Dynamic Contract Pricing with Adverse Selection and Moral Hazard: 
A Structural Analysis of Online Credit Markets

Yi Xin *
Johns Hopkins University
January 15, 2017
PRELIMINARY DRAFT, PLEASE DO NOT CIRCULATE.

Abstract

This paper investigates the effectiveness of reputation/feedback systems in improving welfare in online credit markets when both adverse selection and moral hazard present. Using data from Prosper.com, we find empirical evidence that reputational incentives impact market outcomes. We then develop a finite-horizon dynamic structural model to analyze how borrowers of different types choose effort under the reputation/feedback system, and how funding decisions and equilibrium interest rates are determined. We develop identification strategies to recover the underlying default cost distribution of borrowers, effort choice probabilities, and the utility primitives. In counterfactual analysis, we compare the welfare under three market designs – one with reputation, one without, and one with no asymmetric information. We also consider policy interventions that protect borrowers from accidental loss of reputation. This issue is particularly relevant for small businesses.

*Department of Economics, Wyman Park Building 544E, 3100 Wyman Park Drive, Baltimore, MD 21211, email: yxin4@jhu.edu.
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 lenders, and often require no collateral. Hence, the issue of credit rationing discussed in the seminal work of Stiglitz and Weiss (1981) is likely to be more prevalent compared to traditional markets. Most of these online credit markets rely on a “reputation/feedback” system, which computes and publishes “reputation scores” based on past outcomes to facilitate transactions (see Einav et al, 2015 and Tadelis, 2016). For example, Prosper.com, a leading website for unsecured loans, collects and publishes all borrowers’ credit histories and decides interest rates accordingly. This website now has more than two million members registered and has helped facilitate over five billion dollars in loans over the past ten years. With the rapid emergence of these marketplaces, an understanding of whether and to what extent reputation/feedback systems can improve the total welfare of market participants has become increasingly important. However, these questions have been rarely explored empirically.

Although in theory it is well known that reputation/feedback systems can alleviate inefficiencies through incentivizing agents to exert more effort and gradually revealing their underlying true qualities (Holmstrom, 1999), we seldom observe empirical studies that quantify the effect of this mechanism. The goal of this paper is to investigate the effectiveness of reputation/feedback systems in improving welfare in online credit markets when both hidden information (adverse selection) and hidden actions (moral hazard) present.

To quantify the value of the reputation/feedback system, this paper develops and estimates a dynamic structural model of contract pricing using data from Prosper.com. Our analysis benefits from the setting of this online marketplace in the following ways. First, the reputation/feedback system in this market is representative of other peer-to-peer websites in the sense that it provides clear reputational incentives (or dynamic incentives) through history-dependent pricing schemes. Second, “reputation scores” (credit grade) and “past outcomes” (whether defaults or late payments occur) on this website are more objective compared to reviews or individual rating scores based on the quality of goods or services. Third, the website keeps track of all borrowers’ proposed listings (with borrowers’ detailed characteristics) and

2
the outcomes of each loan for a relatively long time, so that we are able to observe repeated borrowing patterns and how a borrower’s reputation is updated over time. In addition, this website 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 our analysis.

To see the effect of reputational incentives on borrowers’ behavior and market outcomes, we temporarily focus on a group of borrowers who have two overlapping loans. The loans for each borrower are ordered based on their closing dates. Two observations are obtained from the regression analysis on this sample. To begin, defaults for borrower’s first and second closed loans are significantly positively correlated. This finding may be attributed to (a) once default occurs, borrowers knowing that they will not be allowed to enter again anyway, lose reputational incentives of exerting effort in the existing loans, thus leading to a higher chance of default in the second closed loans, and/or (b) there exists an unobserved heterogeneity in borrower’s type that creates intertemporal positive correlation. Our second finding helps us disentangle these two channels. We find that the second closed loans have a significantly higher default rate compared to the first closed loans controlling for many observed characteristics. This result highlights the importance of reputational incentives through channel (a). In other words, if borrowers do not react to reputational incentives and the correlation only comes from unobserved heterogeneity, the default rates for the first and second closed loans should exhibit similar patterns. To summarize, this empirical exercise indicates that borrowers are responsive to different dynamic incentives imposed by the reputation/feedback system, and more importantly, default rates significantly increase when incentives are removed.

Given the model-free evidence, we develop a finite-horizon dynamic structural model to analyze how borrowers of different types choose effort under the reputation/feedback system, and how funding decisions and equilibrium interest rates are determined in online credit markets. Compared with the existing literature on contract pricing which focuses mainly on the role of borrower’s hidden information through screening/signaling mechanisms (see Einav et al, 2012 and Kawai et al, 2016), our paper’s contribution is to incorporate a borrower’s ex-post effort choices. Specifically, we model the process where a borrower with private type information
chooses effort after observing the assigned interest rate for his listing. The interest rate is decided by the website based on the belief about the borrower’s type. If the expected payoff from participation exceeds the outside option draw, the borrower stays and with some probability he is funded by lenders. The effort chosen by the borrower affects the outcomes of the loan, which will further update the beliefs about his type in the future. The motivation of this modeling approach is that we believe borrower’s hidden actions after a loan is funded contribute an additional source of market inefficiency. Moreover, through modeling borrowers’ effort choices under different incentives, we gain a better understanding of to what degree inefficiencies due to moral hazard can be alleviated by reputation/feedback systems.

We specifically consider two scenarios that are associated with different incentives (1) the last period when borrowers have no dynamic concerns, and (2) the preceding periods when borrowers have concerns about future. The main difficulty in solving the model for the preceding periods comes from the fact that borrowers’ strategies need to be consistent with the updated beliefs about their true types in the next period. Intuitively speaking, the continuation value of a borrower depends on the updated beliefs about his type conditional on various outcomes. Given any sets of future beliefs, borrowers of each type choose effort; while the belief is formed through the Bayesian updating process that involves effort choices of all types. Therefore the equilibrium condition requires the effort choices be consistent with the beliefs. In this paper, a numerical algorithm is derived to solve for the equilibrium effort choices for borrowers of different types, as well as the optimal interest rates in a dynamic setting.

In terms of identification, the parameters of interest in our model include the underlying default cost distribution of borrowers, probabilities related to effort choice, and the utility primitives. A three-step identification strategy is developed in the paper to recover the aforementioned unknowns. In the first step, we exploit variations in the transition process of state variables to recover the distribution of the unobserved default cost. In particular, suppose we observe two loans for each borrower, there exist at least three sets of variables that reveal information of a borrower’s type along the timeline. When borrowers first enter the market, their initial characteristics (credit grade etc.) naturally serve as some noisy measures for the true default costs. Across two loans, borrowers’ financial statuses (debt-to-income ratio, home-
ownership, etc.) evolve, and this updating process may be governed by the underlying types and some additional random shocks. At the end of the second loan, realized outcomes also relate to the default costs up to idiosyncratic errors. This is because outcomes of loans are mainly driven by effort choices, which are directly determined by borrowers’ types and cost shifters. Therefore with three pieces of information that are independent conditional on the default cost, we are able to borrow the identification results from the measurement error literature to recover the type distribution. (see Hu, 2008; Hu and Schennach, 2008; Hu and Shum, 2012). In the second step, relying on variations in default and late payment performances, as well as an exogenous variation that shifts borrower’s outside option distribution, we are able to identify the probabilities related to effort choice for any type. Then in the third step, with certain parametric assumptions, from the recovered probabilities of participation and default, we can first pin down the levels of effort and payoffs for borrowers of different types; then with variations in observed interest rates, the utility parameters are identified.

Directly following our identification results, we apply the likelihood-based estimation strategy to the large transaction-level dataset from Prosper.com. Our counterfactual experiments consist of two parts. We first conduct the baseline counterfactuals to quantify the value of the reputation/feedback system in online credit markets. We compare the welfare of market participants under three cases: (1) when the website and lenders perfectly observe borrower’s type; (2) when there is asymmetric information between borrowers and the other parties, but no reputation/feedback system is available; (3) asymmetric information exists, while a reputation/feedback system is implemented. The first two scenarios represent the two extremes, and the difference in welfare levels between them measures the loss from asymmetric information. The extent to which the difference is reduced under Case (3) quantifies the effectiveness of the reputation/feedback system in this market.

The second part of our 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 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, there is still a chance that they obtain unlucky draws of revenues. And if this situation
accidentally occurs, good borrowers are left with no access to future credit and the beliefs about their quality get “stuck”. A natural policy question hence arises – can we find efficient ways to protect borrowers from accidental loss of reputation? If so, what is the monetary value of reputation borrowers would like to pay to avoid being accidentally considered untrustworthy? These questions have strong empirical relevance, especially for small businesses who 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 paper, we focus on three policy interventions that protect borrowers from accidental loss of reputation. First, the website or the government can offer borrowers an option to buy Payment Protection Insurance (PPI), which covers the loan repayments for a set period of time if borrowers are unable to meet them in certain situations. These circumstances usually include “being made redundant or not being able to work because of an accident or illness”.\(^1\) The intuition of this mechanism is straightforward. If a borrower wants to maintain a good reputation (hence credit access in the future), but also worries about future negative shocks, he/she can purchase this insurance to hedge against this risk. Second, the website may set up a Provision Fund by requiring borrowers to pay a risk assurance charge when they apply for a loan. The amount of this charge depends on the characteristics of the borrowers. When a borrower fails to make a payment, the money in the Provision Fund may be used to compensate lenders’ loss. With this ex-ante risk assurance, it may be better for the website to relax its penalties for defaulted borrowers, leading to an improvement in the overall welfare. The last policy is concerned with borrowers’ self-insurance. Borrowers may form a group with collective responsibility, implying that whenever a group member defaults, the reputation of the group is affected. This mechanism provides group borrowers with incentives to help each other and smooth transitory shocks. Our estimates allow us to compare the welfare under these policies, and more interestingly quantify the monetary value of reputation by computing how much individuals would be willing to pay to avoid accidental loss of reputation.

The rest of the paper is organized as follows. Related literature is presented in Section 2. We summarize the data patterns and show empirical evidence of the value of reputation in Section

\(^1\)For details of PPI, see https://www.fca.org.uk/consumers/income-payment-protection.
3. A structural model is provided in Section 4, with the corresponding identification results in Section 5. In Section 6, we present estimation results and details for our counterfactual experiments. Section 7 concludes.

2 Related Literature

First, our paper relates to an empirical literature that uses hedonic regressions to study the value of reputation/review system 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). There is also a small literature studying the value of reputation using structural models. Yoganarasimhan (2013) models buyer’s trade-off between seller’s price and reputation while taking seller’s behavior as given in an online freelance market. Lewis and Zervas (2016) study the hotel industry and find a large welfare loss without the review system. Bai (2016) experimentally demonstrates that reputational incentives induce higher quality and profits, but initial reputation building may need to be subsidized to improve welfare. While qualitatively understanding the effect of reputation is important and necessary for further analysis, estimating a model that explicitly accounts for agents private information and hidden actions under the reputation system can help researchers to quantify the welfare effects of this mechanism and answer additional questions regarding market design and regulations. Our paper will follow the second path.

The second strand of the literature to which our paper is related is the theoretical literature on reputation. The theory begins with Akerlof’s (1970) observation that, in markets where the qualities of a good cannot be observed by the parties involved, potential market failures can be solved by relying on long-term relationships and reputations of sellers. Given different sources of uncertainty to qualities, the survey paper of Bar-Isaac and Tadelis (2008) distinguish between three approaches of modeling reputation: pure hidden information, pure hidden action and mixed models. For more discussion on the hidden action models using repeated game framework, see Klein and Leffler (1981), and a comprehensive survey of Mailath and Samuelson (2006). Our paper is closely related to the third approach, which combines both hidden information and hidden actions. In the signal-jamming model of Holmstrom (1999), the agent
does not know his own type but is motivated to exert effort so that there is a higher chance of success, which will be attributed to a higher probability of good type and be compensated in the future. Bar-Isaac and Tadelis (2008) also provide a variant of this model by allowing the agent to know his own type. Other theoretical papers studying reputation in credit markets include Stiglitz and Weiss (1983), Diamond (1989).

