borrowers are bigger than increases in repayment from low risk ones. It also explains the decrease in aggregate defaults which is quite an accomplishment for a credit market.

This paper contributes to two main strands of literature. The first is the growing literature on multi-sided platforms, including the peer-to-peer platforms that make up the sharing economy. Questions about the effects of different platform design choices on market outcomes are of particular interest. Recent studies include the works of Cullen and Farronato (2015) who focus on matching short-term supply and demand on a platform for domestic tasks, Fradkin (2014) estimates search inefficiencies in a market for short term accommodation rentals, and study seller behavior under different pricing mechanisms in a general marketplace. On a more related

note to my paper, the theoretical work by Hagiu and Wright (2015a) and Hagiu and Wright (2015b) attempt to study the trade-offs faced by such platforms in their choice of operating as marketplaces or resellers.

Among the specific papers on online P2P credit platforms, there has been no attempt to estimate the structural parameters that capture the sensitivity of credit demand and repayment to different contract terms. Estimating such parameters becomes important when one has to estimate the effect of a different pricing mechanism on the aggregate market outcomes. Nonetheless, several reduced form papers on P2P online credit markets provide motivation for this approach. Pope and Sydnor (2011) and Ravina (2013) show how an applicant's personal characteristics (for example outward appearance and skin color) can affect her probability of getting a loan, Iyer et al. (2016) provide evidence that the market is able to determine interest rates that predict defaults better than the finest credit scores do, and Zhang and Liu (2012) provide evidence of investor herding behavior in these markets.

Two closely related papers Meyer (2013) and Wei and Lin (2016) show reduced form evidence of how a change in the pricing mechanism on P2P online credit platform affects lender returns, prices and probability of getting a loan. The contribution of my paper, on the other hand, is to estimate the effects of such a change in pricing mechanism on the demand and repayment behavior of borrowers. Moreover, I use a structural model to explain the channels through which the prices affect borrower choices. To that end I show how interdependencies in borrower choices reveal that full effects can be quite different from partial effects.

The second field this paper contributes to is the empirical literature on consumer and microcredit markets. A classic contribution here is by Karlan and Zinman (2005) who carry out an experiment in a credit market to identify sources of adverse selec-

tion. On the other hand, Einav et al. (2012) and Crawford and Schivardi (2016) use structural approaches to estimate the effects of contract terms on loan demand and repayment in consumer and firm credit markets. My paper builds on the framework proposed by Einav et al. (2012) by introducing loan maturity as an additional choice variable in the specification of loan demand. There are two main reasons to include this choice as part of the model. First, in many credit markets, and particularly in P2P online credit markets, choosing the maturity of a loan is part of the loan demand process, and Hertzberg et. al (2016) show how this choice can be a significant source of adverse selection in online P2P credit markets. They use a natural experiment that took place on Lendingclub.com to identify the effect of loan maturity on default to show that an increase in loan maturity has a negative effect on loan repayment and the magnitude of this effect is much bigger than that of an increase in interest rate. Second, since loan maturity affects both the choice of loan amount and the choice of repayment, a change in loan price has indirect effects on loan amount and repayment because that same change in price also affects loan maturity choice. Thus, the full effect of a change in price on loan size and repayment needs to take this into account and by ignoring it one could bias the price coefficients in the model.

The rest of the paper is structured as follows: Section 1.2 provides an institutional overview of P2P online credit markets with an emphasis on how they differ from traditional credit markets, section 1.3 presents the data and sample selection procedure, section 1.4 section 4 develops the econometric model and its estimation procedure, and presents estimation results, section 1.5 presents the case counterfactual experiment and its results, and section 1.6 gives conclusion.

## 1.2 Institutional Overview

Credit is considered an essential commodity for improving social welfare by allowing consumers to smooth consumption over time and by allowing firms to invest in new projects. Access to credit is often considered one of the hallmarks of a developed economy. However, a traditional credit market today is still plagued by many frictions, some of which have been shown to be reduced greatly within an online P2P credit market.

Over the past decade more than a thousand P2P online credit platforms have opened up across the world.[3] In the three biggest markets, China, United States and United Kingdom, cumulative loan volumes by Dec. 2015 reached $70 billion, $25 billion, and $7 billion, respectively.[4] In 2014 in U.S. alone, the five biggest platforms issued $3.5 billion in loans compared to $1.2 billion in 2013. However, this makes up a sliver of consumer debt in U.S. To put things in perspective, total outstanding credit card debt in the United States grew to $880 billion by July 2014. According to a Fitch report, the market volume in P2P online credit markets may grow to $114 billion in the medium term.[5] The U.S. market is dominated by two competing platforms named Lending Club and Prosper which together have a market share of over 90% in P2P small personal loans.

In a typical online P2P credit market borrowers seek loans from a group of lenders by posting their credit information on the platform website. The platform performs initial screening of borrowers, collects credit information, and sets loan contract terms including loan maturity, interest rate, and transaction fees. Individual and institutional investors decide how much to invest in each loan based on their own

[3] Americas Alternative Finance Benchmarking Report, 2015
[4] Citi Group Report, 2016
[5] Federal government data aggregated by www.nerdwallet.com

preferences. In this market, in its current form, the price vector is controlled by the platform while both borrowers and lenders are price takers and pick their own allocations.

The processes of obtaining and investing in a loan through a P2P online platform are similar across major platforms. In what follows, I will explain these processes in detail for Prosper.com which provided the data used in this paper and it is the second largest P2P online credit platform in the U.S. by loan volume. The main flows of information and funds are depicted in Figure 1.1.

To obtain a loan, a borrower first needs to be accepted by the platform to be able to post a listing for the loan. The platform accepts or rejects a borrower based on a credit check to make sure the borrower meets some basic cut off criterion[6]. If the borrower is accepted, the platform assigns him a credit grade which is a function of an external credit score and the platform's own proprietary credit score. Based on this credit grade, the borrower is offered a menu of loan contracts which differ in maturity, interest rate and loan origination fee. Each borrower is also assigned a loan limit based on his credit grade and this loan limit stays the same regardless of which loan contract the borrower picks. Once the borrower picks a loan contract and loan amount, $L$, a standardized listing for that loan is created on the platform's website which includes detailed information about the loan contract and borrower credit report.

To obtain a loan, a borrower first needs to be accepted by the platform to be able to post a listing for the loan. The platform accepts or rejects a borrower based on a credit check to make sure the borrower meets some basic cut off criterion[7]. If

[6]On Prosper.com a borrower needs to have a minimum Experian Scorex Plus score of 640 to be eligible to get a loan.

[7]On Prosper.com a borrower needs to have a minimum Experian Scorex Plus score of 640 to be eligible to get a loan.

the borrower is accepted, the platform assigns him a credit grade which is a function of an external credit score and the platform's own proprietary credit score. Based on this credit grade, the borrower is offered a menu of loan contracts which differ in maturity, interest rate and loan origination fee. Each borrower is also assigned a loan limit based on his credit grade and this loan limit stays the same regardless of which loan contract the borrower picks. Once the borrower picks a loan contract and loan amount, $L$, a standardized listing for that loan is created on the platform's website which includes detailed information about the loan contract and borrower credit report.

The listing stays open for one to two weeks during which time different lenders invest in the loan with a minimum investment of \$25 per loan per lender, i.e. $l_j \geq \$25$. If the requested loan amount is reached, $\sum_{j=1}^{J} l_j = L$, the listing is closed from further investment and the loan is issued to the borrower. In case the listing period is over but the desired amount is not reached, the listing is termed unsuccessful and the lenders get their investments back.[8]

To repay the loan, the borrower makes monthly payments with an interest rate $r_1$ and if he defaults (i.e. if the he chooses $s < 1$), the platform sells the loan to a debt collection agency and distributes the proceeds among the lenders of that specific loan in the ratio of their investments. A lender $j$ earns a gross return of $(1 + r_2)\, sl_j$ as the borrower repays the loan. The platform charges a loan origination fee, $f_1$, to each borrower which can range from 1% - 5% of loan amount. The platform also charges a 1% loan servicing fee, $f_2$, to lenders which is earned by platform as the loan is repaid.[9]

[8]The platform also allows the borrowers an option to convert the listing to a loan if it receives at least 70% of the requested amount.

[9]Note that $r_1 > r_2$ because $f_1, f_2 > 0$. Also note that there is no charge to a borrower for posting the initial listing

A few important points to note here are that these are small personal loans, the borrower does not need to provide any collateral to take the loan, and lastly the platform does not perform any monitoring of the borrower.[10] However, this does not mean that there is no penalty for the borrower if he defaults. A debt collection agency can take the borrower to court which would cost the borrower fees, time, and effort and eventually the remaining amount owed. Moreover, the borrower would also be penalized with a higher interest rate if he wishes to take a loan in the future because the repayment behavior of a borrower directly affects his credit score which can be accessed by any professional lender.

There are some unique features of a P2P online credit market which reduce certain frictions present in a traditional offline credit market. These are explained as follows:

*Lower search costs for borrower:* A borrower can apply to take a loan from many different lenders at the same time with a single application and save time and effort of contacting several lenders to find the best contract terms.

*Lower search and operational costs for lenders:* A lender does not have to engage in costly marketing activities to promote his loan contracts to the public and wait for borrowers to show in a branch office or a website, both of which require additional start-up and operating costs.

*Lower cost diversification for lenders:* A lender does not have to invest in an entire loan but instead can invest a small amount in a loan and be part of a syndicate for that one loan without incurring the high costs of creating a syndicate. Traditionally, the syndicated loan market was restricted for large corporate loans due to the costs associated with forming a syndicate of multiple lenders. However, in a P2P online credit market, such costs are incurred by the platform which is able to keep costs

[10] The borrower usually states the purpose of the loan in the loan listing, and the most common purpose is to repay previous credit card debt.

low due to innovations in technology and by utilizing economies of scale.

*Access to a significantly bigger credit market:* With the advent of a P2P online credit platforms, small lenders are able to lend to borrowers in geographical locations where these small lenders do not have a physical presence. Given the extremely low cost of transferring funds, the platform is able to create thickness in the market by aggregating thin and local credit markets into one big market for credit. Theoretically, this should give a small local lender access to the entire borrower population of a country which effectively reduces the need for the lender to have a physical presence near its borrowers. This has severe implications for increasing competition in the credit markets.

There are also differences in regulations for a P2P credit platform which allow it to scale up its operations. Since the platform is simply an intermediary that matches borrowers with savers in the credit market, it is not subject to the same set of regulations as a depository institution (e.g. a commercial bank) or an investment company (e.g. a mutual fund). There are two primary reasons for this: First, a crucial point of difference between a P2P credit platform and a traditional bank is that the platform does not solve the problem of liquidity mismatch between savers and borrowers in the same way that a depository institution does. Matching different borrowers and savers/investors based on different maturity preferences is a fundamental function of depository institutions. This also makes them susceptible to bank runs if there is a shock to the liquidity needs of savers (Diamond and Dybvig, 1983) or worse a contagion of bank runs (Allen and Gale, 2000). On the other hand, a P2P credit platform the platform acts purely as a match-maker and does not take any risk on behalf of investors. The investment of savers is not a liability of the platform but of the borrowers only. Hence, in case of a positive shock to investors' demand for

early withdrawal of investment, the investors can simply sell their claim on their loans in a secondary market. This way a bank run can be avoided and this is the reason that P2P credit platforms are not subject to any reserve requirements by the central bank. Second, unlike a depository institution or an investment company, the platform does not make investment decisions on behalf of the lenders. The platform simply provides the information to lenders and facilitates the transfer of funds. As a consequence of this, the platform is much less restricted in forming its ownership structure as it is free from any fiduciary requirements. These two key differences lower the regulatory and legal costs of starting and operating a firm and help to scale up its operations.

*Lenders are Price takers:* One big disadvantage to lenders is that they lose their bargaining power to set their own prices which would effectively mean they would be price takers if they want to participate in this market. The equilibrium prices in this market are not determined by lenders competing with each other in this unified and more competitive credit market but instead the prices are set directly by the platform. This last point raises the question of whether the platform is able to allocate credit as efficiently as one would see if the prices were determined by borrowers and lenders in the market using any bargaining or auction process. On the one hand the platform lowers several different frictions and their associated costs which result in more competition relative to a traditional offline credit market, while on the other hand the platform may theoretically reduce one of the biggest benefit of increasing competition – that of resultant set of prices which increase consumer surplus.

A pioneering feature of online P2P credit platforms, like Prosper and Lending Club, is that they specialize in screening borrowers at scale and then set prices ac-

cordingly. The idea is that the platform uses its proprietary credit scoring methodology, developed using advances in machine learning, to predict the probability of default. Based on that, the platform assigns a high price to a borrower if his probability of default is high. This is not to say that the platform assigns the right price because the ordering of prices set in the market already reflects probabilities of default. Instead the platform maintains the same ordering with a lower set of prices. This way of screening is already a lower cost method of screening, but since the platform is able to do this at scale, the average cost of screening is decreased.

Given the digital nature of such a market where the details of each transaction are recorded by a computer, it provides excellent opportunities for researchers to study consumer financial decision making. The data generated by these markets contain records of several decisions made by a consumer, the details of his choice set, and the market conditions when such decisions are made. Additionally, researchers can study how individual consumer decisions aggregate up to the market level to determine aggregate market outcomes.

## 1.3 Data and Sample Selection

The data for this paper come from Prosper.com which is the second largest P2P online credit platform in the U.S. by loan volume. These data contain all required loan specific and borrower specific variables. For each loan, I observe the amount of loan, maturity period, interest rate, amount repaid (till the end of sample) and time stamps for loan application, issuance and repayment. For each borrower I observe a rich set of credit variables from the Experian credit bureau, Prosper.com's own credit score, credit grade and demographic variables. Identifiers for each loan application, loan and borrower allows for seamless merging of different parts of the database.

Owing to the online nature of the platform, it can implement big changes to the workings of the market very quickly and at scale. To address this issue, I used 54 snapshots of Prosper.com from internet archives to look for changes in borrowing and lending processes over time. These proved to be quite useful in isolating a time period during which no such major changes took place.

For my main estimation sample, I selected all loans issued between May 1st, 2013 to June 30th, 2014 and their repayment data was observed until Feb 29th, 2016. I dropped loans by borrowers of the lowest credit grade since they were offered just one maturity contract. Moreover, I keep only new loans because modeling the evolution of borrower behavior for follow-up loans is outside the scope of this paper. Descriptive statistics are provided in Table 1.1.

Next I highlight the variation in the platform's pricing schedule that was used to identify the main price coefficients in the model to come. Figure 1.2 illustrates two examples of how the platform changes prices for identical borrowers over time. The dotted line (...) shows all borrowers which are observationally identical in terms of their risk measure, the expected loss rate, which is the finest measure of borrower riskiness that the platform uses. The solid line (---) gives another example for another set of identical borrowers but these borrowers which are less risky than the ones represented by the dotted line. The flat part of each line is evidence of the fact that at any given snapshot of time, all borrowers with the same estimated loss rate are considered identical and are assigned the same price. The variation over time in interest rates conditional on this risk measure is what I use in the model in section 4 to identify my coefficients of interest.

This expected loss rate assigned to a borrower can be interpreted as the expected loss on $1 of investment to that borrower or simply the default probability. Notice

that this is independent of loan amount and loan maturity. This measure is simply a function of the borrower's credit and demographic variables. The platform sets prices based on loan term, whether a borrower has taken a prior loan from its the same platform, a measure of borrower riskiness (expected loss rate) and market and macroeconomic factors. Here the identifying assumption is that an individual borrower's loan demand and repayment choices do not depend on those market and macroeconomic factors. Once a borrower is accepted by the platform, he expects to get a loan almost surely (i.e. with probability 0.99), his decision depends only on contract specific variables (price, term, and fees) and his own observed and unobserved demographic and credit characteristics. Hence, when the platform changes prices for identical borrowers over time, as shown in Figure 1.2 for two representative types of borrowers who vary in their expected loss rates, this variation is exogenous to a borrower's decision of loan contract, size and repayment.

## 1.4 An Econometric Model of Loan Demand and Repayment

I specify a model of loan demand and repayment with the objective of quantifying the effects of contract terms on borrower choices while taking into account the interdependencies in those choices. I assume each borrower has a liquidity need and is willing to borrow from a set of lenders on the platform which has been allowed by the platform. Each borrower is assigned a credit grade based on which he is offered a maximum loan amount and a set of two loan maturity contracts from which he can pick only one. The contracts differ in maturity, interest rate and loan origination fee but the loan limit is the same on both contracts. Each borrower then decides which maturity contract to pick, how much loan to take, constrained by the loan limit, and subsequently how much of the borrowed amount to repay in order to maximize his

expected utility from these choices. [11]

The model adapts the framework of Einav et al. (2012) and Crawford and Schivardi (2016) but differs in the specification of loan demand by adding choice of maturity contract. Loan maturity is an integral part of a loan contract and borrowers often face this choice when taking a personal loan. This choice that borrowers face becomes important when other loan contract terms change with the loan maturity, which is the case on P2P lending platforms.

Let there be $i = 1, \ldots, I$ borrowers each of whom picks exactly one loan contract from a set of two contracts indexed by $j = 3\, or\, 5$. The specification of indirect utility for borrower $i$ who picks a $j - year$ loan contract is given by

$$U^*_{ij} = \alpha_{Pj} Price_{ij} + \alpha_{Fj} Fee_{ij} + X'_i \alpha_{Xj} + \varepsilon_{Uij}$$

Here $Price_{ij}$ and $Fee_{ij}$ denote the price (interest rate) and loan origination fee offered to borrower $i$ on contract $j$ which are the only two variables that vary across the maturity contracts. $X_i$ is a vector of borrower specific variables, including credit scores and demographic variables, and $\varepsilon_{Uij}$ is the error term observed by borrower but not by researcher.

Each borrower has a true loan demand representing a liquidity need which he aims to fulfill by taking a loan from the platform. The specification of this unobserved true loan demand need is given by

$$L^*_i = \beta_T Term_i + \beta_P Price_i + \beta_F Fee_i + X'_i \beta_X + \varepsilon_{Li}$$

[11] Since more than 90% (get the exact number) of the loan listings get funded, it is safe to assume a walrasian supply of credit coming from a large number of suppliers in a single market.

Where $Term_i$, $Price_i$ and $Fee_i$ represent the contract-specific variables of the loan contract the borrower ends up picking, and $\varepsilon_{Li}$ is the error term observed by borrower by not by researcher. Note here that the variable $Term_i$ is essentially the same as the binary variable indicating the choice of maturity contract.

Finally, conditional on having a loan of contract variables $Term_i$, $Price_i$, and $Fee_i$ each borrower has a demand to repay a fraction $S_i^*$ of loan principal. The specification of this unobserved demand to repay is given by

$$S_i^* = \gamma_T Term_i + \gamma_P Price_i + \gamma_F Fee_i + \gamma_L L_i + \gamma_H \mathbf{1}\{L_i = \overline{L}_i\} + X_i'\gamma_X + \varepsilon_{Si}$$

Where $L_i$ is the observed loan size, $\overline{L}_i$ is the loan limit assigned to borrower $i$ by the platform, and $\varepsilon_{Si}$ is the error term observed by borrower by not by researcher.

### 1.4.1 Estimation

In this section I explain the estimation strategy which boils down to a full Maximum Likelihood Estimation.

First I assume $(\varepsilon_U, \varepsilon_L, \varepsilon_S)$ are distributed jointly normal with the distribution given by $f(\varepsilon_U, \varepsilon_L, \varepsilon_S) =$

To derive the choice probabilities and the likelihood function, I first rewrite the joint density as the product of two conditional densities and one unconditional density:

$$f(\varepsilon_U, \varepsilon_L, \varepsilon_S) = f(\varepsilon_S \mid \varepsilon_L, \varepsilon_U)\, f(\varepsilon_L \mid \varepsilon_U)\, f(\varepsilon_U)$$

To simplify the notation I define the following matrices of $W_{Ui} = [\Delta Price_i, \Delta Fee_i, X_i]$,

$W_{Li} = [Term_i, Price_i, Fee_i, X_i]$, $W_{Si} = \left[Term_i, Price_i, Fee_i, L_i, \mathbf{1}\{L_i = \overline{L}_i\}, X_i\right]$ and the following sets of parameters $\alpha = \{\alpha_P, \alpha_F, \alpha_X\}$, $\beta = \{\beta_T, \beta_P, \beta_F, \beta_X\}$ and $\gamma = \{\gamma_T, \gamma_P, \gamma_F, \gamma_L, \gamma_H, \gamma_X\}$. Now I can derive the individual choice probabilities. First I consider the choice of loan contract. Define $Q_i$ as

$$Q_i = \begin{cases} 1, \; if \; U^*_{i5} - U^*_{i3} \geq 0 \\ 0, \qquad otherwise \end{cases}$$

The probability that the borrower picks the 5-year loan contract is given by

$$P_{Q_i=1} = \Phi\left(W'_{Ui}\alpha\right)$$

and the probability that a borrower picks the 3-year contract is $P_{Qi=0} = 1 - P_{Q_i=1}$. Here $\varepsilon_{Ui} = \varepsilon_{U3i} - \varepsilon_{U5i}$, $\alpha_X = \alpha_{X5} - \alpha_{X3}$, and for simplification I assume the coefficients on the alternative specific variables to be the same for each alternative i.e. $\alpha_{P5} = \alpha_{P3} = \alpha_P$ and $\alpha_{F5} = \alpha_{F3} = \alpha_F$. Note here that for alternative invariant covariates, only the difference in the coefficients $\alpha_X$ will be identified.

Next I consider the loan size choice. For this, the observable counterpart for $L^*_i$ is $L_i$ defined as follows

$$L_i = \begin{cases} L^*_i = W'_{Li}\beta + \varepsilon_{Li}, \;\; if \; L^*_i < \overline{L}_i \\ \overline{L}_i, \qquad\qquad\qquad otherwise \end{cases}$$

If $L^*_i < \overline{L}_i$, the true loan demand of the borrower, $L^*_i$, is observed since the borrower's loan limit constrained was not binding. The probability of observing such a case is given by

$$P_{L_i=L_i^*|\varepsilon_i^U} = Prob\left(L_i^* = W_{Li}'\beta + \varepsilon_{Li}\right)$$
$$= f_{\varepsilon_L|\varepsilon_U}\left(L_i - W_{L_i}'\beta\right)$$

On the other hand, if the loan limit constraint is binding for borrower $i$, i.e. $L_i^* \geq \overline{L}_i$, then the true loan demand of the borrower is not observed and thus the probability of observing a loan equal to the limit is given by

$$P_{L_i=\overline{L}_i|\varepsilon_i^U} = Prob\left(L_i^* \geq W_{Li}'\beta + \varepsilon_{Li}\right)$$
$$= F_{\varepsilon_L|\varepsilon_U}\left(W_{Li}'\beta - \overline{L}_i\right)$$

Calculation of the moments of the conditional distribution function $F_{\varepsilon_L|\varepsilon_U}$ is complicated because $\rho_{UL} \neq 0$ and $\varepsilon_{Ui}$ is not observed for any borrower. For this reason, I cannot directly calculate the moments of the conditional distribution function $F_{\varepsilon_L|\varepsilon_U}$ and instead must integrate over all $\varepsilon_{Ui}$ that result in the observed loan size. This yields the following expression for likelihood of loan size choice conditional on choosing a loan maturity contract

$$P_{L_i|Q_i=1} = \int_{-\infty}^{W_{Ui}'\alpha} P_{L_i|\varepsilon_i^U} \times f\left(\varepsilon_U\right) d\varepsilon_U$$

and

$$P_{L_i|Q_i=0} = \int_{W_{Ui}'\alpha}^{\infty} P_{L_i|\varepsilon_i^U} \times f\left(\varepsilon_U\right) d\varepsilon_U$$

Next, conditional on the loan size and loan contract choices, I derive the probability of observing loan repayment outcome censored by full payments or end of sample. For this I first define a censoring point $\bar{S}_i \in (0, 1]$ as the fraction of loan that needs

to be repaid by the end of my sample[12]. There are two possibilities for repayment: ***(i)*** default before full repayment or censoring point, ***(ii)*** repayment censored due to full repayment or the end of sample. The observed loan repayment choice is then given by

$$S_i = \begin{cases} S_i^* = W'_{Si}\gamma + \varepsilon_{Si},\ if\ S_i^* < \bar{S}_i \\ \bar{S}_i, \qquad\qquad\qquad\qquad if\ S_i^* \geq \bar{S}_i \end{cases}$$

The probability of observing repayment less than censoring point (analogous to default) is given by

$$P_{S_i = S_i^* | \varepsilon_{L_i}, \varepsilon_{U_i}} = f_{\varepsilon_S | \varepsilon_L, \varepsilon_U} \left( S_i - W'_{S_i}\gamma \right)$$

The probability of observing full or censored repayment is given by

$$P_{S_i = \bar{S}_i | \varepsilon_{L_i}, \varepsilon_{U_i}} = F_{\varepsilon_S | \varepsilon_L, \varepsilon_U} \left( W'_{S_i}\gamma - \bar{S}_i \right)$$

Here again, calculation of the moments of the conditional distribution function $F_{\varepsilon_S|\varepsilon_L,\varepsilon_U}$ is complicated since $\varepsilon_{Ui}$ is not observed. Another problem here is that $\varepsilon_{Li}$ is not observed for any borrower who picked a loan size exactly equal to the limit i.e. $L_i = L_i^*$. For these borrowers, I cannot directly calculate the moments of the conditional distrbution function $F_{\varepsilon_S|\varepsilon_L,\varepsilon_U}$ and instead must integrate over all $\varepsilon_{Li}$ that result in the observed loan size equal to the limit. There are two cases here: For borrowers who choose loan sizes less than their loan limits, I integrate over all possible $\varepsilon_{Ui}$ and for borrowers who choose loan sizes equal to their loan limits I integrate over all possible $\varepsilon_{Ui}$ and all possible $\varepsilon_{Li}$. The expressions for the likelihood of observed

[12]Note that $\bar{S}_i = 1$ for completed loans.

repayment conditional on borrowers choosing loans of sizes less than loan limits are given by:

$$P_{S_i|L_i=L_i^*,Q_i=1} = \int_{-\infty}^{W'_{Ui}\alpha} P_{S_i|\varepsilon_{Li},\varepsilon_{Ui}} \times f\left(\varepsilon_L, \varepsilon_U\right) d\varepsilon_U$$

$$P_{S_i|L_i=L_i^*,Q_i=0} = \int_{W'_{Ui}\alpha}^{\infty} P_{S_i|\varepsilon_{Li},\varepsilon_{Ui}} \times f\left(\varepsilon_L, \varepsilon_U\right) d\varepsilon_U$$

The expressions for the likelihood of observed repayment conditional on borrowers choosing loans of sizes equal to loan limits are given by:

$$P_{S_i|L_i=\overline{L}_i,Q_i=1} = \int_{\bar{L}_i-W'_{Li}\beta}^{\infty} \int_{-\infty}^{W'_{Ui}\alpha} P_{S_i|\varepsilon_{Li},\varepsilon_{Ui}} \times f\left(\varepsilon_L, \varepsilon_U\right) d\varepsilon_U d\varepsilon_L$$

$$P_{S_i|L_i=\overline{L}_i,Q_i=0} = \int_{\bar{L}_i-W'_{Li}\beta}^{\infty} \int_{W'_{Ui}\alpha}^{\infty} P_{S_i|\varepsilon_{Li},\varepsilon_{Ui}} \times f\left(\varepsilon_L, \varepsilon_U\right) d\varepsilon_U d\varepsilon_L$$

To summarize, I observce eight possible mutually exclusive cases observed in the data and I use Maximum Likelihood Estimation to estimate the parameters $\alpha$, $\beta$, $\gamma$, and $\Sigma$.

*Error Structure Discussion*

The correlation parameters $\rho_{US}$, and $\rho_{LS}$ have economic meaning. They characterize the relation between a borrower's unobserved reasons for picking a loan with a longer maturity and loan size, and his repayment behavior. If both these correlation parameters are zero, it means there is no new information in the choice of loan contract and choice of loan size about later repayment. However, if $\rho_{US} > 0$, one should

expect that, all else equal, borrowers who pick loans of longer maturity are likely to repay more and thus are better risks to take. Similarly, if $\rho_{LS} > 0$, one should expect that, all else equal, borrowers who pick loans of larger amounts are expected are likely to repay more.

The correlation parameter $\rho_{US}$ helps with identification – if it is zero, loan size can be considered independent of loan contract choice. Furthermore, the variance parameters, $\sigma_U$, $\sigma_L$, $\sigma_S$ capture the importance of unobserved characteristics relative observed characteristics in borrower decisions.

