Essays on the Economics of Electronic Commerce and Social Networks

by

Zaiyan Wei

A Dissertation Submitted to the Faculty of the

Department of Economics

In Partial Fulfillment of the Requirements
For the Degree of

Doctor of Philosophy

In the Graduate College

The University of Arizona

2 0 1 5

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Zaiyan Wei, titled Essays on the Economics of Electronic Commerce and Social Networks and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

_____ Date: 05/01/2015
Keisuke Hirano

_____ Date: 05/01/2015
Mo Xiao

_____ Date: 05/01/2015
Stanley Reynolds

_____ Date: 05/01/2015
Mingfeng Lin

Final approval and acceptance of this dissertation is contingent upon the candidates submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

_____ Date: 05/01/2015
Dissertation Director: Keisuke Hirano

_____ Date: 05/01/2015
Dissertation Director: Mo Xiao

Statement by Author

This dissertation has been submitted in partial fulfillment of requirements for the degree of Doctor of Philosophy at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Signed: Zaiyan Wei

Acknowledgments

I would like to express my deepest gratitude to my co-advisors, Keisuke Hirano and Mo Xiao, for their guidance, insight, and patience. Working with them has been an amazing intellectual experience. I am truly grateful to Mingfeng Lin for introducing me to pursue the economics of electronic commerce. He has been a constant source of ideas and encouragement. Stanley Reynolds taught me a lot about auction theory. I could not complete the first chapter of this thesis without learning from him. I thank Mo Xiao and Mingfeng Lin for permission to include the joint research in my dissertation.

Qiang Ye and Bin Fang provided me the panel data from Tencent.com used in Chapter 4 of the thesis. I received helpful comments from Paulo Goes, Tanjim Hossain, Christopher Lamoureux, Adair Morse, Chris Parker, Siva Viswanathan, John Wooders, and Dajun Zeng, and seminar participants at the University of Arizona, Purdue University, and at the International Industrial Organization Conference, the Conference on Information Systems and Technology, the Workshop on Information Systems and Economics, the Conference on Internet Search and Innovation by the Searle Center at Northwestern University, and the Academic Symposium on Crowdfunding at the University of California, Berkeley.

I would especially like to thank my parents and my wife Yang Wang, for their support and forbearance. This dissertation is dedicated to my parents.

To my parents

Table of Contents

Table of Contents—*Continued*

List of Figures

List of Figures—*Continued*

List of Tables

List of Tables—*Continued*

List of Tables—*Continued*

Abstract

This dissertation studies the economics of electronic commerce and social networks. The first essay, "Market Mechanisms in Online Crowdfunding," systematically compares auctions with a platform-mandated posted-price mechanism in online crowdfunding. We exploit a regime change from auctions to platform-mandated posted prices on a debt-based crowdfunding platform. We develop a game-theoretic model that yields empirically testable hypotheses, then test them using detailed transactions data. Consistent with our hypotheses, we find that after the regime change, loans are funded with higher probability and higher interest rates. More important, all else equal, loans funded under the posted-price regime are more likely to default. While the posted prices may be faster than auctions that rely on the "crowd" to discover prices, auctions are not necessarily inferior in terms of overall social welfare.

The essay, "'Smart Money': Institutional Investors in Online Crowdfunding," studies institutional investors and their interactions with retail investors in online crowdfunding. Given their expertise, institutional investors are often referred to as "smart money" in financial markets, in the sense that these professional have selection ability and receive higher returns from their securities. We study whether institutional investors are indeed better able to screen borrowers in the marketplace than, and have any significant impacts on the behaviors of, retail investors. We find that although institutional investors indeed behave differently in terms of portfolio size and diversification strategies, their portfolios do not necessarily outperform those of retail investors. Institutional investors' bids have significant influence on the bidding strategies of retail investors, as well as final transaction outcomes. We find that this phenomenon is driven by the designation of "institutional investors" rather than just the size of their portfolios.

The last essay, "For Whom to Tweet? A Study of a Large-Scale Social Network,"

studies the effects of peer-groups sizes on individuals' contributions to public goods - tweets - in a large-scale and influential online social network. We attribute the highly-skewed distribution of tweets, which is observed from the network, to an individual's conflicting incentives of free-riding and maximizing social influence. We exploit the asymmetry of an individual's peer groups (followers and followees, groups of people following and being followed by the individual respectively) to disentangle these incentives, and devise empirical strategies to deal with the endogenous formation of one's peer groups. We find a larger group of followees leads an individual to tweet less, while a larger group of followers leads an individual to tweet more. With randomly generated 1% new links the total tweets will increase by 25% as the estimated follower effects dominate the followee effects.

1. Introduction

Market mechanisms and participant behaviors are two fundamental elements in any market or industry studies in economics. Particularly, transaction mechanisms, which specify rules of matching (between supply and demand) and terms (prices) at which transaction occurs, play an important role in determining market efficiency and total welfare. On the other hand, the behavior of market participants and strategic interactions between them have profound influence on transaction outcomes. This dissertation studies these two aspects in an emerging Internet financial market, as well as individual behaviors in online social networks.

The Internet and related technology adoption have facilitated market transactions that cannot be fulfilled regularly in the offline world. One of the examples is online crowdfunding. Although the earliest recorded crowdfunding practice dates back to 1885, when the project of the Statue of Liberty was funded through newspaper campaigns, it is with the Internet that entrepreneurs and borrowers can reach millions of potential investors, lenders, and donators. The disintermediation of this nascent Internet financial market enlarges the effects of transaction mechanisms and participant behaviors on transaction outcomes and allocative efficiency. This is the path taken in Chapter 2 and 3. These two studies take advantage of two "natural experiments" with sales mechanisms (auctions versus platform-mandated posted prices) and participant identities (institutional investors per se) on a debt-based crowdfunding, or peer-to-peer lending, platform. These switches in platform policies and rules provide unique opportunities to study the fundamental questions articulated above.

Another phenomenon arises with the Internet is online social networking. In addition to forming connections with others, individuals engage in information generation in social networks. The content generated can be viewed as public goods. The study in Chapter 4 examines individuals' conflicting incentives of free-riding and maximizing

social benefits in the provision of the online content, by exploiting the asymmetry in the effects of individuals' peer groups (groups of people following and being followed by an individual).

Market Mechanisms in Online Crowdfunding

The chapter, written jointly with Mingfeng Lin, compares auctions with a platform-mandated posted-price regime on a major peer-to-peer lending platform in the U.S.. The essay takes advantage of a regime change, from auctions to posted prices, to study the consequences on individual behaviors, transaction outcomes, and social welfare.

Online crowdfunding has emerged as an appealing new channel of financing in recent years. Two of the most popular transaction mechanisms are auctions (where the "crowd" determines the price of the transaction through an auction process) and posted prices (where the platform determines the price). While crowdfunding platforms typically use one or the other, there is little systematic research on the implications of such choices for the behavior of market participants, transaction outcomes, and social welfare. In this chapter, we address this question both theoretically and empirically. We first develop a game-theoretic model that yields empirically testable hypotheses, taking into account the incentives of the crowdfunding platform. We then test these hypotheses by exploiting the regime change.

Consistent with our hypotheses, we find that under platform-mandated posted-prices, loans are funded with higher probability, but the pre-set interest rates are higher than borrowers' starting interest rates in auctions. Governed by the higher initial interest rates, the contract interest rates for funded loans are also higher after the regime change. More important, all else equal, loans funded under the posted-price regime are more likely to default, thereby undermining lenders' returns on investment and their surplus from trading. Although platform-mandated posted prices may be faster in originating loans, auctions that rely on the "crowd" to discover prices are

not necessarily inferior in terms of overall social welfare.

"Smart Money": Institutional Investors in Online Crowdfunding

The next chapter, also joint work with Mingfeng Lin, studies institutional investors and their interactions with retail investors on the same peer-to-peer lending platform as in Chapter 2.

Given their expertise and adequate source of capital, institutional investors are often regarded as "smart money" in financial markets. "Smart money" is usually referred to as well-informed investors that exhibit selection ability and can yield higher returns from their choices of investment opportunities. Institutional investors, including mutual funds and pension funds, are shown to have the "smart money" effects (Gruber 1996, Zheng 1999). We test whether institutional investors are also "smart" in online crowdfunding setting, by investigating whether they are better able to screen borrowers in the marketplace.

Unlike traditional financial markets, where retail investors are very few and their trading volumes are largely negligible, the online crowdfunding is characterized by the coexistence of a large number of retail investors and institutional investors. Although institutional investors are studied extensively in the literature, there is little research on mutual influences between them and retail investors. We study these mutual influences by exploiting a "natural experiment" on displaying the identity of institutional investors on the platform. The platform changed from hiding the identity of institutional investors to the designation of "institutional lender." This switch in transaction rule helps us understand how institutional investors and retail investors respond, and the mutual influences between them.

We find that although institutional investors indeed behave differently in terms of portfolio size and diversification strategies, overall their portfolios do not necessarily outperform those of retail investors. Institutional investors' bids have significant

influence on the bidding behavior of retail investors, as well as final transaction outcomes. We find that this phenomenon is driven by the designation of "institutional investors" rather than just the size of their portfolios.

For Whom to Tweet? A Study of a Large-Scale Social Network

The last chapter, written jointly with Mo Xiao, studies the effects of peer-groups sizes on individuals' contributions to public goods - tweets - in a large-scale and influential online social network.

Online social networking, provided by platforms such as Facebook.com and Twitter.com, plays a critical role in one's social lives and has become a huge phenomenon in the last few years. Users of social networking service not only form connections online, but also (perhaps more importantly) engage in online postings routinely. In Twitter-type networks, users can *tweet* (writing short postings with texts and/or multimedia contents) or *retweet* others' tweets.

As public goods, tweets disseminate information and exert social influence. However, 50% of the individuals in the network under investigation post less than 6 tweets per month and contribute to less than 15% of the total tweets in stock, while the top 10% post average 40 tweets per month and contribute to more than 50% of the total tweets. We attribute the highly-skewed distribution of tweets to an individual's conflicting incentives of free-riding and maximizing social influence. We exploit the asymmetry of an individual's peer groups (followers and followees, groups of people following and being followed by the individual respectively) to disentangle these incentives, and devise empirical strategies to deal with the endogenous formation of one's peer groups.

We find that a larger group of followees leads an individual to tweet less, while a larger group of followers leads an individual to tweet more. These asymmetric effects suggest that individuals free ride information from their followees but receive social

benefits from tweeting to their followers. With 1% more links the total tweets in stock will increase by 25%, as the estimated follower effects dominate the followee effects.

2. Market Mechanisms in Online Crowdfunding

Zaiyan Wei and Mingfeng Lin

2.1. Introduction

Online crowdfunding is an emerging Internet market for capital (Agrawal et al. 2013, Burtch et al. 2013, Lin et al. 2013). It broadly refers to the aggregation of funds from individual lenders, investors, or donors to support a person or an organization via the Internet. Unlike traditional markets facilitated by financial intermediaries, online crowdfunding allows the demand and supply of funds to match directly in a many-to-one fashion. A fundamental question in online crowdfunding, and in any other markets, is how price discovery mechanisms affect the matching of market participants, their behaviors, and transaction outcomes. All types of crowdfunding[1] are inevitably built upon certain market mechanisms. The problem of mechanism choice is especially relevant in such burgeoning markets, and has important managerial as well as policy implications (Wu 2013, Biais et al. 2002, Hortaçsu and McAdams 2010).

Auctions and posted prices are two widely adopted transaction mechanisms. In particular, auctions have been extensively used long before the Internet age; but it is with the Internet that online auctions can now help match millions of potential buyers and sellers in highly decentralized virtual markets.[2] Naturally, when debt-based crowdfunding (online market for unsecured personal loans; also known as peer-

[1] In addition to debt-based crowdfunding (also known as peer-to-peer lending) that we study, there are also reward-based (such as Kickstarter.com), donation-based (such as Kiva.org), and equity-based crowdfunding sites.

[2] Auctions have long been a dominant market mechanism in electronic commerce. A prominent example is eBay.com, one of the earliest and most successful examples of e-commerce, where auctions are extensively used to match buyers and sellers through a competitive process and determine the market-clearing price for products. Auctions are also widely used in online business-to-business (B2B) procurements and many other contexts (Lucking-Reiley 2000, Bajari and Hortaçsu 2004, The Economist 2000).

to-peer lending) market emerged in the past decade, auctions were often adopted as the typical funding mechanism to match borrowers and lenders. In fact, one of the earliest such platforms, Prosper.com, was called "the eBay for personal loans" since it used an auction mechanism to aggregate bids from investors.

Auction however, is not the only way to match borrowers and lenders in crowdfunding. Another mechanism is posted prices, where the "price" (interest rate) of each loan is determined by the platform's in-house experts rather than the "crowd." LendingClub.com, another major player in this arena in the US, has long matched borrowers and lenders in that manner. Outside the US in many countries such as China, Canada, and UK, auctions and posted prices also represent the two major market mechanisms in online peer-to-peer lending. To date, however, we still know little about how different mechanisms affect the behavior of participants in this market and transaction outcomes.[3] We study this question in the context of online crowdfunding, exploiting a regime change mandated by Prosper.com.

On December 20, 2010, Prosper.com unexpectedly abandoned its well-known auction model (where each investor bids both a dollar amount and an interest rate) and switched the entire website to a posted-price mechanism (Renton 2010) (where the interest rate is preset and lenders only invest a dollar amount).[4] Since then, all personal loans take place at the new pre-set interest rate setting. This regime change was effective immediately on the whole website, and unanticipated by market participants; it therefore provides an ideal opportunity to investigate how different market mechanisms impact participant behaviors and market efficiency, especially considering the incentives of the platform itself.

[3]Some studies such as Einav et al. (2013) and Hammond (2010, 2013) address a similar question in the context of eBay.com. It should be noted however, that posted-price in eBay is decided by sellers for each auction; whereas in crowdfunding this is a site-wide market mechanism, and the platform determines the interest rate.

[4]Renton (2010) provides evidence that this change was largely unexpected. Prosper.com's corporate blog about the regime change can be found at http://blog.prosper.com/2010/12/30/exciting-new-enhancements-at-prosper/.

Understanding how online market mechanisms affect participant behaviors and transaction outcomes is an important and in fact fundamental question for the nascent but burgeoning industry of online crowdfunding (Agrawal et al. 2013, Lawton and Marom 2010), of which Prosper.com is but one example. Other types of crowdfunding, such as equity-based crowdfunding (e.g. Seedrs.com), long-term debt crowdfunding (e.g. AbundanceGeneration.com), as well as crowd-based alternative finance markets built on different types of account receivables, almost always use one of these two mechanisms. Only if we come to a comprehensive understanding of consequences of different market mechanisms and the platform's incentive, can we design a more efficient and effective marketplace to match demand and supply of funds, and ensure its long-term viability. As the US prepares for further growth in this industry, especially the upcoming equity-based crowdfunding legalized by the 2012 "Jumpstart Our Business Startups" (JOBS) Act,[5] this research question has important and timely policy implications.

To address this research question, we first propose an analytical model comparing the multiunit uniform price auction with the posted-price mechanism in the context of peer-to-peer lending, then test its predictions using data from Prosper.com around the time of the regime change. A novel and unique feature of our context is that after the regime change, Prosper.com (the platform), not the borrowers, sets the interest rates. By specifically taking into account the incentives of the platform, our model predicts that Prosper.com, as the pricing agent in the posted-price regime, will assign higher interest rates compared to what the borrowers would have chosen as their reserve interest rates in auctions. Meanwhile, loans will be funded with higher probability under the posted-price mechanism. By natural extension, the contract interest rates

[5]Specifically, the JOBS Act amends Section 4 of the Securities Act to exempt security issuers from some requirements when they offer and sell up to $1 million in securities, provided that individual investments do not exceed certain thresholds and other conditions are satisfied. One of these conditions is that issuers can use a crowdfunding intermediary. More information can be found on the U.S. Securities and Exchange Commission (SEC) webpage for the Act: http://www.sec.gov/spotlight/jobs-act.shtml.

for loans funded in the posted-price regime will be higher as well, since in auctions the final interest rates cannot be higher than borrowers' reservation interest rates. Further, since loans with higher interest rates are more likely to default (Stiglitz and Weiss 1981, Bester 1985), loans funded under posted prices should be more likely to default.

We empirically test these predictions using data from Prosper.com around the time of the regime change. We focus on listings initiated during a short time period before and after the regime change, specifically from August 20, 2010 to April 19, 2011. We compare the pricing (the initial interest rate of a listing), funding probabilities (the listings' probability of full funding), as well as the contract interest rates of funded loans. Consistent with our theoretical predictions, we find that the initial interest rates under the posted-price mechanism are indeed higher than those in auctions, and listings after the regime change are much more likely to be funded. We also find that the contract interest rates in posted prices are on average higher. In other words, while the regime change indeed led to a higher funding probability, it came at a cost of higher interest rates.

More important, and as a result of the higher contract interest rates, we find evidence that loans funded under the posted-price regime are more likely to default than those funded in auctions. In addition, results from a survival analysis indicate that the hazard rate of default is on average higher for loans initiated after the regime change. These are consistent with the literature that the interest rate is an important predictor of default (Stiglitz and Weiss 1981, Bester 1985), and points to an unintended consequence of eliminating the auctions model.

We also investigate the effect of the mechanism change on lenders' behaviors, including herding and bidding strategies. We find that lenders tend to submit larger bids, and submit those bids sooner, under the posted-price regime than in auctions. This further supports and explains the finding that loans are funded faster for borrowers.

We further study whether the regime change is "better" or "worse" for various stakeholders in this market. We find that, due to the higher probability of funding under posted prices, the platform's profit from loan initiating fees is significantly higher after the regime change. On the other side of the market, lenders' gains from investment are *ex ante* ambiguous due to conflicts between higher interest rates and default rate. We calculate lenders' return on investment, and find that the regime change indeed leads to lower returns. We find further evidence of deteriorations of lenders' wellbeing by comparing lender surplus from the transactions. These findings lend support to the platform's rationale to change mechanisms, but again point to unintended consequences, which suggests that the new mechanism is not always better than the old-fashioned auctions.

Our paper first contributes to the growing literature on crowdfunding in general, and the research on peer-to-peer lending in particular. Recent investigations include Iyer et al. (2010), Zhang and Liu (2012), Rigbi (2013), Lin et al. (2013), Burtch et al. (2013), Hahn and Lee (2013), Kawai et al. (2014), and Kim and Hann (2014). More generally, we also contribute to the literature on crowdsourcing, such as Dellarocas et al. (2010), Hosanagar, Han and Tan (2010), Chen et al. (2014a), and Liu et al. (2014). To our knowledge, our study is the first theoretical and empirical comparison of the two transaction mechanisms exploiting an exogenous change, especially in online crowdfunding. More broadly, our work contributes to a long-term debate over the optimal sales mechanism. For instance, Bulow and Klemperer (1996) compare auctions against negotiations, whereas Wu (2013) provides theoretical comparison between single object auctions and posted-price mechanisms. More recent comparisons of these market mechanisms, particularly between posted prices and auctions, can be found in such diverse fields as treasury auctions (Ausubel and Cramton 2002, Hortaçsu and McAdams 2010) and initial public offering (IPO) (Biais et al. 2002, Zhang 2009). Unlike our setting, in most of these papers they compare auctions with the posted-price mechanisms where sellers, instead of platforms or market makers,

set prices. Our results on lenders' herding behavior contributes to the observational learning literature, as developed theoretically in Banerjee (1992) and Bikhchandani et al. (1992), and empirically in Simonsohn and Ariely (2008) and Zhang and Liu (2012).

2.2. Research Context

Since the inception of Zopa.com in 2005 in the United Kingdom, online peer-to-peer lending has witnessed rapid growth around the globe. In the United States, Prosper.com and LendingClub.com are the two largest platforms. Prosper.com officially opened to the public on February 13, 2006. By the end of 2014,[6] there had been over 2 million registered members (either as a borrower, a lender, or both) on Prosper.com. More than $100,000$ unsecured personal loans, valued over USD 2.4 billion in total, have been funded.

A brief outline of the funding procedure on Prosper.com is as follows.[7] A potential borrower starts the borrowing process by registering on Prosper.com and verifying their identity. After that, the borrower may post an eBay-style listing, describing the purpose, the requested amount of dollars, and the length of the loan (typically 3 years). The *initial* interest rate can be specified in two ways. Before December 20, 2010 the borrower needs to specify a reservation (maximum) interest rate they are willing to accept. After the regime change to posted prices (pre-set interest rate), Prosper.com *presets* an interest rate for the loan based on the borrower's creditworthiness.

Prior to the regime change on December 20, 2010, once the listing is posted with a specified duration (typically 7 days), a multiunit uniform price auction will be

[6]Statistics about Prosper.com are obtained from summarizing our dataset, or the corporate prospectus at https://www.prosper.com/prospectus/.

[7]Our descriptions emphasize website features that are most relevant to our study, but may not cover all institutional details. Interested readers can refer to other studies of Prosper.com such as Zhang and Liu (2012), Lin et al. (2013), and Freedman and Jin (2011) for further details.

conducted until the listing is either fully funded, or expired. Any verified Prosper.com lender can bid in the auction. In their bids, lenders specify the amount of funds that they would like to invest, and the minimum interest rate at which they are willing to lend. A single lender's bidding amount does not have to cover the borrower's requested amount, and multiple lenders are allowed to fund the loan. All lenders can observe previous lenders' identities and their bidding amount. Lender's priority in participating in the loan is ranked by the interest rate that they specified in their bids, where those with lowest interest rates are most likely to participate. During the auction process, the ongoing interest rate for the loan is either the starting interest rate that the borrower sets at the beginning if the full amount of the loan is not yet funded; or, once the funds from all lenders exceed the amount requested by the borrower, the lowest interest rate among all lenders that are outbid (excluded from funding the loan). At the end of the auction process, if the loan receives full funding, the winners will be all lenders who specified the lowest interest rates among all those who bid; the contract interest rate will be the ongoing interest rate at that time, which can be either the lowest losing interest rate or the reserve interest rate. In a sense, the borrower sets the initial interest rate and the auction helps "discover" the contract interest rate.

On December 20, 2010, Prosper.com unexpectedly eliminated this auctions model. Since then, the interest rate is preset by Prosper.com, based on the website's evaluation of the borrower's creditworthiness. The borrower can no longer use the auction format. Lenders now only specify a dollar amount for their investment, implicitly accepting the preset interest rate. Multiple lenders are still allowed to fund a loan. Listings will not be converted into loans unless the full amount requested by the borrower is funded before listing expiration, as no partial funding is allowed during our study period; and the contract interest rate is the rate preset at the beginning. Table 2.1 summarizes the key difference between these two regimes.

When announcing the regime change, Prosper.com argued that the new posted-

Table 2.1: A Comparison of Auctions vs. Price Posting - The Regime Change

	Auction Regime[a]	**Posted-Price Regime**
Initial interest rate	*Chosen by the borrower*	*Preset by Prosper.com*
Contract interest rate	*Prevailing interest rate at the end of auction*	*Initial interest rate*

[a]This table summarizes the content of the regime change. Note that Prosper.com eliminated their auctions model and switched to posted prices on December 20, 2010.

price regime would allow "*...a quicker deployment of funds*," and borrowers would "*likely get their loan listing funded sooner as well*". "Quicker deployment" refers to the fact that while lenders fund their Prosper.com account before they can invest, their Prosper.com account bears no interest. Their funds can only generate returns when invested in a loan that is successfully funded and originated. To understand whether the regime change indeed had these effects, and more important whether there are other consequences of this change, we develop a stylized model to generate several empirically testable hypotheses. We then test these hypotheses using data from Prosper.com around the time of the regime change.

It should be noted that in addition to the stylized analytical model, our hypotheses can be intuitively deducted as well. In other words, even though the model that we will describe in the next section inevitably requires various assumptions to solve, its intuition is in fact straightforward and will be described in detail. This approach (using stylized analytical models to derive predictions or hypotheses, then empirically test them) has been used in many empirical studies in information systems and economics. Some recent examples include Arora et al. (2010), Sun (2012), and Lemmon and Ni (2014). The stylized model, however, highlights key features of the mechanisms and provides a more formal treatment of the hypothesis derivation.

2.3. A Model of Market Mechanisms

In this section, we propose a model to compare the multiunit uniform price auctions with the platform-mandated posted prices, in the context of Prosper.com. Our model is based on the share auction model proposed by Wilson (1979), and further developed in Back and Zender (1993) and Wang and Zender (2002). We develop the following model to highlight the key difference between auctions and the posted-price mechanism (Chen et al. 2014b).

We consider a borrower posting a personal loan request on the platform. In an auction, either the lowest losing interest rate or the borrower's initial interest rate sets the contract interest rate for all winning lenders, if the loan is funded. In the posted-price setting, Prosper.com presets the interest rate for the loan, and the borrower either accepts or rejects this offer. Once the borrower accepts and the listing is created, any lender can "purchase" a portion of the loan at the pre-set interest rate. All lenders will fund the personal loan based on this rate. We denote p as the contract rate of a loan funded from either an auction or a posted-price sale. Finance literature (Stiglitz and Weiss 1981, Bester 1985) suggests that higher contract interest rates cause, in general, higher default rates. We let $\delta(p)$ be the default rate given the contract interest rate is p, therefore we have $\delta^{'}(\cdot) \geq 0$.

On borrower side, there is a variable cost c for each dollar he or she borrows from the site. If the loan is successfully funded, this cost will be deducted before the funds are transferred to the borrower. In the context of Prosper.com, the website charges a closing fee for each funded loan. The borrower may also incur other explicit or implicit costs, such as their time and efforts in creating the loan request. For simplicity, we assume that the borrower is requesting a loan with Q units, and the discount rate for the borrower is τ. This can be interpreted as the interest rate of his or her best outside option, or the lowest interest rate from other channels to obtain the loan. The maximum interest rate that the borrower is willing to pay for this loan

on Prosper.com will then be $\tau - c$.

On the other side of the market, suppose there are N potential lenders for this particular personal loan. We assume $N \gg Q$, i.e., there are always enough lenders to fund the loan if the price is right. This assumption is reasonable considering the large pool of lenders on Prosper.com. We assume that each lender supplies at most one unit of the loan, and has an independent private willingness to lend (WTL). A lender's WTL is the lowest rate at which they are willing to lend the borrower. They will never fund the listing at any interest rate below this WTL. This can equivalently be considered the lender's true valuation of the loan, or the maximum risk-free interest rate from the lender's outside options. The private value assumption is not restrictive given the fact that promissory notes were not allowed to be traded on Prosper.com during our study period. In other words, there was no resale or secondary markets for Prosper loans, thus a lender's WTL is independent of others' valuations. Let W_n denote lender n's WTL, $n = 1, 2, ..., N$. Let w_n denote its realization. We assume that W_n is IID (independent and identically distributed) with CDF $F_W(\cdot)$, and PDF $f_W(\cdot)$. We let $W^{N:k}$ denote the k-th lowest value among N IID willingness-to-lend, and $k = 1, 2, ..., N$. Let $w^{N:k}$ denote its realization. We denote the distribution of $W^{N:k}$ by $G_k(\cdot)$ (or PDF $g_k(\cdot)$).

2.3.1. Auctions

We model a sealed-bid multiunit uniform price auction with single-unit demand. In such an auction, the market clearing interest rate is set by the lowest losing bid or the borrower's reserve rate. The lenders all incur a nonnegative transaction cost, λ. Sources of this cost can be the uncertainty associated with the auction process, i.e., the efforts required to judge a borrower's creditworthiness, whether the ongoing interest rate reflects that quality, the need to observe the behavior of other investors (Zhang and Liu 2012), and the fact that the loan may not be fully funded at the

end. This cost λ therefore increases the lenders' minimum acceptable interest rate to $W_n+\lambda$. In other words, lenders will require higher interest rates than their true values to compensate for the transaction costs associated with the auction mechanism. We normalize λ to zero under posted prices, reflecting the idea that lenders in auctions incur strictly higher transaction costs than in posted prices.

In the private value paradigm, auction theory (Krishna 2009) predicts that the weakly dominant strategy for a lender is to submit their true value $W_n+\lambda$. Compared to models in the auction literature, our context presents two complications. First, lenders' bids are bounded above by the borrower's reserve interest rate. Second, lenders will take into account the loan's potential likelihood of default, which is closely related to the contract interest rate. However, it is straightforward to show that these two factors do not affect lenders' equilibrium bidding strategy, and their weakly dominant strategy is still to bid their true valuation. Thus, the winners will be the Q lenders with the lowest true WTL, and each of them wins one unit of the loan. Knowing the lenders' bidding strategy, the borrower will choose a reserve interest rate, r^*, to maximize his expected payoff. In the context of Prosper.com before the regime change, this reserve interest rate corresponds to the initial interest rate set at the very beginning of the auction process.[8]

To develop the borrower's payoff function, we notice that this is a personal loan market with borrowing and repayment obligations in a future date. Specifically, if the loan is funded, the borrower receives Q units from the winning lenders immediately, but is also committed to pay back the principal and interest within a certain time period. Without loss of generality, we assume that this is a one-period loan. The total repayment amount will be $Q\cdot(1+p_A(r)+c)$, where $p_A(r)$ is the market clearing interest rate if the loan is funded. The subscript indicates that the listing is funded

[8]Recall that we assume the borrower's discount rate to be τ. We can interpret it as the interest rate of the borrower's best outside option, *i.e.*, the lowest rate he or she is offered from other financial institutions. Based on this interpretation, the borrower's reserve interest rate must satisfy $r<\tau$. Our results later verify this observation.

from an auction, and this interest rate is a function of the borrower's reserve interest rate. If the borrower is risk neutral, then their payoff in terms of present value will be $\pi_A = Q - Q \cdot (1 + p_A(r) + c) / (1+\tau)$. We simplify the equation and get the borrower's payoff (conditional on being funded) as a function of r:

$$\pi_A = \frac{Q \cdot (\tau - p_A(r) - c)}{1 + \tau}.$$

It is clear that the market clearing interest rate will vary across listings. Specifically, the market clearing rates will be equal to

$$\begin{cases} w^{N:Q+1} + \lambda, & \text{if } w^{N:Q+1} + \lambda \leq r; \\ r, & \text{if } w^{N:Q} + \lambda \leq r < w^{N:Q+1} + \lambda. \end{cases}$$

The first case corresponds to the scenario where the lowest losing interest rate is less than the borrower's reserve rate, thus setting the price of the loan. The second case is where the reserve rate sets the price since no losing lender is willing to bid less than that rate. Note that the probability of being funded is $\Pr\left(W^{N:Q+1} + \lambda \leq r\right)$ and $\Pr\left(W^{N:Q} + \lambda \leq r < W^{N:Q+1} + \lambda\right)$ respectively. Then the expected market clearing rate $p_A(r)$ conditional on the reserve interest rate r will be $E[W^{N:Q+1} + \lambda | W^{N:Q+1} + \lambda \leq r]$ and r, respectively. Therefore, we can write the borrower's expected payoff as

$$\begin{aligned} E\pi_A = &\frac{Q \cdot \left[\tau - E[W^{N:Q+1} + \lambda | W^{N:Q+1} + \lambda \leq r] - c\right]}{1 + \tau} \cdot \Pr\left(W^{N:Q+1} + \lambda \leq r\right) \\ &+ \frac{Q \cdot (\tau - r - c)}{1 + \tau} \cdot \Pr\left(W^{N:Q} + \lambda \geq r > W^{N:Q+1} + \lambda\right). \end{aligned}$$

The borrower maximizes their expected payoff by choosing the initial reserve interest rate r. Proof in Appendix A.1 shows that the optimal reserve interest rate r^* is defined by the following implicit function,

$$r^* = \tau - c - \frac{G_Q(r^* - \lambda) - G_{Q+1}(r^* - \lambda)}{g_Q(r^* - \lambda)}. \tag{2.1}$$

The numerator of the ratio is exactly the probability that the market clearing interest rate is the reserve interest rate. Thus we can interpret this ratio as the

normalized premium deducted to rule out the possibility that the lowest losing bid is less than the reserve interest rate. We can further simplify Equation (2.1) to get

$$r^* = \tau - c - \frac{F_W(r^* - \lambda)}{Q \cdot f_W(r^* - \lambda)}. \tag{2.2}$$

An important implication of this result is that the optimal reserve interest rate is strictly less than the borrower's underlying maximum acceptable rate, $\tau - c$. This allows the borrower to secure a positive expected profit. We also recall that $F_W(\cdot)$ and $f_W(\cdot)$ are the distribution functions of lenders' values. The result implies that the optimal reserve price is independent of the number of lenders. If W_n has a log-concave distribution, r^* is increasing with the quantity Q.

2.3.2. (Platform-mandated) Posted Prices

We now model the dynamics among borrowers, lenders, and the platform under the posted-price mechanism. Importantly, since Prosper.com sets the initial interest rate (p^*) under this new mechanism, rather than borrowers, we specifically consider its incentive in pricing borrower loans to maximize its own expected profit. The borrower either accepts or rejects the price that Prosper.com sets for their loan. Upon accepting the price, the listing will be posted on the platform, and the interest rate will be fixed at the pre-set level.

Before we model Prosper.com's decision process, we consider a hypothetical setting where the borrower gets to choose the interest rate under posted prices. If the borrower could have chosen the fixed interest rate level, their expected payoff can be written as

$$E\pi_B = \frac{Q \cdot (\tau - p - c)}{1 + \tau} \cdot \Pr\left(W^{N:Q} \leq p - \delta(p)(1 + p)\right),$$

where $\delta(\cdot)$ is the default rate function.The B subscript indicates that it is (for now) the borrower's choice. To see the inequality in the equation, notice that a lender will find the loan profitable if and only if $1 + w_n \leq (1 + p)(1 - \delta(p))$, which can be

simplified to $w_n \leq p - \delta(p)(1+p)$. Let $\gamma(p) = p - \delta(p)(1+p)$, which is the loan's expected rate of return.

The borrower would maximize their revenue by choosing p. The following equation characterizes this optimal price level implicitly,

$$p_B^* = \tau - c - \frac{G_Q(\gamma(p_B^*))}{g_Q(\gamma(p_B^*)) \cdot \gamma'(p_B^*)}. \tag{2.3}$$

It can be shown that the relationship between p_B^* and the reserve interest rate r^* in auctions depends on the expected rate of return function, $\gamma(\cdot)$, as well as the distribution of lenders' valuations. Under mild conditions, p_B^* is strictly less than r^*. This is an extension to the basic model in Einav et al. (2013). Under the usual commodity economy interpretation, therefore, the seller assigns higher "price"—analogous to lower interest rate in the current context—in the posted-price setting, given the common value setup.

We now return to Prosper.com's decision. The platform presets the interest rate p for a particular loan. The borrower's strategy is to pick a threshold or cut-off rate $\tilde{p}$. If p is lower than this cut-off, the borrower will accept the offer. If it is higher, they will reject it and leave the market. In other words, the borrower accepts p if $p \leq \tilde{p}$, and rejects otherwise.

Note that again the probability of being fully funded is $\Pr\left(W^{N:Q} \leq \gamma(p)\right)$. Similar to our analysis of the auctions, the borrower's expected payoff for accepting Prosper.com's pre-set interest rate can be shown to be $\frac{Q \cdot (\tau - p - c)}{1+\tau} \cdot \Pr\left(W^{N:Q} \leq \gamma(p)\right)$, while rejecting the offer generates zero payoff. At the threshold, the borrower is indifferent between accepting and rejecting. That is, $\frac{Q \cdot (\tau - p - c)}{1+\tau} \cdot \Pr\left(W^{N:Q} \leq \gamma(p)\right) = 0$. Then it is easy to show that the cutoff price is $\tilde{p} = \tau - c$.

Suppose now that the platform knows the borrower's true cost c and discount rate τ, and thus $\tau - c$.[9] Prosper.com's profit comes from the fees that it charges on

[9]To see why this is a reasonable assumption, notice that we interpret τ as the borrower's lowest interest rate from their outside options. As a professional financial organization, Prosper.com has

funded loans. We let α denote this closing fee. This fee is fixed, in the sense that Prosper.com does not change it in the short run (consistent with what we observe for our study period). Then we can write down Prosper.com's expected profit as,

$$E\pi_P = \alpha \cdot Q \cdot \Pr\left(W^{N:Q} \leq \gamma(p)\right).$$

Prosper.com chooses an interest rate to maximize this profit given $p \leq \tau - c$ and $p \geq 0$. It can be shown that the following assumption is a sufficient condition under which Prosper.com assigns the highest possible interest rate, *i.e.*, $\tau - c$.

Assumption 1. The hazard rate for the default rate function, $\delta'(p)/(1-\delta(p))$, is bounded above by $1/(1+\tau-c)$, *i.e.*, for all $p \in [0, \tau - c]$,

$$\frac{\delta'(p)}{1-\delta(p)} \leq \frac{1}{1+\tau-c}.$$

Assumption 1 adds some functional form restrictions but are fairly reasonable and intuitive. Under the assumption, we can show that $\Pr\left(W^{N:Q} \leq \gamma(p)\right)$ is a non-decreasing function of p.[10] This implies that Prosper.com will choose an interest rate as high as possible to maximize its expected profit. To summarize, in a posted-price setting Prosper.com presets an interest rate,

$$p^* = \tau - c. \tag{2.4}$$

2.3.3. Comparisons and Predictions

An immediate observation is that $p^* > r^*$. In other words, Prosper.com will preset an interest rate higher than what the borrower would have chosen in auctions. Also note that the probability of being funded in the auctions, $\Pr\left(W^{N:Q} + \lambda \leq r^*\right)$, is strictly less than that under posted prices, $\Pr\left(W^{N:Q} \leq \gamma(p^*)\right)$.[11] Conditional on the loan

the whole set of credit information (including borrower's detailed credit reports) necessary to derive the interest rate that the borrower is able to obtain from other channels.

[10] To see this, note that Assumption 1 is a sufficient condition for the expected rate of return function, $\gamma(\cdot)$, to be nondecreasing on $[0, \tau - c]$.

[11] A necessary condition is $\delta(\tau - c) \leq (\tau - c)/(1 + \tau - c)$, which suggests that unless the default rate is too high, the funding probability under the posted-price regime is strictly higher.

being funded, since the initial interest rate in the posted-price regime is higher than that in the auctions, and the contract rate is identical to the initial rate under posted prices, the contract rate is also strictly higher than the contract rate in auctions. In turn, the loan default rate, which the finance literature has shown to be increasing in contract interest rates, should also be higher in the posted-price regime. We therefore summarize the following hypotheses from the model:

1. Hypothesis 1: All else equal, the initial interest rates assigned by Prosper.com under the posted-price mechanism should be higher than the initial interest rates chosen by borrowers in auctions.

