Asymmetric Information, Reputation, and Welfare in Online Credit Markets

Yi Xin *
Johns Hopkins University

November 7, 2017
(Job Market Paper Draft, Link to Latest Version)

Abstract

This paper studies the impact of reputation/feedback systems on the operation of online credit markets using data from Prosper.com. The ability of lenders to recover their loans is one of the main concerns in these markets, where the problems of asymmetric information are two-fold. On the one hand, borrowers differ in their inherent risks; on the other hand, additional incentives are necessary to motivate borrowers to exert effort. In this paper, I investigate the channels through which reputation/feedback systems improve the total welfare of market participants when both adverse selection and moral hazard are present. A finite-horizon dynamic model of a credit market in which borrowers and lenders interact repeatedly over time is developed and then estimated. I prove the identification of the distribution of borrowers’ private types and utility primitives based on variations in borrowers’ repayment histories, transitions of their characteristics, and interest rates. In the counterfactual analysis, I find that 22 percent of welfare loss from asymmetric information is due to adverse selection, while 78 percent is due to moral hazard. Furthermore, I find that 95 percent of the inefficiency induced by asymmetric information is eliminated by the reputation system. I consider a policy intervention that protects borrowers from accidental loss of reputation. My results suggest that incorporating a payment protection insurance into the market further improves total welfare.

Keywords: asymmetric information, reputation/feedback system, credit markets

JEL Code: L14, D82, G21, C14

*Department of Economics, Johns Hopkins University, Wyman Park Building 544E, 3100 Wyman Park Drive, Baltimore, MD 21218. Email: yxin4@jhu.edu. I thank my dissertation advisors Yingyao Hu, Elena Krasnokutskaya, and Yuya Takahashi for their guidance and encouragement throughout the research process. This work has benefited particularly from the comments of Yuya Sasaki, Robert Moffitt, Jian Ni, Richard Spady, Jorge Balat, Ying Chen, Nicholas Papageorge, Matthew Backus, John Cawley, Ruli Xiao, and Yajing Jiang. I also thank seminar and conference participants at the Johns Hopkins University, WEAI Graduate Student Workshop (2017), NASM (2017), Washington DC IO Conference (2017), and 15th Annual IIOC (2017). The usual disclaimer applies.
1 Introduction

Online credit markets for peer-to-peer lending have developed rapidly over the last several years. These markets attract dispersed and anonymous borrowers and often require no collateral. The ability of lenders to recover their loan amounts is one of the main concerns in markets of unsecured loans, where the problems of asymmetric information are two-fold. On the one hand, borrowers differ in their inherent risk, which is hidden information; on the other hand, borrowers’ actions are hidden as well, so additional incentives are necessary to motivate them to exert effort to repay debts. Most online credit markets rely on a “reputation/feedback” system which computes and publishes “reputation scores” based on past outcomes to facilitate transactions. While the effect of reputation/feedback systems is well-known in theory, there is very little empirical work on the extent to which and the channels through which reputation systems improve the total welfare of market participants. Answers to these questions shed light on optimal mechanism design and regulations not only for fast-growing online marketplaces, but also for traditional credit markets, and thus have become increasingly important.

This paper is, to the best of my knowledge, the first to quantify the extent to which reputation/feedback systems improve the total welfare of market participants when both hidden information (adverse selection) and hidden actions (moral hazard) are present. I develop and estimate a finite-horizon dynamic model of a credit market in which borrowers and lenders interact repeatedly over time. I prove the identification of the distribution of borrowers’ private types and utility primitives based on variations in borrowers’ repayment histories, transitions of their characteristics, and interest rates. My results confirm three important channels through which the welfare is improved by the reputation system. I also provide quantitative results on the welfare loss from adverse selection and moral hazard separately, and I find that moral hazard plays an important role in online credit markets.

In this paper, I use a large transaction-level dataset from Prosper.com. This website is one of the leading peer-to-peer lending marketplaces in the US and provides services that match lenders with borrowers. It collects information on borrowers’ credit profiles and loan histories and decides on interest rates accordingly. My analysis benefits from the setting of Prosper in the following ways. First, the reputation/feedback system in this market is representative of that in other peer-to-peer websites. It provides clear reputational incentives

---

1See the trend of growth in Figure 3.
2For reviews of reputation/feedback systems in online markets, see Einav et al (2015) and Tadelis (2016).
3There is a large theoretical literature on the qualitative effect of reputation systems; see Akerlof (1970), Holmstrom (1999), Bar-Isaac and Tadelis (2008), Stiglitz and Weiss (1983), Diamond (1989).
4Prosper now has more than two million registered members and has helped facilitate over nine billion dollars in loans over the past ten years.
through history-dependent pricing schemes. Second, “reputation scores” (credit grades) and past outcomes (whether defaults or late payments occur) on this website are more objective compared to customer reviews or individual rating scores based on the quality of goods or services.\(^5\) Third, the website keeps track of all borrowers’ proposed listings with borrowers’ detailed characteristics and the outcomes of each loan for a relatively long time. Thus, I am able to observe repeated borrowing patterns and how a borrower’s reputation is updated over time. In addition, Prosper prohibits once-defaulted borrowers from future credit access. The incentive effects of terminations discussed in the work of Stiglitz and Weiss (1983) strengthen the role of the reputation/feedback system in my analysis.

In order to see whether reputational incentives impact borrowers’ behavior and market outcomes, I begin my analysis by focusing on a group of borrowers who have two overlapping loans. I arrange the loans for each borrower based on their closing dates and obtain two results from regression analysis using this sample. First, I find that borrowers who default on the first-closed loans are more likely to default on the second-closed loans. There are two possible explanations for this finding: (1) once they default on one loan, borrowers know that they are not allowed to enter again, and thus lose the reputational incentives of paying off the existing loans; or (2) borrowers’ unobserved types create positive correlation between defaults on the two loans. If only the second channel matters, we would expect that the default rates for the first- and the second-closed loans are the same, since the pool of borrowers remains unchanged. However, I further find that the second-closed loans have a significantly higher default rate compared to the first-closed loans controlling for observables. This confirms the importance of the first channel. In other words, borrowers respond to reputational incentives. When incentives are reduced, default rates significantly increase.

Given the empirical evidence, I develop a finite-horizon dynamic structural model to analyze borrowers’ repayment decisions, lenders’ funding strategies, and websites’ pricing schemes under asymmetric information in online credit markets. In the model, borrowers are heterogeneous in default cost, which constitutes their private type. For each loan, borrowers choose effort levels after observing interest rates assigned by the website and shocks to the cost of effort. The outcomes of each loan are stochastically affected by effort levels. If expected payoffs from participation exceed outside option draws, borrowers stay in the market. On the supply side, lenders make funding decisions after observing borrowers’ participation. Lenders’ payoff from a project depends on the borrower’s type and the effort exerted. Lenders’ decisions are thus impacted by their beliefs about the borrower’s type. The website charges commission fees for each funded loan.

---

\(^5\)Customer review systems are widely adopted by e-commerce platforms such as eBay and Amazon. Other peer-to-peer markets that use individual rating systems include Uber and Airbnb.
While the existing literature on empirical contract models of credit markets mainly focuses on adverse selection in a static environment (see Einav et al, 2012 and Kawai et al, 2016), my paper combines adverse selection and moral hazard in a dynamic setting. Asymmetric information enters my model through two channels. First, lenders and the website do not observe borrowers’ true types. Borrowers with the same observables are charged the same interest rates, which means that “lemon” borrowers are more likely to participate. This is the effect of adverse selection. Second, borrowers’ effort levels are also unobserved. In my model, the impact of moral hazard is multi-layered. For one, lenders face a trade-off between interest rates and default probabilities. When interest rates increase, borrowers exert less effort. Furthermore, borrowers’ effort choices interact with lenders’ beliefs in the dynamic setting.\(^6\) In equilibrium, borrowers’ strategies need to be consistent with lenders’ updated beliefs about borrowers’ true types.

In terms of identification, the key primitive of the model is the joint distribution of the borrower’s type and the unobserved part of his/her cost of effort. A borrower’s type is persistent, while costs of effort are drawn for each loan. To disentangle these two unobserved sources, I first exploit variations across different loans. Following the literature that deals with unobserved heterogeneity in dynamic models (see Hu and Shum, 2012), I use transitions of state variables across loans to pin down the type distribution for borrowers who appear multiple times.\(^7\) I then exploit variations within a loan. Borrowers’ participation, default and late payment performances are independent conditional on effort levels. With these three pieces of information, I recover the distribution of costs of effort conditional on each type following the measurement error literature (see Hu, 2009; Hu and Schennach, 2008). I further identify utility parameters in borrowers’ payoff functions and the outside option distributions for borrowers and lenders using variations in interest rates. In the last step, given other primitives that have been recovered, I identify the original type distribution for all borrowers before any selection occurs.

Directly following my identification results, I apply the likelihood-based estimation strategy to a large transaction-level dataset from Prosper.com. My estimation results indicate that borrowers who have high credit grades and who use loans for debt consolidation are more likely to be “good borrowers”. In addition, borrowers with higher default costs and smaller loan requests are more likely to draw smaller costs of effort. As for the state transition process, I find high debt-to-income ratios are persistent, while the transition process

\(^6\)On the one hand, lenders’ different future beliefs about the borrower’s type lead to different continuation values for borrowers, thus affecting their effort choices. On the other hand, future beliefs are determined by borrowers’ effort through the Bayesian updating process.

\(^7\)For borrowers who have multiple loans, they must have no defaults in their previous loans. Thus, this is a selected sample.
is type-specific – borrowers who have high default costs are more likely to stay with low
debt-to-income ratios.

Using the structural estimates, I conduct counterfactual experiments consisting of three
parts. I first compare welfare under three information structures – one with types and
effort observed (symmetric information), one with only types observed, and one with both
unobserved. This experiment indicates that 22 percent of inefficiency from asymmetric
information is due to adverse selection and 78 percent is due to moral hazard. Furthermore,
I quantify the value of reputation and find that the reputation system recovers 95 percent
of welfare loss from asymmetric information through three important channels. First, the
reputation system helps to refine beliefs about the underlying risks of borrowers, so that
“lemons” are screened out of the market over time. Second, the reputation system creates
additional incentives for borrowers to repay debts, which results in a lower default probability
and an increase in lender’s welfare. Third, with more effort exerted due to reputational
incentives, borrowers have a better chance of having credit access. This result highlights the
effect that reputation systems have on alleviating the welfare loss from credit rationing (see
Stiglitz and Weiss, 1981). I also observe that some good borrowers may accidentally default.
Borrowers that have lost their “reputation” can only take outside options for the rest of time
under the current mechanism. This reduces borrowers’ surplus in future loans.

To address potential long-run inefficiencies due to accidental loss of reputation, the last
part of my counterfactual analysis considers a case where borrowers are offered an option
to buy Payment Protection Insurance (PPI). This insurance covers loan repayments for a
set period of time if borrowers are unable to repay in certain situations. The intuition of
this mechanism is straightforward. If a borrower wants to maintain a good reputation (and
hence credit access in the future), but also worries about future negative shocks, he or she
can purchase this insurance to hedge against that risk. From the counterfactual experiment,
I find that the rate of transaction in the second loans increases, and around 98 percent of
the welfare loss from asymmetric information is recovered by this mechanism. This policy
intervention has strong empirical relevance, especially for small businesses who find peer-to-
peer lending an attractive financing alternative (see Segal, 2015) and who rely heavily on
this form of credit access for their success and growth.

My paper contributes to the literature that studies the value of reputation using structural
models (Yoganarasimhan, 2013; Saeedi, 2014; Lewis and Zervas, 2016; Bai 2016). I am the
first to quantify the extent to which reputation/feedback systems improve the total welfare of

---

8These circumstances usually include being made redundant at one’s job or not being able to work
because of an accident or illness. For details of PPI, see https://www.fca.org.uk/consumers/income-payment-
protection.
market participants when both adverse selection and moral hazard are present. My results confirm three important channels through which the welfare gain is achieved. My paper also relates to the literature that uses hedonic regressions to study the qualitative effect of reputation/review systems on e-commerce platforms (Melnik and Alm, 2002; Eaton, 2005; Jin and Kato, 2006; Lucking-Reiley et al., 2007; Cabral and Hortacsu, 2010), and online labor markets (Lin et al., 2016). I provide new empirical evidence that reputational incentives impact market outcomes and borrowers’ behavior in online marketplaces for consumer loans.

There has been a long discussion on how to test the existence, and further disentangle the effects, of adverse selection and moral hazard in the empirical literature (Chiappori and Salanie, 2000; Chiappori and Salanie, 2002; Abbring et al., 2003; Chiappori et al., 2006). My paper provides quantitative results on the welfare loss from adverse selection and moral hazard separately, and I find that moral hazard plays an important role in credit markets. However, the existing literature on empirical contract models of credit markets mainly focuses on revealing borrowers’ private information through screening/signaling devices (Adams et al., 2009; Einav et al., 2012; Einav et al., 2013; Kawai et al., 2016).

The third contribution of this paper is on identification of contract models. In the existing literature, Perrigne and Vuong (2011) impose a “truth-telling” condition so that there is a one-to-one mapping between private information and observed prices. Kawai et al. (2016) rely on the fact that a borrower’s type and signal have a one-to-one mapping in a separating equilibrium. Gayle and Miller (2015) study models of managerial compensation and assume that some levels of revenue can only be achieved through high effort. The identification strategy in this paper takes advantage of the dynamic structure. I exploit variations in the state transition process to recover borrowers’ unobserved type distribution, which does not require a one-to-one mapping from observables. In addition, I recover probabilities related to unobserved effort levels without assuming that the support of revenue varies with effort. In a separate paper, I develop general identification strategies for dynamic models with unobserved choice variables, which can be easily applied to other types of contract models.

The rest of the paper is organized as follows. I summarize data patterns and show empirical evidence of the value of reputation in Section 2. A structural model is provided in Section 3, with the corresponding identification strategies in Section 4. I present estimation results in Section 5 and the details of counterfactual experiments in Section 6. Section 7 concludes.

---

9A related paper by Bajari et al. (2014) finds adverse selection is an important source of inefficiency in medical insurance markets.
2 Institutional Background and Data Summary

In this paper, I use a large transaction-level dataset from Prosper.com. Institutional details of this website are introduced in this section. I then summarize data patterns and provide empirical evidence that reputational incentives have an impact on borrowers’ behavior and market outcomes.

2.1 Institutional Background

Prosper.com is one of the largest peer-to-peer lending markets in the US. This website aims to provide a platform for individual lenders and borrowers to meet with each other without going through a complicated process as in traditional banking systems. Since its founding in 2005, the website has initiated more than nine billion dollars in loans and has attracted more than two million registered members. On Prosper.com, borrowers list loan requests between $2,000 and $35,000 and individual investors invest as little as $25 in each loan listing they select.\(^\text{10}\) On average, each loan is funded by 43 individual investors, reflecting the crowdfunding feature of this market.\(^\text{11}\) Prosper handles the servicing of the loan on behalf of the matched borrowers and investors; it makes profit by charging both borrowers and lenders service fees proportional to the amount funded.\(^\text{12}\)

The market works in the following manner. To post a listing online, a borrower needs to provide basic information about himself to the website, including his social security number, employment status, whether he is a homeowner, annual income, etc. Prosper hires a third-party credit report agency to verify the applicant’s identity and credit history. The borrower’s FICO score, total number of delinquencies, current number of credit lines and so on are thereby revealed to the website. After the verification stage, the borrower is assigned a credit grade and can post a listing online, specifying the amount he requests and the purpose of the loan.\(^\text{13}\) Then the website decides on the interest rate for each listing posted. After seeing the interest rate, the borrower has the option to withdraw his listing before it is funded. Once the borrower decides to participate, it takes fourteen days for a listing to expire. Before the listing expires, lenders observe all posted information and decide whether or not to fund the loan. An example of a listing is shown in Figure 4. From this example, it is clear that lenders not only observe borrowers’ detailed credit profiles, but also their loan and payment histories. As long as the amount requested is reached, the listing is successfully funded and

---

\(^\text{10}\)For details of the company, see https://www.prosper.com/plp/about/.

\(^\text{11}\)The average number of investors for each loan is calculated using loans originated between January 2011 and December 2014.

\(^\text{12}\)For details of the fee structure, see https://www.prosper.com/help/contextual/fees/.

\(^\text{13}\)The borrower may also write a short paragraph about him/herself or about the description of the loan.
the loan is originated. In the following 12-60 months, the borrower needs to pay back the loan, while it is possible that defaults and/or late payments occur in the repayment process. Note that if borrowers default, only lenders bear the loss in this market.

Compared to traditional lending markets, Prosper, as a representative of other peer-to-peer lending marketplaces, has the following distinct features. From borrowers’ perspective, application requirements are easier to satisfy in online credit markets. As long as borrowers’ basic information is verified by the website, borrowers are allowed to post a listing online. Thus even borrowers with relatively low credit scores may obtain access to credit. For borrowers with good credit scores, they are charged lower interest rates by these markets due to lower operational costs.\textsuperscript{14} In addition, it is more convenient and much faster for borrowers to take a loan online than a personal loan in a bank. This is particularly the case when the amount of the loan is relatively small. From lenders’ perspective, there is no collateral required on Prosper, which may indicate a higher level of risk. However, due to the crowdfunding feature of this market, it is convenient for lenders to diversify their investment portfolios so as to reduce idiosyncratic risks. Moreover, this website adopts a harsh punishment scheme to disincentivize default. That is, borrowers who have defaulted once are not allowed to borrow from the website again.

2.2 Data Summary

The data used in this paper include all listings (some of which become loans) that were originated on Prosper between January 2011 and December 2014. The clean dataset I obtain contains 114,804 listings that come from 102,528 unique borrowers.\textsuperscript{15} Overall, about 67 percent of loans are used for debt consolidation, 7 percent for home improvement, 5 percent for business, and the rest for other purposes. 30 percent of borrowers have FICO scores below 600, and 94 percent of borrowers are employed.\textsuperscript{16} I characterize borrowers into different groups based on their repeated borrowing patterns. Borrowers in Category 1 appeared only once during the period I observe. Borrowers in Categories 2 – 4 appeared twice and are sorted by the different statuses of their first loans at the time the second loans appeared.

