NBER WORKING PAPER SERIES

LEARNING BY DOING WITH ASYMMETRIC INFORMATION: EVIDENCE FROM PROSPER.COM

Seth M. Freedman
Ginger Zhe Jin

Working Paper 16855
http://www.nber.org/papers/w16855

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2011

We owe special thanks to Liran Einav for insightful comments and detailed suggestions on an earlier draft. We have also received constructive comments from Larry Ausubel, Robert Hampshire, John Haltiwanger, Anton Korinek, Phillip Leslie, Russel Cooper, Hongbin Cai, Jim Brickley, Estelle Cantillon, Severin Borenstein, and various seminar attendants at Rochester, Toronto, Northwestern Kellogg, Columbia, University of Maryland Smith School, 2010 NBER IO program meeting, Universiti Libre de Bruxelles, and Katholieke Universiteit Leuven. Chris Larsen, Kirk Inglis, Nancy Satoda, Reagan Murray and other Prosper personnel have provided us data support and tirelessly answered our questions about Prosper.com. Adam Weyeneth and other Prosper lenders have generously shared their prosper experience. We are grateful to the UMD Department of Economics, the Kauffman Foundation, and the Net Institute (www.netinst.org) for their generous financial support. An earlier draft has been circulated under the title "Dynamic Learning and Selection." This paper is independent of Prosper.com, all errors are our own, all rights reserved. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Learning by Doing with Asymmetric Information: Evidence from Prosper.com
Seth M. Freedman and Ginger Zhe Jin
NBER Working Paper No. 16855
March 2011
JEL No. D14,D53,D83,L15,L81

ABSTRACT

Using peer-to-peer (P2P) lending as an example, we show that learning by doing plays an important role in alleviating the information asymmetry between market players. Although the P2P platform (Prosper.com) discloses part of borrowers' credit histories, lenders face serious information problems because the market is new and subject to adverse selection relative to offline markets. We find that early lenders did not fully understand the market risk but lender learning is effective in reducing the risk over time. As a result, the market excludes more and more sub-prime borrowers and evolves towards the population served by traditional credit markets.

Seth M. Freedman
University of Michigan
School of Public Health
1415 Washington Heights
SPH-II M2224
Ann Arbor, MI 48109
sfreedm@umich.edu

Ginger Zhe Jin
University of Maryland
Department of Economics
3105 Tydings Hall
College Park, MD 20742-7211
and NBER
jin@econ.umd.edu

1 Introduction

The Internet as a platform for peer-to-peer (P2P) transactions has extended to job search, dating, social networks, and recently consumer lending. While search cost savings may explain part of the growth, it is puzzling how commodities that feature significant information asymmetry between buyers and sellers can flourish on the Internet. Take consumer lending as an example. Since the P2P platforms keep individual borrowers and lenders anonymous to each other, the extent of information asymmetry is likely to be exaggerated and Akerlof (1970) type adverse selection could be more salient online than offline. Nevertheless, P2P lending has flourished on the Internet, even in the midst of a credit crisis. To what extent and by what channels can P2P lenders overcome the information problems about borrower risk? The answer to this question is not only important for the long-run viability of these online markets, but it will also deepen our understanding on the role of information in markets.

Using transaction level data from the largest P2P lending platform in the US (Prosper.com), we show that learning by doing is an important tool for lenders in alleviating their information problems about borrower risk. Unlike traditional banks, Prosper lenders have access to only part of a borrower's credit history. The online market is so new that it is uncertain to what extent borrowers adversely select Prosper because they cannot get credit from offline lenders. Additionally, every Prosper loan is unsecured, on a fixed length of three years, and non-tradable during our data period. Borrowers were also disallowed to borrow more than one loan on Prosper until the end of our sample. These institutional features restrict a lender's ability to improve actions on a specific borrower. However, lenders may infer market-wide risk by observing outcomes of existing loans. In this sense, the learning by doing in our context is broader than private learning of idiosyncratic individual risk[2] and more similar to market-wide learning concerning the incentive to gather private information and the overall efficiency of a market.

In theory, traders that are eager to know a market-wide parameter (say the return of common stocks) may not have enough incentive to gather private information because market price may reflect all of the information that the rest of the market has gathered (Fama 1970). As shown in Grossman and Stiglitz (1980), when information is costly, such an efficient equilibrium is replaced by an equilibrium where informed traders know more than uninformed traders and price is a noisy signal of the information that the informed traders learn. Both equilibria can be reached

[2]Many empirical studies focus on the information asymmetry on a specific subject – for example, the quality of a particular used car, the risk of an insured individual (see Cohen and Sigelman 2009 for a summary), the default risk of a particular borrower (Sharpe 1990), or the ability of a particular worker (Schonberg 2007) – while assuming the less informed party has a correct understanding on the statistical distribution of the risk.

dynamically if traders start with private information, trade the asset repeatly and observe market price every round. However, the dynamic evolution is slow (Vives 1993), depends on the form of learning,[3] and does not necessarily converge to the informationally efficient equilibrium (Jun and Vives 1996). Since learning can be private, heterogeneous, slow, and divergent, it is important to document the pattern and strength of learning in a new market like P2P lending.[4]

The information to be learned on Prosper.com is the true average risk of borrowers with a specific observable attribute. Since the market is new, signals that most likely inform the risk of a prospective borrower are historical performance of similar loans. Prosper.com publicizes monthly performance of *every* existing loan, but the extent to which a lender understands such information depends on the lender's attention to his own portfolio, awareness of market-wide performance data, as well as his time cost and ability to digest the information. In theory, a lender may learn directly from his own understanding of performance data, or indirectly from market price which aggregates the understanding of historical loan performance of other lenders.[5] Since we observe *every* lender's portfolio and *every* loan's payment history on any day, we can characterize how each lender learns from the pool of loan performance signals without making explicit assumption about the informational role of market price.[6]

We find that lenders, especially those that joined Prosper early, have systematically underestimated borrower risk, even after we account for the unexpected financial crisis beginning in August 2007. But over time, lenders learn vigorously from their own mistakes. We show that a lender is more likely to stop funding any new loans as more of his existing loans are late, and conditional on funding new loans, the new loans shy away from the credit grade (or other observable attributes) of the mis-performing loans in his portfolio.

Interestingly, learning from one's own mistakes is stronger than learning from the portfolio performance of other lenders in the same social group. This suggests that part of the learning

[3]If each trader is endowed with one private signal, the common wisdom is that rational learning (with Bayesian updating) will lead to an equibrium in which the price is fully informative (Townsend 1978, Feldman 1987) but adaptive learning (learning by rule of thumb) may or may not result in such convergence (Blume and Easley 1982 and Marcet and Sargent 1988). When it is costly to acquire private information, Routledge (1999) shows that adaptive learning will lead to an asymptotic convergence to the Grossman-Stiglitz (GS) equilibrium if the process involves monotonic selection.

[4]Many experimental studies of learning focus on a textbook setting of game theory; see Cooper, Gavin and Kagel (1997) for an example of adaptive learning Salmon (2001) for a review of how economists identify reinforcement learning from belief learning in lab experiments.

[5]Note that Prosper loans are not tradable in our sample period. If a lender funds a bad loan by mistake, it does not imply any direct financial loss for those that did not fund the loan at its origination date. In this sense, our context is simpler than what is considered in the noise trader literature (De Long et al. 1990 and follow ups).

[6]To the extent that market price aggregates how other lenders process the loan performance data, it is included in our analysis of market-wide learning.

that drives the better selection of borrower risk over time is private, although newer cohorts of lenders do appear more aware of borrower risk than older cohorts. When we divide a cohort of lenders according to whether the performance of their initial portfolio is above or below the market median, we find that the below-median lenders learn faster, become more similar to the above-median group in terms of loan selection, and close the gap between the two groups after roughly 15 months on Prosper. We rule out mean reversion as the main explanation, thus this finding supports the argument that lenders are heterogeneous in information processing and such heterogeneity gradually declines over time as the less-informed parties learn more about the market-wide risk.

The above-mentioned learning has significant implications for both online and offline markets. As lenders realize the actual risk on the Internet, the P2P market has excluded more and more subprime borrowers and evolved towards the population served by traditional credit markets. This suggests that, unless P2P lenders can find innovative tools to select "diamonds in the rough," P2P lending is likely to compete head-to-head with traditional banks in the future and would not provide a viable alternative for those excluded from traditional credit markets.

Our work contributes to a number of literatures. In addition to the learning literature mentioned above, a large literature views information asymmetry as a source of market failure and argues that the information asymmetry can be alleviated by reputation, third-party certification, or collateral (Akerlof 1970, Stiglitz and Weiss 1981, and their follow-ups). We add learning by doing to this list. Unlike previous studies that document the segmentation between online and offline markets (Jin and Kato 2007, Hendel, Nevo and Ortalo-Magne 2009), we show that Prosper is *converging* with the traditional market. Our work is also complementary to a growing literature on Prosper.com (Ravina 2007, Pope and Sydnor *forthcoming*, Iyer et al. 2009, Rigbi 2008, Hampshire 2008, Freedman and Jin 2008, Lin et al. 2009), which focuses on the relationship between borrower attributes (such as race, gender, age, beauty, credit score, interest rate caps, social network affiliation) and listing outcomes (funded or not, interest rate and default).

The rest of the paper is organized as follows. Section 2 describes the background of Prosper.com and its major competitors in traditional lending. Section 3 describes the data, defines the sample, and summarizes the nature of information asymmetry for Prosper lenders. Section 4 presents basic evidence on how individual lenders learn to cope with the information problem over time. Section 5 quantifies part of the learning by analyizing internal rate of return. Section 6 explores lender heterogeneity in learning and Section 7 sheds light on the market implications of learning by doing. A short conclusion is offered in Section 8.

2 Background

2.1 Market Setup

All Prosper loans are fixed rate, unsecured, three-years in duration, and fully amortized with simple interest. Loan can range from $1,000 to $25,000. There is no penalty for early payment. As of the end of our sample period (July 31, 2008), the loans are not tradable in any financial market,[7] which means a lender that funds a loan is tied up with the loan until full payment or default. Upon default Prosper hires collection agencies and any money retrieved in collections is returned to the loan's lenders.

Before a potential borrower lists a loan application, Prosper authenticates the applicant's social security number, driver's license, and address. Prosper also pulls the borrower's credit history from Experian, which includes the borrower's credit score and historical credit information such as total number of delinquencies, current delinquencies, inquiries in the last six months, etc.[8] If the credit score falls into an allowable range, the borrower may post an eBay-style listing specifying the maximum interest rate she is willing to pay, the requested loan amount, the duration of the auction (3-10 days),[9] and whether she wants to close the listing immediately after it is fully funded (called autofunding). In the listing, the borrower may also describe herself, the purpose of the loan, the city of residence, how she intends to repay the loan, and any other information (including an image) that she feels may help fund the loan. In the same listing, Prosper will post the borrower's credit grade (computed based on credit score), home ownership status, debt-to-income ratio, and other credit history information.[10]

Like borrowers, a potential lender must provide a social security number and bank information for identity confirmation. Lenders can browse listing pages which include all of the information described above, plus information about bids placed, the percent funded, and the listing's current prevailing interest rate. To view historical market data, a lender can download a snapshot of all past Prosper records from Prosper.com (updated daily), use a Prosper tool to query desired statistics, or visit a third party website that summarizes the data. Interviews conducted at the 2008 Prosper Days Conference suggest that there is enormous heterogeneity in lender awareness of the data, ability to process the data, and intent to track the data over time.

[7] In October 2008, Prosper began the process of registering with the SEC in order to offer a secondary market, which was approved in July 2009 and therefore is outside of our sample period.

[8] The credit score reported uses the Experian ScorePLUS model, which is different from a FICO score, because it intends to better predict risks for new accounts.

[9] As of April 15, 2008 all listings have a duration of 7 days.

[10] The debt information is available from the credit bureau, but income is self-reported.

The auction process is similar to proxy bidding on eBay. A lender bids on a listing by specifying the lowest interest rate he will accept (so long as it is below the borrower's specified maximum rate) and the amount of dollars he would like to contribute (any amount above $50[11]). A listing is fully funded if the total amount bid exceeds the borrower's request. If the borrower chooses autofunding, the auction will end immediately and the borrower's maximum interest rate applies. Otherwise, the listing remains open and new bids will compete down the interest rate. Lenders with the lowest specified minimum interest rate will fund the loan and the prevailing rate is set as the minimum interest rate specified by the first lender excluded from funding the loan. We will refer to the resulting interest rate as the contract rate.

Prosper charges fees to both borrowers and lenders. These fees have changed over time, but in general borrowers pay a closing fee when their loan originates ranging from 1% to 3% depending on credit grade (there is no fee for posting a listing). If a borrower's monthly payment is 15 days late, a late fee is charged and transferred to lenders in the full amount. Lenders are charged an annual servicing fee based on the current outstanding loan principal.[12] The lender fee has ranged from 0.5% to 1% depending on credit grade. Prior to April 15, 2008, Prosper was subject to state usury laws which specify the maximum interest rate a lender can charge. The interest rate caps varied from 6% to 36% depending on the borrower's state of residence. On April 15, 2008, Prosper became a partner of WebBank, which allows the site to circumvent most state usury laws. Following this partnership, the interest rate cap became a universal 36% (except for Texas and South Dakota).

Prosper has continually changed the information that it provides lenders. At the beginning of our sample (June 2006), the credit information posted on Prosper includes debt-to-income ratio, credit grade, whether the borrower owns a home and some credit history information about delinquencies, credit lines, public records, and credit inquires. Throughout our sample time, credit grades are reported in categories, where grade AA is defined as 760 or above, A as 720-759, B as 680-719, C as 640-679, D as 600-639, E as 540-599, HR as less than 540, and NC if no credit score is available.[13] The actual numerical credit score is not available to lenders partly because of privacy protection for borrowers,[14] and partly because Prosper has promised to not reveal the numerical credit score in exchange for a deep discount on credit reports from Experian. On February 12, 2007, Prosper began posting more detailed credit information plus self reported income, employment and occupation.[15] Additionally, Prosper tightened the definition of grade

[11]After Prosper registered with SEC in July 2009, the minimum bid was reduced to $25.

[12]This fee is accrued the same way that regular interest is accrued on the loan.

[13]Prosper has refined credit grade definitions since its registration with the SEC in July 2009.

[14]If a borrower volunteers personal-identifiable information in the listing, Prosper personnel will remove such information before posting the listing.

[15]On this date, lenders were also allowed to begin asking borrowers questions and the borrowers had the option

E from 540-599 to 560-599 and grade HR from less than 540 to 520-559 eliminating borrowers that have no score or a score below 520. On October 30, 2007, Prosper began to display a Prosper-estimated rate of return on the bidding page (bidder guidance). Before this change, a lender had to visit a separate page to look for the historical performance of similar loans.[16] These information changes are likely to impact lender selection of loan risks on Prosper.

As detailed in a companion paper (Freedman and Jin 2008), Prosper also facilitates social networking through groups and friends. A non-borrowing individual may set up a group as a group leader, recruit new borrowers or lenders into the group (with a $12 reward when a group member has a loan funded), but has no legal responsibility for the payment of any group loan. A potential imbalance between member recruiting and performance monitoring prompted Prosper to discontinue the group leader reward on September 12, 2007. Starting February 12, 2007, Prosper members were allowed to invite offline friends to join the website. The inviting friend receives a reward when the new member funds ($25) or borrows her first loan ($50).[17] Group leaders, group members and friends can all provide endorsements on a related listing and their bids are highlighted on the listing page.

2.2 Offline Competitors and Macro environment

The main competitors that Prosper faces in the traditional market are credit card debt and unsecured personal loans.[18] In our sample period (June 1, 2006 to July 31, 2008), 36% of Prosper listings have mentioned credit card consolidation, which is higher than the mention of business (23%), mortgage (14%), education (21%), and family purposes (18%) such as weddings.[19] Roughly 6% of Prosper listings mentions that the Prosper loan, if funded, will be used to pay off payday loans in the offline market.[20]

---

to post the Q&A on the listing page.

[16]Prosper also introduced portfolio plans on October 30, 2007, which allow lenders to specify a criterion regarding what types of listings they would like to fund and Prosper will place their bids automatically. These portfolio plans simplified the previously existing standing orders.

[17]Existing Prosper members can become friends as well if they know each other's email address, but the monetary reward does not apply.

[18]According to Federal Reserve G.19 Statistical Release as of April 7, 2008, the total consumer outstanding (excluding mortgages) was valued at $2.54 trillion in February 2008. Within this category, $0.95 trillion was revolving debts primarily borrowed in the form of credit cards. The rest ($1.58 trillion) were non-revolving debts including loans for cars, mobile homes, education, boats, trailers, vacations, etc.

[19]69% of listings mention cars, but this at least partially a result of borrowers listing their car payments as a monthly expense.

[20]Compared to the APR of 528% that Caskey(2005) reports for payday loans, one may argue Prosper could provide a much better alternative to payday loans, given the 3-year duration of Prosper loans and the interest rate cap no higher than 36%. However, lenders must consider the credit risk they face on Prosper. If a payday lender must charge an annual interest rate of 500% to survive competition (Skiba and Tobachman 2007), it is

As shown in Appendix Figure 1, consumer lending has undergone dramatic changes during our sample period, ranging from a calm market with stable monetary policy before August 2007 to the outbreak of the subprime mortgage crisis on August 9, 2007 and gradual spillovers to other types of lending and investment. In light of this, our analysis controls for a number of daily macroeconomic variables, including the bank prime rate,[21] the TED spread,[22] the yield difference between corporate bonds rated AAA and BAA, and S&P 500 closing quotes. According to Greenlaw et al. (2008), the middle two are the strongest indicators of the subprime mortgage crisis. Additionally, we include the unemployment rate reported by the Bureau of Labor Statistics (BLS) by state and month, the housing price index reported by the Office of Federal Housing and Enterprise Oversight (OFHEO) by state and quarter, and the quarterly percentage of senior loan officers that have eased or tightened credit standards for consumer loans, and the foreclosure rate reported by Realtytrac.com by state and month.

We also control for a number of daily Prosper-specific market characteristics, including the total value of active loan requests by credit grade, the total dollar amount of submitted bids by credit grade, and the percentage of funded loans that have ever been late by credit grade. Because the financial turmoil observed in the macro environment is rooted in the subprime mortgage crisis, we control for the interaction of the OFHEO foreclosure rate and the borrower's home owner status and consumer loan easing and tightening with whether the borrower has a credit grade of E or HR. Most of the time-series variables, except for those specific to date, state or credit grade, will be absorbed in year-week fixed effects. Whenever possible, we estimate specifications with and without these fixed effects for robustness.

