Predicting and Deterring Default with Social Media Information in Peer-to-Peer Lending

RUYI GE, JUAN FENG, BIN GU, AND PENGZHU ZHANG

RUYI GE (gery@sbs.edu.cn) is an associate professor in the department of electronic commerce at Shanghai Business School, China. She received her Ph.D. in management information systems from Shanghai Jiao Tong University. Her research interests focus on P2P lending and crowdsourcing. She has published in a variety of journals, and in the proceedings of conferences such as the Hawaii International Conference on System Sciences. She has also presented her research at the Workshop on Information Systems and Economics (WISE).

JUAN FENG (juafeng@gapps.cityu.edu.hk) is an associate professor in the Department of Information Systems, College of Business, City University of Hong Kong. She received her Ph.D. in business administration with a dual title in operations research from Pennsylvania State University. Her research focuses on the economics of information systems. She has published in *Management Science, Information Systems Research, Production and Operations Management, Marketing Science*, and other journals.

BIN GU (bin.gu@asu.edu; corresponding author) is a professor and associate dean at the W.P. Carey School of Business at Arizona State University. He received his Ph. D. in operations and information management from the Wharton School at the University of Pennsylvania. His research interests focus on online social media, user-generated content, mobile commerce, and online platforms. His work has appeared in *Management Science, MIS Quarterly, Information Systems Research, Production and Operations Management*, and other journals. He serves as senior editor for *MIS Quarterly* and an associate editor for *Information Systems Research.*

PENGZHU ZHANG (pzzhang@sjtu.edu.cn) is Chair Professor and Head in the Department of Management Information Systems, Antai College of Economics and Management, Shanghai Jiaotong University. He received his Ph.D. from Xi'an Jiaotong University. His research focuses on decision support in innovation and health management. His research has been published in many academic journals including *MIS Quarterly, Decision Support Systems, Information & Management, Computers in Human Behavior, PLos One, Journal of the American Society for Information Science and Technology, Journal of Nanoparticle Research, International Journal of Information Technology & Decision Making, Government Information Quarterly, Journal of Management Analytics*, as well as some important journals in Chinese.

ABSTRACT: This study examines the predictive power of self-disclosed social media information on borrowers' default in peer-to-peer (P2P) lending and identifies social deterrence as a new underlying mechanism that explains the predictive power. Using

a unique data set that combines loan data from a large P2P lending platform with social media presence data from a popular social media site, borrowers' self-disclosure of their social media account and their social media activities are shown to predict borrowers' default probability. Leveraging a social media marketing campaign that increases the credibility of the P2P platform and lenders disclosing loan default information on borrowers' social media accounts as a natural experiment, a difference-in-differences analysis finds a significant decrease in loan default rate and increase in default repayment probability after the event, indicating that borrowers are deterred by potential social stigma. The results suggest that borrowers' social information can be used not only for credit screening but also for default reduction and debt collection.

Online peer-to-peer lending, also known as online P2P lending, refers to the practice of lending money directly to unrelated individuals through online platforms, without the intervention of traditional financial intermediaries, such as banks. These online platforms usually operate at a much lower cost than traditional lending programs and pass the savings on to borrowers in the form of lower interest rates, and to lenders in the form of higher returns. Since the founding of the first online P2P lending platform, Zopa, in the UK in February 2005, thousands of similar practices have emerged worldwide, with more and more lenders and borrowers using online P2P lending platforms. By the end of 2016, the world's largest online P2P lending platform, LendingClub, had funded over US$20 billion in total loans.

Peer-to-peer loans are usually unsecured personal loans to be repaid by individuals. Due to the lack of effective methods to screen individuals' creditworthiness, traditional financial institutions tend to rely excessively on collateral [5, 6, 34, 44]. With online P2P lending platforms, however, borrowers are not required to provide collateral as a protection against default. This practice, on one hand, makes P2P lending particularly attractive to individuals who might otherwise turn to borrow from lenders with high interest rate alternatives, such as day lenders or credit cards [2], and on the other hand, makes it challenging for lenders who have no access to borrowers' creditworthiness before lending and no collateral to collect after a default event.

As the financial risk in online P2P lending platforms is caused mainly by information asymmetry between lenders and borrowers, borrowers are often encouraged to submit as much relevant information as possible. According to Iyer et al. [29], there are two categories of borrower information: standard "hard" information and nonstandard "soft" information. The former directly reflects borrowers' financial status or creditworthiness, for example, credit score, debt-to-income ratio, and annual income, while the latter has no direct relationship with borrowers' financial status or creditworthiness. This is usually submitted by borrowers voluntarily, for example, a picture of the borrower or the identity of his friends on P2P platforms. While soft information is not directly related

to borrowers' creditworthiness, it can act as a useful supplement in the evaluation process, especially for borrowers with scant or unattractive hard information [29].

Previous studies examine the predictive value of various types of soft information on borrowers' creditworthiness, such as borrowers' pictures, textual descriptions of how the loans are to be used, borrowers' facial features [36, 40, 46], and their social network characteristics on online P2P lending platforms [20, 32]. Extending this stream of research, our study examines a new variety of soft information: borrowers' self-disclosed social media information.

In our study, we refer to social media as web-based communities where users create personal profiles; share information, ideas, and personal messages; and communicate with other users. Importantly, social media are open-access venues rather than private, invite-only territories. Users' profiles on social media can be accessed by anyone, and communication in social media is not limited to that between acquaintances.[1] On a P2P lending platform, if a borrower discloses his social media account information, such as a username or a home-page address, the platform or lenders will be able to access the borrower's profile with the account information,[2] to observe his presence (e.g., how many friends he has, how many messages he has posted, etc.) in the social media, and to interact with him or his friends. Our research questions are thus: Are borrowers who disclose their social media account information more creditworthy? Is borrowers' disclosed social media presence and information a significant predictor of their default probability? And if yes, what is the mechanism behind this phenomenon?