\textsuperscript{14}The average APR for credit cards is around 20%. For borrowers with credit scores higher than 600, 700 and 800, the average interest rates from Prosper are 16.92%, 13.70% and 9.76%, respectively.

\textsuperscript{15}The original dataset contains 192,916 listings. However, by November 8, 2016, there are still 63,790 ongoing loans that come from 62,841 unique borrowers. To ensure all loan outcomes are observed for each individual, I drop borrowers with ongoing loans. I also keep only one listing for each borrower within a short period (one month) to take care of the cases where borrowers may propose multiple listings for one monetary demand. I also drop borrowers with missing information. In addition, I focus on borrowers who have at most three listings, since the proportion of borrowers that have more than three listings is less than 1 percent.

\textsuperscript{16}The distribution of their stated monthly income is shown in Figure\textsuperscript{5}. Summary statistics of other variables used in the regressions or estimation are provided in Table\textsuperscript{8}. 

8
were originated. Specifically, their first loans may be paid off, still ongoing, or not funded (possibly withdrawn by themselves). Borrowers in Category 5 proposed three listings. The percentage of borrowers in each category is summarized in Table 1. I find that 89 percent of the borrowers appeared once, which indicates that the probability of receiving a future money demand shock for borrowers is approximately 11 percent.\textsuperscript{17} There is a small proportion of borrowers that appeared twice during the time period I observe. For borrowers who appeared twice with their first loans funded, about 60 percent of them proposed their second listings when the first loans were still ongoing.

<table>
<thead>
<tr>
<th>Data Category</th>
<th>Note</th>
<th>Freq.</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>appear once</td>
<td>91,891</td>
<td>89.63</td>
</tr>
<tr>
<td>2</td>
<td>appear twice: first loan is paid off</td>
<td>3,247</td>
<td>3.17</td>
</tr>
<tr>
<td>3</td>
<td>appear twice: first loan is ongoing</td>
<td>5,163</td>
<td>5.04</td>
</tr>
<tr>
<td>4</td>
<td>appear twice: first listing is withdrawn or unfunded</td>
<td>597</td>
<td>0.58</td>
</tr>
<tr>
<td>5</td>
<td>appear three times</td>
<td>1,630</td>
<td>1.59</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>102,528</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Figure 1 compares the distributions of credit grades for borrowers’ first and second listings. Prosper characterizes borrowers into seven credit groups, from AA (best) to HR (worst). From this figure, it is clear that a larger proportion of borrowers falls into better credit groups (include AA, A and B) in the second listings. This situation may be attributed to selection of borrowers across two loans. Borrowers I observe in the second listings must have no defaults in their first loans. This group of borrowers may be inherently better borrowers with higher credit grades. Alternatively, the shift of credit grade distribution may be driven by the updating of borrowers’ credit grades after the first loans’ outcomes realize. For instance, after paying back their loans, borrowers are very likely to be characterized into better groups by the reputation system. The selection and updating channels jointly determine the empirical pattern in Figure 1. To gain a better understanding of how the reputation system refines the pool of borrowers over time, in Section 5 I use my structural estimates to disentangle these two channels.

To better understand the differences between credit groups, Table 2 summarizes the average amount requested (in dollars), average interest rates, withdraw and funding probabilities, and default and late payment rates for borrowers by different credit categories.

\textsuperscript{17}According to the rule of this website, borrowers that default once cannot borrow again from this website, so I cannot observe the appearance of those borrowers when they receive future money demand shocks. Moreover, it is hard to rule out the possibility that borrowers may borrow from other places instead even if they need money. Therefore, 11 percent is a lower bound on the arrival rate of future money demand shocks.
Figure 1: Histograms of Credit Grade for Borrowers' First and Second Listings

Table 2: Summary Statistics by Credit Grades

<table>
<thead>
<tr>
<th>Credit Grades</th>
<th>AA</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>HR</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg. Amt. Requested($)</td>
<td>13250.31</td>
<td>12974.22</td>
<td>12982.66</td>
<td>11813.24</td>
<td>9179.09</td>
<td>5106.38</td>
<td>3586.34</td>
<td>10662.36</td>
</tr>
<tr>
<td>Withdraw Prob. (%)</td>
<td>6.39</td>
<td>5.56</td>
<td>5.66</td>
<td>5.52</td>
<td>7.77</td>
<td>6.13</td>
<td>6.13</td>
<td>11.26</td>
</tr>
<tr>
<td>Funding Prob. (%)</td>
<td>89.42</td>
<td>91.11</td>
<td>91.37</td>
<td>92.09</td>
<td>87.7</td>
<td>92.29</td>
<td>71.82</td>
<td>89.51</td>
</tr>
<tr>
<td>Default Prob. (%)</td>
<td>6.37</td>
<td>14.15</td>
<td>21.53</td>
<td>28.55</td>
<td>31.56</td>
<td>33.93</td>
<td>34.19</td>
<td>24.23</td>
</tr>
<tr>
<td>Late Payment Prob. (%)</td>
<td>2.64</td>
<td>5.61</td>
<td>8.58</td>
<td>11.02</td>
<td>15.47</td>
<td>16.5</td>
<td>20.86</td>
<td>10.75</td>
</tr>
<tr>
<td>Number of Obs.</td>
<td>8,231</td>
<td>21,166</td>
<td>22,271</td>
<td>24,964</td>
<td>18,046</td>
<td>12,196</td>
<td>7,930</td>
<td>114,804</td>
</tr>
</tbody>
</table>
Table 3: Regression Results of Interest Rate on Past Outcomes

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>borrower_rate</td>
<td>0.317***</td>
<td>0.303***</td>
<td>0.303***</td>
</tr>
<tr>
<td>(0.000986)</td>
<td>(0.00156)</td>
<td>(0.00156)</td>
<td></td>
</tr>
<tr>
<td>second_loan</td>
<td>-0.00902***</td>
<td>0.00157***</td>
<td>0.00157***</td>
</tr>
<tr>
<td>(0.000336)</td>
<td>(0.000415)</td>
<td>(0.000415)</td>
<td></td>
</tr>
<tr>
<td>overlap</td>
<td>0.00208***</td>
<td>0.00208***</td>
<td></td>
</tr>
<tr>
<td>(0.000409)</td>
<td>(0.000409)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>late_ever</td>
<td>0.00452***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.000663)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>16,820</td>
<td>8,410</td>
<td>8,410</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.932</td>
<td>0.937</td>
<td>0.937</td>
</tr>
</tbody>
</table>

Note: Control for Borrowers’ observables, year dummies and loan characteristics. Standard errors in parentheses, *** p < 0.01, ** p < 0.05, * p < 0.1.

There is a clear pattern for interest rates and default probabilities. Borrowers with lower credit grades are charged higher interest rates. Their default probabilities are also higher. This phenomenon intuitively captures the trade-off faced by lenders – higher risks must be compensated by higher returns.

### 2.3 Empirical Evidence on the Effect of Reputation

In this section, I provide empirical evidence on how market prices, borrowers’ behavior and loan outcomes are affected by the reputation system. I first investigate whether the interest rates charged on borrowers vary with past loan outcomes for a given individual. Table 3 presents regression results of interest rate on past loan outcomes controlling for borrowers’ observables, year dummies, and loan characteristics. Specifically, Column 1 shows that if borrowers have previously funded loans, the interest rates for their second loans are lower. Columns 2–3 further illustrate that the second loans which overlap with first loans or have late payments prior to them are associated with higher interest rates. These results indicate that the interest rate as a pricing device appears to “reward” prior successful loan repayment behavior, and “punish” late payments or uncertainty from overlapping loans.\(^\text{18}\)

Combining this with the institutional feature that once a borrower defaults, he is unable to

\(^{18}\)To alleviate the concern that borrowers may first request and pay off a small loan in order to get better deals for later larger loans, I run a regression of amount requested on whether a loan is the second one, controlling for a set of observables. The result in Table 9 illustrates that borrowers’ second loans do not constitute significantly larger requests.
Table 4: Logit Regression of Default on Whether the Loan is the Second-closed Loan

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>default</td>
<td>default</td>
<td>default</td>
<td>default</td>
</tr>
<tr>
<td>second_closed_loan</td>
<td>0.390***</td>
<td>0.359***</td>
<td>0.481***</td>
<td>0.456***</td>
</tr>
<tr>
<td></td>
<td>(0.0480)</td>
<td>(0.0731)</td>
<td>(0.0863)</td>
<td>(0.149)</td>
</tr>
<tr>
<td>borrower_rate</td>
<td>13.10***</td>
<td>12.66***</td>
<td>16.09***</td>
<td>17.68***</td>
</tr>
<tr>
<td></td>
<td>(1.347)</td>
<td>(1.979)</td>
<td>(2.238)</td>
<td>(3.678)</td>
</tr>
<tr>
<td>Constant</td>
<td>-4.572***</td>
<td>-4.659***</td>
<td>-5.960***</td>
<td>-6.472***</td>
</tr>
<tr>
<td></td>
<td>(0.414)</td>
<td>(0.608)</td>
<td>(0.706)</td>
<td>(1.136)</td>
</tr>
<tr>
<td>Second Loan for Debt Consolidation</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>FICO Below 600</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Long Gap Between Starting Dates</td>
<td></td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>10,166</td>
<td>4,630</td>
<td>3,550</td>
<td>1,696</td>
</tr>
</tbody>
</table>

Note: Control for Borrowers’ observables, year dummies and loan characteristics. Standard errors in parentheses, *** p < 0.01, ** p < 0.05, * p < 0.1.

borrow again, we can see that the reputation/feedback system imposes dynamic incentives on forward-looking borrowers through pricing schemes and entry restrictions.19

To further examine the causal impact of the reputation/feedback system on borrowers’ behavior and market outcomes, ideally we would have an experimental setting with one group given reputational incentives and another serving as a control group. Without such data, I find one institutional feature very useful: borrowers on this website could propose their second listings even when their first loans are still in process. This feature provides me with a group of borrowers who have two overlapping loans. I arrange the two loans for each borrower based on their closing dates. For the first- and the second-closed loans, the composition of borrowers remain the same. If borrowers do not react to reputational incentives (or are myopic), then the default behavior is mainly driven by types. Given the same group of borrowers, we would expect the default rates for the first- and the second-closed loans are the same. However, if borrowers react to dynamic incentives (or are forward-looking), once default occurs, borrowers, knowing that they cannot enter again, lose incentive to pay back their existing loans, which leads to a higher default rate for the second-closed loans.

Under either hypothesis, defaults for the first- and the second-closed loans are positively correlated.20 Column 1 in Table 4 shows that the second-closed loans have a significantly higher default rate after controlling for observables. Since I compare the average default...

19The rule that once-defaulted borrowers are prohibited from future credit access is inline with the incentive effects of terminations discussed in the work of Stiglitz and Weiss (1983).
20The positive correlation between the defaults in two loans is confirmed in Table 10 in the Appendix.
rates on two loans for the same group of borrowers, the higher default rate for the second-closed loans is not driven by the selection of borrowers. This result suggests that borrowers are likely to respond to reputational incentives.\footnote{Another explanation for the higher default rate of the second-closed loans is that borrowers are more likely to leave the market after the second loans, so they have less incentives to repay the second loans. This also suggests that borrowers are responsive to dynamic incentives.} As a robustness check, Column 2 restricts the sample to borrowers whose second loans are not for debt consolidation so as to alleviate concerns for re-financing. Column 3 further restricts the sample to borrowers whose initial FICO scores are below 600 in order to ensure that borrowers’ outside options cannot get much worse after defaulting on their first loans. Finally, Column 4 focuses on the group of borrowers with long gaps between the starting dates of their two loans. For this subsample, the issue that two correlated negative income shocks realize within a short period is mitigated somewhat. To summarize, this empirical exercise provides suggestive evidence that borrowers are responsive to the dynamic incentives imposed by the reputation/feedback system. When the incentives to repay debts are reduced, default rates significantly increase.

3 Model

In this section, I develop a finite-horizon dynamic model of a credit market in which borrowers and lenders interact repeatedly over time. I first describe main players – borrowers, lenders and the website – and then analyze their strategies. The equilibrium of the model is discussed last.

3.1 Basic Set-up

3.1.1 Borrowers

In the model, borrowers are dynamic and discount the future at the rate of $\delta$. Borrowers are heterogeneous in default cost $c \in \Theta_c$, which is their private information.\footnote{There are different ways to introduce borrower heterogeneity in observationally equivalent models. I follow Kawai et al (2016) to interpretate $c$ as borrower’s default cost.} If types are discrete, $\Theta_c = \{c_1, c_2, \cdots, c_J\}$; for continuous types, $\Theta_c = [c, \bar{c}]$. $F_c(\cdot) : \Theta_c \to [0,1]$ represents the cumulative distribution function of the default cost. The distribution is public knowledge. Borrowers may want to take out multiple loans. Let $t = 1, 2, \cdots, T$ index for each loan period of a borrower. Borrowers’ utility associated with a given loan period of a borrower. Borrowers’ utility associated with a given loan period of a borrower. Borrowers’ utility associated with a given loan period of a borrower. Borrowers’ utility associated with a given loan period of a borrower. Let $i = 1, 2, \cdots, N$ index each borrower. The rest of this section considers borrower $i$. To simplify the notation, I drop subscript $i$. 

...
Once the borrower applies for a loan, he is assigned an interest rate $r_t$ by the website and chooses effort $e_t \in [0, \infty]$ to maximize his expected payoff. Let $R_t \in [0, \bar{R}]$ denote the revenue of a given loan. The revenue is scaled by the size of the loan. The borrower’s effort $e_t$ affects the distribution of $R_t$ in a stochastic way. Let $F_{R_t|e_t}(\cdot|e_t)$ represent the revenue distribution conditional on effort. Assume that for any $e_2 > e_1$, $F_{R_t|e_2}(\cdot|e_2)$ first order stochastically dominates $F_{R_t|e_1}(\cdot|e_1)$. The borrower incurs cost $\phi(e_t, \theta_t)$ when exerting effort $e_t$. $\theta_t \in \Theta$ is the transitory shock to the cost of effort that realizes for each loan $t$. The distribution of $\theta_t$ is stationary and related to borrowers’ default costs. Use $F_{\theta_t|c}(\cdot|c) : \Theta \rightarrow [0, 1]$ to represent the distribution of $\theta_t$ conditional on $c$. Assume that $\phi'(\cdot, \theta_t) > 0$, and $\phi''(\cdot, \theta_t) > 0$.

For each loan $t$, the borrower draws an outside option $v_{0,t}$ from $F_v(\cdot)$, the mean of which is $\bar{v}_0$. If the borrower’s expected payoff from having the loan is lower than $v_{0,t}$, he withdraws the listing and takes the outside option. Otherwise the borrower posts the listing online. If his listing is funded by lenders, the borrower exerts effort $e_t$. If not funded, the borrower takes the outside option. Use $W_t = 1$ to represent the case when the borrower withdraws his listing. $W_t = 0$ then represents the case when the borrower stays on the market (or participates) after observing interest rate $r_t$. Let $I_t = 1$ denote the case when the listing is funded by lenders; $I_t = 0$ otherwise. For each funded loan, the outcomes include the borrower’s default and late payment performances. Let $D_t = 1$ denote the case when default occurs for loan $t$; $D_t = 0$ otherwise. Assume the borrower’s default cost is large enough that if $R_t \geq 1 + r_t$, $D_t = 0$, and if $R_t < 1 + r_t$, $D_t = 1$. Let $L_t = 1$ denote the case when late payment occurs for loan $t$; $L_t = 0$ otherwise. Assume that for any $e_2 > e_1$, $Pr(L_t = 1|e_2) < Pr(L_t = 1|e_1)$.

The belief about the borrower’s type when he applies for loan $t$ is public information, which is maintained by the website and is easily accessible by lenders. This belief is an aggregate state variable that depends on the borrower’s history. Assume that its transition follows a first-order Markov process. That is, the belief about the borrower’s type when he applies for loan $t$ depends on the previous belief and his history at loan $t - 1$. Define $\mathcal{H} = \{nl, l, np\}$ as a collection of realized loan outcomes. Let $H_{t-1} \in \mathcal{H}$ denote the borrower’s history at loan $t - 1$. Specifically, $H_{t-1} = nl$, or $l$ represent the cases when the borrower paid off his debts and had no late payments or had late payments at loan $t - 1$, respectively. $H_{t-1} = np$ represents the case when the borrower did not participate (withdrew the listing by himself or was not funded by lenders) at $t - 1$.

### 3.1.2 Lenders and the Website

I summarize the supply side of this market by assuming that lending services are provided by a risk-neutral, representative lender, who is a one-time player and cares about the expected
payoff of each project. The simplification of the lender’s side model is mainly because individual lenders’ investment strategies are not observed. Instead, only the final funding decisions are recorded. The crowdfunding feature of this market provides a rationale for this simplification.\footnote{The final funding decisions are approximated by aggregating the funding decisions of small individual investors. Consider the case where a borrower requests $M$ dollars from the website. There are $\bar{n}$ investors in total, and each investor chooses to invest $m$ dollars or not. As a result, for the project to be funded, we need at least $n = \frac{M}{m}$ lenders to invest. The probability that each lender invests is $p_i$, which depends on the expected payoff from the project monotonically. Let $X$ denote the number of lenders that choose to invest. $X \sim N(\bar{n}p_i, \bar{n}p_i(1 - p_i))$ when $\bar{n}$ is large enough (by Central Limit Theorem). Therefore, $Pr(X \geq n) = 1 - \Phi(\frac{n - \bar{n}p_i}{\sqrt{\bar{n}p_i(1 - p_i)}})$. The final funding probability is again a monotone function of the expected return from the project if $n - \bar{n}p_i > 0$ and $p_i < 0.5$.} Since individual lenders usually invest a small amount (as little as $25$) on a specific project, they are likely to be risk-neutral for each project. As a result, the risk-neutral assumption for the representative lender is not very restrictive.

For each loan $t$, the representative lender draws an outside option $\mu_{0,t}$ from $F_\mu(\cdot)$. The lender does not observe default costs of borrowers, but he has access to public information on borrowers’ history, which allows him to form beliefs about borrowers’ default cost distribution. Denote the lender’s belief about the borrower’s default cost when he applies for loan $t$ as $S_t \in S$. Specifically,

1. When no histories are available ($t = 1$), beliefs coincide with the population distribution of borrowers’ default cost. $S_t = F_c(\cdot) : \Theta_c \to [0, 1]$.