3 Data Description and Evidence of Information Problems

In addition to the macroeconomic indicators described above, we download two Prosper snapshots: one on August 1, 2008 and one on March 1, 2010. The 2008 snapshot provides information on all listings and loans that have appeared on Prosper through July 31, 2008. We choose this cutoff because in October 2008 Prosper started a SEC review and stopped all new listings. When it reopened in July 2009, a number of policies and market features had changed, which makes the post-July 2009 period not comparable to the previous period. However, Prosper continued to service ongoing loans during the review period, hence the Prosper snapshot from March 1, 2010 tracks all loan performance up through February 21, 2010.

For each listing created between June 1, 2006 and July 31, 2008, we observe all of the credit variables posted on the listing from Experian credit reports, the description and image

unclear why Prosper lenders would be willing to support this pool of borrowers with a much lower interest rate.

[21]Bank prime rate tracks the Fed funds rate with a 0.99 correlation.

[22]Defined as the difference between 3-month LIBOR and 3-month Treasury bills.

information that the borrower posts, and a list of auction parameters chosen by the borrower. For those listings that become loans, we observe the full payment history through February 21, 2010. For each Prosper member we observe their group affiliation and network of friends.[23] Finally, data on all Prosper bids allow us to construct each lender's portfolio on any given day. The average lender funds 36 loans worth a total of $3,345 over his lifetime on Prosper, while the median lender funds 12 loans worth a total of $850.

Excluding the few loans that were suspects of identity theft and as a result repurchased by Prosper, Table 1 summarizes listings and loans by quarter from June 1, 2006 through July 31, 2008. This sample includes 293,808 listings and 25,008 loans for $158.27 million. This implies an average funding rate of 8.51%, though this has varied over time ranging from 6.32% to 10.14%. Average listing size and average loan size both increased through the first half of 2007 and have decreased since. Comparing listings and loans, the average listing requests $7,592 and the average loan is worth $6,329. It appears that lenders are wary of listings requesting larger loans and view this as a signal of higher risk. The average listing lists a maximum borrower rate of 19.19% while the average contract rate is 17.90%.[24] This is much higher than the average interest rate for credit card accounts (13.71%).[25] or bank-issued unsecured personal loans (11.40%) as reported by the Federal Reserve as of February 2008.

Table 2 presents the funding rate, interest rate, the percent late, the percent default, and the percent 3-months late or worse (as of February 21, 2010) by the 8 credit grades observable to Prosper lenders. We define a loan as in "default" if it is four or more months late or labeled default by Prosper due to bankruptcy. As expected, a better grade means a higher funding rate, lower interest rate, and better loan performance. The last three columns attempt to compare Prosper loan performance to all the Experian accounts that had a new credit line approved in September 2003. Since the performance of Experian accounts are observed as of September 2005, we summarize the observed 2-year performance for Prosper loans for comparison. While the time horizon of Prosper and Experian loans are not exactly the same, it is clear that Prosper loans perform much worse than the traditional Experian accounts, even after we restrict the sample to borrowers with debt to income ratio less than 20% (thus more comparable to borrowers that can borrow on the traditional market). While part of the stark difference can be driven by an unexpected financial crisis and subsequent recession,[26] it also suggests that the Prosper market

[23] The data dump reflects information about groups and friends as of the download date. Because these characteristics can change over time, we use monthly downloads beginning in January 2007 to identify these characteristics at the closest possible date to the actual listing.

[24] The sharp increase in borrower maximum rates between the first and second quarters of 2008 reflects the April 2008 removal of state specific interest rate caps.

[25] Conditional on the accounts that have been assessed interest.

[26] According to the Federal Reserve, the credit card charge-off rate has increased from 4.3% in the third quarter

involves a large amount of unknown risk.

The ordinal difference in performance across grades remains salient after we run three descriptive regressions that correlate observable listing attributes to the probability of being funded (1_{funded}), the interest rate if funded ($InterestRate$), and whether the loan is default or late as of February 21, 2010 ($1_{defaultorlate}$). If a certain listing attribute (say credit grade) is a well-understood indicator of credit risk, we should see a greater funding probability, a lower interest rate, and better ex-post performance for listings of higher grades. The regression equations are as follows:

$$1_{funded,i} = f_1(ListingAttributes_i, macro, FE_{yw}) + e_{1it} \quad (1)$$

$$ContractRate_i = f_2(ListingAttributes_i, macro, FE_{yw}) + e_{2it} \quad (2)$$

$$1_{defaultorlate,it} = f_3(ListingAttributes, macroFE_{yw}, FE_a) + e_{3it} \quad (3)$$

All three regressions include year-week fixed effects (FE_{yw}) to control for the changing environment on and off Prosper. Equation (3) also includes a full set of monthly loan age dummies (FE_a) to control for the life cycle of loan performance. $ListingAttributes$ include Experian-verified credit history information, borrower-specified loan terms (e.g. amount request and maximum interest rate), borrower self-reported information (e.g. loan purpose, image, description) and social network variables (e.g. whether the borrower belongs to a Prosper group, whether the listing is endorsed and/or bid on by group leaders and friends). Summary statistics of these attributes can be found in Apendix Table 1. The funding rate and performance regressions are estimated by probits and the interest rate regression is estimated by OLS.

According to Table 3, the probability of being default or late increases by credit grade, and in response, interest rate increases and the funding probability decreases. This suggests that credit grade is an important measure of borrower risk and lenders understand these ordinal differences. Similarly, lenders understand that the more a borrower requests to borrow, the higher the risk of mis-performance, and therefore the funding rate decreases and the interest rate increases with loan size.[27] Lenders also foresee the higher risk of autofunded loans and adjust funding rate and interest rate accordingly.

In contrast, the consistency between funding rate, interest rate and loan performance fails to hold for some of the self-reported attributes. For instance, revealing a re-listing increases the funding rate but shows no difference in interest rate or ex post performance. Similar inconsistencies appear in social network variables such as group, group leader endorsement, and friend endorsement without bid. In contrast, mentioning education in a listing description implies

of 2005 to 5.5% in the second quarter of 2008.

[27] Loan size is a typical method of credit rationing (Stiglitz and Weiss 1981).

equal funding rate, equal interest rate but significantly lower probability of default or late. We take these regressions as evidence that, on average, Prosper lenders do not always predict the correct relationship between some observable characteristics and loan performance.

One explanation for why Prosper loans tend to perform worse than the Experian accounts within each credit grade is that Prosper attracts more borrowers towards the lower end of each grade (adverse selection). To detect this, we obtain from Prosper a private data set that includes the number of listings, number of loans, average contract interest rate, percent late at 6 months, and percent late at 12 months by state, month, and "half grade." Except for the two ends of the score distribution, half grades are defined as a 20-point interval of credit scores, for instance, 600-619 (referred to as D-) and 620-639 (D+). In total, we have 20 half grades, which is much more detailed than the 8 credit grades posted on Prosper.com during our sample period.[28] For comparison, this data set also includes Experian data on historical loan performance in these finer credit intervals for offline consumer loans. All half grade performance statistics are observed as of August 1, 2008.

Figure 1A compares the c.d.f. of Prosper listings, Prosper loans, the Experian population, and Experian new accounts across the 20 half grades. By Experian population, we mean all the accounts that have a score by the Experian ScorexPLUS model in December 2003.[29] As a person may have a record in Experian without demanding credit, the Experian population is an imperfect comparison for Prosper listings. The Experian new accounts are defined as above, where the credit could be secured (such as a mortgage) or unsecured (such as a credit card). Even though the Prosper vs. Experian comparison is imperfect,[30] there is no doubt that Prosper listings have a much greater concentration in lower credit intervals. Prosper lenders are able to select better risks from the listing pool, but the overall distribution of Prosper loans is still worse than that of Experian accounts.

Figures 1B and 1C present the p.d.f. of Prosper listings and Prosper loans by the 20 half grades and across time. The loan distribution is also compared with the p.d.f. of Experian new accounts as defined above. Not surprisingly, Prosper attracts listings towards the lowest end of the credit score distribution (Figure 1B) while the traditional lenders tend to focus on the highest end (Figure 1C), probably because traditional lenders cannot satisfy the credit demand

[28]The precise definitions of the 20 half grades are 300-479, 480-499, 500-519, 520-539 (HR-), 540-559 (HR+), 560-579 (E-), 580-599 (E+), 600-619 (D-), 620-639 (D+), 640-659 (C-), 660-679 (C+), 680-699 (B-), 700-719 (B+), 720-739 (A-), 740-759 (A+), 760-779 (AA-), 780-799 (AA+), 800-819, 820-839, 840-900.

[29]"Redeveloped Experian/Fair, Issac Risk Model" (December 2003) accessed at www.chasecredit.com/news/expficov2.pdf on September 5, 2008.

[30]Given the stability of credit markets before the subprime crisis and the credit crunch after August 2007, the Experian distribution is likely to overestimate the traditional credit access in 2006-2008 and therefore constitutes a conservative comparison group against Prosper.

of near or subprime risk and therefore these risks find Prosper an attractive alternative. More interestingly, the Prosper loan distribution is much less smooth than the Experian new accounts. From Figure 1C we see a higher concentration at D- than D+, C- than C+, etc. in the Prosper loans, but not in the Experian accounts. This is consistent with adverse selection towards minus grades. Also note that this pattern does not disappear over time, though the listing and loan distributions are both moving towards the right, which could be due to the credit crunch forcing near prime and prime risks to seek credit on Prosper, Prosper revealing more information hence discouraging subprime risks, or Prosper lenders learning to avoid subprime risks.

To further explore the systematic difference between minus and plus grades, we examine the population of Prosper listings and loans by half-grade (i), census division (c) and month (t)[31] Table 4 reports a set of regressions that evaluate the impact of minus grade on (1) the number of Prosper listings, (2) the number of Prosper loans, (3) the funding rate,[32] (4) the average interest rate of loans, (5) the percent late after 6 months, and (6) the percent late after 12 months, while controlling for year-month fixed effects (μ_t), credit grade fixed effects (μ_{grade}, i.e. one dummy for AA, one for A, etc.), and census division fixed effects (μ_c). Denoting dependent variables as Y, this amounts to the following regression equation in which the coefficient on the dummy of minus grade, β, tells us how minus grades differ from plus grades *within* the same grade.[33] Standard errors are clustered by census division.

$$Y_{ict} = 1_{minusgrade} \cdot \beta + \mu_c + \mu_t + \mu_{grade} + \epsilon_{ict} \tag{4}$$

Table 4 shows evidence of adverse selection consistent with the raw data: compared to plus grades, minus grades have on average 11 more listings and 2 more loans per division-grade-month. Both numbers imply a significant concentration towards minus grades as there are only 30 listings and 6 loans in each cell on average. As expected, the minus grade loans perform significantly worse. The fact that Prosper lenders do not observe credit scores explains why the funding rate is no different between minus and plus grades. However conditional on funding, lenders do charge 0.4 percentage point higher interest rates on the minus grades, which suggests that they may make some inferences as to which loans are minus grades and which are not based on other listing attribute. This is consistent with the findings in Iyer et al. (2009). However, since 93.47% (95.05%) of loans that are late by the 6th (12th) month will eventually default,

[31] We have state level data but some states have too few observations in the count of listings or loans. Aggregation into census division alleviates this problem. We have also tried aggregation into census regions, and results are similar.

[32] Which is literally the number of loans divided by the number of listings in each cell.

[33] Because the February 2007 Prosper policy disallowed any listing with credit score below 520, to facilitate comparison the regression sample excludes credit scores below 520. Results using all the "half-grade" intervals are very similar to the presented results except for the coefficient on the HR dummy.

the 0.4 percentage point higher interest rates is hardly enough to compensate the increased risk of minus grades as shown in the last two columns of Table 4 where minus grade loans are 1.4 percentage points and 2.3 percentage points more likely to be late in 6 or 12 months, respectively. The same specification on the counts of Experian new accounts by half grade finds a close-to-zero coefficient for the minus grade dummy (t=0.25).[34]

Overall, this data summary suggests that Prosper lenders understand the ordinal differences across credit grades, but some listing attributes are related to better funding rates and better interest rates without better loan performance or vice versa. Moreover, the crude definition of credit grade may have resulted in adverse selection towards minus grades. These findings suggest that Prosper lenders face significant information problems. Whether and to what extent they can overcome these problems via learning is an empirical question we will address next.

4 Basic Evidence of Lender Learning

Strictly speaking, lenders may learn from not only their own experience but also market-wide performance. As a start, we focus on the former because it is difficult to disentangle market-wide performance from other unobservable time series that affect the Prosper market at the same time. In this sense, the evidence documented below describes the extra learning that lenders obtain from their own experience *in addition to* their learning from the overall market. We will revisit market-wide learning later when we compare different lender cohorts.

We estimate a series of regressions describing how lender i's choices to fund, amount to fund, and type of loans to fund in week t respond to characteristics and performance of the lender's portfolio up through week $t-1$:

$$FundedALoan_{it} = g_1(PortChar_{it-1}, PortLate_{it-1}) + a_{1it} + \mu_{1i} + \gamma_{1t} + \epsilon_{1it} \tag{5}$$

$$AmountFunded_{it} = g_2(PortChar_{it-1}, PortLate_{it-1}) + a_{2it} + \mu_{2i} + \gamma_{2t} + \epsilon_{2it} \tag{6}$$

$$\begin{aligned} PortComp_{it} = g_3(&PortChar_{it-1}, AtoAALate_{it-1}, BtoDLate_{it-1}, EtoHRLate_{it-1}, \\ &NCLate_{it-1}) + a_{3it} + \mu_{3i} + \gamma_{3t} + \epsilon_{3it} \end{aligned} \tag{7}$$

The first equation is a linear probability model of an indicator that a lender funded at least one loan in a given week.[35] The other two equations only include the sample of lenders who funded at least one loan in week t. In Equation 6, $AmountFunded_{it}$ is the dollar amount invested by an active lender in week t. Equation 7 is run separately for various $PortComp_{it}$ variables,

[34] These findings are all robust to controlling for a polynomial function of the mid-point of each credit interval.

[35] Because we will use a large number of fixed effects, we choose a linear probability model over a probit model for this set of regressions,

which specify the percentage of an active lender's investment in AA to A, B to D, or E to HR loans in week t. $PortChar_{it-1}$ includes lender i's portfolio HHI and portfolio size through the previous week to control for time varying lender characteristics. $PortLate_{it-1}$ reflects the percentage of lender i's portfolio that has ever been late as of the previous week. $AtoAALate_{it-1}$, $BtoDLate_{it-1}$, $EtoHRLate_{it-1}$ are the percentage of lender i's portfolio through the previous week that has ever been late in each of the three respective credit grade categories.[36]

All regressions include lender, week, and lender age fixed effects, with standard errors clustered by lender. With lender fixed effects (μ_{ji}) the coefficients on the ever late variables are identified by *within* lender changes in portfolio performance and investment decisions. Note, these regressions reflect how lenders respond to both late and on-time performance because on-time payment (or early payoff) is by definition the opposite of default or late. Year-week fixed effects (γ_{jt}) controls for changes in the macroeconomic environment and the Prosper market.[37] Monthly lender age fixed effects (a_{jit})[38] capture any general pattern in lenders' choices as they age.

The results of regressions (5)-(7) are reported in Table 5. Lenders show strong responses to poorly performing loans in their portfolios. On average, a ten percentage point increase in the proportion of their portfolio that has ever been late decreases their probability of funding a loan by 0.78 percentage points in Column 1 and decreases the amount they invest in an active week by \$79.5 in Column 3. Columns 2 and 4 show that these two outcomes are sensitive to late loans in all credit grades.

One may argue that there is a mechanical relationship between portfolio performance and new investment because bad past performance implies less money available for new investments. However, this would not explain why a lender changes his portfolio *composition* in response to past performance. Columns 5-7 display the coefficients from the different versions of the $PortComp$ regressions. As lenders observe late loans, they tend to decrease their funding of loans in the grade with the adverse shock and increase their funding of higher quality grades.[39] We take these results as evidence of learning. The high late and default rates of E and HR loans have driven lenders away from these loans and toward higher credit grades as lenders have learned about the dangers of investing in these lower credit grades. It is possible that a lender

[36]We have also tried specifications using the percent of a lender's portfolio (in total or in various categories) that is currently late or in default and the results are very similar.

[37]Results of identical regressions with controls for macro variables and Prosper supply, demand, and market performance instead of week fixed effects are very similar.

[38]We count a lender as joining Prosper when he funds his first loan, and age is defined as weeks since joining Prosper.

[39]Note that when lenders observe late AA to A loans, they do show slight substitution towards the lower credit grade loans.

with less wealth becomes more risk averse and therefore invests in safer credit grades. This is unlikely the driving force, because the amount a typical lender invests on Prosper ($850 at the median) is small as compared to the median household income in the US ($52,175 according to the 2006-2008 American Community Survey).

In results not shown here, we observe similar learning patterns when we use regressions to describe the propensity to fund loans in other categories (including autofunded loans, loans of various sizes, and loans affiliated with specific types of social networks) as a function of late loans in these categories. These results suggest that in this new market, lenders attempt to learn the meaning of many listing attributes, even though some of them have been well understood in traditional off line markets.

Above all, we find evidence that lenders learn from their own portofolio performance on two margins: on the extensive margin, a greater percentage of default or late existing loans triggers less new investment; on the intensive margin, conditional on funding new loans, a lender tends to avoid the listing attributes that led to bad performance in his portofolio and prefer the attributes that led to good performance. Given the large number of listing attributes lenders observe, the next section summarizes a lender's loan choice through one number – internal rate of return (IRR).

5 Measuring the Extent of Learning with IRR

We compute the internal rate of return (IRR) that a sophisticated lender should expect from a Prosper loan as he considers all of the information at the time of the listing and projects loan performance throughout the 36-month loan life. If a lender initially underestimates the risk of a loan with certain attributes (say grade HR) but later learns to either charge a higher contract rate on a similar loan or fund a better-grade loan, this process can be summarized as the IRR improvement from old to new loans. We emphasize that our goal in calculating IRR is not to quantify the absolute level of performance of Prosper loans, but instead to obtain a summary measure that ranks loans by the relationship between their observable characteristics and performance taking interest rate into account.[40]

To calculate IRR, we first use the observed ex post loan performance to predict a relationship between listing attributes and loan performance, and then calculate an annual discount factor (call it R) that equalizes the loan amount to the present value of all the predicted monthly cash inflows. Compounding R monthly, $IRR = (1 + R/12)^{12} - 1$ reflects the annual percentage yield

[40]In finance, a popular measure of asset risk β is the correlation of its return with that of the financial market as a whole. It is impossible to compute β per loan because Prosper loans were not traded after origination in our sample period.

from the loan. We believe this method captures the rate of return that a sophisticated lender *expects* to earn at the start of the loan if he can perfectly predict the statistical distribution of loan performance.