2. Hypothesis 2: Conditional on loans being funded, the contract interest rates under the posted-price mechanism are higher than those under the auction regime.

3. Hypothesis 3: The funding probability under the posted-price mechanism is higher than in auctions.

4. Hypothesis 4: For funded loans, the probability of default under the posted-price mechanism is higher than in auctions.

2.3.4. Intuitions for the Stylized Model

Although we derived the above hypotheses through our stylized model, the intuitions behind them are also straightforward. In this subsection we discuss these intuitions to complement the foregoing model.

Under both regimes, the borrower moves before lenders, as the borrower needs to commit to an initial (asking) interest rate before potential lenders decide whether or not to invest. The website's profit comes from originated loans. In addition, there are two key features for the posted-price mechanism: First, the pricing power lies with

Prosper.com; and second, by serving as the pricing agent, Prosper.com reduces the uncertainty associated with ongoing interest rates in auctions.

Under auctions, the borrower is faced with a tradeoff between the initial (asking) interest rate and probability of funding. The borrower will favor a lower starting interest rate because they cannot revise that rate once the auction starts, and if the rate is unnecessarily high, that rate will be effective for the life of the loan (c.f. "winner's curse"). In fact, if the lower rate does not attract sufficient funding, it is virtually costless to post another request with a higher asking rate. Under posted prices, Prosper.com implicitly ensures that the interest rate reflects borrower quality. In other words, at the same starting interest rate, lenders will be more likely to place bids under posted prices than auctions. Prosper.com is therefore able to extract surplus from the borrower because the borrower has to move first, and by setting the interest rate higher, potential lenders are more likely to participate and fund loans – after which the website is able to charge fees. This is the intuition behind hypothesis 1, and the other hypotheses follow.

In addition to these predictions, for the higher initial interest rate to induce higher funding probabilities, we should be able to observe that lenders are more willing to bid under posted prices. Hence, an auxiliary hypothesis from our model is that lenders should be willing to place bids earlier in auctions, and place larger bids. And because the price information (interest rate) from Prosper.com website should carry more weight than the asking rate of borrowers, lenders should be able to rely less on the behaviors of others to judge borrower quality. Hence, the rational herding behavior documented in Zhang and Liu (2012) should be reduced, if not eliminated, under posted prices. However, our focus remains the four hypotheses above.

We next turn to transactions data from Prosper.com to empirically test these hypotheses. In the next section, we introduce our datasets and discuss the construction of our samples.

2.4. Data and Sample

We obtained data from Prosper.com on January 14, 2013. The dataset contains all transactions since the website's inception in February 2006, including both funded and failed listings. For each listing, we obtain an extensive set of variables including the requested loan amount, initial interest rate, loan term, timestamps (the starting time and ending time), results of the listings, and the repayment status of funded loans as of our data collection date. The borrower's credit information includes their Prosper rating (a letter grade designated by Prosper.com indicating the borrower's creditworthiness), credit score range, debt to income ratio, as well as extended credit information such as the number of credit lines, delinquency history, and bank card utilization. We also obtain detailed information for each bid, including the identity of the bidder (lender), bid amount, timestamp, and outcome (winning or losing). Finally, for successfully funded loans, we have the loan origination date, contract interest rate, repayment status updated monthly, and so on.

Prosper.com eliminated auctions on December 20, 2010. We therefore construct our main sample to include all listings that are posted between August 20, 2010 and April 19, 2011.[12] During this period, the regime change under investigation was the only policy change on the platform. We drop listings that are suspected of borrower identity theft and repurchased by Prosper.com.

Table 2.2 and 2.3 summarize the main sample used in our empirical analysis. There were $13,017$ listings posted during the period. $8,470$ of them began before December 20, 2010, and $4,547$ listings were initiated in the posted-price regime. Out of these listings, $4,446$ were funded and became loans. Among these loans, $1,925$ were funded using auctions, and $2,521$ were funded under posted prices. Thus, roughly 22.7% of

[12]Prior to August 2010, Prosper.com allowed borrowers to use "automatic funding" for their auctions, where the borrower sets a reserve interest rate and the auction process would end as soon as 100% funding was reached. This funding option was discontinued by the website, and during the time that we study, no such auctions existed.

listings in the auction regime were successfully funded, whereas that ratio is about 55.4% under posted prices. The table also shows that on average, the contract interest rates the borrowers pay in the loans are lower in posted-price regime on average. As another summary, we depict the daily average quantities of the four outcome variables in Figure 2.1. We observe that the daily ratio of funded and defaulted loans both increase after the regime change. In contrast, the average asking interest rates and contract interest rates for funded loans both decline under posted prices without controlling for any covariates. To test the hypotheses that we derived in Section 2.3, we seek to identify how funding probabilities, initial (asking) interest rates, contract interest rates, and default rate change as a result of the regime change from auctions to posted prices, holding all observed characteristics of listings and loans constant.

Figure 2.1: Daily Transaction Outcomes before and after the Regime Change

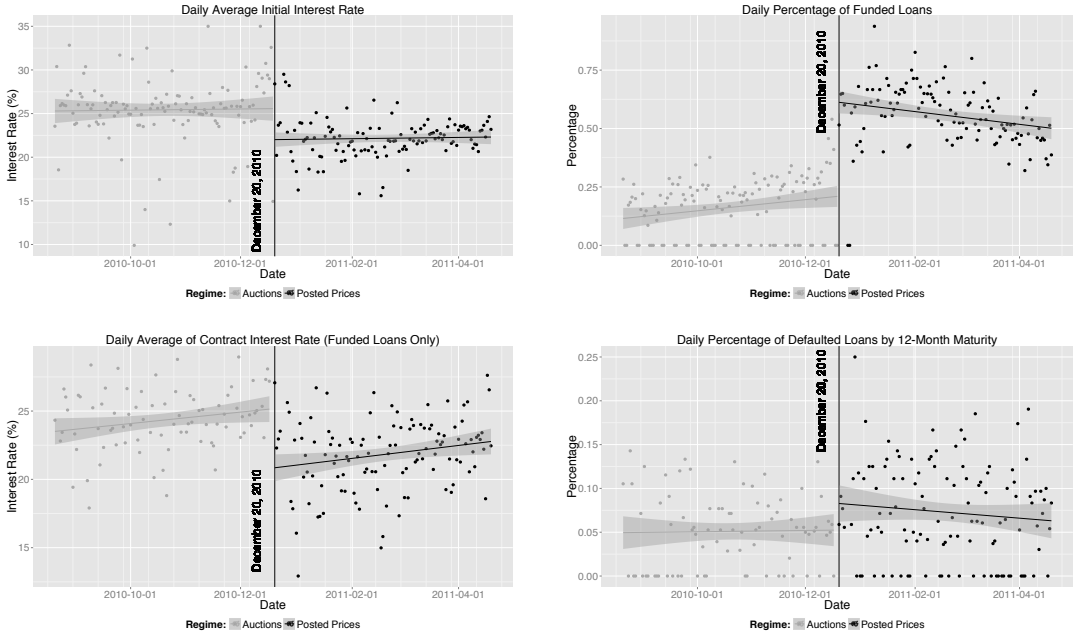


Table 2.2: Summary Statistics: All Listings

	All Listings		Auctions		Posted Prices	
Variable	**Mean**	**sd**	**Mean**	**sd**	**Mean**	**sd**
Listing Characteristics						
# of Bids	70.84	93.88	58.24	94.39	94.32	88.27
1(Loan Funded)	0.34	0.47	0.23	0.42	0.55	0.50
1(Electronic Transfer)	0.99	0.08	0.99	0.10	1	0
1(With Description)	0.10	0.05	0.10	0.04	0.10	0.05
1(Group Member)	0.05	0.21	0.05	0.22	0.04	0.19
1(With Images)	0.16	0.37	0.24	0.43	0.002	0.05
Amount Requested	6,589.45	4,368.21	6,105.62	3,953.51	7,490.71	4,926.06
Estimated Loss (%)	13.66	8.58	15.86	9.29	9.55	4.89
Initial Interest Rate (%)	24.32	9.27	25.58	9.35	21.98	8.64
Listing Effective Days	7.83	6.46	7.05	5.26	9.29	8.05
Loan Term in Months	36.74	6.11	36.20	2.79	37.74	9.53
1(With Friends Bidding)	0.02	0.13	0.02	0.14	0.01	0.12
Length of Description	130.89	80.78	144.91	87.63	104.78	57.71
Credit Profiles						
1(Verified Bank Account)	1	0	1	0	1	0
Debt/Income (DTIR) (%)	21.29	44.72	22.07	46.13	19.82	41.94
1(Top Coded DTIR)	0.001	0.04	0.001	0.04	0.001	0.04
1(Missing DTIR)	0.17	0.38	0.19	0.40	0.13	0.34
1(Homeowner)	0.49	0.50	0.50	0.50	0.47	0.50
Amount Delinquent	1,012.45	6,724.71	996.73	5,719.38	1,041.74	8,278.70
Bankcard Utilization (%)	50.53	33.93	53.12	34.56	45.70	32.15
Current Credit Lines	9.06	5.29	9.08	5.31	9.04	5.26
Current Delinquencies	0.42	1.22	0.46	1.28	0.34	1.10
Delinquencies Last 7 Yrs	3.11	8.25	3.121	8.21	3.10	8.31
Inquiries Last 6 Months	1.35	1.92	1.54	2.12	0.99	1.40
Length Credit History	5,939.49	2,964.23	5,843.95	2,915.02	6,117.46	3,046.12
Open CreditLines	8.00	4.77	8.00	4.78	8.00	4.75
Pub Rec Last 12 Months	0.01	0.14	0.01	0.13	0.02	0.15
Pub Rec Last 10 Years	0.24	0.67	0.24	0.67	0.25	0.66
Revolving Credit Balance	17,200.92	37,234.34	17,794.58	38,372.88	16,095.08	34,992.20
Stated Monthly Income	5,010.90	13,875.29	4,571.00	12,482.46	5,830.32	16,122.23
Total Credit Lines	24.95	13.99	25.07	14.19	24.74	13.59
Total Open Accounts	6.16	4.30	6.10	4.30	6.27	4.31
Macroeconomic Environment						
Unemployment Rate	9.44	1.89	9.62	1.88	9.11	1.86
1(Miss Unemployment)	0.01	0.08	0.01	0.07	0.01	0.09
Zillow Home Value Index	187,424.10	82,477.59	189,662.10	83,724.63	183,255.20	79,945.08
1(Miss Zillow Index)	0.02	0.13	0.02	0.13	0.02	0.14
Observations	13,017		8,470		4,547	

2.5. Empirical Analyses and Results

We now empirically test the hypotheses on transaction outcomes from our theoretic model earlier in the paper. Since funding probability and contract interest rate of funded loans are immediate results of lenders' behaviors, our analyses will also not

Table 2.3: Summary Statistics: Funded Loans

	All Loans		Auctions		Posted Prices	
Variable	**Mean**	**sd**	**Mean**	**sd**	**Mean**	**sd**
Listing Characteristics						
# of Bids	123.20	97.55	140.51	108.45	109.97	86.03
1(Electronic Transfer)	1.00	0.05	1.00	0.07	1.00	0.00
1(With Description)	1.00	0.04	1.00	0.02	1.00	0.05
1(Group Member)	0.06	0.23	0.07	0.26	0.04	0.21
1(Images)	0.12	0.32	0.27	0.44	0.001	0.03
Amount Requested	6108.74	4321.44	5234.43	4057.03	6776.35	4398.17
Contract Interest Rate (%)	22.79	9.08	24.25	9.32	21.68	8.73
Estimated Loss (%)	10.37	6.09	11.61	7.22	9.42	4.85
Initial Interest Rate (%)	23.30	9.24	25.44	9.46	21.68	8.73
Listing Effective Days	7.69	5.93	6.32	3.18	8.74	7.19
Loan Term in Months	36.48	7.03	36.25	3.36	36.66	
1(With Friends Bidding)	0.03	0.17	0.04	0.20	0.02	0.14
Length of Description	127.82	76.75	150.56	90.71	110.46	58.41
Closing Fees	210.82	140.18	180.55	125.32	233.92	146.41
1(Defaulted by 540 Days)	0.09	0.28	0.08	0.26	0.09	0.29
Credit Profiles						
1(Verified Bank Account)	1.00	0.00	1.00	0.00	1.00	0.00
Debt/Income (DTIR) (%)	19.77	33.48	19.68	31.32	19.84	35.05
1(Top Coded DTIR)	6.75e-04	0.03	5.20e-04	0.02	7.93e-04	0.03
1(Missing DTIR)	0.10	0.31	0.11	0.32	0.10	0.30
1(Homeowner)	0.51	0.50	0.52	0.50	0.50	0.50
Amount Delinquent	765.44	6035.19	661.71	4283.06	844.65	7087.08
Bankcard Utilization (%)	50.73	32.59	52.88	33.07	49.09	32.13
Current Credit Lines	9.24	5.18	9.20	5.12	9.27	5.23
Current Delinquencies	0.34	1.13	0.35	1.16	0.33	1.12
Delinquencies Last 7 Years	2.91	7.66	2.67	6.89	3.09	8.20
Inquiries Last 6 Months	0.94	1.38	0.99	1.47	0.91	1.30
Length Credit History	5997.56	2821.95	5945.18	2843.14	6037.56	2805.56
Open Credit Lines	8.15	4.67	8.10	4.64	8.19	4.70
Pub Rec Last 12 Months	0.01	0.120	0.01	0.10	0.02	0.13
Pub Rec Last 10 Years	0.26	0.65	0.26	0.64	0.27	0.67
Revolving Credit Balance	16409.10	34784.77	16615.13	33808.10	16251.77	35518.38
Stated Monthly Income	5533.68	12454.72	5040.51	4108.85	5910.26	16136.82
Total Credit Lines	25.43	13.56	25.21	13.61	25.60	13.53
Total Open Accounts	6.26	4.19	6.21	4.13	6.30	4.23
Macroeconomic Environment						
Unemployment Rate	9.32	1.87	9.59	1.85	9.11	1.86
1(Miss Unemployment)	0.01	0.08	0.01	0.08	0.01	0.09
Zillow Home Value Index	185,355.00	80,456.74	187,407.40	82,348.41	183,787.80	78,962.31
1(Miss Zillow Index)	0.02	0.14	0.02	0.14	0.02	0.14
Observations	4446		1925		2521	

be complete without a finer look into the lenders' investment behaviors, such as the amount of bids, timing of their bids, and if the regime change resulted in any changes in their herding behavior (Zhang and Liu 2012).

2.5.1. Empirical Strategy

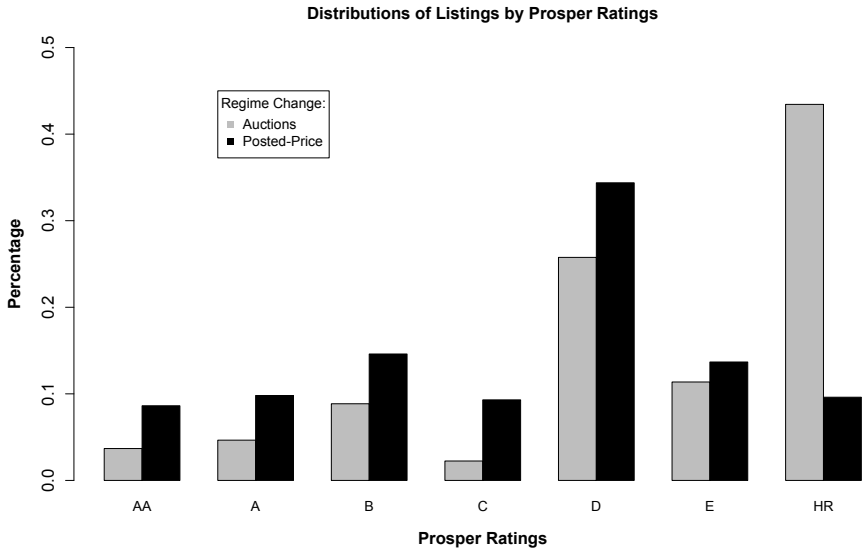
As a short summary, our analytical model predicts that Prosper.com, in the posted-price regime, will assign higher interest rates compared to what the borrower would have chosen as the reserve interest rates in the auction regime *ceteris paribus*. A natural extension of this prediction is that the *contract* interest rates (for funded loans) should be also higher under posted prices. Our model also predicts that the funding probability for loans that are listed under posted prices is higher than in auctions. Further, all else equal, loans funded after the regime change should be more likely to default than those under auctions.

To empirically test these model hypotheses, we first notice that the regime change led to a compositional change in the population of personal loans in terms of borrowers' credit worthiness. Summary statistics in Table 2.2 and 2.3 reveal that after the regime change, the set of borrowers tend to have better credit records. As another piece of evidence showing the compositional change, we depict (in Figure 2.2) distributions of personal loans in terms of borrowers' credit grades. We observe that after the regime change, the fraction of borrowers with better credits is significantly higher than before. We further notice that the drop in the number of posted listings (as in Table 2.2) is mainly due to the decrease in the number of high risk borrowers. Our empirical strategy therefore has to account for such changes by controlling for their credit profile, loan and listing characteristics, and other covariates. In other words, after controlling for all observed characteristics (key variables), the before and after samples should be sufficiently similar to conduct comparative static comparisons.

The key variables we control for in our estimations are summarized in Table 2.2 and 2.3. In addition to these variables, we further control for the creditworthiness categories of borrowers as specified by Prosper.com. The categorical pricing system shown in Figure 2.3[13] illustrates how Prosper.com pre-set the interest rate for

[13]This is a screenshot of Prosper's pricing table on September 26, 2011, taken from Archive.org.

Figure 2.2: Distributions of Personal Loans across Credit Grades before and after the Regime Change



borrowers of different categories, and we control for borrower category in the same manner.

Our main empirical strategy is propensity score matching. Specifically, we consider the posted-price mechanism as the "treatment." Therefore, listings and loans under posted prices are in the "treatment group," whereas those under auctions are in the "control" group. We estimate the treatment effect associated with the sales mechanism switch. We use nearest-neighbor propensity score matching. First, we estimate the following probability of being in the posted-price regime (the treatment) given the set of covariates discussed above, using a logit regression,

$$\rho(\mathbf{x}_i) = \Pr\left(D_i = 1 | \mathbf{X} = \mathbf{x}_i\right), \tag{2.5}$$

where D_i is a dummy variable equal to 1 if the listing is posted after the regime change, and $\mathbf{X}$ includes all the covariates. Then we estimate the average treatment effects (ATE) with bias adjusted and robust standard errors. In other words, in

Prosper.com updates this table from time to time. More information can be found at http://www.prosper.com/loans/rates-and-fees/.

Figure 2.3: A Sample of Prosper.com's Pricing Table after the Regime Change

Prosper Rating	Term (yrs)	# Previous Prosper Loans	Borrower Rates	Borrower APR***
AA	1	1+	5.99%	6.93%
AA	3	1+	7.99%	8.33%
AA	5	1+	10.99%	11.21%
AA	1	0	5.65%	6.59%
AA	3	0	8.99%	9.33%
AA	5	0	11.99%	12.21%
A	1	1+	9.29%	15.08%
A	3	1+	11.29%	13.41%
A	1	0	10.70%	16.51%
A	3	0	13.90%	16.06%
B	1	0	14.99%	20.86%
B	3	0	17.99%	20.20%
C	1	1+	17.49%	26.44%
C	3	1+	18.99%	22.36%
D	1	1+	21.99%	31.04%
D	3	1+	23.99%	27.47%
D	1	0	24.99%	34.11%
D	3	0	26.99%	30.53%
E	1	1+	26.69%	35.84%
E	3	1+	30.99%	34.62%
E	1	0	26.69%	35.84%
E	3	0	31.99%	35.64%
HR	3	1+	31.25%	34.89%
HR	3	0	31.99%	35.64%

matching we first find the nearest "neighbors" in the auction regime (untreated) for each listing under the posted-price mechanism (treated) in terms of propensity score, and vice versa. We then estimate the sample ATE with replacement. We follow Abadie and Imbens (2006) to calculate the standard errors of the matching estimates. All estimations are performed in R (Sekhon 2011).

2.5.2. Results and Discussions

Funding Probability and Interest Rates We report the results of our matching estimations in Table 2.4 through 2.7. Table 2.4 reports the estimates for the comparison of *initial* interest rates. Estimates for the comparison of funding probability are reported in Table 2.5. Table 2.6 presents the results for the comparison of *contract* interest rates for the subsample of funded loans. In addition to the matching method, we also

estimate simple linear regressions using OLS for all outcomes (except contract interest rate where a Heckman model is required since only funded loans report contract interest rates), and Logit regressions for funding probability and default rate. Results (shown in Table 2.10) are qualitatively similar to our matching estimates. In other words, estimates of the coefficients of the key explanatory variable, a dummy variable indicating the regime change, are in the same directions as our matching estimates.

Our empirical results lend supports to our hypotheses about the comparisons of asking interest rates, funding probability, as well as the contract interest rates if funded. The matching estimate for the ATE of the regime change in Table 2.4 shows that in posted-price sales, the initial interest rate is around 2%, or 200 basis points, higher than what the borrower would have set in auctions. This is a nontrivial difference for interest rates (c.f. Saunders, Jr (1993)). Notice that if we did not control for the interest rate categories, the results would have been much smaller and insignificant.

Table 2.4: Matching Estimates for the Comparison of *Initial* Interest Rates

	Spec 1[a]			**Spec 2**		
y = *Initial* interest rate (%)	$M=1$	$M=4$	$M=8$	$M=1$	$M=4$	$M=8$
1(Posted Prices)	-1.119***	-1.036***	-1.067***	1.065***	1.076***	1.056***
	(0.161)	(0.155)	(0.154)	(0.116)	(0.114)	(0.113)
Control covariates:						
Personal loan characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Borrower's credit profile	Yes	Yes	Yes	Yes	Yes	Yes
Macroeconomic environment	Yes	Yes	Yes	Yes	Yes	Yes
Interest rate categories	No	No	No	Yes	Yes	Yes
Num. obs.	13,017	13,017	13,017	13,017	13,017	13,017

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

[a]This table presents the nearest-neighbor propensity score matching estimates of the treatment effect (ATE) of the regime change on the *initial* interest rates. In Specification 1, the matching covariates include all the control covariates except the interest rate categories. M is the number of matches to be searched. Typically the rule of thumb for M is 3 or 4. The standard errors are calculated by the method suggested in Abadie and Imbens (2006). All the estimations are performed in R (Sekhon 2011).

For the funding probability, Table 2.5 shows that compared to using auctions,

the funding probability using the price posting strategy is on average more than 20% higher. The top right panel in Figure 2.1 also displays this apparent trend in funding probability. Notice that there is a kink at the regime change date, and this turns out to be significant even if we control for multiple covariates.[14]

Table 2.5: Matching Estimates for the Comparison of Funding Probability

	Spec 1[a]			**Spec 2**		
$y = 1$(Loan funded)	$M = 1$	$M = 4$	$M = 8$	$M = 1$	$M = 4$	$M = 8$
1(Posted Prices)	0.325***	0.325***	0.326***	0.301***	0.308***	0.306***
	(0.010)	(0.010)	(0.010)	(0.012)	(0.012)	(0.011)
Control covariates:						
Personal loan characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Borrower's credit profile	Yes	Yes	Yes	Yes	Yes	Yes
Macroeconomic environment	Yes	Yes	Yes	Yes	Yes	Yes
Interest rate categories	No	No	No	Yes	Yes	Yes
Num. obs.	13,017	13,017	13,017	13,017	13,017	13,017

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

[a]This table presents the nearest-neighbor propensity score matching estimates of the average treatment effect (ATE) of the regime change on the funding probability. In Specification 1, the matching covariates include all the control covariates except the interest rate categories. M is the number of matches to be searched. Typically the rule of thumb for M is 3 or 4. The standard errors are calculated by the method suggested in Abadie and Imbens (2006). All the estimations are performed in R (Sekhon 2011).

Table 2.6 presents the estimation results for the comparison of contract interest rates. The estimates in the last three columns imply that after the regime change, the *contract* interest rates for funded loans are around 1% higher than the loans funded in the auction regime. These results lend support to our theoretical prediction that the *contract* interest rates should be higher in the posted-price regime. They show that while the loans are more likely to be funded under posted prices, it came at a cost in the form of higher interest rates. All results above are robust to the choice of different matching estimators. Similar results hold when we estimate the average

[14]One may argue that the kink could be driven by seasonality effect since the regime change was made right before the Christmas holiday. We check that possibility by examining the daily ratio of funded loans one year prior to our study period (when there was no regime change), and do not find such an effect of seasonality.

treatment effect on the treated (ATT) for all three outcome variables.

Table 2.6: Matching Estimates for the Comparison of *Contract* Interest Rates

	Spec 1[a]			**Spec 2**		
$y =$ *Contract* interest rate (%)	$M = 1$	$M = 4$	$M = 8$	$M = 1$	$M = 4$	$M = 8$
1(Posted Prices)	0.173	0.078	-0.614	0.738*	0.726**	0.922**
	(1.030)	(0.714)	(0.502)	(0.418)	(0.329)	(0.366)
Control covariates:						
Personal loan characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Borrower's credit profile	Yes	Yes	Yes	Yes	Yes	Yes
Macroeconomic environment	Yes	Yes	Yes	Yes	Yes	Yes
Interest rate categories	No	No	No	Yes	Yes	Yes
Num. obs.	4,446	4,446	4,446	4,446	4,446	4,446

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

[a]This table presents the nearest-neighbor propensity score matching estimates of the average treatment effect (ATE) of the regime change on the *contract* interest rates. In Specification 1, the matching covariates include all the control covariates except the interest rate categories. M is the number of matches to be searched. Typically the rule of thumb for this number is 3 or 4. The standard errors are calculated using the method suggested in Abadie and Imbens (2006). All the estimations are performed in R (Sekhon 2011).

Loan Repayment Prosper.com focuses on faster "fund deployment" as a motivation for the regime change. In the long run, however, it is the repayment of the loans that matters most for investors as uncertainties resolve and returns are made. If lenders' choices turn out to be wrong, that could hurt their incentive to continuously participate in the market. Our Hypothesis 4 is that the default rate will be higher under posted prices, and we test it next.

Since some loans originated in this period have not yet matured as of the time of writing, we record their repayment results as of the 18th month since their origination date. This ensures that we are comparing loans at a similar stage of "maturity." The base rate of default at the 18th month for the whole sample is 9.92%. As a brief contrast in both regimes, we find that 8.62% of loans originated in auction stage default at the 18-month maturity, while this ratio is 10.91% for posted prices. We also investigate whether the default rate varies across credit grading. We find that under

auctions, the default rate is decreasing with borrowers' credit grades and ranges from 1.34% to 15.15%. In contrast, under posted prices, the default rate is also negatively correlated with creditworthiness and ranges from 0.83% to 19.15%.

We then estimate whether the regime change is associated with a higher or lower default rate, and present the estimation results in Table 2.7. Our results are robust to the choice of different repayment dates. Consistent with Hypothesis 4, we find that loans originated after the regime change have *higher* default rates: roughly 4% higher than in auctions (column 5 in Table 2.7). This is an important consequence of the regime change.[15] It suggests that even though lenders see a higher interest rate at the time of loan origination, their overall return may not be strictly higher, due to the increase in *ex post* default probabilities.

Loan default is usually declared after a certain period (more than 4 months) since a borrower stops resuming payments. So far we have examined the loan payment results up to different maturity dates. In addition to comparing the repayment results at certain points in time, it is helpful to explore whether the regime change leads to deteriorations or improvements over the whole payment process. We conduct survival analysis exploring the regime change effects on the longevity of the loan payment process. We hypothetically treat a loan being "dead" if it is defaulted, and "survived" if paid in full. Analogously, we define an observation being "censored" if the loan is still in payment progress up to a certain date. We study the effects up to various maturity dates as before, and results are similar.

Results of the survival analysis are reported in Figure 2.4 and Table 2.8. The figure

[15]One may have a concern that after the regime change the lenders race to secure a fraction of the loan before full funding. This competition effect may encourage risk taking behavior among lenders, which causes the finding of higher default rate. However, we first observe that the median credit grade of lenders' investment was actually improved after the regime change, meaning more funds were invested in less risky loans. In addition, given all observed loan characteristics and borrower's credit profile, the risk taking behavior implicitly suggests that lenders are seeking opportunities with higher expected returns, most likely from unobserved sources. However, they lack of such incentives under posted prices because the interest rates are fixed given the credit profile. Therefore, our empirical specifications accounting for all observables clearly rule out this possibility.

Table 2.7: Matching Estimates for the Comparison of Default Rate

	Spec 1[a]			**Spec 2**		
$y = 1$(Loan defaulted by 12 months)[b]	$M = 1$	$M = 4$	$M = 8$	$M = 1$	$M = 4$	$M = 8$
1(Posted Prices)	0.019*	0.018*	0.018*	0.025**	0.025**	0.024**
	(0.011)	(0.010)	(0.010)	(0.011)	(0.010)	(0.010)
Control covariates:						
Personal loan characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Borrower's credit profile	Yes	Yes	Yes	Yes	Yes	Yes
Macroeconomic environment	Yes	Yes	Yes	Yes	Yes	Yes
Interest rate categories	No	No	No	Yes	Yes	Yes
Num. obs.	4,443[c]	4,443	4,443	4,443	4,443	4,443

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

[a]This table presents the nearest-neighbor propensity score matching estimates of the average treatment effect (ATE) of the regime change on the default rate. In Specification 1, the matching covariates include all the control covariates except the interest rate categories. M is the number of matches to be searched. Typically the rule of thumb for this number is 3 or 4. The standard errors are calculated by the method suggested in Abadie and Imbens (2006). All the estimations are performed in R (Sekhon 2011).

[b]We examine the loan payment results by the 18th cycle (or month) since origination dates. Similar results appear as of 6-cycle and 12-cycle maturity.

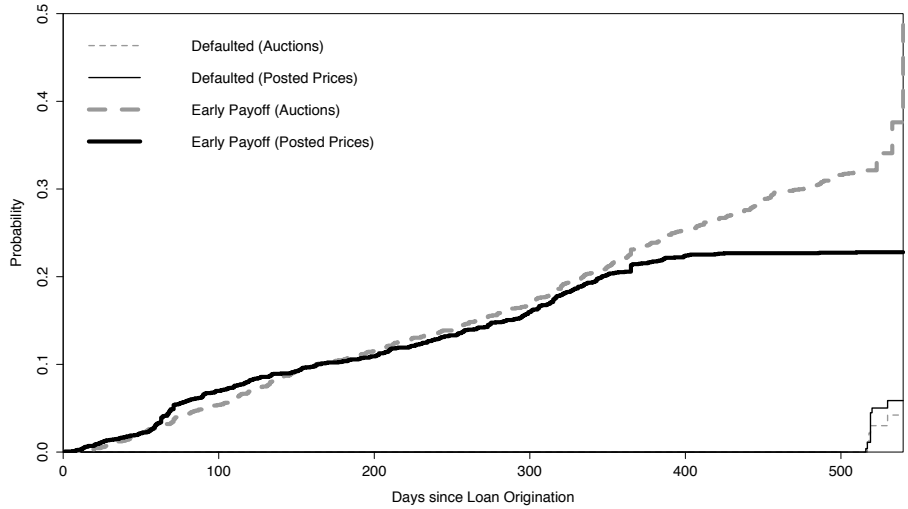
[c]Three observations are dropped because of incomplete records.

depicts the nonparametric likelihood of a loan being defaulted or early paid. We find that, for loans under the posted-price regime, the likelihood of being defaulted is higher than in auctions. In sharp contrast, the probability of early payoff for posted-price loans is significantly lower after the certain period of time (around 360 days). Estimates in Table 2.8 are the exponentiated estimates from a Cox proportional hazard regression. Results from the main specification (Spec 5 in the table) suggest that after the regime change, the hazard rate of being defaulted[16] increases by a factor of 1.504, or equivalently, by 50.4%. The findings from the survival analysis are consistent with our findings on comparisons of payment results. As a short summary,

[16]In the main specification, we define a loan being defaulted if the status shows "4+ months late", "defaulted", or "charge-off." To better understand whether the regime change has significant effects on the overall payment process, we extend the definition of default to being late and beyond, 1 month late and beyond, 2 month late and beyond, or 3 month late and beyond. Cox estimates from all these extensions are qualitatively similar, that the regime change indeed leads to higher hazard rate of default.

our findings of loans' payment results indicate that the regime change not only leads to higher default rate, but also increases the likelihood of being unpaid over the whole payment process.

Figure 2.4: Cumulative Incidence Functions from Payment Results (by 540 Days) of 36-Month Loans



Lenders' Bidding Behaviors We study the change in lender's bidding behaviors from two aspects. The first one uses each bid as the unit of analysis to compare the timing of lenders' bids as well as the amount of their bids. The second one draws on empirical studies of herding in the context of Prosper.com (Zhang and Liu 2012), and tests whether the regime change affects lender's tendency to follow each other.

Figure 2.5 presents the comparison of the amount of dollars submitted in each bid. It shows that in the posted-price regime, the lenders tend to invest more in each of their bids compared to that in the auction regime. Another major reason for Prosper.com to switch to the posted-price mechanism is that the personal loans should be funded quicker. We find that half of the funded loans from the posted-price selling receive full funding within 80 hours, compared to more than 160 hours in the

Table 2.8: Cox Estimates of a Competing Risks Model for the Duration of the Loan Payment Process

	Cox proportional hazard estimates[a]				
	Spec 1	Spec 2	Spec 3	Spec 4	Spec 5
exp (**Coef. of 1(Posted Prices)**)	1.250	1.371	1.396	1.417	1.504
	(0.102)**[b]	(0.108)***	(0.110)***	(0.110)***	(0.111)***
Control covariates:					
Personal loan characteristics	No	Yes	Yes	Yes	Yes
Borrower's credit profile	No	No	Yes	Yes	Yes
Macroeconomic environment	No	No	No	Yes	Yes
Interest rate categories	No	No	No	No	Yes
Num. obs.	4,446	4,446	4,446	4,446	4,446

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

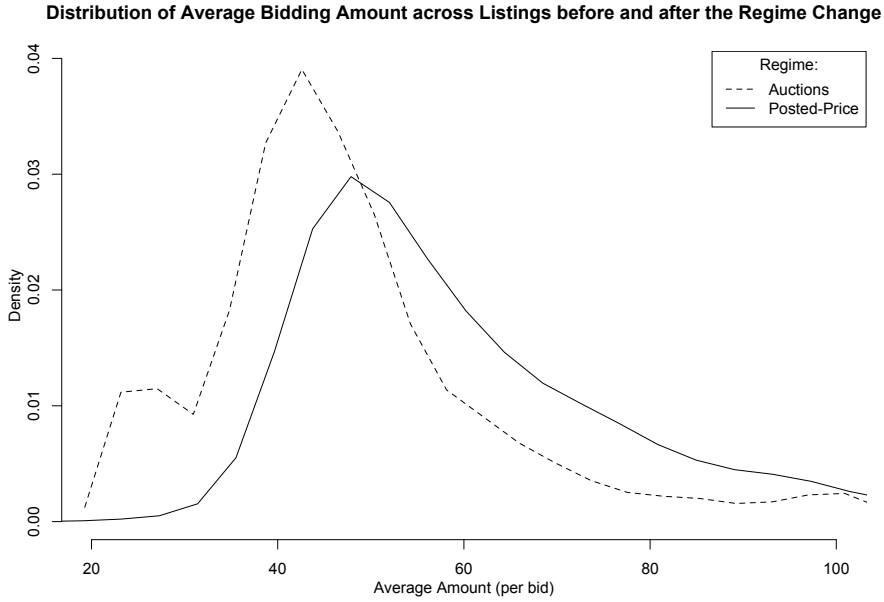
[a]This table presents Cox proportional hazard estimates of a competing risks model for the duration of the funded loans' payment on Prosper.com. Specifically, we are modeling the length of time until a loan being defaulted ("dead"), and define a loan being defaulted if the status shows either over 4 months late, defaulted, or charge-off. We treat early payoff as a competing "risk" of terminating the loan payment process. We artificially compare loans' payment results by 540 days since their origination dates, thus loans still in payoff progress are right-censored. Estimations and Figure 2.4 are performed in R (Scrucca et al. 2007).

[b]The standard errors for the coefficient estimates for the regime dummy 1(Posted Prices).

auction regime. In summary, after the regime change, the lenders indeed invest more and quicker on average, so that the funding procedure is indeed more efficient.

In addition to the timing and amount of bids, another important characterization of lender behavior is herding. We hypothesized that since lenders are more likely to trust the price assigned by the platform than the initial asking rate of borrowers, lenders will have lower needs to rely on observing the behaviors of others, and therefore herding should be less likely to occur under posted prices. To test this, we draw on prior empirical work by Zhang and Liu (2012) who also used data from Prosper.com. They find that lenders exhibit rational herding, in the sense that they gravitate toward listings with more funds received even when those listings may have seemingly "bad" credit. We replicate their empirical strategy to examine the effect of the regime change on this rational herding behavior. Specifically, we estimate the model proposed

Figure 2.5: Distribution of Average Amount across Listings before and after the Regime Change



by Zhang and Liu separately using the listings in the auction regime, and in the posted-price regime. Table 2.9 reports the estimation results. The estimate shows an interesting *reversal* in lenders' herding behavior. In auctions, a listing with USD 100 more funding at the start of a day will receive on average USD 2.7 more funds during the day; while this number under posted prices is negative (-USD 17.2). A full investigation of the herding reversal will be beyond the scope of this paper; however, this result does lend support to our hypotheses.

2.5.3. Robustness

We now turn to robustness checks for our main empirical analyses on the comparisons of transaction outcomes. We consider alternative specifications other than the main matching model, explore a series of confounding factors that may potentially contaminate our main findings, compare the outcomes at different periods of time,

Table 2.9: Comparison of Herding Behavior

	Within Estimates[a]	
Dep. var.: Daily fund received	**Auctions**	**Posted prices**
Lag Cum Amount	0.027*	-0.172***
	(0.014)	(0.018)
Lag Percent Needed	-0.005***	-0.003***
	(0.000)	(0.001)
Lag Min Rate	-0.098***	
	(0.022)	
Lag Bids	-0.001	0.015***
	(0.001)	(0.001)
Lag Cum Amount * Lag Percent Needed	0.001***	0.008***
	(0.000)	(0.000)
Listing FE	Yes	Yes
Day of Listing FE	Yes	Yes
Weekday FE	Yes	Yes
Adj. R^2	0.090	0.252
Num. obs.	24,773	21,366

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

[a]This table presents the within estimates of a panel model that explores the correlation between the funds received in a day, and the total funds received by the end of previous day. The complete specifications can be found in Zhang and Liu (2012). Note that the listings sampled in this table include only those with at least one bid.

and conduct sensitivity analyses for our propensity score matching method.

