New Technologies and the Value of Information Systems

Hossein Ghasemkhani

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2013

Reading Committee:
Yang Tan, Chair
Ming Fan
Hamed Mamani

Program Authorized to Offer Degree:
Business Administration

DEDICATION

To my wife, Ameneh, who has always believed in me and supported me, and to my parents who have always been there for me.

ACKNOWLEDGEMENTS

I would like to take this opportunity to express my gratitude to those who helped me to get where I am today. In the past four and a half years, I have learned so much from the faculty and stuff at the University of Washington. I want to thank Professor Yong Tan, who has been a great advisor, mentor, and most importantly friend for me in the past few years. We have spent numerous hours discussing research and teaching, traveling and eating together, and have formed a bond that I hope will last longer than my academic career. I also want to sincerely thank other members of my supervisory committee, Professors Guozhong Cao, Ming Fan, Jacques Lawarrée, Douglas MacLachlan, Hamed Mamani, and Arvind Tripathi who have been most helpful and supporting of me and my work since I formed my committee.

The faculty members at the ISOM Department of the Foster School of Business have played a fundamental role in my growth as a scholar. Especially, I cannot overstate the value of mentoring and support that I received from Professor Kamran Moinzadeh from whom I have learned so much in the past four years. I would also like to thank Professor Arvind Tripathi again who has taught me a lot and supported me since the day I entered the PhD program, and Professor Atanu Lahiri who has been constantly motivating and challenging me in research and teaching.

I would like to extend special thanks to Shawna Reimers, the ISOM Department Administrative Assistant, and Jaime Banaag, the Assistant Director of the PhD program for their kindness, support and consistently positive attitude. Also my special thanks go to my friends and colleagues, Young Jin Lee and Yi-Chun Ho, who consistently supported me and made my PhD

experience so much better. I feel blessed that I had the opportunity to know these individuals and work with them.

I should thank my Professors in Sharif University of Technology for guiding and supporting me in my undergraduate and master's education. Their teachings laid the foundation based on which I started and finished my doctoral studies. I also learned so much from my work experiences and my colleagues were a huge part of it, for which I am very thankful.

Last but not least, I must thank my family for their unconditional support. My amazing wife, Ameneh, has been my best friend, believer and support in the past nine years, and I would not have been where I am now if I had not had her beside me. I will also be eternally thankful to my parents who have always believed in me and my abilities, and given me unwavering love and support.

University of Washington

ABSTRACT

New Technologies and the Value of Information Systems

Hossein Ghasemkhani

Chair of the Supervisory Committee:
Professor Yong Tan
Foster School of Business
Department of Information Systems and Operations Management

As new forms of information systems emerge, they affect organizations in new ways. Technologies such as cloud computing, virtualization and mobile computing are revolutionizing the way businesses work. So many aspects of the value that is created by these new technologies are not explicitly observable, and capturing them calls for approaches beyond measuring firms' financial outcomes. In this dissertation, I conduct three studies to examine how technologies create tangible and intangible value.

In the first study, we study the intangible value created through availability of information in the context of online peer-to-peer lending. Technological advancements in web-based information systems have given rise to Web 2.0 market mechanisms in which consumers act as buyers, sellers and experts. Participants act based on their perception of trustworthiness of other parties and the risks involved in the associated transactions. We use data from a peer-to-peer lending community to empirically investigate the role of information systems in alleviating adverse selection. We posit that due to the repeated nature of transactions, borrowers develop a reputation which is the collective perception of the lenders about the borrower. We propose a simultaneous-equation model of outcomes and a dynamic latent class model of reputation, and use the novel approach of latent instrumental variable modeling to deal with endogeneity. We show that accounting for reputation improves the explanatory power of the model, and provide a way to empirically model the evolution and impact of reputation in online platforms where repeated transactions are performed.

Retailers are increasingly exploiting online auctions as an effective and low cost distribution channel for disposing large quantities of inventory. To clear large inventories, these sellers conduct many auctions of identical items one after another, termed as sequential online auctions. In such auction environments, bidders have opportunity to participate in many auctions to find good bargains. There has been a host of studies to explore bidder behavior in this environment, but none has ever addressed the underlying reasons of a bidder's decision to stop participating before winning the item. In the second study, we use survival analysis to investigate this phenomenon which we call bidder's attrition.

In the third study, we analyze the competition dynamics of cloud computing providers using two different analytical models. We first use a two-period model of duopoly competition of cloud providers with differentiated products. We show that when capacity is perceived as a quality signal and economies of scale are present, the two markets will be separated such that each firm only serves the customers that have a preference for its product. We then develop a differential game model of duopoly competition, derive the long-run stationary equilibrium, and analyze the short-run dynamic decision making of the firms in the transition period. We observe that although initial market shares do not affect the steady-state market sizes, initial capacities, maximum budget, cost of capacity provision and market switching parameters are all important in determining the steady-state market structure and long-run profit streams of the firms.

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1. Background

There are several papers by information systems researchers on the value of information systems. Many empirical studies try to find a causal relationship between different information technology and information systems variables such as IT investment and IT labor with firms' performance variables such as income and profit. However as new forms of information systems emerge, they affect organizations in new ways. Technologies such as cloud computing, virtualization and mobile computing are revolutionizing the way businesses work, in a way that it is nearly impossible to draw comparisons. So many aspects of the value that is created by these new technologies are not explicitly observable, and capturing them calls for approaches beyond measuring firms' financial outcomes.

Looking at Web 2.0 and mobile technologies, and opportunities for digital business that have been made possible through these technologies, examples of the value created by these new technologies can be clearly seen They are thoroughly reshaping many aspects of businesses' operations and people's lives. They have enabled the creation of social networking websites such as Facebook and Twitter and have resulted in omnipresence of internet connectivity through smartphones, tablets and ultraportable laptops. It is oftentimes hard, if not impossible, to measure the financial impacts of IT/IS investments, due to the fact that they don't simply improve productivity, they change the game. They create possibilities that were otherwise nonexistent, such as the kind of portability and scalability afforded by cloud computing. It was from this perspective that I decided to investigate the indirect and sometimes intangible ways through which information systems create value.

The essence of this dissertation could be captured by the following three sentences. I want to in investigate how information systems create tangible and intangible value for organizations and people. I want to study how information systems, information technology and increasing access to information impact business operations and people's lives, and how they can take advantage of these changes.

In my dissertation, I intend to study three of these new technologies and the ways they create value from different perspectives. In the first essay, I look at peer-to-peer lending markets and perform an empirical analysis of reputation formation and the value created through the information availability. In the second essay, I perform an empirical study of a series of sequential online auctions, and in particular, investigate the reasons bidders stop participating in the auctions before having their demand satisfied. Finally in the third essay, I study cloud computing and use a game-theoretic analytical approach to study the dynamics of competition between the cloud providers.

1.2. Reputation Building in an Online P2P Lending System

Peer-to-peer (P2P) systems facilitate direct communication and resource sharing among peers and are characterized as decentralized, self-organizing distributed systems that exploit and efficiently make use of the untapped resources of the heterogeneous hosts in the P2P networks (Boutaba and Marshall 2006). Traditional online peer-to-peer systems are designed for the sharing of computer resources like content, storage and CPU cycles by direct exchange, rather than requiring the intermediation or support of a centralized server or authority (Androutsellis-Theotokis and Spinellis 2004). These systems are essentially communities of self-governing interconnected individuals with great potential for collaboration. Online peer-to-peer lending systems provide a platform for borrowers to directly communicate with lenders, get loans from them and pay them back. In doing so, online P2P lending communities effectively eliminate financial intermediaries such as banks.

Microfinance as a means to provide funding and financial support to people in need has received considerable attention across the world in the past few decades. The original idea of microfinance is simple; small loans are provided to poor people to start or improve their businesses and help lift them out of poverty (Bruett 2007). Mohammed Yunus, an economics professor at a Bangladesh university, who started making small loans to poor local people in 1970s, is the pioneer of the modern day microfinance (Armendáriz and Morduch 2010). The Grameen Bank, the financial institution founded by Yunus in Bangladesh for the same purpose, and similar microfinance institutions, provide small loans without collateral to customers who have been written off by commercial banks as unprofitable (Armendáriz and Morduch 2010). Yunus believes that credit is a fundamental human right and has advocated micro-lending as an approach to eradicate poverty (Coleman 2007). Some of these institutions use innovative approaches such as group-lending contracts which effectively make a borrower's neighbors co-signers to loans, mitigating problems created by informational asymmetries between lenders and borrowers (Morduch 1999).

Building on the same idea, Prosper.com started the first commercial online P2P lending market in the US, soon followed by Lending Club and other lending platforms (Renton 2012). Prosper started its business with the goal to combine the notion of a marketplace such as Amazon or eBay for money with the idea of microfinance that was already present in developing countries (Prosper.com 2010). In 2012, Lending Club and Prosper are both thriving as the leaders in an industry that is gaining widespread attention from borrowers and investors. With greater than 100% year over year growth, P2P lending is a fast-growing investment platform (Renton 2012). Even some of the large traditional financial institutions have begun to consider peer-to-peer lending, as offering loans through these websites could be a way to bring in more deposits and reach more consumers (Kim 2008). In 2012, the Industry volume is over $50 million in new loans per month and over the 2012 Memorial Day long weekend, total loan volume passed the $1 billion mark since the industry began back in 2006 (Renton 2012). Borrowers are usually able to get loans more quickly and with less paperwork through online P2P lending systems compared to a bank. People with good credit are able to lock in lower rates than they would otherwise have to pay on credit cards or unsecured bank loans, and people with bad credit also get a second chance (Kim 2008).

The lean nature of electronic markets as compared to the traditional markets could lead to transaction risks for reasons such as lack of information on the quality of the services and products, or the identity of online trading parties, as online trading parties can often remain anonymous or change their identities (Ba and Pavlou 2002a). For instance on consumer-to-consumer (C2C) auction platforms such as eBay, individuals could have multiple e-stores or accounts. Neumann (1997) argues that when there is the risk of identity misuse, proper and strong user authentication measures must be put in place to mitigate the risk. These problems are mitigated in online P2P lending communities, as the users are uniquely identified. These communities are very strict in verifying user identities. Users are required to provide key information such as their social security numbers when they register to become borrowers and request a loan or even to become lenders. The provided data will then be verified by the P2P platform, before the users can join. So in contrast to many other e-commerce platforms, users cannot fake their identity and are held responsible for their actions as their actions are directly traced back to the users. P2P lending platforms provide a wealth of information on virtually all of the activities of the borrowers on these platforms since the time of joining in addition to the hard credit information such as credit score, debt to income ratio and number of delinquencies.

Credit bureaus use statistical analysis of people's credit information to build credit scores that are supposed to represent people's credit worthiness. Online P2P lending communities effectively provide platforms on which every lender can individually analyze not only the current state and the evolution of borrowers' credit score and other standard credit information, but also the history of borrowers' activities in the online lending context. Especially that as time goes on and the amount of information available on members' transactions and decisions increases, members of the community could establish themselves as trustworthy, or as having the reputation to behave in certain ways. We posit that this virtual repository of data on money-related activities of the borrowers in the P2P lending community results in the accumulation of a form of social capital that we refer to as a borrowers' reputation in the P2P lending community.

Following the literature in economics and finance on repeated games, we define reputation of a given agent as a form of identity that other agents shape, from learning from observed behavior about some exogenous characteristics of the agent over time (Kreps and Wilson 1982a; Milgrom and Roberts 1982). An issue inherent to P2P lending systems is adverse selection due to the information asymmetry between borrowers and lenders. Borrowers' types and their intention to default are private information of the borrowers. Incentive problems can be most severe for borrowers with very short track records and become less severe for borrowers who manage to acquire a good reputation (Diamond 1989).

Drawing on the idea of wisdom of crowds, we posit that the additional information generated and kept by the online systems could be used as signals to alleviate information asymmetry, and result in collective decisions by lenders that could be potentially more efficient than those driven by algorithmic procedures of financial institutions (Brabham 2008; Surowiecki 2005; Kittur et al. 2007). Therefore we conjecture that for the loans of a similar structure, online P2P lending systems might even be more efficient than the traditional financial intermediaries such as banks. We will not formally investigate this conjecture, as this task will require having access to data to construct comparable measures of efficiency on both markets. We will, however, show that the historical information on borrowers' activities within the community is used by lenders when making decisions, and that this information is helpful in predicting a borrower's risk of defaulting on a loan.

We propose two approaches to model the evolution of reputation, a hidden Markov model and a dynamic latent class model of reputation. In both models, it is assumed that there are a finite number of reputation states, and at a given time, a given borrower resides in one of these states. The impact of reputation is measured via three outcomes; the success of a loan request, the ability to secure a lower interest rate, and the probability of defaulting on a loan. We use a hierarchical Bayesian approach to estimate the models and compare the models with each other. We also estimate a benchmark model with no explicit reputation component. We provide model estimates for each reputation state, and show that the dynamic latent class model performs best among the three.

In Chapter 2, we will first study the existing body of related literature. Then in Section 2.2, we will discuss the conceptual model and elaborate our research question. In Section 2.3, we discuss our empirical model and the equations and parameters that need to be estimated. In

Section 2.4, we will discuss our different approaches to model the evolution and impact of reputation, and in Sections 2.5 and 2.6, we look at the data, variables and measures, and discuss the estimation approach. We provide the results and discussion in Section 2.7, and finally conclude the study in Section 2.8.

1.3. Bidders Attrition in Sequential Online Auctions

Over the last decade as electronic markets have matured, online auctions have emerged as a key component of these markets. Buyers prefer auctions because this market mechanism allows them to collectively determine the price, and creating opportunities for big bargains. Bidder competition and their desire to win help sellers as they can sell an item to highest paying bidder. Overall, online auctions help increase allocation efficiency by matching products to buyers who value them most.

In the beginning, the majority of the auctions were C2C but online auctions have evolved since businesses have developed strategies to use auctions and sell items via multiple channels. While Internet auction houses like eBay facilitate auction platform to sellers of any size, the large retailers, in particular, have benefited by integrating online auctions into their overall selling strategy. First of all, these retailers intend to sell large inventories through auction channel and so they desire to create an auction market that can clear their inventories and fulfill buyers' needs. Further, creating their own auction market allows them to choose from a variety of auction mechanism and allows them more control on the structure of this auction market. Retailers like Dell and Sam's Club who have always dealt with surplus inventories and ensuing revenue losses have now created their own auction markets to clear off their sizable inventories of various items. To clear large inventories, these sellers conduct many auctions selling identical items, one after another, termed as sequential online auctions. These sequences of identical item auctions create a market dynamics where both sellers and buyers have information about market supply/demand, and hence, can derive optimal strategies to maximize their payoffs. Nevertheless, this sequential auction environment brings design and operational challenges for the sellers. For example, a few recent studies (Pinker, Seidmann, & Vakrat, 2010; Tripathi, et al., 2009) have examined sellers' problem of determining number of auctions and number of items (lot-size) in each auction, given an initial inventory of items. This decision making becomes even more challenging with product obsolescence, stochastic demand (Chen, et al., 2011) and strategic bidders (Zeithammer, 2007). Other studies have investigated evolution of bidder behavior in these auctions (Goes, et al., 2010).

This research focuses on the interesting but largely overlooked phenomenon of bidders' attrition in sequential online auctions. In sequential auctions where identical items are sold in successive auctions, it is logical to assume that bidders who didn't win in the previous auctions will avail the opportunity to participate in the future ones to secure a win. However, empirical studies have shown that more than fifty percent unsuccessful bidders do not come back to participate in auctions selling identical items. Figure 1 shows the percentage of bidders who participate in more than one auction (called repeat bidders) of an auction sequence selling identical items. Although, these repeat bidders strive to win by participating in many auctions of an auction sequence, not all of them become successful and many of them leave without winning.

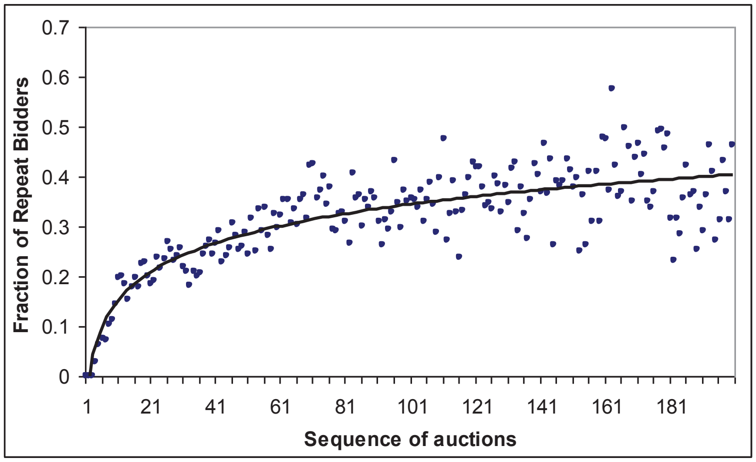

Figure 1: Fraction of Repeat Bidders in Sequential Auctions

Just like online retailers, online auction houses also face stiff competition from other auction platforms. Since the success of online auctions thrives on number of participants, attracting new and retaining existing participants is crucial for the success of online auctions. Online auction platforms, similar to other online retail websites, spend a considerable amount of resources to attract new customers and retain the existing ones. Successful bidder retention lowers the cost of attracting new bidders and further, the increase in the number of bidders brings in more competition among bidders, which eventually leads to higher clearing prices (Pinker et al. 2003). Hence, it is important for the auctioneers to understand the factors that affect bidders' attrition in sequential auctions. We define bidders' attrition as the situation where bidders stop participating in a sequence of consecutive auctions selling identical items (auction sequence), before completely satisfying their demand. Note that bidders are heterogeneous in their demand, such that, some bidders participate to win only one item whereas others may participate to win multiple items (Vincent Lyk-Jensen and Chanel 2007; Goes et al. 2010). Estimation of the bidders' demand is made based on bidders' behavior in sequential auctions.

There is a very rich literature in marketing and economics on customer retention/attrition. For example, Srinivasan, Anderson, & Ponnavolu (2002) investigate the antecedents and consequences of customer loyalty in an online B2C context. Li, Browne, & Wetherbe (2007) study the reasons for internet users to stick with a web site. However to the best of our knowledge, prior research has not examined the factors affecting bidders' attrition in online auction environment. We aim to fill this gap and investigate factors that affect bidders' attrition in sequential auctions using a discrete time hazard model.

In section 3.1 we will summarize the relevant literature. Section 3.2 presents a conceptual model and derives testable hypotheses. Section 3.3 overviews our data collection effort from a popular online auction site. Section 3.4 presents results and discussions while section 3.5 concludes with directions for future research.

1.4. Competition in Cloud Computing Market

Since 2008, cloud computing has appeared among the Gartner's Top 10 Strategic Technologies and Gartner's Hype Cycle due to its disruptive force and potential for broad long-term impact in most industries (Fenn 2012; Gartner 2011). Some analysts and vendors define cloud computing narrowly as an updated version of utility computing. Others say anything you consume outside the firewall is "in the cloud," even including conventional outsourcing (Knorr and Gruman 2008). Based on a recent survey of 1,650 IT and security decision-makers, more than a third of the current IT budgets is being allocated to cloud computing solutions. Also more than half of the respondents expect to increase spending in the next 12 months, such that on average, the organizations will increase spending on cloud computing by 16% (IDG-Enterprise 2012).

Proponents of cloud computing argue that cloud providers are able to provide computing services more efficiently resulting in higher quality services at lower costs, due to specialization and economies of scale. Amazon.com's vice president argues that servers, networking and administration costs for an average enterprise is five to seven times that of a large provider (Hamilton 2010). Other benefits of cloud computing are scalability and flexibility which are especially important for small firms with more demand uncertainty. Taking advantage of this scalability enables firms to quickly scale their computing power up and down to deal with peaks and valleys in their demand for computing. There are several cases of start-up companies that have been able to cope with unexpectedly high surges in demand using the services offered by cloud providers. Even large firms such as Netflix use cloud computing services to satisfy their computing needs. In a business environment characterized by rapid changes in trends and customer expectations, cloud computing enables firms to be more agile in dealing with these changes which in turn fosters innovation and makes even more rapid changes possible.

Forrester Research predicts that the global cloud computing market will grow from $40.7 in 2010 to more than $240 billion in 2020 (Ried and Kisker 2011). Oracle, IBM and SAP all have major initiatives to deliver a broader range of cloud services and even companies such as LG and Samsung are offering cloud services (Fingas 2012). As the current players continue to expand their offerings, and traditional players enter the market, users will see competition heat up (Gartner 2011). Thus it is interesting to study the competition of the small and big players in this market. It is perceivable that given the capital intensive nature of cloud computing and the economies of scale, in the long-run, there will only be a handful of large providers remaining and the other companies will either crash out of the competition, be bought out by larger players or find a niche market to focus on.

Service offered by the cloud providers could be categorized into three broad service models: Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS) (Mell and Grance 2011).

Infrastructure as a Service is the capability provided to the customers to provision processing, storage, networks and other fundamental computing resources where the consumer is able to deploy and run arbitrary software (Mell et al. September 2011). In this offering, the users have the greatest level of control as they get to manage and control the operating systems, storage and the deployed applications and even some networking components such as load balancers and the host firewalls. They however still don't manage the underlying cloud infrastructure. The most prominent example for an IaaS provider is probably Amazon.com's AWS (Amazon Web Services) whose most notable offerings are Elastic Compute Cloud (EC2) and Simple Storage Service (S3).

The NIST defines Platform as a Service as the capability provided to the consumers to deploy onto the cloud infrastructure consumer-created or acquired applications (Mell et al. September 2011). These applications are created using programming languages, libraries, services, and tools supported by the cloud computing provider on a hosted infrastructure delivered as a service. Typical examples of PaaS include middleware, databases and development tools delivered as a service through the internet. In this type of cloud offering, consumers have control over the deployed applications and possibly configuration settings of the application-hosting environment. However they still do not manage or control the underlying infrastructure such as the network, servers and operating systems. Microsoft, Google and Salesforce.com are among the pioneers of PaaS with their respective offerings, Microsoft Azure, Google Apps Engine and Force.com.

SaaS providers offer to the consumers the capability to use the provider’s applications running on a cloud infrastructure (Mell and Grance 2011). The applications are accessed either through a thin client such as a web browser or through a program interface. In this type of cloud offering, the users are only allowed to make limited user-specific application-level configurations, and they do not manage or control the underlying cloud infrastructure such as servers and networking. There are a lot of examples for SaaS such as Google’s email and productivity services, Gmail and Google Docs, QuickBooks online and online services offered by Photoshop.com.

One can also categorize and classify cloud computing services based upon the deployment models utilized. This can also be considered as the degree to which cloud computing is externalized from the organizations taking advantage of the services. There are four types of cloud offerings based on deployment models, i.e. private cloud, community cloud, public cloud and hybrid cloud (Mell et al. September 2011). Private cloud is the cloud computing service that is intended for exclusive use by a single organization. This type of cloud could be owned and managed by the organization or a third party or a combination of these two. In terms of location, private cloud may be on or off the organization’s premises.

There are extensive studies by computer scientists on issues related to cloud computing, such as distributed systems and virtualization, but to the best of our knowledge, there is no work done on the economics of cloud computing and competition dynamics by information systems researchers and economists. We try to address this gap using two different game-theoretic models. A two-period model and a differential game model of duopoly competition. The research agenda that we are trying to address is two-fold. We intend to build a model which enables us to study the dynamics of decision-making by cloud computing providers, given the unique characteristics of this market. We also try to investigate the short-term and long-term outcomes of the competition in this market using our models. In this study, we have presented models through which we try to investigate the forward-looking capacity and pricing decisions of competing cloud providers. We derive the symmetric subgame perfect Nash equilibrium of the two-period game and characterize the open-loop symmetric Markovian Nash equilibrium of the differential game and briefly discuss its properties.

We use the two-period game because it allows us to build a more comprehensive and realistic model and investigate the effects of more parameters of interest. However, due to the fast-paced and dynamic nature of the market and the cloud computing providers' decision making, a differential game model is an appropriate approach to study competition in a cloud computing market. An executive at Amazon.com claims that "every day Amazon Web Services adds enough new capacity to support all of Amazon.com's global infrastructure through the company's first 5 years, when it was a $2.76 billion annual revenue enterprise." (Miller June 2011)

We will see that based on our two-period model, if it is profitable for both firms to produce, each firm will completely capture the market segment which prefers that firm's product, so the market will be separated. This is true as long as there are economies of scale. Through the differential game model, we show that the market will stabilize in the long-run after which the market sizes of the two competitors will stay the same. The sizes of the market shares depend on the initial capacities, the customers' tendency to switch providers, and the firms' budget and capital investment in the short-run, while they are independent of the initial market shares. We also briefly investigate the short-run competition dynamics. In chapter 4, we first review the related literature. In sections 4.2, we will explain and analyze the two-period model and in sections 4.3 and 4.4, we will study the differential game model of competition respectively. Finally we will provide the concluding remarks, and potential areas for future research in section 4.5.

Chapter 2 The Invisible Value of Information Systems: Reputation Building in an Online P2P Lending System

2.1. Literature Review

Webster's Revised Unabridged Dictionary (1913) defines reputation as “the estimation in which one is held; character in public opinion; the character attributed to a person, thing, or action” and “the character imputed to a person in the community in which he lives.” The concept of reputation has been widely studied by researchers in various academic disciplines, such as strategic management, organizational theory, economics, marketing, communications, accounting and finance (Fombrun and van Riel 1997).

From a game-theoretic standpoint, strategies of the agents in repeated games could be significantly different from those of a single game, as the repeated game strategies are a sequence of rules that might depend upon the preceding outcomes (Rubinstein 1979). Because of the repeated plays, the agents are able to respond to each other’s actions, so they must consider this and other players’ reactions before choosing their action (Fudenberg and Maskin 1986). In repeated game settings where information asymmetry is present, role of reputation and reputation building is especially important. In these settings, assuming the agents are heterogeneous in some respect, reputation works as a means for agents to signal their type (Weigelt and Camerer 1988).

The adverse selection approach to modeling reputation is the cornerstone of game-theoretic study of reputation. An agent’s type is private information of the agent which gives rise to the information asymmetry issue. Other agents observe the actions of a given agent, agent A, over time and update their perception of the agent’s type using the Bayes’ rule. The agents’ perception of agent A's type is usually modeled as a probability distribution which changes over time. A way to study the equilibrium outcomes and strategies in these settings is to use Kreps and Wilson's notion of sequential equilibria for games in extensive form (Kreps and Wilson 1982b). Sequential equilibria are further refinements of Nash equilbiria and subgame-perfect Nash equilbria in which the players' strategies are required to be sequentially rational. In order for that, players' beliefs must be considered in addition to their actions. It must be done in such a way that given a certain belief, a certain course of action will be optimal. A player decides on his or her future actions based on prior actions of the other players and his or her beliefs about the possible actions of them in the future.

As long as the long-run player with private type in a sequential game, in our case the borrower, is sufficiently patient, he or she can leverage other players' uncertainty about his or her type and use reputation effects to secure higher payoffs (Atakan and Ekmekci 2012; Fudenberg and Levine 1989, 1992). An agent's patience is defined as the case where the long-run benefits of reputation building outweigh the costs that agent has to incur in the short run. In infinitely repeated games, and under some conditions, the reputation effect disappears over time and the agents' types are eventually revealed (Cripps et al. 2004, 2007). Many of the reputation effect studies assume a setting in which the player with private information plays the game infinitely, while others are short-lived. The short-lived players are, however, aware of all the prior plays. In these settings, results similar to folk theorem are obtained, i.e., each individually rational payoff is obtainable by an equilibrium if the players' discount factor is sufficiently close to one (Fudenberg et al. 1990). In our study, the borrowers could be modeled as potentially long-run players. While at least some of the lenders might be short-lived players, they are nevertheless informed about all prior actions of the borrowers.

Another discipline that has extensively studied firm, brand and product reputation is marketing. Marconi (2002) argues that an agent's target audience act upon what they have seen, heard, and learned. A name, or a single word, can suggest an image, and can speak volumes in terms of the agents' beliefs and perceptions. Gotsi and Wilson (2001) provide a survey on the concept of corporate reputation and its relationship with corporate image in marketing. They categorize the definitions offered for the term corporate reputation by marketing researchers into two schools of thought; the analogous school of thought, which views corporate reputation as synonymous with corporate image, and the differentiated school of thought, which considers them to be different, but interrelated. Dawar and Parker (1994) empirically evaluate the use of seller reputation by customers as a signal of quality along with other marketing variables such as price and physical product characteristics in different cultures. Herbig and Milewicz (1993) study the importance of a firms' reputation in successful performance of its brands, the impact of a firm's reputation decay on the firm's brands and how a brand's reputation can be transferred to other products.

Researchers from management, strategy and organizational behavior have also studied reputation, particularly corporate reputation. They consider corporate reputation to be an intangible organizational asset derived from internal features of a firm (Roberts and Dowling 2002). Weigelt and Camerer (1988) define reputation as a set of attributes ascribed to a firm by its stakeholders that are inferred from the firm's past actions. Some studies argue that the existing measures of firms' performance such as profitability and other financial measures are inadequate and further measures must be developed that value intangible assets like reputation (Chakravarthy 1986; Rindova and Fombrun 1999). Organizational reputation is said to be interconnected with a firm’s culture and the attitude of the employees toward stakeholders (Barney 1986; Dutton and Dukerich 1991).

Our study is directly related to the area of reputation building in the context of consumer-to-consumer online systems and social capital accumulation. Peer-to-peer lending, like most other instances of electronic commerce, is a form of online exchange in which most transactions occur among entities that have never met (Ba and Pavlou 2002a). The existence of information asymmetry means that there is potential for opportunistic behavior, which could lead to loss of trust or even market failure due to the lemons problem (Akerlof 1970). Trust consumers have for an online retailer and the associated value of branding is an important factor in an online retailer’s success and its market power much similar to traditional market settings (Ba et al. 1999; Brynjolfsson and Smith 2000). Several information systems researchers have examined the role of trust in e-commerce transactions. Much of this research has looked at the nature of consumer trust placed in institutions supporting e-commerce (Vance et al. 2008). Trust and trust-building mechanisms are important in electronic commerce, as they can reduce the perceived uncertainty and risk associated with online transactions and help consumers get more actively involved in online activities like exchanging personal information and purchasing goods and services (Greiner and Wang 2010; Gefen et al. 2003; McKnight et al. 2002).

