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

Talal Rahim*

December 18, 2016

Abstract

In many Peer-to-Peer (P2P) online markets the transactions between buyers and sellers are facilitated by platforms that often control the market prices. This restricts a fundamental function of a market: its ability to aggregate information and reflect it in prices. In this paper I use micro data from an online P2P credit market to show evidence of better allocation of credit when prices are set by the platform instead of by competing lenders in an auction. I specify and estimate an econometric model of loan demand and repayment and use unique variation in the platform’s pricing schedule to identify key parameters. I use the estimated parameters to conduct a counterfactual experiment in which borrowers are offered prices determined through an auction among lenders. I find that when lenders set prices, 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.

JEL codes: D14, D40, G14, L19

Keywords: Peer-to-Peer Lending, Platform Marketplaces, Credit Demand, Pricing Mechanism

*Email: rahimt@bu.edu. Department of Economics, Boston University
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 competitive 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

---

[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.
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 pseudo out-of-sample prediction error. 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.

The focus in this paper is on a market for credit, a commodity considered essential 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.

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 price vector is controlled by the platform while both borrowers and lenders are price takers.

\[\text{In section 5 I explain that this approach has several advantages over its alternatives in calculating the counterfactual input prices I need to answer the question motivated above.}\]
and pick their own allocations.

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 accordingly. 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.

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 bank does. Matching different borrowers and savers/investors based on different maturity preferences is a fundamental function of depository institutions like banks. 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 withdrawl 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.

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.

The rest of the paper is structured as follows: Section 2 goes over related literature, section 3 provides an institutional overview of P2P online credit markets with an emphasis on how they differ from traditional credit markets, section 4 presents the data and sample selection procedure, develops the econometric model and its estimation procedure, and presents estimation results, section 5 presents the case counterfactual experiment and its results, and section 6 gives conclusion.
2 Related Literature

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 Einav et al., (2014) 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 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 Syndor (2011) and Ravina (2012) 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 (2015) provide evidence that the market is able to determine interest rates that predict defaults better than the finest credit scores do, Zhang and Liu (2012) provide evidence of investor herding behavior in these markets, and Butler, et. al (2014) show evidence of home bias in investor decisions.

Two closely related papers Meyer (2013) and Wei and Lin (2015) 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 Zimmerman (2009) who carry out an experiment in a credit market to identify sources of adverse selection. On the other hand, Einav et. al (2012) and Crawford et. al (2015) 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.

3 Institutional Overview and Data

3.1 Institutional Overview

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 2013.5 However, this only 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.6 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.

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.

---

3Americas Alternative Finance Benchmarking Report, 2015
4Citi Group Report, 2016
5Fitch Report, 2014
6Federal government data aggregated by www.nerdwallet.com
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. 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.

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.

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

---

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

8The platform also allows the borrowers an option to convert the listing to a loan if it receives at least 70% of the requested amount.
their investments. A lender \( j \) earns a gross return of \( (1 + r_2) s_l \) 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.\(^9\)

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 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

---

\(^9\)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.

\(^10\)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.
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.

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 benefits of increasing competition – that of resultant set of prices which increase consumer surplus.

3.2 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.
Table 1: Descriptive Statistics

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

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.

Now 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 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 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.

Figure 2: Variation in Interest Rates Over Time

![Graph showing variation in interest rates over time](image)

**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.

4 A Model of Loan Demand and Repayment

The model I construct aims to quantify 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.

4.1 Model Setup

The setup adapts the framework of Crawford et. al (2015) and Einav et. al (2012) 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.

Given the above assumptions, let there be \(i = 1, \ldots, I\) borrowers each of whom picks exactly one contract 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 \(j - \text{year}\) 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.

---

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.
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

\[ L^*_i = \beta_T \text{Term}_i + \beta_P \text{Price}_i + \beta_F \text{Fee}_i + X_i'\beta_X + \epsilon_{Li} \]

Where \( \text{Term}_i, \text{Price}_i \) and \( \text{Fee}_i \) represent the contract-specific variables of the loan contract the borrower ends up picking, and \( \epsilon_{Li} \) is the error term observed by borrower by not by reseacher. Note here that the variable \( \text{Term}_i \) is essentially the same as the binary variable indicating the choice of maturity contract.

Finally, conditional on having a loan of contract variables \( \text{Term}_i, \text{Price}_i \), and \( \text{Fee}_i \) 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

\[ S^*_i = \gamma_T \text{Term}_i + \gamma_P \text{Price}_i + \gamma_F \text{Fee}_i + \gamma_L L_i + \gamma_H 1\{L_i = \bar{L}_i\} + X_i'\gamma_X + \epsilon_{Si} \]

Where \( L_i \) is the observed loan size, \( \bar{L}_i \) is the loan limit assigned to borrower \( i \) by the platform, and \( \epsilon_{Si} \) is the error term observed by borrower by not by reseacher.