### 1.4.2 Identification Assumptions

I now highlight and discuss the sources of variation in the data that identify specific parameters of interest in the demand and repayment model. The parameters of interest from the demand model are the price coefficients in all three equations, $\alpha_P, \beta_P, \gamma_P$, loan maturity coefficients in equations 2 and 3, $\beta_T$ and $\gamma_T$, and the loan size coefficient in equation 3, $\gamma_L$.

For the price coefficients, I use variation in the platform's pricing schedule conditional on platform's finest measure of borrower riskiness, the expected loss rate, which can be interpreted as the expected loss on \$1 of investment to the borrower or simply the default probability. The platform sets prices based on loan term, whether a borrower has taken a prior loan from the same platform, expected loss rate and market and macroeconomic factors. The key identifying assumption here is that an individual borrower's loan demand and repayment choices do not depend on market and macroeconomic factors. Once a borrower is accepted by the platform, she expects to get a loan almost surely[13], her decision depends only on contract specific

[13]This is because over 90% of non-cancelled loan applications get funded.

variables (price, term, and fees) and her own observed and unobserved demographic and credit characteristics. Hence, when the platform changes prices for identical borrowers over time, as shown in Figure 1.2 for two representative types of borrowers who vary in their expected loss rates, this variation is exogenous to a borrower's decision of loan contract, size and repayment.

For $\beta_T$ and $\gamma_T$, note that in equation 1, the choice of loan contract depends on the difference in the contract specific variables, not the actual levels of those variables. It becomes clear then that conditional on making the contract choice, the loan size and loan repayment decisions depend on the levels of the chosen contract. Furthermore, I allow the unobservables $\varepsilon_S$ and $\varepsilon_L$ to be correlated with $\varepsilon_U$.

For $\gamma_L$, I highlight that loan limits are artificially imposed by the platform. This induced variation in the loan limits creates variation in loan amounts which helps to identify the coefficient of loan amount in equation 3. By allowing the unobservable in $\varepsilon_S$ to be correlated with $\varepsilon_L$, the identification of a change in repayment to loan size comes from variation in loan limits.

### 1.4.3 Demand Estimates

Table 1.2 provides the estimates of the demand model. The first column in the table provides the marginal effects of variables on the probability of picking the 5-year maturity contract over the 3-year maturity contract. This probability is sensitive to the difference in the interest rates on the two contracts. A one percentage point increase in the difference in interest rates reduces the probability of picking the longer term contract by 5.1 percentage points. Also note that borrowers with high credit scores are more likely to pick the longer term contract.

The second and third columns in Table 1.2 give estimates of the average effects

of variables on loan size choice and loan repayment choice. Loan size is much less sensitive to changes in interest rate than to loan origination fees and loan maturity term. A one percent increase in interest rate decreases loan size by $82. In contrast, a 1 percent increase in loan origination fees decreases loan size by about $2,300. Switching from a 3-year to a 5-year maturity contract increases loan size by about $2,700.

Loan repayment is more sensitive to a change in interest rate and loan maturity than to loan origination fee. A 1 percent increase in interest rate decreases the fraction of loan repaid by 1 percentage point and switching from a 3-year to a 5-year maturity contract decreases the fraction of loan repaid by 3.5 percentage points. Lastly, a $1000 increase in loan size decreases the fraction of loan repaid by 0.04 percentage point. A change in loan origination fee has no significant effect on loan repayment.

It is important to note here that these coefficients measure only the partial (direct) effects of a change in price on each of the three choices. Since borrower choices are interdependent, the full effect of a price change on loan amount and repayment choices would depend on the magnitude of the change in price and also credit and demographic characteristics of the borrower. Consider the loan amount choice: If the price on three year contracts increases by a small amount, a few marginal borrowers would switch to five-year contracts and their new loan term and new loan prices will affect their loan sizes. The borrowers who did not switch away from 3-year contracts will now decrease their loan sizes because now they face a higher price. However, if there is a large increase in the price of 3-year loan contracts, many more borrowers may switch to 5-year contracts and hence the full effect on loan size can be even bigger.

The full effect of a price change on loan repayment can be even more complicated since both loan maturity and loan amount would change with a price change and the new values of both these variables affect loan repayment. Hence, the full effects of price changes can be ambiguous until we can pin down the original change in price for each borrower. This will be explained more in the counterfactual section of this paper where I calculate the full effects of a given change in the price distribution on all three choices of borrowers.

## 1.5 Counterfactual

The main question I want to answer using this counterfactual experiment is how would borrower decisions change if they are offered prices which were determined by market forces of supply and demand within this online P2P credit market? To carry out this experiment, I first need the set of prices (interest rates) for 3-year and 5-year loans that would have been offered to each borrower under this counterfactual scenario. Although one could find out a comparable set of prices for each borrower in an existing offline credit market, such prices would include the costs to lenders which are specific to an offline market. Recall the online P2P market contains at least 1.5 million lenders competing to finance loans in a *single* market. As explained in the institutional background, the lender costs would be different in this online market from those in an offline market.

### 1.5.1 Exploiting the Change in Pricing Mechanism

Fortunately for this counterfactual experiment, Prosper.com used to operate an auction pricing mechanism to determine the price of each loan applicant who would post a listing on its website prior to Dec. 20th, 2010. At that time, the platform

allowed the lenders to collectively determine the price for each loan using a multi-agent auction. Figure 1.3 provides an illustration of the differences in prices for observationally similar borrowers under these two pricing mechanisms around the time when the change was implemented.

Next, I explain the details of the auction pricing mechanism. Each borrower $i$ posts a listing of amount $L_i$ and a reserve price $\bar{P}_i$, which is the maximum interest rate he would be willing to pay for that loan if it gets funded. Then each lender $j$ posts a bid of amount $a_{ij} < L_i$ as investment in the loan to borrower $i$ along with a minimum interest rate that is willing to accept $p_{ij} < \bar{P}_i$. If the desired loan amount of borrower $i$ is raised by the time the listing period of seven or ten days is over, he gets the loan i.e. if $\sum_j a_{ij} \geq L_i$, the loan gets funded and the contracted final price of the loan is determined by the price of the lender who is excluded from the auction. This is explained as follows:

Given an orderd bid profile of prices $\vec{\mathbf{p}_i} = (p_{i1}, \ldots, p_{iJ})$,

let

$$q = min\{r \mid \sum_{j=1}^{r} a_{ij} \geq L_i, r = 1, \ldots, J\}$$

Then the final contracted price for loan to borrower $i$ is given by $P_i = p_{i,q+1}$ and each lender's final investment $l_{ij}$ is given by

$$l_{ij} = \begin{cases} a_{ij,} & if \;\; j < q \\ L_i - \sum_{j=1}^{q-1} a_{ij}, & if \;\; j = q \\ 0, & if \;\; j > q \end{cases}$$

I exploit this unique pricing mechanism for a credit market to estimate the price a

borrower would have to pay when the market determines the price he is charged. To be more specific, I match borrowers under the two pricing mechanism based on a rich set of borrower and market characteristics to find out the prices a borrower under the posted-pricing mechanism would have paid under the auction-pricing mechanism.

On the other hand, under the new posted-pricing mechanism, the platform itself would set the price for each loan $P_i$ and in doing so the platform eliminated the auction pricing mechanism completely. This means that the prices were no longer determined by the market but instead were determined according to the platform's profit maximization condition. Note that now both borrowers and lenders were price takers and each lender only decides how much to invest in each loan by observing the price and riskiness of the loan.

### 1.5.2 Estimating the Pricing Function

Given this nice change in pricing mechanism, I match borrowers under the two mechanisms based on a rich set of credit variables for each borrower and macroeconomic variables at the time a borrower applied to get a loan. Owing to the inefficiency of simple matching procedures in high dimensions, I turn the problem the problem into a prediction problem: I first use machine learning techniques to predict the final auction-determined price, $P$, for each borrower under auction pricing mechanism. This yields a pricing function with a very low pseudo out of sample prediction error (root MSE of 2%). Then I use this estimated pricing function to predict the counterfactual prices for borrowers under the posted-price mechanism. I should note here that when approximating an unknown function from the data, *if the aim is simply to predict well on another sample generated from the same distribution as the original sample one must avoid overfitting and this is where machine learning can be*

*extremely useful.*

Here I explain the methodology of predicting the auction-determined price, $P$, for loans funded in the auction mechanism by using borrower characteristics and macroeconomic variables during that time. I will give a brief overview of sample splitting and random forests, which is a machine learning technique that gave the lowest pseudo-out-of-sample RMSE in this application.

*Sample Splitting:* Let there be let there by $i = 1, \ldots, N$ borrowers with data

$$\{(P_1, X_1), ..., (P_N, X_N)\} = (\mathbf{P}, \mathbf{X}) \sim \mathbf{D}$$

where $P_i$ is the auction-determined price for borrower $i$ and $X_i$ is a vector of $k$ borrower and market specific variables.[14]. The objective here is to estimate $P_i$ as a function of $X_i$ such that the estimated function can predict the prices for a new sample of borrowers *drawn from the the same distribution* $\mathbf{D}$.

To do this as efficiently as possible one must avoid overfitting and simply aim to reduce the out-of-sample prediction error. The problem here is that we can never truly get a precise estimate of this out-of-sample error because we do not observe the outcome variable for the new sample. However, we can use sample splitting to calculate pseudo out-of-sample error as illustrated below. Let

$$P_i = f(X_i)$$

To decide which functional form of $f$ to pick, I designate a *randomly selected* part of the sample as a training sample and the other part as a test sample. The training sample is given by $Z = (P_1, X_1), ..., (P_M, X_M)$, whereas the test sample is given by

[14]Here I have collapsed the state of the world index into the borrower index $i$ for simplicity.

$Z' = (P_{M+1}, X_{M+1}), ..., (P_N, X_N)$. In applied machine learning, a rule of thumb is to use a 2-to-1 split and it works quite well and the results are not very sensitive to small deviations around this rule. The point of sample splitting is to use the training sample to approximate the function, $f$, and use it to predict the outcome for the test set. Since we do have the actual outcome variable for the test set, we can calculate the prediction error as the mean squared error (MSE).

$$MSE_{out-of-sample} = \frac{1}{N-M} \sum_{i=M+1}^{N} \left(P_i - \hat{f}(X_i)\right)^2$$

*Random Forests:* Following the classic text of Hastie et al. (2009), I take the following steps to build the random forest:

(i) Draw a bootstrap sample $Z_b^*$ of size $M$ from the training sample $Z$

(ii) Grow a random-forest tree $T_b$ to the bootstrapped data as follows:

Select $r$ variables at random from the $m$ variables in $X$, where $m \leq k$. Then define a pair of half-planes as follows:

$$R_1(j, s) = X \mid X_j \leq s$$

and

$$R_2(j, s) = X \mid X_j > s$$

where $j$ is the index of the *splitting variable* [15] and $s$ is the point at which the split is made called *split point*. Starting with the base node at the top of the

[15] This is not to be confused with sample splitting.

tree, the rule for that node is formed by the following optimization problem:

$$\min_{j,s}\left[\min \sum_{x_i \in R_1(j,s)} (y_i - c_1)^2_{c_1} + \min_{c_2} \sum_{x_i \in R_2(j,s)} (y_i - c_2)^2\right]$$

The inner optimization problem is solved by setting $c$ equal to the mean of outcome variable for the observations in that partition. The key issue here is picking the right split point, $s$. Once the split point has been found, the same procedure is then performed on each resulting partition until the reduction in squared prediction error falls under a predefined threshold.

(iii) Repeating step 2 across $B$ trees constructed from $B$ bootstraps results in a forest of random trees $\{T_b\}_{b=1}^{B}$. Finally, the regression predictor for the true function is given by:

$$\hat{f}_{rf}^{B}(x) = \frac{1}{B}\sum_{b=1}^{B} T_b(x)$$

There are several benefits to calculating the counterfactual prices this way. First, it does not reflect any markups charged by lenders in a traditional offline credit market which suffers from its own frictions and their associated costs. The ideal price vector that is required here should be determined in the exact same peer-to-peer market where the only difference is that the lenders determine the prices instead of the platform.

Second, the auction pricing mechanism used by Prosper.com prior to Dec 20th, 2010 was successful in the process of price discovery. This highlights the point that the market was able to determine a fair price for each borrower. There is a good explanation for this. Arrow et. al, 2008 present a case for the promise of prediction markets by claiming that to predict the outcome of a future event, a

market can be created in which a commodity is trade whose value depends on the outcome of that future event. This will allow the market traders to aggregate all the available information and reflect it in prices. Keeping in line with this argument, Prosper.com essentially allowed the market to predict the future outcome of each borrower's repayment based on their credit variables. Moreover, Ilyer et. al, 2015 show evidence that under the auction mechanism on Prosper.com, the market the prices determined by the market were better predictors of default than even the finest credit score, even though the lenders could not observe the finest credit score, only a coarser measure of it.

Third, the contracted interest rate from the auction mechanism can be predicted, to a high degree of accuracy, from borrower characteristics using machine learning techniques. If this exercise is done carefully, as shown above, one can get out-of-sample error rate (root MSE) of 2%.

Fourth, it is more efficient to use this "inductive" approach in a data rich environment to approximate the price vector instead of taking a deductive approach of predicting such a price vector from a theoretical model. Given that the price offered to each borrower is determined by the choices of hundreds of lenders who observe borrower quality from about 400 variables, a comparable theoretical model must be able to either predict how each of those 400 variables affects the interest rate based on lenders' expectations of loan outcomes. It can certainly be simplified by a set of assumptions but that may make the theory incomplete.

### 1.5.3 Counterfactual Results

The results from the matching exercise are presented in Figure 1.4. It shows how the change in pricing mechanism affected the prices offered to borrowers based on

their riskiness. For each loan maturity, it shows how the difference in the platform offered price and market determined price changed with the riskiness of borrowers. It is evident from this figure that the risk premium charged by the platform to high risk borrowers was lower than the risk premium charged to similar borrowers by the market. On average, the prices are lower under the platform's posted pricing mechanism than in the auction mechanism of the market. This provides an explanation as to what is driving the results of higher demand and slightly higher repayment when the platform sets prices.

Next, I use the estimates of the demand and repayment model from section 4 and these counterfactual prices to predict the counterfactual choices of borrowers. The three choices that borrowers make are (*i*) choice of loan maturity contract, (*ii*) choice of loan amount, and (*iii*) choice of repayment. Upon getting these predictions, I compare them with the model's prediction given the actual data in which the prices where determined by the platform. Comparisons of these choices are summarized in Table 1.3 and Figures 1.4 to 1.6.

The average differences in the means of each variable under the two pricing mechanisms are shown in Table 1.3. The First thing to notice is that the market assigned prices are on average 3% lower than the auction determined prices for 3-year loans, and 4% lower for 5-year loans. Next, note that the average effect of the change in pricing mechanism on the probability of picking the 5-year loan contract decreased by 0.5 percentage point. Though the average effect is small, later I will discuss how the effect varies with borrower riskiness.

The full effect of the change in pricing mechanism on loan size can be decomposed into a direct and an indirect effect. The direct effect looks only at the partial effect of a change in price on the loan size. This effect ignores the fact that this same change

in price will also affect the loan term choice of borrowers. As seen in Table 2, the choice of loan term also has an effect on loan size. This effect of a price change on loan size through an effect on loan term choice is the indirect effect. This is especially important because depending on the signs and magnitudes of different coefficients, the direct and indirect effect may go in the same or opposite direction. In the case of equation 2 in my model, the two effects have the same sign so the full effect is bigger than either the direct or the indirect effect. Table 1.3 shows the average direct effect on loan size is about -$271 while the average full effect is -$1,137.

Similarly, the full effect of pricing mechanism on fraction of loan repaid is composed of a direct and an indirect effect. The direct effect of a change in prices on the fraction of loan repaid does not take into account the indirect channels of effect of the price change on loan term and loan size, whereas the full effect does take this into account. Table 1.3 reports that the average direct effect is a -0.014 which means that switching from platform prices to market prices decreased the fraction of loan repaid by about 1.4 percentage points. However, when you look at the full effect of -0.008, it is much smaller saying that the switch leads to a decrease in the fraction of loan repaid by only 0.8 percentage point. This is because the indirect effect of an increase in prices on loan repayment would go in the opposite direction. In Table 1.2 we can see that if there is a one unit change in interest rate, it would decrease loan amount by $82 which should ultimately increase loan repayment by 0.03 percentage point. So in this case the indirect effect partially dampens out the direct effect of change in prices on loan repayment. By extension the effect of a change in the pricing mechanism on loan repayment is also reduced. However, this effect changes only slightly with credit score i.e. it is bigger by 0.002 if borrower credit score increases by 1 point.

When you aggregate these effects at the market level you get the total effect of the change in pricing mechanism on the market's performance. The direct effect on credit demand was that credit demand fell by 2.3%, however, this was augmented by a bigger indirect effect leading to the full effect of 9.68% decrease in total demand. The repayment tells a slightly different story since the direct and indirect effects work in opposite directions. The direct effect of the pricing mechanism on fraction of loan repaid was a decrease by 1.71%, however, the full effect, which takes into account the changes in loan size and for a few borrowers a change in loan maturity, was a decrease by 0.8% only.

To delve deeper into the distribution of the effects presented in Table 1.3, I show how these effects change with borrower riskiness as shown in Figures 1.5 to 1.7. Figure 1.5 shows the average effect on the probability of picking the 5-year contract increased with credit score. Overall this effect was positive and small for all types of borrowers, but it was as low as 2 percent to 8 percent depending on your credit score.

Figure 1.6 and 1.7 highlight two important points: The first, Figure 1.6, is that the average partial effect on loan size is linearly decreasing in credit score and second, that it is much smaller than the average full effect across borrower types. The average full effect is in fact largest for borrowers with average credit scores while this effect is smaller for borrowers with lowest and highest credit scores. This figure also hints at what is driving the increased demand for credit under the platform's posted pricing mechanism. We can infer that the increase in total credit demand is coming from borrowers with close to average credit scores. The drastic differences in Figures 1.6 and 1.7 highlight the importance of taking into account the interdependencies in borrower choices, which make the full effects radically different from partial effects

not just in magnitude but also in heterogeneity across borrower riskiness.

Figure 1.8 and 1.9 tell a somewhat different story about the effects of the pricing mechanism on the fraction of loan repaid. Here the average partial effect, as shown in Figure 1.8, is bigger than the average full effect, as shown in Figure1.9, and both these effects are decreasing with credit score. While looking at the previous results of increased loan amounts and increased probabilities of switching to longer contracts would have raised concerns about lower repayment, we find that here the partial effect dominates such that the full effect remains positive. This is quite an achievement for a credit market: The platform's pricing mechanism is able to improve the repayment behavior of the risky borrowers. This is something traditional credit markets have historically struggled with as highlighted in the asymmetric information literature. While the emergence of credit scoring has definitely been helped alleviate this problem, there is definitely room for improvement. As shown here, the platform's pricing mechanism has helped alleviate the problem further.

## 1.6 Conclusion

In this paper I show how different components of loan contracts affect the choice of loan contract, loan demand and subsequent repayment choices. For that, I specify and estimate an econometric model of loan demand and repayment and exploit unique variation in the platform's pricing schedule to identify key parameters. I find that a change in loan maturity has a much bigger effect on loan size and repayment as compared to a change in loan prices. Furthermore, contract terms, including prices, affected all choices and due to interdependencies in these choices, the partial effects of a change in prices were much different from full effects.

Using the estimates of the model and exploiting a change in the pricing mechanism

implemented on the platform, I conducted a counterfactual experiment in which I predicted the loan demand and repayment choices of borrowers under the two pricing mechanisms. I found that when the lenders collectively determine the prices of loans, the prices were on average higher than the prices offered by the platform. This difference was bigger for observably high risk borrowers to whom the market charges a higher risk premium than the platform would. Additionally, under the auction mechanism, the borrowers picked are more likely to pick loans of shorter maturity, or smaller sizes and eventually repay smaller fractions of loans, as compared to when the platform sets prices. Aggregated at the market level, demand for credit and repayment of credit owed fall by 10% and 2% respectively under the auction mechanism.

These results have important implications for how credit markets can be made to price and allocate resources more efficiently. The above results show that when the platform sets the prices, it is able to increase the total demand for credit without decreasing the repayment of credit, but rather increasing the repayment slightly too. Moreover, the benefits of the borrowers with lower credit scores increase their demand more as compared to those with higher credit scores. By reducing the gap between the prices charged to high and low risk borrowers, the platform was able to increase the demand from high risk borrowers which eventually did not lead to more defaults, but rather slightly decreased the defaults.

**Figure 1.1:** The Platform Setup

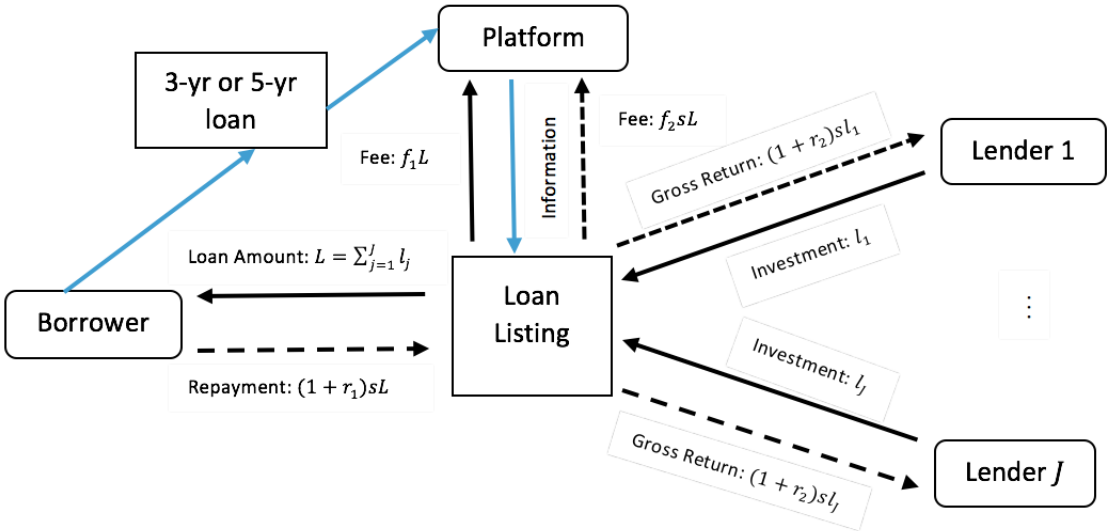


**Table 1.1:** Descriptive Statistics

| | | | Distribution | | |
|---|---|---|---|---|---|
| **Loan Variables** | Mean | SD | 10th | 50th | 90th |
| Loan Amount ($) | 11,989 | 7,147 | 4,000 | 10,000 | 21,749 |
| Interest Rate (%) | 16.03 | 5.52 | 9.20 | 15.35 | 24.24 |
| Frac. of Owed Amt. Repaid | 0.92 | 0.24 | 0.52 | 1 | 1 |
| 1{Loan Maturity = 5 yrs} | 0.36 | | | | |
| 1{Loan Limit Reached} | 0.08 | | | | |
| 1{Default} | 0.13 | | | | |
| **Credit Variables** | | | | | |
| External Credit Score | 708.53 | 54.19 | 645 | 713 | 800 |
| Internal Credit Score (0-11) | 6.10 | 2.48 | 3 | 6 | 10 |
| Estimated Loss Rate | 6.45 | 3.42 | 2.24 | 5.99 | 11.75 |
| No. of Credit Lines | 10.54 | 4.87 | 5 | 10 | 17 |
| Years of Employment | 9.27 | 8.46 | 0.92 | 6.92 | 21.33 |
| Stated Monthly Income ($) | 6,329 | 4,405 | 2,856 | 5,417 | 10,417 |
| External Monthly Debt ($) | 1,115 | 958 | 332 | 948 | 2074 |
| Delinquencies Last 7 Yrs | 3.85 | 9.65 | 0 | 0 | 14.00 |
| Inquiries Last 6 Months | 0.94 | 1.30 | 0 | 1 | 3 |
| Bankcard Utilization | 0.59 | 0.27 | 0.20 | 0.62 | 0.93 |
| 1{Home Owner} | 0.53 | | | | |
| No. of Observations | | | 20,000 | | |

Notes: This table presents summary statistics that were calculated using a random sample of 20,000 observations drawn from the selected sample of 74,168 observations. The external credit score refers to Experian Scorex PLUS. Loan maturity is a binary variable taking a value of 1 if loan maturity is 5 years and 0 if it's 3 years.

**Table 1.2:** Estimation Results

| **Dep. Var** | **1{Maturity=5 yrs}** | **Amount ($1000s)** | **Frac. Repaid** |
|---|---|---|---|
| | Marginal Effect | Coefficient | Coefficient |
| $\Delta$ Interest Rate | -0.051 | | |
| | (0.009) | | |
| $\Delta$ Orig. Fee | 0.229 | | |
| | (0.055) | | |
| Interest Rate (%) | | -0.082 | -0.010 |
| | | (0.031) | (0.002) |
| Orig. Fee (%) | | -2.293 | 0.000 |
| | | (0.444) | (0.000) |
| 1{Maturity=5 yrs} | | 2.780 | -0.035 |
| | | (0.117) | (0.002) |
| Amount ($1000s) | | | -0.004 |
| | | | (0.001) |
| 1{Limit Reached} | | | -0.002 |
| | | | (0.003) |
| No. of Observations | | 20,000 | |
| Controls | Credit Scores, Seasonal Fixed Effects, Demographic vars. | | |

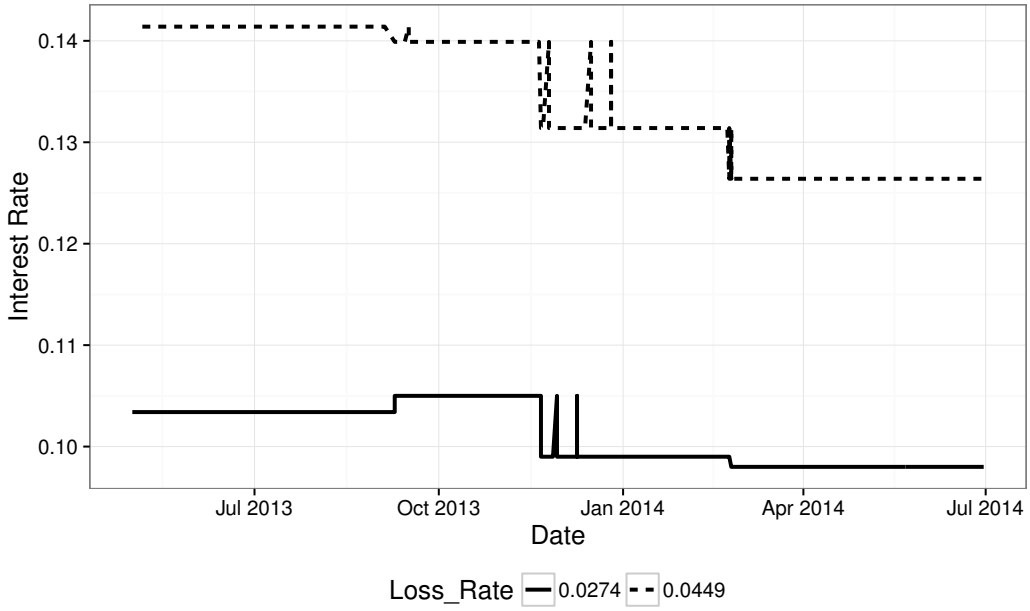
Notes: All estimates are based on the demand and repayment model presented in section 4. The sample used is a random sample of 20,000 observations drawn from a selected sample of 74,000 observations. This was done to ease the computational burden of the estimation procedure discussed in section 4. Estimates in the 2nd column show the marginal effects of a unit change in each of the explanatory variables on the probability of choosing the 5-year contract over the 3-year contract. For dummy variables, this is computed by taking the difference between the probability of contract choice when the variable is equal to 1 and the and the probability when the variable is equal to 0 (holding other variables fixed). For continuous variables, this is computed by taking a numerical derivative of the probability of contract choice with respect to the continuous variable. Estimates in the 3rd column show the effects of a unit change in each explanatory variable on loan size (in $1000s). The 4th column shows the effects of a unit change in each explanatory variable on the fraction of payments made. Standard errors were calculated from the numerical hessian evaluated at the estimated coefficients.