Our study aims to answer the above questions by examining a combined data set obtained from both a leading online P2P lending platform and a popular social media site in China. We first collect loan listings and borrower information from the P2P lending platform and model borrowers' default probability as a function of a dummy variable denoting whether a borrower discloses his social media account information, controlling for relevant factors, such as borrowers' demographic characteristics and identity verifications. The result shows that borrowers who disclose their account information have a significantly lower default probability compared to those who do not. Then, using the disclosed account information, we access the borrowers' home pages on social media and collect their presence data (e.g., their number of friends, the number of messages they posted, etc.). We estimate the effect of the social media presence variables on default probability with a set of models and find that the more substantial borrowers' social media presence is (e.g., the more friends they have, the more messages they post), the less likely they are to default.

The estimation models demonstrate predictive relationships between the social media variables and default probability, but do not identify the reason for the relationship. When we develop hypotheses of the relationships, we conjecture that the deterrence of social-stigma costs acts as an underlying cause of the relationships. Default is often associated with social-stigma costs, both nonpecuniary (e.g., disgrace) and pecuniary (e.g., the consequences of a bad reputation) [22]. For a borrower who discloses his social media account information, the P2P lending platform or lenders can, if they want, increase public knowledge of the borrower's default incident by leaving a message on his home page, or they can interact with the borrower's friends on social media, which

may cause considerable social-stigma costs on the borrower. The potential social-stigma costs deter the borrower from default. Moreover, if the deterrence effect exists, the more substantial the borrower's presence on social media, the more social stigma costs he would incur, and therefore, the less likely the borrower would be to default. To identify the conjectured deterrence effect, we leverage a natural experiment involving a marketing campaign launched by the P2P lending platform. In the campaign, the platform provided a pecuniary incentive for borrowers to disclose their social media account information. The campaign demonstrates that the platform highly values borrowers' social media account information, and one reason for this is certainly the potential to use account information as a default enforcement tool. Therefore, the campaign made the aforementioned deterrence more credible to borrowers. We divide borrowers into two groups: the treatment group, which includes borrowers who disclose their social media account information, and the control group, which includes those who do not, and we use a *difference-in-differences* analysis to compare the default probability of the two borrower groups before and after the marketing campaign. We find that the marketing campaign, which sent a strong deterrence signal, has a significantly negative effect on the default probability of borrowers in the treatment group, who are deterred by the signal. This demonstrates that the conjectured deterrence effect does exist.

This study makes several important contributions to research and practice. It is among the first to identify the predictive value of borrowers' self-disclosed social media information on their default probability. We find that adding a simple dummy variable denoting whether a borrower discloses his social media account information can increase the prediction power of the benchmark model (which includes only traditional borrowers' characteristics and verification variables) by 28 percent, which suggests a very effective and low-cost approach for P2P lending platforms to enhance their ability to screen borrowers. In addition, by including social media presence variables (i.e., the number of borrowers' friends on the social media, the number of messages borrowers post on the social media), the model can obtain an extra 5 percent increase in its predictive power. More important, our study not only examines the predictive relationships but also investigates the underlying reason for the relationship. We identify the deterrence mechanism for the predictors to influence borrowers' default probability, showing that borrowers who disclose their social media account information can be deterred from default by potential social stigma costs. This finding not only explains why the social media predictors can be used to predict borrowers' creditworthiness but also extends the usage of borrowers' social medial information from credit screening to default reduction and even debt collection. With this information, P2P lending platforms or lenders can exert deterrence on borrowers to make them reduce their default probability or repay their debts if a default has occurred.

Background on the China P2P Lending Context

The online P2P lending industry has thrived in China since 2007, and it is the largest P2P lending market in the world. By the end of 2016, China had 2,448 online P2P lending platforms, and the accumulated amount of funded loans had reached RMB

3,000 billion, about US$430 billion. China dominates online lending, occupying three-quarters of the global market. Online P2P lending in China owes its growth to substantial market demand and supply for personal lending. Personal lending services, especially for small personal loan amounts, have not fully developed in China. According to the *Economist* [47, p. 55], "this problem [of banks overlooking small borrowers] is especially acute in China. State-owned banks dominate the financial system, with a preference for lending to state-owned companies. The absence of a mature system for assessing consumer credit-risk adds to banks' reluctance to lend to individuals." Therefore, individuals, especially those who do not reside in large cities or earn enough income, have few options to obtain personal loans.

This deficiency in the traditional financial system creates demands for online P2P lending. On the supply side, many Chinese households have considerable savings sitting in their bank accounts with low interest rates. Chinese savers faced two extreme options for managing their money: stash it in bank accounts with very low interest or punt on the stock market with very high risk [47]. Chinese savers long for investment avenues in the middle ground, and online P2P lending becomes one of them. In 2016, more than 13 million people invested on P2P lending platforms, up by 134 percent from 2015, based on data from wdzj.com (http://www.wdzj.com/news/yanjiu), the largest P2P lending community in China.

Although the demand and supply are substantial, the extent of information asymmetry between lenders and borrowers in Chinese markets is more severe than that in developed markets. There are no personal credit rating agencies in China, such as TransUnion, Equifax, or Experian in the United States, which collect and disseminate personal credit history information. As a result, it is difficult to assess the financial creditworthiness of individuals in China. To cope with the shortage of "hard" financial information, many Chinese online P2P lending platforms encourage borrowers to submit soft information as a supplement. Among all types of soft information, social media information is welcomed as a favorable choice because of its widespread use among Chinese borrowers. Chinese people love online social media and use social media applications extensively. According to a China Internet Network Information Center (CNNIC) report [14] published in January 2017, over 90 percent of Chinese online users adopt at least one social media application, 63.3 percent of them use social media at least once a day, and 46.2 percent spend more than 60 minutes per day on social media. The heavy use of social media makes it a rich source of information on Chinese individuals.

Another important aspect of the Chinese P2P lending market is the lack of enforcement mechanisms on borrowers who default on loans. In the United States, various mechanisms exist to punish borrowers if a default occurs, such as reporting it to credit bureaus or filing lawsuits against the borrowers. In China, it is difficult to punish a borrower who defaults. Given the scale of their operations, online P2P lending platforms especially lack the manpower to ensure the repayment of loans. As a result, deterring borrowers with social stigma and shaming on online social media could be a low-cost enforcement option for Chinese P2P lending platforms.