Our paper is also related to an empirical literature on contract pricing in credit markets. Specifically, Adams et al. (2009) and Einav et al. (2012) study the role of down payment as a screening device for high default risk in automobile loan markets. Einav et al. (2013) quantifies the impact of credit scoring on firm’s profits through screening high-risk borrowers and targeting more generous loans to lower-risk borrowers. Kawai et al. (2016) also use data from prosper.com and focus on reserved interest rate of borrowers as a signaling device. Our paper differs from the aforementioned papers in two ways. First of all, we study contract pricing in a dynamic setting, taking borrowers’ intertemporal choices into account. Moreover, our model incorporates both hidden information and hidden actions, which can potentially shed light on moral hazard issues in credit markets.

Besides the complications arising from modeling, the relatively thin empirical literature on contract pricing, especially in dynamic settings, may be primarily attributed to the difficulties of identifying distribution of private information and the utility primitives when effort choices are also unobserved. Perrigne and Vuong (2011) focus on the false moral hazard model developed by Laffont and Tirole (1986) and impose a “truth-telling” condition so that there is a one-to-one mapping between private information and observed prices; Kawai et al. (2016) also rely on the fact that borrower’s type and signal have a one-to-one mapping in a separating equilibrium. Alternatively, Gayle and Miller (2015) study models of managerial compensation and derive identification results under the assumption that some levels of revenue can only be achieved through high effort. Instead, the identification strategy in our paper takes advantage of the dynamic structure. Borrowing method from measurement error literature (Hu and Shum, 2012), we exploit variations in state transition process to recover the unobserved type distribution of borrowers. In addition, we use independence between various loan outcomes conditional on effort to identify the effort choice probabilities.
3  Institutional Background and Data Summary

In this paper, we use a large transaction-level data set from Prosper.com. Institutional details of this website are first introduced, followed by the summary of the data patterns. In addition, we provide empirical evidence that reputational incentives have an impact on market outcomes.

3.1  Institutional Background

Prosper.com, funded in 2005, is one of the largest peer-to-peer lending markets in the US. The aim of this website is to provide a platform for private lenders and borrowers to meet with each other without going through complicated banking systems. Since funded, Prosper has initiated more than five billion dollars in loans and encouraged more than two million individuals to participate in online credit markets. This website makes profit by charging both borrowers and lenders service fees proportional to the amount funded. The market works in the following manner:

(1) To post a listing online, a borrower needs to provide basic information about himself, including social security number, employment status, whether he is a homeowner, annual income etc., to the website. A third-party agency is hired to verify the applicant’s identity and credit histories, including his total number of delinquencies, current number of credit lines, and so on.

(2) After the verification stage, the borrower will be assigned to a credit grade and can then post a listing online, providing the amount requested and some other characteristics such as the purpose of the loan.

(3) Then the website decides the interest rate for each posted listing. After seeing the interest rate, the borrower has an option to withdraw his listing before it is funded.

(4) Once the borrower decides to participate, it takes fourteen days for a listing to expire. Before it expires, lenders can observe all the posted information\(^2\) and decide whether or not

\(^2\)An example of a listing can be found in Figure 1.
to fund this project. As long as the amount requested is reached, the listing is successfully funded and the loan is originated.

(5) In the following 12-60 months, the borrower needs to repay the loan with some possibilities of default and/or late payments occurring in the middle.

There is no collateral required on Prosper.com, which may indicate a higher level of risk faced by the lenders. 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, that is, once default occurs, the borrower involved will not be allowed to borrow again from the website.

3.2 Data Summary

The data we use for this paper include all listings (some become loans) that were originated between January 2011 and December 2014. The clean dataset we obtain contains 113,871 listings that come from 102,881 unique borrowers. We characterize borrowers into four groups based on their repeated borrowing patterns. Borrowers in Category 1 appear only once during the period we observe; borrowers in Categories 2-4 appear at least twice but with different first loans’ outcomes – their first loans may be paid off, still ongoing, or not funded (possibly withdrawn by themselves). The percentage of each category is summarized in Table 1. From this table, we know that 89% of the borrowers appear once, which indicates that the probability of receiving future money demand shock for borrowers is approximately 11%.

For borrowers

---

3 The original dataset contains 192,916 listings. However by November 8, 2016, we still have 63,790 ongoing loans that come from 62,841 unique borrowers. To ensure all individuals have complete loan outcomes, we drop all borrowers with ongoing loans. We futher clean the data by dropping multiple listings within a short period and borrowers with missing information. In addition, we focus on the first two listings for borrowers that have more than two listings for simplicity.

4 According to the rule of this website, borrowers that default once cannot borrow again from this website, so for those borrowers that receive money demand shock, we cannot observe. Moreover, it is hard to rule out the possibility that borrowers may borrow from other places instead even if they need money. Therefore, 11% may underestimate the arrival rate of future money demand shock.
who appear at least twice with their first loans funded, about 60% of them propose their second listings when the first loans are still ongoing.

Figure 2 compares the distributions of credit grade for borrowers’ first and second listings. We can see from this figure that, more borrowers fall into higher credit groups (include AA, A and B) in the second listings. This situation could be a mixed result of sample selection and endogenous updating of borrowers’ credit grades after one loan. We will use a structural model to further disentangle these two cases. To better understand the differences between credit groups, Table 2 summarizes the average amount requested, average interest rates, withdraw and funding probabilities, default and late payment probabilities for borrowers by different credit categories. There is a clear pattern for interest rates: as the credit grade gets lower, borrowers face a higher interest rate on average. Meanwhile, the probabilities that default or late payments occur are also increasing as the credit grade goes down. This phenomenon intuitively captures the trade-off faced by the lenders – higher risks must be compensated by higher returns. Summary statistics of other variables used in the regressions or estimation are provided in Table 3.

3.3 Empirical Evidence on the Value of Reputation

To study the role of reputation/feedback systems in online credit markets, we are interested in whether market prices, borrowers’ behavior and loan outcomes are affected by this mechanism. In this section, we show that borrowers receive lower interest rates after successful loan repayments, while higher interest rates are charged when late payments occur. In addition, we find empirical evidence that default rates increase when dynamic incentives imposed by the reputation/feedback system are removed in this market.

We first investigate whether the interest rates charged on borrowers vary with past outcomes. Table 4 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 the first ones or have late payments prior to them, are associated with higher interest rates. These results
indicate that interest rate as a pricing device could “reward” prior successful loan repayment behavior, and “punish” late payments or uncertainty from overlapping loans. Combining with the institutional feature that once a borrower defaults, he is unable to borrow again, we clearly see that the reputation/feedback system imposes dynamic incentives on forward-looking borrowers through pricing schemes and entry restrictions.

To further figure out the causal relationship between the reputation/feedback system and market outcomes, ideally we need an experimental setting with one group having reputational incentives and the other one serving as a control group. Without such data availability directly, we find one institutional feature very useful. That is, borrowers on this website could propose their second listings even when their first loans are still in process. This feature provides us with a group of borrowers who have two overlapping loans at the same time. The loans for each borrower are ordered based on their closing dates. Suppose borrowers do not react to reputational incentives, then the default behavior is mainly driven by types, implying that the default probabilities for this group of borrowers’ first and second closed loans should be roughly the same. However, if borrowers are forward-looking and react to dynamic incentives, once default occurs, borrowers knowing that they cannot enter again, will lose incentives of paying back their existing loans, which potentially leads to a higher default rate for the second closed loans.

Having these two hypotheses in mind, we obtain two observations from the data. Table 5 shows that defaults for the first and second closed loans are positively correlated. This observation alone cannot reject either hypothesis. However, Column 1 in Table 6 further reflects that the second closed loans have a significantly higher default rate after conditioning on various control variables. The second observation is only consistent with the forward-looking case, which confirms the importance of reputational incentives on market outcomes. For 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 sample of borrowers with long gaps between the starting dates of their two loans to exclude the possibilities that two correlated shocks appear within a short period. All the estimates
from the different specifications have the same sign and similar magnitudes. To summarize, this empirical exercise indicates that borrowers are responsive to different dynamic incentives imposed by the reputation/feedback system, and more importantly, default rates significantly increase when incentives are removed.

4 Model

In this section, we provide details for the finite-horizon dynamic structural model that captures borrowers’ effort choices, lenders’ funding decisions and website’s pricing schemes under asymmetric information in online credit markets. Specifically, we first describe the timeline and assumptions of the model, then discuss two scenarios: (1) the last period when borrowers do not have dynamic concerns, and (2) the preceding periods when borrowers do have concerns about future. To clarify the notations, we use $F(\cdot)$ to represent the cdf of a distribution and $f(\cdot)$ to represent the corresponding pdf. Subscripts are used to distinguish between different distributions.

4.1 Timeline and Model Assumptions

The timeline of our model is presented in Figure 3. At the beginning of each period, a borrower arrives with some observed characteristics and an unobserved type. Lenders and the website do not know the borrower’s type, but the distribution of his type conditional on observables is public information. The website first decides interest rate based on its belief about the borrower’s type. After observing the realized interest rate while knowing his own type, the borrower receives a shock to his cost of effort, and then he can decide the level of effort to exert if his project is funded. By comparing the expected payoff from participation with his outside option draw, the borrower decides whether or not to withdraw his listing. If the borrower chooses to stay, lenders will fund the project whenever the expected revenue exceeds the outside option. For all funded loans, beliefs about the borrowers’ types will be updated after the outcomes realize. If the borrower defaults, he will have no chance to borrow again from this website. Instead, if he pays off the loan, with some probability, the borrower may
receive another money demand shock and propose a new listing with the updated belief.

From the timeline, we can see that there are three important players in this model: the website, the lenders and borrowers. Next we briefly discuss the objective of each party and the assumptions imposed on them. To begin, the website’s objective is to maximize the successful rate of transactions, since it charges a percentage service fee for each funded loan. When deciding the optimal interest rate $r^*$, the website has to consider the reactions from the other two sides, because the successful rate is jointly determined by borrowers’ participation probability and lenders’ funding probability. To better match with the data, we assume that the website may observe an idiosyncratic shock $\varepsilon$ when choosing the interest rate for each borrower.

Lender side’s model in our paper is simplified, since we don’t observe individual lender’s investment behavior; instead we only observe the final funding decision for each posted listing. To deal with this data limitation, we model that the funding decision for each project is made by a representative lender through comparing the expected return with the outside option drawn from $F_{\mu}(\cdot)$. We argue that, the crowdfunding feature of this market provides a rationale for this simplification. Moreover, it is assumed that the representative lender is a one-time player, since lenders in this market don’t have to form a long-term relationship with the website or the borrowers. Note that, the lender makes his funding decision after a borrower chooses to stay, while participation itself may reveal some additional information on the borrower’s type.

Borrowers in our model are assumed to be risk-averse, and more importantly, have dynamic concerns. Borrowers discount future at a rate of $\delta$, and the probability that they receive a money demand shock in each period is $p_m$. Borrowers choose effort level $e$ to maximize expected payoffs, which depend on utilities from revenues, i.e. $U(x; \alpha)$, with $\alpha$ as a risk-averse parameter, and disutilities from exerting effort, i.e. $\phi(e, \theta)$, with $\theta$ as the cost of effort. Following the literature, we assume $U'(\cdot; \alpha) > 0, U''(\cdot; \alpha) < 0, \phi'(\cdot, \theta) > 0$ and $\phi''(\cdot, \theta) > 0$. Borrower’s default cost $c$ is private information, while distribution $F_c(\cdot)$ is known to other players in the market but not to the econometrician. The cost of effort $\theta$ is a transitive shock that realizes for each project before

---

5For the projects that are funded, the average number of investors is 43. Since individual lenders usually invest a small amount on a specific project, the final funding decision is mainly driven by the expected payoff of each project after integrating out individual lender’s characteristics. Therefore it is reasonable to assume that there exists a representative lender that mainly care about the expected return.
the borrower decides the effort level; and its distribution depends on the borrower’s type, i.e. $F_\theta(\cdot | c)$. In addition, assume the borrower draws his outside option $v_0$ for each loan from $F_\xi(\cdot)$. Given the effort, suppose there are two possible revenues that may realize, $R_h$ and $R_l$, and the probability that $R_h$ is realized depends on borrower’s effort through $p(\cdot; \beta)$. Here it is also assumed that $p(e_h; \beta) > p(e_l; \beta)$ if $e_h > e_l$, implying that the revenue distribution given higher effort first order stochastically dominates the one given lower effort. To focus on “ex-ante” moral hazard issues concerned with unobserved effort choices, it is assumed that borrowers will repay the loan whenever $R_h$ is realized and will default when receive $R_l$. In addition, we argue that borrower’s effort may also have an impact on his late payment performances, and the probability that a late payment does not occur is denoted as $L(\cdot; \gamma)$. As long as the borrower pays off the debt in the end, late payments will not affect the lender’s payoffs, but they can help the lender to update his belief about the borrower’s type.