Alternative specifications: One potential concern for our empirical approach is that the matching model imposes special functional form restrictions on the underlying data generation process. As an alternative, we re-estimate the regime change effects on interest rates, funding probability, and default rates using simple linear models. The specific equation we estimate is,

$$y_{ir} = \nu_r + \beta_1 1(\text{Posted Prices})_{ir} + \beta_2' \mathbf{X}_{ir} + \epsilon_{ir}, \tag{2.6}$$

where y_{ir} are the transaction outcomes of asking rates, the indicator of being funded, the contract interest rates if funded, or the indicator of default for funded loans; ν_r are the interest rate category fixed effects as discussed in Figure 2.3; $\mathbf{X}_{ir}$ are the control variables in Equation 2.5. The coefficient of interest is β_1 of which the estimates are

the regime change effects. One caveat is that the identification of these linear models also relies on the assumption of selection on observables.

We report estimation results using our main sample in columns 1, 2, 4, and 6 of Table 2.10. The estimates are qualitatively consistent with our findings in matching except the result for contract interest rates. However, the linear model for interest rate may have been biased due to a selection process. In auctions, given an interest rate category, loans with higher interest rates are more likely to be funded. In addition, the variation of interest rates within each category is larger under auctions. To control for this issue, we estimate a Heckman selection model using two-step method. Results are reported in column 5 of Table 2.10. The estimate of the regime change effect, β_1, is again positive and consistent with the matching estimate. For comparisons of funding probability and default rate, we also estimate Logit models in view of the dichotomous dependent variables. Results are reported in columns 3 and 7 in Table 2.10. Estimates of the regime effects are in the same direction as our matching estimates.

In our main specifications, we include all observed covariates including a series of loan characteristics. However, one might have a concern that the funding success and transaction interest rates also depend on information available from other channels based on the purpose of the loan. For instance, holding everything else constant, a loan used for an urgent medical treatment is arguably different from that for the house decoration. We explore this information in the textual description and title of a personal loan. We consider a robustness check examining whether our results are consistent with inclusions of dummies for a few text keywords, such as "medical", "doctor", or "real estate." Results are reported in Table 2.11. The matching estimates are qualitatively the same as our findings from the main specifications.

Lenders: Another concern on our main specification is that lenders, particularly the composition of lenders (in terms of investor sophistication, as an example), have non-ignorant impacts on the funding outcomes since they are themselves a generic

Table 2.10: Comparisons of Transaction Outcomes using Linear Regressions

	Transaction Outcomes[a]						
	***Initial* IR**	**Funding probability**		***Contract* IR**		**Default rate**[b]	
	OLS	OLS	Logit	OLS	Heckman[c]	OLS	logit
1(Posted Prices)	1.181*	0.255***	0.210***	-0.414*	0.701**	0.012	0.010
	(0.592)	(0.014)	(0.015)	(0.237)	(0.341)	(0.017)	(0.018)
Control covariates:							
Loan characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Borrower's credit profile	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Macro. environment	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Interest rate categories	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.735	0.242	-	0.953	-	0.107	-
Pseudo R^2	-	-	0.211	-	-	-	0.146
Num. obs.	13,017	13,017	13,017	4,446	13,017	4,443	4,443

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

[a]This table displays estimation results of linear models for all transaction outcomes. For comparisons of funding probability and default rate at 18-cycle maturity, we also estimate Logit models. Similar results show up for probit models. Robust standard errors (in all models) are present in the table.

[b]We examine the loan repayment results at their 18-cycle maturity. Similar results show up for other selections of maturity such as 6-cycle and 12-cycle.

[c]For the comparison of contract rates, there is a selection issue that in auctions the loans with higher reserve rates are more likely to be funded, even if we control for all observed characteristics. Thus we also estimate a Heckman selection model using two-stage least squares.

part. However, the lenders composition is more appropriately treated as outcomes rather than control variables. Lenders arrive at listings after the market mechanism is public knowledge, and after the borrower's listing (including borrower information and loan characteristics) is revealed to all potential investors. To include lender composition into the right-hand-side of our models will introduce endogeneity and bias our results. As a robustness test, we study whether considering the lender composition will alternate our findings. Specifically, we consider the composition by the proportion of participating lenders who had invested in at least 1 listing before the regime change. There are 556 posted-price listings (or 74 funded loans) in which all lenders had participated before the event. We repeat our matching estimations by excluding this subset of loans. Results are presented in Table 2.12, and shown to be

Table 2.11: Estimates of the Regime Change Effect with Additional Text Information

	Matching estimates[a]			
Transaction Outcomes:	Initial Interest Rates	Funding Probability	Contract Interest Rates	Default Rate
1(Posted Prices)	0.880***	0.311***	0.681**	0.022**
	(0.133)	(0.012)	(0.297)	(0.010)
Control covariates:				
Personal loan characteristics	Yes	Yes	Yes	Yes
Borrower's credit profile	Yes	Yes	Yes	Yes
Macroeconomic environment	Yes	Yes	Yes	Yes
Text information	Yes	Yes	Yes	Yes
Interest rate categories	Yes	Yes	Yes	Yes
Num. obs.	13,017	13,017	4,446	4,443

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

[a]This table presents matching estimates of the regime change effects on all transaction outcomes with additional matching covariates. Specifically, we include texts information in listings' titles and descriptions. We create dummies for each of the following words appearing in listing titles: "help, "credit card", "debt", "consolidate", "real estate", "student", "school", "tuition", "medical", "doctor". For all four estimations, we let the number of matches to be 4.

quite consistent with our main findings.

Table 2.12: Estimates of the Regime Change Effect Considering Lender Composition

	Matching estimates[a]			
Transaction Outcomes:	Initial Interest Rates	Funding Probability	Contract Interest Rates	Default Rate
1(Posted Prices)	1.002***	0.367***	0.746	0.025**
	(0.120)	(0.012)	(0.595)	(0.010)
Control covariates:				
Personal loan characteristics	Yes	Yes	Yes	Yes
Borrower's credit profile	Yes	Yes	Yes	Yes
Macroeconomic environment	Yes	Yes	Yes	Yes
Interest rate categories	Yes	Yes	Yes	Yes
Num. obs.	12,461	12,461	4,367	4,364

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

[a]This table presents matching estimates of the regime change effects on all transaction outcomes considering lender compositions. After the regime change, there were listings in which only lenders that invested before (the regime change) participated. Here we exclude this subset of listings. For all four estimations, we let the number of matches to be 4.

We are also interested in to what extent our main findings about transaction outcomes are driven by the behavioral change of lenders, especially their "activeness." We find no evidence that lender participations or investment behaviors changed dramatically after the regime change. We first compare the daily number of newly registered lenders before and after, and a two-sample T-test rejects the null that the means of these two groups are different at any common significance levels. Figure 2.6 shows that the platform had a steady but relatively low level of growth on the supply side (of funds), with on average 18 new lenders joining per day. More important, the figure suggests that the regime change does not have significant effects on the arrival of new lenders. We also compare the daily number of active lenders that actually participated. Instead of dramatically growing, we find that the active lenders dropped under the posted-price regime, which is statistically sound. However, the decrease in active lenders may also be driven by the drop of listings posted on the platform. We thus compare the ratio of active lenders against the listings initiated daily. Figure 2.7 shows that there is no apparent structural change around the regime change. A two-sample one-sided T-test rejects the null that the mean ratio is bigger after the regime change (with p-value $= 1.412e-04$). These findings suggest that the regime change did not cause significant change in participating lenders or their investment behaviors on site.

Funding duration: Along with the regime change, Prosper.com also revised the funding durations for all its listings after December 20, 2010. The lengths raised from 7 days under auctions to 14 days for posted-price loans. There exists a concern that the higher funding probability, an important finding from the model and our empirical analyses, is driven by the longer duration instead of the mechanism change. We first examine whether the extended duration was indeed a "binding constraint" on the funding process, or the extent to which the longer duration actually helped more loans receive full funding. We found that after the regime change, 68% of loans were funded within 7 days even though the allowable duration is 14 days. Further,

Figure 2.6: Daily Number of Newly Registered Lenders before and after the Regime Change

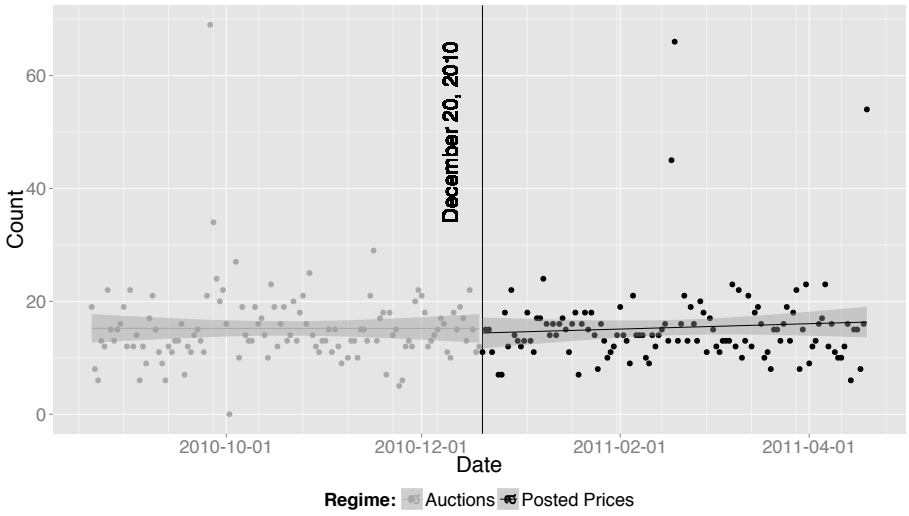
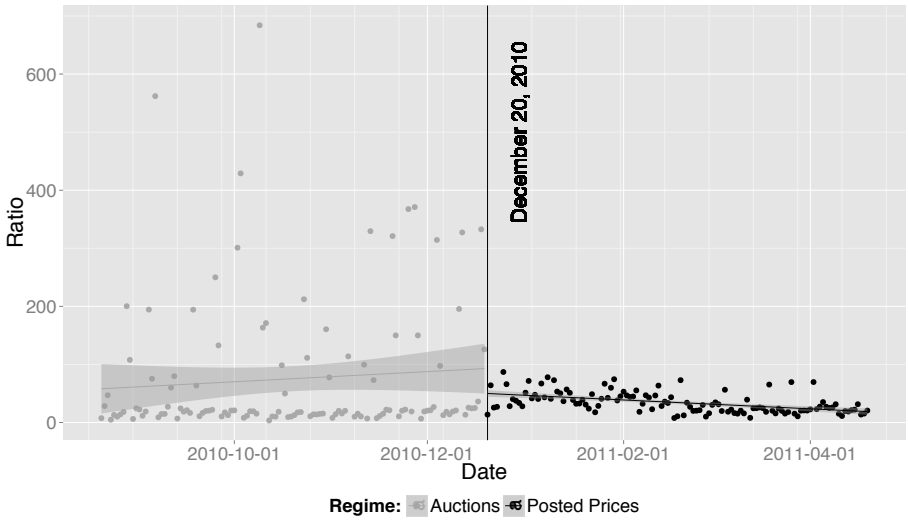


Figure 2.7: Daily Ratio of Lenders/Listings before and after the Regime Change



as we show in Section 2.5.2, more than half of posted-price loans were funded within 80 hours, which is significantly less than 7 days.

To better take into account that the total duration can be (but not necessarily) different under the two transaction mechanisms, we model the funding process using

a duration analysis. Specifically, we estimate a Cox competing risks model where t_0 is the start of a listing, and the dependent variable is the length of time to full funding. We hypothetically let the event of being funded as if "dead," and the failure to be funded as "survived" and right censored. Other events such as cancellation or withdrawn by the borrower are treated as competing risks. Table 2.13 reports the exponentiated estimates from the Cox regressions. The main specification, Spec 5 in the table, suggests that after the regime change, the hazard rate of being funded increases by a factor of 2.422, or, by 142.2%. This significant change in funding rate is consistent with our main findings of higher funding probability, and further implies that the posted-price loans were indeed quicker to fund.

Table 2.13: Cox Estimates of a Competing Risks Model for the Duration of Prosper Loans' Funding Processes

	Cox proportional hazard estimates[a]				
	Spec 1	Spec 2	Spec 3	Spec 4	Spec 5
exp (**Coef. of 1(Posted Prices)**)	2.380 (0.030)***[b]	2.796 (0.035)***	2.627 (0.036)***	2.632 (0.036)***	2.422 (0.037)***
Control covariates:					
Personal loan characteristics	No	Yes	Yes	Yes	Yes
Borrower's credit profile	No	No	Yes	Yes	Yes
Macroeconomic environment	No	No	No	Yes	Yes
Interest rate categories	No	No	No	No	Yes
Num. obs.	13,017	13,017	13,017	13,017	13,017

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

[a]This table presents Cox proportional hazard estimates of a competing risks model for the duration of the listing funding process. Specifically, we are modeling the length of time until a loan being funded ("dead"), and we treat the loans cancelled or withdrawn by the borrowers as competing "risks." The failed loans are regarded as right censored. Estimations are performed in R (Scrucca et al. 2007).

[b]The standard errors for the coefficient estimates of the regime dummy 1(Posted Prices).

Prosper.com implemented a policy change on listing durations from April 15, 2008, which helps us better understand how the durations affect transaction outcomes. Prior to this date, all borrowers have the option to choose among 3, 5, 7, or 10 days as the funding duration. Starting from April 15, 2008, the platform standardized

the expiring days to 7. This provides us an opportunity to evaluate the effects of funding durations on the probability of being funded. We carry out separate analyses by comparing loans with 3- or 5-day durations before the policy change, with those posted after the policy change. We also compare the loans with 10-day duration with those after the policy change. For both sets of empirical investigations, we find insignificant results on the funding probability. These findings support our conjecture that the funding duration does not have significant effects on funding outcomes.

"Interim" loans: We study an over-night change of regimes in this study. All new loans started using the posted-price mechanism from the midnight of December 20, 2010. Nevertheless, there existed loans that were posted before, but stayed unfunded and unexpired until, the regime change. This subsample contains 144 listings, and 40 of them were successfully funded. The interest rates of these loans were frozen at the time of the change, and the funding processes continued under the prevailing rates using the posted-price mechanism. Namely, after the regime change, these loans would be originated immediately upon all funds received. In all previous empirical investigations, we include these loans as if they are under the auction regime. One may worry that our results are attenuated by including this subset of loans. We conduct another robustness check by excluding these "interim" loans. Table 2.14 reports the estimates from our matching method. Results are highly consistent with the findings from our main specifications.

Loan terms: The personal loans in our sample have different maturity lengths including 12, 36, and 60 months. In our main empirical specifications we include a variable of the terms in the matching covariates. However, there is still a question whether the regime change effects are distinct across these subgroups. Lumping all loans together might disguise the potential heterogeneity. Therefore, we consider an additional robustness check of subsample analyses. In our sample, $12,161$ out of $13,017$ Prosper listings assigned 36-month maturity. The rest of loans with different maturity lengths consist of less than 1% of the sample. We repeat all matching

Table 2.14: Estimates of the Regime Change Effect Excluding "Interim" Loans

	Matching estimates[a]			
Transaction Outcomes:	Initial Interest Rates	Funding Probability	Contract Interest Rates	Default Rate
1(Posted Prices)	1.107***	0.308***	0.701	0.026***
	(0.115)	(0.011)	(0.655)	(0.010)
Control covariates:				
Personal loan characteristics	Yes	Yes	Yes	Yes
Borrower's credit profile	Yes	Yes	Yes	Yes
Macroeconomic environment	Yes	Yes	Yes	Yes
Interest rate categories	Yes	Yes	Yes	Yes
Num. obs.	12,873	12,873	4,406	4,403

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

[a]This table presents matching estimates of the regime change effects on all transaction outcomes, using an alternative sample of loans. Specifically, we exclude "interim" loans that were still in fundraising at the time of the regime change. For these loans, the interest rates were frozen after the regime change. For all four estimations, we let the number of matches to be 4.

estimations for the loans with 36-month maturity, and report the results in Table 2.15. We find qualitatively consistent results with those from our main specifications.

Table 2.15: Matching Estimates of the Regime Change Effect using 36-Month Loans Only

	Matching estimates[a]			
Transaction Outcomes:	Initial Interest Rates	Funding Probability	Contract Interest Rates	Default Rate
1(Posted Prices)	1.080***	0.321***	0.761	0.026**
	(0.113)	(0.012)	(0.715)	(0.011)
Control covariates:				
Personal loan characteristics	Yes	Yes	Yes	Yes
Borrower's credit profile	Yes	Yes	Yes	Yes
Macroeconomic environment	Yes	Yes	Yes	Yes
Interest rate categories	Yes	Yes	Yes	Yes
Num. obs.	12,161	12,161	4,063	4,060

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

[a]This table presents matching estimates of the regime change effects on all transaction outcomes, using an alternative sample of loans. Specifically, we focus on loans with 36-month maturity. For all four estimations, we let the number of matches to be 4.

Extended sample: Another potential concern is the choice of sample used in our estimations. Our main sample contains all listings that were posted 4 months immediately before and after the regime change. One may argue that the regime change effect may decline or even disappear eventually. We construct another sample spanning different periods of time. Specifically, we compare loans that were generated 4 months before the regime change, and those posted between April 20, 2011 and August 19, 2011. Matching estimates are reported in Table 2.16. We do not find qualitatively different results from those using our main sample.

Table 2.16: Matching Estimates of the Regime Change Effects using an Extended Sample

	Matching estimates[a]			
Transaction Outcomes:	Initial Interest Rates	Funding Probability	Contract Interest Rates	Default Rate
1(Posted Prices)	3.007***	0.207***	2.511***	0.042***
	(0.270)	(0.045)	(0.612)	(0.007)
Control covariates:				
Personal loan characteristics	Yes	Yes	Yes	Yes
Borrower's credit profile	Yes	Yes	Yes	Yes
Macroeconomic environment	Yes	Yes	Yes	Yes
Interest rate categories	Yes	Yes	Yes	Yes
Num. obs.	16,459	16,459	5,438	5,438

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

[a]This table presents matching estimates of the regime change effects on all transaction outcomes, using an alternative sample of loans. Specifically, we compare loans that were generated 4 months before the regime change, and those posted between April 20, 2011 and August 19, 2011. For all four estimations, we set the number of matches to be 4.

Sensitivity analysis: Finally, the propensity score matching method that we used relies on the assumption of selection on observables. In order to investigate how sensitive our matching estimates are to the possible presence of an unobserved confounding variable, we conduct sensitivity analysis (Rosenbaum 2002). We find that in order for the regime change effect to disappear, the unobserved confounder has to change the odds of selection into the treatment by at least 30% for all outcomes,

lending further support to the propensity score matching method.

2.5.4. Discussions and Welfare Implications

We continue our discussions with implications from our main findings. We first consider the welfare implications on various stakeholders of the market. We then explore "moderating" type of factors that could affect the main findings, and discuss how the embedded "soft information" (Lin et al. 2013) affects the results.

Welfare implications: As a quick summary, our results so far show that funding probability, interest rates, as well as default probability are all higher under posted prices. An important question is whether the regime change is "better" for various stakeholders in this market. From the platform's point of view, its short-term goal is profit improvement. Prosper.com's profits are mainly from the servicing fees charged to the funded loans, according to the corporate annual reports. Our findings show that the regime change leads to significantly higher funding probability. Therefore, the switch in transaction mechanisms should lead to higher short-term profits for the platform. We empirically test this conjecture by calculating the closing fees for loans that were successfully funded. We carry out matching estimations using our main specifications. Table 2.17 reports the estimation results. The estimates show that the regime change indeed causes higher platform profits from loan servicing fees.

Although in the short run the profit is increasing, the platform's long-term profitability is ambiguous. Particularly, as the capital suppliers in the market, lenders' participation is critical for the success of the market. We study how the change in mechanisms affects the investment performance of lenders, which determines their participation and activeness. Higher interest rates and higher likelihood of default are conflicting factors influencing lenders' gains from investment. We calculate the lenders' return on investment (ROI) using the loan performance data. Given a particular loan up to a certain payment cycle, the lenders' ROI is calculated as

Table 2.17: Matching Estimates for the Comparison of Loan Closing Fees

	Spec 1[a]			**Spec 2**		
y = Closing fees (\$)	$M = 1$	$M = 4$	$M = 8$	$M = 1$	$M = 4$	$M = 8$
1(Posted Prices)	22.218***	20.967***	20.450***	23.682***	24.293***	24.031***
	(4.225)	(3.988)	(3.932)	(4.250)	(4.046)	(4.002)
Control covariates:						
Personal loan characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Borrower's credit profile	Yes	Yes	Yes	Yes	Yes	Yes
Macroeconomic environment	Yes	Yes	Yes	Yes	Yes	Yes
Interest rate categories	No	No	No	Yes	Yes	Yes
Num. obs.	4,446	4,446	4,446	4,446	4,446	4,446

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

[a]This table presents the nearest-neighbor propensity score matching estimates of the average treatment effect (ATE) of the regime change on Prosper.com's profits that are from closing fees. In Specification 1, the matching covariates include all the control covariates except the interest rate categories. M is the number of matches to be searched. Typically the rule of thumb for this number is 3 or 4. The standard errors reported in this table are calculated by the method suggested in Abadie and Imbens (2006). All the estimations are performed in R (Sekhon 2011).

(Cumulative payment − Loan amount requested) /Loan amount requested, factoring in the loan's repayment status. The calculation allows us to compare lenders' ROI up to a certain cycle during the loan's payment process. We repeat our matching estimations and report the results in Table 2.18. We find that after the regime change, the lenders' returns are significantly lower by about 3% by the 18-th cycle. The finding suggests that although the higher interest rates are associated with higher payment, the higher probability of default dominates the effects on lenders' ROI and deteriorates their investment performances.

More important, the lender surplus from participating in this market determines the long-term profitability of the platform. From a social planner's point of view, the lender surplus is a generic part of, and has significant influence on, the overall social welfare. The regime change effect on the lender surplus is also ambiguous *ex ante* due to the similar conflict between interest rates and default rate. We empirically study this question by quantifying the lender surplus using a simple method. We first seek to recover a representative lender's supply function. Such a function specifies a

Table 2.18: Matching Estimates for the Comparison of Lender Return on Investment

	Spec 1[a]			**Spec 2**		
y = ROI (%) by 360 days[b]	$M = 1$	$M = 4$	$M = 8$	$M = 1$	$M = 4$	$M = 8$
1(Posted Prices)	-0.035***	-0.031***	-0.030***	-0.030***	-0.029***	-0.027***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Control covariates:						
Personal loan characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Borrower's credit profile	Yes	Yes	Yes	Yes	Yes	Yes
Macroeconomic environment	Yes	Yes	Yes	Yes	Yes	Yes
Interest rate categories	No	No	No	Yes	Yes	Yes
Num. obs.	4,443[c]	4,443	4,443	4,443	4,443	4,443

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

[a]This table presents the nearest-neighbor propensity score matching estimates of the average treatment effect (ATE) of the regime change on Lender ROI. In Specification 1, the matching covariates include all the control covariates except the interest rate categories. M is the number of matches to be searched. Typically the rule of thumb for this number is 3 or 4. The standard errors are calculated by the method suggested in Abadie and Imbens (2006). All the estimations are performed in R (Sekhon 2011).

[b]We examine the ROI by the 18th cycle (or month) since origination dates. Similar results appear as of 6-cycle and 12-cycle maturity.

[c]Three observations are dropped because of incomplete records.

relationship between the lender's willingness-to-lend (measured by the lowest acceptable interest rate) and a personal loan's amount requested. We use $S(\cdot)$ to denote the supply function with $W = S(Q)$. We assume a simple functional form[17] of the supply function as $W = \alpha \cdot Q^{\gamma}$. We take a log-transformation and get,

$$\log W = \log \alpha + \gamma \cdot \log Q. \tag{2.7}$$

With the supply function, the lender's surplus will be simply the area above this supply function but below the market transaction interest rate, factoring in the probability of default.

The sample of bids, which we use to estimate the supply function, are those submitted to the listings posted one year before our original sample (Section 2.4). Specifically, those listings were posted between August 20, 2009 and August 19, 2010.

[17]As robustness checks, we allow for flexible functional forms including linear relations or with higher-order terms of the requested amount. Similar results appear in these cases.

We focus on the open listings where the funding process continues after full funding, since the immediate funding listings mimic posted prices with borrowers the pricing agent. In open auctions, only bids that were outbid were made public. Our estimations are therefore based on these failed bids. The identification of the supply function depends on two assumptions: 1) lenders are truthfully bidding by submitting their lowest acceptable interest rates; and 2) a representative lender's supply function stays the same before and after the regime change. We also assume that the representative lender has different supply functions for loans with different credit grades. Therefore, we estimate the supply function for loans with each of the credit grades, ranging from AA (the best) to HR (the most risky). To estimate the expected probability of default, we run Logit regressions with the dependent variable the dummy for loan default. We then use the predicted value from the Logit regressions as the proxy for the default rate. It is straightforward to calculate the lender surplus for each funded loan after estimating the supply functions. We find that the average lender surplus for loans under auctions is $35.95, while the average surplus for posted-price loans is -$29.49. We repeat matching estimations with the dependent variable the calculated lender surplus for a loan. Results are reported in Table 2.19. Not surprisingly, we find that the regime change leads to lower lender surplus. Consistent with our findings on the lender ROI, these results also imply that the bigger loss from loan default outweighs the gains from higher interest payment for the loans under posted prices. The new regime leads to short-term deterioration for lenders.

From a policy point of view, a critical question is how this change affects overall social welfare. That is, how does the decrease in lender surplus compare to the increase in platform surplus and the decrease (if any) in borrower surplus, and can we say with certainty that this market is better or worse off under the new, posted-price mechanism? Although the posted-price mechanism is associated with higher funding probability, the higher contract interest rates may lead to moral hazards (Stiglitz and Weiss 1981) that increase ex post default rates. The total social surplus

Table 2.19: Matching Estimates for the Comparison of Lender Surplus

	Spec 1[a]			**Spec 2**		
y = Lender surplus	$M = 1$	$M = 4$	$M = 8$	$M = 1$	$M = 4$	$M = 8$
1(Posted Prices)	-44.466***	-46.637***	-50.357***	-42.521***	-43.676***	-45.355***
	(7.218)	(5.721)	(5.074)	(6.358)	(5.193)	(4.813)
Control covariates:						
Loan characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Borrower's credit profile	Yes	Yes	Yes	Yes	Yes	Yes
Macro. environment	Yes	Yes	Yes	Yes	Yes	Yes
Interest rate categories	No	No	No	Yes	Yes	Yes
Num. obs.	4,446	4,446	4,446	4,446	4,446	4,446

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

[a]This table presents the nearest-neighbor propensity score matching estimates of the average treatment effect of the regime change on lender surplus. In Specification 1, the matching covariates include all the control covariates except the interest rate categories. M is the number of matches. Typically the rule of thumb for this number is 3 or 4. The standard errors are calculated by the method suggested in Abadie and Imbens (2006). All the estimations are performed in R (Sekhon 2011).

depends not only on how often the loans are being funded, but also on how often the funded loans are repaid. The welfare comparison is therefore ambiguous. We present a formal development of this observation in Appendix A.2, and we can see that, in fact, under certain conditions, auctions may outperform posted prices in terms of social welfare. In a sense, this result is also reasonable. The business model of Prosper.com (and other major players in the debt-based online crowdfunding arena) dictates that at least in the short term, they need to ensure higher probability of funding, since a significant source of their revenue is from the closing fees at the time of loan origination, not repayment. In addition, defaults likely occur several years after funding, and the platform may not have been fully able to foresee such consequences.

Loan categories: We have explored the overall regime change effects on our main transaction outcomes and a variety of other important outcomes thus far. However, the regime change effects may be heterogeneous across different subgroups of personal loans. An important consideration is the "purpose" of a personal loan. Holding

everything else constant, a personal loan used for debt consolidation is arguably distinct from that for home improvement. We seek answers to the question whether considering "moderating" type of factors, such as the loan category, could affect the primary findings. We conduct subsample analyses by exploring the heterogeneous effects on our main transaction outcomes.

The majority of Prosper loans are used for personal debt consolidation and private home improvement in our study period. These loans comprise of about 80% in our main sample, with $5,762$ loans for debt consolidation and $3,309$ for home improvement. We repeat our matching estimations for each category and report the results in Table 2.20. We first find that the regime change effects are consistent across all loan categories. They are qualitatively the same as our main findings. However, we also find heterogeneous effects across the categories. Specifically, we find that the interest rates increase more dramatically in the category of debt payment. Similarly, the probability of successfully funding and default rate also increase more dramatically in this subcategory.

"Soft information": The P2P social lending allows for incorporations of "soft information" in the funding process (Lin et al. 2013). Examples of the soft information include friends' bidding, offline acquaintances, and so on. The switch in mechanisms has significant effects on how the information affects loans' funding outcomes. Compared to borrower-choosing interest rates under auctions, the posted-price mechanism fixes the interest rate given credit profile. Therefore, the "price" under posted prices does not reflect the borrower's soft information, which is a vital aspect of the personal loan. Among things reflecting borrowers' soft information, the friendship on the platform is an important proxy for borrowers' credibility (Lin et al. 2013). Therefore, given the credit profile, the effect of having friends on the success rate of funding should be more significant under the posted-price regime. We test this auxiliary hypothesis of friendship on the funding probability by conducting subsample analyses. Specifically, we examine the effect of having Prosper friends for loans under auctions

Table 2.20: Matching Estimates of the Regime Change Effects by Loan Categories

	Coef. estimates for 1(Posted Prices)[a]			
Transaction outcomes:	Initial Interest Rates	Funding Probability	Contract Interest Rates	Default Rate
Loan categories:				
Debt consolidation	1.327***	0.351***	0.196	0.035**
	(0.199)	(0.017)	(0.735)	(0.015)
Home improvement	0.681	0.259***	0.667	0.005
	(0.439)	(0.049)	(0.571)	(0.030)
Others	0.725***	0.273***	0.707	0.023
	(0.158)	(0.016)	(0.505)	(0.016)
Control covariates:				
Personal loan characteristics	Yes	Yes	Yes	Yes
Borrower's credit profile	Yes	Yes	Yes	Yes
Macroeconomic environment	Yes	Yes	Yes	Yes
Interest rate categories	Yes	Yes	Yes	Yes

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

[a]This table presents matching estimates of the regime change effects on all transaction outcomes by loan categories. For all estimations, we let the number of matches to be 4.

and posted-price loans separately. Results are reported in Table 2.21. Consistent with our conjecture, we find that the effects under posted prices are indeed larger than in the auctions regime. Numerically, having friends on Prosper.com raises the funding probability by 12% under posted prices, while less than 7% before the regime change.

2.6. Conclusions

The choice of market mechanisms is one of the most fundamental questions in any marketplace. For a nascent industry such as online peer-to-peer lending, or more broadly online crowdfunding, this choice is critical. For market designers, platform owners and policy makers, there are delicate short-term and long-term consequences that should be carefully weighed. Although currently the two major peer-to-peer lending platforms in the US both use posted-prices, many peer-to-peer lending plat-

Table 2.21: Effects of "Having Friends" on the Funding Probabilities

	Before Dec. 20, 2010[a]			**After Dec. 20, 2010**		
$y = 1$(Loan funded)	$M = 1$	$M = 4$	$M = 8$	$M = 1$	$M = 4$	$M = 8$
1(Having Friends)	0.078***	0.069***	0.065***	0.139***	0.122***	0.128***
	(0.025)	(0.020)	(0.020)	(0.038)	(0.035)	(0.034)
Control covariates:						
Personal loan characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Borrower's credit profile	Yes	Yes	Yes	Yes	Yes	Yes
Macroeconomic environment	Yes	Yes	Yes	Yes	Yes	Yes
Interest rate categories	Yes	Yes	Yes	Yes	Yes	Yes
Num. obs.	8,470	8,470	8,470	4,547	4,547	4,547

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

[a]This table presents the nearest-neighbor propensity score matching estimates of the effect of "having friends" (on Prosper.com) on the funding probability. M is the number of matches to be searched. Typically the rule of thumb for M is 3 or 4. The standard errors are calculated by the method suggested in Abadie and Imbens (2006). All the estimations are performed in R (Sekhon 2011).

forms in other countries, as well as other types of crowdfunding platforms, still use auctions instead. More important, there is little systematic empirical evidence in favor of a particular mechanism. Our goal is to take a first step in fulfilling this gap.

We exploit a regime change on an online peer-to-peer lending marketplace, Prosper.com, to answer this question. We develop a stylized game theoretic model to compare the multiunit uniform price auctions and posted-price sales, taking into account the incentive of the platform to maximize its own expected payoff; and test model predictions with detailed transaction data from the market around the time of the regime change. Our empirical results lend support to our hypotheses in terms of initial and contract interest rates, as well as funding and default probabilities.

We find evidence of short-term improvements for both borrowers and lenders. For lenders, our results show that lenders benefit from a quicker "deployment" of funds because loans are funded faster under posted prices. There appeared to be a virtuous cycle: After the regime change, lenders tend to invest more in each bid, and invest sooner. This, in turn, contributes to faster funding of loan requests and deployment

of lenders' funds. Borrowers also seem to benefit from the change, since their requests are funded sooner. Both of these short-term benefits in terms of funding probabilities were noted by Prosper.com when they made the regime change, and rightly so.

On the other hand, however, as our theoretical model predicts and our empirical analysis shows, Prosper.com assigns higher initial interest rates under posted prices than borrowers would have in auctions; in fact, about 100 basis points higher. That is, while lenders enjoy a higher nominal return on loans, it comes at a cost to borrowers. More important, this change has long-term implications for the repayment of loans. As the finance literature has long documented (Stiglitz and Weiss 1981), the interest rate is one of the most critical factors in predicting *ex post* loan default. Our analysis of loans originated around the time of the regime change is consistent with this prediction, and we find that loans are indeed more likely to default under posted prices. These findings, along with our analysis in Appendix A.2, suggest that the switch to posted prices does not necessarily increase overall social welfare.

Our paper is one of the first to systematically document these intended and unintended consequences of the regime change in online peer-to-peer lending. While the traditional "crowd"-based auction mechanism may be slower to fund loans, its long-term welfare effects may not necessarily be worse than posted prices set by experts. Since peer-to-peer lending platforms such as Prosper.com deduct service fees at the time of the loan origination rather than repayment, it *is* in their best interest to ensure a higher funding probability in the short-term. However, they will be well advised to take into account the long-term repayment of loans. In other words, market designers of online crowdfunding, as well as policy makers, should carefully consider the long-term consequences on borrowers' capabilities and willingness to repay. Hence, our study not only has implications for peer-to-peer lending, but for other types of crowdfunding as well: When the marketplace derives revenue from short-term success, such as the origination of a loan or the funding of a business venture, open auctions that rely on "wisdom of the crowd" may (under certain conditions) actually allocate

resources in a more socially desirable fashion. More broadly, our study contributes to the growing literature on electronic market design and online auctions, and also a better understanding of how market participant behaviors change under different market mechanisms.

3. "Smart Money": Institutional Investors in Online Crowdfunding

Mingfeng Lin and Zaiyan Wei

3.1. Introduction

The "crowd" in online crowdfunding is not only comprised of retail investors; in fact, the debt-based crowdfunding or peer-to-peer (P2P thereafter) lending market has long attracted the interest of institutional investors. Particularly in recent years, as the industry matures, it is moving toward relying on institutional instead of retail investors for the supply of capital. LendingClub.com and Prosper.com are the two largest P2P lending platforms for unsecured personal loans in the U.S.. Lending Club in September 2012 introduced a new program named "whole loans," in which institutional investors have the privilege to "purchase" loans in their entirety. Prosper has a similar program launched in April 2013. Since then, the fraction of institutional investment in both platforms has skyrocketed, *e.g.*, two thirds of the total loan volume on LendingClub.com (The Economist 2013).

Given their expertise and adequate source of capital, institutional investors are often referred to as "smart money" in financial markets (Shleifer and Summers 1990), in the sense that these professionals have selection ability and can yield on average higher returns from their choice of investment (Gruber 1996, Zheng 1999). Although institutional investors are studied extensively in the literature (Puckett and Yan (2011), Basak and Pavlova (2013) as examples of recent investigations), there is surprisingly little research on the interactions between them and retail investors. Moreover, in traditional financial markets, retail investors are very few, and their trading volumes are largely negligible (Barber and Odean 2011). In contrast, our context, the emerging debt-based crowdfunding, provides us a unique research opportunity to study the

interactions. We ask whether institutional investors are indeed "smarter money" in the marketplace than, and have any significant impacts on the behaviors of, retail investors. We study these questions by exploiting a "natural experiment" with the identity of institutional investors on Prosper.com.

Prosper.com, the first P2P lending platform in the U.S., allowed institutional investment since its official open to the public in February 2006. Before May 19, 2008, Prosper labeled all investors, either institutional or retail, as generic "Lender" on their profile pages. Starting from that date, Prosper suddenly changed institutional investors' labels to "Institutional Lender."[1] Since all users' profile pages are public on Prosper.com, it is now easy to differentiate the types of investors directly from their labels. This change was immediately effective on the whole site, and unanticipated by either borrowers or lenders. It provides us an opportunity to study how institutional investment affects behaviors of retail investors, transaction outcomes, and overall market efficiency under the context of online consumer loan market.

Understanding the effects of institutional investors on behaviors of smaller investors and overall market efficiency has important managerial as well as policy implications for the general online crowdfunding industry (Agrawal et al. 2013). Particularly, the "Jumpstart Our Business Startups" (JOBS) Act, signed by President Obama, legalized equity-based crowdfunding in the U.S. in 2012. Only if we understand the interactions between institutional and retail investors, especially the effect of financial professionals on individuals, can we design a more efficient online crowdfunding market that matches fund raisers (either individuals in the case of debt-based crowdfunding or enterprises in equity-based crowdfunding) and investors.

