

YALE DEPARTMENT OF ECONOMICS

The Folly of the Crowd:
Market Inefficiencies in Peer-to-Peer Lending

Samuel Shleifer

Thesis Advisor: Vineet Kumar, Yale School of Management

April 1, 2015

Essay submitted to the Faculty of the Department of Economics in partial fulfillment of the requirements for the degree of Bachelor of Arts at Yale University.

Please print in color.

The Folly of the Crowd: Market Inefficiencies in Peer-to-Peer Lending

Samuel Shleifer

Thesis Advisor: Vineet Kumar, Yale School of Management

Abstract

This essay analyzes the effect of market maker interventions on financial market efficiency, using data from Prosper.com, the first p2p lending website in the US.¹ Between 2006 and 2011, Prosper underwent two major policy changes that guided and then forced willing investors to lend at higher interest rates, as well as to reduce their exposure to subprime borrowers.

During the period before either change, investors set interest rates by auction, made avoidably bad mistakes, and earned poor returns. After Prosper began guiding investors towards setting higher interest rates, they made fewer mistakes and returns improved.

Freedman and Jin (2008) use investor learning to explain this improvement. This paper expands on their work, with the advantage of three additional years of data and two major policy changes. I do not find much evidence for investor learning. Rather, I find evidence that Prosper's rule changes guided and then forced willing investors to make better decisions. When investors set interest rates by auction, investment strategies using publicly available information earn excess returns, an inconsistency with the predictions of the Efficient Markets Hypothesis. Once Prosper begins setting rates, it is no longer possible to beat the market.

*I am grateful to Professor Vineet Kumar for his help in developing the ideas and statistical methodologies involved in this paper, Devin Pope and Rajkamal Iyer for help manipulating the Prosper data, and to Andrei Shleifer, Nicholas Barberis and Andrew Metrick for helpful comments and suggestions.

R code available at github.com/sshleifer/ProsperThesis

¹P2P lending is the practice of individual and institutional investors lending money to borrowers over the internet, without the presence of a financial intermediary

Contents

1	Introduction	5
1.1	Background	7
1.2	Auction Design	9
1.3	Early Auction Regime	9
1.4	Late Auction Regime	10
1.5	Prosper-rate Regime	10
2	Literature	11
2.1	Efficient Markets Hypothesis	11
2.2	Soft Information and Credit Screening	11
3	Data	12
3.1	Variable Definitions for the Prosper Data	13
3.2	Measuring Loan Performance with ROI	15
4	Performance Across Regimes	16
4.1	ROI Across Regimes	16
4.2	Interest Rates and Defaults Across regimes	17
4.3	Impact of Subprime Borrower Ban	24
5	Excess Returns	26
5.1	Puzzle: Bad Returns in Early Auction regime	26
5.2	Methodology	27
5.3	In Sample Results	30
5.4	Out of Sample Results	32
5.5	Methodological Expansions	34
6	Controlling for Time Period	34
6.1	Methodology	34
6.2	Results & Discussion	36
6.3	Comparison to Traditional Loans	38
7	Conclusion	39
8	Appendix	42
8.1	Prosper's Charged-Off vs. Default Terminology	42
8.2	Out of Sample Performance by Credit Grade	42
8.3	Regressions to Predict Interest Rate	42

1 Introduction

Historically, there are two major perspectives on efficient market prices. Fama (1970)'s semi-strong efficient markets hypothesis ("EMH") predicts that a strategy using publicly available information should not be able to earn excess risk-adjusted returns. Risk models such as CAPM imply that risky assets should have positive expected returns and that riskier assets should have higher returns than less risky assets.² The information aggregation perspective associated with Hayek (1945) and Stiglitz and Grossman (1980), further predicts that many individuals, each using their own information, collectively arrive at a more efficient price than a central agent who sets prices, since the agent cannot access all the information available to the crowd.³⁴

In the P2P Lending context, the market's price is the interest rate charged on a loan. This paper studies the question of whether interest rates on Prosper.com are set efficiently. Fama (1970)'s semi-strong EMH predicts that if interest rates were set efficiently, then...

1. Loans would have positive expected returns. Investors would be unwilling to lend to Borrowers if they expected to lose money.⁵
2. Riskier loans would have higher expected returns.⁶
3. It is not possible to do better than the average investor using publicly available information.⁷

Hypothesis 1 is tested by measuring historical returns of loans to borrowers of each credit grade. Hypothesis 2 is tested by comparing the returns of high risk loans to low risk loans. Hypothesis 3 is tested by using publicly available information to select loans with a higher than average expected return, without look-ahead bias.

Prosper.com data provide some additional valuable insights into investor behavior. Over the period I study, there are three interest rate determination regimes:

1. Early Auction Regime (April 19, 2006-October 15, 2008): Rates are set by investor auction. Investors have easy access to useful data, but Prosper does not guide them meaningfully.
2. Late Auction Regime: (July 20, 2009-December 19, 2010): Rates are set by investor auction, with Prosper offering valuable guidance in the form of more complex credit grading.
3. Prosper-rate Regime: (December 20, 2010 - August 3, 2011): Prosper sets rates without investor input using its more complex credit grading algorithm.

²William F Sharpe. "Capital asset prices: A theory of market equilibrium under conditions of risk". In: *The journal of finance* 19.3 (1964), pp. 425–442.

³Sanford J. Grossman and Joseph E. Stiglitz. "On the Impossibility of Informationally Efficient Markets". In: Reprint Series 121 (Dec. 1980). URL: <http://www.nber.org/papers/r0121>.

⁴Friedrich A. Hayek. *The Use of Knowledge in Society*. Library of Economics and Liberty, 1945. URL: <http://www.econlib.org/library/Essays/hykKnw1.html>.

⁵Assuming a Risk-Free Rate of 0.

⁶Sharpe, "Capital asset prices: A theory of market equilibrium under conditions of risk".

⁷Eugene F. Fama. "Efficient Capital Markets: A Review of Theory and Empirical Work". In: *The Journal of Finance* 25.2 (1970), pp. 383–417. ISSN: 1540-6261. DOI: 10.1111/j.1540-6261.1970.tb00518.x. URL: <http://dx.doi.org/10.1111/j.1540-6261.1970.tb00518.x>.

Additionally, for regimes (2) and (3), subprime borrowers are not allowed to apply for Prosper loans.

A comparison of loan performance in different regimes allows me to consider two contrasting views of how an internet marketplace such as Prosper works. One view, closely linked to Hayek, suggests that the marketplace would work best with minimal intervention from the market maker Prosper, so many individual lenders incorporate their private knowledge and information in the determination of the interest rate. An opposing view would hold that individuals are not sophisticated or knowledgeable enough to assess loan risk, and some guidance or even interest rate setting from Prosper, the market maker, would improve loan performance. The evidence on loan returns in the three regimes points quite clearly in favor of the market maker hypothesis as Prosper becomes more involved in the rate-setting process, the excess returns available using publicly available information approach 0, while aggregate investor returns exhibit higher mean and lower variance.

A major confound of this analysis is the time period. Perhaps 2006-2008, the early auction period, is just a bad time to invest in consumer credit, whereas 2009-2011 is much better. To address this concern, I estimate the importance of time effects by comparing default rates of loans to similar borrowers in different periods. I then show that time period is not the only explanation of the various changes to the marketplace by estimating what loan performance would have been if pre-shutdown borrowers borrowed at the higher Prosper-set interest rates.

This hypothetical method suffers from the problem that it is impossible to perfectly simulate the effect of charging risky borrowers higher interest rates during a turbulent macroeconomic period, as Section 6 attempts. A perfect dataset would have Prosper randomly assign different rate-setting mechanisms to borrowers, rather than assigning all borrowers in a given time period the same rate-setting mechanism, and then switching it when things aren't going well.

In this spirit, I do not claim that poor investor decision making is the only factor causing poor returns in the Early auction period, or that the impact of the policy changes can be fully isolated by my methodologies. The inconsistencies with EMH predictions 1 and 2 can certainly be partially explained by the same bad luck that explains poor returns on all risky assets during economically turbulent times. I merely argue that returns in the early auction period were likely made worse by poor investor decisions, as evidenced by the predictable mispricings that could be exploited. In Section 5, I show that over a somewhat longer period of time, simple loan selection strategies can use publicly available information to earn higher returns than the benchmark.⁸

Much of the evidence presented in this essay is inconsistent with an explanation that all of the differences between regimes can be explained by differences in time period. For example, in the late auction regime, when investors set interest rates, strategies using publicly available information can earn excess returns. In the Prosper-rate regime, when Prosper sets interest rates directly, they cannot. The two periods are very similar macroeconomically. Perhaps most tellingly, in the period when Prosper sets rates it seems to use similar variables to those that were helpful in selecting the best performing loans in the early auction regime, pointing strongly toward the conclusion that early investors were not using publicly available information effectively.

⁸Benchmark is calculated by assuming \$1 investment in every available loan.

Table 1: Timeline of Events

Apr. 21, 2006	•	First loan of early auction regime.
Mar. 1, 2007	•	Prosper bans borrowers with credit score ≤ 520 .
Oct. 15, 2008	•	Early Auction period ends with SEC enforced shutdown.
Jul. 20, 2009	•	Prosper reopens with more informative credit grades, requires repeat borrowers to have credit score ≥ 600 , and requires new borrowers to have credit score ≥ 640 . Beginning of Early Auction Regime.
Dec. 19, 2010	•	Prosper-rate regime, where Prosper sets rates directly, begins, because “auctions made the deployment of funds more time consuming with little gain in lender returns. - Prosper CEO Chris Larsen. ¹²
Aug. 3, 2011	•	Last loan in Prosper-rate regime originated.

1.1 Background

Peer-to-peer lending (“P2PL”) is the practice of lending money to borrowers over the internet, without going through a traditional financial intermediary.⁹ From April 2006 until December 2010, interest rates were set by lenders¹⁰, who competed for the lowest rate in a reverse auction model.

In October 2008, Prosper was shut down by the SEC because it was designated as a seller of securities,¹¹ and re-opened in July 2009 after a 9 month quiet period with more stringent underwriting standards, and greater guidance for lenders. Importantly, all the information about borrower creditworthiness that is used in this paper is similar to what was available to investors when they bid on loans. Since December 2010, the beginning of the *Prosper-rate* regime, rates have been determined by Prosper, on the basis of a borrower’s credit history, as provided by Experian.

Under all three regimes, lenders could choose to whom to lend and how much to lend, allowing them to diversify across many loans, thereby decreasing the variance of their returns. The minimum investment in a loan was \$25, with no maximum, and lenders could hold as many loans as they wished. Before a potential borrower’s listing is posted on the website, Prosper authenticates their social security number, driver’s license and address, and pulls the borrower’s credit history from Experian, which includes credit score, total number of delinquencies, current delinquencies, and many other historical credit variables.¹³ In exchange for verifying borrowers’ identities and managing the marketplace, Prosper charges a 0.5-3 percentage point origination fee on each loan depending on the borrower’s risk profile and loan duration. This fee, usually 1%, is added to the borrower’s monthly payment. Lenders are also charged a servicing fee of 0.5-1%. There are no prepayment penalties. This rate is lower than that charged by most traditional financial institutions, which have the higher costs of physical branches, and pay for human analysis of credit history.¹⁴ The P2PL

⁹Wikipedia. *Peer-to-Peer Lending*. URL: en.Wikipedia.org/wiki/Peer-to-peer_lending (visited on 01/01/2015).

¹⁰For the purposes of this essay, lender and investor are synonyms.

¹¹Rather than, as it argued, a marketplace connecting buyers and sellers.

¹³Seth Freedman and Ginger Zhe Jin. “Do social networks solve information problems for peer-to-peer lending? evidence from prosper.com”. In: (2008).

¹⁴Samuel G Hanson et al. “Banks as Patient Fixed-Income Investors”. In: (2014).

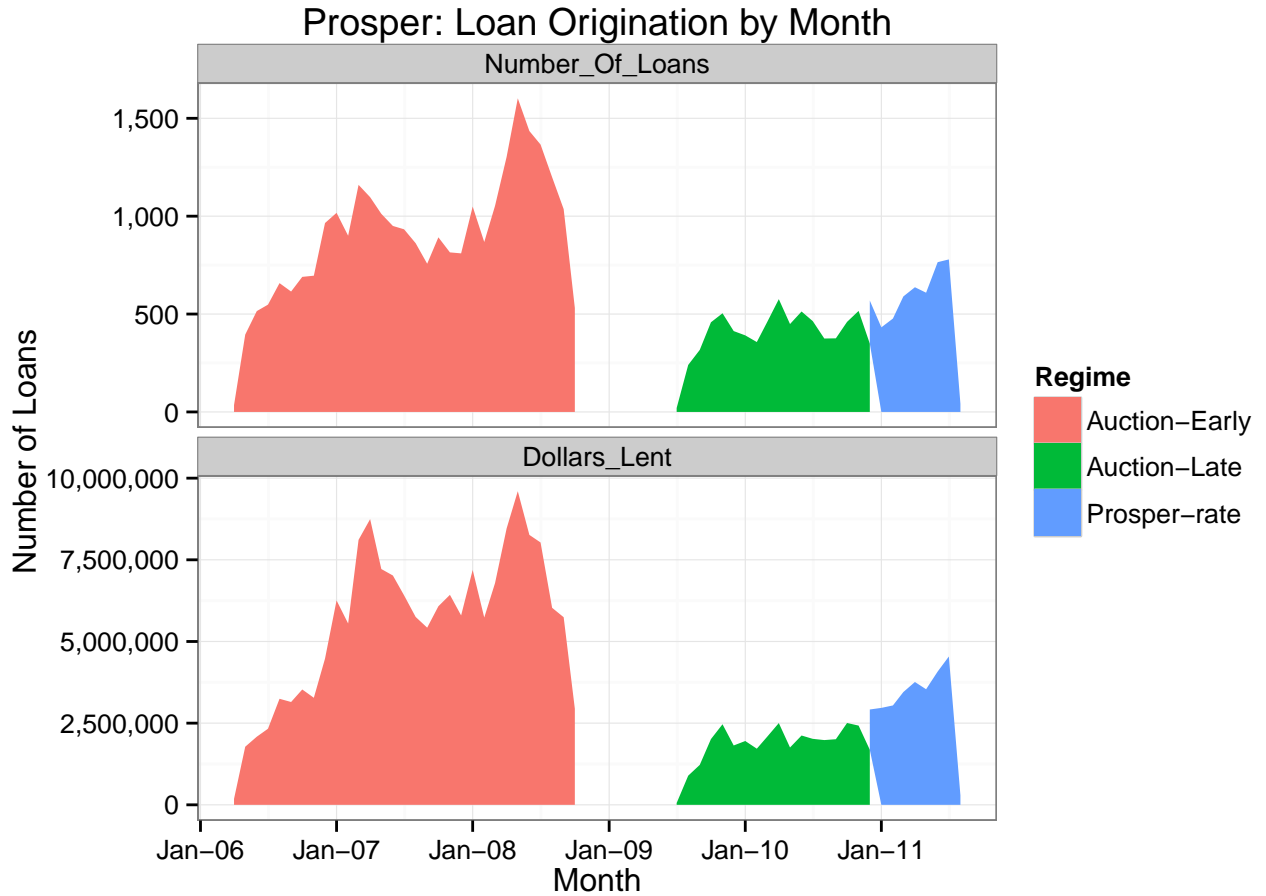


Figure 1: This chart shows two panels. The first panel shows the number of loans funded on Prosper.com in each month of the sample. The second panel shows the total dollars lent to borrowers on Propers.com in each month of the sample.

model relies on lenders being able to infer the riskiness of borrowers: since Prosper gets fees from monthly payments, they too are incentivized to maximize investor returns, both for short term profit, and long term reputation.

