Asymmetric Information, Reputation, and Welfare in Online Credit Markets

Yi Xin *
California Institute of Technology

May 17, 2019

Abstract

This paper studies the welfare impact of reputation/feedback systems in markets where both adverse selection and moral hazard are present. Using a transaction-level dataset from an online credit market, I estimate a dynamic model of borrowers and lenders, where borrowers are subject to reputational incentives. I quantify the welfare loss from adverse selection and moral hazard separately and find that 78 percent of overall inefficiency is induced by moral hazard. When the reputation system is implemented, I find that 95 percent of total welfare loss from asymmetric information is eliminated. I also consider a policy intervention that protects borrowers from accidental loss of reputation. My results suggest that introducing a payment protection insurance into the market further improves total welfare.

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

JEL Code: L14, D82, G21, C14

*Division of Humanities and Social Sciences, California Institute of Technology, 1200 E California Blvd, MC 228-77, Pasadena, CA 91125. Email: yixin@caltech.edu. I thank my dissertation advisors Yingyao Hu, Elena Krasnokutskaya, and Yuya Takahashi for their guidance and encouragement throughout the research process. This work has benefited particularly from the comments of Jorge Balat, Ying Chen, Laura Doval, Michael Gibilisco, Yajing Jiang, Robert Moffitt, Jian Ni, Nicholas Papageorge, Kota Saito, Yuya Sasaki, Matthew Shum, and Ruli Xiao. I also thank seminar and conference participants at Caltech, Cornell, Johns Hopkins, Maryland, Penn State, SUNY-Albany, Toronto, UBC-Sauder, UC-Riverside, UNC-Chapel Hill, WEAI Graduate Student Workshop (2017), NASMES (2017), DC IO Conference (2017), IIOC-Rising Stars (2018), and SEA Conference (2018). The usual disclaimer applies.

1 Introduction

Online credit markets for peer-to-peer lending have developed rapidly over the last several years.[1] These markets attract dispersed and anonymous borrowers and often require no collateral. The ability of lenders to recover their loan amounts is one of the main concerns in markets of unsecured loans, where the problems of asymmetric information are two-fold. On the one hand, borrowers differ in their inherent risks, which is hidden information; "lemons" are more likely to self-select into the markets. On the other hand, borrowers' actions are not observed but impose externalities on lenders; as a result, additional incentives are necessary to motivate borrowers to exert effort to repay their debts. Most online credit markets rely on a reputation/feedback system which computes and publishes reputation scores based on past outcomes to facilitate transactions.[2] To what extent and through which channels do reputation systems improve the total welfare of market participants? Answers to these questions shed light on optimal mechanism design and regulations, not only for fast-growing online marketplaces but also for traditional credit markets, and thus have become increasingly important. Nevertheless, these questions have received little attention in the empirical literature.[3]

This paper quantifies the extent to which reputation/feedback systems improve the total welfare of market participants when both hidden information (adverse selection) and hidden actions (moral hazard) are present. To do so, I develop and estimate a dynamic model of borrowers and lenders, where borrowers are subject to reputational incentives, using a large transaction-level dataset from Prosper.com. I find that 95 percent of total welfare loss from asymmetric information is eliminated when the reputation system is implemented. I quantify the welfare loss from adverse selection and moral hazard separately and find that 78 percent of overall inefficiency is induced by moral hazard in online credit markets. In this paper, I also consider a policy intervention that protects borrowers from accidental loss of reputation. My results suggest that introducing a payment protection insurance into the market further improves total welfare.

Prosper.com is one of the leading peer-to-peer lending marketplaces in the United States and provides services that match lenders with borrowers. It collects information on borrowers' credit profiles and loan histories and decides on interest rates accordingly. My analysis benefits from the setting of Prosper in the following ways. First, the reputation/feedback

[1]See the trend of growth in Figure 1 in the online appendix Section C. The link to the online appendix is https://drive.google.com/open?id=1QSjHR3Y7W4Tsl1lXfTJIgXgXohux4R3q.

[2]For reviews of reputation/feedback systems in online markets, see Einav, Farronato, and Levin (2016) and Tadelis (2016).

[3]There is a theoretical literature on the qualitative effect of reputation systems; see Akerlof (1970), Holmstrom (1999), Bar-Isaac and Tadelis (2008), Stiglitz and Weiss (1983), and Diamond (1989).

system in this market is representative of that in other peer-to-peer websites. It provides clear reputational incentives through history-dependent pricing schemes. Second, "reputation scores" (credit grades) and past outcomes (whether defaults or late payments occur) on this website are more objective compared to customer reviews or individual rating scores based on the quality of goods or services.[4] Third, the website keeps track of all borrowers' proposed listings with detailed loan-level and individual-level characteristics as well as the outcomes of each loan. Having access to all listings enables me to observe repeated borrowing patterns, which are essential for studying reputation mechanisms. In addition, Prosper prohibits once-defaulted borrowers from future credit access. The incentive effects of terminations strengthen the role of the reputation/feedback system in my analysis.[5]

In order to see whether reputational incentives impact borrowers' behavior and market outcomes, I begin my analysis by focusing on a group of borrowers who have two overlapping loans. I arrange the loans for each borrower based on the closing dates. From the regression analysis using this sample, I find that borrowers who default on the first-closed loans are more likely to default on the second-closed loans. There are two possible explanations for this finding: (1) once borrowers default on one loan, knowing that they are not allowed to enter again, they lose the reputational incentives to pay off the existing loans (i.e., incentive effects); and/or (2) borrowers who default on the first-closed loans are on average more risky, and therefore are more likely to default on the other loans (i.e., selection effects). If only the selection story matters, we would expect that the default rates of the first- and the second-closed loans are roughly the same, since the pool of borrowers remains unchanged across the two loans. However, I find that the second-closed loans have a significantly higher default rate compared to the first-closed loans when controlling for observables. This finding highlights the importance of the incentive story: borrowers respond to reputational incentives; when incentives are reduced, default rates significantly increase.

Given the empirical evidence, I develop a finite-horizon dynamic structural model to analyze borrowers' repayment decisions, lenders' funding strategies, and Prosper's pricing schemes under asymmetric information in online credit markets. In the model, borrowers are heterogeneous in default costs, which constitutes their private type. After observing interest rates assigned by the website, borrowers decide whether to withdraw the listings or not; if they stay in the market, borrowers choose optimal effort levels to exert. The outcomes of each loan are stochastically affected by effort levels. On the supply side, lenders make funding decisions after observing borrowers' participation. Lenders' payoffs from a

[4] Customer review systems are widely adopted by e-commerce platforms such as eBay and Amazon. Other peer-to-peer markets that use individual rating systems include Uber and Airbnb.

[5] See Stiglitz and Weiss (1983) for a detailed discussion on the effects of terminations in credit markets.

loan depend on the borrower's type and the effort exerted. Having no access to borrowers' private information, lenders make decisions based on their beliefs about the borrower's type.

While the existing literature on empirical contract models of credit markets mainly focuses on adverse selection in a static environment (see Einav, Jenkins, and Levin, 2012; Kawai, Onishi, and Uetake, 2018), my paper combines adverse selection and moral hazard in a dynamic setting.[6] Asymmetric information enters the model through two channels. First, borrowers' types are private information. Conditional on the same observables, borrowers are charged the same interest rates; as a result, "lemons" are more likely to be adversely selected into the market. Second, effort levels exerted by borrowers are also unobserved. Moral hazard occurs when borrowers tend to shirk because lenders are the ones who bear the risk of default. In a dynamic setting, borrowers' hidden actions create further complications. Lenders' beliefs about borrowers' types are formed through the Bayes updating process, which involves borrowers' effort-exerting decisions. In equilibrium, borrowers' strategies need to be consistent with lenders' beliefs.

In terms of identification, the key is to recover the distribution of the borrower's latent type, the effect of effort on loan outcomes, and utility primitives of borrowers and lenders. In this paper, I provide nonparametric identification of borrowers' default cost distribution following the literature that deals with unobserved heterogeneity in dynamic models (see Hu and Shum, 2012).[7] The intuition is that under certain Markov structure, the *intertemporal* correlation between future loan outcomes and borrowers' characteristics in proceeding loans is induced by the latent type, and thus provides identifying power for the type distribution of borrowers. Once the type distribution is recovered, conditional on the latent type, borrowers' unobserved effort choices generate further correlation between different dimensions of loan outcomes within a period. Leveraging this *intratemporal* variation, I separately identify the effects of effort on defaults and late payments using techniques developed in the measurement error literature (see Hu, 2008; Hu and Schennach, 2008). In the last step, I exploit functional form assumptions and variations in interest rates to identify utility primitives in borrowers' payoff functions and lenders' outside option distributions.

Directly following the identification results, I apply the likelihood-based estimation strategy to a large transaction-level dataset from Prosper.com. My estimation results indicate that borrowers who have high credit grades (assigned by Prosper based on their characteristics) and who use loans for debt consolidation are more likely to be "good borrowers." In

[6]Einav, Jenkins, and Levin (2012) and Kawai, Onishi, and Uetake (2018) model borrowers' dynamic repayment behavior within a loan period; my paper mainly focuses on dynamics across different loans.

[7]Other papers that study identification and estimation of dynamic structural models with serially correlated unobserved heterogeneity include Aguirregabiria and Mira (2007), Houde and Imai (2006), and Kasahara and Shimotsu (2009) etc.

addition, borrowers with higher default costs and smaller loan requests are more likely to draw smaller costs of effort. As for the state transition process of borrowers' characteristics, I find high debt-to-income ratios are persistent, while the transition process is type-specific—borrowers who have high default costs are more likely to stay with low debt-to-income ratios.

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

To address potential long-run inefficiencies due to accidental loss of reputation, the last part of my counterfactual analysis considers a case where borrowers are offered an option to buy Payment Protection Insurance (PPI). This insurance covers loan repayments for a set period of time if borrowers are unable to repay in certain situations.[8] The intuition of this mechanism is straightforward. If a borrower wants to maintain a good reputation (and hence credit access in the future), but also worries about future negative shocks, he or she can purchase this insurance to hedge against that risk. From the counterfactual experiment, I find that around 98 percent of the welfare loss from asymmetric information is recovered by introducing PPI into the market where the reputation system is implemented. This policy intervention has strong empirical relevance, especially for small businesses who find peer-to-peer lending an attractive financing alternative (see Segal, 2015) and who rely heavily on this form of credit access for their success and growth.

[8] These circumstances usually include being made redundant at one's job or not being able to work because of an accident or illness. For details of PPI, see https://www.fca.org.uk/consumers/income-payment-protection.

Related Literature. This paper is related to the literature that studies the value of reputation using structural models (Yoganarasimhan, 2013; Saeedi, 2014; Lewis and Zervas, 2016; Bai, 2016). It is, to the best of my knowledge, the first to quantify the extent to which reputation/feedback systems improve the total welfare of market participants when both adverse selection and moral hazard are present. My results confirm three important channels through which the welfare gain is achieved. This paper also relates to the literature that uses regressions to study the qualitative effect of reputation/review systems on e-commerce platforms (Melnik and Alm, 2002; Eaton, 2005; Jin and Kato, 2006; Lucking-Reiley et al., 2007; Cabral and Hortacsu, 2010; Fan, Ju, and Xiao, 2016) and online labor markets (Lin, Liu and Viswanathan, 2016).[9] I provide new empirical evidence that reputational incentives impact market outcomes and borrower behavior in online marketplaces for consumer loans.

There has been a long discussion of how to test the existence, and further disentangle the effects, of adverse selection and moral hazard in the empirical literature (Chiappori and Salanie, 2000; Chiappori and Salanie, 2002; Abbring et al., 2003; Chiappori et al., 2006). This paper provides quantitative results on the welfare loss from adverse selection and moral hazard separately, and I find that moral hazard plays an important role in credit markets.[10] The existing literature on empirical contract models of credit markets mainly focuses on revealing borrowers' private information through screening/signaling devices (Adams et al., 2009; Einav, Jenkins, and Levin, 2012; Einav, Jenkins, and Levin, 2013; Kawai, Onishi, and Uetake, 2018).

This paper is also related to the literature on identification of contract models. In particular, Perrigne and Vuong (2011) impose a "truth-telling" condition which generates a one-to-one mapping between agents' private information and observed prices. Kawai, Onishi, and Uetake (2018) rely on the fact that a borrower's type and signal have a one-to-one mapping in a separating equilibrium. Gayle and Miller (2015) study models of managerial compensation and assume that some levels of revenue can only be achieved through high effort. Different from the existing literature, the identification strategy in this paper takes advantage of the dynamic structure. I exploit variations in repeated borrowing and repayment patterns to recover the distribution of borrowers' persistent latent types, which does not require a one-to-one mapping between type and observables. In addition, I recover the effect of effort on loan outcomes without assuming that the support of revenue varies with effort. A related paper by Hu and Xin (2019) develops general identification strategies for dynamic models with unobserved choice variables, which can be easily applied to other types

[9]A related paper by Klein, Lambertz, and Stahl (2016) shows that an increase in market transparency in the feedback system on eBay leads to a reduction in moral hazard.

[10]A related paper by Bajari et al. (2014) finds adverse selection is an important source of inefficiency in medical insurance markets.

of contract models.

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

2 Institutional Background and Data Summary

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

2.1 Institutional Background

Prosper.com is one of the largest peer-to-peer lending markets in the United States. This website provides a platform for individual lenders and borrowers to meet without going through a complicated process as in traditional banking systems. Since its founding in 2005, the website has initiated more than nine billion dollars in loans and has attracted more than two million registered members.[11] On Prosper.com, borrowers list loan requests between $2,000 and $35,000 and individuals invest as little as $25 in each loan listing they select.[12] On average, each loan is funded by 43 individual investors, reflecting the crowdfunding feature of this market.[13] Prosper handles the servicing of the loan on behalf of the matched borrowers and investors; it makes a profit by charging both borrowers and lenders service fees proportional to the amount funded.[14]

The market works in the following manner. To post a listing online, a borrower needs to provide basic information about himself to the website, including his social security number, employment status, whether he is a homeowner, annual income, etc. Prosper hires a third-party credit report agency to verify the applicant's identity and credit history. The borrower's FICO score, total number of delinquencies, current number of credit lines, and so on are thereby revealed to the website. After the verification stage, the borrower is assigned a credit grade and can post a listing online, specifying the amount he requests and the purpose of the

[11] The amount of loan initiated is based on the data by the end of 2016.

