## ABSTRACT

**TITLE OF DISSERTATION** Decisions under Uncertainty in Decentralized Online Markets: Empirical Studies of Peer-to-Peer Lending and Outsourcing

**DISSERTATION DIRECTED BY:** Professors Siva Viswanathan and Hank Lucas
Department of Decision, Operations and Information Technologies
Robert H. Smith School of Business
University of Maryland, College Park

Recent developments in information technologies, especially Web 2.0 technologies, have radically transformed many markets through disintermediation and decentralization. Lower barriers of entry in these markets enable small firms and individuals to engage in transactions that were otherwise impossible. Yet, the issues of informational asymmetry that plague traditional markets still arise, only to be exacerbated by the "virtual" nature of these marketplaces. The three essays of my dissertation empirically examine how participants, many of whom are entrepreneurs, tackle the issue of asymmetric information to derive benefits from trade in two different contexts. In Essay 1, I investigate the role of online social networks in mitigating information asymmetry in an online peer-to-peer lending market, and find that the relational dimensions of these networks are especially effective for this purpose. In Essay 2, I exploit a natural experiment in the same marketplace to study the effect of shared geographical ties on investor decisions, and find that "home bias" is not only robust but also has an interesting interaction pattern with rational decision criteria. In Essay 3, I study how the emergence of new contract forms, enabled by new monitoring technologies, changes the effectiveness of traditional signals that affect a buyers' choice of sellers in online outsourcing. Using a matched-sample approach, I show that the effectiveness of online ratings and certifications differs under pay-for-time contracts versus pay-for-deliverable contracts. In all, the three essays of my dissertation present new empirical evidence of how agents leverage various network ties, signals and incentives to facilitate transactions in decentralized online markets, form transactional ties, and reap the benefits enabled by the transformative power of information technologies.

# DECISIONS UNDER UNCERTAINTY IN DECENTRALIZED ONLINE MARKETS: EMPIRICAL STUDIES OF PEER-TO-PEER LENDING AND OUTSOURCING

by

Mingfeng Lin

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2010

Advisory Committee:

Professor Siva Viswanathan, Chair
Professor Hank Lucas, Co-Chair
Professor Ritu Agarwal
Professor Joseph P. Bailey
Professor Il-Horn Hann
Professor Nagpurnanand R. Prabhala
Professor John Rust, Department of Economics, Dean's Representative

# ACKNOWLEDGEMENTS

I thank all my examining committee members for their comments and suggestions, especially the guidance of Professors Siva Viswanathan and Hank Lucas. Financial support from the Ewing Marion Kauffman Foundation, the Economic Club of Washington, and the NET Institute is gratefully acknowledged. All errors remain my own.

# TABLE OF CONTENTS

# LIST OF FIGURES

# List of Tables

# INTRODUCTION TO DISSERTATION

Developments in information technologies, especially Internet and Web2.0 technologies, have not only reduced the costs of communication but also created new mechanisms for individuals and firms to interact with each other. One of the most salient changes is the emergence of Internet-based, highly decentralized marketplaces populated with large numbers of small participants. Without the Internet, these atomistic individuals are likely to find the transaction costs prohibitively high, so they either choose not to enter the market at all, or enter via intermediaries. By reducing the fixed cost of business transactions, the Internet has significantly changed this landscape. A "long tail" now exists not only in product variety, but also in the number of buyers and sellers of goods and services. On the other hand, while more opportunities for trade are likely to increase social welfare, the growing number of suppliers and buyers inevitably increases the costs of effectively matching them. Information asymmetry problems that plague traditional markets still exist in these nascent marketplaces – only to be exacerbated by the anonymity and small-stake nature of individual transactions. A better understanding of the matching process in this marketplace has implications for academic researchers, policymakers, as well as entrepreneurs who seek to leverage the power of these online markets.

With this as the background, I set out to empirically study how individuals in decentralized online marketplaces make decisions under uncertainty to form transaction ties. Two emerging marketplaces serve as the context for my research. The first one is online peer-to-peer lending, where individual lenders make unsecured loans to borrowers. This market has experienced significant growth worldwide in the past few years. Data from

Prosper.com, one of the largest peer-to-peer lending websites in the United States, is used for the first two essays. The second one is online software outsourcing, where buyers and sellers of customized software from around the world can make transactions with each other via the Internet. Proprietary data from a leading online software-outsourcing marketplace are used in the third essay. Both contexts present degrees of asymmetric information that are significantly higher than online product markets such as eBay, due to the nature of financial products (Essays 1 and 2) and software development contracts (Essay 3).

The first essay in my dissertation specifically addresses the role of online social networks in addressing adverse selection in financial lending. I seek to link social network metrics to loan-level transactional outcomes, distinguishing between structural and relational aspects of the network. I emphasize the different identities, roles and actions of a borrower's friend, and whether a particular dimension of social networks can serve as an effective mechanism to mitigate information asymmetry. I test and find that *online* social networks serve as "prisms" that help signal the credibility of a borrower to those outside the network. More importantly, the more verifiable these ties are, the more strongly they are associated with the ex-post riskiness of the loan. This effect survives a large number of robustness tests, including the textual content of descriptions as well as the images used in the loan request. Online social networks indeed can help mitigate asymmetric information and improve transaction efficiency in online peer-to-peer lending.

The second essay delves further into the networks on Prosper.com to study the dyadic relationship among market participants. Drawing on theories of homophily and home bias, I investigate whether investors are more likely to invest in borrowers from their home state. To address this question, I exploit a natural experiment on Prosper.com where lenders were constrained to one state, while borrowers came from almost all states. The start and end of the 10-day window were also largely unexpected. These unique features allow me to circumvent many empirical analysis issues in prior studies, such as endogeneity, strenuous data reduction, and specialized statistical methods. I find that even though this is an online marketplace, investors are still more likely to bid on loan requests from same-state borrowers; however, such benefit only accrues to borrowers with good credit grades. In fact, there is a bias against less creditworthy borrowers. I further show that the economic distance between borrower and lender states has a stronger effect on the decision of lenders than the spatial distance. These results represent very conservative evidence of home bias, and how shared-geographical ties between borrowers and lenders affect lender behavior.

The third essay uses data from the emergent online market for software outsourcing, where software buyers and sellers (developers) from around the world participate in a decentralized online marketplace. The development in technologies allows buyers in such online labor markets to effectively monitor the effort level of sellers, making it possible for buyers and sellers to enter pay-for-time contracts. I study how the change in contract formats – pay-for-time contracts versus pay-for-deliverable contracts – affect how buyers interpret different signals from sellers, and choose the seller to work with. I focus on two signals that the literature has shown to be effective signals in online markets: online reputation and certifications. Data used in this study include comprehensive information about all developers who compete for buyers' contracts, including those who lost in the auctions. While the literature typically holds that the online reputation system has a strong influence on individual choice in e-commerce, in this context, I find that it is only under pay-for-deliverable (PFD) contracts that higher reputations lead to better chances of winning a contract. When the contract format is changed to pay-for-time (PFT), buyers are more likely to take risk, giving new entrants (sellers) more opportunities to grow. In other words, changes in the contract mechanism can help reduce market concentration and increase competition. On the other hand, certifications do not have statistically significant impact on buyers' choice under either contract forms. I further conduct exploratory textual analysis of the private communication between buyers and developers. Results show that different categories of information have different impact on buyers' choice under different contract mechanisms. Whereas prior empirical studies of outsourcing were often restricted to buyer-seller dyads that were ultimately in the contract, this dataset provides insights into the choice process of outsourcing clients. Most importantly, results in this essay show that as technology enables new contracting forms, the "Matthew Effect" in online reputation systems can be mitigated, as buyers substitute second-hand information from other buyers with first-hand information gathered through their own interaction with sellers.

Overall, the three essays of my dissertation investigate two industries where the development of information technologies has not only significantly changed the relationship between trading partners, but also enable new mechanisms that allow efficient matching. These studies contribute to the literature and practice in the following ways.

First, the three studies of my dissertation provide rich empirical evidence on how transactional ties in decentralized online markets are formed. These transactional ties can be borrowing and lending, or software development. Essay 1 focuses on the role of online social networks in this process; Essay 2 emphasizes the role of shared demographic information (geography); and Essay 3 examines the moderating effect of contract mechanism on the relation between various signals and the formation of transaction ties.

Second, my dissertation contributes to a growing literature on trust and reputation mechanisms in online markets. The first essay highlights online social networks as a new mechanism to mitigate information asymmetry and encourage trusting relationships; Essay 2 emphasizes homophily-based trusting behavior; and the third essay, in particular, directly examines the effectiveness of online reputation systems. While online reputation systems have a tendency to create a Matthew Effect whereby larger sellers are more likely to be chosen, we can potentially alleviate this issue by revisiting the contracting relationship between trading partners.

Last but not the least, all three studies share a focus on “small” players in these emerging markets, including entrepreneurs. Entrepreneurs can be on either side of the online lending market, and can also be on either side of the software development market. While the peer-to-peer lending model can serve as a new channel for small business financing, online software outsourcing can help small business buyers reduce the costs for software development, and enable developers to expand their market scopes. With an emphasis on the unique features of these markets, the findings of these studies can potentially increase the efficiency of these markets and further benefit market participants.

# ESSAY 1: JUDGING BORROWERS BY THE COMPANY THEY KEEP: SOCIAL NETWORKS AND INFORMATION ASYMMETRY IN ONLINE PEER-TO-PEER LENDING[1]

## Abstract

I study the online market for peer-to-peer (P2P) lending, in which individuals bid on unsecured microloans sought by other individual borrowers. Using a large sample of consummated and failed listings from the largest online P2P lending marketplace – Prosper.com, I test whether a borrower's online social networks can help mitigate information asymmetry between borrowers and lenders, focusing on the distinction between the structural and relational aspects of networks. While the structural aspects have limited to no significance, the relational aspects are consistently significant predictors of lending outcomes, with a striking gradation based on the verifiability and visibility of a borrower's social capital. Stronger and more verifiable relational network measures are associated with a higher likelihood of a loan being funded and lower interest rates for the borrowers, and lower risks of default for lenders. I discuss the implications of my findings for the disintermediation of financial markets and the design of decentralized electronic markets.

[1] An earlier version of this study was published in the Proceedings of the 30th International Conference on Information Systems (Lin, Prabhala & Viswanathan 2009).

# 1. Introduction

The ability of online markets to efficiently bring together buyers and sellers has transformed businesses, spawned several success stories, and redefined the roles of traditional intermediaries. In this paper, I study the online market for peer-to-peer (P2P) lending, where individuals make unsecured microloans to other individual borrowers. This market was virtually nonexistent in 2005 but has experienced significant growth since then. The biggest of them, Prosper.com, has logged over 200,000 listings seeking $1 billion in funding since its inception. Because the online P2P lending marketplace is highly decentralized with little opportunity for face-to-face contact between borrowers and lenders, the asymmetric information problems in traditional credit markets are especially amplified. My study examines whether social networks alleviate information asymmetry, and if so, what aspects of these networks help. I show that the social networks, specifically the relational aspects of networks based on the roles and identities of the network members, matter. Networks mitigate adverse selection and lead to better outcomes in all aspects of the lending process.

Borrowers in the P2P market can create social networks and make these networks visible to potential lenders. I focus on two such highly prominent networks – a borrower's friendship ties, and a borrower's group membership. I examine whether these social networks – friendship ties and group affiliations – predict loan outcomes and find affirmative evidence. I then focus on the nature of borrowers' social networks, particularly on the roles and identities of the friends that comprise a borrower's social network. For instance, I examine whether a friend has undergone the verification necessary to become a Prosper.com lender and test whether these have more pronounced effects on loan outcomes. Further into the friends “hierarchy”, I observe whether the friend has actually placed bids on the borrower's listing or other listings, and whether these bids were successful or not. In other words, I observe different grades of “the company that borrowers keep.” I examine whether networks explain the probability of attracting funding, the interest rate paid by borrowers, and the ex-post loan defaults and whether these economic effects become more pronounced along the relational hierarchy of the borrower's social network.

I will summarize the results briefly first. My sample comprises 205,132 listings on Prosper.com from January 2007 to May 2008 seeking to borrow an aggregate amount of $1.7 billion. Of these, 16,500 listings for $114 million are successfully funded. I first analyze the probability that a listing is funded. I find that a borrower's friendship ties are significantly related to loan outcomes and the effects show a significant gradation effect based on the roles and identities of the borrower's friends. The stronger a tie and the more verifiable and visible it is to lenders, the greater the probability of attracting funding. Interestingly, non-actions by particularly verifiable relationships, such as non-participation by lender-friends, lead to less favorable outcomes. I find similar effects when social capital is measured using group affiliations: verifiable antecedents matter.

To assess the price effects, I include networks as an explanatory variable in a censored regression that explains loan rates conditional on the loan attracting funding. Social network variables reduce the interest rates on funded loans. While the above mentioned outcomes – probability of a borrower's loan being funded, and the interest rate of funded loans – indicate that specific features of a borrower's social network do indeed affect lenders' decisions, they do not shed light on the rationality of these decisions. To examine whether the funding and interest rate decisions made by lenders are indeed rational, I examine the impact of the social network variables on default probabilities. I use survival models to test whether social networks have information about default probabilities beyond that contained in traditional credit variables. I find that borrowers with stronger and verifiable relationships are less likely to default. The results survive several robustness tests including analyses of subsamples without images, controls for non-linearities, as well as contents of images and text descriptions in listings. I describe these empirical analyses and the results subsequently in the paper.

The theoretical motivation for my tests comes from research in multiple disciplines including sociology, economics, finance and information systems. Section 2 reviews the related work more extensively. One strand of literature focuses on the relation between social capital and economic outcomes (e.g., Granovetter 2005). To this literature, I offer new evidence on the role played by social capital. As Granovetter (1972) writes, social capital is conventionally conceptualized either as an individual attribute that generates an economic benefit or as a group attribute of a collection of individuals that enhances the efficacy of transactions between the individuals for economic gain. Examples in sociology include Coleman (1988), Mizruchi (1992), or Putnam (1993). My results indicate that social capital between individuals plays another role: it serves as an additional source of verifiable information, and generates an informational externality that can be harvested to facilitate transactions with outsiders such as lenders in financial markets.

In the management literature, Podolny (1993, 2001) articulates related resource and informational perspectives of social networks. Podolny argues that ties between two market actors in the network can be understood as “pipes” that convey resources or information between them. Alternatively, the ties can be “prisms,” or informational cues that others rely to draw inferences about the quality of one or both of the market actors (Podolny 2001, page 34). My results represent novel evidence for the latter view. I find that a borrower's social network serves as a prism through which potential lenders deduce which borrowers to fund and at what interest rate.

My findings are of interest from at least two other perspectives. One view of my study is that it represents data from credit markets that have an especially severe Akerlof (1970) style lemons problem. In the P2P marketplace, loan funding is achieved through bids of small lenders who put small sums of money to work. Virtually all lenders are strangers who possess little private information about the credit quality of borrowers. From this viewpoint, my findings represent evidence that agents adapt to mitigate adverse selection in ways remarkably consistent with economic theory. I find a positive role for soft information, i.e., fuzzy, hard-to-quantify information about borrowers beyond hard credit data such as ratings. Small lenders in the P2P market seem to process and use the soft information implied in borrowers' networks in their lending decisions.

While the literature on soft information (Agarwal and Hauswald, 2010; Petersen and Rajan, 2002; Rajan, 2002) emphasizes that it is critical to successful lending outcomes, such information is traditionally regarded as a province of financial intermediaries such as banks (Fama, 1985). It is interesting that soft information even arises in a non-intermediated credit markets in which individuals have small sums of money at stake. In this regard, my results also imply that concerns that electronic markets and disintermediation could lead to significant loss of soft information may be mitigated.

While information technology could subtract some forms of soft information, it could also bring in new forms of soft information such as a borrower's social capital. A related point is that the new sources of soft information need not be intermediary-generated: in my study, borrower-generated soft information usefully adds to agents' decision-making. From a market design perspective, my study emphasizes that such new information is most useful when there are credible mechanisms to enhance its verifiability to outside lenders.

An alternative perspective of my study is that it analyzes the economic value of social networks. While it is widely accepted in economics and sociology that networks matter, especially to the sets of individuals forming the networks and the organizations that employ them, my study quantifies its value and puts additional boundary conditions to the claim. I find that networks are valuable not merely to individuals or organizations forming or containing them, but also to third-party outsiders, by helping mitigate informational asymmetry and adverse selection problems between the individuals in the network and outsiders.

My study is of separate interest because of its focus on the economic value of *online* social networks, an area in which there is little prior work. While online networks may be as valuable as their offline counterparts, this is not obvious given the ease of creating and building them. The P2P lending marketplace is an especially interesting context to study online networks because such networks are integral to the marketplace. Furthermore, my study addresses a major limitation of the received work on online networks: the difficulty of quantifying economic outcomes or the strength of ties. This necessitates costly methods such as surveys or interviews (e.g. Karlan 2007; Moran 2005;

Uzzi 1999), or subjective measures of outcomes (e.g., Bagozzi and Dholakia 2006; Uzzi and Lancaster 2003). In my study of credit markets, both the network itself and the associated economic outcomes are quantifiable using relatively objective measures such as funding probability and interest rates.

The rest of the paper is organized as follows. Section 2 provides theoretical background and reviews the related literature. Sections 3 and 4 describe the research context and the data used in the study. Section 5 describes the empirical methodology, and Section 6 contains the results of the study. Section 7 provides details on the robustness checks. Section 8 discusses the implications of my results and concludes.

## 2. Theoretical Motivation and Literature Review

My study draws on and contributes to multiple streams of research in sociology, economics, finance and information systems. To place my findings in context, I review the related literature.

The literature on social capital originates in sociology, but its role in facilitating economic exchanges and affecting behavior has attracted considerable attention in other disciplines. Granovetter (2005) overviews applications in such diverse areas as labor economics, price setting, production, financial innovation, and entrepreneurship. Recent work on the role of social capital includes Guiso, Sapienza and Zingales (2004), and Sapienza, Toldra and Zingales (2007). One issue that often arises in this literature is the identification of social capital. As Granovetter (2005) writes, social capital is best thought of as being generated by actions, patterns, or processes of people outside the immediate economic context and the issue is whether economic gains arise as a by-product.

Social networks are viewed as promising avenues for identifying social capital, so the literature often identifies an individual's social capital using her social networks. For instance, Burt (1992, page 9) describes an individual's social capital as “friends, colleagues, and more general contacts through whom you receive opportunities to use your financial and human capital.” Portes (1998, page 6) adds that there is growing consensus on social capital being “the ability of actors to secure benefits by virtue of membership in social networks or other social structures.” Durlauf and Fafchamps (2004) survey the methodology used in empirical studies of social capital. They argue that the “most successful theoretical work and the most compelling empirical work” is the role of networks in facilitating economic exchange. My work, which studies how social capital is leveraged through networks, has precisely this focus.

Theory suggests two avenues by which an actor's social network can influence transactional outcomes. Social networks can act as a direct channel for the transfer of information and resources. This role of social networks is termed as “pipes” by Podolny (1993, 2001). As noted by Granovetter (1973), information can flow through links, thereby either reducing the search costs for individuals, or enabling the gathering of heterogeneous information from different parts of the network. Within organizations, individuals occupying certain positions in the network can enjoy better information, easier access to resources, and therefore enjoy more power. Here again, the network ties serve as the channel for the flow of resources and utility accrues to the individuals on the nodes.

Alternatively, Podolny (1993, 2001) argues that social networks can serve as “prisms” that reflect otherwise unobservable characteristics. When networks play this role, it is critical that the social networks be credibly verifiable. Verifiability is particularly important in online networks because the ease of forming online networks may compromise their credibility. The verifiability issue is also important from a behavioral and marketing perspective, rather than a purely economic one. As noted by Rosenthal (1971), a message " must be testable by means independent of its source and available to its receiver"to be verifiable. Such requirement of verifiability applies to online social networks as well. A borrower's social network is also subject to skepticism from lenders; it is credible only to the extent it is verifiable. In the online P2P lending markets, a borrower's social networks can play the role of "pipes" by serving as a conduit for borrowers to obtain financial or informational resources. Alternatively, a borrower's networks can serve as "prisms" to signal the quality of creditworthiness of the borrower to potential lenders. Whether social networks serve as "pipes" or "prisms" in this market is another empirical question that I seek to address in this study.

The social networks literature also offers a useful taxonomy of the different dimensions of social capital (e.g. Granovetter 1992; Moran 2005). Structural embeddedness refers to the position of an actor in the network. Relational embeddedness refers to the quality of the relationship among actors in the network. Empirical evidence on structural aspects includes work on venture capital by Shane and Cable (2002) and Hochberg, Ljungqvist and Yu (2007). Studies in sociology and management show that certain positions on a given network endow individuals control over resources, e.g., individuals in hubs, those with weak ties (Granovetter, 1972) or those occupying structural holes (Burt 1992). Bampo, Ewing, Mather, Stewart, and Wallace (2008) investigate the structure of digital networks on the performance of viral marketing campaigns. Studies of relational embeddedness include Grewal, Lilien, and Mallapragada (2006) for open software projects; Robert, Dennis and Ahuja (2008) for knowledge integration and performance in digitally-enabled teams; and Cowan, Jonard, and Zimmerman (2007) for the networks of collaborators. My study stresses the importance of relational aspects, the roles and identities of the actors on an individual's networks.

My study also adds to an extensive literature in finance and economics on credit markets. A key theme in this literature is information asymmetry, which presents itself through ex ante adverse selection and ex post moral hazard. The ex-ante information asymmetry considers the Akerlof (1970) style adverse selection problems in lending. Social networks can provide information relevant to lending outcomes. If someone who knows the borrower personally can attest to his or her creditworthiness, or even better, participate in lending to the borrower, the loan should be relatively less risky. Obtaining and transforming such information into a usable format was traditionally difficult. Digitization and information technology has helped overcome this constraint. The key issue with such information is its reliability, which can be mitigated if the marketplace has credible verifiability standards.

Empirical studies seek to understand “soft information” in financial intermediation. Using survey data on small business loans, Petersen and Rajan (1994) find that soft information in bank or supplier relationships could increase the supply of credit to small firms. Organizational researchers apply social network theories to the banking sector. Using a social embeddedness approach, Granovetter (1985) and Uzzi (1999) study how bank-borrower relationships affect a firm's acquisition and cost of capital and introduce the idea of networks in these papers. My study adds to this literature by focusing explicitly on the role of social networks as a source of soft information. I illustrate how technology hardens it into usable form for lenders, show its use in lending decisions, and quantify its effect. I show that technological progress can lead to soft information being used even in disintermediated credit markets.

An alternative view of intermediaries is that they economize on search costs of matching borrowers and lenders. The Internet lowers the cost of search (Malone, Yates, and Benjamin 1987), making disintermediated search more viable. Additionally, digitization based on Web 2.0 technologies alters the way in which users connect and interact with each other. This results in new sources of soft information and social capital and new methods of transmitting the information. The net effect of these forces is to facilitate the growth of lending networks that are decentralized.

There is a small but growing body of research on peer-to-peer lending that focuses on the personal characteristics of borrowers to test theories of taste-based discrimination (Pope and Sydnor 2008; Ravina 2008). Pope and Sydnor examine loan listings between June 2006 and May 2007 while Ravina examines listings for a one month period between March 12, 2007 and April 16, 2007. Both papers focus on race and "beauty" of the borrowers, which I control for in my analysis. My main objective is the effect of social networks after controlling for these potential confounding factors.

Online P2P lending can also be viewed as a digitized and somewhat modified version of traditional microfinance programs (see, e.g., Morduch 1999 for a review of this work). Like peer-to-peer lending, microfinance is typically collateral-free and there are similar information asymmetry problems in both settings. My results suggest an additional parallel. P2P lenders also appear to rely on soft collateral implied by social networks or group attributes for repayment, as do lenders in microfinance (Ledgerwood 1999). The scalability of P2P lending across geographic regions and borrower types suggest that digitization and technology could help mitigate the problems that limit the scaling of traditional microfinance programs.

Finally, the term *peer-to-peer* has also been used to contrast different configurations in computer networks. Whereas in a client-server architecture, one server occupies a central position in the network, in a peer-to-peer architecture all computers are on equal footing – there are no hierarchies and all computers are *peer nodes*. The decentralized, distributed nature of P2P networks is usually considered an advantage in computer networking. Over time, the idea of decentralized *peer-to-peer* networks has been adapted to wider contexts. A well-known example is peer-to-peer music sharing (e.g. Bhattacharjee et al, 2007; Asvanund, Clay, Krishnan and Smith 2004). These studies typically investigate the impact of the growth of P2P music sharing networks on sales and consumer welfare, and uses' decision to contribute resources. Krishnan, Smith, Tang and Telang (2006) further explores economic issues related to business models that can be built upon file-sharing P2P services. More recently, IS researchers are also increasingly interested in online social networks, such as co-purchase networks and recommendation networks (e.g. Oestreicher-Singer and Sundararajan 2008, 2009). The unique feature of online P2P lending relative to this work is the availability of objective transactional outcomes of both the structural and relational aspects of networks.

## 3. Institutional Background on P2P lending

My data come from a leading online peer-to-peer lending website, Prosper.com. Prosper.com opened to the public on February 5th, 2006. At the end of 2008, it had

830,000 members and more than $178 million in funded loans. The following paragraphs describe the lending process and the information provided by borrowers seeking loans during the duration of my study.

### *3.1 Verification*

Users join Prosper.com by providing an email address, which is verified by the website. To actually engage in a transaction, users must go through additional verification. Borrowers must reside in the U.S., have a valid social security number, a valid bank account number, a minimum FICO (Fair Isaac Credit Organization) credit score of 520, and a valid driver's license and address. The details are verified by Prosper.com, which also extracts a credit report from Experian, a major credit reporting agency in the US. Loan proceeds are credited to the bank account and funds withdrawn automatically for monthly loan repayments. In the time period I study, borrowers can borrow a maximum of $25,000, and a maximum of 2 concurrent loans. Loans amortize over a 36 month period. Prosper lenders are also subject to verification of the social security number, driver's license number, and bank account number. To protect privacy, the true identity of borrowers and lenders is never revealed in the website. Communication occurs through usernames that are chosen when signing up.

### *3.2 Listing*

A loan request is a listing that indicates the loan amount and the maximum interest rate that the borrower will pay. A borrower can also post images and write a free-format description to accompany the listing. Neither the image nor the text is verified by the website. Borrowers choose an auction format. Closed auctions close as soon as the total amount bid reaches the amount sought at the borrower's asking rate. In the open format, the auction remains open even after the entire amount requested is funded for up to 7 days. During this period, lenders continue to bid down the the interest rate of the loan.

Figure 2 provides a screenshot of a sample listing. The listing displays information from the borrower's credit such as the number of credit inquiries in the 6 months prior to the listing (number of times that the borrower requested loans from banks, or applied for other kinds of credit cards)[2] , the debt-to-income ratio, and a letter credit grade, which is a coarse version of the borrower's FICO score. The credit grade ranges from AA (high quality) to HR (low quality) high risk borrowers. The correspondence between letter credit grades and the actual FICO score is shown in Table 1.1. During my sample period, lenders could not see the borrower's actual score but only the letter grade The purpose of a loan is tagged as a field in the listings. The listing indicates information about the borrower's friends and groups to which he belongs. With all the above information in place, the listing can go live to solicit bids from lenders.

### *3.3 Bidding*

When a lender sees a listing, she can decide whether or not to lend to the borrower. An important feature of online peer-to-peer lending is that an individual lender does not have to finance the entire loan request. A lender can bid an amount of $50 or more and specify the minimum interest rate she desires. The actual bidding process uses a proxy bidding mechanism. If the loan has not yet been funded 100%, the ongoing interest rate will be the borrower's asking rate, even if the lenders' minimum rate is lower. Once 100% of the requested funding has been reached and the format of the auction is open,

[2] Note that this is not the number of prior loan requests on Prosper.com.

the ongoing rate decreases as the lender with the highest rate-bid is competed out. In a sense, the auction is similar to a second-price auction. All bids are firm commitment bids and no withdrawals are allowed. From a lender's viewpoint, a bid could win or be outbid, in which case the lender can place another bid to rejoin the auction. Once the auction ends, the loan could be fully funded. If not, the auction is deemed to have failed and no funds are transferred.

### *3.4 Post-bidding, funding and repayment*

Once the bidding process ends, the listing is closed and submitted to Prosper staff for further review. Sometimes, additional documentation is required of borrowers. Once the review process is completed, funds are collected from the winning bidders' accounts and transferred to the borrower's account after deducting fees of up to 2% of the loan amount.

Loans on Prosper.com have a fixed maturity of 36 months with repayments in equated monthly installments. The monthly repayment is automatically deducted from a borrower's bank account and distributed to lenders' Prosper accounts. If the monthly payment is made in time, the loan status for that month is considered current. If a monthly bill is not paid, the loan status will be changed to “late”, “1 month late”, “2 months late”, etc. If a loan is late for 2 months or more, it is sent to a collection agency. Lenders on prosper.com must agree that the proceeds of the collection represent the full settlement of loans. Delinquencies are reported to the credit report agencies and can affect borrowers' credit scores.