A concept that is closely related to reputation is trust. Trust has been studied from different perspectives and modeled in different ways, both theoretically and operationally, and academics have long acknowledged the confusion surrounding the topic (Gefen et al. 2003; Shapiro 1987). Some define trust as a set of specific beliefs that agents have toward a given agent, mainly about the agent’s integrity, benevolence, and abilities (Doney and Cannon 1997; Sganesan 1994). Others define it as a general belief instead of a specific one that another party can be trusted (Zucker 1986). This is referred to by Mayer et al. as the "willingness of a party to be vulnerable to the actions of another, based on the expectation that the other party will perform a particular action which is important to the trustor, irrespective of the ability to monitor or control that other party" (Mayer et al. 1995). This definition emphasizes that despite the uncertainty surrounding the transaction, one party is willing to take the risk and potentially lose something important by depending on the other party, and this willingness is dependent upon certain expectations or beliefs about the other party (Greiner and Wang 2010).

Most of the existing C2C reputation and trust literature focuses on institution-based trust, reputation systems, or the evolution of reputation and trust in the population of sellers in general (e.g. Ba and Pavlou 2002a; Greiner and Wang 2010; Ba et al. 1999). Xiong and Liu (2003) argue that P2P electronic commerce communities could offer both opportunities and threats, and that a way to minimize threats in such environments is to use community-based reputation mechanisms to help the peers evaluate the trustworthiness of other peers and predict their future behavior. Gefen et al. (2003) find that online trust in the context of e-commerce is built through a belief that the party has nothing to gain by cheating, a belief that there are safety mechanisms built into the system, or by having an intuitive and easy-to-use interface. Vance et al. use an empirical model to study the topic of trust in information technology artifacts and find that system quality constructs significantly predict the extent to which users place trust in mobile commerce technologies (Vance et al. 2008).

Many researchers, particularly in computer science, have studied reputation mechanisms that help participants of online communities gain a better understanding of a given member's type and future actions, based on his or her previous activities (Despotovic and Aberer 2006; Marti and Garcia-Molina 2006; Resnick et al. 2000). Some empirical studies have looked at the impact of reputation systems, and an agent's reputation according to these systems, on the agent's decisions and performance in platforms such as eBay (Melnik and Alm 2002; Resnick and Zeckhauser 2002; Houser and Wooders 2006). Ghose et al. (2005) argue that web-based systems that establish reputation are central to the viability of many electronic markets. They study different dimensions of online reputation and their influence on the pricing power of sellers and provide evidence that existing numeric reputation scores conceal important seller-specific dimensions of reputation. They instead propose a text mining technique that identifies and quantitatively evaluates further dimensions of importance in reputation profiles (Ghose et al. 2005). Ba uses a game-theoretic prescriptive approach to building a community responsibility system as a social structure, and show that under the community responsibility system for trust building, online transactions that are impersonal can be supported (Ba 2001).

A number of researchers from different disciplines have studied P2P lending from different points of view. Existing research on P2P lending can be categorized into three broad areas (Greiner and Wang 2010). Some are qualitative case studies of P2P lending marketplaces, such as Prosper.com and the UK-based company Zopa (Kupp and Anderson 2007; Briceño Ortega and Bell 2008). Second group of studies are in-depth investigations of the social aspects and features of P2P lending systems, such as groups and loan performance, friendship and the importance of the social network structure, intermediaries, and etc. (Berger and Gleisner 2009; Jeong et al. 2012; Lin et al. 2009, 2011). Finally the third category of studies apply existing economic theories to P2P lending, and develop new ones, and explore the effectiveness and efficiency of P2P lending marketplaces for creating a more open and competitive credit market (Garman et al. 2008; Zhang and Liu 2012; Herzenstein et al. 2011).

Greiner and Wang (2010) study borrowers' performance on one of the leading P2P lending communities, based on the elaboration likelihood model to model trust-building. Their results support the importance of the central route, economic status, as the major driver for success and that of peripheral cues, social capital and listing quality, as trust-building mechanisms that influence trust behavior (Greiner and Wang 2010). In our study, we focus on two levels of information, the current information available through the loan request, and the historical information that shapes the borrowers' reputation, including the historical credit information and the record of the borrowers' activities. Greiner and Wang, in contrast, categorize the available information into central and peripheral cues. Also as opposed to their study, we model all of the outcomes of interest simultaneously, include the risk of defaulting, and consider serial correlation and potential endogeneity.

Collier and Hampshire (2010) consider the community reputation system in an online P2P lending service. They argue that by embedding individual reputations within a community reputation, incentives become aligned for peers to select highly qualified borrowers and produce information signals to reduce the information asymmetry issues of any principle-agent relationship. They focus on groups and group membership in a P2P lending system, and utilize agency theory to examine the signals that enhance community reputation. There are also empirical studies on the effect of borrower attributes, such as the images posted by the borrowers, on the outcomes (Duarte et al. 2012; Ravina 2008; Pope and Sydnor 2011). Some studies focus on lenders' behavior and find evidence of rational herding among them when choosing what listings to invest in (Zhang and Liu 2012; Herzenstein et al. 2011).

Berger and Gleisner empirically analyze a P2P lending platform and argue that designated group leaders could be instrumental in screening of potential borrowers and the monitoring of loan repayment and that they somewhat assume the role of intermediaries on electronic markets. They find that these participants act as financial links between borrowers and lenders and can significantly improve the efficiency of the lending market (Berger and Gleisner 2009). Some researchers have studied the role of social networking in online P2P lending markets. Lin et al. analyze a sample of consummated and failed listings from Prosper.com to find that the online friendships of borrowers act as signals of credit quality. They find evidence that friendships increase the probability of successful funding, lower interest rates on funded loans, and are associated with lower ex-post default rates. Also based on their findings, the economic effects of friendships demonstrate a gradation based on the roles and identities of the friends (Lin et al. 2011).

2.2. Conceptual Model

The main goal of this study is to investigate whether the availability of additional information and collective evaluation of the borrowers by a large number of lenders create value. The online P2P lending market provides a platform through which lenders and borrowers can communicate, and provide personal information and descriptions of their needs and expectations. Furthermore the history of borrowers' activities such as their loan requests, their loan repayment records, late payments and changes in their credit information is readily available. This wealth of information could enable a rather personalized evaluation of the borrowers' trustworthiness and creditworthiness by the lenders. This is unprecedented as in the traditional financial markets, a limited number of financial intermediaries perform the task of evaluating potential borrowers. In an online P2P lending system, however, a loan is approved and given if several lenders are willing to participate, and the market decides whether a borrower should get a loan with the requested terms.

Rational members of a community will naturally not trust a member who has had the history of violating his or her agreements. Of course in order for the members to make decisions, the community must be structured in such a way that all the members are informed about others' identities and actions. In the context of infinitely repeated games, agents consider every action in the context of a sequence of social interactions. Therefore, although an agent could enjoy immediate gains by violating the agreements when others perform as expected, ruining one's reputation and losing the opportunity of future agreements with the community members might outweigh the immediate gains from refusing to reciprocate when other agents have kept their end of the deal (Vanderschraaf 2007). So the more transparent the community is and the better informed the community members are, the better incentives for the members will be to consider the long term implications of their actions.

We hypothesize that every borrower establishes some form of identity through participation in the online lending community which works as a signal of trustworthiness. Following the literature in economics, we refer to this directly unobservable identity as the borrower's reputation. This is possible because all of the users in the P2P lending market are uniquely identified and tracked, and a borrower cannot reset his or her history of actions and start afresh. So there are lasting consequences to any decisions made and any actions taken by the members. From a modeling point of view, this unobservable reputation can be modeled as a latent construct, the value of which changes over time as a result of the actions of the borrows. From an empirical point of view, two levels of heterogeneity are relevant here. First, different borrowers have different reputations which also change in time. Secondly a given borrower at a given time might be viewed as having different levels of reputation in the eyes of different lenders. In other words, different lenders might consider different criteria in evaluating a borrower’s reputation, and therefore have different evaluations of his or her level of reputation. For example some lenders might place a greater weight on a borrower’s age and experience in the lending community. Our model does not try to capture this heterogeneity of lenders in defining and measuring reputation. Instead we consider the average reputation of the borrowers within the lending community.

Information systems, in particular web 2.0, have enabled a virtual community with a great deal of information on a large scale, such that every potential lender can gain access to the borrowers’ past information and make informed decisions accordingly. So our main research question is whether lenders consider the borrowers’ reputation, i.e. the available history of actions of the borrowers, when they decide on whether to lend money and whether this reputation helps the borrowers get more favorable outcomes such as better interest rates. We are also interested to investigate whether borrowers’ reputation helps in predicting the risk of defaulting on a loan. We show that borrowers’ reputation is significant in predicting the success of loan requests and risk of defaulting in online P2P lending environments.

When registering to become members of this online community, members choose their roles. They could be borrowers, lenders or both. Registered borrowers provide their personal information including their social security number, employment status and income. All borrowers are verified before being allowed to borrow money. In order to borrow money, registered borrowers should list their loan requests. These loan requests are called listings. The loans created through this platform and similar competing P2P lending communities are unsecured, meaning that they are not backed by the borrowers’ personal assets. The loans have fixed interest rates, and the payback period for all loans is 36 months. The borrowers, however, can decide to pay the loan back in full earlier without incurring any penalties.

At the time of creating the loan listing, the borrowers choose how much money they want to ask for. A borrower cannot have more than $25,000 in loans outstanding within the community at any time. Borrowers also must specify the maximum interest rate they are willing to pay for the loan. They choose the duration for which the listing will be active which could range from three days to ten days. The borrowers also provide additional information such as the title of the listing, the loan category (such as business, debt consolidation, home improvement, etc.) and a description for the listing. They can provide more details on why they need the money, what their financial status is, and on their plan to pay the loan back.

Once the loan listing is created, it will be posted on the website among other listings. Lenders can browse the available listings and decide whether they want to invest on a listing or not. If a lender decides to invest on a listing, he or she should submit a bid on that listing. Bids include an amount, and the minimum interest rate the lender is willing to receive (which cannot be greater than the maximum interest rate posted by the borrower). By making a bid on a listing, a lender effectively commits to invest on the loan. They use the financial and personal information about the borrowers and the information available on borrowers' prior activities on the platform in making these decisions. To mitigate risk, most lenders choose to diversify their investment portfolios and lend only small amounts to any single borrower.

If the aggregate amount of bids submitted by the lenders on a given listing reaches the amount requested by the borrower, new lenders who want to invest in the loan have to compete with the existing ones by lowering the minimum interest rate they submit. Therefore the final interest rate of a loan could be lower than the borrower's specified maximum rate. The final interest rate and the winning lenders are decided through an auction mechanism. The auction mechanism is very similar to a second-price Dutch auction in the sense that the bidders who have submitted the lowest interest rates are selected until the full amount is funded, and all the winning bidders receive the same interest rate, which is the lowest losing interest rate.

When creating the listing, the borrowers have the option to let the listing be active until the end of the selected duration, or have the auction clock stop once the total amount submitted by the lenders reaches the amount requested by the borrower. This choice will be displayed on the listing information page. When one of the conditions to end the listing auction is triggered, i.e., either the listing is fully funded for those who have selected the option or the duration of the listing ends, the listing expires. Then if the total amount requested is covered by the submitted bids, the P2P platform owner will take the money from winning bidders and transfer the money to the borrower after consolidating it. Then the borrower starts paying back the principal and interest on the loan through monthly installments.

2.3. Empirical Model

2.3.1. Listing to Loan Conversion Model

We utilize a binary logit model in order to investigate the probability of a listing being funded and successfully converted to a loan. In the binary logit model, the likelihood of a successful loan depends on the borrower's reputation, borrower-level variables measured at the time of listing, such as the borrower's credit information, and listing-level variables such as the amount of effort put into creating the listing. Each listing by a borrower is an occasion with a binary outcome such that a fully funded listing is a successful listing. Integrating over the extreme value distributed random shock of the logit model, we attain the following expression for the probability of observing the successful conversion of a listing to a loan for borrower i in his or her j^{th} listing:

$$\Pr(u_{ij}=1)=\frac{\exp\left(x_{ij}{}'\beta_u^r + A_{ij}\alpha_u^r + MR_{ij}\gamma_u^r + \xi_u^i + \varepsilon_u^{ij}\right)}{1+\exp\left(x_{ij}{}'\beta_u^r + A_{ij}\alpha_u^r + MR_{ij}\gamma_u^r + \xi_u^i + \varepsilon_u^{ij}\right)} \quad , \; r \in \{1,2,...,R\}. \tag{1}$$

x_{ij} is a vector of n_x exogenous variables, whereas A_{ij}, the loan amount requested, and MR_{ij}, the maximum interest rate, could be potentially endogenous. $\xi_u^i \sim N\left(0,\sigma_{\xi_u}^2\right)$ is a borrower-specific random effect which is included to account for potential unobserved borrower-level heterogeneity (Wooldridge 2010). The model coefficients $\left(\beta_u^{r\prime} \; \alpha_u^r \; \gamma_u^r\right)'$ have reputation state, r, superscripts. This is because we are allowing the coefficients for borrowers at different levels of reputation to be different. $\varepsilon_u^{ij} \sim N\left(0,\sigma_u^2\right)$ is an idiosyncratic random shock. As we will see later, this unobserved part of the latent propensity to succeed is the part which is correlated with the error terms in other models. Based on the logit model of equation (1), we can form the following log likelihood function for this model that we will use to estimate the model.

$$LL_u = \sum_{i=1}^{n_b}\sum_{j=1}^{n_{b_i}}\left[u_{ij}\left(x_{ij}{}'\beta_u^r + A_{ij}\alpha_u^r + MR_{ij}\gamma_u^r + \xi_u^i + \varepsilon_u^{ij}\right) - \log\left[1+\exp\left(x_{ij}{}'\beta_u^r + A_{ij}\alpha_u^r + MR_{ij}\gamma_u^r + \xi_u^i + \varepsilon_u^{ij}\right)\right]\right]. \tag{2}$$

2.3.2. Interest Rate Model

In the online P2P lending community that we study, borrowers post listings in which they state the amount of money they need, and the maximum interest rate they are willing to pay. Then the lenders participate in an auction by submitting bids including the amount they intend to invest, and the minimum interest rate they are willing to accept. If the listing is fully funded, and the amount of bids submitted exceeds the amount of loan requested, the process of selecting lenders to contribute to the loan boils down to a competition based on the minimum interest rate they have submitted. Therefore the final interest rate of a successful listing, IR_{ij}, is ideally lower than the maximum rate posted by the borrower, MR_{ij}, and borrowers are naturally interested in larger reductions in the interest rate of the loans. The interest rate model investigates the size of decrease in the interest rate for a given listing, and we use a linear regression model for this purpose. In order to make sure the normally distributed error term assumption is viable, we use the monotonic logit transformation, $y_{ij} = \text{logit}\left[\left(MR_{ij} - IR_{ij}\right) / \max_{i,j}\left(MR_{ij} - IR_{ij}\right)\right]$, to extend the range of the left-hand side variable to $(-\infty, \infty)$.

Similar to the loan conversion model of (1), the reduction in interest rate is modeled as depending on a vector of exogenous variables x_{ij} and two potentially endogenous variables, A_{ij} and MR_{ij}. ξ_y^i is a mean-zero borrower-specific effect included to account for potential unobserved heterogeneity of borrowers, and $\varepsilon_y^{ij} \sim N\left(0, \sigma_u^2\right)$ is the idiosyncratic error term. Final interest rate is only relevant for listings which are converted to a loan, so the interest rate equation is modeled as conditional on the listing being successfully funded.

$$y_{ij} = x_{ij}{}' \beta_y^r + A_{ij} \alpha_y^r + MR_{ij} \gamma_y^r + \xi_y^i + \varepsilon_y^{ij}, \quad \{i, j \mid u_{ij} = 1\}. \tag{3}$$

2.3.3. Defaulting Model

When making the decision on whether to invest on a loan or not, the lenders are interested in evaluating the risk of defaulting for a given listing by a given borrower. This is especially important in a P2P lending community as the loans are unsecured. So in order for the rational lenders to evaluate the expected payoff from participating in a listing auction and eventually investing in a loan, they have to form an expectation of the borrowers' chance of defaulting on the loan. Similar to the listing to loan conversion model, we use a binary logit to model the borrowers' propensity to default on a loan.

$$\Pr(d_{ij}=1\,|\,u_{ij}=1)=\frac{\exp\left(x_{ij}{}'\beta_d^r+A_{ij}\alpha_d^r+MR_{ij}\gamma_d^r+\xi_d^i+\varepsilon_d^{ij}\right)}{1+\exp\left(x_{ij}{}'\beta_d^r+A_{ij}\alpha_d^r+MR_{ij}\gamma_d^r+\xi_d^i+\varepsilon_d^{ij}\right)}. \tag{4}$$

Similar to the interest rate model, the choice to default is only relevant in the case of a listing which is successfully converted to a loan. Therefore we model the defaulting decision conditional on the listing being successfully funded. Covariates of the model are similar to (1) and (3), and the vector of coefficients, $\left(\beta_d^{r\prime}\ \alpha_d^r\ \gamma_d^r\right)'$, have r superscript implying that they could be different for different borrowers at different times depending on their state of reputation. Similar to the listing conversion and interest rate models, we have included a borrower-level random effect term, ξ_d^i, to account for potential unobserved heterogeneity, and ε_d^{ij} is a mean-zero normal random shock which is potentially correlated with the contemporaneous shocks in the other two models. The log-likelihood function of the defaulting model is similar to that of the loan conversion model in (2).

2.3.4. Interdependence and Error Structure

It is reasonable to assume that the contemporaneous random shocks in the conversion model, the interest rate model and the defaulting model could be correlated, as the unobserved factors conducive to a listing’s success might also help a listing achieve a lower interest rate, and there might both be related to the unobserved factors that drive a borrower to default on a loan. To account for this potential interdependence, we assume the following covariance structure for the normal idiosyncratic error terms in each model, and later estimate the parameters of the covariance matrix along with other model parameters.

$$\begin{pmatrix} \varepsilon_u^{ij} \\ \varepsilon_y^{ij} \\ \varepsilon_d^{ij} \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_u^2 & \sigma_{u,y} & \sigma_{u,d} \\ \sigma_{u,y} & \sigma_y^2 & \sigma_{y,d} \\ \sigma_{u,d} & \sigma_{y,d} & \sigma_d^2 \end{pmatrix}\right]. \tag{5}$$

2.3.5. Latent IV

The requested amount, A_{ij}, and the maximum interest rate the borrower is willing to pay, MR_{ij}, are decision variables of the borrowers when they create listings, and these two variables could potentially be endogenous. One source of potential endogeneity is the omitted variable problem. The decisions of the borrowers on the amount and the maximum interest rate of a listing could be dynamically affected by their observations of the pool of active lenders and their level of activity at the time. However we, as the researchers, do not observe that, and therefore do not control for it. We also do not control for a listing's rank as shown to the lenders among other available listings. This rank could be correlated with A_{ij} and MR_{ij}, and it is reasonable to assume that it might also affect a listing's performance.

Another piece of information which is observable by the borrowers at the time of posting, but not to the researcher, is the current level of competition as measured by the status of other available listings, their amounts and interest rates, their performance, and etc. These could both affect a borrower's choices of amount requested and the maximum interest rate, and the outcomes of a listing. We also have not accounted for the semantic content of the listings' description which could be correlated with these two decision variables and the outcomes. There could also be potential endogeneity especially in the defaulting model as the decisions on the requested amount, maximum rate and defaulting are all made by the borrower. Finally in our interest rate model, we are calculating the left-hand side variable, y_{ij}, as a transformation of the difference between the final interest rate of a listing and MR_{ij} which could give rise to endogeneity.

A common approach to deal with endogeneity is the instrumental variable estimation approach. In this approach, we find variables that satisfy the inclusion and exclusion conditions, i.e., are correlated with the endogenous variables but are uncorrelated with the error term. Angrist and Krueger (2001) provide a good survey on instrumental variable (IV) estimation. This, however, is a challenging task, which is in many situations very hard or impossible to do. Oftentimes additional data is not available. Even if additional data is available, it is hard to find instruments that satisfy both conditions, especially that instruments that are not correlated with the disturbances are usually only weakly correlated with the endogenous variable (Staiger and Stock 1997). The use of weak instruments that explain little of the variation in the endogenous variable lead to large inconsistencies in the IV parameter estimates, even if the instruments and the error term are virtually uncorrelated. Also in finite samples, even with large sample sizes, IV estimates are biased in the same direction as the OLS estimates, and the magnitude of this bias could be even worse than that of the OLS as the R^2 between the instruments and the endogenous variable approaches zero. (Bound et al. 1995; Stock et al. 2002).

In this study, we utilize the Latent Instrumental Variable (LIV) approach, an instrument-free method proposed by Ebbes et al. (2005). Instrument-free or frugal approaches to dealing with the endogeneity problem utilize statistical procedures that do not require observed instruments. They therefore circumvent the instrument availability and validity issues. The LIV approach is a likelihood-based method that assumes we can separate the endogenous variable into two parts, an endogenous part and an exogenous part. The unobserved exogenous part is approximated using a latent discrete variable with a finite number of levels (Ebbes et al. 2009).

The LIV approach has been successfully applied by researchers to account for endogeneity in marketing applications (e.g. Zhang et al. 2009; Rutz et al. 2012; Rutz and Trusov 2011). Applying the LIV method, we can decompose each of the endogenous variables as following:

$$A = Z_A \varphi_A + \varepsilon_A, \tag{6}$$

$$MR = Z_{MR} \varphi_{MR} + \varepsilon_{MR}. \tag{7}$$

Here A and MR are $N \times 1$ vectors, φ_A and φ_{MR} are $C_A \times 1$ and $C_{MR} \times 1$ vectors, and Z_A and Z_{MR} are $N \times C_A$ and $N \times C_{MR}$ matrices respectively. Each of these equations is essentially a latent variable model in which $Z_s \varphi_s$, $s \in \{A, MR\}$ is the systematic part that is uncorrelated with the error terms in our structural models. Conversely, ε_s could potentially be correlated with one or more of the structural random shocks. The original model of Ebbes et al. (2005) was applied in a linear regression setting. Rutz et al. (2012) extend the application of the model to choice models. Building on these previous studies, we extend the application of the LIV approach to a system of simultaneous equations which includes both discrete choice and linear regression models.

In (6) and (7), we are assuming a single unobserved discrete instrument, $z_s^{ij\prime}$, for each observation. φ_s is a vector of category means which essentially works like the vector of coefficients in a linear regression model that need to be estimated. The instrument, $z_s^{ij\prime}$, is a $C_s \times 1$ vector such that one of the elements is 1 and the rest are 0. Each discrete instrument should at least have two categories ($C_s \geq 2$) with different category means for the LIV model to be identified (Ebbes et al. 2005). A C_s-dimensional vector of category indicators, $z_s^{i\prime}$, is assumed to have a multinomial distribution with parameters $\left(n = 1,\ \pi = \left(\pi_1, \ldots, \pi_{C_s}\right)\right)$ where π_c is the probability that the c^{th} latent instrument is 1 and the rest are 0, and $\sum_{c=1}^{C_s}\pi_c = 1$. This means that observation ij belongs to category c in terms of the endogenous variable s.

The endogenous part, ε_s, is assumed to be independent of the latent instrument, and normally distributed with mean zero and variance σ_s^2. The correlation between our potentially endogenous variables and the random shocks of the structural models is captured through covariance terms $\sigma_{p,s}$ where $p \in \{u, y, d\}$ and $s \in \{A, MR\}$. This enables us to obtain unbiased estimates of the impact of our covariates in each of the models. Utilizing the normality assumption of the structural shocks and the endogenous LIV error terms, we can represent the unobserved error structure of the model as following:

$$\begin{pmatrix} \varepsilon_u^{ij} \\ \varepsilon_y^{ij} \\ \varepsilon_d^{ij} \\ \varepsilon_A^{ij} \\ \varepsilon_{MR}^{ij} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_u^2 & \sigma_{u,y} & \sigma_{u,d} & \sigma_{u,A} & \sigma_{u,MR} \\ \sigma_{u,y} & \sigma_y^2 & \sigma_{y,d} & \sigma_{y,A} & \sigma_{y,MR} \\ \sigma_{u,d} & \sigma_{y,d} & \sigma_d^2 & \sigma_{d,A} & \sigma_{d,MR} \\ \sigma_{u,A} & \sigma_{y,A} & \sigma_{d,A} & \sigma_A^2 & \sigma_{A,MR} \\ \sigma_{u,MR} & \sigma_{y,MR} & \sigma_{d,MR} & \sigma_{A,MR} & \sigma_{MR}^2 \end{pmatrix}\right), \text{ where } i = 1, \cdots, n_b \text{ and } j = 1, \cdots, n_{b_i} \quad (8)$$

The elements of this covariance matrix will be estimated along with other parameters through our Bayesian MCMC procedure.

2.4. Modeling Reputation

In this section, we present our approach to model the dynamic reputation building of the borrowers and the effect of reputation on the future performance of the borrowers' listings and the borrowers' likelihood to default. We provide two approaches, a hidden Markov model in which a borrower's latent state of reputation in a period directly depends on his or her previous reputation through a state-dependent transition probability matrix. Our second model is a dynamic latent class model of reputation in which the current state of reputation does not directly depend on the previous states, and there is, rather, a single regime determining the borrowers' reputation at any time. In each of these two models, the reputation model works as an underlying model that sits underneath our structural models and determines the cross-sections of the data in a two-level hierarchical model. The coefficients of each of the first level structural models depend upon the reputation state. We also estimate a benchmark model in which we simply include the variables of our reputation model as covariates in each of the three structural models. We then estimate each of the three reputation models and compare them in terms of their fit.

2.4.1. The Hidden Markov Model

Hidden Markov models (HMMs) are stochastic models in which the distribution that generates an observation depends on the state of an underlying Markov process which is not directly observable (Zucchini and MacDonald 2009). In an HMM, each agent is in one of a set of R distinct states at any time. Over time, the agents could move between states. Although the states are not observable, there are observations which are probabilistic functions of the states (Rabiner 1989). Our main goal in this study is to characterize the evolution and impact of reputation which is by definition a construct that is not directly observable. There might exist some carryover effect in a borrower's reputation over time, meaning that a borrower's reputation at an occasion is dependent on his or her reputation at the previous occasion. HMMs allow for the probability distribution of each observation to depend on the hidden state of a Markov chain and can accommodate serial dependence (Zucchini and MacDonald 2009). A byproduct of the hidden Markov approach is a segmentation of the borrowers in terms of their state of reputation (Singh et al. 2011). In other words, borrowers at different levels of reputation are hypothesized to (i) behave differently in terms of defaulting, and (ii) be treated differently by the lenders.

Furthermore, the HMM allows us to include more information in estimating the model with a structured way of interpreting the results, and yet bypassing multicollinearity issues.

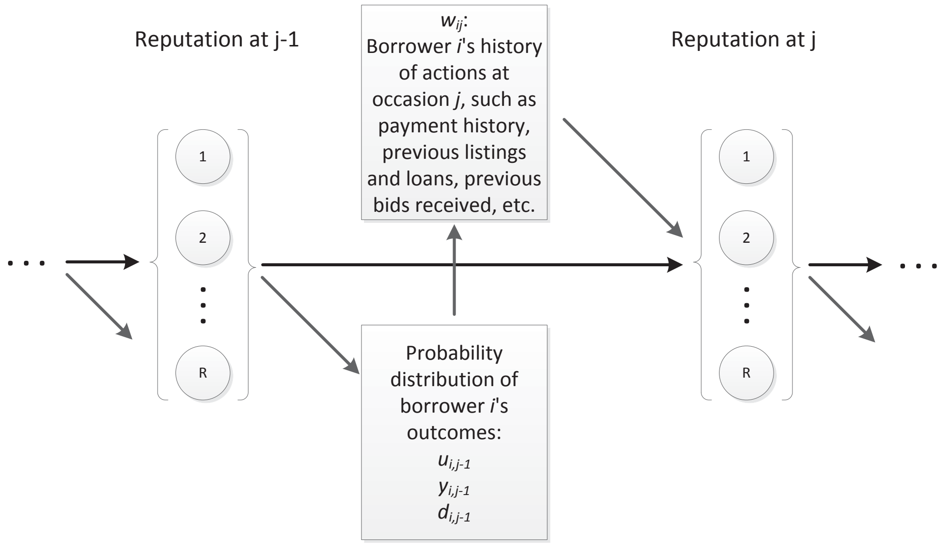

Figure 2: The Hidden Markov Model of Reputation

In our hidden Markov model of reputation, the effects of the covariates on the endogenous variables in structural models are moderated by the borrowers' latent state of reputation in the P2P lending community. The reputation is affected by the borrowers' history of decisions and actions within the community. Reputation is modeled as a dynamic latent discrete state, and there are a total of R hidden states. The borrowers' reputation level changes stochastically as a result of the new actions they take. Reputation levels are not directly observable, and we only observe their impact on the outcomes of interest. Superscripts r for the coefficients in (1), (3) and (4) allow for different coefficients for different reputation levels. Our HMM captures the dynamic evolution of reputation and the impact of reputation on the outcomes of interest. The outcomes at any time period depend on the borrower's reputation only through the borrower's reputation at that given period. A borrower's reputation at any period depends on his or her reputation in the past periods only through the borrower's reputation at the previous period.

An HMM has two main components: a finite set of states, and the observed outcomes. To characterize and estimate an HMM, we have to determine (i) the number of states, (ii) the initial state probability distribution, (iii) the state transition probability distribution, and (iv) the probability distribution of the observed outcomes (Rabiner 1989). The number of states has to be fixed before estimating the model, and we assume that all borrowers are at the lowest reputation state ($r=1$) at their first participation. In other words, the probability of initially being at $r=1$ is 1, and the probability of being at any other state is 0 for all the borrowers.