2. For loan $t \in \{2, \cdots, T\}$, beliefs are updated based on different loan outcomes for any given interest rate $r_{t-1}$. $S_t(S_{t-1}, r_{t-1}, H_{t-1}) = F_c|_{S_{t-1}, r_{t-1}, H_{t-1}}(\cdot|S_{t-1}, r_{t-1}, H_{t-1}) : \Theta_c \to [0, 1]$.

Since the lender makes investment decisions after observing borrowers’ participation, the beliefs are further updated. The loan is funded as long as the expected revenue for the lender exceeds his outside investment option $\mu_{0,t}$.

For all loan listings, the website adopts a decision rule to set interest rates which depends on beliefs about borrowers’ default cost distribution. The pricing rule is public knowledge. Here, I take the website’s pricing scheme as given. Later in the counterfactual analysis, I will be specific about the website’s objective function.

### 3.2 Borrowers’ Effort Choices

For loans $t$, borrowers choose effort levels to maximize their expected payoffs once their loans are originated. Borrowers also need to consider whether or not to withdraw their listings.\footnote{I admit that this simplification ignores potential competition and herding effect on lender’s side. See Zhang and Liu (2012).}
after observing the interest rates assigned by the website. The state of a borrower’s decision problem includes the belief about his default cost distribution $S_t$, interest rate $r_t$, cost of effort $\theta_t$, and an outside option draw $v_{0,t}$.

3.2.1 The Last Loan

In the last loan, borrowers are myopic. For a borrower with default cost $c$, his expected payoff from having loan $T$ with interest rate $r_T$ and exerting effort level $e$ with the cost of effort $\theta_T$ is defined in the following equation.

\[
\tilde{V}_{B,c,T}(e, r_T, \theta_T) = \int_{1+r_T}^{R} U(R - 1 - r_T)dF_{R_t|e}(R|e) + \int_{0}^{1+r_T} U(R - c)dF_{R_t|e}(R|e) - \phi(e, \theta_T).
\] (3.1)

The borrower decides his optimal effort level for loan $T$ through the following payoff maximization problem:

\[
e_{c,T}(r_T, \theta_T) = \arg \max_e \tilde{V}_{B,c,T}(e, r_T, \theta_T).
\] (3.2)

Before exerting effort, the borrower decides whether or not to withdraw the listing he has proposed. The borrower’s choice-specific (withdraw the listing or not) value function is as follows.

\[
\bar{V}_0^{B,c,T}(r_T, \theta_T) = \max_e \tilde{V}_{B,c,T}(e, r_T, \theta_T);
\]

\[
\bar{V}_1^{B,c,T}(v_{0,T}) = v_{0,T}.
\] (3.3)

Equation (3.3) implies that (1) if the borrower stays on the market, he will choose the optimal effort to maximize his expected payoff; (2) if the borrower withdraws his listing, he will get the outside option $v_{0,T}$. For the borrower with type $c$, the probability that he participates is

\[
Pr(W_{c,T} = 0|r_T, \theta_T) = Pr(\bar{V}_0^{B,c,T}(r_T, \theta_T) > v_{0,T}) = F_e(\bar{V}_0^{B,c,T}(r_T, \theta_T)).
\] (3.4)

Even if the borrower decides to stay on the market, he may not be funded by lenders with some probability. If this occurs, the borrower takes the outside option. The ex-ante value function for the borrower before the withdrawal and investment decisions are made is

\[
V_{B,c,T}(S_T, r_T, \theta_T, v_{0,T}) = Pr(I_T = 1|S_T, r_T, W_T = 0) \max_{W_T=0,1} \{\bar{V}_0^{B,c,T}(r_T, \theta_T), \bar{V}_1^{B,c,T}(v_{0,T})\} + Pr(I_T = 0|S_T, r_T, W_T = 0)\bar{V}_1^{B,c,T}(v_{0,T}).
\] (3.5)

In Equation (3.5), $Pr(I_T = 1|S_T, r_T, W_T = 0)$ represents the probability that loan $T$ is funded given the belief $S_T$ and the interest rate $r_T$ if the borrower does not withdraw his
listing. I will derive the funding probability when discussing the lender’s side problem in Section 3.3.

### 3.2.2 Preceding Loans

In loans \( t < T \), borrowers are forward-looking and care about their future continuation values. If a borrower with type \( c \) has no defaults on previous loans, his expected payoff from loan \( t \) with belief \( S_t \), interest rate \( r_t \), and effort level \( e \) with the cost of effort \( \theta_t \) is defined in the following equation.

\[
\tilde{V}_{B,c,t}(e, S_t, r_t, \theta_t) = \int_{1+r_t}^{R} [U(R - 1 - r_t)]
+ \Pr(L_t = 0|e)\delta EV_{B,c,t+1}(S_{t+1}(S_t, r_t, H_t = nl), r_{t+1}, \theta_{t+1}, v_{0,t+1})
+ \Pr(L_t = 1|e)\delta EV_{B,c,t+1}(S_{t+1}(S_t, r_t, H_t = l), r_{t+1}, \theta_{t+1}, v_{0,t+1})]dF_{R|e}(R|e)
+ \int_{0}^{1+r_t} [U(R - c) + \sum_{\tau=t+1}^{T} \delta^{\tau-t}\tilde{v}_0]dF_{R|e}(R|e) - \phi(e, \theta_t) \tag{3.6}
\]

Effort level \( e \) enters the borrower’s expected payoff through three channels. First, it affects the revenue distribution stochastically. If the realized revenue is enough to pay off the loan, the borrower with high default cost will repay his debts. If the revenue is not sufficient to repay the debts, the borrower defaults. If default occurs, the borrower pays the default cost for loan \( t \), and is not allowed to enter the market again. For \( \tau > t \), the borrower takes the average of the outside option distribution in expectation. Second, effort level \( e \) affects the realization of late payments. Conditional on higher effort, there is a lower chance that late payments occur. As another dimension of loan outcomes, late payments affect lenders’ future beliefs about the borrower’s type, which enters the expected value function at \( t + 1 \).

When the borrower does not default and has no late payments at loan \( t \), his history \( H_t = nl \), and the future belief \( S_{t+1}(S_t, r_t, H_t = nl) = F_{c|S_t,r_t,H_t}(\cdot|S_t, r_t, H_t = nl) \). If late payments occur at loan \( t \), the history becomes \( H_t = l \), and the future belief is affected accordingly. Note that the borrower’s cost of effort and outside option draws are assumed independent across different loans, so the expected value function does not depend on \( \theta_t \) and \( v_{0,t} \). Third, the effort level induces cost \( \phi(e, \theta_t) \). The borrower decides his optimal effort level for loan \( t \) through the following payoff maximization problem:

\[
e_{c,t}(S_t, r_t, \theta_t) = \arg\max_e \tilde{V}_{B,c,t}(e, S_t, r_t, \theta_t). \tag{3.7}
\]

Before exerting effort, the borrower decides whether or not to withdraw the listing he
has proposed at $t$. I then define the borrower’s choice-specific (withdraw the listing or not) value function as follows.

$$
\bar{V}_{B,c,t}^0(S_t, r_t, \theta_t) = \max_e \tilde{V}_{B,c,t}(e, S_t, r_t, \theta_t)
$$

$$
\bar{V}_{B,c,t}^1(S_t, r_t, v_{0,t}) = v_{0,t} + \delta \mathbb{E} V_{B,c,t+1}(S_{t+1}(S_t, r_t, H_t = np), r_{t+1}, \theta_{t+1} + v_{0,t+1})
$$

Equation (3.8) implies that (1) if the borrower stays on the market, he will choose the optimal effort to maximize his expected payoff, and (2) if the borrower withdraws his listing, he will get the outside option $v_{0,t}$ plus the future continuation value conditional on having history $H_t = np$. For the borrower with type $c$, the probability that he participates at loan $t$ is therefore

$$
Pr(W_{c,t} = 0|S_t, r_t, \theta_t) = Pr(\bar{V}_{B,c,t}^0(S_t, r_t, \theta_t) > \bar{V}_{B,c,t}^1(S_t, r_t, v_{0,t})).
$$

Again, following the same argument for the last loan, I construct the ex-ante value function for the borrower before the withdrawal and investment decisions are made.

$$
V_{B,c,t}(S_t, r_t, \theta_t, v_{0,t}) = Pr(I_t = 1|S_t, r_t, W_t = 0) \max_{W_t=\{0,1\}} \{\bar{V}_{B,c,t}^0(S_t, r_t, \theta_t), \bar{V}_{B,c,t}^1(S_t, r_t, v_{0,t})\}
$$

$$
+ Pr(I_t = 0|S_t, r_t, W_t = 0) \bar{V}_{B,c,t}^1(S_t, r_t, v_{0,t}).
$$

(3.10)

### 3.3 Lenders’ Investment Decisions

The representative lender makes investment decisions by comparing the expected revenues of loans with his outside options. The lender makes decisions after observing the participation of borrowers. As a result, the lender’s belief about the borrower’s type is updated. Denote the joint density of $(c, \theta_t)$ conditional on observing the borrower’s participation as

$$
f_{c,\theta_t|W_t=0,S_t,r_t}(c, \theta_t) = \frac{Pr(W_{c,t} = 0|S_t, r_t, \theta_t) f_{\theta_t|c}(\theta_t|c) s_t(c)}{\int_c \int_{\theta_t} Pr(W_{c',t} = 0|S_t, r_t, \theta') dF_{\theta_t|c}(\theta'|c') dS_t(c')}.
$$

(3.11)

where the participation probabilities for the borrower at different loan specifications are derived in Equations (3.4) and (3.9). In Equation (3.11), $s_t(c)$ represents the probability density of type $c$ given belief $s_t$. That is, $s_t(c) = S_t(c)$.

For any given belief $s_t$ and interest rate $r_t$, the lender’s revenue conditional on $(\theta_t, c)$ is derived in the following equation:

$$
\bar{\pi}_{L,t}(S_t, r_t, \theta_t, c, W_t = 0) = \int_{1+r_t}^{\hat{R}} (1 + r_t) dF_{R|c,t}(s_t, r_t, \theta_t)(R|c,t)\pi_{c,t}(S_t, r_t, \theta_t)) - 1,
$$

(3.12)

which implies that, whenever the borrower pays back the loan, the lender receives $1 + r_t$. The probability that the borrower pays off his debt depends on the effort level $e_{c,t}(S_t, r_t, \theta_t)$.
Since the lender does not observe the borrower’s default cost \( c \) and cost of effort \( \theta_t \), the lender computes his expected revenue by weighting the revenue for each type in Equation (3.12) using the updated belief about the borrower’s type derived in Equation (3.11).

\[
\pi_{L,t}(S_t, r_t, W_t = 0) = \int_{c'} \int_{\theta'} \tilde{\pi}_{L,t}(S_t, r_t, \theta', c', W_t = 0)f_{c,\theta|W_t=0,S_t,r_t}(c',\theta'|W_t = 0, S_t, r_t)d\theta'dc'.
\]

(3.13)

The lender invests in the project if \( \pi_{L,t}(S_t, r_t, W_t = 0) > \mu_{0,t} \) and the funding probability is therefore \( Pr(I_t = 1|S_t, r_t, W_t = 0) = F_{\mu}(\pi_{L,t}(S_t, r_t, W_t = 0)) \).

### 3.4 Equilibrium

In this section, I characterize the equilibrium of the model. I first summarize the strategies for borrowers and lenders. A pure strategy for the borrower with type \( c \in \Theta_c \) specifies an effort choice and a withdrawal decision. Specifically, \( e_{c,t} : \mathcal{S} \times [0, \bar{r}] \times \Theta_\theta \rightarrow [0, \infty] \) specifies the borrower’s effort choice decision for a belief, an interest rate he is assigned, and a cost of effort draw. \( W_{c,t} : \mathcal{S} \times [0, \bar{r}] \times \Theta_\theta \times \mathbb{R} \rightarrow \{0,1\} \) specifies the borrower’s binary choice of participation for a given state (including a belief, an interest rate, a cost of effort draw, and an outside option). A pure strategy for the lender \( I_t : \mathcal{S} \times [0, \bar{r}] \times \mathbb{R} \rightarrow \{0,1\} \) specifies his binary investment decision for a belief, an interest rate, and an outside investment option.

For any given pricing rule of the website, a pure strategy Bayesian Nash equilibrium for the game consists of borrowers’ effort choice strategies \( e_{c,t} : \mathcal{S} \times [0, \bar{r}] \times \Theta_\theta \rightarrow [0, \infty] \) and participation strategies \( W_{c,t} : \mathcal{S} \times [0, \bar{r}] \times \Theta_\theta \times \mathbb{R} \rightarrow \{0,1\} \) for each type \( c \), and the lender’s investment strategies \( I_t : \mathcal{S} \times [0, \bar{r}] \times \mathbb{R} \rightarrow \{0,1\} \), for each loan \( t \), such that:

1. For the borrower with type \( c \), for any belief \( S_t \), interest rate \( r_t \), and cost of effort \( \theta_t \), he chooses the optimal effort level, and participates only if the expected payoff from participation exceeds the value of withdrawing the listing at \( t \).

2. Given the optimal effort choices and participation strategies for borrowers with each type, for any belief \( S_t \) and interest rate \( r_t \) the lender invests only if the expected revenue of the listing exceeds the draw of his outside investment options.

3. Borrowers’ strategies are consistent with the beliefs. For any \( S_{t-1} \) and \( r_{t-1} \), the belief about the borrower’s type conditional on history \( H_{t-1} \) is formed through a Bayesian updating process that involves borrowers’ optimal effort choices. For each \( c \in \Theta_c \),

\[
f_{c|S_{t-1}, r_{t-1}, H_{t-1}}(c|S_{t-1}, r_{t-1}, H_{t-1}) = \frac{\int_{c'} f_{H_{t-1}|S_{t-1}, r_{t-1}, c'}(H_{t-1}|S_{t-1}, r_{t-1}, c)s_{t-1}(c)}{\int_{c'} \int_{H_{t-1}|S_{t-1}, r_{t-1}, c'}(H_{t-1}|S_{t-1}, r_{t-1}, c')dS_{t-1}(c')}.
\]

(3.14)
where
\[ f_{H_{t-1}|S_{t-1},r_{t-1},c}(H_{t-1}|S_{t-1},r_{t-1},c) = \int_{\theta'} Pr(H_{t-1}|S_{t-1},r_{t-1},\theta',c) dF_{\theta'|c}(\theta'|c). \]

The following equation derives the probability of observing \( H_{t-1} = nl \), which represents the case where \( W_{t-1} = 0, I_{t-1} = 1, D_{t-1} = 0, \) and \( L_{t-1} = 0 \) i.e., the borrower participates, is funded, and pays off the debts with no late payments at loan \( t-1 \).

\[
Pr(H_{t-1} = nl|S_{t-1},r_{t-1},\theta',c) = Pr(W_{t-1} = 0, I_{t-1} = 1, D_{t-1} = 0, L_{t-1} = 0|S_{t-1},r_{t-1},\theta',c) \\
= (1 - F_{R_{t-1}|e_{c,t-1}}(1 + r_{t-1}|e_{c,t-1}(S_{t-1},r_{t-1},\theta'))) Pr(L_{t-1} = 0|e_{c,t-1}(S_{t-1},r_{t-1},\theta')) \\
Pr(W_{c,t-1} = 0|S_{t-1},r_{t-1},\theta') Pr(I_{t-1} = 1|S_{t-1},r_{t-1},W_{t-1} = 0) \\ 
\] (3.15)

In Equation (3.15), the effort level \( e_{c,t-1}(S_{t-1},r_{t-1},\theta') \) drives the realization of loan outcomes, which further affects the updating process of beliefs. The derivation of equilibrium conditions when \( H_{t-1} = l \) or \( H_{t-1} = np \) is similar. In Appendix A, I sketch the proof of the existence of the equilibrium in a simplified setting. The model can be solved using backward induction.

4 Identification

In this section, I provide nonparametric identification strategies to recover the joint distribution of the borrower’s type and the shock to his cost of effort. I also identify utility parameters in borrowers’ payoff functions and the outside option distributions for borrowers and lenders under certain parametric assumptions.

The difficulties in identification are two-fold. First, borrower’s types and effort levels, which jointly determine loan outcomes, are both unobserved. Borrower’s type is persistent across loans, while effort levels are also affected by cost of effort, which realizes for each loan. To disentangle these two unobserved sources, I first use transitions of state variables across periods to pin down the type distribution for borrowers with multiple loans. I then rely on variations in default and late payment performances within a period to recover the distribution of cost of effort. Using variations in interest rates, I am able to further identify utility primitives. The second difficulty comes from the fact that borrowers enter the market again only if they have no defaults in their previous loans. That is, the sample of borrowers with multiple loans is endogenously selected. I match the default probability from the model with that from the data before selection occurs to recover the original type distribution.
4.1 Data and Primitives

I now fix the notation for the observed data and model primitives to be identified. The rest of this section is for borrower \(i = 1, 2, \cdots, N\). I drop the subscript \(i\) to simplify the notation. Denote the vector of observed state variables for loan \(\tau\) as

\[ O_\tau = \{r_\tau, D_\tau, L_\tau, X_\tau, K_\tau, W_\tau, I_\tau\}. \]

\(r_\tau\) represents the interest rate. \(D_\tau = 1\) when the borrower defaults, and 0 otherwise. \(L_\tau = 1\) when late payment occurs, and 0 otherwise. \(X_\tau\) and \(K_\tau\) represent the borrower’s financial status (e.g. debt-to-income ratio, homeownership, etc.) and credit grade, respectively. \(W_\tau = 1\) means that the borrower withdraws the loan, and \(I_\tau = 1\) means that the loan is funded by lenders. The borrower’s history is observed lenders and the website. If the borrower defaults at loan \(\tau\), there will be no future loans. For those who do not default at \(\tau\), define

\[
H_\tau = \begin{cases} 
  nl & \text{if } W_\tau = 0, I_\tau = 1, D_\tau = 0, \text{ and } L_\tau = 0 \\
  l & \text{if } W_\tau = 0, I_\tau = 1, D_\tau = 0, \text{ and } L_\tau = 1 \\
  np & \text{if } W_\tau = 1, \text{ or } W_\tau = 0, I_\tau = 0
\end{cases}.
\]

Let \(\Omega_\tau = \{r_\tau, X_\tau, K_\tau\}\) contain all observables not involved in \(H_\tau\). In the data, I observe \(\{O_\tau\}_{\tau=t-1,t}\) for each borrower.

