

# **The Invisible Value of Information Systems: Reputation Building in an Online P2P Lending System**

**Hossein Ghasemkhani**

Information Systems and Operations Management  
Michael G Foster School of Business  
University of Washington  
hossein@uw.edu

**Yong Tan**

Information Systems and Operations Management  
Michael G Foster School of Business  
University of Washington  
ytan@uw.edu

**Arvind K. Tripathi**

Information Systems and Operations Management  
University of Auckland Business School  
Auckland, New Zealand.  
a.tripathi@auckland.ac.nz

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

## Abstract

*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. We use data from a peer-to-peer lending community to empirically investigate the role of information systems in alleviating adverse selection through information availability in a market where information asymmetry poses the greatest threat. We focus on borrowers and posit that due to the repeated nature of transactions in the online P2P system, borrowers develop a form of identity, the borrower's reputation. A borrower's reputation is the collective perception of the lenders about the borrower and a signal of the trustworthiness of the borrower. We set up a simultaneous equation model of outcomes of interest and a latent class model of reputation that affects the outcomes and is affected by them. Reputation is modeled as a discrete variable with three states, and we use the novel approach of latent instrumental variable modeling to deal with endogeneity. We estimate the model using Bayesian MCMC, compare alternative models, and discuss the impact of the state of reputation on the outcomes through the different reputation-specific coefficients of the models. We show that accounting for reputation improves the explanatory power of the model, and therefore provide a way to empirically model the evolution and impact of reputation in online platforms where repeated transactions are performed.*

**(Peer-to-Peer Lending, Reputation, Latent Instrumental Variable Modeling,  
Latent Class Model, Hidden Markov Model, Bayesian Analysis)**

## 1. Introduction

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 support 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 than at 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 2002). 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 they 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 through and kept by the online systems could be used as signals to alleviate information asymmetry, and result in collective decisions by lenders that are 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 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 that 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 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 latent class model performs best among the three.

In the remainder of the paper, we first study the existing body of related literature in section 2. Then in section 3, we discuss the conceptual model and elaborate our research question. In

section 4, we discuss our empirical model and the equations and parameters that need to be estimated. In section 5, we discuss our different approaches to model the evolution and impact of reputation, and in section 6, we look at the data, variables and measures, and discuss the estimation approach. We provide the results and discussion in section 7, and finally conclude the study in section 8.

## **2. 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, organization 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 equilibria and subgame-perfect Nash equilibria 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 agent's short-run costs. 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 the 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 2002). 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 2002; 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 and predict the future behavior of other peers. 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 on the outcomes, such as the images posted by the borrowers (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. 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).

### **3. Conceptual Model**

The main goal of this paper is to investigate whether the availability of additional information and collective evaluation of the borrowers by a large number of lenders creates 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 the changes in credit information is readily available. This wealth of information could enable a rather personalized evaluation of the borrowers' trustworthiness and creditworthiness by 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, this is done if several lenders are willing to participate, and in fact 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 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 might enjoy immediate gains by violating when others perform as expected, ruining one's reputation and losing the opportunity of future agreements with the community members outweighs immediate gain 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. 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 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 borrowers. 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, compared to other lenders. 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 decision 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 a borrower, a lender 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, 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 detail on why they need the money, what their financial status is, and 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 wants 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. In order to mitigate risk, most lenders diversify their investment portfolios and lend small amounts to any individual borrower.

If the aggregate amount of bids submitted by the lenders on a given listing reaches the amount requested by the borrower, new lenders that 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 which 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, i.e., 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.

## 4. Empirical Model

### 4.1. 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 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. According to the logit model, the success of a listing depends on its propensity to succeed which has two components. One of the components is an unobservable extreme value distributed random shock. Integrating over this random shock, 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:

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

$x_{ij}$  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.  $\xi_u^i \sim N(0, \sigma_{\xi_u}^2)$  is a borrower-specific random effect which is included to account for potential unobserved borrower-level heterogeneity (Wooldridge 2010). The model coefficients  $\left(\beta_u^r \alpha_u^r \gamma_u^r\right)'$  have reputation state,  $r$ , superscripts. This is because we are allowing the coefficients for borrowers at different levels of reputation to be different.  $\varepsilon_u^{ij} \sim N(0, \sigma_{\varepsilon_u}^2)$  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.