[12] For details of the company, see https://www.prosper.com/plp/about/.

[13] The average number of investors for each loan is calculated using loans originated between January 2011 and December 2014.

[14] For details of the fee structure, see https://www.prosper.com/help/contextual/fees/.

loan.[15] Then the website decides on the interest rate for each listing posted. After seeing the interest rate, the borrower has the option to withdraw his listing before it is funded. Once the borrower decides to participate, it takes fourteen days for a listing to expire. Before the listing expires, lenders observe all posted information, including borrowers' detailed credit profiles and their loan and payment histories, and then decide whether or not to fund the loan.[16] As long as the amount requested is reached, the listing is successfully funded and the loan is originated. In the following 12-60 months, the borrower needs to pay back the loan, while it is possible that defaults and/or late payments occur in the repayment process. Note that if borrowers default, only lenders bear the loss in this market.

Compared to traditional lending markets, Prosper, as a representative of other peer-to-peer lending marketplaces, has the following distinctive features. From borrowers' perspectives, application requirements are easier to satisfy in online credit markets. As long as borrowers' basic information is verified by the website, borrowers are allowed to post a listing online. Thus, even borrowers with relatively low credit scores may obtain access to credit. For borrowers with good credit scores, they are charged lower interest rates by these markets due to lower operational costs.[17] In addition, it is more convenient and much faster for borrowers to take a loan online than a personal loan in a bank. This is particularly the case when the amount of the loan is relatively small. From lenders' perspectives, there is no collateral required on Prosper, which may indicate a higher level of risk. However, due to the crowdfunding feature of this market, it is convenient for lenders to diversify their investment portfolios so as to reduce idiosyncratic risks. Moreover, this website adopts a harsh punishment scheme to disincentivize default. That is, borrowers who have defaulted once are not allowed to borrow from the website again.

2.2 Data Summary

The data used in this paper consist of all listings (some of which become loans) that were originated on Prosper between January 2011 and December 2014. The final dataset used for empirical analysis and estimation contains 114,804 listings that come from 102,528 unique borrowers.[18] Overall, about 67 percent of loans are used for debt consolidation, 7 percent

[15]The borrower may also write a short paragraph about him/herself or about the description of the loan.

[16]An example of a listing is shown in Figure 2 in the online appendix Section C.

[17]For borrowers with credit scores higher than 600, 700, and 800, the average interest rates from Prosper are 16.92%, 13.70%, and 9.76%, respectively. These rates are lower than the credit card penalty APR, which is typically above 20%.

[18]The original dataset contains 192,916 listings. However, by the time I collected the dataset (November 8, 2016), there are still 63,790 ongoing loans that come from 62,841 unique borrowers. To ensure all loan outcomes are observed for each individual, I drop borrowers with ongoing loans. I also keep only one listing for each borrower within a short period (one month) to take care of the cases where borrowers may propose

for home improvement, 5 percent for business, and the rest for other purposes. 30 percent of borrowers have FICO scores below 600, and 94 percent of the borrowers are employed.[19] I characterize borrowers into different groups based on their repeated borrowing patterns. The percentage of borrowers in each category is summarized in Table 1. I find that 89 percent of the borrowers appeared only once during the sample period. Among those who appeared twice, the statuses of their first loans could be different when the second loans were originated. Specifically, their first loans may have been paid off, still ongoing, or were not funded (possibly withdrawn by themselves). In Table 1, it shows that conditional on the first loans being funded, about 60 percent of borrowers who appeared twice proposed their second listings when the first loans were still ongoing. The proportion of borrowers who appeared three times is small compared to that of other categories.

Table 1: Repeated Borrowing Pattern

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

Figure 1 compares the distributions of credit grades for borrowers' first and second listings. Prosper characterizes borrowers into seven credit groups, from AA (best) to HR (worst). From this figure, it is clear that a larger proportion of borrowers falls into better credit groups (include AA, A and B) in the second listings. Notice that borrowers I observe in the second listings must have not defaulted in their first loans. This selected group of borrowers may be inherently better, and thus have higher credit grades. Alternatively, the shift of credit grade distribution may be driven by the updating of borrowers' credit grades after the first loans' outcomes realize. For instance, after paying back their loans, borrowers are very likely to be characterized into better groups by the reputation system. The selection and updating channels jointly determine the empirical pattern in Figure 1. To gain a better understanding of how the reputation system improves the pool of borrowers over time, I use the structural estimates to disentangle these two channels in Section 5.

multiple listings for one monetary demand. I drop borrowers with missing information. In addition, I focus on borrowers who have at most three listings, since the proportion of borrowers who have more than three listings is less than 1 percent.

[19]The distribution of their stated monthly income is shown in Figure 3 in the online appendix Section C. Summary statistics of other variables used in the regressions or estimation are provided in Table 9.

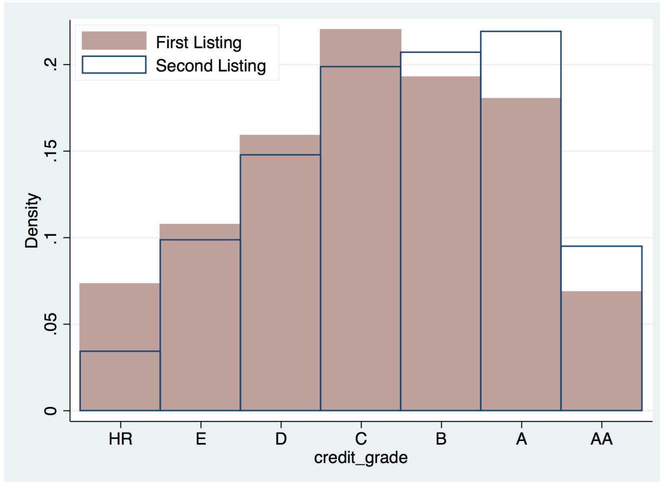

Figure 1: Histograms of Credit Grade for Borrowers' First and Second Listings

Table 2: Summary Statistics by Credit Grades

Credit Grades	AA	A	B	C	D	E	HR	All
Avg. Amt. Requested($)	13250.31	12974.22	12982.66	11813.24	9179.09	5106.38	3586.34	10662.36
Avg. Interest Rate (%)	7.53	10.99	14.72	18.58	23.57	28.2	31.53	18.34
Withdraw Prob. (%)	6.39	5.56	5.66	5.52	7.77	6.13	11.26	6.43
Funding Prob. (%)	89.42	91.11	91.37	92.09	87.7	92.29	71.82	89.51
Default Prob. (%)	6.37	14.15	21.53	28.55	31.56	33.93	34.19	24.23
Late Payment Prob. (%)	2.64	5.61	8.58	11.02	15.47	16.5	20.86	10.75
Number of Obs.	8,231	21,166	22,271	24,964	18,046	12,196	7,930	114,804

To better understand the differences between credit groups, Table 2 summarizes the average amount requested (in dollars), average interest rates, withdrawal and funding probabilities, and default and late payment rates for borrowers by different credit categories. There is a clear pattern for interest rates and default probabilities. Borrowers with lower credit grades are charged higher interest rates. Their default probabilities are also higher. This phenomenon intuitively captures the trade-off faced by lenders—higher risks must be compensated by higher returns.

2.3 Empirical Evidence on the Effect of Reputation

In this section, I provide empirical evidence on how market prices, borrowers' behavior, and loan outcomes are affected by the reputation system. I first investigate whether the interest rates charged on borrowers vary with past loan outcomes for a given individual. Table 3 presents regression results of interest rates 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.

Table 3: Regression Results of Interest Rate on Past Outcomes

VARIABLES	(1) borrower_rate	(2) borrower_rate	(3) borrower_rate
second_loan	-0.00902***		
	(0.000336)		
overlap		0.00208***	0.00157***
		(0.000409)	(0.000415)
late_ever			0.00452***
			(0.000663)
Constant	0.317***	0.303***	0.303***
	(0.000986)	(0.00156)	(0.00156)
Observations	16,820	8,410	8,410
R-squared	0.932	0.937	0.937

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

Columns 2–3 further illustrate that the second loans which overlap with first loans or have prior late payments are associated with higher interest rates. These results indicate that the interest rate as a pricing device appears to "reward" prior successful loan repayment behavior and "punish" late payments or uncertainty from overlapping loans.[20] Combining this finding with the institutional feature that once a borrower defaults, he is unable to borrow again, we can see that the reputation/feedback system imposes dynamic incentives on forward-looking borrowers through history-dependent pricing schemes and entry restrictions.

To further examine the causal impact of the reputation/feedback system on borrowers' behavior and market outcomes, in this section I consider a group of borrowers who have two overlapping loans.[21] I arrange the two loans for each borrower based on their closing dates. From the regression analysis using this sample, I find that defaults for the first- and the second-closed loans are positively correlated, i.e., conditional on defaulting on the first loan, borrowers are more likely to default on the second-closed loans.[22] There are two possible explanations for this finding: (1) once borrowers default on one loan, knowing that they are not allowed to enter again, they lose the reputational incentives to pay off the existing loans; and/or (2) borrowers who default on the first-closed loans are on average more risky, and therefore are more likely to default on the other loans. If borrowers do not react to

[20]To alleviate the concern that borrowers may first request and pay off a small loan in order to get better deals for later larger loans, I run a regression of amount requested on whether a loan is the second one, controlling for a set of observables. The result in Table 1 in the online appendix D illustrates that borrowers' second loans do not constitute significantly larger requests.

[21]On Prosper, borrowers could propose their second listings even when their first loans are still in process. See Table 1 for summary statistics.

[22]The positive correlation between the defaults in two loans is shown in Table 10 in the Appendix.

Table 4: Logit Regression of Default on Whether the Loan is the Second-closed Loan

VARIABLES	(1) default	(2) default	(3) default	(4) default
second_closed_loan	0.390***	0.359***	0.481***	0.635**
	(0.0480)	(0.0731)	(0.0863)	(0.322)
borrower_rate	13.10***	12.66***	16.09***	18.98***
	(1.347)	(1.979)	(2.238)	(5.757)
Constant	-4.572***	-4.659***	-5.960***	-7.502***
	(0.414)	(0.608)	(0.706)	(1.785)
Second Loan Not for Debt Consolidation		Y	Y	Y
FICO Below 600			Y	Y
Long Gap Between Closing Dates				Y
Observations	10,166	4,630	3,550	1196

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

reputational incentives (or are myopic), given the same group of borrowers, we would expect the default rates for the first- and the second-closed loans roughly the same. However, if borrowers react to dynamic incentives (or are forward-looking), we would expect a higher default rate for the second-closed loans.

Column 1 in Table 4 shows that the second-closed loans have a significantly higher default rate after controlling for observables. Since I compare the average default rates on two loans for the same group of borrowers, the higher default rate for the second-closed loans is not driven by the selection of borrowers. This result suggests that borrowers are likely to respond to reputational incentives.[23] As a robustness check, Column 2 restricts the sample to borrowers whose second loans are not for debt consolidation so as to alleviate concerns that individuals borrow again to refinance their first loans. 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. For these borrowers, the higher default rate of the second-closed loans is less likely to be driven by changing outside opportunities. Finally, Column 4 focuses on the group of borrowers with long gaps (one year) between the closing dates of their two loans. For this subsample, the issue that defaults on the two loans may be induced by two correlated negative income shocks received by the borrower within a short period is mitigated to some extent. The results using different subsamples do not change much. To summarize, this model-free evidence suggests

[23] Another explanation for the higher default rate of the second-closed loans is that borrowers are more likely to leave the market after the second loans, so they have fewer incentives to repay the second loans. This also suggests that borrowers are responsive to dynamic incentives.

that borrowers are responsive to the dynamic incentives imposed by the reputation/feedback system. When the incentives to repay debts are reduced, default rates significantly increase.

3 Model

In this section, I develop a finite-horizon dynamic model of a credit market in which borrowers and lenders interact repeatedly over time. Let $t = 1, \cdots, T < \infty$ be the index of each period. On the demand side, forward-looking borrowers at each period t seek funding opportunities for their listings, and the interest rates are assigned by the market intermediator—the website. On the supply side, I assume that the market has a sequence of short-lived lenders who invest in borrowers' listings as long as the expected payoffs from these listings exceed their outside investment options.

I first describe a game played by a single borrower and a lender, taking the website's pricing scheme as given, to illustrate how the credit market works. The optimal strategies of borrowers and lenders are analyzed at the end of this section.

3.1 A Model of a Credit Market

Players. In this game, there are two players: a borrower (agent) and a lender (principal). Before the game starts, the borrower privately observes his default cost $c \in \Theta_c = \{c_1, c_2, \cdots, c_J\}$. The distribution of the default cost is characterized by $F_c(\cdot) : \Theta_c \to [0, 1]$, which is common knowledge. The borrower's utility associated with a given loan is $U(\cdot; \alpha)$, with α as the risk-aversion parameter. Assume $U'(\cdot; \alpha) > 0$ and $U''(\cdot; \alpha) < 0$. The borrower discounts future at the rate of δ. The lender is assumed to be risk-neutral and to live for only one period.