In my sample, Prosper’s origination volume, as shown in Figure 1, peaked in the months before the shutdown, and then slowly resumed after the quiet period. Since 2011, when the sample ends, origination volume has eclipsed its 2008 peak.

Reasons for borrowing on Prosper, as shown in Table 2, have changed relatively little over the site’s 8 year history: the plurality of borrowers aim to consolidate credit, while fewer use the funds to start a business, or pay for a wedding or a car. Morse (2015) finds that “Borrowers are characterized as debt-laden, middle-to-high income, individuals who are consolidating credit cards and other debt.” The average P2P loan face value comprises 20.5% of annual income, and payments absorb 7.5% of monthly income.¹⁵

¹⁵ Adair Morse. “Peer-to-Peer Crowdfunding: Information and the Potential for Disruption in Consumer Lending”. In: Working Paper Series 20899 (Jan. 2015). DOI: 10.3386/w20899. URL: <http://www.nber.org/papers/w20899>.

	Purpose	n	Frequency
1	Debt Consolidation	10531	0.446
2	Other	4167	0.176
3	Business	3016	0.128
4	Personal	2365	0.100
5	Home Improvement	1783	0.075
6	Auto	1017	0.043
7	Student Use	748	0.032

Table 2: Stated reasons for borrowing on Prosper (2006-2001). NAs removed.

1.2 Auction Design

From April 19, 2006 until December 19, 2010. Prosper operated a *variable rate* model, where lenders and borrowers determined interest rates using a Dutch auction-like system, where each investor bid consisted of a (dollar amount, interest rate) pair, such that the dollar amount is greater than \$50, the minimum bid, and the rate is capped at the borrower’s reservation rate.¹⁶ In these auctions, borrowers would declare a maximum reservation interest rate, capped at 36%, where bidding would start. Over the course of a 7-10 day bidding period, that rate would be bid down by investors, with the loan’s interest rate closing at the lowest rate $R_{clearing}$ such that enough investors were willing to fully fund the loan at or below the $R_{clearing}$. All investors who bid at or below the clearing rate would receive the clearing rate. If a borrower could not reach 100% funding at his reservation rate, the loan would not be funded.

As an illustrative example, suppose Bill the borrower needed to borrow \$100, and set his maximum rate at 20%. Let Chris, David and Ellen be investors, with Chris bidding (\$50, 15%), David bidding (\$50, 18%) and Ellen bidding (\$30, 19%). In this scenario, Chris and David would win the auction, and each lend Bill \$50, and receive an 18% interest rate. Bill would pay 18% interest before fees.

1.3 Early Auction Regime

In the early auction regime, loans originated between April 19, 2006 and October 15, 2008, auction participants had access to considerable information from Prosper, including over 60 Borrower characteristics, such as residence, employment status and an Experian credit score bucket, but few intelligent risk estimates. Additionally, Prosper did not regulate who could apply for a loan on the website. Besides basic fraud prevention, any applicant who received 100% funding from investors was eligible, for Prosper’s first year of operation. On March 1, 2007, amidst poor returns, Prosper announced that listings with credit scores below 520 would no longer be offered to investors. This intervention is not discussed in depth, for lack of space, but shows that after a year, more or less, Prosper felt the need to nudge a fully decentralized investment process in the right direction.

In an October 15, 2008 email, Prosper announced that it would temporarily stop accepting new loan applications to register with the SEC, which subsequently released a cease and desist order

¹⁶Freedman and Jin, “Do social networks solve information problems for peer-to-peer lending? evidence from prosper.com”.

characterizing Prosper as a “seller of securities,” rather than a marketplace merely matching lenders and borrowers, as the site had argued.¹⁷

1.4 Late Auction Regime

After a 9 month quiet period, Prosper re-opened, using the same Auction model, with three significant market interventions. First, under-writing standards again became more stringent: new borrowers were required to have a credit score above 640, and repeat borrowers were required to have a score above 600.

Second, whereas previously Prosper used only the credit score to determine a borrower’s Credit Grade (eg. AA), the site began offering a more advanced, multi variable formula to determine each borrower’s credit grade for investors consideration. Finally, the site also partnered with FolioFN, a secondary market where lenders could liquidate their positions. Crucially, Prosper still had no direct role in setting interest rates, investors still used an auction, but with the extra information credit grades that more closely matched default probabilities. Besides credit score, Prosper began using (and still uses) the following “Key variables” to determine its revamped credit grades:¹⁸

- TotalInquiries: The number of times potential lenders requested the borrowers’ credit portfolio from the credit bureau, which is usually Experian or a competitor.
- CurrentDelinquencies: The number of delinquent loans associated with the borrower.
- Credit card utilization: how much of available credit the borrower uses on a monthly basis.
- Number of open credit card accounts
- Debt to income ratio
- Loan payment performance on prior Prosper loans.

This Auction rate-setting mechanism continued for more than a year after the shutdown, until December 19, 2010.

Unsurprisingly, but critically for the interpretation of the evidence, the automated feature selection procedure in section 5, which identifies the subset of variables that best predict default, finds that credit Score, inquiries, delinquencies and number of credit cards are all useful independent variables for predicting default.

1.5 Prosper-rate Regime

Under the current *Prosper-rate* model, enacted December 20, 2010, lenders no longer determine the loan rate in an auction. Instead, Prosper uses the intelligent credit grades given to investors in the late auction regime to set interest rates on loans, with slightly less stringent under-writing standards to determine eligibility. The change was motivated by a desire to accelerate the listing

¹⁷Securities and Exchange Commission. *Order Instituting Cease and Desist Proceedings in the matter of Prosper Marketplace*. 2008. URL: <http://www.sec.gov/litigation/admin/2008/33-8984.pdf> (visited on 01/01/2015).

¹⁸Prosper Blog. *Estimated Loss Rates*. URL: https://www.prosper.com/help/topics/general-estimated-loss_rates/ (visited on 01/01/2015).

process, so that investors could reinvest money more quickly. Moreover, towards the end of the late auction regime, auctions were arriving at interest rates that were almost completely determined by the credit grade. Under the fixed-rate regime, investors choose whether or not to invest in a given loan, at the interest rate set by Prosper, so bids consist only of a dollar amount.

2 Literature

2.1 Efficient Markets Hypothesis

In layman's terms, the efficient markets hypothesis (EMH), states that 'it is impossible to beat the market.' More specifically, Fama explains, market prices should always fully reflect publicly available information. Therefore, EMH rules out the possibility of strategies that earn excess returns using only publicly available information.¹⁹ In the P2P lending context, this market price is the interest rate determined by investors (or accepted by investors in the fixed-rate regime), and paid by the borrower, and returns are the cash flows received from borrowers divided by the amount invested.

Fama, (1970) notes the dependence of most tests of market efficiency on a model of risk and return. The models, including most famously, Sharpe's capital asset pricing model (CAPM), make two predictions: First, since an investor can always hold cash risklessly, the average return on risky investments must be positive. Second, securities with higher systematic risk earn higher average returns.²⁰

The EMH rests on three assumptions. "First, investors are assumed to be rational and hence to value securities rationally. Second, to the extent that some investors are irrational, their decisions are random and uncorrelated. Third, to the extent that some investors are irrational in similar ways, they are met in the market by rational arbitrageurs, who eliminate their influence on prices."²¹

In a reverse interest rate auction, such as Prosper's market design, arbitrageurs cannot short a borrower and increase his interest rate. In the situation where irrational investors under estimate a borrower's riskiness, they will outbid rational bidders, and set the market price. To earn excess returns, then, an investor must select loans for which other investors have demanded too high a rate, so that he can participate in funding.

2.2 Soft Information and Credit Screening

The novel idea of P2P is a large group of investors may be better equipped to assess risk than traditional financial institutions. "Traditionally, the credit score provided by rating agencies has been the main tool banks use to screen smaller borrowers."²² As such, the initial strand of literature focuses largely on the novel sources of information utilized by P2P investors. Gao and Lin (2012) analyze the text written by the borrower in the "description field" of the loan application, and find that descriptions that are easy to read have 2.3% lower default rates than average, while complex

¹⁹Fama, "Efficient Capital Markets: A Review of Theory and Empirical Work".

²⁰Sharpe, "Capital asset prices: A theory of market equilibrium under conditions of risk".

²¹Andrei Shleifer. *Inefficient Markets*. 2000.

²²Rajkamal Iyer et al. "Screening in New Credit Markets: Can Individual Lenders Infer Borrower Creditworthiness in Peer-to-Peer Lending?" In: rwp09-031 (Aug. 2009). URL: <http://ideas.repec.org/p/ecl/harjfk/rwp09-031.html>.

narratives are associated with a 3.6% higher default rate. They do not measure returns.²³ Pope and Sydnor (2008) find that investors reduce their defaults by statistically discriminating against racial minorities, but also do not analyze return on investment. Iyer et. al (2009) find that within the credit score buckets provided by Prosper, investors use soft information to further sort borrowers within buckets, reducing defaults.²⁴

These papers echo Hayek (1945)’s basic argument for market efficiency: that many individuals, each using their own information, collectively arrive at one efficient price. The right price “might have been arrived at by one single mind possessing all the information which is in fact dispersed among all the people involved in the process.”²⁵

Freedman and Jin (2008) are more focused on the failures of the auction regime, Although the main focus of their paper is the role of soft information and social networks on Prosper, they also study how investors learn to weight these different variables, and, like this paper, find negative expected returns in the early auction regime: “lenders, especially those that joined Prosper early, did not fully understand the market risk. We estimate, on average, Lenders would have expected an internal rate of return of between -.62% and 1.38% if they had correctly understood the risk distribution of Prosper loans.” “Vigorous” lender learning, they argue, explains the increase in interest rates and borrower creditworthiness, and associated increase in lender returns.²⁶

This essay continues Freedman and Jin’s critical look at investor decisions in the early auction regime, but with the benefit of three more years of data, and two more major Prosper policy changes: informative grading and centralized rate-setting. Instead of a focus on soft information and social networks, I describe the hard information that went largely ignored by investors until Prosper aggregated it into more informative credit grades. Instead of investor learning as the catalyst that led to a more efficient market, I frame Prosper’s rule changes as catalysts that force investors to make better decisions.

Inconsistent with Hayek, the data suggest that when investors are left on their own to set prices, markets are not nearly as efficient as the prices set by “a single mind”, namely a market maker who aggregates the data. Furthermore, as interest rate setting becomes more centralized, interest rates become more sensitive to the hard credit variables that best predict default.

3 Data

My sample consists of 39,555 three year, unsecured (no collateral) Loans, worth \$231.7 Million, originated from 2006-2011. The data was downloaded off Prosper.com. Borrowers apply for loans by creating a public listing on the Prosper.com website, and can request anywhere from \$1,000 to \$35,000.²⁷ Since our analysis requires measurement of loan performance, which can only take place three years after origination (given 3 year maturity on all loans), the sample is ended in 2011, as we do not know what will end up happening to more recently originated loans. As shown in Table 3,

²³Qiang Gao and Mingfeng Lin. “Linguistic Features and Peer-to-Peer Loan Quality: A Machine Learning Approach”. In: *Available at SSRN 2446114* (2013).

²⁴Iyer et al., “Screening in New Credit Markets: Can Individual Lenders Infer Borrower Creditworthiness in Peer-to-Peer Lending?”

²⁵Hayek, *The Use of Knowledge in Society*.

²⁶Freedman and Jin, “Do social networks solve information problems for peer-to-peer lending? evidence from prosper.com”.

²⁷The minimum increased to \$2,000 effective Nov. 1, 2011

this leaves us with many more loans from the early auction regime than the post shutdown regimes.

	Regime	n	Freq.
1	Auction-Early	27764	0.70
2	Auction-Late	7238	0.18
3	Prosper-rate	4550	0.12

Table 3: Observations per regime. Each observation is one funded loan.

To help investors better assess the creditworthiness of borrowers, for every loan originated in the site’s history, Prosper posts historical data regarding loan performance, such as payments made, interest rate, and principal lost. Additionally, borrower characteristics, including credit variables obtained from Experian and verification of which are usually obtained directly from Experian and verification of income and employment, are available to investors for each listing. Moreover, historical market data is available for download, through a query tool, and through third party websites. Interviews conducted at the 2008 Prosper Days Conference suggest that there was “enormous heterogeneity in lender awareness of the data, ability to process the data, and intent to track the data over time.”²⁸

3.1 Variable Definitions for the Prosper Data

For each loan, Prosper offers 63 variables, I now define an important subset.