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

## 4.2. 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,  $MR_{ij}$ . This model investigates the size of decrease in the interest rate for a given listing. Borrowers are naturally interested in larger reductions in the interest rate of the loans. We use a linear regression model for this purpose. In order to make sure the normal assumption is viable, we use the monotonic logit transformation,  $y_{ij} = \text{logit}\left[\frac{(MR_{ij} - IR_{ij})}{\max_{i,j}(MR_{ij} - IR_{ij})}\right]$ , 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  $MR_{ij}$ .  $\xi_y^i$  is a mean-zero borrower-specific effect included to account for potential unobserved heterogeneity of borrowers, and  $\varepsilon_y^{ij} \sim N(0, \sigma_u^2)$  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.

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

## 4.3. 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 if the listing is funded. 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.

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

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.  $d_{ij} = 1$  denotes defaulting. Covariates of the model are similar to (1) and (3), and the vector of coefficients,  $(\beta_d^r \ \alpha_d^r \ \gamma_d^r)'$ , have  $r$  superscript implying that they could be different for different borrowers at different times depending on the state of reputation. Similar to the listing conversion and interest rate models, we have included a borrower-level random effect term,  $\xi_d^i$ , to account for potential unobserved heterogeneity.  $\varepsilon_d^{ij}$  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).

#### 4.4. 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.

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

#### 4.5. 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 all 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 paper, 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 by 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:

$$A = Z_A \varphi_A + \varepsilon_A \quad (6)$$

$$MR = Z_{MR} \varphi_{MR} + \varepsilon_{MR} \quad (7)$$

Here  $A$  and  $MR$  are  $N \times 1$  vectors,  $\varphi_A$  and  $\varphi_{MR}$  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,  $\varepsilon_s$  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 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,  $z_s^{ij'}$ , for each observation.  $\varphi_s$  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,  $z_s^{ij'}$ , 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,  $z_s^{i'}$ , is assumed to have a multinomial distribution with parameters ( $n=1$ ,  $\pi = (\pi_1, \dots, \pi_{C_s})$ ) where  $\pi_c$  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,  $\varepsilon_s$ , is assumed to be independent from the latent instrument, and normally distributed with mean zero and variance  $\sigma_s^2$ . 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  $s \in \{A, MR\}$  which enable 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:

$$\begin{pmatrix} \varepsilon_u^{ij} \\ \varepsilon_y^{ij} \\ \varepsilon_d^{ij} \\ \varepsilon_A^{ij} \\ \varepsilon_{MR}^{ij} \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_u^2 & \sigma_{u,y} & \sigma_{u,d} & \sigma_{u,A} & \sigma_{u,MR} \\ \sigma_{u,y} & \sigma_y^2 & \sigma_{y,d} & \sigma_{y,A} & \sigma_{y,MR} \\ \sigma_{u,d} & \sigma_{y,d} & \sigma_d^2 & \sigma_{d,A} & \sigma_{d,MR} \\ \sigma_{u,A} & \sigma_{y,A} & \sigma_{d,A} & \sigma_A^2 & \sigma_{A,MR} \\ \sigma_{u,MR} & \sigma_{y,MR} & \sigma_{d,MR} & \sigma_{A,MR} & \sigma_{MR}^2 \end{pmatrix} \right), \text{ where } i=1, \dots, n_b \text{ and } j=1, \dots, n_{b_i} \quad (7)$$

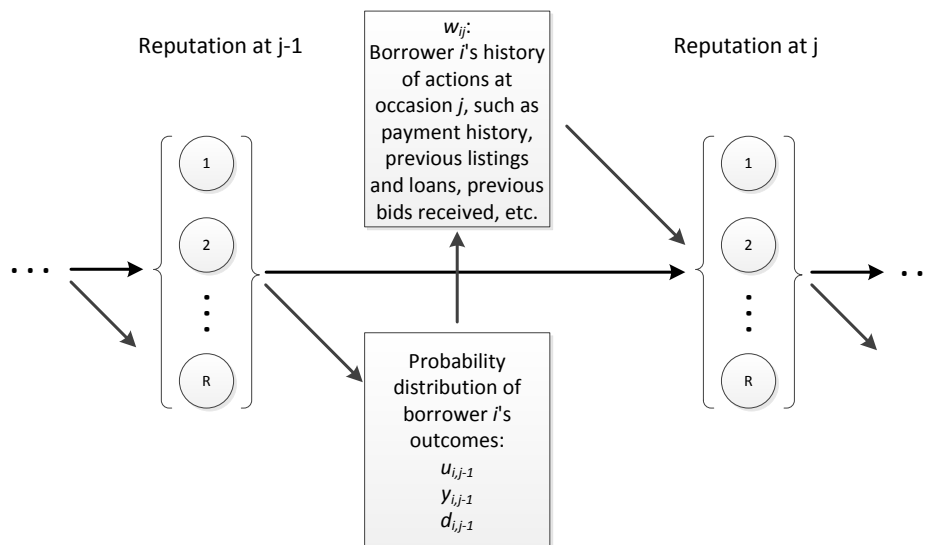
The elements of this covariance matrix will be estimated along with other parameters through our Bayesian MCMC procedure.