To address our research questions, we first characterize the behaviors and investment outcomes of institutional investors on Prosper.com. We focus on transactions

[1]Prosper introduced other improvement of account infrastructure for institutional investors, along with the change in labels. The official announcement of the change can be found at Prosper's corporate blog: http://blog.prosper.com/2008/05/page/2/.

in 2007. In October 2008 Prosper.com was temporarily shut down due to its conflicts with the rules of the U.S. Securities and Exchange Commissions (SEC). After its reopening in April 2009, Prosper made a number of revisions to its funding rules in accordance with SEC requirements. We mainly investigate the growth of institutional investment, compare institutional investors' behaviors (*e.g.*, investment volumes, diversification strategies, and bidding strategies) and their investment outcomes such as the return on investment (ROI) with retail investors. We answer the question whether these professionals are indeed "smart" in using their capital.

We continue with a systematic examination of the effect of labeling institutional investors on their behaviors and investment strategies, on behaviors of retail investors, and on loan-level transaction outcomes. We use the "natural experiment" before and after May 19, 2008 to explore these effects. We construct our main sample including loan requests that were posted on Prosper.com between January 19, 2008 and September, 19, 2008. We first examine how the labeling event affected institutional investors, including their activeness in participating, diversification strategies, portfolio performances, and lending strategies mainly the entry stage in a funding process. We continue with the effects of institutional investment on retail investors' behaviors, focusing on outcomes including the average amount proposed by retail investors, the number of active retail investors, the average time between bids for retail investors, and the percent of retail bids winning. We further examine the effects on loan-level outcomes including whether a request is funded, the contract interest rate (or APR) if funded, whether a funded loan defaulted, and the overall ROI for funded loans.

We find that institutional investors had much larger loan volumes than retail investors, and they invested in more diverse loans in terms of risk levels. However, we do not find that the professionals' returns on investment dominated those of retail investors. These findings suggest that although often referred to as "smart money" in traditional markets, institutional investors are not so smart on Prosper.com. In addition, we do not find evidence that the event of labeling institutional investors

had any significant effects on their investment behaviors (such as bidding volumes and diversification strategies), or investment performance like ROI from their portfolios. However, we find that institutional investors adjusted their lending strategies. Specifically, these professionals entered on average earlier in funding processes after the labeling event.

For the effect of the labeling event on retail investors and loan-level transaction outcomes, our results suggest, first of all, that labeling institutional investors was associated with less retail investors participating after the first institutional investment. However, retail investors, who still participated after the first institutional investor, offered larger amount of dollars, and submitted them quicker. We also observe that the effect on the percent of retail bids winning was positive, perhaps due to the increased intensity of competition after the entry of institutional investors. For the effects on transaction outcomes, we find heterogeneous effects depending on the first institutional investor's entry stage (in terms of percent of funds fulfilled). Specifically, we observe that the early entry of institutional investors (less than 10% funded) was associated with higher contract interest rate, but lower if they entered later. We find that the effect of early entry was positive for funded loans' default rate. We do not find significant effects on loans' overall ROI.

Our paper contributes to the literature on institutional investors (Shleifer and Summers 1990, Sias 2004, Barber and Odean 2008, Stein 2009) by investigating the direct effect of institutional investment on retail investor behaviors using an exogenous change in a consumer loan market. Our results also add to the identification of "herding" behaviors that are observed in various markets (Scharfstein and Stein 1990, Bikhchandani et al. 1992, Wu 2013) by studying whether the phenomenon in consumer loan markets is driven by the existence of financial professionals. We also contribute to an emerging literature on crowdfunding in general (Agrawal et al. 2013, Lin et al. 2013, Rigbi 2013) across multiple disciplines. We start with our theoretical motivations in the next section, and introduce our research background in Section 3.3. We continue

with our data and empirical analyses in Sections 3.4 through 3.6, and conclude in the last section.

3.2. Theoretical Background

Institutional investors are often regarded as sophisticated and well-capitalized investors in financial markets (Shleifer and Summers (1990) among others). Their investment behaviors, associated transaction outcomes and investment performance, are widely studied in various strands of literature such as economics, finance, law, and management. Theoretical models (Stein 2009, Chevalier and Mayzlin 2006, Basak and Pavlova 2013) have been proposed to understand mechanisms underlying their behaviors, and extensive empirical studies (Lawton and Marom 2010, Sias 2004, Puckett and Yan 2011) are conducted to test the theories. Retail investors, not only smaller in budgets but also less informed than their institutional counterparts, are also well studied in the literature (Barber and Odean (2011) provide a thorough review). However, the interactions between these two types of investors are much less studied in the literature (Nofsinger and Sias 1999, Barber and Odean 2008). The direct effect of institutional investment on the behaviors of retail investors, such as whether they are following the professionals' investment decisions, is even rare in the literature. Part of the reason is that in traditional financial markets, retail investors are very few in numbers and small in trading volumes. In contrast, P2P lending markets are non-traditional in the sense that retail investors play important roles in online transactions. Thus the current paper contributes to the literature by providing the empirical evidence of the interactions between institutional and retail investors.

Information asymmetry and thus adverse selection are present in the Internet-based P2P lending markets (Lin et al. 2013). Theory (Wu 2013) predicts that investors, when making investment decisions, optimally ignore their private information and simply imitate previous investors in the presence of information asymmetry. This

phenomenon of mimicrying in financial markets is often called "herding" behavior, or information cascading (Scharfstein and Stein 1990, Banerjee 1992, Bikhchandani et al. 1992, Zhang and Liu 2012, Wu 2013). We empirically observe the "herding" behavior of retail investors in our context. We show that the "herding" phenomenon was driven by the designation of professional titles – the label of "Institutional Lender." We contribute to the literature of information cascading by empirically examining whether the herd behavior in consumer financial markets is driven by the existence of professionals.

3.3. Prosper.com and Investors

Prosper.com, established in February 2006, was the first debt-based crowdfunding marketplace in the U.S.. Prosper provides a platform that matches lenders and individual borrowers all over the country to originate fixed-rate unsecured personal loans. The dollar amount of these loans ranges from $\$1,000$ to $\$35,000$, and types of loans include debt consolidation, home improvement, small businesses, and so on. Up to August 2014, more than $95,000$ borrowers have collected loans with over $1 billion, making Prosper.com the second largest platform of personal loans in the U.S..

Prosper was using an auction-based funding process before December 2010 (called the "eBay for personal loans" during that time), after which the platform switched completely to a fixed interest rate mechanism (Wei and Lin 2013). In this study, we focus on the period before the website's shutdown in October 2008 that was due to its conflicts with the rules of the U.S. Securities and Exchange Commission. Thus during our study period the funding processes were auction-based. A typical loan request starts with the borrower posting a listing on Prosper.com. The borrower indicates the amount requested and the maximum interest rate at which he or she is willing to pay. Prosper adds verified financial information about the borrower collected from Experian - one of the three traditional credit bureaus. Upon observing

the loan request, any lender, institutional or retail, can submit bids by specifying a certain amount of dollars and the lowest interest rate he or she is willing to lend. The amount can be any portion of the requested amount, and the interest rate has to be at most the borrower's pre-determined rate. The auction process finishes after a pre-specified duration expires (typically 7 days before 2009). The winners are the lenders with lowest interest rates, and the total amount proposed by them will cover the borrower's requested amount. The listing can fail to be funded. Once it raises full funding and transfers to a personal loan, it will be fully amortized in monthly payments over three years. The borrower is subject to late fees and can suffer a substantial decrease in his or her credit score. For more details of Prosper auctions and P2P lending in general, Lin et al. (2013) provide more thorough summaries. Here we focus on the most relevant part to our research questions.

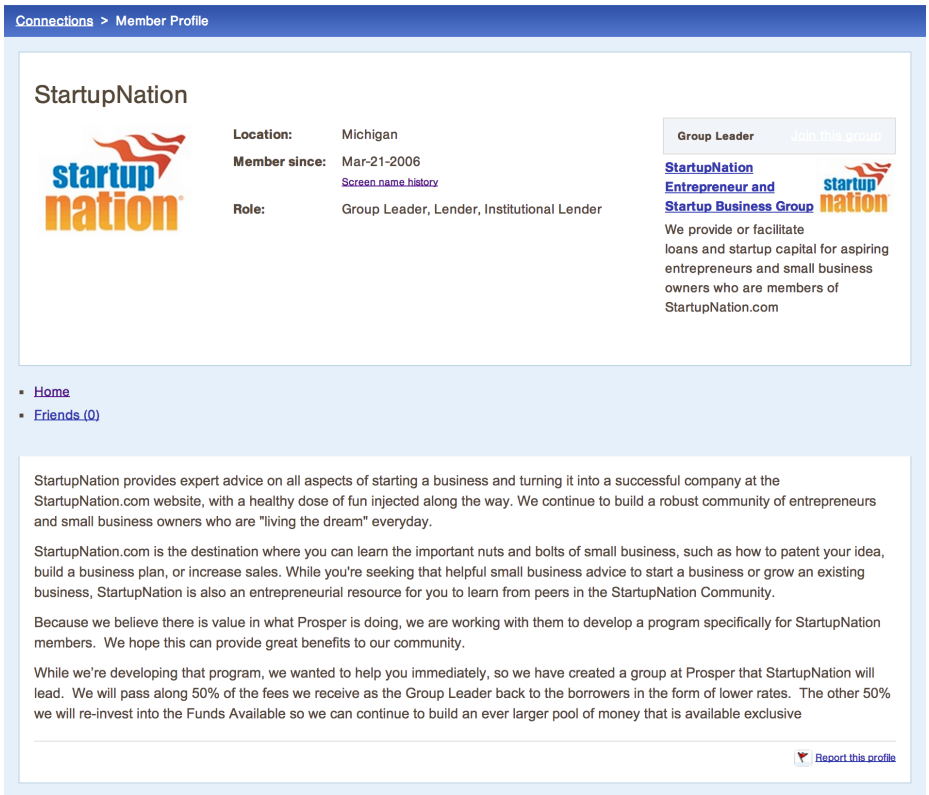
All types of institutional investors have been allowed to lend on Prosper.com since its inception in 2006. According to their company profile, most of the institutional investors are money managers or agents working for pension funds, mutual funds, and life insurance companies. The size of these institutions has a wide range, from small institutions with annual revenue of $\$10,000$, to big companies with \$1.6 trillion in assets under management. Before May 19, 2008, all lenders, either institutional or retail, were labeled as generic "Lender" on the website. Starting from the date, Prosper enhanced the account infrastructure for institutional investors. An important and prominent aspect of the improvement was that Prosper started labeling institutional investors as "Institutional Lender," keeping the label of other investors unchanged.[2]

This change had a substantial impact in the funding processes. The auction mechanism used by Prosper was an open format, in the sense that a potential lender was able to observe all previous bidders' identities, bidding amount, and in some cases

[2]All users of Prosper.com have labels of their identities, either borrower, lender, or both. This identity information is public on the website. Particularly, Prosper changed the label for lenders and differentiated them as either "institutional lender" or generic "lender."

the proposed interest rate.[3] This implies that after May 19, 2008, if an institutional investor participated in an auction, all other lenders would be able to observe its special identity. Figure 3.1 provides an example of the institutional investor's profile page. The label of "Institutional Lender" is now public under the "Role" section, particularly prominent to all other lenders. Potentially, the participation of institutional investors serves as a signal of the borrower's creditworthiness and repayment ability, which obviously has profound impact on other lenders' investment decisions and thus transaction outcomes.

Figure 3.1: A Screenshot of the Profile Page of an Institutional Investor on Prosper.com[4]



[3]Lenders may be outbid in Prosper auctions. It happens when other lenders' proposed rates are strictly lower and their aggregate amount satisfies the borrower's funding needs. Once a lender is outbid, his or her interest rate will be made public on the listing page.

3.4. Data and Sample

Prosper provided public access to its API service containing all its administrative data. The dataset contains all loan requests, both failed and successful listings, that were posted on the website up to February 2013. It also records all bids made in the auction stage, and investment in the fixed interest rate regime. Prosper also makes the verified financial information, collected from Experian, available to Prosper lenders. We thus have access to all the credit information about a borrower, such as the range of credit score, credit usage information, detailed credit history, and so on. For funded loans, we are able to track the monthly payment history and thus payment results (e.g., defaulted or not).

We construct two samples for our empirical analyses from the dataset. For characterizing institutional investors on Prosper.com, we focus on all transactions that were made in the calendar year 2007. An advantage of focusing on this period is that the platform features were largely consistent throughout the period. The website experienced a temporary shutdown after October 2008, due to its violations of SEC rules. During its silence period, Prosper completed its registration with SEC and updated many transaction rules. This event had potentially the most profound influence on all aspects of its transactions in Prosper's history. As an example, Prosper raised the minimum credit score requirement to 640 after reopening. Therefore the borrower population has dramatically changed since the SEC regulations. We thus ignore all loan requests that were posted after Prosper's reopening in April 2009. This sample contains $133,537$ loan requests with a total of $2,665,716$ unique bids submitted.

We mainly study the effects of institutional investment on behaviors of retail investors and transaction outcomes. As mentioned above, we make use of the exogenous change in the label of institutional investors on May 19, 2008. We therefore construct the second sample of transactions within a short period of time before and after the date, which is the main sample used in our estimations. Specifically, the sample

contains all loan requests between January 19, 2008 and September 19, 2008. We delete the listings that were posted before the change and ended afterwards. These listings comprise a negligible subset of our second sample, less than 5% of the total. After the elimination, we have 98,395 loan requests in profile with 2,234,092 bids submitted to these loans. Among these listings, 9,722 were funded and transformed to unsecured personal loans. With monthly payment data, we are able to calculate the return on investment (ROI) for 9,292 funded loans. Less than 5% of personal loans have missing values of ROI due to data incompleteness or data absence. Table 3.1 reports the descriptive statistics for the main variables. It is easy to see that the listing characteristics or borrowers' credit profile did not change after May 19, 2008. In other words, the demand side did not change dramatically before and after Prosper labeling the institutional investors.

Table 3.1: Summary Statistics of All Loan Requests Posted January 19, 2008 - September 19, 2008

	Before May 19, 2008		**After May 19, 2008**			
Variables:[a]	**Mean**	**sd**	**Mean**	**sd**	***t*-Statistic**	***p*-Value**
Amount Requested ($)	7433.320	6438.256	7145.494	6341.980	-0.849	0.408
# Bids	23.280	77.220	22.163	71.337	0.845	0.411
Borrower Maximum Rate (%)	20.928	9.575	26.476	9.599	0.731	0.474
Estimated Loss (%)	15.231	9.649	14.645	9.154	-0.154	0.878
1(Borrower is Homeowner)	0.381	0.486	0.383	0.486	0.922	0.370
1(Funded)	0.094	0.292	0.103	0.304	0.838	0.415
Listing Effective Days	5.769	3.029	5.641	2.390	-0.111	0.912
Bankcard Utilization (%)	63.137	41.915	62.277	41.505	-0.668	0.513
Current Credit Lines	8.840	6.186	8.545	6.118	-0.514	0.612
Current Delinquencies	2.858	4.588	2.594	4.416	-0.724	0.478
Delinquencies Last 7 Years	9.368	15.449	9.117	15.732	0.532	0.601
Inquiries Last 6 Months	3.572	4.344	3.223	3.914	-0.902	0.380
Open Credit Lines	7.757	5.481	7.722	5.564	-0.461	0.649
Public Records Last 10 Years	0.577	1.181	0.521	1.093	-0.007	0.994
Public Records Last 12 Months	0.073	0.336	0.066	0.306	0.863	0.400
Total Credit Lines	26.097	14.829	25.668	14.930	0.845	0.410
Total Open Revolving Accounts	6.002	5.016	5.913	5.059	-0.650	0.523
1(With Institutional Lender)	0.025	0.155	0.120	0.325	0.386	0.703
Num. obs.	47,813		50,582			

[a]This table presents the summary statistics of key variables for loan requests that were posted between January 19, 2008 and September 19, 2008 included. It is easy to see that the listing characteristics and borrower credit profile do not change after the labeling event on May 19, 2008.

3.5. Institutional Investors on Prosper.com

We start our analyses by characterizing institutional investors in the P2P lending market for unsecured personal loans. We examine their participation in the market by exploring the growth of their investment over time. We investigate their investment strategies such as the timing of entry and diversification strategies. We also compare their investment performance, mainly return on investment, with those of retail investors. We answer the question whether institutional investors are indeed "smart" in debt-based crowdfunding.

The number of active institutional investors, who were actively competing in listings, was growing over the period. Figure 3.2 shows the monthly number of active institutional investors before the website's shutdown in October 2008. There is an obvious jump in May 2008, mainly due to Prosper's improvement in account infrastructure for institutional investors. To check if institutional investors were making larger investment compared to retail investors, we compare the ratio between the number of institutional investors and that of retail investors, with the ratio of institutional funds versus retail funds. We find that the funds ratio was significantly larger than the investors count ratio in Figure 3.3. This implies that institutional investments are apparently large in volumes relative to retail investments. They are still "big fish" in the emerging online consumer loan market.

We also study how institutional investors' investment strategies are different from those of retail investors. We first examine how institutional investors' diversification strategies differed from retail investors'. Figure 3.4 shows the distributions of loan requests according to borrowers' credit grades. We do observe that institutional investors were more diversified in their investment. In terms of loan purposes, Figure 3.5 shows that compared to retail investors, institutional investors were more likely to invest in business-purpose loans. We also observe that institutional investors submitted their bids earlier than retail investors overall, in terms of the funded ratio prior to

Figure 3.2: Growth of Active Institutional Investors before October 2008

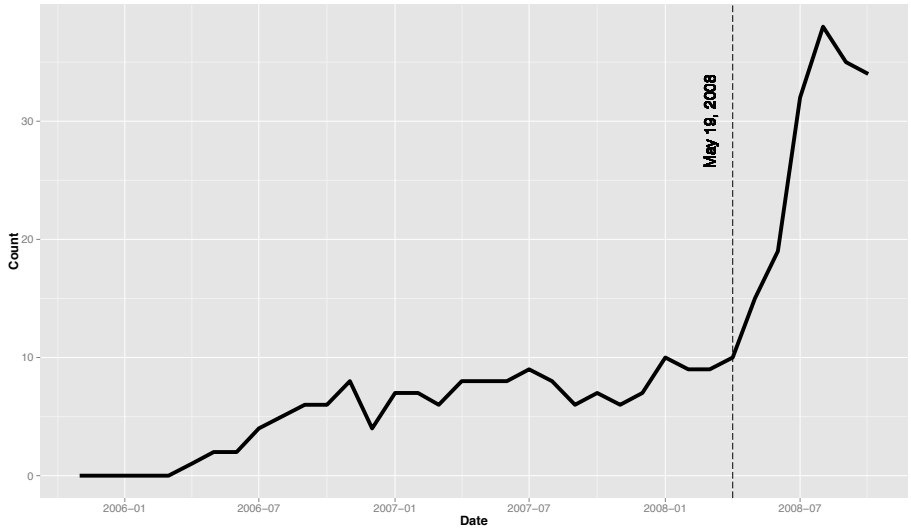
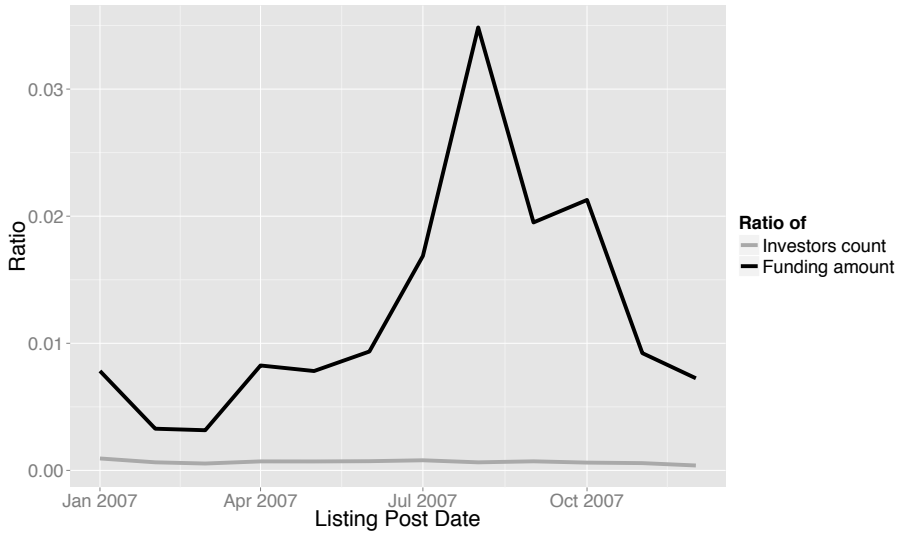


Figure 3.3: The ratio of Institutional and Retail Investors Count vs. the Ratio of Institutional and Retail Investment Amount in 2007



the bidding. Figure 3.6 illustrates that compared to institutional investment, a fairly large fraction of retail investors submitted their bids after the borrower's requested amount was satisfied. The x-axis of the figure evaluates the ratio of funds satisfied at which lenders submit their bids.

Figure 3.4: Distributions of Loan Requests by Borrowers' Credit Grades in 2007

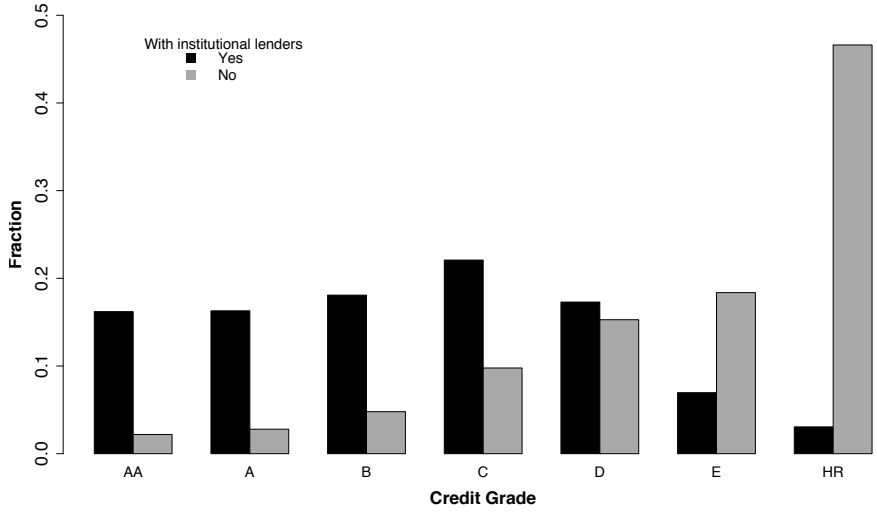
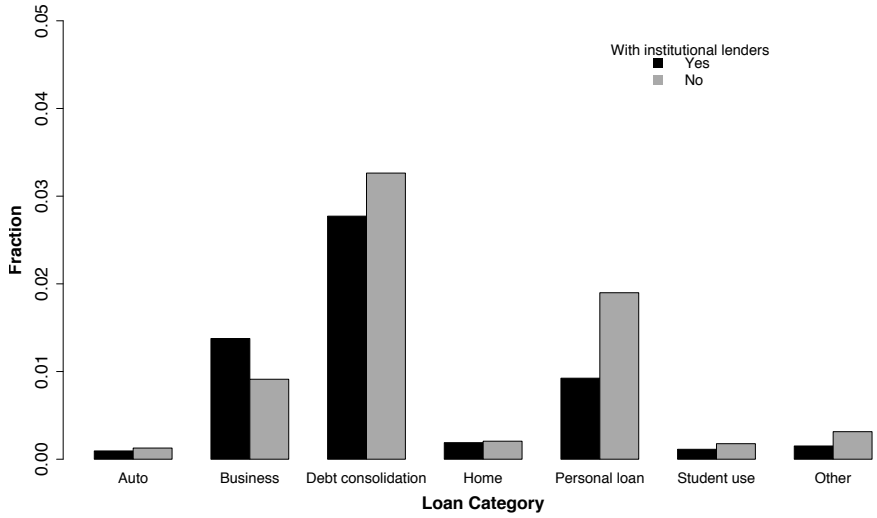
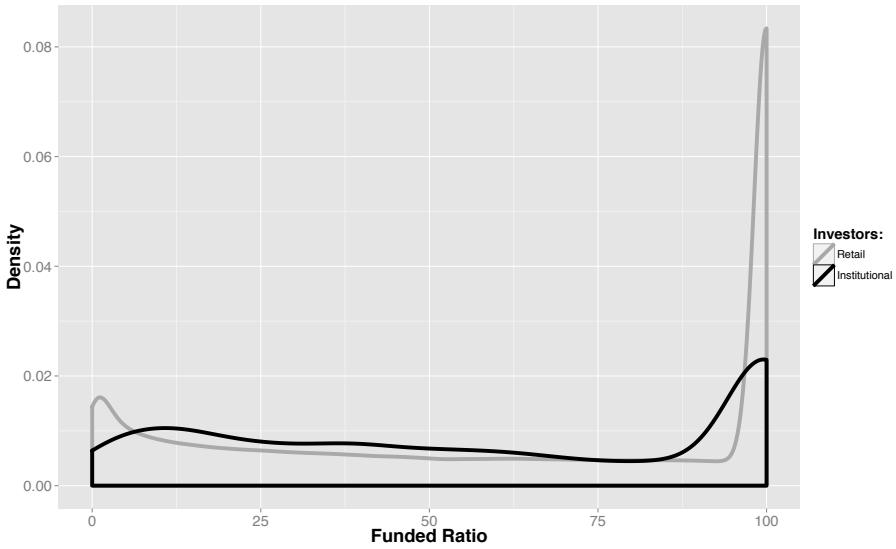


Figure 3.5: Distributions of Loan Requests by Categories (Purposes) in 2007



Institutional investors are professionals in financial markets. Given their expertise, institutional investors often have better performance than retail investors in terms of financial returns such as return on investment (ROI). Prosper's rich dataset allows us to calculate the ROI for most of their funded loans. We compare lenders' ROI

Figure 3.6: Distributions of Bids by Percentage Funded in 2007

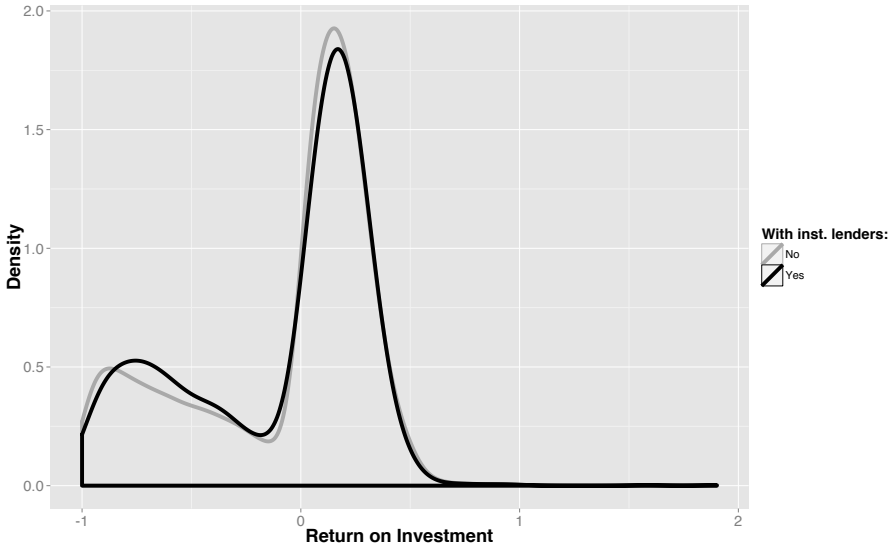


from loans with institutional investment to those without. Figure 3.7 (smoothed with Gaussian kernels) shows the distributions of the two subsets of Prosper loans.[5] It is noteworthy that these distributions are not very different, suggesting that institutional investors' ROI did not dominate that of retail investors. A two-sided t-test also rejects the null of equal means for the two samples.

We further study how institutional investment is correlated with transaction outcomes in the marketplace, including whether a loan request is funded, the contract interest rate or APR if a loan is successfully funded, whether a funded loan defaults, and more important lenders' ROI from a funded loan. We first examine the correlations between having institutional investment and the outcomes. Table 3.2 reports the coefficient estimates for the dummy variable of having institutional investment. The estimates suggest that the requests with institutional investment had significantly higher probability of being funded. Given funded, the loans had on average lower APR as well as lower default rate. The prediction about lenders' ROI is am-

[5]It is noted that both distributions have multiple peaks. This phenomenon is driven by two subsets of loans. One subset of loans defaulted, while the other subset were paid in full. The two modes reflect the difference in loans' returns between these two subsets.

Figure 3.7: Distributions of Funded Loan by Lenders' Return on Investment in 2007



biguous, since the interest rate is positively correlated with ROI while the rate of default is negatively related. Interestingly, our estimates support the conjecture that the loans with institutional investment did not generate higher ROI for the investors. This finding implies that these financial professionals are not necessarily better at picking up investment opportunities than retail investors on Prosper.com. Further investigations of the relationships between the number of participating institutional investors and transaction outcomes yield qualitatively similar results (Table 3.3).

As a short summary, we find that institutional investors are still influential players in the nascent P2P lending market. But as financial professionals, they are not always better at picking up investment opportunities than retail investors. Next we turn to the exploration of the effects of the "natural experiment" on both institutional and retail investors' lending behaviors, as well as transaction outcomes.

Table 3.2: Effects of Having Institutional Investment on Transaction Outcomes

	OLS Estimates			
Dep. var.:	1 (Funded)	Interest rate	1 (Defaulted)	ROI
1 (Having institutional lenders)	0.482***	-0.862***	-0.029***	0.012
	(0.007)	(0.087)	(0.010)	(0.009)
State and Week FE	Yes	Yes	Yes	Yes
Verified Credit Information	Yes	Yes	Yes	Yes
Discretization of Financial Info.	Yes	Yes	Yes	Yes
Non-verified Information	Yes	Yes	Yes	Yes
Adjusted R^2	0.273	0.662	0.158	0.094
Num. Obs.	133,537	10,102	10,102	10,089[a]

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

[a]There are 13 personal loans with no loan performance (repayment) data.

Table 3.3: Effects of the Number of Institutional Investors on Transaction Outcomes

	OLS Estimates			
Dep. var.:	1 (Funded)	Interest rate	1 (Defaulted)	ROI
# Institutional investors	0.340***	-0.727***	-0.020***	0.010
	(0.005)	(0.062)	(0.007)	(0.007)
State and Week FE	Yes	Yes	Yes	Yes
Verified Credit Information	Yes	Yes	Yes	Yes
Discretization of Financial Info.	Yes	Yes	Yes	Yes
Non-verified Information	Yes	Yes	Yes	Yes
Adjusted R^2	0.264	0.663	0.158	0.094
Num. Obs.	133,537	10,102	10,102	10,089[a]

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

[a]There are 13 personal loans with no loan performance (repayment) data.

3.6. Effects of Labeling Institutional Investors

Our empirical analyses are motivated by the absence in the literature on the interactions between institutional and retail investors, as well as predictions about investment behaviors from the theory of information cascading. As systematic investigations of the effects of labeling institutional investors, we start with the effects on themselves. We compare institutional investors' "activeness" and lending strategies before and after the labeling event. Then we study how the change in retail investors' lending strategies before and after observing institutional investment differs after the

labeling event. Last but not least, we compare the transaction outcomes before and after the "natural experiment," including funding outcomes (whether funded and interest rates if funded) and payment results (whether defaulted and lenders' ROI from funded loans).

3.6.1. Effects on Institutional Investors

Institutional investors potentially have strong incentives to adjust their investment and bidding strategies after the platform starts revealing their identities to the public. We start our exploration of the effects of the labeling event on these professionals' investment behaviors. As a short preview, we do not find significant changes in their "activeness" (measured by their frequency of participation), diversification strategies, or portfolio performances. However, institutional investors adjusted their bidding strategies by submitting larger funds earlier than before.

We first notice that after May 19, 2008, the number of registered institutional investors witnessed a rapid growth as illustrated in Figure 3.2. We further find that the fraction of institutional bids increased dramatically after the labeling event (Figure 3.8). These findings seem to suggest that institutional investors became more "active" with their identities being revealed to the public. However, it is observed that these changes were driven by the new institutional investors who registered after May 19, 2008. To check whether the event of labeling has any effects on the activeness of institutional investors, we focus on the bids submitted by the set of professionals that registered before the natural experiment. We do not find huge change in the "activeness" for these institutional investors (Figure 3.9).

We also examine whether institutional lenders' investment strategies in terms of choices and outcomes changed dramatically. We first compare their diversification strategies before and after the labeling event. We find that institutional lenders invested slightly more in loans with higher risk levels. Figure 3.10 shows the compar-

Figure 3.8: Daily Fraction of Bids by Institutional Investors before and after May 19, 2008

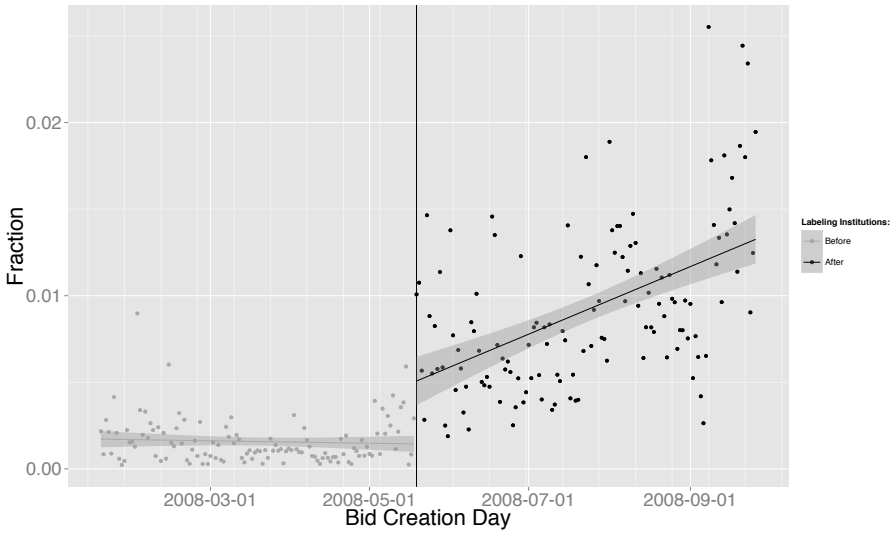
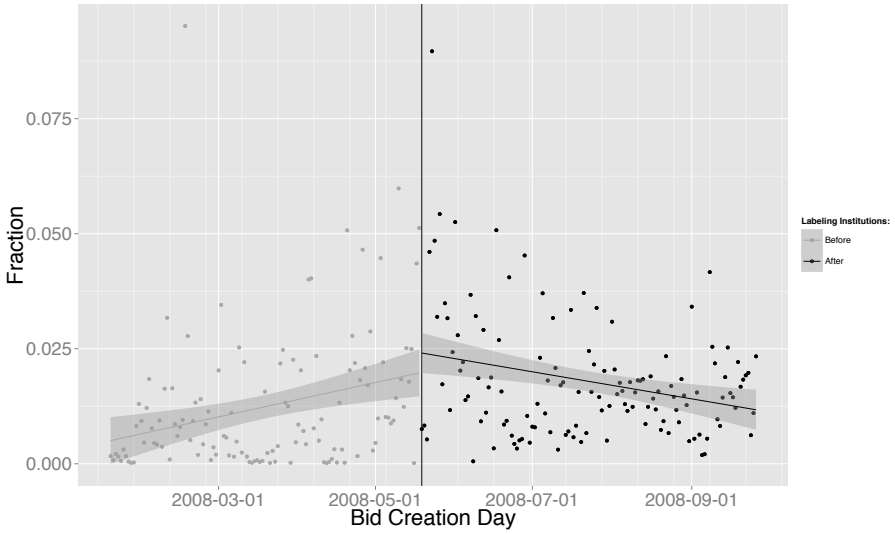


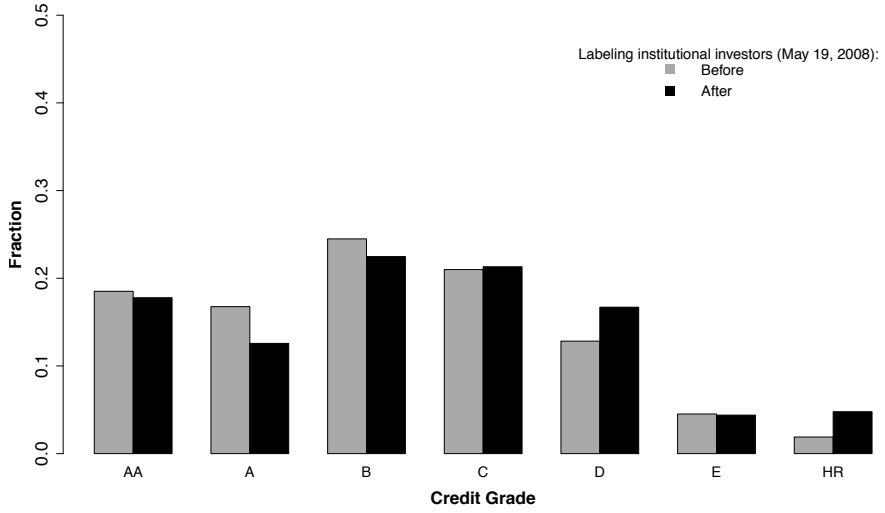
Figure 3.9: Daily Fraction of Bids by Institutional Investors (Registered before Labeling) before and after May 19, 2008



ison of the distributions of funded loans with institutional investment. However, this change in the composition of personal loans was far from significance. On the other hand, we check whether the payment outcomes were dramatically different for loans

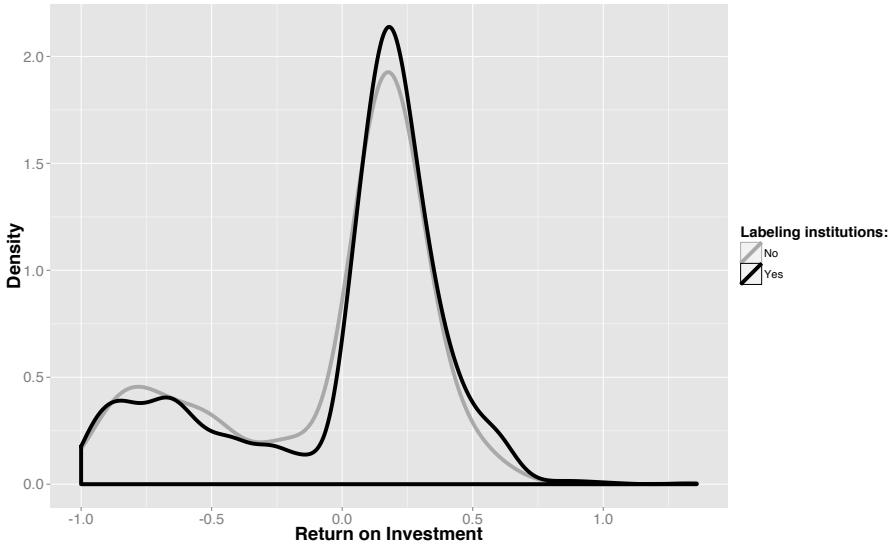
with institutional investors' participations. We compare the distributions of loans with these financial professionals in terms of the return on investment. Figure 3.11 (smoothed with Gaussian kernels) suggests that, on average, institutional investors' gains (or losses) did not change too much after May 19, 2008. A standard two-sided t-test also fails to reject the null of equal means in the two samples.