We use four dummies to measure loan performance: default, default or late, missed payment, and early pay off. Located between the most optimistic (default) and the most pessimistic (default or late), the dummy of missed payment is defined as one if the loan's payment history indicates that the borrower has missed the payment in a specific month. If the borrower misses the payment at month t but makes it up in a later month, we count it as not missing the payment. Early pay off is treated as a bulk of cash flow in the actual month of payment and zero afterwards. This implicitly assumes that the early payoff is reinvested into a loan that is identical to the loan under study.[41] While in reality the payment history can be very complicated, we simplify the predicted cash inflow for loan i at month m as the sum of (1) the probability of early pay off at m times the principal remaining plus interest at m and (2) the probability of on-time payment at m times the monthly payment, net of the lender fee. Monthly payment and principal remaining are computed according to the 3-year amortization table, and probability of on-time payment is computed as one minus the probability of mis-performance and the probability of having paid off before or at month m.

Before we use loan attributes to predict loan performance, it is worth noting that the macroeconomic environment has changed substantially due to the worldwide financial crisis that even the most sophisticated loan officer may not have anticipated. If we do not isolate macroeconomic changes from the realized loan performance, we may mistakenly attribute lender adjustment in response to the unexpected macroeconomic shock as a form of learning that overcomes lender misunderstanding of borrower attribute. To address this problem, we include the realized macroeconomic variables on the right hand side when we predict loan performance. Denoting $Perf_{it}$ as the performance of loan i at calendar month t, $grade_i$ as i's credit grade, X_i as other loan attributes, α_m as loan age fixed effects (by month) and $realmacro_t$ as the realized macroeconomic environment, we estimate the below specification by probit. We choose probit over a duration model because probit allows an event (such as missed payment) to switch on and off over time, and because probit yields fewer prediction errors as compared to a duration model. The Appendix reports results using a duration model and an alternative probit specification.

$$Perf_{it} = \alpha_m + \beta_1 \cdot grade_i + \beta_2 \cdot X_i + \beta_3 \cdot realmacro_t + \beta_4 \cdot grade_i \cdot realmacro_t + \epsilon_{it} \tag{8}$$

[41] The assumption could lead to an over- or under-estimate for the return on investment in a given period. One way to overcome this problem is assuming the reinvestment rate equal to a specific average cost of capital. Here we do not use modified IRR, partly because any choice of the reinvestment rate is arbitrary, and partly because this study is not meant to be an investment guide. We emphasize performance comparison across loans, not whether Prosper loans are financially worth investing in a fixed time period.

A subsequent problem arises in prediction: if lenders did not anticipate the realized macroeconomic shock, what kind of macroeconomic environment did they anticipate? How can we separate the change in lenders' macroeconomic anticipation from the change of their fundamental understanding of borrower attributes? To address these questions, we construct two series of macroeconomic forecasts: one is the quarterly forecast that a lender would have forecast as of June 1, 2006 (the beginning of our sample) given all of the macroeconomic information available then. We make this forecast by fitting a vector autoregressive model from 1997 through the second quarter of 2006.[42] Denote this forecast as $\hat{macro}_m^{6/1/06}$ for loan age m. By construction, it is not updated for later loans as more macro information accumulates over time. In contrast, in the other macro forecast – denoted as $\hat{macro}_{im}^{rolling}$ – we make the forecast on a rolling basis so that a loan (i) funded at calendar month m corresponds to the quarterly forecast that a lender would have made given the available macroeconomic information from 1997 through that quarter. With parameter estimates from the above probit specification (which utilize the actual realized macroeconomic variables in estimation), we have two versions of predicted performance for a loan at age $m = 1, ...36$:

$$
\begin{aligned}
(9)\ \hat{Perf}_{im}^{6/1/06} &= \hat{\alpha}_m + \hat{\beta}_1 \cdot grade_i + \hat{\beta}_2 \cdot X_i + \hat{\beta}_3 \cdot \hat{macro}_m^{6/1/06} + \hat{\beta}_4 \cdot grade_i \cdot \hat{macro}_m^{6/1/06} \\
(10)\hat{Perf}_{im}^{rolling} &= \hat{\alpha}_m + \hat{\beta}_1 \cdot grade_i + \hat{\beta}_2 \cdot X_i + \hat{\beta}_3 \cdot \hat{macro}_{im}^{rolling} + \hat{\beta}_4 \cdot grade_i \cdot \hat{macro}_{im}^{rolling}
\end{aligned}
$$

The rolling forecast better captures the lender's prediction at the time of funding, but an IRR change in a lender's portfolio (based on $\hat{macro}^{rolling}$) reflects both the changes in the rolling forecast and the changes in borrower attributes between new and old loans. In comparison, fixing the macro forecast as of June 1, 2006, changes in the IRR estimate (based on $\hat{macro}^{6/1/06}$) will only reflect changes in loan attributes, which we interpret as a change in lender's understanding of loan characteristics independent of macroeconomic shocks.

For comparison, we also construct $IRR_{nomacro}$ based on a probit specification that excludes all the macroeconomic variables, and $IRR_{realmacro}$ based on the same probit as above but using the realized macroeconomic environment as the prediction. $IRR_{nomacro}$ assumes a sophisticated lender has perfect foresight on the realized distribution of loan performance and attributes all of it to loan attributes. In comparison, $IRR_{realmacro}$ assumes a sophisticated lender has perfect foresight on both the macroeconomic shocks and loan performance distribution separately. Since IRR calculation requires performance prediction from month 1 to month 36, we can only estimate $IRR_{realmacro}$ for loans that have matured as of February 21, 2010 while $IRR_{nomacro}$, $IRR_{forecast6/01/06}$ and $IRR_{rollingforecast}$ can be estimated for all loans in our sample.

Table 6 summarizes 12 versions of IRR estimates depending on which misperformance mea-

[42]The model includes a two and four quarter lag and four quarter fixed effects to account for seasonality.

sure we use and how we treat the macroeconomic forecast. In theory, the present value formula is monotone and should have a unique solution of R that is bounded between -1200% and the contract rate. After monthly compounding, IRR is bounded between -100% and $(1 + \text{contract rate}/12)^{12} - 1$. In practice, we do achieve over 99% of convergence if we do not impose any constraint on IRR. However, since we predict the likelihood of early payoff and misperformance separately,[43] there is a small chance (less than 10%) that the sum of the estimated likelihood is over one in at least one of the 36 months, hence the converged IRR can exceed the theoretical upper bounds. To address this issue, the IRRs reported in Table 6 are estimated with the imposed constraint that they cannot lie outside their theoretical bounds. All of the loan comparisons reported below are robust if we focus on the loans whose unconstrained IRRs do not exceed the bound.

We present summary statistics of these IRR measures at three levels: the individual loan level, the dollar level (weighting IRR by the dollar value of each loan), and the lender level (averaging across lender portfolios). At all three levels – loan, lender or dollar – the comparison of IRR1 to IRR12 (conditional on convergence) is consistent with expectation. For example, using default as the misperformance measure yields higher IRRs than using missed payment or default or late. The average IRR using missed payment is only slightly higher than that of default or late, consistent with the fact that many borrowers that miss payments do not catch up later. The IRR estimates using the realized macroeconomic variables are significantly more negative than other versions using macroeconomic forecasts, partly because the loans in the sample using real macroeconomic values are early loans, and partly because the real macroeconomic version unrealistically assumes lenders have perfect foresight on future macroeconomic shocks. If we focus on the loans that matured before February 21, 2010, the average $IRR_{realmacro}$ is only slightly lower than the average $IRR_{rollingforecast}$ and $IRR_{forecast6/1/06}$. This suggests that unexpected macroeconomic shocks only explain a small fraction of $IRR_{realmacro}$.[44]

As shown in Figure 2, the average $IRR_{rollingforecast}$ is systematically higher than the average $IRR_{forecast6/1/06}$ over time. This suggests that lenders become more optimistic about the macroeconomic environment between June 2006 and July 2008. This is plausible because the financial crisis did not strike the wholesale financial market until August 2007 and it was unclear how soon this crisis would affect the payment ability of Prosper borrowers. $IRR_{rollingforecast}$ does decline after the beginning of 2008, as lenders became less optimistic after the crisis broke out. As mentioned above, $IRR_{forecast6/1/06}$ keeps the macroeconomic forecast fixed at what lenders would have predicted as of June 1, 2006, hence changes of $IRR_{forecast6/1/06}$ only re-

[43]The joint estimation takes an extremely long time and does not yield stable results.

[44]Note that we can only compute $IRR_{realmacro}$ for the earliest cohorts of loans but the macroeconomic shock could have a bigger effect on later cohorts.

flect changes in loan attributes. Because default tends to underestimate loan misperformance and default or late overestimates it, we focus on IRR8, the $IRR_{forecast6/1/06}$ that uses missed payment as the misperformance measure.

According to Table 6, average IRR8 does not differ much across loan (-4.13%), lender (-3.75%) and dollar (-4.03%) levels, suggesting that big lenders and big loans do not earn significantly higher returns than small lenders or small loans. For comparison, the average annual yields of 3-year Treasury Bill and S&P 500 are 3.97% and -0.66% in the same period (June 1, 2006 to July 31, 2008). While part of the low $IRR_{forecast6/1/06}$ can be attributed to lenders expecting better macroeconomic conditions than those of June 1, 2006, the average $IRR_{rollingforecast}$ per loan (0.18%) is still lower than the return of treasury bills or savings account. While we do not emphasize this absolute level, it appears that lenders may have systematically underestimated borrower risk; it is also possible that lenders willingly lend on Prosper for charity or for fun.

Figure 3 presents the kernel density of loan-specific IRR8 by credit grade. Not surprisingly, grades E-HR have the longest left tail and the lowest average IRR8 (-14.38%). Interestingly, the average IRR8 of B-D loans (0.52%) is higher than that of AA-A (-2.39%), probably because borrowers of B-D grades often specify higher maximum interest rates and lenders do not compete down the interest rate of B-D as much as AA-A. Figure 4 plots how IRR8 changes over time by grades, and interestingly the IRR8 within the E-HR category increases over time. This figure and the above-mentioned learning evidence suggest that the marketwide improvement of IRR8 is driven by lenders switching towards better grades and picking better performing loans within the E-HR category.

If lender mistakes are the main explanation for the low IRRs, lenders should learn to choose better loans when they observe their previously funded loans perform poorly. To quantify learning on such an intensive margin, we rerun the learning equation 5 but redefine the dependent variable as the average IRR8 of the loans that lender i funded in week t.

$$AvgIRR8_{it} = g_1(PortChar_{it-1}, PortLate_{it-1}) + a_{1it} + \mu_{1i} + \gamma_{1t} + \epsilon_{1it} \tag{11}$$

The regression results are reported in the last two columns of Table 5. The coefficient in the second to last column suggests that, when the average lender sees a ten percentage point increase in the portion of his portfolio that has been late, his newly funded loans have a 1.98 percentage point higher IRR8. This is a large improvement: by the end of our sample period (July 2008), 39.8% of the loans originated at the beginning of the sample (June 2006) are either late or default. According to the regression, this performance alone would have motivated a 7.88 percentage point improvement in IRR8, a magnitude that is comparable to the largest marketwide improvement of IRR8 (6.51 percentage points) in our sample period. More specifically, Figure 5 plots by loan origination month (1) the average IRR8 per loan, (2) the average contract rate, and (3) the predicted probability of missed payment in month 6. Over time, the average IRR8 first declines from -4.10% in June 2006 to -9.85% in November 2006 and then gradually increases to -3.34% in July 2008. We suspect the initial decline of IRR8 is because Prosper lenders underestimated the market risk, which therefore attracted a large flow of high risk borrowers to enter the market. As the risk unfolded, both lender learning and the more transparent information from Prosper could have contributed to the IRR8 improvement after November 2006. Interestingly, the marketwide IRR8 improvement is mostly driven by lower missed payment rates rather than higher interest rates.

The result that lenders choose loans with higher IRR8 in response to past late loans in their portfolios is consistent with the earlier result that lenders switch towards better credit grades after they observe misperformance in their own portofolio. Additionally, the last column of Table 5 shows that lenders choose loans of higher IRR8 in response to late loans in all four credit grade categories. In addition to substituting *between* credit grades, lenders may choose better performing loans *within* credit grades as they learn. To test this we regress the average IRR8 of loans a lender funds in a given week within each credit grade on the percent late variables across all grade groups defined above. As shown in Appendix Table 2, the results reveal that, except within the AA grade, lenders fund higher IRR8 loans within each credit grade in response to past late loans, and this response is usually concentrated to late loans within the corresponding coarse grade group. This suggests that lender learning is not limited to switching from low credit grades to high credit grades (as shown in Table 5); they also learn to better interpret other listing attributes within each credit grade.

Combining Sections 4 and 5, evidence suggests that lenders learn from their own mistakes in two ways: when their existing loans perform poorly, they fund fewer new loans and the new loans that they fund have observable attributes that predict a higher internal rates of return.

6 Heterogeneous Learning

The literature led by Grossman and Stiglitz (1980) argues that in a market with uncertainty market efficiency depends on not only the degree of uncertainty but also on how different traders acquire or process information to address this uncertainty. It is clear that there is a great deal of heterogeneity across lenders on Prosper. For example, lenders differ greatly in the rate of return of their portfolio: while the mean of per lender IRR8 is -3.75%, it ranges from -99.9% to 32.9% with a standard deviation of 7.0%. If lenders differ in their ability to screen loans and their ability to learn from their mistakes, competition will not lead to equalized expected returns across funded loans. In this section we document significant gaps in lender ability and how learning acts to close these gaps.

In particular, we separate lenders into two groups based on their initial "sophistication." We first calculate each lender's average portfolio IRR8 for the loans chosen during their first month on Prosper and then label a lender "sophisticated" if his first month portfolio IRR8 is above his cohort's median (lender cohort is defined by the week in which a lender joined Prosper).[45] Recall that our measure of IRR8 is based on the *expected* return associated with a given loan's observable characteristics. We split lenders by whether or not they choose loans with observable characteristics that would predict good performance, not whether or not they choose loans that result in good performance *ex post*.

As pointed out by Hotelling (1933), this type of classification is subject to mean reversion. To the extent that lenders misunderstand and/or ignore observable loan characteristics, some lenders may choose good loans not by sophistication but by chance. To mitigate this concern, Table 7 shows that above- and below-median lenders differ systematically in their inital lending strategies. At the end of their first month on Prosper, above-median lenders have on average smaller portfolios, composed of a slightly higher number of loans, and therefore a slightly lower average loan size.[46] More strikingly, these two groups have a much different credit grade distribution with above-median lenders much more likely to choose AA-A and B-D loans at the median and much less likely to invest in E-HR loans.

Additionally, if luck were driving the differences between the two groups (and assuming luck is independent across time and lenders), we would expect, by the law of large numbers, the average IRR8 of the two groups to converge *immediately* after the first month. Figure 6 plots the average new portfolio IRR8 by lender age (in weeks) for the two groups. There appears to be some evidence of mean reversion as average IRR of the sophisticated group drops and the average IRR of the unsophisticated group increases over the first few weeks. However, after this discrete change, a gap remains and then closes gradually. This pattern suggests that the initial gap is not simply due to luck, and that the subsequent convergence is likely due to behavioral changes over the lender's life cycle. This gradual convergence is consistent with learning as the below-median lenders respond to late loans in their portfolio and pick future loans with better observable characteristics. The subsequent curves suggest that learning allows the below-median type to catch up with the above-median type by roughly the 80th week (or the 20th month).

To further support learning as a likely explanation for the eventual convergence, we plot

[45] In all subsequent analysis that separates above- and below-median lenders, we condition the sample on cohorts after June 1, 2006 to focus on lenders who *joined* Prosper during our sample period and because the earlier weekly cohorts are quite small (74 lenders on average compared to 399 for later cohorts) making the above/below-median split less reliable.

[46] These differences are all statistically significant at the 10% level, although the medians are much more similar across groups.

how above- and below-median lenders differ in the grade composition of newly funded loans as a function of a lender's Prosper age (Figure 7 for grades AA-A and Figure 8 for E-HR). Consistent with the within-lender learning regressions, both above- and below-median types increase investment in grades AA-A and decrease investment on E-HR as they age. However, there is a significant gap between the two types until the 80th week: before they converge, above-median lenders always invest more in AA-A and less in E-HR. This suggests that both the inital and subsequent rate of return gap between the two types of lenders as observed in Figure 6 is most likely driven by systematic difference in loan choice instead of luck.

To formalize the above graphs in a regression framework, we first define the unit of observation as lender i of cohort c in type p at age a, where cohort and age are both measured in months. We then regress the average IRR8 of all loans funded by lender i, in group p at age a on cohort fixed effects, age fixed effects, week fixed effects,[47] a dummy indicating the "above-median" type ($1_{abovemedian}$) and interactions of $1_{abovemedian}$ with a linear cohort term, a linear age term and a separate dummy for the second month:

$$\begin{aligned} AverageIRR8_{icpa} = & \mu_{cohort} + \gamma_{age} + \beta_1 \cdot 1_{abovemedian} + \beta_2 \cdot 1_{abovemedian} \cdot cohort \\ & + \beta_3 \cdot 1_{abovemedian} \cdot month2 + \beta_4 \cdot 1_{abovemedian} \cdot age \\ & + \beta_5 \cdot 1_{abovemedian} \cdot cohort \cdot age + FE_{wk} + \epsilon_{cga}. \end{aligned} \tag{12}$$

The coefficient of $1_{abovemedian} \cdot age$ captures the gradual change in the gap between above- and below-median types, while the coefficient of $1_{abovemedian} \cdot month2$ captures the initial shrinkage of the gap due to potential mean reversion. As shown in Table 8, we find $\beta_1 = 0.084$, which indicates that the first-month portfolio of "above-median" lenders has an average IRR 8.4 percentage points higher than the "below-median" lenders. The estimates of β_2 (-0.001) and β_4 (-0.004) suggest that later cohorts are systematically more homogeneous and the IRR difference between above- and below-median lenders declines steadily within each cohort as "below-median" lenders learn. Interestingly, β_4 is much larger than β_2 in magnitude, suggesting that learning by doing is more effective in reducing lender heterogeneity within a cohort than factors that reduce heterogeneity across cohorts, an observation we will return to in the next section. Additionally, the estimate of β_3 (-0.026) suggests that mean reversion explains at most 31% of the initial IRR8 difference between above- and below-median lenders. Taken together, these results suggest significant heterogeneity in the initial level of lender sophistication. The estimate of β_5 is very small (-0.00018), suggesting that the relatively less sophisticated lenders in the later cohorts learn almost as much as the earlier cohorts, even if they are more similar to the sophisticated lenders when they started lending on Prosper. Table 8 also shows results of similar regressions

[47] Cohort, age and week can be simultaneously identified because cohort and age are by month instead of week. Results barely change if we ignore week fixed effects.

where the dependent variable is the fraction of the lender's loans invested in different credit grade categories in a given week. Consistent with Figures 7 and 8, below-median lenders initially fund more E to HR loans and less A to AA loans, but over time their portfolio composition becomes more similar to above-median lenders.