## 5. Modeling Reputation

In this section, we present our approach to model the dynamic reputation building of the borrowers and the effects 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 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.

## 5.1. 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 borrowers' 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 1 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. 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:

$$Prob(Reputation_{ij} = r' | Reputation_{i,j-1} = r) = \begin{cases} p_{ij}^{r'} > 0 & r' = r, r-1, r+1 \\ 0 & otherwise \end{cases} \quad (8)$$

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, so we have:

$$Reputation_{ij} | w_{ij}, \{Reputation_{i,j-1} = r\} = \begin{cases} r+1 & \bar{\mu}_r \leq w_{ij}' \lambda_r + \omega_{ij} \\ r & \underline{\mu}_r < w_{ij}' \lambda_r + \omega_{ij} < \bar{\mu}_r \\ r-1 & w_{ij}' \lambda_r + \omega_{ij} \leq \underline{\mu}_r \end{cases} \quad (9)$$

We have  $\omega_{ij} \sim N(0,1)$ ,  $\underline{\mu}_1 = -\infty$  and  $\bar{\mu}_R = \infty$  as the reputation states are limited to  $\{1, \dots, R\}$ .  $w_{ij}$  is an  $n_w \times 1$  vector of reputation covariates and  $\lambda_r$  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 (9) 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:

$$\begin{aligned} p_{ij}^{r,r+1} &= 1 - \Phi(\bar{\mu}_r - w_{ij}' \lambda_r) \\ p_{ij}^{r,r} &= \Phi(\bar{\mu}_r - w_{ij}' \lambda_r) - \Phi(\underline{\mu}_r - w_{ij}' \lambda_r) \\ p_{ij}^{r,r-1} &= \Phi(\underline{\mu}_r - w_{ij}' \lambda_r) \end{aligned} \quad (10)$$

$\Phi$  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:

$$P_{ij} = \begin{bmatrix} p_{ij}^{1,1} & p_{ij}^{1,2} & 0 & 0 & \dots & 0 \\ p_{ij}^{2,1} & p_{ij}^{2,2} & p_{ij}^{2,3} & 0 & \dots & 0 \\ 0 & p_{ij}^{3,2} & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & p_{ij}^{R-2,R-2} & p_{ij}^{R-2,R-1} & 0 \\ \vdots & \vdots & 0 & p_{ij}^{R-1,R-2} & p_{ij}^{R-1,R-1} & p_{ij}^{R-1,R} \\ 0 & 0 & \dots & 0 & p_{ij}^{R,R-1} & p_{ij}^{R,R} \end{bmatrix} \quad (11)$$

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

$$\Pr(u_{ij} = 1) = \frac{\exp\left(x_{ij}' \beta_u^r + A_{ij} \alpha_u^r + MR_{ij} \gamma_u^r + \xi_u^i + \varepsilon_u^{ij}\right)}{1 + \exp\left(x_{ij}' \beta_u^r + A_{ij} \alpha_u^r + MR_{ij} \gamma_u^r + \xi_u^i + \varepsilon_u^{ij}\right)} \quad (12)$$

$$y_{ij} = x_{ij}'\beta_y^r + A_{ij}\alpha_y^r + MR_{ij}\gamma_y^r + \xi_y^i + \varepsilon_y^{ij} \quad (13)$$

$$\Pr(d_{ij} = 1 | u_{ij} = 1) = \frac{\exp\left(x_{ij}'\beta_d^r + A_{ij}\alpha_d^r + MR_{ij}\gamma_d^r + \xi_d^i + \varepsilon_d^{ij}\right)}{1 + \exp\left(x_{ij}'\beta_d^r + A_{ij}\alpha_d^r + MR_{ij}\gamma_d^r + \xi_d^i + \varepsilon_d^{ij}\right)} \quad (14)$$