**Table 1.3:** Counterfactual Results

| **Variable** | **Platform Prices** | **Market Prices** | |
|---|---|---|---|
| | Mean | Mean | Mean Diff. |
| 3-Year Prices (%) | 15.99 | 19.01 | -3.02*** |
| 5-Year Prices (%) | 16.33 | 20.75 | -4.41*** |
| Pr. of Choosing 5-year Contract | 0.211 | 0.206 | 0.005*** |
| Loan Size Chosen (Partial) ($) | 11,744 | 11,473 | 271*** |
| Loan Size Chosen (Full) ($) | 11,744 | 10,606 | $1,137*** |
| Fraction of Loan Repaid (Partial) | 0.84 | 0.83 | 0.014*** |
| Fractionof Loan Repaid (Full) | 0.833 | 0.825 | 0.008*** |

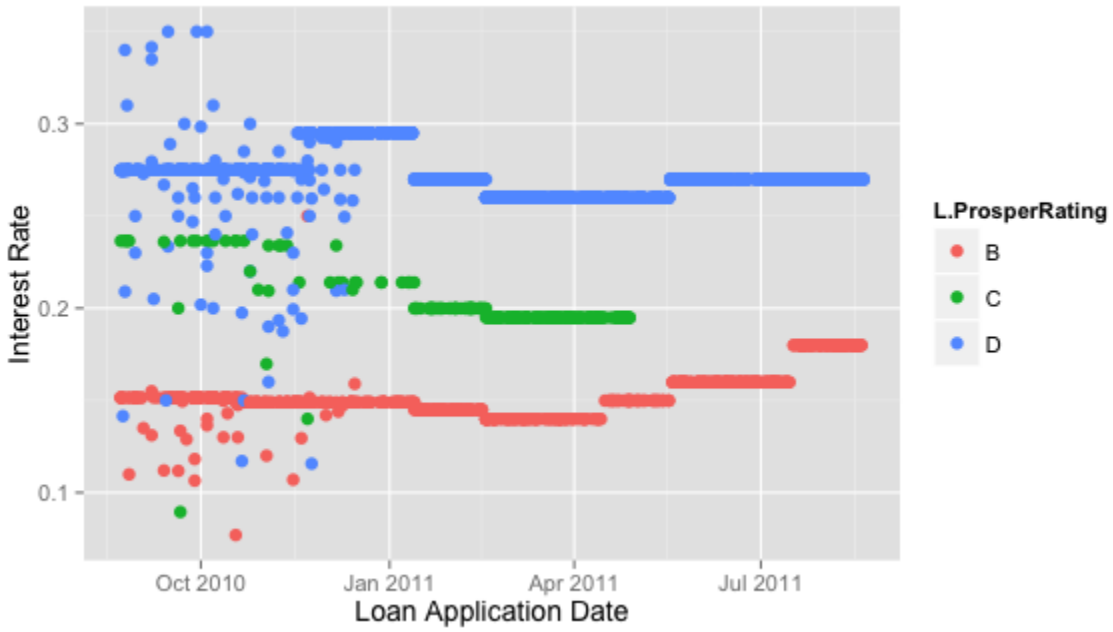
Notes: This table presents the results of the counterfactual experiment. The 2nd column shows the average of each variable when platform sets prices. Here the price averages are coming straight from the data (i.e. rows 1 and 2 of column 2). Rows 3 to 7 of column 2 show the average of the predicted quantities from model fit. The 3rd column shows the same averages when the market sets the prices under the auction mechanism. Here rows 1 and 2 of column 3 show the averages of the counterfactual prices predicted using the high dimensional matching exercise explained in section 5. Rows 3 to 7 of column 3 show the averages of the predicted quantities from the model given these new counterfactual prices determined by the market. The partial quantities in rows 4 and 6 hold fixed the other quantities that change when prices change. Full quantities in rows 5 and 7 do not hold fixed the other quantities that change when prices change. The 3rd column shows the mean difference in the quantities in each row. This can be interpreted as the average effect of a change in pricing mechanism. The significance for each effect was checked by calculating the standard errors of mean difference using paired t-tests. Significance level indicated as *** p ¡ 0.001

**Figure 1.2:** Variation in Interest Rates Over Time



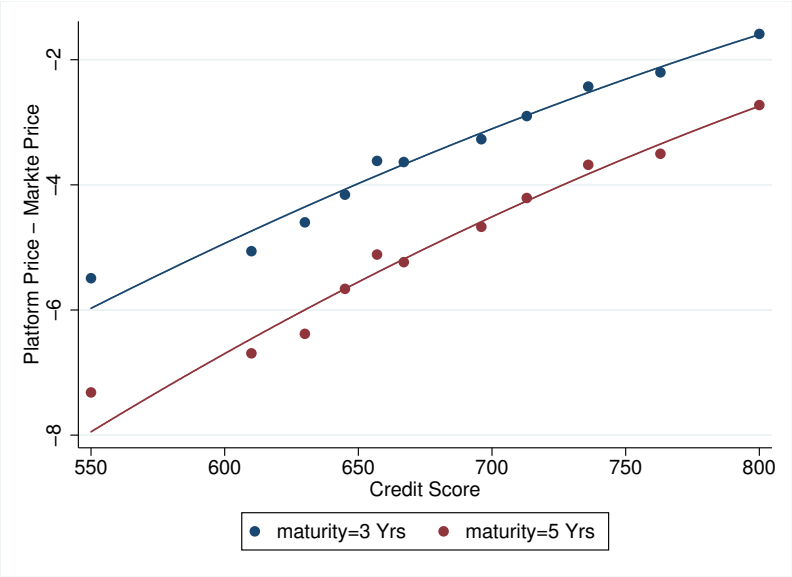

Notes: This figure highlights the variation in the prices charged by the platform for two sets of identical borrowers over time. The dotted line represents borrowers who are riskier than the ones represented by the solid line. The measure of riskiness is the estimated loss rate, a proprietary measure of Prosper.com. The other variables of loan contract, namely loan term and whether a borrower has taken a prior loan, are held fixed for this figure. The flat part of each line is evidence of the fact that at any given snapshot of time, all borrowers with the same estimated loss rate are considered identical and are assigned the same price. The variation over time in interest rates conditional on this risk measure is what I use in the model in section 4 to identify my coefficients of interest.

**Figure 1.3:** Change in Pricing Mechanism



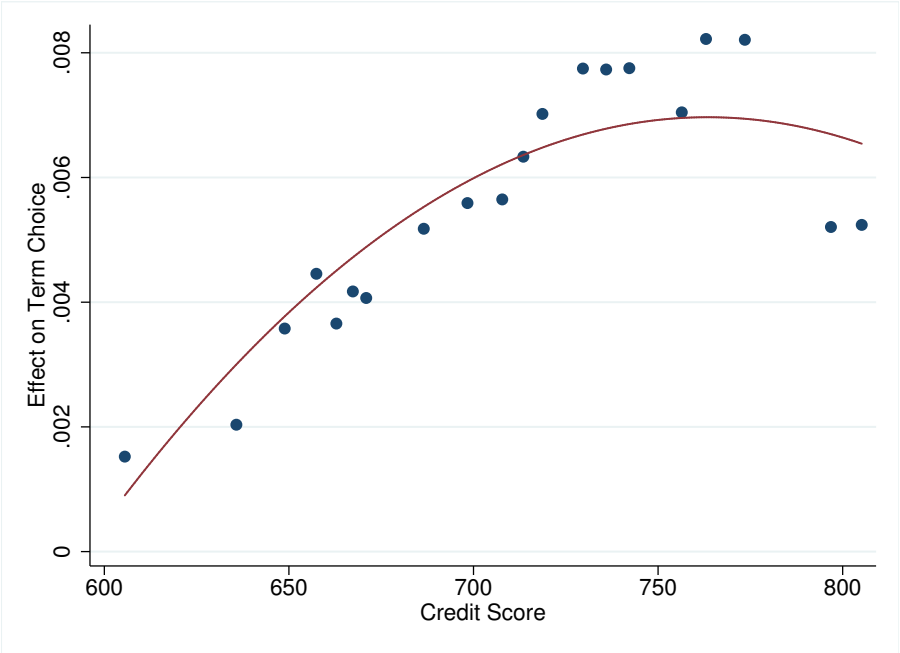

Notes: This figure shows how the prices (interest rates) changed for three narrowly defined credit categories just before and after the change in the pricing mechanism. The new posted-pricing mechanism was implemented on Dec. 20, 2010 and under this mechanism, the platform would set the prices itself. Prior to Dec. 20, 2010, the price for each loan was determined collectively by the lenders using a multi-agent decreasing price auction. Under the auction mechanism, there is huge variation in prices for one type of borrowers, however, the platform assigns the same price to all borrowers in the same credit category under the posted-price mechanism.

**Figure 1.4:** Distribution of Differences in Prices Charged by Platform




Notes: This figure shows how the difference the prices charged by platform and prices charged by market is distributed by borrowers' credit scores. The y-axis shows the platform prices in the actual data minus counterfactual market prices predicted for the same borrower using the the high dimensional matching exercise explained in section 5. In this binned scatter plot, each point represents the average difference in the price offered to borrower in one of the 11 credit score bins. The two graphs represent loans of 3 and 5 year maturities.

**Figure 1.5:** Distribution of the Effects of Change in Pricing Mechanism on Loan Maturity Choice by Credit Score




Notes: This figure shows how the effect of change in pricing mechanism on probability of choosing the 5 − year contract is distributed by borrowers' credit scores. On the y − axis you have the difference in choice probability given platform (model fit) and choice probability given market prices (counterfactual) predicted only by the credit scores. To construct this binned scatter plot, I first residualize the y and x-axis variable with respect to controls, which are year and month dummies (Note this is the first step of the partitioned regression). The I grouped the residualized x-variable into 20 equal-sized bins, computed the mean of the x-variable and y-variable residuals within each bin, and created a scatterplot of these 20 data points. For this I used the visualization method proposed by Chetty et al. (2013)

**Figure 1.6:** Distribution of Partial Effects of Change in Pricing Mechanism on Loan Size Choice by Credit Score

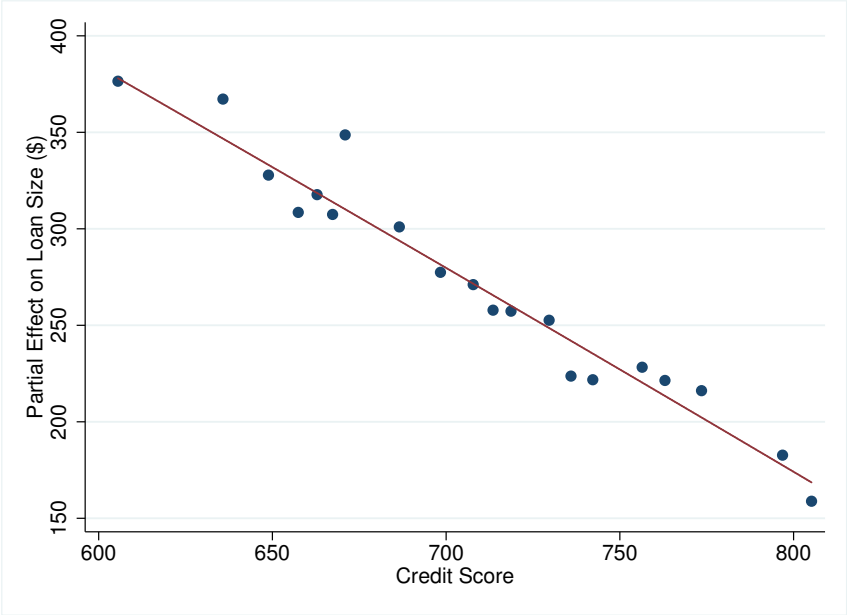
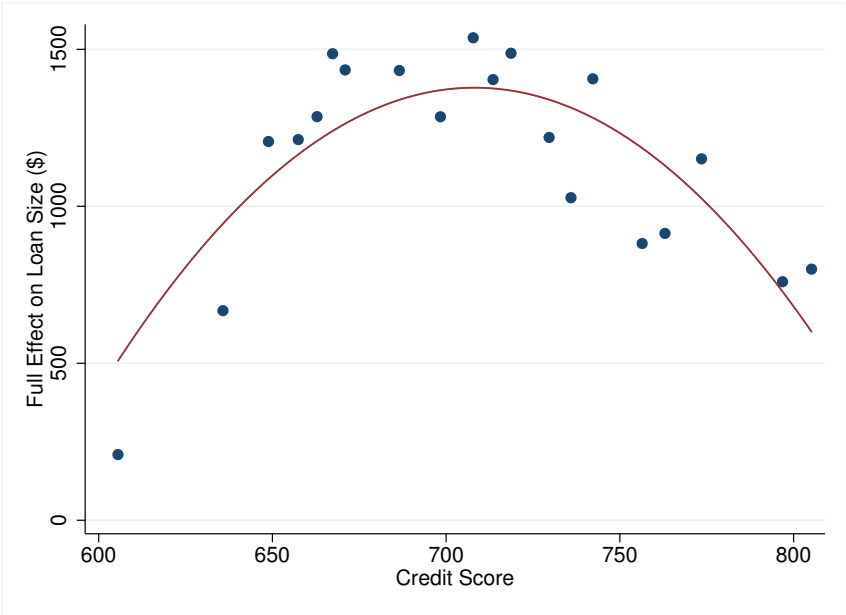


**Figure 1.7:** Distribution of Full Effects of Change in Pricing Mechanism on Loan Size Choice by Credit Score




This figure shows how the effect of change in pricing mechanism on loan size is distributed by borrowers' credit scores. On the y-axis you have the difference in loan size given platform (model fit) and loan size given market prices (counterfactual) predicted only by the credit scores. To construct this binned scatter plot, I first residualize the y and x-axis variables with respect to controls, which are year and month dummies (Note this is the first step of the partitioned regression). Then I grouped the residualized x-variable into 20 equal-sized bins, computed the mean of the x-variable and y-variable residuals within each bin, and created a scatterplot of these 20 data points. For this I used the visualization method proposed by Chetty et al. (2013). Panel (a) shows the distribution of partial effects, that holds constant the effect of price change on loan maturity, while panel (b) shows the distribution of full effects which take into account the effects of price changes on loan maturity.

**Figure 1.8:** Distribution of Partial Effects of Change in Pricing Mechanism on Repayment Choice by Credit Score

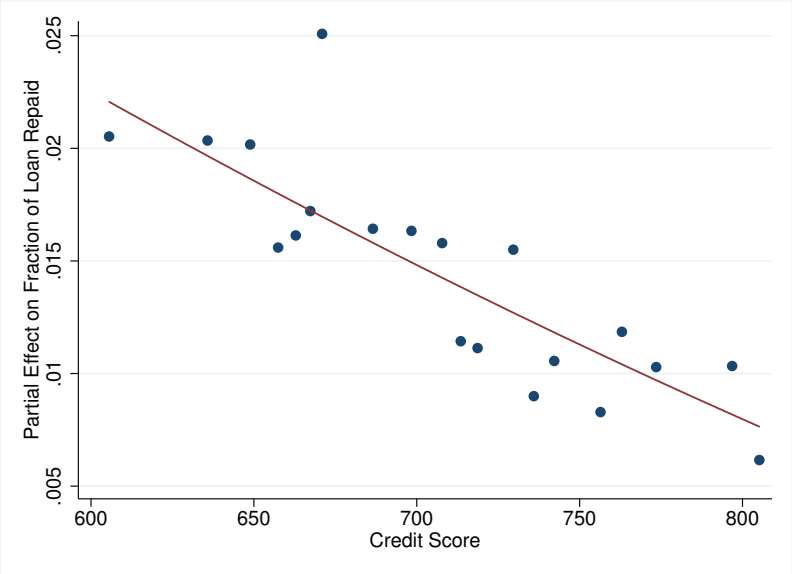
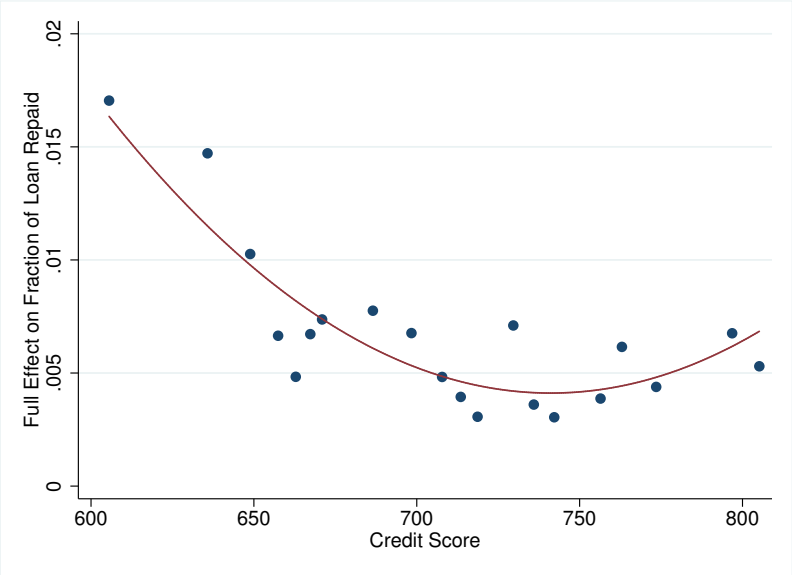


**Figure 1.9:** Distribution of Full Effects of Change in Pricing Mechanism on Repayment Choice by Credit Score




This figure shows how the effect of change in pricing mechanism on fraction of loan repaid is distributed by borrowers' credit scores. On the y − axis you have the difference in fraction repaid given platform (model fit) and fraction repaid given market prices (counterfactual) predicted only by the credit scores. To construct this binned scatter plot, I first residualize the y and x-axis variables with respect to controls, which are year and month dummies (Note this is the first step of the partitioned regression). Then I grouped the residualized x-variable into 20 equal-sized bins, computed the mean of the x-variable and y-variable residuals within each bin, and created a scatterplot of these 20 data points. For this I used the visualization method proposed by Chetty et al. (2013). Panel (a) shows the distribution of partial effects that hold constant the loan maturity and size, while panel (b) shows the distribution of full effects which take into account the effects of price changes on loan maturity and size.

# Chapter 2

# Impact of Finer Credit Scoring on Borrower Behavior

## 2.1 Introduction

Consumer credit scoring and risk-based pricing have been widely adopted to reduce the information asymmetry and its associated problems in credit markets. A credit score, typically calculated by an independent credit bureau, reflects the trustworthiness of an individual in a credit market, it as it allows lenders to quickly evaluate the risk of default (or likelihood of repayment) for a potential borrower and charge an appropriate price. This has significantly reduced screening costs for lenders over the past two decades when lenders relied on costly interview techniques and verification of borrower assets and income to assess his or her trustworthiness. Credit scores are also used by uninformed parties in other markets (labor, housing) as an effort to reduce information asymmetry. Even though credit scores are so widely used, it is very difficult to evaluate the effect of a change in credit score on the behavior of borrowers and lenders. This is because the credit scores are provided by a credit rating agency and lenders are free to interpret the scores, and changes in those scores, in any way they find appropriate. As explained in the paper this creates a fundamental problem in evaluating the effect of credit scoring.

In this paper I quantify the effects of implementation of finer credit-scoring on credit demand, defaults and repayment, in the context of a large Peer-to-Peer (P2P) online credit platform. I exploit a unique credit market setting which provides a major advantage to evaluate the impact of improvements of credit scoring on borrower choices using a policy change in which the platform introduced finer credit scores

relative to its previously coarse credit scores. The data set comes from Prosper.com, which is a large online P2P credit platform. There are two major advantages to using these data for evaluation of credit score improvements or implementation. First, this is a setting in which the platform itself determines the credit risk of all borrowers and determines the appropriate price of loan based on the credit risk and the lenders are price takers. In this unique setting a singular change in the credit score of a borrower results in a singular change in the price charged to that borrower. Second, the change in credit scoring function from coarse to fine provides some much-needed variation in credit scores and prices.

To elaborate on the importance of the first point further, consider that an improvement in credit scoring methodology is implemented in a traditional market for debt in which price setting is decentralized which means each lender sets his or her own price. In this decentralized case, a single change in the credit score (provided by the same credit rating agency) of two identical borrowers will result in multiple different changes in the prices charged by different lenders in the market to those two borrowers. The final change in price will not just depend on the change in the credit score but also depend on the choices made by the two borrowers and those choices will depend on a range of other factors. This makes it highly likely that the two borrowers end up paying two different prices even if they have the same credit score. Consequently, the changes in the two prices originating from a single change in credit score will also be different. In a nutshell, the one-to-one relationship between credit-score and price is broken in a traditional credit market with decentralized pricing but remains intact in the setting of the P2P credit market used in this paper.

In a setting where a single change in credit score results in multiple different prices charged creates a problem to evaluate the effect of a change in credit score on

borrower choices. However, in this paper, we have the centralized case, in which both the credit scoring and price determination is done by a single agent (the platform), and thus, a single change in the credit scores of two identical borrowers will result in a single change in the prices charged to those two borrowers.

The second advantage of the data and setting mentioned above provided enough detail in the data to evaluate the effect of the policy using a continuous treatment variable instead of considering the policy just as a binary treatment. Due to this, I was able to take into account the fact that different agents were impacted differently by the policy when evaluating its effect on borrower outcomes. To do this, I first measured the heterogeneity in the treatment levels for all agents and use that variation in the treatment to estimate its average effect. While earlier literature has mostly treated implementation of credit scoring as a binary policy, I went further to estimate the effect of a unit change in price due to better credit scoring on borrower choices. Essentially, I estimate the effect of the implementation of finer credit scoring in a more realistic way in which treatment is continuous rather than assuming that all borrowers were treated equally and consider the treatment to be binary. Although the study is focused on a single platform which has its own idiosyncrasies, it is more representative of the wider consumer credit market than some of the earlier studies which focused on individual institutional lenders adopting credit-scoring technologies (Einav et. al 2013). By focusing on an entire platform, which works like a marketplace, I was able to study how borrower behavior changes in of a competitive online market.

Theoretically, when information asymmetry in a market is reduced and the prices change to reflect the new set of information, it is not exactly clear how the previously informed set of agents would behave. Similarly, in a credit market, with the imple-

mentation of credit scoring it is not exactly clear how borrowers would respond to changes in prices. A typical borrower is expected reduce the loan amount requested if the interest rate increased for him, given everything else stays constant. However, a risky borrower, who has a higher probability of default, would be less sensitive to price in his choice of loan amount as compared to how sensitive a less risky borrower will be to a change in price. In the case of this paper, price is increasing for risky borrowers and decreasing for less risky borrowers. Hence, the magnitude of the average effect from a unit increase in price will depend on the distribution of different types of borrowers and the levels of price differences they face with the implementation of credit scoring.

The default and repayment choices of borrowers are also going to differ depending on the riskiness of the borrower because, by definition, the riskiness of a borrower reflects the expected ex-ante repayment outcome of the borrower. However, when the prices change in a way that a high-risk borrower's price is increased while a low risk borrower's price is decreased, the average effect could be ambiguous. While a decrease in price would decrease the likelihood of default, it also leads to an increase in requested loan amount which in turn may increase the likelihood of default.

To follow the earlier literature and highlight the importance of using continuous treatment, I also estimated the effects of the change in policy on borrower choices by considering the policy change as a binary treatment. For the binary treatment case, I find that the policy increased the average requested loan amount by about 8.7% ($684) which amounted to about $684 on average, decreased the likelihood of default by about 0.027 and increased the fraction of principal loan amount repaid by 0.02. The main findings of the paper are the ones when the treatment is considered continuous (as argued above). These findings show that a 1% increase in interest

rate due to finer credit scoring leads to a 0.01 increase in the fraction of borrowers who default and about 0.02 increase in the fraction of loan principal repaid.

For the continuous treatment case, since treatment levels can be positive, negative or 0, it makes sense to interpret the findings separately for the two groups of agents depending on the sign of treatment since positive and negative cancel each other out and the average treatment level is closer to 0. For the subset of borrowers who received positive treatment (price increases), the average treatment level was 1.1% which lead to an average decrease in loan amount by 0.29% ($24), an average increase in the fraction of borrowers who defaulted by 0.011 (65 borrowers), and an average increase in the fraction of loan repaid by 0.026 ($196). On the other hand, for the set of borrowers who received negative treatment (price decreases) the average treatment level was -1.5% which lead to an average increase in loan amount by $37, an average decrease in the fraction of borrowers who defaulted by 0.015 (123 borrowers), and an average decrease in the fraction of loan repaid by 0.035 ($290).

The effects on mean loan amount and mean loan repayment appear to be small but statistically significant. However, the effect on the number of people who default on loans is substantial. Comparing the results of the binary treatment with the ones with continuous treatment, one can immediately spot the differences in the effects and their signs. In a case where the two effects were similar, it would not matter much to use either type of treatment variable. However, given that the continuous treatment carries more information and gives a more comprehensive picture of the effects of the treatment, it is clearly the preferred approach. Furthermore, the findings of this paper contribute to the empirical evidence that improves our understanding of how reduction in information asymmetry affects the choices of the previously informed party in the context of a credit market. Although the study is focused on a single

platform which has its own idiosyncrasies, it is more representative of the wider consumer credit market than some of the earlier studies which relied on data from individual lending firms.

### 2.1.1 Related Literature

My paper contributes to two strands of literature on the development of credit markets. First is the impact of credit scoring, either implementation for the first time or improvements in it, on the actions of market participants and eventually the market structure. Earlier papers by Edelberg (2006), and Grodzicki (2012) study the impact of credit scoring adoption on development of credit markets. They find that with the adoption of credit scoring and risk-based pricing in traditional credit markets, the correlation between loan pricing and default risk has increased. However, these papers study these issues in the context of a whole economy and rely on aggregate and survey data to conduct their empirical analysis. In contrast, I use individual decision and transaction level data from an entire marketplace which permits a more detailed analysis. A more closely aligned paper to my paper is by Einav et al. (2013a) in which the authors use data from a large car financing company to show how the adoption of credit scoring for the first time affected the profitability of loans as lenders substituted credit scores for various types of local information. A recent paper by Cox (2017) studies the student loan market in which risk-based pricing by private firms coexists with flat prices for all loans provided by public entities. The paper shows how less risky borrowers move out of the government pool and opt for risk-based pricing which leads to an increase in consumer surplus.

A second strand of literature this paper contributes to is the research on P2P online credit markets and online markets in general. Rahim (2017) makes builds

a model of borrower demand and repayment to show how partial effects of price changes are different from full effects while taking choice of loan maturity into account. He further shows how the platform is able to increase the demand for credit and reduce defaults by setting prices itself which are, on average lower than those set by the market. Papers by Pope and Sydnor (2011) and Ravina (2013) show how an applicant's personal characteristics (for example outward appearance and skin color) can affect her probability of getting a loan. Iyer et al. (2016) provide evidence that the market is able to determine interest rates that predict defaults better than the finest credit scores do, and Zhang and Liu (2012) provide compelling evidence of investor herding behavior in P2P online credit markets.

The rest of the paper is structured as follows: In section 2.3, I give an overview of the institutional background of P2P online credit markets, explain the how the new policy was implemented and what it did exactly, and discuss descriptive statistics from the data. In section 2.4, I explain the two parts of the empirical strategy (with binary and continuous treatment) to answer the questions motivated above and show the results from each of the two parts of the empirical analysis. Section 2.5 provides concluding discussion and policy implications of the findings.

## 2.2 Empirical Strategy

## 2.3 Data and Institutional Background

### 2.3.1 Institutional Background

Over the past decade more than a thousand P2P online credit platforms have opened up across the world.[1] In the three biggest markets, China, United States and United

[1] Americas Alternative Finance Benchmarking Report, 2015

Kingdom, cumulative loan volumes by Dec. 2015 reached $70 billion, $25 billion, and $7 billion, respectively.[2] In 2014 in U.S. alone, the five biggest platforms issued $3.5 billion in loans compared to $1.2 billion in 2013. However, this makes up a sliver of consumer debt in U.S. To put things in perspective, total outstanding credit card debt in the United States grew to $880 billion by July 2014. According to a Fitch report (Fitch (2014)), the market volume in P2P online credit markets may grow to $114 billion in the medium term.[3] The U.S. market is dominated by two competing platforms named Lending Club and Prosper which together have a market share of over 90% in P2P small personal loans.

In a typical online P2P credit market borrowers seek loans from a group of lenders by posting their credit information on the platform website. The platform performs initial screening of borrowers, collects credit information, and sets loan contract terms including loan maturity, interest rate, and transaction fees. Individual and institutional investors decide how much to invest in each loan based on their own preferences. In this market, in its current form, the set of prices is controlled by the platform while both borrowers and lenders are price takers and pick their own allocations. Rahim (2017) provides an extensive discussion about how these platforms operate and how they differ from offline credit markets with a particular focus on the advantages and disadvantages for borrowers and lenders to participate in such a market.