Following Hotelling (1933), further evidence of these initial behavioral differences and subsequent changes can be seen in Appendix Figures 10 and 11. These figures show that the across lender variance in IRR8 decreases as lenders age, suggesting that the convergence of IRR8 is due to lenders choosing more similar loans as they age. Furthermore, below-median lenders are initally a much more heterogenious group with higher variance than the above-median lenders. As they age, both groups become more homogenous, and the variance of below-median lenders eventually converges with the above-median group.

7 Market-Wide Implications

To this point, we have presented evidence that (1) lenders actively respond to the performance of their own portfolios on top of market-wide fluctuation; (2) there is a large amount of heterogeneity within each lender cohort, most of which is driven by relatively more sophisticated lenders selecting loans with better observables; (3) over time, the less sophisticated lenders gradually catch up with the sophisticated lenders in the same cohort, (4) later lender cohorts show less heterogeneity than earlier cohorts, and (5) learning by doing is more effective in reducing lender heterogeneity within a cohort than factors that decrease heterogeneity across cohorts.

While these facts highlight the importance of learning by doing, it also raises more fundamental questions regarding lender information and the overall market evolution. For example, do lenders' responses to their own portfolio imply that they ignore market wide information? Do some lenders over-react? Why are some lenders more sophisticated than others when they join Prosper? Would lenders obtain better information by watching the market before lending themselves? Taking the market as a whole, how does lender sophistication evolve over time and what does it imply for market evolution? Some of these questions are difficult to answer, but a combination of existing and further evidence sheds some light on them.

7.1 Over-reaction?

In a model where price is the only observable market-wide statistic that aggregates private information, there should be a negative relationship between price informativeness and sensitivity to private information. However, in our context, reaction to one's own portfolio could reflect many possibilities, including ignorance of market wide information (to be discussed in the next subsection), noise in market-wide information, lender taste deviating from the market average, and over-reaction.

Empirically, it is difficult to identify over-reaction because we do not know what the correct reaction should be. That being said, we can compare how above- and below-median lenders differ in their investment choices over time. If less sophisticated lenders were too optimistic in funding loans of lower grades (say E-HR) but over-react when these E-HR loans become late or default, we should observe their earlier portfolios to have too many E-HR loans but their later investments to have too few E-HR loans, both relative to the above-median lenders from the same lender cohort. Figures 7 and 8 (and corresponding regression results in Table 8) clearly reject this speculation: the below-median type does invest more in E-HR loans initially and shies away from them later on, but they never devote a significantly smaller proportion of their investment to E-HR than the above-median type. Nor do they invest disproportionally more in A-AA grades. Note that the comparison is only relative: data suggest that below-median lenders do not over-react *more than* above-median lenders; but it is still possible that above-median lenders have over-reacted to either market wide performance or the outcome of their own portfolios.

7.2 Market-wide learning

Since the market-wide performance of Prosper is simply a time-series, it is impossible to disentangle learning of market-wide performance from the impact of other macroeconomic factors. That being said, if we can isolate some part of market performance that seems more relevant to a lender, we can examine how the lender responds to this part of information in addition to market-wide performance or unobservable macroeconomic factors.

Here, we examine whether a lender that belongs to group g adjusts his investment in response to the performance of the loans funded by other lenders in the same group. This test does not formally identify learning from the market, but it does show whether lenders learn from loan performance at a more aggregate level than their own portfolios. More specifically, for lender i at week t, we calculate the percent late for all the loans funded by his group up to week $t-1$ (excluding i's own portfolio) and add this $GroupLate_{g,t-1}$ variable on the right hand side of Equations 5, 6, and 11. In Panel A of Table 9 the coefficient on this group portfolio performance measure is negative and statistically significant in the funding regression, but small and not statistically significant in the regressions describing the amount funded and IRR of funded loans. In addition the own portfolio percent late coefficients are very similar to those initially reported in Table 5 for the latter two regressions. This suggests that, in addition to the marketwide fluctuation (controlled for by week fixed effects), an average group lender does learn on the extensive margin (i.e. propensity to fund new loans) from the performance of other loans funded by his group's members, but conditional on funding, does not change the types of loans funded. While the lack of learning from the group's portfolio in terms of IRR could be explained by a lender viewing group loans as charity lending, results reported in Freedman and Jin (2008) suggest that lenders shy away from borrowers with group affilation when previously funded group loans misperform, both within their own groups and across all group affiliated borrowers.

Our second test to infer market wide learning is to examine whether later cohorts of lenders respond less to the percent late of their own loans than do earlier cohorts. If later cohorts have learned from the market performance that occurred before they joined Prosper, they would have a more precise prior about the meaning of borrower attributes and therefore adjust their posterior belief less when their own portfolios misperform. To identify this effect, we interact $PortLate_{it-1}$ with cohort dummies in Equations 5, 6, and 11. Here cohorts are defined in six month periods. While our sample only includes lending activity after June 2006, it still includes lenders who joined Prosper prior to this date. Therefore, the first cohort, which is the reference group, contains those lenders who began lending on Prosper during the first half of 2006. As shown in Panel B of Table 9, the coefficients on the cohort interaction terms are all of opposite signs to the main effect. This pattern implies that lenders in latter cohorts respond less to the late loans in their portfolios as compared to the first cohort on both the extensive funding margin and the intensive margin regarding the quality of their chosen loans. One caveat in interpreting these results is that the latest cohorts have had less opportunity to observe late loans in their portfolios, which could lead to the decreased relationship between late loans and lending behavior, particularly the large negative coefficient on the interaction between percent late and the last cohort (the second half of 2008) in the IRR regression.

To track across- and within-cohort changes, Figure 9 plots the average lender's IRR8 for loans he funds in a given week by lender cohort. Here lender cohort is defined by the quarter in which the lender funded his first loan. Consistent with the learning evidence, as lenders age they fund loans with a higher rate of return. Moreover, new cohorts pick up the market trend, perhaps responding to information revealed by the market that was not available when older lenders joined Prosper. To push this point further, recall in Table 8 that the gap between above- and below-median lenders is smaller for later cohorts. This suggests that new lenders are not a simple replicate of the old cohorts; new cohorts have less heterogeneity, potentially because they learn some information from prior market performance before funding their first loans.

Overall, the evidence suggests that lenders have watched the market before joining Prosper but once they are on Prosper they respond to the performance of their own portfolio in addition to signals from their own group or the whole market.

7.3 Better watching than doing?

If for behavioral or psychological reasons Prosper lenders are biased by the outcomes of their own portfolios, watching the market (without lending) may allow a prospective lender to better digest market-wide information and obtain better information than active lenders. Conversely, if it is easier to learn from one's own portfolio performance than following a continuous update of market-wide information, learning by doing may be more informative than watching the market.[48] We try to distinguish these two stories by comparing different cohorts at a snapshot in time which includes a distribution of new and old cohorts. Assuming the macroeconomic environment affects them equally, if the new cohort on average makes better choice of loans than the old cohorts, it suggests that watching the market may be more informative than learning by doing.

Specifically, we regress lender i's average IRR8 of all the loans he funded in week t as a function of whether this lender is new at t (i.e. has been a Prosper lender for less than a month and therefore has not observed any performance of his own portfolio), the lender's age (in months) if he is not new, and week fixed effects.

$$AvgIRR8_{it} = \beta_1 \cdot New_{it} + \beta_2 \cdot Age_{it} + FE_t + \epsilon_{it} \tag{13}$$

As shown in Table 10, if we do not control for the age of old cohorts, the coefficient of the new dummy suggests that new lenders on average pick loans that have 0.31 percentage points higher expected rate of return than the average old lender at the same point in time. This suggests that, on average, a new lender watching the market has better information on the meaning of loan attributes than those who learn by doing on Prosper. Note that this could be driven by a selection effect: those who join Prosper later may be more sophisticated than earlier lenders because they are endowed with better information about consumer lending before observing any Prosper activities, or because they have a better ability to extract useful information from Prosper data. Either way, our results do not necessarily imply that the *same* lender would do better by joining Prosper in a later period.

The second column of Table 10 reports both coefficients of New and Age. The coefficient of New remains significant but becomes larger in magnitude (0.74 percentage points vs. 0.31 percentage points), while the coefficient of Age is positive and signficant (0.06 percentage points). These results suggest that watching the market does not always dominate learning by doing: new cohorts pick better loans than existing young cohorts, but when a previous cohort becomes older than 11 months, it may do as well or even better than the new cohort.

[48] If lenders learn about their own preference say risk tolerance) instead of the meaning of loan attributes, learning by doing could be more informative as well.

7.4 Lender evolution

What does the new-versus-old comparison say about the nature of lender heterogeneity and lender evolution? To answer this question, it is necessary to combine the previous result with an intriguing fact mentioned above: Table 8 suggests that the initial gap of IRR8 between above- and below-median lenders in the same cohort will converge to zero in (0.084-0.026)/0.004=14.5 months; by comparison it takes (0.084-0.026)/0.001=58 cohorts for this initial gap to become zero. In other words, below-median lenders become more similar to the above-median lenders within a cohort than they do across cohorts over the same time frame. What patterns of heterogeneity and learning would lead to these facts?

We suspect the reality is that there is a continuum of lender types. Some lenders choose to watch the market first so that they can process more market-wide information before lending money on Prosper. If lenders who spend more time researching the market before lending perform better, new above-median lenders may be on average more sophisticated than old above-median lenders that have aged up to the same calender time and remain active (in funding new loans). To see this, suppose the new above-median lenders were not as sophisticated as an aged above-median lender. To be consistent with the fact that the above/below difference of IRR8 is larger in the new cohort than in an *aged* old cohort, the average IRR8 of the new cohort must be lower than the aged old cohort. However, this contradicts the fact that new lenders on average do better than old cohorts at a specific time. This logic suggests that new above-median lenders must be more sophisticated initially than old above-median lenders on average.

By comparison, the relatively unsophisticated lenders may have difficulty digesting market-wide information but learn from experience. In Panel C of Table 9 we rerun Equations 5, 6, and 11 including an interaction between $1_{abovemedian}$ and $PortLate_{i(t-1)}$. As expected, the above-median type responds significantly less to the lateness of their own portfolio than the below-median lenders in terms of the IRR of funded loans. To the extent that learning by doing is more salient for below-median lenders, aged below-median lenders that remain active on Prosper could become more sophisticated than new below-median lenders. In the meantime, the difficulty of processing market-wide information keeps the new below-median lenders at low sophistication, which explains why the above/below-median difference of IRR8 declines very slowly across cohorts. Interestingly, the regression suggests that above-median lenders respond more on the funding margin than the below-median lenders if they observe the *same* amount of $PortLate$. In reality, above-median lenders observe systematically less default or late in their own portfolios (because of their sophistication and reluctance to invest a lot in the beginning) and therefore slightly surpass below-median lenders in the likelihood of funding any new loans after week 16 (Appendix Figure 12).

8 Conclusion

We examine how online lenders in a peer-to-peer lending market cope with information asymmetry that is likely to be exaggerated on the Internet. Evidence suggests that individual lenders on Prosper.com do face serious information problems because they do not observe borrowers' complete credit history and Prosper borrowers may have adversely selected the website as they have difficulty obtaining credit in offline markets.

Ex-post performance data suggest that many lenders fund loans of low expected returns and they learn to shy away from risky loans over time. While we cannot rule out charity or fun as potential motivations, the most likely explanation is that lenders, especially those who joined Prosper early, lacked expertise in risk evaluation. Learning by doing plays an important role in addressing the problem, but there remains a large amount of lender heterogeneity. Consistent with the literature of inefficient markets, we observe that some lenders are better informed than others. We find that the gap between more and less sophisticated lenders closes gradually over time, both within cohorts and across cohorts. Because convergence within cohorts is faster than the convergence across cohorts, there is still a wide range of heterogeneity in expected returns at the end of our sample.

Our findings are not inconsistent with P2P lending's rapid growth on the Interenet. We suspect part of the observed prosperity was driven by unsophisticated lenders who drastically underestimated borrower risk on the Internet, which in turn has attracted risky borrowers to populate the website. As lenders learn the actual risk, the initial mistake-driven prosperity would not be sustainable in the long run. This explains why over time the P2P market has excluded more and more subprime borrowers and evolved towards the population served by traditional credit markets. In fact, at the end of our sample period, the funding rate of risky listings had become so low that when Prosper reopened in July 2009 (after the SEC review) it began dissallowing any borrower with a credit score below 640 (i.e. grade D or below) to list on the website.[49] The extent to which P2P lending can compete with traditional banks for prime borrowers remains an open question. P2P lending may have advantages in terms of search cost savings or online social networks. We have examined the latter in Freedman and Jin (2008) and will consider the former in future research.

9 References

Akerlof, George A. (1970): "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism" *Quarterly Journal of Economics* 84(3): 488-500.

[49]In July 2009, Prosper also refined the definition of credit grades to be based on predicted rates of return instead of coarse credit score intervals. This change may alleviate the adverse selection within grades.

Blume, L. E. and D. Easley (1982), "Learning to be Rational",*Journal of Economic Theory*, 26, 340-351.

Caskey, John P. (2005): "Fringe Banking and the Rise of Payday Lending", Chapter 2 of *Credit Markets for the Poor*, Russell Sage Foundation, New York.

Cohen, Alma and Peter Siegelman (2009), "Testing for Adverse Selection in Insurance Markets" NBER working paper #15586.

Cooper, D. J., Garvin S., Kagel J. (1997): "Signaling and Adaptive Learning in an Entry Limit Pricing Game" *Rand Journal of Economics* 28, 662-683.

De Long, J. B.; A. Shleifer; L. H. Summers and R.J. Waldmann "Noise Trader Risk in Financial Markets" *The Journal of Political Economy* 98(4): 703-738.

Fama, Eugene F. (1970) "Efficient Capital Markets: A Review of Theory and Empirical Work" *Journal of Finance* 25(2): Papers and Proceedings of the Twenty-Eighth Annual Meeting of the American Finance Association New York, N.Y. December, 28-30, 1969 (May, 1970), pp. 383-417.

Feldman, M. (1987) "An Example of Convergence to Rational Expectations with Heterogeneous Beliefs",*International Economic Review*, 28, 635-650.

Freedman, Seth and Ginger Z. Jin (2008) "Do Social Networks Solve Information Problems for Peer-to-Peer Lending? Evidence from Prosper.com" Net Institute Working Paper 08-43.

Greenlaw, David; Jan Hatzius; Anil Kashyap, Hyun Song Shin (2008) : "Leveraged Losses: Lessons from the Mortgage Market Meltdown", mimeo, accessed at http://www.chicagogsb.edu/usmpf/docs/usmpf2008confdraft.pdf.

Grossman, Sandy Jay and Joseph Stiglitz (1980) "On the impossibility of informationally efficient markets" *American Economic Review* 70: 393-408.

Hampshire, Robert (2008) "Group Reputation Effects in Peer-to-Peer Lending Markets: An Empirical Analysis from a Principle-Agent Perspective" *mimeo*.

Hendel, Igal; Aviv Nevo and Francois Ortalo-Magne (2009) "The Relative Performance of Real Estate Marketing Platforms: MLS versus FSBOMadison.com" *American Economic Review* 99(5), December 2009, 1878-98.

Hotelling, H. (1933) "Review of The triumph of mediocrity in business by Horace Secrist," *Journal of American Statistics Association* 28, 463-465.

Iyer, Rajkamal; Asim Ijaz Khwaja; Erzo F. P. Luttmer; Kelly Shue (2009) "Screening in Alternative Credit Markets: Can Individual Lenders Infer Borrower Credit-worthiness in Peer-to-Peer Lending" mimeo.

Jin, Ginger Z. and Andrew Kato (2007) "Dividing Online and Offline: A Case Study" *Review of Economic Studies* July 2007.

Jun, Byoung and Xavier Vives (1996) "Learning and Convergence to a Full-Information Equilibrium are not Equivalent" *The Review of Economic Studies* 63(4): 653-674.

Lin, Mingfeng; Siva Viswanathan and N.R. Prabhala (2009) "Judging Borrowers By The Company They Keep: Social Networks and Adverse Selection in Online Peer-to-Peer Lending" *SSRN working paper* available at http://ssrn.com/abstract=1355679.

Marcet, A. and T. J. Sargent (1988) "The Fate of Systems with 'Adaptive' Expectations", *American Economic Review*, 78, 168-172.

Pope, Devin and Justin R. Sydnor (forthcoming) "What's in a Picture? Evidence of Discrimination from Prosper.com" *forthcoming Journal of Human Resources*.

Ravina, Enrichetta "Love & Loans: The Effect of Beauty and Personal Characteristics in Credit Markets," Available at SSRN: http://ssrn.com/abstract=972801.

Rigbi, Oren (2008) "The Effects of Usury Laws: Evidence from the Online Loan Market" *Dissertation*, Stanford University.

Routledge, Bryan R. "Adaptive Learning in Financial Markets" *The Review of Financial Studies* 12(5):1165-1202.

Salmon, Timothy C. (2001) "An Evaluation of Econometric Models of Adaptive Learning" *Econometrica* 69(6): 1597-1628.

Schonberg, Uta (2007) "Testing for Asymmetric Employer Learning" *Journal of Labor Economics* 25(4): 651-691.