## 5.2. The Latent Class Model

Our latent class model of reputation is in spirit similar to our 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 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:

$$\text{Prob}\left(\text{Reputation}_{ij} = r' | w_{ij}, \{\text{Reputation}_{i,j-1} = r\}\right) = \text{Prob}\left(\text{Reputation}_{ij} = r' | w_{ij}\right)$$

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, one for each state, there is a single model with  $R-1$  thresholds that governs the dynamic evolution of reputation. Therefore, we have:

$$\text{Reputation}_{ij} | w_{ij} = \begin{cases} R & \mu_{R-1}^{LC} < w_{ij}'\lambda^{LC} + \omega_{ij} \\ r & \mu_{r-1}^{LC} < w_{ij}'\lambda^{LC} + \omega_{ij} \leq \mu_r^{LC} \\ 1 & w_{ij}'\lambda^{LC} + \omega_{ij} \leq \mu_1^{LC} \end{cases} \quad (14)$$

It is worth noting that unlike the HMM, here the vector of coefficients of the underlying stock of reputation,  $\lambda^{LC}$ , 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  $j^{\text{th}}$  occasion as:

$$P_{ij}^r = \Phi\left(\mu_r^{LC} - w_{ij}'\lambda^{LC}\right) - \Phi\left(\mu_{r-1}^{LC} - w_{ij}'\lambda^{LC}\right) \quad (14)$$

We have  $r \in \{1, \dots, R\}$ ,  $\mu_0^{LC} = -\infty$  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 (12), (13) and (14) are unchanged in this model. The 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 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 the goodness of fit of the two.

## 6. Data and Estimation

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 2006, 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 number of members in the community surpassed one million in 2010, and by 2012, more than 350 million dollars in loans 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 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 information on the listing 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 that period. This provides us with around 37 months of data on all the listings, loans created, bids submitted by the lenders.

**Table 1 Description and Summary Statistics of the Listing Variables**

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

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 (10) and (14). 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 2 Description and Summary Statistics of the Reputation Variables**

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

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 most important 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. The details of the estimation steps are provided in the appendix.

## **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 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 10<sup>th</sup> draw to mitigate the autocorrelation issue which is a 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<sup>1</sup> data in our models.

We estimated a model with three simultaneous equations and no explicit reputation component, an HMM and a 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).

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	<i>DIC</i>
M1: The benchmark model with no explicit reputation component	67,460
M2: The hidden Markov model of reputation with 3 states	62,562
<b>M3: The latent class model of reputation with 3 states</b>	<b>61,642</b>

From now on, we focus on the 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 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.

<sup>1</sup> <http://www.djaverages.com/>

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

$R$	$DIC$
2	62,151
<b>3</b>	<b>61,642</b>
4	61,950

LIV approach is generally robust to different specifications 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 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
<b>3</b>	<b>2</b>	<b>61,642</b>
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. Confidence intervals are calculated as the empirical quantiles of the drawn samples and posterior standard deviations of the parameters are also reported. 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 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.

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 to provide some information and create a sense of trust toward the borrower (Duarte et al. 2012).

**Table 6 Loan Conversion Model Estimates ( $\beta_u$ )**

	Reputation State		
	(1)	(2)	(3)
<i>Intercept</i>	-2.711*** (0.310)	-0.084 (0.089)	0.607*** (0.123)
<i>Amount</i>	-0.369*** (0.100)	0.700** (0.164)	0.642*** (0.073)
<i>MR</i>	0.233* (0.159)	0.042*** (0.087)	-0.212 (0.085)
<i>MR<sup>2</sup></i>	-1.100*** (0.400)	0.116 (0.083)	-0.086 (0.040)
<i>FundingOption</i>	0.036*** (0.093)	-1.396 (2.194)	-0.674* (3.283)
<i>CreditGrade</i>	3.390*** (3.054)	-0.510 (2.561)	1.071* (0.454)
<i>ListCharacters</i>	-0.179 (1.237)	-0.973 (2.844)	1.078* (2.813)
<i>ListImage</i>	1.103* (0.723)	1.032 (3.276)	3.020 (2.548)
<i>ListDuration</i>	-2.799* (2.109)	-2.106 (1.964)	-2.636 (0.504)
<i>FriendsBids</i>	0.890*** (0.133)	0.102 (0.137)	0.024 (0.121)
<i>NumQ&amp;A</i>	-1.508*** (0.346)	0.570* (0.105)	0.197*** (0.126)
<i>NumDelin</i>	-0.098** (0.094)	-0.126* (0.119)	0.455 (0.440)
<i>DJIA</i>	0.890* (0.140)	0.001 (0.048)	0.021 (0.125)