Figure 3.10: Distributions of Funded Loans with Institutional Investment before and after May 19, 2008



Although institutional investors' activeness, diversification strategies, or return on investment did not change dramatically after the labeling event, they adjusted their lending strategies in light of the fact that the revelation of their identities convey information to other investors. We first find that institutional investors entered earlier in the funding process after Prosper.com labeling them. We examine their entry stage by the percent of funds satisfied, the number of bids submitted, and the prevailing interest rate when entering. Estimates in the first three columns of Table 3.4 suggest that, on average, after the labeling event institutional investors were more likely to participate when more funds needed, less bids submitted, and with higher prevailing interest rate. All these results imply that these professionals participated earlier than

Figure 3.11: Distributions of Loans (of ROI) with Institutional Investment before and after May 19, 2008



they did without identity revelation. Upon participating, the contents of institutional bids serve as signals of their estimates of the borrower's credit worthiness. Interest rates are not directly observed unless the bids are outbid in the subsequent funding process. But the amount of funds is public information observed by all other investors. We find that institutional investors became more conservative after Prosper labeling their identities, by offering less amount of dollars on average. The last column in Table 3.4 reports the OLS estimates of the effect.

3.6.2. Effects on Retail Investor Behaviors

During the funding process, the participation of institutional investors serve as signals of the borrower's credit worthiness. Retail investors, with limited information and usually biased estimates of the personal loan, will adjust their lending strategies upon observing institutional investment. We focus on how the change in retail lenders' bidding strategies before and after institutional investment differed after the labeling event on May 19, 2008.

Table 3.4: Effects of Labeling Institutional Investors on Their Lending Strategies

	OLS Estimates			
Dep. var.:	Med. cumulative amount funded[a]	Med. # Bids	Med. prevailing interest rate	Med. bidding amount
1 (Labeling institutional lenders)	-2.039e+03***	-26.759***	2.602***	-382.322***
	(308.032)	(3.747)	(0.163)	(51.419)
State and Week FE	Yes	Yes	Yes	Yes
Verified Credit Information	Yes	Yes	Yes	Yes
Discretization of Financial Info.	Yes	Yes	Yes	Yes
Non-verified Information	Yes	Yes	Yes	Yes
Adjusted R^2	0.323	0.314	0.670	0.059
Num. Obs.	7,238	7,238	7,238	7,238

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

[a]The first three dependent variables are evaluated prior to the time of institutional bids. As an example, the variable "med. cumulative amount funded" calculates the median cumulative amount funded right before all institutional bids. The last dependent variable is the median amount in all institutional bids. We repeat all estimates with the mean value (instead of the median) of each variables, and results are fairly close.

Empirical Strategies Since we study the effects of institutional investment on bidding strategies of retail investors, we focus on loan requests with institutional investment during the funding processes. We examine the effects of the "natural experiment" on the changes in retail investors' bidding outcomes before and after the entry of the first institutional investor. The bidding outcomes we examine are the average offering amount by retail investors, the number of participating retail investors and their bids, the average time in hours between retail bids, and the percent of retail bids winning to support.

In our estimation equation below, we include a set of week dummies and state dummies, controlling for any effect that is constant within a week and a state (Rigbi 2013). We also include all observed listing characteristics and borrower's complete credit profile. Prosper provides virtually all the information that lenders observe in the process of making their investment decisions. Therefore, controlling for all listing characteristics and credit profile enables us to account for the selection issue. In Appendix B.1 we provide the detailed list of control variables in our main specifications.

Our baseline model is

$$\Delta y_{ist} = \beta_0 + \beta_i \cdot D_{ist}^{\text{Labeling}} + \mathbf{b}_2' \cdot \mathbf{X}_{ist} + \beta_3 \cdot I_s^{\text{State}} + \beta_4 \cdot I_t^{\text{Week}} + \epsilon_{ist}, \tag{3.1}$$

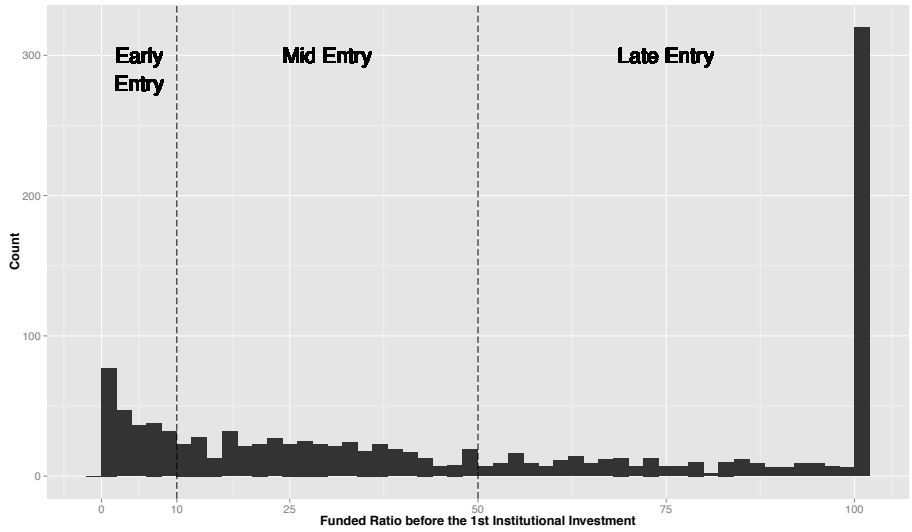
where Δy_{ist} is the difference in the outcome variables for retail investors before and after the first institutional investment. For example, we calculate the average values of retail investors' bidding amount for those before and those after the first institutional investor's entry respectively, and Δy gives the difference between the two means. Other outcome variables are constructed in the similar manner. The dummy variable D^{Labeling} equals 1 if the listing was posted after May 19, 2008. I^{State} and I^{Week} are vectors of state dummies and week dummies. The vector of control variables, $\mathbf{X}$, contains all listing characteristics and borrower's credit profile.

We further explore whether the effects are different when institutional investors enter at different stages. We carry out subsample analysis by dividing the loan requests with institutional investment into three subgroups, based on the entry stage of the first institutional investor in terms of the percent of funds fulfilled. Listings with the first institutional investment made when less than 10% of requests fulfilled are labeled as *Early Entry*, and listings with the first institutional investment made at 10% to 50% and more than 50% are labeled as *Mid Entry* and *Late Entry*. Figure 3.12 shows the distribution of listings and how we divide the subgroups. We estimate Equation (3.2) for this subsample analysis. The coefficients of the interaction terms are of interest and revealing how the effects of labeling institutional investors differ across subgroups.

$$\begin{aligned}\Delta y_{ist} = \beta_0 + \sum_{g \in \{\text{Early,Mid,Late}\}} \beta_{1,g} \cdot D_{ist}^{\text{Labeling}} \cdot D_{ist}^{g} + \sum_{g \in \{\text{Early,Mid}\}} \beta_{2,g} \cdot D_{ist}^{g} \\ + \mathbf{b}_3' \cdot \mathbf{X}_{ist} + \beta_4 \cdot I_s^{\text{State}} + \beta_5 \cdot I_t^{\text{Week}} + \epsilon_{ist},\end{aligned} \tag{3.2}$$

where D^g are dummies variables indicating each subgroup. Since we include time fixed effects in our baseline specification, the term for the labeling dummy D^{Labeling} is excluded from the equation. The term for late entry dummy D^{Late} is also deleted due to multicollinearity.

Figure 3.12: Distribution of Listings with Institutional Investment by the Entry Stage of the First Institutional Investor before May 19, 2008



Findings In our main specification (Equation 3.1), the estimates of which are reported in Table 3.5, the control variables include all listing characteristics such as borrower's asking interest rate and requested amount and all the verified credit information Prosper collected from Experian (e.g., homeownership, the number of delinquencies, the number of credit lines, and so on). We discretize a set of financial information such as the requested amount, bankcard utilization (in percentage) and the number of delinquencies due to the nature of the variables. We also add a set of non-verified information extracted mainly from the title of listings. Examples are whether "credit card", "student", or "small businesses" are mentioned in the title (Rigbi 2013). Appendix B.1 provides a complete list of our control variables.

We find that, first of all, after May 19, 2008 the difference between the average retail investors' bidding amount before and after the first institutional investment was larger. The positive estimates in the first column of Table 3.5 suggests that, on average, the labeling event caused the retail investors to propose larger amount upon observing an institutional investor's participation. This positive effect shows up re-

Table 3.5: Effects of Labeling Institutional Investors on the Retail Investor Behaviors

	Marginal Effects (OLS Estimates)[a]			
Dep. var.:	**Average bid amount**	**Active retail investors**	**Average time between bids**	**Percent of bids winning**
1 (Labeling institutional investors)	40.831*	-17.301	-8.937**	0.068
	(23.002)	(64.129)	(3.853)	(0.068)
State and Week FE	Yes	Yes	Yes	Yes
Verified Credit Information	Yes	Yes	Yes	Yes
Discretization of Financial Info.	Yes	Yes	Yes	Yes
Non-verified Information	Yes	Yes	Yes	Yes
R^2	0.046	0.083	0.191	0.126
Num. Obs.	5,983	5,983	5,983	5,983

[a]Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Marginal effects and standard errors clustered by state and week are presented.

gardless of the entry stage - the percent funded - of the first institutional investment (Table 3.6). We also find that the retail investors submitted their bids quicker after observing the participation of institutional investors. Estimates in column 2 of Table 3.5 says that the difference between the average time (in hours) intervals before and after the first institutional investment for retail investors was smaller after the labeling event on May 19, 2008. This negative effect caused by the designation is also robust across the first institutional investor's entry stage (as in Table 3.6). These findings are consistent with the predictions from the theory of information cascading ("herding"). Less informed (about the borrower's creditworthiness and payment ability) retail investors tend to imitate the investment choices of well-informed institutional investors.

Although the labeling of institutional investors caused the participating retail investors to submit larger amount of dollars and submit them quicker, we find that the change led to less participating retail investors after observing institutional investment on average. Estimates in the third column of Table 3.5 implies that, on average, the difference in the number of active retail investors before and after the first institutional investment was smaller after May 19, 2008. This crowding-out effect is likely due to

Table 3.6: Effects of Labeling Institutional Investors on the Retail Investor Behaviors - by Entry Stages

Treatment Categories	**Marginal Effects (OLS Estimates)**[a]			
1st institutional investment - Percentage funded	**Average bid amount**	**Active retail investors**	**Average time between bids**	**Percent of bids winning**
$<$ 10% Funded	54.401**	-94.892**	-6.784	0.138*
	(22.458)	(48.352)	(5.161)	(0.074)
$<$ 50% Funded	40.866*	-77.666	-4.006	0.148**
	(22.361)	(48.368)	(4.218)	(0.072)
50% More Funded	34.005	-11.914	-3.935	0.125*
	(21.717)	(47.129)	(4.112)	(0.072)
State and Week FE	Yes	Yes	Yes	Yes
Verified Credit Information	Yes	Yes	Yes	Yes
Discretization of Financial Info.	Yes	Yes	Yes	Yes
Non-verified Information	Yes	Yes	Yes	Yes
R^2	0.049	0.219	0.262	0.301
Num. Obs.	5,983	5,983	5,983	5,983

[a]Significance levels: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Marginal effects and standard errors clustered by state and week are presented.

institutional investors' large portfolios that allow them to bear lower interest rates, which leave retail investors too little profits to participate. This crowding-out effect holds across all subgroups by the entry stage of the first institutional investor (Table 3.6).

Apart from amount and time of bidding, we are also interested in whether retail investors competed more aggressively after seeing institutional investment. We find that, in the last column of Table 3.5, the difference between the percent of retail bids winning before and after the first institutional investment was bigger after institutional investors were labeled as "Institutional Lender". Part of the explanation is that since the participation of institutional investors serves as a signal of the borrower's creditworthiness, the retail investors, who decide to enter, are more willing to accept lower interest rates. Thus it led to higher winning probabilities for the retail investors who participate later.

3.6.3. Effects on Transaction Outcomes

Findings in previous sections suggest that the labeling event had significant effects on lending strategies of both institutional and retail investors. These changes directly affect the transaction outcomes, *e.g.*, whether a loan is successfully funded and the interest rate if funded. They have impacts on loan payment outcomes and lenders' portfolio performances indirectly by affecting the loan terms such as interest rates.

Empirical Strategies We are interested to examine how the difference in transaction outcomes between loans with institutional investment and those without changed after the "natural experiment" on institutional investors' identities. As an example, we consider whether a loan request is funded or not. We compare, before and after the labeling event, the difference in whether a request is funded for loans with institutional investment and those without. Other outcomes under considerations include, for funded loans only, the interest rate (or APR), whether a loan is delinquent, and the ROI for lenders.

In the analysis of the effects on retail investor lending strategies as in Section 3.6.2, we focus on the subgroup of loan requests with institutional investment. Instead, due to the nature of comparing loans with and without institutional investors, we examine all requests and funded loans for analyzing the effects on transaction outcomes. The main specification we estimate is hence different from Equation (3.1). The main equation for estimating how the labeling event affects transaction outcomes is,

$$y_{ist} = \beta_0 + \beta_1 \cdot D^I_{ist} + \beta_2 \cdot D^{\text{Labeling}}_{ist} \cdot D^I_{ist} + \mathbf{b}_3' \cdot \mathbf{X}_{ist} + \beta_4 \cdot I^{\text{State}}_s + \beta_5 \cdot I^{\text{Week}}_t + \epsilon_{ist}, \tag{3.3}$$

where y_{ist} are the transaction outcomes mentioned above. The dummy variable D^I_{ist} indicates whether a loan has institutional funds or not. Note that the term for $D^{\text{Labeling}}_{ist}$ is also omitted with the inclusion of week fixed effects. The coefficient, β_2, of the interaction term reveals how the difference in outcomes is correlated with the labeling event. Similar to the analysis of labeling effects on retail investors, we also

conduct subsample analysis by dividing all loans into three subsamples according to the entry stage of the first institutional investor. The extended specification is,

$$y_{ist} = \beta_0 + \sum_{g\in\{\text{Early,Mid,Late}\}} \beta_{1,g} \cdot D^g_{ist} + \sum_{g\in\{\text{Early,Mid,Late}\}} \beta_{2,g} \cdot D^{\text{Labeling}}_{ist} \cdot D^g_{ist} + \mathbf{b}_3' \cdot \mathbf{X}_{ist} + \beta_4 \cdot I^{\text{State}}_s + \beta_5 \cdot I^{\text{Week}}_t + \epsilon_{ist}. \tag{3.4}$$

Similarly, the set of coefficients, $\beta_{2,g}$, are the relations between the difference in outcomes and the labeling event.

Findings We now turn to the effects on transaction outcomes, including whether a loan request is funded, the APR or contract interest rate if funded, the payment result and ROI for funded loans. Estimates of the key coefficients in Equations 3.3 and 3.4 are reported in Tables 3.7 and 3.8. We include the same set of control variables as in the estimation of treatment effect on retail investors' behaviors.

Table 3.7: Effects of Labeling Institutional Investors on Transaction Outcomes

	Marginal Effects[a]			
Dep. var.:	1(**Funded**) Probit Est.	**APR** Tobit Est.	1(**Defaulted**) Probit Est.	**ROI** Tobit Est.
1 (Labeling inst. investors) × 1 (With inst.)	-0.030***	-0.552*	-0.012	0.008
	(0.007)	(0.302)	(0.020)	(0.035)
1 (With inst.)	0.187***	0.949***	0.003	0.030
	(0.006)	(0.252)	(0.023)	(0.029)
State and Week FE	Yes	Yes	Yes	Yes
Verified Credit Information	Yes	Yes	Yes	Yes
Discretization of Financial Info.	Yes	Yes	Yes	Yes
Non-verified Information	Yes	Yes	Yes	Yes
Pseudo R^2	0.261	0.094	0.079	0.030
Num. Obs.	98,395	9,722	9,722	9,292

[a]Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Marginal effects and standard errors clustered by state and week are presented.

In Section 3.6.2, we find that the event of labeling institutional investors caused retail investors to bid larger amount but less participating retail investors. These two conflicting forces make the prediction of funding rate ambiguous. Estimates in the first column of Table 3.7 reveals that, after the labeling event, the difference in funding

probability between loans with institutional investment and those without became smaller. However, this effect does not hold across different entry stages of the first institutional investor. The estimates in Table 3.8 shows that if the first institutional investor entered early (less than 10% funds satisfied), the effect was negative, meaning that loans were less likely to be funded. While if the first institutional investment was made later (after 10% funded), the funding rate was higher although the effect was not statistically significant.

Table 3.8: Effects of Labeling Institutional Investors on Transaction Outcomes - by Entry Stage

Treatment Categories	**Marginal Effects**[a]			
1st institutional investment - Percentage funded	1(**Funded**) Probit Est.	**APR** Tobit Est.	1(**Defaulted**) Probit Est.	**ROI** Tobit Est.
$< 10\%$ Funded	-0.029**	3.765***	0.122*	0.016
	(0.012)	(0.918)	(0.064)	(0.087)
$< 50\%$ Funded	0.014	-1.566***	0.023	0.016
	(0.010)	(0.545)	(0.037)	(0.056)
50% More Funded	0.014	-1.020***	-0.028	-0.002
	(0.010)	(0.318)	(0.028)	(0.039)
State and Week FE	Yes	Yes	Yes	Yes
Verified Credit Information	Yes	Yes	Yes	Yes
Discretization of Financial Info.	Yes	Yes	Yes	Yes
Non-verified Information	Yes	Yes	Yes	Yes
Pseudo R^2	0.287	0.095	0.079	0.047
Num. Obs.	98,395	9,722	9,722	9,292

[a]Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Marginal effects and standard errors clustered by state and week are presented.

It is also critical to examine the effect on the final contract interest rate (equivalently APR) for funded loans. Estimates in column 2 of Table 3.7 suggest that, after labeling institutional investors, the difference in the contract interest rate between funded loans with institutional funds and those without dropped more than 50 basis points. It can be seen that this effect also vary with different entry stages of the first institutional investor. When the first institutional investment was made early, the labeling event led to higher APR if loans were funded, holding everything else equal. While if it was made after 10% of funds satisfied, the effect was negative, leading to

lower APR.

Meanwhile, it is important to examine the effect on loans' payment results. All funded loans in our sample have completed their payment cycles. This allows us to investigate the effect on default rate, as well as the overall ROI for lenders. We first look at the effect on the probability of default, which we define as either "Charge-off" or "Defaulted" in the data. Economic theory suggests that contract interest rates are positively correlated with default rate (Stiglitz and Weiss 1981). Our findings are consistent with this prediction. The effects of the labeling event had a negative correlation with the difference in the probability of default for loans with and without institutional funds (column 3 of Table 3.7). Estimates of the effect in Table 3.8 reveal that the labeling event caused the default rate to be higher for the loans with the first institutional investor entering early, *i.e.*, less than 10% funded. On the other hand, the effect was either not significant or negative if the entry was late.

From the investors' points of views, it is probably most important to examine the effect of labeling institutional investors on the return from their investment - ROI. Prosper provided the detailed monthly payment results as well as each and every additional payment borrowers made for most of the personal loans.[6] We are then able to calculate the detailed ROI at each payment cycle, particularly the overall ROI at the end of the payment period. We find that, on average, the labeling event did not have statistically significant effect on the ROI. Estimates in the last columns of Table 3.7 and 3.8 are not signifiant under common significance levels.

3.6.4. Robustness Checks

There are potentially two major concerns that might deteriorate the validity of our empirical results. One concern is that the introduction of institutional investors label and other aspects of the improvement on account infrastructure were associated with

[6] We are able to calculate the ROI for $9,292$ out of $9,722$ personal loans. The rest are discarded due to data incompleteness or absence.

the entry of new institutional lenders on Prosper.com. We observe such a change in Figure 3.2. This raises the question whether the new set of institutional investors had different investment behaviors and thus distinct effects on retail investors and transaction outcomes. Another concern is related to the elimination of interest rate caps on Prosper.com on April 15, 2008 (Rigbi 2013). Recall that our main sample contains the loan requests that were posted between January 2008 and September 2008, which covers the policy change of usury law restrictions. This policy change might potentially contaminate our empirical results as well. In this section, we conduct robustness checks that deal with these two concerns.

We check the first concern about the incoming of new institutional investors by examining the effect of institutional investors that registered and participated before May 19, 2008. More specifically, we focus on a subsample that contains all listings with these "old" institutional lenders as well as those without any institutional investment. In this subsample, there are altogether 25 institutional investors and $93,718$ loan requests. We estimate the effect on both retail investor behaviors and transaction outcomes as in Equation 3.1, 3.2, 3.3, and 3.4. Estimation results are reported in Table 3.9 through 3.12. We do not find significantly different effects compared to those from our main sample. This result is sensible since the new set of institutional investors are not totally different from those registered earlier.

As a robustness check to the second concern about the change in usury law restrictions, we construct another subsample that contains all loan requests originated in states without interest rate cap restrictions before April 15, 2008. Prosper provided detailed usury law restrictions for all states on their website.[7] There are altogether 12 states always setting the highest interest rate caps at 36%. This leads to a subsample of $31,433$ loan requests from our main sample. We then use this subsample to estimate the effects as above. Results are reported in Table 3.13 to 3.16. Again,

[7]Prosper's webpage listing the usury law restrictions, particularly interest rate caps for all US states, is parsed by Archive.org.

Table 3.9: Robustness I - Effects on the Retail Investor Behaviors with "Old" Institutional Investors

	Marginal Effects (OLS Estimates)[a]			
Dep. var.:	**Average bid amount**	**Active retail investors**	**Average time between bids**	**Percent of bids winning**
1 (Labeling institutional investors)	48.395*	-20.535	-2.078	0.019
	(25.152)	(73.107)	(3.404)	(0.093)
State and Week FE	Yes	Yes	Yes	Yes
Verified Credit Information	Yes	Yes	Yes	Yes
Discretization of Financial Info.	Yes	Yes	Yes	Yes
Non-verified Information	Yes	Yes	Yes	Yes
R^2	0.103	0.145	0.208	0.209
Num. Obs.	2,280	2,280	2,280	2,280

[a]Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Marginal effects and standard errors clustered by state and week are presented. This table presents the effect estimates on the behaviors of retail investors for a subsample of loan requests. In this subsample, we only include the listings with institutional investors who had registered prior to May 19, 2008.

Table 3.10: Robustness I - Effects on the Retail Investment Behaviors with "Old" Institutional Investors - by Entry Stage

Treatment Categories	Marginal Effects (OLS Estimates)[a]			
1st institutional investment - Percentage funded	**Average bid amount**	**Active retail investors**	**Average time between bids**	**Percent of bids winning**
< 10% Funded	49.400*	-38.802	4.033	0.025
	(25.958)	(59.445)	(4.501)	(0.098)
< 50% Funded	47.877*	-35.520	-6.221*	0.116
	(25.633)	(59.210)	(3.524)	(0.095)
50% More Funded	52.316**	-18.386	4.364	0.106
	(24.956)	(57.781)	(3.455)	(0.096)
State and Week FE	Yes	Yes	Yes	Yes
Verified Credit Information	Yes	Yes	Yes	Yes
Discretization of Financial Info.	Yes	Yes	Yes	Yes
Non-verified Information	Yes	Yes	Yes	Yes
R^2	0.106	0.305	0.290	0.364
Num. Obs.	2,280	2,280	2,280	2,280

[a]Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Marginal effects and standard errors clustered by state and week are presented. This table presents the effect estimates on the behaviors of retail investors for a subsample of loan requests. In this subsample, we only include the listings with institutional investors who had registered prior to May 19, 2008.

we do not find significant different treatment effects of labeling institutional investors on retail investor behaviors or transaction outcomes.

Table 3.11: Robustness I - Effects on Transaction Outcomes with "Old" Institutional Investors

	Marginal Effects[a]			
Dep. var.:	1(**Funded**) Probit Est.	**APR** Tobit Est.	1(**Defaulted**) Probit Est.	**ROI** Tobit Est.
1 (Labeling inst. investors) × 1 (With inst.)	-0.005	-0.989*	-0.004	0.006
	(0.007)	(0.327)	(0.027)	(0.038)
1 (With inst.)	0.177***	0.879***	-0.014	0.030
	(0.005)	(0.251)	(0.020)	(0.030)
State and Week FE	Yes	Yes	Yes	Yes
Verified Credit Information	Yes	Yes	Yes	Yes
Discretization of Financial Info.	Yes	Yes	Yes	Yes
Non-verified Information	Yes	Yes	Yes	Yes
Pseudo R^2	0.246	0.093	0.086	0.030
Num. Obs.	93,718	8,147	8,147	7,786

[a]Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Marginal effects and standard errors clustered by state and week are presented. This table presents the effect estimates on loan-level transaction outcomes for a subsample of loan requests. In this subsample, we only include the listings with institutional investors who had registered prior to May 19, 2008 and listings without institutional investment.

Table 3.12: Robustness I - Effects on Transaction Outcomes with "Old" Institutional Investors - by Entry Stage

Treatment Categories	**Marginal Effects**[a]			
1st institutional investment - Percentage funded	1(**Funded**) Probit Est.	**APR** Tobit Est.	1(**Defaulted**) Probit Est.	**ROI** Tobit Est.
< 10% Funded	0.040***	2.021**	0.118*	-0.056
	(0.014)	(0.952)	(0.071)	(0.095)
< 50% Funded	0.022*	-2.418***	0.027	0.023
	(0.012)	(0.587)	(0.045)	(0.067)
50% More Funded	0.020	-1.074***	-0.045	0.011
	(0.012)	(0.368)	(0.033)	(0.043)
State and Week FE	Yes	Yes	Yes	Yes
Verified Credit Information	Yes	Yes	Yes	Yes
Discretization of Financial Info.	Yes	Yes	Yes	Yes
Non-verified Information	Yes	Yes	Yes	Yes
Pseudo R^2	0.253	0.094	0.087	0.059
Num. Obs.	93,718	8,147	8,147	7,786

[a]Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Marginal effects and standard errors clustered by state and week are presented. This table presents the effect estimates on loan-level transaction outcomes for a subsample of loan requests. In this subsample, we only include the listings with institutional investors who had registered prior to May 19, 2008 and listings without institutional investment.

Table 3.13: Robustness II - Effects on the Retail Investor Behaviors for States without Usury Law Restrictions

	Marginal Effects (OLS Estimates)[a]			
Dep. var.:	**Average bid amount**	**Active retail investors**	**Average time between bids**	**Percent of bids winning**
1 (Labeling institutional investors)	12.018*	-1.519	-0.671***	0.153***
	(7.100)	(88.180)	(0.072)	(0.053)
State and Week FE	Yes	Yes	Yes	Yes
Verified Credit Information	Yes	Yes	Yes	Yes
Discretization of Financial Info.	Yes	Yes	Yes	Yes
Non-verified Information	Yes	Yes	Yes	Yes
R^2	0.073	0.111	0.188	0.141
Num. Obs.	2,143	2,143	2,143	2,143

[a]Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Marginal effects and standard errors clustered by state and week are presented. This table presents the treatment effect estimates on the behaviors of retail investors for a subsample of loan requests. In this subsample, we only include the listings in the states with no usual law restrictions. In these states, the interest rate cap is uniformly set at 36%.

Table 3.14: Robustness II - Effects on the Retail Investor Behaviors for States without Usury Law Restrictions - by Entry Stage

Categories	Marginal Effects (OLS Estimates)[a]			
1st institutional investment - Percentage funded	**Average bid amount**	**Active retail investors**	**Average time between bids**	**Percent of bids winning**
< 10% Funded	12.220	-71.213*	2.944	0.235***
	(10.514)	(43.358)	(6.097)	(0.035)
< 50% Funded	12.730	-67.422	-8.430**	0.225***
	(7.837)	(41.400)	(3.840)	(0.025)
50% More Funded	4.711	-6.380	-8.313**	0.224***
	(6.995)	(40.698)	(3.768)	(0.022)
State and Week FE	Yes	Yes	Yes	Yes
Verified Credit Information	Yes	Yes	Yes	Yes
Discretization of Financial Info.	Yes	Yes	Yes	Yes
Non-verified Information	Yes	Yes	Yes	Yes
R^2	0.081	0.247	0.258	0.300
Num. Obs.	2,143	2,143	2,143	2,143

[a]Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Marginal effects and standard errors clustered by state and week are presented. This table presents the treatment effect estimates on the behaviors of retail investors for a subsample of loan requests. In this subsample, we only include the listings in the states with no usual law restrictions. In these states, the interest rate cap is uniformly set at 36%.

Table 3.15: Robustness II - Effects on Transaction Outcomes for States without Usury Law Restrictions

	Marginal Effects[a]			
Dep. var.:	1(**Funded**) Probit Est.	**APR** Tobit Est.	1(**Defaulted**) Probit Est.	**ROI** Tobit Est.
1 (Labeling inst. investors) × 1 (With inst.)	-0.019	-0.493	-0.031	0.003
	(0.012)	(0.479)	(0.039)	(0.060)
1 (With inst.)	0.205***	1.156***	0.028	0.044
	(0.010)	(0.396)	(0.032)	(0.048)
State and Week FE	Yes	Yes	Yes	Yes
Verified Credit Information	Yes	Yes	Yes	Yes
Discretization of Financial Info.	Yes	Yes	Yes	Yes
Non-verified Information	Yes	Yes	Yes	Yes
Pseudo R^2	0.250	0.094	0.086	0.040
Num. Obs.	31,433	3,626	3,626	3,476

[a]Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Marginal effects and standard errors clustered by state and week are presented. This table presents the treatment effect estimates on loan-level transaction outcomes for a subsample of loan requests. In this subsample, we only include the listings in the states with no usual law restrictions. In these states, the interest rate cap is uniformly set at 36%.

Table 3.16: Robustness II - Effects on Transaction Outcomes for States without Usury Law Restrictions - by Entry Stage of the First Institutional Investment

Categories	**Marginal Effects**[a]			
1st institutional investment - Percentage funded	1(**Funded**) Probit Est.	**APR** Tobit Est.	1(**Defaulted**) Probit Est.	**ROI** Tobit Est.
< 10% Funded	-0.020	2.463*	0.030	0.177
	(0.020)	(1.467)	(0.097)	(0.158)
< 50% Funded	0.047***	-0.978	0.038	0.065
	(0.016)	(0.909)	(0.057)	(0.088)
50% More Funded	0.011	-1.042**	0.024	-0.054
	(0.019)	(0.479)	(0.044)	(0.064)
State and Week FE	Yes	Yes	Yes	Yes
Verified Credit Information	Yes	Yes	Yes	Yes
Discretization of Financial Info.	Yes	Yes	Yes	Yes
Non-verified Information	Yes	Yes	Yes	Yes
Pseudo R^2	0.275	0.096	0.091	0.080
Num. Obs.	31,433	3,626	3,626	3,476

[a]Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Marginal effects and standard errors clustered by state and week are presented. This table presents the treatment effect estimates on loan-level transaction outcomes for a subsample of loan requests. In this subsample, we only include the listings in the states with no usual law restrictions. In these states, the interest rate cap is uniformly set at 36%.

3.7. Conclusions

Although institutional investors are well studied in traditional financial markets, there is surprisingly little research on the interactions between them and retail investors. One reason is that retail investors are very few in traditional markets. The nascent online consumer financial market, debt-based crowdfunding, provides a unique research opportunity to study the interactions. We exploit a "natural experiment" with the identity of institutional investors on Prosper.com to answer the research questions.

Although we do not find supporting evidence of the effects on institutional investors' behaviors and investment performance, we are able to see significant effects of labeling institutional investors on both retail investor behaviors and loan-level transaction outcomes. Specifically, we observe that retail investors were crowded out by the participation of institutional investors in the first place; but once participating, their activeness in terms of bidding amount and how quick they submit bids was significantly greater. For transaction outcomes such as funding probability and APR, we find heterogeneous effects across risk levels and treatment groups.

Our paper is among the first to systematically investigate the interactions between institutional and retail investors in the context of debt-based crowdfunding. The behavior of market participants is fundamental to transaction outcomes and market efficiency. Institutional and retail investors are widely existing in all types of financial markets. Thus our results have both policy and managerial implications for other types of financial markets as well, particularly other types of crowdfunding. Especially with the burgeoning of debt-based and reward-based crowdfunding and the upcoming equity-based crowdfunding in the US, this paper has important implications for market makers as well as policy makers.

4. For Whom to Tweet? A Study of a Large-Scale Social Network

Zaiyan Wei and Mo Xiao

4.1. Introduction

Social networks permeate our social and economic lives. They play a key role in many aspects, including consumption (Moretti 2011), employment (Calvó-Armengol and Jackson 2004), employee productivity (Aral, Brynjolfsson, and Van Alstyne 2012, Wu 2013), and technology adoption (Conley and Udry 2010). Online social networking, provided by platforms such as Facebook.com and Twitter.com, facilitates one's social lives and has become a huge phenomenon in the last few years. It is widely adopted by users with different age, gender, income, or from extensive ethnicity groups. Surveys reveal that as of January 2014, 74% of adult Internet users have registered at least one social networking site (SNS).[1] As the largest social networking platform, Facebook.com has attracted 1.18 billion registered users by January 2014. In China, 46% of its population, or more than 600 million people, use social networking service.

Users of social networking service not only form connections online, but also (perhaps more importantly) engage in online postings routinely. In Twitter-type networks, users can *tweet* (writing short postings with texts and/or multimedia contents) or *retweet* others' tweets. Online tweeting, together with other forms of information generation and transmission, are playing an increasingly important role in our society. An example is that the social media usage, Twitter being an example, is shaping the workings of American democracy (Chi and Yang 2011, Peterson 2012, Halberstam and Knight 2014). Individual tweeting can be viewed as providing public goods, since

[1]PewResearch Internet Project annual reports from online surveys can be found at http://www.pewinternet.org/fact-sheets/. Annual survey by "We Are Social" reports the statistics of online social networks: http://wearesocial.net/.

in most Twitter-type SNS, one's tweets are free to access within the community.

When making tweeting decisions, *e.g.*, whether and how much to tweet, individuals are influenced by their "neighbors" (Aral and Walker 2012). Particularly, the sizes of peer groups have prominent effects on the provision of the public good - tweeting in the current context (Bulow and Klemperer 1996, Zhang and Zhu 2011). Different from the structure of bilateral friendship that is prevalent on platforms such as Facebook.com, on Twitter-type SNS, an individual's reference groups include followees (users whom she follows) and followers (users who follow her). On one hand, a user can read all her followees' tweets, therefore her major information source in the network; on the other hand, her tweets can be viewed by all followers, and hence are significantly shaped by this group as well.[2]

The effects of either the followees count or the followers count on one's tweeting are not *a priori* clear-cut, due to her conflicting incentives of free-riding and social effects. In these directed networks, the information transmission has certain directions that the information embedded in tweets is mainly passed from an individual to her followers. On one side, an individual may free ride on others' (including followees' and followers') contributions and tweet *less* with larger population of neighbors. For instance, with an exogenous increase in the population of Weibo users, such as the merge of two large-scale networks, it is likely that she starts following some newly registered users and free rides on their contributions. Similarly, some newcomers may start following her. Since the individual can observe all her followers' tweets, she has incentives to tweet less thereafter.[3] In short, the free-riding incentive indicates negative effects of both followees count and followers count. On the flip side, the social

[2]Although one's tweets appear automatically on all her followers' home pages, her followers' tweets do not necessarily show up on her page. Unless a follower is simultaneously her followee or she clicks on the follower's tweets page intentionally, she does not receive updates automatically.

[3]In addition, the negative effect of followers count may reflect an individual's privacy concern. With more users following her, particularly random accounts in the network, she would be more cautious about tweeting to the public. Therefore she is more selective of what to tweet, and would not tweet as much as she used to.

effects predict that the population size of both followees and followers have *positive* effects on one's tweeting. Specifically, first of all, an individual's followees count has positive effect on her tweeting, merely because the amount of information acquired is positively correlated with the number of followees. Analogously, she may tweet more with more followers, because she gets more social benefits (*e.g.*, warm glow or other altruistic incentives that increase her utility) from larger population of audience (Andreoni 1989, Andreoni and Bernheim 2009), or merely has higher motivations to express herself.

In the current research, we empirically examine the effects of individuals' followees and followers on their contributions to the public good - tweeting - on Tencent Weibo. Tencent Weibo is often viewed as the Twitter counterpart in China.[4] By 2014, there have been over 220 million registered users on Tencent Weibo, making it the largest Twitter-type SNS in China. In this large-scale online network, we observe that, within our samples approximately 10% of the users contribute more than 50% of the tweets in stock, while a half of them provide less than 15% tweets. Similarly, 50% of the users post less than six tweets monthly, while a user at the 90th percentile generates on average 40 tweets a month. We attribute the highly skewed distribution and scarcity of tweets to the users' conflicting incentives of free-riding and maximizing social benefits (or social effects) as summarized before. We explore whether we can distinguish these two conflicting incentives by exploiting the asymmetry in the effects of one's followees count and followers count.

The main challenge for the identification of the neighborhood size effects is the endogeneity of individuals' peer groups in social networks (Manski 1993, Bramoullé, Djebbari, and Fortin 2009, Goldsmith-Pinkham and Imbens 2013). In our context,

[4]The friendship rules and user activities on Tencent Weibo are largely the same as Twitter.First of all, similar to the network structure on Twitter, users of Tencent Weibo can unilaterally form connections with other users by simply following their tweets. Also analogous to activities on Twitter, Weibo users can tweet (less than 140 Chinese characters with optional attachment such as videos and pictures) and retweet any other user's tweets (not restricted to those of followees). The website (in Chinese) can be found at http://t.qq.com/.

it is a particular concern that the unilateral following relationships are established by individuals who deliberately choose to connect. Econometrically, the specific concern is that there exist individual unobserved characteristics correlated with network formations that in turn determine the network sizes, and that these characteristics simultaneously determine her tweeting behavior (Goldsmith-Pinkham and Imbens 2013). For instance, individuals' personal traits are unobserved factors that are correlated with both the tweeting intensity and the number of others she will follow, such as how socially active she is. If that is the case, researchers will find that the tweeting outcomes are correlated with the number of followees even without peer size effects.