Timing and Payoffs. The timing of the game is illustrated in Figure 2. Specifically, in each period t,

- The borrower proposes a listing, draws an outside option $v_{0,t} \in \Theta_v$ according to $F_v(\cdot)$, and a cost of effort $\theta_t \in \Theta_\theta$ according to $F_{\theta|c}(\cdot|c)$, which are only observed by himself.
- An interest rate $r_t \in [0, \bar{r}]$ is assigned by the website, which is publicly observed only for players at period t.
- The borrower decides whether to withdraw his listing ($W_t = 1$) or not ($W_t = 0$). If he withdraws, the borrower gets his outside option $v_{0,t}$.
- If the borrower stays, the lender draws an outside investment option $\mu_{0,t} \in \Theta_\mu$ according to $F_\mu(\cdot)$ privately and decides whether to invest ($I_t = 1$) or not ($I_t = 0$) in

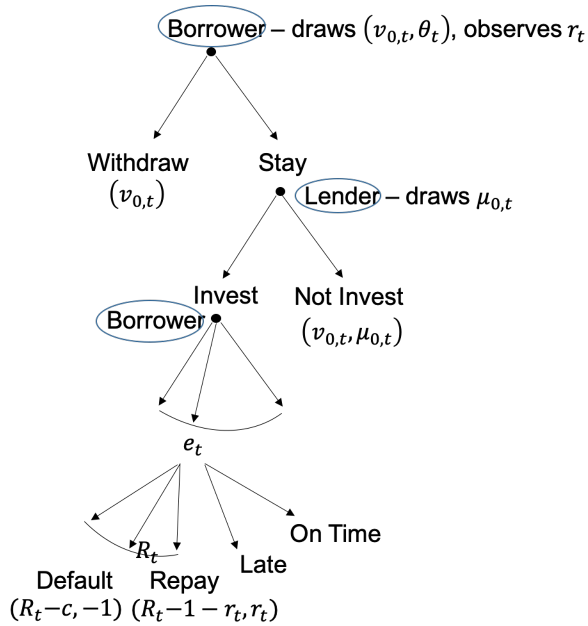

Figure 2: Timing of the Game for Period t.

the listing. If the listing is not funded, the borrower and the lender get their outside options $v_{0,t}$ and $\mu_{0,t}$, respectively.

– If funded, the borrower exerts an effort $e_t \in [0, \infty]$, which is only observed by himself.

– A revenue $R_t \in [0, \bar{R}]$, scaled by the size of the loan, is drawn from $F_{R|e}(\cdot|e_t)$. R_t is privately observed by the borrower. A late payment occurs ($L_t = 1$) with probability $P_L(e_t)$ and L_t is publicly observed.

– If $R_t > 1 + r_t$, the borrower repays ($D_t = 0$) with a monetary payoff $R_t - 1 - r_t$, and the lender receives r_t; otherwise, the borrower defaults ($D_t = 1$) with the payoff $R_t - c$, and the lender receives -1. D_t is publicly observed.

– If $D_t = 0$, the game proceeds to $t+1$, and a new short-lived lender arrives. If $D_t = 1$, in all future periods, the borrower has no access to this market.

When exerting effort e_t, the borrower incurs cost $\phi(e_t, \theta_t)$, which is assumed to be additively separable from the utility the borrower obtains from the monetary payoff of each loan. Assume that $\phi'(\cdot, \theta_t) > 0$ and $\phi''(\cdot, \theta_t) > 0$. The borrower's effort e_t affects the revenue distribution in a stochastic way. Specifically, assume that for any $e^h > e^l$, $F_{R|e}(\cdot|e^h)$ first order stochastically dominates $F_{R|e}(\cdot|e^l)$. The effort also has an impact on the realization of late payments. Assume that for any $e^h > e^l$, $P_L(e^h) < P_L(e^l)$, i.e., the borrower is more likely to make a late payment if e^l is exerted. $F_{R|e}(\cdot|e_t)$ and $P_L(e_t)$ for any given e_t are common knowledge among players.

Histories and Strategies. For $t \geq 2$, the public histories in this game are defined as

$$H^t = (W_1, I_1, L_1, D_1, \cdots, W_{t-1}, I_{t-1}, L_{t-1}, D_{t-1}),$$

where when the borrower withdraws his listing (i.e., $W_t = 1$), $I_t = \emptyset$, $L_t = \emptyset$, $D_t = \emptyset$; when the listing is not funded (i.e., $I_t = 0$), $L_t = \emptyset$, $D_t = \emptyset$. For $t = 1$, $H^t = \emptyset$. Note that the interest rate r_t is not included in the public histories. When a new short-lived lender arrives, he only observes the current period interest rate and the outcomes of the previous loans.[24] Let $\mathcal{H}^t$ denote the set of all period t public histories. In addition to the public histories, the lender observes his outside investment option at each period t. Conditional on observing the borrower staying in the market at t, the lender's investment strategy is defined as $\sigma : \mathcal{H}^t \times [0, \bar{r}] \times \Theta_\mu \to \{0, 1\}$.

A history for the borrower consists of the public history, the interest rates, and his private information. For $t \geq 2$, define the borrower's history as

$$H_B^t = (H^t, r_1, v_{0,1}, \theta_1, e_1, R_1, \cdots, r_{t-1}, v_{0,t-1}, \theta_{t-1}, e_{t-1}, R_{t-1}).$$

Let $\mathcal{H}_B^t(H^t)$ denote the set of all period t private histories for the borrower that are consistent with H^t. In addition to his history, the borrower observes his own type (default cost). In this paper, I focus on borrowers' strategies that depend on public histories. The withdrawal strategy for the borrower with default cost c is defined as $w_{c,t} : \mathcal{H}^t \times [0, \bar{r}] \times \Theta_\theta \times \Theta_v \to \{0, 1\}$. Conditional on staying in the market, the borrower's effort choice strategy is $e_{c,t} : \mathcal{H}^t \times [0, \bar{r}] \times \Theta_\theta \to [0, \infty]$. A belief for the lender at the beginning of period t with history H^t is a distribution $\xi(H^t) \in \Delta(\Theta_c \times \mathcal{H}_B^t(H^t))$.

Equilibrium. Define $w_c = (w_{c,1}, w_{c,2}, \cdots, w_{c,T})$ and $e_c = (e_{c,1}, e_{c,2}, \cdots, e_{c,T})$. A Perfect Bayesian Equilibrium is a tuple $< \sigma^*, (w_c^*, e_c^*)_{c \in \Theta_c}, \xi^* >$ such that the following conditions hold for any H^t:

(1) $\sigma^*(H^t, r_t, \mu_{0,t})$ is sequentially rational given $\xi^*(H^t)$ and $(w_c^*, e_c^*)_{c \in \Theta_c}$.

(2) $w_{c,t}^*(H^t, r_t, \theta_t, v_{0,t})$ and $e_{c,t}^*(H^t, r_t, \theta_t)$ are sequentially rational given σ^*.

(3) $\xi^*(H^t)$ is derived from strategies via Bayes's rule whenever possible.

In the rest of this paper, I focus on the equilibrium where lenders' strategies given public histories $\hat{H}^t$ and $\tilde{H}^t$ are identical whenever the beliefs induced by the histories $\xi(\hat{H}^t)$ and

[24] This setting is more realistic as on the website, the outcomes of previous loans are shown on the listing page, but the previous interest rates are not directly displayed.

$\xi(\tilde{H}^t)$ are the same.[25] I discuss the conditions under which the existence and uniqueness of the equilibrium are achieved in Section E of the online appendix. Optimal strategies for borrowers and lenders are derived and analyzed in Sections 3.2 and 3.3.

3.2 Borrowers' Withdrawal Decisions and Effort Choices

Proceeding Loans. Let $V_{c,t}(H^t, r_t, \theta_t, v_{0,t})$ denote the value function of a borrower with type c at the beginning of period t. If the listing is funded, I use $\tilde{V}_{c,t}(e, H^t, r_t, \theta_t)$ to represent the expected payoff for the borrower given (e, H^t, r_t, θ_t), where e is any effort level in $[0, \infty)$.

$$
\begin{aligned}
&\tilde{V}_{c,t}(e, H^t, r_t, \theta_t) \\
&= \int_{1+r_t}^{\bar{R}} \Bigg[U(R - 1 - r_t) \\
&+ \delta \sum_{l \in \{0,1\}} Pr(L_t = l|e) \, \mathrm{E} \, V_{c,t+1}(H^{t+1}(L_t = l, D_t = 0), r_{t+1}, \theta_{t+1}, v_{0,t+1}) \Bigg] dF_{R|e}(R|e) \\
&+ \int_{0}^{1+r_t} \left[U(R - c) + \sum_{\tau=t+1}^{T} \delta^{\tau - t} \, \mathrm{E}(v_{0,\tau}) \right] dF_{R|e}(R|e) - \phi(e, \theta_t).
\end{aligned}
\tag{3.1}
$$

Note that $H^{t+1}(L_t = l, D_t = 0)$ denotes the public history at $t+1$ consistent with $L_t = l, D_t = 0$.[26] If the borrower defaults, in all future periods, he receives a random draw from the outside option distribution. Once the listing is funded, the borrower chooses his optimal effort level through the following payoff maximization problem.

$$
e^*_{c,t}(H^t, r_t, \theta_t) = \arg \max_{e \in [0, \infty)} \tilde{V}_{c,t}(e, H^t, r_t, \theta_t). \tag{3.2}
$$

Before exerting effort, the borrower decides whether to withdraw the listing or not. I define the borrower's choice-specific value functions given $W_t = 0$ and $W_t = 1$ as follows.

$$
\begin{aligned}
\bar{V}^0_{c,t}(H^t, r_t, \theta_t, v_{0,t}) = {} & P(H^t, r_t) \max_{e \in [0, \infty)} \tilde{V}_{c,t}(e, H^t, r_t, \theta_t) \qquad (3.3) \\
& + (1 - P(H^t, r_t)) \left(v_{0,t} + \delta \, \mathrm{E} \, V_{c,t+1}(H^{t+1}(I_t = 0), r_{t+1}, \theta_{t+1}, v_{0,t+1}) \right). \\
\bar{V}^1_{c,t}(H^t, v_{0,t}) = {} & v_{0,t} + \delta \, \mathrm{E} \, V_{c,t+1}(H^{t+1}(W_t = 1), r_{t+1}, \theta_{t+1}, v_{0,t+1}). \qquad (3.4)
\end{aligned}
$$

[25] This restriction on lenders' strategies also implies that borrowers' strategies given public histories $\hat{H}^t$ and $\tilde{H}^t$ are identical whenever the beliefs induced by the histories $\xi(\hat{H}^t)$ and $\xi(\tilde{H}^t)$ are the same.

[26] In particular, $H^{t+1}(L_t = l, D_t = 0) = (H^t, W_t = 0, I_t = 1, L_t = l, D_t = 0)$.

Note that $H^{t+1}(W_t = 1)$ and $H^{t+1}(I_t = 0)$ denote the public histories at $t+1$ consistent with $W_t = 1$ and $I_t = 0$, respectively.[27] $P(H^t, r_t)$ represents the probability that a listing is funded when borrowers stay in the market, which depends on the lender's optimal investment strategy σ^*. As shown in Equation (3.3), when the borrower chooses to stay in the market, there are two possible outcomes: (1) if the listing is funded, the borrower exerts optimal effort; (2) if not funded, the borrower obtains his outside option and proceeds to the next period with a history of not being funded. When the borrower withdraws his listing, he gets the outside option immediately and the future continuation value depends on listing withdrawal history as shown in Equation (3.4). The optimal withdrawal strategy for the borrower is therefore

$$w^*_{c,t}(H_t, r_t, \theta_t, v_{0,t}) = \begin{cases} 0 & \text{if } \bar{V}^0_{c,t}(H^t, r_t, \theta_t, v_{0,t}) \geq \bar{V}^1_{c,t}(H^t, v_{0,t}) \\ 1 & \text{otherwise} \end{cases}. \tag{3.5}$$

The value function of the borrower at the beginning of period t is

$$V_{c,t}(H^t, r_t, \theta_t, v_{0,t}) = \max\left\{\bar{V}^0_{c,t}(H^t, r_t, \theta_t, v_{0,t}), \bar{V}^1_{c,t}(H^t, v_{0,t})\right\}. \tag{3.6}$$

The Last Loan. When $t = T$, the future continuation value for the borrower is 0. Following Equation (3.1), the borrower's expected payoff for an effort e is

$$\tilde{V}_{c,T}(e, r_T, \theta_T) = \int_{1+r_t}^{\bar{R}} U(R - 1 - r_t) dF_{R|e}(R|e) + \int_0^{1+r_t} U(R - c) dF_{R|e}(R|e) - \phi(e, \theta_t), \tag{3.7}$$

and the optimal effort choice in the last period is

$$e^*_{c,T}(r_T, \theta_T) = \arg\max_{e \in [0,\infty)} \tilde{V}_{c,T}(e, r_T, \theta_t). \tag{3.8}$$

Remark 1. *Note from Equation (3.8) that a borrower's optimal effort-exerting strategy in the final period does not depend on past histories, as histories affect borrowers' payoffs only through future continuation values. Following similar ideas, a borrower's optimal withdrawal strategy in the last period can be easily derived.*

3.3 Lenders' Investment Decisions

At the beginning of period t, a belief for the lender with history H^t is $\xi^*(H^t)$. Lenders make their investment decisions by comparing the expected revenues of loans with their outside

[27] Specifically, $H^{t+1}(W_t = 1) = (H^t, W_t = 1, I_t = \emptyset, L_t = \emptyset, D_t = \emptyset)$, and $H^{t+1}(I_t = 0) = (H^t, W_t = 0, I_t = 0, L_t = \emptyset, D_t = \emptyset)$.

options. In particular, a lender decides whether to fund the listing or not after observing the borrower staying in the market. As a result, the lender's belief about the borrower's type is updated. I denote the updated belief as $\xi^*(H^t, W_t = 0)$.

Given a public history H^t and an interest rate r_t, the lender's revenue from a borrower with (θ_t, c) is derived in the following equation.

$$\tilde{\pi}(H^t, r_t, \theta_t, c) = \int_{1+r_t}^{\bar{R}} (1 + r_t) dF_{R|e}(R|e^*_{c,t}(H^t, r_t, \theta_t)) - 1, \tag{3.9}$$

where $e^*_{c,t}(H_t, r_t, \theta_t)$ represents the borrower's optimal effort-exerting strategy. Let $\pi(H^t, r_t)$ denote the lender's expected revenue after integrating over (θ_t, c) under the belief $\xi^*(H^t, W_t = 0)$. The lender's optimal investment strategy is therefore

$$\sigma^*(H^t, r_t, \mu_{0,t}) = \begin{cases} 1 & \text{if } \pi(H^t, r_t) \geq \mu_{0,t} \\ 0 & \text{otherwise} \end{cases}, \tag{3.10}$$

and the probability that a listing is funded equals $F_\mu(\pi(H^t, r_t))$.