Reputation is an ordinal discrete variable, meaning that among the R reputation levels, R represents the best reputation, while 1 is the lowest level of reputation. Therefore in order to model the dynamic evolution of reputation as a function of variables of interest, we use an ordered probit model. For the reasons of parsimony and simplicity, the evolution of reputation is assumed to follow a random walk meaning that the reputation state could move at most one level at a time. In other words, if borrower i has reputation level $1<r<R$ at occasion $j-1$, at occasion j we will have:

$$Prob(Reputation_{ij}=r' \mid Reputation_{i,j-1}=r)=\begin{cases} p_{ij}^{rr'}>0, & r'=r,r-1,r+1; \\ 0, & otherwise. \end{cases} \tag{9}$$

The random walk assumption can easily be relaxed by allowing the transition probabilities to the non-adjacent states to be non-zero. Borrowers' discrete state of reputation changes dynamically as their underlying continuous stock of reputation changes. The continuous stock of reputation is composed of two parts; an observable part which includes covariates representing a borrower's history of actions, and an unobservable mean-zero random part which is normally distributed. The level and scale of the stock of reputation are not identified. Therefore, we do not include a constant in the observable part, and fix the variance of the random part to 1. There are two thresholds governing the transition at each state of reputation. The discrete state of reputation at an occasion is determined according to the size of the stock of reputation as compared to the two thresholds, therefore we have:

$$Reputation_{ij} \mid w_{ij}, \left\{Reputation_{i,j-1} = r\right\} = \begin{cases} r+1, & \bar{\mu}_r \le \mathrm{w}_{ij}{}'\lambda_r + \omega_{ij}; \\ r, & \underline{\mu}_r < \mathrm{w}_{ij}{}'\lambda_r + \omega_{ij} < \bar{\mu}_r; \\ r-1, & \mathrm{w}_{ij}{}'\lambda_r + \omega_{ij} \le \underline{\mu}_r. \end{cases} \tag{10}$$

We have $\omega_{ij} \sim N(0,1)$, $\underline{\mu}_1 = -\infty$ and $\bar{\mu}_R = \infty$ as the reputation states are limited to $\{1,\cdots,R\}$. w_{ij} is an $n_w \times 1$ vector of reputation covariates and λ_r is an $n_w \times 1$ vector of coefficients when the borrower is at reputation level r. There are a total of $2(R-1)$ threshold parameters to be estimated. Based on the threshold model in (10) and the normality assumption of the unobservable random term, the transition probabilities for borrower i in reputation state r at his or her $(j-1)^{th}$ listing to his or her j^{th} listing are:

$$\begin{aligned} p_{ij}^{r,r+1} &= 1 - \Phi\left(\bar{\mu}_r - \mathrm{w}_{ij}{}'\lambda_r\right); \\ p_{ij}^{r,r} &= \Phi\left(\bar{\mu}_r - \mathrm{w}_{ij}{}'\lambda_r\right) - \Phi\left(\underline{\mu}_r - \mathrm{w}_{ij}{}'\lambda_r\right); \\ p_{ij}^{r,r-1} &= \Phi\left(\underline{\mu}_r - \mathrm{w}_{ij}{}'\lambda_r\right). \end{aligned} \tag{11}$$

Φ is the cdf of the standard normal distribution. Therefore the transition probability matrix for borrower i moving from his or her $(j-1)^{th}$ listing to j^{th} listing is:

$$P_{ij} = \begin{bmatrix} p_{ij}^{1,1} & p_{ij}^{1,2} & 0 & 0 & \cdots & 0 \\ p_{ij}^{2,1} & p_{ij}^{2,2} & p_{ij}^{2,3} & 0 & \cdots & 0 \\ 0 & p_{ij}^{3,2} & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & p_{ij}^{R-2,R-2} & p_{ij}^{R-2,R-1} & 0 \\ \vdots & \vdots & 0 & p_{ij}^{R-1,R-2} & p_{ij}^{R-1,R-1} & p_{ij}^{R-1,R} \\ 0 & 0 & \cdots & 0 & p_{ij}^{R,R-1} & p_{ij}^{R,R} \end{bmatrix}. \tag{12}$$

If borrower i in his or her j^{th} listing is in reputation state $r \in \{1, \cdots, R\}$, we can characterize his or her observed outcomes as following, based on (1), (3) and (4):

$$\Pr(u_{ij} = 1) = \frac{\exp\left(x_{ij}{}'\beta_u^r + A_{ij}\alpha_u^r + MR_{ij}\gamma_u^r + \xi_u^i + \varepsilon_u^{ij}\right)}{1 + \exp\left(x_{ij}{}'\beta_u^r + A_{ij}\alpha_u^r + MR_{ij}\gamma_u^r + \xi_u^i + \varepsilon_u^{ij}\right)}; \tag{13}$$

$$y_{ij} = x_{ij}{}'\beta_y^r + A_{ij}\alpha_y^r + MR_{ij}\gamma_y^r + \xi_y^i + \varepsilon_y^{ij}; \tag{14}$$

$$\Pr(d_{ij} = 1 \mid u_{ij} = 1) = \frac{\exp\left(x_{ij}{}'\beta_d^r + A_{ij}\alpha_d^r + MR_{ij}\gamma_d^r + \xi_d^i + \varepsilon_d^{ij}\right)}{1 + \exp\left(x_{ij}{}'\beta_d^r + A_{ij}\alpha_d^r + MR_{ij}\gamma_d^r + \xi_d^i + \varepsilon_d^{ij}\right)}. \tag{15}$$

2.4.2. The Dynamic Latent Class Model

Our dynamic latent class model of reputation is in spirit similar to our nonhomogeneous HMM and is essentially a constrained HMM. In both models, there is a finite number of states of reputation in which a borrower can reside at any point in time. In both models, the reputation states are not directly observable, but they can be observed through their impact on a set of observable outcomes. Both approaches effectively segment the borrowers at any time into R distinct groups based on their reputation. The state of reputation in both models depends upon the borrowers' stock of reputation at the time, which is a function of the history of actions and decisions made by them up to that time. The difference between the two models is that in our dynamic latent class model, the current reputation state of a borrower depends on the previous reputation states only indirectly through the cumulative effect of the history of the borrower's actions. In other words, as opposed to the HMM, here the transition probability of the states is conditionally independent of the prior states:

$$Prob\left(Reputation_{ij} = r' \mid w_{ij}, \left\{Reputation_{i,j-1} = r\right\}\right) = Prob\left(Reputation_{ij} = r' \mid w_{ij}\right). \tag{16}$$

To characterize the latent class model, we use an ordered probit model similar to our HMM. The difference is that as opposed to the HMM for which we had R different ordered probit models for each state, there is a single model with $R-1$ thresholds that governs the dynamic evolution of reputation. Therefore, we have:

$$Reputation_{ij} \mid w_{ij} = \begin{cases} R, & \mu_{R-1}^{LC} < \mathrm{w}_{ij}'\lambda^{LC} + \omega_{ij}; \\ r \in \{2,\ldots,R-1\} & \mu_{r-1}^{LC} < \mathrm{w}_{ij}'\lambda^{LC} + \omega_{ij} \le \mu_{r}^{LC}; \\ 1, & \mathrm{w}_{ij}'\lambda^{LC} + \omega_{ij} \le \mu_{1}^{LC}. \end{cases} \tag{17}$$

It is worth noting that unlike the HMM, here the vector of coefficients of the underlying stock of reputation, λ^{LC}, does not have an r subscript. This is because the coefficients of the ordered probit model are not state-dependent anymore. Similar to our HMM, $\omega_{ij} \sim N(0,1)$ are iid random shocks. Given this specification, we can write the probability of borrower i being at reputation state r in his or her j^{th} occasion as:

$$p_{ij}^{r} = \Phi\left(\mu_{r}^{LC} - \mathrm{w}_{ij}'\lambda^{LC}\right) - \Phi\left(\mu_{r-1}^{LC} - \mathrm{w}_{ij}'\lambda^{LC}\right) \tag{18}$$

We have $r \in \{1,\cdots,R\}$, $\mu_{0}^{LC} = -\infty$ and $\mu_{R}^{LC} = +\infty$. The moderating impact of the reputation states on the observed outcomes is modeled similar to our HMM through a two-level hierarchical model. Therefore equations (13), (14) and (15) are unchanged in this model. The dynamic latent class model is more parsimonious compared to the HMM, as instead of estimating $2\times(R-1)$ thresholds and $r \times n_{w}$ coefficients, we only estimate $R-1$ thresholds and n_{w} coefficients. The difference in the number of parameters to be estimated in HMM versus the dynamic latent class model could particularly be large when we have a larger number of reputation states. The HMM instead provides a richer specification and directly models the interdependence of the states. We will later estimate both models and compare their goodness of fit.

2.5. Data

The data used in this study was collected from one of the largest online peer-to-peer lending websites in the United States. Starting its operations in 2000s, this P2P lending website provides a platform on which borrowers can request to borrow money from lenders who are willing to invest in personal loans. The community has more than one million members, and more than 350 million dollars in have been were issued within the community.

The P2P platform owner is in charge of receiving the money from the borrowers and dividing it between the lenders. In the event of late payments or defaulting by the borrowers, the P2P platform owner attempts to collect late payments on lenders' behalf through various collection practices (herzenstein et al. 2008). Serious delinquency could result in the account being sent to collections. The platform owner reports the borrowers' activities to major credit bureaus, so late or missed payments may damage borrowers' credit scores and make it more difficult for them to get a loan in the future (Steinisch 2012). The P2P platform earns money through charging a transaction fee on the listings that are fully funded and successfully transformed into loans. The borrowers are charged a one-time origination fee as a percentage of the amount of the loan when a loan is successfully initiated and the lenders pay an annual loan servicing fee as a percentage of the outstanding principal balance of the corresponding loans.

On this P2P lending community, there is a good deal of information available to the lenders when they want to make decisions on investing on a loan and submitting a bid. Other than the listing information such as the amount, loan category and maximum interest rate, lenders observe the borrower's financial information such as credit grade and debt to income ratio. Moreover the history of all of the prior activities of the borrowers including previous listings, successful loans, and the borrowers' financial data in each of those, and payment history are also available to the lenders to help them in decision making. The availability of this information is what streamlines reputation formation.

Our data set includes complete data on all activities, i.e., the listings, bids, loans and members, from the early 2006 to the end of April 2009. We also supplemented the data set with payment data until September 2012 for the loans created in the 2006-2009 period. This provides us with around 37 months of data on all the listings, loans created, and bids submitted by the lenders.

For the purpose of this study, we only considered borrowers who have had at least three of their listings successfully funded and converted into loans. This is done because the notion of reputation building and being able to model and capture it heavily relies on having a reasonable amount of prior activities and available information. So unless historical information is available, reputation effect cannot be identified. Also the size of the data is very large making it computationally prohibitive to conduct an analysis on the whole data set. This criterion narrowed down our dataset to 3,201 listings from 365 unique borrowers. Table 1 shows the listing-level variables and constructs studied in this research accompanied by a short description, and their means and standard deviations for our selected data. The covariates included in our three structural equations, A, MR and the matrix X in equations (1), (3) and (4) are selected from these variables. Table 2 provides the same information on our reputation level variables out of which we have selected the covariates to be included in our underlying model of reputation stock, W, in (11) and (18). Subscripts i for the borrower and j for the listing count number are dropped to keep the table clean, but it is important to note that the values of the variables might change across borrowers and time.

Table 1: Description and Summary Statistics of the Listing Variables

Variable Name	Description	Mean	SD
Amount (A)	Amount requested by the borrower in the listing	$6450	$5270
MR	Maximum interest rate posted by the borrower in the listing	21%	9%
FundingOption	A dummy which is 1 when the auction is open for the duration and 0 if it is closed when fully funded.	0.79	0.41
CreditGrade	Borrowers' credit grade as an ordinal variable ranging from 0 for no credit available to 7 for AA credit	3.79	1.66
ListCharacters	Total number of characters in the description of the listing	1147	765
ListImage	A dummy variable which is 1 if the borrower has at least one picture posted in the listing	0.63	0.48
ListDuration	Duration of the listing auction, ranging between 3 and 10 days	8.22	2.68
FriendsBids	Number of bids by the borrower's friends in a listing	0.07	0.26
NumQ&A	Number of questions and answers in the listing	0.46	1.01
NumDelin	Number of current delinquencies of the borrower	1.38	3.03
Home Owner	A dummy variable which is 1 if the borrower is a home owner.	0.41	0.49
DebtToIncome	Debt to income ratio	0.31	0.63

Table 2: Description and Summary Statistics of the Reputation Variables

Variable Name	Description	Mean	SD
Lender	Whether the borrower is also registered as a lender	0.56	0.49
GroupMember	A dummy which is 1 if the borrower is a member of a group	0.41	0.49
MemberDescription	Total number of characters in the member's profile description	404	538
CumBids2ListRatio	Ratio of cumulative number of bids to cumulative number of listings for a borrower	31.85	53
CumNumLists	Cumulative number of listings that the borrower has posted before	6.55	8.70
CumLoans	The cumulative number of loans that the borrower has had	1.13	.92
CumWithdrawals	The cumulative number of listings that the borrower has withdrawn	4.04	7.80
CumFriendsBids	Cumulative number of bids by the borrower's friends in prior listings	0.06	0.44
CumQ&A	Cumulative number of questions and answers in a borrower's prior listings	1.74	3.82
TimeSinceJoin	The logarithm of the number of days since joining the lending community	450	398
DeltaCredit	Change in the credit grading since last posting	0.08	0.63
Friends	Total number of friends of the borrower	0.99	1.59
Endorsed	Number of endorsements a borrower has received	0.29	0.73
Endorser	Number of people a borrower has endorsed.	0.24	1.91
CumBorrowed	Cumulative amount borrowed by the borrower	$ 5231	4673
Outstanding	The amount of principal outstanding owed by the borrower	$ 1497	1986
Payments	Total number of on-time payments by the borrower so far	12.44	13.72
LatePay	Total number of late payments by the borrower so far	0.36	1.61

2.6. Estimation

We utilize an empirical Bayesian approach and estimate the model using a Markov Chain Monte Carlo (MCMC) estimator with multiple Gibbs and Metropolis steps. Both the LIV method and the hidden Markov model are likelihood-based approaches, so they are amenable to MCMC estimation. A benefit of the Bayesian approach is that it allows us to get draws of the parameters of the model, whereas in the frequentist approach, only point estimates are attained. From the practitioners' point of view, the Bayesian approach allows the platform designers to estimate the distribution of the parameters for each of the borrowers separately. This provides a flexible random component specification that allows us to incorporate both observable and unobserved borrower heterogeneity (Agarwal et al. 2011).

One of the challenges with the Bayesian analysis is deciding on the prior distribution of the parameters, as in many applications, ours included, no useful prior information exists. This issue is more highlighted in the context of hierarchical models when the number of parameters is larger and their interpretation is more complex. One way to deal with this is using empirical Bayes in which the existing dataset is used to get the prior distribution parameters. They are made as uninformative as possible such that the weight of the prior is minimal in determining the posterior distribution. In this study, as we shall see, we use empirical Bayes and in particular a class of empirical Bayes priors that are called g-priors for our second-level coefficients (Hoff 2009).

Although the joint posterior distribution of the model is non-standard, we use an iterative algorithm that constructs a dependent sequence of parameter values whose distribution converges to the parameters' joint posterior distribution (Hoff 2009). At each step, we form the full conditional posterior distribution of a group of parameters by conditioning on the current values of all other parameters that are being estimated.

Following are the steps used in estimating the dynamic latent class model. The MCMC algorithms used to estimate the hidden Markov model and the benchmark model are modifications of this algorithm.

1. Draw β_p^r, where $r \in \{1,..,R\}$ and $p \in \{u,y,d\}$. β_p^r is an n_x vector of coefficients of equation p for a borrower who is at reputation state r.
    1.1. We use a Gibbs step to generate β_y^r. μ_{β_y} and Σ_{β_y} are the second-level within-sample mean and covariance of the coefficients.

$$\beta_y^r \mid y_r, X_r, \sigma_y^2, \mu_{\beta_y}, \Sigma_{\beta_y}, R \sim MVN\left(\tilde{\mu}_{\beta_y^r}, \tilde{\Sigma}_{\beta_y^r}\right), \text{ where } \tilde{\Sigma}_{\beta_r^y} = \left[\left(\Sigma_{\beta_y}\right)^{-1} + \frac{X_r{}'X_r}{\sigma_y^2 - \theta_{\beta_y}}\right]^{-1},$$

$$\tilde{\mu}_{\beta_r^y} = \tilde{\Sigma}_{\beta_y^r}\left[\left(\Sigma_{\beta_y}\right)^{-1}\mu_{\beta_y} + X_r{}'y_r / \sigma_y^2\right]$$

$$\theta_{\beta_y} = \begin{pmatrix}\sigma_{uy} & \sigma_{dy} & \sigma_{Ay} & \sigma_{MR,y}\end{pmatrix}\begin{pmatrix}\sigma_u{}^2 & \sigma_{ud} & \sigma_{uA} & \sigma_{uMR} \\ \sigma_{ud} & \sigma_d{}^2 & \sigma_{dA} & \sigma_{dMR} \\ \sigma_{uA} & \sigma_{dA} & \sigma_A{}^2 & \sigma_{A,MR} \\ \sigma_{uMR} & \sigma_{dMR} & \sigma_{A,MR} & \sigma_{MR}{}^2\end{pmatrix}\begin{pmatrix}\sigma_{uy} \\ \sigma_{dy} \\ \sigma_{Ay} \\ \sigma_{MR,y}\end{pmatrix}$$

    1.2. In order to generate β_u^r and β_d^r, We use a random walk Metropolis-Hastings step (Rossi et al. 2005) with the log likelihood function of (2) combined with the second-level between-sample mean and covariance of the coefficients.

$$\beta_p^r \mid y_{r,p}, X_{r,p}, \sigma_p^2, \mu_{\beta_p}, \Sigma_{\beta_p} \propto \left(LL_p^r\right) + log\left(\beta_y^r \sim MVN(\mu_{\beta_p}, \Sigma_{\beta_p})\right), \text{ where } r \in \{1,..,R\} \text{ and } p \in \{u,d\}$$

2. Draw the first-level between-sample mean and covariance of the coefficients, μ_{β_p} and Σ_{β_p} for model p where $p \in \{u, y, d\}$. We use a Gibbs step with diffuse conjugate normal priors for the means and inverse Wishart priors for the covariances. For priors, we use the one-level model estimates of means and covariances.

$$\mu_{\beta_p} \mid \beta_p^1, \ldots, \beta_p^R, \Sigma_{\beta_p} \sim MVN\left(\tilde{\mu}_{\beta_p}, \tilde{\tilde{\Sigma}}_{\beta_p}\right), \text{ where } \tilde{\tilde{\Sigma}}_{\beta_p} = \left[\left(\Sigma_{\beta_p}^0\right)^{-1} + R\left(\Sigma_{\beta_p}\right)^{-1}\right]^{-1},$$

$$\tilde{\mu}_{\beta_r^y} = \tilde{\tilde{\Sigma}}_{\beta_p}\left[\left(\Sigma_{\beta_p}^0\right)^{-1} \mu_{\beta_p}^0 + R\left(\Sigma_{\beta_p}\right)^{-1} \bar{\beta}_p\right]$$

$$\Sigma_{\beta_p} \mid \sim IW\left[\eta_0^\Sigma + R, \left(S_0^\Sigma + \sum_{r=1}^R \left(\beta_p^r - \mu_{\beta_p}\right)\left(\beta_p^r - \mu_{\beta_p}\right)'\right)^{-1}\right]$$

3. Generate the normal error terms, ε_q, where $q \in \{u, y, d, A, MR\}$ in each of the three main models and the two LIV equations.

   3.1. The error terms for the interest rate reduction equation, ε_u is calculated as:

   $\varepsilon_y^{ij} = y_{ij} - \beta_y^{r\prime} x_{ij} -, \xi_y^i$, where ε_y^{rj} is the ij^{th} element of the error vector and $r \in \{1, \ldots, R\}$, $j \in \{1, \ldots n_r\}$

   3.2. The endogenous error terms for the Amount LIV equation, ε_A, and the maximum interest rate LIV equation, ε_{MR}, are generated as follows:

$$\varepsilon_A = A - Z_A \varphi_A, \quad \varepsilon_{MR} = MR - Z_{MR} \varphi_{MR}$$

   3.3. We use a random-walk Metropolis-Hastings step to generate ε_u and ε_d, based on the log likelihoods as given in (2), and a multivariate normal sampling distribution which acts like a prior:

$$\begin{pmatrix} \varepsilon_u^{rj} \\ \varepsilon_d^{rj} \end{pmatrix} | \varepsilon_y, \varepsilon_A, \varepsilon_{MR}, \Lambda \propto \left(LL_u\right) + \left(LL_d\right) + log\left[\begin{pmatrix} \varepsilon_u^{rj} & \varepsilon_y^{rj} & \varepsilon_d^{rj} & \varepsilon_A^{rj} & \varepsilon_{MR}^{rj} \end{pmatrix}' \sim MVN\left(0, \Lambda\right)\right],$$

where $r \in \{1,...,R\}$, $j \in \left\{1,...n_r^p\right\}$

4. Draw the 5x5 model covariance matrix. We do this in two steps. We first draw the covariance matrix of ε_u, ε_y, ε_A and ε_{MR} for each of which we have N values. Then we draw the variance of ε_d and its covariance with other variables, as we only have N_d values of it. We use a diffuse conjugate inverse Wishart prior. The draw for the second step is:

$$\Lambda | E \sim IW\left[\eta_0^\Lambda + N_d, \left(S_0^\Lambda + \mathrm{E}'\mathrm{E}\right)^{-1}\right], \text{ where } \mathrm{E} = \left(\varepsilon_u, \varepsilon_y, \varepsilon_d, \varepsilon_A, \varepsilon_{MR}\right)$$

5. Draw Category memberships for each of the two endogenous variables. We generate an $N \times C_A$ matrix, Z_A, for the amounts and an $N \times C_{MR}$ matrix, Z_{MR}, for the maximum interest rates as two categorical variables with the posterior probability given below (Rutz et al. 2012):

$$\Pr(Z_s^i = c_s) = \frac{L\left(s, Z_s^i(c_s), \varphi_s, \mathrm{E}, \Lambda\right) \times \pi_{c_s}}{\sum_{k=1}^{C_s} L\left(s, Z_k^i(c_k), \varphi_k, \mathrm{E}, \Lambda\right) \times \pi_k}, \text{ where } i = 1,...,N,\ s \in \{A, MR\}$$

$Z_s^i = c_s$ and $Z_s^i(c_s)$ for the i^{th} observation mean that the element on the $c_s{}^{th}$ column of row i of Z_s is 1 and the rest are 0. π_{c_s} is the sampling probability of membership in category c_s which acts like a prior and $L\left(s, Z_s^i(c_s), \varphi_s, \mathrm{E}, \Lambda\right)$ is the likelihood of the whole LIV model evaluated at $Z_s^i = c_s$.

6. Generate category membership probabilities for the two latent IVs, π_A and π_{MR}, using a Dirichlet draw as following (Rutz et al. 2012):

$$\pi_s \mid Z_s \sim Dirichlet\left(1+\mathrm{K}_1,\ldots,1+\mathrm{K}_{C_s}\right),\ s \in \{A, MR\}$$

K_c denotes the number of observations for the latent IV s that belong to category c, i.e., $\mathrm{K}_c = \sum_{i=1}^{N} Z_s^i \left(Z_s^i = c\right)$.

7. Generate the discrete set of means for the two latent IV's using a Gibbs step similar to (Rutz et al. 2012) with a diffuse multivariate normal prior. These means are essentially the coefficients in a linear regression of the endogenous variable on the category memberships. We use the quantiles of the endogenous variables as the prior means.

$$\varphi_s \mid s, Z_s, \Lambda \sim MVN\left(\tilde{\mu}_{\varphi_s}, \tilde{\Sigma}_{\varphi_s}\right), \text{ where } s \in \{A, MR\},\ \tilde{\Sigma}_{\varphi_s} = \left(\left(\Sigma_0^{\varphi_s}\right)^{-1} + \frac{1}{\sigma_s^{\,2} - \theta_s} Z_A{}' Z_A\right)^{-1},$$

$$\tilde{\mu}_{\varphi_s} = \tilde{\Sigma}_{\varphi_s}\left(\left(\Sigma_0^{\varphi_s}\right)^{-1} \mu_0^{\varphi_s} + \frac{1}{\sigma_s^{\,2} - \theta_s} Z_A{}' A\right)^{-1},$$

$$\text{and, e.g., for } s = A,\ \theta_A = \begin{pmatrix} \sigma_{uA} & \sigma_{yA} & \sigma_{dA} & \sigma_{A,MR} \end{pmatrix} \begin{pmatrix} \sigma_u^{\,2} & \sigma_{uy} & \sigma_{ud} & \sigma_{uMR} \\ \sigma_{uy} & \sigma_y^{\,2} & \sigma_{yd} & \sigma_{yMR} \\ \sigma_{ud} & \sigma_{yd} & \sigma_d^{\,2} & \sigma_{dMR} \\ \sigma_{uMR} & \sigma_{yMR} & \sigma_{dMR} & \sigma_{MR}^{\,2} \end{pmatrix} \begin{pmatrix} \sigma_{uA} \\ \sigma_{yA} \\ \sigma_{dA} \\ \sigma_{A,MR} \end{pmatrix}$$

8. Draw the coefficients for the equation determining the underlying stock of reputation. Conditioning on the previous draws of the latent continuous reputation stocks and using a diffuse normal g-prior (Hoff 2009; Zellner 1986) with mean zero for the coefficients, we draw the coefficients similar to a regular draw for a linear regression model. As the scale of the reputation equation is not identified, we fix it to 1.

$$\beta_R \mid W, Rep^* \sim MVN\left(\tilde{\mu}_{\beta_R}, \tilde{\Sigma}_{\beta_R}\right), \text{ where } \tilde{\Sigma}_{\beta_R} = \frac{N}{N+1}\left(W'W\right)^{-1} \text{ and } \tilde{\mu}_{\beta_R} = \tilde{\Sigma}_{\beta_R} W' Rep^*$$

9. Draw the latent continuous reputation stocks. We use a data augmentation approach to generate the reputation stock of a given borrower at a given time, conditional on the current draw of reputation state, the reputation equation coefficients, and the current values of the thresholds. Given these draws, the reputation stock is truncated normally distributed.

$$Rep_i^* \mid w_i, \beta_R, r_i, \sim N\left(\beta_R' w_i, 1\right) \times I_{\left(\delta_L(r_i), \delta_H(r_i)\right)}\left(Rep_i^*\right), \text{ where } I_{(a,b)}\left(w\right) = \begin{cases} 1 & a \le w \le b \\ 0 & otherwise \end{cases}$$

Given r_i we know that Rep_i^* must lie between $\delta_L(r_i)$ and $\delta_H(r_i)$.

10. Draw the thresholds of the reputation ordered probit model. Assuming a diffuse uniform prior for the thresholds, we generate each threshold as a uniform draw in the range allowed based on the current values of reputation stock and reputation state.

$$\delta_k \mid Rep^*, R \sim U\left(\max\left\{Rep_i^* \mid r_i = k-1\right\}, \min\left\{Rep_i^* \mid r_i = k\right\}\right)$$

11. Draw the reputation states for each borrower at each listing occasion. We draw the state for each observation as a categorical variable such that the posterior probability of belonging to each category is calculated based on the posterior likelihood of each state. We are assuming a discrete uniform prior for the states.

$$\Pr(R_i = r \mid Rep^*, \beta_u, \beta_y, \beta_d, \varepsilon_u^i, \varepsilon_u^d, \sigma_y, \xi_p^i) = \frac{L_u\left(\beta_u^r, \varepsilon_u^i, \xi_u^i\right) L_y\left(\beta_y^r, \sigma_y, \xi_y^i\right) L_d\left(\beta_d^r, \varepsilon_d^i, \xi_d^i\right) \times \left(\Phi\left(\delta_{r+1} - Rep_i^*\right) - \Phi\left(\delta_r - Rep_i^*\right)\right)}{\sum_{k=1}^{R} L_u\left(\beta_u^k, \varepsilon_u^i, \xi_u^i\right) L_y\left(\beta_y^k, \sigma_y, \xi_y^i\right) L_d\left(\beta_d^k, \varepsilon_d^i, \xi_d^i\right) \times \left(\Phi\left(\delta_{k+1} - Rep_i^*\right) - \Phi\left(\delta_k - Rep_i^*\right)\right)},$$

where $i = 1, \ldots, N$, and Φ is the CDF of the standard normal distribution.

12. Draw the random effects for each borrower in each of the models. Conditioning on the current draws of the coefficients and the current values of the error terms, and using the sampling distribution of the random effects' variances that works similar to a prior, we

generate the random effects. We use Metropolis-Hastings steps for loan conversion and defaulting models. The draw for the interest rate model is done using a Gibbs step for the within-sample coefficients of a two-level model.

$$\xi_p^i \mid \beta_p, \sigma^2_{\xi_p}, Rep_i, \varepsilon_p^i \propto \left(LL_p\right) log\left(\xi_p^i \sim N\left(0, \sigma^2_{\xi_p}\right)\right),$$

where $p \in \{u, y, d\}$, and LL_p is the log likelihood function calculated according to equation (2).

$$\xi_p^i \mid \beta_p, \sigma^2_{\xi_p}, Rep_{ij} \sim N\left(\tilde{\mu}_{\xi_p}, \tilde{\sigma}^2_{\xi_p}\right), \text{ for } i = 1,..,n_b \text{ and } j = 1,..,n_{b_i},$$

where $\tilde{\sigma}^2_{\xi_p} = \left(1/\sigma^2_{\xi_p} + n_{b_i}/\sigma_p^2\right)^{-1}$ and $\tilde{\mu}_{\xi_p} = \tilde{\sigma}^2_{\xi_p} \sum_{j=1}^{n_{b_i}} \left(y_{ij} - \beta_p^{ij} x_{ij}\right)$, n_{b_i} is the number of observations for borrower i, p_{ij} is an element of p, x_{ij} is vector of covariates and β_p^{ij} is the state-dependent vector of coefficients.