### 2.3.2 Implementation of finer credit scores

The platform I study operates as a marketplace for small personal loans to prime borrowers within the United States. The loans do not require a collateral and they

[2]Citi Group Report, 2016

[3]Federal government data aggregated by www.nerdwallet.com

are not monitored by the platform or the lenders. As discussed earlier, the platform used risk-based pricing for all loans. It used data from a credit bureau as an input to the its proprietary credit scoring algorithm to determine a borrower's default risk. Different borrowers were assigned to different risk categories and at any given time, all borrowers with the same risk score were charged the same price. The policy in question made these risk categories smaller.

On July 20, 2012, the platform implemented finer credit scoring in which it increased the number of risk categories from 8 to 49. Figures 2.1 and 2.2 depict this change in the platform's risk-to-price function before and after the policy. Each point in these figures represents a price-risk pair with price on the vertical axis and risk on the horizontal axis. Each level of risk represents a separate risk category and is assigned a separate price. In the pre-policy period (Figure 2.1), the risk categories were fewer and hence coarser than they became the time period after the policy was changed (Figure 2.2). With this policy, the borrowers in each older risk category were further classified into several risk categories.

A crucial thing to note in figures 2.1 and 2.2 is that price is a linear function of the risk level and this function remained unchanged after the under the new credit scoring policy. As explained later in the empirical strategy section, this fact helps with the identification strategy in the estimation procedure. Further evidence of the implementation of this policy can be found in the tabulated pricing functions of the platform reported in Figures 2.4 and 2.5 in the appendix. Figure 2.4 shows that prior to the implementation of the new policy for each narrowly defined credit category, the platform assigned a single price for all borrowers in that category. However, Figure 2.5 shows that after the new policy was implemented, the platform would assign a range of prices (denoted by minimum and maximum) for each narrowly

defined category. With some data mining, I was able to figure out that within each credit category, the platform assigns a single price to all borrowers with the same risk score.

### 2.3.3 Data and Summary Statistics

The data for this paper come from Prosper.com which is the second largest peer-to-peer internet credit platform in the U.S. by loan volume. These data contain all required loan specific and borrower specific variables. For each loan, I observe the amount of loan, maturity period, interest rate, loan amount repaid (till the end of sample) and time stamps for loan application, issuance and repayment. For each borrower I observe a rich set of credit variables from the Experian credit bureau, Prosper.com's own credit score, credit grade and demographic variables. Identifiers for each loan application, loan and borrower allows for seamless merging of different parts of the database. Owing to the online nature of the platform, it can implement big changes to the workings of the market very quickly and at scale. To address this issue, I used 54 snapshots of Prosper.com from internet archives to look for changes in borrowing and lending processes over time. These proved to be quiet useful in isolating a time period no major changes took place. Furthermore, I used some macroeconomic variables taken from the website Federal Reserve Economic Data (FRED) maintained by the Federal Reserve Bank of St. Louis.

In Table 2.1, I report a comparison of the summary statistics of loan contract and credit variables as well as those of loan outcomes for a period of four months before the policy was implemented and four months after it. Since the policy was implemented on July 20, 2012, I select this sample of eight months to conduct my empirical analysis similar to a window study analysis. Table 2.1 shows significant

changes in basic choices of borrowers: the average requested loan amount is about 700 dollars larger after the policy, the average probability of default decreased by about 2 percent and the average fraction of loan repaid increased by about 1.7 percent after the policy. Furthermore, the average external credit score is only about 1 percent higher for borrowers after the policy and the average internal loss rate, which is analogous to the proprietary risk score of the platform, is only about 0.1 percent smaller for borrowers after the policy. Note that both the internal and external credit scores represent borrower riskiness in a way that the higher the score the less risky the borrower is. Table 2.1 also shows that the average difference in the interest rates is very low at about -0.3 percent and the average difference in loan maturity is less than a month. All these differences, except the home-owner status, are statistically significant at the 95 percent confidence level.

## 2.4 Empirical Strategy

### 2.4.1 ATE with Binary Treatment

I use propensity score matching as a first step in my empirical strategy to estimate the effects of implementing the policy discussed above. For this, I first create an eight-month time window in a way that the date for policy change falls in the middle of this time window. I select all the loan applications made between the dates of March 20, 2012 and November 20, 2012. The treatment is defined as the implementation of the finer credit scoring function. Since this function was implemented on July 20, 2012, all the observations between March 20, 2012 and July 19, 2012 make up the control group, while all the observations between July 20, 2012 and November 20, 2012 make up the treatment group.

To implement the propensity score matching method to estimate the Average Treatment Effect (ATE), I first estimate the probability for each observation to belong to the treatment group given the set of covariates on which the loan applications need to be matched. This is done by logit regression as follows

$$p(x_i) = \Lambda(D_i = 1 \mid X_i = x_i)$$

where $D_i$ is an indicator equal to 1 if borrower $i$ belongs to the treatment group and it equals 0 if borrower $i$ belongs to the control group. $X$ is a matrix of the matching covariates and $p()$ is the propensity score function that needs to be estimates. The ATE is then estimated by using a matching algorithm in which the matching is done on the propensity score.

#### 2.4.1.1 Results from Propensity Score Matching

Tables 2.2 presents the results from Propensity Score Matching estimation. The effect of finer credit scoring on requested loan amount is positive and statistically significant as shown in column 1. I find that requested loan amount increased by 8.7 percent due to the change in policy, given everything else as equal. Given the average requested loan amount was 7,862 dollars, the policy increased the average requested loan amount by 684 dollars.The set of matching variables used included the loan contract variables, interest rate and contract length, as well as a large set of variables from the borrower credit profile. The matching procedure was successful in creating a balanced sample and the bias from each matching covariate was less that 5 percent in the balanced sample. Details about the efficiency of the matching process are given in Tables A.1 to A.4 of the appendix.

To estimate the effect on defaults and percent of loan principal repaid, the sample

was reduced to only those borrowers who were actually issued loans because these borrowers were able to raise the required amount by the lenders to convert the loan applications into loans. In column 2, I report the effect of the policy on the likelihood of default is negative and statistically significant. I find that the likelihood of default decreased by 2.7 percent due to the implementation of new policy. Another measure of loan repayment is the fraction of loan principal amount repaid. In column 3, I report the estimate of the effect of finer credit scoring on fraction of loan principal repaid and the effect is positive and statistically significant. The average borrower repaid 2.4 percent more due to the implementation of the new policy and this finding further supports the earlier finding of a negative effect of the policy on the likelihood of default.

As a robustness check that I used for was to use an alternative estimation technique. For this I used ordinary least squares (OLS) to regress each outcome variable above on the treatment and all matching variables. The results from these estimations are shown in Table 2.3. Although the estimates from OLS and Propensity Score Matching are not directly comparable since the two techniques are, by definition, different and they estimate slightly different functions, the magnitudes of the effects are close enough and their signs are also the same. Hence, these results further support the results from the main specification of propensity score matching.

### 2.4.2 ATE with Continuous Treatment

In the previous section, I assumed that all borrowers were treated equally by the policy which essentially means that all agents received the same level of treatment. The purpose of the previous section was to relate this paper to the existing literature in which such a treatment was considered binary, and also establish a baseline case

for the findings when the assumption of binary treatment is relaxed. In this section I take into account the fact that different agents were impacted differently by the policy when evaluating its effect on borrower outcomes. To do this, I first measure the heterogeneity in the treatment levels for all agents and use that variation in the treatment to estimate its average effect. Essentially, I estimate the effect of the implementation of finer credit scoring in a more realistic way in which treatment is continuous.

Due to the variation in the treatment levels in terms of changes in prices, one should expect variation in the responses to treatment levels. The nature of the policy change was such that some borrowers were better off while others were worse off with the new policy. This means that some borrowers faced price increases, making them worse off, while others faced price decreases, making them better off. It is important to note here that this situation is representative of a setting in which the implementation or improvement of credit scoring helped to distinguish agents based on the credit-worthiness. Despite the heterogeneity in treatment levels, earlier literature mostly considered this a binary treatment. A key contribution in this paper is that I consider this to be a continuous treatment with the intention of estimating more realistic and precise average treatment effects.

To estimate the average effect of the difference in price on borrower outcomes of interest, I first aggregated the data at the level of risk category defined by the level of Estimated Loss Rate (ELR) under finer (new) credit scoring policy. Define $\Delta Y_j$ as the mean difference in the outcome variable within category $j$, before and after the implementation of policy. Within each category, the main outcomes of interest were the number of borrowers who applied for loans, the mean loan amount, the fraction of borrowers who defaulted, and the mean fraction of loan repaid. Similarly, define

$\Delta P_j$ as the mean difference in the prices charged to all identical borrowers within category $j$. Once they are constructed (as explained below), the average treatment effect of the change in prices can be estimated with OLS where each observation is weighted by the number of borrowers of category j in the market. To be precise, I estimate the following regression using weighted OLS for different outcome variables of interest:

$$\Delta Y_j = \beta_0 + \beta_1 \times \Delta P_j + \varepsilon_j$$

Where $\Delta P_j = P_j^{new} - P_j^{old}$ and $P_j^t$ denotes the price charged to all borrowers in category $j$ in time period $t$ for $t \in \{old, new\}$ denoting pre-policy and post policy time periods respectively. Even though the price is the same for every borrower in the same category and same time period, this price may vary over time. In cases where this price varied over time, $P_j^t$ was constructed by taking the weighted average of the prices over time within the given category $j$ for the given time period $t$. However, to actually get to the final measure of $\Delta P_j$ and subsequent measures of $\Delta Y_j$, I first classified all borrowers under the old credit scoring policy according to the new credit scoring policy and the details of it can be found in the next section.

As noted above, $\Delta Y_j$ is defined as the mean difference in the outcome variable within category $j$, before and after the implementation of policy. For each outcome variable, $\Delta Y_j$ is defined as follows for each of the three estimations:

**Difference in the mean loan amount requested within category $j$:**

$$\Delta Y_j = \frac{1}{n_j^{new}} \sum_{i=1}^{n_j^{new}} \left(log\left(Loan\ Amount_{ij}^{new}\right)\right) - \frac{1}{n_j^{old}} \sum_{i=1}^{n_j^{old}} \left(log\left(Loan\ Amount_{ij}^{old}\right)\right)$$

Where $Loan\ Amount_{ij}$ is the requested loan amount by borrower $i$ in credit category $j$, and $n_j^t$ denotes the number of borrowers in category $j$ in time period $t$. Time period *new* is defined as the period after the new policy was implemented and time period *old* is defined as the period before the new policy was implemented.

**Difference in the fraction of borrowers who defaulted within category $j$ :**

$$\Delta Y_j = \frac{1}{n_j^{new}} \sum_{i=1}^{n_j^{new}} \left(Default_{ij}^{new}\right) - \frac{1}{n_j^{old}} \sum_{i=1}^{n_j^{old}} \left(Default_{ij}^{old}\right)$$

Where $Default_{ij}^t$ is equal to 1 if borrower $i$ in category $j$ repaid less for a loan that was issued in period $t$ and $Default_{ij}^t$ is equal to 0 if that borrower repaid in full.

**Difference in mean fraction of loan repaid within category j:**

$$\Delta Y_j = \frac{1}{n_j^{new}} \sum_{i=1}^{n_j^{new}} \left(Repayment_{ij}^{new}\right) - \frac{1}{n_j^{old}} \sum_{i=1}^{n_j^{old}} \left(Repayment_{ij}^{old}\right)$$

Where $Repayment_{ij}^t \in [0, 1]$ is the exact fraction of loan principal repaid by borrower $i$ in category $j$ for a loan that was issued in time period $t$. Finally, to account for the variation in the number of borrowers within each risk category, I weighted each observation by total the number of borrowers in category $j$ as a fraction of the total number of borrowers in the market. This weight is defined below:

$$Weight_j = \frac{n_j^{new} + n_j^{old}}{\sum_{j=1}^{J} \left(n_j^{new} + n_j^{old}\right)}$$

Where $J$ is the total number of credit categories.

**Matching Borrowers Under Two Credit Scoring Regimes**

As mentioned earlier, to construct the variables defined above and to measure the different treatment levels of all borrowers, I need to classify these borrowers into different credit score categories according which as defined by the finer (newer) credit scoring function. For the borrowers who were issued loans under this new regime, this is simply the credit category they were assigned by the platform so it is already defined in the data. However, for borrowers under the old regime, I conducted a matching exercise. To explain this, I turn to the details of how the platform's its internal risk scores changed and how the platform used these scores to determine prices.

Let there be two types of borrowers: a low risk type denoted by $L$ and a high risk type denoted by $H$. Before the policy, both these sets of borrowers were considered identical in terms of their repayment probabilities and so their estimated loss rate was the same, denoted by $E\bar{L}R$. Based on this, these borrowers also faced the same price denoted by $\bar{P}$. After the policy was implemented, the platform could distinguish between $L - type$ and $H - type$ borrowers and hence assigned $ELR^h$ to high risk type and $ELR^l$ to the low risk type. The corresponding prices for these types were $P^h$ and $P^l$. Here $ELR^h > E\bar{L}R > ELR^l$ and $P^h > \bar{P} > P^l$.

With the impact of policy, the $H - type$ borrowers face a higher price after the policy and this difference is calculated as $P^h – \bar{P} > 0$. Similarly, the $L – type$ borrowers face a lower price after the policy and this difference is calculated as $P^l – \bar{P} < 0$. Once the change in price for each type of borrowers is calculated, it can be used as the treatment level for that type and estimate the average treatment effect of the policy more realistically.

The key challenge in calculating $\Delta P_i$ is that $ELR_{new}$ is not observable for bor-

rowers who applied for loans before the policy was implemented. To address this challenge, I use machine learning to estimate a function that predicts $ELR_{new}$ from borrower credit variables and macroeconomic variables at the time of loan application. It is essential to note here that the platform itself uses these same variables to assign $ELR_{new}$ to each borrower and since I observe all these, I estimate this function directly from the data. This can be done using OLS too, but since it is a pure prediction problem, I expanded the set of techniques to include some from the machine learning literature and evaluated the performance of each technique using pseudo-out-of-sample Root Mean Squared Error (RMSE). The random forest algorithm gave the lowest RMSE of 0.0054 and also gave out-of-sample R-Squared of 0.98 which is why I picked it as the final estimation technique for predicting $ELR_{new}$. Alternative techniques, like Lasso and Ridge gave slightly higher RMSE of 0.0066 and 0.0076, respectively.

With this estimated function I predicted the $ELR_{new}$ for borrowers in the pre-policy period and this gave me a single measure of borrower type according to the platform's new policy. Furthermore, I used this $ELR_{new}$ to find the closest match for each borrower from the pre-policy period to a set of borrowers in the post-policy period to assign $P_{new}$ to each borrower in the pre-policy period. Similarly, I used this matching variable to assign $\bar{P}$ to each borrower in the post-policy period based on the closest match for that borrower from the pre-policy period. As defined above, I calculated the $\Delta P_i$ for each borrower given the complete sets of prices.

Figure 2.3 shows the empirical distribution of $\Delta P_i$ and it can be seen that for a large number of borrowers, the difference in price was close to zero while there was non-trivial mass on either side of zero. This Figure highlights my point earlier that not all borrowers were equally treated by the policy. Instead, some borrowers were

better off from this policy since they received a price decreased while some others received a price increase. Intuitively, the borrowers who received a price decrease would be the ones who were of lower risk type but the platform assigned grouped them with higher risk types and thus charged a higher price. Similarly, the borrowers who faced an increase in the price were the ones who were benefitting from a lower price before the policy because the platform grouped them with lower risk type borrowers. This highlights an important feature of the variation generated in prices due to this policy: the direction and magnitude of the price difference is not random but is determined by the difference in the platform's estimates of the two risk scores (loss rates) for each borrower.

Given this variation in prices it is not exactly clear how borrowers would respond to changes in prices. A typical borrower should reduce the loan amount requested if the interest rate increase, given everything else stays constant. However, a risky borrower, who has a higher probability of default, may also be less sensitive to price in his choice of loan amount as compared to a less risky borrower. In the case of this paper, price is increasing for risky borrowers and decreasing for less risky borrowers. In this case the magnitude of the average effect from a unit increase in price will depend on the distribution of different types of borrowers and the levels of price differences they face with the implementation of this policy. Furthermore, the default and repayment choices of borrowers are also going to differ depending on the riskiness of the borrower because the by definition, the riskiness of a borrower reflects the expected ex-ante repayment outcome of the borrower. However, when the prices are change in a way that a high risk borrower's price is increased while a low risk borrower's price is decreased, the average effect could be ambiguous. While a decrease in interest decrease the likelihood of default, it also leads to an increase

in requested loan amount which in turn may increase the likelihood of default.

### 2.4.3 Estimation Results

As discussed earlier, the main coefficient of interest is the coefficient on $\Delta P_i$ which the average treatment effect using a *continuous measure of treatment*. Earlier discussion implies that the sign of this coefficient, in the cases for all three outcome variables, could be positive or negative depending on the composition of borrower applicant pool and their sensitivity to price changes. In Table 2.4, I report the results for the average treatment effect on mean loan amount. Column 1 reports the results with just one covariate which is $\Delta P_i$, column 2 reports the results with an indicator for when $\Delta P_i$ is greater than 0 and column 3 reports the results when both these covariates are included in the regression. The purpose of including the indicator for when $\Delta P_i$ is greater than 0 is to consider relate these results to the binary treatment case of the previous section. Column 1 shows that a 1% increase in interest rate lead to a 0.29% decrease in Mean Loan Amount and this coefficient is statistically significant[4]. Given the average requested loan amount before the new policy was introduced was \$7,862, a 1% increase in interest rate lead to an average decrease of \$24[5]. Column 2 shows that the binary treatment indicator is statistically insignificant from 0 and column 3 shows that the inclusion of this binary indicator together with the $\Delta P_i$ does not make either of the two coefficients much different from their values and significance levels from columns 1 and 2.

To estimate the average treatment effect on the fraction of borrowers who defaulted and on the mean fraction of loan repaid, the sample is reduced to include

[4] $exp(25.796/100)–1 = 0.294$

[5] Note a 1 unit increase in price difference is like the interest rate (price) on a loan increased from 7% to 107%. Therefore, I interpret the coefficients with a 1% increase in interest rate which is equivalent of going from 7% to 8%.

only those borrowers who were actually issued loans because they were able to raise the required amount by the lenders. Table 2.5 shows the treatment effects on the fraction of borrowers who defaulted. Column 1 shows that a 1% increase in interest rate lead to an increase of 0.01 in the fraction of borrowers within a bin who defaulted. Lastly, I report the results for the effects on mean fraction of loan repaid in Table 2.6. Column 1 shows that a 1% increase in interest rate lead an increase of 0.02 in the mean fraction of loan principal repaid and this result is statistically significant. Given the average fraction repaid is about 0.83, a 1% increase in interest rate lead that fraction to increase to 0.85. Note that this is well below full repayment rate of 1 so we still see defaults. Note that this variable is a fraction of the loan amount. On the other hand, the fraction of borrowers who defaulted is fraction of the number of people in their credit category.

Since treatment levels can be positive, negative or 0, it makes sense to interpret the findings separately for the two groups of agents depending on the sign of treatment since positive and negative cancel each other out and the average treatment level is closer to 0.

Table 2.7 shows the summary statistics of the treatment level and the outcome variables for two subsets of the sample based on the sign of the treatment level i.e. those borrowers who received price increases ($\Delta P_i > 0$) and those who received price decreases ($\Delta P_i < 0$). Note that about 41.9% of borrowers received a price increases with an average increase of 1.1% while 58.1% of borrowers received price decreases with an average decrease of 1.5%. Hence, the distribution of treatment levels is skewed to the left and this can also be seen in figure 2.3.

For the subset of borrowers who faced price increases, the average increase in price was 1.1% and given an average loan amount of $7,536, it led to an average

decrease in the loan amount of approximately $24. Furthermore, a treatment level of 1.1% lead to an average increase in default fraction by 0.011 i.e. from 0.137 to 0.148. Given the number of borrowers in this subset was about 5,944, this increase of 1.1% in price lead to an increase in the number of borrowers who defaulted by 65. In the proper context, this seems to be quiet significant. Lastly, this 1.1% treatment level lead to an average increase in fraction of loan repaid by 0.026 i.e. from 0.859 to 0.885. Given an average loan amount, this effect was of an increase in repayment amount of approximately $196.

For the subset borrowers who received price decreases, the average decrease in price was 1.5% and given an average loan amount of $8,277, it lead to an average increase in loan amount of approximately $37. Furthermore, an average decrease in price by 1.5% lead to an average decrease in default fraction by 0.015 i.e. from 0.213 to 0.198. Given the number of borrowers in this subset were 8,228, this 1.5% decrease in price lead to a decrease in the number of borrowers who defaulted by 123. Finally, a decrease in price by 1.5% lead to an average decrease in fraction of loan repaid by 0.035 i.e. from 0.757 to 0.722. Given an average loan amount, this effect amounted to an average decrease in repayment amount by approximately $290. The effects on loan amount and fraction of borrowers who defaulted have the expected signs but the positive sign on the effect of price increase on fraction of loan repaid is positive which is surprising. However, note that even though an increase in price leads to an increase in the fraction of loan repaid, this effect is small to change the default outcome of the borrower. The borrower may pay a little more, but eventually will default on the loan since default is defined is paying strictly less than the entire loan amount. In case of a price increase, the average repayment increases from 0.859 to 0.885 which still well below 1. One could argue that since the loan is smaller, the

fraction of repaid is higher because the borrower might be paying approximately the same amount, but this seems unlikely since the decrease in loan amount is of only $24 for borrowers who faced price increases. The effect on fraction of borrowers who defaulted may seem small in absolute terms, but it is essentially the most significant when interpreted in the proper context. This effect is in terms of number of people who defaulted instead of the monetary loan value. For the subset of borrowers who received price increases, the average price increase of 1.1% lead to 65 more borrowers to default. Likewise, for the set of borrowers who face price decreases, the average decrease of 1.5% lead to 123 fewer borrowers to default.

## 2.5 Conclusion

In this paper I show evidence of how borrower loan outcomes change due to the implementation of finer credit scoring in the context of a large P2P online credit platform. I used micro data at the level of individual borrower decisions and loan transactions and exploited a change in the credit scoring policy of the platform to conduct my empirical analysis. In the first part of my analysis showed that, if one simply considers the policy as a binary treatment in this market, its effects of on borrower outcomes can be estimated using propensity score matching in which the matching was done on a rich set of borrower and loan characteristics. I find that the policy increased the average requested loan amount by about 8.7% which amounted to about $684 on average. Additionally, I find that the policy decreased the likelihood of default by about 2.7% and increased the fraction of principal loan amount repaid by 2.4%. The matching process was successful in creating balanced matched samples and these results were shown to be robust to alternative estimation technique.

In the second part of my empirical analysis, I relax the assumption that the policy

treated all borrowers equally by considering the treatment variable to be continuous. By examining the implementation of the policy in detail, I classified the borrowers by their risk levels as measured by the finer credit scoring implemented by the platform. The findings from this part of my analysis showed that for the subset of borrowers who received price increases, the average treatment level was 1.1% which lead to an average decrease in loan amount by 0.29% ($24), an average increase in the fraction of borrowers who defaulted by 0.011 (65 borrowers), and an average increase in the fraction of loan repaid by 0.026 ($196). On the other hand, for the set of borrowers who received price decreases the average treatment level was -1.5% which lead to an average increase in loan amount by $37, an average decrease in the fraction of borrowers who defaulted by 0.015 (123 borrowers), and an average decrease in the fraction of loan repaid by 0.035 ($290).

The effects on mean loan amount and mean loan repayment appear to be small but statistically significant. However, the effect on the number of people who default on loans is substantial. Comparing the results of the binary treatment with the ones with continuous treatment, one can immediately spot the differences in the effects and their signs. In a case where the two effects were similar, it would not matter much to use either type of treatment variable. However, given that the continuous treatment carries more information and gives a more comprehensive picture of the effects of the treatment, it is clearly the preferred approach. Furthermore, the findings of this paper contribute to the empirical evidence that improves our understanding of how reduction in information asymmetry affects the choices of the previously informed party in the context of a credit market. Although the study is focused on a single platform which has its own idiosyncrasies, it is more representative of the wider consumer credit market than some of the earlier studies which relied on data from

**Table 2.1:** Descriptive Statistics

| | Before | | After | | | |
|---|---|---|---|---|---|---|
| **Loan Apps.** | **Mean** | **SD** | **Mean** | **SD** | **Mean Diff.** | **p-value** |
| Loan Amount | 7862 | 5404.99 | 8580 | 6159.34 | 717.56 | ¡0.00001 |
| Interest Rate | 0.23 | 0.07 | 0.22 | 0.08 | -0.003 | 0.002198 |
| Ext. Credit Score | 714 | 50.27 | 716 | 49.05 | 1.310 | 0.03475 |
| Loss Rate | 0.094 | 0.049 | 0.093 | 0.051 | -0.001 | 0.01964 |
| Listing Term | 43.347 | | 42.963 | | -0.385 | 0.01619 |
| 1(Prior Loans) | 0.195 | 0.396 | 0.184 | 0.388 | -0.011 | 0.02752 |
| 1(Home Owner) | 0.492 | 0.500 | 0.497 | 0.500 | 0.005 | 0.4273 |
| # Loan Apps. | 12,633 | | 13,004 | | | |
| **Loans Issued** | | | | | | |
| Default rate | 0.234 | 0.424 | 0.211 | 0.408 | -0.023 | 0.001 |
| Fraction repaid | 0.832 | 0.327 | 0.849 | 0.314 | 0.017 | 0.002 |
| No. of Loans | 6,977 | | 7,195 | | | |

individual lending firms.