4 Identification

4.1 Data and Primitives

Data. In the dataset, I observe a group of borrowers indexed by $i = 1, 2, \cdots, N$ for $t = 1, 2, \cdots, T$. To simplify the notations, I drop subscript i in the rest of this section. Denote the vector of observed outcomes of a loan at t by $O_t = (W_t, I_t, L_t, D_t)$ and the public histories at the beginning of t by $H^t = (O_1, \cdots, O_{t-1})$.[28] In addition to the public history, at each period, the interest rate r_t and borrowers' characteristics are also available to lenders who arrive at t. Let $X_t \in \Theta_X$ and $K_t \in \Theta_K$ represent a borrower's financial status and credit grade, respectively. The following assumption is invoked on the transition of the borrower's observed characteristics.

Assumption 1. *(i) The distribution of X_{t+1} depends on (X_t, c) and is denoted by $F_{X_{t+1}|X_t,c}$; $X_{t+1}|X_t, c$ is independent of other variables. (ii) The distribution of K_{t+1} depends on (K_t, O_t) and is denoted by $F_{K_{t+1}|K_t,O_t}$; $K_{t+1}|K_t, O_t$ is independent of other variables.*

Assumption 1 implies that the transition of borrowers' observed characteristics has a first-order Markov property. In addition to that, the transition of a borrower's financial status X_t is related to his private type c, while the updating of his credit grades on this

[28] The definitions for (W_t, I_t, L_t, D_t) remain the same as in Section 3. $W_t = 1$ denotes that the borrower withdraws the listing, and 0 otherwise. $I_t = 1$ represents the case where the loan is funded, and 0 otherwise. L_t and D_t are indicators for late payment and default performances, respectively.

website K_t is only based on realized loan outcomes in the last period. To summarize, for each borrower i in the dataset, I observe $\{O_t, r_t, X_t, K_t\}_{t=1}^T$.

Primitives. On the borrower's side, the parameters of interests include utility function, $U(\cdot;\alpha)$, the cost function, $\phi(\cdot,\theta_t)$, the joint distribution of default cost and cost of effort, $F_{c,\theta}$, the distribution of outside option, F_v, the distribution of revenue given effort, $F_{R|e}$, the probability that a late payment occurs given effort, $P_L(\cdot)$, and the type-specific transition rule of borrower's financial status, $F_{X_{t+1}|X_t,c}$.[29] On the lender's side, the distribution of outside investment option, F_μ, needs to be identified. The website adopts a pricing rule $g : \Theta_X \times \Theta_K \times \mathcal{H}^t \to \mathbb{R}$, which maps borrowers' characteristics and public histories to an interest rate. Assume that the observed interest rates are based on the pricing rule subject to idiosyncratic errors, i.e., $r_t = g(X_t, K_t, H^t) + \varepsilon_t$. The pricing rule can be nonparametrically identified from the data.

In the rest of this section, I first provide nonparametric identification results for $F_{c,\theta}$ and $F_{X_{t+1}|X_t,c}$. Specifically, a case where borrowers have discrete types $c \in \Theta_c = \{c_1, c_2, \cdots, c_J\}$, and the cost of effort θ_t takes value from $\Theta_\theta = \{\theta_1, \theta_2, \cdots, \theta_S\}$ is considered.[30] For other primitives, I impose the following parametric assumptions. I assume that the borrower has a CARA utility function, $U(x;\alpha) = 1 - \exp(-\alpha x)$, where α is the risk-aversion parameter. I specify the borrower's cost function given an effort e as $\phi(e,\theta_t) = \theta_t e^2$. For each loan, assume that there are two possible revenues that may realize, R_h and R_l. The probability that R_h is realized given an effort e is specified as $P_R(e) = 1 - \exp(-\beta e)$, and the probability that a late payment occurs takes the form $P_L(e) = \exp(-\gamma e)$. β and γ measure the "effectiveness" of the effort level on preventing defaults and late payments. For the outside option distributions, let $v_{0,t}$ follow a normal distribution with mean equal to $v(X_t, K_t) = v_0 + v_x X_t + v_k K_t$ and variance equal to σ_v; lenders' outside options are drawn from $N(\mu_0, \sigma_\mu)$. To summarize, the primitives to be identified include $\{F_{c,\theta}, F_{X_{t+1}|X_t,c}, \alpha, R_h, R_l, \beta, \gamma, v_0, v_x, v_k, \sigma_v, \mu_0, \sigma_\mu\}$.

4.2 Nonparametric Identification of the Type Distribution

In this section, I first show the nonparametric identification of the borrower's default cost distribution. The intuition is as follows. The loan outcome O_t at each period depends on the borrower's withdrawal and effort exerting decisions, as well as the lender's investment strategy. All of these strategies are based on the public history H^t, the interest rate r_t, and the borrower's observed characteristics (X_t, K_t). If the latent type does not exist, $O_t|H^t, r_t, X_t, K_t$ should be independent of the observed characteristics in previous periods,

[29]The transition probabilities for credit grades K_t only involve observables, thus can be recovered directly from the data.

[30]The nonparametric identification results can be generalized for the cases where c and θ_t are continuous. However, for illustration purposes, in this paper, I stick to discrete types.

such as X_{t-1} and K_{t-1}. In other words, the *intertemporal* correlation between the loan outcomes and the lagged characteristics is induced by the latent type, and thus provides identifying power for the distribution of the borrower's unobserved type. For identification purposes, hereafter I make the following assumption.

Assumption 2. *(i) $O_t|r_t, X_t, K_t, H^t, \theta_t, c$ is independent of other variables. (ii) $\varepsilon_t|X_t, K_t$ is independent of other variables. (iii) $\theta_t|c$ is independent of other variables.*

Assumption 2 has several implications. First, given the set of variables $(r_t, X_t, K_t, H^t, \theta_t, c)$ that are associated with borrowers and lenders' decisions, loan outcomes are randomly drawn. Second, the website does not have additional information that is correlated with borrowers' hidden type when assigning interest rates. Assumption 2(iii) implies that the correlation between θ_t and θ_{t-1} is caused by unobserved heterogeneity rather than state dependence.[31] Under Assumptions 1 and 2, the joint distribution of $(O_t, r_t, X_t, K_t, H^t, X_{t-1}, K_{t-1})$ is decomposed in the following equation.[32]

$$
\begin{aligned}
& f_{O_t,r_t,X_t,K_t,H^t,X_{t-1},K_{t-1}} \\
&= \sum_c \sum_{\theta_t} f_{O_t|r_t,X_t,K_t,H^t,\theta_t,c} \cdot f_{r_t|X_t,K_t,H^t} \cdot f_{X_t|X_{t-1},c} \cdot f_{K_t|K_{t-1},O_{t-1}} \cdot f_{\theta_t|c} \cdot f_{H^t,X_{t-1},K_{t-1},c} \\
&= \left(\sum_c f_{O_t|r_t,X_t,K_t,H^t,c} \cdot f_{X_t|X_{t-1},c} \cdot f_{H^t,X_{t-1},K_{t-1},c} \right) \cdot f_{r_t|X_t,K_t,H^t} \cdot f_{K_t|K_{t-1},O_{t-1}}.
\end{aligned}
\tag{4.1}
$$

In Equation (4.1), $f_{O_t,r_t,X_t,K_t,H^t,X_{t-1},K_{t-1}}$, $f_{r_t|X_t,K_t,H^t}$, and $f_{K_t|K_{t-1},O_{t-1}}$ are only related to observables, thus can be nonparametrically identified from the data. However, the three densities that are associated with the borrower's private type c need to be recovered separately. Following Hu and Shum (2012), I use $(O_t, X_t, X_{t-1}, K_{t-1})$ as "measurements" for the latent type c and construct eigenvalue-eigenvector decompositions based on the matrix form of Equation (4.1). The details of constructing the matrix decompositions are provided in Appendix A.1. I make the following assumption to guarantee the existence and uniqueness of the eigenvalue-eigenvector decomposition for each (r_t, K_t, H^t).

Assumption 3. *For all values of $(r_t, K_t, H^t, X_t, X_{t-1})$, the following conditions are satisfied: (i) $F_{O_t|r_t,X_t,K_t,H^t,c=c^l}$ first order stochastically dominates $F_{O_t|r_t,X_t,K_t,H^t,c=c^h}$, $\forall c^h > c^l$; (ii) The distribution of X_t given (X_{t-1}, c) has a positive density $\forall c \in \Theta_c$; (iii) $F_{c|H^t,X_{t-1},K_{t-1}=K^h}$ first order stochastically dominates $F_{c|H^t,X_{t-1},K_{t-1}=K^l}$, $\forall K^h > K^l$.*

[31] Assumption 2(iii) can be relaxed to allow that the distribution of θ depends on some observables. In the estimation, it is related to the loan amount.

[32] In the rest of this paper, I drop the realization of the random variables in the joint pdf's to simplify the notations. Specifically, for any vector of random variables $Y = (Y_1, Y_2)$, let f_{Y_1,Y_2} and $f_{Y_1|Y_2=y_2}$ denote the joint density of Y and the density of Y_1 conditional on $Y_2 = y_2$, respectively.

Intuitively, Assumption 3(i) implies that borrowers with lower default costs are more likely to produce adverse loan outcomes (i.e., defaults and late payments). It is required by Assumption 3(ii) that the probability density of each financial status is nonzero. Assumption 3(iii) is satisfied when borrowers' credit grades are informative predictors for their true types.[33] Overall, Assumption 3 requires that the underlying latent type c plays an important role in generating the intertemporal correlation between future loan outcomes and borrowers' characteristics in the previous loans.[34]

Theorem 1 (Identification)**.** *If Assumptions 1–3 are satisfied, densities* $f_{O_t|r_t,X_t,K_t,H^t,c}$, $f_{X_t|X_{t-1},c}$, *and* $f_{H^t,X_{t-1},K_{t-1},c}$ *are identified for any* $(O_t, r_t, X_t, K_t, H^t, X_{t-1}, K_{t-1}, c)$.

The formal proof of Theorem 1 follows Hu (2008), and hence is skipped in this paper. When c is continuous, see Hu and Schennach (2008).

Remark 2. *Notice that Theorem 1 requires borrowers at least have two loans. The identified density* $f_{H^t,X_{t-1},K_{t-1},c}$ *represents the type distribution for a selected group of borrowers who share the history* H^t *in which no defaults have ever occurred in their previous loans. To identify the original type distribution, I consider the first loans for all borrowers in the sample. As shown in the following equation, the probability that borrowers pay off their first loans conditional on observed characteristics, which can be estimated directly from the data, is a finite mixture of the component distributions given different type* c *with* $f_{c|X_1,K_1}$ *serving as the mixing weight.*

$$f_{D_1=0,I_1=1,W_1=0|X_1,K_1} = \sum_c f_{D_1=0,I_1=1,W_1=0|X_1,K_1,c} \cdot f_{c|X_1,K_1}. \tag{4.2}$$

If all structural utility primitives are known, $f_{D_1=0,I_1=1,W_1=0|X_1,K_1,c}$ *can be easily computed using the model—then the only unknowns in Equation (4.2) are* $f_{c|X_1,K_1}$*. In Sections 4.3 and 4.4, I show that utility parameters, which are the key ingredients for identifying the original type distribution, are identified from borrowers who have multiple loan listings. The implicit assumption made here is that for borrowers who have different histories, although the type distribution may differ, the utility primitives are the same.*

4.3 Identification of the Effects of Effort on Loan Outcomes

In this section, I focus on identifying the effects of effort on loan outcomes. The intuition is as follows. Once the type distribution is recovered in Section 4.2, conditional on the

[33]That is, when a borrower has a high credit grade, he is more likely to have a high default cost compared to the case in which he has a low credit grade.

[34]Assumption 3 imposes restrictions on the distributions related to the latent type c, thus cannot be directly tested by the data. However, this assumption is motivated by the economic intuition of the market and is consistent with the setup of the theoretical model in Section 3.

latent type, the unobserved effort generates further correlation between different dimensions of outcomes within a loan period. Using this *intratemporal* variation, I separately identify the effects of effort on defaults and late payments. The following assumption is imposed to simplify the process in which loan outcomes are realized.

Assumption 4. *Let $e_t = e_t^*(r_t, X_t, K_t, H^t, \theta_t, c)$ denote the optimal effort level, the following conditions are satisfied: (i) $D_t \perp L_t | e_t$; (ii) $\mu_{0,t}$ is independent of other variables; (iii) $v_{0,t}|X_t, K_t$ is independent of other variables.*

Assumption 4 imposes conditional independence on different dimensions of loan outcomes. Specifically, Assumption 4(i) implies that conditional on the optimal effort levels chosen by the borrower, the realizations of defaults and late payments are driven by independent shocks.[35] Assumption 4(ii)–(iii) are supported by the model—the outside options of lenders and borrowers are randomly drawn.[36] Under Assumption 4, the following equation is obtained when $t = T$.[37]

$$\frac{f_{D_T,L_T,I_T=1,W_T=0|r_T,X_T,K_T,H^T,c}}{f_{I_T=1|r_T,X_T,K_T,H^T}} = \sum_{\theta_t} f_{D_T|r_T,\theta_T,c} \cdot f_{L_T|r_T,\theta_T,c} \cdot f_{W_T=0,\theta_T|r_T,X_T,K_T,c}. \tag{4.3}$$

In Equation (4.3), defaults, late payments, and participation probabilities all relate to borrowers' latent effort choices. Identifying the effects of effort levels on loan outcomes leverages the intratemporal correlations between different dimensions of loan outcomes. Equation (4.3) again leads to an eigenvalue-eigenvector decomposition under certain rank conditions, and therefore $f_{D_T|r_T,\theta_T,c}$, $f_{L_T|r_T,\theta_T,c}$, and $f_{W_T=0,\theta_T|r_T,X_T,K_T,c}$ are identified. I provide the details in Appendix A.2.

4.4 Identification of Utility Primitives

Last, I discuss the identification of utility primitives from the densities $f_{D_T|r_T,\theta_T,c}$, $f_{L_T|r_T,\theta_T,c}$, and $f_{W_T=0,\theta_T|r_T,X_T,K_T,c}$ recovered in Section 4.3. Under Assumption 2,

$$f_{W_T=0,\theta_T|r_T,X_T,K_T,c} = f_{W_T=0|r_T,X_T,K_T,\theta_T,c} \cdot f_{\theta_T|c}. \tag{4.4}$$

[35] This assumption could be relaxed to allow the case where defaults and late payments are independent conditional on effort and some observables. For illustration of the key identifying restrictions, I stick to Assumption 4 in this section.