Sharpe, Steven A. (1990) "Asymmetric Information, Bank Lending, and Implicit Contracts: A Stylized Model of Customer Relationships." *Journal of Finance* 45 (September): 1069-87.

Skiba, Paige and Jeremy Tobacman (2007) "Profitability of Payday Loans" Working paper available at http://www.law.vanderbilt.edu/faculty/faculty-personal-sites/paige-skiba/index.aspx.

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

Townsend, R.M. (1978) "Market Anticipations, Rational Expectations, and Bayesian Analysis", *International Economic Review*, 19, 481-494.

Vives, Xavier (1993) "How Fast do Rational Agents Learn?" *The Review of Economic Studies* 60(2): 329-347.

Table 1: Summary of Listings and Loans by Quarter (Sample period: June 1, 2006 – July 31, 2008)

Listings:				
Quarter	Number	Total Amount Requested ($100,000)	Mean Amount Requested ($)	Mean Borrower Max Interest Rate
20062	5,375	26.65	4,957.22	16.86%
20063	19,771	107.25	5,424.63	18.15%
20064	31,629	196.57	6,214.85	17.45%
20071	31,373	263.22	8,389.94	16.72%
20072	37,505	331.62	8,841.98	17.51%
20073	39,353	328.79	8,355.00	18.06%
20074	41,585	334.23	8,037.29	18.41%
20081	33,485	250.14	7,470.30	19.24%
20082	43,371	318.53	7,344.20	24.50%
20083	10,361	73.48	7,092.42	26.40%
Total	293,808	2230.48	7,591.62	19.19%
Loans:				
Quarter	Number	Total Amount Funded ($100,000)	Mean Amount Funded ($)	Mean Contract Interest Rate
20062	385	1.47	3,822.17	19.03%
20063	1,934	9.37	4,844.63	19.41%
20064	2,403	11.54	4,804.05	18.97%
20071	3,079	19.93	6,472.60	17.37%
20072	3,118	23.47	7,527.98	17.42%
20073	2,671	18.43	6,900.12	17.31%
20074	2,593	18.98	7,320.17	17.11%
20081	3,074	20.47	6,658.94	17.37%
20082	4,344	26.33	6,061.10	17.98%
20083	1,407	8.27	5,877.70	19.39%
Total	25,008	158.27	6,328.65	17.90%

Note: Authors' tabulations from Prosper listing and loan data.

Table 2: Summary of Listings and Loans by Credit Grade

	Prosper Listings			Prosper Loans							Experian Accounts opened in Sept. 2003
						As of 2/21/2009			Conditional on loan life = 24 months, observed by 2/21/09		observed on 9/2005
Credit Grade	Number of listings	Mean Borrower Maximum Interest Rate	Funding Rate	Number of loans	Mean Contract Interest Rate	% Late	% Default	% 3m late or worse	% 3m late or worse	% 3m late or worse (DTI<0.2)	% 3m late or worse (loan life = 2 years)
2A	9321	11.57%	32.08%	2990	9.70%	1.61%	11.81%	12.54%	8.70%	4.48%	0.89%
A	11099	14.05%	25.45%	2825	12.29%	1.70%	19.61%	20.35%	16.70%	9.23%	3.33%
B	17211	16.47%	21.88%	3766	15.00%	2.79%	26.50%	27.54%	22.55%	19.55%	6.04%
C	30843	18.57%	15.76%	4862	17.49%	2.98%	30.65%	31.45%	26.70%	23.44%	9.44%
D	43282	20.08%	10.35%	4479	20.66%	3.35%	33.53%	34.49%	30.34%	27.56%	15.29%
E	52000	20.65%	5.56%	2891	24.82%	2.59%	44.97%	45.66%	39.13%	37.69%	24.25%
HR	128633	19.83%	2.39%	3077	24.52%	2.34%	56.55%	57.56%	51.21%	50.66%	34.40%
NC	1419	17.66%	8.32%	118	22.06%	0.00%	72.03%	72.03%	59.32%	60.32%	
Total	293808	19.19%	8.51%	25008	17.90%	2.57%	32.08%	32.94%	28.95%	26.86%	6.94%

Note: Funding rate refers to the percentage of listings that become funded loans. Default refers to loans that are 4 months or more late or considered default due to bankruptcy. DTI stands for the borrower's debt to income ratio as reported in the listing.

Table 3: Funding Rate, Interest Rate and Default or Late (June 1, 2006 – July 31, 2008)

	Funded?	Contract interest rate	Default or late as of 2/21/2010
	Probit (marginal effects)	OLS	Probit (marginal effects)
Listing attributes available before Feb 2007			
Grade=AA	0.696*	-0.032*	-0.330*
	(21.041)	(-8.258)	(-12.222)
Grade=A	0.409*	-0.026*	-0.320*
	(14.144)	(-8.326)	(-14.602)
Grade=B	0.252*	-0.021*	-0.307*
	(11.146)	(-6.855)	(-11.137)
Grade=C	0.095*	-0.016*	-0.298*
	(8.135)	(-5.175)	(-9.139)
Grade=D	0.033*	-0.008*	-0.280*
	(6.020)	(-2.597)	(-8.667)
Grade=E	0.001	-0.003	-0.218*
	(0.578)	(-0.889)	(-6.302)
Grade=HR	-0.005*	-0.003	-0.142**
	(-2.711)	(-1.068)	(-3.037)
Amountrequested	-0.000*	0.000*	0.000*
	(-32.188)	(17.345)	(18.182)
Autofunded	0.011*	0.036*	0.079*
	(20.910)	(90.511)	(9.539)
borrowermaximumrate	0.702*	0.458*	3.705*
	(38.743)	(17.009)	(10.329)
borrowermaximumrate2	-1.161*	0.484*	-5.392*
	(-34.968)	(7.978)	(-7.036)
Yeshomeowner	0.001***	0.001	-0.047**
	(1.729)	(0.608)	(-2.559)
Debt to income ratio	-0.017*	0.004*	0.042*
	(-15.799)	(5.433)	(3.818)
debt-to-income * homeowner	0.001**	-0.000	0.010
	(2.156)	(-0.711)	(1.513)
having an image	0.005*	-0.001*	-0.014***
	(16.152)	(-3.449)	(-1.873)
length of description	0.000*	-0.000*	-0.000**
	(18.177)	(-5.380)	(-2.510)
mention debtconsolidation	0.001**	0.000	-0.018**
	(2.372)	(0.174)	(-2.556)
mention business	-0.001*	0.000	0.050*
	(-3.199)	(1.062)	(6.470)
mention car	-0.000	0.001**	0.011
	(-1.207)	(2.524)	(1.289)
mention mortgage	0.000	0.000	-0.011
	(0.315)	(0.026)	(-1.191)
mention health	0.001*	0.001**	0.024*
	(2.647)	(1.997)	(2.817)

mention education	0.000	-0.000	-0.030*
	(0.016)	(-0.132)	(4.058)
mention family	0.001*	0.001**	0.024*
	(2.887)	(2.054)	(2.828)
mention retirement	-0.001**	-0.001	-0.020
	(-2.053)	(-1.343)	(-1.266)
mention pay-day loan	0.003*	0.003*	0.074*
	(5.157)	(3.113)	(4.852)
Saidrelisting	0.008*	0.002	-0.011
	(4.951)	(1.570)	(-0.542)
count of relisting	-0.000*	0.001*	0.005*
	(-5.666)	(6.828)	(4.402)
Currentdelinquencies	-0.001*	0.000*	0.014*
	(-18.347)	(4.454)	(9.932)
delinquencies in past 7 yrs	-0.000*	0.000*	-0.001*
	(-11.790)	(6.091)	(-3.564)
length of credit history	-0.000*	0.000*	0.000
	(-7.724)	(3.973)	(1.251)
totalcreditlines	0.000	0.000**	-0.000
	(0.438)	(2.171)	(-1.610)
In public records in past 10 yrs	-0.001*	0.000	0.013*
	(-8.563)	(0.126)	(3.808)
# of inquiries in past 6m	-0.001*	0.000*	0.015*
	(-13.047)	(6.197)	(14.854)
missing credit info	0.002	-0.005	0.076
	(0.957)	(-1.359)	(1.229)
in_a_grp_borrower	0.004*	-0.004*	-0.007
	(10.464)	(-9.299)	(-0.756)
have endorsement + nobid by group leader	0.017*	-0.003*	0.020
	(8.118)	(-2.798)	(0.992)
have endorsement + bid by group leader	0.096*	-0.005*	0.015
	(21.708)	(-7.294)	(1.301)
have endorsement + nobid by friend	0.002*	0.001**	0.014
	(5.304)	(2.367)	(1.528)
have endorsement + bid by friend	0.050*	-0.007*	-0.086*
	(12.454)	(-6.150)	(-5.713)
Year-week FE	Yes	Yes	Yes
N	293,802	25,008	24,995
Adjusted R2	0.375	0.855	0.163

The sample includes all the listings and loans between June 1, 2006 and July 31, 2008. T-statistics are in parenthesis. * p<0.01, ** p<0.05, *** p<0.1. All regressions control for state dummies, year-week FE, macro variables, variables for prosper environment, duration of auction, and indicators for missing debt-to-income ratio and other credit attributes. All regressions do not include the new credit variables added after Feb. 12, 2007. In an unreported table, we show that regressions including these variables and condition on the post-Feb-2007 sample generate similar results.

Table 4: Half Grade Regressions

Unit of observation = census division by month by half-grade interval
Sample: the half-grade intervals that have credit scores at or above 520

	# of listings	# of loans	Funding rate	Average contract interest rate	% late in 6m	% late in 12m
Dummy of minus grade	11.381*	1.908*	0.006	0.004*	0.014**	0.023*
	(4.962)	(4.124)	(0.671)	(2.989)	(2.213)	(2.825)
N	3,978	3,978	3,779	3,357	2,776	2,006
R2	0.679	0.552	0.269	0.795	0.145	0.208

T-statistics are in parenthesis. * p<0.01, ** p<0.05, *** p<0.1. All regressions control for credit grade FE, year-month FE, and census-division FE, standard errors clustered by census division. Performance data are as of August 1, 2008.

Table 5: Lender Responses to Ever Late Loans

				Conditional on Funding a Loan in Week t					
					% of Investment in:				
	Funded a Loan		Amount Funded		AA to A	B to D	E to HR	Mean IRR	
	coef/t	coef/t	coef/t	coef/t	coef/t	coef/t	coef/t	coef/t	coef/t
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
% of Portfolio Ever Late	-0.078*		-795.256*					0.198*	
	(-16.401)		(-10.999)					(39.164)	
% of NC Loans Ever Late		-0.116*		-183.335*	0.062*	0.046*	-0.078*		0.041*
		(-13.269)		(-3.521)	(5.655)	(3.487)	(-8.662)		(13.288)
% of E to HR Loans Ever Late		-0.080*		-125.933*	0.159*	0.052*	-0.211*		0.038*
		(-21.487)		(-5.833)	(17.567)	(5.380)	(-30.943)		(19.043)
% of B to D Loans Ever Late		-0.105*		-405.820*	0.475*	-0.416*	-0.061*		0.071*
		(-24.743)		(-9.870)	(24.223)	(-21.191)	(-6.273)		(18.048)
% of A to AA Loans Ever Late		-0.126*		-255.409*	-0.146*	0.087*	0.060*		0.016*
		(-22.661)		(-5.956)	(-10.693)	(5.541)	(5.436)		(4.045)
N	2,564,481	2,564,481	553,117	553,117	553,117	553,117	553,117	550,789	550,789

T-statistics are in parenthesis. * p<0.01, ** p<0.05, *** p<0.1. Column 1 is a linear probability model and all other columns are OLS regressions. Standard errors are clustered at the lender level.

Table 6: Summary of IRR Measures

				per loan					per dollar	per lender
	misperformance measure	macro	# of loans	mean	standard deviation	min	max	mean of loans with real macro	mean	mean
IRR1	default	no macro	24996	-0.68%	10.89%	-94.92%	35.21%	-3.05%	-1.86%	-1.61%
IRR2	default	real macro	5736	-3.40%	13.26%	-91.82%	28.88%	-3.40%	-2.47%	-2.27%
IRR3	default	rolling forecast	24995	4.70%	17.94%	-99.84%	41.20%	-1.48%	2.36%	2.27%
IRR4	default	forecast 6/1/06	24997	2.40%	12.13%	-99.70%	36.82%	-0.06%	2.71%	2.72%
IRR5	miss payment	no macro	24996	-6.38%	13.12%	-96.12%	29.33%	-9.64%	-7.63%	-7.19%
IRR6	miss payment	real macro	5735	-9.80%	15.60%	-93.07%	2.56%	-9.80%	-8.66%	-8.04%
IRR7	miss payment	rolling forecast	24995	0.18%	18.85%	-100.00%	41.20%	-9.10%	-2.51%	-1.03%
IRR8	miss payment	forecast 6/1/06	24996	-4.13%	13.99%	-99.99%	32.86%	-7.32%	-4.03%	-3.75%
IRR9	default or late	no macro	24996	-7.35%	13.49%	-96.13%	29.33%	-10.59%	-8.71%	-8.24%
IRR10	default or late	real macro	5736	-10.68%	15.86%	-93.09%	25.59%	-10.68%	-9.68%	-8.98%
IRR11	default or late	rolling forecast	24995	-0.93%	20.14%	-100.00%	41.20%	-9.86%	-4.01%	-3.52%
IRR12	default or late	forecast 6/1/06	24997	-4.98%	14.37%	-100.00%	32.86%	-8.31%	-4.92%	-4.60%

Note: All numbers reported are conditional on convergence in the calculation of IRR. Our full sample has 25008 loans, about 10 of them drop out of the probit regression of loan performance because one of the listing attributes predicts the performance perfectly. We do not calculate IRR for these loans. Conditional on have predicted monthly performance, at most 1 loan does not meet the convergence criteria of IRR. IRR2, IRR6 and IRR10 have significant fewer observations because they are conditional on the loans that have observed real macro variables in all the 36 months of loan life.

Table 7: Above and Below Median Lender First Month Characteristics

				Fraction of Portfolio by Grade		
	Total Portfolio Size	Number of Loans	$ per Loan	AAtoA	BtoD	EtoHR
Below Cohort Median IRR8						
Mean	860.27	9.08	98.65	0.36	0.47	0.17
Median	300.00	5.00	51.25	0.25	0.50	0.00
SD	2553.03	14.16	366.19	0.37	0.35	0.29
SE	16.95	0.09	2.43	0.0024	0.0023	0.0019
N	22696	22696	22696	22696	22696	22696
Above Cohort Median IRR8						
Mean	823.84	9.91	93.38	0.38	0.57	0.05
Median	300.00	4.00	50.00	0.33	0.60	0.00
SD	2148.36	17.14	323.29	0.35	0.35	0.15
SE	14.22	0.11	2.14	0.0023	0.0023	0.0010
N	22840	22840	22840	22840	22840	22840
t test p-value	0.099	0.000	0.104	0.000	0.000	0.000

Table provides summary statistics of loans made by lenders within their first 4 weeks on Prosper. Above and below median split is determined within each weekly cohort of lenders.

Table 8: Detection of Lender Heterogeneity in Initial Sophistication

	IRR8	% of new loans in AAtoA	% of new loans in BtoD	% of new loans in EtoHR
	coef/t	coef/t	coef/t	coef/t
Dummy of Above-median IRR in first-month portfolio	0.084*	0.211*	0.010	-0.212*
	(56.167)	(30.016)	(1.473)	(-39.379)
Dummy of above-median * Second Month	-0.026*			
	(-25.692)			
Dummy of above-median * lender age	-0.004*	-0.012*	0.000	0.011*
	(-28.500)	(-11.952)	(0.413)	(30.610)
Dummy of above-median * lender cohort	-0.001*	-0.012*	0.004*	0.008*
	(-13.179)	(-25.978)	(8.418)	(27.679)
Dummy of above-median * lender cohort * lender age	-0.00018*	0.00100*	-0.00074*	-0.00024*
	(-12.887)	(9.232)	(-7.411)	(-6.613)
N	513,603	514,636	514,636	514,636
Adjusted R2	0.080	0.056	0.023	0.099

Unit of observation is defined by lender-week. The sample includes all the lenders that started on Prosper since June 1, 2006. Cohort is defined as the count of months from June 2006. Lender age is measured in month since the first day of investment on Prosper. Regression controls for cohort fixed effects, lender age fixed effects and year-week fixed effects. T-stat in parentheses. *p<0.01, **p<0.05, ***p<0.1.

Table 9: Heterogeneous Learning

A: Response to Group Performance				B: By Lender Cohort				C: By Lender Sophistication			
	Funded A Loan	Amount Funded	Mean IRR8		Funded A Loan	Amount Funded	Mean IRR8		Funded A Loan	Amount Funded	Mean IRR8
	coef/t	coef/t	coef/t		coef/t	coef/t	coef/t		coef/t	coef/t	coef/t
% of Portfolio Ever Late	-0.031*	-617.450*	0.204*	% of Portfolio Ever Late	-0.167*	-1,161.596*	0.237*	% of Portfolio Ever Late	-0.058*	-709.569*	0.246*
	(-4.27)	(-6.201)	(16.262)		(-9.340)	(-4.102)	(16.675)		(-10.70)	(-11.556)	(49.2)
% of Group Portfolio Ever Late	-0.101*	-27.157	-0.022	X 2006 Half 2	0.073*	384.006	-0.019	X Above-median	-0.031*	-41.606	-0.164*
	(-3.43)	(-0.204)	(-1.283)		(3.89)	(1.403)	(-1.306)		(-3.38)	(-0.706)	(-23.429)
				X 2007 Half 1	0.108*	453.434***	-0.056*				
					(5.74)	(1.659)	(-3.679)				
				X 2007 Half 2	0.142*	572.299**	-0.100*				
					(7.07)	(1.987)	(-5.124)				
				X 2008 Half 1	0.136*	331.377	-0.241*				
					(4.75)	(0.969)	(-6.254)				
N	499,575	81,394	80,905	N	2,564,481	553,117	550,789	N	2,371,593	514,636	513,603

T-statistics are in parenthesis. * p<0.01, ** p<0.05, *** p<0.1. The Funded a Loan columns are linear probability model and all other columns are OLS regressions. The Amount Funded and IRR8 regression samples are conditional on funding a loan in week t. Standard errors are clustered at the lender level.