- **BorrowerRate:** Annual interest rate paid by the borrower.
- **TotalCreditLinespast7years:** Number of credit lines, including cards and loans, used by the borrower in the seven years before the prosper loan was originated.
- **Inquirieslast6Months:** Number of times a lender inquired about the Borrower’s credit report, in the 6 months before the loan was originated. Traditional lenders make inquiries when a borrower applies for a loan, so this serves as a proxy for the number of times someone has borrowed in the past 6 months.
- **CurrentDelinquencies:** Number of loans on which the borrower is delinquent at the origination date.
- **DelinquenciesLast7Years:** The number of loans which the borrower has not repaid fully over the 7 years before the origination date.
- **MonthlyLoanPayment:** The borrower’s scheduled monthly payment on the Prosper loan.
- **CreditScoreRangeLower:** The lower bound of the credit score range (which bounds his raw credit score) , that bounds a borrower’s credit grade, as provided by Experian. Prosper provides the range instead of the raw score to preserve borrower anonymity.

²⁸Freedman and Jin, “Do social networks solve information problems for peer-to-peer lending? evidence from prosper.com”.

	Auction-Early	Auction-Late	Prosper-rate
roi.mean	-0.03754	0.15381	0.13493
roi.sd	0.4152	0.3226	0.3756
default.rate.mean	0.3700	0.1585	0.2112
default.rate.sd	0.4828	0.3652	0.4082
LenderYield.mean	0.1730	0.1939	0.2205
LenderYield.sd	0.07266	0.09655	0.08532
BorrowerRate.mean	0.1833	0.2040	0.2305
BorrowerRate.sd	0.07391	0.09654	0.08531
LoanOriginalAmount.mean	6166	4759	5956
LoanOriginalAmount.sd	5562	3820	3613
Investors.mean	118.13	144.17	86.91
Investors.sd	117.74	125.89	74.38
CreditScoreEst.mean	653.2	714.2	705.1
CreditScoreEst.sd	88.71	58.85	53.29
DebtToIncomeRatio.mean	0.3284	0.2292	0.2458
DebtToIncomeRatio.sd	0.9423	0.2681	0.3769
InquiriesLast6Months.mean	2.879	1.031	1.132
InquiriesLast6Months.sd	3.911	1.532	1.640

Table 4: This table presents means (.mean) and standard deviations (.sd) for relevant variables by regime.

- CreditScoreEst: The average of the lower and upper bounds of the credit score range.
- PublicRecordsLast10Years: Number of public records of borrowing activity in the past 10 years at the origination date.
- DebtToIncomeRatio: The debt to income ratio of the borrower at the origination date. This value is capped at 10.01 (any debt to income ratio larger than 1000% will be coded as 1001%).
- TotalProsperPaymentsBilled: Number of on time payments the borrower made on Prosper loans at the time they created the listing. This value is 0 if the borrower had no prior loans.
- LenderYield: The lender’s yield on the loan. Equal to the interest rate minus servicing fee.
- BorrowerRate: The borrower’s interest rate for the loan.
- Investors: The number of investors that funded the loan.
- *.sc: The Variable * was scaled to have 0 mean and unit variance. (found only in regression results)

3.2 Measuring Loan Performance with ROI

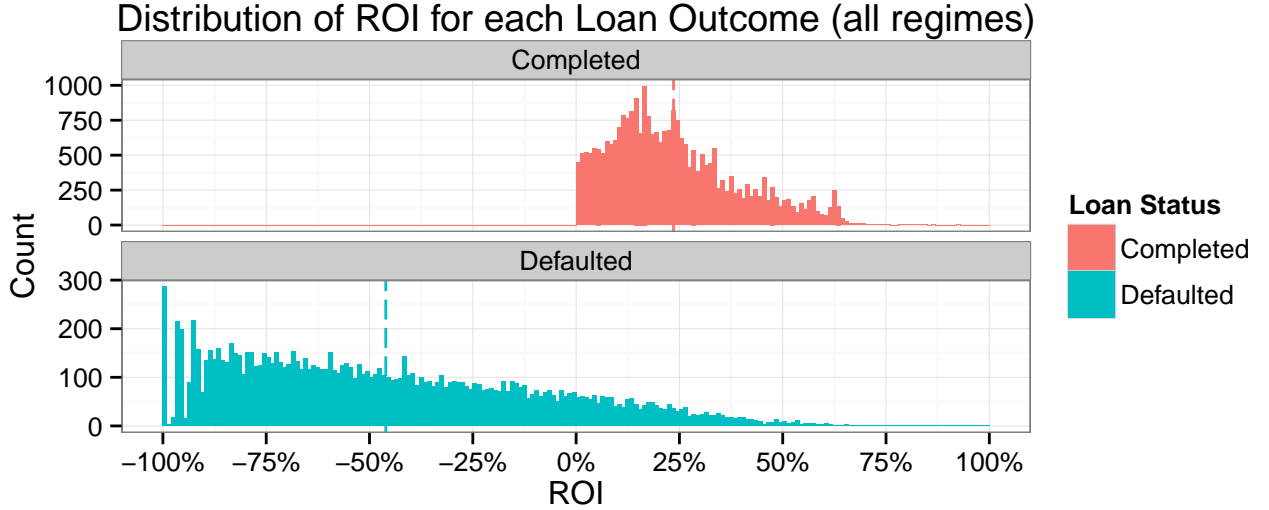


Figure 2: This chart shows the distribution of ROI for all completed and defaulted loans in the sample. A continuous ROI variable offers more information about the quality of an investment than a binary default indicator. Dotted lines indicate mean ROI for category.

In order to capture the actual impact of a loan on an investor’s portfolio, we measure the ROI for a given loan as the sum of the borrower’s payments / amount borrowed -1.

$$ROI = \frac{amount.repaid}{amount.borrowed} - 1 \quad (1)$$

As shown in Figure 2, a return on investment variable offers much more granularity than a binary default variable for both defaulted and repaid loans. In many observations, a borrower paid back over half his principal before defaulting. A binary indicator would treat this loan, which only cost the investor half his money, as identical to a loan that repaid him nothing. Additionally, a default indicator fails to account for the higher interest payments associated with riskier loans. Still, measuring ROI and default lead to similar conclusions, as the correlation between ROI and default is -.80 over the full sample.

I do not discount or annualize ROIs, because in the period of low interest rates, the difference between default and repayment is large enough that a few percentage points don’t matter: loans that default average -46% ROI, loans that repay average + 24% ROI. Furthermore, my ROI results are very similar to Freedman and Jin’s, who use discounting.²⁹ Although their sample is shorter in length, using discounted ROIs for our sample would only make the results more extreme, as risk-free interest rates were highest during the early auction regime.^{30 31}

²⁹Freedman and Jin, “Do social networks solve information problems for peer-to-peer lending? evidence from prosper.com”.

³⁰Board of Governors of the Federal Reserve System (US). *Effective Federal Funds Rate [FEDFUNDS]*. URL: <https://research.stlouisfed.org/fred2/series/FEDFUNDS/>.

³¹The fed funds rate’s peak during our sample, is at 5% in 2007, during the early auction regime. It then falls

The rest of this paper presents my results, which are divided into three sections. Section 4 presents facts about the performance of loans across regimes which suggest that early auction regime investors were not sufficiently sensitive to risk, and that this partly explains the poor returns of the period. Section 5 tests and rejects the EMH prediction that it should be impossible to earn excess returns using publicly available information. Section 6 attempts to isolate the impact of Prosper’s policy changes by controlling for time period effects, and finds that they explain at least some of the improvement in investor returns.

4 Performance Across Regimes

This section begins with an examination of ROIs over time and across borrower credit buckets. I then discuss regime differences in default rates and interest rates, the two features that largely determine returns. The discussion of interest rates includes measurement of investors’ varying sensitivity to risk across regimes. The section closes by examining the impact on investor returns of Prosper’s interventions, which reduce the valid borrower pool to a less-risky subset and assign borrowers more informative credit grades.

I find the following: (1) the mean ROI of early auction regime loans is negative, (2) in the early auction regime, average ROI is lower for higher risk loans, (3) interest rates are insufficiently sensitive to risk in the early auction period and (4) investors made sub-optimal choices, given the information they had at their disposal, and (5) In the period when Prosper intervened, interest rates and investor returns were much higher, and part of the increase is likely due to Prosper’s intervention.

These findings suggests that although investors were also unlucky in the early auction regime, they made avoidably bad decisions that hurt their returns. More specifically, Finding (1) is inconsistent with the EMH prediction that risky assets like Prosper loans should offer investors a positive expected return.³² Finding (2) is inconsistent with the prediction that riskier loans should have higher expected returns to compensate investors for taking risks. Finding (3) describes the sub-optimal investor decisions that led to (1) and (2), and in part addresses arguments that investors chose an unlucky time period. Finding (4) is inconsistent with Hayek’s notion that a free market will arrive at accurate prices that reflect the information available to market participants, without the meddling of a market maker.

As the default data demonstrate, Finding (3) is only partially responsible for Findings (1) and (2); investors were incredibly unlucky to have chosen to invest in risky assets during turbulent macroeconomic times that many of the most sophisticated investors did not see coming. In section 6, I attempt to control for this time period effect to isolate the impact of the policy changes.

4.1 ROI Across Regimes

Figure 3 shows the performance of loans originated in each month of the sample. The blurred lines represent the mean return across all loans for each month, while the solid lines represents the mean return to borrowers in the 640-659 credit score bucket, which was chosen because it is the most

to near 0 for the duration of the later two regimes. Discounting would thus lead to significantly lower early auction regime ROIs, but only slightly lower ROIs in the later two periods.

³²Otherwise, investors would just hold cash.

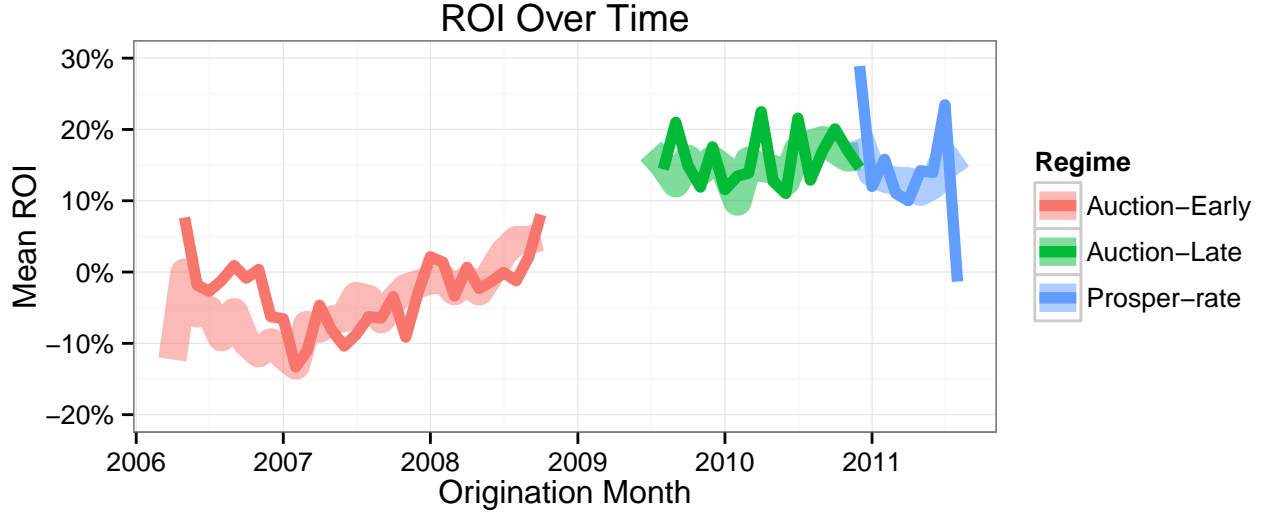


Figure 3: This chart shows the mean ROI of loans originated each month in our sample. The thick blurred lines show the average ROI across all loans. The solid lines show is the average ROI for loans to borrowers in the 640-660 credit score bucket. The color of each line represents the regime. Loans funded after the shutdown are associated with higher ROI, even controlling for borrower creditworthiness.

common in the full sample. The average ROI of all loans increases from -3.8343% in the early auction period to 15.2047% in the late auction period. The average return for loans originated in each of the site's first 27 months is negative, rejecting the EMH prediction of positive expected returns for risky assets. Of course, the time period is unusual. I address this possibility in section 6.

Figure 4 shows that lower credit score loans earn lower average returns in the early auction regime, inconsistent with the EMH prediction of a risk/reward tradeoff. The expected pattern is shown in the two later regimes' curves: riskier loans earn higher average returns. Figure 4 also reveals early auction regime returns are worse than the other two regimes' returns for each bucket, but that the gap is much greater for riskier loans. Again, time period may be a confounding factor, as riskier investments often perform worse in bad economic times, a possibility I discuss later. We now examine default rates and interest rates, the two features that explain the variation in returns.

4.2 Interest Rates and Defaults Across regimes

Figure 5 presents the median interest rate³³ for each month of the sample. The solid line again represents borrowers in the 640-660 credit score bucket, while the blurred line represents all borrowers. Overall (blurred line), the median interest rate paid by borrowers increases from 17.4% to 19.6% to 22.5% over the early auction, late auction and Prosper-rate regimes. When we control for borrower selection by only considering the 640-660 credit score bucket, however, the climb in median interest rate (from 17.1% in the early auction period to 26.4% in the late auction period) is more dramatic. Interestingly, rates for the 640-660 bucket are almost identical directly before and right after the

³³Before Fees.