To simplify the notations, let us consider a case where borrowers’ default costs take two values $c_h$ and $c_l$, and costs of effort take two values $\theta_h$ and $\theta_l$ for illustration in the rest of this section. For each borrower, let $W_t = 1$ represent the case in which the borrower withdraws his listing and 0 otherwise; $I_t = 1$ means the lender invests in the project and 0 otherwise; $D_t = 1$ when the borrower defaults at period $t$ and 0 otherwise; in addition, $L_t = 1$ when late payments occur and 0 otherwise. Let $t = 1, 2, \cdots, T$.

### 4.2 Last Period: No Dynamic Concerns

We start with the last period problem when borrowers do not have dynamic concerns. At the beginning of period $T$, suppose the lender believes that the probability that the borrower has $c_h$ as his default cost is $p_T$. For each borrower with type $c = c_h, c_l$ and cost of effort $\theta_T = \theta_h, \theta_l$, after observing interest rate $r_T$, he can decide the optimal effort level through the following

---

6This assumption rules out the possibility of “ex-post” moral hazard in which case borrowers can choose to default even if they have enough money to pay off the debt. Theoretically these two sources of moral hazard cannot be separated empirically without extra information, since they may generate observationally equivalent outcomes; to better understand how the hidden actions are chosen, especially in a dynamic context, our paper adopts the first modeling approach.
payoff maximization problem:

$$e^*_T(r_T, \theta_T, c) = \arg \max_e x_T(e, r_T, \theta_T, c)$$  \hspace{1cm} (4.1)$$
$$x_T(e, r_T, \theta_T, c) = p(e)U(R_h - 1 - r_T) + (1 - p(e))U(R_l - c) - \phi(e, \theta_T)$$

We can further compute borrower’s expected utility given his optimal effort $$\tilde{x}_T(r_T, \theta_T, c) = x_T(e^*_T(r_T, \theta_T, c), r_T, \theta_T, c)$$, as well as lender’s payoff $$\tilde{\pi}_T(r_T, \theta_T, c) = p(e^*_T(r_T, \theta_T, c))(1 + r_T) - 1.$$ Once the borrower has figured out his payoff from participation, if he draws an outside option $$v_0 \leq \tilde{x}_T(r_T, \theta_T, c)$$, he will stay for sure; on the other hand, if the borrower draws a larger outside option, he will withdraw the listing. To summarize, the participation probability for the borrower in the last period is $$F_{\xi}(\tilde{x}_T(r_T, \theta_T, c))$$.

The lender decides whether or not to fund the project after observing the participation of borrowers. As a result, the lender can further update his belief from $$Pr(c_h) = p_T$$ and $$Pr(c_l) = 1 - p_T$$ to

$$Pr(c, \theta_T|W_T = 0, r_T) = \frac{Pr(W_T = 0|c, \theta_T, r_T)Pr(\theta_T|c)Pr(c)}{\sum_{c', \theta'} Pr(W_T = 0|c', \theta_T', r_T)Pr(\theta_T'|c')Pr(c')}.$$  \hspace{1cm} (4.2)$$

where $$Pr(W_T = 0|c, \theta_T, r_T) = F_{\xi}(\tilde{x}_T(r_T, \theta_T, c))$$ is the participation probability for borrowers with $$(r_T, \theta_T, c)$$. The lender will then calculate his expected payoff using this updated belief and the funding probability is therefore derived as a function of the interest rate $$r_T$$ and the original belief $$p_T$$

$$p_f(r_T, p_T) = F_{\mu} \left( \sum_{c', \theta'} Pr(c', \theta'|W_T = 0, r_T)\tilde{\pi}_T(r_T, \theta', c') \right).$$  \hspace{1cm} (4.3)$$

After understanding how borrowers and the lender react to interest rate $$r_T$$, the website can maximize the rate of successful transactions for each belief of the type distribution. Note that, the website decides the interest rate before $$\theta_T$$ is realized and the borrower participates, so it can only take the expectation over the original belief $$p_T$$. Furthermore, a transaction happens only when the borrower participates and gets funded, thus the maximization problem of the website is as follows.

$$\max_r p_f(r, p_T) \left[ p_T \int Pr(W_T = 0|r, \theta', c_h)dF_\theta(\theta'|c_h) + (1 - p_T) \int Pr(W_T = 0|r, \theta', c_l)dF_\theta(\theta'|c_l) \right].$$  \hspace{1cm} (4.4)$$
After solving this problem, the website can decide their last period’s optimal interest rate \( r_T^*(p_T) \) as a function of \( p_T \). Since the website may receive some idiosyncratic shocks for each borrower, the realized interest rate \( r_T \) may deviate from \( r_T^*(p_T) \) up to an error term \( \varepsilon \).

When the borrower calculates his expected payoff of the last loan before entering period \( T \), he needs to integrate out three dimensions of randomnesses. First, the borrower does not know the shock to the website when deciding interest rates; second, he does not know the realization of his cost of effort; third, the borrower has uncertainty about his and the lender’s future outside options (whether he will participate and get funded). Specifically, for any realized \( \theta_T \) and \( r_T = r_T^*(p_T) + \varepsilon \), the borrower with type \( c \) is able to compute his payoff from participation \( \tilde{x}_T(r_T, \theta_T, c) \) and the probability he gets funded \( p_f(r_T, p_T) \) given the belief \( p_T \). Therefore the expected payoff for this borrower integrating over the outside option distribution is derived in the following equation.

\[
\tilde{V}_{Tb}(r_T, p_T, \theta_T, c) = \int_0^{\tilde{x}_T(r_T, \theta_T, c)} [p_f(r_T, p_T) \tilde{x}_T(r_T, \theta_T, c) + (1 - p_f(r_T, p_T))v_0]dF_\xi(v_0) \\
+ \int_{\tilde{x}_T(r_T, \theta_T, c)}^\infty v_0dF_\xi(v_0)
\] (4.5)

The rationale behind Equation (4.5) is that, if the draw of \( v_0 \) is small, then the borrower participates and gets the loan with probability \( p_f(r_T, p_T) \). Given that he is funded, the borrower achieves \( \tilde{x}_T(r_T, \theta_T, c) \); if rejected, the borrower takes the outside option. On the other hand, if \( v_0 \) is large enough, the borrower will directly withdraw his listing. After further integrating out \( \theta \) and \( \varepsilon \), we obtain the borrower’s expected payoff before starting his last loan \( V_{Tb}(p_T, c) \), which is a function of the belief \( p_T \) and his own type \( c \).

### 4.3 Preceding Periods: with Dynamic Concerns

In all preceding periods, when borrowers decide their effort levels, they need to take future continuation values into account. Specifically, borrowers of each type need to solve their effort choices given the beliefs about their types after outcomes of the loans are realized. At the equilibrium, borrowers’ strategies have to be consistent with the beliefs. In this section, we consider how borrowers choose optimal efforts and how interest rates are determined in period \( T - 1 \). The results can be easily generalized for other preceding periods.
We first write down the expected revenue a borrower with \((r_{T-1}, \theta_{T-1}, c)\) can obtain given effort level \(e\), future beliefs \((\hat{p}_T, \tilde{p}_T)\) as follows.

\[
x_{T-1}(e, r_{T-1}, \theta_{T-1}, c, \hat{p}_T, \tilde{p}_T) = p(e)L(e) \left[ U(R_h - 1 - r_{T-1}) + \delta p_m V_{Tb}(\hat{p}_T, c) \right]
+ p(e)(1 - L(e)) \left[ U(R_h - 1 - r_{T-1}) + \delta p_m V_{Tb}(\tilde{p}_T, c) \right]
+ (1 - p(e)) \left[ U(R_l - c) + \delta p_m V_0 \right] - \phi(e, \theta_{T-1}),
\]

where \(p(e)\) and \(L(e)\) represent the probabilities that the borrower pays off the loan and has no late payments respectively. \(V_0\) represents the average value of the outside option a borrower receives once he defaults and is not allowed to borrow from this website again. \(V_{Tb}(\hat{p}_T)\) and \(V_{Tb}(\tilde{p}_T)\) denote the borrower’s expected future continuation value given beliefs \(\hat{p}_T\) and \(\tilde{p}_T\). To be more specific about the beliefs,

\[
\hat{p}_T = Pr(c_h | D_{T-1} = 0, L_{T-1} = 0) \quad \text{and} \quad \tilde{p}_T = Pr(c_h | D_{T-1} = 0, L_{T-1} = 1).
\]

Through the following payoff maximization problem, the borrower can derive his optimal effort level at period \(T - 1\)

\[
e^{*}_{T-1}(r_{T-1}, \theta_{T-1}, c, \hat{p}_T, \tilde{p}_T) = \arg \max_e x_{T-1}(e, r_{T-1}, \theta_{T-1}, c, \hat{p}_T, \tilde{p}_T). \quad (4.7)
\]

Again we can further compute borrower’s expected utility given his optimal effort

\[
\tilde{x}_{T-1}(r_{T-1}, \theta_{T-1}, c, \hat{p}_T, \tilde{p}_T) = x_{T-1}(e^{*}_{T-1}(r_{T-1}, \theta_{T-1}, c, \hat{p}_T, \tilde{p}_T), r_{T-1}, \theta_{T-1}, c, \hat{p}_T, \tilde{p}_T),
\]

as well as lender’s payoff \(\tilde{x}_{T-1}(r_{T-1}, \theta_{T-1}, c, \hat{p}_T, \tilde{p}_T) = p(e^{*}_{T-1}(r_{T-1}, \theta_{T-1}, c, \hat{p}_T, \tilde{p}_T))(1 + r_{T-1}) - 1\).

In order to derive the borrower’s participation probability at period \(T - 1\), we need to first figure out the borrower’s future payoff if he has no loans at \(T - 1\). This situation is realized either when the borrower withdraws his listing \((W_{T-1} = 1)\) or when he decides to participate but is not funded \((W_{T-1} = 0, I_{T-1} = 0)\). Let us denote the union of these two cases as \(O_{T-1, \text{nopar}}\) and suppose the updated belief after observing \(O_{T-1, \text{nopar}}\) is \(\tilde{p}_T\), then the probability that the borrower participates at \(T - 1\) is

\[
Pr(W_{T-1} = 0 | r_{T-1}, \theta_{T-1}, c, \hat{p}_T, \tilde{p}_T) = Pr(\tilde{x}_{T-1}(r_{T-1}, \theta_{T-1}, c, \hat{p}_T, \tilde{p}_T) > v_0 + \delta p_m V_{Tb}(\tilde{p}_T, c)) = F_\xi(\tilde{x}_{T-1}(r_{T-1}, \theta_{T-1}, c, \hat{p}_T, \tilde{p}_T) - \delta p_m V_{Tb}(\tilde{p}_T, c)), \quad (4.8)
\]
where \( v_0 \) represents a draw from the borrower’s outside option distribution. As discussed in Section 4.2, the lender makes funding decision after observing the borrower’s participation, thus the lender can update his belief about the borrower’s type when computing expected revenues. Similarly in this period, if we plug the borrower’s participation probability (as shown in Equation (4.8)) into Equations (4.2) and (4.3), we are able to compute the lender’s funding probability \( p_{T-1}(r_{T-1}, p_{T-1}, \hat{p}_T, \tilde{p}_T, \check{p}_T) \). Notice that \( p_{T-1} \) represents the belief about the borrower’s type at the beginning of period \( T-1 \). The rest of this section illustrates how beliefs are updated after different loan outcomes realize at the end of period \( T-1 \).