Notes: The first numbers are the empirical means of the draws for each coefficient, and the numbers in the parentheses are the empirical standard errors for the estimated parameters.  
\* significant at 0.1; \*\* significant at 0.05; \*\*\* significant at 0.01 (empirical confidence intervals)

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 7 Interest Rate Model Estimates ( $\beta_y$ )**

	Reputation State		
	(1)	(2)	(3)
<i>Intercept</i>	-6.938*** (0.171)	-0.195* (0.046)	0.975*** (0.059)
<i>Amount</i>	-0.370*** (0.051)	0.139 (0.109)	0.025 (0.038)
<i>MR</i>	0.023*** (0.042)	0.060 (0.085)	-0.144 (0.043)
<i>MR<sup>2</sup></i>	0.235*** (0.125)	0.053 (0.025)	-0.025* (0.018)
<i>FundingOption</i>	0.156*** (0.046)	-1.893 (5.791)	-0.810 (2.602)
<i>CreditGrade</i>	0.817*** (2.704)	-0.681 (2.585)	-4.557 (6.916)
<i>ListCharacters</i>	0.851 (0.894)	-0.618 (3.878)	0.386 (3.368)
<i>ListImage</i>	1.327* (0.533)	3.371 (3.666)	0.400 (0.518)
<i>ListDuration</i>	0.168*** (0.210)	0.521*** (1.365)	-1.141 (0.430)
<i>FriendsBids</i>	0.026* (0.105)	-0.041 (0.128)	0.122 (0.106)
<i>NumQ&amp;A</i>	-1.703** (1.176)	-0.110* (0.088)	-0.041 (0.100)
<i>NumDelin</i>	-0.118* (0.101)	0.038 (0.100)	0.162 (0.370)
<i>DJIA</i>	0.004*** (0.008)	-0.014 (0.039)	-0.049* (0.031)

Notes: The first numbers are the empirical means of the draws for each coefficient, and the numbers in the parentheses are the empirical standard errors for the estimated parameters.  
\* significant at 0.1; \*\* significant at 0.05; \*\*\* significant at 0.01 (empirical confidence intervals)

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, higher loan amount on average has a negative impact on the chances of defaulting. As opposed to the interest rate model, 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 8 Defaulting Model Estimates ( $\beta_d$ )**

	Reputation State		
	(1)	(2)	(3)
<i>Intercept</i>	0.106* (0.098)	-0.262* (0.189)	-0.227 (0.318)
<i>Amount</i>	0.060*** (0.019)	0.246* (0.326)	-0.478*** (0.077)
<i>MR</i>	0.203*** (0.138)	-0.606* (0.406)	-0.070 (0.098)
<i>MR<sup>2</sup></i>	-1.304 (1.411)	-0.094 (0.070)	-0.383 (0.158)
<i>FundingOption</i>	-0.064* (0.061)	-0.985** (0.448)	1.884* (1.046)
<i>CreditGrade</i>	-7.684** (5.624)	-6.049*** (4.473)	-24.650 (20.446)
<i>ListCharacters</i>	-1.043*** (0.841)	-3.407 (4.615)	-9.733 (11.224)
<i>ListImage</i>	-1.108* (0.997)	3.143 (6.617)	-8.295 (11.752)
<i>ListDuration</i>	-11.280 (22.640)	-4.578 (5.570)	-0.731 (1.175)
<i>FriendsBids</i>	1.124** (0.478)	-0.711* (0.595)	-0.218 (0.507)
<i>NumQ&amp;A</i>	-2.106** (0.941)	-0.525* (0.318)	-0.528*** (0.466)
<i>NumDelin</i>	0.281*** (0.081)	0.077* (0.027)	-0.111 (1.020)
<i>DJIA</i>	0.896* (0.541)	0.133 (0.160)	0.892 (0.635)