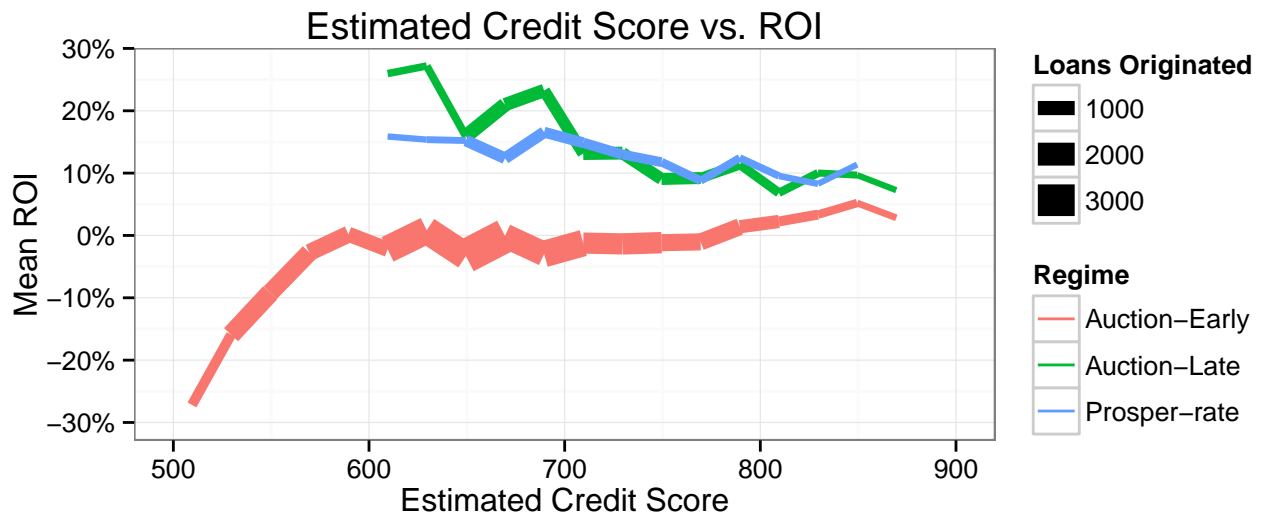


Figure 4: This chart shows the mean ROI by borrower credit bucket, under each regime. The thickness of the line represents the number of loans originated to borrowers in the credit bucket. In the early auction regime, higher risk loans earn lower average returns.

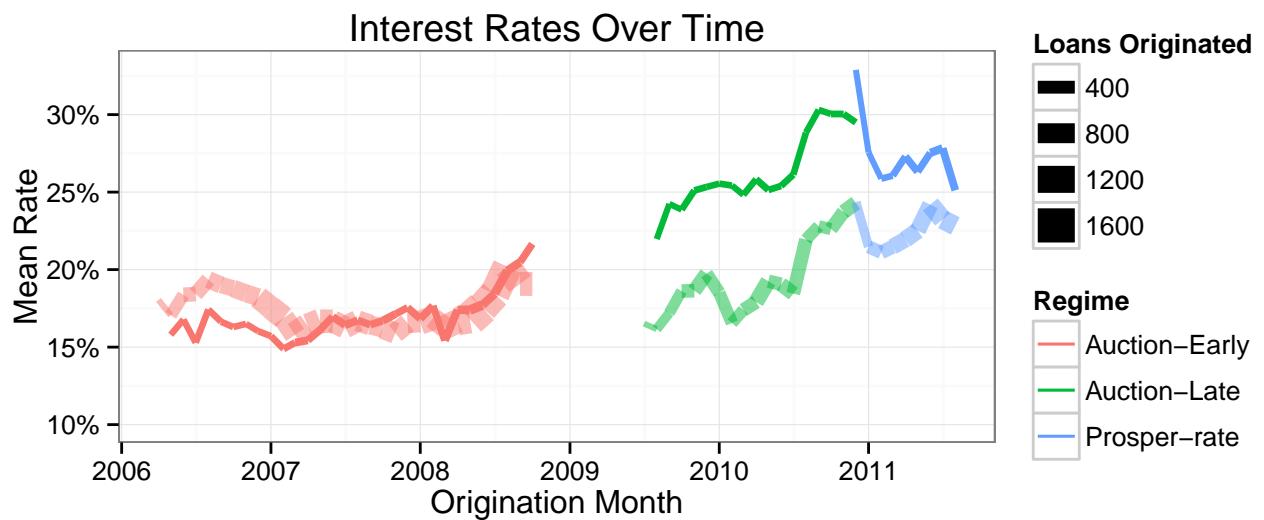


Figure 5: The blurred lines show the median interest rate of loans originated each month. The solid lines represent the median rate for loans to borrowers in the 640-660 credit score bucket for each month. The size of each line represents the number of loans originated in the appropriate month. The color of the line represents the regime. Loans funded after the shutdown are associated with higher rates after controlling for borrower credit score bucket.

shutdown but the upwards trend, from from 15% at the beginning of 2007 to 30% at the end of 2010 is clear.

Figure 6 shows that majority of the difference in interest rates between regimes can be attributed to the rates paid by borrowers in risky credit buckets, with credit scores in the 620-700 range. Investors across regimes demand similar interest rates from low-risk (high credit) borrowers. For the 760-780 bucket, early auction regime investors charged 9.0%, late auction regime investors charged 9.4%. However, for higher risk buckets, the prices are extremely different; for the 640-660 bucket, early auction regime investors charged 16.0%, while late auction investors charged 26.3%.

More generally, the slope of the curves in Figure 6 show that late auction regime investors were much more sensitive to changes in borrower credit score than early auction regime investors were. In the early auction regime, a 100 point increase in credit score was associated with a 4% reduction in interest rate.³⁴ In the late auction regime, a 100 point increase in credit score is associated with an 11% reduction. In the Prosper-rate regime, a 100 point increase in credit score is associated with a 10% reduction in interest rate.

Freedman and Jin find that early auction regime borrowers “understand ordinal difference across credit buckets.” It appears, however, that investors were almost completely insensitive to changes in credit score below the 600-620 buckets. In fact, any sensitivity they might have shown was in the wrong direction: for early regime auction borrowers in credit buckets below 600, the correlation between credit score and interest rate is .097, meaning borrowers with higher credit scores paid higher interest rates. This pattern cannot be dismissed as an artifact of small sample size, the 6,504 loans to borrowers in these buckets account for 23% of early auction regime loans. Early auction regime investors’ insensitivity to credit scores among risky borrowers can be understood in the context of Iyer et. al (2009)’s finding that investors relied more heavily on “soft” information when evaluating borrowers with low credit scores.³⁵

Lower sensitivity to credit scores, and total disregard for differences in credit score below 600 did not work out well for early auction regime investors. The relationship between credit score and default probability was even stronger in that regime than in the other regimes. Figure 7 presents the relationship between credit score bucket and default rate, and shows that the probability of default decreases (increases) with higher (lower) credit score, for all three regimes. In the early auction regime, a 100 point increase (5 buckets) in credit score is associated with a 15% reduction in default probability. In the late auction regime, a 100 point increase is associated with a 7% reduction in default probability. In the Prosper-rate regime, an 100 point increase is associated with a 12% increase in default probability. The data that underly Figure 7 can be viewed in tabular form in Table 5.

Overall, 37.1% of borrowers in the early auction regime defaulted, many more than the 16.2% default rate of the late auction period. However, much of this effect is caused by the fact that borrowers in the late auction regime were in higher credit buckets. Inside of each credit bucket, where each regime has borrowers, the difference between default rates is, on average, 10.2%. Just as with interest rates, however, the gap between the early and late auction regime default rates is wider for risky borrowers.³⁶

³⁴This analysis uses regressions that predict Interest rate, not the chart itself. The regressions can be found in the appendix.

³⁵Zhang and Liu (2012) also find more herding in this population of borrowers. (Juanjuan Zhang and Peng Liu. “Rational Herding in Microloan Markets”. In: *Management Science* 58.5 [2012], pp. 892–912)

³⁶For borrowers in the 640-660, 740-760 and 840-860 buckets default rates were 13%, 9% and 2% higher, respectively,

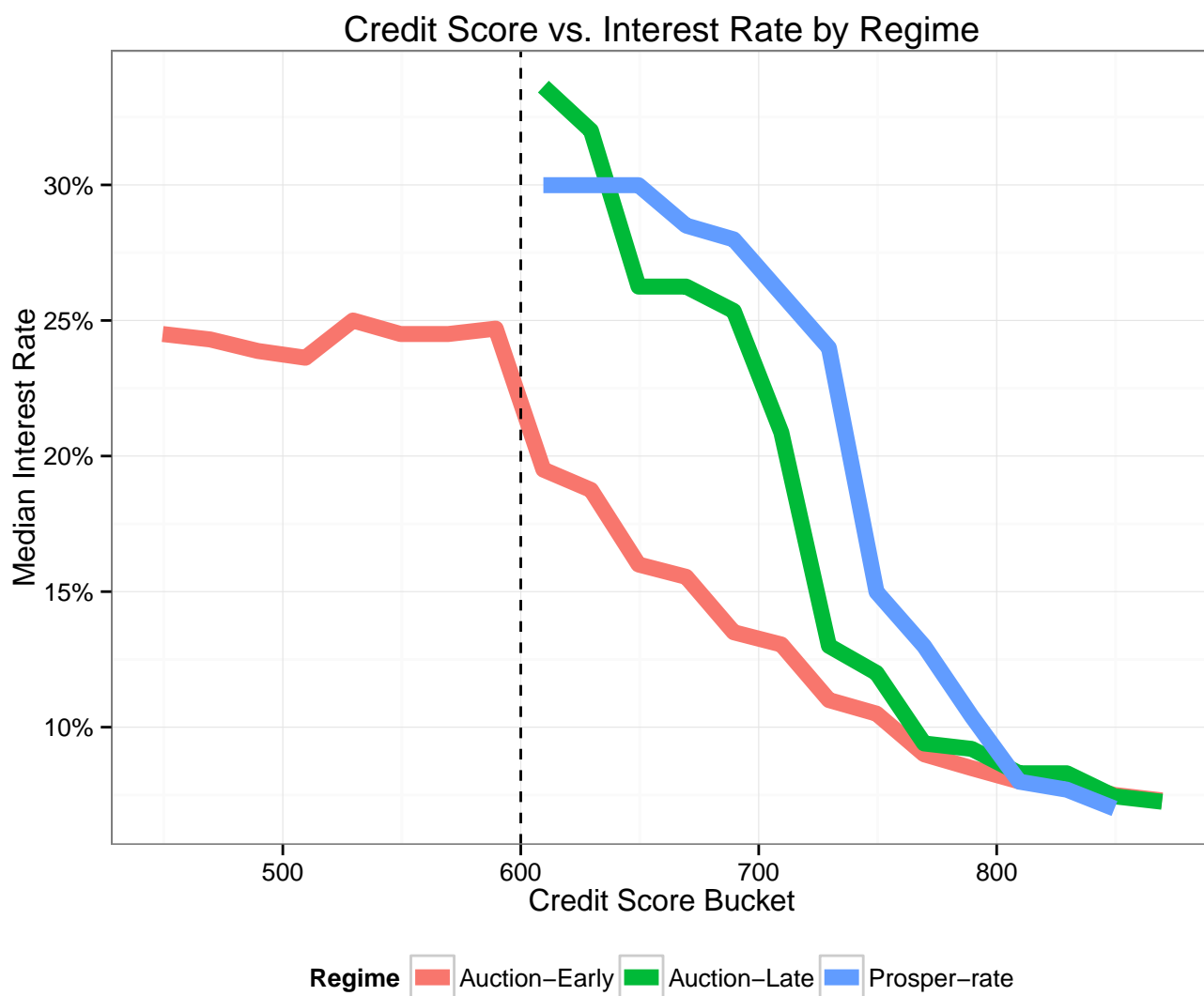


Figure 6: This chart shows the median interest rate borrowers from each credit score bucket paid during each period. Under the early auction regime, interest rates were lower, especially for risky borrowers, and lenders were less sensitive to risk.

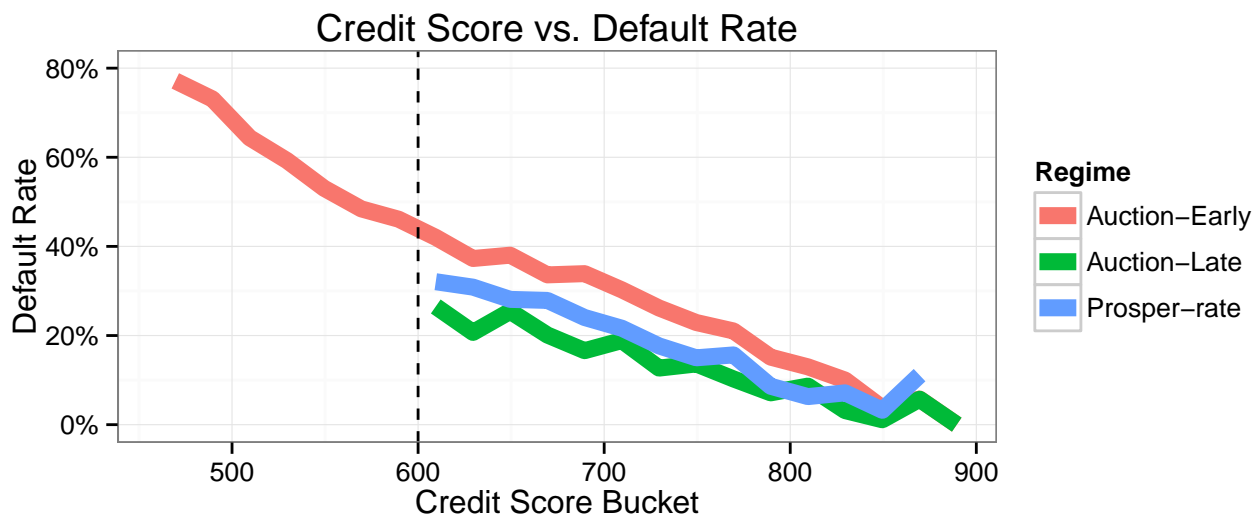


Figure 7: This chart shows the default rate of the borrowers in each credit score bucket. Controlling for credit score bucket, early regime borrowers were more likely to default.

For credit buckets above 600, the relationship between credit score and default rate appears to be fairly linear, for all three regimes, suggesting that borrowers should be sensitive to credit scores when setting interest rates, since they are predictive of default. Early auction regime investors, then, were not sufficiently sensitive to the hard credit variables that are predictive of default, especially when evaluating risky, low-credit borrowers, suggesting that the poor returns of the early auction regime were not caused exclusively by the time period, but also by poor investor decisions. We now look at the more informative grading algorithm Prosper began using after the shutdown, which made investors more sensitive to hard credit variables.