There are two groups of primitives to be identified. The parameters in the first group, including the joint distribution of the borrower’s private type and cost of effort as well as the state transition probabilities conditional on each type, are identified nonparametrically. In this section, I consider a case where borrowers have discrete types \(c \in [c_1, c_2, \cdots, c_J]\) and the cost of effort \(\theta_\tau\) takes value from \([\theta_1, \theta_2, \cdots, \theta_J]\). It is easy to generalize the identification strategies for the cases where \(c\) and \(\theta_\tau\) are continuous. However, for illustration purposes, I stick to discrete types in this section. Let \(F_{X_{\tau+1}|X_\tau,c}\) denote the type-specific transition probabilities of the borrower’s financial status, which are also recovered nonparametrically.\(^{25}\)

For the second group, which includes the borrower’s utility primitives and the outside option distributions of borrowers and lenders, I impose the following parametric assumptions. Specifically, I assume the borrower has a CARA utility function, \(U(x) = 1 - \exp(-\alpha x)\), where \(\alpha\) is the risk-aversion parameter. The borrower’s cost function is \(\phi(e, \theta_\tau) = \theta_\tau e^2\), where \(e\) is the effort level exerted. The following assumption simplifies the revenue realization process:

\(^{25}\)In the rest of this paper, I drop the realization of the random variables in the joint pdf’s to simplify the notation. Specifically, for any vector of random variables \(Y = (Y_1, Y_2)\), let \(f_{Y_1, Y_2} = f_{Y_1, Y_2}(y_1, y_2)\) denote the joint density of \(Y\). Moreover, for the density of \(Y_1\) conditional on \(Y_2 = y_2\), use \(f_{Y_1|Y_2=y_2}\) (or \(f_{Y_1|y_2}\)) to represent \(f_{Y_1|Y_2=y_2}(y_1|y_2)\).
I assume that for each loan there are two possible revenues that may realize, $R_h$ and $R_l$. The probability that $R_h$ is realized is $p(e) = 1 - \exp(-\beta e)$, and the probability that no late payments appear is $L(e) = 1 - \exp(-\gamma e)$. Assume the borrower’s outside option $v_{0,\tau}$ follows a normal distribution with mean $v(X_{\tau}, K_{\tau}) = \bar{v}_0 + v_x X_{\tau} + v_k K_{\tau}$ and variance equal to 1. The lender’s outside options are drawn from $N(\mu_0, \sigma_0)$. To summarize, the parameters of interest in the model include $\{F_c, F_{\theta_\tau} | c, F_{X_{\tau+1} | X_{\tau}, c}, \alpha, R_h, R_l, \beta, \gamma, \bar{v}_0, v_x, v_k, \mu_0, \sigma_0\}$.

4.2 Nonparametric Identification of Private Information Distribution

In this section, I first use transitions of state variables across periods to pin down the type distribution for borrowers with multiple loans. From the empirical setting, I find three pieces of information that are related to the borrower’s unobserved and persistent type, namely, borrower’s initial characteristics, transition of his financial status, and outcomes of the loan. These three measurements are assumed independent conditional on the latent type. I use the dynamic structure of the model in Section 3 to justify this assumption. Following similar strategies to those found in Hu and Shum (2012), I recover the underlying type distribution for the selected group of borrowers. I then rely on variations in default and late payment performances to recover the distribution of cost of effort conditional on each type.

For identification purposes, I impose the following assumption on the dynamic process.

**Assumption 1.** The dynamic process of $\{O_t, \theta_t, c\}$ satisfies

(1) the first order Markov process;

(2) conditional independence:

\[
\begin{align*}
\int_{D_t, L_t, I_t=1, W_t=0, \Omega_t, H_{t-1}, \theta_{t-1}, c} f_{D_t, L_t, I_t=1, W_t=0, \Omega_t, H_{t-1}, \theta_{t-1}, c} &= \int_{D_t} f_{D_t} | r_t, \theta_t, c \cdot \int_{L_t} f_{L_t} | r_t, \theta_t, c \cdot \int_{I_t=1} f_{I_t=1} | W_t=0, \Omega_t, H_{t-1} \\
&\quad \cdot \int_{W_t=0} f_{W_t=0} | \Omega_t, \theta_t, c \cdot \int_{X_t} f_{X_t} | X_{t-1}, c \cdot \int_{K_t} f_{K_t} | K_{t-1}, H_{t-1} \cdot \int_{\theta_t} f_{\theta_t} | c \\
&= \int_{D_t} f_{D_t} | r_t, \theta_t, c \cdot \int_{L_t} f_{L_t} | r_t, \theta_t, c \cdot \int_{I_t=1} f_{I_t=1} | W_t=0, \Omega_t, H_{t-1} \cdot \int_{W_t=0} f_{W_t=0} | \Omega_t, \theta_t, c \cdot \int_{X_t} f_{X_t} | X_{t-1}, c \cdot \int_{K_t} f_{K_t} | K_{t-1}, H_{t-1} \cdot \int_{\theta_t} f_{\theta_t} | c.
\end{align*}
\]

Assumption 1 has several important implications. First, conditional on effort level for loan $t$, $D_t$ and $L_t$ are independent. This assumption is directly motivated by the model, since conditional on effort levels, the realizations of defaults and late payments are driven by independent shocks.\(^\text{26}\) Second, interest rates and probabilities that projects are funded are only associated with borrowers’ characteristics and histories. In other words, $\int_{I_t=1} f_{I_t=1} | W_t=0, \Omega_t, H_{t-1}$

\(^\text{26}\)The probabilities that default and late payments appear are allowed to be related to some other observables. For illustration of the key identifying restrictions, in this section I focus on the case when the realization of defaults and late payments are only driven by effort and some independent shocks.
and $f_{rt|X_t, K_t, H_{t-1}}$ can be directly estimated from the data. This assumption is consistent with the institutional setting. Under asymmetric information, the website and the lender make decisions only based on observables. Moreover, the probability that a borrower participates depends on his expected payoff from the loan and the outside option distribution, the mean of which is shifted by observed characteristics. The transition of the borrower’s financial status $X_t$ is governed by $c$, but the update of his credit grade $K_t$ on this website is mainly driven by his history. That is to say, $f_{K_t|K_{t-1}, H_{t-1}}$ can be also recovered from the data. In addition, based on the model in Section 3, borrower’s type relates to the distribution of the cost of effort $\theta_t$.

Assumption 1 also implies that $\theta_{t-1}$ does not enter the Markov transition kernel. As a result, $\theta_t$ can be integrated out of $f_{D_t, L_t, I_t=1, W_t=0, \Omega_t, \theta_t|H_{t-1}, X_{t-1}, K_{t-1}, c}$ without losing Markovian properties.

$$f_{D_t, L_t, I_t=1, W_t=0, \Omega_t|H_{t-1}, X_{t-1}, K_{t-1}, c} = \sum_{\theta_t} f_{D_t, L_t, I_t=1, W_t=0, \Omega_t, \theta_t|H_{t-1}, X_{t-1}, K_{t-1}, c}$$

$$= f_{D_t, L_t, I_t=1, W_t=0|\Omega_t, H_{t-1}, c} \cdot f_{r_t|X_t, K_t, H_{t-1}}$$

$$\cdot f_{X_t|X_{t-1}, c} \cdot f_{H_t-1|X_{t-1}, K_{t-1}, c}. \quad (4.1)$$

Then the joint distribution of observables for borrowers who have loans at $t-1$ and $t$ (i.e. borrowers with $I_t = 1, W_t = 0$ who have no defaults at loan $t-1$) helps to identify the default cost distribution and the state transition probabilities through the following equation.

$$\frac{f_{D_t, L_t, I_t=1, W_t=0, \Omega_t|H_{t-1}, X_{t-1}, K_{t-1}, c}}{f_{r_t|X_t, K_t, H_{t-1}} \cdot f_{K_t|K_{t-1}, H_{t-1}}} = \sum_c f_{D_t, L_t, I_t=1, W_t=0|\Omega_t, H_{t-1}, c} \cdot f_{X_t|X_{t-1}, c} \cdot f_{H_t-1|X_{t-1}, K_{t-1}, c}. \quad (4.2)$$

Equation (4.2) serves as the key identifying equation for kernels related to the latent type. Intuitively, this equation provides three pieces of information that are independent conditional on $c$. Specifically, type is related to the outcome of the second loan after integrating out effort, type affects the transition process of financial status, and the borrower’s initial characteristics also reveal some information about his type. These three pieces of information are treated as “measurements” of the unobserved type.

Suppose $D_t$, $K_{t-1}$, and $c$ are all discrete random variables. Use $j_d = 1, \cdots, J_d$, $j_k = 1, \cdots, J_k$ and $j_c = 1, \cdots, J_c$ to index the values of these three variables. I consider the case where $I_t = 1, W_t = 0$, and $\{r_t, K_t, L_t, H_{t-1}\}$ are fixed, and define the following matrices for
a pair of \((X_t, X_{t-1})\).

\[
M_{D_t,K_{t-1},X_t,X_{t-1}} = \left[ \frac{f_{D_t,L_t,t=1,W_t=0|\bar{r}_t,X_t,K_t,H_{t-1},X_{t-1},K_{t-1}}}{f_{\bar{r}_t|X_t,K_t,H_{t-1}}} \right]_{j_d,j_k},
\]

\[
M_{D_t,c,X_t} = \left[ f_{D_t,L_t,t=1,W_t=0|\bar{r}_t,X_t,K_t,R_{t-1},c} \right]_{j_d,j_c},
\]

\[
M_{c,X_t,X_{t-1}} = \text{diag}\left\{ f_{X_t|X_{t-1},c} | c=j_c \right\}_{c=1,2,\ldots,J_c},
\]

\[
M_{c,K_{t-1},X_{t-1}} = \left[ f_{H_{t-1},X_{t-1},K_{t-1},c} | c=j_c, K_{t-1}=j_k \right]_{j_c,j_k}.
\]

The matrix form of Equation (4.2) for any \((X_t, X_{t-1})\) is therefore

\[
M_{D_t,K_{t-1},X_t,X_{t-1}} = M_{D_t,c,X_t} M_{c,X_t,X_{t-1}} M_{c,K_{t-1},X_{t-1}}.
\]

Given four combinations of \((X_t, X_{t-1})\), namely \((\bar{X}_t, \bar{X}_{t-1})\), \((\hat{X}_t, \hat{X}_{t-1})\), \((\tilde{X}_t, \tilde{X}_{t-1})\), \((\bar{X}_t, \bar{X}_{t-1})\), I construct the following equations following Hu and Shum (2012).

\[
\left( M_{D_t,K_{t-1},X_t,\bar{X}_{t-1}} \cdot M_{D_t,K_{t-1},X_t,\bar{X}_{t-1}}^{-1} \right) \left( M_{D_t,K_{t-1},\hat{X}_t,\bar{X}_{t-1}} \cdot M_{D_t,K_{t-1},\hat{X}_t,\bar{X}_{t-1}}^{-1} \right)^{-1}
= M_{D_t,c,\bar{X}_t} \left( M_{c,\bar{X}_t,\bar{X}_{t-1}} \cdot M_{c,\hat{X}_t,\bar{X}_{t-1}}^{-1} \cdot M_{c,\bar{X}_t,\bar{X}_{t-1}} \cdot M_{c,\hat{X}_t,\bar{X}_{t-1}}^{-1} \right)^{-1}
= M_{D_t,c,\bar{X}_t} M_{c,\bar{X}_t,\bar{X}_{t-1}} \cdot M_{D_t,c,\bar{X}_t}^{-1}
\]

(4.5)

and

\[
\left( M_{D_t,K_{t-1},X_t,\bar{X}_{t-1}} \cdot M_{D_t,K_{t-1},X_t,\bar{X}_{t-1}}^{-1} \right) \left( M_{D_t,K_{t-1},\bar{X}_t,\bar{X}_{t-1}} \cdot M_{D_t,K_{t-1},\bar{X}_t,\bar{X}_{t-1}}^{-1} \right)^{-1}
= M_{D_t,c,\bar{X}_t} \left( M_{c,\bar{X}_t,\bar{X}_{t-1}} \cdot M_{c,\hat{X}_t,\bar{X}_{t-1}}^{-1} \cdot M_{c,\bar{X}_t,\bar{X}_{t-1}} \cdot M_{c,\hat{X}_t,\bar{X}_{t-1}}^{-1} \right)^{-1}
= M_{D_t,c,\bar{X}_t} M_{c,\bar{X}_t,\bar{X}_{t-1}} \cdot M_{D_t,c,\bar{X}_t}^{-1}
\]

(4.6)

provided that the following assumption is satisfied.

**Assumption 2 (Invertibility).** Matrices \(M_{D_t,c,X_t}, M_{c,X_t,X_{t-1}}, M_{c,K_{t-1},X_{t-1}}\) are invertible for the four combinations of \((X_t, X_{t-1})\).

Assumption 2 requires \(j_d = j_c = j_k\). For discrete case, it is feasible to regroup different values to make the numbers of categories equal for \(D_t, c\) and \(K_{t-1}^\cdot\)\(^{27}\). The economic meaning of Assumption 2 is as follows. Suppose there are only two types, \(c \in \{c_h, c_l\}\). For \(M_{D_t,c,X_t}\) to be invertible, it is sufficient to have

\[
f_{D_t=1,\bar{L}_t,I_t=1,W_t=0|\bar{r}_t,X_t,K_t,H_{t-1},c_l} < f_{D_t=1,\bar{L}_t,I_t=1,W_t=0|\bar{r}_t,X_t,K_t,H_{t-1},c_h},
\]

\(^{27}\)Admittedly, this may take less advantage of the variations in the data.
This inequality implies that borrowers with low types are more likely to default. Suppose there are two categories of credit score, $K_1 > K_2$. It is sufficient to have $f_{H_{t-1},X_{t-1},K_1,c_h} > f_{H_{t-1},X_{t-1},K_2,c_h}$ and $f_{H_{t-1},X_{t-1},K_1,c_l} < f_{H_{t-1},X_{t-1},K_2,c_l}$ to ensure the invertibility of $M_{c,K_{t-1},X_{t-1}}$. The economic intuition behind these inequalities is that high-type borrowers are more likely to have higher credit grades and low-type borrowers are more likely to have lower grades. Finally, if the transition probabilities for all combinations of $(X_{t-1}, X_t)$ conditional on different types are non-zero, $M_{c,X_{t-1},X_{t-1}}$ is invertible. With Assumption 2 satisfied, Equations (4.5) and (4.6) lead to eigenvalue-eigenvector decompositions of the matrices on the left hand side. To guarantee the uniqueness of the decomposition, I invoke the following assumption.

**Assumption 3 (Uniqueness).** The densities in $M_{D_t,c,X_t}$ and $M_{c,K_{t-1},X_{t-1}}$ satisfy

1. $f_{D_t=1,L_t,I_t=1,W_t=0|r_t,X_t,K_{t-1},H_{t-1},c}$ decreases with $c$;
2. $f_{H_{t-1},X_{t-1},K_{t-1},c}$ increases with $K_{t-1}$ for $c = c_h$ and decreases with $K_{t-1}$ for $c = c_l$;

**Theorem 1 (Identification).** If Assumptions 1 – 3 are satisfied, $\{O_r\}_{r=1,t}$ for borrowers with two loans identifies $f_{X_t|X_{t-1},c}$, $f_{D_t,L_t,I_t=1,W_t=0|\Omega_t,H_{t-1},c}$ and $f_{H_{t-1},X_{t-1},K_{t-1},c}$ for different combinations of $(X_{t-1}, X_t)$ given any fixed values of $(r_t, K_t, L_t, H_{t-1})$.

The formal proof of Theorem 1 follows Hu (2008). For the continuous case, see Hu and Schennach (2008). However, from $f_{H_{t-1},X_{t-1},K_{t-1},c}$, I only recover the type distribution conditional on the borrower having two loans and $H_{t-1} \in \{nl, l, np\}$. This is due to the selection problem in the data. If borrowers fail to pay off their debts at $t - 1$, they are not allowed to enter at $t$. As a result, no dynamic variations are observed for this group of borrowers. Further analysis is presented at the end of Section 4.3 to show how to recover the original type distribution before selection occurs.

Next, I use variations in default and late payment performances within a period to identify the distribution of $\theta_t$ conditional on each type. Following Assumption 1 I decompose the identified kernel $f_{D_t,L_t,I_t=1,W_t=0|\Omega_t,H_{t-1},c}$ with respect to $\theta_t$:

$$
\frac{f_{D_t,L_t,I_t=1,W_t=0|\Omega_t,H_{t-1},c}}{f_{I_t=1|W_t=0,\Omega_t,H_{t-1}}} = \sum_{\theta_t} \left( f_{D_t|r_t,\theta_t,c} \cdot f_{L_t|r_t,\theta_t,c} \cdot f_{W_t=0,\theta_t|\Omega_t,c} \right). \tag{4.7}
$$

The kernels on the left hand side of Equation (4.7) are either identified or directly estimable from the data. On the right hand side, default and late payment performances are independent conditional on the effort level. Furthermore, the borrower’s participation probability is determined by the effort level, his type, and the covariates that shift his outside option distribution. Following a similar strategy, this equation leads to a unique eigenvalue-eigenvector decomposition under certain assumptions, and therefore $f_{D_t|r_t,\theta_t,c}$, $f_{L_t|r_t,\theta_t,c}$ and $f_{W_t=0,\theta_t|\Omega_t,c}$ are identified. I provide the theorem and the required assumptions in Appendix B.
4.3 Identification of Utility Primitives

Borrowers make withdrawal decisions by comparing utilities from participation with the random draws from the outside option. Thus only the utility levels relative to the outside option can be identified. Without loss of generality, I normalize the location of the outside option distribution by fixing the value of \( \tilde{v}_0 \). The borrower’s expected payoff \( \bar{V}_{B,c,T}^0(r_T, \theta_T) \) enters his participation probability through \( f_{W_t=0|\Omega_t, \theta_t, c} = \Phi(\bar{V}_{B,c,T}^0(r_T, \theta_T) - v(X_t, K_t)) \), where \( \Phi(\cdot) \) represents the cdf of the standard normal distribution. Using the variation in \( X_t \) when \( K_t = 0 \), I obtain the following equation,

\[
\frac{f_{W_t=0, \theta_t|X_t=x_1, K_t=0, r_t, c}}{f_{W_t=0, \theta_t|X_t=x_2, K_t=0, r_t, c}} = \frac{f_{W_t=0|X_t=x_1, K_t=0, r_t, c} \cdot f_{\theta_t|c}}{f_{W_t=0|X_t=x_2, K_t=0, r_t, c} \cdot f_{\theta_t|c}} = \frac{\Phi(\bar{V}_{B,c,T}^0(r_T, \theta_T) - \tilde{v}_0 - v_x X_1)}{\Phi(\bar{V}_{B,c,T}^0(r_T, \theta_T) - \tilde{v}_0 - v_x X_2)}.
\]

(4.8)

The left hand side of Equation (4.8) is identified from previous steps. If I further normalize \( v_x \), \( \bar{V}_{B,c,T}^0(r_T, \theta_T) \) is identified. With variations in \( K_t \), I am able to recover \( v_k \), which leads to the identification of \( f_{\theta_t|c} \).