First, let us derive \( \hat{p}_T \) and \( \check{p}_T \), the updated probabilities of \( c_h \) given the loan is funded at \( T-1 \) but with different outcomes. To be more specific, with \( Pr(c_h) = p_{T-1} \) and \( Pr(c_l) = 1 - p_{T-1} \), for any interest rate \( r_{T-1} \),

\[
\hat{p}_T = Pr(c_h|D_{T-1} = 0, L_{T-1} = 0, W_{T-1} = 0, I_{T-1} = 1, r_{T-1}) \\
= \frac{Pr(D_{T-1} = 0, L_{T-1} = 0, W_{T-1} = 0, I_{T-1} = 1|c_h, r_{T-1}) Pr(c_h)}{\sum_{c'=c_h,c_l} Pr(D_{T-1} = 0, L_{T-1} = 0, W_{T-1} = 0, I_{T-1} = 1|c', r_{T-1}) Pr(c')},
\]

(4.9)

where

\[
Pr(D_{T-1} = 0, L_{T-1} = 0, W_{T-1} = 0, I_{T-1} = 1|c', r_{T-1}) \\
= \sum_{\theta'} Pr(D_{T-1} = 0|e_{T-1}^*(r_{T-1}, \theta', c, \hat{p}_T, \check{p}_T)) Pr(L_{T-1} = 0|e_{T-1}^*(r_{T-1}, \theta', c, \hat{p}_T, \check{p}_T))
\]

\[
\left(p_{T-1}(r_{T-1}, p_{T-1}, \hat{p}_T, \tilde{p}_T) Pr(W_{T-1} = 0|r_{T-1}, \theta', c', \hat{p}_T, \check{p}_T) Pr(\theta'|c') \right)
\]

(4.10)

Combining information from Equations (4.9) and (4.10), we easily see that for any given \( r_{T-1} \) and \( p_{T-1} \), \( \hat{p}_T \) is a function of \( \hat{p}_T, \tilde{p}_T \) and \( \check{p}_T \). Similar results applies to \( \check{p}_T \) when \( L_{T-1} = 1 \).

Second, we derive \( \tilde{p}_T \), the updated probability of \( c_h \) when there is no loan at \( T-1 \). Again, for any interest rate \( r_{T-1} \),

\[
\tilde{p}_T = Pr(c_h|O_{T-1,\text{nopar}}, r_{T-1}) = \frac{Pr(O_{T-1,\text{nopar}}|c_h, r_{T-1}) Pr(c_h)}{\sum_{c'=c_h,c_l} Pr(O_{T-1,\text{nopar}}|c', r_{T-1}) Pr(c')},
\]

(4.11)

where

\[
Pr(O_{T-1,\text{nopar}}|c', r_{T-1}) \\
= \sum_{\theta'} Pr(W_{T-1} = 1|r_{T-1}, \theta', c', \hat{p}_T, \check{p}_T) Pr(\theta'|c')
\]

\[
+ \sum_{\theta'} (1 - p_{T-1}^f(r_{T-1}, p_{T-1}, \hat{p}_T, \tilde{p}_T, \check{p}_T)) Pr(W_{T-1} = 0|r_{T-1}, \theta', c', \hat{p}_T, \check{p}_T) Pr(\theta'|c').
\]

(4.12)
Equations (4.11) and (4.12) illustrate the third relationship between $\bar{p}_T$ and $(\hat{p}_T, \tilde{p}_T)$.

In the equilibrium, borrowers’ strategies should be consistent with the lender’s belief, which leads to a fixed-point problem for $(\hat{p}_T, \tilde{p}_T, \bar{p}_T)$ for any given interest rate $r_{T-1}$ and initial belief $p_{T-1}$. The changes of interest rate result in the changes of equilibrium beliefs, at the same time also affecting borrowers’ effort choices. Therefore, the website can maximize its objective function, similar to the one in Equation (4.4), by searching for the optimal interest rate after taking borrowers and lenders’ reactions into consideration.

5 Identification

In this section, we study identification issues of the key primitives, which include borrower’s type distribution, effort choice probabilities, and other utility parameters in borrower’s payoff function. Difficulties in identification mainly come from three aspects. First, borrowers’ types and effort choices, that jointly determine the equilibrium outcomes, are both unobserved. Second, we only observe dynamic behaviors for a selected group of borrowers that have paid off their previous loans and choose to participate again. Moreover, primitives in borrower’s payoff function include not only the risk-averse parameter, but also levels of unobserved types, revenues and costs of effort. Given the time range observed from the data, we consider a two-period model in this section and a three-step identification strategy is summarized as follows.

1. Exploit variations in the transition process of state variables to recover the distribution of the unobserved default cost.

2. Use variations in default and late payment performances, as well as an exogenous variation that shifts borrower’s outside option distribution to identify the probabilities related to effort choice for any type.

3. With certain parametric assumptions, from the recovered probabilities of participation and default, first pin down the levels of borrowers’ efforts and payoffs for different types; then with variations in observed interest rates, other utility parameters are identified.

Before presenting details in each step, we fix the notations for the observed data and the
parameters to be recovered. We have a vector of observed state variables for two periods

$$O_{\tau} = \{r_{\tau}, D_{\tau}, L_{\tau}, X_{\tau}, K_{\tau}, W_{\tau}, I_{\tau}\}, \text{ for } \tau = t - 1, t,$$

where $r_{\tau}$ represents interest rate, $D_{\tau} = 1$ represents default and 0 otherwise, $L_{\tau} = 1$ when late payment occurs and 0 otherwise, $X_{\tau}$ and $K_{\tau}$ represent 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 the lender. To further simplify the notation, define the state variables that are not involved in the selection mechanism as

$$S_{\tau} = \{r_{\tau}, L_{\tau}, X_{\tau}, K_{\tau}\} \text{ for } \tau = t - 1, t.$$  

Therefore $O_{\tau} = \{S_{\tau}, D_{\tau}, W_{\tau}, I_{\tau}\}$. In the data, we observe $O_{i\tau}$ for $\tau = t - 1, t$ and $i = 1, 2, \ldots, N$.

In the dynamic process, let $e_{\tau}$ denote borrower’s effort choice at each period and $c$ denote borrower’s default cost, which is a fixed type across time periods. Borrower’s effort is uniquely determined for any given realization of the cost of effort $\theta_{\tau}$, type $c$ and the interest rate $r_{\tau}$ through the borrower’s payoff maximization problem. Thus, the distribution (including levels) of the unobserved private information $\{\theta_{\tau}, c\}_{\tau=t-1,t}$ is the main focus of our identification strategy. For illustration, let us consider the case where $c = c_h, c_l$ and $\theta_{\tau} = \theta_h, \theta_l$ for $\tau = t - 1, t$. For other primitives in the model, we impose the following parametric assumptions. Assume the borrower has a CARA utility function, $U(x) = 1 - \exp(-\alpha x)$, where $\alpha$ is the risk-averse parameter. Borrower’s cost function is $\phi(e) = \theta e^2$. Assume for each loan, there are two possible revenues that may realize, $R_h$ and $R_l$. The probability that $R_h$ is realized is $p(e) = 1 - \exp(-\beta e)$; the probability that a late payment does not occur is denoted as $L(e) = 1 - \exp(-\gamma e)$. Other unknown distributions include borrower’s outside option distribution $F_{\xi}(\cdot)$ and that of lender’s, $F_{\mu}(\cdot)$. To summarize, we need to recover $\{Pr(c_h), Pr(\theta_h|c_h), Pr(\theta_l|c_l), \alpha, \beta, \gamma, R_h, R_l, F_{\xi}(\cdot), F_{\mu}(\cdot)\}$ based on the assumption that the conditional distribution of $\theta_{\tau}$ on $c$ is time-invariant.

For the rest of this paper, we 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 \( f_{Y_1|Y_2=y_2}(y_1|y_2) \).

5.1 Identification of the Type Distribution

We first impose assumptions on the dynamic process.

Assumption 1 (Conditional Independence). The dynamic process \( \{O_t, \theta_t, c_t\} \) satisfies

1. the first order Markov process;

2. for borrowers who do not default at \( t - 1 \) and get funded at \( t \), the Markov kernel can be decomposed as

\[
\bar{f}_{D_t,I_t=1,W_t=0,S_t,\theta_t|D_{t-1}=0,I_{t-1}=1,W_{t-1}=0,S_{t-1},\theta_{t-1},c} = \bar{f}_{D_t|I_t=1,W_t=0,S_t,\theta_t|D_{t-1}=0,I_{t-1}=1,W_{t-1}=0,S_{t-1},\theta_{t-1},c} \\
\times \bar{f}_{I_t=1,W_t=0,S_t,\theta_t|D_{t-1}=0,I_{t-1}=1,W_{t-1}=0,S_{t-1},\theta_{t-1},c} \\
\times \bar{f}_{X_t|X_{t-1},c} \times \bar{f}_{K_t|K_{t-1},D_{t-1}=0,L_{t-1}=0,L_{t-1}=0} \times \bar{f}_{\theta_t|c} \\
= \bar{f}_{D_t,I_t=1,W_t=0,S_t,\theta_t|D_{t-1}=0,I_{t-1}=1,W_{t-1}=0,S_{t-1},\theta_{t-1},c} \\
\times \bar{f}_{I_t=1,W_t=0,S_t,\theta_t|D_{t-1}=0,I_{t-1}=1,W_{t-1}=0,S_{t-1},\theta_{t-1},c} \\
\times \bar{f}_{X_t|X_{t-1},c} \times \bar{f}_{K_t|K_{t-1},D_{t-1}=0,L_{t-1}=0,L_{t-1}=0} \times \bar{f}_{\theta_t|c}.
\]

Assumption 1 has several important implications. First, conditional on effort \( e_t = \tilde{e}_t^*(r_t, \theta_t, c) \), \( D_t \) and \( L_t \) are independent. This assumption is directly motivated by the model, since the probabilities of defaults and late payments are only related to effort levels up to some idiosyncratic shocks. Second, interest rates and probabilities that projects are funded are only related to observables, so that \( f_{I_t=1|W_t=0,r_t,X_t,K_t,D_{t-1}=0,L_{t-1}} \) and \( f_{r_t|X_t,K_t,D_{t-1}=0,L_{t-1}} \) can be directly estimated from the data. This assumption is consistent with the institutional setting—under asymmetric information, the website and the lender can only make decisions based on the observables. Moreover, the probability that the borrower participates is related to his expected payoff \( \tilde{x}_t(r_t, \theta_t, c) \) and his outside option distribution, which depends on his characteristics. The transition of borrower’s financial status \( X_t \) is governed by \( c \); but the transition of credit grade on this website \( K_t \) is purely determined by the previous outcomes, meaning that \( f_{K_t|K_{t-1},D_{t-1}=0,L_{t-1}} \) can be also directly recovered from the data. In addition, borrower’s type relates to the distribution of \( \theta_t \), which is assumed to be time-invariant. Note that, \( D_{t-1} = 0 \) directly indicates that \( I_{t-1} = 1, W_{t-1} = 0 \), so we may drop these two terms in \( f_{D_t,I_t=1,W_t=0,S_t,\theta_t|D_{t-1}=0,L_{t-1},X_{t-1},K_{t-1},c} \).
Assumption 1 also implies that, $\theta_{t-1}$ does not enter the Markov transition kernel, thus we can integrate out $\theta_t$ from $f_{D_t,I_t=1,W_t=0,S_t,\theta_t|D_{t-1}=0,L_{t-1},X_{t-1},K_{t-1},c}$ without losing Markovian properties. In particular,

$$f_{D_t,I_t=1,W_t=0,S_t|D_{t-1}=0,L_{t-1},X_{t-1},K_{t-1},c} = \sum_{\theta_t} f_{D_t,I_t=1,W_t=0,S_t,\theta_t|D_{t-1}=0,L_{t-1},X_{t-1},K_{t-1},c} = \sum_{c} f_{D_t,I_t=1,W_t=0|r_t,X_t,K_t,D_{t-1}=0,L_{t-1},c} \cdot f_{r_t|X_t,K_t,D_{t-1}=0,L_{t-1}} \cdot f_{X_t|X_{t-1},c} \cdot f_{K_t|K_{t-1},D_{t-1}=0,L_{t-1},X_{t-1},K_{t-1},c}. \tag{5.1}$$

Then the joint distribution of observables for borrowers who have loans in $t-1$ and $t$ (i.e. $I_t = 1, W_t = 0, D_{t-1} = 0$) helps to identify the distribution of default cost through the following equation,

$$f_{r_t|X_t,K_t,D_{t-1}=0,L_{t-1}} \cdot f_{K_t|K_{t-1},D_{t-1}=0,L_{t-1}} = \sum_{c} f_{D_t,I_t=1,W_t=0|r_t,X_t,K_t,D_{t-1}=0,L_{t-1},c} \cdot f_{X_t|X_{t-1},c} \cdot f_{K_t|K_{t-1},D_{t-1}=0,L_{t-1},X_{t-1},K_{t-1},c}. \tag{5.2}$$

Equation (5.2) serves as the key identifying equation for kernels related to the unobserved type, i.e. $f_{D_{t-1}=0,L_{t-1},X_{t-1},K_{t-1},c}$. This equation provides three pieces of information that are independent conditional on $c$. Specifically, type is related to the outcomes of the second loan after integrating out effort; and type governs the transition process of financial status; borrower’s initial characteristics also reveal some information about his type. However, from $f_{D_{t-1}=0,L_{t-1},X_{t-1},K_{t-1},c}$, we can only recover the type distribution conditional on $D_{t-1} = 0$, but not when $D_{t-1} = 1$. This is due to the selection problem in the data – at period $t$, we only observe borrowers who get and pay off their debt at $t-1$. Further analysis will be presented in Section 5.3 to show how to recover the the original type distribution before selection.