13. Draw the variance of the borrower-level random effects term. We generate the variance for each model using a Gibbs step on the current draws of the random effects and a diffuse conjugate inverse gamma prior.

$$1/\sigma^2_{\xi_p} \mid \xi_p \sim gamma\left(\tilde{v}_p/2, \tilde{v}_p \tilde{\sigma}^2_{\xi_p}/2\right),$$

where $p \in \{u, y, d\}$ And $\tilde{v}_p = v_0 + n_b$ and $\tilde{\sigma}^2_{\xi_p} = \left(\tilde{v}_p\right)^{-1}\left(v_{0_b}\sigma^2_{0_b} + \sum_{i=1}^{n_b}\left(\xi_p^i\right)^2\right)$,

and $1/\sigma^2_{\xi_p} \sim gamma\left(\frac{v_{0_b}}{2}, \frac{v_{0_b}}{2}\sigma^2_{0_b}\right)$ is the prior with $v_{0_b} = 1$ and $\sigma^2_{0_b} = 1$.

2.7. Results and Discussion

In order to get initial values for the parameters, we first drew 50,000 samples for each of the three equations separately. Then we used the mean of the last 20,000 draws as initial values for the coefficients in each of the models. In order to estimate models M1 (the benchmark model), M2 (the HMM) and M3 (the dynamic latent class model), we implemented the MCMC code in R and for each model, ran the code for 250,000 iterations. We only recorded every 10th draw to mitigate the autocorrelation issue which is a common consequence of the MCMC simulation (Hoff 2009). To insure that the effect of initial values has dissipated, out of the 25,000 resulting samples, we discarded the first 5,000 as the burn-in samples and used the remaining 20,000 samples to draw inference. In order to reduce correlation, and also to make the comparison of the estimated coefficients easier, all of the variables except for dummies and count variables are mean-centered and scaled by dividing them by their standard deviation. As our data spans several months, in order to control for the effects of the state of the economy, we include the daily Dow Jones Industrial Average[1] data in our models.

We estimated a model with three simultaneous equations and no explicit reputation component, an HMM and a dynamic latent class model. For each of these models, we calculated DIC, the deviance information criteria (Spiegelhalter et al. 2002). DIC is a measure of fit for Bayesian model estimation that values both fit and parsimony by penalizing models that have a larger number of effective parameters. In each of the models, we use latent IVs with the same number of categories to account for potential endogeneity. As we see in Table 3, our latent class model, M3, with 3 states (the best latent class model in terms of fit) outperforms the benchmark model and the HMM with 3 states (the best HMM).

---

[1] http://www.djaverages.com/

The HMM has more effective parameters, so is naturally expected to provide a better fit. The improvement in fit, however, is not sufficient to warrant the additional parameters. This lends support to the notion that the lenders' perception of a borrower's reputation is determined solely through the available information on the borrower, and the stochastic updating of the reputation does not explicitly depend on the current state of reputation. This is reasonable, as there are a large number of lenders in the community, each of which evaluates the given borrower's reputation at a given time using a fixed set of standards, regardless of the borrower's reputation in the prior period.

Table 3: Model Comparison

Model	***DIC***
M1: The benchmark model with no explicit reputation component	67,460
M2: The hidden Markov model of reputation with 3 states	62,562
M3: The dynamic latent class model of reputation with 3 states	**61,642**

From now on, we focus on the dynamic latent class model of reputation, and report the estimation results for this model. In order to estimate a latent class model, we need to determine the number of states as an input to the model. Table 4 shows the deviance information criterion for dynamic latent class models with 2, 3 and 4 reputation states. The model with 3 reputation states provides the best comparative fit as it has the lowest DIC among the three models. The three reputation states will hence be labeled as low, medium and high reputation levels.

Table 4: Comparison of Latent Class Models for Different Number of States

R	*DIC*
2	62,151
3	**61,642**
4	61,950

LIV approach is generally robust to the specification of the number of categories as long as there are at least two categories (Ebbes et al. 2005; Rutz et al. 2012). However, in order to decide on the number of categories for each of the latent instrumental variables, we estimated our three-state dynamic latent class model with combinations of 2, 3 and 4 categories for each of the latent IVs. As we can see in Table 5, the model with 3 categories for the listing amount LIV and 2 categories for the maximum interest rate LIV outperforms the other models.

Table 5: Comparison of LIV Latent Class Models for Different Number of Categories

C_A	C_{MR}	*DIC*
2	2	61,881
2	3	61,973
3	**2**	**61,642**
4	2	61,987
3	4	62,157

After deciding on the number of reputation states, and the number of latent categories for each of the instruments, we estimate the model and provide the results in Tables 6 to 11. Table 6 shows the posterior means of the coefficients of the logit model of loan conversion. Credible intervals (Edwards et al. 1963) are calculated as the empirical quantiles of the drawn samples. Each column represents the coefficients for each state of reputation. The coefficients have logit scale, so a negative coefficient represents a lower chance of conversion as the variable increases, while a positive coefficient represents a higher chance. The signs of the estimated coefficients are in general as expected and in accordance with prior research on P2P lending.

Careful inspection of the coefficients for each reputation state reveals some interesting patterns. The intercept for low-reputation borrowers is negative and significant, while the intercept for the high-reputation borrowers is positive and significant. It means that a low-reputation borrower's listing has a lower chance of getting funding, ceteris paribus. Also the coefficient for amount for low-reputation borrowers is negative and significant meaning that the listings that ask for a higher amount are less likely to be funded. However, interestingly, the opposite is true for the medium and high reputation borrowers.

Table 6: Loan Conversion Model Estimates (β_u)

	Reputation State		
	(1)	(2)	(3)
Intercept	-2.711***	-0.084	0.607***
Amount	-0.369***	0.700**	0.642***
MR	0.233*	0.042***	-0.212
MR^2	-1.100***	0.116	-0.086
FundingOption	0.036***	-1.396	-0.674*
CreditGrade	3.390***	-0.510	1.071*
ListCharacters	-0.179	-0.973	1.078*
ListImage	1.103*	1.032	3.020
ListDuration	-2.799*	-2.106	-2.636
FriendsBids	0.890***	0.102	0.024
NumQ&A	-1.508***	0.570*	0.197***
NumDelin	-0.098**	-0.126*	0.455
DJIA	0.890*	0.001	0.021

Notes: The numbers are the empirical means of the draws for each coefficient. * significant at 0.1; ** significant at 0.05; *** significant at 0.01 (empirical credible intervals)

This could be because lenders are more comfortable trusting their money with a borrower with higher reputation, and as they regard the listing as a good investment opportunity, they are even more willing to invest if the amount requested is higher. Whilst insignificant for medium and high reputation borrowers, the coefficient for inclusion of images in listings is positive and significant for low-reputation borrowers and improves their chances of getting a loan. One way to explain this is that due to the lack of further useful information, images could serve as signal that provides some information and creates a sense of trust toward the borrower (Duarte et al. 2012).

Table 7: Interest Rate Model Estimates (β_y)

	Reputation State		
	(1)	(2)	(3)
Intercept	-6.938^{***}	-0.195^{*}	0.975^{***}
Amount	-0.370^{***}	0.139	0.025
MR	0.023^{***}	0.06	-0.144
MR^2	0.235^{***}	0.053	-0.025^{*}
FundingOption	0.156^{***}	-1.893	-0.810
CreditGrade	0.817^{***}	-0.681	-4.557
ListCharacters	0.851	-0.618	0.386
ListImage	1.327^{*}	3.371	0.400
ListDuration	0.168^{***}	0.521^{***}	-1.141
FriendsBids	0.026^{*}	-0.041	0.122
NumQ&A	-1.703^{**}	-0.11^{*}	-0.041
NumDelin	-0.118^{*}	0.038	0.162
DJIA	0.004^{***}	-0.014	-0.049^{*}

Tables 7 and 8 present state-level posterior means of the coefficients for interest rate and defaulting models respectively. Trends similar to what we discussed for the loan conversion model could be spotted in these two tables. The intercepts in the interest rate model, for instance, are all significant, such that low-reputation borrowers have the least reduction in the final interest rate and the high-reputation borrowers have the highest. Also interestingly, credit grade and the number of delinquencies of the borrowers are only significant in the interest rate model for the low-reputation borrowers. Another surprising finding is the significant and negative sign of the coefficients for the number of questions and answers for low and medium reputation borrowers.

This could be because listings that receive more questions are listings that provide less information, so fewer lenders are willing to invest in them, resulting in higher interest rates.

Table 8: Defaulting Model Estimates (β_d)

	Reputation State		
	(1)	(2)	(3)
Intercept	0.106^{*}	-0.262^{*}	-0.227
Amount	0.060^{***}	0.246^{*}	-0.478^{***}
MR	0.203^{***}	-0.606^{*}	-0.07
MR^2	-1.304	-0.094	-0.383
FundingOption	-0.064^{*}	-0.985^{**}	1.884^{*}
CreditGrade	-7.684^{**}	-6.049^{***}	-24.65
ListCharacters	-1.043^{***}	-3.407	-9.733
ListImage	-1.108^{*}	3.143	-8.295
ListDuration	-11.28	-4.578	-0.731
FriendsBids	1.124^{**}	-0.711^{*}	-0.218
NumQ&A	-2.106^{**}	-0.525^{*}	-0.528^{***}
NumDelin	0.281^{***}	0.077^{*}	-0.111
DJIA	0.896^{*}	0.133	0.892

According to the estimates in Table 8, all else the same, low-reputation borrowers are more likely to default as the intercept for them is positive and significant in the logit model of defaulting. Higher amount loans with higher interest rates by the low-reputation borrowers are more likely to default. Conversely, for high-reputation borrowers, a higher loan amount on average has a negative impact on the chances of defaulting. As opposed to the interest rate model, a greater number of questions and answers is associated with lower chances of defaulting for borrowers at all reputation states. One explanation is that with more questions and answers, the lenders manage to acquire better information about the borrower's risk of defaulting on a loan. As expected, credit grade and the number of delinquencies are both strong predictors of the risk of defaulting.

Table 9 shows the posterior means and statistical significance of the ordered probit model of reputation evolution. Some insignificant variables are also included to provide insight. Ratio of cumulative number of bids to the number of listings, number of characters in the member's profile description, number of endorsements received by the borrower, borrower's experience as measured through the logarithm of the time since joining, the change in credit grade and the number of on-time payments all have positive and significant estimated coefficients. This means that a borrower with a larger value of these variables is likely to have a higher reputation. The estimated coefficient for the total number of late payments is negative and highly significant, meaning that late payments have a large adverse impact on borrowers' reputation. The estimated thresholds of the ordered probit model are also reported in Table 9.

Table 9: Reputation Model Estimates (λ^{LC})

Lender	0.004
GroupMember	0.013
MemberDescription	0.064^{**}
CumBids2ListRatio	0.051^{*}
Endorsed	0.110^{**}
TimeSinceJoin	0.118^{***}
CumLoans	0.138^{***}
CumWithdrawals	0.305^{***}
DeltaCredit	0.113^{**}
CumFriendsBids	0.153^{**}
CumQ&A	-0.012
CumBorrowed	0.037^{*}
Outstanding	0.017
Payments	0.103^{***}
LatePay	-0.560^{***}
μ_1	-0.272^{***}
μ_2	0.088^{*}

Table 10 reports the estimated covariance matrix for the joint distribution of the error terms of each of the three structural models and the endogenous variables. The estimated error variances are much smaller as compared to the model with no reputation component (M1), as more of the variation in data is explained by the dynamic latent class model. The significance of the structural covariance terms lends support to the joint estimation of the models, while significance of the covariance terms σ_{dA}, σ_{uMR} and σ_{uMR} provides support for the endogeneity of the two variables and the virtues of using the LIV approach. The significant and well-separated estimates of the LIV category means in Table 11 provide evidence that the LIV model is well-identified.

Table 10: Posterior Mean and Standard Deviation of the Error Covariance Matrix

	ε_u	ε_y	ε_d	ε_A	ε_{MR}
ε_u	0.90 (0.03)	-0.08 (0.06)	0.008 (0.003)	0.01 (0.04)	-0.01 (0.002)
ε_y		1.92 (0.54)	-0.002 (0.03)	0.005 (0.006)	-0.03 (0.01)
ε_d			1.13 (0.02)	0.004 (0.002)	0.01 (0.003)
ε_A				0.561 (0.073)	0.05 (0.02)
ε_{MR}					0.31 (0.02)

Table 11: LIV Parameter Estimates

	Amount	**Maximum Interest**
LIV category means (φ)	-0.524***	-0.629***
	0.087***	1.093***
	0.410***	
LIV category probabilities (π)	0.284***	0.615***
	0.444***	0.385***
	0.261***	

2.8. Conclusion

Technological advancements in web-based information systems in the last decade has given rise to Web 2.0 market mechanisms, where individuals take up active roles and their participation is key to survival of these markets. Many of these markets lean heavily towards C2C transactions where consumers act as buyers, sellers and even independent experts. Participants in these markets act based on their perception of trustworthiness or reputation of other parties and the risks involved in the associated transactions. In the online peer-to-peer lending community under study, borrowers are uniquely identified and the history of their activities within the lending platform is available to all registered lenders. So in contrast with many other e-commerce platforms, users cannot fake their identity and they are held accountable for their actions as these actions are directly traced back to them.

P2P lending platforms provide the lenders with the borrowers' hard credit information, similar to what financial institutions look at when considering a loan application. There is also additional information available as the borrowers' activities within the P2P lending community is recorded and is made available to the lenders to help them in decision making. This Information includes the details of the borrowers' prior loan requests, successful loans, loan payback information, social networks and the progression of the borrowers' hard credit information. In this study, we focused on the value created by this information which is made possible through the information systems deployed by the P2P lending platforms. In particular, we drew on the repeated games literature in economics and posited that borrowers and lenders are engaged in interactions that could be modeled as a repeated game. A borrower's type, i.e., his or her creditworthiness, is the private information of the borrower so information asymmetry is present. In order to mitigate the information asymmetry issue, rational lenders use the available information on the borrowers as a signal to infer the borrowers' types.

Therefore over time, as a borrower performs actions, the lenders' beliefs about him or her are updated. We refer to this form of identity as a borrower's reputation, and posit that a borrower's reputation affects his or her outcomes in the P2P community. In order to investigate this, we set up a system of three simultaneous equations to model a listing's conversion to loan, the final interest rate of a loan and the ex-ante probability of defaulting. We used the novel latent instrumental variable approach to account for potential endogeneity in the amount requested and the maximum interest rate posted by the borrowers. We propose two ways to model a borrower's reputation making, a hidden Markov model and a latent class model. The difference between the two is that the probability of moving from a state in the HMM is directly dependent on the current state.

We utilized a Markov chain Monte Carlo estimator with Gibbs and Metropolis steps and estimated the two models using a pseudo-hierarchical Bayesian approach. We also estimated a benchmark model with no explicit reputation component in which the reputation variables were included in the structural equations as covariates. We used DIC, the deviance information criterion (Spiegelhalter et al. 2002) to measure the goodness of fit of the models and found that the latent class model of reputation with three states outperforms others in explaining the data. We provided parameter estimates for each reputation state, and showed that our modeling approach provides a means to tease out the varying effects of covariates on the outcomes of interest for borrowers in different reputation states. One way to interpret the findings is to assume that the estimated coefficients for each state represent a model through which the outcomes for a borrower at that stage are decided. The lenders have a belief about a borrowers' private state. As more information about the borrower becomes available, this belief is updated and the borrower's reputation evolves. The outcomes of the borrower are then essentially determined by calculating expectations that are taken over the probability distribution of the borrower's type. These expectations are weighted averages of the outcomes for every state. Our model is able to recover the distribution of each borrower's perceived type at any time through its Bayesian draws of the borrower's type.

In this study, we proposed a novel empirical approach to measure the evolution of reputation and its impact on outcomes of interest in the peer-to-peer lending context. This approach is readily extendable to any other environment in which repeated transactions are performed and information signals are available resulting in reputation effects. Using this approach, we investigated and provided evidence for the intangible value created by the information systems in online P2P lending. Our approach could also be used to improve the classification of borrowers in P2P lending platforms or agents in other similar platforms, and therefore help users make better decisions. Furthermore, we adopt and extend the latent instrumental approach, a new approach to deal with endogeneity. As there are usually numerous issues with traditional IV estimation, LIV approach could provide a viable alternative to investigate empirical questions in the presence of endogeneity.

Chapter 3 Bidders Attrition in Sequential Online Auctions

3.1. Literature Review

Consumers spend significant amount of time and effort to search for information related to product quality/service and prices before making purchase decisions. Understanding consumers' desire and effort for information in any market is critical because it eventually affects the market structure (Nelson, 1970). Unprecedented proliferation of electronic markets in last decade has exponentially increased number of available options to consumers. These markets offer a variety of market mechanisms and purchase options for a wide range of consumer demographics. Electronic markets have also lowered the search cost and hence it is expected that cost to find vendors offering lowest price is almost negligible or very low. Due to lower search costs in these markets, it was predicted that consumers will switch to low cost providers which would lead to almost zero price differential (Reibstein, (2002). However, results of empirical studies were contrary to the popular belief. These studies found much wider price dispersions online compared to traditional brick and mortar retailing (Brynjolfsson and Smith 2000). Explosive growth in number of available options in online markets is one of the reasons that consumers are not able to search for lowest price vendors without incurring a high search cost. We argue that, instead of engaging in exhaustive and costly market search, consumers often focus on smaller niche markets. In these markets, along with product quality and price information, consumers also observe market supply and demand which can affect their purchase decisions.

Marketing literature is rich with studies on customer acquisition and retention. Most of these studies look at the repeat purchasing behavior of a firm's customers, the value that the customers bring to the firm, and the costs to acquire and retain them. It is true that similar to brick and mortar retailing, in online markets, a critical measure of success is businesses' ability to retain its customer base (Wang and Splegel 1994). In online markets where search costs are low and customer switching is high, it is becoming increasingly essential for all online businesses to understand the value of their repeat customers and develop efficient strategies to retain them for a long period of time (Wang and Splegel 1994). One simple reason for business to focus on retention is because it costs less than acquiring new customers. The general estimate is that : "It costs five times more to acquire a new customer than to retain an existing one" (Pfeifer 2005).

The current literature on customer acquisition and retention looks from a firm's perspective where the firm wants to attract more customers and keep them, especially customers who create more value. Customer acquisition and retention policies are based on customers' lifetime value (CLV) which is estimated based on their expected financial contribution over a period of time (Kumar, 2008). In contrast, our study focuses on the issue of customer retention in a sequential auction environment. We argue that in auction environment, CLV shouldn't be based only on their purchases because, even those customers who do not win, provide a value to the auctioneer by sheer participation. A higher number of participating bidders in an auction increases bidding competition. In auction environment, where final price is determined by competing bidders, a higher completion leads to higher winning bids (Pinker et al., 2003) (Pinker et al. 2003) (Pinker et al. 2003) (Pinker et al. 2003) (Pinker et al. 2003) (Pinker et al. 2003) (Pinker et al. 2003) (Pinker et al. 2003) (Pinker et al. 2003) (Pinker et al. 2003) (Pinker et al. 2003) (Pinker et al. 2003) (Pinker et al. 2003) (Pinker et al. 2003). Hence, the auctioneer wants the customers, i.e. bidders, to keep participating as long as they have not satisfied their demand for the product. This research can help auctioneers estimate the retention and attrition probabilities of bidders which in turn enable them to estimate customers' projected value and make more informed decisions.

Customer loyalty plays an important role in the customers' repeat purchase behavior. Although most of the literature on loyalty has focused on frequency of purchased packaged goods which measures brand loyalty, loyalty concept is critical to traditional and online retail establishments (Dick et al., 1994) to retain their customer base. Consumers who are satisfied with the service quality (and information quality in online markets) are more likely to be loyal. These loyal customers create positive word of mouth (Hagel III and Armstrong, 1997) which leads to higher profits. Loyal customers have lower price elasticity and are willing to pay higher to reduce their search cost of shopping around (Reichheld and Sasser, 1990). Prior studies have also validated that customer satisfaction, customer loyalty and profitability are positively related (Hallowell, 1996).

Customer loyalty in electronic markets (called e-loyalty) is more challenging than its counterpart in traditional markets because with lower switching cost, competition is just a few clicks away. Srinivasan et al. (2002) identify eight factors- customization, contact interactivity, convenience, cultivation, choice and character that can affect e-loyalty. Word of mouth by loyal customers has much wider impact on Internet via blogs, social media and online communities. Hence, customer loyalty and retention are highly desirable in electronic markets. In online auction markets, auctioneers' choice of auction market design can affect consumers' convenience and ease of interactivity which affects their loyalty to the auction platform.

We extend the notion of consumer retention or consumer attrition to the online auction environment. In the online auction environment, sellers don't offer items at fixed prices, and instead consumers compete with each other and determine the final prices. In this market, where price is not fixed, consumers have to balance the trade-off between the cost of participation and expected payoff upon winning in current or the future auctions. Hence, auction design parameters, consumers' perception of market supply and demand, significantly affect their decision to continue participating in auctions hosted on an auction platform. Further, consumers' understanding of the final equilibrium prices in auctions and their own private values also play a significant role on consumer attrition in online auction environment.

3.2. Conceptual Model and Hypothesis Development

It is important to understand consumer attrition in online markets, as competition is only a click away in these markets. Consumers always search for higher quality and lower price; however their search behavior varies and depends on market structure. Nelson (1970) in a seminal paper argued that market information (price and quality of the product), or lack of it, affects consumer behavior and eventually market structure. Consumers desire to obtain price and quality information which affects their purchase decision and eventually the market structure. In sequential auctions, the quality of the item auctioned doesn't change from auction to auction, and price is determined by the competing bidders. In this environment, prospective bidders focus on market information that drives bidding competition. Hence, we argue that in online auction markets, consumers' perceived information on market supply and demand, irrespective of its accuracy, plays a crucial role in their purchase decision and may eventually affect the market structure.

In fact, market demand and supply information has always played a key role in sellers' decision making, such as market mechanism design and estimation of prices (Engelbrecht-Wiggans and Kahn, 1999), however, not so much in consumer decision making, partially because of unavailability of mechanisms to estimate this information for consumers. We posit that many online market mechanisms, for example, online auctions, allow consumers to estimate market supply and demand information. Since bidders consider “auction as a competition” (Lee and Malmendier, 2006) and other bidders as their competitors (Ariely and Simonson, 2003), market supply and demand information allows them to estimate their cost (participation) and payoffs (winning) in sequential auctions environment. It is logical to argue that bidders’ future participation depends on their perception of market supply and demand of the product of their interest.

Prior work has argued that bidders’ perceived demand and supply information affects their bidding strategies (Goes et al., 2012), we extend that line of research and argue that perceived supply and demand information also impacts participation in these auctions. Research studies investigating consumer attrition in online retailing markets found that consumer online search, consideration and evaluation play a larger role in shopping cart abandonment (Kukar-Kinney and Close, 2010). We extend that line of research for online auctions and argue that bidders’ perception of market supply and demand affect their chance of winning in sequential auctions and hence determine their decisions to continue or leave the auction sequence.

3.2.1. Drivers of Bidders Attrition

3.2.1.1. Bidders’ Perception of Market Supply

Sellers conducting a sequence of auctions are generally endowed with a sizable inventory. These sellers determine the number of items (lot size) to be sold in each auction to maximize their expected payoff over the auction sequence. The primary reason of conducting auctions in a sequence is to better manage the release of these items in the auction market while maintaining a sustainable demand. There are various factors that affect sellers’ decision of lot size, such as product obsolescence, stochastic nature of demand, etc. to name a few. Hence, managing the pace of inventory clearance in sequential auctions is anything but trivial (Tripathi et al. 2009; Chen et al. 2011).

For strategic reasons, sellers don't reveal the size of available inventory to participating bidders. Bidders try to estimate the available inventory by observing sellers' actions in sequential auctions. We argue that bidders perceive a larger lot size in an auction as a signal for higher available inventory. A higher lot size in an auction could signal the seller's desire to clear the inventory quickly for various reasons, such as decreasing demand or product obsolescence, and has a positive and significant effect on auction entry (Bapna et al. 2003b). Hence, it might raise bidders' expectations of winning the product for a cheaper price in the future. Lower expected prices translate into higher expected payoff (Bapna et al. 2003a) for the bidders. Vakrat and Seidmann (1999a) also find that the consumer surplus is non-decreasing with the increase in the lot size offered in a single auction. Therefore, we argue that an increase in perceived supply should lead to lower risk of attrition. Formally, we hypothesize that:

H1: Increase in perceived supply reduces the odds of bidders' attrition.

The successive auctions in a sequence can begin before or after the first one ends. If the successive auctions overlap, bidders in the first auction can see an upcoming auction for the identical item whereas if successive auctions have a finite time gap between them (Time between auctions), bidders in the first auction are uncertain about the upcoming auction for the identical item. As the time gap between successive auctions increases, unsuccessful bidders from the previous auction are more likely to drift away from that auction site because they don't know about upcoming auctions for identical items and start searching for the item at competing sites. We argue that uncertainty about any more auctions selling identical items affects bidders' perception of market supply. Since bidders don't know if there would be another auction, they may perceive the previous auction as the end of the auction sequence. This perception about the end of supply becomes stronger as time gap between successive auctions increases.

Longer time between auctions could result in an increase in bidders' foraging behavior outside the auction sequence. Similar to the way humans and other species look for food, foraging behavior could be defined for the way customers browse and shop the goods they need online or offline (DiClemente and Hantula 2003; Rajala and Hantula 2000). In this context, we argue that bidders might find the product or an alternative for it outside the auction sequence and quit participating. Longer time gap between auctions is akin to waiting cost for the bidders (Vakrat et al., 1999) who stay around and expect another auction for the identical item. If auctions overlap, auctioneers can attract bidders from previous auctions to the next auction. Hence, we argue that the longer time between successive auctions increases uncertainty about upcoming auctions and increases the odds of bidders' attrition. Formally, we hypothesize that:

H2: Longer time between successive auctions increases the odds of bidders' attrition.

3.2.1.2. Bidders' Perception of Market Demand

Consumers compare prices of product/services with an initial anchor or an internal reference price that they estimate based on their prior experience (Winer 1986). Reference price affects consumers' willingness-to-pay and plays a critical role in consumers' purchase decisions. Drawing on Prospect Theory (Kahneman and Tversky 1979), transaction utility, the discrepancy between reference and actual price, is considered by customers when evaluating potential transactions and making decisions (Thaler 1985). However, there is no simple explanation of how consumers determine reference prices of products/services.

Consumer decision models have shown that consumers often struggle with the value assessment of product and services (Simonson & Tversky, 1992) and refine their assessment over time (Bettman, Luce & Payne, 1998). Along with these lines, we argue that consumers' reference price for a product/service may depend on a variety of market cues they acquire over time from multiple sources. In the context of sequential auctions, where bidders can estimate market supply and demand by observing sellers' and buyers' behavior, they continuously refine their reference prices as they acquire additional explicit and subjective information about the product/services. Although the items sold in multiple auctions of a sequence are identical, empirical evidences show that the final prices change over time, partially due to dynamic mix of bidders and ensuing bidding competition. Hence, we argue that even in sequential auctions where each auction sells identical items, consumers' reference prices for these items may change as they observe others' bids (Kauffman and Wood 2006) over time. Similar to other online shopping environments, bidders in online auctions prefer to bid closer to their reference prices. In sequential auctions of identical items, there is a significant fluctuation in winning bids and hence, bidders track many auctions in a sequence, and expect to win at prices closer to their reference prices. A larger the gap between the price and the external reference price increases the customers' perception of value which in turn results in an enhanced willingness to buy (Urbany et al. 1988), so bidders with higher reference prices see more opportunities compared to bidders with lower reference prices. Therefore, we argue that bidders with higher reference prices are likely to stay longer in the auction sequence compared to the ones with lower reference prices. Formally, we hypothesize that:

H3: A higher reference price reduces the odds of bidders' attrition.

In sequential auctions where bidders expect to see many auctions for identical items, they shade their bids and do not place their highest bid in their initial auctions. According to Assimilation and Contrast Theory (Sherif 1963), consumers have ranges of acceptable prices for contemplated purchases and they contrast observed prices with their acceptable range (Monroe 1971). Based on their valuation for the product, and the auction sequence dynamics, bidders decide the maximum they want to bid in each auction. A bidder's maximum bid in any auction is also termed as her final bid. This maximum or final bid often changes from one auction to another, based on bidders' demand, their experience (Goes et al., 2010) and market information. For example, a bidder's final bids may go from $80 in the first auction, to $120 in the second and $140 in the third auction that she participated. Bidders follow a variety of bidding strategies in an auction (Bapna, et al., 2004). Some bidders start with a lower bid and bid up to their final bid whereas others place only one bid which is their final bid. Whatever strategy they follow, they retire after they reach their maximum bid (final bid) for that auction.

Bidder's final bid in an auction is their best effort to win an item in that auction. Relatively speaking, a higher final bid (compared to some fixed value) implies that the bidder has tried harder to win the item. If he or she still does not win in spite of submitting a relatively higher bid, this could cause disappointment and frustration and results in a lower perceived chance of winning. Bidders may attribute this lower expected winning probability to increase in market demand estimated based on others' bids among other market parameters. A lower perceived chance of winning translates into lower expected returns from participating in the auction sequence, so it may discourage the bidder from participating in the future auctions. Therefore, we hypothesize that as bidders place higher bids and are still unsuccessful, they are more likely to leave the auction sequence. Formally, *H4: Higher unsuccessful bids increase the odds of bidders' attrition.*

Number of bidders has a positive impact on bidding competition in any auction. Prior research has shown that a higher number of bidders leads to a higher number of bids and eventually higher winning bids. This increased market competition due to increase in number of bidders influences all the participating bidders (Ariely and Simonson 2003) as they contemplate to continue their participation in the current or future auctions. For an individual bidder, a higher number of bidders in an auction could be a signal of higher market demand and higher competition. Arora et al. (2004) argue that the number of bidders plays a critical role in bidders' decision to continue participating in auctions. They argue that bidders compute an option value for future auctions. This option value measures the expected payoff from future auctions and depends on various factors including the number of competing bidders in future auctions. For example, if bidders anticipate possible increase in number of bidders (market demand) over time, they may decide to quit participating in auction sequence. Formally, we hypothesize that:

H5: Higher number of bidders in an auction increases the odds of bidders' attrition.