We address the endogeneity problem by devising different identification strategies for two datasets from Tencent.com. The first dataset we obtain is a proprietary longitudinal dataset containing observations of a sample of Weibo users over a four-month period between August 2011 and December 2011. To approach the endogeneity problem, we first consider a panel data method that incorporates individual and time fixed effects, to control for the unobserved individual heterogeneity that is constant over time, as well as other unobservables that affect all individuals but differ in time.

In our panel data analysis, some concerns on endogeneity are alleviated by including individual and time fixed effects; however, there might still exist correlations between unobservables and group sizes. To further alleviate such concerns, we obtain a second dataset, which is a cross-sectional sample with bigger network size and more complete network structure. The dataset contains the snapshot of a random sample of Weibo users as of a date in January 2012, with their activities and all followees and followers identities. We therefore consider a second approach - the instrumental variable (IV) method - with our cross-sectional data. We propose to use the average observed characteristics (including age and gender) of one's second-order followees and those of second-order followers as instruments for all endogenous variables. Our instruments are relevant, because they are related to an individual's neighborhood

sizes by influencing her immediate neighbors' tweeting activities, which directly affect her utilities of forming connections. Under some assumptions we explicitly list, our instruments do not directly influence, and are not correlated with unobservables determining, one's tweeting.

We find consistent results from both longitudinal and cross-sectional analysis. Specifically, our first set of results suggest that a user is more likely to tweet with more followers, but less likely to tweet if she is following more people. Similarly, further findings from both panel data approach and cross-sectional analysis suggest that a greater number of followees leads to fewer tweets while more followers cause individuals to tweet more.[5] These asymmetric effects, particularly the positive impact of followers count, lend support to the hypothesis that individuals obtain benefits or utilities from their private contributions, which is formulated as social effects in the literature (Andreoni 1989, 1990, Zhang and Zhu 2011). From the platform's point of view, our results imply that user recommendations for others to follow will be an effective tool to promote more active contributions in the network. Based on our estimates, a user with 2 followers (the median of followers count distribution) will tweet 65.3% more with one extra follower. A simulation exercise suggests that the randomly generated 1% new connections will increase the total tweets by 25% as the estimated follower effects dominate the followee effects.

We first contribute to the literature on social networks by studying asymmetric influences of followees and followers in directed networks on individuals' online activities (Aral and Walker 2011, Chi and Yang 2011, Bond et al. 2012). To our knowledge, the paper is among the first to document the asymmetric peer effects in directed social networks (see Borgatti et al. (2009) for a thorough review of this literature), and helps us better understand how people tweet and engage in such networks. In a recent

[5]In addition, we observe biases by not including individual and time fixed effects in the case of panel data, or not instrumenting for endogenous variables in cross-sectional analyses. We discuss the potential causes of these biases later in the paper.

investigation, Toubia and Stephen (2013) conduct field experiments to explore the effects of the followers count on individuals' incentives to contribute on Twitter.com. We take a step further by showing that not only the number of followers influences individuals' tweeting decisions, the followees count has significant effects. Similar to Toubia and Stephen (2013), we confirm the positive effects of the followers count. We add to the literature by documenting a negative effects of the number of followees. We contribute to a second strand of literature about group size effects on the provision of public goods in social networks (Olson 1965, Andreoni 2007, Bulow and Klemperer 1996, Chen et al. 2010, Zhang and Zhu 2011). We first add to the literature by providing evidence from a well-defined large-scale online social network. More importantly, following recent development in this literature that supports the social effect hypothesis (*i.e.*, contributors not only receive utilities from the total volume of public goods provision but also gain social benefits from private contributions), our results further suggest that when modeling people's behaviors in social networks (Bulow and Klemperer 1996), the distinct effects of followees and followers should be considered.

The paper is structured as follows. We introduce our research context in Section 4.2. Section 4.3 describes our data and samples used in estimations. We then present our empirical strategies and results in Section 4.4. We start with the longitudinal analysis in subsection 4.4.1, and continue to cross-sectional analysis in subsection 4.4.2. We perform discussions based on our results subsequently, and Section 4.5 concludes.

4.2. Tencent Weibo

Tencent Weibo is a micro-blogging platform, mainly in Chinese, providing Twitter-type social networking service. It was launched on April 1, 2010 as an affiliated

platform to Tencent.com.[6] By January 2014, there had been over 220 million registered users on Tencent Weibo, and the average number of daily active users had reached a peak of 100 million.

Users of Tencent Weibo can either "tweet" or "retweet" other users' tweets. We characterize tweeting as providing public goods, since an individual's tweets are free to access within the community. In Tencent Weibo case, an individual's tweets or retweets can be viewed by all registered Weibo users, not restricted to her followers.[7] The maximum length of a tweet is 140 words or Chinese characters. Tweets can be attached with multimedia content such as pictures and videos. A retweet is a re-posting of others' tweets, also constrained by the maximum length of texts.

In addition to those two main activities, Weibo users can communicate with others by, first of all, "mentioning" them in a tweet or retweet. Specifically, a user can notify others, not necessarily her followees or followers, of some tweets (or "comments") by typing their user names following the "@" symbol directly in the texts. Alternatively, any user can post comments below a tweet unless the setting is customized. By default the comments, together with the tweets, will be listed publicly on Weibo pages. Any registered Weibo user will be able to view the content. Table 4.1 summarizes these activities and compares with those on Twitter.

In addition to providing the micro-blogging service summarized above, Weibo users form a *directed* social network that is connected by "following" relationships. A Weibo user can choose to follow another user's tweets without the other party's consent unless the setting is changed. As mentioned, we define the user as the "follower", and the other party the "followee". This connection can be established unilaterally,

[6]Tencent is the largest networking service provider in Mainland China. The major services they provide include instant messaging, personal space, and micro-blogging. Tencent's flagship products are QQ and WeChat. By the end of 2013, there had been over 0.8 billion registered users on Tencent; they had reached a peak of 0.18 billion users online simultaneously. For more information about Tencent, visit their official website at http://www.tencent.com/en-us/index.shtml.

[7]This is slightly different from the setting on Twitter.com. On Twitter, a user has the option to make her tweets private and available to followers only. In contrast, Weibo users do not have a similar privacy setting, and all their tweets are accessible by any Weibo users.

Table 4.1: A Comparison of User Activities on Tencent Weibo versus Twitter

User Behaviors		Tencent Weibo	Twitter
Tweets -	Texts of less than 140 words, with optional attachment such as pictures and videos.	Yes[a]	Yes
Retweets -	Re-post tweets by others.	Yes	Yes
Mentions -	Notifying others of certain tweets (or comments) by typing their user names following the "@" symbol.	Yes	Yes
Comments -	"Commenting" on any tweet.	Public[b]	Private

[a]If the texts are in Chinese, the restriction will be 140 characters instead of words.

[b]Twitter users can also comment on a tweet by "replying" to it. However, unlike the public comments on Weibo, the replies on Twitter are kept private between the communicating parties.

meaning that the followee does not have to follow the follower's tweets. Therefore, the Weibo network is directed. With this direction feature, followers of a user are also called her "audience". Once a following relationship is established, a follower can read all her followee's tweets. This Weibo network is identical to that on Twitter, while different from other online social networks such as Facebook and LinkedIn, where friendship cannot be established without mutual consent of both parties, and is thus undirected.

We focus on the directed network structure and individuals' tweeting activities on Tencent Weibo in this research. Starting from the next section, we introduce our data, and in Section 4.4 discuss our empirical strategies to identify the neighborhood size effects on individuals' tweeting behavior.

4.3. Data and Samples

We obtain two datasets from Tencent.com. The first one is a proprietary dataset that consists of 29,956 Weibo users with their Tencent identities,[8] tweeting, and networking information between August 2011 and December 2011. Specifically, Tencent kept track of these users' number of followees and followers (but not their identities), the number of tweets, retweets, and other activities such as comments and private messages. Demographic information such as user location and job information verification status are also included.

The specific sampling process of this longitudinal dataset is as follows. A sample of active users who registered Weibo before August 29, 2011 was selected on January 4, 2012. Tencent reports their records at six different dates between August 15, 2011 and December 11, 2011. From this dataset, we construct a sample of 20,289 users with complete records at each of the six observation dates, *i.e.*, constituting a balanced panel. This is the main sample used in our estimations.[9] Table 4.2 summarizes this longitudinal sample.

We obtain another dataset from Tencent.com. The firm publishes a random sample of 1,392,873 users from their Weibo user pool on a webpage.[10] For these users, Tencent provides their networks, tweeting, and demographic information up to a date in January 2012 (not explicitly mentioned by the firm). Particularly, the dataset contains the *coded* identities (different from their "QQ numbers" - users' unique identities on Tencent) of these users' followees and followers. For all original 1,392,873 users

[8]Each user has a unique identity on Tencent. The identity is a number with 5 to 11 digits. Users can use any Tencent service with this identity, called "QQ Number."

[9]We focus on the balanced panel mainly because the data provider failed to collect the missing records, but not because those records do not exist or are missing randomly. We conduct a robustness check using the original sample of 29,985 users.

[10]Tencent hosted the 2012 Knowledge Discovery and Data Mining (KDD) Cup competition in 2012. This is an annual global competition of data mining that targets at all data scientists all over the globe. The competition website is https://www.kddcup2012.org/. In one of the two tracks, Tencent made friendship connecting recommendations to the 1,392,873 Weibo users, and kept track of their decisions over a period of time. The task for participants was to develop statistical models to predict users' decisions of accepting or rejecting the recommendations.

and their immediate followees and followers, the dataset contains all their number of tweets, retweets, other activities such as comments, and demographic information including age and gender. Compared to the panel data, this cross-sectional dataset documents a larger network size and more complete network structure with identities of all followees and followers, although it does not keep track of higher-order neighbors.

From the original random sample, we construct a set of $402,470$ users[11] as the main sample in our estimations. These users have at least one second-order followee and one second-order follower. As mentioned in the introduction, the instruments we propose for the endogenous neighborhood sizes are the average observed characteristics of second-order neighbors, which is the reason why we focus on this subset of users. Table (4.3) summarizes this cross-sectional sample.

4.4. Empirical Strategies and Results

Those two complementary datasets offer us a unique opportunity to investigate the causal peer effects on an individual's contributions to the public good, Weibo tweets in the current context. A major challenge of causality is that an individual's network or peer group - followees and followers on Weibo - is generally endogenous (Manski 1993, Bramoullé, Djebbari, and Fortin 2009, Goldsmith-Pinkham and Imbens 2013). The specific concern is that individuals have unobserved characteristics that are associated with their tweeting behavior, and also correlated with their decisions to form links (determining the number of followees) and others' decisions to follow (determining the number of followers). We propose two methods to approach this endogeneity issue. One is based on a panel data method, and the other an instrumental variable approach.

[11] In Appendix C.1 we compare this constructed sample with the original dataset consisting of $1,392,873$ users. We find that this sample is representative of the original dataset in terms of demographic information.

In the first subsection, we present our panel data approach and report estimation results. We discuss additional concerns of endogeneity that survive this method. Then we move on to our IV approach in Section 4.4.2.

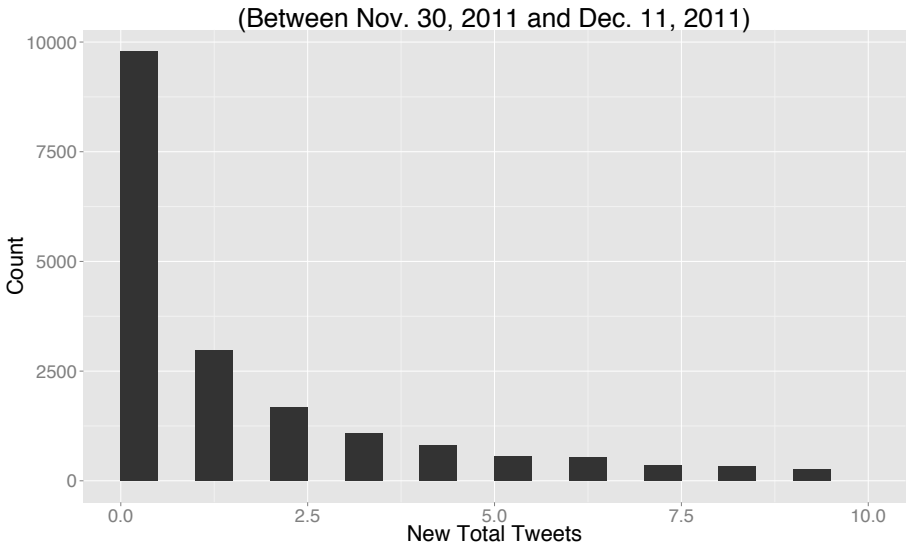
4.4.1. Panel Data Approach and Results

Unobserved individual heterogeneity that is constant over time is an important source of endogeneity in social networks. Personality, as an example, significantly influences one's tweeting behavior, and determines her peer groups simultaneously. Besides this, exogenous shocks that affect all Weibo users, *e.g.*, the introduction of a new private messaging tool that is available to all Weibo users, are another cause of correlations between neighborhood sizes and the error term. To account for these two endogeneity sources, we adopt a panel data method that incorporates both individual and time fixed effects.

Empirical Strategy With the longitudinal data structure we explore the effects of the number of followees and that of followers at date $t-1$ on an individual's total number of tweets between date $t-1$ and t. A particular data phenomenon is worth noting prior to presenting our empirical models. In the longitudinal sample obtained from Tencent, we observe that, at each cross section, there exists a significant fraction of users who did not have any new tweets since the last period. For instance, $9,788$ out of $20,289$ users had no new tweets between November 30, 2011 and December 11, 2011. Figure 4.1 shows the fractions of individuals with no new tweets in this period.

This pattern indicates that one's propensity to tweet is left censored at zero. A user may have great disincentives to tweet, but all we observe is that she has not tweeted in a period of time. To deal with this dichotomy that governs one's tweeting behavior, we consider two tweeting decisions facing each Weibo use: 1) the decision whether to tweet; 2) if an individual decides to tweet, how much she would tweet. Thus, we first study the effects of the followees count and the followers count on the

Figure 4.1: Panel Data - Distribution of Users with respect to the New Total Tweets



decisions to tweet or not. In addition, with skewed distributions of the followees count and the followers count (Table 4.2 shows that they are both skewed to the right.), we take log-transformations of these variables. Our main empirical specification is,[12]

$$1\left(\Delta Y_{it} > 0\right) = \beta_1 \cdot \log\left(N^e_{i,t-1} + 1\right) + \beta_2 \cdot \log\left(N^r_{i,t-1} + 1\right) + \mathbf{X}'_{it}\beta_3 + \mu_i + \nu_t + \epsilon_{it}, \quad (4.1)$$

where ΔY_{it} is the user i's total number of tweets (including both tweets and retweets) between date $t-1$ and t; $N^e_{i,t-1}$ and $N^r_{i,t-1}$ are her number of followees and that of followers up to date $t-1$; $\mathbf{X}_{it}$ includes the number of days between date $t-1$ and t; μ_i are individual user fixed effects, and ν_t the time fixed effects; ϵ_{it} is the idiosyncratic error term.

Given a user decides to tweet ($\Delta Y_{it} > 0$), we further explore the effects of her

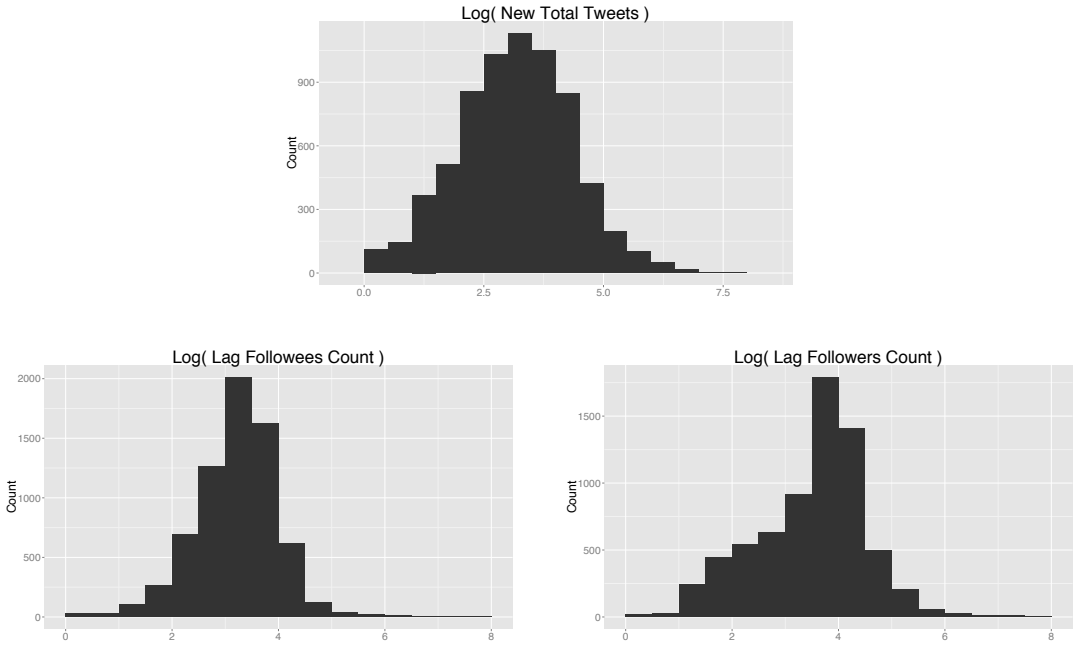
[12]In principle, we could estimate nonlinear models, such as Poisson regressions, to explore the nature of tweets as count data. However, the data suggests that the group of users that did not tweet in each period were very different from those who tweeted. For example, we find that the "silent" individuals had significantly smaller number of total tweets in stock, less followees, and less followers. Thus, we explore the two decisions facing each individual, which corresponds the hurdle model that is also standard in the literature on count data (Duan et al. 1983). In addition, given the large numbers of zeros observed in the data, nonlinear models such as a Poisson specification may not work well (Cameron and Trivedi 2005).

neighborhood sizes on how much she would contribute to the public good - tweets - in the network. Table 4.2 shows that the distribution of ΔY_{it} is also skewed. Similarly, we take log-transformation of this variable. As an illustration, Figure 4.2 displays the distributions of of these variables up to September 15, 2011 after log-transformations. Similar patterns are observed at other dates. Our main regression equation is,

$$\log(\Delta Y_{it}) = \beta_1 \cdot \log\left(N^e_{i,t-1} + 1\right) + \beta_2 \cdot \log\left(N^r_{i,t-1} + 1\right) + \mathbf{X}'_{it}\beta_3 + \mu_i + \nu_t + \epsilon_{it}, \quad (4.2)$$

In both Equation (4.1) and (4.2), estimates of β_1 and β_2 will be the effects of followees count and followers count respectively.

Figure 4.2: Panel Data - Distributions of Log-Transformations for Key Variables on September 15, 2011



Results The main estimation results are reported in Table 4.5 and 4.6. Estimates of our main specifications, Equation (4.1) and (4.2), are presented in the fourth column

of both tables.[13] The results suggest that, in general, the number of followees in the previous period has negative effects on both decisions whether and how much to tweet; while the lag number of followers has a positive effect on both decisions.

More specifically, for the neighborhood size effects on whether to tweet, the results suggest that with a 1% increase in the number of followees at the last stage, on average the probability of an individual tweeting during the current period will be around 2.5% less (Table 4.5). In contrast, an increase in the lag number of followers increases the probability of tweeting (Table 4.5). The neighborhood size effects on the quantity of tweets exhibit the same pattern. Specifically, a 1% increase in the number of followees reduces, on average, 0.08% of tweets up to the next observation date (Table 4.6). But a 1% increase in the number of followers tends to increase the number of tweets by 0.5 percentage point (Table 4.6). Note that if we do not include individual or time fixed effects, both the number of followees and that of followers are positively related to both decisions whether and how much to tweet (by comparing Spec 1 with Spec 4 in both Table 4.5 and 4.6). This illustrates the potential bias due to the endogeneity of one's networks or peer groups.

There are several potential concerns about the main specification and the sample used in estimations. In subsection 4.4.3, we perform several checks to demonstrate that these results are robust using alternative specifications or with different samples, that is, with the panel data we observe negative effects of the followees count and positive effects of the followers count on both decisions whether and how much to tweet.

As briefly mentioned, some concerns on the endogeneity of group sizes are alleviated by including individual and time fixed effects, such as the correlation between

[13]The simultaneity between tweeting and network sizes causes concerns about serial correlations in error terms. For all fixed effects regressions in our panel data analysis, we carry out tests for serial correlations, which are suggested by Wooldridge (2002). These tests all strongly reject the null hypothesis of no serial correlations. This suggests that we should be careful in inferences. For example, we report the standard errors clustering at individual user level. In addition, robust standard errors and those from bootstrapping are generally smaller.

the endogenous variables and individual unobserved characteristics that are constant over time. However, there might still exist other types of correlations between the error term and endogenous variables. As an illustration, the event of an individual's marriage during our study period has long lasting effects on both network connection decisions and tweeting behavior, which cannot be captured by either individual or time fixed effects. Thus we need more evidence from our cross-sectional data, which allows us to use an instrumental variable approach.

4.4.2. Cross-Sectional Data and Results

We continue our discussions on the endogeneity of peer groups and motivate our instrumental variable approach. The model prevalent in peer effects literature is the linear-in-means model (Manski 1993, Bramoullé, Djebbari, and Fortin 2009, Goldsmith-Pinkham and Imbens 2013). We combine this model adapted to our situation with a strategic network formation process for our purpose. Consider first, an individual's tweeting equation on Tencent Weibo has the linear-in-means form,

$$\log(Y_i) = \beta_0 + \beta_1 \cdot \log(N_i^e) + \beta_2 \cdot \log(N_i^r) + \beta_{\bar{Y}} \cdot \bar{Y}_{(i)} + \bar{\mathbf{X}}_i' \beta_{\bar{\mathbf{X}}} + \mathbf{X}_i' \beta_3 + \epsilon_i, \quad (4.3)$$

where Y_i is user i's number of tweets; N_i^e and N_i^r are her number of followees and followers respectively; $\mathbf{X}_i$ is a vector of her observed characteristics including age and gender; $\bar{Y}_{(i)}$ is the average quantity of tweets by all user i's followees and followers, and $\bar{\mathbf{X}}_i'$ are the means of observed characteristics of all her followees and followers. Particularly, the error term ϵ_i captures all remained factors that affects one's tweeting, such as user i's unobserved (to econometricians) characteristics and idiosyncratic shocks. Let $\mathbf{X} = [\mathbf{X}_i]$ be the observed characteristics of all users, and ϵ the vector of all user' error terms that capture their unobservables. We assume that they are independent. Specifically,

Assumption 2. The observed characteristics are orthogonal to unobserved characteristics, *i.e.*, $\mathbf{X} \perp \epsilon$.

We are mainly interested in the effects of group sizes, the number of followees and followers. The corresponding coefficients measure how an individual's tweeting behavior reacts to different sizes of neighborhood. A major concern of causality arises due to the endogeneity of one's peer groups, for both followees and followers in our context. Under our current context, two issues are most relevant: simultaneity problem (Manski 1993) and unobserved heterogeneity (Goldsmith-Pinkham and Imbens 2013). For the first aspect, an individual's tweeting behavior not only is affected by the size of her neighborhood, but also determines the number of other users that form connections with her. For the other, one's unobservables affect her tweeting behavior as well as network formation decisions. Since a user's peer groups are endogenous, in Equation (4.3) N_i^e, N_i^r, $\bar{Y}_{(i)}$, and $\bar{\mathbf{X}}_i$ are all endogenous variables. Econometrically, these endogenous variables are correlated with the error term ϵ_i.

We propose to use a set of instruments (Berry, Levinsohn, and Pakes 1995) to approach the endogeneity issue that is due to the unobserved heterogeneity. Specifically, the IVs are the average observed characteristics, including age and gender, of an individual's second-order followees and second-order followers that are otherwise not her immediate neighbors.

We will get the ideas for our IV strategy from a network formation process. Suppose user i is determining whether to follow another user j. Many (but not restricted to) factors may influence her connection decision. First, since her followees' tweets are the major source of information on Tencent Weibo, user i cares about what and how much j tweets. She may also examine whether j has similar (both observed and unobserved) characteristics. In social networks, an individual has the tendency to follow others that are similar (the phenomenon of homophily, McPherson,Smith-Lovin, and Cook (2001)). On top of these, idiosyncratic shocks may also affect user i's utility of following j. For example, the occurrence of a natural disaster makes a user more likely to follow news reporters or charity organizations. We thus formalize

i's utility of following j as, for any $j \neq i$,

$$u_{ij} = f\left(Y_i, Y_j; \mathbf{X}_i, \mathbf{X}_j; \epsilon_i, \epsilon_j\right) + \eta_{ij}, \tag{4.4}$$

where $f(\cdot)$ is a function of the tweets quantity and all characteristics for both i and j; η_{ij} is the idiosyncratic error term representing shocks that are i-j pair specific.

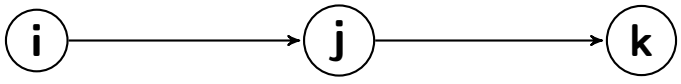
By our earlier discussions, the function, $f(\cdot)$, is flexible enough to capture all factors related to its coordinates. As examples, the Euclidean distances between characteristics, *e.g.*, $d(\mathbf{X}_i, \mathbf{X}_j)$ and $d(\epsilon_i, \epsilon_j)$, are negatively related to i's utility of following j (homophily); the unobservables may have heterogeneous effects across age or gender, thus the interaction terms $\mathbf{X}_i\epsilon_j$ and $\mathbf{X}_j\epsilon_i$ may enter $f(\cdot)$ as well. Given i's utility of following j, her decision to follow j obeys a simple rule that,

$$\begin{cases} i \text{ follows } j, & \text{if } u_{ij} \geq 0; \\ i \text{ does not follow } j, & \text{if } u_{ij} < 0. \end{cases}$$

As a consequence, these network formation processes lead to variations in peer effect variables in Equation (4.3) including N_i^e, N_i^r, $\bar{Y}_{(i)}$, and $\bar{\mathbf{X}}_i$.

Validity of Instruments Based on the model in the last section, we discuss the validity of our instruments. Figure 4.3 is a parsimonious social network with only three users. Our discussions will be illustrated by this example. Specifically, suppose there are three Weibo users, i, j, and k, and two edges (following relations) connecting them. The arrows in Figure 4.3 show the following directions, *e.g.*, user i is a follower of user j and thus j is i's followee.

Figure 4.3: A Simple Network with Three Users and Two Edges



Relevance: We begin with discussing how our instruments, the average observed characteristics of one's second-order neighbors, are correlated with the endogenous

peer effect variables. The first set of instruments contains all the average characteristics of second-order followees. In Figure 4.3, k's characteristics will be used as instruments for i's quantity of followees. To see how they are relevant, they first enter user j's tweeting equation (Equation (4.3) for j) since k is j's immediate followee. This implies that the instruments has direct effects on j's actions, tweeting, which in turn determines i's utility of following j since Y_j enters Equation (4.4) for i. In words, variations in the average observed characteristics of second-order followees (our IVs) lead to changes in an individual's utilities of following immediate followees, by changing the first-order followees' tweeting actions, which governs the number of followees one might have (N_i^e).

We then consider the average observed characteristics of second-order followers as being correlated with the quantity of one's followers. In Figure 4.3, now we focus on using i's observed characteristics as instruments for k's number of followers. By analogy to our analysis above, the instruments enter k's potential follower j's tweeting equation, thus have direct effects on j's tweeting action. In turn, when j is making the connecting decision to k, her own tweeting determine the utility directly. As a result, this process determines whether j will follow k. The consequence of similar processes, from k's all potential followers, is the variation in the number of followers k might have (N_k^r). Similarly, variations in the average observed characteristics of second-order followers lead to changes in the immediate followers' tweeting actions that, in turn, affect the follower group a user may have.

Correlations between the average observed characteristics of an individual's second-order neighbors and the tweets and characteristics of her immediate neighbors ($\bar{Y}_{(i)}$ and $\bar{\mathbf{X}}_i$ in Equation (4.3)) are easier to establish, since these two parties are mutually first-order neighbors.

Exclusion restriction: Clearly, the validity of our IV strategy requires a certain degree of shortsightedness when a user is tweeting. Specifically, we require the assumption that the users are not perfectly forward-looking, in the sense that when

tweeting they are only affected by the characteristics of immediate followees and followers, but not of neighbors' neighbors (implicitly assumed in Equation (4.3)). In other words, information about second-order neighbors is simply too much or too noisy to influence one's tweeting decisions. As a result, the average observed characteristics of second-order neighbors do not enter an individual's tweeting equation - Equation (4.3).

In addition, we also assume the shortsightedness when an individual is making network formation decisions. As implicitly assumed in Equation (4.4), the characteristics of neighbors' neighbors (second-order followees or followers) are excluded from the utility function. By similar arguments, the information about the indirect connections is hardly affecting one's connection decisions.

Exogeneity: It is left to show that our instruments are exogenous, *i.e.*, the average observed characteristics of second-order neighbors are uncorrelated with the error term ϵ_i in Equation (4.3). It turns out that Assumption 2 is a necessary condition for the exogeneity. To motivate this assumption, consider ϵ being the individual idiosyncratic component influencing her network formation utilities. It captures all remaining effects that cannot be explained by one's observed characteristics, $\mathbf{X}$. In other words, the observed component, $\mathbf{X}$ in Equation (4.3), captures all effects on tweeting that are related to users' observed characteristics; while ϵ picks up everything left out. It is often assumed in the literature (Berry, Levinsohn, and Pakes 1995, Xiao 2010) that the observed and unobserved components capture separate effects on outcomes, thus are orthogonal to each other.

Relaxing Assumption 2 makes our instruments, the average observed characteristics of second-order neighbors, correlated with one's unobservables, because our instruments are first correlated with immediate neighbors' characteristics (both observed and unobserved) that are in turn related to one's unobserved characteristics. Take Figure 4.3 as an example, we first consider $\mathbf{X}_k$ as instruments for i's number of followees. Suppose Assumption 2 does not hold, $\mathbf{X}_k$ will be first correlated with

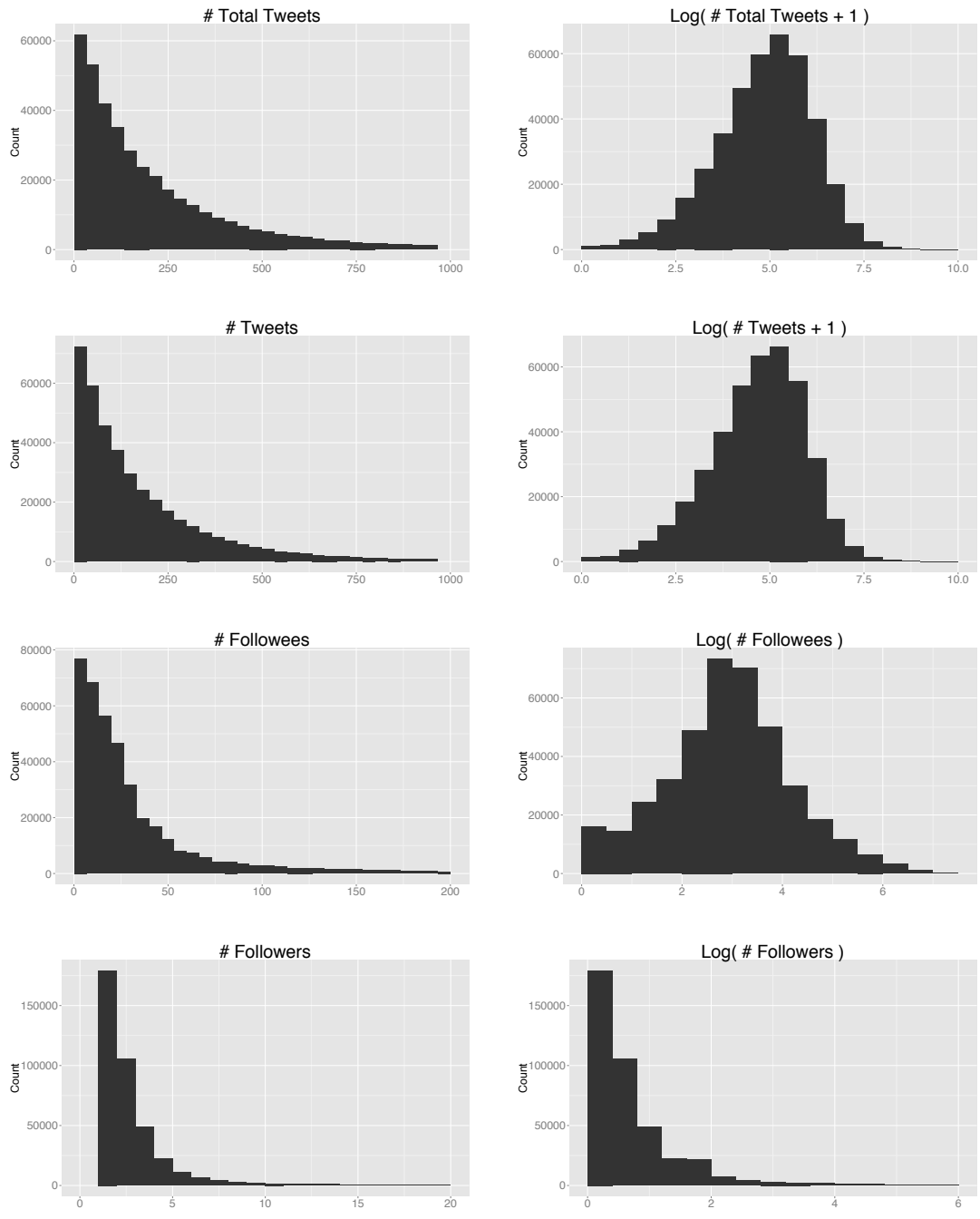
$\mathbf{X}_j$ and ϵ_j governed by j's decision to follow k (Equation (4.4) for j). Furthermore, in some cases (homophily) ϵ_j is fairly similar to ϵ_i, which makes our instruments $\mathbf{X}_k$ indirectly correlated with ϵ_i. We then discuss using $\mathbf{X}_i$ to instrument for k's number of followers. Analogously, $\mathbf{X}_i$ will be correlated with $\mathbf{X}_j$ and ϵ_j by i's connection utility Equation (4.4). The resemblance of ϵ_j and ϵ_k makes our instruments problematic without Assumption 2.

As a temporal summarization, so far we have discussed the validity of our instruments. We combine a linear-in-means peer effects model with a strategic network formation process to illustrate our arguments. Particularly, we explicitly outline assumptions, including the shortsightedness and Assumption 2, under which our instruments - the average observed characteristics of second-order followees and followers - are valid. In the next subsection, we present our empirical strategy and estimation results with the cross-sectional sample.

Empirical Strategy and Results Our main empirical specification for cross-sectional analyses is based on the linear-in-means model, Equation (4.3). From the cross-sectional sample, analogous to the longitudinal sample, we also observe skewed distributions for the quantity of tweets and the neighborhood sizes (Table 4.3 and Figure 4.4). Therefore, we take log-transformations of these key variables (Figure 4.4). The main specification we will estimate is thus Equation (4.3). Again, estimates of β_1 and β_2 will be the effect of peer group sizes on an individual's quantity of tweets as of the data collection date with an elasticity interpretation.

We report our main results in Table 4.10. We use the average observed characteristics of second-order followees and those of second-order followers (but not immediate neighbors) as instruments for all the endogenous variables, $\log(N_i^e)$, $\log(N_i^r)$, $\bar{Y}_{(i)}$, and $\bar{\mathbf{X}}_i$. More specifically, the observed characteristics contain the user age, gender dummy for female, dummies for missing values of age and gender separately, and a

Figure 4.4: Cross-Sectional Data - Distributions of Key Variables and Log-Transformations



dummy for age outliers.[14] Then for each user, we calculate the mean value of each characteristic of her second-order followees and second-order followers correspondingly. For the average tweets and characteristics of one's reference group ($\bar{Y}_{(i)}$ and $\bar{\mathbf{X}}_i$), we calculate these (endogenous) variables by stacking her immediate followees and followers together.

We report the two-stage least squares (2SLS) estimates in the last column of Table 4.10. Results of all first-stage regressions are reported in Table C.3 in Appendix C.2. Consistent with our findings from the panel data analyses, we also observe negative effect of the number of followees and positive effect of the followers count. However, the magnitudes are different. Particularly, our estimates suggest that a 1% increase in the number of followees will reduce an individual's quantity of total tweets by about 0.2% as of our data collection date. On the other hand, a 1% increase in the follower count raises the number of total tweets by more than 1.3%. Both effects are statistically significant. Interestingly, our results suggest that ignoring the endogeneity of an individual's peer groups (Spec 3 in Table 4.10) will overestimate the effect of followee count, while underestimate the effect of the number of followers.[15]

In addition to these main results, we also conduct several robustness checks in the next subsection. Our dataset allows us to distinguish each user's number of tweets and retweets. We conduct the first two robustness checks by formulating our dependent variables (Equation (4.3)) based on tweets and retweets respectively. From the data, we also find that a small set (992 out of 402,470) of users had not tweeted as of the data collection date. We perform another check by excluding these "inactive"

[14] We define users with year of birth being before 1940 or 2000 as age outliers. We include this dummy variable to control for cases of "fake" year of birth, a variable that is reported by users.

[15] We attribute these bias corrections to the correlations between the endogenous group sizes and the error term. For example, personality is an important factor that is omitted and potentially picked up by the error term. For a typical user, one wants more information from her followees, which leads to a positive correlation between the error term and the endogenous followees count. On the other hand, an individual is typically more cautious about expressing herself with a large group of audience, which implies a negative relationship between the followers count and the error term. Arguably, the bias corrections can be partially explained by these two conflicting factors.

users. Estimates from these robustness checks are reported in Table 4.11, and not qualitatively different from our main results.[16]

4.4.3. Robustness Checks

In this section, we discuss several concerns about our empirical specifications and main samples. We conduct robustness checks according to these concerns.

Panel Data We start with concerns on the panel data analysis.

"Mismatched" cross-section: The first concern about our panel data, as mentioned in a footnote under Table 4.2, there exists a cross-section at which the network sizes (followee and follower quantities) were collected on September 15, 2011, while the variables of activities including tweets and retweets were recorded on October 10, 2011. Our results may be attenuated by the inclusion of this mismatched cross section. We examine this concern by repeating estimations of Equation 4.1 and 4.2 without this observation date. Results are reported in the panel (I) of Table 4.7, and found to be qualitatively consistent with our main findings in Table 4.5 and 4.6.