Table 10: Cohort Difference within a Snapshot of Time

	IRR8	**IRR8**
	coef/t	**coef/t**
=1 if lender is new on Prosper in his first month	0.003*	0.007*
	(7.139)	(15.569)
# of months that the lender has been on Prosper		0.001*
		(14.898)
N	550,789	550,789
Adjusted R2	0.025	0.026

Unit of observation is defined by lender-week. Cohort is defined as the count of months from June 2006. Lender age is measured in month since the first day of investment on Prosper. Regression controls for year-week fixed effects. T-stat in parentheses. *p<0.01, **p<0.05, ***p<0.1.

Figure 1A: CDF of Prosper and Experian Listings

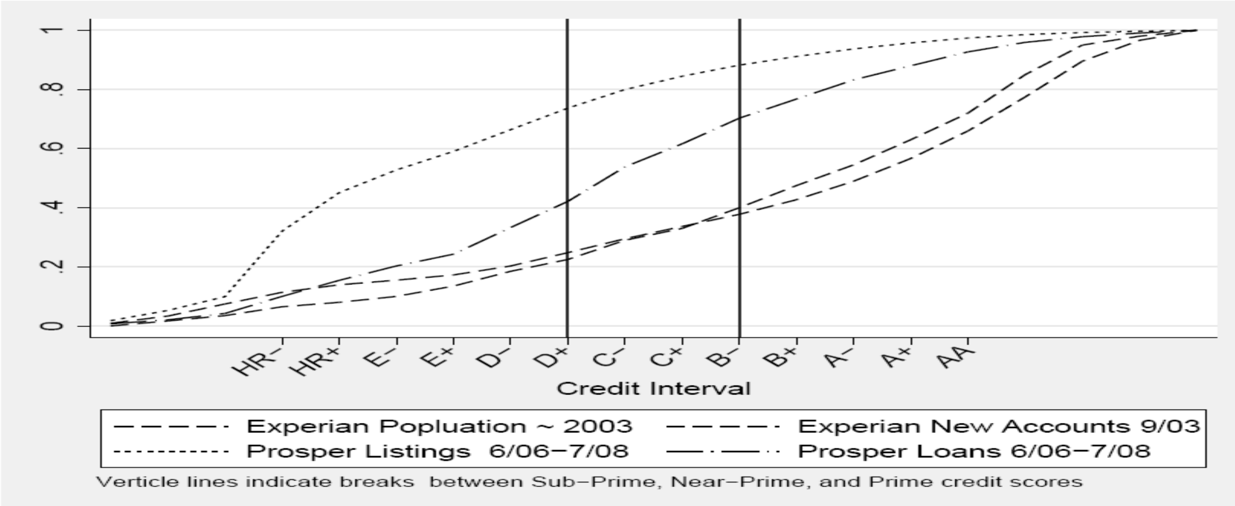


Figure 1B: PDF of Prosper Listings by Time

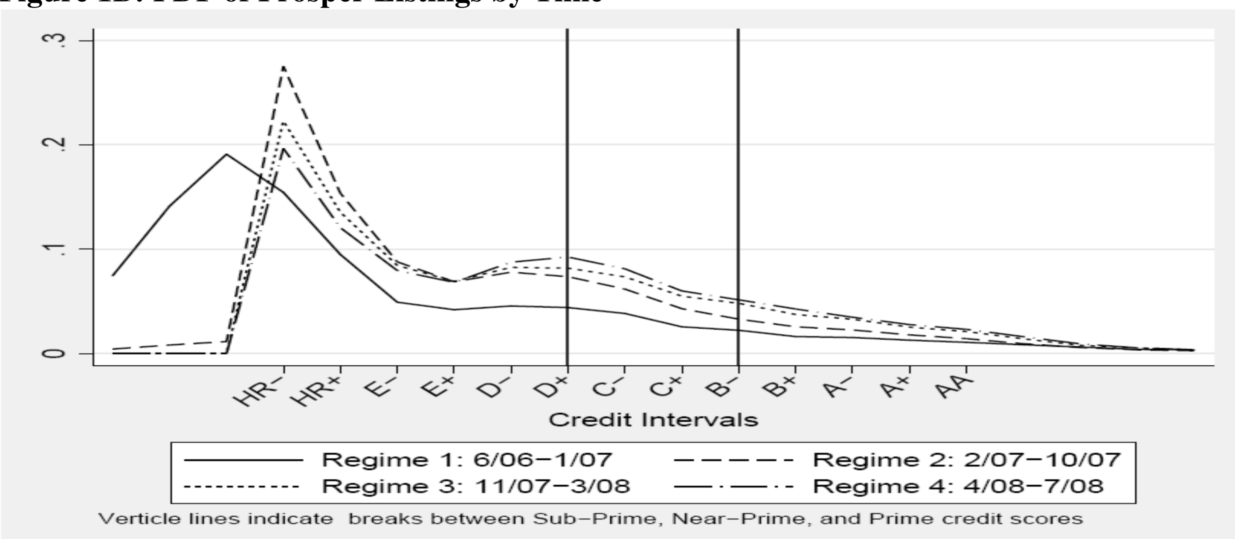


Figure 1C: PDF of Prosper and Experian Loans by Time

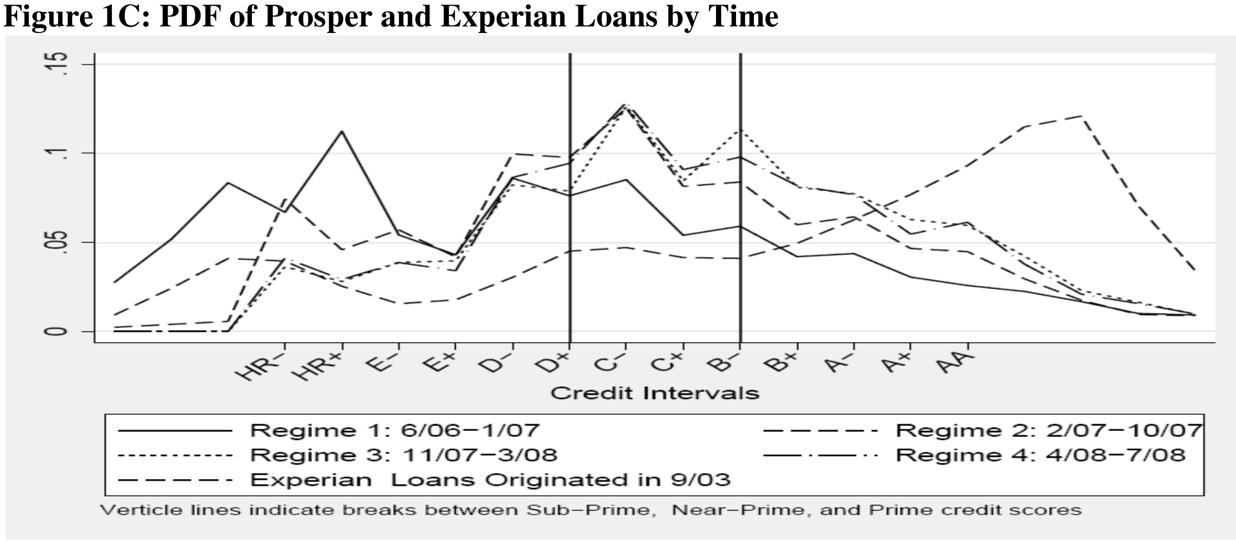
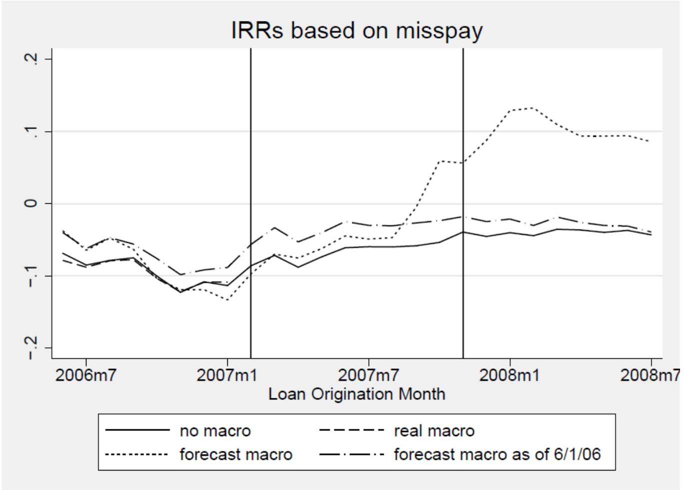


Figure 2: IRRs by Macro Controls



Vertical lines indicate Prosper's Feb. 12, 2007 policy of redefining E and HR plus posting more credit information and Oct. 30, 2007 introduction of bidder guidance.

Figure 3: IRR8 Distribution by Grade

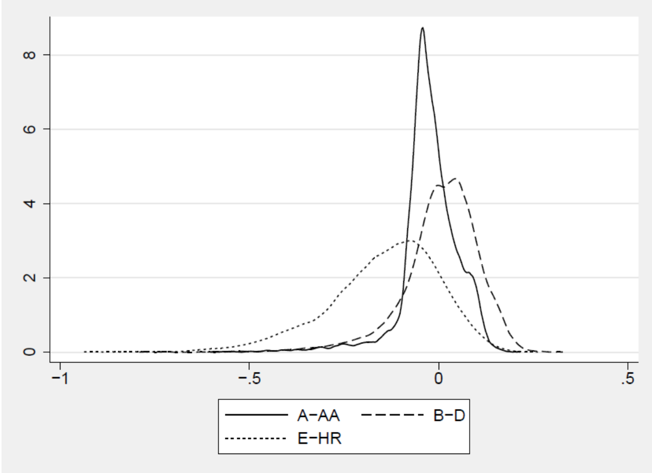
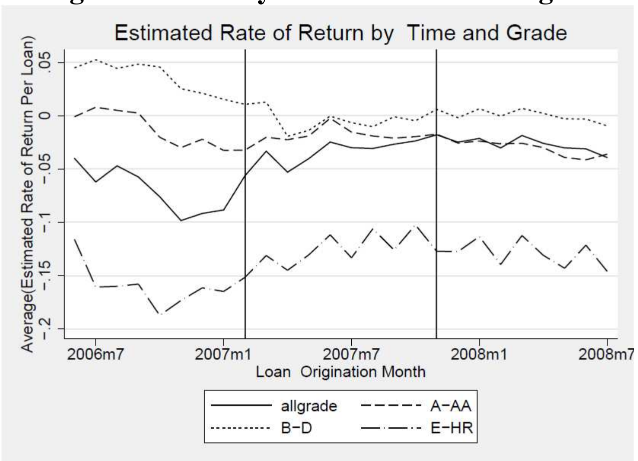


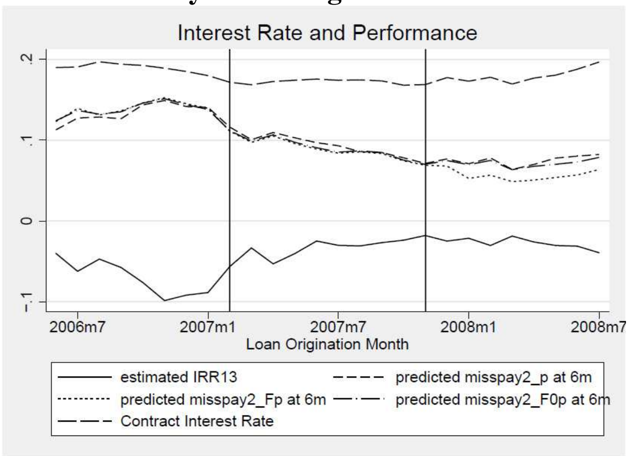
Figure presents kernel densities of IRR for each grade category.

Figure 4: IRR8 by Grade and Loan Origination Month



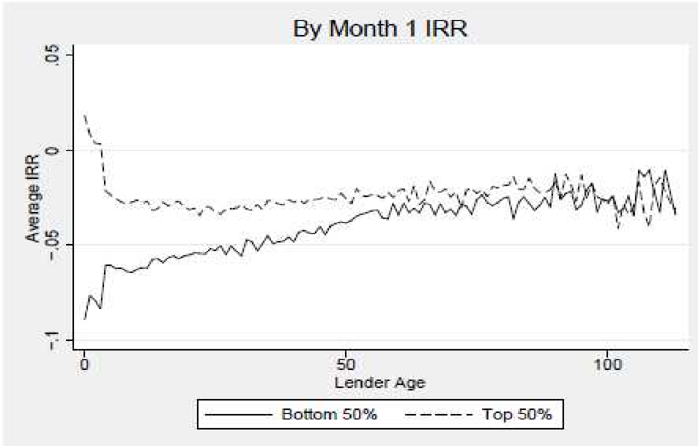
Vertical lines indicate Prosper's Feb. 12, 2007 policy of redefining E and HR plus posting more credit information and Oct. 30, 2007 introduction of bidder guidance.

Figure 5: Mean of IRR8, Contract Rate and Predicted Performance by Loan Origination Month



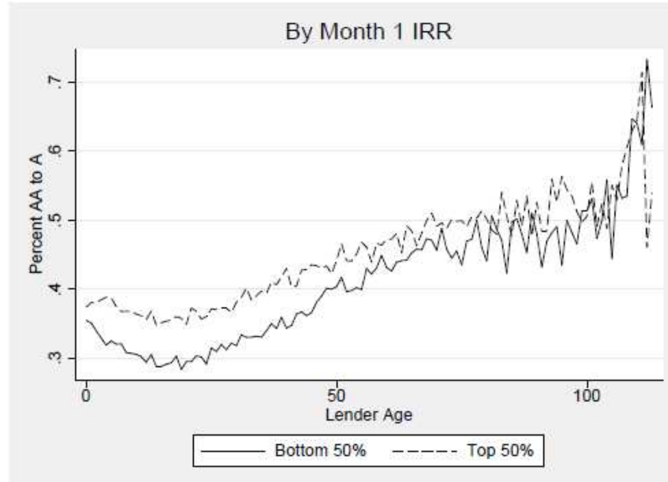
Vertical lines indicate Prosper's Feb. 12, 2007 policy of redefining E and HR plus posting more credit information and Oct. 30, 2007 introduction of bidder guidance.

Figure 6: Average IRR8 of New investments by Lender Age and Initial Portfolio IRR8



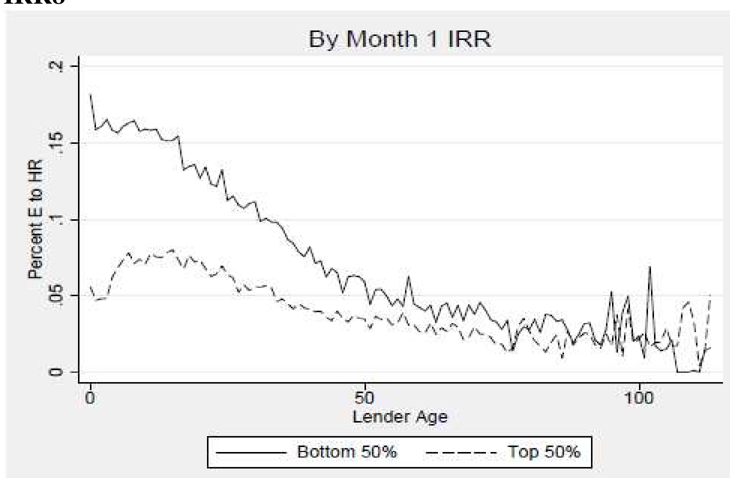
Above and below median split is determined within each weekly cohort of lenders.

Figure 7: Percentage of AA to A loans by Age and Initial Portfolio IRR8



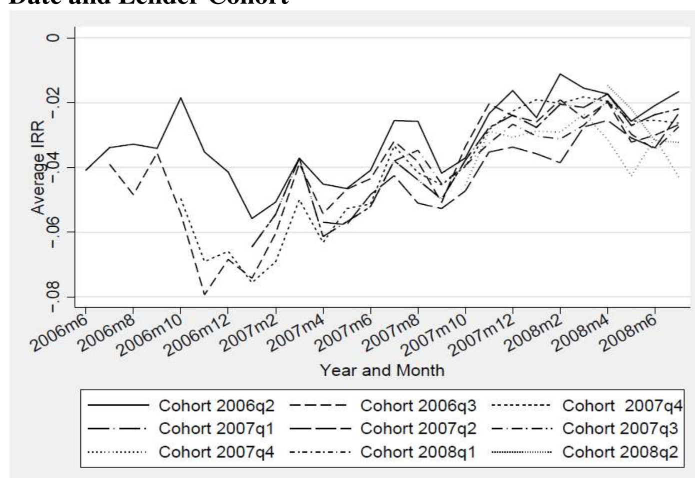
Above and below median split is determined within each weekly cohort of lenders.

Figure 8: Percentage of E to HR Loans by Age and Initial Portfolio IRR8



Above and below median split is determined within each weekly cohort of lenders.

Figure 9: Average IRR8 of New Investments by Investment Date and Lender Cohort



Cohorts refer to 6 month period in which a lender funds his first loan on Prosper.

Appendix – Not for Publication – IRR algorithm and robustness checks

As described in the main text, we compute the expected IRR by first using the observed ex post loan performance to predict a relationship between listing attributes and loan performance, and then calculating a compounded annual discount factor that equalizes the loan amount to the present value of all the predicted monthly cash inflow. In this Appendix, we present evidence on the fit of our main probit model in performance prediction, describe how IRR changes when we use duration or an alternative probit model to predict loan performance, and finally discuss the potential bias in the absolute measure of IRR.

Fit of the Main Probit Model To predict monthly cash inflow (for a lender), we need to estimate whether a loan will misperform or pay off the whole loan in a specific month. Appendix Figure 2 contrasts the real data with the predicted probability of miss payment under real macro, rolling forecast of macro, and the macro forecast as of 6/1/2006. From loan month 1 to 36, the prediction with the realized macroeconomic variables closely fits the real data, with close to zero prediction error for each month. Miss payment predicted with rolling forecast and forecast as of 6/1/2006 are systematically lower than the real performance because lenders did not expect the macroeconomic downturn. This point is more obvious in Appendix Figure 3 which presents the same prediction by calendar month instead of loan life. It seems that lender forecast started to deviate from the real macro after July 2008, which is reasonable given the fact that the subprime crisis broke out in the wholesale financial market in August 2007 and it took time for it to spread to consumer lending. Appendix Figures 4 and 5 show the predicted probability of a loan being paid off by a specific month. Consistent with lender optimism, the forecast probabilities of paid off are systematically higher than the real data and the prediction made from real macro.