**Table 2.2:** Propensity Score Matching Results

| | **Ln (Loan Amt.)** | **Default (0/1)** | **Fraction repaid** |
|---|---|---|---|
| Policy (ATE) | 0.087*** | -0.027** | 0.020** |
| | (0.011) | (0.009) | (0.007) |
| Controls | Loan Contract, Credit, Macroeconomic variables | | |
| Num. of Obs | 25,637 | 14,172 | 14,172 |

Notes: Standard Errors are given in parentheses. Significance level indicated as $*p < 0.05$, $**p < 0.01$, $***p < 0.001$

**Table 2.3:** OLS Estimates of the Effect of Policy

| | **Log (Loan Amt.)** | **Default (0/1)** | **Fraction repaid** |
|---|---|---|---|
| Policy (ATE) | 0.066*** | -0.015* | 0.011* |
| | (0.007) | (0.007) | (0.005) |
| Controls | Loan Contract, Credit, Macroeconomic variables | | |
| Num. of Obs | 25,637 | 14,172 | 14,172 |

Notes: Standard Errors are given in parentheses. Significance level indicated as $*p < 0.05$, $**p < 0.01$, $***p < 0.001$

**Table 2.4:** Effect of Difference in Price on Difference in Mean Loan Amount

| | **ΔMean Log (Loan Amt.)** | **ΔMean Log (Loan Amt.)** | **ΔMean Log (Loan Amt.)** |
|---|---|---|---|
| Constant | 0.005*** | 0.007*** | 0.006 |
| | (0.001) | (0.002) | (0.002) |
| $\Delta P_j$ | -25.796* | - | -23.623* |
| | (10.823) | - | (11.653) |
| $1\{\Delta P_j > 0\}$ | - | -0.003 | -0.001 |
| | - | (0.002) | (0.003) |
| N | 235 | 235 | 235 |
| $Adj.\, R^2$ | 0.020 | 0.003 | 0.017 |

Notes: Standard Errors are given in parentheses. Significance level indicated as $*p < 0.05$, $**p < 0.01$, $***p < 0.001$

**Table 2.5:** Effect of Difference in Price on Fraction of Defaulted Borrowers

| | **ΔFrac. Defaulted** | **ΔFrac. Defaulted** | **ΔFrac. Defaulted** |
|---|---|---|---|
| Constant | 0.000 | 0.000 | 0.000 |
| | (0.000) | (0.000) | (0.000) |
| $\Delta P_j$ | 1.080*** | - | 1.078** |
| | (0.317) | - | (0.329) |
| $1\{\Delta P_j > 0\}$ | - | 0.000 | 0.000 |
| | - | (0.000) | (0.000) |
| N | 231 | 231 | 231 |
| $Adj.\, R^2$ | 0.044 | -0.001 | 0.040 |

Notes: Standard Errors are given in parentheses. Significance level indicated as $*p < 0.05$, $**p < 0.01$, $***p < 0.001$

**Table 2.6:** Effect of Difference in Price on Mean Fraction of Loan Repaid

| | **ΔMean Frac. Repaid** | **ΔMean Frac. Repaid** | **ΔMean Frac. Repaid** |
|---|---|---|---|
| Constant | 0.001*** | 0.001*** | 0.001*** |
| | (0.000) | (0.000) | (0.000) |
| $\Delta P_j$ | 2.342*** | - | 2.553*** |
| | (0.589) | - | (0.608) |
| $1\{\Delta P_j > 0\}$ | - | 0.000 | 0.000 |
| | - | (0.000) | (0.000) |
| N | 231 | 231 | 231 |
| $Adj.\, R^2$ | 0.061 | -0.004 | 0.065 |

Notes: Standard Errors are given in parentheses. Significance level indicated as $*p < 0.05$, $**p < 0.01$, $***p < 0.001$

**Table 2.7:** Summary Statistics Based on Treatment Signs

| | $\Delta P_i > 0$ | $\Delta P_i < 0$ |
|---|---|---|
| Mean $\Delta P_i$ | 0.011 | 0.015 |
| Mean Loan Amount | $7,536 | $8,277 |
| Mean Frac. of Borrowers Defaulted | 0.137 | 0.213 |
| Mean Frac. of Loan Repaid | 0.859 | 0.757 |
| N (%) | 5,944 (41.9%) | 8,228 (58.1%) |

**Figure 2.1:** Pricing Function Before the Policy Change

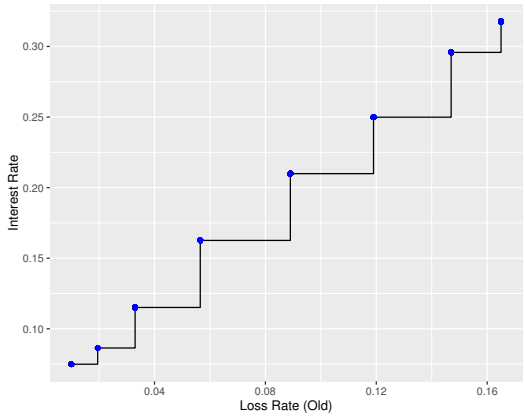



**Figure 2.2:** Pricing Function After the Policy Change

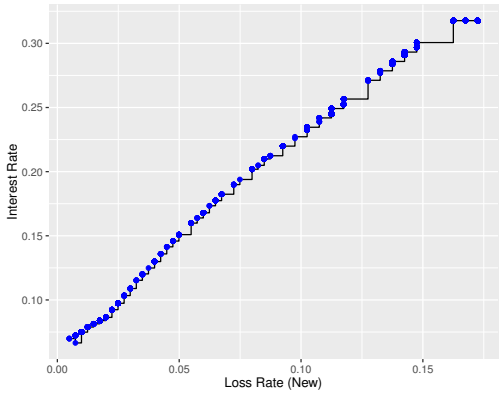


**Figure 2.3:** Kernel Density Plot of $\Delta P_i$

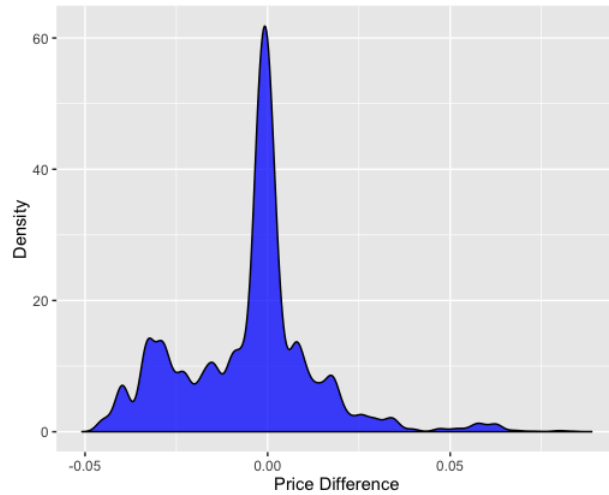



**Figure 2.4:** Pricing Table Before Implementation of Finer Credit Scoring

| Prosper Rating | Term (yrs) | # Previous Prosper Loans | Borrower Rates |
|---|---|---|---|
| AA | 1 | 1+ | 5.65% |
| AA | 3 | 1+ | 6.49% |
| AA | 5 | 1+ | 9.76% |
| AA | 1 | 0 | 5.65% |
| AA | 3 | 0 | 7.49% |
| AA | 5 | 0 | 10.85% |
| A | 1 | 1+ | 6.28% |
| A | 3 | 1+ | 9.64% |
| A | 5 | 1+ | 13.85% |
| A | 1 | 0 | 8.17% |
| A | 3 | 0 | 11.51% |
| A | 5 | 0 | 15.51% |
| B | 1 | 1+ | 10.10% |
| B | 3 | 1+ | 14.69% |
| B | 5 | 1+ | 18.30% |
| B | 1 | 0 | 12.24% |
| B | 3 | 0 | 16.26% |
| B | 5 | 0 | 20.85% |
| C | 1 | 1+ | 13.04% |
| C | 3 | 1+ | 19.02% |
| C | 5 | 1+ | 22.20% |
| C | 1 | 0 | 15.76% |
| C | 3 | 0 | 20.99% |

**Figure 2.5:** Pricing Table After Implementation of Finer Credit Scoring

| Prosper Rating | Loan Term (yrs) | # Previous Prosper Loans | Borrower Rate | |
|---|---|---|---|---|
| | | | Min | Max |
| AA | 1 | 0 | 5.65% | 5.65% |
| AA | 3 | 0 | 7.49% | 7.49% |
| AA | 5 | 0 | 10.71% | 10.71% |
| AA | 1 | 1+ | 5.65% | 5.65% |
| AA | 3 | 1+ | 7.49% | 7.49% |
| AA | 5 | 1+ | 10.71% | 10.71% |
| A | 1 | 0 | 9.05% | 9.81% |
| A | 3 | 0 | 11.99% | 12.99% |
| A | 5 | 0 | 16.19% | 17.28% |
| A | 1 | 1+ | 7.69% | 9.05% |
| A | 3 | 1+ | 9.74% | 11.99% |
| A | 5 | 1+ | 13.64% | 16.19% |
| B | 1 | 0 | 11.47% | 12.79% |
| B | 3 | 0 | 15.09% | 16.79% |
| B | 5 | 0 | 19.47% | 21.24% |
| B | 1 | 1+ | 10.72% | 12.15% |
| B | 3 | 1+ | 14.14% | 15.99% |
| B | 5 | 1+ | 18.52% | 20.39% |
| C | 1 | 0 | 13.46% | 15.87% |
| C | 3 | 0 | 17.74% | 20.99% |
| C | 5 | 0 | 22.18% | 24.98% |
| C | 1 | 1+ | 13.46% | 15.87% |
| C | 3 | 1+ | 17.74% | 20.99% |

# Chapter 3

# What Drives the Expansion of the Peer-to-Peer Credit?[1]

## 3.1 Introduction

First peer-to-peer (P2P) credit platforms, Zopa, Prosper and Lending Club, have been launched in 2005-2007 in the UK and the US. These online lenders directly match savers with borrowers who need personal and business loans.[2] Although, online credit amounts to a small share of total credit, it has been growing rapidly (Figure 3.1) and in 2015, the flow of US online consumer credit was equivalent to 12.5% of traditional consumer credit (Wardrop et al., 2016). Not surprisingly, the emergence of online lenders, which are a part of the wider FinTech movement, has provoked a debate about their ability to disrupt traditional banking (Philippon (2015); The Economist, 2015; Wolf, 2016; Citi, 2016). Haldane (2011) suggests that the entry of new FinTech players could diversify the intermediation between savers and borrowers, which would make the financial sector more stable and efficient and could ensure greater access to financial services.

The objective of this paper is to provide the first exploration of the main drivers of the expansion of the P2P Credit in the US. Is rapid development of online lenders due to structural factors in the brick-and-mortar banking, such as weak competition in the consumer credit market due to high switching costs or barriers to entry? Has

[1]This chapter is based on a joint work with Olena Havrylchyk, Carlotta Mariotto and Marianne Verdier

[2]Peer-to-peer credit was born to match directly lenders and borrowers without the use of the intermediation of banks. However, as the market expanded, a large part of it has been funded not by individual lenders, but traditional banks, hedge funds and other financial institutions. Hence, the name Peer-to-peer credit has been changing to marketplace credit. In this paper we use terms Peer-to-peer credit platforms, marketplace lenders and online lenders interchangeably.

it been spurred by the Great Recession, bank failures, banks' deleveraging and credit crunch? Could the timing of the P2P Credit be explained by the spread of Internet, sophistication of Internet users and trust in new technologies? What role do social networks play? What are the socio-economic and demographic characteristics of online borrowers? Ultimately, we would like to get closer to understanding whether online lenders could be potentially disrupt the traditional banking sector.

In light of these questions, we outline three main hypotheses for the expansion of online lenders. Our first hypothesis is that P2P Credit development could be related to the nature of the banking competition. The banking sector is characterized by monopolistic competition due to high entry barriers, switching costs and strong brand loyalty (Claessens and Laeven (2004); Shy (2002); Kim et al. (2003)). Philippon (2015) demonstrates that the cost of financial intermediation in the US have remained unchanged since the 19 century. This fact is astonishing in the context of rapid progress in the communication and information technologies that should have driven down the price of financial services for end users. Hence, the entry of new Fintech players could be needed to improve the provision of financial services and disrupt traditional players. Indeed, online lenders argue that their operating expenses are much lower than those of brick-and-mortar banks due to the extensive use of new technologies as well as absence of legacy problems and costly branch networks.[3] We test the impact of the market structure on the expansion of online lenders and refer to these explanations as *competition-based hypotheses*.[4]

[3]Operating expenses include the costs of originating the loan, processing payments, collection and bad debt expenses.

[4]The existing literature finds weak conclusions on the relationship between innovation and market structure (see the survey of Cohen and Levin, 2010). A number of theoretical studies (e.g., Gilbert, 2006) show that the competition innovation is monotonic only under restrictive conditions. On the one hand, innovation incentives should be lower in more concentrated markets because of the replacement effect identified by. On the other hand, innovation incentives should be lower in more competitive environments because aggregate industry profits are lower. Aghion et al. (2005)

The expansion of online lenders might have been spurred by the financial crisis and the Great Recession. On the credit supply side, as interest rates approached zero, new lenders entered the market, attracted by the higher return (and risk) available from exposure to P2P assets. On the credit demand side, a wider and more creditworthy pool of potential borrowers appeared as the banking sector was weak, regulation has tightened, banks have deleveraged and mistrust in the banks has spread (Atz and Bholat, 2016). As shown by figure 3.2 below, total consumer credit significantly decreased in the years 2008-2011. The credit rationing may have spurred the demand for alternative forms of financing.

For example, Hasan et al. (2009) show that bank instability in Germany has pushed businesses to use equity crowdfunding as a source of external finance. We refer to this explanation as crisis-based hypothesis.

It is also possible that the surge in P2P Credit is not caused by problems in the banking sector. Online lenders claim to harness big data innovations to revolutionize credit risk assessment and efficiently match lenders with borrowers. Furthermore, the entry of online lenders reflects the readiness of the society to embrace internet to perform financial transactions. Indeed, Fintech is part of the larger revolution as new internet platforms (Amazon, Uber, BlaBlaCar and AirBnB) are on the way to disrupt other service markets, such as retail trade, transport and accommodation. Similar to previous financial innovation, online lenders could expand and cheapen access to financial services (Einav et al. (2013b)). We refer to this explanation as innovation-based hypothesis.

Sorting out these three competing hypotheses is difficult because the expansion of the P2P Credit has coincided with the post-crisis period, increased concentration of

demonstrate that the relationship between competition and innovation should have a nonlinear inverted U-pattern. Other studies include measures of entry and exit in the market (Geroski (1989).

the banking sector and the diffusion of communication and information technologies (e.g., smartphones, broadband). Our identification strategy relies on the exploration of the geographic heterogeneity of the P2P Credit expansion at the county level. The choice of the local dimension of a market is relevant for consumer and SME credit that are targeted by online lenders. The county unit is the standard definition of the local banking market in the literature (e.g., Hannan and Prager (1998); Berger et al. (1999); Rhoades et al. (2000); and Black and Strahan (2002)).

Since the expansion of the P2P Credit is similar to the diffusion of other technologies, it could be explained by spatial network effects due to human interactions (Comin et al. (2012)). Notwithstanding the online nature of the P2P Credit, geography might still play a crucial role in its diffusion. Indeed, we document an important spatial correlation, as P2P Credit per capita is higher in counties close to California, New York and Florida. Hence, our econometric approach relies on incorporating a spatial lag variable in our model. [5]

This paper contributes to the nascent literature on the Peer-to-peer credit. The largest strand of this literature explores how borrower characteristics affect loan outcomes and how lenders on P2P platforms mitigate informational frictions (see the literature review by Morse (2015)).[6] The only paper that explores how borrowers choose between traditional and alternative sources of finance is Butler et al. (2016),

[5]This hypothesis is different from but related to the study by Agrawal et al. (2011) who find that crowdfunding largely overcomes the distance-related economic frictions as the average investor is not in the local market but is 3,000 miles away. Our hypothesis that the expansion of the P2P Credit exhibits spatial correlation does not contradict the fact that investors could be located far away.

[6]Morse (2015) provides a literature survey of papers that study how P2P Credit mitigates information frictions by relying on real world social connections (Freedman and Jin (2008); Everett, 2010), textual analysis of successful funding bids (Mitra and Gilbert (2014)), psychology text mining techniques to uncover deception (Gao and Lin (2013)), identity claim methodology to identify trustworthy and hardworking borrowers (Herzenstein et al. (2011)) as well as discrimination (Ravina (2013); Pope and Sydnor (2011); Duarte et al. (2012)).

who show that borrowers who reside in areas with good access to bank finance request loans with lower interest rates.

This paper makes the first attempt to analyze the expansion patterns of online lenders. For the first time, we aggregate data for the two leading P2P consumer credit platforms in the US - Prosper and Lending Club – and study the geography of online lenders. We measure the expansion of the P2P Credit by aggregating the number and the volume of loans provided by these two online lenders. As early as 2007, 1183 counties had P2P borrowers, and their number has increased to 2609 in 2013. We then use this data to relate the amount of P2P Credit to a wide range of county level determinants that could affect the speed of its penetration.

By focusing on the expansion of a new technology, our paper is related to the literature on the diffusion of innovation (Bass (1969) and Rogers (2003)).[7] The literature on financial innovation is scarce and focuses on the new products and distribution channels in the traditional banking (Frame and White (2014)). Most of these studies have focused on users' incentives to adopt innovations according to their individual characteristics.[8] DeYoung et al. (2007) and Hernando and Nieto (2007) analyze the impact of the adoption of online banking on banks' profitability and find that the Internet channel is a complement to rather than a substitute for physical branches.

The paper is structured as follows. In section 3.2, we describe the institutional environment in which Peer-to-peer credit platforms evolve. In section 3.3, we explain how we assemble our data set, provide data sources and variable definition. In section

[7]Rogers (2003) argues that the more people that use a technology, the more non-users are likely to adopt.

[8]Frame and White (2014) mention three different types of innovations: products and services (e.g., subprime mortgages, new means of payment and online banking), production processes (such as Automated Clearing Houses, small business credit scoring, asset securitization, risk management), organizational forms (such as Internet only banks).

3.4, we explain our identification strategy and provide empirical results. In section 3.5, we conclude.

## 3.2 Institutional Environment of Peer-to-Peer Credit Platforms in the United States

Online credit marketplaces are platforms that connect individuals or businesses wishing to obtain a loan with individuals and institutions willing to commit to fund this loan. Marketplace credit encompasses P2P Credit platforms, which offer credit-based crowdfunding for consumers and small businesses, and online credit platforms by large institutions (e.g., OnDeck Capital, Kabbage), which offer credit exclusively to businesses, rather than consumers.[9] In our paper, we focus on P2P Credit platforms, on which multiple lenders lend small sums of money online to consumers or small businesses with the expectation of periodic repayment.

Prosper Marketplace and Lending Club launched the first online P2P Credit platforms in the United-States respectively in 2006 and 2007, followed by other companies such as Upstart, Funding Circle, CircleBack Credit or Peerform. Between 2006 and 2015, the two most important platforms, Prosper and Lending Club, have facilitated approximately $8.7 billion loans.[10] Both platforms believe that their online marketplace model has key advantages relative to traditional bank credit both for borrowers and investors, among which convenience of online operations, automation, reduced cost and time to access credit.

Consumer loan amounts vary between a minimum loan of $1,000 for Prosper and $500 for Lending Club and a maximum loan of $35,000 for both platforms ($300,000

[9] Other types of crowdfunding include donation or reward-based crowdfunding.

[10] The figures and information of this paragraph is based on the study of Prosper and Lending Club annual reports, which can be found on the companies' websites.

for businesses). They fund various types of projects ranging from credit card debt consolidation to home improvement, short-term and bridge loans, vehicle loans or engagement loans.[11]

Prosper and Lending Club rely on a partnership with WebBank, an FDIC-insured, Utah-chartered industrial bank that originates all borrower loans made through their marketplaces. In December 2014, Lending Club became the first publicly traded online Peer-to-peer credit company in the United-States, after its Initial Public Offering on the New York Stock Exchange.

As in many other two-sided markets (Rysman (2009)), online credit marketplaces try to attract two different groups of users, namely borrowers and investors, by choosing an appropriate structure of fees that increases the size of network effects. On the borrower side of the market, both companies compete with banking institutions, credit unions, credit card issuers and other consumer finance companies. They also compete with each other and with other online marketplaces such as Upstart or Funding Circle. Platforms claim that their prices are lower on average than the ones consumers would pay on outstanding credit card balances or unsecured installment loans funded by traditional banks.[12] Online marketplaces perform the traditional screening function of banks by defining various criteria that must be met by borrowers. Any U.S. resident aged at least 18 with a U.S. bank account and a social security number may apply and request a loan, provided that the platform is authorized in her/his state. Platforms collect online some information about the applicant (i.e., FICO score, debt-to-income ratio, credit report variables, etc.), which is used

[11]Consumer credit does not include credit for purchase of a residence or collateralized by real estate or by specific financial assets like stocks and bonds.

[12]This view is confirmed by a study conducted by Demyanyk and Kolliner at the Federal Reserve Bank of Cleveland. They offer time-series evidence that, on average, marketplace loans carry lower interest rates than credit cards and perform similarly.

to compute a proprietary credit score. Some additional enquiries may also be performed offline (e.g., employment verification). Consumers are divided into several rating segments, which correspond to different fixed interest rates ranging from 6% to 26% for Lending Club in 2014. Origination fees paid to the platform depend on the consumer's level of risk.

On the investor side, online credit marketplaces face potential competition from investment vehicles and asset classes such as equities, bonds and commodities. Prosper claims to offer an asset class that has attractive risk adjusted returns compared to its competitors. Investors can be divided into two different populations: individuals and institutions. Both populations are subject to different requirements. Individual investors must be U.S. residents aged at least 18, with a social security number, and sometimes a driver's license or a state identification card number. Institutional investors must provide a taxpayer identification number and entity formation documentation. Investors' annual income must exceed a floor defined by platforms' rules. Prosper and Lending Club issue a series of unsecured Notes for each loan that are sold to the investors (individual or institutional), and recommend that each investor diversifies his/her portfolio by purchasing small amounts from different loans.[13] Each investor is entitled to receive pro-rata principal and interest payments on the loan, net of a service charge paid to the platform. In addition to the "Note Channel", Prosper has designed specifically a "Whole Loan Channel" for accredited investors (according to the definition set forth in Regulation D under the Securities Act of 1933), which must be approved by the platform. Accredited Investors can purchase a borrower loan in its entirety directly from Prosper.

The credit market in the United-States is subject to many regulations, which are

[13]Notes can be viewed as debt-back securities.

changing continuously (e.g., State Usury Laws, State Securities Laws, Dodd-Frank Wall Street Reform and Consumer Protection Act, Truth-in-Credit Act...). Online credit platforms need to obtain a license to operate in a given state and comply with all existing regulations on consumer credit. For example, currently, Lending Club does not facilitate loans to borrowers in Idaho, Iowa, Maine, Nebraska and North Dakota, but has obtained a license in all other jurisdictions. Furthermore, state and local government authorities may impose additional restrictions on their activities (such as a cap on the fees charged to borrowers) or mandatory disclosure of information. In some states, platforms are opened to borrowers but not to investors, or vice versa. Authorizations can also differ for Prosper and Lending Club.

An important issue is the potential violation of states' usury laws. The interest rates charged to borrowers are based upon the ability under federal law of the issuing bank that originates the loan (i.e., WebBank) to "export" the interest rates of its jurisdiction (i.e., Utah) to other states. This enables the online marketplace to provide for uniform rates to all borrowers in all states in which it operates. Therefore, if a state imposes a low limit on the maximum interest rates for consumer loans, some borrowers could still borrow at a higher rate through an online marketplace since the loan is originated in Utah.[14] Some states have opted-out of the exportation regime, which allows banks to export the interest rate permitted in their jurisdiction, regardless of the usury limitations imposed by the borrower's state.

[14]Of the fourty-six jurisdictions whose residents may obtain loans in the United-States, only seven states have no interest rate limitations on consumer loans (Arizona, Nevada, New Hampshire, New Mexico, South Carolina, South Dakota and Utah), while all other jurisdictions have a maximum rate less than the maximum rate offered by WebBank through online marketplaces.

## 3.3 Data

To construct variables about the diffusion of P2P Credit, we rely on loan book data from Lending Club and Prosper Marketplace. For Lending Club we have 376 261 observation points, corresponding to a total volume of funded loans equal to $3.2 billion, starting from January 2007 to December 2013. This amounts to 99.25% of the Lending Club portfolio. For Prosper we have 88 988 observation points, corresponding to a total volume of originated loans equal to $662 million, starting from January 2006 to 30 October 2013. This amounts to 100% of the total Prosper portfolio. There are 313 counties with zero P2P loans in our final dataset.

Since loan book data provides information about each borrower's city, we can assign a county name to each borrower by matching with an official data containing US States, cities and counties.[15] Our analysis ends in 2013, because platforms have stopped providing city names afterwards. Due to missing values and mistakes in city names, we lose 4.8% of the volume of funded loans in the Lending Club dataset and 10% from the Prosper dataset. Next, we aggregate this data at the year-county level to construct two measures of P2P Credit diffusion: number of P2P loans per capita and volume of P2P Credit per capita. For large cities belonging to multiple counties, we split the total data between counties weighted by total income per county. Table 3.1 shows the total volume of funded loans, the number of counties and the total number of loans that we have in our dataset.

We can now map the depth of the P2P development at the county level for each year (Figures 3.3 to 3.). As early as 2007, 1183 counties had P2P borrowers, and their number has increased to 1881 in 2010 and to 2609 in 2013. For cross-sectional regres-

[15] We use the Americas Open Geocode (AOG) database. Source: http:/ www.opengeocode.org/download.php.

sions, we aggregate yearly data for each county and, then, merge our dataset with other datasets that contain our explanatory variables. Our specification accounts for a large number of county characteristics that could influence the expansion of the P2P Credit.

***Crisis variables***

To measure the effects of the financial crisis on the penetration of the P2P Credit, we rely on two types of variables. First, we compute the share of deposits in each county affected by bank failures during the analyzed period. To do this, we merge FDIC Failed Bank List with the data on branches of these banks in each county from the FDIC Summary of Deposits. This is an exhaustive database about all branches of deposit taking institutions in the US, providing data on the amount of deposits at the branch level. We then compute the share of deposits held by failed banks in a county i in the total amount of deposits held by all banks in a county i as of 31 December, 2013. As shown by Aubuchon et al. (2010), there is a wide geographic heterogeneity with respect to bank failures in the US and it is possible that customers from counties that have been the most affected by the crisis have relied more on alternative credit providers. If our crisis-related hypothesis is confirmed, we expect a positive sign on this variable.

Our second measure of the depth of the financial crisis relies on the FDIC Summary of Deposits to identify the presence of branches in each county that we merge with information on capital at the bank consolidated level, taken from Call Reports. This measure is based on the assumption that banks' capital management is done at the consolidated level (Haas and van Lelyveld, 2010). We rely on two measures of capital (unweighted leverage ratio and risk-weighted tier 1 capital ratio) computed

during the crisis period 2009-2010.[16] Solvency ratio of a county i is computed as an average capital ratios of banks present in a county i weighted by deposits of their branches in county i. If our crisis-related hypothesis is confirmed, we expect a negative sign on this variable.

***Measuring competition and brand loyalty***

Ideally, we would like to explore banking competition, but this is notoriously difficult to measure, particularly at the county level. The FDIC Summary of Deposits allows us to compute concentration measures, such as HHI and C3 indices, as well as branch density per 10000 population. To eliminate any endogeneity due to reverse causality, we estimate these variables in 2007. Since some studies show that market structure could be unrelated to the banking competition (Claessens and Laeven (2004)), we prefer to refer to these measures as market structure or concentration measures.

Market structure measures could be correlates of bank quality and brand loyalty. In particular, branch density measures the outreach of the financial sector in terms of access to banks' physical outlets (**?**; Beck et al. (2007)). Branch density is also a measure of the quality of the overall bank network and could play an important role in the bank's advertising strategy to develop brand loyalty (Dick (2007)). Indeed, branches are a form of advertising for banks. Dick (2007) provides plenty of anecdotal evidence on how banks hope to attract customers using their branches, usually with stylish merchandising and customer service. Banks become more visible to consumers through their branches; in fact, banks are known to put clocks outside their branches for this reason. Importantly, there is evidence that banks open branches mostly in

[16]We define these two years as crisis-years because bank capital ratios and loan growth were at their lowest and bank failures and credit-card delinquencies at the highest during this period. This allows us to capture the severity of the crisis.

response to their own market targets, as opposed to their existing customers' needs.

Banking sector is a highly concentrated market with high switching costs. If bank customers wanted to switch to P2P Credit, they would need to incur learning costs about P2P platforms, transaction costs to set up their profile, describe their loan (a task that is performed by their credit officer in a bank), as well as to overcome brand loyalty. Since our study is done in the homogeneous institutional environment in the context of switching to one of the two very similar credit platforms, learning and transaction costs should be similar across counties. We control for educational attainment and age, which could be correlated with learning costs. The remaining geographic heterogeneity in banking concentration could be a subjective measure of brand loyalty.

In light of this discussion, the impact of the concentration measures on the expansion of the P2P Credit could be interpreted differently. A positive correlation between market concentration and P2P Credit platforms could signal that customers from highly concentrated markets try to switch to alternative, less costly providers. A negative correlation, on the contrary, could signal that high market concentration reflects high brand loyalty, which slows down the penetration of the P2P Credit.

Finally, since credit marketplaces operate online, their entry decision at the county level is exogenous and it is not correlated to the density of bank branches.

***Measuring openness to innovation and new communication and informational technologies***

To proxy for openness to innovation, we use U.S. Patent and Trademark Office data to compute the number of patents per capita. This measure is often used as a measure of innovation and, as such, it has a number of shortcomings, since some

innovations are not patented and patents differ enormously in their economic impact. Nonetheless, our objective is not to measure innovation per se, but rather to account for a local culture that has a high propensity to generate innovative ideas and, hence, accept innovative ideas of others. Such culture could be more open to new forms of financing though P2P Credit.

To measure the penetration of internet at the county level, we rely on the NTIA's State Broadband Initiative that allows us to compute the following measures: 1) percent of county population with access to any broadband technology (excluding satellite); 2) percent of county population with access to Mobile Wireless (Licensed) technology; 3) percent of county population with access to upload speed 50 mbps or higher. Each measure is computed as an average between 2010 and 2013, the only data available at the county level. All these variables should have an expected positive sign if our innovation-based hypothesis is confirmed.

***Socio-economic characteristics***

We control for the socio-economic characteristics, such as age, education attainment, population density, poverty level, race etc. We expect that counties with higher educational attainment, higher population density and higher proportion of young people, should have higher levels of P2P Credit penetration because human capital and network effects of urban areas are significant predictors of the technological diffusion. These characteristics could also be correlates with brand loyalty.[17]

As to poverty rate and race, we have no theoretical priors about the sign of their impact. Racial minorities might be less familiar with online credit opportunities, but their demand could be higher because race identification is no longer possible

[17]Surveys have found that consumer credit use is greatest in early family life stages when the rate of return of additional goods that might be financed using credit is high.

on P2P Credit platforms.[18] Interestingly, racial identification was possible during earlier years of the P2P Credit when borrowers had the possibility to post a picture. This has led to the well documented discrimination of racial minorities on the Prosper credit platform (Pope and Sydnor (2011); Ravina (2013); Duarte et al. (2012)). Consequently, platforms have removed the possibility of posting a photo which has made the identification of borrowers' race impossible. This could incentivise racial minorities to turn to the P2P platforms to avoid discrimination that is well documented in traditional credit markets (see a literature review by Pagern and Shepherd, 2008).