4.2.1 More Informative Grading, Higher Interest Rates

In Figure 8, each $(row, column)$ entry is the distribution of borrower credit scores that were assigned grade *column* during the *row* regime. For example the top row, left most entry presents the credit score buckets of the borrowers who were assigned grade AA in the early auction regime. In the early auction regime, all borrowers in a given credit bucket were assigned the same grade. Since all borrowers in the 640-680 buckets were assigned grade C, we can see two adjacent spikes, one for the 640-660 bucket, and another for the 660-680 bucket, in the top row of the grade C column of Figure 8. No other credit grades in the top row have any borrowers from the 640-680 buckets.

In the late auction regime, Prosper began using multiple variables to form its credit grade, and borrowers in each credit score bucket are spread across grades based on other credit variables. Correspondingly, in the bottom row of the grade C column of Figure 8, we can see that late auction regime, grade C borrowers came from 11 different credit buckets. When Prosper began aggregating hard credit variables into more informative grades, investors began relying heavily on these grades when determining interest rates.

in the early auction regime than in the late auction regime.

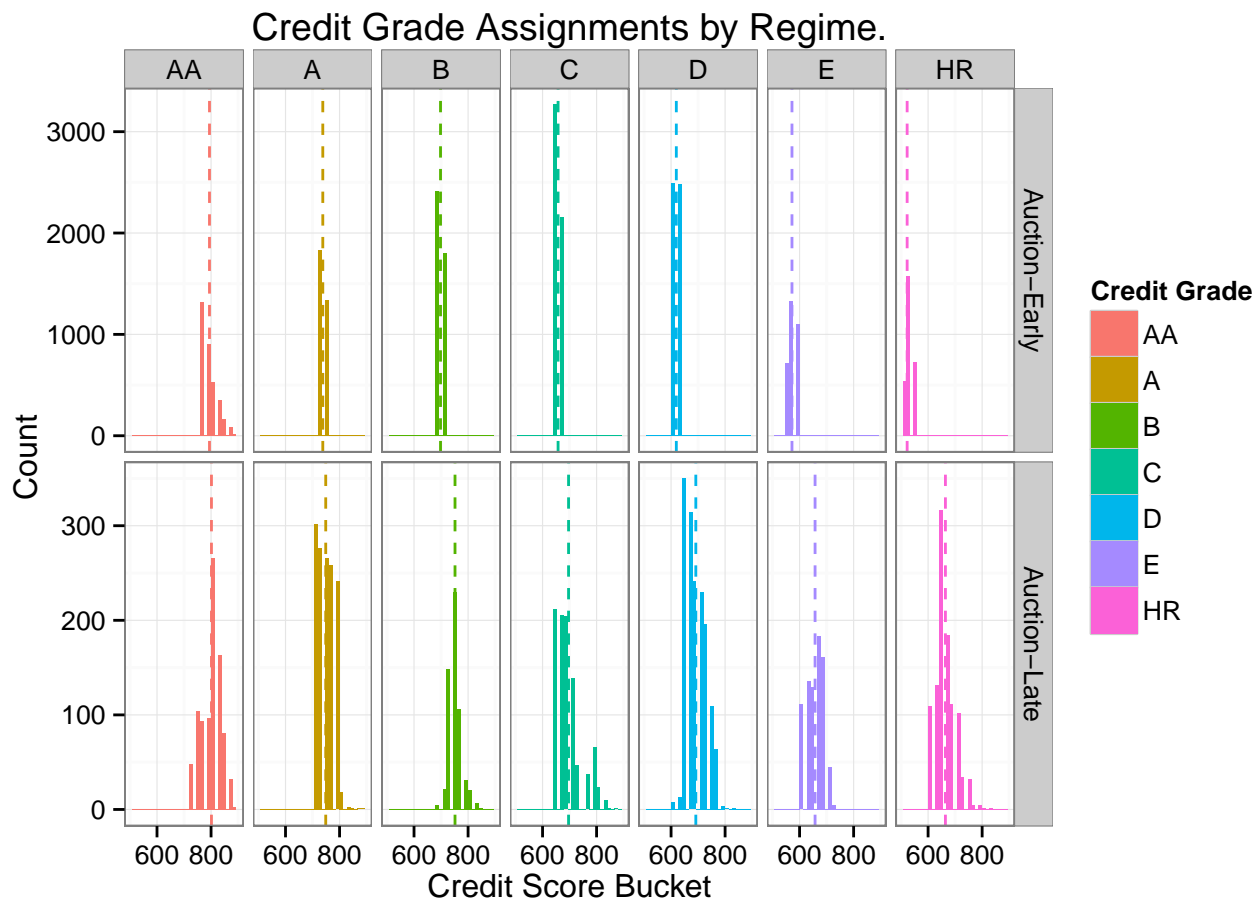


Figure 8: Another way to visualize the credit scores of borrowers in each grade. Dotted Lines indicate mean credit bucket for each grade. Each (row,column) entry is the distribution of borrower credit scores that were assigned grade (column) during the (row) regime.

	Credit Score Bucket	Auction-Early	Auction-Late	Prosper-rate
1	609.50	0.42	0.27	0.32
2	629.50	0.37	0.21	0.31
3	649.50	0.38	0.25	0.28
4	669.50	0.34	0.20	0.28
5	689.50	0.34	0.17	0.24
6	709.50	0.30	0.19	0.22
7	729.50	0.26	0.13	0.18
8	749.50	0.23	0.14	0.15
9	769.50	0.21	0.10	0.16
10	789.50	0.15	0.07	0.09
11	809.50	0.13	0.09	0.06
12	829.50	0.10	0.03	0.07
13	849.50	0.04	0.01	0.03
14	869.50	0.06	0.06	0.11

Table 5: Default rates for borrowers in each credit score for each regime. Credit score buckets are denoted by their midpoint. EG, the 600-619 bucket is in the first row, 609.5. Each column then represents the default rates borrowers in the (column name) regime, in the (row-name) credit bucket. Default rates are much higher for the early auction regime, especially among risky borrowers. This data underlies the previous figure. Early auction regime borrowers in buckets below 500 have even higher default rates.

	Regime	Fit	Number of Loans
1	Auction-Early	0.49	27764
2	Auction-Late	0.87	7238
3	Prosper-rate	0.97	4550

Table 6: This table documents how much of the variance in interest rates in each of the three regimes can be explained by variance in hard credit variables, including grades. 'Fit' is measured as the Adjusted R^2 of a regression with interest rate as the dependent variable and standard credit variables as independent variables, using every loan in the relevant regime.

The tension between the importance of the credit score in the credit grade and the importance of the credit score in the interest rate is resolved by the fact that credit grades are much more predictive of interest rates in the later two regimes. In Table 6, the *Fit* column shows the fraction of the variance in interest rate that can be described by variance in hard credit variables, as measured by the adjusted R^2 of a regression with interest rate as the dependent variable and credit score bucket, current delinquencies and credit grade as independent variables. This regression yields adjusted R^2 of 49%, 88% and 97% for the early auction, late auction and Prosper-rate regimes, respectively, suggesting that Prosper's more informative grades are immediately used by borrowers to determine interest rates.³⁷

Since we do not know Propser's exact grading formula, it may include information that was not previously available to borrowers. Table 7 addresses this concern by running the same procedure as used for Table 6, but with credit score bucket as the only independent variable. The direction of

³⁷The full regression results are available in the appendix.

	Regime	Fit	Number of Loans
1	Auction-Early	0.37	27764
2	Auction-Late	0.47	7238
3	Prosper-rate	0.39	4550

Table 7: This table documents how much of the variance in interest rates can be explained by borrowers’ credit score buckets. ‘Fit’ is measured as the Adjusted R^2 with interest rate as the dependent variable and credit score bucket as the independent variable, using every loan in the relevant regime.

the results is the same: Prosper’s change to more informative grading did not only lead to a larger role for those grades in investors’ interest rates, but increased the weight of credit score bucket, a variable that investors had access to in the same format in both periods. The difference between the early and late auction regimes’ Fit is much larger in Table 6 than in Table 7, which suggests that part of the increased predictability of interest rates in the Late Auction regime comes from hard credit variables besides the credit score bucket. We cannot determine whether these hard credit variables were included in the grade.

If the crowd of investors were as wise as Hayek hopes, Prosper’s repackaging of multiple variables into a more informative credit grade should not affect the weight of a single variable like credit score in investors’ risk estimates. Furthermore, Figure 7 makes it clear that investors would have experienced fewer defaults, and thus earned higher returns, if they had used credit buckets more heavily in their risk assessments, given the poor returns associated with low credit borrowers. We can thus attribute some of the poor returns of the early auction regime to a lower level of reliance on hard information.

In the next sub section, I estimate the impact of Prosper’s other interventions, which ban low-credit borrowers.

4.3 Impact of Subprime Borrower Ban

We now turn to Prosper’s policy changes, which restrict the types of borrowers who investors are allowed to lend to. An auction perspective predicts that the crowd of lenders should be able to effectively screen all candidates, not just low risk candidates. On multiple occasions, however, Prosper has intervened to restrict the borrower pool to high credit borrowers, and thereby improved investor returns. Figure 9 shows the mean borrower credit score for each month in the sample, with the grey clouds showing one standard deviation bounds. Overall, for the three samples in full, the average borrower credit scores increases from 653.2 before the shutdown to 714.2 after the shutdown, then dips to 705 for the Prosper-rate period. For the first 10 months of Prosper’s history, investors choose to lend to borrowers with an average estimated credit score of 609. Prosper’s policy change, on March 1, 2007, that prevented loans to “low-credit borrowers,” defined as those with credit score below 520, showed an immediate positive impact on investor decision making and returns.

Prosper’s reasoning for the change was to protect investors: by March 1, 2007, 14.9% of loans to low-credit borrowers had already defaulted, more than triple the 4.6% of loans to other borrowers that had already defaulted.³⁸ Theoretically, however, it is more difficult to understand why Prosper needed to step in. From the EMH perspective, investors would have reacted to the news of the

³⁸The 10% gap in default rates is even more striking when the average age of the two groups of loans is considered. On March 1, 2007, the loans to low-credit borrowers averaged 100 days since origination, while loans to other

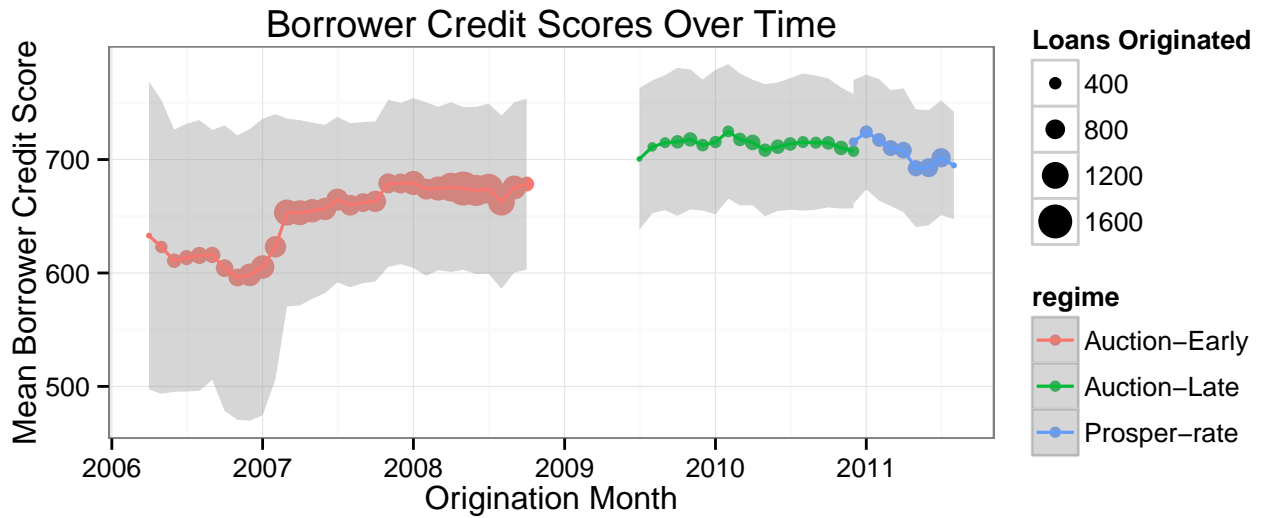


Figure 9: This figure presents average estimated credit score of loans originated each month. The size of each dot represents the number of loans originated each month. The average credit score (using bucket midpoint) increases from 653 in the early auction regime to 714 in the late auction regime, before returning to 705 in the Prosper-rate regime. Grey clouds indicate one standard deviation.

defaults on low credit loans by decreasing their exposure to low credit loans, or increasing the interest rates they bid on them. However, the median interest rate on low-credit, actually decreased from 24.23% for the month of July, 2006 to 23.4% for February, 2007, the month before the intervention. Similarly, exposure to low credit loans did not substantially decrease. On December 1, 2006, 13.3% of these bad loans had already, after an average of only 3 monthly payments, gone into default, as compared to 3.7% of other loans. Yet over the month of December 2006, loans to borrowers with credit score below 520 accounted for 22% of originations, the highest proportion in the site's history. Hayek's wisdom of the crowd theory expects market participants to adjust their beliefs to new information more rapidly and flexibly than Prosper. The data, however, show that investors continued to make avoidably bad decisions until Prosper disallowed them.

After the intervention, returns improved substantially. For the loans originated in February 2007, the month before the change, the mean borrower credit score was 620, with standard deviation 117, while the mean ROI on a loan was -13.1%. For loans originated in March 2007, the month after the change, the mean borrower credit score was 650, with standard deviation of 83, and mean ROI -7.4%. The 30 point month to month increase in mean borrower credit score is the largest in the sample, and the associated ROI change is not caused by time differences: the default rate to borrowers in the 640-660 bucket, a representative example, was 42.7% in February, and actually rose to 44.8% in March. Had Prosper enforced this rule from the site's inception, the default rate on loans originated before the rule change would have been reduced by 6%, and mean ROI would have increased from -9.1% to -4%, a savings of roughly \$2 million for investors.