Notes: The first numbers are the empirical means of the draws for each coefficient, and the numbers in the parentheses are the empirical standard errors for the estimated parameters.  
\* significant at 0.1; \*\* significant at 0.05; \*\*\* significant at 0.01 (empirical confidence intervals)

Table 9 shows the posterior means, standard deviations and statistical significance of the ordered probit model of reputation evolution. 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 ( $\lambda^{LC}$ )**

<i>Lender</i>	0.004 (0.055)
<i>GroupMember</i>	0.013 (0.056)
<i>MemberDescription</i>	0.064** (0.029)
<i>CumBids2ListRatio</i>	0.051* (0.034)
<i>Endoresed</i>	0.110** (0.053)
<i>TimeSinceJoin</i>	0.118*** (0.041)
<i>CumLoans</i>	0.138*** (0.045)
<i>CumWithdrawals</i>	0.305*** (0.045)
<i>DeltaCredit</i>	0.113** (0.045)
<i>CumFriendsBids</i>	0.153** (0.076)
<i>CumQ&amp;A</i>	-0.012 (0.010)
<i>CumBorrowed</i>	0.037* (0.033)
<i>Outstanding</i>	0.017 (0.035)
<i>Payments</i>	0.103*** (0.037)
<i>LatePay</i>	-0.56*** (0.030)
$\delta_1$	-0.272*** (0.057)
$\delta_2$	0.088* (0.055)

Notes: The first numbers are the empirical means of the draws for each coefficient, and the numbers in the parentheses are the empirical standard errors for the estimated parameters.  
\* significant at 0.1; \*\* significant at 0.05; \*\*\* significant at 0.01 (empirical confidence intervals)

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

	$\varepsilon_u$	$\varepsilon_y$	$\varepsilon_d$	$\varepsilon_A$	$\varepsilon_{MR}$
$\varepsilon_u$	0.90 (0.03)	-0.08 (0.06)	0.008 (0.003)	0.01 (0.04)	-0.01 (0.002)
$\varepsilon_y$		1.92 (0.54)	-0.002 (0.03)	0.005 (0.006)	-0.03 (0.01)
$\varepsilon_d$			1.13 (0.02)	0.004 (0.002)	0.01 (0.003)
$\varepsilon_A$				0.561 (0.073)	0.05 (0.02)
$\varepsilon_{MR}$					0.31 (0.02)

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 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  $\sigma_{dA}$ ,  $\sigma_{uMR}$  and  $\sigma_{uMR}$  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 11 LIV Parameter Estimates**

	Amount	Maximum Interest Rate
LIV category means ( $\varphi$ )	-0.524 (0.22)	-0.629 (0.020)
	0.087 (0.036)	1.093 (0.027)
	0.410 (0.142)	
LIV category probabilities ( $\pi$ )	0.284 (0.017)	0.615 (0.013)
	0.444 (0.021)	0.385 (0.013)
	0.261 (0.018)	

## 8. Conclusion

Technological advancements in Internet technologies 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/reputation of other parties and the risk 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 hard credit information of the borrowers, 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, the rational lenders use the available information on the borrowers as a signal to infer the borrowers' types.

Therefore through 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 posited 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 as the expected outcomes, where the expectation is taken over the probability distribution of the borrower's type. This expectation is a weighted average 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. The approach is readily extendable to any other environment in which repeated transactions are done 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.

## 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.
- 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.
- Armendáriz, B., and J. Morduch. 2010. *The Economics of Microfinance*. 2 ed: The MIT Press.
- 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. 2002. 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.
- Barney, J. B. 1986. Organizational Culture: Can It Be a Source of Sustained Competitive Advantage? *The Academy of Management Review* 11 (3):656-665.
- 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.
- 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.
- Chakravarthy, B. S. 1986. Measuring Strategic Performance. *Strategic Management Journal* 7 (5):437-458.
- 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.
- 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.
- 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.