We introduce state level dummies to control for differences in state-level regulation of consumer credit and P2P Credit platforms, as well as other state characteristics that are not captured by our county-level variables. These dummies account for the fact that Iowa was closed for borrowers from both Lending Club and Prosper platforms, while Maine and North Dakota were closed for Prosper platform.

***Spatial relations***

Our data contain explicit spatial relationships, as counties are likely to be subject to observable and unobservable common disturbances which will lead to spatial correlation. This could be explained by various channels of interdependence due to regional business cycles and economic shocks, technology diffusion, access to bank branches, policy coordination, regional disparities for which we do not control with our right-hand variables (see e.g. Garrett et al. 2005 for the importance of spatial correlation in state branching policy). Spatial correlation could also occur because of the boundary mismatch problems when the economic notion of a market does

[18] However, the platforms have removed the possibility of posting the photo, which has made the identification of borrowers' race impossible.

not correspond well with the county boundaries (Rey and Montouri (1999)). Spatial correlation is particularly important for the diffusion of technology due to a theory of human interactions (Comin et al. (2012)). Borrowers from P2P Credit platform require acquiring knowledge about their existence, as well as trust in their reliability, which often comes from interactions with other agents. The frequency and success of these interactions is likely to be shaped by geography. Hence, we expect that knowledge about P2P potential is likely to be more easily transmitted between agents in counties that are close than between counties that are far apart. Figure 3 also attest to this hypothesis. To account for spatial correlation, we introduce a spatial lag in our model.

Overall, we have sufficient cross-sectional data for 3,059 out of 3,144 counties and county equivalents. Table 3.2 provides summary statistics.

## 3.4 Methodology

### A. Model specification: a spatial autoregressive model

Our objective is to test

(i) The three hypothesis on the adoption of P2P lending (See Section 3)

(ii) Whether adopting P2P lending in a county has a positive impact on the adoption of P2P lending in neighboring counties

We specify the following regression model, also known as a SARAR model in the literature (See Anselin et al. (1980)):

$$y_i = \beta_0 + \lambda W y_{-i} + \beta_1 \times Competition_i + \gamma \times Crisis_i + \delta \times Innovation_i + X_i'\alpha + u_i$$

where

$$i, j = 1, ..., n$$

and

$$u_i = \rho \sum_{j=1}^{n} w_{ij} u_j + \varepsilon_i$$

with

$$\varepsilon_i \sim N\left(0, \sigma^2 I\right)$$

Where $i$ and $j$ denote one of the $n$ counties, $y_i$ is the log of our observed dependent variable, that is either the volume of P2P Credit per county per capita or the number of P2P loans per county per capita. $Wy_{-i} = \sum_{j \neq i}^{n} w_{ij} y_j$ is a weighted average of P2P Credit per county per capita of other counties, known as a spatial lag, where the weights are determined by an $N \times N$ spatial weights contiguity matrix $W = \sum_{j \neq i}^{n} w_{ij}$ where each element $w_{ij}$ denotes the degree of spatial proximity between county $i$ and county $j$.[19] $\lambda$ is the spatial autoregressive coefficient; $\beta_1$ is the coefficient of our observed independent variables regarding competition and market structure; $\gamma$ is the coefficient of variables regarding the credit rationing; $\delta$ is the coefficient of our variables regarding the innovation and internet variables; $\alpha$ is a vector of coefficients for our socio-economic and demographic variables ; $\rho$ is the spatial autoregressive coefficient as, in our model. We allow the error term to be affected

[19] The matrix $W$ we use is a "minmax-normalized" matrix where the $(i, j)^{th}$ element of $W$ becomes $w_{ij} = \mathbf{w_{ij}}/m$, where $m = \{max_i\left(r_i\right), max_i\left(c_i\right)\}$. Here $max_i(r_i)$ is the largest row sum of $W$ and $max_i(c_i)$ is the largest column sum of $W$. We also use the inverse-distance matrix composed of weights that are inversely related to the distances between the units, and we obtain similar results in our regression. Obtaining similar results with an inverse-distance and a contiguity matrix is consistent with the findings of LeSage and Pace, 2010.

by the disturbances of neighbors; $\varepsilon_i$ and $u_i$ are unobserved error terms.

Thus, this model specification accounts not only for spatial correlation of the dependent variable, but also for spatial correlation within the error terms, which could be affected by unobservable factors such as regional economic cycles. Ignoring spatial relation, in this case, could potentially lead to inconsistency in the standard errors.

Our main parameters of interest are the coefficients $\beta, \gamma, \delta, \alpha$ and $\lambda$ . The parameters $\beta, \gamma$ and $\delta$ measure the marginal impact of market structure variables, crisis variables, innovation variables as well as socio-economic and demographic variables on the adoption of P2P Credit in each county. When the dependent variable is the volume of P2P loans per capita, the magnitude of the coefficients $\beta, \gamma, \delta, \alpha$ predict of how many dollars the volume of P2P loans will increase or decrease for a one unit increase of the control variable. When the dependent variable is the number of loans, the magnitude of the coefficients $\beta, \gamma, \delta, \alpha$ predict how many additional or less loans there will be following a one unit increase of the control variable. Finally, $\lambda$ measures how the adoption of P2P Credit in a given county positively impacts neighbor counties. If this coefficient is significantly greater than 0, we can conclude that there is a correlation between the adoption of P2P Credit between neighboring counties.

To compute our cross-sectional spatial regressions, we use the Maximum-Likelihood Estimator method[20], as the OLS estimation will be biased and inconsistent due to simultaneity bias (See Anselin, 2003 and LeSage and Pace, 2009 for a theoretical explanation on why MLE solves the simultaneity bias). As a matter of fact, the spatial lag term must be treated as an endogenous variable since the volumes of loans in

[20]The maximum likelihood estimator method relies on the assumption that the error terms are normally distributed.

contingent counties are simultaneously impacting one another.

Our findings are presented in Tables 3.3, 3.4, B.1 and B.2 and they all show that we always reject the null hypothesis that the spatial lag lambda is greater or equal to 0. Spatial lag is always positive and statistically significant, pointing to the existence of strong spatial effects. In other words, the higher the level of P2P loans in one county, the higher it is going to be in the contingent counties.

**OLS vs. SARAR**

Since from the SARAR model the estimates for the coefficients $\rho$ and $\lambda$ are significantly different from zero, ordinary least-squares may lead to inconsistent estimations. Table B.3 in the Appendix shows the estimates from the OLS regression model. If we compare these estimates to the output from our SARAR model, we realize that they are mostly biased up words as in LeSage (2008).

#### 3.4.0.1 Estimation Results

The SARAR model estimates cannot be interpreted as partial derivatives like in the typical regressions (see LeSage and Pace (2009)). Therefore the coefficients cannot be interpreted as marginal effects of the explanatory variable on the dependent variable in one region, because a change in the explanatory variable is likely to impact the dependent variable in all neighboring regions too. In subsection A we will discuss the short-run impacts of a change in the explanatory variables on the volume and number of P2P Credit per capita in each county. In subsection B, we will compute the average total direct impact (ATDI), the average total indirect impacts (ATII) and the average total impact (ATI) which is the sum of the direct and indirect impacts.

**Empirical results: short run impacts of the explanatory variables on the dependent variable**

Table 3.3 and table 3.4 present our empirical findings for the P2P expansion (in terms of volume and number of loans respectively) as a function of different county characteristic, with a particular focus on crisis and competition characteristics.

Among socio-demographic variables, higher population density, higher educational attainment, lower levels of poverty, lower levels of income and higher share of Hispanic and Black minorities have a positive and significant impact on the expansion of the P2P Credit. An increase of population density by one standard deviation significantly increases the volume of the P2P Credit and the number of loans. An increase of bachelor graduates by one standard deviation significantly increases the volume of the P2P Credit and the number of loans. An increase of the share of Hispanic minorities by one standard deviation increases the volume of the P2P Credit and the number of loans. As reported in table B.3, this result is driven by Lending Club. Also, an increase of the share of Black minorities by one standard deviation increases the number of the P2P loans, but does not affect the volume. An increase in the percentage of people leaving under the poverty line decreases the volume of P2P Credit and the number of P2P loans. Also, the income per capita affects negatively only the volume of loans: an increase in the income per capita decreases the volume of P2P loans. The variables measuring the age of the population are never significant for these specifications.

Our finding that the expansion of the P2P Credit is faster in counties with higher share of Black and Hispanic minorities could be a sign of higher demand from these areas to escape discrimination in traditional credit markets. As online lenders have removed the possibility to post a photo, identifying the race of the borrower has

become much more difficult. During our sample period, 2007-2013, investors had access to the information on the location of borrowers. Although this information could have been used by institutional investors as a proxy for race, it is unlikely that retail investors would do that. Recently, any information on the location of the borrower has been removed, which makes the identification of the race completely impossible. Hence, racial discrimination is not anymore possible in the online credit.

The positive effect of the higher educational attainment is consistent with the fact that human capital is a significant predictor of the technological diffusion and could diminish switching costs due to lower cost of learning. A positive effect of population density reflects the existence of network effects in urban areas that is another well-known predictor of the diffusion of new technologies.

As to the crisis variables, our findings show that in both specifications of tables 3.3 and 3.4, the leverage ratio is statistically significant and has a negative effect on P2P Credit expansion both in terms of volume and loans. A decrease of the leverage ratio during the financial crisis increases the volume of credit, and increases the number of loans. The share of deposits affected by failed banks and the Tier 1 capital ratios during the crisis did not have an impact on the diffusion of P2P Credit. This finding is consistent with the idea that leverage ratios appear to be better predictors of future banks' performance and problems (Blundell-Wignall and Roulet (2013); Haldane (2011) and Haldane (2012)) with respect to weighted leverage ratios, since weights may be inconsistent and subject to manipulations (Mariathasan and Merrouche (2014)); Le Lesl and Avramova, 2012; Haldane (2012); FSA, 2010).

Most of P2P borrowers use credit platforms to consolidate and manage their credit card debt and a minority borrow for business purposes. To account for difficulties in the credit card market, we test the robustness of our results by constructing two

additional crisis variables: percentage change in credit card debt balance per capita and percent of credit card debt balance with more than 90 days of delinquency during crisis years. The data comes from the New York Fed Consumer Credit Panel / Equifax that is available only for 2220 counties. None of these variables turns out to be statistically significant. Results are available upon request.

In addition, Table 3.3 and table 3.4 present the empirical findings for the P2P expansion as a function of market structure variables. Our findings demonstrate that low branches density in 2007 is a statistically significant driver of the P2P Credit. We interpret this result as a suggestion that customers living in counties with low outreach of traditional banks and low quality of financial services are more likely to turn to P2P Credit due to weaker brand loyalty. Counties that had one standard deviation less branches in 2007, experienced an increase in the average volume of P2P Credit and an increase in the number of P2P loans.

Turning our attention to concentration measures, both our concentration measures C3 and HHI have a negative and statistically significant sign. In other words, P2P Credit penetrates fewer counties with higher concentration of the largest three banks and with a higher overall traditional banking market concentration. This is consistent with the interpretation of the high market concentration as an outcome of high switching costs due to strong brand loyalty. An increase of the concentration of the three biggest banks by one standard deviation diminishes the average amount of the P2P Credit and the number of loans, whereas an increase in the concentration of the whole traditional banking market in one county diminishes the average amount of P2P Credit and the number of loans.

We additionally test the impact of the alternative consumer credit providers, such as payday loans. To do so, we use County Business Patterns to construct

the ratio of non-bank establishments that are related to consumer credit and credit intermediation per capital (Bhutta, 2013). We find that P2P Credit is more diffused in counties with a higher number of payday loan establishments. In particular, an increase in the number of payday loans establishments increases the volume of P2P Credit at a 10% level of significance. This might reflect a higher demand for alternative consumer credit.

Table B.1 and table B.2 present results with variables that capture the geographic heterogeneity of the innovation, measured by the quality of Internet connection and by the number of patents issued by each county. Since the variable which measures the number of patents is correlated to the level of education, we performed one specification excluding the level of education, and found that it is statistically significant and with a positive sign. Counties with density of patents that is one standard deviation above the average exhibit a higher volume of P2P Credit and an increase in the number of loans. Among the variables describing the quality of Internet, only broadband and mobile are statistically significant and have a negative sign only when the dependent variable is the volume of loans. High Internet quality and speed do not impact the number of P2P borrowers.

To compare the expansion patterns of different online platforms, we estimate the model separately for Prosper Marketplace and Lending Club. The results, presented in Table B.3, show that not all local characteristics play a similar role in the case of both online lenders. The market structure variables (HHI and Branches) played a similar role for the two platforms, whereas payday loan establishments have a strong and positive impact only on Prosper's volume of loans and a negative but small impact on the number of Lending Club borrowers. Moreover, the leverage ratio during the crisis played a role in the case of Prosper but is not significant for

Lending Club. Interestingly, broadband access plays a positive role for the Prosper credit, and a negative one for Lending Club volume of loans. To understand this difference, one should remember that Proper platform had an earlier start than the Lending Club. A large part of the Prosper's credit in our sample has been done in 2006-2008 and it has experienced a sharp decline in 2008-2009 due to regulatory uncertainty about its legal status, followed by a slow expansion since 2010. The finding that broadband access plays a role for the Prosper credit is likely to reflect this earlier period when there was still an important geographic heterogeneity in access to Internet. This intuition is reinforced by the estimates of the SARAR model regressions performed each year separately, as shown in table B.4. As a matter of fact, the negative and significant effect of broadband is present only starting from the year 2012, whereas it is positive and significant on the year 2008 and otherwise it is never significant.

The age structure only plays a role for Lending Club: a higher percentage of population aged between 20 and 34 increases the volume of P2P loans but decreases the number of loans. With respect to the minorities, counties with a higher share of Hispanic population have a higher number of P2P loans on both platforms but only a higher volume of Lending Club loans.

Finally, the spatial lag is always positive and significant in all the regressions, suggesting the presence of positive spatial relations among contingent counties. It is interesting to note from table B.4, that, starting from 2008, this coefficient increased systematically during the years, going from 0.3777 in 2008 to 0.915 in 2013.

**Computing Marginal Effects**

Following the method proposed by Drukker et al. (2013), we manually compute the average total direct and indirect impacts of the explanatory variables (crisis, competition, innovation and socio-economic and demographic variables) on the dependent variable (either volume or number of P2P loans per capita per county) using the reduced-form predictors coming from the SARAR regression. Doing so allows us to understand the magnitude of these effects. For example, as shown in table B.6, an increase by one standard deviation of the number of branches in a given region decreases the average volume of P2P Credit per capita of all regions by 0.0013% (ATDI). Similarly, an increase by one standard deviation of the number of branches in all neighboring regions, reduces by 0.0004 % the volume of P2P Credit per capita in that one region (ATII). The signs of the coefficients are the same as the short-run impacts shown in tables 3.3, 3.4, B.1, B.2 and B.3, and in general the direct impacts are stronger than the indirect ones, which leads to the fact that total impacts are composed mainly by direct impacts in our main sample.

## 3.5 Concluding Remarks and Future Extensions

This paper is a first attempt to explore the drivers of the expansion of online lenders. We have proposed three hypotheses related to (1) the competition in the brick-and-mortar banking sector and switching costs to online lenders, (2) the consequences of the financial crisis and (3) the innovation and internet expansion. We also account for spatial effects and socio-economic and demographic characteristics.

Our findings suggest that online lenders have made inroads into counties that have a poor branch network. This suggests that borrowers that either live far away

**Table 3.1:** Growth of P2P Credit

| | Lending Club | | | Prosper | | |
|---|---|---|---|---|---|---|
| **Year** | **Vol. ($M)** | **# Counties** | **# Loans** | **Vol. ($M)** | **# Counties** | **# Loans** |
| 2006 | 0 | 0 | 0 | 29 | 673 | 6,145 |
| 2007 | 2 | 110 | 246 | 81 | 1,175 | 11,592 |
| 2008 | 13 | 379 | 1,488 | 69 | 1,377 | 11,683 |
| 2009 | 46 | 676 | 4,500 | 9 | 631 | 2,118 |
| 2010 | 116 | 987 | 10,594 | 27 | 1,029 | 5,864 |
| 2011 | 257 | 1,359 | 19,861 | 75 | 1,397 | 11,508 |
| 2012 | 718 | 1,836 | 49,811 | 154 | 1,739 | 20,054 |
| 2013 | 2,064 | 2,384 | 137,824 | 217 | 1,721 | 21,990 |

from a brick and mortar bank branch or have a poor branch experience due to long waiting times are more likely to turn to online lenders due to lower brand loyalty. We also find that counties with a more concentrated banking structure have witnessed slower growth of online lenders, which is also consistent with the idea of higher brand loyalty. Higher education and higher propensity to innovate play a significant and positive role, possibly because these characteristics diminish the costs of learning about online lenders. Our results show that the leverage ratio during the crisis has affected the demand for online credit. Despite the online nature of the P2P Credit, spatial effects play a crucial role, which could be interpreted as an important role of social interactions in building trust in online markets.

Our analysis could be extended in a number of ways. First, we would like to use the panel nature of the data to estimate Bass model of the innovation diffusion. Second, we would like to explore the balancing of demand and supply in the P2P Credit. This is possible due to the information in our dataset about loan demand that has not been met because loans have been rejected by online lenders or have failed to attract potential lenders.

**Figure 3.1:** P2P Credit growth in the US (in billions of dollars)

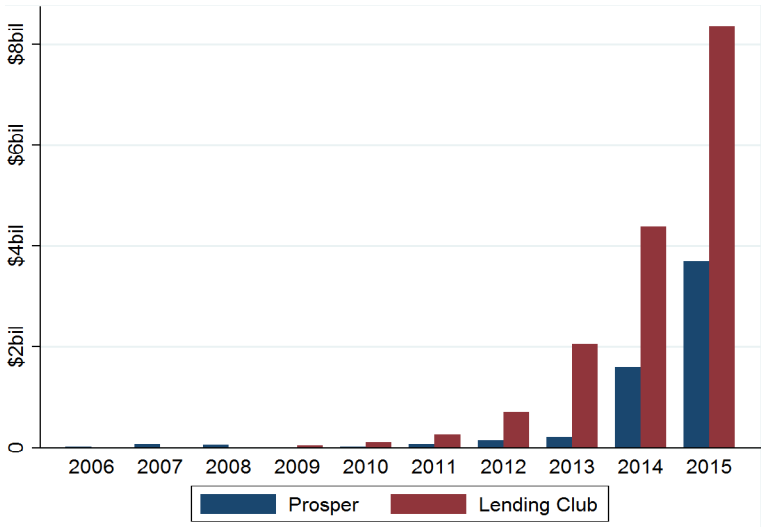



**Figure 3.2:** Total consumer loans in the USA in billions of dollars

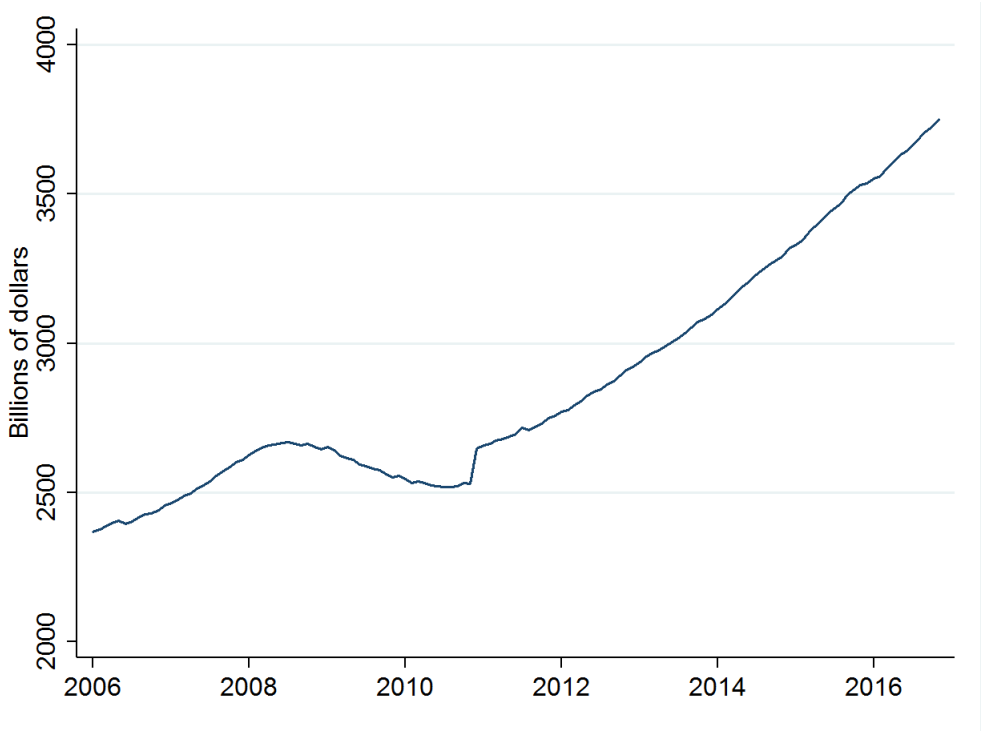


**Table 3.2:** Summary Statistics

| **Variable** | **Obs** | **Mean** | **Std.Dev.** | **Min** | **Max** |
|---|---|---|---|---|---|
| Prosper volume | 3,059.0 | 13,930.0 | 28,786.0 | 0.00 | 777,512.0 |
| Lending Club volume | 3,059.0 | 81,080.0 | 147,689.0 | 0.00 | 4,517,468.0 |
| Volume of P2P loans | 3,059.0 | 95,010.0 | 171,766.0 | 0.00 | 5,294,980.0 |
| Number of P2P loans | 3,059.0 | 6.0 | 11.6 | 0.00 | 451.3 |
| **Crisis variables** | | | | | |
| Failed | 3,059.0 | 0.02 | 0.08 | 0.00 | 1.00 |
| Crisis Tier1 | 3,059.0 | 0.14 | 0.08 | 0.06 | 3.99 |
| Crisis leverage | 3,059.0 | 0.09 | 0.02 | 0.04 | 0.33 |
| **Competition variables** | | | | | |
| C3 | 3,059.0 | 0.77 | 0.19 | 0.28 | 1.00 |
| HHI | 3,059.0 | 0.31 | 0.21 | 0.05 | 1.00 |
| Branches | 3,059.0 | 15.68 | 17.18 | 0.61 | 216.74 |
| Payday | 3,059.0 | 1.01 | 1.25 | 0.00 | 8.67 |
| **Innovation variables** | | | | | |
| Mobile | 3,059.0 | 0.95 | 0.11 | 0.00 | 1.00 |
| Broadband | 3,059.0 | 0.98 | 0.05 | 0.01 | 1.00 |
| Speed50000k | 3,059.0 | 0.42 | 0.35 | 0.00 | 1.00 |
| Speed10000k | 3,059.0 | 0.23 | 0.35 | 0.00 | 1.00 |
| Patents | 3,059.0 | 8.60 | 19.32 | 0.00 | 372.86 |
| Other variables | | | | | |
| Density | 3,059.0 | 77.00 | 473.00 | 0.00 | 18,354.00 |
| Age 20 to 34 | 3,059.0 | 0.19 | 0.02 | 0.09 | 0.32 |
| Bachelor | 3,059.0 | 0.17 | 0.08 | 0.04 | 0.61 |
| Income | 3,059.0 | 34,733.90 | 8,860.97 | 14,885.43 | 158,212.10 |
| Poverty | 3,059.0 | 0.17 | 0.06 | 0.03 | 0.50 |
| Asian | 3,059.0 | 0.01 | 0.02 | 0.00 | 0.58 |
| Hispanic | 3,059.0 | 0.05 | 0.08 | 0.00 | 0.49 |
| Black | 3,059.0 | 0.08 | 0.15 | 0.00 | 0.88 |

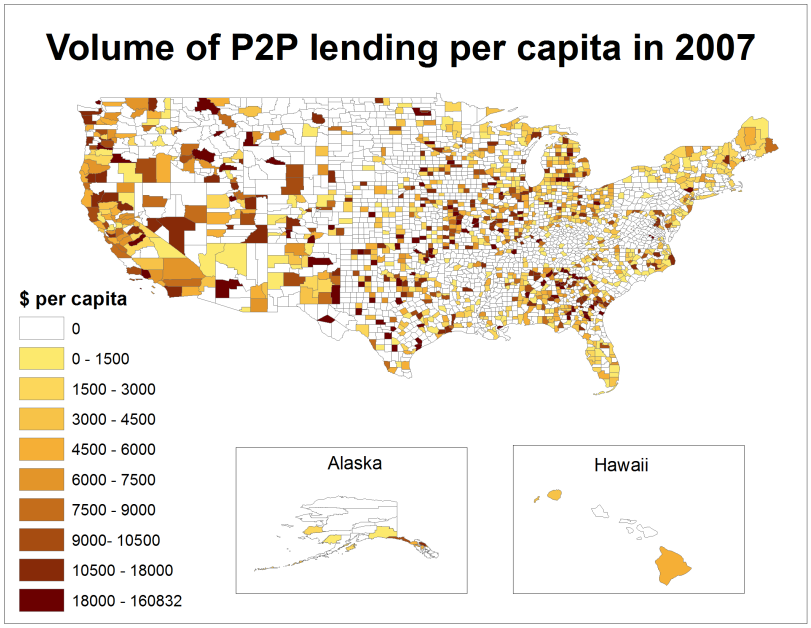
**Table 3.3:** SARAR model of loans per capita as a dependent variable

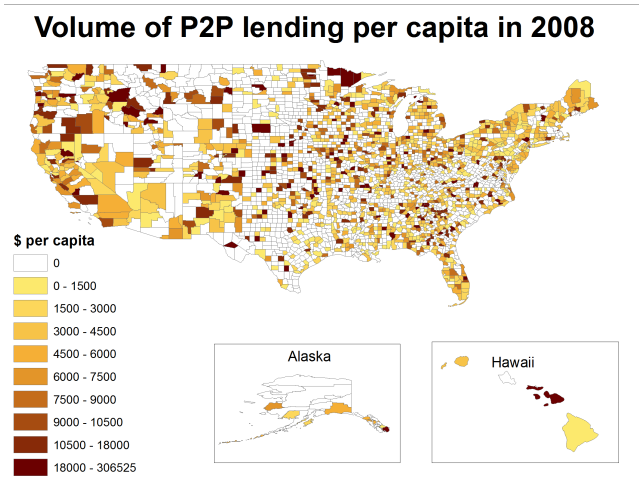
| | | | | | | |
|---|---|---|---|---|---|---|
| Competition variables | | | | | | |
| Branches | -0.0138*** | -0.0135*** | -0.0149*** | -0.0150*** | -0.0150*** | -0.0139*** |
| | -0.004 | -0.004 | -0.004 | -0.004 | -0.004 | -0.004 |
| C3 | -0.800** | | | | | |
| | -0.390 | | | | | |
| HHI | | -1.997*** | -2.029*** | -2.035*** | -2.038*** | |
| | | -0.330 | -0.331 | -0.331 | -0.331 | |
| Payday | 0.119** | 0.0848* | 0.0898* | 0.0900* | 0.0900* | 0.134*** |
| | -0.048 | -0.048 | -0.048 | -0.048 | -0.048 | -0.048 |
| Crisis variables | | | | | | |
| Crisis leverage | -11.31*** | -10.75*** | | | | -11.55*** |
| | -3.442 | -3.426 | | | | -3.442 |
| Capital_crisis | | | -0.441 | | | |
| | | | -1.030 | | | |
| Tier1_crisis | | | | -0.084 | | |
| | | | | -0.663 | | |
| Failed banks | | | | | 0.018 | |
| | | | | | -0.558 | |
| Other variables | | | | | | |
| Density_log | 0.546*** | 0.466*** | 0.480*** | 0.480*** | 0.480*** | 0.592*** |
| | -0.047 | -0.046 | -0.046 | -0.046 | -0.046 | -0.041 |
| Broadband | -3.626*** | -4.395*** | -4.532*** | -4.534*** | -4.536*** | -3.495*** |
| | -0.918 | -0.924 | -0.924 | -0.924 | -0.925 | -0.916 |
| Income_log | -1.755*** | -1.914*** | -1.984*** | -1.989*** | -1.989*** | -1.691*** |
| | -0.412 | -0.411 | -0.411 | -0.411 | -0.412 | -0.411 |
| Poverty | -7.070*** | -6.524*** | -6.649*** | -6.662*** | -6.664*** | -7.387*** |
| | -1.425 | -1.417 | -1.419 | -1.419 | -1.419 | -1.417 |
| Bachelor | 2.672*** | 3.291*** | 3.581*** | 3.598*** | 3.599*** | 2.830*** |
| | -1.001 | -0.996 | -0.993 | -0.993 | -0.993 | -0.998 |
| Black | -0.240 | 0.049 | 0.043 | 0.043 | 0.043 | -0.399 |
| | -0.461 | -0.458 | -0.458 | -0.459 | -0.459 | -0.454 |