The identified levels of \( \bar{V}_{B,c,T}^0(r_T, \theta_T) \) are associated with unknown parameters \( \alpha, R_h \), and the levels of private information \( (\theta_T, c) \). The intuition for identification of these parameters is as follows. The marginal effect of interest rate on \( \bar{V}_{B,c,T}^0(r_T, \theta_T) \) depends only on the shape of the utility function and the value of high revenue. Under the assumption of CARA utility, I show that two values of interest rates identify \( \alpha \) and \( R_h \) in Appendix C.1. Once I normalize \( \gamma \), the effort levels are identified from the probability that late payment occurs. Effort enters borrower’s payoff function through two channels – it affects default probabilities and induces cost. After plugging the effort levels into the payoff function, I obtain enough restrictions to pin down the levels of default cost and transitory cost shocks. The details are provided in Appendix C.2.

In the last part of this section, I focus on the identification of the original type distribution \( f_{c|X_{t-1},K_{t-1}} \) before selection takes place. After recovering all other primitives in the model, I am able to match the observed probability of default for all borrowers at period \( t - 1 \) with the prediction from the model:

\[
\begin{align*}
& f_{D_{t-1}=0|I_{t-1}=1, W_{t-1}=0, \Omega_{t-1}} \\
& = \sum_{\theta_{t-1}, c} \left( f_{D_{t-1}=0|r_{t-1}, \theta_{t-1}, c} \cdot f_{I_{t-1}=1|W_{t-1}=0, \Omega_{t-1}} \cdot f_{W_{t-1}=0|\Omega_{t-1}, \theta_{t-1}, c} \cdot f_{\theta_{t-1}|c} \cdot f_{c|X_{t-1}, K_{t-1}} \right). \tag{4.9}
\end{align*}
\]

In Equation (4.9), I estimate the funding probability \( f_{I_{t-1}=1|W_{t-1}=0, \Omega_{t-1}} \) at period \( t-1 \) directly from the data. \( f_{W_{t-1}=0|\Omega_{t-1}, \theta_{t-1}, c} \) represents the participation probability of borrowers with \( \{\Omega_{t-1}, \theta_{t-1}, c\} \), which can be computed using borrowers’ optimal effort choices at loan \( t-1 \).
Given all other parameters identified in the previous steps, I solve the borrower’s dynamic optimization problem. The participation probability is therefore identified with the optimal level of effort. The only unknowns in Equation (4.9) are the original probabilities of different default costs conditional on observed characteristics. $f_{\theta|X_{t-1},K_{t-1}}$ is therefore identified with this linear restriction.

5 Estimation Results

5.1 Likelihood-Based Estimation Method

Following the identification strategies in Section 4, I construct the likelihood for all individual borrowers in the sample. Let $i = 1, \cdots, N$ be the index for each borrower, and $t-1$ and $t$ represent two loans. $O_{it} = \{r_{it}, D_{it}, L_{it}, X_{it}, K_{it}, A_{it}, DC_{it}, W_{it}, I_{it}\}_{\tau=t-1,t}$ contains all observables for $i$ at loan $t$. The notation for interest rate ($r_{it}$), default ($D_{it}$), late payment ($L_{it}$), withdraw ($W_{it}$) and funding decisions ($I_{it}$) remains the same. I incorporate some other characteristics of borrowers into the estimation in addition to their financial statuses ($X_{it}$) and credit grades ($K_{it}$). Specifically, I allow the borrower’s type and the outside option distribution be associated with $X_{it}$, $K_{it}$ and the purpose of the loan (whether the loan is used for debt consolidation, denoted as $DC_{it}$). The realization of $DC_{it}$ depends on the borrower’s type $c$. The distribution of $\theta_{\tau}$ conditional on $c$ relies on the amount requested by the borrower, denoted as $A_{it}$. The realization of $A_{it}$ depends on borrower’s credit grade and the previous loan outcomes.\footnote{I characterize the amount requested by borrowers into two groups. The means of the high and low amount group are $13,000$ and $8,700$ respectively.}

In the estimation, the borrower’s default cost $c$ takes a value from \{c_{l}, c_{h}\}; the cost of effort $\theta_{\tau}$ takes a value from \{\theta_{l}, \theta_{h}\}. All other parameters follow the assumptions imposed in Section 4. With $\Theta$ denoting the vector of parameters identified and to be estimated, I
construct the log-likelihood for the whole sample as follows.

\[
LL(\Theta) = \sum_{i}^{N} \log(f_{O_{it}, O_{it-1}; \Theta})
\]

\[
= \sum_{i}^{N} \log(\sum_{c} f_{c|x_{it-1}, k_{it-1}, dc_{it-1}, \theta_{it-1}} \times f_{x_{it-1}, k_{it-1}, a_{it-1}, dc_{it-1}}
\times \left[ \sum_{\theta_{it}} f_{D_{it-1}, L_{it-1}, W_{it-1}, r_{it-1}, \theta_{it-1}|c, x_{it-1}, k_{it-1}, a_{it-1}, dc_{it-1}} \right]
\times f_{x_{it}|x_{it-1}, c} \times f_{K_{it}|k_{it-1}, d_{it-1}, L_{it-1}} \times f_{A_{it}|k_{it-1}, d_{it-1}} \times f_{D_{it}|c}
\times \left[ \sum_{\theta_{it}} f_{D_{it}, L_{it}, W_{it}, r_{it}, \theta_{it}|c, x_{it}, k_{it}, a_{it}, dc_{it}, d_{it-1}, L_{it-1}} \right]).
\] (5.1)

\[
f_{D_{it-1}, L_{it-1}, I_{it-1}, W_{it-1}, r_{it-1}, \theta_{it-1}|c, x_{it-1}, k_{it-1}, a_{it-1}, dc_{it-1}}
\]
represents the joint likelihood of the borrower’s observables (and \(\theta_{it-1}\)) at loan \(t - 1\). Given the conditional independence assumed earlier, I decompose this likelihood as follows. Notice that default and late payment probabilities enter the likelihood only for borrowers who participate and are funded.

\[
f_{D_{it-1}, L_{it-1}, I_{it-1}, W_{it-1}, r_{it-1}, \theta_{it-1}|c, x_{it-1}, k_{it-1}, a_{it-1}, dc_{it-1}}
\]

\[
\prod_{j=0,1} \left( f_{D_{it-1}=j|r_{it-1}, \theta_{it-1}, c} \right)^{1{D_{it-1}=j, I_{it-1}=1, W_{it-1}=0}}
\times \prod_{j=0,1} \left( f_{L_{it-1}=j|r_{it-1}, \theta_{it-1}, c} \right)^{1{L_{it-1}=j, I_{it-1}=1, W_{it-1}=0}}
\times \prod_{j=0,1} \left( f_{I_{it-1}=j|W_{it-1}=0, r_{it-1}, x_{it-1}, k_{it-1}, a_{it-1}, dc_{it-1}} \right)^{1{I_{it-1}=j, W_{it-1}=0}}
\times \prod_{j=0,1} \left( f_{W_{it-1}=j|x_{it-1}, k_{it-1}, r_{it-1}, \theta_{it-1}, c} \right)^{1{W_{it-1}=j}}
\times f_{r_{it-1}|x_{it-1}, k_{it-1}, a_{it-1}, dc_{it-1}} \times f_{\theta_{it}|c, a_{it}}.
\] (5.2)

My identification strategy does not rely on the optimality conditions of the interest rate to recover borrowers’ and lenders’ primitives. Thus, when constructing the likelihood, I am able to treat \(f_{r_{it-1}|x_{it-1}, k_{it-1}, a_{it-1}, dc_{it-1}}\) as an observed density from the data. To compute \(f_{D_{it-1}|r_{it-1}, \theta_{it-1}, c}, f_{I_{it-1}|r_{it-1}, \theta_{it-1}, c}\) and \(f_{W_{it-1}=0|x_{it-1}, k_{it-1}, r_{it-1}, \theta_{it-1}, c}\), borrowers’ optimal effort levels for the first loans are required. When solving the dynamic optimization problem for borrowers, I first compute their continuation values conditional on different observables and past outcomes. To do this, I estimate the second loan’s interest rate distribution and the funding probabilities nonparametrically from the data, and then compute continuation values using Equation (3.5). By taking different future consequences into account, the borrower
chooses his first loan’s optimal effort, which determines his participation probability and the probabilities that defaults and late payments occur. With the optimal effort level of borrowers conditional on different types, I can also compute the lender’s expected payoff given the updated beliefs derived in Equation (3.13). $f_{\theta_{it-1} | W_{it-1} = 1, r_{it-1}, x_{it-1}, K_{it-1}, A_{it-1}, DC_{it-1}}$ is thus obtained by comparing the expected payoff with the lender’s outside option. In addition, $f_{\theta_{it} | c, A_{it}}$ are primitives to be estimated. For the borrower’s joint likelihood of observables at loan $t$, a similar approach applies once I obtain the beliefs about the borrower’s type at the time he applies for the second loan. Beliefs are updated following Equation (3.14). In the estimation, I solve a three-period dynamic model and use the results from the first two periods to match the data, since borrowers with more than three listings are rarely observed in the sample.29

5.2 Results

The estimation results are shown in Table 11 with standard errors computed using a two-step maximum likelihood variance-covariance estimator (Murphy and Topel, 2002).30 I separate the estimates into four panels, where Panel (A) focuses on utility primitives and Panels (B) – (D) show the estimates of probabilities of low effort cost, transition probabilities of state variables, and the probabilities of high type. The risk-aversion parameter I get from the estimation is around 1.5 using a CARA utility function.31 The difference between the two categories of cost shocks is large, implying that transitory shocks are important when borrowers make their effort choices. My estimates also show that, when defaults occur, high-type borrowers achieve negative payoffs and low type borrowers achieve positive payoffs. If no defaults occur, the average net revenue borrowers can get out of a loan is around 49%. The estimates for $v_x$ and $v_d$ suggest that borrowers who have lower debt-to-income ratios or request loans for debt consolidation are faced with worse outside options. The last two lines of Panel (A) represent the estimated mean and variance of the lender’s outside option distribution. My estimate of the average return is slightly lower than the risk-free rate in the outside markets. This may be because lenders’ outside options are not restricted to investment options.

Panel (B) in Table 11 shows that borrowers who have higher default costs or request loans of a lower amount are more likely to draw smaller cost shocks. From Panel (C), it is

---

29 One caveat here is that we do not know the exact $T$ for each borrower, but my model is flexible enough to account for more periods if borrowers with more listings are available.

30 In the estimation, I normalize $\gamma = 4.5$, $v_0 = -1$, and $v_k = 0.1$.

31 When computing the actual monetary value for each loan, I multiply the revenue $R_h - r_t$ or $R_l - c$ by the amount requested (divided by $10^5$). So the risk-aversion parameter I get is essentially around $1.5 \times 10^{-5}$, which is within the range obtained in the literature using a CARA utility function; see Cohen and Einav (2007).
clear that when borrowers have high debt-to-income ratios, there is a high chance for them to stay in that bad situation. Borrowers with $c_h$ are more likely to stay with low debt-to-income ratios (about 64%). Panel (D) presents the estimates of original type distribution given different combinations of observables. It is clear that borrowers with high credit grades and who use the loans for debt consolidation are more likely to be “good borrowers”; that is, they have high default cost. Conditional on different debt-to-income ratios, the proportions of high type do not seem to vary much. In the estimation, I also allow the probability of borrowing for debt consolidation to be correlated with borrowers’ unobserved types. I find high type borrowers are more likely to propose loans that are used for debt-consolidation. This result is intuitive in the sense that borrowers who are attempting to pay off existing loans are likely to have higher default cost and may be more responsive to the dynamic incentives imposed by the reputation system.

5.3 Model Fit

I check the model fit by comparing borrowers’ participation probabilities, lenders’ funding probabilities, and default and late payment rates generated by the model with the ones observed in the data. This exercise is done for borrowers with different combinations of observables. The results are shown in Tables 12 and 13. Note that funding probabilities are computed using listings that are not withdrawn by borrowers. Default and late payment rates are computed for loans that are funded. In addition, I use the model and my estimates to predict the shift of credit grade distribution across two listings and compare the results with the patterns observed in the data. Since I classify the seven credit grades into two groups in the empirical estimation, I reproduce the histograms in Figure 1. The simplified version is provided in the left panel of Figure 2. To recap, the shift of the credit grade distribution may

---

32In the empirical setting, high credit groups include borrowers with credit grade AA, A and B. Other grades are characterized as low credit group.
occur for two reasons. First, borrowers who pay off the first loans and propose the second listings may have higher credit grades (selection channel). An alternative explanation would be that borrowers’ credit grades are updated after the first loan so that the distribution shifts (update channel). One advantage of the structural model is that I can decompose the two channels by plotting the distribution of credit grades first for all borrowers after the first loan, and then only for borrowers that pay off their loans and are selected in the second period. The shift of credit grade distribution under three stages is shown in the right panel of Figure 2. It is clear from the figure that the update channel is important and the selection channel further shifts the distribution to the right. Combining all results, my estimates arguably match the data patterns well.

6 Counterfactual Analysis
Given that the utility primitives and the distribution of borrowers’ private information have been recovered, I conduct three sets of counterfactuals in this section. I first compare welfare under three information structures – one with types and effort observed (symmetric), one with only types observed, and one with both unobserved (asymmetric). These counterfactuals allow me to (1) quantify the total welfare loss from asymmetric information, and (2) decompose the sources of inefficiency into adverse selection and moral hazard. I further consider the factual scenario, where the market with asymmetric information is imposed with a reputation system. By comparing the welfare gain from reputation with the total welfare loss, I am able to quantify the extent to which efficiency is restored by the reputation system. To alleviate long-run inefficiencies induced by the reputation/feedback system, I further study the implications of offering Payment Protection Insurance to borrowers.

For all counterfactual experiments, I assume that the website chooses the optimal interest rate to maximize the rate of successful transaction, since it charges commission fees for each successfully funded loans. I solve borrowers’ optimal effort levels, lenders’ funding decisions and the website’s pricing rule for each market design. Table 5 presents rates of successful transaction, borrower’s participation and lender’s funding probabilities, default and late payment probabilities, and average utilities for borrowers and lenders under the five market designs. A transaction occurs when a borrower participates and lenders invest. Hence, the rate of successful transaction equals the multiplication of borrowers’ participation rate and lenders’ funding probability for listings not withdrawn by borrowers. Default and

33The website may also be concerned about other aspects of market performance. For example, the website may prefer to keep the overall default rate below a certain level, it may prefer to attract more borrowers with high credit grades, etc. For this paper, I focus on the main channel through which the website makes profit.
Table 5: Counterfactual Results: Compare Welfare under Five Scenarios

<table>
<thead>
<tr>
<th>Scenarios</th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>S4</th>
<th>S5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Symmetric</td>
<td>Observe type</td>
<td>Asy. Info.</td>
<td>Reputation</td>
<td>Rep.+PPI</td>
</tr>
<tr>
<td>First Loan</td>
<td>Rate of transaction</td>
<td>0.9282</td>
<td>0.5577</td>
<td>0.4657</td>
<td>0.9255</td>
</tr>
<tr>
<td></td>
<td>Participation prob.</td>
<td>0.9293</td>
<td>0.9359</td>
<td>0.9329</td>
<td>0.9267</td>
</tr>
<tr>
<td></td>
<td>Funding prob.</td>
<td>0.9988</td>
<td>0.5959</td>
<td>0.4992</td>
<td>0.9988</td>
</tr>
<tr>
<td></td>
<td>Default prob.</td>
<td>0.1165</td>
<td>0.1007</td>
<td>0.2090</td>
<td>0.1563</td>
</tr>
<tr>
<td></td>
<td>Late prob.</td>
<td>0.0302</td>
<td>0.0228</td>
<td>0.1108</td>
<td>0.0596</td>
</tr>
<tr>
<td></td>
<td>Borrower’s avg. util.</td>
<td>0.2487</td>
<td>0.2658</td>
<td>0.2420</td>
<td>0.2497</td>
</tr>
<tr>
<td></td>
<td>Lender’s avg. util.</td>
<td>0.0816</td>
<td>0.0628</td>
<td>0.0550</td>
<td>0.0801</td>
</tr>
</tbody>
</table>

| Second Loan| Rate of transaction | 0.9186 | 0.5495 | 0.4451 | 0.7413 | 0.7676 |
|           | Participation prob. | 0.9198 | 0.9236 | 0.9239 | 0.7493 | 0.9249 |
|           | Funding prob. | 0.9987 | 0.5949 | 0.4817 | 0.9893 | 0.8299 |
|           | Default prob. | 0.1100 | 0.1007 | 0.2099 | 0.1481 | 0.1553 |
|           | Late prob. | 0.0305 | 0.0216 | 0.1178 | 0.0595 | 0.0636 |
|           | Borrower’s avg. util. | 0.2586 | 0.2722 | 0.2392 | 0.2642 | 0.2752 |
|           | Lender’s avg. util. | 0.0836 | 0.0619 | 0.0577 | 0.0665 | 0.0573 |

| Total Surplus | 0.3894 | -0.3208 | -0.5237 | 0.3444 | 0.3779 |

Note: Borrowers’ and lenders’ average utilities are computed for funded loans.

late payment rates and average utilities for borrowers and lenders are computed for funded loans. The upper and lower panels of Table 5 show results for the first and the second loans, respectively. Table 14 in the Appendix summarizes market outcomes and prices by type.