We provide assumptions required to identify the type distribution using Equation (5.2). Suppose $D_t$, $K_{t-1}$, $c$ are all discrete random variables. Use $j_d = 1, \ldots, J_d$, $j_k = 1, \ldots, J_k$ and $j_c = 1, \ldots, J_c$ to index the values of these three variables. Then we can rewrite Equation (5.2) in matrix form for any fixed values of $\{r_t, K_t, L_t, L_{t-1}, I_t, W_t, D_{t-1}\}$. To first fix notations, for $r_t = \bar{r}_t, K_t = \bar{K}_t, L_t = \bar{L}_t, L_{t-1} = \bar{L}_{t-1}, I_t = 1, W_t = 0$ and $D_{t-1} = 0$, we define the following
matrices for \((X_t, X_{t-1})\),

\[
M_{D_t,K_{t-1},X_t,X_{t-1}} = \left[ \begin{array}{c} f_{D_t,L_t=1,W_t=0,\bar{r}_t,\bar{L}_t,X_t,K_t,D_{t-1}=0,\bar{L}_{t-1},X_{t-1},K_{t-1}} \end{array} \right]_{j_d,j_k} \]

\[
M_{D_t,c,X_t} = \left[ \begin{array}{c} f_{D_t,L_t=1,W_t=0,\bar{r}_t,\bar{L}_t,X_t,K_t,D_{t-1}=0,\bar{L}_{t-1},c} \end{array} \right]_{j_d,j_c} \]

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

\[
M_{c,K_{t-1},X_t,X_{t-1}} = \left[ f_{D_{t-1}=0,X_t=0,\bar{X}_t=0,K_{t-1},L_{t-1},c} \right]_{c=j_c,K_{t-1}=j_k} \]

The matrix form of Equation (5.2) for any \((X_t, X_{t-1})\) is therefore

\[
M_{D_t,K_{t-1},X_t,X_{t-1}} = M_{D_t,c,X_t}M_{c,X_t,X_{t-1}}M_{c,K_{t-1},X_{t-1}}. \tag{5.4}
\]

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

\[
\left( M_{D_t,K_{t-1},\bar{X}_t,\bar{X}_{t-1}} \cdot M_{D_t,K_{t-1},\hat{X}_t,\hat{X}_{t-1}}^{-1} \right) \left( M_{D_t,K_{t-1},\bar{X}_t,\bar{X}_{t-1}} \cdot M_{D_t,K_{t-1},\hat{X}_t,\hat{X}_{t-1}} \right)^{-1} = M_{D_t,c,\bar{X}_t} \left( M_{c,\bar{X}_t,\bar{X}_{t-1}} \cdot M_{c,\hat{X}_t,\hat{X}_{t-1}} \right) \left( \bar{M}_{D_t,c,\bar{X}_t} \right)^{-1} \tag{5.5}
\]

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 \((X_t, X_{t-1})\).

It is straightforward to see that Assumption 2 requires \(j_d = j_c = j_k\). For discrete case, we can always regroup different values to make the numbers of categories equal for \(D_t, c\) and \(K_{t-1}\). However we admit that, this may take less advantage of the data variation. We now consider the economic meaning of Assumption 2. Suppose there are only two types, \(c = c_h, c_l\). For \(M_{D_t,c,X_t}\) to be invertible, it is sufficient to have the following inequalities satisfied,

\[
f_{D_t=1,L_t=1,W_t=0,\bar{r}_t,\bar{X}_t,K_t,D_{t-1}=0,\bar{L}_{t-1},c_h} < f_{D_t=1,L_t=1,W_t=0,\bar{r}_t,\bar{X}_t,K_t,D_{t-1}=0,\bar{L}_{t-1},c_l} < f_{D_t=1,L_t=1,W_t=0,\bar{r}_t,\bar{X}_t,K_t,D_{t-1}=0,\bar{L}_{t-1},c_h} < f_{D_t=1,L_t=1,W_t=0,\bar{r}_t,\bar{X}_t,K_t,D_{t-1}=0,\bar{L}_{t-1},c_l} \]

These inequalities intuitively imply that borrowers with low types are more likely to default. Suppose there are two categories of credit score, \(K_1 > K_2\). Then it is sufficient to have
\[ f_{D_{t-1}=0,X_{t-1},K_{t-1},L_{t-1},c_h} > f_{D_{t-1}=0,X_{t-1},K_{2},L_{t-1},c_h} \text{ and } f_{D_{t-1}=0,X_{t-1},K_{1},L_{t-1},c_l} < f_{D_{t-1}=0,X_{t-1},K_{2},L_{t-1},c_l}, \]

which imply that high type borrowers are more likely to have a higher credit grade and low type borrowers are more likely to have a lower grade, to ensure the invertibility of \( M_{c,K_{t-1},X_{t-1}} \).

Finally, if the transition probabilities for all combinations of \((X_{t-1}, X_{t})\) conditional on different types are non-zero, \( M_{c,X_{t},X_{t-1}} \) is invertible. With Assumption 2 satisfied, Equation (5.5) leads to an eigenvalue-eigenvector decomposition of the matrix on its left hand side. To guarantee the uniqueness of the decomposition, we invoke the following assumption.

**Assumption 3 (Uniqueness).** The following statements are equivalent for uniqueness:

(i) \( f_{D_{t}=1,L_{t}=1,W_{t}=0 | r_{t}, X_{t}, K_{t}, D_{t-1}=0, L_{t-1}, c} \) decreases with \( c \);

(ii) \( f_{D_{t-1}=0,X_{t-1},K_{t-1},L_{t-1},c} \) increases with \( K_{t-1} \) for \( c = c_h \) and decreases with \( K_{t-1} \) for \( c = c_l \);

(iii) \( f_{X_{t} | X_{t-1},c} \) increases with \( c \) if \( X_{t} > X_{t-1} \) and vice versa.

**Theorem 1 (Identification).** If Assumptions 1, 2 and one of (i)-(iii) in Assumption 3 are satisfied, \( \{O_{\tau}\}_{\tau=t-1,t} \) for borrowers with \( I_{t} = 1, W_{t} = 0 \) and \( D_{t-1} = 0 \) identifies \( f_{X_{t} | X_{t-1},c} \).

The formal proof of Theorem 1 follows Hu (2008). For continuous case, see Hu and Schennach (2008).

### 5.2 Identification of the Effort Choice Probabilities

Since the effort choice \( e_{t} \) is uniquely determined by \( r_{t}, \theta_{t} \) and \( c \), to identify the effort choice probabilities, we essentially need to identify the distribution of \( \theta_{t} \) using variations in loan outcomes given different types \( c \). Specifically, we can decompose \( f_{D_{t},L_{t},I_{t}=1,W_{t}=0 | r_{t}, X_{t}, K_{t}, D_{t-1}=0, L_{t-1}, c} \) with \( \theta_{t} \) following Assumption 1.

\[
\begin{align*}
 f_{D_{t},L_{t},I_{t}=1,W_{t}=0 | r_{t}, X_{t}, K_{t}, D_{t-1}=0, L_{t-1}, c} \\
 = f_{I_{t}=1 | W_{t}=0, r_{t}, X_{t}, K_{t}, D_{t-1}=0, L_{t-1}, c} \cdot \sum_{\theta_{t}} (f_{D_{t} | r_{t}, \theta_{t}, c} \cdot f_{L_{t} | r_{t}, \theta_{t}, c} \cdot f_{W_{t}=0 | X_{t}, r_{t}, \theta_{t}, c} \cdot f_{\theta_{t} | c})
\end{align*}
\]
which is equivalent to
\[
\frac{f_{D_t,L_t,K_t,t=1,W_t=0|r_t,X_t,K_t,D_t-1=0,L_t-1,c}}{f_{t=1|W_t=0,r_t,X_t,K_t,D_t-1=0,L_t-1}} = \sum_{\theta_t} \left(f_{D_t|r_t,\theta_t,c} \cdot f_{L_t|r_t,\theta_t,c} \cdot f_{W_t=0,\theta_t|X_t,r_t,c}\right). 
\]
(5.6)

The kernels on the left hand side of Equation (5.6) are either identified or directly estimable from the data. On the right hand side, we have default and late payment behaviors that are independent conditional on the effort \(e_t^*(r_t, \theta_t, c)\); furthermore, borrower’s participation probability is affected by the effort and the covariates that are related to his outside option distribution. Following similar strategy, this equation leads to a unique eigenvalue-eigenvector decomposition under certain assumptions, therefore we can identify \(f_{D_t|r_t,\theta_t,c}\), \(f_{L_t|r_t,\theta_t,c}\) and \(f_{W_t=0,\theta_t|X_t,r_t,c}\) for any fixed values of \(\{r_t, K_t, L_t, L_{t-1}, c\}\). We provide the theorem and the assumptions required to identify the kernels related to effort choice in Appendix A.

### 5.3 Identification of Other Parameters

After we have recovered \(f_{W_t=0,\theta_t|X_t,r_t,c}\), our goal is to pin down the unknown parameters in borrower’s payoff function. Borrowers’ participation decisions are made by comparing utilities from participation with the draws of the outside option, thus we can only identify the utility levels relative to the outside option. We assume borrower’s outside option follows an exponential distribution with the mean related to the observable \(X_t\).  