Hypotheses Development

The Predicting Effect of Disclosed Social Media Information on Default Probability

By disclosing social media account information on an online P2P lending platform, borrowers enable lenders or the platform to locate their social media home pages. On their home pages, lenders or the platform can access the borrowers' social network and post public replies, which makes it possible for default behavior to be revealed to their friends. Social psychological studies show that being honest and trustworthy is important for a positive social image in the eyes of others [7, 17]. Moral failure harms a person's social image, thus damaging their social bonds [4, 7, 17] and yielding social punishment in the forms of being marginalized, ostracized, or excluded [9, 17]. The social capital literature finds that as a valuable resource [3, 18], social capital lies not only in the structure and content of social relations but also in trust [30, 31, 37]. In the economics literature, damage to a person's social image or social capital is referred as social stigma; theories on social stigma show that a social-stigma cost is imposed on borrowers who default on loans [15, 16, 48]. Revealing a loan default on a borrower's social media is thus very likely to cause him substantial social-stigma costs. Accordingly, we propose

> *Hypothesis 1 (Borrowers' Social Media Account Information Disclosure Hypothesis): Borrowers who voluntarily disclose their social media account information on the online P2P lending platform are less likely to default.*

For borrowers who disclose their social media account information, social media presence data can be collected from their social media home pages. We focus on two categories of data: the scope of borrowers' social networks and the messages borrowers post on the social media site, both of which illustrate the value of social media to borrowers and thus the potential social stigma costs that may accrue.

The scope of borrowers' social networks refers to how many friends or acquaintances are listed on their social media home page. This measurement is related to how much damage a default reveal could inflict on borrowers' social image and social capital. Specifically, the larger a borrower's social network, the more damage or costs a default will be inflicted on the borrower. The number of messages a borrower posts is closely related to the time and effort he expends on the social media site. These inputs are usually aimed at establishing a positive reputation [45]. Therefore, a default reveal would cause a bigger loss to a borrower who has expended more time and effort on the site than to a borrower who has expended less. In other words, the more a borrower engages in building up his reputation, the more he values the reputation and is subsequently more reluctant to damage it with a default. Therefore, we propose:

> *Hypothesis 2a (Borrowers' Social Media Network—Default Hypothesis): Borrowers who have a larger social network on the social media site are less likely to default on the online P2P lending platform.*

Hypothesis 2b (Borrowers' Message Posting—Default Hypothesis): Borrowers who have posted more messages on the social media site are less likely to default on the online P2P lending platform.

The Deterrence Effect of Disclosed Social Media Information

A borrower who defaults on a loan or files bankruptcy bears social-stigma costs [16, 48], and considering the costs, borrowers tend to be more responsible about their debts [15]. In other words, borrowers are deterred from default by the threat of social-stigma costs. To test whether a threat has deterrence effect, people's behavior should be compared before and after the threat is posed or becomes credible [1]. In the context of our study, we need to find an event that makes the threat of social-stigma cost to borrowers become credible. If a deterrence effect exists, borrowers who disclose their social media account should be less likely to default after the event. Therefore, we propose:

Hypothesis 3 (Deterrence Effect Hypothesis): After the event that increases the likelihood of borrowers' default incident being revealed on their social media, borrowers who disclose their social media information are less likely to default.

Data

Loan Listing and Borrower Data from the Online P2P Lending Platform

The online P2P lending platform, from which we source our loan listing data, is one of the largest online P2P lending platforms in China. By the end of 2016, it had attracted over 30 million members and facilitated about US$4 billion in loans, based on data at Web Financial Data (http://shuju.wdzj.com/). To match borrowers with lenders, the platform requires borrowers to submit both a loan application and a personal profile for initial screening. In the loan application, borrowers list the expected loan amount, the time for repayment, the purpose of the loan, and so on. In the personal profile, they are required to provide demographic information as well as their education background, income status, and any other information that they are willing to provide. As a loan passes the initial screening, the platform posts the loan information on the platform's website for lenders to examine. After reviewing the information, lenders decide whether to invest in a certain loan, and how much to invest. The platform does not provide any guarantee of loan repayments; therefore, lenders bear all the risks.

Our data samples cover all loan listings posted on the platform between January 2011 and August 2013. They consist of 35,457 loan applications and 11,047 borrower profiles. The loan data include loan amount, interest rate, repayment

term, and the status of repayment. The borrower data contain the borrower's demographic characteristics, such as age, gender, education, and marital status, and verification methods, including national identification verification, education certificate verification, phone number verification, and image verification. We also collect one important dummy variable that indicates whether the borrowers disclose their social media account on Weibo.com, one of the largest social media sites in China.

Social Media Presence Data from the Social Media Site

Borrowers' social media presence data are crawled from Weibo.com. In our data sample, some borrowers disclose their account information involving more than one website. Among these websites, however, only Weibo.com strictly meets our definition of social media. Therefore, our study focuses on borrowers who disclose their Weibo.com. account information. Weibo.com is the mirror application of Twitter. Users can post messages and interact with others on their home page at Weibo.com. It is one of the most popular social media sites in China, with nearly 300 million users by September 2016, according to Weibo. In our data sample, over 40 percent of borrowers of the P2P lending platform disclose their Weibo account on the platform. For those borrowers who disclose their Weibo account, we access their home page and collect data about their social network scope and their engagement in Weibo from their Weibo home page.

Operationalization of Variables

Dependent Variable

Default. The dependent variable is a dummy variable, whose value is 1 if the borrower defaulted on a loan (loans) and 0 if the borrower never defaulted on any loans. Following the rule adopted by the online P2P lending platform, we define a default event as occurring when the payment of a loan is more than 120 days late. Very few borrowers default on two or more loans, because borrowers are not allowed to apply for a new loan once they default.