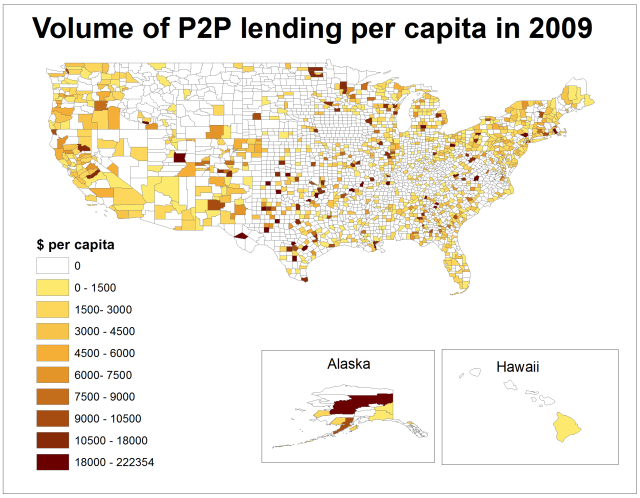
**Table 3.4:** SARAR model of loans per capita as a dependent variable (continued))

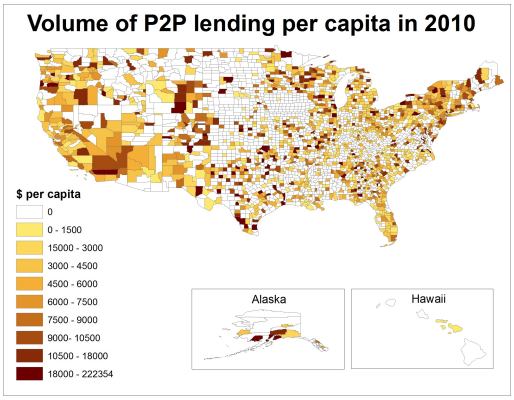
| | | | | | | |
|---|---|---|---|---|---|---|
| Hispanic | 6.362*** | 6.583*** | 6.691*** | 6.684*** | 6.677*** | 6.539*** |
| | -1.045 | -1.037 | -1.038 | -1.038 | -1.042 | -1.042 |
| Age 20 to 34 | 4.058 | 2.225 | 2.053 | 2.061 | 2.067 | 4.474 |
| | -3.133 | -3.133 | -3.137 | -3.137 | -3.142 | -3.128 |
| Constant | 31.93*** | 34.36*** | 34.37*** | 34.38*** | 34.37*** | 30.52*** |
| | -4.554 | -4.525 | -4.532 | -4.532 | -4.551 | -4.504 |
| Lambda | 0.571*** | 0.557*** | 0.558*** | 0.557*** | 0.557*** | 0.581*** |
| | -0.035 | -0.034 | -0.035 | -0.035 | -0.035 | -0.034 |
| Sigma2 | 8.078*** | 7.998*** | 8.023*** | 8.024*** | 8.024*** | 8.085*** |
| | -0.207 | -0.205 | -0.205 | -0.206 | -0.206 | -0.207 |
| Number of counties | 3,059 | 3,059 | 3,059 | 3,059 | 3,059 | 3,059 |
| State dummies | Yes | Yes | Yes | Yes | Yes | Yes |

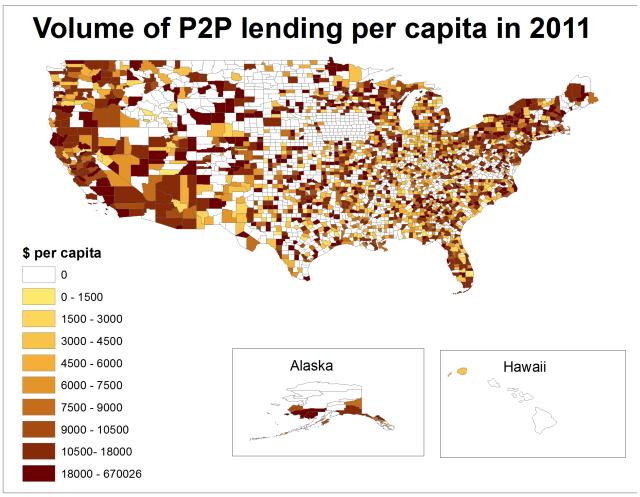
Notes: We estimate cross-sectional models of the geographic expansion of the P2P Credit during the period 2006-2013. Dependent variable is the amount of P2P Credit per capital in a county. Variable definitions are provided in Table 2. Models are estimated with maximum likelihood approach while controlling for the spatial dependence with a spatial lag term (lambda). State dummies are not shown. Standard errors are in parentheses. ***, **, * denote significance at the 1, 5 and 10 percent level, respectively.

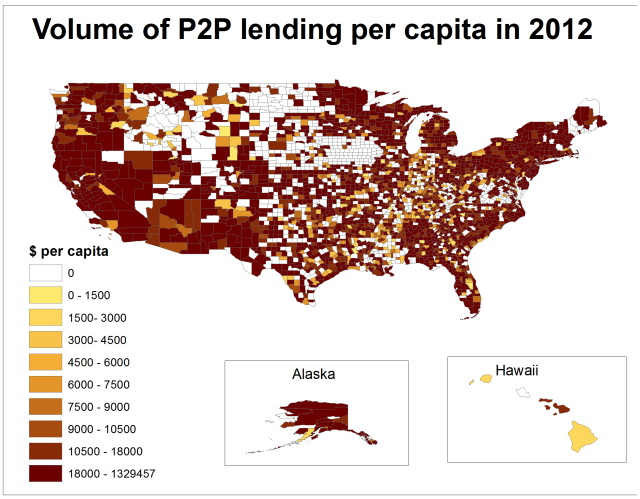


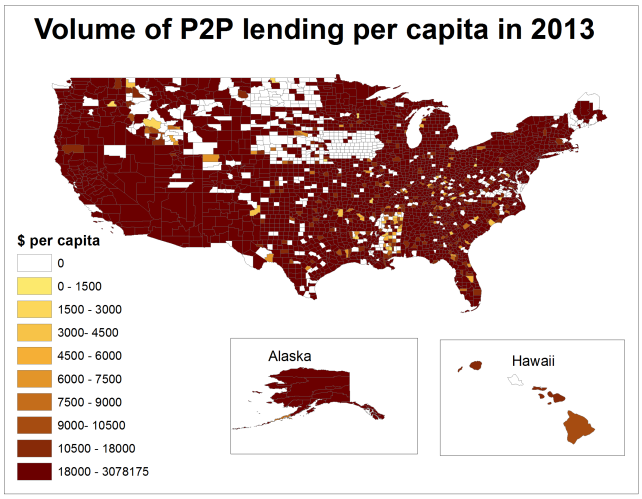

Volume of P2P lending per capita in 2008
$ per capita
0
0 - 1500
1500 - 3000
3000 - 4500
4500 - 6000
6000 - 7500
7500 - 9000
9000 - 10500
10500 - 18000
18000 - 306525
Alaska
Hawaii


Volume of P2P lending per capita in 2009
$ per capita
0
0 - 1500
1500- 3000
3000 - 4500
4500 - 6000
6000- 7500
7500 - 9000
9000 - 10500
10500 - 18000
18000 - 222354
Alaska
Hawaii


Volume of P2P lending per capita in 2010
$ per capita
0
0 - 1500
15000 - 3000
3000 - 4500
4500 - 6000
6000 - 7500
7500 - 9000
9000- 10500
10500 - 18000
18000 - 222354
Alaska
Hawaii


Volume of P2P lending per capita in 2011
$ per capita
0
0 - 1500
1500 - 3000
3000 - 4500
4500 - 6000
6000 - 7500
7500 - 9000
9000 - 10500
10500- 18000
18000 - 670026
Alaska
Hawaii


Volume of P2P lending per capita in 2012
$ per capita
0
0 - 1500
1500- 3000
3000- 4500
4500 - 6000
6000 - 7500
7500 - 9000
9000 - 10500
10500 - 18000
18000 - 1329457
Alaska
Hawaii


Volume of P2P lending per capita in 2013
$ per capita
0
0 - 1500
1500 - 3000
3000- 4500
4500 - 6000
6000 - 7500
7500 - 9000
9000- 10500
10500 - 18000
18000 - 3078175
Alaska
Hawaii

# Appendix A

# Additional Tables for Chapter 2

**Table A.1:** Matching Efficiency

| Variable | Group | Treated | Control | %bias | t-test | p¿t |
|---|---|---|---|---|---|---|
| Interest rate | Unmatched | 0.22 | 0.23 | -3.80 | -3.06 | 0.00 |
| | Matched | 0.22 | 0.22 | -0.90 | -0.66 | 0.51 |
| Loan term | Unmatched | 43.35 | 42.96 | 3.00 | 2.41 | 0.02 |
| | Matched | 43.35 | 42.92 | 3.30 | 2.65 | 0.01 |
| Listing category | Unmatched | 4.83 | 5.54 | -12.50 | -9.98 | 0.00 |
| | Matched | 4.83 | 5.05 | -3.80 | -3.19 | 0.00 |
| # of credit inquiries | Unmatched | 3.95 | 4.08 | -3.50 | -2.77 | 0.01 |
| | Matched | 3.95 | 4.10 | -4.10 | -3.33 | 0.00 |
| # of traded items | Unmatched | 25.43 | 24.96 | 3.50 | 2.79 | 0.01 |
| | Matched | 25.43 | 25.33 | 0.80 | 0.61 | 0.54 |
| # of satisfactory accounts | Unmatched | 22.90 | 22.33 | 4.30 | 3.48 | 0.00 |
| | Matched | 22.90 | 22.76 | 1.10 | 0.86 | 0.39 |
| # of Delinq. Accts. Now | Unmatched | 0.40 | 0.42 | -1.00 | -0.81 | 0.42 |
| | Matched | 0.40 | 0.39 | 1.30 | 1.10 | 0.27 |
| # of Delinq. Accts. Before | Unmatched | 2.13 | 2.21 | -2.40 | -1.95 | 0.05 |
| | Matched | 2.13 | 2.18 | -1.60 | -1.31 | 0.19 |
| Delinquencies ¿ 30 days | Unmatched | 3.95 | 4.15 | -2.70 | -2.17 | 0.03 |
| | Matched | 3.95 | 3.91 | 0.50 | 0.38 | 0.70 |
| Delinquencies ¿ 60 days | Unmatched | 1.72 | 1.77 | -1.40 | -1.08 | 0.28 |
| | Matched | 1.72 | 1.76 | -0.90 | -0.68 | 0.50 |
| Delinquencies ¿ 90 days | Unmatched | 3.74 | 3.72 | 0.20 | 0.19 | 0.85 |
| | Matched | 3.74 | 3.67 | 0.70 | 0.58 | 0.56 |
| Install. Balance | Unmatched | 24865.00 | 23589.00 | 3.50 | 2.76 | 0.01 |
| | Matched | 24865.00 | 24768.00 | 0.30 | 0.20 | 0.84 |

**Table A.2:** Matching Efficiency (Continued)

| Variable | Group | Treated | Control | %bias | t-test | p¿t |
|---|---|---|---|---|---|---|
| Real Estate Balance | Unmatched | 100,000 | 110,000 | -2.30 | -1.80 | 0.07 |
| | Matched | 100,000 | 110,000 | -1.00 | -0.85 | 0.40 |
| Revolving Balance | Unmatched | 18,523 | 17,208 | 3.80 | 3.02 | 0.00 |
| | Matched | 18,523 | 18,730 | -0.60 | -0.45 | 0.65 |
| Real Estate Pmt. | Unmatched | 788.82 | 813.19 | -2.20 | -1.74 | 0.08 |
| | Matched | 788.82 | 796.97 | -0.70 | -0.60 | 0.55 |
| Rev. Bal. Available | Unmatched | 54.37 | 54.62 | -0.90 | -0.69 | 0.49 |
| | Matched | 54.37 | 54.42 | -0.20 | -0.16 | 0.88 |
| Pub. Records 10 Yrs. | Unmatched | 0.25 | 0.27 | -2.20 | -1.80 | 0.07 |
| | Matched | 0.25 | 0.26 | -1.10 | -0.92 | 0.36 |
| Pub. Records 12 Yrs. | Unmatched | 0.01 | 0.01 | -2.30 | -1.83 | 0.07 |
| | Matched | 0.01 | 0.01 | -1.80 | -1.50 | 0.14 |
| Inquiries in 6 mos. | Unmatched | 1.16 | 1.21 | -3.30 | -2.62 | 0.01 |
| | Matched | 1.16 | 1.20 | -2.90 | -2.31 | 0.02 |
| Bank card trades w/ 6 | Unmatched | 3.94 | 3.73 | 8.20 | 6.53 | 0.00 |
| | Matched | 3.94 | 3.88 | 2.30 | 1.81 | 0.07 |
| Avail. Credit w/6 | Unmatched | 11991 | 11750 | 1.10 | 0.91 | 0.37 |
| | Matched | 11991 | 11845 | 0.70 | 0.61 | 0.54 |
| Bal. on bank cards w/6 | Unmatched | 49.23 | 47.73 | 4.70 | 3.75 | 0.00 |
| | Matched | 49.23 | 48.30 | 2.90 | 2.36 | 0.02 |
| Bal. on open trades | Unmatched | 19545 | 18484 | 3.60 | 2.85 | 0.00 |
| | Matched | 19545 | 19376 | 0.60 | 0.44 | 0.66 |
| # of Real Estate Trades | Unmatched | 1.64 | 1.62 | 0.80 | 0.66 | 0.51 |
| | Matched | 1.64 | 1.64 | -0.20 | -0.15 | 0.88 |
| # of propert trades | Unmatched | 2.18 | 2.20 | -0.70 | -0.54 | 0.59 |
| | Matched | 2.18 | 2.18 | -0.10 | -0.05 | 0.96 |
| # of Derog. Trades | Unmatched | 1.01 | 1.05 | -1.70 | -1.39 | 0.16 |
| | Matched | 1.01 | 1.02 | -0.40 | -0.32 | 0.75 |

**Table A.3:** Matching Efficiency (Continued)

| Variable | Group | Treated | Control | %bias | t-test | p¿t |
|---|---|---|---|---|---|---|
| Monthly Debt | Unmatched | 856.10 | 820.45 | 4.30 | 3.46 | 0.00 |
| | Matched | 856.10 | 853.39 | 0.30 | 0.28 | 0.78 |
| # of trades | Unmatched | 21.94 | 21.51 | 3.60 | 2.87 | 0.00 |
| | Matched | 21.94 | 21.86 | 0.70 | 0.56 | 0.58 |
| # of open trades w/6 | Unmatched | 0.80 | 0.77 | 2.10 | 1.66 | 0.10 |
| | Matched | 0.80 | 0.79 | 0.70 | 0.53 | 0.60 |
| # of rep. open trades | Unmatched | 8.30 | 7.80 | 10.50 | 8.40 | 0.00 |
| | Matched | 8.30 | 8.17 | 2.90 | 2.27 | 0.02 |
| # of trades w/ due bal. | Unmatched | 0.14 | 0.14 | -1.20 | -0.95 | 0.34 |
| | Matched | 0.14 | 0.13 | 1.10 | 0.95 | 0.34 |
| # of open trades w/ due bal. | Unmatched | 0.04 | 0.04 | -0.60 | -0.51 | 0.61 |
| | Matched | 0.04 | 0.04 | 0.00 | 0.00 | 1.00 |
| # of trades reptd. | Unmatched | 9.29 | 8.82 | 8.90 | 7.13 | 0.00 |
| | Matched | 9.29 | 9.18 | 2.00 | 1.62 | 0.11 |
| # of trades reptd. Derog. | Unmatched | 0.16 | 0.16 | -1.30 | -1.02 | 0.31 |
| | Matched | 0.16 | 0.16 | -0.30 | -0.23 | 0.82 |
| # of trades 90 ¿ derog. | Unmatched | 1.22 | 1.27 | -2.30 | -1.81 | 0.07 |
| | Matched | 1.22 | 1.22 | -0.40 | -0.29 | 0.77 |
| Ag. Bal. open trades | Unmatched | 130,000 | 140,000 | -0.40 | -0.32 | 0.75 |
| | Matched | 130,000 | 140,000 | -0.50 | -0.43 | 0.67 |
| Ag. Mon. pmt. On trades | Unmatched | 1548.40 | 1530.40 | 1.20 | 0.96 | 0.34 |
| | Matched | 1548.40 | 1556.80 | -0.60 | -0.45 | 0.66 |
| Age of oldest trade | Unmatched | 213.20 | 207.79 | 5.40 | 4.33 | 0.00 |
| | Matched | 213.20 | 210.49 | 2.70 | 2.17 | 0.03 |
| # of inquiries w/6 | Unmatched | 0.86 | 0.89 | -2.50 | -1.98 | 0.05 |
| | Matched | 0.86 | 0.89 | -2.50 | -2.02 | 0.04 |
| % of trades no. delinq. | Unmatched | 89.89 | 89.26 | 5.00 | 3.98 | 0.00 |
| | Matched | 89.89 | 89.64 | 2.00 | 1.63 | 0.10 |

**Table A.4:** Matching Efficiency (Continued)

| **Variable** | **Group** | **Treated** | **Control** | **%bias** | **t-test** | **p¿t** |
|---|---|---|---|---|---|---|
| Income Range | Unmatched | 4.09 | 4.09 | -0.10 | -0.11 | 0.91 |
| | Matched | 4.09 | 4.09 | -0.30 | -0.23 | 0.82 |
| Debt /Income w. Prosper | Unmatched | 120,000 | 130,000 | -3.80 | -3.00 | 0.00 |
| | Matched | 120,000 | 130,000 | -2.50 | -2.06 | 0.04 |
| Stated Mon. Income | Unmatched | 6,653 | 6,234 | 1.10 | 0.86 | 0.39 |
| | Matched | 6,653 | 7,525 | -2.20 | -1.00 | 0.32 |
| Empl. Length | Unmatched | 99.42 | 97.45 | 2.00 | 1.61 | 0.11 |
| | Matched | 99.42 | 98.28 | 1.20 | 0.94 | 0.35 |
| Prior Loans Outs. | Unmatched | 623.12 | 560.25 | 3.20 | 2.52 | 0.01 |
| | Matched | 623.12 | 718.68 | -4.80 | -3.67 | 0.00 |
| Prior Bal. Outs. | Unmatched | 628.72 | 565.33 | 3.20 | 2.52 | 0.01 |
| | Matched | 628.72 | 725.27 | -4.80 | -3.68 | 0.00 |
| Prior Late loans | Unmatched | 0.145 | 0.179 | -2.40 | -1.90 | 0.06 |
| | Matched | 0.145 | 0.183 | -2.70 | -2.19 | 0.03 |
| Prior Late Pmts. | Unmatched | 0.010 | 0.018 | -2.40 | -1.91 | 0.06 |
| | Matched | 0.010 | 0.009 | 0.20 | 0.23 | 0.82 |
| Prior cycles 31+ | Unmatched | 0.003 | 0.006 | -4.00 | -3.24 | 0.00 |
| | Matched | 0.003 | 0.003 | -0.30 | -0.32 | 0.75 |
| Prior cycles 61+ | Unmatched | 0.001 | 0.002 | -2.20 | -1.79 | 0.07 |
| | Matched | 0.001 | 0.001 | 0.00 | 0.00 | 1.00 |
| Estimated Loss Rate | Unmatched | 0.093 | 0.092 | 1.70 | 1.38 | 0.17 |
| | Matched | 0.093 | 0.094 | -2.40 | -1.84 | 0.07 |

# Appendix B

# Additional Tables for Chapter 3

**Table B.1:** SARAR model with with number of loans per capita as dependent variable (continued))

| | | | | | | |
|---|---|---|---|---|---|---|
| Black | 0.346*** | 0.355*** | 0.355*** | 0.355*** | 0.351*** | 0.291** |
| | -0.119 | -0.12 | -0.12 | -0.12 | -0.12 | -0.119 |
| Hispanic | 0.982*** | 1.044*** | 1.051*** | 1.052*** | 1.038*** | 1.020*** |
| | -0.195 | -0.195 | -0.195 | -0.195 | -0.196 | -0.195 |
| Constant | 2.596** | 2.644** | 2.650** | 2.651** | 2.548** | 2.102* |
| | -1.197 | -1.196 | -1.197 | -1.197 | -1.201 | -1.19 |
| Lambda | 0.960*** | 0.964*** | 0.965*** | 0.966*** | 0.964*** | 0.973*** |
| | -0.0402 | -0.0401 | -0.0401 | -0.0401 | -0.0401 | -0.04 |
| Sigma2 | 0.577*** | 0.576*** | 0.577*** | 0.577*** | 0.577*** | 0.579*** |
| | -0.0149 | -0.0149 | -0.0149 | -0.0149 | -0.0149 | -0.0149 |
| Number of counties | 3,059 | 3,059 | 3,059 | 3,059 | 3,059 | 3,059 |
| State dummies | Yes | Yes | Yes | Yes | Yes | Yes |

Notes: We estimate cross-sectional models of the geographic expansion of the P2P Credit during the period 2006-2013. Dependant variable is the amount of P2P Credit per capital in a county. Variable definitions are provided in Table 2. Models are estimated with maximum likelihood approach while controlling for the spatial dependence with a spatial lag term (lambda). State dummies are not shown. Standard errors are in parentheses. ***, **, * denote significance at the 1, 5 and 10 percent level, respectively.

**Table B.2:** SARAR model with with number of loans per capita as dependent variable

| | | | | | | |
|---|---|---|---|---|---|---|
| Competition variables | | | | | | |
| Branches | -0.00301*** | -0.00295*** | -0.00321*** | -0.00319*** | -0.00314*** | -0.00278*** |
| | -0.000997 | -0.000995 | -0.000992 | -0.000992 | -0.000989 | -0.000996 |
| C3 | -0.346*** | | | | | |
| | -0.1 | | | | | |
| HHI | | -0.329*** | -0.337*** | -0.336*** | -0.341*** | |
| | | -0.0865 | -0.0866 | -0.0866 | -0.0866 | |
| Payday | -0.0105 | -0.0123 | -0.0113 | -0.0113 | -0.0107 | -0.00307 |
| | -0.0128 | -0.0128 | -0.0128 | -0.0128 | -0.0129 | -0.0126 |
| Crisis variables | | | | | | |
| Crisis leverage | -1.813** | -1.789* | | | | -1.944** |
| | -0.921 | -0.92 | | | | -0.921 |
| Capital_crisis | | | 0.0483 | | | |
| | | | -0.276 | | | |
| Tier1_crisis | | | | -0.00377 | | |
| | | | | -0.178 | | |
| Failed banks | | | | | 0.155 | |
| | | | | | -0.15 | |
| Other Variables | | | | | | |
| Density | 0.0557*** | 0.0542*** | 0.0563*** | 0.0563*** | 0.0559*** | 0.0717*** |
| | -0.011 | -0.011 | -0.0109 | -0.0109 | -0.0109 | -0.00996 |
| Broadband | -0.192 | -0.324 | -0.352 | -0.352 | -0.37 | -0.0818 |
| | -0.32 | -0.325 | -0.325 | -0.325 | -0.325 | -0.319 |
| Income_log | -0.0936 | -0.101 | -0.114 | -0.114 | -0.103 | -0.0835 |
| | -0.108 | -0.108 | -0.108 | -0.108 | -0.108 | -0.108 |
| Poverty | -2.146*** | -2.157*** | -2.181*** | -2.180*** | -2.167*** | -2.317*** |
| | -0.384 | -0.383 | -0.383 | -0.383 | -0.383 | -0.381 |
| Bachelor | 1.486*** | 1.620*** | 1.674*** | 1.672*** | 1.661*** | 1.595*** |
| | -0.257 | -0.255 | -0.254 | -0.253 | -0.254 | -0.255 |
| Age 20 to 34 | -0.174 | -0.348 | -0.389 | -0.389 | -0.343 | 0.0615 |
| | -0.778 | -0.782 | -0.782 | -0.782 | -0.784 | -0.776 |

**Table B.3:** SARAR model with volume of loans per capita as dependent variable

| | | | | | | |
|---|---|---|---|---|---|---|
| Innovation variables | | | | | | |
| Patents_log | 0.118** | | | | | |
| | -0.0566 | | | | | |
| Broadband | | -4.395*** | | | | |
| | | -0.924 | | | | |
| Optical fiber | | | -0.348 | | | |
| | | | -0.255 | | | |
| Mobile | | | | -1.278*** | | |
| | | | | -0.348 | | |
| Speed10000k | | | | | -0.0316 | |
| | | | | | -0.168 | |
| Speed50000k | | | | | | 0.0367 |
| | | | | | | -0.169 |
| Other variables | | | | | | |
| Branches | -0.0121*** | -0.0135*** | -0.0126*** | -0.0135*** | -0.0131*** | -0.0132*** |
| | -0.00382 | -0.00382 | -0.00385 | -0.00382 | -0.00385 | -0.00384 |
| HHI | -1.630*** | -1.997*** | -1.738*** | -1.978*** | -1.744*** | -1.744*** |
| | -0.327 | -0.33 | -0.327 | -0.333 | -0.327 | -0.327 |
| Payday | 0.0668 | 0.0848* | 0.0685 | 0.0837* | 0.0702 | 0.0689 |
| | -0.0482 | -0.048 | -0.048 | -0.0481 | -0.0481 | -0.0481 |
| Crisis levergae | -12.19*** | -10.75*** | -11.34*** | -11.70*** | -11.55*** | -11.49*** |
| | -3.431 | -3.426 | -3.437 | -3.428 | -3.436 | -3.441 |
| Density | 0.449*** | 0.466*** | 0.446*** | 0.465*** | 0.449*** | 0.444*** |
| | -0.0476 | -0.0461 | -0.0461 | -0.0462 | -0.0468 | -0.0488 |
| Income_log | -1.460*** | -1.914*** | -1.899*** | -1.955*** | -1.962*** | -1.964*** |
| | -0.371 | -0.411 | -0.415 | -0.411 | -0.412 | -0.412 |
| Poverty | -5.722*** | -6.524*** | -5.865*** | -6.481*** | -5.882*** | -5.852*** |
| | -1.431 | -1.417 | -1.415 | -1.422 | -1.417 | -1.417 |
| Bachelor | | 3.291*** | 3.454*** | 3.313*** | 3.439*** | 3.392*** |
| | | -0.996 | -0.999 | -0.997 | -1.007 | -1.004 |

**Table B.4:** SARAR model with volume of loans per capita as dependent variable (continued)

| | | | | | | |
|---|---|---|---|---|---|---|
| Age 20 to 34 | 4.916* | 2.225 | 2.075 | 2.552 | 1.944 | 1.927 |
| | -2.981 | -3.133 | -3.145 | -3.142 | -3.145 | -3.148 |
| Black | -0.0367 | 0.0493 | -0.107 | 0.0653 | -0.115 | -0.135 |
| | -0.463 | -0.458 | -0.458 | -0.46 | -0.462 | -0.46 |
| Hispanic | 6.690*** | 6.583*** | 6.419*** | 6.668*** | 6.451*** | 6.424*** |
| | -1.04 | -1.037 | -1.04 | -1.04 | -1.045 | -1.041 |
| Constant | 25.08*** | 34.36*** | 29.82*** | 31.71*** | 30.50*** | 30.51*** |
| | -4.048 | -4.525 | -4.494 | -4.471 | -4.468 | -4.468 |
| Lambda | 0.550*** | 0.557*** | 0.549*** | 0.548*** | 0.551*** | 0.550*** |
| | -0.0346 | -0.0344 | -0.0346 | -0.0345 | -0.0345 | -0.0345 |
| Sigma2 | 8.079*** | 7.998*** | 8.055*** | 8.025*** | 8.059*** | 8.059*** |
| | -0.207 | -0.205 | -0.206 | -0.206 | -0.206 | -0.206 |
| Number of counties | 3,059 | 3,059 | 3,059 | 3,059 | 3,059 | 3,059 |
| State dummies | Yes | Yes | Yes | Yes | Yes | Yes |

Notes: We estimate cross-sectional models of the geographic expansion of the P2P Credit during the period 2006-2013. Variable definitions are provided in Table 2. Models are estimated with maximum likelihood approach while controlling for the spatial dependence with a spatial lag term (lambda). State dummies are not shown. Standard errors are in parentheses. ***, **, * denote significance at the 1, 5 and 10 percent level, respectively.