Without loss of generality, we normalize the location of the outside option distribution by setting the mean to be \(\xi X_t\) with \(\xi > 0\). This normalization also implies that borrowers with better financial status have a better outside option. Then borrower’s expected payoff \(\bar{x}_t(r_t, \theta_t, c)\) determines his participation probability through \(f_{W_t=0|X_t,r_t,\theta_t,c} = 1 - \exp\left(-\frac{1}{\xi X_t}\bar{x}_t(r_t, \theta_t, c)\right)\). The following equation that involves \(\xi\) and \(\bar{x}_t(r_t, \theta_t, c)\) as unknowns is obtained subsequently. Suppose \(X_t\) takes two values \(X_1\) and \(X_2\),
\[
\frac{f_{W_t=0,\theta_t|X_1,r_t,c}}{f_{W_t=0,\theta_t|X_2,r_t,c}} = \frac{f_{W_t=0|X_1,r_t,\theta_t,c}}{f_{W_t=0|X_2,r_t,\theta_t,c}} = \frac{1 - \exp\left(-\frac{1}{\xi X_1}\bar{x}_t(r_t, \theta_t, c)\right)}{1 - \exp\left(-\frac{1}{\xi X_2}\bar{x}_t(r_t, \theta_t, c)\right)}. 
\]
(5.7)

It is straightforward to see that Equation (5.7) identifies \(\frac{\bar{x}_t(r_t, \theta_t, c)}{\xi}\); thus to further normalize the scale, we set \(\xi = 1\). To summarize, the ratio of \(f_{W_t=0,\theta_t|X_t,r_t,c}\) for different \(X_t\)’s creates variations

---

8 The data we have provide no information for recovering borrower’s outside option distribution nonparametrically, thus we impose some parametric assumptions.
to identify borrower’s payoff level given \((r_t, \theta_t, c)\). Once \(\tilde{x}_t(r_t, \theta_t, c)\) is recovered, we can easily compute the participation probability of borrowers, which further leads to the identification of the conditional distribution of \(\theta_t\),

\[
f_{\theta_t|c} = \frac{f_{W_{t=0,0.01|x_t,r_t,c}}}{1 - \exp(-\frac{1}{X_t}\tilde{x}_t(r_t, \theta_t, c))}. \tag{5.8}
\]

Next, we consider the unknown parameters in \(\tilde{x}_t(r_t, \theta_t, c)\). Explicitly,

\[
\tilde{x}_t(r_t, \theta_t, c) = p(\tilde{e}_t^*(r_t, \theta_t, c); \beta)U(R_h - 1 - r_t; \alpha) + (1 - p(\tilde{e}_t^*(r_t, \theta_t, c); \beta))U(R_l - c; \alpha) - \phi(\tilde{e}_t^*(r_t, \theta_t, c), \theta_t), \tag{5.9}
\]

where \(\tilde{e}_t^*(r_t, \theta_t, c)\) is the level of optimal effort given \((r_t, \theta_t, c)\). 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. Even though we have recovered the probabilities of each type in Section 5.1 and the conditional probabilities of \(\theta_t\) in Equation (5.8), we still don’t know the levels of \(\theta_t\) and \(c\), i.e., \(\{\theta_h, \theta_l, c_h, c_l\}\). Another important unknown parameter is \(\alpha\), which characterizes the absolute risk-aversion in a CARA utility function.

Using variations from observed interest rates, the following theorem shows the identification of \(R_h\) and \(\alpha\).

**Theorem 2** (Identification of the risk-averse parameter). Given two values of observed interest rate, \(\bar{r}_t\) and \(\hat{r}_t\), \(\alpha\) and \(R_h\) are identified.

The intuition of Theorem 2 is as follows. With \(\tilde{x}_t(r_t, \theta_t, c)\) recovered, its derivative with respect to \(r_t\) is also recovered. It is shown that \(\frac{\partial \tilde{x}_t(r_t, \theta_t, c)}{\partial r_t}\) uniquely determines \(U'(R_h - 1 - r_t)\). Then with CARA utility, when we have two different values of interest rate, the ratio of marginal utilities at different \(r_t\) identifies \(\alpha\) and \(R_h\). The proof is in Appendix B.

To identify the levels of \(c\) and \(\theta_t\), relate the identified density \(f_{D_t=0|r_t, \theta_t, c}\) with the effort level through \(p(\cdot; \beta)\)

\[
f_{D_t=0|r_t, \theta_t, c} = p(\tilde{e}_t^*(r_t, \theta_t, c); \beta) = 1 - \exp(-\beta \tilde{e}_t^*(r_t, \theta_t, c)) \tag{5.10}
\]

\(^9\)In our model we assume that whenever \(R_h\) is realized, the borrower will pay off his loan and \(p(e; \beta)\) represents the probability of \(R_h\) is realized given effort level \(e\).
Equation (5.10) for \( c = c_h, c_l \) uniquely determines \( \frac{\tilde{e}_t^*(r, \theta_h, c_h)}{\tilde{e}_t^*(r, \theta_l, c_l)} \). If we rewrite Equation (5.9) for \( c_h \) and \( c_l \) and take the ratios, we are able to construct the following two equations for \( \theta_h \) and \( \theta_l \):

\[
\left( \frac{\tilde{e}_t^*(r, \theta_h, c_h)}{\tilde{e}_t^*(r, \theta_l, c_l)} \right)^2 = \frac{p(\tilde{e}_t^*(r, \theta_h, c_h))U(R_h - 1 - r_t) + (1 - p(\tilde{e}_t^*(r, \theta_h, c_h)))U(R_l - c_h) - \tilde{x}(r_t, \theta_h, c_h)}{p(\tilde{e}_t^*(r, \theta_l, c_l))U(R_h - 1 - r_t) + (1 - p(\tilde{e}_t^*(r, \theta_l, c_l)))U(R_l - c_l) - \tilde{x}(r_t, \theta_l, c_l)}. \tag{5.11}
\]

The two equations in (5.11) provide us with a system of two linear equations of \( U(R_l - c_h) \) and \( U(R_l - c_l) \). To ensure identification of levels of \( R_l - c \)\(^{10}\), we need the following rank condition to hold.

**Assumption 4 (Rank Condition).**

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

where \( \delta_1 = \left( \frac{\tilde{e}_t^*(r, \theta_h, c_h)}{\tilde{e}_t^*(r, \theta_l, c_l)} \right)^2 \) and \( \delta_2 = \left( \frac{\tilde{e}_t^*(r, \theta_l, c_l)}{\tilde{e}_t^*(r, \theta_h, c_h)} \right)^2 \), \( p_1 = p(\tilde{e}_t^*(r, \theta_h, c_h)), p_2 = p(\tilde{e}_t^*(r, \theta_l, c_l)), p_3 = p(\tilde{e}_t^*(r, \theta_l, c_l)), p_4 = p(\tilde{e}_t^*(r, \theta_h, c_h)). \)

Note that all terms in Assumption 4 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 the levels of \( \theta_h \) and \( \theta_l \), it is necessary to normalize \( \beta \).\(^{11}\)

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

\[
\begin{align*}
\hat{f}_{D_{t-1}=0|X_{t-1}=0, W_{t-1}=0, r_{t-1}, X_{t-1}, K_{t-1}} &= \sum_{\theta_{t-1}, c} \left( f_{D_{t-1}=0|r_{t-1}, \theta_{t-1}, c} \cdot f_{W_{t-1}=0|X_{t-1}, r_{t-1}, \theta_{t-1}, c} \cdot f_{\theta_{t-1}|c} \cdot f_{c|X_{t-1}, K_{t-1}} \right) \tag{5.12}
\end{align*}
\]

\(^{10}\)We are not able to separately identify the levels of \( R_l \) and \( c \), since only their difference matters in the model.

\(^{11}\)We can easily identify the ratio between \( \theta_h \) and \( \theta_l \) without knowing the exact effort level. However, since \( \beta \) and \( \theta_l \) enter into the model with effort level in a non-separable way, we cannot recover all of them simultaneously. We decide to normalize \( \beta \) and identify the costs of effort for better economic interpretation.
where the funding probability $f_{it-1=1|W_{t-1}=0,r_{t-1},X_{t-1},K_{t-1}}$ at period $t-1$ can be directly estimated from the data, $f_{W_{t-1}=0|X_{t-1},r_{t-1},\theta_{t-1},c}$ represents the participation probability of borrowers with \{X_{t-1},r_{t-1},\theta_{t-1},c\}, which can be computed using borrowers’ optimal effort choices at period $t-1$. Specifically, we need to solve borrower’s optimization problem when there are dynamic concerns given all other parameters identified in the previous steps. The only unknowns we have in Equation (5.12) are the original type distribution conditional on observables $f_{c|X_{t-1},K_{t-1}}$.

6 Estimation Results and Counterfactuals

Following our identification strategies in Section 5, we construct the likelihood for all individuals in the sample. Let $i = 1, \cdots, N$ be the index for each individual borrower; and $t - 1$ and $t$ represent two loans. $O_{i\tau} = \{r_{i\tau}, D_{i\tau}, L_{i\tau}, X_{i\tau}, K_{i\tau}, W_{i\tau}, I_{i\tau}\}_{\tau=t-1,t}$ contains all observed data for individual $i$, where $r_{i\tau}$ is the interest rate for $i$ at period $\tau$, $D_{it}$ and $L_{it}$ represent the default and late payment performances, $X_{i\tau}$ and $K_{i\tau}$ denote the borrower’s observed financial status and credit grade from the website, $W_{i\tau}$ represents the withdraw decision made by the borrower, and $I_{i\tau}$ represents funding decision made by the lender. With $\Theta$ denoting the vector of parameters identified and to be estimated, we construct the log-likelihood for the whole sample as follows.

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

$$= \sum_{i}^{N} \log(\sum_{c} f_{c|X_{it-1},K_{it-1}} \cdot f_{X_{it-1},K_{it-1}} \cdot \left[ \sum_{\theta_{it-1}} f_{D_{it-1},L_{it-1},I_{it-1},W_{it-1},r_{it-1},\theta_{it-1}|c,X_{it-1},K_{it-1}} \right]$$

$$\cdot f_{X_{t1}|X_{it-1},c} f_{K_{it}|K_{it-1},D_{it-1},L_{it-1}} \cdot \left[ \sum_{\theta_{it}} f_{D_{it},L_{it},I_{it},W_{it},r_{it},\theta_{it}|c,X_{it},K_{it},D_{it-1},L_{it-1}} \right]).$$

(6.1)

where $f_{D_{it-1},L_{it-1},I_{it-1},W_{it-1},r_{it-1},\theta_{it-1}|c,X_{it-1},K_{it-1}}$ represents the joint likelihood of borrower’s observables (and $\theta_{t-1}$) at period $t-1$. Given the conditional independence assumed earlier, we can decompose this likelihood for the borrowers with $I_{it-1} = 1, W_{it-1} = 0$ as,

$$f_{D_{it-1},L_{it-1},I_{it-1},W_{it-1},r_{it-1},\theta_{it-1}|c,X_{it-1},K_{it-1}}$$

$$= f_{D_{it-1}|r_{it-1},\theta_{it-1},c} \cdot f_{L_{it-1}|r_{it-1},\theta_{it-1},c} \cdot f_{I_{it-1}=1|W_{it-1}=0,r_{it-1},X_{it-1},K_{it-1}} \cdot f_{W_{it-1}=0|X_{it-1},r_{it-1},\theta_{it-1},c}$$

$$\cdot f_{r_{it-1}|X_{it-1},K_{it-1}} \cdot f_{\theta_{it}|c}$$

29
where \( f_{I_{it-1}=1|W_{it-1}=0,r_{it-1},X_{it-1},K_{it-1}} \) and \( f_{r_{it-1}|X_{it-1},K_{it-1}} \) are directly estimable from the data; \( f_{D_{it-1}|r_{it-1},\theta_{it-1},c}, f_{L_{it-1}|r_{it-1},\theta_{it-1},c} \) and \( f_{W_{it-1}=0|X_{it-1},r_{it-1},\theta_{it-1},c} \) can be computed using the optimal effort level solved from the model given \( \Theta \); and \( f_{\theta_{it}|c} \) are primitives to be estimated. For borrowers with different participation and funding patterns, we need to slightly modify the likelihood. For the borrower’s joint likelihood of observables at period \( t \), similar approach applies. To control for other characteristics, we let \( R_h \) be associated with some loan characteristics, such as the purposes of the loan (whether the loan is used for debt consolidation); we also allow the distribution of \( \theta_t \) conditional on \( c \) to rely on some observables. In addition, we multiply the revenue \( R_{h-1} - 1 - r_t \) with the amount requested to account for the actual monetary value from each loan.

The estimation results are shown in Table 7 with stand errors computed using bootstrap. We separate the estimates into two panels, where Panel (A) focuses on utility primitives and Panel (B) shows the estimates of type distribution and transition probabilities of state variables. The risk-averse parameter we get is essentially around \( 2.4 \times 10^{-5} \), since we divide the amount with \( 10^5 \) when computing the revenue from each loan. This normalization is implemented purely to improve the precision of estimating the risk-averse parameter. Our estimates are within the range obtained in the literature using CARA utility function, see Cohen and Einav (2007). The last line of Panel (A) also shows that, if the loan is used for debt consolidation, the average revenue is higher. Panel (B) in Table 7 shows that, borrowers who have higher default cost and request a larger amount are more likely to draw a smaller cost of effort. Moreover, when borrowers have high debt-to-income ratios, there is a high chance for them to stay in that bad situation (borrowers with \( c_h \) have 80% probability to stay in \( X_2 \), and the chance is above 96% for borrowers with \( c_l \)); and borrowers with \( c_h \) are more likely to stay with low debt-to-income ratios. The last four rows in Panel (B) present the estimates of original type distribution given different combinations of observables. It is clear to see that, borrowers with high credit grade and low debt-to-income ratio are more likely to be “good” borrowers, that is, they have high default cost.