*3.5 Social networks*

Any Prosper.com member with a verified email account can create or join a friendship network. Friendships are typically created through the following process. After the inviting member fills out the friend's email address and a short message, Prosper.com generates an email message that contains a link to join Prosper. The recipient can click on the link contained in the email to sign up, or use the link to establish friendship with an already-registered account. In this regard, the presence of a friendship tie suggests that the two individuals have at least some offline, non-public information about each other, such as an email address.

Any member, including these friends, may elect to have no roles or elect to be verified as lenders or borrowers. Friendship ties are bi-directional. From an empirical viewpoint, a member's friendship network is visible on the profile page or a loan listing page. Other members can click through the link to see the profile information of those friends. Friends who bid on a listing are also tagged very clearly by a unique icon in the list of bids, so they are readily visible to other potential bidders.

A second type of social network on Prosper.com is a group. There are 4,139 groups in my sampling period. Any member can create a group and a member can typically join any group whose membership criteria are met. However, in my sample, each individual can be a member of only one group at a time. Entry or exit into a group is free but this bar is raised for borrowers. If a borrower is a member of a group when requesting a loan, the borrower cannot leave the group or join any other group until the outstanding loan is repaid in full. The leader of each group can determine the rules regarding who can become group members and how others may join. Some groups, such as alumni groups, typically require a high degree of verification. Applicants must prove that they are indeed affiliated with the institution. Other groups require very little verification.

## 4. Data

My sample comprises all listings that seek funding on Prosper.com between January 2007 and May 2008. I obtained information regarding borrower's credit history, their unique Prosper ID, the social network variables, features of their auctions, and outcome of their loan listings using an API provided by Prosper.com. I ensure that the descriptive fields in my analysis are in the information set of potential lenders. For instance, I gathered information on loan requests over time, so that information about borrowers' social networks is current at the time of the loan requests. I will now describe the variables used in my analysis and discuss some descriptive statistics.

### *4.1 Social networks*

In my sample, 56,584 listings report friends, and 41% of all listings are associated with a group. The friendship ties and group-based affiliations capture soft information in a borrower's social network. As can be seen from the screenshot of listings, information about friends that borrowers have a direct link to (first-degree neighbors on the friendship network) is much more salient than those who are two-degrees away and beyond. Since my focus is the “signaling” value of networks, remote neighbors are unlikely to affect the decision of lenders when their information is not directly observable. Therefore in the main body of the paper, I focus on network measure calculations based on these neighbors.

The literature provides two perspectives to understand the effect of social networks. The “structural” view emphasizes the topology of the network, while the “relational” view emphasizes the nature and quality of ties. Structural aspects of networks have received much attention from researchers. While many structural metrics can be calculated from my network, only some of them are theoretically justifiable as predictors for the three outcomes that I study. The basic metric of degree centrality (undirected) captures the number of friends that a borrower has at the time of the listing. It is possible that having more friends can be associated with a higher social pressure to repay loans. It is also conspicuously displayed on listing pages, hence likely to influence lenders' decision to bid. I therefore consider this metric as the basis for my analysis of the relational dimension of networks. One other category of structural metrics that could have an impact on loan outcomes are those related to network closure, or the degree to which friends of a borrower know each other. These ego-network metrics include constraint, efficiency, and effective size (Hanneman and Riddle 2005). These measures are also likely to impact a borrower's loan outcomes because the news of a default can easily diffuse throughout a borrower's social networks, and reputational consequences can motivate borrowers to honor their debt. I calculated these metrics in UCINET. However, as of the time of my study, the degree of closure in the network is quite low, and none of these metrics is consistently significant on three outcomes that I study. This structure is also confirmed in exploratory analysis of the largest few components of the network. Therefore, I focus on the degree centrality of borrowers as I delve into the nature of ties, or the relational aspect of their networks.

The relational social network measures concern the roles and identities of the members in the network. Figure 1 describes the hierarchical levels underlying my analysis. Level 1 distinguishes friends according to whether their identities are verified on Prosper versus individuals who have merely registered and are thus little more than a verified email address. Level 2 categorizes the verified friends based on their specific roles – whether these friends are borrowers or lenders. Lenders are individuals with extra financial capital while borrowers are likely to be facing financial constraints. On the other hand, borrowers are subject to greater scrutiny as they have verifiable credit grades that form a backbone of their listings. Level 3 further differentiates between *real lender* friends – those who have lent prior to the current listing; and *potential lender* friends – those who have provided enough information to Prosper to be listed as lenders but have yet to participate in any loan. Level 4 differentiates real lender friends according to whether they bid on the specific borrower's listing or not. Level 5, the finest classification, distinguishes between lender friends who bid on the borrower's listing and won and those who bid but did not win. As I progress from Level 1 to Level 5, the relationship between the borrower and lender becomes more actionable, verifiable and more strongly embedded.

In addition to a borrower's friendship ties, I consider membership of groups. I manually coded all groups that have at least 3 members and are active in the generation of loans. I categorize all groups into one or more of several categories and include dummies for groups. Groups can be based on self-identified categorization such as membership of a religious denomination or having hobbies of a specific kind. Alternatively, they could be based on verifiable antecedents, such as being an employee at a particular company or living in a particular geographic region such as the greater Washington DC area. Borrowers can choose what group to belong to but group membership cannot be reversed until loans are repaid. I also include controls for group size. However, its sign in the regressions is not clear because while larger groups involve more peer pressure they also result in less oversight of any one member. An interesting group variable is whether the leader of a group is rewarded for listings of group members that are successfully funded. These rewards create incentives similar to the originate-sell model of intermediaries held responsible for the 2008 financial crisis. This reward structure was discontinued by Prosper.com in October 2007. I include a dummy for group leader incentives in my analysis.

Friendship and group networks can benefit the borrowers in two different ways. If these networks serve the role of "pipes" (Podolny 2001), borrowers should be able to directly obtain resources from their friends or peer group members. These resources could be financial (i.e. direct funding from friends or group members), or informational (i.e. helping to make the loan request more appealing). If instead, borrowers social networks function as "prisms", then these networks should help the borrowers' signal their creditworthiness to potential lenders outside the borrowers' immediate social network. Summary statistics of my sample shows that the financial aspect of "pipe" effect is likely minimal: the funds contributed by friends or group members to a borrower's loan account for less than 5% for over 95% of loans associated with these networks, whereas most funding comes from lenders outside the borrowers' social networks. On the other hand, to control for the informational aspect of the "pipe" effect, I specifically address the potential confounding effects of texts and images in later parts of the paper.

### *4.2 Hard credit information*

Prosper.com provides a letter grade for each borrower ranging from AA to HR, which correspond to the credit scores listed in Table 1.1. In addition, I include the other hard credit information provided by the website, including a borrowers' debt-to-income ratio and the number of credit inquires (at banks or credit card companies) in the six months prior to the listing. I include these variables as additional credit indicators to allow for the possibility that the letter grade itself is not a sufficient statistic for credit risk. Rather than a numerical score (e.g., AA=1, A=2, etc.), I include a full set of dummy variables for each letter grade.

### *4.3 Other control variables*

In addition to the above soft and hard credit information, I also gather information on whether the auction is conducted via the open or closed format. The latter closes as soon as it is funded 100% and perhaps indicates borrowers with more urgent financial need. I include a dummy for the auction type. I also considered maximum auction duration, which could range between 3 and 10 days but has been since standardized to 7 days. This variable showed little significance in any of my models and I omit it.

Some states in the US have usury laws that enforce a cap on the allowable interest rate on consumer loans. While usury laws intend to protect customers, they could reduce the chances of successful funding if the supply curve for funds intersects the demand curve at a rate above a state's usury limit. Whether the laws have this bite or not is an empirical issue. After April 15 2008, Prosper started collaboration with a bank in an effort to circumvent that limit. My sample spans both periods, so I include a control for usury laws.

Each borrower indicates a maximum borrowing rate that she is willing to pay. I include this variable in quadratic format. While low rates indicate less profitable loans, high rates could also indicate less profitable loans because the effect of higher rates could be swamped by the greater likelihood of default for borrowers willing to pay high rates (Stiglitz and Weiss 1981). The setting of an intermediary and the sophisticated reasoning modeled in Stiglitz and Weiss is probably far from my setting of atomistic lenders bidding for a small piece of a listing. However, I include the quadratic term as a hypothetical possibility.

To control for broad lending rates, I purchased a proprietary dataset from a professional company. The data include the average interest rate for borrowers in each credit grade in each regional market for each month for 36-month loans. This variable not only serves as a proxy for the "outside option" of borrowers and lenders, but also reflects temporal and regional variations in consumer lending.

I further control for the purposes of loans. Borrowers indicate several types of needs, including debt consolidation, home improvement, business loans, personal loans for a variety of purposes (including vacations), or student or auto loans. The loan purpose is self-indicated by borrowers and can thus be thought of as cheap talk. However, potential lenders often communicate with borrowers during the auction process and seek more tangible details. In balance, there may still be some information in the loan purpose, so I include this in the regressions.

As a new business model, Prosper.com has received significant media exposure since its inception. Articles in the media make it more likely to attract new borrowers and lenders to the website after their publication. To control for the potential influence of such news, I include an additional variable to absorb these exogenous shocks to this marketplace. I download the search volume on Google for Prosper.com and construct a dummy variable based on whether there is a significant change in search volume, which I call *spikedays*. Finally, I include quarterly dummy variables to control for unobserved time effects.

## 5. Empirical Modeling and Identification

The theoretical motivation for my study comes from Akerlof (1970) style adverse selection models. Because both borrowers and lenders in the P2P market are small and must make their lending decisions in the face of considerable informational asymmetry, lenders are likely to use several informational cues or “signals” for their lending decisions. While hard credit variables (e.g., a borrower's credit score) are well-known signals of a borrower's quality, it is not known whether (and if so, what aspects of) a borrower's social network can serve as an additional signal of quality in this market. The main hypothesis that I examine is whether a borrower's social network helps agents adapt to this environment by acting as informational cues of borrower quality. If so, the basic testable hypothesis is that social networks should be associated with an increased chance of a successful listing and should reduce the interest rates of consummated loans. Affirmative evidence would be consistent with the joint hypothesis of an economic model in which (i) investors rationally adapt to informational asymmetry by relying on other informational cues of borrower credit quality; and (ii) networks provide such a cue.

The first set of tests described above rely on differences between borrowers with friendship networks and those without. If the hypothesis that friendship networks act as informational cues of borrower quality is correct, transactional outcomes should not depend on just friends per se, but also on the *type* of individuals that comprise the friendship network. The empirically testable proposition is that the network effects on both transactional outcomes should be especially pronounced when borrowers have better "quality" friends. My tests on the friendship hierarchy of borrowers represent precisely this test of differences in the "quality" of friends a borrower has. I test whether the network results are stronger when friends have verifiable antecedents, whether the antecedents reflect roles as lenders, the subset of lenders who actually bid, and the subset who bid and lend. I test whether the relation to transactional outcomes – successful listings and lower interest rates – strengthens as I progress down the hierarchy of friendship quality. In sum, the adverse selection theories predict not only that networks should matter but also that they should do so for stronger cues of quality, a proposition that I take to the data in a sequence of specifications

Furthermore, the hypothesis that networks act as informational cues to mitigate adverse selection results in yet another hypothesis. If networks are incrementally informative about borrower credit quality, the testable implication is that they should also be associated with lower ex-post defaults in funded loans. Furthermore, the effects of friendships on loan defaults should also follow the gradation discussed in the funding probability and interest rate tests. Better quality friendships that delve deeper into the hierarchy in Figure 1 should have greater effects in reducing loan defaults. I test both implications by tracking ex-post repayment histories of borrowers with successful listings and model the time to default using a flexible Cox hazard model. This test has the major empirical advantage that it is an out of sample test of the adverse selection hypothesis. The test is implemented on an entirely disjoint database far removed from the transactional data on loan funding or interest rates.

Endogeneity is a common concern in empirical modeling. One source of such concerns is reverse causality. This is unlikely in my context. Many prior studies of peer effects (e.g. Kremer and Levy 2003) reflect the “pipe” effects of networks: outcome such as drinking behavior could be a direct result of peers' actions, either through imitation or social pressure. Reverse causality is plausible in these studies because of homophily: individuals with the same behavioral patterns are also more likely to form ties. By contrast in my study, the outcome of interest is largely a result of the actions of those outside the agent's immediate network (friendship ties). Furthermore, because all operational transactions such as fund transfers and ex post monitoring of repayments are carried out by Prosper.com, there is little interaction between borrowers and lenders after loans are originated. Friendship motivation for lending is unlikely. Empirically, reverse causality is not a concern either. All loans on Prosper.com are three-year loans and most in my sample are first-time borrowers on Prosper. Even when borrowers and lenders do become friends online, it is not likely the result of previously repaid loans, as the time frame in my anaylsis does not allow such reciprocity to occur.

A second concern is the role of unobservables. Do my estimates truly reflect the effect of friendship variables or other unobservables? My research context makes this about as unlikely as reasonably possible in social studies with nonexperimental data. In my study, loans are funded by aggregated contributions from many small lenders, virtually all of whom are strangers with no private information about the creditworthiness of the borrowers. Hence, if a certain variable is unobservable to me, it is also unobservable to potential lenders and will not affect their behaviors. This identification strategy is consistent with Angrist (1998), who writes that he has access to information about "most of the characteristics used by the military to screen applicants," and is thus able to eliminate bias due to unobservables. My study has a similar or even stronger settings given that I observe and control for all information available to potential lenders.

As abundance of caution, I consider additional tests to further mitigate concerns about unobservables. I not only include an exhaustive set of controls that could predict loan defaults, but also go beyond and include non-standard variables derived from the text and images of loan listings. Loan listings on Prosper.com often contain images and some descriptive text attached to the loan listings. Because these variables are self-reported, it is not immediately clear that they could subsume the information content of friendships, and particularly the grades of friendships based on the roles of friends. Those variables are either verified by the website or based on a clearly verifiable trail such as bids or participation in prior listings. Nevertheless, I consider the effect of text and images in my analysis through two approaches. First, I separately analyze listings with or without images on Prosper.com. About half the listings in prosper post no images. To assess the sample with images, I randomly sample 10% of listings with images, and manually code the image content. Finally, I examine subsamples with descriptive text in listings and code the text content using disambiguation routines in the text analysis literature. These additional variables have little effect on my key results on networks or their relational grades based on the hierarchy of roles and identities of friends.

## 6. Results and Discussion

From the raw data described above, I construct two datasets for the three outcomes of my interest. For the funding probability and interest rate, I construct a dataset where each record is a loan request[3]. For the riskiness of loans, I construct a panel dataset where the unit is a funded loan, and each time period is one billing cycle. I then track the repayment history of all funded loans in the first dataset at each billing cycle. Overall, my sample has 205,132 listings with an average loan amount of $6,973. Of these, 56,584 (27.58%) report friends while 148,548 (71.42%) report no friends. The group of listings in which borrowers have friends is spread across the Prosper.com credit grade spectrum. For instance, of the 6,523 AA listings, 1,881 or 28.84% have friends, while of the 33,068 D grade listings, 9,462 or 28.61% have friends. In the high risk, or HR category, 22,556 out of 62,904 listings, or 26.39%, are associated with a friend. Listings in which borrowers have no friends have mean debt-to-income ratios of 58% while listings of borrowers with friends have debt to income ratios of 57%. Borrowers with no friends have about 4.17 credit inquiries (credit applications to banks or credit card companies) in the six-month period prior to the listing date against 4.22 inquiries for borrowers with friends.

In my data, 16,500 (8.04%) listings attract full funding. For the sample of borrowers with no friends, 10,410 out of 148,548 listings, or 7.04% are successfully funded, while 6,090 or 10.76% of listings where borrowers have friends are successfully funded. My other social network variable is group membership. 29% of all requests or 59,978 listings indicate a group affiliation. Of this sample, 28,006 listings, or 46.63%, are

[3] I subjected this dataset to different structures in my robustness tests, which I will describe later.

associated with an incentive structure in which group leaders are rewarded with a fee for successful listing.Likewise, 7.09% of listings not affiliated with a group are successfully funded, while 10.36% of listings with a group affiliation are funded.

To reiterate, my main research question is whether borrowers' online social networks can help mitigate information asymmetry in the P2P market. To this end, I need to study how network metrics relate to the three outcomes and if these relationships taken together are indeed consistent. I examine loan funding probability, interest rates and loan default in multivariate specifications. Table 1.2 provides a full list of the explanatory variables. Tables 1.3 and 1.4 describe the different models that I report in the paper and the set of variables used in each specification. For instance, specification P1 is a model of funding probability that uses variable set 1 (Table 1.3). From Table 1.4 I can see that variable set 1 corresponds to the root level of the friendship hierarchy and uses "ttlfriends", or the number of friends plus the "common variables" listed in Table 1.2 as explanatory variables. Section 5.1 reports the probit results while Sections 5.2 and 5.3 model the interest rate of funded loans and the probability of default. Each section focuses on the results relating to social network variables. Section 5.4 discussed the coefficients for control variables.

### *6.1 Funding probability*

Table 1.5 reports estimates of a probit model for the probability that a listing is successfully funded. I report six sets of results that include the social network variables in Figure 1 and common controls. I discuss the social network coefficients first and then examine the results for the control variables. As I will discuss further in the "robustness" section, results are consistent when I use a logit regression. Probit results are reported here because I can easily compare with the first stage of the Heckman model for interest rates.

### 6.1.1 Social networks: Friendship network

I will start by considering structural measures of networks. The degree centrality measures a borrower's position in the friendship network. Specification P1 in Table 1.5 shows that degree centrality is positively related to the probability of being funded. As I shall explain below, this relation reflects a more extensive relation caused by the roles and identities of the members of the friendship network. And as discussed earlier, other structural network measures such as effective size of network and efficiency have no significant effects on the probability of funding, largely due to the low degree of closure in the friendship network.

Specification P2 in Table 1.5 distinguishes friends according to whether their identities are verified on Prosper or not. This process effectively decomposes a borrower's degree centrality into two orthogonal pieces, friends who are verified and those who are not. I find that unverified connections, i.e., connection that merely signify a valid email address, represent insignificant cheap talk or even negative signals at the 10% significance level. In contrast, TTLROLE, which denotes friends with verified roles, has a positive coefficient that is significant at 1%. These results constitute the first evidence that roles and identities, or the nature of the company that borrowers keep, matter.

I next categorize friends based on their roles on Prosper.com. To this end, I decompose the verified friends into two orthogonal and additive pieces: friends with roles as borrowers and those with roles as lenders, both adding up to the total number of friends with roles. In addition to these two components of friends with roles, I also include the total number of friends with no roles. Specification P3 gives the results. Friends with no verified roles have negative effects as before. Connections to borrowers have insignificant effects while having friends with roles as lenders increases the probability of the loan being funded.

Specification P4 further differentiates between *real lender* friends, who have made loans on Prosper.com prior to the current listing, and *potential lender* friends who have not yet made loans on Prosper.com prior to the start of the current listing. I continue to include the excluded variables, all friends with no roles, and friends who are borrowers but not lenders as control variables. There is a continued gradation of the friendship effects. Having lender-friends matters only to the extent that the friends are real lenders who have already bid. The coefficient almost doubles relative to that of the total number of lender friends.

Specification P5 reports results when I decompose the real lender-friends into the ones who bid on the specific borrower's listing and ones who do not. At this level, it is also possible that a potential lender who has not lent in the past now chooses to initiate bidding with the current loan. Thus, I decompose both potential lenders and real (past) lenders into ones who bid on the current listing and ones who do not. I find positive and significant effects for potential lenders who bid on the current listing. Interestingly, borrowers with potential lender-friends who do not bid on the listing are less likely to get funded. In contrast, having real lender-friends bid on a listing elevates the chances of a successful funding. The funding probability equation does not decompose real bidders into those who win and those who lose because whether a bidder wins or not is observable only after the outcome is known.

In sum, I find that social capital, as reflected in borrowers' social networks, matters in attracting outside financial capital. In this regard, the structural aspects of the social networks are not necessarily critical. Rather, the role and identity of the network members are important. It is interesting that social capital matters even when the outside lenders are atomistic individuals participating in arm's-length transactions with the individuals possessing the social capital.

### 6.1.2 Groups

I next test whether group variables matter. Proceeding as before, I first consider group size, measured as the natural logarithm of number of members in a group. In specifications P1 through P4, larger groups are less likely to result in a successful listing. Perhaps the effect of default of a group member on the overall group credit quality declines when the membership is very high. Alternatively, members who choose to belong to a larger group are voluntarily foregoing membership of a smaller group, recognizing which potential lenders may become less willing to fund a listing. The group size variable loses significance in specifications P5 and P6. In terms of group type, university alumni groups and geography-based groups increase chances of funding. Membership of these groups is based on verifiable antecedents rather than self-reported identification. Interestingly, borrowers with religious affiliation are more likely to be funded. Perhaps individuals with religious affiliations default less. Alternatively, this could reflect tastes for lending to people with religious affiliations, in the spirit of the taste-based discrimination hypothesis (Becker 1971). I examine this issue in Section 5.3, which deals with loan defaults.

### *6.2 Interest Rates*

Section 6.1 shows that social network variables increase the probability of successful funding. I next examine whether these variables have complementary price effects. I regress interest rates on social network variables and controls. I control for the fact that interest rates are only observed when listings are successfully funded by using the familiar two-step method of Heckman (1979). The model can be identified either through the non-linearity intrinsic to selection models (an example can be found in Uzzi 1999) or through exclusion restrictions, and the results are similar. For exclusion restriction, I consider the variable SPIKEDAYS: when Prosper.com received media exposures, traffic volume to the site will increase significantly. Borrowing activities will increase, but lending will respond slower as lenders need to verify their bank accounts and transfer funds (which takes up to a week) before they can place bids. Hence SPIKEDAYS should be negatively associated with funding probability; meanwhile, interest rate should be more stable and determined by borrowers' credit circumstances. Empirically, this variable has an $F$-statistic exceeding 50, well above the cutoff of 10 for a strong instrument suggested by Staiger and Stock (1997). The results with the exclusion restriction are reported in Table 1.6; they are also highly consistent with those estimated without using any variable as exclusion restrictions.

The interest rate results in Table 1.6 are remarkably consistent with those for funding probability and default rates. The variables reflecting the role and identity of network members show a direction and gradation consistent with the results for funding probability. Connections to friends not verified by Prosper.com tend to increase interest rates, as reflected by the coefficient for the variable ttlNoRole. More importantly, connections to verified friends with lender roles have the opposite effect. Both connections to real lenders and those to potential lenders lower interest rates. Interest rates fall the most when real lender-friends who have participated in past loans on Prosper.com and also participate in the current listings. The effects are significant regardless of whether they win in the listing or not.

Interestingly, group variables also explain interest rates in a fashion largely consistent with the funding probability model. The group size itself has marginal statistical significance and little economic significance in explaining interest rates. The nature of the group matters more. Groups with a religious motif enjoy lower interest rates by between 70 and 200 basis points. Geography-based groups are consistent in models H1--H4 but not in models H5--H6. Groups based on business or university alumni affiliations, which tend to have verifiable criteria, show strong effects, lowering interest rates by close to 120 basis points.

*6.3 Loan defaults*

Prosper.com records the status of loans in each month, or payment cycle. Loans are current if repayments occur on time. Otherwise, loans can be “late,” “1 month late,” “2 months late,” and so on. I model a default as occurring if a payment is late by at least two months. As in the consumer finance literature (e.g, Gross and Souleles, 2002), I estimate survival models. I employ a Cox proportional hazards model (e.g. Grover, Fiedler and Teng 1997; Bhattacharjee, Gopal, Lertwachara, Marsden and Telang 2007; Cleves, Gould, Gutierrez and Marchenko 2008) in which the hazard $h(t)$ is specified as

$$h(t \mid x) = h_0(t) exp(x\beta) \tag{1}$$

where $h_0(t)$ is a baseline hazard rate, and $x$ denotes a vector of covariates. For each covariate $x_j$ in the Cox model, I report the exponentiated form of the coefficient $\beta$, which is called the *hazards ratio*, whose standard error is obtained using the Delta method (Cleves et al 2008, page 133). A hazards ratio greater than 1.0 for variable $x_j$ indicates that it increases the probability of default, while a ratio less than 1.0 indicates that $x_j$ decreases the probability of default. The smaller the ratio, the greater the effect on reducing the risks of default. The Cox hazards model estimates of $\beta$ can be used to recover estimates of the baseline hazard function (Kalbfleisch and Prentice 2002; Cleves et al 2008). As shown in Figure 3, the baseline hazard of default increases sharply at the beginning, reaching a peak at about 10 months, and then slowly wears off. This pattern is remarkably consistent with consumer lending delinquencies reported in Gross and Souleles (2002, page 327).

Table 1.7 reports exponentiated estimates of coefficients $\beta$ in equation (1). In specification C1, the total number of friends is insignificant as a predictor of default. Specification C2 decomposes friends into those with verified identities as lenders or borrowers and friends with no verification. Having more unverified friends increases the odds of default, as indicated by a hazards ratio of 1.05, while friends with verified identity decrease the odds of default. However, neither variable is significant. Specification C3 shows statistically significant effects for verified lender friends in a borrower's social network. The hazards ratio of 0.91 suggests that having lender-friends decreases default risk by 9% on average.

Specification C4 includes the number of lender-friends but controls for whether they actually participated in lending prior to the borrower's listing. The hazards ratio for real lender-friends is 0.88, indicating that having real lender-friends decreases the odds of default. Likewise, the hazards ratio is 0.86 when when I consider lender-friends who bid on the borrower's listing. Both coefficients are significant at 1%. The hazards ratio for friends who bid on and win a listing is 0.79 and is significant at 1%. Thus, the odds of default are significantly lower when lender-friends bid and win on the borrower's listing.

The result for the real lenders who bid indicate that financial stakes taken by friends are strong signals for outside lenders that a borrower is creditworthy. Alternatively, perhaps peer pressure is generated when friends take stakes in a borrower's listing. The data suggest that this is not a first order force because the median contribution of friends to a listing is less than 5%. The evidence is more consistent with a prism effect in which borrowers' attributes are reflected in the nature of the company they keep, i.e., serve as a source of soft information about borrower quality. Equivalently, the positive social capital communicated by friends who bid appears to be the major reason why social networks reduce defaults.

In terms of group characteristics, Table 1.7 shows that only two matter for loan performance, alumni groups and geography-based groups. Being members of these groups reduced the probability of default. Interestingly, both groups do not rely on self-categorization for membership but rely on verification to establish membership. None of the other groups are related to default risk. Group size is also unrelated to loan defaults.

### *6.4 Controls*

While Sections 5.1--5.3 focus on the role of social networks, I now turn to the major results for control variables. In terms of hard credit variables, credit ratings, the number of credit inquiries, and the debt to income ratio have the expected sign. For instance, listings with more recent credit inquiries (applications to credit card companies or banks), or with lower credit grades, are less likely to be funded, attract higher interest rates, and are more likely to default. Bank card utilization has a positive coefficient while its square has a negative coefficient in all three specifications. Some card utilization is beneficial as it signals creditworthiness. Very high utilization is undesirable because it signals stretched borrowers vulnerable to shocks and leads to lower funding probability and higher interest rates.

Auctions that close immediately when funding reaches 100% can encourage aggressive early bidding, enhancing funding probability but result in higher interest rates because there are no opportunities for lenders to bid down the interest rate. The results in Tables 1.5 and 1.6 support this view: closed auctions result in higher funding probability and higher interest rates. A closed auction may be indicative of weaker borrowers who are willing to forgo price competition, which could result in increased default rates. However, the hazard ratio for auction format is not significantly different from 1.0 in any of my specifications.

In the funding probability models, three types of loan purpose variables are significant. Business loans (listingcatg4) appear to be viewed as being more risky. These are less likely to be funded and when funded, attract higher interest rates. These loans are also about 24% more likely to default, though the result is only significant at the 10% level. Debt consolidation loans (listingcatg1) are more likely to be funded than other loans at lower interest rates, indicating that lenders value the fact that borrowers are using Prosper.com to shop interest rates or limit credit card debt. However, there is no guarantee that borrowers will necessarily adhere to their plans successfully. Debt consolidation loans are about as likely to default as other loans. Specifying some purpose, category increases the overall probability of funding and lowers interest rates but has little effect on default.

Borrowers willing to pay low rates may be less profitable to lenders and may be less likely to be funded. However, in the spirit of credit rationing theories, high rates may signal risky borrowers, who may also be less likely to be funded. The results in Table 1.5 supports this view. The linear term has a positive coefficient and the quadratic term has a negative sign, as predicted. While rationing theories argue for linear and quadratic terms in the funding probability equation, it is less obvious that there is a similar implication for the loans that are actually funded. The linear interest rate term is negative in four specifications and positive in two others while the squared term is consistently positive in all models. In unreported results where I estimate with the linear term alone, I find a positive and significant coefficient.

I also examine the effect of usury laws. In states with usury law limits on interest rates, riskier borrowers screened out of other credit markets may seek to come to Prosper.com, in which case lenders may perceive these borrowers as being riskier. The results in Tables 1.5 and 1.6 suggest that this is the case. Lenders are wary of borrowers from usury law states, who are less likely to get funded and when funded, pay higher interest rates. While the survival model point estimate for the usury law state coefficient is greater than 1.0, the difference is not significant. Group leader incentives matter. When group leaders have financial incentives for promoting listings, the listings are more likely to be funded, face lower interest rates, but are not less likely to default. I also test for potential effects of other variables that have marginal significance at best. An interesting variable is the number of years since a borrower's first credit line, a proxy for the borrower's age and credit experience. It has small effects on the funding probability and interest rate and no effect on default.