- 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.
- 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.
- Fombrun, C. J., and C. B. M. van Riel. 1997. The Reputational Landscape. *Corp Reputation Rev* 1 (1):5-13.
- 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.
- 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.
- 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.
- Gotsi, M., and A. M. Wilson. 2001. Corporate Reputation: Seeking a Definition. *Corporate Communications: An International Journal* 6 (1):24-30.
- 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.
- 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. I. 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.
- Houser, D., and J. Wooders. 2006. Reputation in Auctions: Theory, and Evidence from eBay. *Journal of Economics & Management Strategy* 15 (2):353-369.
- 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.
- Kim, J. J. 2008. Where Either a Borrower Or a Lender Can Be. In *The Wall Street Journal*.

- 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).
- 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.
- 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>.
- Marconi, J. 2002. *Reputation Marketing : Building and Sustaining Your Organization's Greatest Asset*. McGraw-Hill.
- 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.
- 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.
- Milgrom, P., and J. Roberts. 1982. Predation, reputation, and entry deterrence. *Journal of Economic Theory* 27 (2):280-312.
- 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.
- 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](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.
- 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>.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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/](http://www.consumer-action.org/news/articles/2012_p2p_lending_survey/).
- 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.
- 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.
- 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.
- 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.
- 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.



## Appendix

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

1. Draw  $\beta_p^r$ , where  $r \in \{1, \dots, R\}$  and  $p \in \{u, y, d\}$ .  $\beta_p^r$  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  $\beta_y^r$ .  $\mu_{\beta_y}$  and  $\Sigma_{\beta_y}$  are the second-level within-sample mean and covariance of the coefficients.

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

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

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

- 1.2. In order to generate  $\beta_r^u$  and  $\beta_r^d$ , 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.

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

2. Draw the first-level between-sample mean and covariance of the coefficients,  $\mu_{\beta_p}$  and  $\Sigma_{\beta_p}$  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.

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

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

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

3. Generate the normal error terms,  $\varepsilon_q$ , 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,  $\varepsilon_u$  is calculated as:

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

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

$$\varepsilon_A = A - Z_A \phi_A, \quad \varepsilon_{MR} = MR - Z_{MR} \phi_{MR}$$

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

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

$$\text{where } r \in \{1, \dots, R\}, \quad j \in \{1, \dots, n_r^p\}$$

4. Draw the 5x5 model covariance matrix. We do this in two steps. We first draw the covariance matrix of  $\varepsilon_u$ ,  $\varepsilon_y$ ,  $\varepsilon_A$  and  $\varepsilon_{MR}$  for each of which we have  $N$  values. Then we draw the variance of  $\varepsilon_d$  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:

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

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):

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

$Z_s^i = c_s$  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.  $\pi_{c_s}$  is the sampling probability of membership in category  $c_s$  which acts like a prior and  $L(s, Z_s^i(c_s), \varphi_s, E, \Lambda)$  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,  $\pi_A$  and  $\pi_{MR}$ , using a Dirichlet draw as following (Rutz et al. 2012):

$$\pi_s | Z_s \sim \text{Dirichlet}(1 + K_1, \dots, 1 + K_{C_s}), s \in \{A, MR\}$$

$K_c$  denotes the number of observations for the latent IV  $s$  that belong to category  $c$ , i.e.,

$$K_c = \sum_{i=1}^N Z_s^i(Z_s^i = c).$$

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.

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

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

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

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.

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

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.

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

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.

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

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.

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

where  $i = 1, \dots, N$ , and  $\Phi$  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.

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

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

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

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

$$1/\sigma_{\xi_p}^2 | \xi_p \sim \text{gamma}\left(\tilde{\nu}_p/2, \tilde{\nu}_p \tilde{\sigma}_{\xi_p}^2 / 2\right),$$

where  $p \in \{u, y, d\}$  And  $\tilde{\nu}_p = \nu_0 + n_b$  and  $\tilde{\sigma}_{\xi_p}^2 = \left(\tilde{\nu}_p\right)^{-1} \left(\nu_0 \sigma_{0_b}^2 + \sum_{i=1}^{n_b} \left(\xi_p^i\right)^2\right)$ ,

and  $1/\sigma_{\xi_p}^2 \sim \text{gamma}\left(\frac{\nu_{0_b}}{2}, \frac{\nu_{0_b}}{2} \sigma_{0_b}^2\right)$  is the prior with  $\nu_{0_b} = 1$  and  $\sigma_{0_b}^2 = 1$