Our plan for counterfactual analysis using the estimates in this paper consists of two parts. We will firstly conduct the baseline counterfactuals to quantify the value of reputation/feedback
system in online credit markets. Given that the utility primitives and the distribution of
borrowers’ types have been recovered, we compare the welfare of market participants under three
cases: (1) when the website and the lender perfectly observe borrower’s type; (2) when there is
asymmetric information between borrowers and the other parties, but no reputation/feedback
system is available; (3) asymmetric information exists, while a reputation/feedback system is
imposed. The first two scenarios represent the two extremes, and the difference in welfare
levels between them measures the loss from asymmetric information. The extent to which
the difference is reduced under Case (3) quantifies the effectiveness of the reputation/feedback
system in this market.

The second part of our 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 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, there is still a chance that they obtain unlucky draws of revenues. And if this situation
accidentally occurs, good borrowers are left with no access to future credit and the beliefs about
their quality get “stuck”. A natural policy question hence arises – can we find efficient ways
to protect borrowers from accidental loss of reputation? If so, what is the monetary value of
reputation borrowers would like to pay to avoid being accidentally considered untrustworthy?
These questions have strong empirical relevance, especially for small businesses who 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 paper, we focus on three policy interventions that protect borrowers from accidental
loss of reputation. First, the website or the government can offer borrowers an option to buy
Payment Protection Insurance (PPI), which covers the loan repayments for a set period of
time if borrowers are unable to meet them in certain situations. These circumstances usually
include being made redundant 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 (hence credit access in the future), but also worries about future negative shocks,
he/she can purchase this insurance to hedge against this risk. Second, the website may set up
a Provision Fund by requiring borrowers to pay a risk assurance charge when they apply for a loan. The amount of this charge depends on the characteristics of the borrowers. When a borrower fails to make a payment, the money in the Provision Fund may be used to compensate lenders’ loss. With this ex-ante risk assurance, it may be better for the website to relax its penalties for defaulted borrowers, leading to an improvement in the overall welfare. The last policy is concerned with borrowers’ self-insurance. Borrowers may form a group with collective responsibility, implying that whenever a group member defaults, the reputation of the group is affected. This mechanism provides group borrowers with incentives to help each other and smooth transitory shocks. Our estimates allow us to compare the welfare under these policies, and more interestingly to quantify the monetary value of reputation by computing how much individuals would be willing to pay to avoid accidental loss of reputation.

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 present. Using data from Prosper.com, we find empirical evidence that reputational incentives have an impact on market outcomes. We then develop a finite-horizon dynamic structural model to analyze how borrowers of different types choose effort under the reputation/feedback system, and how funding decisions and equilibrium interest rates are determined. We develop identification strategies to recover the underlying default cost distribution of borrowers, effort choice probabilities, and the utility primitives. We also implement a likelihood-based estimation strategy to get point estimates of the key parameters. In counterfactual analysis, we plan to compare the welfare under three market designs – one with reputation, one without, and one with no asymmetric information. Moreover, using estimated parameters, we study policy interventions that protect borrowers from accidental loss of reputation. Specifically, we analyze the effect of providing Payment Protection Insurance to borrowers, setting up Provision Fund and forming collective responsibilities among borrowers. We can also quantify the monetary value of reputation through computing how much individuals would be willing to pay to avoid accidental loss of reputation. These policies are particularly relevant to the development of small businesses.
References


A Identification of the Effort Choice Probabilities

Suppose $D_t, \theta_t$ and $X_t$ are all discrete 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. Then we can rewrite Equation (5.6) in matrix form for any fixed values of $\{r_t, K_t, L_t, L_{t-1}, I_t, W_t, D_{t-1}, c\}$. To first fix notations, for $r_t = \hat{r}_t, K_t = K_t, L_t = L_t, L_{t-1} = L_{t-1}, I_t = 1, W_t = 0, D_{t-1} = 0$, and $c = \bar{c}$ we define the following matrices,

$$M_{D_t, L_{t-1}=1, X_t} = \left[ \frac{f_{D_t, L_{t-1}=1, X_t}}{f_{D_t=1}} \right]_{j_d, j_x}$$

$$M_{D_t, \theta_t} = \left[ f_{D_t, \theta_t} \left| D_t = j_d, \theta_t = j_\theta \right. \right]_{j_d, j_\theta}$$

$$M_{\theta_t, L_{t-1}=1} = \text{diag} \left\{ f_{L_{t-1}=1} \mid \theta_t = j_\theta \right\}_{j_\theta=1,2,\ldots,J_\theta}$$

$$M_{\theta_t, X_t} = \left[ f_{W_t=0, \theta_t} \mid X_t, \bar{c}, \theta_t = j_\theta, X_t = j_x \right]_{j_\theta, j_x}.$$ 

The matrix form of Equation (5.6) for $L_t = 1$ is therefore

$$M_{D_t, L_{t-1}=1, X_t} = M_{D_t, \theta_t} M_{\theta_t, L_{t-1}=1} M_{\theta_t, X_t}. \quad (A.2)$$

Similarly for $L_t = 0$, we have

$$M_{D_t, L_{t-1}=0, X_t} = M_{D_t, \theta_t} M_{\theta_t, L_{t-1}=0} M_{\theta_t, X_t}. \quad (A.3)$$

Combining Equations (A.2) and (A.3), we are able to construct the following equation

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

$$= M_{D_t, \theta_t} M_{\theta_t, L_{t-1}=1} M_{\theta_t, L_{t-1}=0}^{-1} M_{D_t, \theta_t}^{-1} \quad (A.4)$$

provided that the following assumption is satisfied.

**Assumption 2’ (Invertibility).** Matrices $M_{D_t, \theta_t}, M_{\theta_t, L_{t-1}=1}, M_{\theta_t, X_t}$ are invertible.

The economic intuition behind Assumption 2’ is as follows. Consider a case when $\theta_t = \theta_{h_t}, \theta_t$, $X_t = X_1, X_2$. We need to have $f_{D_t=0} \mid r_t, \theta_{h_t}, c < f_{D_t=0} \mid r_t, \theta_t, c$ 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 $M_{\theta_t, L_{t-1}=1}$ to be invertible, we require the probability that late payment
occurs to be non-zero, which also leads to the invertibility of $M_{\theta_t, L_t=0}$. To take a further look at the matrix $M_{\theta_t, X_t}$, since $f_{W_t=0, \theta_t|X_t, r_t, c} = f_{W_t=0|X_t, r_t, \theta_t, c} \cdot f_{\theta_t|c}$,

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

As a result, to achieve the invertibility of $M_{\theta_t, X_t}$, we need to have:

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

or alternatively:

$$\frac{f_{W_t=0|\theta_h, X_1, r_t, c}}{f_{W_t=0|\theta_l, X_1, r_t, c}} \neq \frac{f_{W_t=0|\theta_h, X_2, r_t, c}}{f_{W_t=0|\theta_l, X_2, r_t, c}}. \quad (A.5)$$

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, thus the ratio of the participation probabilities given $\theta_h$ and $\theta_l$ gets closer to 1. For example, consider the extreme case when $X_t = X_2$, the mean of the outside option is larger than the maximum payoff borrowers can get from the website, then both types will be willing to stay out and the ratio between their participation probabilities equals to 1. When other values of $X_t$ is lower enough to generate possibility that borrowers could still be better off by participating, the condition in Equation (A.5) holds. Similarly, with Assumption 2′ satisfied, Equation (A.4) leads to an eigenvalue-eigenvector decomposition of the matrix on its left hand side. To guarantee the uniqueness of the decomposition, we invoke the following assumption.

**Assumption 3′ (Uniqueness).** The following statements are equivalent for uniqueness:

(i) $f_{D_t=0|r_t, \theta_t, c}$ decreases with $\theta_t$;

(ii) $f_{L_t=1|r_t, \theta_t, c}$ increases with $\theta_t$;

**Theorem 3 (Identification).** In addition to the assumptions in Theorem 1, if Assumptions 2 and one of (i)-(ii) in Assumption 3′ are satisfied, $f_{D_t|r_t, \theta_t, c}, f_{L_t|r_t, \theta_t, c}$ and $f_{W_t=0, \theta_t|X_t, r_t, c}$ are identified for any fixed values of $\{r_t, K_t, L_t, L_{t-1}, c\}$. 

37
B Proof of Theorem 2

Since we have recovered $\tilde{x}_t(r_t, \theta_t, c)$ for any $r_t$, the derivative of $\tilde{x}_t(r_t, \theta_t, c)$ with respect to $r_t$,

$$
\frac{\partial \tilde{x}_t(r_t, \theta_t, c)}{\partial r_t} = \frac{\partial \tilde{e}^*_t(r_t, \theta_t, c)}{\partial r_t} (p'(\tilde{e}^*_t(r_t, \theta_t, c))(U(R_h - 1 - r_t) - U(R_i - c)) - \phi'(\tilde{e}^*_t(r_t, \theta_t, c)))
$$

$$
- p(\tilde{e}^*_t(r_t, \theta_t, c))U'(R_h - 1 - r_t)
$$

$$
= - p(\tilde{e}^*_t(r_t, \theta_t, c))U'(R_h - 1 - r_t)
$$

(B.1)

is also identified. Notice that, the second equality in Equation (B.1) holds because $\tilde{e}^*_t(r_t, \theta_t, c)$ satisfies the first order condition in borrower’s optimization problem. Observe also that, $p(\tilde{e}^*_t(r_t, \theta_t, c)) = f_{D_t=0|r_t, \theta_t, c}$, which has been recovered from the matrix decomposition in Section 5.2. Therefore Equation (B.1) identifies $U'(R_h - 1 - r_t)$, which is only related to $R_h$ and $\alpha$ given any observed $r_t$. With the assumption of CARA utility, $U'(R_h - 1 - r_t) = \alpha \exp(-\alpha(R_h - 1 - r_t))$; then with two observed interest rates $\bar{r}_t$ and $\hat{r}_t$, we are able to identify $\alpha$ through

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

(B.2)

Once $\alpha$ is recovered, plug it back to $U'(R_h - 1 - r_t) = \alpha \exp(-\alpha(R_h - 1 - r_t))$, we can further identify $R_h$. \qed


### Table 1: Repeated Borrowing Pattern

<table>
<thead>
<tr>
<th>Data Category</th>
<th>Note</th>
<th>Freq.</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>borrowers who only appear once</td>
<td>91,891</td>
<td>89.32</td>
</tr>
<tr>
<td>2</td>
<td>borrowers who appear twice: first loan is paid off</td>
<td>4176</td>
<td>4.06</td>
</tr>
<tr>
<td>3</td>
<td>borrowers who appear twice: first loan is ongoing</td>
<td>6038</td>
<td>5.87</td>
</tr>
<tr>
<td>4</td>
<td>borrowers who appear twice: first listing is withdrawn or unfunded</td>
<td>776</td>
<td>0.76</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td>102,881</td>
<td>100</td>
</tr>
</tbody>
</table>
Table 2: Summary of Average Amount, Average Interest Rate, Withdraw and Funding Probabilities, Default and Late Payment Probabilities by Category

Panel (a)

<table>
<thead>
<tr>
<th>Credit Grade</th>
<th>Avg. Amount Requested</th>
<th>Avg. Interest Rate</th>
<th>Withdraw Prob. (%)</th>
<th># of Obs</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>13297.23</td>
<td>0.0754</td>
<td>6.50</td>
<td>8,062</td>
</tr>
<tr>
<td>A</td>
<td>12975.14</td>
<td>0.1098</td>
<td>5.60</td>
<td>20,923</td>
</tr>
<tr>
<td>B</td>
<td>12981.03</td>
<td>0.1473</td>
<td>5.79</td>
<td>22,045</td>
</tr>
<tr>
<td>C</td>
<td>11792.36</td>
<td>0.1859</td>
<td>5.58</td>
<td>24,748</td>
</tr>
<tr>
<td>D</td>
<td>9145.873</td>
<td>0.2359</td>
<td>7.85</td>
<td>17,993</td>
</tr>
<tr>
<td>E</td>
<td>5103.02</td>
<td>0.2821</td>
<td>6.25</td>
<td>12,174</td>
</tr>
<tr>
<td>HR</td>
<td>3585.395</td>
<td>0.3153</td>
<td>11.32</td>
<td>7,926</td>
</tr>
<tr>
<td>Overall</td>
<td>10641.77</td>
<td>0.1838</td>
<td>6.52</td>
<td>113,871</td>
</tr>
</tbody>
</table>