3.2.1.3. Bidders' Participation Cost

Assuming bidders continuously arrive on auction platforms, auctions with a longer duration are exposed to higher number of potential bidders. Partly due to this reason, we have observed that many online auctions now run for longer durations. Since these auctions are in open ascending price format, longer duration results in higher number of bidders and more competition because it gives the bidders more time to place higher bids. In essence, longer auction duration has a positive impact on prices (Bapna et al. 2003a). However, from an individual bidder's perspective, longer auction duration changes auction dynamics and expected payoffs. Longer duration results in bidding wars and higher prices which adversely affects bidder's payoff. Longer auction duration also increases the bidders' participation costs, as they need to spend more time monitoring the auction, and it also takes longer to receive the item if she actually wins (Weinberg 2000; Davis and Maggard 1990). This in turn reduces the bidders' expected return from participating in the sequence. Hnece, longer auction duration is considered as a waiting cost and because of that bidders prefer shorter auction durations (Vakrat and Seidmann 1999b). Generally, in electronic markets, longer time to complete a transaction adversely affects buyers' decision to continue with online shopping (Rajamma et al. 2009). We argue that longer auction duration in online auctions will impose higher costs and reduce bidders' expected payoffs, therefore will adversely affect their decision to continue participating in an auction sequence. Formally, we hypothesize that:

H6: Longer Auction duration increases the odds of bidders' attrition.

In sequential auctions, bidders observe/participate-in many auctions, to gain experience and estimate market demand and supply of these items based on auction lot sizes, bidding competition and other relevant market data. In this process, bidders trade their time for market information and expect to use that market information to find better bargains, or in other words, expect to win in auctions, when competition and winning bids are lower. However, bidders incur participation cost to find bargains in these auctions. We argue that bidders' participation cost increases with the time they stay active in an auction sequence. Time already spent in the system can be viewed either as a positive investment to attain the goal (a reward perspective) or as a wasteful expense (a cost perspective or sunk cost attribution) (Meyer 1994). We hypothesize that as time since becoming active (placing a bid) in an auction sequence for a bidder increases, her participation cost also increases. This increase in participation cost leads to decrease in bidders' expected payoff from future participation. Therefore:

H7: Longer duration of being active in an auction sequence increases the odds of bidders' attrition.

3.2.1.4. Winning Effect

Bidders in online auctions are heterogeneous in their demand (Goes et al. 2010). For example, some of the bidders are traders, desiring multiple units of items for reselling, whereas others are individual buyers looking for one unit for their own consumption. The payoff of these resellers/traders comes from market arbitrage, buying items at bargain prices in auctions and selling them in secondary markets for a higher price (reference). These bidders are referred to as multi-unit demand (MU) bidders, as compared to single-unit demand bidders whose goal is to just buy one unit of the auctioned item (Goes et al., 2010). In sequential auctions these resellers are able to spread their demand over multiple auctions and they are generally likely to reduce their bids to decrease their marginal cost. We posit that participating in the auctions is less costly for MU bidders as they generally have clearer objectives, more experience and more knowledge in bidding and purchasing through auctions (Goes et al. 2010; Vincent Lyk-Jensen and Chanel 2007). So they are more likely to continue participating in the auctions until their demand is fulfilled, and they are at a lower risk of leaving the sequence of auctions at any time compared to single-unit demand (SU) bidders who are probably ordinary buyers looking for a bargain.

H8: The attrition risk for Multi-unit demand bidders is on average lower than that of single-unit bidders.

Multi-unit demand bidders who participate in multiple auctions of identical items do so in order to find good bargains. These bidders might continue to participate in the auction sequence after winning, and desire to win more items. However consumers' utility of owning and consuming the same product is often modeled in the literature as a concave function, using functional forms such as quadratic or Cobb-Douglas function (Amir and Jin 2001; Damania et al. 2004; Häckner 2000; Singh and Vives 1984). This implies diminishing utility of consuming the same product. We extend this concept to the context of sequential auctions, and argue that the odds of attrition for the bidders who have previously won an item are higher compared to the ones who have never won.

H9: Odds of bidders' attrition increase after bidders win in a previous auction.

3.3. Data

3.3.1. Data Collection

The data used in this study was collected from a reputable online auctions website. To clear large inventories, this auction house conducts many auctions for identical items, we term them as a sequence of auctions. Number of auctions in any auction sequence depends on the total initial inventory to be cleared. The auctioneer determines the number of units (lot size) to offer in each auction of a sequence, which in itself is a challenging problem (Tripathi et al., 2009; Pinker et al., 2010; Chen et al., 2011) and is not the focus of this research Since all the auctions hosted on this auction site are from single seller, this dataset allows us to control for seller heterogeneity and consequences of sellers' reputation, thus removing the effect of trust in the seller (Ba and Pavlou 2002b). Similar to most other reputable auction sites, bidders on this auction site need to register and which allows us to track every bidder and their bidding patterns across multiple auctions.

This auction house uses the open ascending uniform price auction mechanism. In uniform price auctions, multiple units of the same item might be offered in a single auction, and all the winners pay the same price. This price is equal to the marginal bid, defined as the lowest winning bid. For example, if an auction is selling three identical items and at the close of the auction the top three bids are $200, $180, and $170 from three different bidders. In this case, all the three winning bidders pay only $170 each. This type of auction is sometimes called a "Dutch" auction in the online auctions context (Lucking-Reiley 2000; Parsons et al. 2011). In this open auction format bids are observable and bidders can enter the auction at any stage

We collected data using an automated agent from the online auction website. The auction house conducts multiple auctions for the same product one after another in what we define as an auction sequence. The number of auctions in a sequence varies due to the number of units in stock for each product. Our dataset set includes two groups of data. Auction-specific data contains data on the description and quantity of the product offered, auction start and end time, final price and minimum bid increment for every auction in the sequence. Bid-related data includes data on the amount of bid, bid time, bidder ID, quantity bid and the quantity won for all the bids in any auction.

For the purpose of this study, we selected an auction sequence conducted to sell an electronic consumer product, a digital camera. This auction sequence was one of the largest auction sequences providing rich data for our analysis. A total of 318 auctions were conducted in this sequence and attracted more than 1000 unique bidders who submitted over 3300 bids across all the auctions in this sequence. Table 12 shows the summary statistics of our dataset.

Table 12: Overall Summary statistics of the data set

Statistic	Value
Total Number of Auctions in the Sequence	318
Average Winning Bid	109.11
Total number of bids submitted	3330
Average number of bidders per auction	10.4
Average Lot Size per Auction	1.2
Total Number of Unique Bidders	1126
Average. Number. Of Auctions a Bidder Participated in	2.96

3.3.2. Variables and Measures

We collected data at the bidder level and for each bidder, we tracked all the auctions in a sequence she has participated. Let a bidder i, $i \in \{1,2,\ldots,I\}$, participates in a total of J_i auctions. For the j^{th} auction in this bidder's sequence, $j \in \{1,2,\ldots,J_i\}$, we recorded the following variables: time when bidder placed her final bid ($BiddingTime_{ij}$), the amount of final bid $\left(FinalBid_{ij}\right)$, the duration of auction j ($AuctionDuration_{ij}$), the lot size of auction j ($LotSize_{ij}$), total number of bidders in auction j (NB_{ij}), and the auction j's winning bid which is the price winners paid (P_{ij}). Some of these variables (like auction duration and lot size) are auction-specific variables. However due to the fact that we are looking at the data from the bidders' perspective and because of the way we have defined j, we have kept the bidder subscript for these variables.

Table 13 shows the summary statistics of the measures used in this study. D_{ij} is a dummy variable that is used to demonstrate the sequence of participation (waves of data in panel data terms), or the experience level of a bidder in the auction sequence. For bidder i, $D_{ij} = 1$ if she participated in j^{th} auction and zero otherwise. We will see how we have used this set of dummy variables to capture the effect of time or participation experience on the risk of attrition. $WonDummy_{ij}$ is a dummy variable that is zero if bidder i has not won any items in auctions $1,2,\ldots,j$. If she wins an item in an auction it changes to 1 and will remain 1 in all of the subsequent auctions for this bidder.

MU_i is a dummy variable used to identify bidders based on their demand. Prior research has shown that bidders in sequential auctions are heterogeneous in their demand-single-unit (SU) demand bidders, who wish to win only one unit and multi-unit (MU) demand bidders, who are likely to be resellers and desire to win multiple units to resell in secondary markets (Goes, et al., 2010). Following prior research (Goes et al,(2010), we categorized bidders as SU or MU *ex-post*, based on their revealed demand. Bidders who bid for at most one unit and drop out of the auction sequence before or after winning are categorized as single-unit (SU) demand bidders. On the other hand, bidders who either bid for more than one unit at least once or continued participating in the sequence after winning, reveal their demand for multiple units, are categorized as multi-unit (MU) demand bidders. Since this is an ex-post analysis, a bidder who participated in 20 auctions of a sequence, won in the 10th auction and continued to bid in subsequent auctions after winning, is categorized as MU in all of these 20 auctions.

Following previous research, we assume that a bidder's maximum bid in any auction sequence represents her willingness-to-pay (WTP) and bidders' WTP bounds their valuations from below (Bapna, Goes, Gupta and Karuga, 2008; Jeitschko, 1998). WTP_i is constructed ex post as the maximum bid submitted by a bidder in all of her participations in the sequence. It is calculated as $WTP_i = Max(FinalBid_{i1}, FinalBid_{i2}, \ldots, FinalBid_{iJ_i})$.

$ReferencePrice_{ij}$ is calculated as the minimum winning price observed by the buyer throughout the auctions in which he or she has participated. It can be interpreted as a proxy for a bidder's reference price for the item in the context of the auction sequence. We will use this measure to investigate the impact of reference price on bidders' risk of attrition. It is calculated as $ReferencePrice_{ij} = Min(P_{i1}, P_{i2}, ..., P_{ij})$.

$BidderDuration_{ij}$ is the time since a bidder started participation. It measures the number of days a bidder has been involved with the auction sequence. It is calculated as $ActualCloseTime_{ij} - BidTime_{i1}$. $ActualCloseTime$ is the time when an auction is closed and no further bids are accepted. We use this variable as it is the last point in time a bidder can participate in an auction.

Table 13: Summary Statistics of the Measures

Variable	Mean	SD	Min	Max
D_{ij}	-	-	0	1
$BiddingTime_{ij}$	84.26	22.17	0	129.52
$FinalBid_{ij}$	69.83	37.54	1	196
$AuctionDuration_{ij}$	0.57	0.41	0.083	2.5
$LotSize_{ij}$	1.36	1.14	1	5
NB_{ij}	11.87	4.78	3	31
P_{ij}	110.92	20.86	$ 62.00	$ 196.00
$WonDummy_{ij}$	0.17	0.38	0	1
MU_i	0.30	0.46	0	1
WTP_i	81.51	38.19	$ 1.00	$ 196.00
$ReferencePrice_{ij}$	105.68	19.98	$ 62.00	$ 196.00
$BidderDuration_{ij}$	5.02	11.32	0	100.87
$NormalizedLotSize_{ij}$	0.89	0.27	0.2	1
TBA_{ij}	-0.11	1.57	-1.20	18.79

$NormalizedLotSize_{ij}$ is a measure calculated by dividing the lot size offered in the bidder i 's j^{th} auction by the maximum lot size, this bidder has observed in all of her participations, including auction j. It is calculated as:, $NormalizedLotSize_{ij} = \mathrm{LotSize_{ij}} / \mathrm{MaxLotSize_{ij}}$, where $\mathrm{MaxLotSize_{ij}} = \max(LotSize_{i1}, LotSize_{i2}, \ldots, LotSize_{ij})$. This variable ranges from zero to one and represents the relative size of the current auction's lot size for bidder i compared to the largest lot size he or she has seen. This measure captures the effect of changes in market supply, as perceived by the bidders compared to what they have observed before.

TBA_{ij} is the time gap between the end time of bidder i 's j^{th} auction and the beginning time of the next auction in the sequence, regardless of the bidder's participation in the next auction. Larger time gap between successive auctions increases uncertainty about available supply of items.

3.4. Methodology

In order to study factors that influence bidders' decision to leave an auction sequence before completely satisfying their demand, we utilize survival analysis methodology. Survival analysis or duration modeling is used in situations in which the time to occurrence of some event is of interest for some population of individuals (Lawless 2002). We define a bidder's decision to leave the auction sequence before satisfying her demand as an event and we want to study the variables affecting the occurrence of this event in time. A research question must have three substantive features to make it a candidate for survival analysis. These features are: a well-defined event whose occurrence is being explored, a clearly defined beginning of time, and a substantively meaningful metric for clocking time (Singer and Willett 2003). In addition, survival analysis provides the means to deal with the common problem of censoring in lifetime data. As we discuss in later sections, our data and research question satisfy all of the above-mentioned characteristics that justify and call for the use of survival analysis.

In the previous section, we discussed two types of bidders- SU bidders who aspire to win only one unit of an item and MU bidders who desire to win multiple units of an item. An example of a bidder leaving before completely satisfying her demand is a SU bidder leaving the sequence before a win. Similarly, when a MU bidder comes back to auction sequence after securing a win but later leaves the auction sequence without winning again, we consider this as a case of bidder leaving before completely satisfying her demand. We will elaborate bit more on this in next section. A dummy variable ($Event_{ij}$) is used to denote whether a record of data (which includes data on a bidder's participation in an auction) represents an event or not.

We have the exact time the bids are submitted. However, we have used a discrete-time hazard model. We argue that using a discrete-time model is not only preferred, using the continuous-time model with exact bid times could be erroneous. The reason is that although bid time is seemingly continuous, from the researcher's perspective, it is actually discrete by nature. First, often there is a time gap between the end of one auction and the beginning of next auction in a sequence. So, there are some points in time during which bid submission is not possible. Second, the occurrence of an event happens at the close of an auction because, a bidder, who is not winning, has up until the close of the auction to bid or not bid and possibly leave the sequence. The order of participation in a sequence by a bidder is used as a time metric. Based on this time metric, the beginning of time in our study would be a bidder's first auction participation in the auction sequence.

3.4.1. Identification and Treatment of Censoring

A key characteristic that distinguishes survival analysis from other areas in statistics is that survival data are usually censored (Leung et al. 1997). Individuals, who do not experience the target event during the period of data collection, are tagged as individuals with unknown event times, and are censored (Singer and Willett 2003). There are two major reasons for this: (1) some individuals never experience the event, and (2) some experience the event, but do so after the end of the data collection period (Singer and Willett 2003).In this research, we have collected data on all the auctions in the sequence focused in this study which includes all the bids submitted in these auctions. Therefore, censoring due to the second reason is irrelevant in this data set. Since, we have the data from the beginning of the sequence; this is a case for right-censoring, because bidders do not know the ending time of the auction sequence. For example, if an auction sequence has 318 auctions and a bidder has participated in 316th auction of the sequence. Let say that this bidder might be interested to participate in more auctions, but when she comes back, the sequence has ended. Since we observe all the bids of the bidders and particularly their last participation in the sequence (316th auction from the example discussed above), we can use that data for censoring. Next we discuss our censoring approach in this study.

For the bidders whose last participation coincides with winning the item, we assume they are censored at their last participation. They might have satisfied their demand and left, or they wanted to come back for more (MU bidders) but the auction sequence has ended. In either case, these observations ought to be censored (Singer and Willett 2003). As for the bidders who do not win an item in their last participation, we have used the available data to statistically infer whether a bidder has actually experienced an event or not. We argue that if the time gap between a bidder's last participation and the end of the auction sequence is statistically large enough, one can conclude that the bidder has experienced the target event in his last participation.

We call a bidder a $n\text{-}time$ bidder if she has participated in a total of n auctions in the auction sequence. Let say, a bidder i is a $n\text{-}time$ bidder. The time gap between this bidder's last participation (J_i^{th} auction) and the end of auction sequence can be calculated as: $O_i = EndingTimeOfAuctionSequence - Biddingtime_{iJ_i}$. We now look at $(n+1)-time$ bidders and calculate the time gaps between their n^{th} and $(n+1)^{th}$ auction participation. We assume that these time gaps are generated using the same underlying distribution (F_{n+1}), which can be estimated using data from all $(n+1)-time$ bidders. Following this argument, if the bidder i, who is a $n\text{-}time$ bidder, actually wanted to participate one more time, she would become a $(n+1)-time$ bidder, and then the time gap between her n^{th} and $(n+1)^{th}$ participation would also come from same underlying distribution (F_{n+1}). Our null-hypothesis is that a $n\text{-}time$ bidder could have become a $(n+1)-time$ bidder if the sequence had gone on long enough.

We have only considered 8 data series, namely $2-time$ bidders up to $9-time$ bidders. This is because as n increases, the number of data points in a category decreases, therefore for $n > 9$, it is not possible to fit a distribution to these points. Therefore the bidders with more than eight participations are censored at their 8th participation. Using this conservative approach, we make sure the last participation by these bidders is not wrongly labeled as a target event, at the expense of losing some information. We fitted several distributions to all of the eight data series using maximum likelihood and for each distribution, we used three common tests of goodness of fits of distributions, namely Kolmogorov-Smirnov test, Anderson-Darling test and Chi-squared test (Snedecor and Cochran 1989; Lawless 2002) to assess and compare the fit of the distributions. The Weibull distribution proved to collectively provide the best fit across all eight data series. Weibull is an extremely flexible distribution which can possess either positive or negative skewness depending on the values of its shape and scale parameters (Pinder et al. 1978). We fitted a Weibull distribution to each of the eight data series. We call these distributions $F_k \ (k=1,2,..,8)$. In order to determine whether bidder i who is a $J_i - time$ bidder has experienced an event, we calculated the variable $O_i = EndingTimeOfAuctionSequence - Biddingtime_{iJ_i}$. Adopting a conservative approach, our null hypothesis is that O_i is not large enough, so the bidder has not experienced the target event and has to be censored. To evaluate this hypothesis, we need to calculate $\hat{F}_{J_i+1}\left(O_i\right) = 1 - F_{J_i+1}(O_i)$ and compare this probability to a defined threshold. If it is small enough (smaller than the threshold), we can reject the null and claim that bidder i has had enough time to come back, yet he or she decided not to and so has experienced an event. If not, we cannot make such a claim due to lack of evidence and then bidder i is censored. We have chosen the threshold to be 0.01, so we can claim with 99% probability that the bidder has actually experienced an event.

3.4.2. Discrete Hazard Model

After defining time metric, events and censoring, we now discuss the rationale behind our choice of the survival model. Since we are using a discrete time metric, we need to use a discrete-time hazard model. We have chosen to use the logit link due to its interpretive convenience (Hosmer and Lemeshow 2000). In section 6, we show the robustness of our approach. We show that parameter estimates do not change significantly if we use other links like the complementary log-log link. Using the logit link, the hazard rate for bidder i in his or her j^{th} auction participation, given that she has not yet experienced an event, is defined as:

$$h\left(t_{ij}\right)=\frac{e^{\left[\alpha_1 D_{1ij}+\alpha_2 D_{2ij}+\ldots+\alpha_8 D_{8ij}\right]+[\beta_1 X_{1ij}+\beta_2 X_{2ij}+\ldots+\beta_p X_{pij}]}}{1+e^{\left[\alpha_1 D_{1ij}+\alpha_2 D_{2ij}+\ldots+\alpha_8 D_{8ij}\right]+[\beta_1 X_{1ij}+\beta_2 X_{2ij}+\ldots+\beta_p X_{pij}]}} \tag{19}$$

In equation (19), we see two groups of variables: eight dummy variables, D, for each auction participation and a set of explanatory variables, X. There is no constant term as we include a dummy for each time period. As discussed earlier, in any auction, only one of the time dummies is one and the rest of them are zero. This results in a completely general and flexible baseline hazard model in which the effect of each period of time is captured by a different intercept (α_1 to α_8). The explanatory variables are chosen based on the conceptual model and their estimated coefficients will be used to test the proposed research hypotheses.

Our goal is to estimate the parameters in equation (19) to evaluate the effect of the variables of interest on the hazard rate. In order to estimate the parameters, we use the logit transformation to transfer this equation into the linear regression model below:

$$Logit\, h\left(t_{ij}\right)=\left[\alpha_1 D_{1ij}+\alpha_2 D_{2ij}+\ldots+\alpha_8 D_{8ij}\right]+[\beta_1 X_{1ij}+\beta_2 X_{2ij}+\ldots+\beta_p X_{pij}] \tag{20}$$

Before proceeding with estimation and analysis, we discuss some of the potential challenges in analysis and model estimation and propose ways to deal with them.

3.4.3. Multicollinearity

Multicollinearity could result in inflated variance of the estimated coefficients and produce implausible magnitudes or even wrong signs in small samples (Greene 2007). In order to reduce correlation, we have standardized all the covariates except for the dummies (Marquardt 1980). However mean centering it does not guarantee the elimination of multicollinearity in moderated regressions (Dunlap and Kemery 1987).

We have calculated the variance inflation factors (VIF) and tolerance for all the covariates included in our full model. As we can see in Table 14, the greatest VIF is 5.4 and the average VIF is 2.35. Researchers are not unanimous in determining the existence and severity of multicollinearity. Some researchers consider values of VIF greater than 10 as a sign of problematic multicollinearity (Marquardt 1970; Hair et al. 1998). According to this rule, multicollinearity is not a serious issue with our data. Others suggest that VIF values of as low as 5 could also call for more investigation (Stine 1995; Menard 2001). In order to further investigate the severity of multicollinearity among our covariates, we divided our sample into two rather equal subsamples based on a randomization of bidders. Existence of multicollinearity results in swings in parameter estimates. The estimates of our subsamples were qualitatively similar to the full sample estimates, thus further assuring us that multicollinearity is not an issue in our empirical analysis.

Table 14: Variance Inflation Factors

Variable	VIF	Tolerance
MU	2.3	0.4351
TBA	1.03	0.9679
NormalizedLotSize	1.05	0.9569
AuctionDuration	1.14	0.8742
BidderDuration	1.15	0.8676
ReferencePrice	1.38	0.7249
WTP	5.39	0.1855
WonDummy	3.27	0.3061
FinalBid	5.4	0.1853
FinalBid_WonDummy	2.67	0.3739
NB	1.04	0.9577

3.4.4. Unobserved Heterogeneity

In building equation (19), we have assumed that the bidders' with the same values of the included covariates are homogeneous and have the same attrition risk, i.e., all the differences between the hazard rates in individuals are explained by the explanatory variables. However, there could be some heterogeneity in bidders' tendency to leave the auction sequence which is unexplained by the model. Ignoring such heterogeneity and naïve acceptance of the observed population patterns may lead to drawing erroneous conclusions (Vaupel and Yashin 1985). Greater unobserved heterogeneity across individuals makes it more likely to incorrectly estimate a decreasing hazard rate (Follmann and Goldberg 1988). This is because individuals who are inherently more likely to experience the event, do so in earlier stages, so the proportion of these individuals goes down in time. Figure 2 shows the average hazard rate of bidders' attrition as a function of the time dummies controlling for other variables. Based on Figure 3, unobserved heterogeneity does not seem to be a serious problem in our empirical model, as unobserved heterogeneity always leads to hazard functions that appear to decline over time, an effect that is not observed here (Singer and Willett 2003).

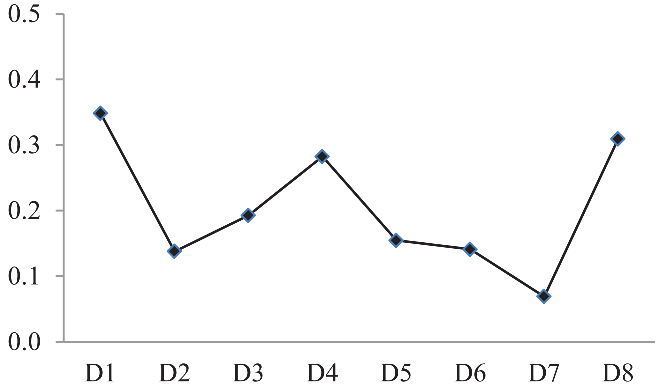

Figure 3: Average Baseline Hazard Rate of Attrition over Time

To further investigate the existence of unobserved heterogeneity in our model and to capture such possible effects we used a random effects model. Adding bidder-specific fixed effects leads to biased estimates as we are dealing with a survival model with non-repeated events (Iyengar et al. 2011). Thus, we estimated the hazard model with the logit link and flexible baseline hazard specification with normally distributed random effects on the intercept, as shown below in equation (21) where $u_i \sim N(0, S_u^2)$:

$$Logit\, h\left(t_{ij}\right) = \left[\alpha_1 D_{1ij} + \alpha_2 D_{2ij} + \ldots + \alpha_8 D_{8ij}\right] + \left[\beta_1 X_{1ij} + \beta_2 X_{2ij} + \ldots + \beta_p X_{pij}\right] + u_i \tag{21}$$

To estimate this model, we used the NLMIXED procedure in SAS which tries to maximize the likelihood function directly by numerical integration methods using adaptive Gaussian quadrature (Pinheiro and Bates 1995). However, the results of the estimation show that the variance of the random effect, S_u^2, is very small and statistically very insignificant. This implies that with the rich set of covariates included, the model is doing a satisfactory job in capturing the existing heterogeneity in bidders and bidders' behavior.

3.5. Empirical Analysis

3.5.1. Model Estimation and Discussion

Based on the research hypotheses, we considered different measures constructed using the available data that could serve to measure the effects of interest. Model (1) in Table 15 shows the full model with variables and measures that help us evaluate the research hypotheses. Whenever there was more than one variable related to a research hypothesis, the one that captured the impact better, both substantively and empirically, is chosen. All the variables except for dummy variables are mean centered and standardized to have a variance of 1. So it is possible to compare the size of the coefficients. Maximum likelihood method is used to estimate the coefficients.

Table 15: Estimation Results for Discrete Hazard Model with Logit Link and Flexible Baseline

	(1) Full model with random effects	(2) Full Model without Random Effects	(3) Parsimonious Model
D1	-0.63**	-0.63**	-0.67**
	(0.10)	(0.10)	(0.10)
D2	-1.84**	-1.84**	-1.83**
	(0.16)	(0.16)	(0.16)
D3	-1.43**	-1.44**	-1.40**
	(0.19)	(0.19)	(0.19)
D4	-0.93**	-0.93**	-0.87**
	(0.21)	(0.21)	(0.20)
D5	-1.70**	-1.70**	-1.63**
	(0.28)	(0.28)	(0.28)
D6	-1.81**	-1.81**	-1.71**
	(0.347)	(0.348)	(0.34)
D7	-2.61**	-2.61**	-2.51**
	(0.50)	(0.50)	(0.49)
D8	-0.81*	-0.81*	-0.71*
	(0.40)	(0.40)	(0.36)
MU	-2.87**	-2.87**	-2.86**
	(0.43)	(0.43)	(0.43)
TBA	0.15**	0.15**	0.15**
	(0.051)	(0.051)	(0.05)
NormalizedLotSize	-0.19**	-0.19**	-0.18**
	(0.057)	(0.057)	(0.06)
AuctionDuration	0.43**	0.43**	0.43**
	(0.062)	(0.062)	(0.06)
BidderDuration	0.11	0.11	
	(0.08)	(0.08)	
ReferencePrice	-0.48**	-0.48**	-0.49**
	(0.07)	(0.07)	(0.07)
WTP	-1.81**	-1.80**	-1.78**
	(0.22)	(0.22)	(0.22)
WonDummy	2.21**	2.21**	2.20**
	(0.51)	(0.51)	(0.51)
FinalBid	1.64**	1.64**	1.60**
	(0.22)	(0.22)	(0.22)
*FinalBid*WonDummy*	-2.22**	-2.22**	-2.21**
	(0.38)	(0.38)	(0.38)
NB	0.059	0.059	
	(0.057)	(0.057)	
S_u	0.006		
	(0.052)		
LL	-995.506	-995.507	-996.9093
AIC	2031.013	2029.013	2027.819
BIC	2302.21	2286.649	2258.335

Notes: The numbers in the parentheses are the standard errors for the estimated parameters.
* Significant at 0.05; ** significant at 0.01

Regarding the effect of time or in other words, the effect of auction participation on the attrition risk, we see that the estimated coefficients for the time dummies, though significant, do not suggest any specific pattern, and some of these changes in hazard rate could also be attributed to sampling variation. Given that we have mean centered all the variables, the coefficient estimates of these dummies represent the baseline logit hazard for an average SU bidder who has not won an item yet. We have included the bidder-level variable WTP_i as a control. WTP_i represents a bidder's willingness to pay, or his or her valuation for the product. The coefficient for this variable is significant and negative as expected, meaning that bidders with a greater willingness-to-pay are less likely to leave compared to bidders with lower WTP. Table 16 summarizes our empirical findings with regards to the research hypotheses.

Table 16: Summary of the Analysis of Hypotheses

	Hypotheses	Result
H1:	*Increase in Perceived supply reduces the odds of bidders' attrition.*	*Supported*
H2:	*Longer time between successive auctions increases the odds of bidders' attrition.*	*Supported*
H3:	*A higher reference price reduces the odds of bidders' attrition.*	*Supported*
H4:	*Higher unsuccessful bids increase the odds of bidders' attrition.*	*Supported*
H5:	*Higher number of bidders in an auction increases the odds of bidders' attrition.*	*Not supported*
H6:	*Longer Auction duration increases the odds of bidders' attrition.*	*Supported*
H7:	*Longer duration of being active in an auction sequence increases the odds of bidders' attrition.*	*Not supported*
H8:	*The attrition risk for Multi-unit bidders is on average lower than that of single-unit bidders.*	*Supported*
H9:	*Odds of bidders' attrition increase after bidders win in a previous auction.*	*Supported*

The negative and significant coefficient of $NormalizedLotSize_{ij}$ provides support for hypothesis 1, that increase in perceived supply reduces the odds of attrition. This measure is calculated by dividing the lot size in an auction by the largest lot size the bidder has seen before. An interesting issue to point out is that the variable $LotSize_{ij}$ which is simply the number of items offered in the j^{th} auction participation of bidder i was not significant on its own. This implies that the history of lot sizes offered in previous auctions and the changes made, and not just the current auction lot size, are considered by the bidders in forming their expected payoff from participation.