## 7. Robustness Tests

To examine the robustness of my findings, I conducted a large number of additional tests. I discuss the main results, especially the implications for social network variables, below. A full set of results is available from the authors but the results are qualitatively discussed below.

### *7.1 Additional Specifications*

I first consider different alternative specifications for the funding probability model. A logit regression yields highly consistent results as the reported Probit regressions. In addition, borrowers whose listings have failed can relist on Prosper.com with a fresh request, inducing correlations across listings. I thus constructed a panel dataset with each member as a unit, and each listing as a time period. The resulting panel is highly unbalanced because 53% of members post only one listing. Due to the incidental parameters, I cannot estimate a fixed effect conditional probit model (Greene 2002). Hence I test a random effects probit model. An alternative specification is a fixed-effect logit model. Both specifications yield highly consistent results. I further consider a "survival" specification in which I model the number of listings that a borrower needs to post before getting funded for the first time: Reversing survival terminology, a "failure" occurs when the borrower is able to obtain their first loan on Prosper.com. I estimate the time to failure using a Cox proportional hazards model. My main results do not change under this approach either.

Results are also qualitatively consistent when I only analyze the subsample of borrowers who only have one loan. This provides further evidence that the number of friends is not likely a result of previous loan outcomes on Prosper.com, which could have been a concern for endogeneity. Another potentail concern is the performance of past loans on Prosper.com. However, most loans are still being repaid, and those who have defaulted on a Prosper.com loan will not be allowed to borrow again. Adding an additional variable for the performance of past loans does not meaningfully change my results either.

While Prosper.com requires loan requests to be 100% funded to be considered successful (i.e. no partial funding), I nonetheless tested "extent of funding" as an alternative dependent variable to funding probabilities. This is the ratio of total (committed) bid amount received to the requested loan amount. Since this is a proportion censored at 0 and 1, I estimated a Tobit model. Results are also highly consistent with those reported in my paper.

For the interest rate of funded loans, I also tested "interest rate discount" as an alternative dependent variable for open-format auctions. This is measured as the difference between the maximum interest rate that the borrowers specified in the listing, and the actual interest rate that they paid when loans are originated. Results are also highly consistent with those reported previously.

I further consider other social capital variables. The online network formed by borrowers has many disjoint components with little overlap between friendship and group networks. The friendship network is star shaped and exhibits little closure considered important in structural analyses of networks (Coleman 1988; Burt 2005), strengthening the case for looking at the roles and identities to gauge the effects of social capital. I also consider the number of endorsements received by borrowers. This variable is cheap talk. Unsurprisingly, it is insignificant. My previous work examines bids from friends. It is also possible that bids from other members in a borrower's group can help reduce the risk of loans. I do not find evidence for these group member bidding effects. One variable that does matter is the number of friends' defaults in a borrower's neighborhood (ego network). The results indicate that a higher number of defaults in a neighborhood of a borrower is associated with higher risk of the ego's loan (Cohen-Cole and Duygan-Bump, 2008).

### *7.2 Images and Text*

Individuals seeking funding on prosper.com can upload images and add descriptive text to their listings. A priori, it is not clear whether the social network variables I study should be entirely subsumed by image and text data. The social network variables are verified to varying degrees while the image and text data are self-reported fields not authenticated by Prosper.com. On the other hand, borrowers with higher quality friends (e.g. those more familiar with Prosper.com) may leverage friendships to post more persuasive text or images that might do a better job at attracting funding. The actual role played by text or images is an empirical issue that I examine next. To maintain focus, this section only report the coefficients for the social network variables. The coefficients for the text and image variables are available to the reader upon request.

#### 7.2.1 No-Image Sample

Close to half the listings on Prosper.com post no images. I consider the role of the social network variables in the subsample without image data. All key social network variables – friends, friends who are potential lenders, real lenders, or real lenders who bid on listings – remain significant in the no-image subsample. Due to page limitations, I present the results for this subsample in a supplemental appendices (available upon request).

### 7.2.2 Subsample With Images

While the no-image sample results are quite suggestive, it is still perhaps useful to estimate the effects of social networks in samples with images. I experiment with but discard results from automatic image processing software because these are not reliable. I manually code the data. Because of the high costs of manually coding the entire sample, I focus on subsamples. One subsample comprises a random 10% of the funded and unfunded listings. To ensure representativeness, I preserve the proportions of successful listings, credit grades, and the degrees of relations depicted in Figure 1. A second subsample consists of all 16,500 funded listings. In the 10% random sample of 20,513 listings, 15,928 post images, of which 7,986 contain images of adult humans. In the sample of 16,500 funded loans, 10,198 listings have images, of which 8,279 listings contain images of adult humans. I hire assistants to code objective aspects of the data including race, age, and gender. I implement extensive screens to ensure output quality, details of which are available to readers upon request.

To provide context, I discuss univariate statistics and then turn to the regression results. 14.55% of the random 10% subsample of all listings have images of blacks, while the proportion of blacks in the funded loans is only about 8.79%. Thus, blacks are less likely to be funded, as in Pope and Sydnor (2008) and Ravina (2008). The differences for other minority racial groups are less significant. 6.20% of listings are Asian and 4.75%

are Hispanic, while these populations represent 6.91% and 4.21% of loans funded. Females form 30% of the listings but 37% of all funded loans, suggesting that women are more likely to attract funding. Young people below 25 form 23% of all listings but 19.33% of all funded loans. Older people of age 50+ form 6.65% of the listings but only 6.24% of the funded loans. These univariate statistics may reflect unobserved correlations. For instance, younger people may have less credit history or lower credit grades. I consider multivariate specifications to evaluate these issues. More importantly, these models test the effect of images on the coefficients for the social network variables.

I briefly discuss the key results for the image variables first. Listings with images of older people of age 50+ and those with images of black adults are less likely to be funded at the 5% and the 10% levels, respectively. Blacks pay between 40 and 50 basis points more in interest rates, which is slightly lower than the point estimate of 60--80 basis points reported in Pope and Sydnor (2008). My estimate is not significant, a finding similar to that in Ravina (2008), perhaps because there are fewer observations in the sample with race data. As in Pope and Sydnor, I find that blacks are significantly more likely to default with a hazards ratio of 1.20 that is significant at 1%. The more interesting question is whether images subsume the content of social network variables. I find that the standard errors in this set of results exceed the corresponding numbers in Figure 4, reflecting a smaller sample size. Nevertheless, the key coefficients are similar and show similar gradation depending on the verifiability and visibility of the social network variables to outside lenders. Friends with verified roles in Prosper.com, especially verified roles as lenders, matter; among these lender friends, those who have participated in prior loans matter more; and the lender friends who bid on the current listing matter even more.

### 7.2.3 Descriptive Text

Over 99% of the listings in my funded sample of 16,500 listings and in the 10% subsample of 20,513 listings have additional descriptive text. I examine the role played by text, and in particular whether it explains some of the content of the social network variables.

Following Tetlock (2007), I use a disambiguation routine to classify text. I employ the program LIWC (Linguistics Inquiry and Word Count) for this purpose. LIWC classifies words into five broad categories, which are further divided into 80 (overlapping) sub-categories such as basic counts of words, long words, or punctuatation marks as well as more complex psychological, social, and personal categories. The classification is based on an extensively validated and updated dictionary of words and word stems from psychology and linguistics (Slatcher and Pennebaker 2006; Cohn et al 2004; Friedman et al 2004). I experimented with several ways of using the LIWC categories and settled on using a set of 12 LIWC that represented a non-trivial fraction of the word count and that seemed most relevant to lending outcomes.

On average, funded listings are likely to have more words per listing, shorter sentences, more non-dictionary words, use more numerals, more words in the “money” subcategory, positive emotion words, more words of certainty and fewer tentative words. Most variables, however, do not survive in the multivariate specifications. For instance, “money” words are more likely to result in funding and lead to lower interest rates but have an insignificant effect on default. On the flip side, quantifiers such as “few” or "many" lower defaults, and "certainty" words increase defaults but these do not matter in the funding equation. The key result is that text variables show little of the consistency that the social network variables have across the funding probability, interest rate, and default specifications. Thus, it is not surprising that even after controlling for text descriptions, they key social network coefficients display similar gradation across the roles and identities of the members in borrowers' networks.

I do not necessarily view the text results as a comprehensive verdict on the role of linguistic content in determining lending outcomes. Rather, the results suggest that there is a difference in how investors process different types of soft information. Self-reported information in the text descriptions, which is not authenticated or verifiable by Prosper.com, appears to be processed unevenly and less rationally than information in social networks, which is perhaps more credible given the extensive verification process put in place by Prosper.com.

## 8. Conclusions and Implications

Developments in Web 2.0 technologies have significantly altered the way in which individuals interact and connect with each other. Perhaps the most significant outcome of this change is the growth in social networks. Online networks such as facebook.com have become ubiquitous in a very short span of time since their inception. I study one of the first attempts to build businesses based on networks. I study peer-to-peer lending, in which individuals make unsecured loans to other individuals without the intervention of financial intermediaries. I find that social networks, especially their relational aspects, can help mitigate the information asymmetry between borrowers and lenders, and lead to better outcomes in all aspects of the lending process. While borrowers are not required to create and maintain these networks online, my results show that the overall performance of the marketplace will be improved if networks are better integrated into the financing process.

My findings are of interest from a number of viewpoints. One perspective of my study is that it represents data from a credit market in which there is an especially severe problem of adverse selection. An interesting question is what mechanisms individual agents use to adapt, given that they lack the sophisticated risk assessment methodologies, scale economies, or soft information in lending from a broader vector of banking relationships that is available to traditional financial intermediaries such as banks. My study suggests that soft information is sought and used in credit decisions. Social networks act as a new source of “soft” information.

My study also sheds light on the role of soft information in credit markets. An extensive literature in finance argues that credit markets suffer from a problem of adverse selection that can be mitigated by soft information. The literature traditionally views financial intermediaries as the producers and repositories of soft information. As financial markets undergo disintermediation driven by information technology, a natural concern is that the loss of soft information produced by traditional intermediaries could adversely affect credit flows. My results highlight that this concern may at least be partially mitigated. While information technology could supplant some sources of soft information, it could also increase its supply by hardening new sources of soft information and making it available to lenders. The use of social networks may be seen as one manifestation of such an effect. The realization of such benefits, however, depends on the ability to make the new information credible and verifiable by lenders in the credit market.

My study can also be viewed as new evidence on whether social capital facilitates economic exchange, a question of growing interest in information systems, economics, finance and management. The P2P lending marketplace offers micro-level data on this issue with two significant empirical advantages. One, I have relatively well defined measures of social capital, which is identified through social networks. In addition, I also have well defined measures of transactional outcomes, viz., funding, interest rates, and ex-post default. My results are consistent with the view that relational aspects of online social networks can help mitigate information asymmetry, since they are consistently associated with these three outcomes.

My results also point to the avenue by which social capital facilitates economic transactions. As Podolny (2001) writes, an individual's network not only acts as “pipes” or channels through which information and resources flow, but also as “prisms” or informational cues that outsiders use to judge the quality of the individual. In my context, the “pipe” effect of social networks can emerge in two ways: (1) Funds directly flow from friends to the borrower; and (2) Friends help the borrower to better “package” the request for loans, such as using proper descriptions and images in the requests. Neither is a valid alternative explanation for my findings: Firstly, for 95% of the loans associated with friends or groups in my sample, funds provided by friends or peer group members accounted for less than 4.4% of the total loan amount. In other words, most of the funds that borrowers obtain are from lenders outside their social networks. Secondly, my results hold even after controlling for textual contents and image characteristics. Hence, networks matter not because friends provide the funds or information directly; but rather, the roles, identities and actions of friends serve as an important signal for other lenders –

strangers – to determine the creditworthiness of the borrowers. Borrowers are judged by the nature of the company they keep.

Finally, my results have implications for the design of businesses based on online networks. One implication is that such businesses could incorporate and facilitate ease of using multiple social network metrics for end-users. In particular, the number of connections and their structure may not be sufficient, but the nature of the relations also matters. In fact, one could make a reasonable case that P2P markets should incorporate functionalities that not only promote interactions among members, but also enable borrowers to credibly signal their embeddedness in their social networks to lenders. Indeed, Prosper.com has taken steps in this direction in more recent listings where social network information is given greater prominence in listings seeking funding. In a similar vein, my results show that in addition to friendship networks, groups can also play a valuable role in reducing information asymmetries. Thus, increasing the interdependence among group members and making these ties verifiable can facilitate capital flows, thus enlarging the scope and applicability of microfinance-style mechanisms. ■

# ESSAY 2: GEOGRAPHY AND ONLINE INVESTMENT DECISIONS: EVIDENCE OF "HOME BIAS" FROM AN ONLINE NATURAL EXPERIMENT


## Abstract

The effect of geography on individual and organizational behaviors has long fascinated researchers in many disciplines. One consistent pattern is "home bias" – interactions are more likely to occur between those who are from the same geographical area. Empirical evidence of this phenomenon, however, has been mostly cross-sectional. Studies on dyadic level interactions are further plagued by the sparseness of transaction networks, and many papers resort to specialized statistical procedures or data reduction techniques. Endogeneity in the geography variable is also a potential concern. In this paper, I exploit a unique natural experiment in an online peer-to-peer lending market, where one side of the market (lenders) was artificially constrained to one geographic area, while the other side (borrowers) was open to almost all US states. This provides a unique context to study how geography affects investor decisions on the dyadic level. In particular, I study whether lenders tend to lend to borrowers from their home state. I find that while on average, having such geographical ties with lenders increases the chances of receiving a bid, such benefits only accrue to high-quality borrowers. I attribute this "bias" to the emotional side of investor decisions. Nonetheless, such bias is also intricately related to rational lending criteria. I further show that the effect of economic distance dominates that of spatial distance. The virtual nature of electronic markets does not completely eliminate the role of geography.

## 1. Introduction

Are economic transactions *more likely* to occur between entities within a geographical area? This question has been widely studied in economics, finance and management. Many empirical studies have documented that such “home bias” exists not only on the national level – as documented in the international economics literature – but also within national boundaries such as trading relationship across states. In many cases, home bias is considered an example of market inefficiency or individual irrationality. The presence of home bias in investors’ equity holdings, for example, suggests that investors do not follow rationality assumptions of classical economics: If they decrease their holdings of domestic equity, the return to the overall portfolio could be higher. A better identification and understanding of home bias, therefore, has significant implications for market efficiency, competition, and individual behaviors.

However, most of these studies are conducted on the macro-level, considering only the aggregated volume of trade or portfolio holdings. Very little attention has been given to the “alternatives” that exist before these firms and individuals made their selections. The micro-level decision processes are usually subsumed by the aggregated data. Recent studies (Sorenson & Stuart, 2001) delve into organizational dyad level and are able to show a home bias in firms’ funding decisions. Meanwhile, a typical challenge in these dyad-level studies is the sparseness of the transaction network compared to all potential dyads. For instance, if there are 100 buyers and 100 sellers in the market, there are over 10,000 possible transaction dyads. Yet the number of actual transactions is usually very small, resulting in a lack of power in statistical modeling. Using the terminology of network analysis, the “density” of the transaction network is very low.

Scholars offer two strategies to deal with this issue. The first is to reduce the number of potential transaction dyads (Sorenson & Stuart, 2001). However, such data reduction needs to be carefully justified – if a venture capitalist is faced with requests from 100 possible entrepreneurs, and the investment is made only to 2, why should I only consider several out of the 98 who do not get funded? An alternative is to resort to statistical procedures such as rare-event logit (Kirsch, Goldfarb, & Gera, 2009; Stuart & Sorenson, 2003). Nevertheless, rare-event logit requires knowledge of census-level proportions to correct for sampling biases (King & Zeng, 2001; Tomz, King, & Zeng, 2003). Relying on stringent statistical assumptions and data reduction techniques can unfortunately cast doubt on the robustness of home bias in dyad-level transactions.

To further investigate whether "home bias" is a robust phenomenon, I exploit a unique natural experiment in the online peer-to-peer lending marketplace, Prosper.com. This is the same context as the first essay of the dissertation. Due to government regulations, Prosper.com was temporarily closed to borrowers and lenders in May 2008 as they sought regulatory approval from SEC (Securities and Exchanges Commission). In April 2009, Prosper.com obtained SEC's approval; however it was also required to obtain further approval from each state before they could service borrowers or lenders from that state. California was the first to grant them approval. On April 28th, 2009, Prosper.com exited this "quiet period"[4], but only to a limited scale: While borrowers from virtually all over the United States could participate and request loans, only lenders located in California were allowed to bid on loan requests. This situation lasted for about 10 days, however. On May 8th, 2009, Prosper.com suddenly announced that it would halt all

[4] During the "quiet period", Prosper.com does not accept borrower or lender registrations; existing borrowers cannot request loans, and lenders cannot bid on loans. All operations are put on hold. Unfinished auctions are cancelled and bids returned to lenders. Prosper.com will also not respond to media requests.

operations, enter the “quiet period” again, and request approval from local regulatory bodies across the US. Prosper.com users in many third-party discussion boards referred to this short period as “mini-Prosper” because of its limited scale.

This essay of my dissertation exploits this period of time on Prosper.com as a natural experiment to test the hypothesis of “home bias” in investors' decision making. More specifically, I am interested in whether individual investors in California are more likely to lend to borrowers from their home state. If there is a “home bias”, I should expect to find that lenders (who are all from California) are more likely to place bids on requests from California borrowers.

This event is an ideal context to study my research question for a number of reasons. First, both the opening and closing of this 10-day window is rather unexpected. Therefore, the geography (location) of the participating borrowers and lenders is exogenously constrained. Unlike other studies, I need not be concerned about the movement of individuals in response to regulations because this was a very short period of time.

Second, unlike lab experiments, borrowers and lenders in this market participate in auctions with their real credit needs and their real money (where all bids are committed bids and cannot be withdrawn), which helps reveal the underlying drivers of their behaviors. These borrowers and lenders have also been verified as “real persons” with a valid Social Security Number and other identifications.

Third, as discussed earlier, a common challenge for empirical studies is the sparseness of transaction networks, especially when compared to all potential connections. In my context, this artificially constrained 10-day marketplace creates a small cohort of borrowers and lenders who interacted and transacted in the online market. I observe all incidences of bids between all potential borrower-lender pairs – including borrowers who never received any bids. This allows me to very conservatively investigate the robustness of home bias on the dyadic level, without resorting to data reductions or specialized statistical procedures.

Last but not least, unlike market access studies (Redding & Sturm, 2008), this natural experiment creates an artificial separation of the borrower-side and lender-side of the marketplace to study the effect of geography on individual behavior. Geography is only constrained on one side of the market, allowing me to attribute the difference to the choice of lenders instead of the borrowers.

For the “home bias” hypothesis, I only focus on the dichotomous measure of whether the borrower and lender are from the same state. Continuous measures of spatial distance sometimes also yield interesting empirical patterns (e.g. Figure 1 of Sorenson et al 2001). Consistent with this literature, I further investigate whether lenders’ likelihood of placing a bid decreases as the borrower's distance from California increases. I measure borrowers' distance from investors in two ways. The first one is the intuitive geographical distance, measured as the driving distance among state capitals. The second approach, inspired by Tsang and Yip (2007), is to measure the *economic distance* between the borrower's state and California. This is the difference in per capita GDP. Details of these metrics follow later in the paper.

My results show that “home bias”, a motivation better classified as an “emotional” driver for investors, does seem to be robust in my analysis. This result is very conservative both because of the “virtual” nature of this online market, as well as the fact that my analysis does not resort to data deductions or sophisticated statistical models with stringent assumptions. More interestingly, I find that "home bias" actually exhibits an interaction effect with rational aspects of the decision process. Being from the same state as the lenders increases the chances of receiving bids for credit-worthy borrowers, but surprisingly reduces the chances of receiving bids for low credit-grade borrowers. The interplay between emotional and rational motives of investment is a pattern that has not been previously identified in the "home bias" literature.

The rest of the paper is organized as follows. I briefly review related literature in the next section, followed by a description of the context and data gathered from the "mini-Prosper" natural experiment. I then describe the empirical models used in my analysis. I also report a number of robustness tests, followed by a discussion of some limitations of the study. I conclude with a discussion of the implications of the findings and some directions for future research.

## 2. Related Literature

### *2.1. Distance and home bias*

I draw from two related but largely distinct streams of literature: studies in psychology and management where geographic propinquity has been mostly considered an emotional driver of decision making; and economic studies of "home bias" in trade that attempt to rationalize this phenomenon with transaction costs or other theories. Despite different disciplinary jargons, the pattern that interactions (trade, social interactions, even referee behavior) are indeed more likely to occur among individuals or organizations that are situated within the same boundary.

The first stream of research relates to the concept of “homophily” in psychology and sociology. Homophily is typically referred to the phenomenon that that individual tends to associate, favor, or trust those who are similar to themselves. Its empirical evidence has been found on many dimensions of similarity. McPherson, Smith-Lovin and Cook (2001) provide an extensive review of studies on “homophily”. Geographic propinquity is one of the contexts through which homophily can represent itself: “*Perhaps the most basic source of homophily is space: We are more likely to have contact with those who are closer to us in geographic location than those who are distant*.” (page 429) Some studies even show that referees for sporting events favor their home teams (Pettersson-Lidbom & Priks, 2010; Sutter & Kocher, 2004). Geographical proximity has also been shown to affect organizational innovation (Whittington, Owen-Smith, & Powell, 2009). It should be noted, however, that much empirical evidence on the effect of propinquity has been static and correlational. An important contribution of my study is to present further evidence that propinquity, or geographic proximity, indeed has a robust effect on outcomes of interest.

The second stream of literature that informs my empirical analysis is the concept of “home bias” in economic geography, as well as related work in finance and international economics. Economists have long documented a pattern of transactions: Trade, investment, venture capital funding and so on, all tend to occur more frequently among those who are within the same border, a phenomenon broadly termed “home bias”. More specifically, the role of geography has been studied in a wide range of areas: international trade (Boulhol & De Serres, 2010; Disdier & Head, 2008; Overman, Redding, & Venables, 2003); intranational trade, or trade within a country’s border

(Hillberry & Hummels, 2003; Wolf, 2000); equity investment decisions of individual investors(Graham, Harvey, & Huang, 2009; Karlsson & Nordén, 2007), fund managers (Cooper & Kaplanis, 1994), venture capitalists (Sorenson & Stuart, 2001), and even governments (Ahearne, Griever, & Warnock, 2004). More recently, some researchers have started looking beyond the “incidence” of transactions and examine the *price effect* of home bias (Carey & Nini, 2007). As I will describe later, even though I do not have complete information about the actual price (interest rate of bids) because Prosper.com does not reveal that information unless a bid has been outbid, I am able to examine several interesting second-stage outcomes in the bidding process. These include the amount of bid that the lender placed on the loan and the timing of the bid (early vs. late).

Most of these studies, however, use data from offline contexts. Will I observe the same “home bias” or effect of distance in the context of electronic commerce, where no face-to-face interactions exist? Some studies on the effect of geography touches upon the subject by studying how changes in technology may make geography less relevant in transaction relationships, such as banks’ lending decisions (Petersen & Rajan, 2002). Yet those relations are usually not entirely offline. Another study looks at the effect of geography in the online auctions(Hortacsu, Martinez-Jerez, & Douglas, 2009) and found that, even though online markets such as eBay help mitigates the deterrence of spatial distance, “home bias” persists – even though to a lesser degree compared to prior studies of intra-national trade bias (Wolf, 2000).

*2.2. "Gaps" in the empirics*

I now turn to the technical details about the empirical modeling strategies of the above two streams of literature, and identify possible gaps that I seek to fill with the current study. These "gaps" include (1) using data only on consummated transactions, thereby losing information about alternatives of matching trading partners; (2) aggregated transaction volume instead of individual choices; (3) reliance on data reduction techniques; (4) reliance on rare-event logit model; and last but not least, (5) potential endogeneity in the geography variable.

The economics geography literature and studies built upon it, such as Hortacsu et al (2009), typically draws from the gravity equation (Bergstrand, 1985) in international trade. While there are some variations, a typical gravity equation takes the following form (Bergstrand, 1985):

$$PX_{ij} = \beta_0 Y_i^{\beta_1} Y_j^{\beta_2} D_{ij}^{\beta_3} A_{ij}^{\beta_4} u_{ij}$$

where the dependent variable is the aggregated volume of trade from region *i* to region *j*. $Y_i$ and $Y_j$ are the economy volume of two entities (e.g. two countries), respectively. $D_{ij}$ is the spatial distance between these two entities, and $A_{ij}$ refers to other factors that facilitate or deter trade. $U_{ij}$ is the error term. The estimation is typically done by taking a logarithm of the equation.

The "equity home bias" literature in finance, as well as the related "consumption home bias" literature, takes a slightly different approach. A good review of this macro-economic view of home bias is Lewis (1999). Research in this literature typically derives testable models from the Capital Asset Pricing Model (CAPM); specifically, "home bias" is said to exist when "*the proportion of foreign assets held by domestic investors is too* *small relative to the predictions of standard portfolio theory*" (Levy & Sarnat, 1970; Lewis, 1999). Researchers have also claimed home bias in a descriptive manner, for example, when there are only a small proportion of investors that own foreign assets. Graham et al (2009) is one such example.

While both the gravity equation and the CAPM are well-accepted foundations to study home bias or the effect of distance, these models lack the micro foundations of economic transactions. More specifically, both the aggregated trade volume and the holding of foreign equity are based on observed transactions that are actually completed. They fail to consider the choice process of the decision makers, especially what alternatives that the buyer/investor had. These alternatives are typically unobservable by econometricians. My data, on the other hand, consist of all potential borrowers that the investors can lend to during the time window, and what their actual choices are. This allows me to conduct a finer level of analysis. Additionally, aggregated data do not allow me to distinguish whether the choice is made by buyers or sellers. The unique context of my study allows me to focus on just the choice from one side (investors, or buyers of a claim) because they are artificially constrained to one geographic area only.

On the other hand, the management and strategy literature has generally taken a different approach in the empirical studies of home bias and the effect of geography. As discussed earlier, two approaches are often used by empirical researchers who are interested in home bias on dyad-level transactions. The first one is data reduction (Sorenson & Stuart, 2001), whereby the researchers strive to filter out at least some possible combinations of transaction ties. To illustrate the empirical method used in Sorenson (2001), let's suppose there are M investors ($A_1$, $A_2$, … $A_M$) and N entrepreneurs seeking funding ($B_1$, $B_2$, … $B_N$). Suppose $A_1$ invests in $B_1$ and $A_2$ invests in $B_2$, and these are the only actual ties that occurred in the data. Conceptually, ($B_1$, $B_2$, … $B_N$) are all possible candidates for $A_1$ and $A_2$, so there are a total of M*N possible dyads. To reduce this number, the authors keep only $A_1$-$B_1$, $A_1$-$B_2$, $A_2$-$B_1$, and $A_2$-$B_2$: only the start-ups that received funding from another investor are considered as alternatives. This allows the authors to reduce the total number of possible dyads from M*N to 4. While this dramatically increases the density of transaction network (from 2/M*N to 2/4), a great deal of information was lost, and the effects of geography variables could have been artificially inflated in this process.

A second approach often used in the strategic management literature is rare-event logit model (Kirsch et al., 2009; Sorenson & Stuart, 2001). Rare-event logit was originally developed to facilitate the sampling of rare subjects in the population (King & Zeng, 1999). For instance, suppose from the census data, I know a priori that a certain ethnic group has about 0.005% of the population. So when I conduct surveys about them, it will be a lot more economical to over-sample the minority, and then adjust for small sample and rare events (e.g. activities within that ethnic group) using this procedure. The weighing (parameters in this procedure) should adjust the 0s and 1s to the proportion in the population. Unfortunately, many studies that use rare-event logit do not specify the weights used. By contrast, "mini Prosper" allows me to use parsimonious logistic models to study the incidence of transactions, eliminating the need to resort to rare-event logit since I will be using the "population", the entire set of possible connections.

Endogeneity, on the other hand, can be common to both the management and the economics literature, but has received very little attention in those empirical studies. One possible reason is that geography is typically considered as exogenously given. This assumption is questionable, however. Specifically, individuals and organizations can strategically choose where to locate their business activities. A recent empirical study that highlights this possibility is Parwada (2008). Parwada (2008) examines the determinants of fund managers' location choices – that is, where they choose to locate their firms. Hence, geographic information is not entirely exogenous, especially when I study funds or venture capitalists. More generally, in the long run, economic production factors (labor, capital and so on) tend to gravitate to a location where the marginal productivity is highest. This point is echoed in Redding et al (2008) as well.  By contrast, my study exploits a natural experiment in which the decision makers are individuals (who are unlikely to relocate just to be close to certain borrowers) faced with a very small time window (they cannot move even if they wanted to). Concerns for endogeneity can be significantly mitigated, if not eliminated.