Independent Variables

Borrowers' disclosure of social media account. As a dummy variable, *Weibo_disclosed* equals 1 if the borrower disclosed his Weibo account information to the online P2P lending platform, and 0 otherwise. In our sample, we find that 5,239 (47.42 percent) borrowers disclose their Weibo account.

Borrowers' social media presence. #Followers, #Friends, and *#Fans* are used to represent the scope of borrowers' social network on Weibo.com, while *#Messages* is used to represent borrowers' engagement on Weibo.com.

Followers are those who subscribe to the borrower's home page and follow all the borrowers' updates. The number of followers (*#Followers*) can act as a proxy for the scope of borrowers' social network on the Weibo site. Moreover, followers can be divided into two groups. If a follower is also followed by the borrower, they are probably friends and know each other in real life, or they are interested in each other and want to be friends. We define this group of followers (who follow each other) as borrowers' "friends." If a follower is not followed by the borrower, he is probably a fan rather than a friend of the borrower. We thus define them as "fans." The *number of friends* (*#Friends*) and the *number of fans* (*#Fans*) provide proxies for borrowers' social network scope.

Since posting messages is the major way for borrowers to describe and showcase themselves on the Weibo site, and posting more messages takes more time and effort, we use the *number of messages* (*#Messages*) that borrowers post on their home page to measure their social media engagement.

Table 1 provides full descriptive statistics of the variables measuring borrowers' social media presence.

Control Variables

Borrower's demographic characteristics. This set of control variables consists of borrowers' age, gender, marriage, and education. *Gender* equals 0 if a borrower is male and 1 if female. *Marriage* equals 0 if a borrower is single, and 1 otherwise. The value of the *Education* variable ranges from 1 (under middle school) to 6 (PhD), corresponding to the highest degree a borrower has obtained. The higher the borrower's degree, the larger the value of his education. Tables 2 and 3 show the descriptive statistics of the control variables.

Preverification processes. These are a series of dummy variables. The P2P lending platform encourages borrowers to go through a variety of verification processes before submitting a loan application. Borrowers' national identification number,

Table 1. Descriptive Statistics of Social Media Presence Data

Variable	Mean	Std. dev.	Min	Max
#Followers	826.50	9,2550	0	346,337
#Friends	25.26	102.6	0	2,355
#Fans	803.30	9,245.0	0	346,242
#Messages	213.40	1,146.0	0	45,836

Table 2. Frequency of Gender and Marriage

Variable	0	1
Gender, %	86.56	13.44
Marriage, %	48.59	51.41

Table 3. Descriptive Statistics of Age and Education

Variable	Mean	Std. dev.	Min	Max
Age	30.20	5.769	20	63
Education	3.64	1.165	1	6

Table 4. Proportion of Verified Borrowers, %

Verification	1	0
Identity	74.78	25.22
Education	27.47	72.53
Phone #	79.26	20.74
Image	57.05	42.95

education certificate, phone number, and picture (i.e., online visual verification) can be verified in such processes. We use dummy variables to denote the processes. If the borrower has gone through a specific verification process, the corresponding dummy variable equals 1, and 0 otherwise. Table 4 shows the proportion of borrowers who go through each verification process.

Model Estimation and Results

The Effect of Social Media Account Disclosure on Default Probability

In this section, we first use a logistic regression model to analyze the relationship between borrowers' default on loans and their social media account disclosure. Then, we employ the *propensity score matching* (PSM) technique to address the selection bias caused by unbalanced covariates and use the *instrument variable* (IV) technique to establish the causality of the relationship.

Logistic Regression Model

In Equation 1, we use a logistic model to estimate the influence of borrowers' Weibo account disclosure (*Weibo_disclosed$_i$*) and borrowers' characteristic factors (*Controls$_i$*) on their default probability.

$$logit(Default_i) = \beta_0 + \beta_1 Weibo\_disclosed_i + \delta \ \text{Controls}_i + \varepsilon_i. \tag{1}$$

Table 5 shows that the coefficient of *Weibo_disclosed* is negative and significant at the $p < 0.01$ level. This indicates that borrowers who disclose their Weibo accounts are less likely to default on their loan. Thus, the Borrowers' Social Media Account Information Disclosure Hypothesis (H1) is supported.

Table 5. Logistic Regression Model of Borrower Default

Variable	Parameter	Std. error
Weibo_disclosed	−0.748***	0.055
Education	−0.150***	0.024
Marriage	−0.158***	0.059
Gender	−0.593***	0.086
Age	−0.012**	0.005
Image_verified	0.090	0.056
Education_verified	−0.614***	0.073
Phone_verified	−0.069	0.063
Identity_verified	−0.159***	0.061
Constant	0.134	0.167
N	11,047	
Log likelihood	−5,027.467	

Notes: *$p < 0.1$; **$p < 0.05$; *** $p < 0.01$.

Propensity Score Matching

Although the result of the logistic regression model is supportive, we find that for covariates in the model, the differences in averages by Weibo account disclosure status are significant (see Table 6). This indicates that the data are unbalanced in covariates between the group that disclosed the Weibo account and the group that did not. The unbalanced data have a *selection bias* and weaken the reliability of the regression result [28]. Therefore, we use the *propensity score matching* (PSM) technique to adjust for the differences in covariates [41].

The objective of PSM is to create a statistical equivalence between the treatment and control groups by selecting borrowers from the two groups who resemble each

Table 6. Differences in Averages of Covariates by Weibo Account Disclosure Status

| | Mean | | *t*-test | |
Variable	Account disclosed	Account not disclosed	*t*	*p-value*
Education	3.629	3.651	−0.99	0.323
Marriage	1.464	1.564	−10.86	0.000
Gender	1.126	1.142	−2.47	0.014
Age	29.362	30.93	−15.01	0.000
Image_verified	0.683	0.466	23.61	0.000
Education_verified	0.287	0.263	2.88	0.004
Phone_verified	0.861	0.729	17.29	0.000
Identity_verified	0.849	0.654	24.21	0.000

Table 7. Logistic Regression Model of Weibo Account Disclosure

Variable	Parameter	Std. error
Education	0.037*	0.020
Marriage	−0.235***	0.047
Gender	−0.037	0.059
Age	−0.046***	0.004
Image_verified	0.595***	0.043
Education_verified	−0.170***	0.052
Phone_verified	−0.563***	0.053
Identity_verified	−0.851***	0.052

Notes: *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

other on all covariates. We begin with a logistic model to derive *propensity scores* (PS). The model weighs differences in the covariates between borrowers who disclose their Weibo account and those who do not. With the weights, that is, the coefficients shown in Table 7, we can construct a propensity score for each treated and control case. The propensity score summarizes several characteristics of each subject into a single index, making matching subjects on an *n*-dimensional vector of characteristics feasible.