A similar line of analysis can be applied to examine the impact of Prosper's second intervention.

borrowers averaged 122 days since origination. 14.9% of low credit borrowers defaulted over an average time since origination of 100 days, whereas only 4.6% of other borrowers defaulted over an average time since origination of 122 days.

Upon re-opening after the shutdown, borrowers were required to have a credit score ≥ 640 and returning borrowers were required to have a credit score ≥ 600 . This policy change is associated with a similar shift in average borrower credit score (measured as bucket mid-point) from 678 in October, 2008 to 701 for July, 2009. The contemporaneous decrease in default rates makes it more difficult to explain the ROI jump exclusively through changes to the borrower population. But like in the previous intervention, low-credit loans continued until they were no longer allowed. By September 2008, the last full month before the shutdown, 19.8% of loans to borrowers in the 520-640 buckets had already defaulted, compared with 8.1% of loans to more creditworthy borrowers, yet 32% of originated loans for that month were to borrowers in the 520-640 range. August, 2009, the first full month after the rule change, only 7.5% of loans were to borrowers in that range, all of whom had borrowed with Prosper before. To reiterate, even with access to public information that showed that their lending strategies were not profitable, investors did not adjust. In the next section, we show that this lack of adjustment allows for the possibility of using public information to earn excess returns during the early auction period.

5 Excess Returns

5.1 Puzzle: Bad Returns in Early Auction regime

Although Figure 6 shows that investors charge higher rates to riskier borrowers, it also shows that early auction regime investors were insensitive to differences between low-credit borrowers, and, more generally, less sensitive to credit score than borrowers in subsequent regimes. The high risk, low-reward data pattern of the early auction regime, shown in Figures 4 and 7 show that this sensitivity to credit score was far too low, and does not come close to making up for the difference in default likelihood. These claims, however, are both vulnerable to the critique of an unlucky time period. In this section, I present this essay's key test of market efficiency: can a simple loan selection strategy earn excess returns using publicly available information?

The answer is simple, but powerful: when investors set interest rates by auction, in the early auction and late auction regimes, it is possible to earn excess returns. Once Prosper starts setting rates, my methodology is no longer able to earn excess returns.

The rest of this section discusses the methods used to create such a strategy, and the results of the strategy. In the methodology sub-section, I define publicly available information as 'the information that would have been available at origination date, the day a given loan was originated,' and discuss the the principal challenge with complying with this definition:

On origination day, Investors only know which loans have *already* defaulted. Many loans will end up defaulting later, but to investors, they appear to be performing well.

This insight creates four major challenges. First, we must create a variable, for each loan, that indicates whether the loan had already defaulted on Origination date, and use that as the dependent variable in a regression, rather than using the natural defaulted or ROI variables used in section 4. Second, we must fix a *cut date* at which we use all publicly available information to train a linear model, and then use this model to select from the pool of loans that have not been originated.³⁹ Thirdly and most subtly, an investor would not have known which variables would be

³⁹This approach also carries the standard cross-validation benefit, by penalizing overfitting.

useful predictors of future loan performance, only which variables had been past predictors of loan performance. For this reason, I do not choose the independent variables used by the linear model, an automated feature selection procedure makes this decision. Finally, during some month after the *cut date* investors did not know whether the listings in the future would be more profitable than the listings in the present, so the strategy must choose 10% of the loans offered in a given month in each credit grade, rather than considering the full pool of loans and picking the best 10%.⁴⁰

The methodology sub-section which follows further explores the nature of these challenges, the implementation of the solutions, and new challenges created by the solutions. The section then continues with In Sample results, which show that the independent variables chosen by the feature selection protocol are very similar to the variables Prosper subsequently used to make its more informative grading, and that mispricings were most predictable in the early auction regime. The section concludes with Out-of-Sample results, which show that such a strategy can earn excess out-of-sample returns in the early auction and late auction regimes, when investors set interest rates, but fails to earn excess returns in the Prosper-rate regime, when Prosper sets interest rates directly.

5.2 Methodology

5.2.1 Cross Validation Procedure

	Sample	Loans	Start	End	Days
1	Auction-Early.train	13841	2006-04-21	2007-10-01	528
2	Auction-Early.test	13923	2007-10-02	2008-10-16	380
3	Auction-Late.train	3579	2009-07-20	2010-04-26	280
4	Auction-Late.test	3659	2010-04-27	2010-12-17	234
5	Prosper-rate.train	2265	2010-12-20	2011-04-28	129
6	Prosper-rate.test	2285	2011-04-29	2011-08-03	96

Table 8: This table documents number of loans and dates of the 6 samples used for training and testing regression strategies.

Regime changes split the data into early auction, late auction and Prosper-rate regimes. To test our predictions out of sample, we split each of these thirds in half. For those familiar with cross validation, regressions are trained on the first half of loans originated under each regime, and tested on the second half. Table 8 describes the size and periods of the six sub samples. For those less familiar with cross validation, I now describe the six sub samples in detail.

1. 13841 loans originated between April 21, 2006 and October 2nd, 2007, are used to generate coefficients for the early auction regime regressions.
2. 13923 loans originated after October 2nd, 2007, but before October 16, 2008, are used to test the predictions of the early auction regime regressions.⁴¹

⁴⁰The credit grade constraint is to ensure that the models do not use our hindsight knowledge that certain credit grades performed poorly to improve performance.

⁴¹Between October 10, 2008 and July 20, 2009, no loans are originated, because Prosper is in the SEC-imposed Quiet Period

3. 3579 loans originated between July 20, 2009 and April 26, 2010, are used to generate coefficients for the late auction regime regressions.
4. 3659 loans originated after April 26, 2010, but before December 17, 2010, are used to test the predictions of the late auction regime regressions.
5. 2265 loans originated between Dec 20,2010 and April 28,2011 are used to generate coefficients for the Prosper-rate regime regressions.
6. 2285 loans originated between April 29,2010 and August 3rd 2011 are used to generate coefficients for the Prosper-rate regime regressions.

5.2.2 Avoiding Look-ahead Bias with Known-Default

Unfortunately, since our time series is so short, the challenge of training models on the limited information a lender would have is considerable; the early Auction regime starts in April 2006, and ends in October 2008, and all loans last 36 months, so in October 2007, an investor only knows 18 months of payment results of the earliest origination. To address this challenge, we develop a binary Known-Default variable. For a given loan, Known-Default equals 1 if the ex-post sum of the payments of a loan is less than the percentage of 36 months that had elapsed by the end of the training period.

As an illustrative example, suppose Bill borrows \$100 in April 2006, and that the training period ends December 2007. If Bill ends up paying back \$60 to investors over the full 36 months, his Known-default = 0 (even though he ended up defaulting) because the training period ends after 55.6% of his three years have elapsed, but he paid back 60% of his loan. If, however, Bill only ends up paying back \$50 to investors, Known Default=1, because Bill MUST have already missed at least one payment by the end of the training period.

Using Known-Default⁴² as our left hand side variable instead of observed defaults or ROI allows us to restrict the information available to models to the same set of information that investors had at their disposal at the end of the training period, namely borrower characteristics of each loan that has already been originated, and whether it has *already* gone into default. One downside of this approach is that for each period, many loans were originated in the month before the end of the training period, and almost none of these investors miss their first monthly payment. To avoid these young loans, about whose performance almost nothing is known, from exerting a large impact on the model's coefficients, I weight observations by *observation maturity*, the amount of time that has elapsed between a loan's origination and the end of the training period, in a weighted linear regression, using the training sample data.

5.2.3 Feature Selection

The following procedure is executed on the training data for each of the three regimes to determine the 8 features⁴³ to use in the regression. An automated⁴⁴ feature selection prevents me from being

⁴²It would be difficult to create known-ROI because there is no data on the magnitude of each monthly payment

⁴³I use the terms 'feature' and 'Independent Variable' interchangeably.

⁴⁴Thomas Lumley using Fortran code by Alan Miller. *leaps: regression subset selection*. R package version 2.9. 2009. URL: <http://CRAN.R-project.org/package=leaps>.

able to add variables to the regression that I know to be powerful, a subtler form of look-ahead bias.

1. Start with 45 features, chosen by me because they don't have missing data.⁴⁵
2. Scale the features to have 0 mean and unit variance, by subtracting the mean from each observation and then dividing by the standard deviation.
3. For each possible 8-feature subset of the original features, run a linear regression predicting Known-Default, and recover the Bayesian Information Criterion.⁴⁶
4. Choose the subset of features with the highest BIC. These 8 features are then used in the training regressions.

5.2.4 Out of Sample Loan Selection

After training linear models to predict Known-Default, we use the coefficients to predict known-default as a proxy for real default, for each possible loan in the testing sample, and choose the 10% of loans from each credit grade with the lowest predicted default likelihood.⁴⁷

5.2.5 Summary of Procedure

For each regime ...

1. Divide the loans into training and testing samples of equal size, with all loans in the training sample originated before the first loan in the test sample was originated.
2. Avoiding Look-ahead Bias: create variable Known-Default = (0||1) for each loan, depending on whether a loan had **already** defaulted at the end of the training sample.
3. Feature Selection: start with 45 Features, select the 8 feature subset that best predicts Known-default.⁴⁸
4. Cross-Validation Part 1: Run Linear Regression to predict Known-Default using training period data and selected variables.
5. Cross-Validation Part 2: Using trained coefficients, generate predicted Known-Default values for loans in the test period.
6. Cross-Validation Part 3: Select the 10% of Loans from each credit grade with the lowest predicted Known-default probability.
7. Measurement of Returns: Compare ROI of selected loans against the benchmark of a strategy that chooses every loan.

⁴⁵A complete list of features can be found on github.com/sshleifer/ProsperThesis

⁴⁶For more on BIC, http://en.wikipedia.org/wiki/Bayesian_information_criterion

⁴⁷Loans are selected from each credit grade to ensure that strategies in the later periods can actually earn positive returns. Grade A loans in later periods return less than average.

⁴⁸Fortran code by Alan Miller, *leaps: regression subset selection*.

5.2.6 Assumptions

Two assumptions underly the out-of sample loan selection technique which selects from the pool of loans that were funded by Prosper’s investors, at the rate they received. First, in the auction regime, by successfully bidding on a loan, investors may decrease the interest rate. This analysis assumes that a strategy does not move the interest rates of the loans it selects. Second, a strategy never chooses to fund loans that will end up unfunded, since we cannot observe their ROIs.

5.3 In Sample Results

The Linear Regression results shown in Table 9 suggest that returns are much more predictable before the shutdown than after: the Adjusted R^2 for each regime’s regression fall from 13.3171% for the early auction regime to 1.15% for the late auction regime and 3.5% for the prosper rate regime. The coefficients for all regressions are so close to zero because we are predicting Known-Default, which occurred in 11% of Auction Regime, training sample observations, but only 1.5% in the late auction regime training data and 1.0% of observations in the Prosper-rate regime data. The discrepancy in Known-default rates is caused by both a shorter period of training data and a lower actual default rate for the later regimes.

More interestingly, we can see that for the early auction regime, the automated feature selection procedure chosen many of the same variables, like Inquiries, Credit Score and Delinquencies, that were used by Prosper to determine its more informative grades, which had the effect of guiding lenders towards setting higher interest rates.

The finding that the features selected to predict superior returns through my purely mechanical procedure end up being very similar to the features subsequently used by Prosper to set credit grades supports the view that investors failed to attend to these variables, and that Prosper as a market maker fixed their errors. Prosper’s subsequent reliance on the information I use to beat the market, supports the argument that the information was, in fact valuable, while the facts of Section 4, namely Figure 6, aim to show that the information was largely ignored. Even more importantly, this analysis is NOT vulnerable to the critique that my tests of market efficiency are undermined by an unlucky time period. In particular, Prosper used these same variables to guide investors, and improve market efficiency, at a very different time.

	Auction-Early	Auction-Late	Prosper-rate
(Intercept)	0.0869*** (0.0042)	0.0162*** (0.0046)	0.0059 (0.0040)
BorrowerRate.sc	0.0511*** (0.0049)		
TotalCreditLinespast7years.sc	-0.0187*** (0.0037)	-0.0070* (0.0029)	
OpenRevolvingAccounts.sc	-0.0189*** (0.0037)	-0.0014 (0.0031)	-0.0028 (0.0023)
InquiriesLast6Months	0.0063*** (0.0007)		0.0161*** (0.0019)
CurrentDelinquencies.sc	0.0455*** (0.0025)		
DelinquenciesLast7Years.sc	-0.0160*** (0.0027)		-0.0045 (0.0032)
MonthlyLoanPayment.sc	0.0185*** (0.0030)		
CreditScoreEst.sc	-0.0347*** (0.0036)	-0.0068 (0.0038)	-0.0045 (0.0037)
TotalInquiries		0.0022*** (0.0005)	-0.0020* (0.0008)
PublicRecordsLast10Years.sc		-0.0029 (0.0039)	-0.0033 (0.0028)
DebtToIncomeRatio.sc		0.0324** (0.0119)	
TotalProsperPaymentsBilled.sc		-0.0068** (0.0024)	
ProsperPrincipalBorrowed.sc		0.0031 (0.0021)	
TotalProsperLoans.sc			-0.0027 (0.0016)
StatedMonthlyIncome.sc			-0.0011 (0.0016)
R ²	0.1337	0.0140	0.0384
Adj. R ²	0.1332	0.0115	0.0350
Num. obs.	13458	3189	2265

Y = Known-Default

Table 9: Coefficients extracted from Least Squares Regression to predict Known-Default in early auction, late auction, and Prosper-rate regime training data.