### *2.3. Non-dichotomous measures of distance*

The discussions above largely focus on a dichotomous measure – that is, whether or not the two parties to the transaction are from the same geographical area. Many of these studies of geography also consider the actual distance, a continuous variable. For instance, the gravity equation in international trade argues that the volume of trade is inversely proportional to the physical distance between the nations (Bergstrand, 1985), which is a continuous measure. Most strikingly, Sorenson and colleagues (2001) uses nonparametric methods to show that (Figure 1 of their paper) the probability of investment decreases as the spatial distance between the investor and the start-up increases, albeit at a decreasing rate. These studies motivate me to also examine the role of physical distance in the online P2P lending market.

In addition to the spatial distance, another measure of distance that can play a role in determining transaction incidence is *Economic Distance* (Tsang & Yip, 2007). Economic distance is concerned with the difference in the level of economic development between two entities. In their paper, Tsang and Yip found evidence that the greater the economic distance between two countries, the less likely for foreign investments to succeed. It is also plausible that economic distance between states can affect investor decisions in the P2P context. For instance, while California is thousands of miles away from east coast states, they are still likely to share many similarities such as population compositions, political views, and the acceptance toward internet technologies. Hence, solely relying on geographical distance may not be sufficient. I therefore also test the hypothesis that shorter economic distance between investors and borrowers can increase the chances of transaction incidence. In addition, since this is an online context, the effects of economic distance can potentially outweigh those of the geographical distance.

### *2.4. Explanations for home bias*

Many studies also attempt to explain the origin of home bias in the empirical studies. There are largely two schools of thought: a rational explanation, and an emotional/behavioral explanation. Many economists attempt to explain home bias using rational criteria, such as transaction costs that include shipping costs and cultural differences, cost of information acquisition for international equity investments, and even new conceptual framework of the production process that responds to trade costs (Yi, 2010). Some economists also started to investigate home bias using behavioral approaches, and found over-optimism toward home equity markets in survey data (Lai & Teo, 2008; Strong & Xu, 2003). On the other hand, many studies in sociology and management often attribute home bias to factors such as homophily (McPherson et al., 2001).

These studies in different disciplines suggest that rational and emotional forces can both contribute to home bias, yet there has been little research on how these two dimensions interact with each other. It is possible that in most studies it is highly unlikely to differentiate these two dimensions of factors. My research context presents some unique advantages that allow me to study them at the same time. One such advantage is that the proportion of investors who have private information about a potential borrower is very small – and it is also captured in my data through borrowers' social networks. For lenders who are strangers to the borrowers, the dataset captures all information that was presented to them at the time of the loan request. Rational information such as credit grade and so on can be easily quantified, which in turn makes it plausible to attribute the remaining, unexplained variance to emotional factors. I will then be able to study how these factors interact and affect investor decisions.

## 3. Context and the "Natural Experiment"

### *3.1. An overview of Prosper.com*

The natural experiment that I exploit in this paper occurred in the same context as the first essay of my dissertation, Prosper.com. The first essay of the dissertation has provided a detailed description of how Prosper.com works. It should be noted that geographic information about the borrower is prominently displayed on the web page requesting bids. Each lender decides whether or not to contribute funds to a request. All bids are committed bids: lenders cannot withdraw from the auction process. In other words, this is a two-sided market with significant levels of information symmetry.

The asymmetric information issue inherent in financial lending is especially severe in this marketplace because individual lenders do not possess the type of sophisticated evaluation techniques that traditional intermediaries have. Hard credit information such as credit grades (letter grades indicating the range of FICO scores of the borrower at the time of the loan request) is certainly an important criterion that lenders will use to screen borrowers. But other information, such as geographic location of borrowers, can also play a role. Consistent with the literature surveyed above, I hypothesize and test that geographical information also has an impact on the investor's decision making, and it can potentially interact with rational factors. In the following, I will describe the details of the natural experiment where I collect the data to test the propositions above.

### *3.2. A Natural Experiment for Geographic Information*

Prosper.com started lending in 2006 in the United States. Since then, this novel business model has increasingly attracted significant media attention and they rapidly increase in their size. From the very beginning, the management team at Prosper.com has maintained that this website is not a bank or a securities company, but just to provide a platform for investors and borrowers to engage in transactions (much like eBay); therefore, it does not need to be regulated by the Securities and Exchange Commission (SEC). However SEC disagreed. It ordered Prosper.com to shut down in October 2008, and fined the website for about $1 Million. Prosper.com complied and entered a “quiet period” to register with SEC. All operations (borrowing and lending) were suspended on

October 15th, 2008 and the company declined any media contacts, including requests from technology bloggers.

In late-April 2009, Prosper.com obtained the permission from SEC, but was also required to further obtain approvals from each and every state in the United States for both borrowing and lending operations in its marketplace. On April 28th, 2009, without prior notice, Prosper.com re-opened its doors to borrowers from most states in the United States. On the lending side however, only lenders from California were allowed to participate as California was the only state to grant their request at that time. Loan requests started to appear from all over the U.S., and Californian lenders were allowed to bid on the auctions (loan requests). This “Mini-Prosper” continued to operate for about 10 days when Prosper.com abruptly decided to re-enter the “Quiet Period” on May 8th, 2009 to obtain further approvals from other states.

### *3.3. Evidence of Home Bias across Prosper.com*

Empirical motivation for my study comes from some high-level statistics on Prosper.com, not just activities during the “mini Prosper” period. Between the beginning of Prosper.com and the end of March 2009 (before the start of “Mini Prosper”), there were 3.7 million bids that became part of actual loans to borrowers. Among these bids, about 7% occur between borrowers and lenders of the same state[5]. Considering there are 50 states, the “naïve” average likelihood that the borrower and lender are from the same state is 2% (1/50). Hence, the actual occurrence of same-state lending relationship is

---

[5] There are 53 state designations on Prosper.com. In addition to the 50 states, AA, AE and AP refer to Armed Forces stationed outside of the US. No lenders come from these “states”, and borrowers are very few. Bids on these loan requests account for only 0.14% of all bids. I hence remove them from the analyses.

about 3.5 times the likelihood in a random network. There does appear to be significant home-bias in lending relationships.

The amount of bid also shows a similar pattern. I first categorize all bids into same-state bids versus different-state bids. A t-test with unequal variance across these two groups shows that on average, same-state lenders bid $18 more (where the average bid amount in the whole dataset is $88), and the difference is also statistically significant.

This pattern persists even when I consider the fact that each state has a different number of lenders, and the total investments from lenders of each state are also different. Virtually in all states (except SD where there is no borrower recorded in the data), the share of home-state lender in the amount received by the state's borrowers exceeds these lenders' investment share in the entire marketplace (Table 2.1). For instance, up until April 2009, Texas lenders contributed 8.24% of all loans made on Prosper.com. They, however, account for 10.82% of all loans made to Texas borrowers. The same pattern exists when I calculate the share of bids (count) placed by home-state lenders, and compare it to the ratio in the overall market (Table 2.2).

These macro-level statistics, however, are not definitive evidence of home bias – they do not consider the choice faced by investors at the time of the loan (other borrowers that they choose not to lend to), and there could be potential herding among lenders across different states. In addition, studying dyad level interactions on the entire market will be virtually impossible. With over 200,000 borrowers and a similar number of lenders, the total number of potential combinations will be astronomical. The "natural experiment" of mini-Prosper, however, provides an ideal context to study home bias on a finer level.

## 4. Data and Empirical Models

During this 10-day window, there were 547 borrowers seeking loans in 701 listings, and 656 lenders (all from California) placing bids. Out of the total 358,832 possible bid-dyads (547 * 656), 3540 bids actually occur. When a borrower posts multiple loan requests during this window, I combine information from these listings for statistical analyses.

I first present some “macro”-level evidence of “home-bias” in investor decisions. Out of the 701 listings, 94 listings or 13.4% were from California borrowers. By comparison, during this window, 29 loans reached 100% funding. Five of these (17.24%) were for California borrowers. If I look at the total amount funded in these loan requests, the total amount funded is $84236, out of which $19636 (23.3%) was for California borrowers. Furthermore, in terms of interest rate, CA borrowers' average interest rate on these loans is 11.58%, while borrowers from other states pay an average of 15.21% in interest rate.

These macro-level summary statistics are similar to the “trade volume” descriptions in international trade literature: It appears that California borrowers benefitted disproportionately due to the market constraint. This however, only gives me a high-level overview without consideration of the characteristics of individual borrowers and lenders. I now delve into this issue, and conduct analysis at the level of borrower-lender dyads instead of aggregated bidding activities described above.

The main outcome variable that I study in this paper is whether or not a lender places a bid on a borrower's loan request. As mentioned earlier, all bids placed on Prosper.com are committed bids – they cannot be withdrawn by lenders. Hence, even though some bids were refunded to lenders when “mini Prosper” was terminated, the action of the lenders to place bids is still of significant interest. My goal is to look at each possible borrower-lender dyad, and test whether geography (whether the borrower is from the same state as the lender – California) has an impact on the probability of a bid being placed. The focus on this “Same-State” dichotomous variable is highly consistent with the literature (Graham et al., 2009; Hortacsu et al., 2009).

More specifically, my level of analysis is a borrower-lender dyad, and the main outcome of interest is the probability that a transaction occurs between that dyad (a bid being placed):

$$Prob\left(Lender_i \; bids \; on \; Borrower_j\right) = f\left(BorrowerInfo_i, LenderInfo_j, AuctionInfo\right) + \epsilon_{ij}$$

If a lender places a bid in the borrower's auction (loan request), then the outcome variable takes on the value of 1. Otherwise it takes on 0. I gather all available information about the borrowers and lenders, and use them as explanatory variables. The key variable is whether the borrower is a resident of California. If this variable is equal to 1, then the borrower is from the same state as the lender.

The main explanatory variable of interest is whether the borrower and lender are from the same state. In the main empirical model, I include this as a dummy variable that takes on a value of 1when borrowers and lenders are from the same state. This is one of the variables in the vector of “BorrowerInfo” in the equation above. Other explanatory variables associated with borrower characteristics and auction characteristics are similarly defined in Lin et al (2009). I also include variables of whether the borrower has friends, whether the lender has friends, and whether or not they are affiliated with groups. I pursue these further in the robustness tests.

I further derive two continuous measures for distance. The first one is the *spatial distance* between California and the other states, defined as the driving distance between capitals of each state. The driving distances are calculated based on Google Map data, where I use an automated agent to submit the name of the cities to Google Maps, and then parse the resulting webpage. This should be a more relevant measure than the “great circle distance” because I am studying individual behavior within a national boundary. Exceptions are Alaska and Hawaii, for which I use great circle distance, because no driving routes are available. I take logarithm on this variable before using it in the regression models.

The second measure of distance is *Economic Distance* (Tsang & Yip, 2007) between borrowers and lenders. The economic distance between the borrower's and the lenders' states were measured as the difference in the logarithm of the GDP per capita of their states in 2009:

$$EconDist_{ij} \equiv \log\left(GDP_{California}\right) - \log(GDP_{BorrowerState})$$

where GDP information about each state was obtained from the US Bureau of Economic Analysis. This value is equal to 0 for California borrowers, less than 0 for borrowers from states that are economically more advanced than California, and greater than 0 for borrowers from “richer” states.

## 5. Findings

### *5.1. Main Findings*

I report my main results from the analysis in Table 2.3. The first column shows the model estimated on the overall sample. There is a statistically significant “home bias” in the overall sample – an exhaustive list of potential dyads – using only a parsimonious logistic model. Standard errors are also conservatively estimated using clustered sandwich estimators to allow for intra-state correlation (since there are multiple borrowers from the same state). I do not have to resort to special statistical assumptions such as those in Rare-event logit, or special data reduction techniques. In terms of economic significance, being a Californian borrower increases the odds ratio (probability of being funded vs. not funded) by 13%. Unfortunately I am unable to find an effect size in the literature to directly compare with, as most studies either use aggregate trading volume (Hortacsu et al., 2009) or using continuous measures of distance (Sorenson & Stuart, 2001).

Many studies have shown that there are both rational and emotional components to home bias; hence, I next investigate how these two dimensions interact with each other. Since it is not possible to measure the "emotional" factors on each potential dyad, I conduct a stratified analysis instead.

Out of the many factors that capture the "rational" part of decision making, the credit grade of the borrower (letter grades reflecting the range of borrower's FICO scores) is by far the most important one. Hence I stratify borrowers by their credit grade. More specifically, I split the borrowers into two groups according to their credit grades: those with credit grades of AA or A (FICO scores of 720 and up), and those with B, C, or D (600-719)[6]. I then estimate the same model on these subsamples (removing dummy variables for credit grades), and report the results in the second and third columns of Table 2.3. Results show that good borrowers from the same state as the lenders are more likely to receive bids than non-residents. Quite surprisingly, less creditworthy borrowers

[6] In a robustness test, I remove borrowers of credit grade B, and compare AA/A versus C/D and obtained very consistent results. These stratification analyses are consistent with the empirical literature when there are multiple categories. An example is Hsu (2004).

are actually penalized. There seems to be an interesting interaction effect between “reason” and emotions of investor decisions. In addition, these effects are both significant on the 1% level. In terms of economic significance, results suggest that, if the borrower is a good candidate in terms of hard credit information, the chances of them obtaining a bid is almost 70% more than not receiving a bid. On the other hand, if the borrower is not as creditworthy, the probability of receiving a bid is less than half of the probability that they will not receive a bid.

The above results from the stratified analysis suggest that not only is there a home bias in the online investors’ decisions, such a bias is especially prominent when a decision to lend can be supported by rational criteria. Otherwise, when there is lack of rational appeal, home bias in fact works against a potential trading partner who is from the same geographical area: there may be a bias, but not in the sense of a preferential treatment.

### *5.2. Geographic Distance vs. Economic Distance: Dyad-Level Analysis*

A natural extension from the analyses above is to look at borrowers' distance from California lenders, and how that may affect investors' decisions. The literature proposes two ways of measuring the distance: Geographical, or Economic. As mentioned previously, I measure the geographic distance between the borrower and California lenders as the driving distance between California state capital (Sacremento) to the capital of the borrowers' state of residence (in miles). I test the effect of this variable in a number of ways: (1) the actual distance; (2) logarithm of this distance; (3) the inverse of the square of the distance (consistent with the gravity equation in astronomy). Results reported in the Tables use the logarithm of distance, but are qualitatively similar with other measurements. I then measure the *economic distance* between the borrower and the lender using the difference in the logarithm of their respective state's per capita GDP in 2009. A greater economic distance reflects larger differences of economic development between borrower state and lender state.

These results are reported in the last column of Table 2.3. When I incorporate both the economic and spatial distance metrics in the model of bidding probabilities, the impact of economic distance unequivocally dominates that of geographical distance: While an increase in distance from CA always decreases the probability of receiving a bid, the effect of economic distance is much more salient than that of the geographic distance. This striking difference can be seen both in terms of the level of statistical significance and economic difference – an increase in the economic distance *reduces* the chance of receiving a bid much faster than an increase in spatial distance.

### *5.3. Geographic Distance vs. Economic Distance: State-Level Analysis*

The effect of the economic distance is also obvious on a state-level analysis, where I aggregate all loan requests from each state. Figure 1 shows an interesting pattern: as the economic distance increases in either direction (away from zero), the volume of transactions decreases. On the other hand, when I measure by geographical driving distance, the difference is mostly dichotomous: all non-California states seem to cluster together, instead of displaying any gradual spatial patterns (see Figure 2). These figures not only show a macro-level concentration of lending activities toward California (home-state) borrowers, but also lend support to the notion that economic distance can provide more interesting insights than geographic distance alone. In particular, the result on the spatial distance differs strikingly from the smoothly-decreasing pattern depicted in Figure 1 of Sorenson and Stuart (2001).

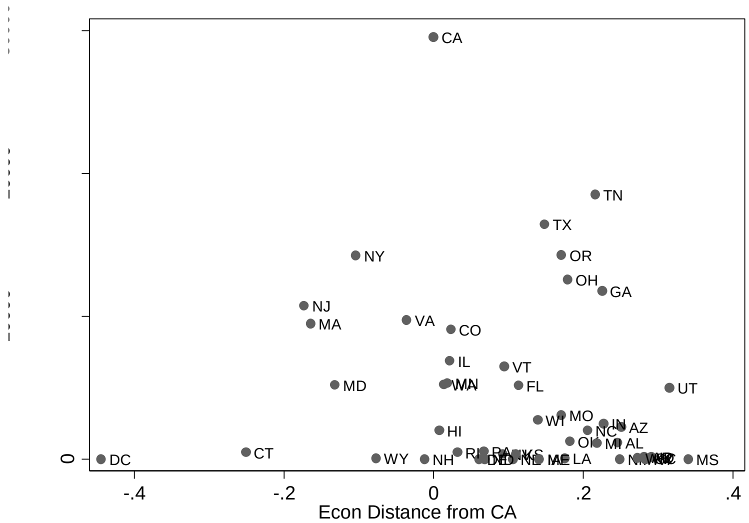



**Figure 1: Investment Amount vs. Econ distance from CA**

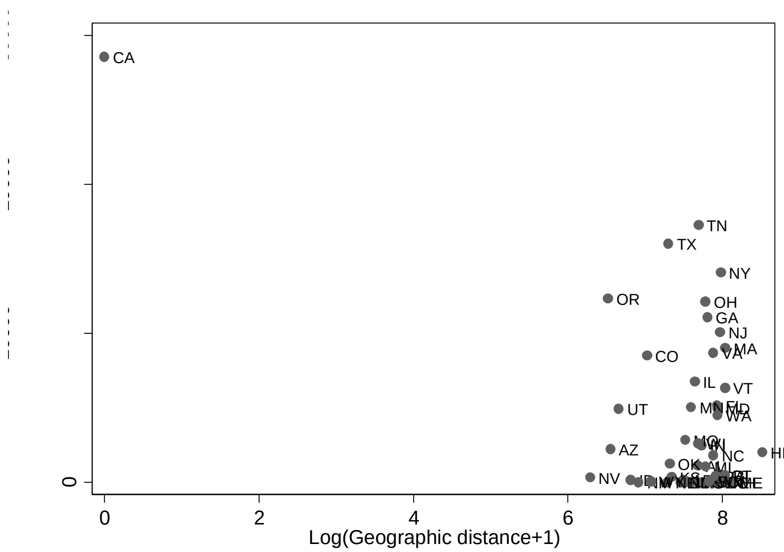



**Figure 2: Investment Amount vs. Geographic Distance from CA**

What are the implications of the comparison between geographical distance and economic distance? The relatively low effect of geographic distance suggests that other than home bias, the effect of spatial distance is in fact quite small. This is possible due to the nature of the marketplace. Prosper.com is an online marketplace without face-to-face interactions between borrowers and lenders, and lenders cannot monitor these loans after they are originated, and therefore, for California lenders, whether a borrower is in Seattle or New York does not make a difference in terms of how long it takes to get there.

On the other hand, the results on the economic distance tell a different story. It is highly likely that, similar to home bias, economic distance points to the emotional side of investor decisions. It is not because borrowers from less developed states are more likely to default: The above results are robust even when I control for state-level default information of Prosper.com loans in the past (see the robustness tests below). It is more likely that economic distance reflects prejudices or stereotypes against certain states, not financial creditworthiness.

## 6. Robustness and Additional Tests

I now turn to a number of robustness tests and address some possible alternative explanations.

### *6.1. Friends, Friend-of-a-Friend, Groups, and other Social Ties: Is it "Home Bias", or is it Private Information?*

Is it possible that the "home bias" I observe is in fact only due to friendship ties? If all lenders are merely investing in their friends, and all friends are in the same state, then I will still have the results shown in Table 2.3. In that case, it is really not about "home bias", but rather the fact that investors simply have more information about their friends and act (bid) accordingly. Bids could be simply a reflection of such private information. Indeed, the overall Prosper.com social networks data shows that friends tend to come from the same geographic region.

To test the validity of this alternative explanation, I use the friendship network on Prosper.com at the start and the end of the California window, and then submit all potential dyads to these graphs to identify possible connections. Out of over 300,000 possible transaction dyads in this time period, only 3 borrower-lender pairs occur. In other words, only on three 3 occasions during the “mini Prosper” do I see friends bid on a borrower. While friendship may have played a role for other borrowers (cf. Lin et al 2009), those borrowers may not be active in this time window.

There were also no “friend-of-a-friend” (FOAF) relationships between any potential dyads in this time period. None of the potential borrower-lender dyads in this time period shares a common friend. Moreover, I do not observe friends on the third (friend of a friend of a friend), fourth, up to the 7$^{th}$ degree (i.e. separated by 6 degrees) in the data. Therefore, potential transitivity of network ties does not explain the home bias in my sample either.

Another social relationship that could convey private information is group membership. Members can join groups and could have some private interactions. However, most groups on Prosper.com do not have much private communications; in many cases the communication is from the group leader to the members; much less occurs among members[7]. Furthermore, if an investor and a borrower both support certain groups[8], that may also have an effect on the investor as that shows a common ideology. However, much like the friendship ties, shared memberships of groups (being a member

---

7 Many groups are inactive after Prosper.com discontinues the incentives for the leaders. A few groups are indeed active and members have a lot of interactions outside Prosper.com, such as on third-party discussion forums. However, most of these groups are lender groups, and may not be California lenders. Cases where lender belong to a group and lend to a borrower in CA are virtually nonexistent.

8 A member can join one group at a time. They can, however, “support” multiple groups. This is a very cheap signal, however, as there are no opportunity costs for supporting any group.

or simply showing support for a group) are negligible between the potential borrowers and lenders in my small time window. Those with strong ties usually have created loans before Prosper.com was shut down; it is highly unlikely for them to predict the "Mini Prosper" window, let alone set up loan requests during this period. There is no benefit to do so whatsoever.

I can therefore rule out the possibility that the "home bias" in the results reported earlier is a reflection of private information. It is, in fact, an investor's preferential treatment toward a borrower who is a stranger to the investor – except that they came from the same state. This shared geographical tie is more likely emotional rather than rational.

### *6.2. Loan defaults on the state level*

Is it possible that investors' preference toward California borrowers is simply a reflection of past repayment history of California borrowers? If California borrowers have been systematically less likely to default than other states, then home bias may still be attributed to rational decision making. This, however, does not turn out to be the case.

I test this alternative explanation by aggregating the loan information up until the start of the "mini Prosper" period. For each state, I calculate the number of loans that were defaulted, percentage of loans in default, and the amount lost in investment (default amount). I then add these variables as additional covariates to my main model. They do not, however, change the results on the dummy variable of "same-state", or the results on economic versus spatial distances.

### *6.3. How much to lend, when to lend*

If I consider the lenders' decisions in light of the choice model literature in marketing, a natural extension of the above analysis will be the following. Conditional on choosing a Californian borrower, how much does the investor lend? Furthermore, given the open-bidding nature of the process, when did the investor actually place the bid? Did they bid early, or did they wait until others have bid?

To answer these questions I use the Heckman (1979) two-stage selection model, where the first stage outcome is whether a bid was placed. For the second stage, I consider several different outcome variables: (1) Amount of bid; (2) Winning amount of bid; and (3) The order of bids. I use the same set of explanatory variables as before (auction characteristics, lender characteristics and bidder characteristics) for consistency. I find that while the sign of the dummy variable for California borrowers are largely in a direction consistent with the "home bias" argument (larger bid amount, larger participation amount, as well as earlier bids), in most cases these coefficients are not statistically significant. This may be the result of a small sample: only 461 bids go to California borrowers. The second stage analysis is based on these actual bids on California borrowers, a very small sample.

## 7. Discussions, Limitations and Implications

Before I discuss potential limitations of the current study, I will first summarize the main findings from analyzing the natural experiment of "Mini-Prosper":

(1) Home-state bias is indeed robust, despite the fact that I only use conservative data and models are used.

(2) Home-bias shows an interaction effect with rational criteria of lender decisions. The benefit of home bias accrues mostly to borrowers with good credit grades. Those less creditworthy are in fact biased against.

(3) Economic distances among states have higher impacts than spatial distances on investor decisions. These effects cannot be rationalized by prior Prosper.com loan performance data from each state.

One possible objection to this study is that I only have data on California lenders who may or may not be representative of other lenders on Prosper.com. However, the main goal of my paper to leverage an interesting natural experiment to convincingly identify that such biases still exist under conservative data, conservative estimation procedures, and a much lower probability of endogeneity in geographical variables. Given a suitable natural experiment on other states, a replication of these findings will certainly be desirable.

A possible extension of this study is to examine the borrowing-lending relationship in the overall market. This, however, is subject to many constraints. First, if I consider the entire population of borrowers and lenders, the sample of all possible dyads are very likely unmanageable. This, in turn, will call for data reduction techniques that can potentially weaken the plausibility of home bias. Second, a unique feature of my analysis is that only CA lenders are participating. While this may sound restrictive, it is in fact an ideal feature to avoid the confounding factor of lender interactions in the auction process. For instance, the bids placed by earlier bidders may affect another lender. This is not an issue in my analyses, as all lenders are from the same state.

One other objection is that “state” is still a rather broad geographical concept. There could be significant difference in Northern California versus Southern California. Hence, a natural follow-up study is to replicate the analysis on the city level. Although I have information about the city of the borrower, that information about lenders has not been made available. It will certainly be interesting to see what happens within California itself.

It is also worth pursuing whether “home bias” is rational ex-post, as measured by the actual outcome of the loans that are generated. Unfortunately, many borrowing requests made during this window never reached the end of the duration, and therefore never became loans. Even though some of them reach 100% funding, the borrower may have chosen the “open” format for their auctions, and the auctions may still be ongoing when the site shut down. Only 13 loans were actually generated in this time period, hence there is no data to analyze the ex post rationality of “home bias”.

Nevertheless, the present analysis of “Mini Prosper” has a number of important theoretical and practical implications.

First of all, this essay of my dissertation represents the first empirical study of home bias in the online peer-to-peer lending marketplace. While there is a growing literature using data from online peer-to-peer lending websites, especially Prosper.com, to my knowledge this is the first empirical study of investors’ home bias in this marketplace.

Second, my data and analyses also represent important contributions to empirical studies of home bias in economic transactions. The natural experiment on Prosper.com creates a short time window where there are a manageable number of borrowers and lenders, and lenders are also artificially constrained to one state only. I have information on each individual level, which allows me to conduct the analysis on the level of all possible transaction dyads. Constraining the lenders to one state eliminates confounding effects from lenders in other states. Most important, the unexpected beginning and end of this window eliminates the possibility of individuals strategically choosing a location, which could have led to endogeneity in the geography variables. I am also able to rule out alternative explanations such as private information embedded in social networks or state-level default history.

More broadly, my study shows that emotions, especially those induced by shared geographical ties and shorter economic distance, do have an impact on individual decision making. Furthermore, there is an interesting interplay between emotional and rational drivers of decisions. Geography affects how individual investors utilize and process hard credit information. In other words, emotional factors interact and strengthen the effects of hard credit information on investor choices.

Last but not least, my study also contributes unique evidence to the literature on market access (Redding & Sturm, 2008), and therefore has significant policy implications. Even though e-commerce has been hailed as “borderless” or “frictionless” (Brynjolfsson & Smith, 2000), results from this study suggest that geography still matters for online transactions. Market access does matter for online commerce, a point that could have significant regulatory implications. Investors are more likely to place bids on same-state creditworthy borrowers, and this is likely to increase the chances of those loans to be funded, and also decrease their interest rates. While my study does not directly address the issue of social welfare, it does point to the fact that well-qualified borrowers, at least, suffers when there is an absence of home-state lenders. Artificially constraining one side of the market can hurt the other side of the market as well, potentially decreasingly the overall social welfare. Policymakers, especially those at the state level, should carefully consider these implications as they evaluate the request from peer-to-peer lending websites. As of the time of writing, regulators are largely willing to allow borrowers to use peer-to-peer lending websites, but much less so when they consider lender participations. $\blacksquare$

# ESSAY 3: AN EMPIRICAL STUDY OF ONLINE OUTSOURCING: SIGNALS UNDER DIFFERENT CONTRACT REGIMES

## Abstract

Technology enables new ways of monitoring worker efforts, new ways of contracting, and consequently new ways of sharing risks in outsourcing arrangements. In this essay, I study whether and how new contractual arrangements (pay-for-deliverable to pay-for-time contracts) impact the efficacy of signals that proved to be such as online reputation, certification, and language characteristics, on the chances of virtual sellers winning outsourcing contracts. Using a comprehensive dataset from an online outsourcing marketplace, I model how buyers choose among bidding sellers, and how the efficacy of these signals change under different contract forms. My results show that online reputation is an important predictor of success only for pay-for-deliverable contracts, but not significant for pay-for-time contracts. In other words, contract forms can potentially mitigate the typical Matthew Effect ("the rich get richer") of online reputation systems. Contrary to popular belief, certification does not increase the chances of winning regardless of the contract forms. This study is one of the first to study the interaction between contract formats and different signals that vendors can reveal to buyers in the competitive bidding process. More broadly, as technology provides outsourcing buyers with greater control and easier monitoring of vendors, it enables them to substitute second-party historical information (for e.g., "online reputation") with first-hand information based on their own interactions with the seller.