[36] It is reasonable to assume that lenders' outside options are independent of observables such as interest rates, histories, and borrowers' characteristics. However, borrowers' outside options may be related to his own characteristics. This is allowed in Assumption 4(iii).

[37] Borrowers' effort choices in the last period do not depend on their histories or observed characteristics, as these will enter borrowers' optimization problems only through future continuation values. See Remark 1 for more details.

It is clear from Equation (4.4) that using variations in X_T and K_T, we are only able to identify the ratio between borrower's participation probabilities given different observed characteristics. Therefore, in addition to the standard location and scale normalizations on borrowers' outside option distribution, I further normalize v_k to identify the borrower's value from participation, i.e., $\tilde{V}_{c,T}(e_T, r_T, \theta_T)$. With variations in X_T, v_x is identified, which leads to the identification of $f_{\theta_T|c}$.

Borrower's value from participation is associated with the risk-aversion parameter α, revenues R_h and R_l, as well as the levels of the private information (θ_T, c). In Appendix B.1, I show that variations in interest rates identify the risk-aversion parameter and the high level of revenue. I then normalize γ to recover the effort levels from the probabilities that late payments occur, which further lead to the identification of β in the default realization process. In Appendix B.2, I show that once the identified effort levels are plugged into the borrower's payoff function, there exists a sufficient number of restrictions to pin down the levels of default cost and transitory cost shocks of exerting effort. Once borrowers' primitives are identified, the mean and variances of lenders' outside option distribution can be recovered from the funding probabilities.

5 Estimation Results

5.1 Likelihood-Based Estimation

I estimate the model using a likelihood-based estimation approach. Denote the vector of observables and public history of borrower i at period t by $\Omega_{it} = (O_{it}, r_{it}, X_{it}, K_{it})$ and $H_i^t = (O_{i1}, \cdots, O_{it-1})$, respectively. Let Θ represent the vector of parameters identified in Section 4. The log-likelihood of observables is

$$\begin{aligned} LL(\Theta) &= \sum_{i=1}^{N} \log \left[f_{\Omega_{i1},\cdots,\Omega_{iT}} \right] \\ &= \sum_{i=1}^{N} \log \left[\sum_{c} \sum_{\theta_{i1}} \cdots \sum_{\theta_{iT}} f_{\Omega_{i1},\theta_{i1},\cdots,\Omega_{iT},\theta_{iT},c} \right] \\ &= \sum_{i=1}^{N} \log \left[\sum_{c} \sum_{\theta_{i1}} \cdots \sum_{\theta_{iT}} f_{\Omega_{i1},\theta_{i1},c} \cdot \left(\prod_{t=2}^{T} f_{\Omega_{it},\theta_{it}|H_i^t,X_{it-1},K_{it-1},c} \right) \right], \end{aligned} \tag{5.1}$$

where the last equation holds under Assumptions 1 and 2. In Equation (5.1), $f_{\Omega_{i1},\theta_{i1},c}$ relates to borrowers' utility primitives and the original distribution of their default costs. For $t > 1$, $f_{\Omega_{it},\theta_{it}|H_i^t,X_{it-1},K_{it-1},c}$ is associated with both utility parameters and type-specific transition rules.[38] Notice that in Equation (5.1), if the borrower i only has one listing observed during the sample period, the log-likelihood simplifies to $\log[\sum_c \sum_{\theta_{i1}} f_{\Omega_{i1},\theta_{i1},c}]$.

The estimation procedure takes two steps. First, I estimate the conditional distribution of the interest rate and the funding probability given borrowers' characteristics and loan outcomes in the past from the data.[39] In the second step, I search for the values of parameters that maximize the log-likelihood function in Equation (5.1). Within each loop, I solve the dynamic optimization problem for borrowers of different types, taking the equilibrium interest rates and funding probabilities estimated in the first step as given. Borrowers' optimal effort levels determine their participation probabilities and the likelihood of different loan outcomes. Note also that in the equilibrium, borrowers' strategies are consistent with lenders' belief about their types. Given the optimal effort levels, the equilibrium belief can be derived (via the Bayes' Rule), which further leads to the funding probability of the loan. Integrating out private types (default cost c and cost of effort θ_t), I obtain the joint likelihood of observables for the given set of parameters.

In the estimation, I use *all* borrowers, including the ones who only appeared once during the sample period—these borrowers also have dynamic concerns and they may return to the market if they receive monetary demand shocks in the future. Since borrowers with more than three listings are rarely observed in the sample as shown in Table 1, I solve a three-period dynamic model and use the results from the first two periods to match the data.[40] Two additional observed characteristics of borrowers are incorporated into the estimation. They are (1) whether the loan is used for debt consolidation, and (2) the amount requested by the borrower.[41] I allow the borrower's outside option distribution to depend on the purpose of the loan; the probability that a loan is used for debt consolidation relates to the private type c. To focus on borrowers' participation and effort exerting strategies, in this paper I do not model amount as an endogenous choice made by the borrower. Instead, I assume that at each period, if the borrower receives a demand shock, the amount they need is a draw from a distribution that depends on their credit grades and the loan outcomes in the past. Borrowers' cost of effort distribution is potentially associated with the amount.

[38] $f_{\Omega_{it},\theta_{it}|H_i^t,X_{it-1},K_{it-1},c}$ can be further simplified as shown in the second line of Equation (4.1).

[39] I parameterize the conditional distribution of the interest rate by a normal distribution. I estimate the mean of interest rates given observables and also estimate the variance of the normal error. The estimation of the funding probability is based on a logit model.

[40] That is, $T = 3$ for all borrowers.

[41] I characterize the amount requested by borrowers into two groups. The means of the high and low amount groups are \$13,000 and \$ 8,700, respectively.

5.2 Results

The estimation results are shown in Table 5 with standard errors computed using a two-step maximum likelihood variance-covariance estimator (Murphy and Topel, 2002).[42]

I separate the estimates into four panels, where Panel (A) focuses on utility primitives and Panels (B)–(D) show the estimates of the distribution of the cost of effort, the probabilities of high type conditional on observables, and the type-specific transition rules of the state variables. The risk-aversion parameter I get from the estimation is around 1.49×10^{-4} using a CARA utility function.[43,44] The difference between the two categories of cost shocks is large, implying that transitory shocks are important when borrowers make their effort-exerting decisions. My estimates also show that, when defaults occur, high-type borrowers achieve negative payoffs and low type borrowers achieve positive payoffs. If no defaults occur, the average net revenue borrowers can get out of a loan is around 49 percent. The estimated coefficients in the mean of borrowers' outside option distribution suggest that individuals who have lower debt-to-income ratios or request loans for debt consolidation are faced with worse outside options. The last two lines of Panel (A) represent the estimated mean and variance of the lender's outside option distribution. My estimate of the average return is slightly lower than the risk-free rate in the outside markets. This may be because lenders' outside options are not restricted to investment options.

Panel (B) in Table 5 shows that borrowers who have higher default costs or request loans of a lower amount are more likely to draw smaller cost shocks. Panel (C) presents the estimates of original type distribution given different combinations of observables. It is clear that borrowers with high credit grades and who use the loans for debt consolidation are more likely to be "good borrowers" (i.e., have high default costs). Conditional on different debt-to-income ratios, the proportions of high type do not seem to vary much. From Panel (D), it

[42] In the estimation, I normalize $\gamma = 4.5$, $v_0 = -1$, and $v_k = 0.1$. The discount rate $\delta = 0.95$, and the probability that a future monetary demand shock arrives is calibrated to be 0.11 based on the summary statistics in Table 1. According to the rules of Prosper, borrowers who default once are not allowed to borrow again from this website. As a result, I cannot observe the appearance of those borrowers when they receive future money demand shocks. Moreover, it is hard to rule out the possibility that borrowers may borrow from other places instead if they need money. Therefore, 11 percent is a lower bound on the arrival rate of future money demand shocks.

[43] In the estimation, when computing the actual monetary value for each loan, I multiply the revenue $R_h - 1 - r_t$ and $R_l - c$ by the amount requested by borrowers (scaled by 10^4). As a result, the estimated risk-aversion parameter shown in Table 5 should be divided by 10^4.

[44] The risk-aversion parameter estimated in this paper is comparable to the estimates in Cohen and Einav (2007). The mean and median individual in their CARA specification has an absolute risk aversion equal to 3.1×10^{-3} and 3.4×10^{-5}, respectively. In other papers using data from peer-to-peer lending markets, the authors mainly focus on estimating the risk-aversion parameter for lenders (investors). For example, in Kawai, Onishi, and Uetake (2018), lenders' risk aversion parameter range from 3.49×10^{-2} to 5.71×10^{-2}; in Paravisini, Rappoport, and Ravina (2017), the estimated risk aversion parameter for the participants is around 3.68×10^{-2}.

Table 5: Estimation Results

Parameters	Estimates	Std. Err.
Panel (A): Utility Primitives		
Risk averse parameter (α)	1.4981	0.0717
Effectiveness of effort parameter (β)	2.3738	0.0189
Low cost of effort (θ_1)	0.0657	0.0038
High cost of effort (θ_2)	0.8090	0.0322
High default cost ($R_l - c_h$)	-0.5321	0.0190
Low default cost ($R_l - c_l$)	0.2370	0.0460
High revenue (R_h)	1.4949	0.0059
Coef. of dti ratio in b's outside option dist. (v_x)	0.0027	0.0125
Coef. of loan purpose in b's outside option dist. (v_d)	-0.4797	0.0127
Mean of lender's outside option (μ_0)	0.0005	0.0010
Std. err. of lender's outside option (σ_μ)	0.0261	0.0005
Panel (B): Distribution of the Cost of Effort		
Pr(low cost of effort|high type, large amount)	0.8913	0.0126
Pr($\sim$ |high type, small amount)	0.9960	0.0006
Pr($\sim$ |low type, large amount)	0.6663	0.0123
Pr($\sim$ |low type, small amount)	0.6734	0.0118
Panel (C): Distribution of the Default Cost (Type)		
Pr(high type|low dti, low credit grade, other purpose)	0.2335	0.0092
Pr($\sim$ |high dti, low credit grade, other purpose)	0.2461	0.0092
Pr($\sim$ |low dti, high credit grade, other purpose)	0.5326	0.0099
Pr($\sim$ |high dti, high credit grade, other purpose)	0.5870	0.0122
Pr($\sim$ |low dti, low credit grade, debt consolidation)	0.3219	0.0095
Pr($\sim$ |high dti, low credit grade, debt consolidation)	0.3311	0.0094
Pr($\sim$ |low dti, high credit grade, debt consolidation)	0.7320	0.0099
Pr($\sim$ |high dti, high credit grade, debt consolidation)	0.7323	0.0110
Panel (D): Type-Specific State Transition Probabilities		
Pr(low dti |high type, low dti)	0.6430	0.0099
Pr(high dti |high type, high dti)	0.8973	0.0089
Pr(low dti |low type, low dti)	0.4979	0.0102
Pr(high dti |low type, high dti)	0.9170	0.0079
Pr(loan used for other purpose|high type)	0.4314	0.0059
Pr(loan used for other purpose|low type)	0.4997	0.0077

is clear that when borrowers have high debt-to-income ratios, there is a high chance for them to stay in that adverse situation. Borrowers with c_h are more likely to stay with low debt-to-income ratios (about 64 percent). In addition, I find high type borrowers are more likely to propose loans that are used for debt-consolidation. This result is intuitive in the sense that borrowers who attempted to pay off existing loans are likely to have higher default cost and may be more responsive to the dynamic incentives imposed by the reputation system.

5.3 Model Fit

I check the model fit by comparing borrowers' participation probabilities, lenders' funding probabilities, and default and late payment rates generated by the model with the ones observed in the data. This exercise is done for borrowers with different combinations of observables. The results are shown in Table 11.[45] In addition, I use the model and my estimates to predict the shift of credit grade distribution across two listings and compare the results with the patterns observed in the data. Since I classify the seven credit grades into two groups in the empirical estimation, I reproduce the histograms in Figure 1 for high and low credit grades.[46] The simplified version is provided in the left panel of Figure 3. To recap, the shift of the credit grade distribution may occur for two reasons. First, borrowers who pay off the first loans and propose the second listings may have higher credit grades (selection channel). An alternative explanation would be that borrowers' credit grades are updated after the first loan so that the distribution shifts (update channel). Using the structural model, I decompose the two channels by plotting the distribution of credit grades first for all borrowers after the first loan, and then only for borrowers who pay off their loans and are selected in the second period. The shift of credit grade distribution under the three stages is shown in the right panel of Figure 3. It is clear from the figure that the updating channel is important and the selection channel further shifts the distribution to the right. Combining all results, my estimates arguably match the data patterns well.

6 Counterfactual Analysis

Given that the utility primitives and the distribution of borrowers' private information have been recovered, I conduct three sets of counterfactuals in this section. I first compare welfare under three information structures—one with types and effort observed (symmetric),

[45]Note that funding probabilities are computed using listings that were not withdrawn by borrowers. Default and late payment rates are computed for loans that were funded.

[46]In the empirical setting, high credit groups include borrowers with credit grade AA, A, and B. Other grades are characterized as a low credit group.

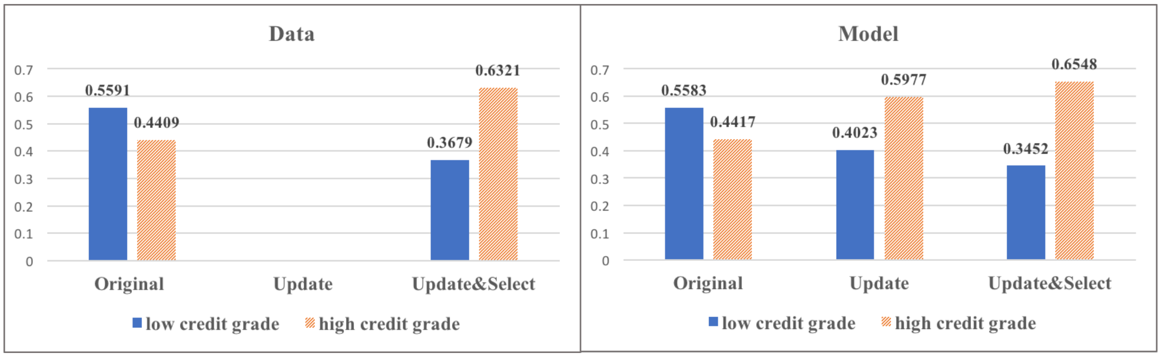

Figure 3: Compare the Shift of Credit Grade Distribution: Data v.s. Model

one with only types observed, and one with both unobserved (asymmetric). These counterfactuals allow me to (1) quantify the total welfare loss from asymmetric information and (2) decompose the sources of inefficiency into adverse selection and moral hazard. I further consider the *factual* scenario, where a reputation system is implemented in the market with asymmetric information. By comparing the welfare gain from reputation with the total welfare loss, I am able to quantify the extent to which efficiency is restored by the reputation system. To alleviate long-run inefficiencies induced by the reputation/feedback system, I further study the implications of offering a payment protection insurance to borrowers.