6.1 Welfare Loss from Asymmetric Information

I first compare the welfare of market participants under three scenarios. The first one features symmetric information, so that the website and lenders perfectly observe and price on borrowers’ types and effort. In the second scenario, only borrowers’ types are observed and priced. The website has to take borrowers’ incentive constraints into account when choosing optimal interest rates. The third one is when both types and effort are unobserved to the website and lenders, and no other mechanisms are imposed. The difference in welfare levels between S1 and S3 measures the total loss from asymmetric information. The difference
between S1 and S2 is only due to unobserved effort, which quantifies the welfare loss from moral hazard. By comparing these three market designs, I find that 22 percent of the inefficiency from asymmetric information is due to adverse selection, and 78 percent is due to moral hazard.

From Column 3 in Table 5, it is clear that the presence of asymmetric information leads to an inefficient market outcome. When both types and effort are unobserved, borrowers exert less effort and low-type borrowers are more likely to participate. This results in a higher default rate (around 20%) and a lower funding probability (49%). Due to asymmetric information, some potential good borrowers have difficulties in getting funded and the transaction rate for the website is also low (around 46%). In the case of symmetric information, the website and lenders perfectly observe and price on borrower’s type and effort. Borrowers are charged with type-specific interest rates. For borrowers that have low default cost and draw large cost shocks of exerting effort, their interest rates are around 35%, which are very close to the upper limit set by the Usury law. On the other hand, interest rates for good borrowers are around 10%. Under this scenario, lenders have a very high probability to invest and the transaction rate is about 92%. When only borrower’s types are observed, this market suffers from moral hazard. I find from the analysis that, for borrowers that have low default costs and draw large cost shocks, the website cannot find an interest rate at which lenders are willing to invest. In other words, the market collapses for this group of “lemon” borrowers. The transaction rate under this scenario is 55%, higher than that in S3 but lower than that in S1. This also implies that reducing adverse selection helps to improve market outcomes, while moral hazard still plays an important role in creating inefficiencies. In Scenarios 1-3, borrowers have no dynamic concerns. The results for the first and the second loans are very similar.

6.2 The Value of the Reputation System

I now consider the factual scenario, where the market with asymmetric information is imposed with a reputation system. I follow the rules imposed by Prosper.com and compute equilibrium interest rates and beliefs about borrower’s type over time. From the experiment, I find that 95 percent of welfare loss from asymmetric information is recovered by the reputation system through the following channels. First, the reputation system helps to refine beliefs about borrowers’ types, so that “lemon” borrowers are excluded from the market gradually. Table 6 presents the proportion of high-type borrowers conditional on observables

\[34\]

When information is symmetric, borrowers have no incentive to exert more effort in earlier periods. When there is asymmetric information but no reputation system is imposed, lenders do not update their beliefs based on past loan outcomes. Borrowers again have no dynamic incentives.
Table 6: Belief of High Type Proportion

<table>
<thead>
<tr>
<th>Observables</th>
<th>Original</th>
<th>After 1st Loan</th>
<th>After 2nd Loan</th>
</tr>
</thead>
<tbody>
<tr>
<td>other purpose</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>low dti, low credit grade</td>
<td>0.2335</td>
<td>0.4754</td>
<td>0.6396</td>
</tr>
<tr>
<td>high dti, low credit grade</td>
<td>0.2461</td>
<td>0.3841</td>
<td>0.5086</td>
</tr>
<tr>
<td>low dti, high credit grade</td>
<td>0.5326</td>
<td>0.6072</td>
<td>0.6811</td>
</tr>
<tr>
<td>high dti, high credit grade</td>
<td>0.5870</td>
<td>0.5202</td>
<td>0.5554</td>
</tr>
<tr>
<td>debt cons.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>low dti, low credit grade</td>
<td>0.3219</td>
<td>0.5436</td>
<td>0.7016</td>
</tr>
<tr>
<td>high dti, low credit grade</td>
<td>0.3311</td>
<td>0.4517</td>
<td>0.5774</td>
</tr>
<tr>
<td>low dti, high credit grade</td>
<td>0.7320</td>
<td>0.6759</td>
<td>0.7404</td>
</tr>
<tr>
<td>high dti, high credit grade</td>
<td>0.7323</td>
<td>0.5813</td>
<td>0.6201</td>
</tr>
</tbody>
</table>

across loans. Borrowers that have defaults in the previous loans are not allowed to enter in the future. This selection process makes the pool of borrowers better. Second, when the reputation system is implemented, borrowers are incentivized to exert more effort, which leads to a lower default rate (around 15%) and higher average utilities for lenders. Third, from Table 14, I find that the funding rate for all types under reputation is high (99%). That is, even low type borrowers now have access to credit, which reduces welfare loss from credit rationing observed in S2 and S3. Given higher rate of matching (92%), the website receives more profit when the reputation system is implemented.

I then compare the outcomes of the first and the second loans under the reputation system. Different from the cases in Section 6.1, borrowers have dynamic concerns in this scenario. There are two counter forces that impact outcomes in the second loans. On the one hand, borrowers that have survived are more likely to have better types. On the other hand, since they are closer to the final period, borrowers have less incentive to exert effort. The lower panel of Table 5 shows that default and late payment rates are slightly lower for second loans, which indicates that the positive effect of selection dominates the negative effect of decreasing incentives. Another interesting point from this table is that borrowers’ participation rate in second loans under the reputation system is significantly lower than that in first loans. This is because borrowers who have defaulted are not allowed to participate in the second period and can only take random draws from the outside option distribution. Some of the borrowers may be of good types but receive bad shocks. The accidental loss of reputation prevents good borrowers from having future credit access, and thus may lead to potential long-run inefficiencies.

35In the counterfactual analysis, I also require the reputation system to not allow borrowers with late payments to enter again. This helps to reduce computational burden.
6.3 Payment Protection Insurance

To address the long-term inefficiencies due to accidental loss of reputation, the last part of my counterfactual analysis considers remedies for the welfare loss that may occur when incorrect beliefs about borrowers’ quality persist in credit markets. In the current setting of the website, once he defaults, the borrower loses his reputation immediately and is unable to borrow again. However, even if borrowers have good types and exert high levels of effort, it is still possible that they obtain unlucky draws of revenues. And if this situation accidentally occurs, good borrowers are left with no access to future credit and beliefs about their quality get “stuck”. This situation has strong empirical relevance, especially for small businesses that find peer-to-peer lending an attractive financing alternative (Segal, 2015) and rely heavily on this form of credit access for their success and growth.

In this section, I consider the policy implication of offering borrowers an option to buy Payment Protection Insurance (PPI), which covers loan repayments for a set period of time if borrowers are unable to make them in certain situations. These circumstances usually include being made redundant at one’s job or not being able to work because of an accident or illness. The intuition of this mechanism is straightforward. If a borrower wants to maintain a good reputation (and hence credit access in the future), but also worries about future negative shocks, he/she can purchase this insurance ex-ante to hedge against the risk.

Adding PPI to the main model in Section 3 induces some new challenges. First, I need to model borrowers’ insurance purchase decisions and how premiums are determined by the insurer. Second, the insurance purchase decisions affect lenders’ beliefs about borrowers’ types, and thus affecting the funding decisions and equilibrium interest rates. Now borrowers have to also decide whether or not to purchase the insurance when making participation decisions. If borrowers buy the insurance, they have to pay premiums at the current period, but the insurer will cover the repayments whenever bad shocks realize. No defaults or late payments occur for a loan that is insured. If borrowers do not have insurance, there is a possibility that they will default or have late payments. I assume that once borrowers default or have late payments, they are not allowed to enter again. Moreover, to simplify the model, I assume that the website and lenders do not observe previous insurance purchase decisions. This implies that future beliefs about borrowers’ types only depend on default or late payment performances. Future beliefs enter into borrowers’ optimization problems with or without insurance purchase.

I now sketch the model in which the borrower is offered an option to purchase payment protection insurance. Consider a borrower with default cost $c$. At the time he applies for loan $t$, lenders’ and insurers’ belief about his type is $S_t$. His expected payoff of having loan $t$ with insurance purchase, given interest rate $r_t$, effort level $e$ and cost of effort $\theta_t$ is defined
in the following equation.

\[ \tilde{V}_{B,c,t}^{\text{ins}}(e, S_t, r_t, \theta_t, B_0) = \int_{1+r_t}^{R} U(R - 1 - r_t) dF_{R|e}(R|e) + \int_{0}^{1+r_t} U(R - c) dF_{R|e}(R|e) \]

+ \delta E V_{B,c,t+1}(S_{t+1}(S_t, r_t, H_t = n l), r_{t+1}, \theta_{t+1}, v_{0,t+1}) - \phi(e, \theta_t) - B_0, \]

(6.1)

where \( B_0 \) represents the premium of the insurance plan. Assume that borrowers need to pay the default cost if they default at period \( T - 1 \) even with insurance purchase, but the insurer will cover their repayments to the lender, so that the future continuation values will not be affected.\(^{36}\) The borrower decides his optimal effort level for loan \( t \) with insurance through the following payoff maximization problem:

\[ e_{c,t}^{\text{ins}}(S_t, r_t, \theta_t, B_0) = \arg \max_{e} \tilde{V}_{B,c,t}^{\text{ins}}(e, S_t, r_t, \theta_t, B_0). \]

(6.2)

If the borrower does not purchase insurance for loan \( t \), his expected payoff and the optimal effort is the same as derived in Equations (3.6) and (3.7). I now construct the borrower’s choice-specific (purchase insurance or not) value function in the following equation.

\[ V_{B,c,t}^{\text{ins}}(S_t, r_t, \theta_t) = \max_{e} \tilde{V}_{B,c,t}^{\text{ins}}(e, S_t, r_t, \theta_t) \]

(6.3)

The borrower purchases the insurance whenever \( \tilde{V}_{B,c,t}^{\text{ins}}(S_t, r_t, \theta_t, B_0) > \tilde{V}_{B,c,t}^{\text{ins}}(S_t, r_t, \theta_t) \). The value from participation at loan \( t \) is therefore

\[ \tilde{V}_{B,c,t}^{0}(S_t, r_t, \theta_t, B_0) = \max\{\tilde{V}_{B,c,t}^{\text{ins}}(S_t, r_t, \theta_t, B_0), \tilde{V}_{B,c,t}^{\text{nins}}(S_t, r_t, \theta_t)\}. \]

Then following a procedure similar to that in Section 3.2.2, it is straightforward to compute the participation probabilities and insurance purchase decisions for borrowers with each type.

I then focus on how the equilibrium premiums are decided. Assuming the insurer charges a premium to break even, the equilibrium condition for the premium is as follows.

\[ B_0 = (1 + r_t) \sum_{e,\theta_t} Pr(D_t = 1|e_{c,t}^{\text{ins}}(S_t, r_t, \theta_t, B_0)) Pr(e, \theta_t| \text{par. and buy ins.}). \]

(6.4)

The RHS of Equation (6.4) represents the cost to the insurer – whenever borrowers default, the insurer needs to cover the repayment to lenders. Note also that the insurer does not know the exact types of borrowers who participate and purchase the insurance, so he can

---

\(^{36}\)This assumption is to capture the idea that in the case when the borrower defaults, the insurer will keep a record.
only compute his expected cost using updated beliefs. Given future beliefs about borrowers’ types, I am able to solve borrowers’ optimal effort levels and the premium charged. In the equilibrium, borrowers’ strategies should be consistent with the beliefs. Following procedures similar to those discussed in Section 3, lenders make funding decisions using updated beliefs and the website chooses the optimal interest rates.

In the last of column of Table 5, I compute market outcomes when the reputation system is imposed with an option to purchase Payment Protection Insurance. When borrowers have access to insurance, the risk faced by lenders is reduced, and thus their average utility increases. From borrowers’ point of view, once they purchase the insurance, they get access to future credit. By offering PPI, I find that borrowers have higher chance to participate in second loans and the rate of transaction also increases. This finding highlights the effect of this intervention on alleviating long-run inefficiencies caused by accidental loss of reputation. This exercise also shows that only borrowers with high probabilities of being good type purchase insurance. This is because after taking adverse selection into account, the insurer cannot find premiums to break even for the remaining borrowers. Overall, I find providing PPI to the market further increases the total welfare of market participants. 98 percent of the welfare loss from asymmetric information is eliminated under this mechanism.

6.4 Impact on Market Size

To get a rough idea about the impact of different mechanisms on the size of online credit markets, I do a simple calculation for Prosper.com. The probabilities that transactions occur shown in Table 5 are used to approximate market sizes. Table 7 summarizes the market sizes under different market designs. The current market size for Prosper.com is around 9 billion dollars. If the reputation system on this market is removed, then the market size could shrink to 4.6 billion dollars. If the market only suffers from moral hazard, the market size is larger (5.5 $bn), but still much lower than the symmetric information scenario (9.2 $bn). From this exercise, we can again see that the reputation system helps to restore a large proportion of efficiency in online credit markets. Adding payment protection insurance can further increase total welfare.

7 Conclusion

This paper investigates the effectiveness of reputation/feedback systems in improving welfare in online credit markets when both adverse selection and moral hazard are present. I develop a finite-horizon dynamic structural model to analyze borrowers’ repayment decisions, lenders’ investment strategies, and the website’s pricing schemes. I prove the identification
Table 7: Market Size under Different Scenarios

<table>
<thead>
<tr>
<th>Scenarios</th>
<th>Market Size ($bn)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symmetric Information</td>
<td>9.18</td>
</tr>
<tr>
<td>Reputation+PPI</td>
<td>9.02</td>
</tr>
<tr>
<td>Reputation (Factual)</td>
<td>9.00</td>
</tr>
<tr>
<td>Under Moral Hazard</td>
<td>5.51</td>
</tr>
<tr>
<td>Under Moral Hazard and Adverse Selection</td>
<td>4.59</td>
</tr>
</tbody>
</table>

of the distribution of borrowers’ private types and utility primitives based on variations in borrowers’ repayment histories, transitions of their characteristics, and interest rates, and then estimate the model using a large transaction-level dataset from Prosper.com.

In this paper, I separate the effect of adverse selection and moral hazard, and find that moral hazard plays an important role in online credit markets. This result has strong empirical relevance, since the policy interventions that can be used to alleviate welfare loss from adverse selection and moral hazard are very different. My result suggests that imposing some “ex-post” monitoring mechanisms on credit markets may be more effective.

The main contribution of this paper is to quantify the welfare gain from reputation systems in online credit markets. I find that reputation matters to a large extent through refining beliefs about borrowers’ types, incentivizing effort exertion, and expanding credit access to low-type borrowers. These results may have implications for other settings, including fast-growing online marketplaces that widely use review systems to facilitate transactions and traditional credit markets that rely heavily on the credit rating system. My paper also considers a policy intervention that protects borrowers from accidental loss of reputation. I find that incorporating payment protection insurance into the market further improves total welfare. This exercise is related to the optimal “forgiveness” mechanisms that are considered in the existing credit rating system. From a methodological perspective, this paper provides new identification strategies for empirical contract models. In a separate paper, I develop general identification and estimation results for dynamic models with unobserved choices.
References


40


A Existence of the Equilibrium Beliefs

In this section, I provide a sketch of the proof for the existence of equilibrium beliefs about borrowers’ type distribution. For illustration, I simplify the model in Section 3 by keeping only the essence of the dynamic updating process. In particular, I assume all borrowers participate and are funded for their loans at $T - 1$. Moreover, I suppose there are no late payments. Future beliefs about borrowers depend only on whether they default at $T - 1$. For illustration, I consider a case in which $c$ takes values from $\{c_l, c_h\}$. The belief about the borrower’s type can be captured by $s_t \in [0, 1]$, which represents the probability of the high type at the beginning of each loan $t$. For this section, I fix the value of $\theta_t = \bar{\theta}$. I also assume the realized revenue has only two values $R_h$ and $R_l$, and the probability that $R_h$ realizes depends on effort level $e$ through $p(e)$. At loan $T - 1$, if the borrower defaults, he obtains $\bar{v}_0$ at $T$. If the borrower pays back the loan, the probability that he is a high-type borrower is updated.

At $T - 1$, the borrower’s optimization problem is as follows.

$$e_{c,T-1}(s_{T-1}, r_{T-1}) = \arg \max_e \tilde{V}_{B,c,T-1}(e, s_{T-1}, r_{T-1}).$$

$$\tilde{V}_{B,c,T-1}(e, s_{T-1}, r_{T-1}) = p(e)[U(R_h - 1 - r_{T-1})$$

$$+ \delta E V_{B,c,T}(s_{T-1}, r_{T-1}, H_{T-1} = nl), r_T, \theta_T, v_{0,T})]$$

$$+ (1 - p(e)) [U(R_l - c) + \delta \bar{v}_0] - \phi(e, \bar{\theta}).$$

(A.1)

Let the belief about the borrower’s type distribution at the beginning of loan $T$ to be $s^*_T = s_T(s_{T-1}, r_{T-1}, H_{T-1} = nl)$ if he has paid off the loan at $T - 1$. Then essentially the borrower’s effort level and expected payoff at loan $T - 1$ depends on $s^*_T$ and $r_{T-1}$, i.e. $e_{c,T-1}(s^*_T, r_{T-1}), \tilde{V}_{B,c,T-1}(e, s^*_T, r_{T-1})$.

The belief is formed through the Bayesian updating process.

$$s^*_{T} = \frac{p(e^*_{c_h,T-1}(s^*_T, r_{T-1}))s_{T-1}}{p(e^*_{c_h,T-1}(s^*_T, r_{T-1}))s_{T-1} + p(e^*_{c_l,T-1}(s^*_T, r_{T-1}))s_{T-1} - 1} \equiv B(s^*_T).$$

(A.2)

Equation (A.2) leads to a fixed-point problem for $s^*_T$. To prove the existence of the equilibrium belief, by Brouwer’s Fixed-Point Theorem, it is sufficient to show that $B(s^*_T) : [0, 1] \rightarrow [0, 1]$ is a continuous function.