## 1. Introduction

Recent developments in Internet technologies have transformed many industries, and the market for labor is a vivid example (Autor, 2001). While previously only large businesses could outsource or offshore software development activities, online markets now allow small buyers and sellers to engage in transactions and even build long-term relationships. The lower barriers to enter the market allow projects of much smaller sizes to be effectively outsourced online. Meanwhile, despite the rapid growth of these markets that bring together atomistic buyers and sellers, the "virtual" and "small stake" nature of these markets exacerbates issues of information asymmetry and the likelihood of opportunistic behaviors. My goal in this essay of the dissertation is to better understand the process through which buyers and sellers[9] are matched in this marketplace. In particular, as new technologies emerge to allow buyers more effectively monitor the efforts of sellers, new contract forms emerge. I study whether and how new micro-contracting mechanisms[10] change the value of various signals in affecting buyers' choices.

The issue of asymmetric information in traditional outsourcing leads to an extensive literature on contract forms. Some of the popular forms of contracts in outsourcing include fixed-price contracts, time-and-materials contracts, and pay-for-performance contracts (Dey, Fan and Zhang 2010). Different contract forms stipulate different schemes of risk sharing. For instance, under fixed-price contracts, if software

[9] In this essay, "buyers," "employers" and "clients" refer to individuals and companies who have software needs. "Sellers," "vendors" and "developers" refer to programmers who work for the buyers. For consistency, I will mostly use "buyers" versus "sellers."

[10] As Section 3 will discuss in greater detail (page 6), I study two contract forms in this essay. Pay-for-deliverable (PFD) contracts are similar to the "fixed-price" contracts in outsourcing literature. Pay-for-time (PFT) contracts are comparable to "time-and-materials" contracts in the literature, but with one important difference: PFT contracts in my data allow buyers to very effectively monitor the effort level of sellers, which was not possible in traditional outsourcing.

developer (seller) underestimates the effort required for a project, the seller will bear the additional costs. Under time-and-materials contracts, when project scopes increase, buyers will have to absorb additional costs. Pay-for-performance contracts allow sellers earnings to be contingent upon the profits that buyers can derive from the software. Different allocations of risks lead to different incentives of buyers and sellers, thereby affecting their behaviors as well.

Fixed-price contracts have long found a counterpart in online outsourcing markets, typically referred to as “pay-for-deliverable” (PFD) contracts. As online outsourcing markets develop, new ways of monitoring seller efforts are emerging in recent years, which further give rise to new ways of contracting. One of the most exciting developments is the emergence of “pay-for-time” (PFT) contracts. PFT contracts are similar to time-and-materials contracts in the outsourcing literature, but with one important difference: online outsourcing markets allow buyers to effectively monitor the effort level of sellers under PFT contracts, which is not available in the traditional time-and-materials contracts. As I will describe in greater detail later, monitoring technologies can include keystroke logging and webcam image capturing.

Such technology-enabled monitoring provides a greater degree of control and information transparency to buyers in the outsourcing process. Buyers now have a much lower exposure to seller opportunism than under a traditional time-and-materials contract. I hypothesize that these advances in monitoring and contract formats will induce a change in the value of signals such as online reputation of sellers, which represent historical information about sellers from other buyers. Another signal that has been shown to be a significant predictor of success in outsourcing is certification, and I also test for its effectiveness in this context. As an exploration, I further investigate whether and how the initial communications from sellers to a potential buyer affect sellers' chances of winning the outsourcing contract.

This study builds on the vast literature in outsourcing from economics, information systems, and other disciplines. While the choice of contract forms is a popular topic in the extant literature, much less is known about how buyers and sellers came into contact in the first place – the process through which the buyer is matched with a seller. One possible reason is that such data is typically very hard to obtain. My study attempts to fulfill this gap in the literature by using a comprehensive dataset from a large online outsourcing marketplace, where all transactions are archived and made available. I focus on two signals in the seller's bidding process that the literature in outsourcing has shown to be effective: (1) the reputation of the seller, reflecting his or her prior experience with other trading partners; and (2) third-party certification. While the literature has documented the impact of both signals, most studies use information about buyers and sellers who actually engaged in transactions, and very little is known about those who failed to obtain the contract in the first place. There has been no study on how their effects change under different contract mechanisms. Moreover, while PFT contracts bear some similarity to time-and-material contracts, the underlying monitoring technology is entirely new. My study is the first to examine the effect of such new technology on the value of signals in the buyers' choice process.

To summarize, the main research questions that I address in the paper are:

*(1) What variables affect a buyer's choice among potential sellers;* and

*(2) How do the efficacy of the following traditional signals of quality change under different contracting mechanisms?*

*(a) Reputation mechanism reflecting a vendors' past performance with other buyers;*

*(b) Third-party certification that can potentially serve as a signal.*

These variables are identified from a review of the literature on reputation systems, contracts, and outsourcing. My working hypothesis is that, not only should these factors play a role in the buyer's decision process, but their effects should differ under different contractual arrangements because of the difference in risk sharing. The most important hypothesis is that as technology allows buyers to more effectively monitor the effort level of sellers, they should be able to substitute “old” information from others (ratings) with their first-hand experiential information with a particular worker. The value of traditional information will be significantly discounted.

## 2. Literature Review

This study is positioned at the intersection of reputation systems, certification, and contract formats, especially in the context of outsourcing. While there is a vast stream of research on these topics, I focus on research that directly relates to the current study[11].

Outsourcing has attracted significant interest in many disciplines, including economics and information systems (Koh, Soon, & Straub, 2004; Levina & Ross, 2003; Tanriverdi, Konana, & Ling, 2007). Two most popular forms of contracting are Fixed Price (FP) Contracts and Time-and-Materials (T&M) Contracts. Fixed Price contracts

[11] To avoid confusion, I will use pay-for-time (PFT) and pay-for-deliverable (PFD) to refer to the contract forms in the online outsourcing markets. Fixed-price (FP) and time-and-materials (T&M) contracts should refer to the contract forms used in traditional outsourcing literature. While FP is similar to PFD, T&M and PFT are different as PFT allows effective monitoring through the online platform.

specify a fixed price for an outsourcing project, and the vendor will be paid the agreed amount upon satisfactory delivery of projects. The risk is on the vendors: if they underestimate the cost of development, they cannot charge a higher price later. Time-and-Materials contract, on the other hand, is more flexible and shifts the risk to the buyer. Sellers are paid by time and the cost incurred, instead of the pre-specified amount.

Many theoretical and empirical studies have examined the distinction between Fixed Price and Time-and-Materials contracts. Gopal, Sivaramakrishnan et al. (2003) investigated the determinants of contract choice, and further related the choice to project outcomes using data from vendors located in India. Hasija, Pinker and Shumsky (2008) employed data from an outsourcing vendor to investigate the effect of different combinations of contract features. Through content analyses of actual contracts, Chen and Bharadwaj (2009) found that contract format has a moderating effect on the relationship between contract provisions and transactional characteristics. The stylized models of Dey, Fan and Zhang (2010) suggested that Fixed-Price contracts are better for simple outsourcing projects, while Time-and-Materials contracts are better suited for complex ones. These results are echoed in Bajari and Tadelis (2001).

The control and enforcement issues in outsourcing have also attracted researchers' attention. For instance, Kirsch (1997) proposed that a portfolio of control modes could be adapted to outsourcing. Rustagi, King and Kirsch (2008) studied variables that lead to the use of formal controls. Meanwhile, one of the classical issues in outsourcing is the hold-up problem, where the party that makes buyer-specific investments will be at a disadvantage during negotiations. Susarla, Subramanyam and Karhade (2010) studied IT outsourcing service contracts and found that contract extensiveness could mitigate the hold-up problem, but this is moderated by the complexity of tasks.

Reputation is another important subject widely studied in outsourcing. Vendor (seller) reputation has been linked to contractual performance (Banerjee & Duflo, 2000; Lewis, 1986), since the concern for reputation can potentially “outweigh the temptation to renege on a given contract” (Tykvová, 2007). Jensen and Roy (2008) modeled the choice a trading partner as a two-stage process, in which reputation helps to decide among a bracket of alternatives.

While the issue of reputation systems in electronic commerce have been extensively studied in the context of product exchanges (such as those on eBay), there has been relatively little empirical study of reputation in the offline outsourcing context. One possible reason is that firms rarely share their outsourcing experience with others, and there is also no central platform for them to do so even if they wish to. By contrast, online outsourcing markets often extensively use such reputation systems to document sellers’ performances. They thus provide an ideal context to study the use of reputation systems in the choice process of buyers in outsourcing.

A typical issue in traditional Time-and-Materials contracts is that the effort level of sellers cannot be easily monitored or verified. This has been significantly changed in the online outsourcing marketplace because of new technologies that allow buyers to effectively monitor the effort level of sellers. As buyers need to approve the billing hours submitted by sellers, they can accurately evaluate sellers’ efforts if necessary. This makes it possible for them to cautiously take some risks and conduct business with sellers who have lesser experience on the marketplace, but are able to complete the task at lower costs. If they do turn out to be of low quality, the buyers will be able to terminate the contract, instead of having to wait until the deadline under a pay-for-deliverables contract, thereby reducing the loss. Therefore, in the online outsourcing market, I should expect to see the value of online reputations of a seller (ratings) as a signal change from being a significant predictor of winning a contract under PFD schemes, to being insignificant under PFT schemes. This is a major hypothesis that I will test in the data.

***Hypothesis 1****: A seller's online rating (volume and valence) should be a significant predictor for winning an outsourcing contract under pay-for-deliverable (PFD) schemes, but insignificant under pay-for-time (PFT) contracts.*

Another topic that has received significant attention is the role of certifications, especially those from third-parties. One such study in the context of outsourcing is Gopal and Gao (2009), who studied the effect of ISO certification on outsourcing vendors. Similar to ISO certifications, the online outsourcing marketplace where I gathered the data provides links to a third-party certification website that tests the sellers' skills on different subjects. When sellers pass these exams, an icon will be displayed next to their ID and prominently displayed to buyers when the seller places a bid. Along the same vein as Hypothesis 1, I propose that the effect of certification on winning a contract should be significant under pay-for-deliverable contracts, but not so under pay-for-time contracts. .

***Hypothesis 2****: A seller should have higher chances of winning an outsourcing contract when he or she has been certified under pay-for-deliverable (PFD) contracts, but insignificant under pay-for-time (PFT) contracts.*

Figure 1 illustrates the conceptual model for the current study.

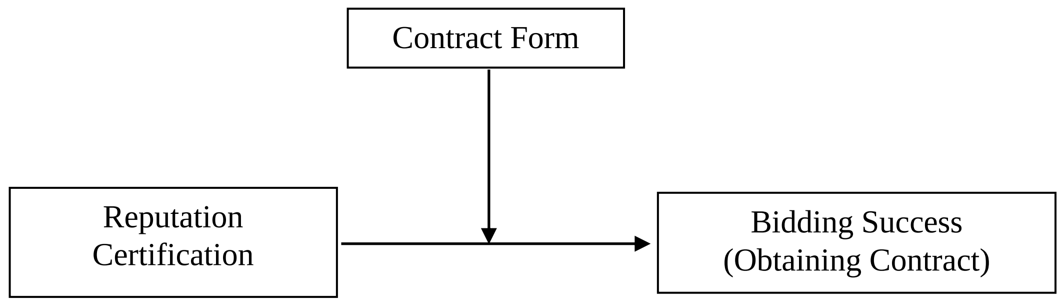



**Figure 1: Conceptual Model**

## 3. Context

In this section, I describe major features of the online labor outsourcing marketplace that provided the data for my analyses, the introduction of a new contracting mechanism, the construction of matched samples for analyses, and the derivation of various variables used in my statistical models.

I obtained data for this study from one of the leading online software outsourcing marketplaces. This marketplace is headquartered in the United States, but buyers and sellers of the market come from all over the world. The largest proportion of work done on this site is customized software development, although more recently there has also been a growing need for graphic design and other tasks. Software development programs include designing a website, enhancing e-commerce website features, file format conversions, and so on. Some examples of typical requests posted on the website are[12]:

> *"We need to integrate our website with Google Checkout … If client selects "credit card" as the payment method, he arrives to a custom payment page where he can enter credit card information. This card information is sent to bank processor, if approved, then a receipt page is displayed and order data is updated. If ……"*

[12] Due to non-disclosure agreements, these descriptions have been slightly changed to protect the privacy of website users.

> *"Small site design needs to be re-coded/optimized for newer standards. Requirements: 1.Convert to XHTML 1.0 strict; 2.Convert all of site design to use CSS; 3. MUST be W3C validated..."*

> *"I need a PHP script that will read every SQLite database in a directory. These databases all have the same schema. Insert their entries into a MySQL table with a similar schema, plus a column for file names."*

### *3.1 Overview*

This proprietary dataset contains extensive information about software buyers, sellers, their transactions, communications, mutual ratings and other information from October 2001 to October 2010 (as I will discuss shortly, my analyses will not use the full sample). This is a marketplace of over 250,000 software developers (sellers) and more than 120,000 software buyers from around the world. These are typically small software development projects, mostly ranging between $150 and $300. Prior to September 2009, all projects are "fixed-price contracts": Sellers are paid when they deliver satisfactory products according to buyer requirements. If the sellers underestimate the time and cost involved, they will have to bear the extra cost.

### *3.2 A Typical Process of Transaction*

This section describes the typical process of transaction on this market as of the time of writing.

Buyers and sellers first sign up with an email address. Before they enter any contracts, the website will verify their identity. For instance, US buyers are verified by phone, credit card, and driver's license information. Once the verification is complete, buyers post "request-for-bids" on the site. A typical request includes a title, a summary of requirements, and a rough budget (e.g., maximum $500). It should be noted that it is buyers who determine the form of the contract (pay-for-time or pay-for-deliverables).

Software developers (sellers) can browse the requests, search for keywords, and they can be notified of new projects should they choose to receive such alert emails from the site. When they find a project of interest, they can post a bid, which is the amount they will charge for the delivered product. Along with the bid, they can (optionally) submit a text message trying to convince the buyer that they are a good candidate. This is very similar to the “pitch” that entrepreneurs make to venture capitalists when they seek funding. It should be noted that these are sealed bid auctions, in the sense that only the buyer can see the bids placed; peer sellers cannot observe each other’s bids.

Buyers can communicate with the sellers, and then choose a seller to work with by accepting his or her bid. Buyers can choose any bid he or she wants, and lower price bids do not necessarily win. This is an important feature of the online labor market that distinguishes itself from websites such as eBay.

Once the bid is accepted, the buyer will first pay the amount of the bid by credit card or electronic check into an escrow account of the site. Then, the site sends a notification to the seller that they can start working on the project. A contract is thus created.

The buyer and seller communicate with each other through the website to clarify requirements and other details. When the seller finishes the project, he or she uploads programs to the site, and the buyer can download it to test whether the requirements are met. If so, buyers accept the project as 100% complete, and the funds are released from the escrow account to the seller. Buyers and sellers can then rate each other on a scale of 1 to 10, and also provide comments on each other. These ratings become public information on their profiles, and form the “reputation” system on the marketplace.

The website deducts a percentage of fees from the escrowed amount when it is released to the vendor. These fees are not only for the provision of an infrastructure and possible arbitrations (see next paragraph), but also for taking care of paperwork related to taxes and other issues involved in paying another person, especially those in a foreign country.

If the project is not completed due to any reason, it typically enters arbitration. The arbitrator is a staff member of the site and the arbitration process can be initiated by either the buyer or the seller. The arbitrator will review all communications on site, including the original requirements, and will contact both parties. Offline communications, if any, are not considered in the process. If either party fails to respond, he or she receives a low rating, and loses.

*3.3 Emergence of Pay-for-Time Contracts*

Up until September of 2009, all projects in this online marketplace use the "pay-by-deliverables" (PFD) format; that is, the buyer and seller agree on the requirements of the project at the beginning, and the cost of the project is fixed. This is comparable to the "fixed price" contracts in outsourcing, where the buyer's obligation is limited a priori and the burden of risks falls on the vendors. In September 2009, the website started to allow buyers and sellers to enter "pay-for-time" contracts. Under PFT arrangements, sellers are merely paid by the number of hours they work on the project without guaranteeing the outcome, and the buyer can terminate the contract at any time. If mutually agreed, the contract can extend at the agreed hourly rate. This is made possible only because the website created an application to allow the buyers to closely monitor the efforts of the sellers. When the sellers start working on a pay-for-time project, they will log into the monitoring software, which will take random screenshots, keystrokes as well as webcam pictures at certain intervals. The buyers can also manually take additional pictures or keystroke recordings as required by the contract. These records are kept for the purpose of arbitration; if the buyer believes that the seller has inflated the number of hours, the arbitrator can use these recordings as evidence.

This change in technology provides an interesting context for me to study how different contract formats change the effect of various seller signals (reputation and certification) on winning a contract.

### *3.4 Reputation Systems*

Much like eBay, the website has developed extensive reputation systems for the sellers so as to assist buyer's choice among candidates. When buyers and sellers first sign up, they have no ratings. When a project is completely successfully, buyers and sellers can rate each other. The rating has a numeric part that ranges from 1 to 10 stars, as well as a textual part that they can comment on the rating. Information about number of ratings that the seller has received up to the time of the bids, as well as the average of those ratings, are displayed prominently to future buyers when they look at the list of sellers who placed bids.  Just as in eBay and other e-commerce markets, online reputation systems is a reflection of a seller’s past performance with another individual. My working hypothesis is that such information can be rendered less relevant in buyers’ decision process when they have access to the efficient monitoring technologies.

### *3.5 Certifications*

The website also works with a third party provider who allows sellers to take exams online on different subjects. As of the time of writing, these exams are free of charge. If they fail an exam, they can wait a few days before making another attempt. After they pass an exam, the website will display an “Expert” icon next to their bids when the category of their exam matches the category of the project, such as the programming language required by the buyer.

## 4. Sample Construction

To understand the factors affecting buyers' choices of vendors and rule out alternative explanations, I constructed two samples of auctions that consummated in actual contracts (i.e., buyers matched to a seller). I then extracted all bids placed in those auctions, information about buyers and sellers as of the time of the contract, as well as project descriptions and communications. The next paragraph describes additional details about how these two samples are constructed. I will test whether effects of various variables on a bidder’s success change under different contract regimes using these samples, both jointly and separately.

*Sample #1: Pay-for-time (PFT) contracts*

Although PFT is a promising new mechanism, it has not yet gained traction in this online marketplace. One year after PFT contracts are allowed, there were still less than 200 such contracts that were actually created between buyers and sellers. To allow for meaningful statistical analysis, I try to retain as many PFT contracts as reasonably possible.

I first removed PFT requests that did not result in an actual contract. While I have data on bid requests (auctions) that do not have any winning bids at all, I exclude them because a contract failing to consummate may be due to unrealistic requirements of the buyers, instead of any characteristics of the vendors. Focusing on contracts that were actually created eliminates confounding factors from the buyer's side.

Less than 1% of the PFT contracts are between buyer-seller pairs who had prior relationships. I removed these observations as buyers are faced with much lower levels of uncertainty in those cases; this is also to ensure consistency with sample #2.

*Sample #2: Pay-for-deliverable (PFD) contracts*

Pay-for-deliverable contract is the original format used on the website, and it is also what most users are accustomed with. Hence, even after pay-for-time contracts are made available; many users (buyers and vendors) continue to use pay-for-deliverable contracts. This is especially true among buyer-vendor pairs that already have repeated transactions. Almost all pay-for-time contracts (one year after the implementation of the new method) are between buyer-seller dyads that do not have prior transactions. Hence, for Sample #2 (PFT contracts), I also removed the contracts between parties with prior experience. This ensures that buyers in these two samples face comparable degrees of uncertainty when they choose among the vendors.

I then retained only PFD contracts in the three months prior to the introduction of PFT contracts (June - August 2009). This was to ensure that vendors did not face resource constraints and had to choose between PFD and PFT auctions posted at the same period of time.

Subsequently, I took a random sample of PFD between June and August 2009 so that there was approximately the same number of contracts in Sample #2 as in Sample #1.

Once both samples were constructed, I extracted all bids related to those auctions, including information about certifications, ratings and so on at the time that the seller placed the bid.

*4.1 Level of Analysis and Dependent Variable*

The level of analysis in my model is each bid; I study how characteristics of sellers' (reputations; certifications) and their comments are associated with the outcome of their bids. The dependent variable is a dichotomous variable that takes on a value of 1 when a bid wins the buyer's contract; 0 otherwise.

*4.2 Independent variables:*

(1) *ExpertCertification*: Indicates whether or not there is an "Expert Certification" icon next to the sellers' bids. Unlike certifications in traditional outsourcing contexts which are much more difficult to obtain, certification on this market is free of charge, and sellers are allowed to re-take exams. In other words, this is a relatively cheap signal. Whether or not this is a useful signal is an empirical question.

(2) *noRating*: An indicator variable that the seller has not yet received any ratings.

(3) *AvgRating*: The mean of ratings that the seller has received up to the time that the bid was placed.

(4) *logRatingsCount*: Logarithm of the number of ratings that the seller has received up to the time of the bid.

(5) *BuyerSellerSameCountry*: An indicator variable that the buyer and seller are residents of the same country. The literature typically suggests that buyers and sellers in the same country are more likely to interact with each other (Hillberry & Hummels, 2003), either due to homophily (Reagans, 2005) or lower transaction costs (Redding & Sturm, 2008).

This is, in fact, closely related to the “home bias” concept that I studied in the second essay of this dissertation.

(6) *BuyerSellerBothUS*: An indicator that both parties are from the United States. This is a special case for “BuyerSellerSameCountry”.

(7) *logSellerMonth*: Logarithm of the number of months that the seller has signed up on this market.

(8) *logExpertiseLength*: Each seller has a “resume” page where they can post their resumes or further describe their experiences and expertise (which are not verified by the site). This variable captures the length of the document.

(9) *logBidAmount*: Logarithm of the dollar amount of the bid.

(10) *logBidOrder*: Logarithm of the order in which the bid was placed. A larger number suggests that the bid was placed later. Since the bids are displayed in the order they are received by default, earlier bids are more likely to be noticed and accepted.

(11) *noCommentBid*: An indicator variable that the bid does not come with a message.

(12) *ProjectAmtRange*: controls for the size of projects, I first calculated the final project cost that the buyer actually paid for each project in the sample. For pay-for-time contracts, this is the hourly rating that the seller bid, multiplied by the estimated number of hours. This amount is then “binned” into different intervals: 1 if it's lower than \$100, 2 if it's between \$100 and \$200, 3 if between \$200 and \$300, 4 if between \$300 and \$400, and 5 for \$400 and above.[13] These are then included in the estimation as a series of dummies in a saturated model specification.

[13] The examples of projects on page 10 cover different ranges of the project costs.

## 5. Model

The main goal of this study is to understand how buyers' choice criteria of potential sellers change under different contract mechanisms. One of the major hypotheses that I test is whether the new monitoring technologies in PFT contracts reduce the effectiveness of traditional signals such as online reputations and certifications, as compared to PFD contracts. The dataset used in this study has some very appealing features that suit this purpose very well. First, since I focused on buyers and sellers who do not have prior relationships, all information that led to a buyer's decision is captured in the data. Unobserved factors that may have contributed to these choices are minimized, and can be considered orthogonal to my variables of interest. This is consistent with the identification strategy in Angrist (1998). Second, whereas most prior studies only have information about sellers that were ultimately chosen, I have information about others who are rejected by the buyer. Such information sheds light on how buyers made their decisions. Third, this website uses a sealed auction format; only the buyers see who the bidders are, and how much the bid amount is. This ensures that the bids among sellers are largely independent of each other, allowing for proper statistical modeling.

Given these features, I used maximum likelihood estimation of logistic models to estimate the probability that the buyer accepts a bid. Independent variables are virtually all the information that buyers had access to when they decide whom to contract. Although all bids placed in these auctions are independent of each other, I estimated the standard errors using clustered sandwich estimators to allow for intragroup correlation, where a cluster is specified to be an auction (a request for bid).

*5.1 Main model of buyers' choice: Full sample analysis*

My main working hypothesis is that the effect of reputation and certification (hypotheses 1 and 2) changes when contract forms change. Hence, I incorporated in the logistic regression a dummy variable indicating a Pay-for-Time contract, which was then interacted with other variables of interest. For the overall sample – which includes bids from bidders who have no ratings, no expert certifications, or bids that were placed without textual comments – I estimated the following model:

$$\begin{aligned}\text{Prob}(BidWins = 1 \mid x) = {} & \beta_0 + \beta_1 noRating + \beta_2 PFT * noRating \\ & + \beta_3 ExpertCertified + \beta_4 PFT * ExpertCertified \\ & + \beta_5 noBidComment + \beta_6 BuyerCoderSameCountry \\ & + \beta_7 \log BidAmout + \beta_8 \log BidOrder \\ & + \beta_9 \log CoderMonths + \beta_{10} \log ExpertiseLength \\ & + \beta_{11} PFT + \varepsilon\end{aligned}$$

In other words, I multiplied the PFT dummy with dichotomous variables that indicated no rating bids, no comment bids, and no certification bids – respectively. Results of this model are shown in Table 3.1. We can see that PFT itself is statistically significant, suggesting that the intercept term for the PFT and PFD are different. Its interaction with the no-rating dummy is also significant; the other two interactions are however not. It appears that ratings play a different role under different contract forms, but the effect of certification is insignificant.

To delve deeper into the differences across these contract forms, I then excluded interaction terms and estimate the model separately on the PFT contract subsample (Sample #1), and the PFD contract subsample (#2). Unreported results show highly consistent patterns with Table 3.1: dummy variables indicating whether or not there is no comment, and whether or not there is no certification, are statistically insignificant for both subsamples. The dummy variable for no-Rating also shows a pattern consistent with my hypothesis: sellers (developers) with no ratings are significantly disadvantaged under Pay-for-Deliverables contracts, but only marginally significant for PFT contracts. Therefore, Hypothesis #1 is supported.

Hypothesis 2 (certifications), on the other hand, is only partially supported. While certification is insignificant under pay-for-time contracts (consistent with Hypothesis 2), it is also insignificant under pay-for-deliverable contracts. A possible reason is that the threshold of such certifications in this market is relatively low: Exams are available online for free, and sellers can take exams multiple times until they pass. Hence, even under pay-for-deliverable contracts, the effect of such certification may not be significant.

Some auxiliary results are also interesting. I found evidence that on average, buyers prefer sellers who are from their same country, a phenomenon consistent with the "home bias" literature and also the findings from the second essay of this dissertation. The tendency to offshore is actually less than what mass media would have us believe: The odds that a same-country vendor is chosen are actually over 250% of that of someone in a foreign country. This pattern persists in many more specifications that I shall discuss, and is robust to the inclusion of variables such as the time zone difference and whether English is the official language. In addition, when I replaced this variable with a dummy that took the value of 1 when both buyer and seller are from the United States, I obtain the same result. In other words, under comparable degrees of uncertainty (first time interactions), US buyers also prefer domestic sellers rather than foreign ones.

Some auction variables are also significant predictors of bidding outcomes, and their results are largely to be expected: bids placed earlier are more likely to be successful, and higher amount of bids are less likely to be chosen.

I further investigated the interaction effect between some other variables. For instance, even though the model above suggests that having no ratings is a bad signal, it could be much worse if the seller has been on the market for a long time. The fact that a seller has been on the market for a long time but has obtained no contracts can indicate bad quality. Future buyers can simply "herd" and choose to ignore those sellers.

To test this effect, I included another interaction term between the indicator variable for "*no rating,*" and "*number of months since vendor signed up.*" While "no rating" is shown to be negatively associated with the chances of winning, it is significantly worse for sellers who are on the market longer. In other words, between two sellers who are not rated, the ones who joined the site earlier are even less likely to win a contract. No-rating suggests that the seller has not been chosen by any other buyer so far. The longer they stay in that situation, the less attractive they become.

### *5.2 Modeling buyers' choices: Volume and valence of ratings*

The above analyses, however, only use dichotomous variables for rating and certification. This may be sufficient for certification (bids either have an "expert" icon next to it, or it does not), but it is certainly worth exploring the actual level of rating and the number of ratings.

I first analyze the number of ratings as well as the average rating of sellers when they place bids. These variables are displayed prominently to buyers when bids are placed. I replaced the dummy variable of "no rating" with two new variables: (1)

logarithm of the number of ratings that the seller has at the time of the bid; and (2) the average of ratings that the seller has received at the time of the bid. To study whether the effect of the volume and valence of ratings change under different contract formats, I first conduct an interaction analysis by creating interaction terms between PFT (dummy) and the logarithm of number of ratings, as well as the interaction term between PFT (dummy) and the average rating. I then run the overall model with these interaction terms in the combined sample. I also excluded auctions that choose sellers who did not have ratings at the time of the request in this estimation. Results are shown in Table 3.2. It can be seen that the volume of rating at the time of bid has a significant interaction effect with the PFT dummy variable, suggesting that at least the effect of rating volumes could change when contract forms change. To further illustrate how the effect of ratings change under different contract formats, I then ran the same analysis on these two subsamples sequentially, and present these results in Table 3.3. Table 3.3 shows that variables examined previously display very consistent results: certification is insignificant, while bidding order and bid amount matters. On the other hand, ratings variables show some interesting patterns. For instance, while having a large number of ratings in pay-for-deliverable (PFD) contracts significantly increases the chances of securing the contract, the effect is statistically insignificant in pay-for-time (PFT) contracts. Meanwhile, while the average rating has a positive and statistically effect on the chances of winning a contract under PFD arrangements, the effect is insignificant for PFT contracts as well. In fact, the magnitudes of these coefficients are also smaller in PFT contracts.