Panel (b)

<table>
<thead>
<tr>
<th>Credit Grade</th>
<th>Funding Prob. (%)</th>
<th>Default Prob. (%)</th>
<th>Late Prob. (%)</th>
<th># of Obs</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>89.23</td>
<td>6.38</td>
<td>2.59</td>
<td>8,062</td>
</tr>
<tr>
<td>A</td>
<td>91.01</td>
<td>14.07</td>
<td>5.58</td>
<td>20,923</td>
</tr>
<tr>
<td>B</td>
<td>91.16</td>
<td>21.47</td>
<td>8.50</td>
<td>22,045</td>
</tr>
<tr>
<td>C</td>
<td>91.99</td>
<td>28.34</td>
<td>10.93</td>
<td>24,748</td>
</tr>
<tr>
<td>D</td>
<td>87.58</td>
<td>31.27</td>
<td>15.38</td>
<td>17,993</td>
</tr>
<tr>
<td>E</td>
<td>92.17</td>
<td>33.56</td>
<td>16.39</td>
<td>12,174</td>
</tr>
<tr>
<td>HR</td>
<td>71.66</td>
<td>33.87</td>
<td>20.74</td>
<td>7,926</td>
</tr>
<tr>
<td>Overall</td>
<td>89.36</td>
<td>24.10</td>
<td>10.70</td>
<td>113,871</td>
</tr>
</tbody>
</table>
Table 3: 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>total_list</td>
<td>1.1930</td>
<td>0.3947</td>
<td>1</td>
<td>2</td>
<td>113,871</td>
</tr>
<tr>
<td>total_success</td>
<td>1.0764</td>
<td>0.5182</td>
<td>0</td>
<td>2</td>
<td>113,871</td>
</tr>
<tr>
<td>term</td>
<td>42.3036</td>
<td>11.2950</td>
<td>12</td>
<td>60</td>
<td>113,871</td>
</tr>
<tr>
<td>debt_consolidation</td>
<td>0.6724</td>
<td>0.4694</td>
<td>0</td>
<td>1</td>
<td>113,871</td>
</tr>
<tr>
<td>home_improvement</td>
<td>0.0723</td>
<td>0.2590</td>
<td>0</td>
<td>1</td>
<td>113,871</td>
</tr>
<tr>
<td>business</td>
<td>0.0501</td>
<td>0.2182</td>
<td>0</td>
<td>1</td>
<td>113,871</td>
</tr>
<tr>
<td>fico_score_below600</td>
<td>0.3126</td>
<td>0.4636</td>
<td>0</td>
<td>1</td>
<td>113,871</td>
</tr>
<tr>
<td>debt_to_income_high</td>
<td>0.5004</td>
<td>0.5000</td>
<td>0</td>
<td>1</td>
<td>113,871</td>
</tr>
<tr>
<td>home_owner</td>
<td>0.5120</td>
<td>0.4999</td>
<td>0</td>
<td>1</td>
<td>113,871</td>
</tr>
<tr>
<td>employed</td>
<td>0.9423</td>
<td>0.2332</td>
<td>0</td>
<td>1</td>
<td>113,871</td>
</tr>
<tr>
<td>is_group_member</td>
<td>0.0121</td>
<td>0.1093</td>
<td>0</td>
<td>1</td>
<td>113,871</td>
</tr>
<tr>
<td>current_credit_lines</td>
<td>10.7161</td>
<td>5.2831</td>
<td>0</td>
<td>64</td>
<td>113,871</td>
</tr>
<tr>
<td>delinquency_over_30days</td>
<td>3.6435</td>
<td>6.8463</td>
<td>0</td>
<td>99</td>
<td>113,871</td>
</tr>
<tr>
<td>VARIABLES</td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>borrower_rate</td>
<td>borrower_rate</td>
<td>borrower_rate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>second_loan</td>
<td>-0.00800***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.000317)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>overlap</td>
<td>0.00142***</td>
<td>0.001000**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.000390)</td>
<td>(0.000396)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>late-ever</td>
<td></td>
<td>0.00403***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.000680)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>amount_request</td>
<td>1.76e-07***</td>
<td>2.53e-09</td>
<td>1.04e-09</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.69e-08)</td>
<td>(3.38e-08)</td>
<td>(3.37e-08)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>debt_to_income_high</td>
<td>0.00151***</td>
<td>0.00165***</td>
<td>0.00162***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.000317)</td>
<td>(0.000413)</td>
<td>(0.000413)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.316***</td>
<td>0.306***</td>
<td>0.306***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.000924)</td>
<td>(0.00156)</td>
<td>(0.00155)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Control for Borrowers’ Char. | Y | Y | Y |
Control for Year Dummies    | Y | Y | Y |
Control for Loan Char.      | Y | Y | Y |
Observations                | 20,428 | 10,214 | 10,214 |
R-squared                   | 0.929 | 0.927 | 0.927 |
Table 5: Logit Regression of Default on Whether the First Closed Loan is Defaulted

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>default</th>
</tr>
</thead>
<tbody>
<tr>
<td>first_close_loan_default</td>
<td>5.023***</td>
</tr>
<tr>
<td></td>
<td>(0.136)</td>
</tr>
<tr>
<td>borrower_rate</td>
<td>14.25***</td>
</tr>
<tr>
<td></td>
<td>(2.785)</td>
</tr>
<tr>
<td>amount_request</td>
<td>4.53e-5***</td>
</tr>
<tr>
<td></td>
<td>(9.45e-06)</td>
</tr>
<tr>
<td>debt_to_income_high</td>
<td>0.329***</td>
</tr>
<tr>
<td></td>
<td>(0.109)</td>
</tr>
<tr>
<td>Constant</td>
<td>-7.000***</td>
</tr>
<tr>
<td></td>
<td>(0.890)</td>
</tr>
</tbody>
</table>

Control for Borrowers’ Char.  Y
Control for Year Dummies  Y
Control for Loan Char.  Y
Observations  5,554
Table 6: 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>second_close_loan</td>
<td>0.397***</td>
<td>0.384***</td>
<td>0.513***</td>
<td>0.495***</td>
</tr>
<tr>
<td></td>
<td>(0.0464)</td>
<td>(0.0707)</td>
<td>(0.0824)</td>
<td>(0.143)</td>
</tr>
<tr>
<td>borrower_rate</td>
<td>14.58***</td>
<td>14.37***</td>
<td>17.99***</td>
<td>19.71***</td>
</tr>
<tr>
<td></td>
<td>(1.273)</td>
<td>(1.856)</td>
<td>(2.081)</td>
<td>(3.511)</td>
</tr>
<tr>
<td>amount_request</td>
<td>3.90e-05***</td>
<td>5.97e-05***</td>
<td>6.31e-05***</td>
<td>8.03e-05***</td>
</tr>
<tr>
<td></td>
<td>(4.62e-06)</td>
<td>(6.94e-06)</td>
<td>(8.60e-06)</td>
<td>(1.30e-05)</td>
</tr>
<tr>
<td>debt_to_income_high</td>
<td>0.246***</td>
<td>0.237***</td>
<td>0.241***</td>
<td>0.224*</td>
</tr>
<tr>
<td></td>
<td>(0.0490)</td>
<td>(0.0724)</td>
<td>(0.0832)</td>
<td>(0.133)</td>
</tr>
<tr>
<td>Constant</td>
<td>-4.975***</td>
<td>-5.178***</td>
<td>-6.577***</td>
<td>-7.109***</td>
</tr>
<tr>
<td></td>
<td>(0.392)</td>
<td>(0.574)</td>
<td>(0.664)</td>
<td>(1.087)</td>
</tr>
</tbody>
</table>

Control for Borrowers’ Char.  | Y          | Y          | Y          | Y          |
Control for Year Dummies     | Y          | Y          | Y          | Y          |
Control for Loan Char.        | Y          | Y          | Y          | Y          |
Observations                  | 11,812     | 5,418      | 4,330      | 1,958      |

Column 1: full sample
Column 2: borrowers whose second loans are not for debt consolidation
Column 3: borrowers whose initial FICO scores are below 600
Column 4: borrowers with long gaps between the starting dates of their two loans
### Table 7: Estimation Results

#### (A) Utility Primitives

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Estimates</th>
<th>Std</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>2.4228</td>
<td>0.0013</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>3.5975</td>
<td>0.0023</td>
</tr>
<tr>
<td>$\theta_l$</td>
<td>0.1122</td>
<td>0.0023</td>
</tr>
<tr>
<td>$\theta_h$</td>
<td>0.9720</td>
<td>0.0019</td>
</tr>
<tr>
<td>$R_l - c_h$</td>
<td>-0.3581</td>
<td>0.0027</td>
</tr>
<tr>
<td>$R_l - c_l$</td>
<td>0.3713</td>
<td>0.0019</td>
</tr>
<tr>
<td>$R_h$ intercept</td>
<td>4.0013</td>
<td>0.0018</td>
</tr>
<tr>
<td>$R_h$ slope</td>
<td>0.4144</td>
<td>0.0012</td>
</tr>
</tbody>
</table>

#### (B) Other Probabilities

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Estimates</th>
<th>Std</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Pr(\theta_l</td>
<td>c_h, A_1)$</td>
<td>0.9189</td>
</tr>
<tr>
<td>$Pr(\theta_l</td>
<td>c_h, A_2)$</td>
<td>0.8979</td>
</tr>
<tr>
<td>$Pr(\theta_l</td>
<td>c_l, A_1)$</td>
<td>0.2785</td>
</tr>
<tr>
<td>$Pr(\theta_l</td>
<td>c_l, A_2)$</td>
<td>0.1782</td>
</tr>
<tr>
<td>$Pr(X_1</td>
<td>c_h, X_1)$</td>
<td>0.6753</td>
</tr>
<tr>
<td>$Pr(X_2</td>
<td>c_h, X_2)$</td>
<td>0.8799</td>
</tr>
<tr>
<td>$Pr(X_1</td>
<td>c_l, X_1)$</td>
<td>0.4809</td>
</tr>
<tr>
<td>$Pr(X_2</td>
<td>c_l, X_2)$</td>
<td>0.9198</td>
</tr>
<tr>
<td>$Pr(c_h</td>
<td>X_1, K_1)$</td>
<td>0.5690</td>
</tr>
<tr>
<td>$Pr(c_h</td>
<td>X_2, K_1)$</td>
<td>0.5321</td>
</tr>
<tr>
<td>$Pr(c_h</td>
<td>X_1, K_2)$</td>
<td>0.8112</td>
</tr>
<tr>
<td>$Pr(c_h</td>
<td>X_2, K_2)$</td>
<td>0.7614</td>
</tr>
</tbody>
</table>

Note: We normalize $\beta = 2, \xi = 0.5$ for this estimation. $A_1$ represents that borrower requests a larger amount; $X_1$ means the borrower has low debt-to-income ratio; $K_1$ means the borrower has low credit grade.
Figures

Figure 1: An Example of a Listing
Figure 2: The Histograms of Credit Grade for Borrowers' First and Second Listings
Figure 3: The Timeline of the Model

- Website with prior $p_0$ decides $r_1$; borrower decides $e_1$ if participates
- Borrower makes withdraw decision; Lenders make funding decisions
- Outcome of the first loan realizes; belief updated
- Demand shock arrives with probability $p_m$, second loan begins

- Shock to cost of effort realizes
- Shock to borrower’s outside option realizes
- Shock to lender’s outside option realizes
- $D_t = 1$: no entry again,
  $D_t = 0, L_1 = 0$
  $D_t = 0, L_1 = 1$
- Second loan follows similar procedure