For all counterfactual experiments, I assume that the website chooses the optimal interest rate to maximize the rate of successful transaction, since it charges commission fees for each successfully funded loan.[47] I solve borrowers' optimal effort levels, lenders' funding decisions, and the website's pricing rule for each market design. Table 6 presents the rates of successful transaction, borrower's participation and lender's funding probabilities, default and late payment probabilities, and average utilities for borrowers and lenders under the five scenarios. Note that a transaction occurs when a borrower participates and lenders invest. Hence, the rate of successful transaction equals the multiplication of the borrowers' participation rate and lenders' funding probability (computed for listings not withdrawn by borrowers). Default and late payment rates and average utilities for borrowers and lenders are computed for funded loans. The upper and lower panels of Table 6 show results for the first and second loans, respectively. Table 12 in the Appendix summarizes market outcomes and prices by type.

[47]The website may also be concerned about other aspects of market performance. For example, the website may prefer to keep the overall default rate below a certain level, it may prefer to attract more borrowers with high credit grades, etc. For this paper, I focus on the main channel through which the website makes a profit.

Table 6: Counterfactual Results: Comparing Welfare under Five Scenarios

Scenarios	S1	S2	S3	S4	S5
	Symmetric	Observe type	Asymmetric	Reputation	Rep.+PPI
			First Loan		
Rate of transaction	0.9282	0.5577	0.4657	0.9255	0.9252
Participation prob.	0.9293	0.9359	0.9329	0.9267	0.9266
Funding prob.	0.9988	0.5959	0.4992	0.9988	0.9985
Default prob.	0.1165	0.1007	0.2090	0.1563	0.1256
Late prob.	0.0302	0.0228	0.1108	0.0596	0.0505
Borrower's avg. util.	0.2487	0.2658	0.2420	0.2497	0.2578
Lender's avg. util.	0.0816	0.0628	0.0550	0.0801	0.0895
			Second Loan		
Rate of transaction	0.9186	0.5495	0.4451	0.7413	0.7676
Participation prob.	0.9198	0.9236	0.9239	0.7493	0.9249
Funding prob.	0.9987	0.5949	0.4817	0.9893	0.8299
Default prob.	0.1100	0.1007	0.2099	0.1481	0.1553
Late prob.	0.0305	0.0216	0.1178	0.0595	0.0636
Borrower's avg. util.	0.2586	0.2722	0.2392	0.2642	0.2752
Lender's avg. util.	0.0836	0.0619	0.0577	0.0665	0.0573
Total Surplus	0.3894	-0.3208	-0.5237	0.3444	0.3779

6.1 Welfare Loss from Asymmetric Information

I first compare the welfare of market participants given different information structures. The first features symmetric information, so that the website and lenders perfectly observe and price based on borrowers' types and effort.[48] In the second scenario, only borrowers' types are observed and priced. The website has to take borrowers' individual rationality (IR) and incentive compatibility (IC) constraints into account when choosing optimal interest rates. The third structure is when both types and effort are unobserved by the website and lenders, and no other mechanisms are imposed. The difference in welfare levels between S1 and S3 measures the total loss from asymmetric information. The difference between S1 and S2 is only due to unobserved effort, which quantifies the welfare loss from moral hazard. By comparing these three market designs, I find that 22 percent of the inefficiency from

[48]Specifically, in the first scenario, I assume that once a loan originates, the borrower is enforced to exert the effort level required by the website and lenders. When maximizing the profit, the website chooses the optimal combination of interest rate and effort level, taking only the borrower's IR constraint into account.

asymmetric information is due to adverse selection, and 78 percent is due to moral hazard.

From Column 3 in Table 6, it is clear that the presence of asymmetric information leads to an inefficient market outcome. When both types and effort are unobserved, borrowers exert less effort and low-type borrowers are more likely to participate—these result in a higher default rate (around 20 percent) and a lower funding probability (49 percent). Due to asymmetric information, some potentially good borrowers have difficulties in getting funded and the transaction rate for the website is also low (around 46 percent). In the case of symmetric information, the website and lenders perfectly observe and price based on borrower's type and effort. Borrowers are charged with type-specific interest rates. For borrowers that have a low default cost and draw large cost shocks of exerting effort, their interest rates are around 35 percent, which is very close to the upper limit set by the usury law. On the other hand, interest rates for good borrowers are around 10 percent. Under this scenario, lenders have a very high probability to invest and the transaction rate is about 92 percent. When only borrower's types are observed, this market suffers from moral hazard. I find from the analysis that, for borrowers that have low default costs and draw large cost shocks, the website cannot find an interest rate at which lenders are willing to invest. In other words, the market collapses for this group of "lemon" borrowers. The transaction rate under this scenario is 55 percent, which is higher than that in S3 but lower than that in S1. This finding also implies that reducing adverse selection helps to improve market outcomes, while moral hazard still plays an important role in creating inefficiencies. In Scenarios 1–3, borrowers have no dynamic concerns.[49] The results for the first and second loans are very similar.

6.2 The Value of the Reputation System

I now consider the *factual* scenario, where a reputation system is implemented in the market with asymmetric information. In the analysis, I closely follow the rules imposed by Prosper.com and also require the reputation system to not allow borrowers with late payments to enter again in order to reduce computational burden. From the experiment, I find that 95 percent of welfare loss from asymmetric information is recovered by the imposed reputation system through the following channels. First, the reputation system helps to refine beliefs about borrowers' types, so that "lemons" are excluded from the market gradually. Table 7 presents the proportion of high-type borrowers conditional on observables across loans. Borrowers who have adverse loan outcomes (such as defaults and late payments) in the previous loans are not allowed to enter in the future—this selection process improves

[49] When information is symmetric, borrowers have no incentive to exert more effort in earlier periods. When there is asymmetric information but no reputation system is imposed, lenders do not update their beliefs based on past loan outcomes. Borrowers again have no dynamic incentives.

the pool of borrowers. Second, when the reputation system is implemented, borrowers are incentivized to exert more effort, which leads to a lower default rate (around 15 percent) and higher average utilities for lenders. Third, from Table 12, I find that the funding rate for all types under reputation is high (99 percent). That is, even low-type borrowers now have access to credit, which reduces welfare loss from the credit rationing observed in S2 and S3. Given the higher rate of matching (92 percent), the website receives more profit when the reputation system is implemented.

Table 7: Proportion of High-Type Borrowers Over Time

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

I then compare the outcomes of the first and second loans under the reputation system. As opposed to the cases in Section 6.1, borrowers now have dynamic concerns. There are two counter forces that impact outcomes in the second loans. On the one hand, borrowers who have survived are more likely to have better types. On the other hand, since they are closer to the final period, borrowers have less incentive to exert effort. The lower panel of Table 6 shows that default and late payment rates are slightly lower for second loans, which indicates that the positive effect of selection dominates the negative effect of decreasing incentives. Another interesting point from this table is that borrowers' participation rate in second loans under the reputation system is significantly lower than that in first loans. This is because borrowers who default in the past are not allowed to participate in the second period and can only take random draws from the outside option distribution. Some of the borrowers may be of good types but receive bad shocks. The accidental loss of reputation prevents good borrowers from having future credit access, and thus may lead to potential long-run inefficiencies.

6.3 Payment Protection Insurance

To address long-term inefficiencies due to accidental loss of reputation, the last part of my counterfactual analysis considers remedies for welfare loss induced by imperfect monitoring of the reputation system. In the current setting of the website, once in default, the borrower loses his reputation immediately and is unable to borrow again. However, even if borrowers have good types and exert high levels of effort, it is still possible that they obtain unlucky draws of revenue. Because of imperfect monitoring, the reputation system cannot differentiate between "lemons" and good borrowers with unlucky draws. This situation has strong empirical relevance, especially for small businesses that find peer-to-peer lending an attractive financing alternative (Segal, 2015) and rely heavily on this form of credit access for their success and growth.

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

Adding PPI to the main model in Section 3 creates new challenges. Borrowers now have an additional decision to make—whether to purchase the insurance or not. On top of that, I need to model how premiums of the plan are determined by the insurer. The key features of the model with PPI are summarized as follows. First, borrowers make insurance purchase decisions after participation but before exerting effort. If the borrower purchases the insurance, he has to pay the premium ex ante, but the insurer will cover the repayments whenever bad shocks occur in the future.[50] If the borrower is not covered by PPI, there is a possibility that he will default or have late payments.[51] Second, to simplify the model, I assume that insurance purchase decisions are not publicly available to the website or lenders—their beliefs about borrowers' private information only depend on observed outcomes, such as defaults and late payments. Third, the insurer charges a premium to break even. Like the website and lenders, the insurer also does not observe borrowers' private information. In the equilibrium, strategies of all players, beliefs about borrowers' types, and insurance premiums are all consistent with each other.

In the last of column of Table 6, I compute market outcomes when the reputation system is imposed with an option for borrowers to purchase the payment protection insurance.

[50] In the counterfactual, I assume that if the borrower is covered by PPI, there are no defaults or late payments "on file" from the lender's point of view—the borrower will have access to credit in the future. However, the borrower still has to pay a default cost if he files a claim with the insurance company. This assumption makes intuitive sense because in reality, filing a claim leads to a potential increase in premiums.

[51] In this counterfactual, I maintain the assumption that once defaults or late payments occur, borrowers are not allowed to enter again.

When borrowers have access to insurance, the risk faced by lenders is reduced, and thus their average utility increases. From the borrowers' point of view, once they purchase the insurance, they get access to future credit. I find that with PPI introduced, borrowers have higher chance to participate in second loans and the rate of transaction also increases. This finding highlights the effect of this intervention on alleviating long-run inefficiencies caused by accidental loss of reputation. This exercise also shows that only borrowers with high chances of being a good type purchase the insurance. After taking adverse selection and moral hazard into account, the insurer cannot find premiums to break even for borrowers who are more likely to be "lemons". Overall, I find adding PPI to the market further increases the total welfare of market participants. 98 percent of the welfare loss from asymmetric information is eliminated under this mechanism.

6.4 Impact on Market Size

To get a rough idea about the impact of different mechanisms on the size of online credit markets, I do a simple calculation for Prosper.com. The probabilities that transactions will occur shown in Table 6 are used to approximate market sizes. Table 8 summarizes the market sizes under different market designs. The current market size for Prosper.com is around 9 billion dollars. If the reputation system on this market is removed, then the market size could shrink to 4.6 billion dollars. If the market only suffers from moral hazard, the market size is larger (5.5 \$bn), but still much lower than the symmetric information scenario (9.2 \$bn). From this exercise, we can again see that the reputation system helps to restore a large proportion of efficiency in online credit markets. Adding payment protection insurance can further increase total welfare.

Table 8: Market Size under Different Scenarios

Scenarios	Market Size (\$bn)
Symmetric Information	9.18
Reputation+PPI	9.02
Reputation (Factual)	9.00
Under Moral Hazard	5.51
Under Moral Hazard and Adverse Selection	4.59

7 Conclusion

This paper investigates the effectiveness of reputation/feedback systems in improving welfare in online credit markets when both adverse selection and moral hazard are present. I develop a finite-horizon dynamic structural model to analyze borrowers' repayment decisions, lenders' investment strategies, and the website's pricing schemes. I prove the identification of the distribution of borrowers' private types and utility primitives based on variations in borrowers' repayment histories, transitions of their characteristics, and interest rates, and then estimate the model using a large transaction-level dataset from Prosper.com.

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

The main contribution of this paper is to quantify the welfare gain from reputation systems in online credit markets. I find that reputation matters to a large extent through refining beliefs about borrowers' types, incentivizing effort exertion, and expanding credit access to low-type borrowers. These results may have implications for other settings, including fast-growing online marketplaces that widely use review systems to facilitate transactions and traditional credit markets that rely heavily on the credit rating system. My paper also considers a policy intervention that protects borrowers from accidental loss of reputation. I find that introducing a payment protection insurance into the market further improves total welfare. This exercise is related to the optimal "forgiveness" mechanisms that are considered in the existing credit rating system.

References

Adams, William, Liran Einav, and Jonathan Levin. 2009. "Liquidity Constraints and Imperfect Information in Subprime Lending." *The American Economic Review,* 99(1): 49–84.

Abbring, J. H., P. A. Chiappori, J. H. Heckman, and J. Pinquet. 2003. "Adverse Selection and Moral Hazard in Insurance: Can Dynamic Data Help to Distinguish?" *Journal of the European Economic Association,* 1(2–3): 512–521.

Aguirregabiria, Victor, and Pedro Mira. 2007. "Sequential Estimation of Dynamic Discrete Games." *Econometrica,* 75(1): 1–53.

Akerlof, George. 1970. "The Market for Lemons: Qualitative Uncertainty and Market Mechanism." *Quarterly Journal of Economics,* 89.

Bai, Jie. 2016. "Melons as Lemons: Asymmetric Information, Consumer Learning and Seller Reputation." *Working Paper.*

Bajari, Patrick, Christina Dalton, Han Hong, and Ahmed Khwaja. 2014. "Moral Hazard, Adverse Selection, and Health Expenditures: A Semiparametric Analysis." *The RAND Journal of Economics,* 45(4): 747–763.

Bar-Isaac, Heski, and Steven Tadelis. 2008. *Seller Reputation.* Now Publishers Inc.

Cabral, Luis, and Ali Hortacsu. 2010. "The Dynamics of Seller Reputation: Evidence from eBay." *The Journal of Industrial Economics,* 58(1): 54–78.