Moreover, we need to make sure that our treated and control units share the same support before matching. Treated cases off the common support need to be excluded. In our data set, one treated unit is excluded.

A variety of matching methods are available, such as *nearest neighbor matching, radius matching*, and *kernel matching*. The primary selection criterion is to select the method that yields the best balance [23, 26, 43]. After trying all the aforementioned methods, we find that kernel matching is the best matching method for our study. Table 8 shows the reduction in bias on the covariates achieved through kernel matching. It is evident that the matching achieves an appreciable reduction in bias. The absolute bias of every covariate is less than 5 percent, and all the *p*-values are larger than 0.05.

Table 9 shows the *average treatment effect on the treated group* (ATT) obtained before and after matching, where we find significant differences in the default rate between treated and control groups. The *t*-statistics obtained after matching are still significant at the $p = 0.01$ level, and thus the Borrowers' Social Media Account Information Disclosure Hypothesis (H1) is supported.

Finally, a sensitivity analysis is conducted to check for hidden bias. Since matching is based on the conditional independence or unconfoundedness assumption, if there are unobserved variables that simultaneously affect the treatment (Weibo account disclosure) and the dependent variable (borrower's default on loan), a *hidden bias* might arise [42]. For estimating the magnitude of hidden bias with nonexperimental data, we address this problem with the *bounding approach* proposed by Rosenbaum [42]. *The bounds provide evidence*

Table 8. Summary Statistics and Covariate Comparison Before and After Matching

| Variable | Sample | Mean | | | % Reduced | t-test | |
		Treated	Control	%bias	|bias|	t	p > |t|
Education	Unmatched	3.629	3.651	−1.9		−0.99	0.323
	Matched	3.629	3.625	0.3	82.5	0.17	0.865
Marriage	Unmatched	1.461	1.564	−20.7		−10.86	0.000
	Matched	1.461	1.464	−0.4	97.9	−0.22	0.822
Gender	Unmatched	1.126	1.142	−4.7		−2.47	0.014
	Matched	1.126	1.128	−0.6	88.2	−0.29	0.770
Age	Unmatched	29.360	30.993	−28.6		−15.01	0.000
	Matched	29.360	29.531	−3.0	89.4	−1.69	0.091
Image_verified	Unmatched	0.683	0.466	45.0		23.61	0.000
	Matched	0.683	0.068	0.6	98.7	0.32	0.752
Edu_verified	Unmatched	0.287	0.262	5.5		2.88	0.004
	Matched	0.288	0.297	−2.2	60.4	−1.10	0.270
Phone_verified	Unmatched	0.861	0.729	33.0		17.29	0.000
	Matched	0.861	0.861	−0.1	99.8	−0.04	0.968
Identity_verified	Unmatched	0.849	0.654	46.3		24.21	0.000
	Matched	0.849	0.848	0.1	99.7	0.07	0.944

Table 9. Comparisons of Average Treatment Effect on Treated Group (ATT)

Variable	Sample	Treated	Controls	ATT	Std. error	t-statistic
Default	Unmatched	0.131	0.235	−0.104	0.0073	−14.27
	Matched	0.131	0.237	−0.106	0.0081	−13.04

Table 10. Sensitivity Analysis: Rosenbaum Critical p-Values for Treatment Effect

Δ	p-value
1.8	< 0.01
1.9	< 0.10
2.0	> 0.10

on the degree of sensitivity to which any results hinge on the unconfoundedness assumption. Our result is not sensitive to hidden bias until the bias almost doubles the odds of the borrower's default (see Table 10).

Identification Through Instrumental Variable

In both the logistic and PSM models, we find that borrowers' default probability is negatively related to their decisions on whether to disclose their Weibo account. We

interpret the causality of this relationship as follows. When borrowers disclose their Weibo account, they expect extra social stigma costs in case they default on a loan, which deters them from defaulting. Specifically, once borrowers disclose their Weibo account, staff or lenders can contact borrowers' friends and fans listed on their home page; thus, if a default event happens, the message could quickly spread in their social network and cause them social stigma. In a society where social reputation is highly valued, such as Chinese society, concerns about social stigma could deter default.

A few alternative explanations for the observed relationship are possible. A major set of alternative explanations that need to be ruled out is that there may be unobserved differences in the nature of borrowers who disclose their Weibo account that explain their low default rate. For example, borrowers who disclose their Weibo account may be more active and diligent than others. This, in turn, could affect borrowers' default probability, because active and diligent borrowers may have better financial status. To rule out this and similar alternative explanations of our results, we introduce an instrumental variable for Weibo account disclosure.

A suitable instrument for Weibo account disclosure should be exogenously related to borrowers' decision about disclosing accounts but not affect the likelihood of default. We notice that the online P2P lending platform began to provide a function on its registration page to help borrowers disclose Weibo accounts on June 30, 2011. This date can be used to set an instrumental variable (i.e., a dummy variable whose value is 0 if borrowers' registration date is before the date, and 1 otherwise), because borrowers who registered on the platform after this date should be more likely to disclose their Weibo account than borrowers who registered before this date.