In other words, the results above suggest that sellers who entered the market earlier, established good reputations and accumulated a long work history had a significant advantage when they competed in the market, especially under pay-for-deliverables (PFD) contracts. This in turn gave them more opportunities to increase the volume of their ratings. The cumulative advantage for established sellers in the online market can be very significant, consistent with the predictions of Matthew Effects (Merton, 1968). This effect exists both for the volume of ratings that the seller has, as well as the valence of ratings. The more jobs you did in the past, and the better you did on those jobs, the more likely you are going to get jobs in the future. While this is not entirely surprising, it does have a detrimental effect on the competitiveness of the market. New entrants will find it very difficult to compete with the incumbents. This can also be harmful for the development of the marketplace itself, as it competes with other online outsourcing platforms to attract new users.

My results under pay-for-time (PFT) contracts, by contrast, show that it is possible to mitigate the market's tendency to polarize by implementing new contract forms. By redistributing the burden of risk between buyers and sellers, pay-for-time contracts allows buyers to "experiment" with sellers who are less experienced on the market,[14] giving them a chance at the competition. From another perspective, the new PFT contract, made possible by the advancement in monitoring technologies, allows buyers to substitute old information from others (ratings) with first-hand experiential information derived from their own interaction with a seller.

[14] It should be noted, however, that a lack of experience on this market does not mean the developer him/herself lacks experience. They may simply be new to this marketplace, and it is difficult for these online markets to verify the validity of their resumes.

## 6. An exploratory analyses: Textual contents of seller communications

Textual content of sellers' communications to potential buyers could play an important role in buyers' choice of sellers. What they place in their first communication to the buyer can affect their chances of winning due to the limited amount of information that a buyer has of the seller. Much like entrepreneurs "pitching" their ideas at venture capitalists, these vendors only have a limited opportunity to convince a potential buyer. Moreover, since buyers have different priorities when deciding between PFT versus PFD contracts, it is very likely that some language features are important for pay-for-time contracts, while others are more important for pay-for-deliverable contracts. So far, no empirical studies have examined the effect of textual comments, largely due to the sensitive nature of such data. To fill this gap, I conduct exploratory analyses of how the characteristics of textual communications affect sellers' chance of winning a contract. I estimated the model on the subsample of auctions that do not choose a bid without comments.

The first variable that I examined is the number of typos. It is possible that typos make it difficult to communicate; therefore, a higher number of typos can make a seller less attractive. Meanwhile it is also likely that buyers can be tolerant of these typos in search of a good deal. I use open-source software GNU Aspell to achieve this by submitting these text files (via Perl scripts) to an English dictionary[15] associated with GNU Aspell, and compared each word against the dictionary. The number of typos was recorded for each comment associated with bids.

[15] Copyrighted by Kevin Atkinson, http://wordlist.sourceforge.net/

I studied two alternative metrics for typos: the total number of typos in the vendor's message, and the ratio of this number to the total number of words. While the odds ratios associated with these variables were indeed smaller than 1, they were not statistically significant in all specifications described earlier. It thus appears that buyers in this market seemed to be tolerant of typos; they either did not consider them a signal of communication difficulties, or fully anticipated them in this market.

The second set of variables is generated from LIWC (*Linguistic Inquirer and Word Count*), a computerized linguistic analysis program that I also used in the first essay of the dissertation. The two main variables that I am interested in are "Time" and "Money" variables, as they represent two of the main dimensions that buyers consider when they choose a vendor. I find that "time" words are a statistically significant predictor of bid success only for pay-for-time (PFT) contracts, but insignificant in pay-for-deliverable (PFD) contracts. For PFT contracts, a larger number of "time" words are associated with higher chances of winning. "Money" words, on the other hand, are also significant only for PFT contracts. The difference is that a larger number of "Money" words are associated with *lower* chances of winning.

Results reported in Table 3.4 suggest that buyers are more concerned about contents in the sellers' messages when screening transaction partners under PFT contracts rather than PFD contracts, consistent with the risk that they are bearing. Under pay-for-time (PFT) contracts, buyers will pay for the number of hours that the developer will be working. Therefore, a more detailed discussion of time will alleviate buyers' concerns on that dimension, and can increase a seller's chances of winning the contract. On the other hand, when the contract is "pay-for-deliverables," (PFD) buyers seem to be place less attention to the content of the text messages. The price of the project is the main information that a buyer is concerned about – most categories of the text characteristics do not significantly predict winning a contract, either statistically or economically.

A few other variables also show some noteworthy patterns. I found that the number of words with more than 6 letters is negatively associated with chances of winning, although the odds ratio is relatively small in scale. “We” words (including “we,” “our,” and so on), on the other hand, were not significant predictors in either PFT or PFD contracts. While this contrasts with the literature (Bagozzi & Dholakia, 2006; Hardy, Lawrence, & Grant, 2005; Kilker, 1999; Levina 2006), this result may simply reflect the nature of jobs being outsourced on this market, which are a lot smaller than the multimillion dollar projects reported in prior empirical studies. Many of these results are robust to specifications.

## 7. Implications

Sociologists have long identified the Matthew Effects (Merton, 1968) in economic life: in a competitive environment, individuals, organizations and entities that were previously in an advantageous position can continue to enjoy their advantage. This is similar to the idea of “preferential attachment” (Hills, Maouene, Maouene, Sheya, & Smith, 2009), or the phenomenon of “the rich grows richer, while the poor grows poorer.” Sociologists also refer to this as Cumulative Advantage (DiPrete, Eirich, Cook, & Massey, 2006). For electronic commerce websites, such tendency may not be ideal as it is likely to drive away new vendors, yet it is indeed happening: Consumers are more likely to buy from sellers with more ratings and higher ratings.  This essay of my dissertation reveals an analogy in online outsourcing: Pay-for-deliverables (PFD) contracts, a dominant form of contract on this marketplace, very much favors sellers (developers) with longer job history on the marketplace, and those with higher average ratings. The cumulative advantage inherent in these online rating systems has significant implications for market design, competition, and public policies as well, as it can easily lead to higher market concentrations over time, and limit the competitiveness of new entrants.

One implication of this study is that changing contractual forms, made possible by the emergence of new monitoring technologies in online outsourcing markets, can partially mitigate Matthew Effects. By redistributing the burden of risk between buyers and sellers, buyers will have an incentive to take some risks and hire less-known, less-experienced sellers under pay-for-time contracts. This is because they are allowed to stop the transactions if the sellers turn out to be of low-ability. In other words, buyers are substituting second-hand historical information, with their own first-hand experiential information. Broadly speaking, electronic commerce websites concerned about expanding their customer base can consider incentive mechanisms to redistribute the risks between buyers and sellers.

The second implication from the findings discussed above is that certifications may not always be effective. It is possible that this is unique only to this website, and only to the particular types of certifications. However, given the popularity of third-party certifications in decentralized online markets such as eBay, the current study suggests that we should not take certifications' effectiveness for granted. Although such certifications do provide an extra “icon,” they do not necessarily increase vendors' chances of obtaining contracts.

Last but not least, my exploratory analyses show some interesting potentials for computerized text analysis. While the analyses I reported here is largely exploratory, the results do show that buyers take into account what was written by the vendors, especially under pay-for-time contracts where the relationship is more persistent – rather than a one-shot exchange under PFD contracts. For platforms such as online outsourcing markets, implementing automatic text analyses programs can potentially help buyers increase the efficiency of screening vendors, especially as when I am able to link textual cues to project outcomes. This will be addressed in a separate paper.

## 8. Limitations and Future Research

The current study has a few limitations that readers should be aware of. Software projects in this context are much smaller compared to typical outsourcing contracts, and this could constrain the generalizability of my findings. Indeed, many outsourcing contracts that we read about in the mass media involved millions of dollars over multiple years. By contrast, the contracts in this marketplace represents the “long tail” of online outsourcing. Therefore, readers have to take these findings with a grain of salt when they apply them to larger contracts. Replicating my analyses on datasets of larger vendors and clients will be certainly highly desirable, especially if researchers can also obtain information about reputation, certifications, and communications between vendors and their clients, including vendors who are unsuccessful in their bids for contracts.

My primary goal in this paper is not about the choice of contract formats. By constructing two non-overlapping subsamples of different contractual forms, I sought to understand how contract forms moderate the relationship between a vendor's reputation and their chance of winning a contract. A natural extension of this analysis is certainly to go beyond dyads of first-time interactions and better understand the endogenous choice of contractual forms in this context, especially between buyers and sellers who have repeated interactions. As described in the paper, the proportion of buyer-seller pairs that switched from pay-for-deliverables to pay-for-time contracts is very small. It is possible, however, that as users become more familiar with this new arrangement, I will observe more "switching" of contract forms. At that time, I will be able to extend the analyses in this paper to model the contract choice endogenously.

A second extension of the current study will be incorporating some metrics for the outcome of projects. This will be the focus of a separate paper.

Another limitation of the current study is that, while I focus on buyer-seller pairs that have no prior contacts so as to reduce the confounding effects of endogeneity, the analyses are still inherently cross-sectional. My ongoing research using this dataset will leverage natural experiments and microeconometric techniques to better identify the effects of some variables, especially the causal effect of certifications.

My analysis of the textual comments is one of the first efforts to study the effect of written language on buyer choice of vendors. Though there are much more advanced text mining techniques available, LIWC has been broadly used in psychology and management, and it yields similar results to other packages such as *General Inquirer* (Tetlock, Saar-Tsechansky, & Macskassy, 2008). Yet another valid critique of my analysis is comparable to the "Lucas Critiques" in economics: when vendors realize how buyers respond to their wordings in the communications, they may accordingly change how they write in the future, which can potentially change how buyers screen vendors. These are certainly interesting dynamic interactions that can be explored in future research. However, it does not affect the validity of the current research; these communications are private between buyers and vendors. These communications are only obtained under a non-disclosure agreement, and neither this website nor its competitors has done analyses like this before. At least in the time frame that I studied, no such results were revealed to vendors.

## 9. Conclusions

The advancement of information technologies, especially Internet technologies, promises to change the landscape of labor markets forever (Autor, 2001). How buyers and sellers (workers) are matched, how services are delivered, and how efforts are monitored, will be dramatically different in the online market. The monitoring technologies used on the website that I described here is but one such development, yet has already brought about a new contract form (pay-for-time contracts) that dramatically changes how buyers use traditional signaling mechanisms. As my results show, with pay-for-time (PFT) contracts, buyers can forego second-hand, past information about a potential trading partner (online ratings) and instead use first-hand information through their own interaction with a seller. Further developments of internet technologies will continue to change the way transactional ties are formed in decentralized online markets, creating a rich area for empirical research. ∎

# DISSERTATION SUMMARY AND CONCLUSION

The growth of online marketplaces populated with atomistic individuals and small firms facilitates transactions that were infeasible in traditional settings. While the sheer number of smaller participants exacerbates issues of information asymmetry, web 2.0 technologies have also enabled the growth of innovative mechanisms, such as social networks, that can potentially help mitigate informational problems in these decentralized markets.

Online peer-to-peer lending and software outsourcing are two examples of such markets and provide the contexts for my dissertation. Drawing on theories from information systems and other disciplines, I empirically study how participants in these markets establish transaction relationships to realize the gains from trade. Each of these essays emphasizes one particular factor that determines who is matched to whom in these online markets. In the first essay, strangers who are not part of the borrowers' network can screen borrowers by the nature of ties that the borrowers possess, and the presence or absence of certain ties can serve as informational cues about the borrower's credibility. In the second essay, I find evidence that even though the context is electronic commerce, geographic information still has an impact on individual decision making. Finally in the third essay, I find that contract mechanisms moderate the effect of various signals on buyers' choice of vendors.

In summary, the three essays of my dissertation provide a better understanding of how online social networks, geographic information, contracting forms and various other signals affect the formation of transactional ties in these highly decentralized, "virtual" marketplaces. Findings from these studies not only are beneficial to market participants, but even regulators who will find it increasingly necessary to regulate these markets. Moreover, my studies also contribute to the growing IS literature of empirical studies in electronic commerce, especially those related to online social networks, outsourcing, contracting, and reputation systems.

# APPENDICES FOR ESSAY 1

**Figure 3**
"Hierarchy of Friends"

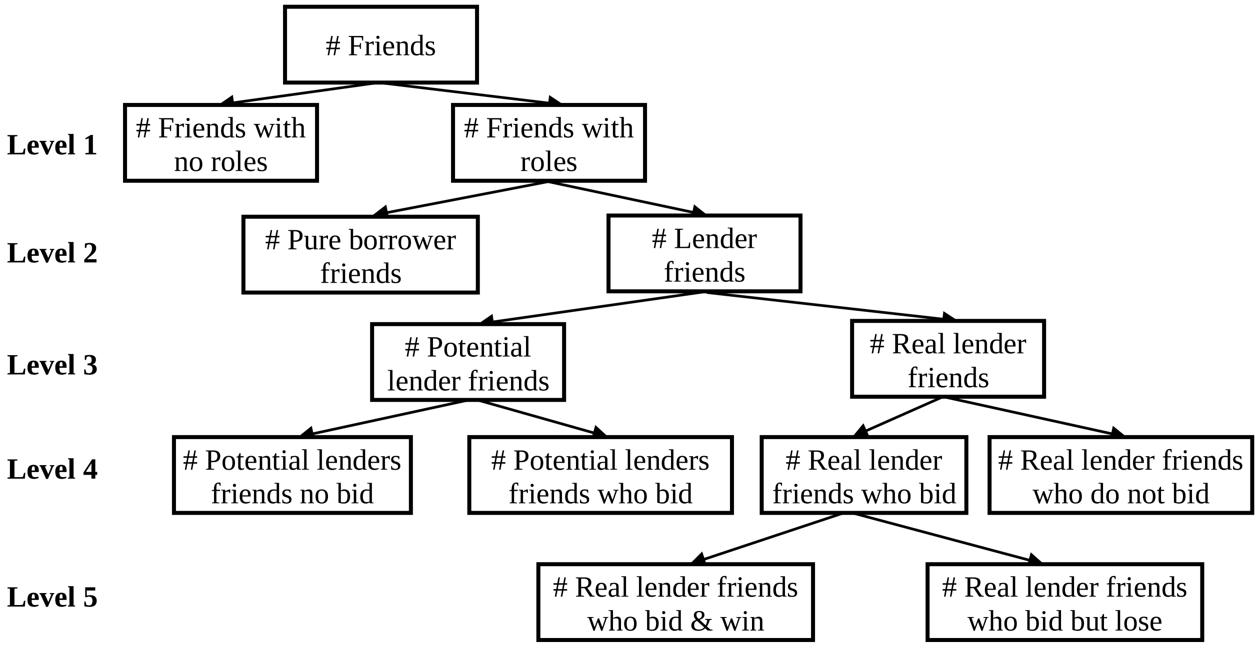



**Figure 4**
Screenshot of a Prosper listing

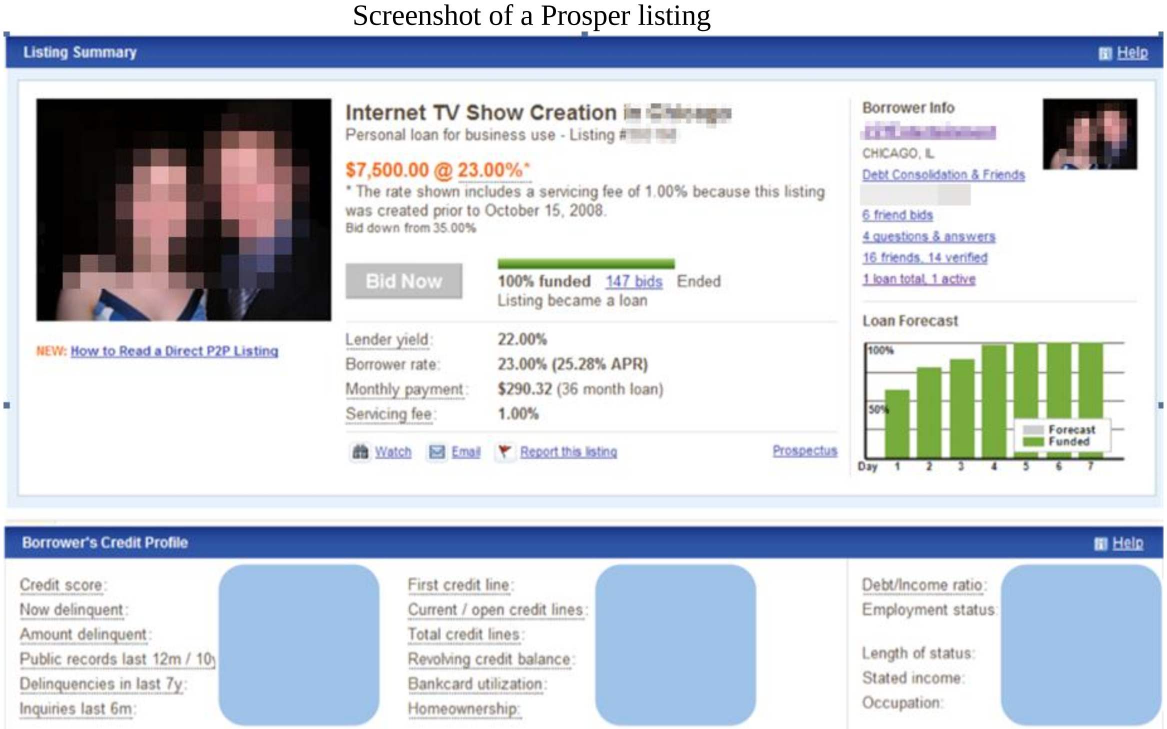


**Figure 5**
Smoothed Baseline hazard function

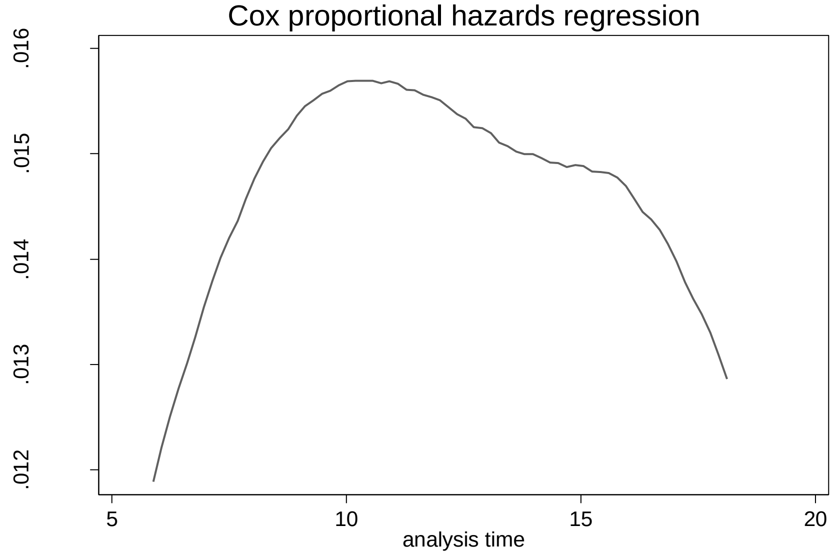
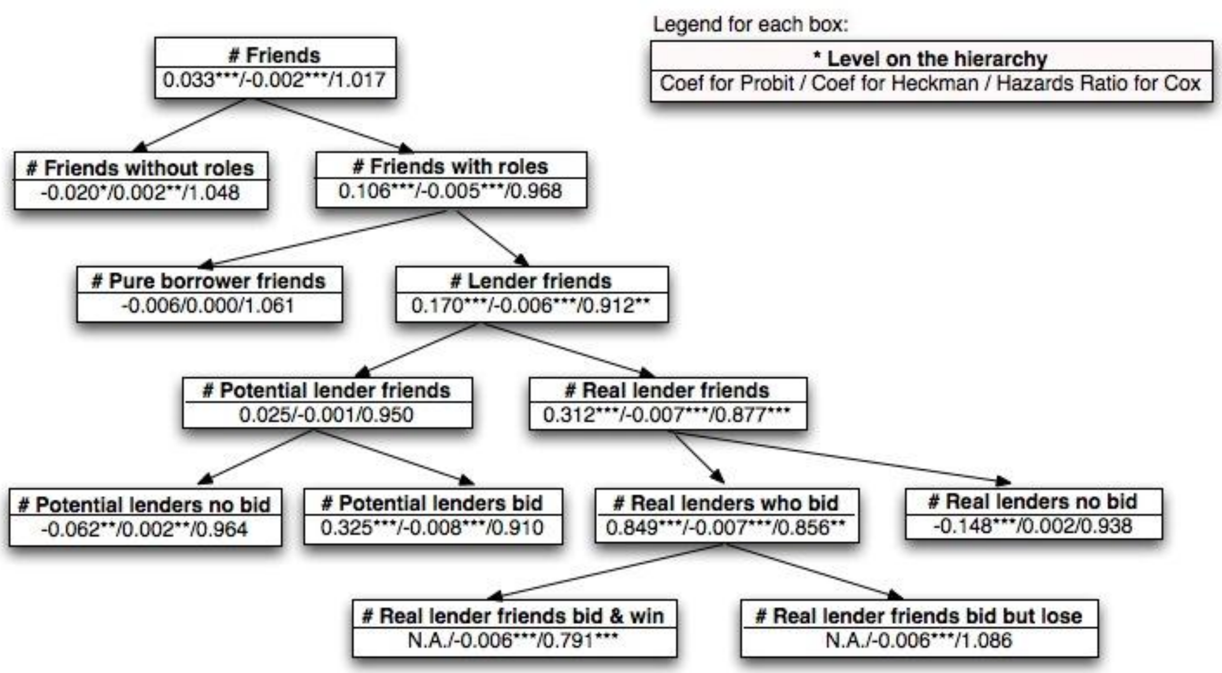



**Figure 6**
"The Hierarchy of Friends": Results




The three numbers in each box are the coefficient for funding probability, coefficient for interest rate on funded loans, and the hazards ratio in the Cox model, respectively. * p<0.1; ** p<0.05; *** p<0.01.

## Table 1.1: Borrowers' FICO score and Prosper credit grades

| Prosper Credit Grade: | AA | A | B | C | D | E | HR |
|---|---|---|---|---|---|---|---|
| Borrower's FICO Score: | 760 and up | 720-759 | 680-719 | 640-679 | 600-639 | 560-599 | 520-559 |

This table reports the correspondence between the letter ratings assigned by Prosper.com to a listing and the listing borrower's Fair Isaac Credit Score.

## Table 1.2: Variables and their descriptions

Table 1.2 reports descriptive statistics for the independent variables used in the paper. My sample comprises 205,131 borrower listings on Prosper.com that have a listing date between January 2007 and May 2008.

| Variable Name | Variable Description | Variable Name | Variable Description |
|---|---|---|---|
| | *Hard Credit Information* | | |
| CreditGradeAA | 1 if borrower's credit grade at time of listing is in grade AA; 0 otherwise. This is the baseline grade, not included in estimation. | CreditGradeHR | 1 if borrower's credit grade at time of listing is in grade HR; 0 otherwise |
| CreditGradeA | 1 if borrower's credit grade at time of listing is in grade A; 0 otherwise | DebtToIncomeRatio | Debt-to-income ratio of borrower at listing |
| CreditGradeB | 1 if borrower's credit grade at time of listing is in grade B; 0 otherwise | BankCardUtilization | Bank Card Utilization of borrower at time of listing, or the percentage of credit line issued by the bank that has been utilized |
| CreditGradeC | 1 if borrower's credit grade at time of listing is in grade C; 0 otherwise | BankCard2 | Quadratic term of BankCardUtilization |
| CreditGradeD | 1 if borrower's credit grade at time of listing is in grade D; 0 otherwise | InquiriesLast6months | Number of credit inquiries in the prior 6 months before listing |
| CreditGradeE | 1 if borrower's credit grade at time of listing is in grade E; 0 otherwise | YearsSinceFirstCredit | Number of years between the the borrower's first credit line and the time of listing |
| | *Auction Characteristics* | | |
| AuctionFormat | Dummy: 1 for a close auction | ListingCat5 | 1 if the borrower chooses "Student Loans" as the listing category |
| BorrowerMaxRate | Borrower's asking interest rate on the listing | ListingCat6 | 1 if the borrower chooses "Auto Loans" as the listing category |
| BorrowerMaxRate2 | Quadratic term of borrower's max rate | ListingCat7 | 1 if the borrower chooses "Other Loans" as the listing category |
| AmountRequested | Amount requested by borrower in listing | Duration3 | 1 if the duration of the listing is 3 days. This is the baseline duration |
| TotalText | Total length of texts provided in borrower profile and listing descriptions | Duration5 | 1 if the duration of the listing is 5 days, 0 otherwise |
| ListingCat0 | 1 if the listing category information is unavailable; this is the baseline category and is not included in the estimation. | Duration7 | 1 if the duration of the listing is 7 days, 0 otherwise |

| Variable Name | Variable Description | Variable Name | Variable Description |
|---|---|---|---|
| ListingCat1 | 1 if the borrower chooses "Debt Consolidation" as the listing category | Duration10 | 1 if the duration of the listing is 10 days, 0 otherwise |
| ListingCat2 | 1 if the borrower chooses "Home Improvement Loans" as the listing category | BorrowerFee | Borrower closing fee charged by Prosper.com at the time of listing |
| ListingCat3 | 1 if the borrower chooses "Business Loans" as the listing category | LenderFee | Lender service fee charged by Prosper.com at the time of listing |
| ListingCat4 | 1 if the borrower chooses "Personal Loans" as the listing category | | |
| *Social Network Information - Groups* | | | |
| GroupSize | Number of members of the group where the borrower is a member. | _Medical | 1 if the borrower belongs to a group specifically mentioning helping with medical needs (e.g. medical costs financing); 0 otherwise |
| Grouplederrewarded | 1 if the borrower's group leader is awarded when loans are generated; 0 otherwise | _Demographic | 1 if the borrower belongs to a group targeting at particular demographic groups, such as Hispanics, Vietnamese, or single parents; 0 otherwise |
| _Alumni | 1 if the borrower belongs to an alumni group - groups targeting at alumni of universities or companies; 0 otherwise | _Hobbies | 1 if the borrower belongs to a group targeting at people with specific hobbies or careers; 0 otherwise |
| _Geography | 1 if the borrower belongs to a geographically-oriented group - groups targeting at members of certain geographical regions; 0 otherwise | _Religion | 1 if the borrower belongs to a religious group. 0 otherwise |
| _Military | 1 if the borrower belongs to a group targeting at military members or their families; 0 otherwise | _Business | 1 if the borrower belongs to a group specifically with the goal of helping small businesses or business developments; 0 otherwise |
| *Social Network Information - Friendship Network* | | | |
| ttlFriends | Total number of friends of the borrower. This is the simplest measure of degree centrality in the friendship network, regardless of their roles or actions | ttlPotentBid | Total number of borrower's potential lender friends who bid on the borrower's listing. This equals the difference between ttlPotentLend and ttlPotentNoBid |
| ttlRole | Total number of friends of the borrower who are either borrowers or lenders (i.e. have their identities verified) | ttlRealBid | Total number of borrower's real lender friends who bid on the borrower's listing |
| ttlNoRole | Total number of friends of the borrower who are neither borrowers nor lenders. This equals the difference between ttlFriends and ttlRole | ttlRealNoBid | Total number of borrower's real lender friends who did not bid on the borrower's listing. Equals the difference between ttlRealLend and ttlRealBid |
| ttlPureBorrow | Total number of borrower's friends who are borrowers but not lenders. | ttlRealBidWin | Total number of borrower's real lender friends who bid on the borrower's listing and win |

| Variable Name | Variable Description | Variable Name | Variable Description |
|---|---|---|---|
| ttlLend | Total number of borrower's friends who are lenders. This equals the difference between ttlRole and ttlPureBorrow | ttlRealBidLose | Total number of borrower's real lender friends who bid on the borrower's listing but lost |
| ttlRealLend | Total number of borrower's lender friends who are “real lenders”, or those who have already made loans prior to the time that the borrower (ego) posts the listing | ttlPotentBidWin | Total number of borrower's potential lender friends who bid on the borrower's listing and win |
| ttlPotentLend | Total number of borrower's lender friends who are “potential lenders”, or those who has not made any actual loans prior to the time that the borrower (ego) posts the listing. This equals the difference between ttlLend and ttlRealLend | ttlPotentBidLose | Total number of borrower's potential lender friends who bid on the borrower's listing and lost |
| ttlPotentNoBid | Total number of borrower's potential lender friends who did not bid on the borrower's listing | ttlPotentBid | Total number of borrower's potential lender friends who bid on the borrower's listing. This equals the difference between ttlPotentLend and ttlPotentNoBid |
| *Additional Control Variables* | | | |
| UsuryState | 1 if borrower resides in a state with usury laws; 0 otherwise | LenderRole | 1 if the borrower has a lender role; 0 otherwise |
| BankRate | The average interest rate on a 36-month consumer loan from a bank in the same market as the borrower, in the same month as the time of listing, and in the same credit grade of the borrower. | LeaderRole | 1 if the borrower is also a group leader; 0 otherwise |
| SpikeDays | 1 if there is abnormal search activities on Google for Prosper.com; 0 otherwise | | |

**Table 1.3: Estimated models**

| Model | Variable Set | | | | | |
|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** | **6** |
| Funding Probability | Spec. P1 | Spec. P2 | Spec. P3 | Spec. P4 | Spec. P5 | |
| Interest Rate | Spec. H1 | Spec. H2 | Spec. H3 | Spec. H4 | Spec. H5 | Spec. H6 |
| Loan Default | Spec. C1 | Spec. C2 | Spec. C3 | Spec. C4 | Spec. C5 | Spec. C6 |

* The sets of variables used in each model are described in Table 1.4.