**Table B.5:** SARAR model with with number of loans per capita as dependent variable

| | | | | | | |
|---|---|---|---|---|---|---|
| Innovation variables | | | | | | |
| Patents_log | 0.0252* | | | | | |
| | -0.0149 | | | | | |
| Broadband | | -0.324 | | | | |
| | | -0.325 | | | | |
| Optical fiber | | | -0.109 | | | |
| | | | -0.0663 | | | |
| Mobile | | | | -0.0313 | | |
| | | | | -0.148 | | |
| Speed10000k | | | | | -0.0412 | |
| | | | | | -0.0515 | |
| Speed50000k | | | | | | 0.0239 |
| | | | | | | -0.0517 |
| Other variables | | | | | | |
| Branches | -0.00229** | -0.00295*** | -0.00277*** | -0.00295*** | -0.00286*** | -0.00298*** |
| | -0.000996 | -0.000995 | -0.001 | -0.000995 | -0.001 | -0.000999 |
| HHI | -0.280*** | -0.329*** | -0.310*** | -0.316*** | -0.315*** | -0.307*** |
| | -0.0856 | -0.0865 | -0.0848 | -0.0871 | -0.085 | -0.0854 |
| Payday | -0.0161 | -0.0123 | -0.0137 | -0.0129 | -0.0136 | -0.0133 |
| | -0.0129 | -0.0128 | -0.0128 | -0.0129 | -0.0128 | -0.0128 |
| Crisis leverage | -2.348** | -1.789* | -1.782* | -1.828** | -1.872** | -1.786* |
| | -0.922 | -0.92 | -0.92 | -0.919 | -0.921 | -0.924 |
| Density | 0.0620*** | 0.0542*** | 0.0533*** | 0.0537*** | 0.0551*** | 0.0516*** |
| | -0.0113 | -0.011 | -0.0109 | -0.011 | -0.0111 | -0.0117 |
| Income_log | 0.178* | -0.101 | -0.0844 | -0.105 | -0.104 | -0.106 |
| | -0.0984 | -0.108 | -0.109 | -0.108 | -0.108 | -0.108 |
| Poverty | -2.119*** | -2.157*** | -2.098*** | -2.119*** | -2.132*** | -2.091*** |
| | -0.385 | -0.383 | -0.379 | -0.384 | -0.38 | -0.38 |
| Bachelor | | 1.620*** | 1.651*** | 1.629*** | 1.664*** | 1.615*** |
| | | -0.255 | -0.255 | -0.255 | -0.257 | -0.257 |

**Table B.6:** SARAR model with with number of loans per capita as dependent variable (continued)

| | | | | | | |
|---|---|---|---|---|---|---|
| Age 20 to 34 | 0.685 | -0.348 | -0.355 | -0.381 | -0.419 | -0.405 |
| | -0.766 | -0.782 | -0.781 | -0.784 | -0.781 | -0.781 |
| Black | 0.363*** | 0.355*** | 0.335*** | 0.344*** | 0.353*** | 0.334*** |
| | -0.12 | -0.12 | -0.119 | -0.12 | -0.12 | -0.119 |
| Hispanic | 1.116*** | 1.044*** | 1.017*** | 1.032*** | 1.044*** | 1.027*** |
| | -0.195 | -0.195 | -0.194 | -0.196 | -0.195 | -0.194 |
| Constant | -0.531 | 2.644** | 2.145* | 2.391** | 2.374** | 2.360** |
| | -1.072 | -1.196 | -1.172 | -1.169 | -1.164 | -1.164 |
| Lambda | 0.977*** | 0.964*** | 0.963*** | 0.963*** | 0.964*** | 0.962*** |
| | -0.0401 | -0.0401 | -0.0401 | -0.0401 | -0.0401 | -0.0401 |
| Sigma2 | 0.583*** | 0.576*** | 0.576*** | 0.577*** | 0.577*** | 0.577*** |
| | -0.015 | -0.0149 | -0.0149 | -0.0149 | -0.0149 | -0.0149 |
| Number of counties | 3,059 | 3,059 | 3,059 | 3,059 | 3,059 | 3,059 |
| State dummies | Yes | Yes | Yes | Yes | Yes | Yes |

Notes: We estimate cross-sectional models of the geographic expansion of the P2P Credit during the period 2006-2013. Variable definitions are provided in Table 2. Models are estimated with maximum likelihood approach while controlling for the spatial dependence with a spatial lag term (lambda). State dummies are not shown. Standard errors are in parentheses. ***, **, * denote significance at the 1, 5 and 10 percent level, respectively.

**Table B.7:** SARAR model for the expansion of Prosper and Lending Club

| | Vol. of loans per capita | | of loans per capita | |
|---|---|---|---|---|
| | Lending Club | Prosper | Lending Club | Prosper |
| Branches | -0.0177*** | -0.00816* | -0.00325*** | -0.00118 |
| | -0.00406 | -0.00471 | -0.00105 | -0.000865 |
| HHI | -1.157*** | -4.339*** | -0.284*** | -0.0352 |
| | -0.349 | -0.407 | -0.0915 | -0.0751 |
| Payday | -0.0796 | 0.273*** | -0.0281** | -0.0124 |
| | -0.051 | -0.0591 | -0.0136 | -0.0112 |
| Crisis leverage | -5.11 | -24.37*** | -0.657 | -1.546* |
| | -3.643 | -4.228 | -0.974 | -0.8 |
| Density | 0.540*** | 0.637*** | 0.019 | 0.0443*** |
| | -0.0494 | -0.0574 | -0.0116 | -0.00952 |
| Broadband | -6.414*** | 3.009*** | -0.546 | 0.382 |
| | -0.981 | -1.139 | -0.344 | -0.283 |
| Income_log | -0.601 | -2.197*** | -0.0349 | -0.230** |
| | -0.435 | -0.507 | -0.114 | -0.0934 |
| Poverty | -3.414** | -8.524*** | -2.068*** | -1.978*** |
| | -1.504 | -1.756 | -0.404 | -0.333 |
| Bachelor | 2.473** | 4.979*** | 0.995*** | 1.709*** |
| | -1.058 | -1.227 | -0.269 | -0.222 |
| Age 20 to 34 | -3.879 | 14.23*** | -0.227 | -1.290* |
| | -3.332 | -3.864 | -0.828 | -0.68 |
| Black | -0.447 | -0.0792 | 0.440*** | 0.408*** |
| | -0.487 | -0.566 | -0.127 | -0.104 |
| Hispanic | 6.677*** | 2.032 | 0.912*** | 0.627*** |
| | -1.106 | -1.275 | -0.207 | -0.169 |
| Constant | 19.80*** | 28.79*** | 1.825 | 2.829*** |
| | -4.797 | -5.587 | -1.259 | -1.034 |
| Lambda | 1.169*** | 0.378*** | 1.033*** | 0.679*** |
| | -0.0293 | -0.0446 | -0.0436 | -0.0525 |
| Sigma2 | 9.045*** | 12.16*** | 0.646*** | 0.436*** |
| | -0.233 | -0.311 | -0.0167 | -0.0112 |
| Number of counties | 3,059 | 3,059 | 3,059 | 3,059 |
| State dummies | Yes | Yes | Yes | Yes |

We estimate cross-sectional models of the geographic expansion of the P2P Credit during the period 2006-2013. Variable definitions are provided in Table 2. Models are estimated with maximum likelihood approach while controlling for the spatial dependence with a spatial lag term (lambda). State dummies are not shown. Standard errors are in parentheses. ***, **, * denote significance at the 1, 5 and 10 percent level, respectively.

**Table B.8:** SARAR model for the expansion of P2P Credit year by year

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Branches | -0.00557 | 0.0049 | 0.00236 | -0.0038 | -0.0109** | -0.0111** | -0.0166*** |
| | -0.00448 | -0.00457 | -0.00391 | -0.00451 | -0.00489 | -0.0048 | -0.00422 |
| HHI | -1.426*** | -1.981*** | -0.718** | -2.186*** | -2.923*** | -3.413*** | -2.180*** |
| | -0.388 | -0.396 | -0.34 | -0.39 | -0.421 | -0.415 | -0.365 |
| Payday | 0.0639 | 0.0706 | -0.00717 | 0.0668 | 0.148** | 0.109* | 0.0899* |
| | -0.0532 | -0.0559 | -0.0491 | -0.0571 | -0.062 | -0.0568 | -0.0512 |
| Crisis leverage | -18.14*** | -15.27*** | -8.714** | -13.22*** | -14.52*** | -11.41*** | -2.877 |
| | -4.052 | -4.129 | -3.546 | -4.075 | -4.421 | -4.339 | -3.826 |
| density_log | 0.561*** | 0.728*** | 0.678*** | 0.767*** | 0.787*** | 0.531*** | 0.530*** |
| | -0.0541 | -0.056 | -0.0481 | -0.0558 | -0.0611 | -0.0591 | -0.0523 |
| Broadband | 1.281 | 1.997* | -1.088 | -0.882 | -0.73 | -3.385*** | -5.488*** |
| | -1.093 | -1.11 | -0.955 | -1.096 | -1.187 | -1.169 | -1.03 |
| income_log | -0.123 | -1.301*** | 0.3 | 0.388 | -1.075** | -1.762*** | -1.447*** |
| | -0.534 | -0.496 | -0.459 | -0.496 | -0.471 | -0.442 | -0.384 |
| Poverty | -2.52 | -7.419*** | -1.526 | -3.929** | -5.620*** | -6.741*** | -6.658*** |
| | -1.699 | -1.724 | -1.475 | -1.678 | -1.795 | -1.741 | -1.542 |
| Bachelor | 9.016*** | 9.365*** | 9.182*** | 6.777*** | 6.128*** | 5.310*** | 1.803* |
| | -1.245 | -1.218 | -1.046 | -1.176 | -1.248 | -1.228 | -1.063 |
| Black | -0.743 | -0.726 | 0.275 | -0.940* | -0.581 | 0.347 | 0.00288 |
| | -0.548 | -0.553 | -0.476 | -0.542 | -0.586 | -0.574 | -0.505 |
| Hispanic | 6.900*** | 4.552*** | 6.814*** | 6.204*** | 5.107*** | 7.809*** | 7.375*** |
| | -1.253 | -1.249 | -1.082 | -1.236 | -1.331 | -1.314 | -1.157 |
| Age 20 to 34 | 8.764** | 10.72*** | 2.962 | 10.15*** | 7.376* | 5.554 | -0.575 |
| | -3.757 | -3.78 | -3.252 | -3.708 | -4.008 | -3.95 | -3.466 |
| Constant | 2.268 | 15.47*** | -0.129 | 0.468 | 18.02*** | 28.71*** | 28.68*** |
| | -5.853 | -5.46 | -5.056 | -5.458 | -5.228 | -4.941 | -4.266 |
| Lambda | 0.742*** | 0.377*** | 0.568*** | 0.565*** | 0.616*** | 0.754*** | 0.915*** |
| | -0.0481 | -0.0521 | -0.0502 | -0.0481 | -0.0465 | -0.0411 | -0.0339 |
| Sigma2 | 11.21*** | 11.57*** | 8.556*** | 11.29*** | 13.24*** | 12.82*** | 9.934*** |
| | -0.288 | -0.296 | -0.22 | -0.29 | -0.34 | -0.329 | -0.255 |
| No. of Obs. | 3,059 | 3,059 | 3,059 | 3,059 | 3,059 | 3,059 | 3,059 |

Notes: We estimate cross-sectional models of the geographic expansion of the P2P Credit during the period 2007-2013 for each year. Variable definitions are provided in Table 2. Models are estimated with maximum likelihood approach while controlling for the spatial dependence with a spatial lag term (lambda). State dummies are not shown. Standard errors are in parentheses. ***, **, * denote significance at the 1, 5 and 10 percent level, respectively.

**Table B.9:** OLS regressions

| **VARIABLES** | **Total Volume** | **Total No. of Loans** |
|---|---|---|
| Branches | -0.0110*** | -0.00211* |
| | -0.004 | -0.0011 |
| HHI | -2.632*** | -0.465*** |
| | -0.344 | -0.0955 |
| Payday | 0.113** | -0.016 |
| | -0.0503 | -0.0142 |
| Crisis leverage | -10.80*** | -2.304** |
| | -3.594 | -1.017 |
| Density | 0.499*** | 0.0700*** |
| | -0.0483 | -0.0121 |
| Broadband | -3.843*** | -0.0923 |
| | -0.968 | -0.359 |
| income_log | -3.083*** | -0.445*** |
| | -0.424 | -0.118 |
| Poverty | -8.178*** | -3.124*** |
| | -1.483 | -0.421 |
| Bachelor | 3.899*** | 2.142*** |
| | -1.044 | -0.281 |
| Black | 0.228 | 0.512*** |
| | -0.48 | -0.132 |
| Hispanic | 7.831*** | 1.703*** |
| | -1.084 | -0.214 |
| Age 20 to 34 | 2.981 | -0.42 |
| | -3.286 | -0.865 |
| Constant | 48.71*** | 6.739*** |
| | -4.655 | -1.309 |
| Observations | 3,059 | 3,059 |
| R-squared | 0.192 | 0.15 |

**Table B.10:** Marginal Effects

| | **Volume of P2P loans** | | | **Number of P2P loans** | | |
|---|---|---|---|---|---|---|
| | **ATDI** | **ATII** | **ATI** | **ATDI** | **ATII** | **ATI** |
| Branches | -0.0013 | -0.0004 | -0.0018 | -0.0022 | -0.0016 | -0.0038 |
| HHI | -0.1963 | -0.0628 | -0.2591 | -14,631 | -0.468 | -19,312 |
| Broadband | -0.4321 | -0.1382 | -0.5703 | -32,202 | -10,300 | -42,502 |
| Poverty | -0.6414 | -0.2052 | -0.8466 | -47,799 | -15,290 | -63,089 |
| Hispanic | 0.6472 | 0.207 | 0.8542 | 48,232 | 15,428 | 63,660 |
| Income_log | -0.1882 | -0.0602 | -0.2484 | -14,023 | -0.4486 | -18,509 |
| Payday | 0.0083 | 0.0027 | 0.011 | 0.0621 | 0.0199 | 0.082 |
| Education | 0.3235 | 0.1035 | 0.427 | 24,111 | 0.7712 | 31,823 |
| Black | 0.0048 | 0.0016 | 0.0064 | 0.0361 | 0.0116 | 0.0477 |
| Age | 0.2187 | 0.07 | 0.2887 | 16,300 | 0.5214 | 21,514 |
| Crisis leverage | -10,565 | -0.3379 | -13,944 | -78,735 | -25,185 | -103,921 |
| Density | 0.0458 | 0.0147 | 0.0605 | 0.3417 | 0.1093 | 0.4509 |

**Table B.11:** Marginal Effects (Continued)

| | **Lending Club Volume** | | | **Prosper Volume** | | |
|---|---|---|---|---|---|---|
| | **ATDI** | **ATII** | **ATI** | **ATDI** | **ATII** | **ATI** |
| Branches | -0.002 | -0.0021 | -0.0041 | -0.0011 | -0.0002 | 0.00 |
| HHI | -0.1318 | -0.1352 | -0.267 | -0.5643 | -0.1117 | -0.676 |
| Broadband | -0.7305 | -0.7494 | -14,799 | 0.3914 | 0.0775 | 0.4688 |
| Poverty | -0.3888 | -0.3988 | -0.7876 | -11,087 | -0.2195 | -13,282 |
| Hispanic | 0.7605 | 0.7801 | 15,405 | 0.2643 | 0.0523 | 0.3166 |
| Income_log | -0.0684 | -0.0702 | -0.1386 | -0.2857 | -0.0566 | -0.3423 |
| Payday | -0.0091 | -0.0093 | -0.0184 | 0.0355 | 0.007 | 0.0426 |
| Education | 0.2817 | 0.289 | 0.5707 | 0.6477 | 0.1282 | 0.7759 |
| Black | -0.0509 | -0.0522 | -0.1031 | -0.0103 | -0.002 | -0.0123 |
| Age | -0.4418 | -0.4532 | -0.895 | 18,512 | 0.3665 | 22,176 |
| Crisis leverage | -0.582 | -0.597 | -11,790 | -31,695 | -0.6275 | -37,970 |
| Density | 0.0615 | 0.0631 | 0.1246 | 0.0829 | 0.0164 | 0.0993 |

**Table B.12:** Marginal Effects (Continued)

| | **Lending Club Number** | | | **Prosper Number** | | |
|---|---|---|---|---|---|---|
| | **ATDI** | **ATII** | **ATI** | **ATDI** | **ATII** | **ATI** |
| Branches | -0.0035 | -0.0028 | -0.0063 | -0.0026 | -0.0011 | -0.0036 |
| HHI | -0.3026 | -0.2445 | -0.5471 | -0.0767 | -0.032 | -0.1088 |
| Broadband | -0.5816 | -0.47 | -10,516 | 0.833 | 0.3479 | 11,809 |
| Poverty | -22,045 | -17,814 | -39,859 | -43,074 | -17,989 | -61,064 |
| Hispanic | 0.9721 | 0.7855 | 17,576 | 13,654 | 0.5702 | 19,356 |
| Income_log | -0.0372 | -0.03 | -0.0672 | -0.5017 | -0.2095 | -0.7113 |
| Payday | -0.0299 | -0.0242 | -0.0541 | -0.027 | -0.0113 | -0.0383 |
| Education | 10,607 | 0.8571 | 19,178 | 37,217 | 15,543 | 52,760 |
| Black | 0.4693 | 0.3793 | 0.8486 | 0.8891 | 0.3713 | 12,604 |
| Age | -0.2423 | -0.1958 | -0.4382 | -28,091 | -11,732 | -39,822 |
| Crisis leverage | -0.7002 | -0.5658 | -12,660 | -33,679 | -14,066 | -47,745 |
| Density | 0.0203 | 0.0164 | 0.0367 | 0.0966 | 0.0403 | 0.1369 |

## Bibliography


Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 84(3):488–500.

Anselin, L. et al. (1980). Estimation methods for spatial autoregressive structures. *Estimation methods for spatial autoregressive structures.*, (8).

Aubuchon, C. P., Wheelock, D. C., et al. (2010). The geographic distribution and characteristics of us bank failures, 2007-2010: do bank failures still reflect local economic conditions? *Federal Reserve Bank of St. Louis Review*, 92(5):395–415.

Bass, F. M. (1969). A new product growth for model consumer durables. *Management science*, 15(5):215–227.

Beck, T., Demirguc-Kunt, A., and Peria, M. S. M. (2007). Reaching out: Access to and use of banking services across countries. *Journal of Financial Economics*, 85(1):234–266.

Berger, A. N., Demsetz, R. S., and Strahan, P. E. (1999). The consolidation of the financial services industry: Causes, consequences, and implications for the future. *Journal of Banking & Finance*, 23(2):135–194.

Black, S. E. and Strahan, P. E. (2002). Entrepreneurship and bank credit availability. *The Journal of Finance*, 57(6):2807–2833.

Blundell-Wignall, A. and Roulet, C. (2013). Business models of banks, leverage and the distance-to-default. *OECD Journal: Financial Market Trends*, 2012(2):7–34.

Butler, A. W., Cornaggia, J., and Gurun, U. G. (2016). Do local capital market conditions affect consumers [U+FFFD] rrowing decisions? *Management Science*.

Chetty, R., Friedman, J. N., and Saez, E. (2013). Using differences in knowledge across neighborhoods to uncover the impacts of the eitc on earnings. *American Economic Review*, 103(7):2683–2721.

Claessens, S. and Laeven, L. (2004). What drives bank competition? some international evidence. *Journal of Money, credit, and Banking*, 36(3):563–583.

Comin, D. A., Dmitriev, M., and Rossi-Hansberg, E. (2012). The spatial diffusion of technology. Technical report, National Bureau of Economic Research.

Cox, N. (2017). Pricing, selection, and welfare in the student loan market: Evidence from borrower repayment decisions. Manuscript, University of California, Berkeley.

Crawford, G., N. P. and Schivardi, F. (2016). Asymmetric information and imperfect competition in lending markets. Working Paper, University of Zurich.

DeYoung, R., Lang, W. W., and Nolle, D. L. (2007). How the internet affects output and performance at community banks. *Journal of Banking & Finance*, 31(4):1033–1060.

Dick, A. A. (2007). Market size, service quality, and competition in banking. *Journal of money, credit and banking*, 39(1):49–81.

Drukker, D. M., Peng, H., Prucha, I. R., Raciborski, R., et al. (2013). Creating and managing spatial-weighting matrices with the spmat command. *Stata Journal*, 13(2):242–286.

Duarte, J., Siegel, S., and Young, L. (2012). Trust and credit: the role of appearance in peer-to-peer lending. *The Review of Financial Studies*, 25(8):2455–2484.

Edelberg, W. (2006). Risk-based pricing of interest rates for consumer loans. *Journal of Monetary Economics*, 53(8):2283–2298.

Einav, L., Jenkins, M., and Levin, J. (2012). Contract pricing in consumer credit markets. *Econometrica*, 80(4):1387–1432.

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

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

Fitch (2014). Peer to peer lending,. *Global Industry Overview*. Fitch Ratings Report.

Fradkin, A. (2014). Auctions versus posted prices in online markets. NBER Working Paper.

Frame, W. S. and White, L. J. (2014). Technological change, financial innovation, and diffusion in banking.

Freedman, S. and Jin, G. Z. (2008). Do social networks solve information problems for peer-to-peer lending? evidence from prosper. com.

Gao, Q. and Lin, M. (2013). Linguistic features and peer-to-peer loan quality: A machine learning approach. *Available at SSRN*.

Geroski, P. A. (1989). Entry, innovation and productivity growth. *The Review of Economics and Statistics*, pages 572–578.

Grodzicki, D. (2012). The evolution of competition in the credit card industry. Manuscript, Stanford University.

Hagiu, A. and Wright, J. (2015a). Marketplace or reseller? *Management Science*, 61(1):184–203.

Hagiu, A. and Wright, J. (2015b). Multi-sided platforms. *International Journal of Industrial Organization*, 43(C):162–174.

Haldane, A. (2012). The dog and the frisbee, speech given at the federal reserve bank of kansas city [U+FFFD] 6th economic policy symposium [U+FFFD] e changing policy landscape [U+FFFD] *Jackson Hole, Wyoming*.

Haldane, A. G. (2011). Capital discipline based on a speech given at the american economic association.

Hannan, T. H. and Prager, R. A. (1998). The relaxation of entry barriers in the banking industry: An empirical investigation. *Journal of Financial Services Research*, 14(3):171–188.

Hasan, I., Koetter, M., and Wedow, M. (2009). Regional growth and finance in europe: Is there a quality effect of bank efficiency? *Journal of Banking & Finance*, 33(8):1446–1453.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning: data mining, inference and prediction*. Springer, 2 edition.

Hernando, I. and Nieto, M. J. (2007). Is the internet delivery channel changing banks [U+FFFD] rformance? the case of spanish banks. *Journal of Banking & Finance*, 31(4):1083–1099.

Herzenstein, M., Sonenshein, S., and Dholakia, U. M. (2011). Tell me a good story and i may lend you money: the role of narratives in peer-to-peer lending decisions. *Journal of Marketing Research*, 48(SPL):S138–S149.

Iyer, R., Khwaja, A. I., Luttmer, E. F. P., and Shue, K. (2016). Screening peers softly: Inferring the quality of small borrowers. *Management Science*, 62(6):1554–1577.

Karlan, D. and Zinman, J. (2005). Observing unobservables: identifying information asymmetries with a consumer-credit field experiment. Proceedings 961, Federal Reserve Bank of Chicago.

Kim, M., Kliger, D., and Vale, B. (2003). Estimating switching costs: the case of banking. *Journal of Financial Intermediation*, 12(1):25–56.

LeSage, J. P. (2008). An introduction to spatial econometrics. *Revue d'économie industrielle*, (3):19–44.

LeSage, J. P. and Pace, R. K. (2009). *Introduction to Spatial Econometrics (Statistics, textbooks and monographs)*. CRC Press.

Mariathasan, M. and Merrouche, O. (2014). The manipulation of basel risk-weights. *Journal of Financial Intermediation*, 23(3):300–321.

Meyer, A. (2013). Pricing mechanisms in peer-to-peer online credit markets. Manuscript, Stanford University.

Mitra, T. and Gilbert, E. (2014). The language that gets people to give: Phrases that predict success on kickstarter. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pages 49–61. ACM.

Morse, A. (2015). Peer-to-peer crowdfunding: Information and the potential for disruption in consumer lending. *Annual Review of Financial Economics*, 7:463–482.

Philippon, T. (2015). Has the us finance industry become less efficient? on the theory and measurement of financial intermediation. *The American Economic Review*, 105(4):1408–1438.

Pope, D. G. and Sydnor, J. R. (2011). What?s in a picture?: Evidence of discrimination from prosper.com. *Journal of Human Resources*, 46(1):53–92.

Rahim, T. (2017). Can peer-to-peer online platforms improve market outcomes by controlling prices? Manuscript, Boston University.

Ravina, E. (2013). Love and loans: The effect of beauty and personal characteristics in credit markets. Working paper, Columbia Business School.

Rey, S. J. and Montouri, B. D. (1999). Us regional income convergence: a spatial econometric perspective. *Regional studies*, 33(2):143–156.

Rhoades, S. A. et al. (2000). Bank mergers and banking structure in the united states, 1980-98. Technical report, Board of Governors of the Federal Reserve System (US).

Rogers, E. M. (2003). Diffusion of innovations. free press. *New York*, page 551.

Rysman, M. (2009). The economics of two-sided markets. *The Journal of Economic Perspectives*, 23(3):125–143.

Shy, O. (2002). A quick-and-easy method for estimating switching costs. *International Journal of Industrial Organization*, 20(1):71–87.

Wei, Z. and Lin, M. (2016). Market mechanisms in online peer-to-peer lending. *Management Science*, 0(0):null.

Zhang, J. and Liu, P. (2012). Rational herding in microloan markets. *Management Science*, 58(5):892–912.

# CURRICULUM VITAE

## Talal Rahim

### EDUCATION

Ph.D. in Economics, Boston University, Boston MA, 2018 (expected)
Dissertation Title: *Essays in Industrial Organization of Credit Markets*
Dissertation Committee: Marc Rysman, Berardino Palazzo, Hiroaki Kaido

M.Sc., Economics,
London School of Economics and Political Science, London, United Kingdom, 2011

M.Sc., Economics and Finance,
Lahore School of Economics, Lahore, Pakistan, 2009

### FIELDS OF INTEREST

Industrial Organization, Financial Economics, Applied Econometrics, Computational Economics

### FELLOWSHIPS AND AWARDS

Department Fellowship, Boston University Economics Department, 2012 - 2017
Initiative of Cities Fellowship, Boston University, 2015
Co-Recipient of First Prize, Smart City Track, Thingworx Internet of Things Hackathon, 2015
Co-Recipient of Runner Up Prize, Wearables Track, MIT Medicine Grand Hackathon, 2015
University Merit Scholarship, Lahore School of Economics, 2005 - 2009

### WORK EXPERIENCE

Research Assistant, Boston University, Boston, MA, 09/2016 - 05/2017
Data Science Fellow, Office of Mayor, City of Providence, Summer 2015

## WORKING PAPERS

"Can Peer-to-Peer Platforms Improve Market Outcomes by Controlling Prices?" (Job Market Paper), July 2017
"Impact of Finer Credit Scoring on Borrower Loan Outcomes", July 2018
"What Drives the Expansion of Peer-to-Peer Lending?", May 2018
"Determinants of Bank Branching Decisions in an Emerging Market" December 2016

## TEACHING EXPERIENCE

Teaching Fellow, Financial Economics (Master's level), Boston University
Teaching Fellow, Public Control of Business (Master's level), Boston University
Teaching Fellow, Environmental Economics (Master's. level), Boston University
Teaching Fellow, Topics in Macro and Monetary Theory (Master's. level), Boston University
Introduction to Macroeconomics, Boston University

## CONFERENCES AND EXTERNAL PRESENTATIONS

Industrial Organization Society, 15th Annual Conference, April 2017
George Mason University Antonin Scalia Law School, Digital Information Policy Scholars Conference, April 2017
Empirical Microeconomics Seminar, Boston University, April 2015, May 2016, October 2016

**Computer Skills:** R, Matlab, Stata, Python, Ampl and LaTeX

**Languages:** English (native), Urdu/Hindi (native)

**Citizenship/Visa:** Pakistan / F-1