To ensure that the instrument also meets the exclusion restriction, we carefully examine the changes made on the registration page after the event and find that no changes were made until October 1, 2012. Being cautious, we only use the data sample of 4,827 borrowers who registered before October 1, 2012, to conduct the IV test. For these borrowers, the instrumental variable can hardly affect the likelihood of default through any direct channel that is independent of Weibo account disclosure. The results of the IV model are reported in Table 11, suggesting that even when using exogenous variation in Weibo account disclosure as an explanatory variable, our result still holds that borrowers' default probability is lower if they choose to disclose their Weibo account (see Table 11).

The results of underidentification test (Kleibergen–Paap statistics = 28.386, $p = 0.000$) and weak identification test (Kleibergen–Paap statistics = 28.682, $p = 0.000$) show that the model is well-identified. Moreover, the result of the Hausman test ($\chi^2 = 2.18$, $p = 0.128$) shows that *Weibo_disclosed* can be treated as exogenous.

Table 11. Results of Original and IV Model

Variables	Original	IV
Weibo_disclosed	−0.103***	−0.340**
	(0.012)	(0.161)
Education	−0.019***	−0.014**
	(0.006)	(0.007)
Marriage	−0.015	−0.023
	(0.013)	(0.015)
Gender	−0.081***	−0.085***
	(0.017)	(0.019)
Age	−0.002**	−0.005**
	(0.001)	(0.002)
Image_verified	0.054***	0.078***
	(0.013)	(0.022)
Education_verified	−0.078***	−0.093***
	(0.015)	(0.018)
Phone_verified	−0.014	0.024
	(0.019)	(0.032)
Identity_verified	−0.022	0.017
	(0.017)	(0.031)
Constant	0.444***	0.555***
	(0.043)	(0.088)
N	4,827	4,827

Notes: Standard errors in parentheses. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

The Effect of Social Media Presence Information on Default Probability

In our data set, 5,239 borrowers disclosed their Weibo account. We further examine these borrowers by integrating their social media presence data crawled from Weibo.com with their P2P lending data. We use a logistic model to estimate the effect of social media presence on the borrowers' default probability.

$$logit(Default_i) = \beta_0 + \beta \text{Social\_Media\_Presence}_i + \delta \text{Controls}_i + \varepsilon_i. \qquad (2)$$

As mentioned earlier, the social media presence information consists of the scope of the social network *(#Followers, #Friends*, and *#Fans)* and the extent of engagement in social media (*#Messages*). Because of the large variance and scale of the social media presence variables, we take the natural log of them in the model. Table 12 shows the results of models with different variables.

Models 1 and 2 analyze the effect of *#Messages* and *#Followers* on borrowers' default probability, respectively. The coefficients of both variables are negative and significant at the $p < 0.01$ level, supporting both the Borrowers' Social Media Network—Default Hypothesis (H2a) and the Borrowers' Message Posting—Default Hypothesis (H2b). They demonstrate that the larger the scope of a borrower's self-disclosed social network,

Table 12. Results of Logistic Models

Variables	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
#Messages	−0.121***		−0.024	−0.043*	−0.042	
	(0.020)		(0.032)	(0.026)	(0.030)	
#Followers		−0.153***	−0.132***			
		(0.022)	(0.035)			
#Friends				−0.174***		−0.153***
				(0.039)		(0.038)
#Fans					−0.117***	−0.079***
					(0.034)	(0.028)
Controls			(Included in estimation)			
Log likelihood	−1,978.2	−1,971.0	−1,970.7	−1,968.0	−1,972.6	−1,965.2

Notes: Standard errors in parentheses. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

the less likely the borrower defaults on a loan, and that the more heavily the borrower engages in the social media site, the less likely the borrower defaults.

The effect of #Messages is no longer significant when both *#Followers* and *#Messages* are included in Model 3, indicating that the two variables may be closely related. As mentioned before, #Followers can be divided into #Friends and #Fans (with #Followers = #Friends + #Fans). Since fans are attracted mainly by the messages posted by borrowers, #Fans probably has a strong relationship with *#Messages*. Friends, in contrast, are more likely to be borrowers' acquaintances in real life; thus, #Friends is likely to have a weaker relationship with *#Messages*. These arguments are confirmed by the results of Models 4 and 5.

We next investigate the effects of the two different types of followers (friends and fans) on borrowers' default probability. For borrowers, both friends and fans are sources of their social capital. Revealing a borrower's default behavior to either the borrower's friends or fans could damage borrowers' social image and raise social stigma costs. Moreover, the more people know, the higher the costs are. Therefore, both *#Friends* and *#Fans* should have effects on borrowers' default probability. Moreover, as previous studies [8, 13] demonstrate that close friends have a stronger behavioral effect on each other than strangers do, we expect that the effect of *#Friends* on borrowers' default probability is more intense than that of *#Fans*. The results of Model 6 show that #Friends and #Fans are both negatively related to default probability with $p < 0.01$, and the coefficient of #Friends is almost twice that of #Fans. The results indicate that although both variables are predictors of default probability, *#Friends* is a stronger predictor than *#Fans.*

Predictive Performance of the Models with Social Media Information

For the variables we propose as predictors of borrowers' default likelihood, we evaluate their prediction power with the area under the *receiver operating characteristic* (ROC) *curve* (AUC). AUC is a standard metric used to assess models that

Table 13. AUCs of Models

| | Weibo_disclosed | | Social Media Presence | | | | | |
Model	Benchmark	Proposed	Benchmark	Model 1	Model 2	Model 4	Model 6	Integrated model
AUC	0.623	0.657	0.625	0.652	0.646	0.653	0.656	0.664

predict classification probabilities [27]. A model that yields a higher AUC generally offers greater predictive power than a model with a lower AUC [19].

Table 13 shows the AUCs of the models with proposed predictors and those of the benchmark models (the models without the proposed variables). The integrated model in the last column includes the variables of both *Weibo_disclosed* and social media presence via *#Friends* and *#Fans*.