### 4.2 Estimation

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

First I assume \( (\epsilon_U, \epsilon_L, \epsilon_S) \) are distributed jointly normal with the distribution given by

\[ f(\epsilon_U, \epsilon_L, \epsilon_S) = N(0, \Sigma). \]

The only restriction on \( \Sigma \) is that it is normalized by the first variance which is \( \sigma^2_U \). All the other parameters, covariances and variances, are allowed to vary and will be estimated by this procedure.

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:

\[ f(\epsilon_U, \epsilon_L, \epsilon_S) = f(\epsilon_S | \epsilon_L, \epsilon_U) f(\epsilon_L | \epsilon_U) f(\epsilon_U) \]

To simplify the notation I define the following matrices of \( W_{Ui} = [\Delta \text{Price}_i, \Delta \text{Fee}_i, X_i] \), \( W_{Li} = [\text{Term}_i, \text{Price}_i, \text{Fee}_i, X_i] \), \( W_{Si} = [\text{Term}_i, \text{Price}_i, \text{Fee}_i, L_i, 1\{L_i = \bar{L}_i\}, X_i] \) 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
\[ Q_i = \begin{cases} 1, & \text{if } U_{i5}^* - U_{i3}^* \geq 0 \\ 0, & \text{otherwise} \end{cases} \]

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

\[ P_{Q_i = 1} = \Phi (W'_{Ui} \alpha) \]

and the probability that a borrower picks the 3-year contract is

\[ P_{Q_i = 0} = 1 - P_{Q_i = 1}. \]

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

\[ L_i = \begin{cases} L_i^* = W'_{Li} \beta + \varepsilon_{Li}, & \text{if } L_i^* < \bar{L}_i \\ \bar{L}_i, & \text{otherwise} \end{cases} \]

If \( L_i^* < \bar{L}_i \), 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

\[ P_{L_i = L_i^* | \varepsilon_{Ui}} = \text{Prob}(L_i^* = W'_{Li} \beta + \varepsilon_{Li}) = f_{\varepsilon_{Li} | \varepsilon_{Ui}} (L_i - W'_{Li} \beta) \]

On the other hand, if the loan limit constraint is binding for borrower \( i \), i.e. \( L_i^* \geq \bar{L}_i \), 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

\[ P_{L_i = \bar{L}_i | \varepsilon_{Ui}} = \text{Prob}(L_i^* \geq W'_{Li} \beta + \varepsilon_{Li}) = F_{\varepsilon_{Li} | \varepsilon_{Ui}} (W'_{Li} \beta - \bar{L}_i) \]

Calculation of the moments of the conditional distribution function \( F_{\varepsilon_{Li} | \varepsilon_{Ui}} \) 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 \( F_{\varepsilon_{Li} | \varepsilon_{Ui}} \) 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.
\[ P_{L_i|Q_i=1} = \int_{-\infty}^{W_{i'}\alpha} P_{L_i|\varepsilon} \times f(\varepsilon_U) \, d\varepsilon_U \]

and

\[ P_{L_i|Q_i=0} = \int_{W_{i'}\alpha}^{\infty} P_{L_i|\varepsilon} \times f(\varepsilon_U) \, d\varepsilon_U \]

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\(^\text{[12]}\). 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

\[ S_i = \begin{cases} S_i^* = W_{S_i^*} + \varepsilon_{S_i^*}, & \text{if } S_i^* < \bar{S}_i \\ \bar{S}_i, & \text{if } S_i^* \geq \bar{S}_i \end{cases} \]

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

\[ P_{S_i=S_i^*|\varepsilon_{S_i},\varepsilon_U} = f_{\varepsilon_{S_i},\varepsilon_U}(S_i - W_{S_i^*}) \]

The probability of observing full or censored repayment is given by

\[ P_{S_i=S_i|\varepsilon_{S_i},\varepsilon_U} = F_{\varepsilon_{S_i},\varepsilon_U}(W_{S_i^*} - \bar{S}_i) \]

Here again, calculation of the moments of the conditional distribution function \( F_{\varepsilon_{S_i}|\varepsilon_U,\varepsilon_L} \) is complicated since \( \varepsilon_{U_i} \) is not observed. Another problem here is that \( \varepsilon_{L_i} \) 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 distribution function \( F_{\varepsilon_{S_i}|\varepsilon_U,\varepsilon_L} \) and instead must integrate over all \( \varepsilon_{L_i} \) 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 \( \varepsilon_{U_i} \) and for borrowers who choose loan sizes equal to their loan limits I integrate over all possible \( \varepsilon_{U_i} \) and all possible \( \varepsilon_{L_i} \). The expressions for the likelihood of observed repayment conditional on borrowers choosing loans of sizes less than

\(^{[12]}\text{Note that } \bar{S}_i = 1 \text{ for completed loans.}\)
loan limits are given by:

\[ P_{S_i|L_i=L_i^*,Q_i=1} = \int_{-\infty}^{\infty} P_{S_i|\varepsilon_L,\varepsilon_U} \times f(\varepsilon_L,\varepsilon_U) d\varepsilon_U \]

\[ P_{S_i|L_i=L_i^*,Q_i=0} = \int_{W_{U_i}}^{\infty} P_{S_i|\varepsilon_L,\varepsilon_U} \times f(\varepsilon_L,\varepsilon_U) d\varepsilon_U \]