Unbalanced panel: As mentioned in the data section, the original sample of our panel data consists of $29,956$ users. For some of them, observations are missing at some cross sections. This missingness is due to the data provider failing to collect those records. Some may suspect that these "incomplete" users may carry extra information that is not captured by our main sample, potentially attenuating our results. We conduct our second robustness check with the original unbalanced panel, and report the estimates of Equation 4.1 and 4.2 in the panel (III) of Table 4.7. The results are qualitatively identical to our main results.

[16]There might exist concerns on the log-linear functional form of our main specification, Equation (4.3). We therefore estimate an alternative specification with the squared log of the followees count and the squared log of the followers count as additional independent variables. We observe qualitatively similar results that the estimated marginal effects of the followees count are negative, and those of the followers count are positive. However, none of the estimates are statistically significant because of the collinearity of the log of group-size variables and the squared terms. Therefore, we only include the first-order terms, $\log(N_i^e)$ and $\log(N_i^r)$, in the baseline specifications.

Inactive users: From the panel data, we observe a small set of "inactive" users. Specifically, 716 out of 20,289 users did not have any new tweets during our sampling period from August 15, 2011 to December 11, 2011. Among these inactive users, 5 of them never tweeted, *i.e.*, had zero total tweet since the beginning of our study period. We conduct another robustness check to investigate whether our results will be significantly different by excluding these users. The panel (II) of Table 4.7 displays the estimation results that are, again, qualitatively similar to our main results from the panel data.

Tweets only: In the main specification and all robustness checks, the dependent variables are based on a user's number of total tweets that is the sum of tweets and retweets. One may worry that the behavior of writing tweets is quite different from that of retweeting others' tweets. We conduct our fourth robustness check of panel data analyses by constructing the dependent variables based on tweets only, but not retweets. We report the estimation results in Table 4.8 and find qualitatively similar effects with our main results again.

New neighbors: A fifth concern is about the empirical specifications used in our panel data analyses. One may argue that not only the number of followees and followers in the last period, but also the number of new followees and new followers between date $t-1$ and t, have impacts on an individual's tweeting behavior. We thus add two additional controls to Equation 4.1 and 4.2 and estimate,

$$1\left(\Delta y_{it}>0\right)=\beta_1\cdot\log\left(N^e_{i,t-1}+1\right)+\beta_2\cdot\log\left(N^r_{i,t-1}+1\right)+\beta_3\cdot\log\left(\Delta N^e_{it}\right)+\beta_4\cdot\log\left(\Delta N^r_{it}\right)+\mathbf{X}^{'}_{it}\beta_5+\mu_i+\nu_t+\epsilon_{it}, \tag{4.5}$$

and,

$$\log\left(\Delta y_{it}\right)=\beta_1\cdot\log\left(N^e_{i,t-1}+1\right)+\beta_2\cdot\log\left(N^r_{i,t-1}+1\right)+\beta_3\cdot\log\left(\Delta N^e_{it}\right)+\beta_4\cdot\log\left(\Delta N^r_{it}\right)+\mathbf{X}^{'}_{it}\beta_5+\mu_i+\nu_t+\epsilon_{it}, \tag{4.6}$$

where ΔN^e_{it} and ΔN^r_{it} are the change in the number of followees and followers respectively. Estimates are reported in Table 4.9. Not surprisingly, our main findings on

the neighborhood size effects are robust to this specification concern.

Cross-Sectional Sample Concerns on our cross-sectional analysis are as follows.

Tweets/retweets only: We perform similar robustness checks for our analyses of the cross-sectional sample. Specifically, we also examine whether using tweets (or retweets) instead of the total number of tweets and retweets would alter our results, which arises due to concerns about different behaviors underlying tweeting and retweeting. We thus calculate the dependent variables in Equation (4.3) based on the number of tweets (or retweets), and report the 2SLS estimates in the first two columns of Table 4.11. The effects of followee count and follower count are shown to be not significantly different from those in our main results.

Inactive users: Another robustness check for the cross-sectional analyses is to examine the effects of "no tweeting". We find that, in the cross-sectional sample, 992 (out of 402,470) users did not have any tweet as of the data collection date. We check whether the behavior of "active" users is significantly different by excluding these inactive users. We report the 2SLS estimates in the third column of Table 4.11. Fortunately, the peer group size effects are qualitatively similar to those from our main sample.

"Outliers": There might exist another concern about the sample of users in our cross-sectional analyses. Users with large numbers of followers are usually "VIPs" or celebrities in offline world. Their incentives of using the online networking tool may be quite different from an "average" person. For instance, a politician in campaign may tweet to attract more followers and then potentially more supporters in offline elections. We examine whether our main results are mainly driven by these "outliers" by performing another robustness check. We check whether the main results still hold if these users are excluded. Specifically, we get rid of the users with more than 200 followers,[17] the set of which contains 2,093 users. Results are reported in column (4) of Table 4.11. Both the effect of followees and that of followers are slightly smaller in absolute values, which implies that these VIP users' tweeting is more "elastic" with respect to neighborhood sizes.

4.4.4. Discussions

Comparing Panel with Cross-Section Analysis Comparing results from the panel data analysis and those from the cross-sectional data, we find that the directions of group size effects are the same while the magnitudes differ. The inconsistency in magnitudes is partly driven by the relatively small sample size of our panel data. In our panel sample, we have records of 20,289 individuals, which is fairly small compared to the Weibo population of over 500 million registered users. In contrast, we have over 2 million users in our cross-sectional data, which comprises of a non-ignorable subsample

[17] We also considered other thresholds such as 100, 500, and 1,000. The 2SLS estimates of our key variables are qualitatively similar to our main results.

of the Weibo population. More important, as we argue in Section 4.4.1, there exist extra correlations between the endogenous variables and the error term even after we incorporate both individual and time fixed effects. Due to these extra concerns on endogeneity, our results indicate that the panel data analysis overestimates the effect of followees count while underestimates that of followers count, which is consistent with our discussions of bias corrections in cross-sectional analysis.[18]

In addition, the empirical specifications in panel data analysis and cross-sectional analysis are different. In particular, the network structure in cross-sectional analysis allows us to incorporate the peer effects terms ($\bar{Y}_{(i)}$ and $\bar{\mathbf{X}}_i$ in Equation 4.4.2), which are absent in the panel analysis since we do not observe the identities of one's neighbors. Literature (Manski 1993, Goldsmith-Pinkham and Imbens 2013) establishes important effects of peers' activities and characteristics on one's behavior, which should not be omitted. Due to these reasons, the results from the cross-sectional analysis are more trustworthy. Our counterfactual exercises later will be based on these results.

Implications on Public Goods Provision Although differing in magnitudes, from both analyses, we observe that the relationship between the number of tweets and the followers count is positive, while the followees effect is negative. These results reveal asymmetric effects of the followees count and the followers count. In previous studies of peer effects, attention has been focused on undirected networks where there is no need to distinguish the effects of those two groups (Manski 1993, Bramoullé, Djebbari, and Fortin 2009, Goldsmith-Pinkham and Imbens 2013). We are among the first to provide analyses of the distinct effects of the two peer group sizes.

These results, first of all, have implications on how individuals' incentives of contributions to public goods differ when facing changes in distinct groups of peers. Given the population of followers, the set of receivers of one's tweets is fixed. Her social benefits (warm glow or personal satisfactions) from contributions will be unchanged accordingly. In this case, our findings of negative followees effects suggest that individuals are most likely to free ride on the contributions by the information senders in the network, who are not simultaneously her information receivers. Conversely, keeping the followees population unchanged fixes the information source one has. The results of the positive effect of followers count lend support to the social effect conjecture (Andreoni 1989, 1990, Zhang and Zhu 2011). Specifically, the theory implies that a larger population of followers increases individuals' utilities from

[18]These findings provide further evidence that the error term is positively correlated with the number of followees while negatively correlated with the followers count. As an example, suppose the level of an individual's dependence on the platform (varying across individuals and over time) enters the error term. This factor predicts a positive relationship between the number of followees and the error term, since more dependent users normally require more information from the platform. In contrast, such users are generally more cautious about their images on the platform, therefore more cautious about tweeting with larger population of followers.

private provision of the public good - tweets, and that they tweet more accordingly, which is supported by our findings.

Estimated Marginal Effect of Peer Group Sizes From more practical point of view, especially the platform's purpose, Tencent Weibo wants to increase the users' "activeness" - tweeting more. A significant fraction of the platform's profits come from the click-through rate of the advertisement (Xiang 2012). Higher level of user participation leads to more click-throughs of the advertisement. Our results have managerial implications for the social network to promote more active contributions.

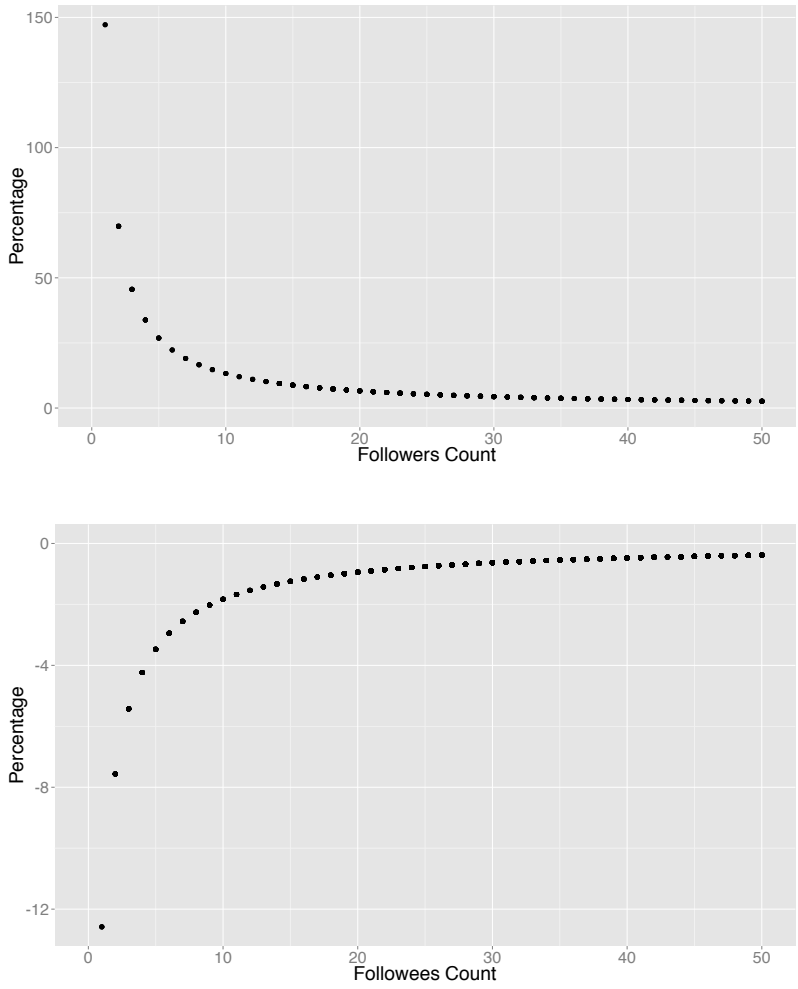
For assessing the impact of new connections on tweeting, we first study the marginal effects of followers count on an individual's tweets. Based on our estimates, with the majority of users having one follower (180, 225 out of 402, 470 in the cross-sectional sample), hypothetically adding one additional follower will more than double their quantity of tweets (130.6% increase) holding everything else constant. Similarly, a user with 2 followers (the median of followers count distribution) will tweet 65.3% more with 1 additional follower. We calculate the estimated marginal effects based on our main specification (Equation (4.3)) and coefficient estimates shown in column 4 of Table 4.10.[19] The top panel of Figure 4.5 displays the relationship between the number of followers and the estimated marginal effects (in percentage) on the quantity of tweets with one additional follower. It clearly shows that the estimated effects are positive and decreasing with the number of followers.

Although the positive effect of followers count on an individual's tweeting is both economically and statistically significant, the effect of additional links on overall tweeting in the network is still ambiguous. With one extra link in the network, an individual has one more follower while the other has an additional followee. The negative effect of followees count on one's contributions masks the effect on overall contributions in the network. We thus calculate the estimated marginal effects of followees count. The bottom panel of Figure 4.5 presents the relationship between the number of followees and the marginal percentage decrease in tweets count from one additional followee. It is to note that the estimated effects are negative for any sizes of followees population.

Simulated Effects of Random New Links In order to determine whether new connections (based on our cross-sectional sample) cause higher levels of "activeness," we conduct a counterfactual analysis. We use the total contributions, more specifically

[19]Specifically, we first calculate the fitted value of tweets count for each observation in the cross-sectional sample, based on our main specification. Then we calculate the hypothetical values of predicted tweets count with one extra follower for all observations. The estimated marginal effects are based on the differences between the fitted values. It is straightforward to show that this estimated effect is equivalent to $\left((N_i^r+1)^{\hat{\beta}_2}-(N_i^r)^{\hat{\beta}_2}\right)/(N_i^r)^{\hat{\beta}_2}$, where $\hat{\beta}_2$ is the 2SLS estimate of β_2 as in Equation 4.3. Note that the effects of an user's characteristics cancel out in the ratio because of the log-log specification.

Figure 4.5: Estimated Marginal Effects of Neighborhood Sizes on an Individual's Tweets



the total number of tweets by all individuals, as a measure of "activeness." We also use other statistics about the distribution of tweets, such as mean and median, for similar comparisons. In our counterfactual exercises, we first randomly generate new links to the existing network. For instance, we randomly draw 1 million pairs of connections from approximately 2 trillion possible links.[20] There exist more than 50 million links in our original sample, which means that the draw of simulated links comprise of less than 2% of existing links. For each draw of a simulated network, we are able to trace

[20]Our original sample contains $1,392,873$ individual users. The number of potential connections between users is thus $1,392,873 \times 1,392,872$ since the network is directed.

out the effect of changes in neighborhood sizes (followees and followers counts) on the total tweets for each individual. To see this, we denote $\tilde{Y}_i$ the number of total tweets the individual i would tweet in the simulated network. Suppose the number of followees and followers are $N_i^e + \Delta N_i^e$ and $N_i^r + \Delta N_i^r$ respectively in the simulated network. By our main specification, $\tilde{Y}_i$ will be,

$$\log\left(\tilde{Y}_i\right) = \beta_0 + \beta_1 \cdot \log\left(N_i^e + \Delta N_i^e\right) + \beta_2 \cdot \log\left(N_i^r + \Delta N_i^r\right) + \beta_{\bar{Y}} \cdot \bar{Y}_{(i)} + \bar{\mathbf{X}}_i' \beta_{\bar{\mathbf{X}}} + \mathbf{X}_i' \beta_3 + \epsilon_i. \tag{4.7}$$

By comparing Equation (4.3) with (4.7), we show that the change in tweets, $\tilde{Y}_i - Y_i$, is

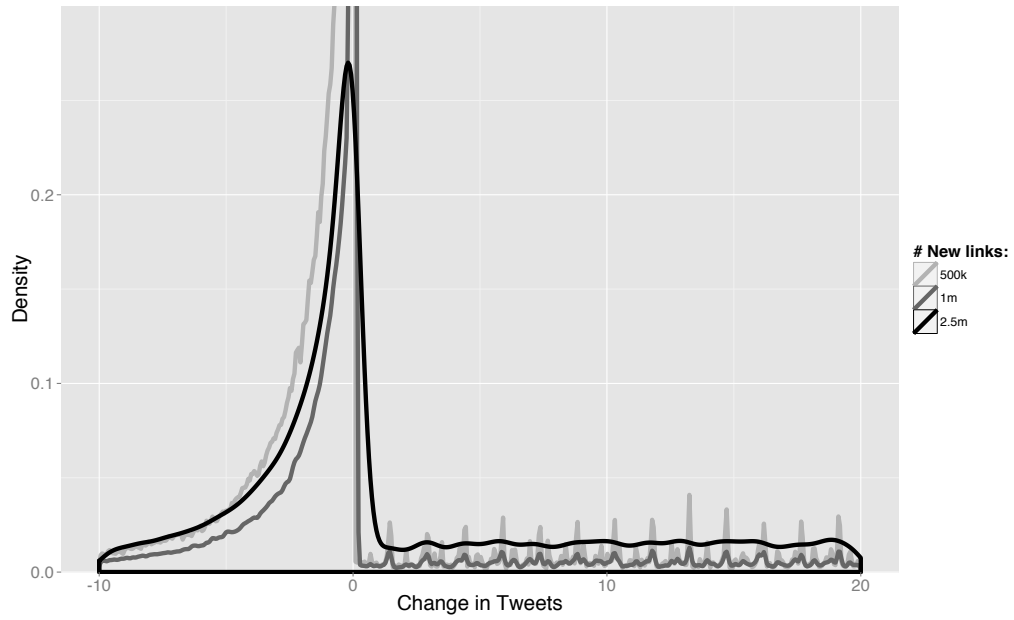
$$\tilde{Y}_i - Y_i = Y_i \cdot \frac{\left(N_i^e + \Delta N_i^e\right)^{\beta_1} \cdot \left(N_i^r + \Delta N_i^r\right)^{\beta_2}}{\left(N_i^e\right)^{\beta_1} \cdot \left(N_i^r\right)^{\beta_2}} - Y_i. \tag{4.8}$$

Then based on the 2SLS estimates $\hat{\beta}_1$ and $\hat{\beta}_2$, we can calculate the estimated change in total tweets for each user i. This allows us to back out the empirical distribution of $\tilde{Y}_i - Y_i$, and examine whether random new connections between individuals lead to higher levels of contributions in the network.

We draw three different simulated networks, with 0.5 million, 1 million, and 2.5 million new links respectively. Different draws with a given number of new links yield overlapping results. Therefore in Figure 4.6 we depict, for each given number of extra links, the empirical density function of $\tilde{Y}_i - Y_i$ from one simulation. We first notice that with more simulated links the distribution has fatter tails. Unlike what the figures might suggest, the distributions are in fact highly skewed to the right. For example, with 2.5 million new links, the median is 129.82 while the mean is 343.62. Table 4.12 summarizes the quantiles and the total changes in tweets, a column corresponding to a draw with a given number of new connections. It is observed that with greater number of simulated links, the total contribution measured by the total quantity of tweets increases significantly. Particularly, with less than 1% increase in the number of links (0.5 million), the total tweets increase by about 25.22%. More dramatically, the 2.5 million new links (about 5% larger) more than double the total number of tweets (increase by 133.75%). These dramatic effects are mainly driven by the users with very small number of followers. We notice that, for all three simulations, more than 97% of the total change in tweets are attributed to the contributions by the subgroup of individuals with less than 5 followers. The marginal effects of extra followers for these individuals are large as discussed in Section 4.4.4, contributing to those dramatic changes. It is also clear, by comparing the quantiles, that the overall contributions are growing with more links. As an example, 2.5 million new links cause 50% of the individuals to post more than 129 additional tweets, while the median is 0 for the distribution with 0.5 million extra links and 4.19 with 1 million extra links. Similar patterns are observed for other quantiles.

These findings, by comparing the distributions of the change in tweets, provide supporting evidence that managerial strategies promoting new connections in the

Figure 4.6: Distributions of the Change in Tweets $\tilde{Y}_i - Y_i$ from Three Simulated Networks



network will be effective in stimulating users' activeness. The platform may randomly recommend users to establish new connections, replicating the simulation exercises above. Furthermore, the platform may favor recommendations to the users with small number of followers. For example, an effective way is to recommend users with smaller groups of followers to those who have already followed large groups of users. Since the marginal effect of extra followers is large and dominant at a smaller number of followers (relative to the followees effect at a large number of followees), this method will generate even higher levels of contribution than purely random recommendations.

4.5. Concluding Remarks

In this paper we study the asymmetric effects of individuals' followees count and followers count on their tweeting intensity in a large-scale online social network. We propose two methods to approach the issue of endogenous peer groups. The first one is based on a panel data method, in which we incorporate both individual and time fixed effects. Extra concerns about endogeneity cannot be resolved by this longitudinal analysis. Therefore, we propose an IV approach in which we use the average observed characteristics of second-order neighbors as instruments. We observe similar asymmetric effects of followees and followers from both methods. Specifically, a greater population of followers causes individuals to tweet more while the followees

group sizes have negative effects on tweeting.

Our findings speak to the question: for whom the individuals tweet? The estimates indicate that they tweet for followers. Our findings, especially the positive effect of the followers count, indicate that individuals receive benefits or utilities from contributing to online public goods - tweeting in our case. Therefore, our paper provides supporting evidence of social effects that explain the existence of many public goods with a large number of contributors in online communities, such as Wikipedia (Zhang and Zhu 2011), online reviews (Chevalier and Mayzlin 2006), and online Q&A platforms (Xu, Nian, and Cabral 2014). We also make managerial recommendations for the platform to promote more active contributions to public goods.

As an extension, the platform may aim at promoting not only more active but also more equal contributions to public goods. Theory (Bulow and Klemperer 1996) predicts specialization in social networks being the only equilibrium that is stable over time. The question, whether incorporating social effects and asymmetric effects of neighbors in modeling individuals' incentives will alleviate the skewness, is worth pursuing in the future both analytically and empirically. Another future direction is to explore whether the provision of public goods in social networks, tweeting being an example, can reach socially optimal level in equilibrium, unlike situations where free-riding incentives lead to market failure in public goods provision.

Table 4.2: Summary Statistics of the Panel Data

	Total Tweets[a]			**New Total Tweets**			**# Followees**			**# Followers**				
Dates	Med.	Mean	s.d.	Med.	Mean	s.d.	Med.	Mean	s.d.	Med.	Mean	s.d.	**Length**[b]	**# Days**[c]
2011-08-15	25	74.014	188.054	-	-	-	20	34.428	746.238	34	43.960	68.083	209.301	-
2011-09-15[d]	36	96.222	221.432	7	22.208	56.191	25	39.269	724.470	36	46.994	70.986	240.301	31
2011-10-31	40	104.361	247.062	2	8.139	48.340	29	47.167	774.478	39	50.440	76.163	286.301	46
2011-11-15	42	110.165	268.005	1	5.804	33.710	31	49.884	788.067	40	51.675	78.252	301.301	15
2011-11-30	47	119.694	299.714	2	9.529	48.046	33	52.511	818.095	41	52.697	77.944	316.301	15
2011-12-11	49	123.502	318.153	1	3.808	27.998	34	55.223	876.927	41	53.888	80.157	327.301	11
Users								20,289						

[a]The total number of tweets is the sum of tweets and retweets, not including other activities as comments and private messages.

[b]The length on site is the number of days since a user's registration date.

[c]The number of days since the last observation date.

[d]For this cross-section, the network size information was collected on September 15, 2011, while the tweeting variables were recorded on October 10, 2011. In a robustness check (Table 4.7), we examine our empirical results by excluding this cross-section.

Table 4.3: Summary Statistics of the Original Sample

	Summary Statistics				
Variables	Med.	Mean	s.d.	Min.	Max.
# Total Tweets[a]	143	256.913	447.180	0	65,518
# Tweets	120	208.704	371.798	0	65,506
# Retweets	4	48.209	177.289	0	21,780
# Comments	1	6.458	36.653	0	13,384
Network Information					
# Followees	19	42.427	80.475	1	5,188
# Followers	2	17.203	692.496	1	159,453
# Second-Order Followees	623	935.278	1,162.658	1	38,424
# Second-Order Followers	4	845.956	13,683.390	1	1,195,098
# All Neighbors	22	59.631	698.584	2	159,496
# Friends[b]	1	1.522	5.671	0	454
Log-Transformations					
log (**# Total Tweets** + 1)	4.970	4.842	1.288	0	11.090
log (**# Followees**)	2.944	2.918	1.292	0	8.554
log (**# Followers**)	0.693	0.679	0.913	0	11.980
Observed Characteristics					
Age	22	24.123	17.147	0	123
1(Missing Year Birth)	0	0.014	0.117	0	1
1(Female)	1	0.511	0.500	0	1
1(Missing Gender)	0	0.007	0.085	0	1
1(Age Outliers)[c]	0	0.094	0.291	0	1
Users			402,470		

[a]The total number of tweets is the sum of tweets and retweets.

[b]Define a "friend" being a user who is both one's followee and follower.

[c]Define users with year of birth being before 1940 or after 2000 as "outliers."

Table 4.4: Summary Statistics of the Cross-Sectional Data - Peer Groups

	All Neighbors		2nd-Order Followees		2nd-Order Followers	
Variables[a]	Mean	s.d.	Mean	s.d.	Mean	s.d.
# Total Tweets	540.086	443.226	595.012	213.273	431.540	516.571
Age	30.186	7.051	32.401	3.066	25.302	11.251
1(Missing Year Birth)	0.071	0.075	0.076	0.039	0.011	0.071
1(Female)	0.409	0.188	0.351	0.089	0.451	0.331
1(Missing Gender)	0.063	0.067	0.077	0.031	0.006	0.051
1(Age Outliers)	0.149	0.111	0.167	0.045	0.093	0.180
Users			402, 470			

[a]For each variable, we calculate the average value among the corresponding peer group for each user. This table reports the summary stats of these mean values.

Tables: Estimation Results

Table 4.5: Panel Data - Neighborhood Size Effects on Whether to Tweet

	OLS Results[a]			
Dep var.: 1(Tweeted)	Spec 1	Spec 2	Spec 3	Spec 4
log (**Lag # Followees**)	0.067***	0.093***	-0.344***	**-0.025****
	(0.002)	(0.002)	(0.007)	(0.008)
log (**Lag # Followers**)	0.037***	0.037***	-0.005	**0.003**
	(0.001)	(0.001)	(0.004)	(0.004)
# Days Btw Dates	0.006***		0.002***	
	(1.137e-04)		(1.144e-04)	
1(Certified)	-0.050	-0.120***		
	(0.031)	(0.032)		
(Intercept)	0.193***	0.249***	1.789***	0.403***
	(0.008)	(0.007)	(0.027)	(0.031)
User FE	No	No	Yes	Yes
Collection Date FE	No	Yes	No	Yes
Adj. R^2	0.044	0.115	0.335	0.383
Users	20,289	20,289	20,289	20,289
Num. obs.	101,445	101,445	101,445	101,445

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

[a]Robust standard errors are reported. GLS estimates controlling for heteroscedasticity yield similar results.

Table 4.6: Panel Data - Neighborhood Size Effects on the Quantity of Tweets

	OLS Results[a]			
Dep var.: log (ΔTotal Tweets)	Spec 1	Spec 2	Spec 3	Spec 4
log (**Lag # Followees**)	0.233***	0.335***	-1.435***	**-0.076****
	(0.007)	(0.007)	(0.027)	(0.025)
log (**Lag # Followers**)	0.113***	0.120***	0.003	**0.005**
	(0.005)	(0.005)	(0.015)	(0.013)
# Days Btw Dates	0.014***		0.002***	
	(3.735e-04)		(3.546e-04)	
1(Certified)	0.051	-0.211		
	(0.146)	(0.146)		
(Intercept)	0.163***	0.155***	6.431***	2.696***
	(0.029)	(0.025)	(0.101)	(0.081)
User FE	No	No	Yes	Yes
Collection Date FE	No	Yes	No	Yes
Adj. R^2	0.047	0.129	0.538	0.628
Users	19,568	19,568	19,568	19,568
Num. obs.	68,954	68,954	68,954	68,954

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

[a]An unbalanced panel containing all cross sections with at least one tweet. GLS estimates controlling for heteroscedasticity yield similar results.

Table 4.7: Robustness Check (Panel) I, II, and III - Alternative Samples

	OLS Results					
	(I)[a]: Excluding Sep. 15		**(II)**[b]: No "Inactive" Users		**(III)**[c]: Unbalanced Panel	
Dep var.:	1(Tweeted)	log (ΔTotal Tweets)	1(Tweeted)	log (ΔTotal Tweets)[d]	1(Tweeted)	log (ΔTotal Tweets)
log (**Lag # Followees**)	**-0.035*****	**-0.043***	**-0.031*****	**-0.076*****	**-0.165*****	**-0.371*****
	(0.008)	(0.027)	(0.008)	(0.025)	(0.005)	(0.021)
log (**Lag # Followers**)	**0.004**	**0.012**	**3.829e-04**	**0.005**	**0.007***	**0.004**
	(0.005)	(0.014)	(0.004)	(0.013)	(0.003)	(0.011)
(Intercept)	1.021***	2.926***	1.037***	2.696***	1.359***	3.679***
	(0.026)	(0.088)	(0.026)	(0.081)	(0.017)	(0.069)
User FE	Yes	Yes	Yes	Yes	Yes	Yes
Collection Date FE	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2	0.384	0.660	0.337	0.628	0.378	0.602
Users	20,289	19,544	19,573	19,536	29,956	28,426
Num. obs.	81,156	55,568	97,865	68,902	139,007	92,518

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

[a]In this panel, we present the estimates of Equation 4.1 and 4.2 using an alternative sample by excluding the cross-section on Sep. 15, 2011. For this cross-section, the network size information was collected on Sep. 15, 2011, while the tweeting variables were recorded on Oct. 10, 2011.

[b]In this panel, we present the estimation results of Equation 4.1 and 4.2 by excluding a subset of users who did not tweet at all during our study period between Aug. 15, 2011 and Dec. 11, 2011.

[c]In this panel, we present the estimation results of Equation 4.1 and 4.2 using the unbalanced panel.

[d]Note that the estimates are exactly the same as in Table 4.6 Spec 4, since the samples are identical.

Table 4.8: Robustness Check (Panel) IV - Using the Number of Tweets Only (Excluding Retweets)

	OLS Results[a]	
Dep var.	1(Tweeted)	log (ΔTweets)
log (**Lag # Followees**)	**-0.034*****	**-0.079*****
	(0.008)	(0.025)
log (**Lag # Followers**)	**0.005**	**0.009**
	(0.004)	(0.013)
(Intercept)	0.982***	2.495***
	(0.025)	(0.080)
User FE	Yes	Yes
Collection Date FE	Yes	Yes
Adj. R^2	0.396	0.594
Users	20,289	19,059
Num. obs.	101,445	64,292

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

[a]Estimation results of Equation 4.1 and 4.2 based on the number of tweets only, not including retweets.

Table 4.9: Robustness Check (Panel) V - An Alternative Specification with Changes in Neighborhood Sizes

	OLS Results[a]	
Dep var.	1(Tweeted)	log (ΔTotal Tweets)
log (**Lag # Followees**)	**-0.023*****	**-0.070*****
	(0.008)	(0.025)
log (**Lag # Followers**)	**0.267e-04**	**0.013**
	(0.004)	(0.014)
log (New # Followees)	0.078***	0.238***
	(0.016)	(0.076)
log (New # Followers)	0.082***	0.270*
	(0.028)	(0.160)
(Intercept)	-0.232	-1.183
	(0.229)	(1.250)
User FE	Yes	Yes
Collection Date FE	Yes	Yes
Adj. R^2	0.383	0.628
Users	20,289	19,568
Num. obs.	101,445	68,954

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

[a]Estimates of Equation 4.5 and 4.6 are reported in this table.

Table 4.10: Cross-Sectional Data - Main Estimation Results

	OLS Results[a]			**2SLS Results**
Dep var.: log (Total Tweets)	Spec 1	Spec 2	Spec 3[b]	
log (**# Followees**)	0.358***		0.272***	**-0.194***
	(0.002)		(0.002)	(0.105)
log (**# Followers**)		0.480***	0.363***	**1.306*****
		(0.003)	(0.003)	(0.186)
Characteristics of All Neighbors				
Avg. Total Tweets	0.346e-03***	3.226e-04***	3.251e-04***	-1.164e-04
	(0.590e-07)	(0.595e-07)	(0.558e-07)	(3.161e-04)
Avg. Age	-0.018***	0.002***	-0.010***	-0.035***
	(3.237e-04)	(3.152e-04)	(0.307e-03)	(0.004)
Avg. 1(Missing Year Birth)	-0.340***	0.752***	0.115***	-9.166**
	(0.036)	(0.038)	(0.036)	(4.021)
Avg. 1(Female)	0.010	0.168***	0.123***	-0.540
	(0.012)	(0.012)	(0.012)	(0.604)
Avg. 1(Missing Gender)	-0.538***	0.827***	-0.033	13.033**
	(0.041)	(0.043)	(0.040)	(6.177)
Avg. 1(Age Outliers)	-0.263***	0.371***	-0.004	-0.952
	(0.019)	(0.020)	(0.019)	(0.845)
Individual Characteristics				
Age	-0.001***	-0.002***	-0.002***	-0.002***
	(1.199e-04)	(1.212e-04)	(0.116e-03)	(0.424e-03)
1(Missing Year Birth)	-0.236***	-0.259***	-0.204***	-0.110***
	(0.018)	(0.019)	(0.018)	(0.040)
1(Female)	0.105***	0.090***	0.120***	0.245***
	(0.004)	(0.004)	(0.004)	(0.037)
1(Missing Gender)	-0.149***	-0.188***	-0.148***	-0.146***
	(0.026)	(0.026)	(0.025)	(0.041)
1(Age Outliers)	0.026***	-0.045***	-0.011*	-0.135***
	(0.007)	(0.007)	(0.007)	(0.022)
(Intercept)	4.206***	4.071***	3.869***	5.776***
	(0.013)	(0.014)	(0.013)	(0.568)
Instruments	No	No	No	Yes
Adj. R^2	0.132	0.131	0.190	-
Num. obs.	402,470	402,470	402,470	402,470

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

[a]We present OLS estimates of various specifications of Equation 4.3 in columns 2 to 4. We also report the two-stage least squares (2SLS) estimates in the last column using the average characteristics of second-order followees and followers as instruments for the endogenous variables. In all regressions, the dependent variable is log (# Total Tweets + 1) since there exist observations with no tweets at all. A GMM method considering heteroscedasticity yields similar estimates and statistical inferences.

[b]Comparing estimates in Spec 3 with those from 2SLS estimation, we find that ignoring the endogeneity overestimates the followee size effect, while underestimates the effect of followers.

Table 4.11: Cross-Sectional Data - Robustness Checks

	2SLS Results[a]			
	Alternative Specs		**Alternative Samples**:	
			No "Inactive" Users[b]	No "Outliers"[c]
Dep var.:	(1) log (Tweets)	(2) log (Retweets)	(3) log (Total Tweets)	(4) log (Total Tweets)
log (**# Followees**)	**-0.116**	**-0.749*****	**-0.233****	**-0.141**
	(0.095)	(0.253)	(0.108)	(0.094)
log (**# Followers**)	**0.970*****	**2.566*****	**1.359*****	**1.215*****
	(0.173)	(0.395)	(0.187)	(0.165)
Characteristics of All Neighbors				
Avg. Total Tweets			-1.737e-04	-2.627e-04
			(3.318e-04)	(2.961e-04)
Avg. Tweets	-2.433e-04			
	(3.436e-04)			
Avg. Retweets		0.011***		
		(0.004)		
Avg. Age	-0.042***	0.021**	-0.034***	-0.037***
	(0.004)	(0.010)	(0.004)	(0.004)
Avg. 1(Missing Year Birth)	-7.472**	17.353	-8.893**	-8.419**
	(3.312)	(14.149)	(4.110)	(3.718)
Avg. 1(Female)	-1.070*	2.167	-0.672	-1.004*
	(0.557)	(1.372)	(0.607)	(0.589)
Avg. 1(Missing Gender)	7.687	26.590**	12.782**	8.339
	(5.766)	(12.946)	(6.184)	(5.510)
Avg. 1(Age Outlier)	-1.588**	8.572***	-0.644	-1.038
	(0.699)	(3.035)	(0.892)	(0.780)
Individual Characteristics				
Age	-0.002***	2.272e-04	-0.002***	-0.001***
	(3.649e-04)	(0.001)	(4.292e-04)	(3.856e-04)
1(Missing Year Birth)	-0.121***	-0.044	-0.110***	-0.140***
	(0.036)	(0.086)	(0.040)	(0.035)
1(Female)	0.266***	-0.124	0.256***	0.246***
	(0.029)	(0.131)	(0.038)	(0.035)
1(Missing Gender)	-0.160***	-0.419***	-0.146***	-0.066*
	(0.036)	(0.090)	(0.042)	(0.036)
1(Age Outlier)	-0.129***	-0.144***	-0.139***	-0.077***
	(0.020)	(0.049)	(0.022)	(0.015)
(Intercept)	6.346***	-4.279***	5.859***	6.269***
	(0.498)	(1.455)	(0.593)	(0.537)
Instruments	Yes	Yes	Yes	Yes
Num. obs.	402,470	402,470	401,478	400,377

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

[a]This table presents four robustness checks of the cross-sectional analyses. Two-stage least squares estimates of our main specification, Equation (4.3), are shown here. GMM estimations yield qualitatively similar results.

[b]In this robustness check, we delete the "inactive" users that had zero tweet by the data collection date.

[c]We get rid of "outliers" with more than 200 followers in this robustness check.

Tables: Simulation Results

Table 4.12: Comparing the Sample Quantiles of the Distributions of Changes in Tweets

	# New Links		
Quantiles	**0.5 Million**[a]	**1 Million**	**2.5 Million**
5%	-3.900	-5.440	-4.031
10%	-1.565	-2.327	-1.042
15%	-0.635	-1.121	0.000
20%	-0.119	-0.470	9.153
25%	0.000	-0.201	24.539
30%	0.000	0.000	40.780
35%	0.000	0.000	58.981
40%	0.000	0.000	79.549
45%	0.000	0.000	102.846
50%	0.000	4.187	129.817
55%	0.000	22.569	161.584
60%	0.000	42.683	199.957
65%	0.000	66.623	246.326
70%	0.000	95.729	304.668
75%	29.437	133.109	379.488
80%	64.972	182.758	480.711
85%	114.802	253.320	628.520
90%	194.004	370.051	866.620
95%	358.042	617.494	1.378e+03
Sum of Changes in Y_i	2.608e+07	5.300e+07	1.383e+08
Sum of Y_i	1.034e+08	1.034e+08	1.034e+08
Percentage Change[b]	25.222%	51.256%	133.749%

[a] In this table, we present the sample quantiles of the distributions of changes in total tweets, $\tilde{Y}_i - Y_i$, from three different simulations. Each simulation has different number of new links.

[b] The variable calculates the ratio of the sum of changes in Y_i relative to the sum of Y_i, *i.e.*, the percentage change of total simulated tweets relative to the observed total tweets.