We hypothesized that longer time between successive auctions increases the odds of attrition. The variable TBA_{ij} measures the time gap between bidder i 's current auction and the next auction in the sequence. The coefficient for this variable is positive and significant meaning that our model provides support for hypothesis 2. Given the definition of this variable, it takes negative values for overlapping auctions. It can thus be inferred that reducing the gap between the auctions in the sequence and having overlapping auctions could help the auctioneer reduce the attrition odds. In fact a 1 standard deviation (1.57 days from Table 13) increase in the time between auctions is associated with a 16% ($e^{0.15}-1$) increase in attrition risk ceteris paribus.

The estimated coefficient for $ReferenePrice_{ij}$ is negative and significant. This variable is time varying and is constructed as the minimum winning price seen by bidder i in his or her past participations. Based on Helson's Adaptation-Level Theory (Helson 1964), reference price is an internal standard against which observed prices are compared. Considerable theoretical justification for consumers' use of psychological reference points exists and customers rely on past prices as part of the reference price formation process (Kalyanaram and Winer 1995). It is found that reference prices also impact customers’ decisions in auction settings (Kamins et al. 2004; Rosenkranz and Schmitz 2007), and our study provides evidence that reference prices also significantly affect the bidders' risk of attrition, as a $20 increase in the bidders' reference price is on average associated with 39% decrease in the odds of attrition.

Hypothesis 4 posits that higher unsuccessful bids increase a bidder's risk of attrition. Our results in Table 15 lends strong support for this hypothesis, as the estimated coefficient for $FinalBid_{ij}$ is positive, statically significant and quite large. A $37 higher unsuccessful bid is associated with an almost five times increase in a bidder's attrition, everything else held constant. This could be attributed to the bidders' frustration which could significantly impact their perceived chance of winning a future auction. Personal and psychological factors are among the most important factors affecting customer behavior (Keisidou et al. 2011). It is important to note that we are controlling for the bidders' willingness to pay in the model. Interestingly, this effect is reverse for the bidders who have won at least one item before. The estimated coefficient of the interaction term between $FinalBid_{ij}\ \ WonDummy_{ij}$ is negative and significant. Its absolute value is also larger than that of the coefficient for $FinalBid_{ij}$ suggesting that when similar bidders with winning experience submit a higher unsuccessful bid, they are actually less likely to leave. This could possibly be due to the fact that higher bids represent a bidder's stronger intent on winning the item.

The fifth hypothesis was about the effect of the total number of bidders in an auction. The results of model estimation in Table 15 however show that the coefficient for variable NB_{ij} is not statistically significant. We tried various measures and different functional forms that could reflect the effect of number of bidders on the attrition risk, but none of them was statistically significant. So interestingly, based on our analysis, the total number of bidders is not an important factor affecting the bidders' risk of attrition when we control for other factors.

In hypothesis 6, we posited that longer auction duration increase the attrition risk. Looking at Table 15, we see that the estimated coefficient for the variable $AuctionDuration_{ij}$ is positive and highly significant. Controlling for the effect of other variables, the odds of attrition for a bidder participating in an auction which is one standard deviation (9.8 hours) longer is 54% more than the shorter auction. Longer auction durations could be interpreted as longer correctional delays for bidders trying to achieve their goal of winning. There are two types of delay for customers in attaining their goal, procedural delays and correctional delays. Correctional delays are those that threaten the very purpose of the service encounter (Hui et al. 1998). This is the type of delay, the bidders are concerned with. Also time is a resource and waiting time might be perceived by customers as a waste of time (Larson 1987). So customers have to make decisions regarding the use of their time in shopping (Leclerc et al. 1995). Our study provides support for the hypothesis that customers care about the auction duration in the context of online sequential auctions, and it affects their decision to participate in the sequence or leave. On the other hand, other studies, e.g. Lucking-Reiley et al. (2007), show that longer auction durations result in an increase in the auction prices. Therefore, there is an interesting tradeoff for the auctioneers that need to be taken into account. Vakrat and Seidmann (1999a) indeed find that the auctioneer's profit function is unimodal concave with respect to auction duration.

In H7, we hypothesized that longer duration of activity in an auction sequence increases the attrition risk. We were not however able to find support for this hypothesis. Being involved in an auction sequence for a longer time means more experience and knowledge of the sequence and also more sunk cost. This sunk cost could potentially have both positive and negative impacts on a bidder's risk of attrition. Longer time of being active might negatively impact a bidder's perception of the expect payoff for future participation. On the other hand, longer activity could be conceptualized as a typical entrapment situation in which people are committed to an escalation process (Meyer 1994). The more time invested, the higher their motivation to confirm their prior decision to join the line and the more difficult it becomes for them to leave it. These effects could be counterbalanced by each other (Meyer 1994) and this could be the reason no significant effect is observed.

Our findings also support H8 about the multi-unit bidders suggesting that multi-unit bidders are at a lower risk of leaving the auction sequence as compared to the single unit bidders. Finally in hypothesis 9, we considered the effect of winning experience in the auction sequence on the chances of leaving and hypothesized that the multi-unit demand bidders who have already won at least one item are at a greater risk of leaving the sequence when they are unsuccessful in an auction, compared to the bidders who have never won. Very large and highly significant estimated coefficient of $WonDummy_{ij}$ supports our hypothesis.

3.5.2. Robustness Checks

In this section, we will evaluate the sensitivity of our model to some of the assumptions we made before, and through this we will gauge the validity of our findings.

3.5.2.1. Continuous Time Metric

As we argued before, due our lack of information on the exact timing of the bidders' decision to leave, and also due to the nature of the timing of the auctions in the sequence, it is more appropriate to use a discrete time hazard model. Nevertheless, in order to investigate the robustness of our results to the choice of time metric, we run two continuous time hazard models in this section. We define the beginning of time for a bidder as the opening time of the first auction a bidder participates in. So the length of time till event for a bidder is measured as the duration between the last bidding time and the beginning of time for him or her.

Table 17: Continuous Time Hazard Models

	(1) Cox Proportional Hazard Model	(2) Weibull Hazard Model
D1	3.26**	0.07
	(0.36)	(0.09)
D2	0.22	-3.30**
	(0.34)	(0.16)
D3	0.08	-3.56**
	(0.34)	(0.19)
D4	0.26	-3.45**
	(0.34)	(0.19)
D5	-0.41	-4.14**
	(0.39)	(0.26)
D6	-0.52	-4.26**
	(0.42)	(0.31)
D7	-1.87**	-5.66**
	(0.65)	(0.59)
D8		-3.65**
		(0.32)
MU	-2.11**	-2.08**
	(0.41)	(0.41)
TBA	0.14**	0.14**
	(0.03)	(0.03)
NormalizedLotSize	-0.09*	-0.09*
	(0.04)	(0.04)
AuctionDuration	0.12	0.08
	(0.06)	(0.05)
ReferencePrice	-0.08	-0.11*
	(0.06)	(0.06)
WTP	-1.67**	-1.69**
	(0.19)	(0.19)
WonDummy	1.58**	1.56**
	(0.48)	(0.48)
FinalBid	0.97**	1.05**
	(0.19)	(0.19)
*FinalBid*WonDummy*	-1.67**	-1.68**
	(0.31)	(0.31)
LL	-3172.29	-1491.59

Notes: The numbers in the parentheses are the standard errors for the estimated parameters.
* Significant at 0.05; ** significant at 0.01

Two continuous time models are fit to the data, a Cox proportional hazard model and a Weibull model. To be able to compare these models with the discrete time model, we have also included the time dummies. From Table 17, we can see that except for some of the time dummies, all other coefficients have the same signs as in the discrete time model. In the proportional hazard model, all the estimated coefficients except for those for *AuctionDuration* and *ReferencePrice* are significant, while in the Weibull model only *AuctionDuration* is not significant (with a p-value of 0.15). This ensures that our results our quite robust to the specification of time metric.

3.5.2.2. Alternative Censoring Approaches

In this section, we set up the model using two alternative censoring approaches. In the first approach, we only censor bidders who leave after winning, and assume that all bidders whose last participation does not coincide with winning have experienced the target event. This is an extreme approach that ignores the possibility of having some borrowers missing the opportunity of participating again due to the end of auction sequence. In the second approach we choose a different threshold for our statistical identification of censoring, i.e. instead of 0.01, we use 0.05 as the cutoff point. This is also a less conservative approach than the one we took before which identifies more events.

Table 18 provides the estimation results for models set up using these two approaches. Discrete time hazard models with a flexible baseline hazard are used. Most of the estimates are qualitatively the same as our main censoring scheme, except for *TBA* which is insignificant in (1) and (2) and *AuctionDuration* which is not significant in (1).

Table 18: Alternative Censoring Approaches

	(1) All non-winners~event	(2) Cutoff point=0.05
D1	0.01	-0.16
	(0.09)	(0.09)
D2	-0.08	-0.41**
	(0.12)	(0.12)
D3	-0.54**	-0.76**
	(0.16)	(0.17)
D4	-0.45*	-0.69**
	(0.20)	(0.20)
D5	-0.90**	-1.10**
	(0.25)	(0.25)
D6	-1.11**	-1.54**
	(0.32)	(0.33)
D7	-0.66*	-1.77**
	(0.33)	(0.41)
D8	-0.12	-0.43
	(0.37)	(0.37)
MU	-3.29**	-3.09**
	(0.37)	(0.37)
TBA	0.08	0.06
	(0.05)	(0.05)
NormalizedLotSize	-0.21**	-0.18**
	(0.05)	(0.05)
AuctionDuration	0.01	0.27**
	(0.06)	(0.06)
ReferencePrice	-0.14**	-0.18**
	(0.06)	(0.06)
WTP	-1.65**	-1.44**
	(0.17)	(0.17)
WonDummy	2.65**	2.39**
	(0.43)	(0.43)
FinalBid	1.45**	1.30**
	(0.17)	(0.17)
*FinalBid*WonDummy*	-2.50**	-2.42**
	(0.29)	(0.30)
LL	-1148.18	-1161.99

Notes: The numbers in the parentheses are the standard errors for the estimated parameters.
* Significant at 0.05; ** significant at 0.01

3.5.2.3. Alternative Link Functions

We used a discrete time hazard model with a logit link function in our main analysis. In this section, we will investigate the sensitivity of our findings to this assumption, by estimating the hazard model using two other link functions, i.e. the complementary log-log link and the probit link. Unlike the logit link, the complementary log-log link is asymmetric and has the advantage that it has a proportional hazard assumption, as opposed to a proportional odds assumption of the logit link (Singer and Willett 2003).

Table 19: Alternative specifications of the discrete hazard model

	(1) CLog-log Link Model	(2) Probit Link Model
D1	-0.93**	-0.38**
	(0.08)	(0.05)
D2	-1.95**	-1.01**
	(0.14)	(0.08)
D3	-1.53**	-0.80**
	(0.16)	(0.11)
D4	-1.15**	-0.47**
	(0.16)	(0.12)
D5	-1.73**	-0.93**
	(0.24)	(0.15)
D6	-1.82**	-0.97**
	(0.30)	(0.18)
D7	-3.12**	-1.71**
	(0.58)	(0.29)
D8	-1.03**	-0.44*
	(0.30)	(0.21)
MU	-2.59**	-1.55**
	(0.41)	(0.20)
TBA	0.11**	0.10**
	(0.03)	(0.03)
NormalizedLotSize	-0.15**	-0.10**
	(0.04)	(0.03)
AuctionDuration	0.34**	0.25**
	(0.05)	(0.04)
ReferencePrice	-0.41**	-0.27**
	(0.06)	(0.04)
WTP	-1.65**	-0.81**
	(0.20)	(0.10)
WonDummy	2.02**	1.14**
	(0.48)	(0.24)
FinalBid	1.49**	0.71**
	(0.20)	(0.10)
*FinalBid*WonDummy*	-2.06**	-0.91**
	(0.33)	(0.18)
LL	-989.74	-1008.88

Notes: The numbers in the parentheses are the standard errors for the estimated parameters.
* Significant at 0.05; ** significant at 0.01

Comparing the estimates in Tables 16 and 19, it is clear that the estimates are qualitatively similar, even the significance levels and relative sizes of the estimated coefficients are similar. Therefore we can safely conclude that our findings are robust to the specification of the discrete time hazard model.

Table 20: Discrete Hazard Models with Logit Link and Polynomial Time

	(1) Linear Time	(2) Quadratic Time	(3) Cubic Time	(4) Quartic Time
Intercept	-0.73**	-0.23	0.39	4.54**
	(0.11)	(0.18)	(0.33)	(0.76)
T	-0.16**	-0.60**	-1.40**	-8.27**
	(0.04)	(0.13)	(0.38)	(1.19)
T^2		0.06**	0.31**	3.66**
		(0.02)	(0.11)	(0.56)
T^3			-0.02*	-0.63**
			(0.01)	(0.10)
T^4				0.04**
				(0.01)
MU	-2.81**	-2.80**	-2.80**	-2.86**
	(0.43)	(0.43)	(0.43)	(0.43)
TBA	0.16**	0.15**	0.15**	0.15**
	(0.05)	(0.05)	(0.05)	(0.05)
NormalizedLotSize	-0.17**	-0.17**	-0.17**	-0.18**
	(0.05)	(0.06)	(0.06)	(0.06)
AuctionDuration	0.40**	0.41**	0.42**	0.43**
	(0.06)	(0.06)	(0.06)	(0.06)
ReferencePrice	-0.42**	-0.46**	-0.47**	-0.49**
	(0.07)	(0.07)	(0.07)	(0.07)
WTP	-1.83**	-1.74**	-1.71**	-1.76**
	(0.22)	(0.22)	(0.22)	(0.22)
WonDummy	2.23**	2.15**	2.14**	2.19**
	(0.50)	(0.50)	(0.50)	(0.51)
FinalBid	1.64**	1.56**	1.52**	1.58**
	(0.22)	(0.21)	(0.21)	(0.22)
*FinalBid*WonDummy*	-2.22**	-2.15**	-2.14**	-2.20**
	(0.37)	(0.37)	(0.37)	(0.38)
LL	-1028.57	-1022.24	-1019.73	-999.18
AIC	2079.14	2068.48	2065.46	2026.36
BIC	2228.30	2231.20	2241.74	2216.20

Notes: The numbers in the parentheses are the standard errors for the estimated parameters.
* Significant at 0.05; ** significant at 0.01

3.5.2.4. Polynomial Modeling of Time

In Table 20, we have provided the estimates for models in which the effect of discrete time is captured through polynomial time functions instead of a flexible characterization. Comparing the AIC and BIC information criteria of the four models, model (4) provides the best fit even accounting for the additional number of parameters required. This was to be expected given the form of the baseline hazard model in Figure 2. More importantly, the model estimates in all four models are virtually the same as those of the flexible baseline hazard model confirming that our findings are not affected by the specification of the time metric.

3.6. Conclusion

Our research is the first study to empirically detect and investigate the phenomenon of bidders attrition in sequential online auctions. Although in sequential auctions, bidders who do not win in an auction have the option to participate in future auctions for the item, only less than fifty percent of unsuccessful bidders return to try again. Similar to other online retailers, there is a fierce competition between online auctioneers, and their profitability depends upon the number of participating bidders. Therefore attracting new bidders and retaining the existing ones is critical for the success of online auction houses. Auction houses that do a better job in retaining their current bidders save on the costs of the usually more expensive activity of brining in new bidders.

In this research, we used survival analysis to investigate the impact of bidders' perception of market supply, bidders' perception of market demand, bidders' expected cost of participation and the effect of winning on the risk of bidder attrition. We found support that all these categories of factors could impact a bidder's decision to leave the sequence of auctions before having his or her demand satisfied. The bidders' perception of the environmental variables shaped through their previous participations seems to be particularly important. Some of the findings of our study can be generalized to the business-to-consumer sub-sector of the online auction market.

Chapter 4 An Analytical Study of Competition in Cloud Computing Market

4.1. Literature Review

This study is related to the existing research on cloud computing, outsourcing of information technology, oligopoly competition and differential games. Although cloud computing has received a lot of attention from practitioners, to the best of our knowledge, the existing body of literature is mostly created by computer scientists and not much has been done by information systems and economics researchers. There are studies on the economics of cloud computing (Hamilton 2010; IDG-Enterprise 2012; Harms and Yamartino 2010), but they are mostly surveys or subjective views by practitioners.

Youseff et al (2008) propose a detailed ontology for the cloud and Armbrust et al (2009) provide a holistic view of cloud computing and discuss various issues that are relevant such as the history, definition, types and applications of cloud computing. Some computer scientists have worked to provide the architecture for creating clouds and present a vision for the creation of global Cloud exchange for trading services (Buyya et al. 2009). There has also been research on comparing cloud computing with grid computing (Foster et al. 2008).

Cloud computing is essentially a new approach to technology outsourcing. In their seminal paper, Loh and Venkatraman integrate business and IT perspectives and develop a model of the determinants of IT outsourcing and they empirically test their model (Loh and Venkatraman 1992a). In another study, they conduct an empirical analysis using diffusion modeling to show that the most important motivator for the adoption of IT outsourcing is internal influence (Loh and Venkatraman 1992b). Lacity and Willcocks (1998) conduct an extensive case study on 61 IT outsourcing projects and identify some best practices such as importance of selective outsourcing decisions or the higher success rate of short-term contracts as opposed to long-term ones.

Numerous economists and researchers in information systems and marketing have studied various aspects of firms’ competition. Many of them are based on the seminal works of Bertrand (Bertrand 1883), Cournot (Cournot 1897) and Hotelling (Hotelling 1929) among others. Moorthy (1988) studies the role of consumer preferences and price competition in determining the competitive strategies of firms and Barua et al utilize a duopoly model of quality competition to analyze the strategic impact of IT investments (Barua et al. 1991).

Differential games belong to a subclass of dynamic games called state space games in which the modeler introduces a set of variables to describe the state of a dynamic system at any particular instant during play (Sethi and Thompson 2005). They originated as an extension of optimal control and have been used in various areas such as economics and management science. We use a dynamic model with two state variables, market share and capacity, an accumulative state variable, similar to the goodwill accumulation model of Horsky (Horsky 1977). The difference however is that in our model, capacity is not only a driver for market share, but also a constraint for the amount of demand that each firm can satisfy. Our model is also similar to capital accumulation games (Spence 1979; Fundenberg and Tirole 1991) in that we consider the effect of accumulated capacity as a quality signal that has an impact on the competition for market share.

4.2. Two-Period Model of Competition

4.2.1. Model Settings

In this benchmark model, we model the competition between two cloud computing providers, firms 1 and 2. Each of the firms offers a differentiated product (service), which are imperfect substitutes. There are two types of customers. Type 1 customers have an inherent preference for firm 1's product while type 2 customers prefer the product of firm 2. There are two stages to the game. There is a market of size 1 at the first stage, supplemented by another market of size 1 at the second stage. The percentage of type 1 customers in periods 1 and 2 are assumed to be β_1 and β_2 respectively, so we have $0 < \beta_1, \beta_2 < 1$.

Customers perceive a provider's capacity as a signal of quality, while different customers value it differently. It is captured through a disutility term which is small for the product of a firm with a large capacity. We assume that customers of each type in each period are uniformly distributed on a scale between 0 and 1, in which 0 means they don't value quality and 1 represents the greatest valuation for quality. Capacity is defined as the underlying cloud infrastructure including networking, servers, server racks and storage, and the staff and administration effort required to run and support the cloud computing services. Here we are not taking into account the location of the data centers, as they are capital-intensive investments that are best modeled as being made at discrete points in time. Our focus is on smaller investments that could be captured as being done continuously over time.

Capacity directly affects the quality of service. As compared to organizational data centers, cloud computing providers enjoy less variation in demand, thanks to the large number of customers that they have which helps to smooth out the overall demand and reduce the uncertainty. Nevertheless, there is still considerable uncertainty in the demand for computational power that they observe. In order to deal with this issue, cloud computing providers have to over-provide capacity. A larger capacity means the providers are able to cope with unexpected peaks in demand without having some of the services facing disruption or slowing down which is a major concern of the customers. Having a larger capacity also means that the firms are able to provide a reliable and consistently fast service. Although the customers might not care so much about a provider's capacity per se, they value the quality of service resulting from it. To simplify the model, we don't explicitly model the link between capacity and quality. Instead, we model the customers' valuation for quality through the valuation placed on capacity. So the cloud computing providers face a trade-off between the cost of capacity building and the two benefits of being able to serve more customers and also offering a higher quality of service that helps the providers expand their market.

The evolution of the game is such that knowing the structure of the market in each period, the firms first simultaneously deicide on their capacities. Next, each firm learns about the other firm's capacities, and then they simultaneously decide on prices for their products. Finally, the customers observe the capacities and the prices, and select the product which provides them with a higher net utility. The utility of a type i customer, $i \in \{1,2\}$, acquiring the services of each of the two firms is defined as:

$$U_i^i = u - \frac{\alpha x}{\mathrm{K}_i} - p_i \tag{22}$$

$$U_j^i = u' - \frac{\alpha x}{\mathrm{K}_j} - p_j \,,\; i \neq j \tag{23}$$

We have $u > u'$ meaning that a type i customer has an inherent preference for firm i's product. K_i and p_i are respectively firm i's price and capacity, and α is a scaling factor. $0 \leq x \leq 1$

models a customer's valuation for quality. We assume that when a type i customer buys firm j's product in period 1, her type changes to j for the second period. This is because of the switching cost associated with buying the other firm's product in period 2. The firms' profits in each period are calculated through multiplying their market size by their marginal profit net of the cost of capacity. The first stage and second stage profit functions of firm 1 are:

$$\pi_{11} = (D_{11}{}^{1} + D_{11}{}^{2})(p_1 - c_{11}) - c_{12} f(\mathrm{K}_{11}) \tag{24}$$

$$\pi_{12} = (D_{11}{}^{1} + D_{11}{}^{2} + D_{12}{}^{1} + D_{12}{}^{2})(p_1 - c_{11}) - \delta c_{12} f(\mathrm{K}_{12}) \; , \; 0 < \delta < 1 \tag{25}$$

$D_{ij}{}^{k}$ is the market share of firm i from the segment of type k customers in period j. The parameter δ captures the well-known effect of Moore's law (Moore 1965). According to Moore's law, hardware computing power almost doubles approximately every two years, an observation that has been more or less holding since 1965 and is predicted to hold for the next few years. A consequence of the Moore's law is that the cost of computing capacity goes down over time, which we capture using δ in our model. δ is assumed to be between 0 and 1 which means that the cost of capacity provision decreases in the second period.

4.2.2. Equilibrium and Analysis

There are three potential scenarios in the equilibrium at each period. The first scenario is a situation in which one of the firms is dominant. Without loss of generality we assume that firm 1 is the dominant firm. In this scenario, which we call the domination scenario, firm 1 serves all of the type 1 customers and also a fraction of the type 2 customers. Those are customers that have a higher valuation for quality of service. In the second scenario, which we call the segmentation scenario, firm 1 serves all of the type 1 customers and firm 2 serves all of the type 2 customers. In other words, none of the firms manages to gain market share from the market that has an inherent preference for the other firm's product. Finally in the last possible scenario, the total domination scenario, one of the firms is pushed out of the market while the other firm captures the whole market. It requires the dominant firm's offering to be so attractive that customers in both markets choose it. Since in this scenario, the model reduces to the trivial case of a monopoly, we don't consider it.

The derivation of the symmetric subgame Nash equilibrium is through the standard backward induction procedure, i.e. we first consider the customers' decision in period 2. Then derive the firms' pricing decision given the customers' decision. Next, we derive the capacity decisions of the firms in the second period and using the second period best responses, we repeat the same procedure for the first period. Proposition 1 provides the main result of the analysis. The proof which is omitted is a set of steps of algebraic analysis (the proof is available upon request).

Proposition 1: In the presence of economies of scale, i.e. when we have non-increasing marginal cost of capacity provision, even in the case when the cost of capacity provision $f(\mathrm{K})$ is linear in the size of capacity, the only viable duopoly equilibrium is the segmentation scenario in which firm 1 serves all type 1 customers and all type 2 customers are served by firm 2. The domination scenario, in which one of the firms manages to capture a part of the other firm's market, is only realized when the cost is a strictly convex function of the capacity implying diseconomies of scale.

4.3. Basic Differential Game Model of Competition

4.3.1. Model Settings and Assumptions

In the previous section, we built a benchmark, two-period model of competition. In this section, the competition between the two cloud computing providers is modeled as a dynamic game. The firms are assumed to be forward-looking, in the sense that they not only consider the impact of their decisions on the current-period payoff, but also take into account the long-term implications of their decisions and the effect of these decisions on their future payoffs. We use differential games to study the dynamic and forward-looking behavior of the two firms. In order to capture the most salient features of the cloud computing market with a model that is also tractable, we have made simplifications compared to the previous model, and focused on the most important aspects. Therefore, we are assuming that the firms' only decision variable is the rate at which they acquire additional capacity.

We have used a model of duopoly competition in which the two firms, firm 1 and firm 2, each offer cloud computing services that are horizontally differentiated. As an example, and in order to help better understand the model, one can think of the firms as being Amazon.com and Microsoft and the products as being an IaaS offering of Amazon.com and a PaaS offering by Microsoft. Prices of the cloud offerings are assumed to be exogenous and determined by the market through competition. The model can to some extent be thought of as an extension of the Cournot competition model (Singh and Vives 1984) in which firms decide on their capacity, while prices are determined according to the demand and supply, so that the market would clear. There are however fundamental differences between our model and the Cournot competition model, as in our model, total demand does not depend on the suppliers' capacity, rather capacity is a constraint that the firms have to satisfy in order to be able to fulfill the market demand. We also assume that the marginal profit for each provider, m_i which is calculated as price less the marginal cost of service provision, is a constant.

There is a market for the cloud services, $D(t)$, which for now we assume is fixed and normalized to one. We also assume that all of the customers subscribe to the cloud services, so the whole market is served by the two firms. Over time, some customers switch to the competing provider. The decision to switch is governed by a stochastic process and the probabilities are affected by the quality of service (capacity) of each firm and two additional parameters α_{ij} and α_{ji}. The size of these two parameters is determined by factors such as switching costs for the customers.

In a differential game, the state of the game at each time is determined through the values of some variables referred to state variables. The state variables that represent the state of the our model at each point in time, t, are capacity and market shares of each firm at t. We use a Lanchester model (Kimball 1957) for the market share competition similar to the one Horsky uses to model competition and goodwill accumulation (Horsky 1977). The way to think about market share changes in this model is that the changes in market shares are stochastic and the transition probabilities depend on capacity and firm specific variables α_{ij} and α_{ji}. The state variables evolve over time according to the following two equations for firm i:

$$\dot{D}_i = \alpha_{ji}\mathrm{K}_i(t)D_j(t) - \alpha_{ij}\mathrm{K}_j(t)D_i(t) \rightarrow \dot{D}_i = \alpha_{ji}\mathrm{K}_i(t)\left(1 - D_i(t)\right) - \alpha_{ij}\mathrm{K}_j(t)D_i(t) \tag{26}$$

$$\dot{\mathrm{K}}_i = \kappa_i(t) \tag{27}$$

We also have the following initial values for the state variables which are the market shares and capacities for the two firms at time zero:

$$\mathrm{K}_i(0) = \mathrm{K}_{i,0} \ , \ D_i(0) = D_{i,0} \tag{28}$$

4.3.2. Model and Equilibrium Derivation

At each point in time, each of the two firms continuously makes a decision about the rate of capacity provision. In other words according to the different games terminology, the capacity provision rate is the control variable while capacity and market shares of each of the firms at that time are the state variables. The control variable at each time is chosen to maximize the profit of the firm at the current period and in future. The profit function for firm i is modeled using the functional of equation (29). Each of the firms maximizes this function subject to constraints (30), (31), (32) and (33).

$$\Pi_i = \int_0^\infty e^{-\rho_i t}\left(m_i D_i(t) - c_i \kappa_i(t)\right)dt \qquad \text{for } i = 1,2 \tag{29}$$

s.t.

$$C_i(t)\kappa_i(t) \le \bar{\kappa}_i \tag{30}$$

$$K_i(t) \ge \sigma D_i(t) \tag{31}$$

$$\kappa_i(t) \ge 0 \tag{32}$$

$$K_i(t) \ge 0 \tag{33}$$

We intend to determine the optimal trajectory of the capacity provision rate, $\kappa_i^*(t)$, for each firm in a way that it maximizes the profit functional (29) at any point in time subject to the constraints, and is the best response to the other firm's trajectory. In other words, if each of the firms chooses $\kappa_i^*(t)$, none of them will have an incentive to deviate. These rates constitute what is referred to as the open-loop Nash equilibrium of this game. An open-loop Nash equilibrium means that the choice of the control variables at each point in time is only a function of time. Intuitively this implies that the firms decide their capacity provision rates at the beginning of the game and commit to the optimal control trajectory throughout the game. This type of equilibrium is especially suitable for games in which it is hard for the participants to measure the state variables over time (Dockner et al. 2001).