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

\[ P_{S_i|L_i=L_i^*,Q_i=1} = \int_{L_i-W_{L_i\beta}}^{\infty} \int_{-\infty}^{W_{U_i}^{\alpha}} P_{S_i|\varepsilon_L,\varepsilon_U} \times f(\varepsilon_L,\varepsilon_U) d\varepsilon_U d\varepsilon_L \]

\[ P_{S_i|L_i=L_i^*,Q_i=0} = \int_{\tilde{L}_i-W_{L_i\beta}}^{L_i-W_{L_i\beta}} \int_{W_{U_i}^{\alpha}}^{\infty} P_{S_i|\varepsilon_L,\varepsilon_U} \times f(\varepsilon_L,\varepsilon_U) d\varepsilon_U d\varepsilon_L \]

To summarize, I observe 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.
4.3 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\footnote{This is because over 90% of non-cancelled loan applications get funded.} her decision depends only on contract specific 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 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.

4.4 Demand Estimates

Table 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 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.
Table 2: Estimates of Borrower Demand and Repayment Model

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

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.

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.

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 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.

Figure 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.

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 ordered bid profile of prices $\vec{p}_i = (p_{i1}, \ldots, p_{iJ})$, let

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

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

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

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.

### 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 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 function 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 $i = 1, \ldots, N$ borrowers with data

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

where $P_i$ is the auction-determined price for borrower $i$ and $X_i$ is a vector of $k$ borrower and market specific variables. 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 same distribution $D$.

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 $Z = (P_1, X_1), \ldots, (P_M, X_M)$, whereas the test sample is given by $Z' = (P_{M+1}, X_{M+1}), \ldots, (P_N, X_N)$. 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.

---

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

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

**Random Forests:** Following the classic text of Hastie, Tibshirani and Friedman (2009), I take the following steps to build the random forest:

1. Draw a bootstrap sample $Z^*_b$ of size $M$ from the training sample $Z$

2. 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:

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

   and

   $$R_2(j, s) = X | X_j > s$$

   where $j$ is the index of the *splitting variable* and $s$ is the point at which the split is made called *split point*. Starting with the base node at the top of the tree, the rule for that node is formed by the following optimization problem:

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

   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.

3. 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:

---

15This is not to be confused with sample splitting.
\[ \hat{f}_v^B(x) = \frac{1}{B} \sum_{b=1}^{B} T_b(x) \]

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.

5.3 Counterfactual Results

The results from the matching exercise are presented in Figure 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.

Figure 4: Distribution of Differences in Prices Charged by Platform Vs. Market by Credit Score

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.

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 were determined by the platform. Comparisons of these choices are summarized in Table 3 and Figures 4 to 6.

Table 3: Counterfactual Results
<table>
<thead>
<tr>
<th>Variable</th>
<th>Platform Prices</th>
<th>Market Prices</th>
<th>Mean Diff.</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-Year Prices (%)</td>
<td>15.99</td>
<td>19.01</td>
<td>-3.02***</td>
</tr>
<tr>
<td>5-Year Prices (%)</td>
<td>16.33</td>
<td>20.75</td>
<td>-4.41***</td>
</tr>
<tr>
<td>Pr. of Choosing 5-year Contract</td>
<td>0.211</td>
<td>0.206</td>
<td>0.005***</td>
</tr>
<tr>
<td>Loan Size Chosen (Partial) ($)</td>
<td>11,744</td>
<td>11,473</td>
<td>271***</td>
</tr>
<tr>
<td>Loan Size Chosen (Full) ($)</td>
<td>11,744</td>
<td>10,606</td>
<td>$1,137***</td>
</tr>
<tr>
<td>Fraction of Loan Repaid (Partial)</td>
<td>0.84</td>
<td>0.83</td>
<td>0.014***</td>
</tr>
<tr>
<td>Fraction of Loan Repaid (Full)</td>
<td>0.833</td>
<td>0.825</td>
<td>0.008***</td>
</tr>
</tbody>
</table>

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

The average differences in the means of each variable under the two pricing mechanisms are shown in Table 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 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 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 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 3, I show how these effects change with borrower riskiness as shown in Figures 5 to 7. Figure 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 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).

The two panels in Figure 6 highlight two important points: The first, Figure 6a, 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 panels (a) and (b) of Figure 6 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 6: Distribution of Partial and Full Effects of Change in Pricing Mechanism on Loan Size Choice by Credit Score

Panel (a)

Notes: 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.

The two panels in Figure 7 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 panel (a) is bigger than the average full effect, as shown in panel (b), 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.

Figure 7: Distribution of Partial and Full Effects of Change in Pricing Mechanism on Repayment Choice by Credit Score

Panel (a)
Notes: 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.

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.
References