A. Appendices to Chapter 2

A.1. Derivation of Equation (2.2)

Recall that under the auction mechanism, the borrower's expected profit is

$$E\pi_A = \frac{Q \cdot \left[\tau - E[W^{N:Q+1} + \lambda | W^{N:Q+1} + \lambda \leq r] - c\right]}{1+\tau} \cdot \Pr\left(W^{N:Q+1} + \lambda \leq r\right) + \frac{Q \cdot (\tau - r - c)}{1+\tau} \cdot \Pr\left(W^{N:Q} + \lambda \geq r > W^{N:Q+1} + \lambda\right).$$

Or,

$$\begin{aligned} E\pi_A = \frac{Q \cdot (\tau - \lambda - c)}{1+\tau} \cdot \Pr(W^{N:Q+1} \leq r - \lambda) - \frac{Q \cdot E[W^{N:Q+1} | W^{N:Q+1} \leq r - \lambda]}{1+\tau} \\ \cdot \Pr(W^{N:Q+1} \leq r - \lambda) + \frac{Q \cdot (\tau - r - c)}{1+\tau} \cdot \left[\Pr(W^{N:Q} \leq r - \lambda) - \Pr(W^{N:Q+1} \leq r - \lambda)\right]. \end{aligned} \tag{A.1}$$

To see that Equation (A.1) holds, note that[1]

$$\begin{aligned} \Pr(W^{N:Q} + \lambda \leq r < W^{N:Q+1} + \lambda) &= \Pr(W^{N:Q} \leq r - \lambda > W^{N:Q+1}) \\ &= \binom{N}{Q} \cdot F_W(r-\lambda)^Q \cdot (1 - F_W(r-\lambda))^{N-Q}, \end{aligned}$$

and

$$\begin{aligned} \Pr(W^{N:Q} \leq r - \lambda) - \Pr(W^{N:Q+1} \leq r - \lambda) &= G_Q(r-\lambda) - G_{Q+1}(r-\lambda) \\ &= \binom{N}{Q} \cdot F_W(r-\lambda)^Q \cdot (1 - F_W(r-\lambda))^{N-Q}. \end{aligned}$$

We can rewrite Equation (A.1) as

$$\begin{aligned} E\pi_A \propto (\tau - \lambda - c) \cdot \Pr(W^{N:Q+1} \leq r - \lambda) - E[W^{N:Q+1} | W^{N:Q+1} \leq r - \lambda] \cdot \Pr(W^{N:Q+1} \leq r - \lambda) \\ + \cdot (\tau - r - c) \cdot \left[\Pr(W^{N:Q} \leq r - \lambda) - \Pr(W^{N:Q+1} \leq r - \lambda)\right], \end{aligned} \tag{A.2}$$

where we omit the constant parts. The conditional expectation in Equation (A.2) can be written as $E[W^{N:Q+1} | W^{N:Q+1} \leq r - \lambda] = \frac{\int_{-\infty}^{r-\lambda} w g_{Q+1}(w) dw}{G_{Q+1}(r-\lambda)}$. And since $\Pr(W^{N:Q} \leq r - \lambda) = G_Q(r-\lambda)$ and $\Pr(W^{N:Q+1} \leq r - \lambda) = G_{Q+1}(r-\lambda)$, (A.2) can be further simplified to

$$\begin{aligned} E\pi_A \propto (\tau - \lambda - c) \cdot G_{Q+1}(r-\lambda) - \int_{-\infty}^{r-\lambda} w g_{Q+1}(w) dw \\ + \cdot (\tau - r - c) \cdot [G_Q(r-\lambda) - G_{Q+1}(r-\lambda)] \end{aligned} \tag{A.3}$$

[1] See the section of order statistics in Casella and Berger (2001).

The first order necessary condition to the maximization of $E\pi_A$ is then

$$\frac{dE\pi_A}{dr} = G_{Q+1}(r-\lambda) - G_Q(r-\lambda) + (\tau - r - c)g_Q(r-\lambda) = 0$$

Then the optimal reserve price satisfies

$$r^* = \tau - c - \frac{G_Q(r^*-\lambda) - G_{Q+1}(r^*-\lambda)}{g_Q(r^*-\lambda)}. \tag{A.4}$$

The second part in the RHS of (A.4) can be extended as

$$\frac{\binom{N}{Q} \cdot F_W(r^*-\lambda)^Q \cdot (1-F_W(r^*-\lambda))^{N-Q}}{\frac{N!}{(N-Q)!(Q-1)!} \cdot f_W(r^*-\lambda) \cdot F_W(r^*-\lambda)^{Q-1} \cdot (1-F_W(r^*-\lambda))^{N-Q}} = \frac{F_W(r^*-\lambda)}{Q \cdot f_W(r^*-\lambda)}.$$

So the following holds

$$r^* = \tau - c - \frac{F_W(r^*-\lambda)}{Q \cdot f_W(r^*-\lambda)}. \tag{A.5}$$

(A.5) is Equation (2.2).

A.2. Welfare Comparison

In this appendix, we formally develop our discussion on the comparison of total welfare under the two market mechanisms. Mathematically, we decompose the total surplus of a particular loan into borrower surplus and lenders surplus, and compare them.

In the model described in Section 2.3, we assume that lenders in an auction incur positive transaction cost (λ), relative to the posted-price regime. An immediate observation is that this cost may induce efficiency loss, as the cost raises the lenders' minimum acceptable interest rates, which in turn lowers the funding probability. In other words, holding everything else equal (such as the borrower profile and lender population), the transaction cost associated with the auction mechanism reduces the total social surplus, as it raises the lenders' willingness to lend (supply function). Another key note is that in the posted-price environment, the borrower surplus is squeezed to zero: our model shows that under posted prices, Prosper.com assigns $p^* = \tilde{p}$, the maximum interest rate at which a borrower is willing to post the listing. Consequently, the lenders surplus is maximized under posted prices.

Let us first look at the total surplus for a listing under the auction mechanism. We consider a successfully funded loan with a default rate, δ. Let p denote the contract interest rate again, and then the default rate $\delta = \delta(p)$. We know that the borrower's highest acceptable interest rate is $\tilde{p} = \tau - c$ from our earlier results. We also know that with probability $\Pr\left(W^{N:Q+1} + \lambda \leq r^*\right)$ the contract interest rate is $p = W^{N:Q+1} + \lambda$ (where the lowest losing bid sets the contract rate), and with probability $\Pr\left(W^{N:Q} + \lambda \leq r^* < W^{N:Q+1} + \lambda\right)$ the contract rate is $p = r^*$.

We calculate the lenders surplus first. For a winner, since his minimum acceptable interest rate is generally lower than the contract rate, the individual surplus will be $\left(1 + w^{N:Q+1} + \lambda\right)(1 - \delta(p)) - \left(1 + w^{N:k} + \lambda\right)$ that is $w^{N:Q+1} + \lambda - \delta(p)(1 + w^{N:Q+1} + \lambda) - \lambda - w^{N:k}$, for k among the Q winners (lowest lenders) if the lowest losing bid sets the contract rate (Scenario I); $(1 + r^*)(1 - \delta(p)) - \left(1 + w^{N:k} + \lambda\right)$ that is $r^* - \delta(p)(1 + r^*) - \lambda - w^{N:k}$ for all k if the borrower's reserve interest rate sets the contract rate (Scenario II). All the other lenders fail to fund the loan, and thus have zero surplus. Then the total lenders surplus will be the sum over of all winners' surplus. The *ex ante* total lenders surplus (LS_A) can be shown to be equal to,

$$\begin{aligned}\mathrm{LS}_A = {} & E\left[\sum_{k=1}^{Q}\left[\gamma(W^{N:Q+1} + \lambda) - \lambda - W^{N:k}\right] | W^{N:Q+1} \leq r^* - \lambda\right] \times \Pr\left(W^{N:Q+1} \leq r^* - \lambda\right) \\ & + E\left[\sum_{k=1}^{Q}\left[\gamma(r^*) - \lambda - W^{N:k}\right] | W^{N:Q} \leq r^* - \lambda < W^{N:Q+1}\right] \times \Pr\left(W^{N:Q} \leq r^* - \lambda < W^{N:Q+1}\right).\end{aligned}$$

We then turn to borrower side. The borrower surplus, if the loan is successfully funded, will be $Q \times \left[(1 + \tilde{p}) - \left(1 + w^{N:Q+1} + \lambda\right)\right]$ that is $Q \times \left(\tilde{p} - w^{N:Q+1} - \lambda\right)$ in scenario I; and $Q \times [(1 + \tilde{p}) - (1 + r^*)]$ that is $Q \times (\tilde{p} - r^*)$ in scenario II. Then the

borrower's *ex ante* surplus (BS_A) will be

$$\begin{aligned}\text{BS}_A = &Q \times E\left[\tilde{p} - W^{N:Q+1} - \lambda | W^{N:Q+1} \leq r^* - \lambda\right] \times \Pr\left(W^{N:Q+1} \leq r^* - \lambda\right) \\ &+ Q \times E\left[\tilde{p} - r^* | W^{N:Q} \leq r^* - \lambda < W^{N:Q+1}\right] \times \Pr\left(W^{N:Q} \leq r^* - \lambda < W^{N:Q+1}\right).\end{aligned}$$

The *ex ante* social surplus (TS_A) will be the sum of the total lenders surplus (LS_A) and the total borrower surplus (BS_A). It can be shown that the social surplus is equal to

$$\begin{aligned}\text{TS}_A = E\left[\sum_{k=1}^{Q}\left[\tilde{p} - \delta(W^{N:Q+1} + \lambda)(1 + W^{N:Q+1} + \lambda) - \lambda - W^{N:k}\right] \middle| W^{N:Q+1} \leq r^* - \lambda\right] \\ \times \Pr\left(W^{N:Q+1} \leq r^* - \lambda\right) + E\left[\sum_{k=1}^{Q}\left[\tilde{p} - \delta(r^*)(1 + r^*) - \lambda - W^{N:k}\right] \middle| W^{N:Q} \leq r^* - \lambda < W^{N:Q+1}\right] \\ \times \Pr\left(W^{N:Q} \leq r^* - \lambda < W^{N:Q+1}\right)\end{aligned} \tag{A.6}$$

Let us now conduct similar analysis for the posted-price mechanism. Previous result tells us that Prosper.com presets the interest rate that is equal to the borrower's highest acceptable rate, which is $p^* = \tilde{p} = \tau - c$. In this case the personal loan is funded if and only if $w^{N:Q} \leq \gamma(p^*) = \gamma(\tau - c)$, or the Q-th lowest lender is willing to fund the loan. So a winning lender's surplus will be equal to $(1 + p^*)(1 - \delta(p)) - (1 + w^{N:k}) = \gamma(p^*) - w^{N:k}$, for all $k = 1, 2, ..., Q$ among the lowest lenders. Then the total lenders surplus is the sum of all lenders' surplus, which is $\sum_{k=1}^{Q}\left[\gamma(p) - w^{N:k}\right]$. The *ex ante* total lenders surplus (LS_P) is thus

$$\text{LS}_P = E\left[\sum_{k=1}^{Q}\left[\gamma(\tau - c) - W^{N:k}\right] | W^{N:Q} \leq \gamma(\tau - c)\right] \times \Pr\left(W^{N:Q} \leq \gamma(\tau - c)\right),$$

where we define $\gamma(p) = p - \delta(p)(1 + p)$ in Section 2.3.

As mentioned earlier, the borrower surplus in the posted-price mechanism is zero since Prosper.com presets the interest rate equal to his maximum willingness to borrow. Thus in this case the social surplus (TS_P) will be equal to the total lenders surplus, which we can show is equal to

$$\text{TS}_P = E\left[\sum_{k=1}^{Q}\left[\gamma(\tau - c) - W^{N:k}\right] | W^{N:Q} \leq \gamma(\tau - c)\right] \times \Pr\left(W^{N:Q} \leq \gamma(\tau - c)\right). \tag{A.7}$$

First of all, given assumption 1, it can be shown that $\text{LS}_p \geq \text{LS}_A$. That is, the lenders surplus under the posted-price mechanism is at least as great as that in the auctions. In other words, the lenders are better off after the regime change.

More importantly, however, the regime change does not necessarily lead to higher total social welfare. It can be shown that under the following assumption, the total surplus in the posted-price selling (TS_P) is smaller than that in the auctions (TS_A).

Assumption 3. Let $p_A = \min\{r^*, w^{N:Q+1} + \lambda\}$, $\delta(\tilde{p})(1 + \tilde{p}) - \delta(p_A)(1 + p_A) \geq \lambda$.

Note that $\delta(p)(1+p)$ measures a lender's expected loss , due to loan defaulting, at contract interest rate p. Thus, the left-hand-side of the inequality measures the difference in the expected loss under the two regimes. While the right-hand-side is the transaction cost that induces efficiency loss in the auctions. This assumption suggests that if the expected loss is reasonably high in the posted-price regime, the total social surplus will be less than that in the auctions. In this situation, the posted-price mechanism is dominated by the auctions.

A.3. List of Control Variables

In this appendix, we list the control variables in all our estimations of the regime change effects, from Table 2.4 through 2.15.

Listing characteristics:

1. Loan term in months
2. 1(Electronic transfer), 1(With images)
3. Borrower's requested amount of dollars
4. 1 (With friends in participating bidders (lenders))
5. **Interest rate categories**
6. 1(Group loan)
7. 1(With descriptions)

Credit information:

1. Whether the borrower is a home owner
2. Amount delinquent
3. Bankcard utilization (%)
4. Current credit lines
5. Current delinquencies
6. Delinquencies in last 7 yeas
7. Inquiries in last 6 months
8. Open credit lines
9. Public records in last 10 years
10. Public records in last 12 months
11. Revolving credit balance
12. Stated monthly income
13. Total credit lines
14. Length credit history

Macroeconomic environment:

1. Unemployment rate by month by state
2. 1(Miss unemployment rate) (Districts such as army areas do not have unemployment data.)
3. Zillow home value index[2] by month by state
4. 1(Miss Zillow home value index) (Zillow.com calculates the index in 48 states during our study period.)

[2]Zillow.com calculates and publishes the data on a monthly basis. The data are available at http://www.zillow.com/research/data/.

B. Appendices to Chapter 3

B.1. List of Control Variables

In this appendix, we list the control variables in our estimations of Equations 3.1, 3.2, 3.3, and 3.4.

Listing characteristics:

1. Borrower's requested amount of dollars
2. Borrower's asking interest rate
3. Whether the listing is open for its duration

Verified credit information:

1. Whether the borrower is a home owner
2. Amount delinquent
3. Current delinquencies
4. Delinquencies in last 7 yeas
5. Public records last year
6. Public records in last 10 years
7. Inquiries in last 6 months
8. Bankcard utilization (%)
9. Current credit lines
10. Revolving credit balance

11. Total credit lines

Discritization of financial information: We allow for nonlinear effects of several control variables. The thresholds are based on median or quantiles values. Specifically, the variables and thresholds are:

1. Amount requested: thresholds at $\$3,000$, $\$5,000$ and $\$10,000$
2. Current delinquencies: a threshold at 1 current delinquency
3. Delinquencies in last 7 years: a threshold at 3 delinquencies
4. Inquiries in last 6 months: a threshold at 2 recent inquiries
5. Bankcard utilization: a threshold at a utilization ratio of 0.75
6. Current credit lines: thresholds at 4, 8 and 12 current credit lines

Non-verified information: We include a set of dummies for the inclusion of each of the following words/phrases in listings' titles: "help, "credit card", "debt", "consolidate", "start business", "real estate", "student", "school", "tuition", "medical", "doctor", "fresh start", "good guy". We also include another set of dummies indicating occupations of the borrowers.

C. Appendices to Chapter 4

C.1. Cross-Sectional Analysis - Sample Comparisons

In this appendix, we present an investigation of our main sample in the cross-sectional analyses. Recall that we construct a sample of $402,270$ users from the original dataset consisting of $1,392,873$ individuals. We compare these two samples and confirm the representativeness of our main sample. Table below shows summary statistics. We observe that our main sample is largely representative in terms of demographic characteristics, although these users have on average larger number of followees and that of followers and tweet more. This is mainly driven by our IV strategy that focuses on individuals with at least one second-order followees and one second-order followers.

Table C.1: Summary Statistics of the Cross-Sectional Data - Sample Comparisons

	Main Sample			**Original Sample**				
Variables	Med.	Mean	s.d.	Med.	Mean	s.d.	***t*-stats**[a]	***p*-value**
# Total Tweets	143	256.913	447.180	62	142.074	283.663	154.203	< 2.2e-16
# Followees	19	42.427	80.475	11	24.130	53.396	135.864	< 2.2e-16
# Followers	2	17.203	692.496	0	5.162	373.047	10.596	< 2.2e-16
Age	22	24.123	17.147	22	24.571	17.124	-14.589	1
1(Missing Year Birth)	0	0.014	0.117	0	0.016	0.127	-11.004	1
1(Female)	1	0.511	0.500	1	0.504	0.500	7.413	6.173e-14
1(Missing Gender)	0	0.007	0.085	0	0.009	0.096	-12.261	1
1(Age Outliers)	0	0.094	0.291	0	0.098	0.297	-7.977	1
Users		402,470			1,392,873			

[a]We conduct two-sample t-test for each variable. The alternative hypothesis is the mean of the corresponding variable in the main sample is *greater* than that in the original sample.

C.2. First-Stage Regressions and F-Statistics

In this section, we report the results from the first-stage regressions for our main specification, as shown in the column 4 of Table 4.10. The OLS estimates, along with F-statistics, of all first-stage regressions are presented in Table C.3. The endogenous variables include the neighborhood sizes (followees and followers count) and the average tweets and observed characteristics of all neighbors. We also report the values of F-statistics for all 2SLS regressions in the robustness check (as in Table 4.11). The results are shown in Table C.2. The specifications are corresponding to those in Table 4.11.

Table C.2: Values of F-Statistic for 2SLS Regressions in Robustness Checks

	F-stats for robustness checks[a]			
Specification:	(1)	(2)	(3)	(4)
log (**# Followees**)	2.179e+03	2.179e+03	2.182e+03	2.154e+03
log (**# Followers**)	1.134e+03	1.134e+03	1.132e+03	1.128e+03
Characteristics of All Neighbors				
Avg. Total Tweets			1.836e+03	1.837e+03
Avg. Tweets	1.470e+03			
Avg. Retweets		1.445e+03		
Avg. Age	2.205e+03	2.205e+03	2.229e+03	2.204e+03
Avg. 1(Missing Year Birth)	0.642e+03	0.642e+03	0.641e+03	0.631e+03
Avg. 1(Female)	2.119e+03	2.119e+03	2.128e+03	2.108e+03
Avg. 1(Missing Gender)	0.261e+03	0.261e+03	0.261e+03	0.257e+03
Avg. 1(Age Outlier)	1.201e+03	1.201e+03	1.202e+03	1.209e+03
Num. obs.	402,470	402,470	401,478	400,377

[a] This table presents the F-statistics from the first-stage regressions of our 2SLS regressions in robustness checks. The columns are corresponding to the specifications in Table 4.11.

Table C.3: First-Stage Regressions Results for the 2SLS Estimations of Equation (4.3) - Main Specification

	OLS estimates[a]							
			Avg. characteristics of all neighbors					
Dep. var.:	log (Followees)	log (Followers)	Ttl. Tweets	Age	1(Miss Birth)	1(Female)	1(Miss Gender)	1(Age Outlier)
Individual Characteristics								
Age	3.545e-04***	0.001***	0.578***	0.044***	-1.645e-04***	-1.299e-04***	-0.406e-04***	0.599e-04***
	(1.228e-04)	(1.059e-04)	(0.044)	(0.00)	(0.685e-07)	(0.174e-04)	(0.643e-07)	(0.111e-04)
1(Missing Year Birth)	-0.216***	-0.156***	25.374***	1.892***	-0.009***	-0.007**	-0.004***	0.006***
	(0.018)	(0.012)	(6.506)	(0.114)	(0.001)	(0.003)	(0.001)	(0.002)
1(Female)	0.042***	-0.037***	34.993***	0.020	0.010***	0.037***	0.002***	-0.007***
	(0.004)	(0.003)	(1.412)	(0.022)	(2.447e-04)	(0.001)	(0.222e-03)	(0.366e-03)
1(Missing Gender)	-0.133***	-0.033	-19.583**	-0.104***	0.003*	0.015***	-0.001	0.001
	(0.026)	(0.023)	(8.214)	(0.149)	(0.002)	(0.004)	(0.001)	(0.003)
1(Age Outliers)	-0.103***	0.085***	20.848***	-0.015	-0.005***	-0.002**	-0.003***	0.007***
	(0.007)	(0.006)	(2.481)	(0.040)	(4.044e-03)	(0.001)	(3.724e-04)	(0.001)
Average Characteristics of Second-Order Followees								
Age	-0.047***	-0.034***	1.916***	0.503***	-0.001***	-0.006***	0.001***	-0.001***
	(0.001)	(0.001)	(0.380)	(0.009)	(0.673e-04)	(2.032e-04)	(0.599e-04)	(1.286e-04)
1(Missing Year Birth)	-2.186***	-1.765***	-1.877e+03***	-20.569***	0.178***	0.389***	0.091***	-0.247***
	(0.138)	(0.056)	(44.140)	(1.076)	(0.010)	(0.024)	(0.008)	(0.017)
1(Female)	-4.103***	-0.927***	-59.135***	-10.141***	-0.003	0.400***	-0.036***	-0.115***
	(0.053)	(0.024)	(14.685)	(0.317)	(0.003)	(0.008)	(0.002)	(0.005)
1(Missing Gender)	-4.634***	-2.983***	-1.482e+03***	14.918***	0.180***	0.452***	0.085***	-0.141***
	(0.164)	(0.069)	(49.624)	(1.242)	(0.012)	(0.027)	(0.010)	(0.018)
1(Age Outliers)	-0.872***	-1.287***	0.245e+03***	-4.632***	-0.087***	-0.079***	-0.008**	0.318***
	(0.100)	(0.042)	(24.753)	(0.575)	(0.004)	(0.013)	(0.004)	(0.009)
Average Characteristics of Second-Order Followers								
Age	0.001***	0.002***	0.710***	0.049***	-0.124e-03***	-1.623e-04***	-0.512***	1.194e-04***
	(1.767e-04)	(0.971e-04)	(0.062)	(0.001)	(0.116e-04)	(0.274e-04)	(0.104e-04)	(0.171e-04)
1(Missing Year Birth)	-0.045	-0.054***	19.407*	1.854***	-0.005***	-0.012***	-0.004**	0.007**
	(0.031)	(0.014)	(10.447)	(0.181)	(0.002)	(0.005)	(0.002)	(0.003)
1(Female)	-0.023***	-0.066***	-11.507***	-0.168***	0.002***	0.009***	0.002***	-0.001
	(0.006)	(0.003)	(2.005)	(0.033)	(3.847e-04)	(0.001)	(3.463e-04)	(0.001)
1(Missing Gender)	-0.048	-0.019	-28.046**	-0.705***	-0.778e-07	0.012*	-0.002	-0.002
	(0.042)	(0.020)	(13.320)	(0.242)	(0.003)	(0.007)	(0.002)	(0.004)
1(Age Outliers)	0.032***	0.058***	14.911***	0.208***	-0.001*	-0.001	-0.001*	0.005***
	(0.011)	(0.005)	(3.649)	(0.065)	(0.001)	(0.002)	(0.001)	(0.001)
(Intercept)	6.499***	2.661***	0.667e+03***	16.348***	0.084***	0.403***	0.042***	0.186***
	(0.064)	(0.034)	(16.636)	(0.431)	(0.003)	(0.009)	(0.003)	(0.006)
***F*-stats**	1.617e+03	0.839e+03	1.296e+03	2.174e+03	0.743e+03	2.378e+03	0.190e+03	1.081e+03
Num. obs.	402,470	402,470	402,470	401,478	400,377	402,470	402,470	402,470

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

[a]This table presents the estimates from the first-stage regressions of the 2SLS estimation of Equation (4.3), for the main specification as shown in Table 4.10. Columns are corresponding to all endogenous variables including neighborhood sizes and average observed characteristics among all neighbors.

References

Abadie, A., and G. Imbens. 2006. Large Sample properties of Matching Estimators for Average Treatment Effects. *Econometrica* **74**(1): 235–267.

Agrawal, A., C. Catalini, and A. Goldfarb. 2013. The Simple Economics of Crowdfunding. *Innovation Policy and the Economy* Volume 14. University of Chicago Press.

Andreoni, J. 1989. Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence. *Journal of Political Economy* **97**(6): 1447–1458.

Andreoni, J. 1990. Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving. *Economic Journal* **100**(401): 464–477.

Andreoni, J. 2007. Giving Gifts to Groups: How Altruism Depends on the Number of Recipients. *Journal of Public Economics* **91**(9): 1731–1749.

Andreoni, J., and B. D. Bernheim. 2009. Social Image and the 50-50 Norm: A Theoretical and Experimental Analysis of Audience Effects. *Econometrica* **77**(5): 1607–1636.

Aral, S., E. Brynjolfsson, and M. Van Alstyne. 2012. Information, Technology, and Information Worker Productivity. *Information Systems Research* **23**(3): 849–867.

Aral, S., and D. Walker. 2011. Creating Social Contagion through Viral Product Design: A Randomized Trial of Peer Influence in Networks. *Management Science* **57**(9): 1623–1639.

Aral, S., and D. Walker. 2012. Identifying Influential and Susceptible Members of Social Networks. *Science* **337**(6092): 337–341.

Arora, A., R. Krishnan, R. Telang, and Y. Yang. 2010. An Empirical Analysis of Software Vendors Patch Release Behavior: impact of Vulnerability Disclosure. *Information Systems Research* **21**(1): 115–132.

Ausubel, L. M., and P. Cramton. 2002. Demand Reduction and Inefficiency in Multi-Unit Auctions. *Working Paper*.

Back, K., and J. F. Zender. 1993. Auctions of Divisible goods: On the Rationale for the Treasury Experiment. *Review of Financial Studies* **6**(4): 733–764.

Bajari, P., and A. Hortaçsu. 2004. Economic Insights from Internet Auctions. *Journal of Economic Literature* **42**(2): 457–486.

Banerjee, A. V.. 1992. A Simple Model of Herd Behavior, *Quarterly Journal of Economics* **75**(4): 797–817.

Barber, B. M., and T. Odean. 2008. All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors. *Review of Financial Studies* **21**(2): 785–818.

Barber, B. M., and T. Odean. 2011. The Behavior of Individual Investors, *Handbook of the Economics of Finance* **2**: 1533–1570.

Basak, S. and A. Pavlova. 2013. Asset Prices and Institutional Investors. *American Economic Review* **103**(5): 1728–1758.

Berry, S., J. Levinsohn, and A. Pakes. 1995. Automobile Prices in Market Equilibrium. *Econometrica* **63**(4): 841–890.

Bester, H.. 1985. Screening vs. Rationing in Credit Markets with Imperfect Information. *American Economic Review* **75**(4): 850–855.

Biais, B., P. Bossaerts, and Jean-Charles Rochet. 2002. An optimal IPO mechanism. *Review of Economic Studies* **69**(1): 117–146.

Bikhchandani, S., D. Hirshleifer, and I. Welch. 1992. A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades. *Journal of Political Economy* **100**(5): 992–1026.

Bond, R. M., C. J. Fariss, J. J. Jones, A. D. Kramer, C. Marlow, J. E. Settle, and J. H. Fowler. 2012. A 61-Million-Person Experiment in Social Influence and Political Mobilization. *Nature* **489**(7415): 295–298.

Borgatti, S. P., A. Mehra, D. J. Brass, and G. Labianca. 2009. Network Analysis in the Social Sciences. *Science* **323**(5916): 892–895.

Bramoullé, Y., Djebbari, H., and B. Fortin. 2009. Identification of Peer Effects Through Social Networks. *Journal of Econometrics* **150**(1): 41–55.

Bramoullé, Y. and R. Kranton. 2007. Public Goods in Networks. *Journal of Economic Theory* **135**(1): 478–494.

Bulow, J., and P. Klemperer. 1996. Auctions versus Negotiations. *American Economic Review* **86**(1): 180–194.

Burtch, G., A. Ghose, and S. Wattal. 2013. An Empirical Examination of the Antecedents and Consequences of Contribution Patterns in Crowd-Funded Markets. *Information Systems Research* **24**(3): 499–519.

Calvó-Armengol, A., and M. O. Jackson. 2004. The Effects of Social Networks on Employment and Inequality. *American Economic Review* **94**(3): 426–454.

Cameron, A. C., and P. K. Trivedi. 2005. *Microeconometrics: Methods and Applications*. Cambridge University Press.

Carlin, B. I. and G. Manso. 2011. Obfuscation, Learning, and the Evolution of Investor Sophistication. *Review of Financial Studies* **24**(3): 754–785.

Casella, G., and R. L. Berger. 2001. *Statistical Inference*. Duxbury Press.

Chen, H., P. De, J. Hu, and B. Hwang. 2014. Wisdom of Crowds: The Value of Stock Opinions Transmitted through Social Media. *Review of Financial Studies* **27**(5): 1367–1403.

Chen, N., A. Ghosh, and N. S. Lambert. 2014. Auctions for Social Lending: A Theoretical Analysis. *Games and Economic Behavior* **86**: 367–391.

Chen, Y., F. M. Harper, J. Konstan, and S. X. Li. (2010). Social Comparisons and Contributions to Online Communities: A Field Experiment on Movielens. *American Economic Review* **100**(4): 1358–1398.

Chevalier, J. A., and D. Mayzlin. (2006). The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of Marketing Research* **43**(3): 345–354.

Chi, F., and N. Yang. 2011. Twitter Adoption in Congress. *Review of Network Economics* **10**(1).

Conley, T. G., and C. R. Udry. 2010. Learning about a New Technology: Pineapple in Ghana. *American Economic Review* **100**(1): 35–69.

Dellarocas, C., T. W. Malone, and R. Laubacher. 2010. Harnessing crowds: Mapping the genome of collective intelligence. *Sloan Management Review* **51**(3): 21–31.

Duan, N., W. G. Manning, C. N. Morris, and J. P. Newhouse. 1983. A Comparison of Alternative Models for the Demand for Medical Care. *Journal of Business and Economic Statistics* **1**(2): 115–126.

Einav, L., C. Farronato, J. Levin, and N. Sundaresan. 2013. Sales Mechanisms in Online Markets: What Happened to Internet Auctions? *NBER Working Paper No. 17802*.

Freedman, S., and G. Z. Jin. 2011. Learning by Doing with Asymmetric Information: Evidence from Prosper.com. *NBER Working Paper No. 16855*.

Goldsmith-Pinkham, P., and G. W. Imbens. 2013. Social Networks and the Identification of Peer Effects. *Journal of Business and Economic Statistics* **31**(3): 253–264.

Gruber, M. J. 1996. Another Puzzle: The Growth in Actively Managed Mutual Funds. *Journal of Finance* **51**(3): 783–810.

Hahn, J., and G. Lee. 2013. Archetypes of Crowdfunders Backing Behaviors and the Outcome of Crowdfunding Efforts: An Exploratory Analysis of Kickstarter. *Working Paper*.

Halberstam, Y., and B. Knight. 2014. Homophily, Group Size, and the Diffusion of Political Information in Social Networks: Evidence from Twitter. *NBER Working Paper* (No. w20681).

Hammond, R. G. 2010. Comparing Revenue from Auctions and Posted Prices. *International Journal of Industrial Organization* **28**(1): 1–9.

Hammond, R. G. 2013. A structural Model of Competing Sellers: Auctions and Posted Prices. *European Economic Review* **60**(1): 52–68.

Hortaçsu, A., and D. McAdams. 2010. Mechanism Choice and Strategic Bidding in Divisible Good Auctions: an Empirical Analysis of the Turkish Treasury Auction Market. *Journal of Political Economy* **118**(5): 833–865.

Hosanagar, K., P. Han, and Y. Tan. 2010. Diffusion Models for Peer-to-Peer (P2P) Media Distribution: On the Impact of Decentralized, Constrained Supply. *Information Systems Research* **21**(2): 271–287.

Iyer, R., A. Ijaz, K. Erzo, F. P. Luttmer, and K. Shue. 2010. Inferring Asset Auality: Determining Borrower Creditworthiness in Peer-to-Peer Lending Markets. *Working Paper*.

Kawai, K., K. Onishi, and K. Uetake. 2014. Signaling in Online Credit Markets. *Working Paper*.

Kim, K., and I-H. Hann. 2014. Crowdfunding and the Democratization of Access to Capital: A Geographical Analysis. *Working Paper*.

Krishna, V. 2009. *Auction Theory*. Academic Press.

Lakonishok, J. and E. Maberly. 1990. The Weekend Effect: Trading Patterns of Individual and Institutional Investors. *Journal of Finance* **45**(1): 231–243.

Lawton, K., and D. Marome. 2010. *The Crowdfunding Revolution: Social Networking Meets Venture Financing*. www.thecrowdfundingrevolution.com.

Lemmon, M., and S. X. Ni. 2014. Differences in Trading and Pricing Between Stock and Index Options. *Management Science* Forthcoming.

Lin, M., N. R. Prabhala, and S. Viswanathan. 2013. Judging Borrowers by the Company They Keep: Friendship Networks and Information Asymmetry in Online Peer-to-Peer Lending. *Management Science* **59**(1): 17–35.

Liu, T. X., J. Yang, L. Adamic, and Y. Chen. 2014. Crowdsourcing with All-Pay Auctions: A Field Experiment on Taskcn. *Management Science* Forthcoming.

Lucking-Reiley, D. 2000. Auctions on the Internet: What are Being Auctioned, and How? *Journal of Industrial Economics* **48**(3): 227–252.

Manski, C. F.. 1993. Identification of Endogenous Social Effects: The Reflection Problem. *Review of Economic Studies* **60**(3): 531–542.

McPherson, M., L. Smith-Lovin, and J. M. Cook. 2001. Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology* **27**: 415-444.

Moretti, E.. 2011. Social Learning and Peer Effects in Consumption: Evidence from Movie Sales. *Review of Economic Studies* **78**(1): 356–393.

Nofsinger, J. R. and R. W. Sias. 1999. Herding and Feedback Trading by Institutional and Individual Investors. *Journal of Finance* **54**(6): 2263–2295.

Olson, M. 1965. The Logic of Collective Action: Public Goods and the Theory of Groups. Harvard University Press.

Peterson, R. D. 2012. To Tweet or not to Tweet: Exploring the Determinants of Early Adoption of Twitter by House Members in the 111th Congress. *Social Science Journal* **49**(4): 430–438.

Puckett, A. and X. Yan. 2011. The Interim Trading Skills of Institutional Investors. *Journal of Finance* **66**(2): 601–633.

Renton, P. 2010. *Prosper.com Ending Their Auction Process Dec 19th*. www.lendacademy.com.

Rigbi, O. 2013. The Effects of Usury Laws: Evidence from the Online Loan Market. *Review of Economics and Statistics* **95**(4): 1238–1248.

Rosenbaum, P. R. 2002. *Observational Studies*. NY: Springer.

Saunders Jr, E. M. 1993. Stock Prices and Wall Street Weather. *American Economic Review* **83**(5): 1337–1345.

Scharfstein, D. S. and J. C. Stein. 1990. Herd Behavior and Investment. *American Economic Review* **80**(3): 465–479.

Scrucca, L., A. Santucci, and F. Aversa. 2007. Competing Risk Analysis Using R: An Easy Guide for Clinicians. *Bone Marrow Transplantation* **40**(4): 381–387.

Sekhon, J. S. 2011. Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching Package for R. *Journal of Statistical Software* **42**(7): 1–52.

Shleifer, A. and L. H. Summers. 1990. The Noise Trader Approach to Finance. *Journal of Economic Perspectives* **4**(2): 19–33.

Sias, R. W. 2004. Institutional Herding. *Review of Financial Studies* **64**(4): 1517–1548.

Simonsohn, U., and D. Ariely. 2008. When Rational Sellers Face Nonrational Buyers: Evidence from Herding on eBay. *Management Science* **54**(9): 1624–1637.

Stein, J. C. 2009. Sophisticated Investors and Market Efficiency. *Journal of Finance* **17**(1): 165–206.

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

Sun, M. 2012. How Does the Variance of Product Ratings Matter? *Management Science* **58**(4): 696–707.

Tiebout, C. M. 1956. A Pure Theory of Local Expenditures. *Journal of Political Economy* **64**(5): 416-424.

The Economist. 2000. A Survey of E-commerce: in the Great Web Bazaar. *The Economist* February 24, 2000.

The Economist. 2013. Crowdfunding in America - End of the Peer Show. *The Economist* June 1, 2013.

Toubia, O., and A. T. Stephen. 2013. Intrinsic vs. Image-Related Utility in Social Media: Why do People Contribute Content to Twitter? *Marketing Science* **32**(3): 368–392.

Wang, J. JD, and J. F. Zender. 2002. Auctioning Divisible Goods. *Economic Theory* **19**(4): 673–705.

Wang, R. 1993. Auctions versus Posted-Price Selling. *American Economic Review* **83**(4): 838–851.

Wei, Z., and M. Lin. 2013. Auction vs. Posted-Price: Market Mechanism, Lender Behaviors, and Transaction Outcomes in Online Crowdfunding. *Working Paper.*

Welch, I. 1992. Sequential Sales, Learning, and Cascades. *Journal of Finance* **47**(2): 695–732.

Wilson, R. 1979. Auctions of Shares. *Quarterly Journal of Economics* **93**(4): 675–689.

Wooldridge, J. 2002. Econometric Analysis of Cross Section and Panel Data. The MIT Press.

Wu, L. 2013. Social Network Effects on Productivity and Job Security: Evidence from the Adoption of a Social Networking Tool. *Information Systems Research* **24**(1): 30–51.

Xiang, T. 2012. *How Does Douban Make Money?* http://www.technode.com/2012/12/26/how-does-douban-make-money/

Xiao, M. 2010. Is Quality Certification Effective? Evidence from the Childcare Market. *International Journal of Industrial Organization* **28**(6): 708–721.

Xu, L., T. Nian, and L. Cabral. 2014. What Makes Geeks Tick? A Study of Stack Overflow Careers. *Working paper.*

Zhang, J., and P. Liu. 2012. Rational Herding in Microloan Markets. *Management Science* **58**(5): 892–912.

Zhang, P. 2009. Uniform Price Auctions and Fixed Price Offerings in IPOs: An Experimental Comparison. *Experimental Economics* **12**(2): 202–219.

Zhang, X., and F. Zhu. 2011. Group Size and Incentives to Contribute: A Natural Experiment at Chinese Wikipedia. *American Economic Review* **101**(4): 1601–1615.

Zheng, L. 1999. Is Money Smart? A Study of Mutual Fund Investors' Fund Selection Ability. *Journal of Finance* **54**(3): 901–933.