I first prove that $e^*_{c_h,T-1}(s^*_T, r_{T-1})$ and $e^*_{c_l,T-1}(s^*_T, r_{T-1})$ are continuous functions of $s^*_T$ by Berge’s Theorem of Maximum. Assume the support of $e$ is bounded and closed, thus compact. $\tilde{V}_{B,c,T-1}(e, s^*_T, r_{T-1})$ is jointly continuous in $e$ and $s^*_T$. To see the continuity of $EV_{B,c,T}(s^*_T, r_T, \theta_T, v_{0,T})$, note that Equation (3.10) shows that the value function is continuous in the funding probability and Equation (3.13) further shows that the lender’s expected
payoff (and hence funding probability) is a continuous function of $s_T^*$. Then by Berge’s Theorem of Maximum, $e^{c_{T-1}}(s_T^*, r_{T-1})$ and $e^{c_T}(s_T^*, r_{T-1})$ are upper hemicontinuous at $s_T^*$. I impose the following assumption to guarantee the uniqueness of the optimal effort in the borrower’s optimization problem.

**Assumption A1** (Second Order Condition).

$$\frac{\partial^2 \tilde{V}_{B,c,T-1}(e, s_{T-1}, r_{T-1})}{\partial e^2} < 0$$

Since a single-valued upper-hemicontinuous correspondence must be a continuous function, $e^{c_{T-1}}(s_T^*, r_{T-1})$ and $e^{c_T}(s_T^*, r_{T-1})$ are continuous in $s_T^*$. $B(s_T^*)$ is therefore continuous, since the composition of two continuous functions is also continuous.
Identification of the Effort Choice Probabilities

Suppose $D_t, \theta_t$ and $X_t$ are all discrete random variables. Use $j_d = 1, \ldots, J_d$, $j_\theta = 1, \ldots, J_\theta$ and $j_x = 1, \ldots, J_x$ to index the values for these three variables. I consider the case where $I_t = 1, W_t = 0$ and the values of \{r_t, K_t, H_{t-1}, c\} are fixed. To rewrite Equation (4.7) in matrix form, I define the following matrices for $L_t = 1$,

$$M_{D_t, L_t=1, X_t} = \frac{f_{D_t, L_t=1, X_t=0} \theta_t}{f_{D_t, L_t=1, X_t=0} \theta_t} \begin{bmatrix} f_{D_t, L_t=1, X_t=0} \theta_t \end{bmatrix}_{j_d \theta}$$

$$M_{D_t, \theta_t} = \begin{bmatrix} f_{D_t, \theta_t, \theta_t} \big| D_t=j_d \big| \theta_t=j_\theta \end{bmatrix}_{j_d \theta}$$

$$M_{\theta_t, L_t=1} = \text{diag} \left\{ \begin{bmatrix} f_{L_t=1} \big| \theta_t, \theta_t \big| \theta_t=j_\theta \big| \theta_t=0 \big| \theta_t=1, \ldots, J_\theta \end{bmatrix} \right\}$$

$$M_{\theta_t, X_t} = \begin{bmatrix} f_{X_t=0, \theta_t} \big| X_t, \theta_t, \theta_t \big| \theta_t=j_\theta \big| X_t=j_x \big| \theta_t=0 \big| \theta_t=1 \big| \theta_t=1, \ldots, J_\theta \end{bmatrix}_{j_\theta \theta}$$

The matrix form of Equation (4.7) when $L_t = 1$ is therefore

$$M_{D_t, L_t=1, X_t} = M_{D_t, \theta_t} M_{\theta_t, L_t=1} M_{\theta_t, X_t}. \quad (B.1)$$

Similarly, when $L_t = 0$,

$$M_{D_t, L_t=0, X_t} = M_{D_t, \theta_t} M_{\theta_t, L_t=0} M_{\theta_t, X_t}. \quad (B.3)$$

Combining Equations (B.2) and (B.3), I construct the following equations.

$$M_{D_t, L_t=1, X_t} M_{D_t, L_t=0, X_t}^{-1} = (M_{D_t, \theta_t} M_{\theta_t, L_t=1} M_{\theta_t, X_t}) (M_{D_t, \theta_t} M_{\theta_t, L_t=0} M_{\theta_t, X_t})^{-1}$$

$$= M_{D_t, \theta_t} (M_{\theta_t, L_t=1} M_{\theta_t, L_t=0}) M_{D_t, \theta_t}^{-1} \quad (B.4)$$

and

$$M_{D_t, L_t=1, X_t} M_{D_t, L_t=0, X_t}^{-1} = (M_{D_t, \theta_t} M_{\theta_t, L_t=1} M_{\theta_t, X_t})^{-1} (M_{D_t, \theta_t} M_{\theta_t, L_t=0} M_{\theta_t, X_t})$$

$$= M_{\theta_t, X_t} (M_{\theta_t, L_t=1} M_{\theta_t, L_t=0}) M_{\theta_t, X_t}^{-1} \quad (B.5)$$

provided that the following assumption is satisfied:

**Assumption B.1 (Invertibility).** Matrices $M_{D_t, \theta_t}, M_{\theta_t, L_t=1}, M_{\theta_t, L_t=0}, M_{\theta_t, X_t}$ are invertible.

The economic intuition behind Assumption B.1 is as follows. For illustration, I consider a case where $\theta_t \in \{\theta_h, \theta_l\}, X_t \in \{X_1, X_2\}$. It is sufficient to have $f_{D_t=0} | r_t, \theta_h, c < f_{D_t=0} | r_t, \theta_l, c$ to ensure the invertibility of $M_{D_t, \theta_t}$. This inequality implies that, conditional on receiving low cost of effort, borrowers are more likely to pay off the debt. For $M_{\theta_t, L_t=1}$ and $M_{\theta_t, L_t=0}$ to be invertible, the probability that late payment occurs must be in $(0, 1)$. For matrix $M_{\theta_t, X_t}$,
since \( f_{W_t=0,\theta_t|X_t,K_t,r_t,c} = f_{W_t=0|X_t,K_t,r_t,\theta_t,c} \cdot f_{\theta_t|c} \),

\[
M_{\theta_t,X_t} = \begin{bmatrix}
    f_{W_t=0|\theta_h,X_1,K_t,r_t,c} \cdot f_{\theta_h|c} & f_{W_t=0|\theta_h,X_2,K_t,r_t,c} \cdot f_{\theta_h|c} \\
    f_{W_t=0|\theta_l,X_1,K_t,r_t,c} \cdot f_{\theta_l|c} & f_{W_t=0|\theta_l,X_2,K_t,r_t,c} \cdot f_{\theta_l|c}
\end{bmatrix}.
\]

As a result, to achieve the invertibility of \( M_{\theta_t,X_t} \), the following inequality is required:

\[
f_{W_t=0|\theta_h,X_1,K_t,r_t,c} \cdot f_{W_t=0|\theta_l,X_2,K_t,r_t,c} \neq f_{W_t=0|\theta_h,X_2,K_t,r_t,c} \cdot f_{W_t=0|\theta_l,X_1,K_t,r_t,c},
\]

or alternatively:

\[
\frac{f_{W_t=0|\theta_h,X_1,K_t,r_t,c}}{f_{W_t=0|\theta_l,X_1,K_t,r_t,c}} \neq \frac{f_{W_t=0|\theta_h,X_2,K_t,r_t,c}}{f_{W_t=0|\theta_l,X_2,K_t,r_t,c}}.
\]

(B.6)

The intuition behind this assumption is that, as the distribution of the outside option gets better, the impact of receiving different cost draws becomes smaller, and thus the ratio of the participation probabilities given \( \theta_h \) and \( \theta_l \) becomes closer to 1. For example, consider the extreme case in which the mean of the outside option is larger than the maximum payoff borrowers can get from the website when \( X_t = X_2 \), so both types will be willing to stay out and the ratio of their participation probabilities is 1. When the other value of \( X_t \) is lower enough to attract borrowers to participate, the condition in Equation (B.6) holds.

With Assumption [B.1] satisfied, Equations (B.4) and (B.5) lead to eigenvalue-eigenvector decompositions of the matrices on their left hand sides. To guarantee the uniqueness of the decomposition, I invoke the following assumption.

**Assumption [B.2] (Uniqueness).** \( f_{L_t=1|r_t,\theta_t,c} \) increases with \( \theta_t \).

**Theorem [B.1] (Identification).** In addition to the assumptions in Theorem [1] if Assumptions [B.1] and [B.2] are satisfied, \( f_{D_t|r_t,\theta_t,c}, f_{L_t|r_t,\theta_t,c} \) and \( f_{W_t=0,\theta_t|\Omega_t,c} \) are identified for any fixed values of \( \{r_t,K_t,H_{t-1},c\} \).
Consider the unknown parameters in $\bar{V}^0_{B,c,T}(r_T, \theta_T)$. Explicitly,

$$
\bar{V}^0_{B,c,T}(r_T, \theta_T) = p(e_{c,T}(r_T, \theta_T); \beta)U(R_h - 1 - r_t; \alpha) + (1 - p(e_{c,T}(r_T, \theta_T); \beta))U(R_t - c; \alpha) - \phi(e_{c,T}(r_T, \theta_T), \theta_T),
$$
(C.1)

where $e_{c,T}(r_T, \theta_T)$ is the level of optimal effort for borrowers with type $c$ given $(r_T, \theta_T)$. The level of effort enters into the payoff function through two channels: (1) it affects the probability that high revenue $R_h$ is realized through $p(\cdot; \beta)$ with parameter $\beta$ measuring the effectiveness of the effort; and (2) it induces cost through $\phi(e, \theta_T) = \theta_T e^2$, where $\theta_T$ represents the cost of effort. In Equation (C.1), the unknowns include $\beta$, $\alpha$, $R_h$ and the levels of $c$ and $\theta_T$.

## C.1 Identification of Risk-Aversion Parameter

Using variations in observed interest rates, the following theorem identifies $R_h$ and $\alpha$.

**Theorem C.1.** Given two values of observed interest rate, $\bar{r}_t$ and $\hat{r}_t$, $\alpha$ and $R_h$ are identified.

A sketch of the proof of Theorem C.1 is as follows. With $\bar{V}^0_{B,c,T}(r_T, \theta_T)$ recovered, its derivative with respect to $r_T$,

$$
\frac{\partial \bar{V}^0_{B,c,T}(r_T, \theta_T)}{\partial r_T} = \frac{\partial e_{c,T}(r_T, \theta_T)}{\partial r_T} [p'(e_{c,T}(r_T, \theta_T))(U(R_h - 1 - r_T) - U(R_t - c)) - \phi'(e_{c,T}(r_T, \theta_T))] - p(e_{c,T}(r_T, \theta_T))U'(R_h - 1 - r_T)
$$

$$
= -p(e_{c,T}(r_T, \theta_T))U'(R_h - 1 - r_T),
$$
(C.2)

is also identified. Notice that the second equality in Equation (C.2) holds because $e_{c,T}(r_T, \theta_T)$ satisfies the first order condition in borrower’s optimization problem. Observe also that $p(e_{c,T}(r_T, \theta_T)) = \frac{f_{D_T=0|r_T, \theta_T,c}}{f_{D_T=0}}$, which has been recovered from the matrix decomposition in Section 4.2. Therefore Equation (C.2) identifies $U'(R_h - 1 - r_T)$. With the assumption of CARA utility, $U'(R_h - 1 - r_T) = \alpha \exp(-\alpha(R_h - 1 - r_T))$. With two observed interest rates $\bar{r}_T$ and $\hat{r}_T$, $\alpha$ is identified through the following equation

$$
\frac{U'(R_h - 1 - \bar{r}_T)}{U'(R_h - 1 - \hat{r}_T)} = \frac{\alpha \exp(-\alpha(R_h - 1 - \bar{r}_T))}{\alpha \exp(-\alpha(R_h - 1 - \hat{r}_T))} = \exp(-\alpha(\bar{r}_T - \hat{r}_T)).
$$
(C.3)

Plugging $\alpha$ back to $U'(R_h - 1 - r_T)$, $R_h$ is identified. ■
C.2 Identification of Levels of Private Information

To identify the levels of \( c \) and \( \theta_t \), I relate the identified default probability \( f_{D_t=0|\tau,\theta_t,c} \) to the effort level through \( p(\cdot; \beta) \).

\[
f_{D_t=0|\tau,\theta_t,c} = p(\varepsilon_{c,T}(r_T, \theta_T); \beta) = 1 - \exp(-\beta \varepsilon_{c,T}(r_T, \theta_T)). \tag{C.4}
\]

Equation (C.4) for \( c = c_h, c_l \) uniquely determines \( \varepsilon_{c_h,T}(r_T, \theta_T) \) and \( \varepsilon_{c_l,T}(r_T, \theta_T) \). I rearrange Equation (C.1) for \( c_h \) and \( c_l \), and construct the following two equations for \( \theta_h \) and \( \theta_l \).

\[
\left( \frac{\varepsilon_{c_h,T}(r_T, \theta_h)}{\varepsilon_{c_l,T}(r_T, \theta_h)} \right)^2 = \frac{p(\varepsilon_{c_h,T}(r_T, \theta_h))U(R_h - 1 - r_t) + (1 - p(\varepsilon_{c_h,T}(r_T, \theta_h)))U(R_l - c_h) - V^0_{B,c_h,T}(r_T, \theta_h)}{p(\varepsilon_{c_l,T}(r_T, \theta_h))U(R_h - 1 - r_t) + (1 - p(\varepsilon_{c_l,T}(r_T, \theta_h)))U(R_l - c_l) - V^0_{B,c_l,T}(r_T, \theta_h)}
\]

\[
\left( \frac{\varepsilon_{c_h,T}(r_T, \theta_l)}{\varepsilon_{c_l,T}(r_T, \theta_l)} \right)^2 = \frac{p(\varepsilon_{c_h,T}(r_T, \theta_l))U(R_h - 1 - r_t) + (1 - p(\varepsilon_{c_h,T}(r_T, \theta_l)))U(R_l - c_h) - V^0_{B,c_h,T}(r_T, \theta_l)}{p(\varepsilon_{c_l,T}(r_T, \theta_l))U(R_h - 1 - r_t) + (1 - p(\varepsilon_{c_l,T}(r_T, \theta_l)))U(R_l - c_l) - V^0_{B,c_l,T}(r_T, \theta_l)}. \tag{C.5}
\]

(C.5) provides a system of two linear equations of \( U(R_l - c_h) \) and \( U(R_l - c_l) \). To ensure identification of levels of \( R_l - c \), I invoke the following rank condition.

Assumption C.1 (Rank Condition).

\[|\Delta| = -\delta_2(1 - p_1)(1 - p_4) + \delta_1(1 - p_2)(1 - p_3) \neq 0 \]

where \( \delta_1 = \left( \frac{\varepsilon_{c_h,T}(r_T, \theta_h)}{\varepsilon_{c_l,T}(r_T, \theta_h)} \right)^2 \), \( \delta_2 = \left( \frac{\varepsilon_{c_h,T}(r_T, \theta_l)}{\varepsilon_{c_l,T}(r_T, \theta_l)} \right)^2 \), \( p_1 = p(\varepsilon_{c_h,T}(r_T, \theta_h)) \), \( p_2 = p(\varepsilon_{c_l,T}(r_T, \theta_h)) \), \( p_3 = p(\varepsilon_{c_h,T}(r_T, \theta_l)) \), and \( p_4 = p(\varepsilon_{c_l,T}(r_T, \theta_l)) \).

Note that all terms in Assumption C.1 have been recovered, so the rank condition is directly testable. This condition guarantees the unique solution of \( R_l - c_h \) and \( R_l - c_l \) given the identified risk-averse parameter \( \alpha \). To further pin down \( \theta_h \) and \( \theta_l \), it is sufficient to normalize \( \gamma \).

---

37In the model, I assume that whenever \( R_h \) is realized, the borrower pays off his loan and \( p(e; \beta) \) represents the probability \( R_h \) is realized given effort level \( e \).

38The levels of \( R_l \) and \( c \) are not separately identified, since only their difference matters in the model.