5.4 Out of Sample Results

	Regime	BM.mean	BM.sd	BM.def	LM.mean	LM.sd	LM.def	Excess
1	Auction-Early	-0.001	0.031	0.323	0.029	0.032	0.211	0.030
2	Auction-Late	0.159	0.024	0.130	0.216	0.033	0.078	0.058
3	Prosper-rate	0.133	0.024	0.181	0.123	0.037	0.206	-0.009

Table 10: This table compares the out of sample performance of the linear model (LM) strategies to the benchmark (BM). As shown in the next figure. The column names are in the format (Strategy).(statistic), where (LM) for linear model and (BM) for benchmark are the two strategies, and mean return of selected loan (mean), standard deviation of monthly returns (sd) and mean default rate of selected loans (def) are the statistics. For example, the LM.mean column shows the average ROI of a loan selected (out of sample, of course) by the linear model in each of the three regimes. The Excess column is calculated as LM.mean - BM.mean, and thus represents the difference in returns between the loans selected by the linear model and the benchmark over each regime. Notice that the Excess returns are positive for the first two regimes, where investors were involved in setting interest rates, and roughly 0 in the most recent regime, where Prosper set rates directly.

Table 10 shows the mean and standard deviation of the returns of the loans selected by the regression strategy. For the early auction regime, when investors focused less on hard variables, set low rates, and after observing high default rates among similar borrowers, the model beats the benchmark by 3%, with similar variance⁴⁹. Moreover, a T test comparing the 1434 loans selected by the early auction regime model and the 13,923 loans it could have selected, rejects the null hypothesis that the two groups have the same mean ROI at the .1% significance level. This finding is inconsistent with the EMH prediction that it should be impossible to earn excess returns using publicly available information. As discussed in the methodology section, the model is trained on the observed defaults in the training sample, and chooses the loans its predicts to have the lowest probability of default. As such it ends up minimizing the riskiness of loans it selects, rather than trying to optimize a risk reward calculation. Early auction period loans selected by the model have a default rate of 21% , compared to the 32% default rate for all loans. Figure 10, which compares the various strategies out-of-sample performance over time, reveals that most of the model's outperformance is during the initial 6 months of testing.

For the late auction regime, when investors largely used the ratings prosper gave them to form interest rates, the model beats the benchmark by 5.8%, with a slightly higher standard deviation. A T test comparing the 360 loans selected by the late auction regime model and the sample of 3569 loans it chose from rejects the hypothesis that the two groups have the same mean ROI at the 1% significance level. It is again clear that the superior returns of the loans chosen by the model were generated by a lower default rate: 7.8% compared to 13.0% for the full basket of loans. This result should interpreted carefully, however, given the noticeably higher variance and small number of selected loans.

For the Prosper-rate regime, when Prosper set interest rates, the model is unable to beat the benchmark: it chooses loans with worse ROI, higher default rate and identical returns.

In summary, over the initial two regimes where investors were involved in the process of setting interest rates, via auction, it was possible to use publicly available information to earn excess

⁴⁹Variance is calculated as the standard deviation of monthly returns



Figure 10: This table presents the out of sample returns for the loans chosen by the linear model (blue) and all loans (red) for each month in the three testing periods. The linear model, uses only publicly available information to select loans, and is described in detail in the methodology section. Horizontal lines indicate each strategy's mean return.

returns.⁵⁰ In the final regime, where the market maker sets the interest rate, it is much more difficult, if not impossible, to earn excess returns using publicly available information. This likely happens because the variables that investors neglected in the first two regimes, which we use to earn superior returns, and the variables that Prosper eventually uses to set interest rates are very similar.

5.5 Methodological Expansions

An expansion of this project could retrain the model every month with the benefit of newly observed defaults and newly originated loans, and expect performance improvements.

Engineering a known-ROI dependent variable that equals the percentage of expected repayment of principal an investor received from each borrower up to the relevant date, would allow the model to make risk-reward calculations when deciding between loans, as opposed to just choosing the least risky loans.

6 Controlling for Time Period

The first two inconsistencies with EMH predictions: that early auction regime loans have negative average ROI, and that riskier early auction regime loans earn investors negative ROI, are both vulnerable to time period. Specifically, maybe the early period just happened to be bad for loan repayments. In this section, I aim to show that the decisions made by early investors, namely to lend to low credit borrowers and to demand far lower interest rates than hard variables would suggest, partly explain the difference in ROI. Thus, the question guiding this section is: **what would have happened to investor returns in the early auction regime if investors only lent to borrowers that would have qualified for loans in the late auction regime, and demanded the rates charged by investors in the late auction period?** This will help to understand whether the change in returns between the early and late auction regimes was merely a consequence of an unpredictably higher default rate.

6.1 Methodology

In order to estimate the hypothetical returns in an early auction regime where investors charged late auction regime rates to borrowers who would have met the late auction regime criteria, we must (1) reduce early auction borrowers to the subset that would have been acceptable in the late auction regime, (2) estimate what interest rate late auction regime investors would have charged them, and (3) then use this estimated rate to predict a new ROI.

The first step, restricting the borrower pool, is simple. In the late auction regime, Prosper required that new borrowers have a credit score greater than 640, and that returning borrowers have a credit score greater than 600 in order to apply for a loan. Applying this rule reduces the number of early auction regime loans whose rate we will adjust by 40.4%, from 27,764 to 16,545.⁵¹

The second step of the method involves finding the hypothetical interest rate an early auction regime borrower would have paid in the late auction regime. For simplicity, this *adjusted rate* is

⁵⁰Note that excess returns are as compared to other Prosper loans, not treasuries or another asset class.

⁵¹This may not go far enough, as late auction regime investors refused to fund many borrowers applications.

computed by finding the median interest rate paid by late auction regime borrowers in the same credit score bucket, *EG* 700-720.

To simplify the third step, predicting a hypothetical ROI given actual loan performance and a new, adjusted interest rate, I assume that borrowers would have made the same number of monthly payments, but that each payment would be slightly bigger, given the new interest rate. This adjusted roi is calculated by multiplying the payments the borrower made under the old regime by the ratio of $\frac{(1+adjustedrate)}{(1+oldrate)}$.

As an instructional example, suppose Bill is an early auction regime borrower in the 700-719 credit score bucket. Since Bill's credit score is above the 640 cutoff, he remains in the sample.

Under the early-auction regime, Bill borrowed \$100 from investors at a 13% interest rate, and returned \$113, over the one year course of his loan.⁵²

In the late auction regime, the median borrower 700-720 borrower paid 23% interest, so, Bill's Adjusted-Rate is 23%.

To calculate Bill's Adjusted ROI, we multiply the monthly payments he actually made, by the ratio of the interest rates. And then divide the result by the principal and subtract 1. This reduces to:

$$ADJ.ROI = \frac{Payments}{Amount.Borrowed} * \frac{ADJ.Rate}{Actual.Rate} \quad (2)$$

$$ADJ.ROI = \frac{113}{100} * \frac{1.23}{1.13} - 1 = 23\% \quad (3)$$

Suppose, alternatively, Bill had repaid only \$45. So Bill's actual ROI was -55%. In this second case, Bill's Adjusted ROI = $\frac{45}{100} * \frac{1.23}{1.13} - 1 = -51\%$.

Notice that even when Bill defaulted, his Adjusted ROI is higher than his actual ROI. This is because the methodology assumes that Bill makes the same number of monthly payments after his interest rate is adjusted upwards, but that each monthly payment is bigger.

6.1.1 Assumptions

Three assumptions facilitate this methodology, but also cause bias. Most importantly, the methodology assumes that borrowers would have made the same number of monthly payments if their interest rate were higher. This assumption is inconsistent with the idea of cost-induced default, as discussed by Stiglitz and Weiss (1976): forcing the borrower to pay more per month might push him over the default tipping point.⁵³ Empirical support for the theory of cost-induced default can be found in Wei and Lin (2013)'s examination of Prosper's transition from the late auction regime to the Prosper-rate regime, which shows that when borrowers were assigned higher interest rates, they defaulted more frequently.⁵⁴ With this in mind, my adjusted ROI serves as an **upper bound** because it assumes that each borrower makes the same number of payments, even though each payment is higher. More realistically, we might expect Bill to make fewer payments on his loan if he found it more difficult to afford his new interest rate.

⁵²Neither maturity nor principal matter, for our purposes

⁵³Joseph E Stiglitz and Andrew Weiss. "Credit rationing in markets with imperfect information". In: *The American economic review* (1981), pp. 393–410.

⁵⁴Zaiyan Wei and Mingfeng Lin. "Auction vs. Posted-Price: Market Mechanism, Lender Behaviors, and Transaction Outcomes in Online Crowd-Funding". In: (Sept. 2013). URL: <http://ssrn.com/abstract=2328468>.

Two other assumptions work in the other direction, however. First, we have assumed that all applications that satisfied Prosper’s minimum borrower standards would have been funded by investors. In reality, many applications met these criteria but were not funded, and it is likely that later regime lenders were more selective than early regime borrowers. Secondly, we have assumed that in our hypothetical world early auction regime borrowers would pay the same rate as late auction regime borrowers in their credit bucket. In reality, however, they would probably pay higher rates, because controlling for credit score bucket, they have worse average delinquencies, credit card utilization and other hard variables that would be used for grading.

To summarize, the total.ROI reported using the adjusted borrower population is probably a little lower than it should be, and the total.ROI reported using adjusted interest rates is probably higher than it should be.

6.2 Results & Discussion

Figure 11 documents the relationship between credit score bucket and returns for early auction regime, late auction regime, and late auction regime borrowers with adjusted rates, where the rate adjustment and subsequent ROI adjustment was computed using the *upper bound* methodology described above. At low credit scores, where the difference in interest rates between the two regimes is largest, the gap in ROI is also the largest. With adjusted rates, riskier assets have higher average return than less risky assets, and average returns are positive, suggesting that part of the reason the actual data is inconsistent with the EMH predictions may be the low interest rates and exposure to subprime borrowers.

	regime	total.ROI
1	Auction-Early	-0.0383
2	Auction-Early.Adj-borrowers	-0.0136
3	Auction-Early.Adj-rate	0.0439
4	Auction-Late	0.1520

Table 11: The aggregated results of the Rate Adjustment Experiment.

At each credit score, an upper bound estimate of the fraction of the ROI gap that would be accounted for by adjusted interest rates can be measured as:

$$rate.factor = \frac{mean(adj.roi) - mean(early.roi)}{mean(late.roi) - mean(early.roi)} \quad (4)$$

This *rate factor*⁵⁵ is, by construction, heavily correlated with the difference in interest rates for each regime. At the lowest included estimated credit score, 649.5, the difference in average interest rate is 9.3%, and that difference explains 42.1% of the 19.2% difference in average ROI. At the higher buckets, the difference in interest rates for borrowers in the 800-820 credit score bucket was 0.3%, and this difference explained only 3% (not half) of the 6% difference in average ROI.

The aggregated returns of the regimes, as shown in Table 11, indicate that if early auction regime borrowers had self enforced the minimum standards Prosper introduced in the second period,

⁵⁵Visually, think of rate factor as if we are walking up the plot from the red line to green line, what percentage of the way would we make it before we encountered the blue line.



Figure 11: This chart displays the results of our rate-adjustment experiment. The blue line is the result of our rate-adjustment experiment, and displays the hypothetical ROI if late auction regime borrowers had paid early auction interest rates. The grey cloud represents a 2 percent adjustment for cost-induced default.

investor returns would have improved by at least 2.8%. If investors had charged these borrowers the rates similar borrowers were charged in the late auction regime, returns might have improved by another 5.75%.

These results suggest that although much of the difference in performance between regimes is due to an especially bad time period, roughly $\frac{1}{3}$ of the difference in returns between the early auction and late auction regimes can be explained by poor investor decisions.

6.3 Comparison to Traditional Loans

	Origination	CapitalOne.default	Regime	Prosper.default	N.Prosper
1	2006.00	0.13	Auction-Early	0.28	5117
2	2007.00	0.17	Auction-Early	0.33	11205
3	2008.00	0.16	Auction-Early	0.30	11442
4	2009.00	0.12	Auction-Late	0.16	1952
5	2010.00	0.08	Auction-Late	0.17	5286
6	2011.00	0.07	Prosper-rate	0.22	4330

Table 12: This table documents the default rate of Prosper Loans (Prosper.Default) with the default rate in the first three years of consumer credit loans securitized by Capital One (CapitalOne.Default). Additionally, the regime and number of Prosper loans originated in each period are reported.

	Credit.Score	Percentage.of.Sample
1	601-660	29.69%
2	661-720	30.48%
3	>720	39.16%

Table 13: The Portion of borrowers from each credit score group in the Capital One data.

By comparing the performance of Prosper loans originated in each period to the performance of traditional consumer credit loans originated in each period, controlling for credit score as best we can⁵⁶ we can gain a better understanding of the time period. The data presented in Table 12 show that for all credit scores, default rates are higher for Prosper loans than Capital One loans, but that the difference is much larger in the early auction regime. There may again be flawed assumptions, including adverse selection of borrowers on the internet, other credit variables besides score, and the different maturities of the loans, but the data again suggest that early auction regime investors did a particularly bad job.

⁵⁶Method: Given the default rates of some consumer loans, and a table of the frequency of a bunch of credit scores, I created the same credit score groups in the Prosper data, took each group's mean default rate, and took the weighted average default rate for each year, with the group's representation in the Capital One borrower pool, shown in Table 13 as the weights. Since the Capital One loans are of different maturities, I measured the portion of defaults within three years of origination.

7 Conclusion

This paper examined interest rate setting and lender returns on Prosper.com. It took advantage of three policy regimes over the course of 2006-2011. In the first regime, investors set interest rates through a decentralized reverse auction; in the second regime they continued to use the auction, but with considerably more guidance from Prosper; and in the third regime Prosper set the rates directly. I find that compared with the later two periods, interest rates during the early auction regime were much lower and less sensitive to risk, and that investor returns were much lower than in the later two regimes.

I further examined these findings and argued that the evidence from the early auction regime is inconsistent with the predictions of the EMH. In particular, I reported three findings: average returns in the early auction period are negative, riskier loans earn lower average returns than more risky ones, and –most importantly – one can use loan selection strategies without look-ahead bias to select well-performing loans in the early auction period. The third finding is inconsistent with the most fundamental prediction of the EMH, that an investor cannot earn superior risk-adjusted returns based on publicly available information.

The first two findings are vulnerable to the critique that my results depend on the particularly bad time period (2006-2008) during the early auction years. I show that, even correcting for the time period, returns would have been poor because interest rates were set too low and insufficiently sensitive to risk. My third finding is more difficult to criticize based on the time period, since it shows that investors could have done better in the same period, the second half of the early auction regime, based on publicly available information.