Chiappori, Pierre-Andre, and Bernard Salanie. 2000. "Testing for Asymmetric Information in Insurance Markets." *Journal of Political Economy,* 108(1): 56–78.

Chiappori, Pierre-Andre, and Bernard Salanie. 2002. "Testing Contract Theory: A Survey of Some Recent Work." *CESifo Working Paper*, No. 738.

Chiappori, Pierre-Andre, Bruno Jullien, Bernard Salanie, and Francois Salanie. 2006. "Asymmetric Information in Insurance: General Testable Implications." *The RAND Journal of Economics,* 37(4): 783–798.

Cohen, Alma, and Liran Einav. 2007. "Estimating Risk Preferences from Deductible Choice." *The American Economic Review,* 97(3): 745–788.

Diamond, Douglas W. 1989. "Reputation Acquisition in Debt Markets." *The Journal of Political Economy,* 97(4): 828–862.

Eaton, David H. 2005. "Valuing Information: Evidence from Guitar Auctions on eBay." *Journal of Applied Economics & Policy,* 24(1): 1.

Einav, Liran, Mark Jenkins, and Jonathan Levin. 2012. "Contract Pricing in Consumer Credit Markets." *Econometrica,* 80(4): 1387–1432.

Einav, Liran, Mark Jenkins, and Jonathan Levin. 2013. "The Impact of Credit Scoring on Consumer Lending." *The RAND Journal of Economics,* 44(2): 249–274.

Einav, Liran, Chiara Farronato, and Jonanthan Levin. 2016. "Peer-to-Peer Markets." *Annual Review of Economics,* 8: 615–635.

Fan, Ying, Jiandong Ju, and Mo Xiao. 2016. "Reputation Premium and Reputation Management: Evidence from the Largest E-Commerce Platform in China." *International Journal of Industrial Organization,* 46: 63–76.

Gayle, George-Levi, and Robert A. Miller. 2015. "Identifying and Testing models of Managerial Compensation." *The Review of Economic Studies,* 82(3): 1074–1118.

Houde, J. F., and S. Imai. 2006. "Identification and 2-step Estimation of DDC models with Unobserved Heterogeneity." *Working Paper,* Queen's University.

Holmström, Bengt. 1999. "Managerial Incentive Problems: A Dynamic Perspective." *The Review of Economic Studies,* 66(1): 169–182.

Hu, Yingyao. 2008. "Identification and Estimation of Nonlinear Models with Misclassification Error Using Instrumental Variables: A General Solution." *Journal of Econometrics,* 144(1): 27–61.

Hu, Yingyao, and Susanne M. Schennach. 2008. "Instrumental Variable Treatment of Nonclassical Measurement Error Models." *Econometrica,* 76(1): 195–216.

Hu, Yingyao, and Matthew Shum. 2012. "Nonparametric Identification of Dynamic Models with Unobserved State Variables." *Journal of Econometrics,* 171(1): 32–44.

Hu, Yingyao, and Yi Xin. 2019. "Identification and Estimation of Dynamic Structural Models with Unobserved Choices." *Working Paper.*

Jin, Ginger Zhe, and Andrew Kato. 2006. "Price, Quality, and Reputation: Evidence from an Online Field Experiment." *The RAND Journal of Economics,* 37(4): 983–1005.

Kasahara, Hiroyuki, and Katsumi Shimotsu. 2009. "Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices." *Econometrica,* 77(1): 135–175.

Kawai, Kei, Ken Onishi, and Kosuke Uetake. 2018. "Signaling in Online Credit Markets." *Working Paper.*

Klein, Tobias J., Christian Lambertz, and Konrad O. Stahl. 2016. "Market Transparency, Adverse Selection, and Moral Hazard." *Journal of Political Economy,* 124(6): 1677–1713.

Klein, Benjamin, and Keith B. Leffler. 1981. "The Role of Market Forces in Assuring Contractual Performance." *The Journal of Political Economy,* 89(4): 615–641.

Lewis, Gregory, and Georgios Zervas. 2016. "The Welfare Impact of Consumer Reviews: A Case Study of the Hotel Industry." *Working Paper.*

Lin, Mingfeng, Yong Liu, and Siva Viswanathan. 2016. "Effectiveness of Reputation in Contracting for Customized Production: Evidence from Online Labor Markets." *Management Science,* 64(1): 345–359.

Lucking-Reiley, David, Doug Bryan, Naghi Prasad, and Daniel Reeves. 2007. "Pennies from eBay: The Determinants of Price in Online Auctions." *The Journal of Industrial Economics,* 55(2): 223–233.

Mailath, George J., and Larry Samuelson. 2006. *Repeated Games and Reputations: Long-Run Relationships.* Oxford university press.

Melnik, Mikhail I., and James Alm. 2002. "Does a Seller's Ecommerce Reputation Matter? Evidence from eBay Auctions." *The Journal of Industrial Economics,* 50(3): 337–349.

Murphy, Kevin M., and Robert H. Topel. 2002. "Estimation and Inference in Two-Step Econometric Models." *Journal of Business & Economic Statistics,* 20(1): 88–97.

Paravisini, Daniel, Veronica Rappoport, and Enrichetta Ravina. 2017. "Risk Aversion and Wealth: Evidence from Person-to-Person Lending Portfolios." *Management Science,* 63(2):279–297.

Perrigne, Isabelle, and Quang Vuong. 2011. "Nonparametric Identification of a Contract Model with Adverse Selection and Moral Hazard." *Econometrica,* 79(5): 1499–1539.

Saeedi, Maryam. 2014. "Reputation and Adverse Selection: Theory and Evidence from eBay." *Working Paper.*

Segal, M. 2015. "Peer-to-Peer Lending: A Financing Alternative for Small Businesses." *Issue Brief,* 10.

Stiglitz, Joseph E., and Andrew Weiss. 1981. "Credit Rationing in Markets with Imperfect Information." *The American Economic Review,* 71(3): 393–410.

Stiglitz, Joseph E., and Andrew Weiss .1983. "Incentive Effects of Terminations: Applications to the Credit and Labor Markets." *The American Economic Review,* 73(5): 912–927.

Tadelis, Steven. 2016. "Reputation and Feedback Systems in Online Platform Markets." *Annual Review of Economics,* 8(1): 321–340..

Yoganarasimhan, Hema. 2013. "The Value of Reputation in an Online Freelance Marketplace." *Marketing Science,* 32(6): 860–891.

A Constructing Eigenvalue-Eigenvector Decompositions

A.1 Identifying the Latent Type Distribution

I now consider identification of the latent type distribution given fixed values of (r_t, K_t, H^t). To simplify notations, I drop (r_t, K_t, H^t) in the subscripts of densities and rewrite Equation (4.1) in a way that highlights the key conditional independence assumptions.

$$f_{O_t,X_t,X_{t-1},K_{t-1}} = \sum_c f_{O_t|X_t,c} \cdot f_{X_t|X_{t-1},c} \cdot f_{X_{t-1},K_{t-1},c}. \tag{A.1}$$

For fixed values of (X_t, X_{t-1}), the LHS of Equation (A.1) represents the joint distribution of O_t and K_{t-1}, which is a finite mixture of component distributions for different type c with $f_{X_{t-1},K_{t-1},c}$ as the mixing weight. Note that (1) the lagged characteristics (X_{t-1}, K_{t-1}) provide an exogenous source of variations in the mixing weights; (2) X_t relates to the conditional distribution of O_t given c; and (3) X_t and X_{t-1} are connected only through the type-specific transition process.

I stack Equation (A.1) for different values of (O_t, c, K_{t-1}) given fixed values of (X_t, X_{t-1}). For illustration, suppose all variables take binary values from $\{0, 1\}$. Let $O_t = 1, 0$ denote the case where adverse or good loan outcomes occur, respectively. For borrowers with high default cost, $c = 1$; otherwise $c = 0$. $K_{t-1} = 1$ if borrowers have high credit grades, and 0 otherwise. Define

$$M_{X_t,X_{t-1}} = \begin{bmatrix} f_{1,X_t,X_{t-1},1} & f_{1,X_t,X_{t-1},0} \\ f_{0,X_t,X_{t-1},1} & f_{0,X_t,X_{t-1},0} \end{bmatrix}, \quad A_{X_t} = \begin{bmatrix} f_{1|X_t,1} & f_{1|X_t,0} \\ f_{0|X_t,1} & f_{0|X_t,0} \end{bmatrix},$$
$$Q_{X_t,X_{t-1}} = \begin{bmatrix} f_{X_t|X_{t-1},1} & 0 \\ 0 & f_{X_t|X_{t-1},0} \end{bmatrix}, \quad B_{X_{t-1}} = \begin{bmatrix} f_{X_{t-1},1,1} & f_{X_{t-1},0,1} \\ f_{X_{t-1},1,0} & f_{X_{t-1},0,0} \end{bmatrix}.$$

The matrix form of Equation (A.1) is simply

$$M_{X_t,X_{t-1}} = A_{X_t} \cdot Q_{X_t,X_{t-1}} \cdot B_{X_{t-1}}, \tag{A.2}$$

and our goal is to pin down all elements in A_{X_t}, $Q_{X_t,X_{t-1}}$ and $B_{X_{t-1}}$. I now explore variations in X_t and X_{t-1}. Specifically, if $A_{X_t}, Q_{X_t,X_{t-1}}$, and $B_{X_{t-1}}$ are invertible for any (X_t, X_{t-1}),

$$\begin{aligned}\left(M_{1,1} \cdot M_{0,1}^{-1}\right) \cdot \left(M_{1,0} \cdot M_{0,0}^{-1}\right)^{-1} &= \left(A_1 \cdot Q_{1,1} \cdot Q_{0,1}^{-1} \cdot A_0^{-1}\right) \cdot \left(A_1 \cdot Q_{1,0} \cdot Q_{0,0}^{-1} \cdot A_0^{-1}\right)^{-1} \\ &= A_1 \left(Q_{1,1} \cdot Q_{0,1}^{-1} \cdot Q_{0,0} \cdot Q_{1,0}^{-1}\right) A_1^{-1} \equiv A_1 Q A_1^{-1}.\end{aligned} \tag{A.3}$$

The construction of Equation (A.3) follows Hu and Shum (2012). Though the algebra seems to be complicated, the intuition for constructing this equation is straightforward. For fixed values of X_{t-1}, using variations in X_t cancels out $B_{X_{t-1}}$ as shown in the parentheses in the first row. Combing the cases in which X_{t-1} equals 0 and 1 leads to an eigenvalue-eigenvector decomposition of the observed matrix on the LHS of Equation (A.3). A_1 represents the matrix of eigenvectors and the diagonal elements of Q are eigenvalues. Assumption 3 in Section 4.2 guarantees the invertibility of $A_{X_t}, Q_{X_t,X_{t-1}}$, and $B_{X_{t-1}}$, as well as the uniqueness of the eigenvalue-eigenvector decomposition.

A.2 Identifying the Effects of Effort on Loan Outcomes

I now provide details of identifying the effects of effort on loan outcomes using Equation (4.3) for fixed values of $(r_T, X_T, H^T, c, W_T = 0, I_T = 1)$. To simplify notations, I drop $(r_T, X_T, H^T, c, W_T, I_T)$ in the subscripts of densities and rewrite Equation (4.3) as follows.

$$f_{D_T,L_T|K_T} = \sum_{\theta_T} f_{D_T|\theta_T} \cdot f_{L_T|\theta_T} \cdot f_{\theta_T|K_T} \tag{A.4}$$

I stack Equation (A.4) for different values of (D_T, θ_T, K_T) given fixed values of L_T. For illustration, suppose all variables take binary values form $\{0, 1\}$. The definitions of (D_T, L_T, K_T) remain the same. Let $\theta_T = 1$ when borrowers have high cost of effort, and 0 otherwise. Define

$$M_{L_T} = \begin{bmatrix} f_{1,L_T|1} & f_{1,L_T|0} \\ f_{0,L_T|1} & f_{0,L_T|0} \end{bmatrix}, \quad A = \begin{bmatrix} f_{D_T|\theta_T}(1|1) & f_{D_T|\theta_T}(1|0) \\ f_{D_T|\theta_T}(0|1) & f_{D_T|\theta_T}(0|0) \end{bmatrix},$$
$$Q_{L_T} = \begin{bmatrix} f_{L_T|1} & 0 \\ 0 & f_{L_T|0} \end{bmatrix}, \quad B = \begin{bmatrix} f_{\theta_T|K_T}(1|1) & f_{\theta_T|K_T}(1|0) \\ f_{\theta_T|K_T}(0|1) & f_{\theta_T|K_T}(0|0) \end{bmatrix}.$$

The matrix form of Equation (A.4) is simply

$$M_{L_T} = A \cdot Q_{L_T} \cdot B, \tag{A.5}$$

and our goal is to pin down all elements in A, Q_{L_T}, and B. I now explore variations in L_T. If A, Q_{L_T}, and B are invertible for $L_t = \{0, 1\}$,

$$M_1 \cdot M_0^{-1} = \Big(A \cdot Q_1 \cdot B\Big) \cdot \Big(A \cdot Q_0 \cdot B\Big)^{-1} = A \cdot \Big(Q_1 \cdot Q_0^{-1}\Big) \cdot A^{-1} \equiv A\tilde{Q}A^{-1}. \tag{A.6}$$

Equation (A.6) leads to an eigenvalue-eigenvector decomposition of matrix $M_1 \cdot M_0^{-1}$, where A is the matrix of eigenvectors and the diagonal elements of $\tilde{Q}$ are corresponding eigenvalues.

The following assumption guarantees the invertibility of A, Q_{L_T}, and B, as well as the uniqueness of the eigenvalue-eigenvector decomposition.