39The ratio of \( \theta_h \) and \( \theta_l \) can be easily recovered without knowing the exact effort level. However, since \( \beta \) and \( \theta_l \) enter the model with effort level in a non-separable way, I cannot recover all of them simultaneously. I decide to normalize \( \gamma \) to pin down the levels of effort. The model then identifies \( \beta \) and the levels of cost shocks for better economic interpretations.
Table 8: Summary Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
<th># of Obs</th>
</tr>
</thead>
<tbody>
<tr>
<td># of listings</td>
<td>1.1197</td>
<td>0.3706</td>
<td>1</td>
<td>3</td>
<td>102,528</td>
</tr>
<tr>
<td># of loans originated</td>
<td>1.0023</td>
<td>0.4959</td>
<td>0</td>
<td>3</td>
<td>102,528</td>
</tr>
<tr>
<td>term (# of months)</td>
<td>42.3681</td>
<td>11.2994</td>
<td>12</td>
<td>60</td>
<td>114,804</td>
</tr>
<tr>
<td>borrow for debt consolidation</td>
<td>0.6716</td>
<td>0.4696</td>
<td>0</td>
<td>1</td>
<td>114,804</td>
</tr>
<tr>
<td>borrow for home improvement</td>
<td>0.0731</td>
<td>0.2602</td>
<td>0</td>
<td>1</td>
<td>114,804</td>
</tr>
<tr>
<td>borrow for business</td>
<td>0.0504</td>
<td>0.2187</td>
<td>0</td>
<td>1</td>
<td>114,804</td>
</tr>
<tr>
<td>FICO score below 600</td>
<td>0.3058</td>
<td>0.4607</td>
<td>0</td>
<td>1</td>
<td>114,804</td>
</tr>
<tr>
<td>home owner</td>
<td>0.5124</td>
<td>0.4998</td>
<td>0</td>
<td>1</td>
<td>114,804</td>
</tr>
<tr>
<td>employed</td>
<td>0.9424</td>
<td>0.2329</td>
<td>0</td>
<td>1</td>
<td>114,804</td>
</tr>
<tr>
<td>is the borrower a group member</td>
<td>0.0124</td>
<td>0.1105</td>
<td>0</td>
<td>1</td>
<td>114,804</td>
</tr>
<tr>
<td># of current credit lines</td>
<td>10.7388</td>
<td>5.2876</td>
<td>0</td>
<td>64</td>
<td>114,804</td>
</tr>
<tr>
<td># of delinquencies over 30 days</td>
<td>3.6347</td>
<td>6.8248</td>
<td>0</td>
<td>99</td>
<td>114,804</td>
</tr>
</tbody>
</table>

Table 9: Regression of Amount Requested on Whether the Loan is the Second Loan

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>amount_request</th>
</tr>
</thead>
<tbody>
<tr>
<td>second_loan</td>
<td>127.2</td>
</tr>
<tr>
<td></td>
<td>(91.44)</td>
</tr>
<tr>
<td>debt_to_income_high</td>
<td>987.7***</td>
</tr>
<tr>
<td></td>
<td>(92.55)</td>
</tr>
<tr>
<td>Constant</td>
<td>-1,372***</td>
</tr>
<tr>
<td></td>
<td>(268.0)</td>
</tr>
<tr>
<td>Observations</td>
<td>16,820</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.250</td>
</tr>
</tbody>
</table>

Note: Control for Borrowers' observables, year dummies and loan characteristics. Standard errors in parentheses, *** p < 0.01, ** p < 0.05, * p < 0.1.
Table 10: Logit Regression of Default on Whether the First Closed Loan is Defaulted

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>default</td>
<td></td>
</tr>
<tr>
<td>first_close_loan_default</td>
<td>5.008***</td>
</tr>
<tr>
<td></td>
<td>(0.139)</td>
</tr>
<tr>
<td>borrower_rate</td>
<td>9.418***</td>
</tr>
<tr>
<td></td>
<td>(3.126)</td>
</tr>
<tr>
<td>amount_request</td>
<td>3.97e-05***</td>
</tr>
<tr>
<td></td>
<td>(1.02e-05)</td>
</tr>
<tr>
<td>debt_to_income_high</td>
<td>0.313***</td>
</tr>
<tr>
<td></td>
<td>(0.118)</td>
</tr>
<tr>
<td>Constant</td>
<td>-5.475***</td>
</tr>
<tr>
<td></td>
<td>(0.987)</td>
</tr>
</tbody>
</table>

Control for Borrowers’ Char.  | Y                        |
Control for Year Dummies      | Y                        |
Control for Loan Char.        | Y                        |
Observations                   | 4,822                    |

Standard errors in parentheses, *** p < 0.01, ** p < 0.05, * p < 0.1
### Table 11: Estimation Results

#### Panel(A) Utility Primitives

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Notations</th>
<th>Estimates</th>
<th>Std. Err.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk aversion parameter</td>
<td>$\alpha$</td>
<td>1.4981</td>
<td>0.0717</td>
</tr>
<tr>
<td>Effectiveness of effort parameter</td>
<td>$\beta$</td>
<td>2.3738</td>
<td>0.0189</td>
</tr>
<tr>
<td>Cost of effort (low)</td>
<td>$\theta_1$</td>
<td>0.0657</td>
<td>0.0038</td>
</tr>
<tr>
<td>Cost of effort (high)</td>
<td>$\theta_2$</td>
<td>0.8090</td>
<td>0.0322</td>
</tr>
<tr>
<td>Default cost (high)</td>
<td>$R_l - c_h$</td>
<td>-0.5321</td>
<td>0.0190</td>
</tr>
<tr>
<td>Default cost (low)</td>
<td>$R_l - c_l$</td>
<td>0.2370</td>
<td>0.0460</td>
</tr>
<tr>
<td>High revenue</td>
<td>$R_h$</td>
<td>1.4949</td>
<td>0.0059</td>
</tr>
<tr>
<td>Coef. of dti ratio in b’s outside option dist.</td>
<td>$\xi_x$</td>
<td>0.0027</td>
<td>0.0125</td>
</tr>
<tr>
<td>Coef. of loan purpose in b’s outside option dist.</td>
<td>$\xi_d$</td>
<td>-0.4797</td>
<td>0.0127</td>
</tr>
<tr>
<td>Mean of lender’s outside option</td>
<td>$\mu_0$</td>
<td>0.0005</td>
<td>0.0010</td>
</tr>
<tr>
<td>Std. err. of lender’s outside option</td>
<td>$\sigma_0$</td>
<td>0.0261</td>
<td>0.0005</td>
</tr>
</tbody>
</table>

#### Panel(B) Probabilities of Low Cost of Effort

<table>
<thead>
<tr>
<th>Conditional on</th>
<th>Estimates</th>
<th>Std. Err.</th>
</tr>
</thead>
<tbody>
<tr>
<td>high type, high amount</td>
<td>0.8913</td>
<td>0.0126</td>
</tr>
<tr>
<td>high type, low amount</td>
<td>0.9960</td>
<td>0.0006</td>
</tr>
<tr>
<td>low type, high amount</td>
<td>0.6663</td>
<td>0.0123</td>
</tr>
<tr>
<td>low type, low amount</td>
<td>0.6734</td>
<td>0.0118</td>
</tr>
</tbody>
</table>

#### Panel(C) State Transition Probabilities

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Estimates</th>
<th>Std. Err.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Pr$(low dti</td>
<td>high type, low dti)</td>
<td>0.6430</td>
</tr>
<tr>
<td>$Pr$(high dti</td>
<td>high type, high dti)</td>
<td>0.8973</td>
</tr>
<tr>
<td>$Pr$(low dti</td>
<td>low type, low dti)</td>
<td>0.4979</td>
</tr>
<tr>
<td>$Pr$(high dti</td>
<td>low type, high dti)</td>
<td>0.9170</td>
</tr>
</tbody>
</table>

#### Panel(D) Probabilities of High Type

<table>
<thead>
<tr>
<th>Conditional on</th>
<th>Estimates</th>
<th>Std. Err.</th>
</tr>
</thead>
<tbody>
<tr>
<td>other purpose</td>
<td></td>
<td></td>
</tr>
<tr>
<td>low dti, low credit grade</td>
<td>0.2335</td>
<td>0.0092</td>
</tr>
<tr>
<td>high dti, low credit grade</td>
<td>0.2461</td>
<td>0.0092</td>
</tr>
<tr>
<td>low dti, high credit grade</td>
<td>0.5326</td>
<td>0.0099</td>
</tr>
<tr>
<td>high dti, high credit grade</td>
<td>0.5870</td>
<td>0.0122</td>
</tr>
<tr>
<td>debt consolidation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>low dti, low credit grade</td>
<td>0.3219</td>
<td>0.0095</td>
</tr>
<tr>
<td>high dti, low credit grade</td>
<td>0.3311</td>
<td>0.0094</td>
</tr>
<tr>
<td>low dti, high credit grade</td>
<td>0.7320</td>
<td>0.0099</td>
</tr>
<tr>
<td>high dti, high credit grade</td>
<td>0.7323</td>
<td>0.0110</td>
</tr>
</tbody>
</table>
Table 12: Model Fit: Participation and Funding Probabilities

<table>
<thead>
<tr>
<th>Observables</th>
<th>par_prob</th>
<th>fund_prob</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>data</td>
<td>est</td>
</tr>
<tr>
<td>low amount, other purpose</td>
<td></td>
<td></td>
</tr>
<tr>
<td>low dti, low credit grade</td>
<td>0.8828</td>
<td>0.861</td>
</tr>
<tr>
<td>high dti, low credit grade</td>
<td>0.8967</td>
<td>0.8837</td>
</tr>
<tr>
<td>low dti, high credit grade</td>
<td>0.8906</td>
<td>0.8736</td>
</tr>
<tr>
<td>high dti, high credit grade</td>
<td>0.9179</td>
<td>0.8705</td>
</tr>
<tr>
<td>high amount, other purpose</td>
<td></td>
<td></td>
</tr>
<tr>
<td>low dti, low credit grade</td>
<td>0.8984</td>
<td>0.9035</td>
</tr>
<tr>
<td>high dti, low credit grade</td>
<td>0.9046</td>
<td>0.9011</td>
</tr>
<tr>
<td>low dti, high credit grade</td>
<td>0.8726</td>
<td>0.8896</td>
</tr>
<tr>
<td>high dti, high credit grade</td>
<td>0.9151</td>
<td>0.8857</td>
</tr>
<tr>
<td>low amount, debt cons.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>low dti, low credit grade</td>
<td>0.9346</td>
<td>0.9540</td>
</tr>
<tr>
<td>high dti, low credit grade</td>
<td>0.9472</td>
<td>0.9528</td>
</tr>
<tr>
<td>low dti, high credit grade</td>
<td>0.9560</td>
<td>0.9472</td>
</tr>
<tr>
<td>high dti, high credit grade</td>
<td>0.9655</td>
<td>0.9460</td>
</tr>
<tr>
<td>high amount, debt cons.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>low dti, low credit grade</td>
<td>0.9398</td>
<td>0.9611</td>
</tr>
<tr>
<td>high dti, low credit grade</td>
<td>0.9557</td>
<td>0.9600</td>
</tr>
<tr>
<td>low dti, high credit grade</td>
<td>0.9434</td>
<td>0.9520</td>
</tr>
<tr>
<td>high dti, high credit grade</td>
<td>0.9565</td>
<td>0.9509</td>
</tr>
<tr>
<td>Weighted Average</td>
<td>0.9311</td>
<td>0.9314</td>
</tr>
</tbody>
</table>

Note: The funding probabilities are computed conditional on the listings are not withdrawn by borrowers.
Table 13: Model Fit: Default and Late Payment Probabilities

<table>
<thead>
<tr>
<th>Observable</th>
<th>default_prob</th>
<th>late_prob</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>data</td>
<td>est</td>
<td>data</td>
<td>est</td>
</tr>
<tr>
<td>low amount, other purpose</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>low dti, low credit grade</td>
<td>0.2920</td>
<td>0.3284</td>
<td>0.1859</td>
<td>0.1814</td>
</tr>
<tr>
<td>high dti, low credit grade</td>
<td>0.3227</td>
<td>0.3273</td>
<td>0.1581</td>
<td>0.1815</td>
</tr>
<tr>
<td>low dti, high credit grade</td>
<td>0.1312</td>
<td>0.1793</td>
<td>0.0732</td>
<td>0.0775</td>
</tr>
<tr>
<td>high dti, high credit grade</td>
<td>0.1976</td>
<td>0.1654</td>
<td>0.0712</td>
<td>0.0696</td>
</tr>
<tr>
<td>high amount, other purpose</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>low dti, low credit grade</td>
<td>0.3321</td>
<td>0.3049</td>
<td>0.1631</td>
<td>0.1616</td>
</tr>
<tr>
<td>high dti, low credit grade</td>
<td>0.3944</td>
<td>0.3039</td>
<td>0.1665</td>
<td>0.1618</td>
</tr>
<tr>
<td>low dti, high credit grade</td>
<td>0.1761</td>
<td>0.1681</td>
<td>0.0875</td>
<td>0.0702</td>
</tr>
<tr>
<td>high dti, high credit grade</td>
<td>0.2377</td>
<td>0.1552</td>
<td>0.0761</td>
<td>0.0633</td>
</tr>
<tr>
<td>low amount, debt cons.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>low dti, low credit grade</td>
<td>0.2844</td>
<td>0.2845</td>
<td>0.1464</td>
<td>0.1504</td>
</tr>
<tr>
<td>high dti, low credit grade</td>
<td>0.3226</td>
<td>0.2840</td>
<td>0.1228</td>
<td>0.1507</td>
</tr>
<tr>
<td>low dti, high credit grade</td>
<td>0.1240</td>
<td>0.1222</td>
<td>0.0555</td>
<td>0.0443</td>
</tr>
<tr>
<td>high dti, high credit grade</td>
<td>0.1855</td>
<td>0.1226</td>
<td>0.0529</td>
<td>0.0447</td>
</tr>
<tr>
<td>high amount, debt cons.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>low dti, low credit grade</td>
<td>0.2906</td>
<td>0.2643</td>
<td>0.1221</td>
<td>0.1340</td>
</tr>
<tr>
<td>high dti, low credit grade</td>
<td>0.3369</td>
<td>0.2638</td>
<td>0.1130</td>
<td>0.1343</td>
</tr>
<tr>
<td>low dti, high credit grade</td>
<td>0.1555</td>
<td>0.1158</td>
<td>0.0627</td>
<td>0.0413</td>
</tr>
<tr>
<td>high dti, high credit grade</td>
<td>0.1912</td>
<td>0.1162</td>
<td>0.0592</td>
<td>0.0416</td>
</tr>
<tr>
<td>Weighted Average</td>
<td>0.2470</td>
<td>0.2226</td>
<td>0.1079</td>
<td>0.1099</td>
</tr>
</tbody>
</table>

Note: The default and late payment probabilities are computed conditional on the loans are funded.
Table 14: Counterfactual Results: Compare Welfare by Borrowers’ Types

<table>
<thead>
<tr>
<th>Scenarios</th>
<th>Outcome Variables</th>
<th>((\theta_1, c_l))</th>
<th>((\theta_2, c_l))</th>
<th>((\theta_1, c_h))</th>
<th>((\theta_2, c_h))</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>S1: Symmetric</strong></td>
<td>Participation prob.</td>
<td>0.9436</td>
<td>0.9093</td>
<td>0.9337</td>
<td>0.8236</td>
</tr>
<tr>
<td></td>
<td>Funding prob.</td>
<td>0.9992</td>
<td>0.9983</td>
<td>0.9991</td>
<td>0.9975</td>
</tr>
<tr>
<td></td>
<td>Default prob.</td>
<td>0.0936</td>
<td>0.3601</td>
<td>0.0367</td>
<td>0.2067</td>
</tr>
<tr>
<td></td>
<td>Late prob.</td>
<td>0.0112</td>
<td>0.1446</td>
<td>0.0019</td>
<td>0.0510</td>
</tr>
<tr>
<td></td>
<td>Borrower’s avg. util.</td>
<td>0.3409</td>
<td>0.0850</td>
<td>0.2679</td>
<td>-0.4007</td>
</tr>
<tr>
<td></td>
<td>Lender’s avg. util.</td>
<td>0.0840</td>
<td>0.0728</td>
<td>0.0839</td>
<td>0.0644</td>
</tr>
<tr>
<td></td>
<td>Avg. interest rate</td>
<td>0.1351</td>
<td>0.3546</td>
<td>0.1018</td>
<td>0.1769</td>
</tr>
</tbody>
</table>

| S2: Observe Type | Participation prob. | 0.9454 | 0.0000 | 0.9335 | 0.8198 |
|                  | Funding prob.          | 0.3796 | 0.0000 | 0.9991 | 0.9970 |
|                  | Default prob.          | 0.2783 | N/A    | 0.0437 | 0.2540 |
|                  | Late prob.             | 0.0886 | N/A    | 0.0027 | 0.0760 |
|                  | Borrower’s avg. util.  | 0.3768 | N/A    | 0.2647 | -0.3942 |
|                  | Lender’s avg. util.    | -0.0065 | N/A | 0.0831 | 0.0676 |
|                  | Avg. interest rate     | 0.1481 | N/A    | 0.1054 | 0.1988 |

| S3: Asy. Info.  | Participation prob.   | 0.9437 | 0.9400 | 0.9262 | 0.8310 |
|                 | Funding prob.          | 0.4984 | 0.4984 | 0.4984 | 0.4984 |
|                 | Default prob.          | 0.3167 | 0.7904 | 0.0449 | 0.2470 |
|                 | Late prob.             | 0.1138 | 0.6406 | 0.0029 | 0.0719 |
|                 | Borrower’s avg. util.  | 0.3605 | 0.3521 | 0.2142 | 0.1436 |
|                 | Lender’s avg. util.    | -0.0097 | -0.2174 | 0.1378 | 0.1346 |
|                 | Avg. interest rate     | 0.1775 | 0.1777 | 0.1783 | 0.1742 |

| S4: Reputation | Participation prob.   | 0.9374 | 0.9150 | 0.9245 | 0.8148 |
|                | Funding prob.         | 0.9987 | 0.9987 | 0.9987 | 0.9987 |
|                | Default prob.         | 0.1219 | 0.5100 | 0.0416 | 0.2228 |
|                | Late prob.            | 0.0186 | 0.2794 | 0.0025 | 0.0589 |
|                | Borrower’s avg. util. | 0.3137 | 0.2893 | 0.2203 | 0.1451 |
|                | Lender’s avg. util.   | 0.1085 | -0.1064 | 0.1307 | 0.1053 |
|                | Avg. interest rate    | 0.1837 | 0.1833 | 0.1604 | 0.1516 |

| S5: Rep. + PPI | Participation prob. | 0.9387 | 0.9216 | 0.9249 | 0.8200 |
|                | Insurance purchase prob. | 0.3273 | 0.3273 | 0.3273 | 0.3273 |
|                | Funding prob. — w/ ins. | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
|                | Funding prob. — w/o ins. | 0.9980 | 0.9980 | 0.9980 | 0.9980 |
|                | Default prob. — w/ ins. | 0.1293 | 0.5311 | 0.0434 | 0.2304 |
|                | Late prob. — w/o ins.  | 0.0207 | 0.3016 | 0.0027 | 0.0626 |
|                | Borrower’s avg. util.  | 0.3187 | 0.3068 | 0.2279 | 0.1547 |
|                | Lender’s avg. util.    | 0.1154 | -0.0633 | 0.1352 | 0.1067 |
|                | Avg. interest rate     | 0.1820 | 0.1820 | 0.1479 | 0.1330 |

Note: (1) Borrowers’ and lenders’ average utilities and interest rates are computed for funded loans. (2) Default and late payment probabilities are computed conditional on the loans are funded. Under the case where only types are observed by lenders, borrowers with \((\theta_2, c_l)\) are not funded, so default and late payment probabilities are N/A.
Figures

**New loans issued, $bn**

![Bar chart showing growth of Peer-to-Peer Lending Markets](image)

Source: Goldman Sachs; company reports

Figure 3: Growth of Peer-to-Peer Lending Markets
Figure 4: An Example of a Listing
Figure 5: The Histograms of Borrowers’ Stated Monthly Income