An Alternative Probit Model One caveat of the main probit model is that it predicts paid off and misperformance separately, although they are exclusive to each other. This opens a possibility that the predicted probability of on time payment, defined by 1-prob(payoff)-prob(misperformance), may be below zero. In an alternative probit model we make an additional prediction on the probability of [ever paid off by or misperformance in month t]. This does not solve the problem completely, as other parts of the cash flow formula uses the probability of paid off separately. However, if the IRRs computed with the additional probit prediction are similar to what we have reported in the main text, it shows that our results are robust to this specification.

Appendix Table 3 reports the IRRs estimated under this alternative probit. To facilitate comparison with the duration model to be described below, we measure misperformance by default or [default or late]. Miss payment is excluded because the duration model does not allow the studied event to switch on and off over time, and ever missing a payment is equivalent to counting both default and late as misperformance. As shown in Appendix Table 3, the alternative probit model yields slightly different IRRs as compared to Table 6, but IRRs by misperformance measure and macro treatment follow the same pattern as before: counting late as misperformance predicts much lower IRRs and $IRR_{realmacro} < IRR_{forecast060106} < IRR_{rollingforecast}$. The distribution of these IRRs by grade, time, cohort and lender age is similar to what is reported in the main text.

Duration model At the first glance, the time passed until the event of pay off or misperformance occurs lends itself to a duration model. A typical duration model does not allow borrowers to switch between misperformance and on-time payment, but it is likely a small problem given the fact that most borrowers who missed a payment are likely to default eventually. A more compelling reason that motivates us to choose a probit specification over a duration model is that the duration model does not predict the real data as well as probit.

More specifically, we use a Cox proportional hazard model to predict the probability of paid off by month t and the probability of [paid off or misperformance by month t]. Appendix Figures 6-9 compare these two predicted probabilities using probit and duration to the real data. It is clear that probit predictions fit the real data very well, but duration predictions have more errors, especially at later calendar times and later loan lives (when we have fewer observations). Since the duration model requires us to infer the probability of a loan being paid off at exactly month t, an error in the prediction of paid off duration will be carried over to that inference. In comparison, in the main (and alternative) probit model we run a separate probit regression directly to predict this probability and therefore avoid the sequential prediction error.

Appendix Table 3 also presents the IRRs based on the duration model. While the absolute magnitude of the duration model is somewhat different from that of the main (and alternative) probit model, the relationship across misperformance measures and different macro treatments is the same as before. Again, the distribution of duration-based IRRs by grade, time, cohort and lender age is similar to what we have reported in the main text.

Potential bias No matter which prediction model we use, our IRR algorithm is subject to potential bias in both directions. On the one hand, our IRR estimates may be downward biased because we try to be conservative in the calculation of cash flows. Specifically, we assume away any loss recovery from default loans, and we do not account for the late fees that a lender may receive from a late-but-non-defaulting borrower. When we count early payoff as a bulk cash flow that arrives in the paid-off month, it effectively assumes that the paid off amount is reinvested in a loan that is identical to the loan under study. This assumption may be conservative because lenders may learn to fund better loans over time.

On the other hand, our IRR estimates may have overestimated the return on investment because we do not consider any cost that lenders may incur in processing Prosper information. The time that lenders spend on screening listings and digesting Prosper history could be long and stressful. Lastly, our IRR estimates are based on the average loan performance observed from June 1, 2006 to February 21, 2010, a period that stretches from the end of a boom to an economy-wide recession. If the future economy becomes worse than we forecast, the reported IRR will overestimate the actual rate of return for on-going loans.

Appendix Table 1: Summary Statistics of Listing Attributes (June 1, 2006 – July 31, 2008)

	Listings			Loans		
	Mean	STD	N	Mean	STD	N
Information available before Feb. 12, 2007						
Grade=AA	0.032	0.175	293808	0.120	0.324	25008
Grade=A	0.038	0.191	293808	0.113	0.317	25008
Grade=B	0.059	0.235	293808	0.151	0.358	25008
Grade=C	0.105	0.307	293808	0.194	0.396	25008
Grade=D	0.147	0.354	293808	0.179	0.383	25008
Grade=E	0.177	0.382	293808	0.116	0.320	25008
Grade=HR	0.438	0.496	293808	0.123	0.328	25008
Grade=NC	0.005	0.069	293808	0.005	0.069	25008
amountrequested	7592	6388	293808	6329	5679	25008
autofunded	0.311	0.463	293808	0.263	0.441	25008
borrowermaximumrate	0.192	0.084	293808	0.209	0.074	25008
yeshomeowner	0.327	0.469	293808	0.441	0.497	25008
debt-to-income (DTI) ratio	0.505	1.359	293808	0.330	0.978	25008
missing DTI	0.068	0.251	293808	0.035	0.183	25008
DTI topcoded if DTI>=10	0.083	0.275	293808	0.044	0.205	25008
have image	0.515	0.500	293808	0.659	0.474	25008
length of listing desc (in chars)	1058	772	293808	1295	866	25008
mention debt consolidation	0.358	0.480	293808	0.375	0.484	25008
mention business loan	0.231	0.421	293808	0.271	0.444	25008
mention car	0.689	0.463	293808	0.626	0.484	25008
mention mortgage	0.139	0.346	293808	0.187	0.390	25008
mention health	0.721	0.449	293808	0.790	0.407	25008
mention education	0.211	0.408	293808	0.248	0.432	25008
mention family	0.179	0.383	293808	0.189	0.392	25008
mention retirement	0.030	0.171	293808	0.041	0.199	25008
mention pay-day loan	0.057	0.233	293808	0.057	0.231	25008
concede relisting	0.008	0.089	293808	0.021	0.144	25008
# of listings (incld current one)	2.811	3.361	293808	2.912	2.863	25008
interest rate cap	0.243	0.093	293808	0.273	0.082	25008
borrower fee	1.800	0.794	293808	1.548	0.781	25008
lender fee	0.852	0.231	293808	0.790	0.258	25008
amountdelinquent ($)	3516	12374	221618	1176	6257	18618
missing amountdelinquent	0.004	0.066	221618	0.001	0.037	18618
currentdelinquency	3.833	5.303	293808	1.454	3.400	25008
delinquency in 7yrs	11.022	16.450	293808	5.800	12.356	25008
lengthcredithistory (in days)	152.208	84.472	293808	158.049	87.107	25008
totalcreditlines	24.354	14.393	293808	23.964	14.424	25008
in public records in past 10 years	0.657	1.395	293808	0.405	0.936	25008
# of inquiries in past 6 months	4.153	4.959	293808	2.927	3.979	25008

Appendix Table 1 Continued: Summary Statistics of Listing Attributes (June 1, 2006 – July 31, 2008)

	Listings			Loans		
	Mean	STD	N	Mean	STD	N
Credit info added after Feb. 12, 2007						
currentcreditlines	8.230	6.001	221618	9.566	5.931	18618
opencreditlines	7.224	5.303	221618	8.165	5.223	18618
band card utilization rate	0.629	0.431	221618	0.547	0.373	18618
revolving balance ($)	12087	31802	221618	16326	39388	18618
in publice records in past 1 year	0.075	0.346	221618	0.040	0.237	18618
working full time	0.821	0.383	221618	0.859	0.348	18618
working part time	0.040	0.196	221618	0.038	0.192	18618
income 25-75 K	0.670	0.470	202271	0.651	0.477	17782
income > 75K	0.144	0.351	202271	0.220	0.415	17782
missing income	0.297	0.457	293808	0.284	0.451	25008
no employment or income reported	0.014	0.119	293808	0.005	0.072	25008
missing new credit info posted after 2/07	0.000	0.013	293808	0.000	0.018	25008
missing credit info posted bef 2/07	0.008	0.087	293808	0.004	0.062	25008
Social network variables						
borrower in a group	0.288	0.453	293808	0.421	0.494	25008
borrower having any friend	0.191	0.393	293808	0.249	0.432	25008
listing with endorsement+nobid by group leader	0.010	0.098	293808	0.027	0.162	25008
listing with endorsement+nobid by friend	0.120	0.325	293808	0.153	0.360	25008
listing with endorsement+bid by group leader	0.022	0.148	293808	0.117	0.322	25008
listing with endorsement+bid by friend	0.010	0.101	293808	0.041	0.198	25008

Appendix Table 2: Within Grade Learning

	Mean IRR of Loans in Grade:						
	AA	A	B	C	D	E	HR
	coef/t	coef/t	coef/t	coef/t	coef/t	coef/t	coef/t
Panel A							
% of Portfolio Ever Late	-0.031*	0.091*	0.123*	0.139*	0.095*	0.083*	0.122*
	(-7.904)	(7.872)	(14.838)	(14.886)	(7.756)	(4.428)	(4.673)
N	168,400	175,748	205,632	183,744	130,983	57,765	41,166
Panel B							
% of NC Loans Ever Late	-0.004**	-0.002	0.007**	0.004	0.007	0.001	0.005
	(-2.542)	(-0.442)	(2.305)	(1.167)	(1.446)	(0.195)	(0.599)
% of E to HR Loans Ever Late	-0.002***	-0.000	0.003	0.008**	0.010***	0.060*	0.105*
	(-1.773)	(-0.111)	(1.242)	(2.532)	(1.932)	(5.149)	(6.026)
% of B to D Loans Ever Late	-0.012*	0.041*	0.107*	0.136*	0.102*	0.037**	0.000
	(-4.112)	(4.838)	(15.505)	(15.416)	(8.364)	(2.040)	(0.010)
% of A to AA loans Ever Late	-0.020*	0.066*	0.014**	0.015**	-0.006	-0.002	-0.008
	(-5.291)	(6.305)	(2.205)	(2.388)	(-0.750)	(-0.157)	(-0.396)
N	168,400	175,748	205,632	183,744	130,983	57,765	41,166

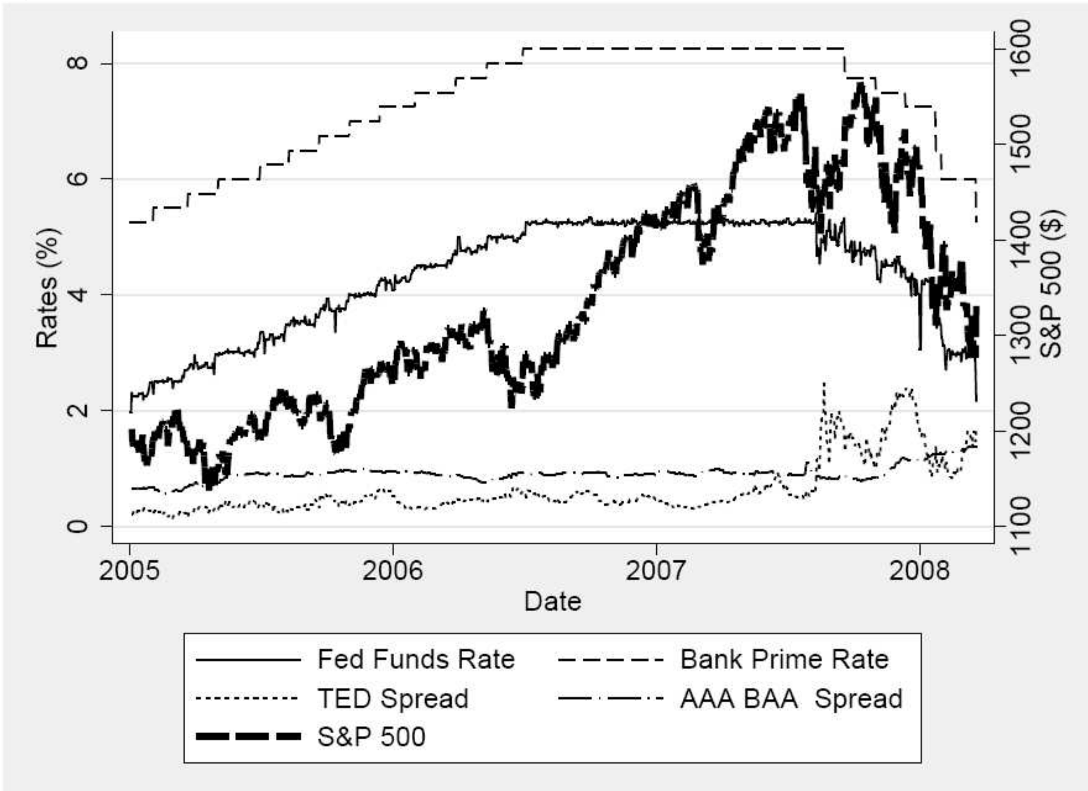
T-statistics are in parenthesis. * p<0.01, ** p<0.05, *** p<0.1. All regression samples are conditional on funding at least one loan in the given grade in week t. Standard errors are clustered at the lender level. Samples in Panel C includes cohorts entering Prosper after June 1, 2006.

Appendix Table 3: Alternative calculation of IRR (conditional on convergence)

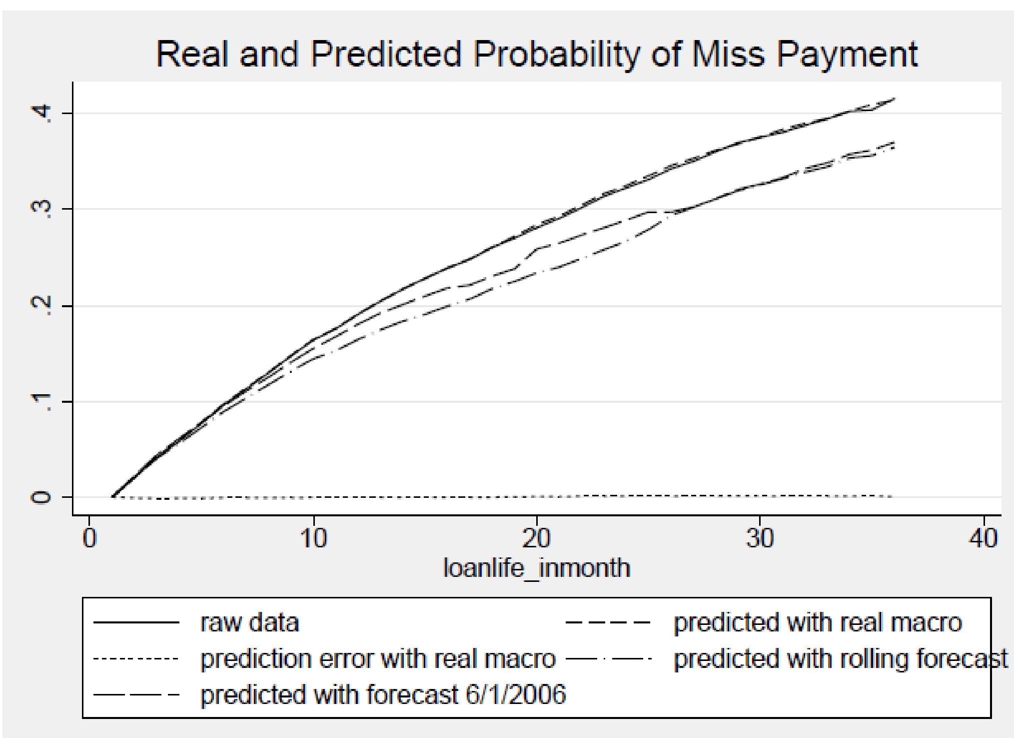
				per	loan		
misperformance measure	model	macro	# of loans	mean	standard deviation	min	max
default	probit	no macro	24997	-0.93%	9.83%	-90.63%	37.56%
default	probit	real macro	5736	-3.67%	12.83%	-90.81%	29.33%
default	probit	rolling forecast	24990	6.60%	15.68%	-99.99%	41.20%
default	probit	forecast 6/1/06	24996	2.23%	12.91%	-99.99%	40.45%
default	duration	no macro	25000	-4.46%	21.18%	-92.46%	28.60%
default	duration	real macro	5631	-12.81%	28.94%	-91.27%	27.75%
default	duration	rolling forecast	24879	8.54%	12.82%	-91.92%	41.20%
default	duration	forecast 6/1/06	24881	4.70%	12.36%	-92.31%	37.35%
default or late	probit	no macro	24979	-12.14%	14.11%	-95.65%	25.28%
default or late	probit	real macro	5731	-16.47%	17.36%	-92.91%	24.36%
default or late	probit	rolling forecast	24983	-1.05%	20.20%	-99.99%	41.20%
default or late	probit	forecast 6/1/06	24986	-8.26%	15.87%	-99.99%	32.86%
default or late	duration	no macro	24991	-10.29%	13.97%	-97.34%	21.94%
default or late	duration	real macro	5628	-10.65%	17.30%	-94.32%	24.02%
default or late	duration	rolling forecast	24878	-0.40%	16.41%	-95.19%	41.19%
default or late	duration	forecast 6/1/06	24879	-4.41%	13.88%	-95.03%	29.56%

In all versions of IRR, the convergence rate is over 99%. Duration model has fewer loans with valid IRR because some loan characteristics perfectly predict the outcome and therefore these loans are dropped in the duration prediction.