**Table 1.4: Variable Sets used in the Models**

| Variable sets | Corresponding level of the friendship hierarchy | Common variables (see Table) | Additional variables |
|---|---|---|---|
| **1** | Root level | | ttlFriends |
| **2** | 1 | Hard credit information | ttlNoRole, ttlRole |
| **3** | 2 | Auction characteristics | ttlNoRole, ttlPureBorrow, ttlLend |
| **4** | 3 | Social network info – Groups; | ttlNoRole, ttlPureBorrow, ttlPotentLend, ttlRealLend |
| **5** | 4 | Additional control variables | ttlNoRole, ttlPureBorrow, ttlPotentLend, ttlRealNoBid, ttlRealBid |
| **6** | 5 | | ttlNoRole, ttlPureBorrow, ttlPotentLend, ttlRealNoBid, ttlRealBidWin, ttlRealBidLose |

## Table 1.5: Probability of Funding

The table reports estimates of a probit specification in which the dependent variable is one if a listing on prosper.com is funded and zero otherwise. The explanatory variables include a borrower's hard credit variables, social network variables, group affiliation, and other characteristics of the loan, the loan domicile, and the borrower plus quarterly time period fixed effects. Table 2 gives the detailed definitions of the variables. Robust standard errors are in parentheses. * p<0.1, ** p<0.05, *** p<0.01

| | Spec. P1 | Spec. P2 | Spec. P3 | Spec. P4 | Spec. P5 |
|---|---|---|---|---|---|
| ttlFriends | 0.033*** | | | | |
| | (0.008) | | | | |
| ttlNoRole | | -0.020* | -0.020** | -0.017* | -0.017 |
| | | (0.010) | (0.010) | (0.010) | (0.010) |
| ttlRole | | 0.106*** | | | |
| | | (0.017) | | | |
| ttlPureBorrow | | | -0.006 | -0.002 | 0.018 |
| | | | (0.023) | (0.021) | (0.018) |
| ttlLend | | | 0.170*** | | |
| | | | (0.023) | | |
| ttlPotentLend | | | | 0.025 | |
| | | | | (0.022) | |
| ttlRealLend | | | | 0.312*** | |
| | | | | (0.055) | |
| ttlPotentNobid | | | | | -0.062** |
| | | | | | (0.028) |
| ttlPotentBid | | | | | 0.325*** |
| | | | | | (0.050) |
| ttlRealBid | | | | | 0.849*** |
| | | | | | (0.044) |
| ttlrealnobid | | | | | -0.148*** |
| | | | | | (0.022) |
| ttlRealBidWin | | | | | |
| ttlRealBidLose | | | | | |
| bankrate | -0.698 | -0.752 | -0.662 | -0.662 | -0.558 |
| | (1.693) | (1.682) | (1.685) | (1.682) | (1.697) |
| borrowerFee | -5.722*** | -5.629*** | -5.732*** | -5.648*** | -5.672*** |
| | (1.775) | (1.737) | (1.763) | (1.743) | (1.682) |
| lenderFee | -29.932*** | -29.540*** | -29.387*** | -29.036*** | -30.203*** |
| | (4.583) | (4.508) | (4.512) | (4.514) | (4.592) |
| usurystate | -0.077** | -0.075** | -0.074** | -0.071** | -0.072** |
| | (0.034) | (0.034) | (0.034) | (0.033) | (0.034) |
| loggroupsize | -0.005** | -0.004** | -0.004* | -0.004** | -0.003 |
| | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) |
| leaderdummy | 0.054** | 0.047** | 0.048** | 0.043** | 0.073*** |
| | (0.022) | (0.022) | (0.022) | (0.022) | (0.022) |
| creditgrdA | -0.375*** | -0.372*** | -0.372*** | -0.373*** | -0.381*** |
| | (0.037) | (0.037) | (0.037) | (0.037) | (0.036) |
| creditgrdB | -0.806*** | -0.805*** | -0.805*** | -0.805*** | -0.814*** |
| | (0.062) | (0.061) | (0.061) | (0.061) | (0.063) |
| creditgrdC | -1.457*** | -1.455*** | -1.456*** | -1.457*** | -1.470*** |
| | (0.056) | (0.056) | (0.056) | (0.056) | (0.057) |
| creditgrdD | -2.105*** | -2.102*** | -2.107*** | -2.109*** | -2.133*** |
| | (0.094) | (0.094) | (0.093) | (0.094) | (0.094) |
| creditgrdE | -2.831*** | -2.825*** | -2.833*** | -2.837*** | -2.867*** |

| | | | | | |
|---|---|---|---|---|---|
| | (0.139) | (0.138) | (0.138) | (0.139) | (0.139) |
| creditgrdHR | -3.310*** | -3.304*** | -3.312*** | -3.318*** | -3.354*** |
| | (0.132) | (0.131) | (0.131) | (0.131) | (0.131) |
| bankcardutilization | 0.359*** | 0.356*** | 0.357*** | 0.355*** | 0.357*** |
| | (0.105) | (0.104) | (0.103) | (0.103) | (0.102) |
| bankcard2 | -0.203** | -0.201** | -0.200** | -0.200** | -0.199** |
| | (0.096) | (0.096) | (0.095) | (0.096) | (0.095) |
| inquirieslast6months | -0.020*** | -0.019*** | -0.019*** | -0.019*** | -0.019*** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| debttoincomeratio | -0.102*** | -0.102*** | -0.102*** | -0.103*** | -0.106*** |
| | (0.005) | (0.005) | (0.005) | (0.005) | (0.005) |
| listingcat4 | -0.164*** | -0.162*** | -0.164*** | -0.158*** | -0.162*** |
| | (0.033) | (0.034) | (0.034) | (0.033) | (0.033) |
| listingcat2 | 0.145*** | 0.146*** | 0.146*** | 0.147*** | 0.148*** |
| | (0.016) | (0.017) | (0.016) | (0.016) | (0.017) |
| listingcat1 | 0.176*** | 0.176*** | 0.175*** | 0.174*** | 0.168*** |
| | (0.024) | (0.024) | (0.024) | (0.024) | (0.025) |
| logamount | -0.706*** | -0.707*** | -0.708*** | -0.709*** | -0.714*** |
| | (0.012) | (0.012) | (0.012) | (0.012) | (0.012) |
| borrowermaximumrate | 24.308*** | 24.307*** | 24.406*** | 24.558*** | 24.844*** |
| | (1.099) | (1.095) | (1.093) | (1.094) | (1.092) |
| borrowermaxrate2 | -37.911*** | -37.904*** | -38.057*** | -38.357*** | -38.849*** |
| | (2.234) | (2.220) | (2.212) | (2.217) | (2.220) |
| auctionformat | 0.122*** | 0.125*** | 0.126*** | 0.127*** | 0.133*** |
| | (0.017) | (0.017) | (0.017) | (0.017) | (0.018) |
| grpleaderrewarded | 0.141*** | 0.145*** | 0.146*** | 0.145*** | 0.141*** |
| | (0.026) | (0.025) | (0.025) | (0.025) | (0.025) |
| _Religion | 0.280*** | 0.287*** | 0.283*** | 0.273*** | 0.246*** |
| | (0.089) | (0.088) | (0.087) | (0.085) | (0.084) |
| _Geography | 0.572** | 0.536** | 0.512** | 0.467*** | 0.411** |
| | (0.241) | (0.220) | (0.200) | (0.176) | (0.161) |
| _Alumni | 0.529*** | 0.526*** | 0.519*** | 0.519*** | 0.527*** |
| | (0.098) | (0.096) | (0.098) | (0.097) | (0.100) |
| logttltext | 0.225*** | 0.228*** | 0.227*** | 0.228*** | 0.222*** |
| | (0.009) | (0.009) | (0.009) | (0.009) | (0.009) |
| borrowerintermediary | 0.164*** | 0.142*** | 0.132*** | 0.137*** | 0.136*** |
| | (0.014) | (0.016) | (0.017) | (0.018) | (0.016) |
| _cons | 2.494*** | 2.473*** | 2.474*** | 2.458*** | 2.528*** |
| | (0.192) | (0.190) | (0.192) | (0.193) | (0.190) |
| **N** | **205131** | **205131** | **205131** | **205131** | **205131** |
| **pseudo R-sq** | **0.322** | **0.323** | **0.324** | **0.325** | **0.331** |

## Table 1.6: Interest Rate on Funded Listings

The table reports two-stage estimates of a model in which the dependent variable is the interest rate on Prosper.com listings that are successfully funded. The probit selection equation models the probability of a listing being successfully funded. The explanatory variables include a borrower's hard credit variables, social network variables, group affiliation, and other characteristics of the loan, the loan domicile, and the borrower plus quarterly time period fixed effects. I report all estimated coefficients for the interest rate equation but suppress coefficients for all probit variables that are included in Table 5. The coefficients for all suppressed variables in the selection equation are consistent with the probit model in Table 5. Robust standard errors are in parentheses.
* p<0.1; ** p<0.05; *** p<0.01

| | **Spec. H1** | **Spec. H2** | **Spec. H3** | **Spec. H4** | **Spec. H5** | **Spec. H6** |
|---|---|---|---|---|---|---|
| ttlFriends | -0.002*** | | | | | |
| | (0.001) | | | | | |
| ttlNoRole | | 0.002*** | 0.002*** | 0.002*** | 0.001*** | 0.001*** |
| | | (0.001) | (0.001) | (0.000) | (0.000) | (0.000) |
| ttlRole | | -0.005*** | | | | |
| | | (0.001) | | | | |
| ttlPureBorrow | | | 0.000 | -0.000 | -0.001 | -0.001 |
| | | | (0.001) | (0.001) | (0.001) | (0.001) |
| ttlLend | | | -0.006*** | | | |
| | | | (0.001) | | | |
| ttlPotentLend | | | | -0.001 | | |
| | | | | (0.001) | | |
| ttlRealLend | | | | -0.007*** | | |
| | | | | (0.001) | | |
| ttlPotentNobid | | | | | 0.002** | 0.002** |
| | | | | | (0.001) | (0.001) |
| ttlPotentBid | | | | | -0.008*** | -0.008*** |
| | | | | | (0.001) | (0.001) |
| ttlRealBid | | | | | -0.007*** | |
| | | | | | (0.001) | |
| ttlrealnobid | | | | | 0.002 | 0.001 |
| | | | | | (0.001) | (0.001) |
| ttlRealBidWin | | | | | | -0.006*** |
| | | | | | | (0.001) |
| ttlRealBidLose | | | | | | -0.006*** |
| | | | | | | (0.002) |
| bankrate | 0.104* | 0.106** | 0.102** | 0.102*** | 0.100*** | 0.099*** |
| | (0.053) | (0.048) | (0.042) | (0.032) | (0.027) | (0.027) |
| lenderservicing100 | 0.024*** | 0.021*** | 0.019*** | 0.015*** | 0.008*** | 0.008*** |
| | (0.003) | (0.003) | (0.003) | (0.002) | (0.001) | (0.001) |
| borrowerclosing100 | 0.003** | 0.002* | 0.002* | 0.001 | -0.001* | -0.001** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| usurystate | 0.005*** | 0.005*** | 0.004*** | 0.003*** | 0.002*** | 0.001*** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.000) | (0.000) |
| loggroupsize | 0.000 | -0.000 | -0.000 | -0.000 | -0.000* | -0.000* |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| leaderdummy | -0.005** | -0.004** | -0.004** | -0.003** | -0.003*** | -0.002*** |
| | (0.002) | (0.002) | (0.002) | (0.001) | (0.001) | (0.001) |
| creditgrdA | 0.026*** | 0.024*** | 0.021*** | 0.017*** | 0.009*** | 0.008*** |
| | (0.003) | (0.003) | (0.002) | (0.002) | (0.001) | (0.001) |
| creditgrdB | 0.058*** | 0.053*** | 0.048*** | 0.038*** | 0.020*** | 0.019*** |
| | (0.005) | (0.004) | (0.004) | (0.003) | (0.002) | (0.001) |
| creditgrdC | 0.103*** | 0.094*** | 0.085*** | 0.066*** | 0.034*** | 0.031*** |

| | | | | | | |
|---|---|---|---|---|---|---|
| | (0.008) | (0.007) | (0.006) | (0.004) | (0.002) | (0.002) |
| creditgrdD | 0.150*** | 0.138*** | 0.124*** | 0.097*** | 0.049*** | 0.045*** |
| | (0.011) | (0.010) | (0.008) | (0.006) | (0.003) | (0.003) |
| creditgrdE | 0.209*** | 0.192*** | 0.173*** | 0.137*** | 0.072*** | 0.066*** |
| | (0.015) | (0.014) | (0.012) | (0.008) | (0.005) | (0.004) |
| creditgrdHR | 0.244*** | 0.223*** | 0.201*** | 0.158*** | 0.082*** | 0.075*** |
| | (0.018) | (0.016) | (0.013) | (0.009) | (0.005) | (0.005) |
| bankcardutilization | -0.024*** | -0.021*** | -0.019*** | -0.014*** | -0.005*** | -0.005*** |
| | (0.003) | (0.003) | (0.002) | (0.002) | (0.001) | (0.001) |
| bankcard2 | 0.015*** | 0.013*** | 0.012*** | 0.009*** | 0.004*** | 0.003*** |
| | (0.002) | (0.002) | (0.002) | (0.001) | (0.001) | (0.001) |
| inquirieslast6months | 0.002*** | 0.001*** | 0.001*** | 0.001*** | 0.001*** | 0.000*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| debttoincomeratio | 0.007*** | 0.006*** | 0.006*** | 0.004*** | 0.002*** | 0.002*** |
| | (0.001) | (0.001) | (0.001) | (0.000) | (0.000) | (0.000) |
| listingcat4 | 0.010*** | 0.008*** | 0.007*** | 0.005*** | 0.001 | 0.001 |
| | (0.002) | (0.002) | (0.002) | (0.001) | (0.001) | (0.001) |
| listingcat2 | -0.009*** | -0.008*** | -0.007*** | -0.005*** | -0.002** | -0.001* |
| | (0.002) | (0.002) | (0.001) | (0.001) | (0.001) | (0.001) |
| listingcat1 | -0.016*** | -0.015*** | -0.013*** | -0.011*** | -0.007*** | -0.006*** |
| | (0.003) | (0.002) | (0.002) | (0.002) | (0.001) | (0.001) |
| logamount | 0.047*** | 0.043*** | 0.038*** | 0.030*** | 0.014*** | 0.012*** |
| | (0.004) | (0.003) | (0.003) | (0.002) | (0.001) | (0.001) |
| borrowermaximumrate | -1.008*** | -0.854*** | -0.688*** | -0.368*** | 0.206*** | 0.255*** |
| | (0.135) | (0.119) | (0.102) | (0.071) | (0.038) | (0.036) |
| borrowermaxrate2 | 2.749*** | 2.507*** | 2.246*** | 1.746*** | 0.846*** | 0.769*** |
| | (0.221) | (0.195) | (0.167) | (0.117) | (0.068) | (0.064) |
| auctionformat | 0.030*** | 0.031*** | 0.031*** | 0.033*** | 0.035*** | 0.036*** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| grpleaderrewarded | -0.005*** | -0.004*** | -0.003*** | -0.002* | 0.002** | 0.002** |
| | (0.002) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| _Religion | -0.020*** | -0.019*** | -0.017*** | -0.014*** | -0.007*** | -0.007*** |
| | (0.006) | (0.005) | (0.004) | (0.003) | (0.003) | (0.002) |
| _Geography | -0.031*** | -0.026*** | -0.021*** | -0.014*** | -0.002 | -0.001 |
| | (0.008) | (0.007) | (0.006) | (0.004) | (0.003) | (0.003) |
| _Alumni | -0.036*** | -0.033*** | -0.030*** | -0.024*** | -0.013*** | -0.012*** |
| | (0.006) | (0.006) | (0.005) | (0.004) | (0.003) | (0.002) |
| logttltext | -0.015*** | -0.014*** | -0.013*** | -0.010*** | -0.004*** | -0.004*** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.000) | (0.000) |
| borrowerintermediary | -0.011*** | -0.010*** | -0.008*** | -0.007*** | -0.004*** | -0.003*** |
| | (0.002) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Inverse Mills Ratio | -0.081*** | -0.073*** | -0.064*** | -0.047*** | -0.016*** | -0.013*** |
| | (0.007) | (0.006) | (0.005) | (0.003) | (0.002) | (0.002) |
| _cons | -0.083*** | -0.073*** | -0.065*** | -0.050*** | -0.027*** | -0.024*** |
| | (0.010) | (0.009) | (0.008) | (0.006) | (0.004) | (0.004) |
| *Selection Equation: All variables used but not reported for conciseness* | | | | | | |
| spikedays | -0.050** | -0.053*** | -0.053*** | -0.054*** | -0.052*** | -0.051** |
| | (0.020) | (0.020) | (0.020) | (0.020) | (0.020) | (0.020) |
| N | 205,132 | 205,132 | 205,132 | 205,132 | 205,132 | 205,132 |

## Table 1.7: Time to Default of Funded Loans

The table reports hazards ratio estimates of a Cox proportional hazards model of the time to default for borrower listings that are successfully funded on prosper.com. The explanatory variables include a borrower's hard credit variables, social network variables, group affiliation, and other characteristics of the loan, the loan domicile, and the borrower plus quarterly time period fixed effects. Table 2 gives the detailed definitions of the variables. The table reports the exponentiated coefficients (hazards ratio), where values greater than 1 suggest that a higher value of the explanatory variable increases the risk of default. Robust standard errors are in parentheses. * p<0.1, ** p<0.05, *** p<0.01

| | **Spec. C1** | **Spec. C2** | **Spec. C3** | **Spec. C4** | **Spec. C5** | **Spec. C6** |
|---|---|---|---|---|---|---|
| ttlFriends | 1.017 | | | | | |
| | (0.021) | | | | | |
| ttlNoRole | | 1.048 | 1.048 | 1.047 | 1.047 | 1.047 |
| | | (0.031) | (0.031) | (0.031) | (0.031) | (0.031) |
| ttlRole | | 0.968 | | | | |
| | | (0.027) | | | | |
| ttlPureBorrow | | | 1.061 | 1.061 | 1.058 | 1.055 |
| | | | (0.055) | (0.055) | (0.055) | (0.053) |
| ttlLend | | | 0.912** | | | |
| | | | (0.034) | | | |
| ttlPotentLend | | | | 0.950 | | |
| | | | | (0.061) | | |
| ttlRealLend | | | | 0.877*** | | |
| | | | | (0.044) | | |
| ttlPotentNobid | | | | | 0.964 | 0.964 |
| | | | | | (0.073) | (0.071) |
| ttlPotentBid | | | | | 0.910 | 0.916 |
| | | | | | (0.150) | (0.150) |
| ttlRealBid | | | | | 0.856** | |
| | | | | | (0.052) | |
| ttlrealnobid | | | | | 0.938 | 0.938 |
| | | | | | (0.113) | (0.113) |
| ttlRealBidWin | | | | | | 0.791*** |
| | | | | | | (0.062) |
| ttlRealBidLose | | | | | | 1.086 |
| | | | | | | (0.146) |
| bankrate100 | 0.958 | 0.958 | 0.958 | 0.958 | 0.958 | 0.957 |
| | (0.034) | (0.033) | (0.033) | (0.033) | (0.033) | (0.033) |
| usurystate | 1.102 | 1.100 | 1.098 | 1.098 | 1.098 | 1.100 |
| | (0.100) | (0.100) | (0.099) | (0.099) | (0.099) | (0.099) |
| loggroupsize | 1.005 | 1.005 | 1.004 | 1.004 | 1.004 | 1.004 |
| | (0.008) | (0.008) | (0.008) | (0.008) | (0.008) | (0.008) |
| leaderdummy | 1.043 | 1.050 | 1.051 | 1.051 | 1.049 | 1.048 |
| | (0.084) | (0.083) | (0.082) | (0.082) | (0.082) | (0.083) |
| bankcard100 | 0.996*** | 0.996*** | 0.996*** | 0.996*** | 0.996*** | 0.996*** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| bankcard2_100 | 1.000*** | 1.000*** | 1.000*** | 1.000*** | 1.000*** | 1.000*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| inquirieslast6months | 1.037*** | 1.037*** | 1.037*** | 1.037*** | 1.037*** | 1.037*** |
| | (0.006) | (0.006) | (0.006) | (0.006) | (0.006) | (0.006) |
| dti10 | 1.004*** | 1.004*** | 1.004*** | 1.004*** | 1.004*** | 1.004*** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| listingcat4 | 1.238* | 1.237* | 1.244* | 1.241* | 1.240* | 1.241* |
| | (0.145) | (0.145) | (0.147) | (0.146) | (0.146) | (0.146) |

| | | | | | | |
|---|---|---|---|---|---|---|
| listingcat2 | 0.991 | 0.989 | 0.992 | 0.992 | 0.992 | 0.994 |
| | (0.112) | (0.111) | (0.111) | (0.111) | (0.111) | (0.111) |
| listingcat1 | 1.102 | 1.101 | 1.102 | 1.103 | 1.102 | 1.102 |
| | (0.130) | (0.131) | (0.128) | (0.128) | (0.128) | (0.127) |
| logamount | 1.329*** | 1.332*** | 1.336*** | 1.337*** | 1.338*** | 1.339*** |
| | (0.058) | (0.058) | (0.059) | (0.059) | (0.059) | (0.059) |
| borrowerrate100 | 1.088*** | 1.088*** | 1.087*** | 1.087*** | 1.087*** | 1.087*** |
| | (0.008) | (0.008) | (0.007) | (0.007) | (0.007) | (0.007) |
| auctionformat | 1.073 | 1.070 | 1.069 | 1.069 | 1.069 | 1.072 |
| | (0.074) | (0.074) | (0.074) | (0.074) | (0.074) | (0.075) |
| grpleaderrewarded | 1.152*** | 1.151*** | 1.152*** | 1.153*** | 1.153*** | 1.148*** |
| | (0.047) | (0.047) | (0.047) | (0.047) | (0.047) | (0.046) |
| _Religion | 0.758 | 0.760 | 0.766 | 0.768 | 0.769 | 0.767 |
| | (0.180) | (0.181) | (0.182) | (0.182) | (0.182) | (0.185) |
| _Geography | 0.404** | 0.416** | 0.426** | 0.430** | 0.430** | 0.433** |
| | (0.166) | (0.166) | (0.170) | (0.171) | (0.170) | (0.170) |
| _Alumni | 0.406*** | 0.407*** | 0.408*** | 0.409*** | 0.408*** | 0.409*** |
| | (0.120) | (0.120) | (0.122) | (0.123) | (0.122) | (0.123) |
| logttltext | 0.990 | 0.989 | 0.990 | 0.990 | 0.991 | 0.991 |
| | (0.039) | (0.039) | (0.039) | (0.039) | (0.039) | (0.039) |
| borrowerintermediary | 0.768*** | 0.778*** | 0.785*** | 0.784*** | 0.783*** | 0.780*** |
| | (0.041) | (0.043) | (0.042) | (0.043) | (0.042) | (0.040) |
| creditgrdA | 1.696*** | 1.692*** | 1.692*** | 1.692*** | 1.694*** | 1.691*** |
| | (0.204) | (0.204) | (0.205) | (0.205) | (0.205) | (0.205) |
| creditgrdB | 2.009*** | 2.009*** | 2.007*** | 2.005*** | 2.009*** | 2.008*** |
| | (0.233) | (0.233) | (0.233) | (0.233) | (0.235) | (0.235) |
| creditgrdC | 2.329*** | 2.330*** | 2.333*** | 2.334*** | 2.341*** | 2.342*** |
| | (0.333) | (0.336) | (0.337) | (0.337) | (0.338) | (0.338) |
| creditgrdD | 2.461*** | 2.463*** | 2.476*** | 2.479*** | 2.489*** | 2.496*** |
| | (0.466) | (0.465) | (0.468) | (0.467) | (0.470) | (0.472) |
| creditgrdE | 3.055*** | 3.064*** | 3.097*** | 3.106*** | 3.122*** | 3.139*** |
| | (0.896) | (0.895) | (0.902) | (0.903) | (0.908) | (0.911) |
| creditgrdHR | 4.550*** | 4.567*** | 4.638*** | 4.655*** | 4.685*** | 4.710*** |
| | (1.394) | (1.393) | (1.408) | (1.408) | (1.421) | (1.427) |

# APPENDICES FOR ESSAY 2

### Table 2.1: Evidence of Home Bias – lending amount (overall market)

*This table presents "macro" evidence of investors' home bias in all lending activities on Prosper.com up until April, 2009. Home bias exists for virtually all states: the percentage of lending amount from home-state lenders (Column 3) exceeds the share of home-state lenders' investment in the entire marketplace (Column 2)* [16].

| State | Total investment of lenders in this state / total investment from all states | Funding amount from home-state lenders / total amount to borrowers in this state |
|---|---|---|
| AK | 0.37% | 1.20% |
| AL | 0.47% | 0.60% |
| AR | 0.40% | 2.40% |
| AZ | 1.75% | 2.20% |
| CA | 22.39% | 24.11% |
| CO | 2.01% | 2.54% |
| CT | 1.10% | 2.03% |
| DC | 0.48% | 0.52% |
| DE | 0.26% | 3.36% |
| FL | 5.53% | 6.37% |
| GA | 2.45% | 3.17% |
| HI | 0.61% | 8.32% |
| IA | 0.50% | 2.11% |
| ID | 0.39% | 1.17% |
| IL | 5.75% | 6.81% |
| IN | 0.78% | 1.28% |
| KS | 0.53% | 3.66% |
| KY | 0.45% | 2.53% |
| LA | 0.68% | 2.18% |
| MA | 2.69% | 3.90% |
| MD | 2.98% | 4.08% |
| ME | 0.15% | 0.44% |
| MI | 1.94% | 2.59% |
| MN | 1.37% | 2.28% |
| MO | 0.84% | 1.55% |
| MS | 0.19% | 0.71% |
| MT | 0.24% | 0.99% |
| NC | 2.17% | 2.90% |
| ND | 0.09% | 0.20% |
| NE | 0.39% | 1.93% |

[16] The only state not reported here is SD, where there is no loans made to that state recorded in the Prosper database as of April 2009. Lenders from SD account for 0.14% of all loaned amount on Prosper.com.

| | | |
|---|---|---|
| NH | 0.66% | 1.13% |
| NJ | 3.24% | 4.40% |
| NM | 0.67% | 1.09% |
| NV | 1.11% | 2.10% |
| NY | 7.08% | 8.96% |
| OH | 1.80% | 2.47% |
| OK | 0.53% | 1.18% |
| OR | 1.55% | 2.56% |
| PA | 2.81% | 4.88% |
| RI | 0.17% | 0.17% |
| SC | 0.60% | 1.02% |
| TN | 0.78% | 1.57% |
| TX | 8.24% | 10.82% |
| UT | 1.11% | 2.52% |
| VA | 5.00% | 6.51% |
| VT | 0.13% | 3.95% |
| WA | 3.55% | 4.66% |
| WI | 1.02% | 1.72% |
| WV | 0.17% | 0.38% |
| WY | 0.07% | 0.78% |

## Table 2.2: Evidence of Home Bias – number of bids (overall market)

*Same comparison as Table 1a, except using the number of bids instead of the amount – for the top borrowing states*

| State | share of bid count of the state's lenders in the entire dataset | share of bid count from same-state lenders |
|---|---|---|
| CA | 20.40% | 21.10% |
| GA | 2.66% | 2.90% |
| FL | 5.61% | 5.90% |
| IL | 5.07% | 5.40% |
| TX | 7.56% | 8.30% |
| NY | 7.18% | 7.90% |
| WA | 3.85% | 4.30% |
| MI | 2.05% | 2.40% |
| MD | 3.08% | 3.30% |
| OH | 1.98% | 2.20% |
| AZ | 1.95% | 2.20% |
| NC | 2.40% | 2.60% |
| MO | 0.97% | 1.40% |
| OR | 1.60% | 1.90% |
| MN | 1.60% | 1.90% |