First, the AUCs of all the proposed models are greater than 0.5, which suggests that the predictive power of the models is higher than a random guess [19]. Second, the AUC of the benchmark model is 0.6231, and the AUC of the model with *Weibo_disclosed* is higher at 0.6573. Note that a 0.01 improvement in AUC is considered a noteworthy gain in the credit scoring industry [29]. By the AUC metric, the proposed model predicts default with 28 percent greater accuracy than the benchmark model.[3] Finally, the integrated model with both account self-disclosure and social media presence variables has the largest AUC, 0.664, which implies that its predictive power is 33 percent higher than the model without these predictors. All these results illustrate the usefulness of borrowers' social media information in loan default prediction in online P2P lending.[4]

Natural Experiment and the Difference-in-Differences Model

Another way to demonstrate the validity of the deterrence effect is to identify an event whose occurrence makes the deterrence more credible to borrowers without affecting their creditworthiness. By comparing borrowers' default behavior before and after the event, we can establish the deterrence effect. In April 2013, the P2P lending platform launched a marketing campaign to encourage borrowers to disclose their Weibo accounts. It was the first campaign of its kind. The campaign demonstrates that the platform highly values access to borrowers' Weibo accounts. The platform's interest in identifying a borrower's Weibo account is a credible threat, and the threat is further increased by the possibility that the platform may disclose default incidents to the borrowers' social networks, if a default were ever to happen. This potential threat increases the social-stigma cost of default without affecting other possible causes of the predicting effect of borrowers' Weibo account disclosure.

Leveraging this natural experiment, we first analyze the default repayment probability before and after the event. Here, the default repayment probability denotes the percentage of the loans being repaid among all the defaulted loans. If the deterrence

Table 14. The Default Repayment Probability of Treatment vs. Control Groups

Borrower Group	Before, %	After, %	Increment, %
Weibo_disclosed	2.74	4.86	2.12
Weibo_not_disclosed	2.43	1.50	−0.93

effect exists, the default repayment probability of borrowers who disclose their Weibo account should increase significantly compared to those who do not disclose them, because they are deterred by the credible social stigma costs of default. We observe 2,341 loans in our sample that had been defaulted before the campaign and carefully examine the default repayment probability of these loans. Table 14 shows the default repayment probability of two borrower groups before and after the campaign.

It is normal for the default repayment probability to decrease after the campaign (as the rates of the group of *Weibo_not_disclosed* show), since the longer the default duration is, the less likely it will be repaid. The default repayment probability, however, actually increases for the group of *Weibo_disclosed* after the campaign. This, we believe, provides evidence for the deterrence effect being proposed.

Besides the default repayment probability, the deterrence effect could also be observed on loan default probability. After the campaign, the loan default probability of borrowers who disclose their Weibo accounts should decrease significantly compared to those who do not disclose them. We formally examine the interaction between the marketing campaign and Weibo account disclosure on the loan default probability with a *difference-in-differences* (DID) model. The estimated model is:

$$logit(Default_{ijt}) = \beta_0 + \beta_1 Weibo\_disclosed_{ij} + \beta_2 Campaign_{ijt} + \beta_3 Weibo\_disclosed_{ij} \times Campaign_{ijt} + \delta\, Controls_{ij} + \varepsilon_{ijt.}$$

(3)

In Equation (3), i denotes a loan, j denotes a treatment group or control group borrower, and t denotes the time period. The dummy variable $Weibo\_disclosed_{ij}$ equals 1 if the borrower discloses his Weibo account, and otherwise 0. $Campaign_{ijt}$ is a dummy variable whose value is 0 if the loan is listed by the borrower before the campaign, and 1 if the loan is listed by the borrower after the campaign. $Controls_{ij}$ represents both a vector of loan characteristics, such as loan amount, interest rate, and repayment term, and a vector of borrower characteristics, including borrowers' demographic data and which verification processes they have completed. The main coefficient of interest is β_3, which captures the deterrence effect of social media on borrowers' default probability.

To benchmark the results and statistical fit of our DID model, we estimate a series of alternative models (see Table 15). Model 1 is the basic DID model without any

Table 15. The Effect of Deterrence on Loan Default Probability

	Model 1 No controls	Model 2 Control for loan characteristics	Model 3 Control for loan characteristics and borrower clustered effects	Model 4 Control for loan and borrower characteristics and clustered effects
Weibo_disclosed	−0.450***	−0.658***	−0.658***	−0.732***
	(0.049)	(0.050)	(0.069)	(0.070)
Campaign	−0.110**	−0.221***	−0.221***	−0.342***
	(0.048)	(0.049)	(0.060)	(0.060)
Interaction	−0.294***	−0.262***	−0.262***	−0.214**
	(0.085)	(0.086)	(0.099)	(0.099)
Amount		−0.149***	−0.149***	−0.179***
		(0.027)	(0.034)	(0.035)
Interest rate		0.217***	0.217***	0.186***
		(0.010)	(0.014)	(0.015)
Term		0.096***	0.096***	0.099***
		(0.006)	(0.008)	(0.008)
Age				−0.015***
				(0.005)
Gender				−0.401***
				(0.076)
Marriage				−0.089
				(0.055)
Image_verified				−0.006
				(0.043)
Edu_verified				−0.710***
				(0.072)
Phone_verified				−0.273***
				(0.065)
Constant	−2.028***	−5.672***	−5.672***	−3.263***
	(0.026)	(0.360)	(0.495)	(0.546)
N	35,457	35,457	35,457	35,457

Notes: Standard errors in parentheses. *$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

control variables. In Model 2 we include loan characteristics variables as controls. Model 3 adds borrower clustered effects. Model 4 is an augmented version of Model 3 that adds some borrower characteristic variables.

We find that the coefficient of the interaction ($Weibo\_disclosed_{ij} \times Campaign_{ijt}$) is consistently negative and significant across the different models, suggesting that although all the loans are less likely to be defaulted on after the campaign, the effect of the marketing campaign is more significant on borrowers who disclose their Weibo account. In other words, as the deterrence becomes more credible, borrowers

who disclose their social media account are less likely to default. The Deterrence Effect Hypothesis (H3) is supported. We believe this clearly demonstrates that these borrowers are deterred by social-stigma costs.