An alternative equilibrium concept is the closed-loop Nash equilibrium or the Markov perfect Nash equilibrium. In the Markov perfect Nash equilibrium, when making their decisions, each player considers not only time, but also the values of the state variables in making the optimal decisions. In this section, we focus on the class of symmetric open-loop equilibria and use Pontryagin's maximum principle (Shell 1968; Dockner et al. 2001) to analyze the game. In order to do so, we form the Hamiltonian function for firm 1 as following. The analysis for firm 2 is similar due to symmetry and is therefore omitted.

$$\begin{aligned} &H(D_i(t), \mathrm{K}_i(t), \kappa_1(t), \lambda_1(t), \lambda_2(t), t) = \\ &\quad m_1 D_1(t) - c_1 \kappa_1(t) + \lambda_1(t)\left[\alpha_{21}\mathrm{K}_1(t) D_2(t) - \alpha_{12}\mathrm{K}_2(t) D_1(t)\right] + \lambda_2(t)\left[\kappa_1(t)\right] \end{aligned} \tag{34}$$

The necessary condition of optimality is satisfied when we maximize the Hamiltonian with respect to the control variable, while the sufficient condition holds since the Hamiltonian is concave in the control variable. We can maximize the Hamiltonian by taking the first-order derivative of it with respect to the control variable. Therefore we will have the optimal rates of capacity provision as characterized in Proposition 2.

Proposition 2: As the Hamiltonian in equation (34) is linear in $\kappa_1(t)$, the optimal trajectory of the control variable is:

$$\rightarrow \kappa_1^*(t) = \begin{cases} 0 & , if\, \lambda_2 < c_1 \\ \bar{\kappa}_1 / c_1 & , if\, \lambda_2 > c_1 \\ Singular & , if\, \lambda_2 = c_1 \end{cases} \tag{35}$$

λ_2 could be interpreted as the shadow price of an extra unit of capacity, so if this shadow price is greater than the cost of capacity, the firm builds as much capacity as possible, while it won’t purchase any if the shadow price net of the capacity cost is negative. The optimal solution however has to also satisfy the constraints (30) to (33). We know that we have $\mathrm{K}_i(0)=\mathrm{K}_{i,0}\geq 0$. Also the structure of the optimal solution in Proposition 2 guarantees that constraint (32) is satisfied. These two coupled with each other guarantee that constraint (33) will be automatically satisfied. However, to ensure that the other constraints are satisfied, we need to form the Lagrangian (Sethi and Thompson 2005). Constraint (30) is the budget constraint which we can directly include in the Lagrangian along with constraint (32). However as for constraint (31), the capacity constraint, because it's a pure constraint that does not include the control variable, we have to make a transformation before we can include it in the Lagrangian function.

The parameter σ $(\sigma \geq 1)$ in the capacity constraint captures the fact that the firms have to overprovision in order to deal with the uncertainty in demand. In other words, we are not explicitly modeling the probabilistic nature of the demand uncertainty, but are using a scale factor to account for it in a deterministic way. Consequently σ is a function of the standard deviation of the demand, and the higher the demand uncertainty, the greater will σ be. To make sure that (31) always holds, we have to have $\kappa_i(t)-\sigma\dot{D}_i(t)\geq 0$ whenever $K_i(t)-\sigma D_i(t)=0$ (Sethi and Thompson 2005).

Now we are ready to form the Lagrangian function as following:

$$\begin{aligned} L = H + \mu g &= m_1 D_1(t) - c_1\kappa_1(t) + \lambda_1\left[\alpha_{21}\mathrm{K}_1(t)D_2(t) - \alpha_{12}\mathrm{K}_2(t)D_1(t)\right] + \lambda_2\left[\kappa_1(t)\right] \\ &+\mu_1\left(\bar{\kappa}_1 - c_1\kappa_1(t)\right) + \mu_2\kappa_i(t) + \mu_3\left(\kappa_1(t) - \sigma\left(\alpha_{21}\mathrm{K}_1(t)D_2(t) - \alpha_{12}\mathrm{K}_2(t)D_1(t)\right)\right) \end{aligned} \tag{36}$$

s.t.

$$\mu_1 \geq 0, \mu_1\left(\bar{\kappa}_1 - c_1\kappa_1(t)\right) = 0 \tag{37}$$

$$\mu_2 \geq 0, \mu_2\kappa_1(t) = 0 \tag{38}$$

$$\mu_3 \geq 0, \mu_3\left(\mathrm{K}_i(t) - \sigma D_i(t)\right) = 0, \mu_3\left(\kappa_1(t) - \sigma\left(\alpha_{21}\mathrm{K}_1(t)D_2(t) - \alpha_{12}\mathrm{K}_2(t)D_1(t)\right)\right) = 0 \tag{39}$$

Based on the Lagrangian function above, we will define the following adjoint first-order differential equations for the adjoint variables, λ_1 and λ_2 (Dockner et al. 2001):

$$\dot{\lambda}_1(t) = \rho_1\lambda_1(t) - m_1 + \lambda_1(t)\left[\alpha_{21}\mathrm{K}_1(t) + \alpha_{12}\mathrm{K}_2(t)\right] - \mu_3\sigma\left[\alpha_{21}\mathrm{K}_1(t) + \alpha_{12}\mathrm{K}_2(t)\right], \tag{40}$$

$$\lim_{t\to\infty}\lambda_1(t) = 0$$

$$\dot{\lambda}_2(t) = \rho_1\lambda_2(t) - \lambda_1(t)\alpha_{21}D_2(t) + \sigma\mu_3\alpha_{21}D_2(t), \tag{41}$$

$$\lim_{t\to\infty}\lambda_2(t) = 0$$

The necessary condition of optimality dictates that the derivative of the Lagrangian with respect to the control variables must be zero, so we have:

$$\partial L / \partial\kappa_1 = \mu_2 + \mu_3 + \lambda_2 - c_1 - c_1\mu_1 = 0 \;\rightarrow \mu_2 + \mu_3 + \lambda_2 = c_1 + c_1\mu_1 \tag{42}$$

4.3.3. Analysis and Discussion

The model presented in the previous section is an autonomous or stationary model, because none of the state equations nor the profit function explicitly depend on time variable (Dockner et al. 2001). Therefore, there the two firms will eventually reach a steady state after which the market shares and capacities will not change.

Proposition 3: Our model is autonomous and has the stationary property, meaning that the system will enter a stationary state in the long-run and remain there. In other words, in the long-run we will have:

$$\dot{D}_i = \alpha_{ji}\mathrm{K}_i\left(1 - D_i\right) - \alpha_{ij}\mathrm{K}_j D_i = 0 \rightarrow D_i = \frac{\alpha_{ji}\mathrm{K}_i}{\alpha_{ji}\mathrm{K}_i + \alpha_{ij}\mathrm{K}_j} \tag{43}$$

$$\dot{\mathrm{K}}_i = \kappa_i(t) = 0 \tag{44}$$

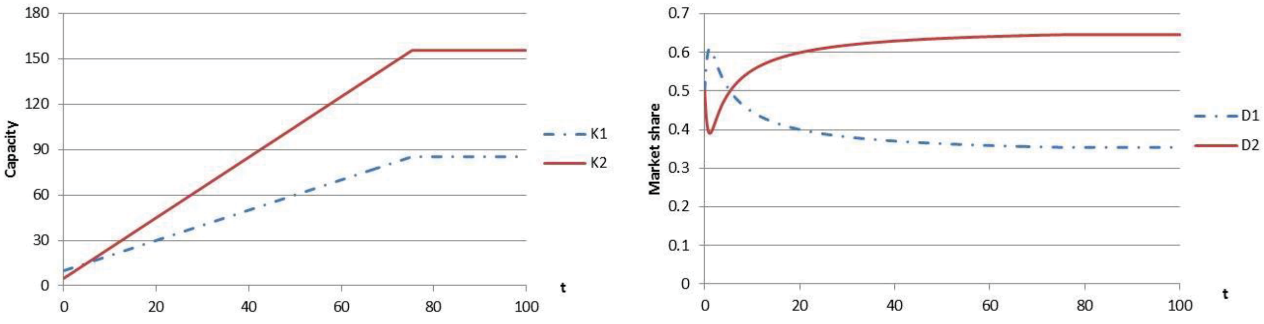

Figure 4: Capacity and Market Share Evolution Over Time

Once the system enters the steady state, it will remain there and as the market shares do not change, all the problem constraints will also remain satisfied. It is important to note that in this model we are assuming that the total market size is fixed, an assumption that is critical to have an autonomous model. This can be further confirmed considering the optimal trajectory (35) and the limiting condition (41). For values of t large enough, we will eventually get to a point in which $\lambda_2 < c_1$, so the firms stop building capacity. Values of the Lagrange multipliers could also be determined through equation (42). Next we provide the results of the analysis for the optimal values of the state variables in the transient period. To do so, we use the structure of the optimal rate of capacity provision and divide the transient period for each firm into provision periods and non-provision periods.

Proposition 4: If we assume that we have $\kappa_1^*(t) = 0$ for all $\underline{\tau}_j^1 \le t \le \overline{\tau}_j^1$ (the j^{th} period of non-provision), and $\kappa_1^*(t) = \overline{\kappa}_1 / c_1$ for $\overline{\tau}_j^1 \le t \le \underline{\tau}_{j+1}^1$ (the j^{th} period of maximum capacity provision), we will have:

$$\dot{\mathrm{K}}_1 = 0 \qquad \rightarrow \mathrm{K}_1^*(t) = \mathrm{K}_1(\underline{\tau}_j^1) \qquad \text{for } \underline{\tau}_j^1 \le t \le \overline{\tau}_j^1 \tag{45}$$

$$\dot{\mathrm{K}}_1 = \frac{\overline{\kappa}_1}{c_1} \qquad \rightarrow \mathrm{K}_1^*(t) = \frac{\overline{\kappa}_1}{c_1} t + \mathrm{K}_1^*(\overline{\tau}_j^1) - \frac{\overline{\kappa}_1}{c_1} \overline{\tau}_j^1 \qquad \text{for } \overline{\tau}_j^1 \le t \le \underline{\tau}_{j+1}^1 \tag{46}$$

Due to symmetry, the same two equations hold for firm 2. Now if we want to derive the market shares at a point in time for each firm, we need to plug (45) and (46) into the state equation (26). Depending on the value of t, we have four different situations. These are the combinations of the situations in which each of the firms is either building maximum capacity or not building capacity at all. In the next two propositions, we have provided the solutions for the case in which both firms are not provisioning and the case in which firm 1 is not provisioning while firm 2 is provisioning at the maximum rate.

Proposition 5: For $\underline{\tau}_j^1 \le t \le \overline{\tau}_j^1$ and $\underline{\tau}_k^2 \le t \le \overline{\tau}_k^2$, the market size of firm 1 is:

$$D_1(t) = \frac{\alpha_{21} \mathrm{K}_1(\underline{\tau}_j^1)}{\left(\alpha_{21} \mathrm{K}_1(\underline{\tau}_j^1) + \alpha_{12} \mathrm{K}_2(\underline{\tau}_k^2)\right)} + cons_1 e^{-\left(\alpha_{21} \mathrm{K}_1(\underline{\tau}_j^1) + \alpha_{12} \mathrm{K}_2(\underline{\tau}_k^2)\right)t} \tag{47}$$

Proposition 6: For $\underline{\tau}_j^1 \le t \le \overline{\tau}_j^1$ and $\overline{\tau}_k^2 \le t \le \underline{\tau}_{k+1}^2$, the market size of firm 1 is:

$$D_1(t) = \sqrt{\frac{\pi c_2}{2\alpha_{12}\overline{\kappa}_2}} \alpha_{21} \mathrm{K}_1(\underline{\tau}_j^1) e^{\frac{c_2\left(\alpha_{21}\mathrm{K}_1(\underline{\tau}_j^1)+\alpha_{12}\left(+\mathrm{K}_2{}^*(\overline{\tau}_k^1)-\frac{\overline{\kappa}_2}{c_2}\overline{\tau}_k^1\right)\right)^2}{2\alpha_{12}\overline{\kappa}_2}} * \tag{48}$$

$$(erf(\sqrt{\frac{\alpha_{12}\overline{\kappa}_2}{2c_2}} + \frac{\alpha_{21}\mathrm{K}_1(\underline{\tau}_j^1) + \alpha_{12}\left(\mathrm{K}_2{}^*(\overline{\tau}_k^1) - \frac{\overline{\kappa}_2}{c_2}\overline{\tau}_k^1\right)}{\sqrt{2\alpha_{12}\overline{\kappa}_2 / c_2}}) + cons2)e^{\frac{t}{2}\left(\frac{\overline{\kappa}_2}{c_2}t+2\left(\alpha_{21}\mathrm{K}_1(\underline{\tau}_j^1)+\alpha_{12}\left(+\mathrm{K}_2{}^*(\overline{\tau}_k^1)-\frac{\overline{\kappa}_2}{c_2}\overline{\tau}_k^1\right)\right)\right)}$$

Where $erf(x)$ is the error function defined as $erf(x) = 2/\sqrt{\pi}\int_0^x e^{-t^2}dt$. Now based on propositions 1 to 5, we are in a position to propose the following observations.

Observation 1: Assuming that the initial capacities and the budget constraints are such that the firms never have a problem satisfying the optimal demands, the initial market shares of the firms are irrelevant in determining the steady-state market share structures and steady-state capacities of the firms.

Observation 2: Initial capacities of the firms, the budget constraint to capacity cost ratio and market switching capacities are the factors that affect and determine the steady-state values of the state variables.

Observation 3: Assuming that during the transient period, both firms have at least one full-provision period, the end points of the last full-provision periods of the two firms coincide. In other words, there is a point in time, τ, at which both firms stop provisioning. τ is the beginning of the steady-state for the market after which the market shares and capacities of the two firms will not change.

4.4. Comprehensive Differential Game Model of Competition

In this section, we relax some of the restricting assumptions that we made in section 4.3 to build a more realistic model of competition. In our basic model, we assumed that the total market size is fixed and it is covered at all times by the two firms. So the firms just compete and try to grab the other firm's customers. In our comprehensive model, we relax this assumption and assume that the total number of customers adopting cloud computing could change over time. We do it in the same spirit as Fruchter (1999), such that we assume there is a saturation level of the market which is the largest possible market size, which we normalize it to 1 without loss of generality. A part of this market is covered by the two competing cloud computing providers, $D_1 + D_2$, and a part of it up, $1-(D_1 + D_2)$, for grabs.

There are three ways for firm *i*'s total demand to change. It changes when some of the other firm's customers switch to firm *i*'s product, it changes when firm *i* loses customers to the other firm, and it changes when firm *i* manages to sell its product to some customers who are not using cloud computing services yet. Therefore we can rewrite the state equation for changes in the market size of firm *i* as following:

$$\dot{D}_i = \alpha_i(\mathrm{K}_i(t) - \sigma D_i(t))(1 - D_i(t) - D_j(t)) + \alpha_{ji}\mathrm{K}_i(t)D_j(t) - \alpha_{ij}\mathrm{K}_j(t)D_i(t) \tag{49}$$

Another change in this comprehensive model is that here, we assume that capacity of each firm depreciates over time. This is an important assumption as it changes the structure of the model. As capacity depreciates over time, the total capacity of the firms changes which means that the model will no longer be stationary. It is however an essential assumption as in practice, the hardware and equipment depreciate, and not modeling it may result in deriving naïve and unrealistic results. With this assumption, the state equation for capacity changes to the following equation:

$$\dot{\mathrm{K}}_i = \kappa_i(t) - \delta\mathrm{K}_i(t) \tag{50}$$

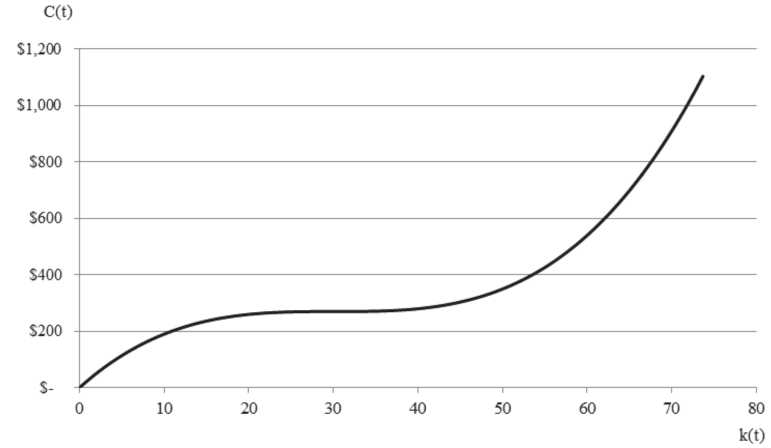

Figure 5: An Example of the Total Cost of Capacity at Time t

Another important change in our comprehensive model is that we relax the assumption that cost of capacity provision is constant over time. To take in to account the consequences of Moore's law and accommodate this important feature of technology markets, we assume that the cost of capacity provision falls over time. Moreover, we assume that the marginal cost of capacity provision at any point in time decreases for up to some level of capacity, and then starts increasing. In other words, there is a threshold below which, there exists economies of scale, and above it, we have diseconomies of scale. Figure 5 provides an example of such a cost function. We assume the following cost function which incorporates both of these notions if all of the parameters have non-negative values:

$$C_i(t) = \left(\theta_{i1} + \theta_{i2} e^{-\theta_{i3} t}\right)\left(\left(\kappa_i(t) - \omega_i\right)^3 + \omega_i^3\right) \tag{51}$$

Finally, we also allow the time horizon to be finite. Given these new assumptions, the profit function of firm *i* needs to be updated as following:

$$\Pi_i = \int_0^T e^{-\rho_i t}\left(m_i D_i(t) - C_i(t)\right)dt + e^{-\rho_i T}(g(K_i(T)) + h(D_i(T))) \tag{52}$$

The same four constraints that we had for our basic model still hold. We can derive the optimal trajectory of the rate of capacity provision by forming the Hamiltonian function, similar to what we did for the basic model. Doing this, we will have:

$$\rightarrow \kappa_1^*(t) = \begin{cases} 0 & , if\ \lambda_{i1} < \left(\theta_{i1} + \theta_{i2} e^{-\theta_{i3} t}\right)\omega_i \\ \bar{\kappa}_1 / c_1 & , if\ \lambda_{i1} \geq \left(\theta_{i1} + \theta_{i2} e^{-\theta_{i3} t}\right)\omega_i \\ Singular & , if\ \lambda_2 = c_1 \end{cases} \tag{53}$$

The optimal rate of provision depends on the values of the adjoint variables. However given the more complicated structure of the problem, a closed-form solution for these variables does not exist. So we have to resort to numerical analysis to get some insight. The algorithm used to perform the numerical analysis is quite straightforward and has the following steps:

1. Set initial values for controls and calculate the states over time.
2. Using the transversality conditions and the current values of the states, calculate the adjoint variables backward.
3. Using the values of the adjoint variables, calculate the new controls and states, and calculate the profits.
4. Iterate 3 and 4, until the profits stop improving

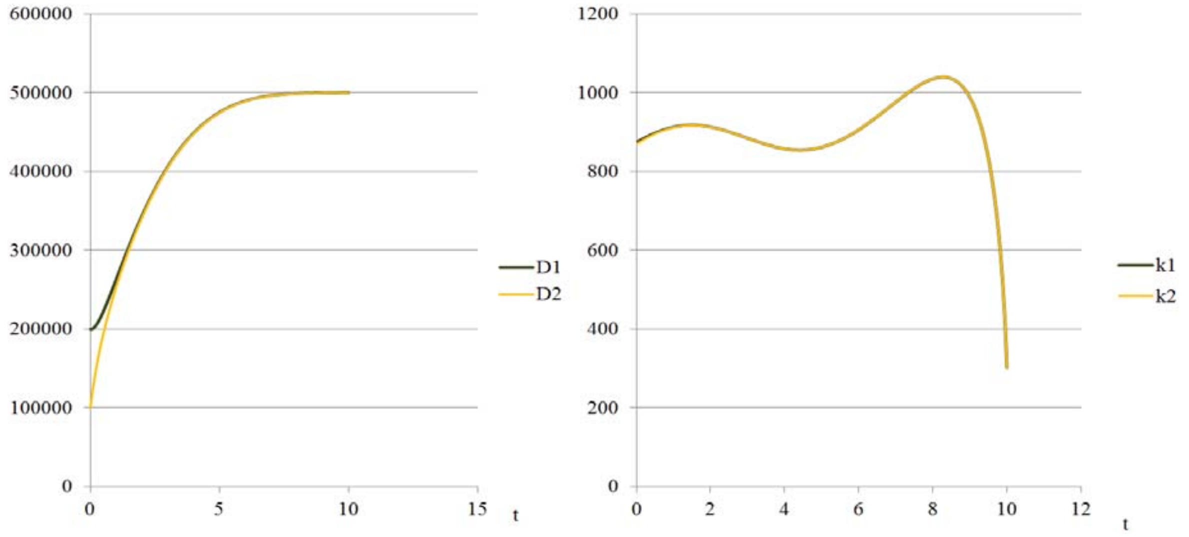

Figure 6: Firm 1 with More Initial Market Share

Using the above algorithm, we solved a sample problem numerically. Here, we will provide the results using graphs, and briefly discuss them. Figure 6 depicts the optimal trajectory of the rate of capacity provision and the evolution of market shares in a case that the two firms are similar in every aspect, but firm 1 has a higher initial market share. As we can see, in this case, the final market sizes will eventually converge and the initial difference in the market sized disappear.

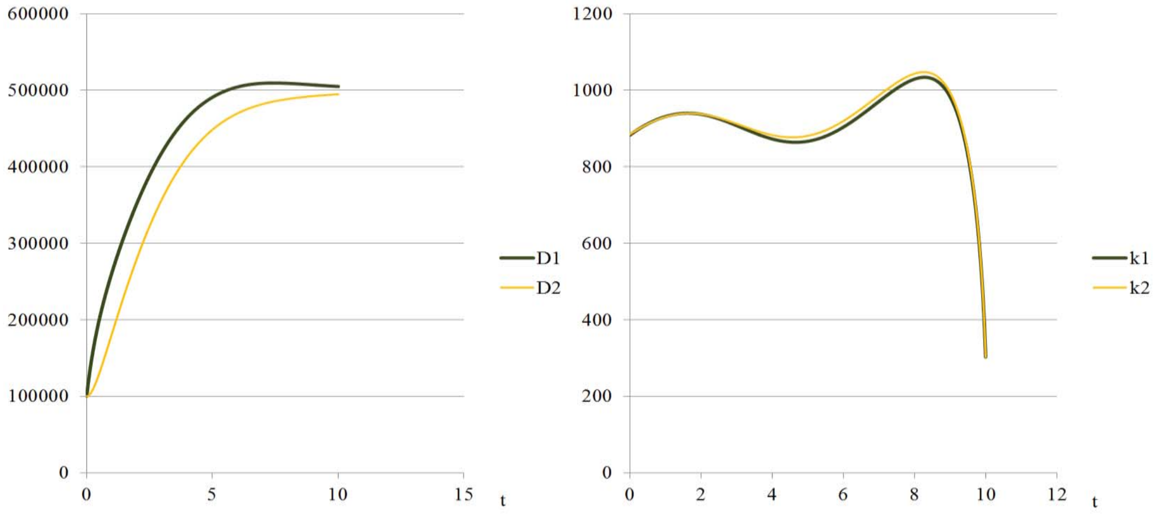

Figure 7: Firm 1 with More Initial Capacity

Figure 7 shows the optimal capacity provision rates and market sizes of the two firms when firm 1 has a higher initial capacity compared to firm 2. In this scenario, firm 2 will have a slightly higher capacity provision rate over time and the market sizes of the two firms will eventually start converging. However the rate of convergence is slower than the case where firm 1 had a higher initial market size.

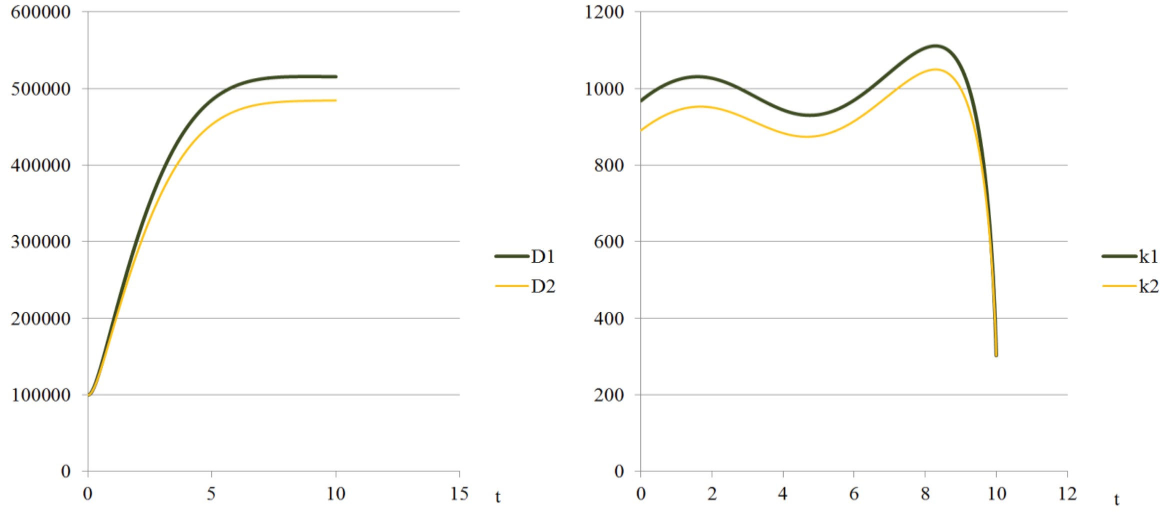

Figure 8: Firm 1 More efficient

In Figure 8, We have the results for the case in which firm 1 is more efficient, meaning that it has a higher marginal profit compared to firm 2. In this case, firm 1 will have higher rates of capacity provision and its market size gets larger. However after some time, the rates of capacity provision begin to get closer to each other and the difference in market sizes seems to stabilize. Therefore, when the whole market is saturated, the more efficient firm will end up having a larger market size compared to the less efficient firm, and the two firms will reduce the amount of capacity they purchase.

Finally in Figure 9, we see the situation in which it is more costly for the customers of firm 1 to switch to firm 2's product. This seems to have the largest impact as the difference in the market sizes steadily increases over time. Firm 1 provisions capacity more aggressively in the early periods, but as time progresses, the rates of capacity provision get closer to each other.

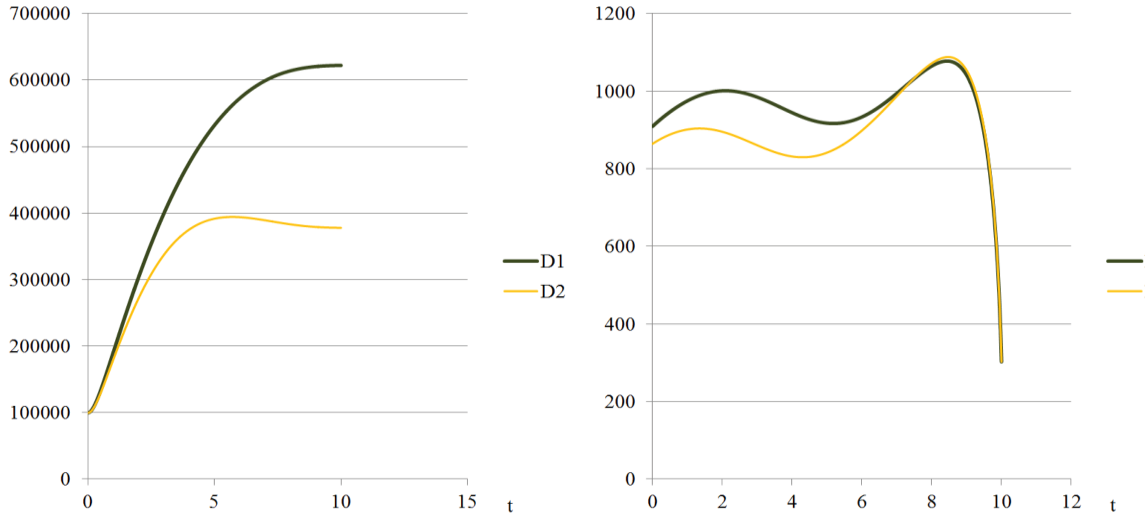

Figure 9: Switching More Costly for Firm 1 Customers

4.5. Conclusion

In this study, we used a two-period game of duopoly competition with price and capacity as decision variables, and two differential game models with rate of capacity provision as the control and market share and capacity as state variables to analyze the competition of forward-looking cloud computing providers. For the two-period game, we formulated the model and presented the equilibrium. We found that if economies of scale are present, the market will be fully segmented, i.e., each firm only serves its own customer type. This result holds even when the cost is a linear function of capacity. This is an interesting finding and implies that as long as it is beneficial for the firms to serve, they will build enough capacity to serve all the customers who prefer the firm’s product. The domination scenario only occurs when there is diseconomies of scale which is not a reasonable assumption given the nature of cloud computing.

For the basic differential game model, we derived the steady-state characteristics of the system and through propositions 2-6, we analyzed the way the system transitions toward this steady state. We also discussed some observations based on the propositions. Most notably, the initial market share structure does not affect the steady-state outcomes and both firms stop their last cycle of provisioning at the same time.

In our comprehensive differential game model, we relaxed many of the restrictive assumptions of the basic model to build a more realistic model which better reflects the salient features of the competition in could computing market. We changed the cost function, assumed decreasing capacity costs over time a la Moore's law (Moore 1965), incorporated depreciation of capacity over time and relaxed the fixed market size assumption. Since the comprehensive model is not computationally tractable, we used numerical analysis to gain insight into the equilibrium result. In different scenarios, we compared the outcomes of the two competing firms. We observed that initial market sizes and initial capacities are not important determinants of the long-term market sizes. However the firm that is more efficient, or manages to make it more costly for its customers to switch to the other firm's product has a clear advantage will have a larger market size in the long run.

References

Agarwal, A., K. Hosanagar, and M. D. Smith. 2011. Location, Location, Location: An Analysis of Profitability of Position in Online Advertising Markets. *Journal of Marketing Research* 48 (6):1057-1073.

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

Amir, R., and J. Y. Jin. 2001. Cournot and Bertrand equilibria compared: substitutability, complementarity and concavity. *International Journal of Industrial Organization* 19 (3–4):303-317.

Androutsellis-Theotokis, S., and D. Spinellis. 2004. A survey of peer-to-peer content distribution technologies. *ACM Comput. Surv.* 36 (4):335-371.