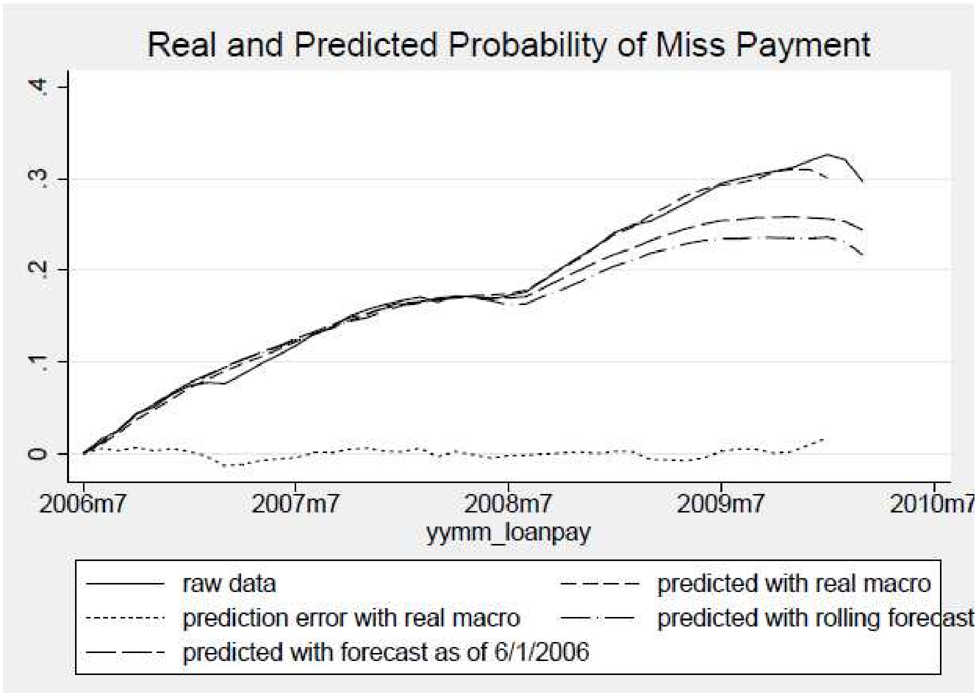
Appendix Figure 1: Macro Economic Indicators (2005 – June 2008)



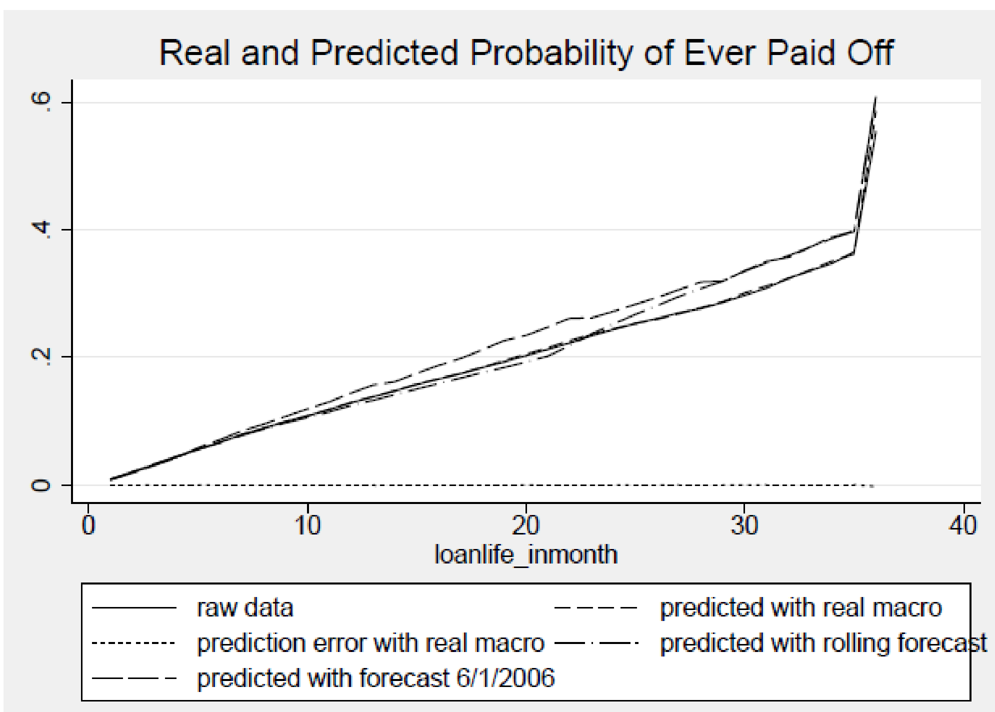
Appendix Figure 2: Fit of the main probit model, in the prediction of miss payment, by loan life



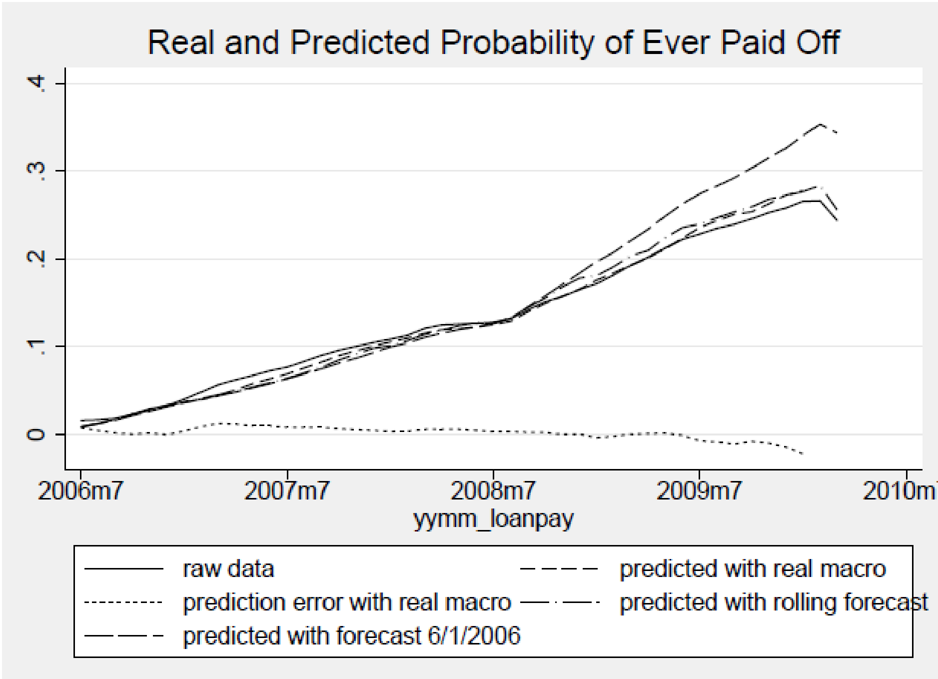
Appendix Figure 3: Fit of the main probit model, in the prediction of default or late, by calendar month



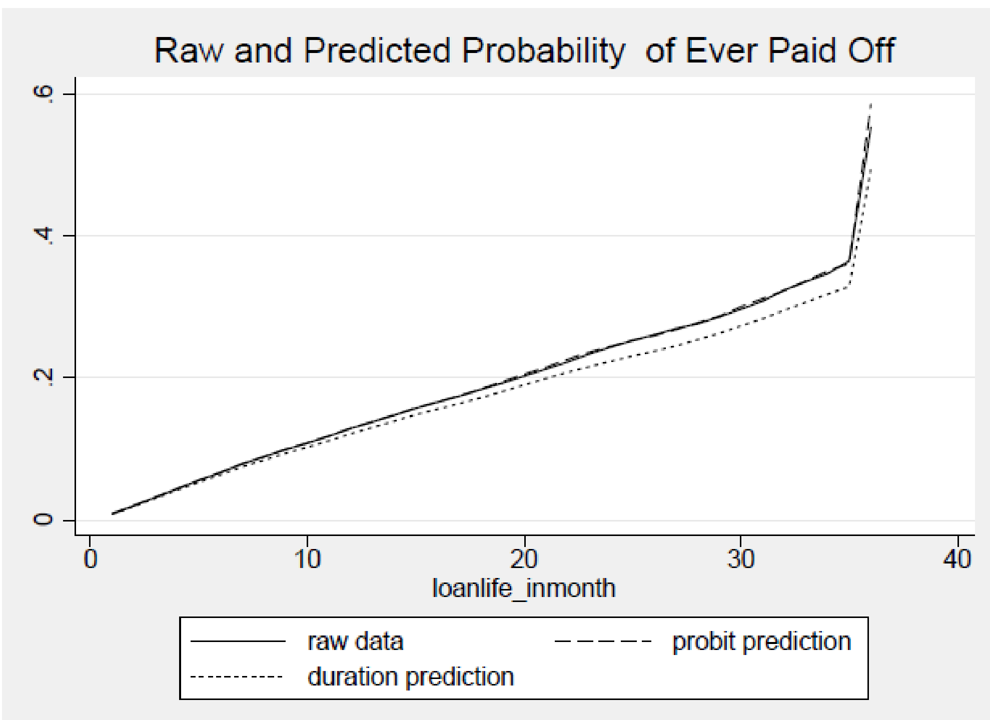
Appendix Figure 4: Fit of the main probit model, in the prediction of ever paid off, by loan life



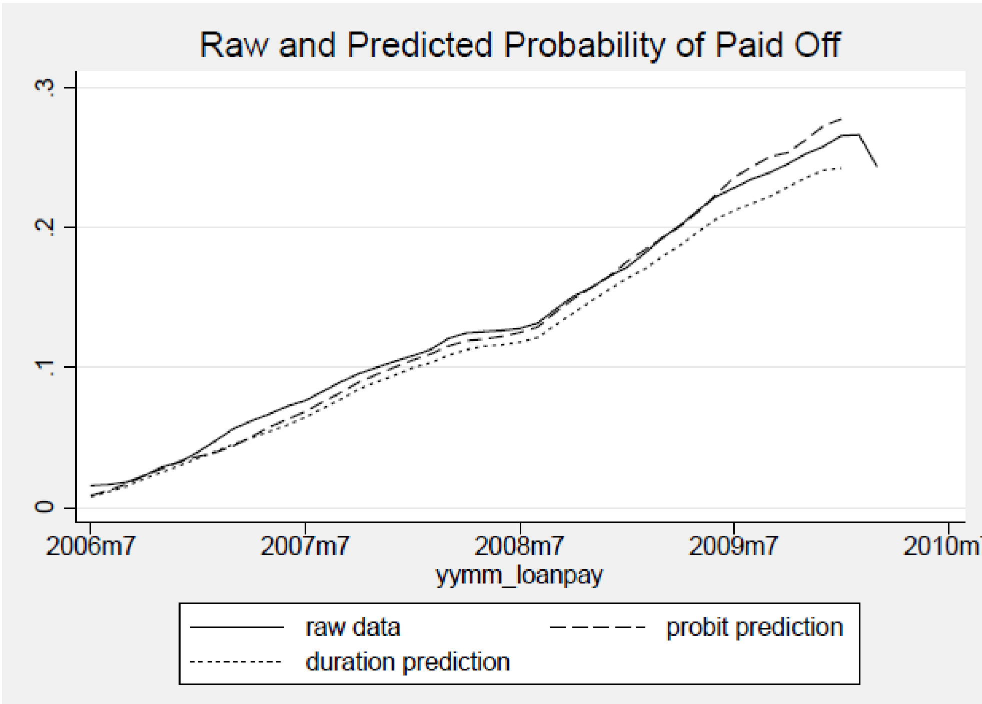
Appendix Figure 5: Fit of the main probit model, in the prediction of ever paid off, by calendar month



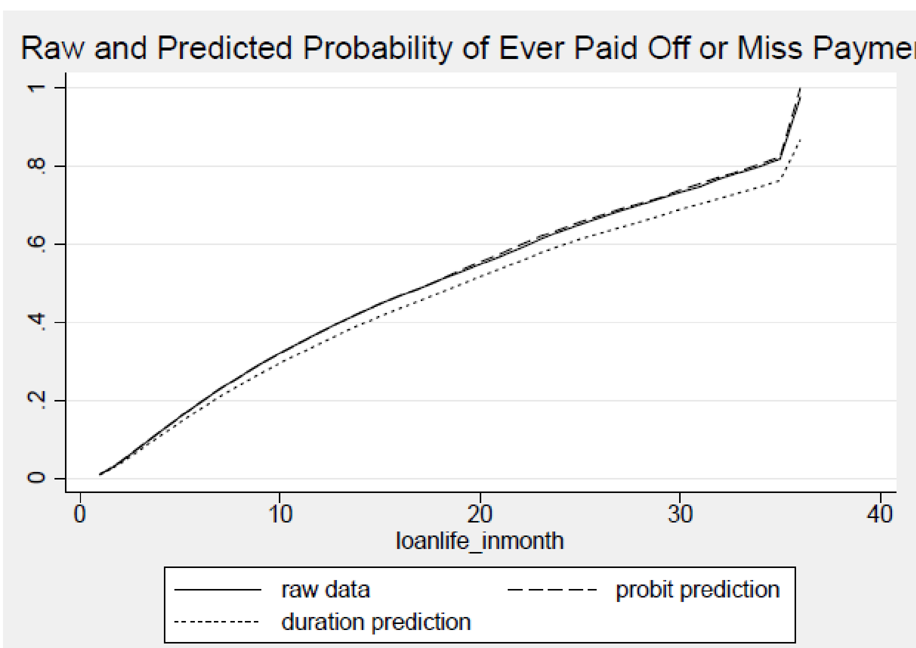
Appendix Figure 6: Compare probit and duration model fit, in the prediction of ever paid off, by loan life



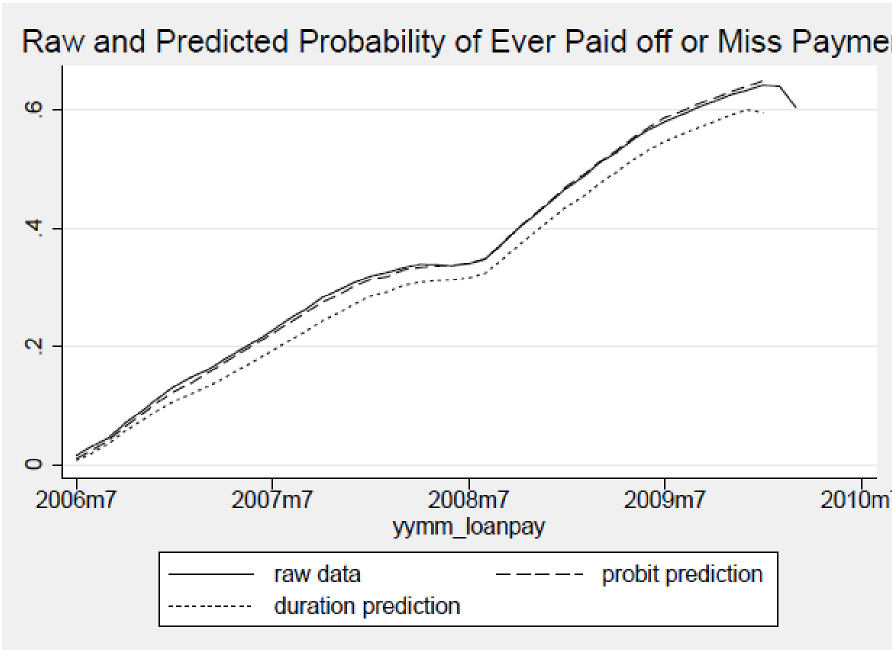
Appendix Figure 7: Compare probit and duration model fit, in the prediction of ever paid off, by calendar month



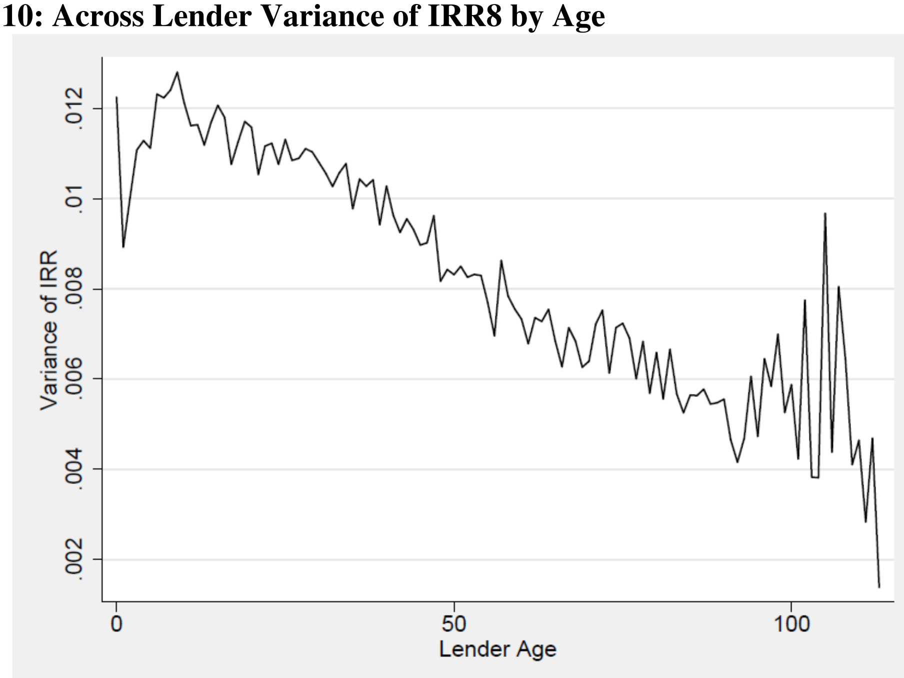
Appendix Figure 8: Compare probit and duration model fit, in the prediction of ever paid off or miss payment, by loan life



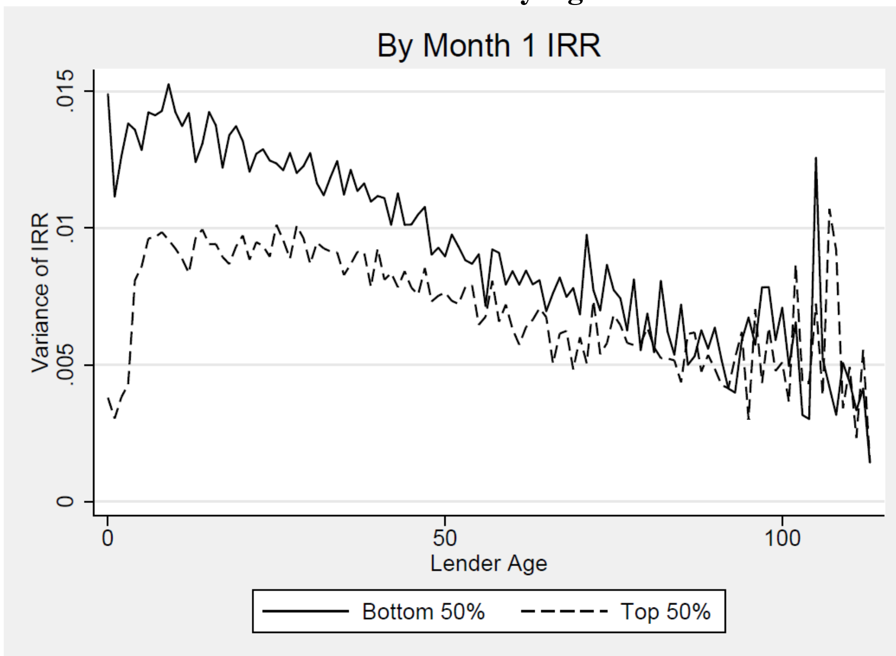
Appendix Figure 9: Compare probit and duration model fit, in the prediction of ever paid off or miss payment, by calendar time



Appendix Figure 10: Across Lender Variance of IRR8 by Age



Appendix Figure 11: Across Lender Variance of IRR8 by Age and Initial Portfolio IRR8



Appendix Figure 12: Whether to fund new loan(s) by lender age (excluding month 1), above- and below-median lenders separately