## Table 2.3: Main results

*For easier interpretations, I report odds ratio in the table instead of coefficients. An odds ratio greater than 1 means the variable has a positive effect on the probability of occurrence. Dependent variable: 1 if a bid is placed, 0 otherwise.*

| | **All Borrowers** | **Creditworthy borrowers (AA and A)** | **Less creditworthy borrowers (B, C and D)** | **Economic vs. Geographic distance** |
|---|---|---|---|---|
| Dummy for Same-State borrowers | 1.130* | 1.698*** | 0.492*** | |
| | (0.058) | (0.097) | (0.068) | |
| Geographic distance (borrower state to CA) | | | | 0.988 |
| | | | | (0.007) |
| Economic distance (Borrower state to CA) | | | | 0.582*** |
| | | | | (0.078) |
| Loan amount requested by borrower | 0.592*** | 0.793*** | 0.333*** | 0.586*** |
| | (0.013) | (0.022) | (0.013) | (0.013) |
| Borrower number of requests | 1.170*** | 1.048 | 1.343*** | 1.169*** |
| | (0.021) | (0.027) | (0.034) | (0.021) |
| Borrower asking interest rate | 0.087*** | 0.040*** | 0.007*** | 0.100*** |
| | (0.019) | (0.012) | (0.002) | (0.021) |
| Borrower auction format | 0.232*** | 0.033*** | 0.394*** | 0.232*** |
| | (0.044) | (0.020) | (0.077) | (0.044) |
| Borrower associated with group | 0.673*** | 0.645*** | 0.670*** | 0.663*** |
| | (0.052) | (0.069) | (0.070) | (0.051) |
| Bidder experience (log of # months on Prosper) | 1.624*** | 2.055*** | 1.151 | 1.624*** |
| | (0.077) | (0.128) | (0.083) | (0.077) |
| Bidder is a trader (re-sells loans) | 1.192*** | 1.274*** | 1.071 | 1.192*** |
| | (0.057) | (0.076) | (0.083) | (0.057) |
| Bidder is a group leader | 1.178* | 0.869 | 1.732*** | 1.178* |
| | (0.079) | (0.081) | (0.169) | (0.079) |
| Bidder is a borrower | 0.650*** | 0.569*** | 0.788** | 0.650*** |
| | (0.032) | (0.037) | (0.060) | (0.032) |
| Bidder # of friends | 1.003 | 1.003 | 1.003 | 1.003 |
| | (0.002) | (0.003) | (0.003) | (0.002) |
| Dummy for grade A borrowers | 2.922*** | | | 2.894*** |
| | (0.250) | | | (0.249) |
| Dummy for AA borrowers | 4.215*** | | | 4.185*** |
| | (0.373) | | | (0.373) |
| Dummy for B borrowers | 1.469*** | | | 1.463*** |
| | (0.129) | | | (0.129) |
| Dummy for C borrowers | 1.108 | | | 1.083 |
| | (0.099) | | | (0.098) |
| Loan purpose dummy: debt consolidation | 0.885** | 0.676*** | 1.374*** | 0.891** |
| | (0.040) | (0.040) | (0.107) | (0.040) |
| Loan purpose dummy: home improvement | 0.741*** | 0.700*** | 0.674** | 0.748*** |
| | (0.049) | (0.054) | (0.092) | (0.049) |
| Loan purpose dummy: Business loan | 0.428*** | 0.310*** | 0.816 | 0.434*** |
| | (0.033) | (0.030) | (0.107) | (0.033) |
| Loan purpose dummy: Student loan | 0.513*** | 0.078*** | 0.956 | 0.509*** |
| | (0.053) | (0.026) | (0.126) | (0.053) |
| **N** | **358832** | **110208** | **248624** | **358832** |
| **Chi-squared** | **5165.160** | **1520.606** | **1385.802** | **5586.844** |

Odds ratios reported; heteroskedasticity-consistent standard errors in parentheses. (* p<0.05 ** p<0.01 *** p<0.001)

# APPENDICES FOR ESSAY 3

## Table 3.1: Full model with interactions

Dependent variable is whether a bid was successfully chosen by the buyer. Modeled with a logistic regression, with standard errors estimated using clustered sandwich estimators to allow for intra-auction correlation. Odds ratio (exponentiated coefficients) reported as they are easier to interpret for binary variables; standard errors in parentheses. An odds ratio greater than 1 suggests that a higher value of the explanatory variable is positively associated with the probability of winning. Some variables, including dummies for project amount range suppressed for brevity. * p<0.1; ** p<0.05; *** p<0.001

| **Variable** | **Odds Ratio** |
|---|---|
| ExpertCertified | 1.195 |
| | (0.768) |
| noRating | 0.339*** |
| | (0.093) |
| noCommentBid | 0.899 |
| | (0.250) |
| BuyerSellerSameCountry | 2.036*** |
| | (0.391) |
| logBidAmount | 0.636*** |
| | (0.038) |
| logBidOrder | 0.628*** |
| | (0.029) |
| logSellerMonths | 1.592*** |
| | (0.132) |
| logExpertiseLength | 1.038 |
| | (0.043) |
| PFT | 2.846*** |
| | (0.363) |
| PFT*Expert | 1.025 |
| | (0.683) |
| PFT*Rating | 2.380*** |
| | (0.793) |
| Intercept | 0.116*** |
| | (0.049) |
| **N (number of bids)** | **5670** |

**Table 3.2: Full model with interaction terms: Number and Average of Ratings**

Dependent variable is whether a bid was successfully chosen by the buyer. Modeled with a logistic regression, with standard errors estimated using clustered sandwich estimators to allow for intra-auction correlation. Odds ratio (exponentiated coefficients) reported as they are easier to interpret for binary variables; robust standard errors are in parentheses. An odds ratio greater than 1 suggests that a higher value of the explanatory variable is positively associated with the probability of winning. Some variables, including dummies for project amount range suppressed for brevity. * p<0.1; ** p<0.05; *** p<0.001

| **Variables** | **Odds Ratio** |
|---|---|
| ExpertCertified | 1.054 |
| | (0.691) |
| logRatingsCount | 1.290*** |
| | (0.091) |
| AvgRating | 1.242* |
| | (0.151) |
| BuyerSellerSameCountry | 2.036*** |
| | (0.459) |
| logBidAmount | 0.632*** |
| | (0.043) |
| logBidOrder | 0.650*** |
| | (0.035) |
| logCoderMonths | 1.422*** |
| | (0.157) |
| PFT | 10.422 |
| | (15.429) |
| Certification X PFT | 1.059 |
| | (0.731) |
| noComment X PFT | 1.720 |
| | (0.780) |
| logAvgRating X PFT | 0.941 |
| | (0.146) |
| logRateCount X PFT | 0.832* |
| | (0.082) |
| **N** | **3581** |

### Table 3.3: Effect of Ratings on PFD and PFT samples

Dependent variable is whether a bid was successfully chosen by the buyer. Modeled with a logistic regression, with standard errors estimated using clustered sandwich estimators to allow for intra-auction correlation. Odds ratio (exponentiated coefficients) reported as they are easier to interpret for binary variables; standard errors in parentheses. An odds ratio greater than 1 suggests that a higher value of the explanatory variable is positively associated with the probability of winning. Some control variables, including dummies for project amount range suppressed for brevity. * p<0.1; ** p<0.05; *** p<0.001

| **Variable** | **Odds Ratio under Pay-for-Delivery Contracts** | **Odds Ratio under Pay-for-Time Contracts** |
|---|---|---|
| ExpertCertified | 1.017 | 1.170 |
| | (0.651) | (0.256) |
| logRatingsCount | 1.229*** | 1.094 |
| | (0.095) | (0.097) |
| AvgRating | 1.255* | 1.166 |
| | (0.154) | (0.117) |
| BuyerSellerSameCountry | 1.395 | 3.378*** |
| | (0.470) | (1.103) |
| logBidAmount | 0.665*** | 0.553*** |
| | (0.061) | (0.058) |
| logBidOrder | 0.725*** | 0.553*** |
| | (0.050) | (0.052) |
| logSellerMonths | 1.616** | 1.354** |
| | (0.341) | (0.184) |
| logExpertiseLength | 1.093 | 0.975 |
| | (0.068) | (0.074) |
| Intercept | 0.004*** | 0.400 |
| | (0.005) | (0.449) |
| **N (number of bids)** | **2607** | **974** |

## Table 3.4: Effect of Communication under Pay-for-Deliverable vs. Pay-for-Time Contracts

Dependent variable is whether a bid was successfully chosen by the buyer. Modeled with a logistic regression, with standard errors estimated using clustered sandwich estimators to allow for intra-auction correlation. Odds ratio (exponentiated coefficients) reported as they are easier to interpret for binary variables; robust standard errors are in parentheses. An odds ratio greater than 1 suggests that a higher value of the explanatory variable is positively associated with the probability of winning. Some variables, including dummies for project amount range suppressed for brevity. * $p<0.1$; ** $p<0.05$; *** $p<0.001$

| Variables | Odds Ratio Under Pay-for-Deliverable Contracts | Odds Ratio Under Pay-for-Time Contracts |
|---|---|---|
| ExpertCertified | 0.988 | 1.300 |
| | (0.658) | (0.260) |
| noRating | 0.450*** | 0.708 |
| | (0.126) | (0.161) |
| BuyerSellerSameCountry | 1.861** | 2.538*** |
| | (0.506) | (0.811) |
| logBidAmount | 0.632*** | 0.591*** |
| | (0.063) | (0.062) |
| logBidOrder | 0.707*** | 0.545*** |
| | (0.047) | (0.042) |
| logSellerMonths | 2.152*** | 1.340*** |
| | (0.350) | (0.138) |
| logExpertiseLength | 1.099* | 1.033 |
| | (0.060) | (0.072) |
| Sixltr | 0.969*** | 0.987 |
| | (0.011) | (0.012) |
| we | 0.972 | 0.994 |
| | (0.027) | (0.039) |
| auxverb | 0.959*** | 0.971 |
| | (0.014) | (0.021) |
| time | 0.981 | 1.049** |
| | (0.017) | (0.020) |
| money | 1.005 | 0.897** |
| | (0.030) | (0.049) |
| noTypo | 0.745 | 0.843 |
| | (0.190) | (0.230) |
| Intercept | 0.080*** | 1.348 |
| | (0.062) | (0.965) |
| **N (number of bids)** | **3653** | **1490** |

# REFERENCES


Agarwal, S., & Hauswald, R. B. 2007. The Choice between Arm's-Length and Relationship Debt: Evidence from Eloans: SSRN.

Ahearne, A., Griever, W., & Warnock, F. 2004. Information costs and home bias: an analysis of US holdings of foreign equities. ***Journal of International Economics***, 62(2): 313-336.

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

Angrist, J. 1998. Estimating the labor market impact of voluntary military service using social security data on military applicants. ***Econometrica***, 66(2): 249-288.

Asvanund, A., Clay, K., Krishnan, R., & Smith, M. D. 2004. An Empirical Analysis of Network Externalities in Peer-to-Peer Music-Sharing Networks, ***Information Systems Research***, Vol. 15: 155-174: INFORMS: Institute for Operations Research.

Autor, D. H. 2001. Wiring the Labor Market. ***Journal of Economic Perspectives***, 15(1): 25-40.

Bagozzi, R. P., & Dholakia, U. M. 2006. Open Source Software User Communities: A Study of Participation in Linux User Groups, ***Management Science***, Vol. 52: 1099-1115: INFORMS: Institute for Operations Research.

Bajari, P., & Tadelis, S. 2001. Incentives versus transaction costs: a theory of procurement contracts. ***RAND Journal of Economics***, 32(3): 387-407.

Bampo, M., Ewing, M. T., Mather, D. R., Stewart, D., & Wallace, M. 2008. The Effects of the Social Structure of Digital Networks on Viral Marketing Performance, ***Information Systems Research***, Vol. 19: 273-290.

Banerjee, A. V., & Duflo, E. 2000. Reputation Effects and the Limits of Contracting: A Study of the Indian Software Industry. ***Quarterly Journal of Economics***, 115(3): 989-1017.

Bergstrand, J. 1985. The gravity equation in international trade: some microeconomic foundations and empirical evidence. ***The Review of Economics and Statistics***: 474-481.

Bhattacharjee, S., Gopal, R. D., Lertwachara, K., Marsden, J. R., & Telang, R. 2007. The Effect of Digital Sharing Technologies on Music Markets: A Survival Analysis of Albums on Ranking Charts, ***Management Science***, Vol. 53: 1359-1374.

Boulhol, H., & De Serres, A. 2010. Have developed countries escaped the curse of distance? ***Journal of Economic Geography***, 10(1): 113-139.

Brynjolfsson, E., & Smith, M. D. 2000. Frictionless Commerce? ***Management Science***.

Burt, R. 2005. ***Brokerage and Closure: An Introduction to Social Capital***: Oxford University Press.

Burt, R. S. 1992. ***Structural Holes***. Cambridge, MA: Westview Press.

Carey, M., & Nini, G. 2007. Is the Corporate Loan Market Globally Integrated? A Pricing Puzzle. ***The Journal of Finance***, 62(6): 2969-3007.

Chen, Y., & Bharadwaj, A. 2009. An Empirical Analysis of Contract Structures in IT Outsourcing. ***Information Systems Research***, 20(4): 484-506.

Cleves, M., Gould, W., Gutierrez, R., & Marchenko, Y. 2008. ***An Introduction to Survival Analysis Using Stata***: Stata Press.

Cohen-Cole, E., & Duygan-Bump, B. 2008. Household Bankruptcy Decision: the role of social stigma vs. information sharing, ***working paper***: Federal Reserve Bank of Boston.

Cohen, L., Frazzini, A., & Malloy, C. 2008. The Small World of Investing: Board Connections and Mutual Fund Returns. ***Journal of Political Economy***, 116(5): 951-979.

Cohen, L., Frazzini, A., & Malloy, C. 2010. Sell-side school ties. ***The Journal of Finance***, 65(4): 1409-1437.

Cohn, M., Mehl, M., & Pennebaker, J. 2004. Linguistic markers of psychological change surrounding September 11, 2001. ***Psychological Science***, 15(10): 687.

Coleman, J. S. 1988. Social Capital in the Creation of Human Capital. ***American Journal of Sociology***, 94(s1): S95.
Cooper, I., & Kaplanis, E. 1994. Home bias in equity portfolios, inflation hedging, and international capital market equilibrium. ***Review of Financial Studies***, 7(1): 45-60.
Cowan, R., Jonard, N., & Zimmermann, J.-B. 2007. Bilateral Collaboration and the Emergence of Innovation Networks, ***Management Science***, Vol. 53: 1051-1067.
Dey, D., Ming, F., & Conglei, Z. 2010. Design and Analysis of Contracts for Software Outsourcing. ***Information Systems Research***, 21(1): 93-114.
DiPrete, T. A., Eirich, G. M., Cook, K. S., & Massey, D. S. 2006. Cumulative Advantage as a Mechanism for Inequality: A Review of Theoretical and Empirical Developments. ***Annual review of sociology***, 32(1): 271-297.
Disdier, A., & Head, K. 2008. The puzzling persistence of the distance effect on bilateral trade. ***Review of Economics and Statistics***, 90(1): 37-48.
Drucker, S., & Puri, M. 2007. Banks in capital markets. In B. E. Eckbo (Ed.), ***Handbook of corporate finance: empirical corporate finance***, Vol. 1: North Holland.
Durlauf, S., Fafchamps, M., & Aghion, P. 2004. Social Capital. In S. Durlauf, & P. Aghion (Eds.), ***Handbook of Economic Growth***.
Fama, E. 1985. What’s different about banks. ***Journal of Monetary Economics***, 15: 29-39.
Friedman, R., Brett, J., Anderson, C., Olekalns, M., Goates, N., & Lisco, C. 2004. The positive and negative effects of anger on dispute resolution: Evidence from electronically mediated disputes. ***Journal of Applied Psychology***, 89(2): 369-376.
Gary, B. 1971. ***The economics of discrimination***. Chicago, IL: University of Chicago Press.
Gopal, A., & Gao, G. 2009. Certification in the Indian Offshore IT Services Industry. ***Manufacturing & Service Operations Management***, 11(3): 471-492.
Gopal, A., Sivaramakrishnan, K., Krishnan, M. S., & Mukhopadhyay, T. 2003. Contracts in Offshore Software Development: An Empirical Analysis. ***Management Science***, 49(12): 1671-1683.
Graham, J. R., Harvey, C. R., & Huang, H. 2009. Investor Competence, Trading Frequency, and Home Bias. ***Management Science***, 55(7): 1094-1106.
Granovetter, M. 1985. Economic Action and Social Structure: The Problem of Embeddedness. ***American Journal of Sociology***, 91(3): 481.
Granovetter, M. 1992. Problems of explanation in economic sociology. ***Networks and Organizations***: 25-56.
Granovetter, M. 2005. The Impact of Social Structure on Economic Outcomes. ***Journal of Economic Perspectives***, 19(1): 33-50.
Granovetter, M. S. 1972. The Strength of Weak Ties. ***American Journal of Sociology***, 78(6): 1360.
Greene, W. 2002. ***Econometric analysis***: prentice Hall Upper Saddle River, NJ.
Grewal, R., Lilien, G. L., & Mallapragada, G. 2006. Location, Location, Location: How Network Embeddedness Affects Project Success in Open Source Systems, ***Management Science***, Vol. 52: 1043-1056: INFORMS: Institute for Operations Research.
Gross, D. B., & Souleles, N. 2002. An Empirical Analysis of Personal Bankruptcy and Delinquency. ***Review of Financial Studies***.
Grover, V., Fiedler, K., & Teng, J. 1997. Empirical Evidence on Swanson's Tri-Core Model of Information Systems Innovation. ***Information Systems Research***, 8(3): 273.
Guiso, G., Sapienza, P., & Zingales, L. 2004. Role of Social Capital in Financial Development. ***American Economic Review***, 94(3): 31.
Hanneman, R., & Riddle, M. 2005. ***Introduction to social network methods***. Riverside, CA: University of California, Riverside.
Hardy, C., Lawrence, T. B., & Grant, D. 2005. Discourse and Collaboration: The Role of Conversations and Collective Identity. ***Academy of Management Review***, 30(1): 58-77.

Hasija, S., Pinker, E. J., & Shumsky, R. A. 2008. Call Center Outsourcing Contracts Under Information Asymmetry. ***Management Science***, 54(4): 793-807.
Heckman, J. 1979. Sample Selection Bias as a Specification Error. ***Econometrica***, 47(1): 153-161.
Hillberry, R., & Hummels, D. 2003. Intranational Home Bias: Some Explanations. ***Review of Economics & Statistics***, 85(4): 1089-1092.
Hills, T. T., Maouene, M., Maouene, J., Sheya, A., & Smith, L. 2009. Longitudinal Analysis of Early Semantic Networks: Preferential Attachment or Preferential Acquisition? ***Psychological Science***, 20(6): 729-739.
Hochberg, Y. V. 2009. Individual lenders infer borrower credit worthiness in peer-to-peer lending? Working paper, Kennedy School, Harvard University.
Hochberg, Y. V., Ljungqvist, A., & Lu, Y. 2007. Whom You Know Matters: Venture Capital Networks and Investment Performance. ***The Journal of Finance***.
Hortacsu, A., Martinez-Jerez, F. A., & Douglas, J. 2009. The Geography of Trade in Online Transactions: Evidence from eBay and MercadoLibre. ***American Economic Journal: Microeconomics***, 1(1): 53-74.
Hsu, D. 2004. What Do Entrepreneurs Pay for Venture Capital Affiliation? ***The Journal of Finance***, 59(4): 1805-1844.
Jensen, M., & Roy, A. 2008. Staging Exchange Partner Choices: When do Status and Reputation Matter? . ***Academy of Management Journal***, 51(3): 495-516.
Kalbfleisch, J., & Prentice, R. 2002. ***The statistical analysis of failure time data***: Wiley-Interscience.
Karlan, D. 2007. Social connections and group banking. ***The Economic Journal***, 117(517): 52-84.
Karlsson, A., & Nordén, L. 2007. Home sweet home: Home bias and international diversification among individual investors. ***Journal of Banking & Finance***, 31(2): 317-333.
Kilker, J. 1999. Conflict on Collaborative Design Teams: Understanding the Role of Social Identities. ***IEEE Technology & Society Magazine***, 18(3): 12-12.
King, G., & Zeng, L. 1999. Logistic regression in rare-events data. ***Department of Government, Harvard University***, http://GKing.harvard.edu.
King, G., & Zeng, L. 2001. Logistic regression in rare events data. ***Political analysis***, 9(2): 137.
Kirsch, D., Goldfarb, B., & Gera, A. 2009. Form or substance: The role of business plans in venture capital decision making. ***Strategic Management Journal***, 30(5): 487-515.
Kirsch, L. J. 1997. Portfolios of Control Modes and IS Project Management. ***Information Systems Research***, 8(3): 215.
Koh, C., Soon, A., & Straub, D. W. 2004. IT Outsourcing Success: A Psychological Contract Perspective. ***Information Systems Research***, 15(4): 356-373.
Kremer, M., & Levy, D. 2008. Peer effects and alcohol use among college students. ***The Journal of Economic Perspectives***, 22(3): 189-183A.
Krishnan, R., Smith, M., Tang, Z., & Telang, R. 2007. Digital business models for peer-to-peer networks: Analysis and economic issue. ***Review of Network Economics***, 6(2): 194–213.
Lai, S., & Teo, M. 2008. Home-Biased Analysts in Emerging Markets. ***Journal of Financial & Quantitative Analysis***, 43(3): 685-716.
Ledgerwood, J. 1999. ***Microfinance handbook***: The World Bank.
Levina, N. 2006. ***Collaborating Across Boundaries in a Global Economy: Do Organizational Boundaries and Country Contexts Matter?*** . Paper presented at the 26th International Conference on Information Systems, Milwaukee, WI.
Levina, N., & Ross, J. W. 2003. From the Vendor's Perspective: Exploring the Value Proposition in Information Technology Outsourcing. ***MIS Quarterly***, 27(3): 331-364.
Levy, H., & Sarnat, M. 1970. Inernational diversification of investment portfolios. ***American Economic Review***, 60(4): 668-675.

Lewis, K. 1999. Trying to explain home bias in equities and consumption. ***Journal of Economic Literature***, 37(2): 571-608.
Lewis, T. R. 1986. Reputation and contractual performance in long-term projects. ***RAND Journal of Economics***, 17(2): 141-157.
Lin, M., Prabhala, N. R., & Viswanathan, S. 2009. Can social networks help mitigate information asymmetry in online markets? . ***Proceedings of the 30th International Conference on Information Systems***.
Malone, T. W., Yates, J., & Benjamin, R. I. 1987. Electronic Markets and Electronic Hierarchies. ***Communications of the ACM***, 30(6): 484-497.
McPherson, M., Smith-Lovin, L., & Cook, J. 2001. Birds of a feather: Homophily in social networks. ***Annual review of sociology***, 27(1): 415-444.
Merton, R. K. 1968. The Matthew Effect in Science. ***Science***, 159(3810): 56-63.
Mizruchi, M. S. 1992. ***The Structure of Corporate Political Action: Inter-firm Relations and their Consequences***: Harvard University Press.
Moran, P. 2005. Structural vs. Relational Embeddedness: Social Capital and Managerial Performance. ***Strategic Management Journal***, 26(12): 1129.
Morduch, J. 1999. The Microfinance Promise. ***Journal of Economic Literature***.
Oestreicher-Singer, G., & Sundararajan, A. The visible hand of social networks in electronic markets. ***http://ssrn.com/abstract=1268516***.
Oestreicher-Singer, G., & Sundararajan, A. 2009. Recommendation Networks and the Long Tail of Electronic Commerce. ***http://ssrn.com/abstract=1324064***.
Overman, H. G., Redding, S. J., & Venables, A. J. 2003. ***The Economic Geography of Trade, Production, and Income: A Survey of Empirics***. Oxford: Blackwell.
Parwada, J. T. 2008. The Genesis of Home Bias? The Location and Portfolio Choices of Investment Company Start-Ups. ***Journal of Financial & Quantitative Analysis***, 43(1): 245-266.
Petersen, M. A. 2002. Information: Hard and Soft, ***Northwestern University, Chicago IL***.
Petersen, M. A., & Rajan, R. G. 1994. The Benefits of Lending Relationships: Evidence from Small Business Data. ***The Journal of Finance***, 49(1): 3-37.
Petersen, M. A., & Rajan, R. G. 2002. Does Distance Still Matter? The Information Revolution in Small Business Lending. ***Journal of Finance***, 57(6): 2533-2570.
Pettersson-Lidbom, P., & Priks, M. 2010. Behavior under social pressure: Empty Italian stadiums and referee bias. ***Economics Letters***, 108(2): 212-214.
Podolny, J. 1993. A status-based model of market competition. ***American Journal of Sociology***, 98(4): 829-872.
Podolny, J. 2001. Networks as the pipes and prisms of the market. ***American Journal of Sociology***, 107(1): 33-60.
Pope, D. G., & Sydnor, J. R. 2008. What's in a Picture? Evidence of Discrimination from Prosper.com. ***http://ssrn.com/abstract=1220902***.
Portes, A. 1998. Social Capital: Its Origins and Applications in Modern Sociology. ***Annual Review of Sociology***, 24: 1-24.
Putnam, R. 1993. The Prosperous Community: Social Capital and Economic Growth. ***American Prospect***, 4(13): 35–42.
Ravina, E. Love & Loans: The Effect of Beauty and Personal Characteristics in Credit Markets. ***http://ssrn.com/abstract=1107307***.
Reagans, R. 2005. Preferences, Identity, and Competition: Predicting Tie Strength from Demographic Data. ***Management Science***, 51(9): 1374-1383.
Redding, S. J., & Sturm, D. M. 2008. The Costs of Remoteness: Evidence from German Division and Reunification. ***American Economic Review***, 98(5): 1766-1797.
Robert, J. L. P., Dennis, A. R., & Ahuja, M. K. 2008. Social Capital and Knowledge Integration in Digitally Enabled Teams, ***Information Systems Research***, Vol. 19: 314-334.

Rosenthal, P. I. 1971. Specificity, Verifiability, and Message Credibility. ***Quarterly Journal of Speech***, 57: 393-401.
Rustagi, S., King, W. R., & Kirsch, L. J. 2008. Predictors of Formal Control Usage in IT Outsourcing Partnerships. ***Information Systems Research***, 19(2): 126-143.
Sapienza, P., Toldra, A., & Zingales, L. 2007. Understanding Trust. ***NBER WORKING PAPER SERIES***, 13387.
Shane, S., & Cable, D. 2002. Network Ties, Reputation, and the Financing of New Ventures, ***Management Science***, Vol. 48: 364-381: INFORMS: Institute for Operations Research.
Slatcher, R., & Pennebaker, J. 2006. How do I love thee? Let me count the words: The social effects of expressive writing. ***Psychological Science***, 17(8): 660-664.
Sorenson, O., & Stuart, T. E. 2001. Syndication Networks and the Spatial Distribution of Venture Capital Investments. ***American Journal of Sociology***, 106(6): 1546-1588.
Staiger, D., & Stock, J. H. 1997. Instrumental Variables Regression with Weak Instruments. ***Econometrica***, 65(3): 557-586.
Stiglitz, J. E., & Weiss, A. 1981. Credit Rationing in Markets with Imperfect Information. ***The American Economic Review***, 71(3): 393-410.
Strong, N., & Xu, X. 2003. Understanding the equity home bias: evidence from survey data. ***Review of Economics and Statistics***, 85(2): 307-312.
Stuart, T., & Sorenson, O. 2003. Liquidity events and the geographic distribution of entrepreneurial activity. ***Administrative Science Quarterly***, 48(2): 175-201.
Susarla, A., Subramanyam, R., & Karhade, P. 2010. Contractual Provisions to Mitigate Holdup: Evidence from Information Technology Outsourcing. ***Information Systems Research***, 21(1): 37-55.
Sutter, M., & Kocher, M. 2004. Favoritism of agents-The case of referees' home bias. ***Journal of Economic Psychology***, 25(4): 461-469.
Tanriverdi, H., Konana, P., & Ling, G. 2007. The Choice of Sourcing Mechanisms for Business Processes. ***Information Systems Research***, 18(3): 280-299.
Tetlock, P. 2007. Giving content to investor sentiment: The role of media in the stock market. ***The Journal of Finance***, 62(3): 1139-1168.
Tetlock, P. C., Saar-Tsechansky, M., & Macskassy, S. 2008. More Than Words: Quantifying Language to Measure Firms' Fundamentals. ***The Journal of Finance***.
Tomz, M., King, G., & Zeng, L. 2003. RELOGIT: Rare events logistic regression. ***Journal of statistical software***, 8(2): 137-163.
Tsang, E., & Yip, P. 2007. Economic distance and the survival of foreign direct investments. ***The Academy of Management Journal (AMJ)***, 50(5): 1156-1168.
Tykvová, T. 2007. Who chooses whom? Syndication, skills and reputation. ***Review of Financial Economics***, 16(1): 5-28.
Uzzi, B. 1999. Embeddedness in the Making of Financial Capital: How Social Relations and Networks Benefit Firms Seeking Financing. ***American Sociological Review***, 64: 481-505.
Uzzi, B., & Lancaster, R. 2003. Relational Embeddedness and Learning: The Case of Bank Loan Managers and Their Clients, ***Management Science***, Vol. 49: 383-399: INFORMS: Institute for Operations Research.
Whittington, K. B., Owen-Smith, J., & Powell, W. W. 2009. Networks, Propinquity, and Innovation in Knowledge-intensive Industries. ***Administrative Science Quarterly***, 54(1): 90-122.
Wolf, H. 2000. Intranational home bias in trade. ***Review of Economics and Statistics***, 82(4): 555-563.
Yi, K.-M. 2010. Can Multistage Production Explain the Home Bias in Trade? ***American Economic Review***, 100(1): 364-393.