Assumption A.1. *For all values of (r_T, X_T, c), the following conditions are satisfied: (i) $F_{D_T|r_T,\theta_T=\theta^h,c}$ first order stochastically dominates $F_{D_T|r_T,\theta_T=\theta^l,c}$, $\forall\theta^h > \theta^l$; (ii) The distribution of L_T given (r_T, θ_T, c) has a positive density $\forall\theta_T \in \Theta_\theta$; (iii) $g(K^h) \neq g(K^l)$, where $g(K_T) = \frac{f_{W_T=0|r_T,X_T,K_T,\theta_T=\theta^h,c}}{f_{W_T=0|r_T,X_T,K_T,\theta_T=\theta^l,c}}$, $\forall K^h > K^l$.*

Assumption A.1(i) implies that borrowers with higher cost of effort draws are more likely to default. This assumption is consistent with the theoretical model, as borrowers with higher marginal cost exert lower level of effort, which results in a higher chance of default. Assumption A.1(ii) requires that the probability of paying late lies in $(0, 1)$. Assumption A.1(iii) implies a monotone likelihood ratio property in K_T—the ratio of borrowers' participation probabilities given different cost of effort varies with the outside option distribution. Intuitively, when borrowers have low credit grades (i.e., K^l), the mean of their outside options is small; no matter the cost of effort draws are small or large, borrowers prefer to participate, so the ratio of participation probabilities is close to 1. When borrowers have high credit grades (i.e., K^h), the draw of θ could play a big role for borrowers to decide whether to participate or not. Borrowers are less likely to participate when θ^h is drawn, so the ratio $\frac{f_{W_T=0|r_T,X_T,K_T,\theta_T=\theta^h,c}}{f_{W_T=0|r_T,X_T,K_T,\theta_T=\theta^l,c}}$ is less than 1.

Theorem A.1 (Identification)**.** *Suppose assumptions in Theorem 1 hold. If Assumptions 4 and A.1 are satisfied, densities $f_{D_T|r_T,\theta_T,c}$, $f_{L_T|r_T,\theta_T,c}$, and $f_{W_T=0,\theta_T|r_T,X_T,K_T,c}$ are identified for any $(D_T, r_T, X_T, K_T, \theta_T, c)$.*

Again, the formal proof of Theorem A.1 follows Hu (2008), and hence is skipped here.

B Identification of Utility Primitives in Borrower's Payoff Functions

As discussed in Section 4.4, once we normalize v_0, σ_v, and v_x, variations that shift the mean of borrowers' outside option distribution identify borrower's value from participation. Consider the unknown parameters in $\tilde{V}_{c,T}(e, r_T, \theta_T)$. Explicitly,

$$\tilde{V}_{c,T}(e, r_T, \theta_T) = p(e; \beta)U(R_h - 1 - r_T; \alpha) + (1 - p(e; \beta))U(R_l - c; \alpha) - \phi(e, \theta_T). \tag{B.1}$$

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 β measuring the effectiveness of the effort; and (2) it induces cost through $\phi(e, \theta_T) = \theta_T e^2$, where θ_T

represents the cost of effort. In Equation (B.1), the unknowns include β, α, R_h, R_l, and the levels of c and θ_T.

B.1 Identification of the Risk-Aversion Parameter

In this section, I focus on the identification of R_h and α using variations in interest rates. With $\tilde{V}_{c,T}(e^*_{c,T}(r_T, \theta_T), r_T, \theta_T))$ recovered, its derivative with respect to r_T,

$$\begin{aligned}
&\frac{\partial \tilde{V}_{c,T}(e^*_{c,T}(r_T, \theta_T), r_T, \theta_T))}{\partial r_T} \\
&= \frac{\partial e^*_{c,T}(r_T, \theta_T)}{\partial r_T} \left[p'(e^*_{c,T}(r_T, \theta_T))(U(R_h - 1 - r_T) - U(R_l - c)) - \phi'(e^*_{c,T}(r_T, \theta_T))\right] \\
&\quad - p(e^*_{c,T}(r_T, \theta_T))U'(R_h - 1 - r_T) \\
&= -p(e^*_{c,T}(r_T, \theta_T))U'(R_h - 1 - r_T),
\end{aligned} \tag{B.2}$$

is also identified. Notice that the second equality in Equation (B.2) holds because $e^*_{c,T}(r_T, \theta_T)$ is the optimal effort level and thus satisfies the first order condition in borrower's optimization problem. Observe also that $p(e^*_{c,T}(r_T, \theta_T)) = f_{D_T=0|r_T,\theta_T,c}$, which is recovered from the eigen-decomposition in Section 4.3. Therefore Equation (B.2) identifies $U'(R_h - 1 - r_T)$. Under the assumption of CARA utility, with two observed interest rates $\bar{r}_T$ and $\hat{r}_T$, α is identified through the following equation

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

Plugging α back to $U'(R_h - 1 - r_T)$, R_h is identified.

B.2 Identification of Levels of Private Information

To identify the levels of c and θ_t, I relate the identified default probability $f_{D_t=0|r_t,\theta_t,c}$ to the effort level through $p(\cdot;\beta)$.[52]

$$f_{D_t=0|r_t,\theta_t,c} = p(e^*_{c,T}(r_T, \theta_T); \beta) = 1 - \exp(-\beta e^*_{c,T}(r_T, \theta_T)). \tag{B.4}$$

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

Equation (B.4) for $c = c_h, c_l$ uniquely determines $\frac{e^*_{c_h,T}(r_T,\theta_T)}{e^*_{c_l,T}(r_T,\theta_T)}$. From Equation (B.1), I construct the following two equations for θ_h and θ_l.

$$\begin{aligned}
\left(\frac{e^*_{c_h,T}(r_T,\theta_h)}{e^*_{c_l,T}(r_T,\theta_h)}\right)^2 &= \frac{p(e^*_{c_h,T}(r_T,\theta_h))U(R_h-1-r_T)+(1-p(e^*_{c_h,T}(r_T,\theta_h)))U(R_l-c_h)-\tilde{V}_{c_h,T}(e^*_{c_h,T}(r_T,\theta_h),r_T,\theta_h)}{p(e^*_{c_l,T}(r_T,\theta_h))U(R_h-1-r_T)+(1-p(e^*_{c_l,T}(r_T,\theta_h)))U(R_l-c_l)-\tilde{V}_{c_l,T}(e^*_{c_l,T}(r_T,\theta_h),r_T,\theta_h)}\\
\left(\frac{e^*_{c_h,T}(r_T,\theta_l)}{e^*_{c_l,T}(r_T,\theta_l)}\right)^2 &= \frac{p(e^*_{c_h,T}(r_T,\theta_l))U(R_h-1-r_T)+(1-p(e^*_{c_h,T}(r_T,\theta_l)))U(R_l-c_h)-\tilde{V}_{c_h,T}(e^*(c_h,r_T,\theta_l),r_T,\theta_l)}{p(e^*_{c_l,T}(r_T,\theta_l))U(R_h-1-r_T)+(1-p(e^*_{c_l,T}(r_T,\theta_l)))U(R_l-c_l)-\tilde{V}_{c_l,T}(e^*(c_l,r_T,\theta_l),r_T,\theta_l)}.
\end{aligned} \tag{B.5}$$

(B.5) provides a system of two linear equations of $U(R_l - c_h)$ and $U(R_l - c_l)$. To ensure the identification of $R_l - c_h$ and $R_l - c_l$, I invoke the following rank condition.[53]

Assumption B.1 (Rank Condition)**.**

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

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

Note that all terms in Assumption B.1 have been recovered, so the rank condition is directly testable. This condition guarantees the unique solution of $R_l - c_h$ and $R_l - c_l$ given the identified risk-averse parameter α. To further pin down θ_h and θ_l, it is sufficient to normalize γ.[54]

[53] The levels of R_l and c are not seprately identified, since only their difference matters in the model.

[54] The ratio of θ_h and θ_l can be easily recovered without knowing the exact effort level. However, β and θ_t enter the model together with the effort level $e^*_{c,T}(r_T,\theta_T)$ in a nonseparable way—only $\beta e^*_{c,T}(r_T,\theta_T)$ and $\theta_T {e^*}^2_{c,T}(r_T,\theta_T)$ are identified in the model. As a result, β, θ, and the level of effort choices cannot be identified simultaneously if no further normalizations are made. Normalizing γ pins down the level of effort. The model then identifies β and the level of cost shocks, i.e., θ_h and θ_l.

Tables

Table 9: Summary Statistics

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

Table 10: Logit Regression of Default on Whether the First Closed Loan is Defaulted

VARIABLES	(1) default
first_close_loan_default	5.008***
	(0.139)
borrower_rate	9.418***
	(3.126)
amount_request	3.97e-05***
	(1.02e-05)
debt_to_income_high	0.313***
	(0.118)
Constant	-5.475***
	(0.987)
Control for Borrowers' Char.	Y
Control for Year Dummies	Y
Control for Loan Char.	Y
Observations	4,822

Note: Standard errors in parentheses, *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 11: Model Fit: Participation, Funding, Default, and Late Payment Probabilities

observables		par_prob		fund_prob		default_prob		late_prob	
		data	est	data	est	data	est	data	est
low amount, other purpose	low dti, low credit grade	0.8828	0.8861	0.8945	0.8900	0.2920	0.3284	0.1859	0.1814
	high dti, low credit grade	0.8967	0.8837	0.9092	0.9096	0.3227	0.3273	0.1581	0.1815
	low dti, high credit grade	0.8906	0.8736	0.9529	0.9522	0.1312	0.1793	0.0732	0.0775
	high dti, high credit grade	0.9179	0.8705	0.9808	0.9806	0.1976	0.1654	0.0712	0.0696
high amount, other purpose	low dti, low credit grade	0.8984	0.9035	0.8672	0.8811	0.3321	0.3049	0.1631	0.1616
	high dti, low credit grade	0.9046	0.9011	0.9103	0.9014	0.3944	0.3039	0.1665	0.1618
	low dti, high credit grade	0.8726	0.8896	0.8333	0.8726	0.1761	0.1681	0.0875	0.0702
	high dti, high credit grade	0.9151	0.8857	0.9297	0.9339	0.2377	0.1552	0.0761	0.0633
low amount, debt cons.	low dti, low credit grade	0.9346	0.9540	0.9657	0.9720	0.2844	0.2845	0.1464	0.1504
	high dti, low credit grade	0.9472	0.9528	0.9715	0.9780	0.3226	0.2840	0.1228	0.1507
	low dti, high credit grade	0.9560	0.9472	0.9892	0.9909	0.1240	0.1222	0.0555	0.0443
	high dti, high credit grade	0.9655	0.9460	0.9946	0.9930	0.1855	0.1226	0.0529	0.0447
high amount, debt cons.	low dti, low credit grade	0.9398	0.9611	0.9664	0.9584	0.2906	0.2643	0.1221	0.1340
	high dti, low credit grade	0.9557	0.9600	0.9743	0.9667	0.3369	0.2638	0.1130	0.1343
	low dti, high credit grade	0.9434	0.9520	0.9484	0.9532	0.1555	0.1158	0.0627	0.0413
	high dti, high credit grade	0.9565	0.9509	0.9710	0.9624	0.1912	0.1162	0.0592	0.0416
Weighted Average		0.9311	0.9314	0.9503	0.9523	0.2470	0.2226	0.1079	0.1099

Note: The funding probabilities are computed for the listings that are not withdrawn by borrowers. The default and late payment probabilities are computed conditional on the loans being funded.

Table 12: Counterfactual Results: Compare Welfare by Borrowers' Types

Scenarios	Outcome Variables	(θ_1, c_l)	(θ_2, c_l)	(θ_1, c_h)	(θ_2, c_h)
S1: Symmetric	Participation prob.	0.9436	0.9093	0.9337	0.8236
	Funding prob.	0.9992	0.9983	0.9991	0.9975
	Default prob.	0.0936	0.3601	0.0367	0.2067
	Late prob.	0.0112	0.1446	0.0019	0.0510
	Borrower's avg. util.	0.3409	0.0850	0.2679	-0.4007
	Lender's avg. util.	0.0840	0.0728	0.0839	0.0644
	Avg. interest rate	0.1351	0.3546	0.1018	0.1769
S2: Observe Type	Participation prob.	0.9454	0.0000	0.9335	0.8198
	Funding prob.	0.3796	0.0000	0.9991	0.9970
	Default prob.	0.2783	N/A	0.0437	0.2540
	Late prob.	0.0886	N/A	0.0027	0.0760
	Borrower's avg. util.	0.3768	N/A	0.2647	-0.3942
	Lender's avg. util.	-0.0065	N/A	0.0831	0.0676
	Avg. interest rate	0.1481	N/A	0.1054	0.1988
S3: Asy. Info.	Participation prob.	0.9437	0.9400	0.9262	0.8310
	Funding prob.	0.4984	0.4984	0.4984	0.4984
	Default prob.	0.3167	0.7904	0.0449	0.2470
	Late prob.	0.1138	0.6406	0.0029	0.0719
	Borrower's avg. util.	0.3605	0.3521	0.2142	0.1436
	Lender's avg. util.	-0.0097	-0.2174	0.1378	0.1346
	Avg. interest rate	0.1775	0.1777	0.1783	0.1742
S4: Reputation	Participation prob.	0.9374	0.9150	0.9245	0.8148
	Fund prob.	0.9987	0.9987	0.9987	0.9987
	Default prob.	0.1219	0.5100	0.0416	0.2228
	Late prob.	0.0186	0.2794	0.0025	0.0589
	Borrower's avg. util.	0.3137	0.2893	0.2203	0.1451
	Lender's avg. util.	0.1085	-0.1064	0.1307	0.1053
	Avg. interest rate	0.1837	0.1833	0.1604	0.1516
S5: Rep. + PPI	Participation prob.	0.9387	0.9216	0.9249	0.8200
	Insurance purchase prob.	0.3273	0.3273	0.3273	0.3273
	Funding prob. — w/ ins.	1.0000	1.0000	1.0000	1.0000
	Funding prob. — w/o ins.	0.9980	0.9980	0.9980	0.9980
	Default prob. — w/o ins.	0.1293	0.5311	0.0434	0.2304
	Late prob. — w/o ins.	0.0207	0.3016	0.0027	0.0626
	Borrower's avg. util.	0.3187	0.3068	0.2279	0.1547
	Lender's avg. util.	0.1154	-0.0633	0.1352	0.1067
	Avg. interest rate	0.1820	0.1820	0.1479	0.1330

Note: (1) Borrowers' and lenders' average utilities and interest rates are computed for funded loans. (2) Default and late payment probabilities are computed conditional on the loans being funded. Under the case where only types are observed by lenders, borrowers with (θ_2, c_l) are not funded, so default and late payment probabilities are not available (denoted by "N/A").