I offer two critical pieces of corroborating evidence. First, when Prosper began setting interest rates, it appears to have used publicly available variables to set interest rates that are very similar to those selected by my Feature Selection algorithm totally independently, and then used to earn superior returns by taking advantage of the fact that investors neglected them. Second, I present some evidence that, in the early auction period, Prosper loans performed considerably worse than Capital One loans with similar risk characteristics. In short, all the evidence suggests that the decentralized unguided mechanism of the early auction period did not lead to efficient interest rates for Prosper loans.

The last finding has implications for financial economics and market design. It suggests that contrary to Hayek’s vision that the wisdom of the crowds leads to efficient price setting, the crowds did not utilize all the public information available to them, and made significant mistakes. Prosper learned this quite quickly, and intervened. The market maker’s guidance, and eventually price control, likely improved returns. The evidence cautions against the assumption of universal investor rationality, and thus suggests that a market design that protects investors from their errors and limited attention, either through educational guidance or even price setting, may improve investor returns, and also benefit the market maker.

References

- [1] Friedrich A. Hayek. *The Use of Knowledge in Society*. Library of Economics and Liberty, 1945. URL: <http://www.econlib.org/library/Essays/hykKnw1.html>.
- [2] William F Sharpe. “Capital asset prices: A theory of market equilibrium under conditions of risk”. In: *The journal of finance* 19.3 (1964), pp. 425–442.
- [3] Eugene F. Fama. “Efficient Capital Markets: A Review of Theory and Empirical Work”. In: *The Journal of Finance* 25.2 (1970), pp. 383–417. ISSN: 1540-6261. DOI: 10.1111/j.1540-6261.1970.tb00518.x. URL: <http://dx.doi.org/10.1111/j.1540-6261.1970.tb00518.x>.
- [4] Sanford J. Grossman and Joseph E. Stiglitz. “On the Impossibility of Informationally Efficient Markets”. In: Reprint Series 121 (Dec. 1980). URL: <http://www.nber.org/papers/r0121>.
- [5] Joseph E Stiglitz and Andrew Weiss. “Credit rationing in markets with imperfect information”. In: *The American economic review* (1981), pp. 393–410.
- [6] Andrei Shleifer. *Inefficient Markets*. 2000.
- [7] Seth Freedman and Ginger Zhe Jin. “Do social networks solve information problems for peer-to-peer lending? evidence from prosper.com”. In: (2008).
- [8] Securities and Exchange Commission. *Order Instituting Cease and Desist Proceedings in the matter of Prosper Marketplace*. 2008. URL: <http://www.sec.gov/litigation/admin/2008/33-8984.pdf> (visited on 01/01/2015).
- [9] Thomas Lumley using Fortran code by Alan Miller. *leaps: regression subset selection*. R package version 2.9. 2009. URL: <http://CRAN.R-project.org/package=leaps>.
- [10] Rajkamal Iyer et al. “Screening in New Credit Markets: Can Individual Lenders Infer Borrower Creditworthiness in Peer-to-Peer Lending?” In: rwp09-031 (Aug. 2009). URL: <http://ideas.repec.org/p/ecl/harjfk/rwp09-031.html>.
- [11] Juanjuan Zhang and Peng Liu. “Rational Herding in Microloan Markets”. In: *Management Science* 58.5 (2012), pp. 892–912.
- [12] Qiang Gao and Mingfeng Lin. “Linguistic Features and Peer-to-Peer Loan Quality: A Machine Learning Approach”. In: *Available at SSRN 2446114* (2013).
- [13] Zaiyan Wei and Mingfeng Lin. “Auction vs. Posted-Price: Market Mechanism, Lender Behaviors, and Transaction Outcomes in Online Crowd-Funding”. In: (Sept. 2013). URL: <http://ssrn.com/abstract=2328468>.
- [14] Samuel G Hanson et al. “Banks as Patient Fixed-Income Investors”. In: (2014).
- [15] Adair Morse. “Peer-to-Peer Crowdfunding: Information and the Potential for Disruption in Consumer Lending”. In: Working Paper Series 20899 (Jan. 2015). DOI: 10.3386/w20899. URL: <http://www.nber.org/papers/w20899>.
- [16] Prosper Blog. *Charge-Offs Explained*. URL: <http://blog.prosper.com/2008/09/16/charge-offs-explained> (visited on 01/01/2015).
- [17] Prosper Blog. *Estimated Loss Rates*. URL: https://www.prosper.com/help/topics/general-estimated_loss_rates/ (visited on 01/01/2015).

- [18] Board of Governors of the Federal Reserve System (US). *Effective Federal Funds Rate [FED-FUNDS]*. URL: <https://research.stlouisfed.org/fred2/series/FEDFUNDS/>.
- [19] Peter Renton. *Prosper.com Has a New Look and a New Business Model*. URL: <http://www.lendacademy.com/prosper-com-has-a-new-look-and-a-new-business-model> (visited on 01/01/2015).
- [20] Wikipedia. *Peer-to-Peer Lending*. URL: en.Wikipedia.org/wiki/Peer-to-peer_lending (visited on 01/01/2015).

8 Appendix

8.1 Prosper’s Charged-Off vs. Default Terminology

After the shutdown, Prosper began using the term “Charged-Off” to mean that a loan is delinquent past 120 days. We observe more defaults than charge-offs before the shutdown, but more charge-offs than defaults after, likely because of the shift in terminology. According to Prosper’s website:

In general, a debt or account is considered charged off when it is unlikely that further payments will be received. Debts are usually charged off after they remain unpaid for a period of time (e.g., 90 to 180 days). Prosper uses the 120 days as the charge off threshold because loans that become over 120 days past due are eligible for sale to a debt buyer, and we have found that there is a steep drop-off in likelihood of further payments after 120 days of delinquency.⁵⁷

8.2 Out of Sample Performance by Credit Grade

	regime	AA	A	B	C	D	E	HR
1	Auction-Early	0.050	0.037	0.033	0.016	-0.001	0.014	0.098
2	Auction-Late	0.022	0.032	0.057	0.032	0.119	0.062	0.045
3	Prosper-rate	0.022	-0.040	-0.180	0.037	-0.004	0.006	0.051

Table 14: The excess returns earned by each regression model on loans in each Credit grade

8.3 Regressions to Predict Interest Rate

The rate regs shown in Table 15 show the increasing importance of both grade and hard credit variables in the later regimes, as discussed in Table 6.

⁵⁷Prosper Blog. *Charge-Offs Explained*. URL: <http://blog.prosper.com/2008/09/16/charge-offs-explained> (visited on 01/01/2015).

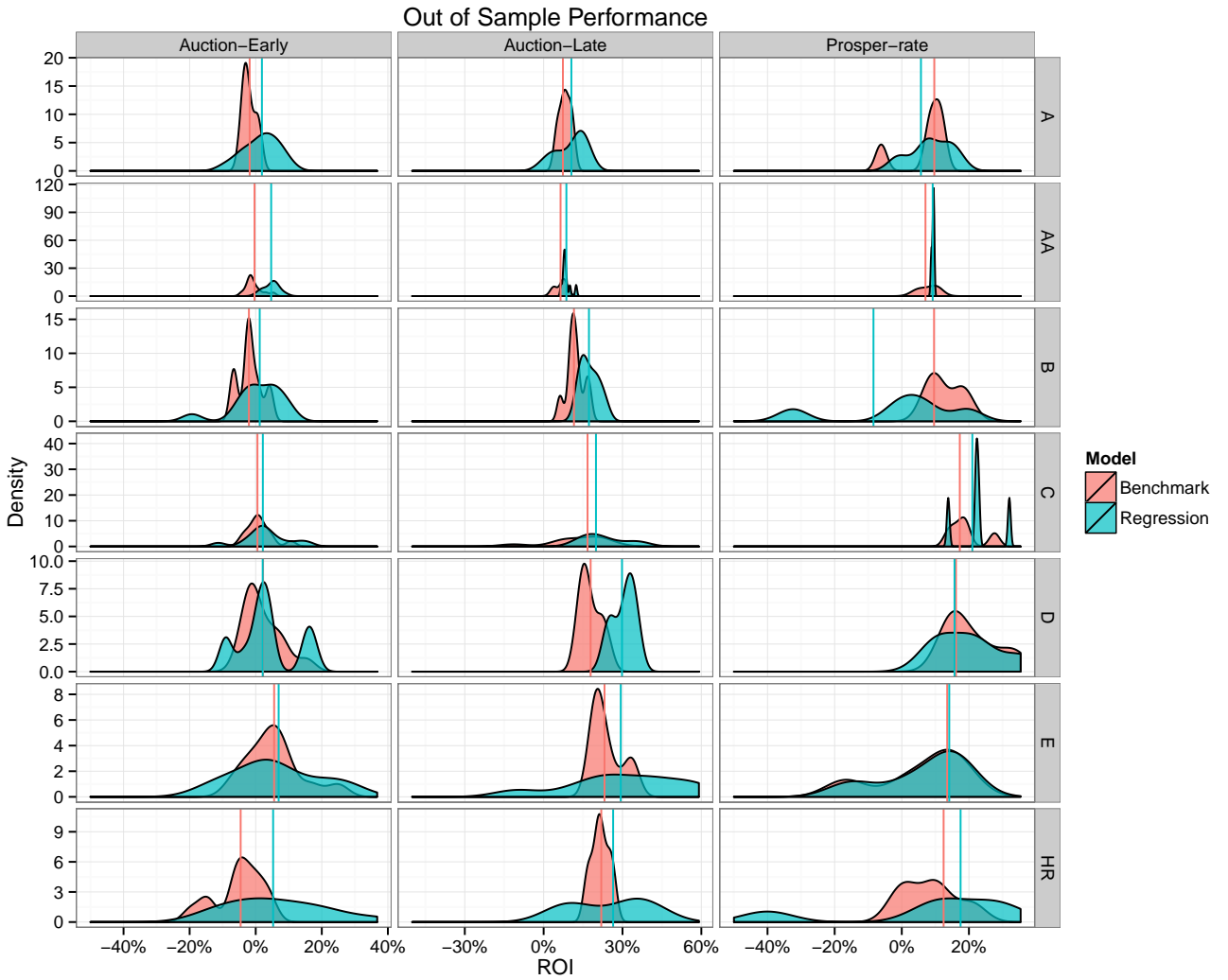


Figure 12: The rois of the loans selected by Each strategy by Credit Grade and Regime. Dotted lines indicate means. For example, In the Prosper-rate (2nd) column in the D (5th) row, the green distribution represents the ROIs of all the Grade D loans selected by the Linear Model

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
(Intercept)	0.11*** (0.02)	0.11*** (0.01)	0.17*** (0.01)	0.48*** (0.00)	1.00*** (0.01)	0.92*** (0.01)
CreditScoreEst	0.00 (0.00)	0.00*** (0.00)	0.00*** (0.00)	-0.00*** (0.00)	-0.00*** (0.00)	-0.00*** (0.00)
CurrentDelinquencies	0.00*** (0.00)	0.00*** (0.00)	0.00 (0.00)			
gradeAA	-0.02*** (0.00)	-0.09*** (0.00)	-0.14*** (0.01)			
gradeB	0.03*** (0.00)	-0.02*** (0.00)	-0.07*** (0.01)			
gradeC	0.05*** (0.00)	0.04*** (0.00)	-0.02* (0.01)			
gradeD	0.09*** (0.00)	0.10*** (0.00)	0.05*** (0.01)			
gradeE	0.13*** (0.00)	0.16*** (0.00)	0.10*** (0.01)			
gradeHR	0.12*** (0.00)	0.16*** (0.00)	0.11*** (0.01)			
gradeNC	0.11*** (0.02)					
gradeA		-0.06*** (0.00)	-0.12*** (0.01)			
R ²	0.48	0.88	0.97	0.37	0.47	0.39
Adj. R ²	0.48	0.87	0.97	0.37	0.47	0.39
Num. obs.	27659	7238	4550	27764	7238	4550

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 15: Regressions to predict interest rate. Columns (1) and (4) use Early Auction Regime data. Columns (2) and (5) use Late Auction regime data, columns (3) and (6) use Prosper-rate regime data.

	Auction-Early	Auction-Late	Prosper-rate
(Intercept)	0.0636*** (0.0038)	0.0155*** (0.0040)	0.0028 (0.0037)
BorrowerRate.sc	0.0370*** (0.0046)		
TotalCreditLinespast7years.sc	− 0.0187*** (0.0033)	− 0.0061* (0.0026)	
OpenRevolvingAccounts.sc	− 0.0126*** (0.0033)	−0.0015 (0.0027)	− 0.0055* (0.0022)
InquiriesLast6Months	0.0054*** (0.0006)		0.0144*** (0.0018)
CurrentDelinquencies.sc	0.0441*** (0.0024)		
DelinquenciesLast7Years.sc	− 0.0158*** (0.0025)		−0.0049 (0.0030)
MonthlyLoanPayment.sc	0.0129*** (0.0025)		
CreditScoreEst.sc	− 0.0357*** (0.0035)	−0.0065 (0.0034)	−0.0027 (0.0035)
TotalInquiries		0.0014*** (0.0004)	−0.0013 (0.0008)
PublicRecordsLast10Years.sc		−0.0018 (0.0034)	−0.0035 (0.0026)
DebtToIncomeRatio.sc		0.0344*** (0.0077)	
TotalProsperPaymentsBilled.sc		− 0.0054* (0.0021)	
ProsperPrincipalBorrowed.sc		0.0022 (0.0020)	
TotalProsperLoans.sc			−0.0023 (0.0015)
StatedMonthlyIncome.sc			−0.0006 (0.0009)
R ²	0.1237	0.0150	0.0372
Adj. R ²	0.1232	0.0125	0.0338
Num. obs.	13458	3189	2265

Y = Known-Default

Table 16: Coefficients extracted from Least Squares Regression to predict Known-Default in early auction, late auction, and Prosper-rate regime training data. Whereas the previous specification weights observations by maturity, this specification does not include weights, as a robustness check.