Angrist, J. D., and A. B. Krueger. 2001. Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. *The Journal of Economic Perspectives* 15 (4):69-85.

Ariely, D., and I. Simonson. 2003. Buying, Bidding, Playing, or Competing? Value Assessment and Decision Dynamics in Online Auctions. *Journal of Consumer Psychology* 13 (1/2):113.

Armbrust, M., A. Fox, R. Griffith, A. D. Joseph, R. H. Katz, A. Konwinski, G. Lee, D. A. Patterson, A. Rabkin, I. Stoica, and M. Zaharia. 2009. Above the Clouds: A Berkeley View of Cloud Computing: Electrical Engineering and Computer Sciences

University of California at Berkeley.

Armendáriz, B., and J. Morduch. 2010. *The Economics of Microfinance*. 2 ed: The MIT Press.

Arora, A., H. Xu, R. Padman, and W. Vogt. 2004. Optimal Bidding in Sequential Online Auctions http://heinz.cmu.edu/research/121full.pdf: Carnegie Mellon University.

Atakan, A. E., and M. Ekmekci. 2012. Reputation in Long-Run Relationships. *Review of Economic Studies* 79 (2):451-480.

Ba, S. 2001. Establishing online trust through a community responsibility system. *Decision Support Systems* 31 (3):323-336.

Ba, S., and P. A. Pavlou. 2002a. Evidence of the Effect of Trust Building Technology in Electronic Markets: Price Premiums and Buyer Behavior. *MIS Quarterly* 26 (3):243-268.

Ba, S., and P. A. Pavlou. 2002b. Evidence of the Effect of Trust Building Technology in Electronic Markets: Price Premiums and Buyer Behavior. *MIS Quarterly* 26(3):243-268.

Ba, S., A. B. Whinston, and H. Zhang. 1999. Building trust in the electronic market through an economic incentive mechanism. In *Proceedings of the 20th international conference on Information Systems*. Charlotte, North Carolina, United States: Association for Information Systems, 208-213.

Bapna, R., P. Goes, and A. Gupta. 2003a. Analysis and Design of Business-to-Consumer Online Auctions. *Management Science* 49 (1):85-101.

———. 2003b. Replicating Online Yankee Auctions to Analyze Auctioneers' and Bidders' Strategies. *Information Systems Research* 14 (3):244-268.

Barney, J. B. 1986. Organizational Culture: Can It Be a Source of Sustained Competitive Advantage? *The Academy of Management Review* 11 (3):656-665.

Barua, A., C. H. Kriebel, and T. Mukhopadhyay. 1991. An Economic Analysis of Strategic Information Technology Investments. *MIS Quarterly* 15 (3):313-331.

Berger, C. S., and F. Gleisner. 2009. Emergence of Financial Intermediaries in Electronic Markets: The Case of Online P2P Lending. *BuR - Business Research* 2 (1):39-65.

Bertrand, J. L. F. 1883. Theorie Mathematique de la Richesse Sociale. *Journal des Savant*:499-508.

Bound, J., D. A. Jaeger, and R. M. Baker. 1995. Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogeneous Explanatory Variable is Weak. *Journal of the American Statistical Association* 90 (430):443-450.

Boutaba, R., and A. Marshall. 2006. Management in peer-to-peer systems: Trust, reputation and security. *Computer Networks* 50 (4):469-471.

Brabham, D. C. 2008. Crowdsourcing as a Model for Problem Solving. *Convergence: The International Journal of Research into New Media Technologies* 14 (1):75-90.

Briceño Ortega, A., and F. Bell. 2008. Online social lending: Borrower-generated content. In *14th Americas Conference on Information Systems- AMCIS 2008*. Toronto, Canada.

Bruett, T. 2007. Cows, Kiva, and Prosper.Com: how disintermediation and the internet are changing microfinance. *Community Development Investment Review* 3 (2):44-50.

Brynjolfsson, E., and M. D. Smith. 2000. Frictionless Commerce? A Comparison of Internet and Conventional Retailers. *Management Science* 46 (4):563-585.

Buyya, R., C. S. Yeo, S. Venugopal, J. Broberg, and I. Brandic. 2009. Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility. *Future Generation Computer Systems* 25 (6):599–616.

Chakravarthy, B. S. 1986. Measuring Strategic Performance. *Strategic Management Journal* 7 (5):437-458.

Chen, X., A. Ghate, and A. Tripathi. 2011. Dynamic lot-sizing in sequential online retail auctions. *European Journal of Operational Research* 215 (1):257-267.

Coleman, R. 2007. Is the Future of the Microfinance Movement to be Found on the Internet?: International Trade and Finance Association.

Collier, B., and R. Hampshire. 2010. Sending Mixed Signals: Multilevel Reputation Effects in Peer-to-Peer Lending Markets. Paper read at The ACM Conference on Computer-Supported Cooperative Work, at Savannah, GA.

Cournot, A. A. 1897. Recherches sur les Principes Mathematiqlles de la TIeuorie des Richesses. Paris: 1838 English translation: (N. Bacon, trans.), Researches into the Mathematical Principles of the Theory of Wealth. New York: Macmillan & Company.

Cripps, M. W., G. J. Mailath, and L. Samuelson. 2004. Imperfect Monitoring and Impermanent Reputations. *Econometrica* 72 (2):407-432.

———. 2007. Disappearing private reputations in long-run relationships. *Journal of Economic Theory* 134 (1):287-316.

Damania, R., P. G. Fredriksson, and J. A. List. 2004. The multiplier effect of globalization. *Economics Letters* 83 (3):285-292.

Davis, M. M., and M. J. Maggard. 1990. An analysis of customer satisfaction with waiting times in a two-stage service process. *Journal of Operations Management* 9 (3):324-334.

Dawar, N., and P. Parker. 1994. Marketing Universals: Consumers' Use of Brand Name, Price, Physical Appearance, and Retailer Reputation as Signals of Product Quality. *Journal of Marketing* 58 (2):81-95.

Despotovic, Z., and K. Aberer. 2006. P2P reputation management: Probabilistic estimation vs. social networks. *Computer Networks* 50 (4):485-500.

Diamond, D. W. 1989. Reputation Acquisition in Debt Markets. *Journal of Political Economy* 97 (4):828-862.

DiClemente, D. F., and D. A. Hantula. 2003. Optimal foraging online: Increasing sensitivity to delay. *Psychology and Marketing* 20 (9):785-809.

Dockner, E. J., S. Jorgensen, N. V. V. Long, and G. Sorger. 2001. *Differential Games in Economics and Management Science*: Cambridge University Press.

Doney, P., and J. Cannon. 1997. An examination of the nature of trust in buyer-seller relationships. *Journal of Marketing* 61 (2):35-51.

Duarte, J., S. Siegel, and L. Young. 2012. Trust and Credit: The Role of Appearance in Peer-to-peer Lending. *Review of Financial Studies* 25 (8):2455-2484.

Dunlap, W. P., and E. R. Kemery. 1987. Failure to detect moderating effects: Is multicollinearity the problem? *Psychological Bulletin* 102 (3):418-420.

Dutton, J. E., and J. M. Dukerich. 1991. KEEPING AN EYE ON THE MIRROR: IMAGE AND IDENTITY IN ORGANIZATIONAL ADAPTATION. *Academy of Management Journal* 34 (3):517-554.

Ebbes, P., M. Wedel, and U. Böckenholt. 2009. Frugal IV alternatives to identify the parameter for an endogenous regressor. *Journal of Applied Econometrics* 24 (3):446-468.

Ebbes, P., M. Wedel, U. Böckenholt, and T. Steerneman. 2005. Solving and Testing for Regressor-Error (in)Dependence When no Instrumental Variables are Available: With New Evidence for the Effect of Education on Income. *Quantitative Marketing and Economics* 3 (4):365-392.

Edwards, W., H. Lindman, and L. J. Savage. 1963. Bayesian statistical inference for psychological research. *Psychological Review* 70 (3):193-242.

Fenn, J. 2012. The Gartner Hype Cycle: What's Hot for 2012. In *Gartner*.

Fingas, J. *LG Cloud launches, guns after iCloud on mobile, PC, TV*. http://www.electronista.com/articles/12/04/30/lg.cloud.kicks.off.in.korea/ 2012 [cited.

Follmann, D. A., and M. S. Goldberg. 1988. Distinguishing Heterogeneity from Decreasing Hazard Rates. *Technometrics* 30 (4):389-396.

Fombrun, C. J., and C. B. M. van Riel. 1997. The Reputational Landscape. *Corp Reputation Rev* 1 (1):5-13.

Foster, I., Y. Zhao, I. Raicu, and S. Lu. 2008. Cloud Computing and Grid Computing 360-Degree Compared. Paper read at Grid Computing Environments Workshop, GCE '08.

Fudenberg, D., D. M. Kreps, and E. S. Maskin. 1990. Repeated Games with Long-Run and Short-Run Players. *The Review of Economic Studies* 57 (4):555-573.

Fudenberg, D., and D. K. Levine. 1989. Reputation and Equilibrium Selection in Games with a Patient Player. *Econometrica* 57 (4):759-778.

———. 1992. Maintaining a Reputation when Strategies are Imperfectly Observed. *The Review of Economic Studies* 59 (3):561-579.

Fudenberg, D., and E. Maskin. 1986. The Folk Theorem in Repeated Games with Discounting or with Incomplete Information. *Econometrica* 54 (3):533-554.

Fundenberg, D., and J. Tirole. 1991. *Game Theory*. Cambridge, MA: MIT Press.

Garman, S. R., R. C. Hampshire, and R. Krishnan. 2008. Person-to-Person Lending: The Pursuit of (More) Competitive Credit Markets. In *ICIS*: Association for Information Systems, 160.

Gartner. 2011. Gartner Identifies the Top 10 Strategic Technologies for 2012. In *Gartner Newsroom*.

Gefen, D., E. Karahanna, and D. W. Straub. 2003. Trust and TAM in online shopping: an integrated model. *MIS Q.* 27 (1):51-90.

Ghose, A., P. G. Ipeirotis, and A. Sundararajan. 2005. Reputation premiums in electronic peer-to-peer markets: analyzing textual feedback and network structure. In *Proceedings of the 2005 ACM SIGCOMM workshop on Economics of peer-to-peer systems*. Philadelphia, Pennsylvania, USA: ACM, 150-154.

Goes, P. B., G. G. Karuga, and A. K. Tripathi. 2010. Understanding Willingness-to-Pay Formation of Repeat Bidders in Sequential Online Auctions. *Info. Sys. Research* 21 (4):907-924.

Gotsi, M., and A. M. Wilson. 2001. Corporate Reputation: Seeking a Definition. *Corporate Communications: An International Journal* 6 (1):24-30.

Greene, W. H. 2007. *Econometric Analysis*. 6 ed: Prentice Hall.

Greiner, M. E., and H. Wang. 2010. Building Consumer-to-Consumer Trust in E-Finance Marketplaces: An Empirical Analysis. *International Journal of Electronic Commerce* 15 (2):105-136.

Häckner, J. 2000. A Note on Price and Quantity Competition in Differentiated Oligopolies. *Journal of Economic Theory* 93 (2):233-239.

Hair, J. F., R. L. Tatham, R. E. Anderson, and W. Black. 1998. *Multivariate Data Analysis*. 5 ed: Prentice Hall.

Hamilton, J. 2010. Cloud Computing Economies of Scale. In *Mix 10*.

Harms, R., and M. Yamartino. 2010. The Economics of the Cloud: Microsoft, 1-21.

Helson, H. 1964. Adaptation-level theory; an experimental and systematic approach to behavior. New York: Harper & Row.

Herbig, P., and J. Milewicz. 1993. The Relationship of Reputation and Credibility to Brand Success. *Journal of Consumer Marketing* 10 (3):18 - 24.

herzenstein, M., R. l. Andrews, U. m. Dholakia, and E. lyandres. 2008. The Democratization of Personal Consumer Loans? Determinants of Success in Online Peer-to-Peer Lending Communities. *papers.ssrn.com*.

Herzenstein, M., U. M. Dholakia, and R. Andrews. 2011. Strategic Herding Behavior in Peer-to-Peer Loan Auctions. *Journal of Interactive Marketing* 25 (1):27-36.

Hoff, P. D. 2009. A First Course in Bayesian Statistical Methods: Springer.

Horsky, D. 1977. An Empirical Analysis of the Optimal Advertising Policy. *Management Science* 23 (10):1037-1049.

Hosmer, D. W., and S. Lemeshow. 2000. *Applied logistic regression.* 2 ed. New York: Wiley-Interscience Publication.

Hotelling, H. 1929. Stability in Competition. *The Economic Journal* 39 (153):41-57.

Houser, D., and J. Wooders. 2006. Reputation in Auctions: Theory, and Evidence from eBay. *Journal of Economics & Management Strategy* 15 (2):353-369.

Hui, Michael K., Mrugank V. Thakor, and R. Gill. 1998. The Effect of Delay Type and Service Stage on Consumers' Reactions to Waiting. *Journal of Consumer Research* 24 (4):469-480.

IDG-Enterprise. *Research Indicates that Cloud Increases Short Term Costs for Long Term Gains* 2012 [cited 04/29/2012. Available from http://www.idgenterprise.com/press/research-indicates-that-cloud-increases-short-term-costs-for-long-term-gains.

Iyengar, R., C. Van den Bulte, and T. W. Valente. 2011. Opinion Leadership and Social Contagion in New Product Diffusion. *Marketing Science* 30 (2):195-212.

Jeong, G., E. Lee, and B. Lee. 2012. Does borrowers' information renewal change lenders' decision in P2P lending?: an empirical investigation. In *Proceedings of the 14th Annual International Conference on Electronic Commerce*. Singapore, Singapore: ACM, 83-86.

Kahneman, D., and A. Tversky. 1979. Prospect Theory: An Analysis of Decision under Risk. *Econometrica* 47 (2):263-291.

Kalyanaram, G., and R. S. Winer. 1995. Empirical Generalizations from Reference Price Research. *Marketing Science* 14 (3):G161-G169.

Kamins, Michael A., X. Drèze, and Valerie S. Folkes. 2004. Effects of Seller-Supplied Prices on Buyers’ Product Evaluations: Reference Prices in an Internet Auction Context. *Journal of Consumer Research* 30 (4):622-628.

Kauffman, R. J., and C. A. Wood. 2006. Doing their bidding: An empirical examination of factors that affect a buyer's utility in Internet auctions. *Information Technology and Management* 7 (3):171-190.

Keisidou, E., L. Sarigiannidis, and D. Maditinos. 2011. Consumer characteristics and their effect on accepting online shopping, in the context of different product types. *International Journal of Business Science and Applied Management* 6 (2):31-51.

Kim, J. J. 2008. Where Either a Borrower Or a Lender Can Be. In *The Wall Street Journal*.

Kimball, G. E. 1957. Some Industrial Applications of Military Operations Research Methods. *Operations Research* 5 (2):201-204.

Kittur, A., E. Chi, B. A. Pendleton, B. Suh, and T. Mytkowicz. 2007. Power of the Few vs. Wisdom of the Crowd: Wikipedia and the Rise of the Bourgeoisie. *World Wide Web* 1 (2).

Knorr, E., and G. Gruman. 2008. What cloud computing really means. In *InfoWorld*.

Kreps, D. M., and R. Wilson. 1982a. Reputation and imperfect information. *Journal of Economic Theory* 27 (2):253-279.

———. 1982b. Sequential Equilibria. *Econometrica* 50 (4):863-894.

Kupp, M., and J. Anderson. 2007. Zopa: Web 2.0 meets retail banking. *Business Strategy Review* 18 (3):11-17.

Lacity, M. C., and L. P. Willcocks. 1998. An Empirical Investigation of Information Technology Sourcing Practices: Lessons from Experience. *MIS Quarterly* 22 (3):363-408.

Larson, R. C. 1987. Perspectives on Queues: Social Justice and the Psychology of Queueing. *Operations Research* 35 (6):895-905.

Lawless, J. F. 2002. *Statistical Models and Methods for Lifetime Data*. 2 ed: Wiley-Interscience.

Leclerc, F., B. H. Schmitt, and L. Dubé. 1995. Waiting Time and Decision Making: Is Time like Money? *Journal of Consumer Research* 22 (1):110-119.

Leung, K.-M., R. M. Elashoff, and A. A. Afifi. 1997. Censoring issues in survival analysis. *Annual Review of Public Health* 18:83-104.

Li, D., G. J. Browne, and J. C. Wetherbe. 2007. Online Consumers' Switching Behavior: A Buyer-Seller Relationship Perspective: IGI Global, 30-42.

Lin, M., N. R. Prabhala, and S. Viswanathan. 2009. Can Social Networks Help Mitigate Information Asymmetry in Online Markets? Paper read at ICIS 2009 Proceedings.

———. 2011. Judging Borrowers by the Company They Keep: Friendship Networks and Information Asymmetry in Online Peer-to-Peer Lending. In *Western Finance Association 2009 Annual Meeting Paper*. http://dx.doi.org/10.2139/ssrn.1355679.

Loh, L., and N. Venkatraman. 1992a. Determinants of Information Technology Outsourcing: A Cross-Sectional Analysis. *Journal of Management Information Systems* 9 (1):7-24.

———. 1992b. Diffusion of Information Technology Outsourcing: Influence Sources and the Kodak Effect. *Information Systems Research* 3 (4):334-358.

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

Lucking-Reiley, D., D. Bryan, N. Prasad, and D. Reeves. 2007. Pennies From EBAY: The Determinents of Price in Online Auctions. *The Journal of Industrial Economics* 55 (2):223-233.

Marconi, J. 2002. Reputation Marketing : Building and Sustaining Your Organization's Greatest Asset: McGraw-Hill.

Marquardt, D. W. 1970. Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation. *Technometrics* 12 (3):591-612.

———. 1980. A Critique of Some Ridge Regression Methods: Comment. *Journal of the American Statistical Association* 75 (369):87-91.

Marti, S., and H. Garcia-Molina. 2006. Taxonomy of trust: Categorizing P2P reputation systems. *Computer Networks* 50 (4):472-484.

Mayer, R., J. Davis, and D. Schoorman. 1995. An Integrative Model of Organizational Trust. *The Academy of Management Review* 20 (3):709-734.

McKnight, D. H., V. Choudhury, and C. Kacmar. 2002. Developing and Validating Trust Measures for e-Commerce: An Integrative Typology. *Information Systems Research* 13 (3):334-359.

Mell, P., and T. Grance. 2011. The NIST Definition of Cloud Computing. In *Recommendations of the National Institute of Standards and Technology*. National Institute of Standards and Technology.

Melnik, M. I., and J. Alm. 2002. Does a Seller's eCommerce Reputation Matter? Evidence from eBay Auctions. *The Journal of Industrial Economics* 50 (3):337-349.

Menard, S. 2001. *Applied Logistic Regression Analysis*. 2 ed: Sage Publications, Inc.

Meyer, T. 1994. Subjective Importance of Goal and Reactions to Waiting in Line. *Journal of Social Psychology* 134 (6):819-827.

Milgrom, P., and J. Roberts. 1982. Predation, reputation, and entry deterrence. *Journal of Economic Theory* 27 (2):280-312.

Miller, R. June 2011. A Look Inside Amazon's Data Centers. In *Data Center Knowledge*. http://www.datacenterknowledge.com/archives/2011/06/09/a-look-inside-amazons-data-centers/.

Monroe, K. B. 1971. Measuring Price Thresholds by Psychophysics and Latitudes of Acceptance. *Journal of Marketing Research* 8 (4):460-464.

Moore, G. E. 1965. Cramming more components onto integrated circuits. *Electronics* 38 (8):114-117.

Moorthy, K. S. 1988. Product and Price Competition in a Duopoly. *Marketing Science* 7 (2):141-168.

Morduch, J. 1999. The Microfinance Promise. *Journal of Economic Literature* 37 (4):1569-1614.

Neumann, P. G. 1997. Identity-related misuse. *Commun. ACM* 40 (7):112.

Parsons, S., J. A. Rodriguez-Aguilar, and M. Klein. 2011. Auctions and bidding: A guide for computer scientists. *ACM Comput. Surv.* 43 (2):1-59.

Pfeifer, P. E. 2005. The optimal ratio of acquisition and retention costs. *Journal of Targeting, Measurement and Analysis for Marketing* 13 (2):179-188.

Pinder, J. E., J. G. Wiener, and M. H. Smith. 1978. The Weibull Distribution: A New Method of Summarizing Survivorship Data. *Ecology* 59 (1):175-179.

Pinheiro, J. C., and D. M. Bates. 1995. Approximations to the Log-Likelihood Function in the Nonlinear Mixed-Effects Model. *Journal of Computational and Graphical Statistics* 4 (1):12-35.

Pinker, E. J., A. Seidmann, and Y. Vakrat. 2003. Managing Online Auctions: Current Business and Research Issues. *Management Science* 49 (11):1457-1484.

Pope, D. G., and J. R. Sydnor. 2011. What's in a Picture?: Evidence of Discrimination from Prosper.com. *Journal of Human Resources* 46 (1):53-92.

Prosper.com. 2010. Prosper Member Base Tops One Million. In *Press Releases*. http://www.prosper.com/about/media_press_releases.aspx?t=Prosper_Member_Base_Tops_One_Million.

Rabiner, L. R. 1989. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* 77 (2):257-286.

Rajala, A. K., and D. A. Hantula. 2000. Towards a behavioral ecology of consumption: delay-reduction effects on foraging in a simulated Internet mall. *Managerial and Decision Economics* 21 (3-4):145-158.

Rajamma, R. K., A. K. Paswan, and M. M. Hossain. 2009. Why do shoppers abandon shopping cart? Perceived waiting time, risk, and transaction inconvenience. *Journal of Product & Brand Management* 18 (3):188 - 197.

Ravina, E. 2008. Love & Loans: The Effect of Beauty and Personal Characteristics in Credit Markets. In *Working paper, Columbia University*. http://dx.doi.org/10.2139/ssrn.1107307.

Reibstein, D. 2002. What attracts customers to online stores, and what keeps them coming back? *Journal of the Academy of Marketing Science* 30 (4):465-473.

Renton, P. 2012. Peer To Peer Lending Crosses $1 Billion In Loans Issued. In *TechCrunch*. http://techcrunch.com/2012/05/29/peer-to-peer-lending-crosses-1-billion-in-loans-issued/.

Resnick, P., K. Kuwabara, R. Zeckhauser, and E. Friedman. 2000. Reputation systems. *Commun. ACM* 43 (12):45-48.

Resnick, P., and R. Zeckhauser. 2002. Trust among strangers in internet transactions: Empirical analysis of eBay' s reputation system. *The Economics of the Internet and E-Commerce, Advances in Applied Microeconomics* 11:127-157.

Ried, S., and H. Kisker. 2011. Sizing the Cloud: Understanding And Quantifying The Future Of Cloud Computing. In *Forrester*.

Rindova, V. P., and C. J. Fombrun. 1999. Constructing competitive advantage: the role of firm–constituent interactions. *Strategic Management Journal* 20 (8):691-710.

Roberts, P. W., and G. R. Dowling. 2002. Corporate reputation and sustained superior financial performance. *Strategic Management Journal* 23 (12):1077-1093.

Rosenkranz, S., and P. W. Schmitz. 2007. Reserve Prices in Auctions as Reference Points*. *The Economic Journal* 117 (520):637-653.

Rossi, P. E., G. M. Allenby, and R. McCulloch. 2005. *Bayesian Statistics and Marketing*. 1 ed: Wiley.

Rubinstein, A. 1979. Equilibrium in supergames with the overtaking criterion. *Journal of Economic Theory* 21 (1):1-9.

Rutz, O. J., R. E. Bucklin, and G. P. Sonnier. 2012. A Latent Instrumental Variables Approach to Modeling Keyword Conversion in Paid Search Advertising. *Journal of Marketing Research* 49 (3):306-319.

Rutz, O. J., and M. Trusov. 2011. Zooming In on Paid Search Ads—A Consumer-Level Model Calibrated on Aggregated Data. *Marketing Science* 30 (5):789-800.

Sethi, S. P., and G. L. Thompson. 2005. Optimal Control Theory: Applications to Management Science and Economics. 2nd ed: Springer.

Sganesan, H. 1994. Determinants of long-term orientation in buyer-seller relationships. *The Journal of Marketing* 58 (2):1-19.

Shapiro, S. 1987. The Social Control of Impersonal Trust. *American Journal of Sociology* 93 (3):623.

Shell, K. 1968. Applications of Pontryagin's Maximum Principle of Economics.

Sherif, C. W. 1963. Social categorization as a function of latitude of acceptance and series range. *The Journal of Abnormal and Social Psychology* 67 (2):148-156.

Singer, J. D., and J. B. Willett. 2003. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. 1 ed. USA: Oxford University Press.

Singh, N., and X. Vives. 1984. Price and Quantity Competition in a Differentiated Duopoly. *The RAND Journal of Economics* 15 (4):546-554.

Singh, P. V., Y. Tan, and N. Youn. 2011. A Hidden Markov Model of Developer Learning Dynamics in Open Source Software Projects. *Information Systems Research* 22 (4):790-807.

Snedecor, G. W., and W. G. Cochran. 1989. *Statistical Methods*. 8 ed: Iowa State University Press.

Spence, A. M. 1979. Investment Strategy and Growth in a New Market. *The Bell Journal of Economics* 10:1-19.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Van Der Linde. 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64 (4):583-639.

Srinivasan, S. S., R. Anderson, and K. Ponnavolu. 2002. Customer loyalty in e-commerce: an exploration of its antecedents and consequences. *Journal of Retailing* 78 (1):41-50.

Staiger, D., and J. H. Stock. 1997. Instrumental Variables Regression with Weak Instruments. *Econometrica* 65 (3):557-586.

Steinisch, M. 2012. Peer-to-Peer Lending Survey. In *Consumer Action*. http://www.consumer-action.org/news/articles/2012_p2p_lending_survey/.

Stine, R. A. 1995. Graphical Interpretation of Variance Inflation Factors. *The American Statistician* 49 (1):53-56.

Stock, J. H., J. H. Wright, and M. Yogo. 2002. A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments. *Journal of Business & Economic Statistics* 20 (4):518-529.

Surowiecki, J. 2005. *The Wisdom of Crowds*: Anchor.

Thaler, R. 1985. Mental Accounting and Consumer Choice. *Marketing Science* 4 (3):199-214.

Tripathi, A. K., S. K. Nair, and G. G. Karuga. 2009. Optimal Lot Sizing Policies For Sequential Online Auctions. *Knowledge and Data Engineering, IEEE Transactions on* 21 (4):554-567.

Urbany, J. E., W. O. Bearden, and D. C. Weilbaker. 1988. The Effect of Plausible and Exaggerated Reference Prices on Consumer Perceptions and Price Search. *Journal of Consumer Research* 15 (1):95-110.

Vakrat, Y., and A. Seidmann. 1999a. Analysis and design models for online auctions. Paper read at Proceedings of INFORMS Conference on Information Systems and Technology (CIST), at Cincinnati, OH.

———. 1999b. Can online auctions beat online catalogs? In *Proceedings of the 20th international conference on Information Systems*. Charlotte, North Carolina, United States: Association for Information Systems, 132-143.

Vance, A., C. Elie-Dit-Cosaque, and D. W. Straub. 2008. Examining Trust in Information Technology Artifacts: The Effects of System Quality and Culture. *Journal of Management Information Systems* 24 (4):73-100.

Vanderschraaf, P. 2007. Covenants and Reputations. *Synthese* 157 (2):167-195.

Vaupel, J. W., and A. I. Yashin. 1985. Heterogeneity's Ruses: Some Surprising Effects of Selection on Population Dynamics. *The American Statistician* 39 (3):176-185.

Vincent Lyk-Jensen, S., and O. Chanel. 2007. Retailers and consumers in sequential auctions of collectibles

Détaillants et consommateurs dans les enchères séquentielles d'objets de collection. *Canadian Journal of Economics/Revue canadienne d'économique* 40 (1):278-295.

Wang, P., and T. Splegel. 1994. Database marketing and its measurements of success. Designing a managerial instrument to calculate the value of a repeat customer base. *Journal of Direct Marketing* 8 (2):73-81.

Webster's Revised Unabridged Dictionary. 1913.

Weigelt, K., and C. Camerer. 1988. Reputation and corporate strategy: A review of recent theory and applications. *Strategic Management Journal* 9 (5):443-454.

Weinberg, B. D. 2000. Don't keep your Internet customers waiting too long at the (virtual) front door. *Journal of Interactive Marketing* 14 (1):30-39.

Winer, R. S. 1986. A Reference Price Model of Brand Choice for Frequently Purchased Products. *Journal of Consumer Research* 13 (2):250-256.

Wooldridge, J. M. 2010. Econometric Analysis of Cross Section and Panel Data. 2 ed: The MIT Press.

Xiong, L., and L. Liu. 2003. A reputation-based trust model for peer-to-peer ecommerce communities [Extended Abstract]. In *Proceedings of the 4th ACM conference on Electronic commerce*. San Diego, CA, USA: ACM, 228-229.

Youseff, L., M. Butrico, and D. D. Silva. 2008. Toward a Unified Ontology of Cloud Computing. *Grid Computing Environments Workshop, GCE '08*:1 - 10.

Zellner, A. 1986. On assessing prior distributions and Bayesian regression analysis with g-prior distributions. In Bayesian Inference and Decision Techniques: Essays in Honor of Bruno De Finetti (Studies in Bayesian Econometrics and Statistics), edited by P. K. Goel and A. Zellner: North-Holland, 233–243.

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

Zhang, J., M. Wedel, and R. Pieters. 2009. Sales Effects of Attention to Feature Advertisements: A Bayesian Mediation Analysis. *Journal of Marketing Research* 46 (5):669-681.

Zucchini, W., and I. L. MacDonald. 2009. *Hidden Markov Models for Time Series: An Introduction Using R*. 1 ed: Chapman and Hall/CRC.

Zucker, L. G. 1986. Production of trust: Institutional sources of economic structure, 1840–1920. *Research in Organizational Behavior* 8:53-111.