Conclusion

Key Findings

In this article, we study the relationship between social media information and borrowers' creditworthiness in online P2P lending. Specifically, we examine the predicting effect of the borrowers' self-disclosure of both their social media account and the disclosed social media presence information on borrowers' default probability. We further investigate the deterrence effect of these predictors on borrowers' possible default on loans. We draw on soft information and social stigma literature to formulate our hypotheses and test them using a novel data set that combines borrowers' loan listing information with data collected from a social media site for the borrowers who disclose their social media account. We adopt PSM and IV techniques to rule out self-selection issues concerning borrowers' self-disclosure behavior and their default probability. Moreover, we rely on a natural experiment with a DID analysis to establish the deterrence effect of borrowers' social media information.

Our results show that social media information predicts a borrower's creditworthiness on two levels. First, for all borrowers in P2P lending, the decision on whether to disclose their social media accounts can be used as a predictor of their default probability. Second, for borrowers who choose to disclose their social media accounts, their social media presence, such as their social network scope and the number of messages they post on the social media site, can predict the probability of default. More importantly, we identify the deterrence of social stigma costs as an underlying reason for the predicting effect of the social media information. Although our study adopts a data set collected in China, which has a long, well-established tradition of reputation consciousness, most borrowers in our data are in their thirties. We believe Chinese individuals at this age have an attitude toward social reputation and stigma similar to that of their western counterparts. For this reason, we believe the Chinese context of our study does not necessarily jeopardize the generality of our results.

Research Contribution and Practical Implications

We contribute to the literature that examines the role of soft information in screening borrowers' creditworthiness in P2P lending [29, 36, 40, 46]. We extend this stream of literature by not only examining a new type of soft information (social media information) but also identifying the underlying reason for the predicting effect. As social network information is becoming a means to understand people's certain

behavior or predict market events [11, 35, 38, 39], it is also employed for credit screening in P2P lending. Our study goes beyond borrowers' social networks on online P2P lending platforms [20, 32] to a broad social media presence on a major social media platform that is widely used by borrowers. This work also contributes to the literature on personal credit approvals [10, 24, 49, 50], which uses different statistical methods, such as logistic regression, decision tree, nearest neighbor, and neural networks, to distinguish between "good" and "bad" borrowers. The results of our study suggest that social media-related variables can be added to the classification models as account-specific covariates [12] to improve the performance of the models.

Our work complements the literature that examines the influence of social stigma on loan default or bankruptcy [15, 16, 48]. We reveal that in P2P lending, social-stigma costs can be exerted on borrowers with the aid of social media, and borrowers are deterred from default by the potential social-stigma costs. Our work also complements recent literature that inspects the business value of social media, which so far has been limited primarily to examining social media from a firm–customer viewpoint, such as regarding it as a marketing tool [21, 51], a venue gathering users' ideas [25], or a predictor of firm value [33]. We examine the effect of social media on alleviating information asymmetry in markets and discover its screening and deterrence value in personal credit markets.

For P2P lending platforms, our study discovers a set of effective, low-cost tools that can be used to screen borrowers' creditworthiness. For example, the platforms can give options for borrowers to choose whether to disclose their social media accounts when registering on the platform; for borrowers who choose to disclose their account information, the platforms can crawl borrowers' presence data automatically from the social media site. All these data can be integrated into credit grading models used by the platforms. Moreover, our study suggests that having access to a borrower's social media account is a good means to deter borrowers from default and to raise the probability of a defaulted loan being repaid.

More generally, our study shows the value of social media information in personal credit scoring and enforcement against defaulters, especially in less-regulated markets, such as Chinese markets. In these markets, the establishment of regulations usually lags behind technology development. Therefore, innovative approaches that make good use of technologies and eliminate regulatory loopholes are especially favored.

Limitations and Future Work

Our study has some limitations that point to exciting lines for future research that we strongly encourage. First, social media are very rich information sources, and we only use a very small part of them in our models. As an extension, it would be interesting to investigate other genres of borrowers' social media information, such

as the messages they post, their interactions with others, and so on. This may lead to some new predictors of default probability. Second, our model is time-independent —the social media information was obtained at the time borrowers disclosed their accounts. In fact, the information changes with the activity of borrowers on the social media site. It would be worthwhile to extend our static model to a dynamic one, which could examine how the change of borrowers' social media information predicts the varying hazard rate of borrowers' default.

Funding

NOTES

1. According to our definition, Facebook and Twitter are social media, while WhatsApp is not.
2. Although in social media users' home pages can be accessed by anyone, it is difficult to locate the home page of a certain user without his username or home-page address.
3. We calculated the percentage improvement as $0.657 - 0.5 / 0.623 - 0.5 = 1.28$, where 0.5 is subtracted from both AUCs because 0.5 is the AUC under a noninformative (random) system.
4. We also calculated the AUC values by using 60 percent of randomly-selected samples as training data and the rest 40 percent of samples as test data. The predictive powers of the proposed model and the integrated models are 34 percent and 39 percent higher than the benchmark model, respectively.

REFERENCES

1. Abrams, D.S. Estimating the deterrent effect of incarceration using sentencing enhancements. *American Economic Journal: Applied Economics*, *4*, 4 (October 2012), 32–56.
2. Adams, W.; Einav, L.; and Levin, J. Liquidity constraints and imperfect information in subprime lending. *American Economic Review*, *99*, 1 (March 2009), 49–84.
3. Adler, P.S., and Kwon, S.W. Social capital: Prospects for a new concept. *Academy of Management Review*, *27*, 1 (January 2002), 17–40.
4. Ahmed, E.; Harris, N.; Braithwaite, J.; and Braithwaite, V. *Shame Management through Reintegration*. Cambridge, UK: Cambridge University Press, 2001.
5. Ang, J.S.; Lin, J.W.; and Tyler, F. Evidence on the lack of separation between business and personal risks among small businesses. *Journal of Entrepreneurial Finance*, *4*, 2 (Summer 1995), 197–210.

6. Avery, R.B.; Bostic, R.W.; and Samolyk, K A. The role of personal wealth in small business finance. *Journal of Banking and Finance*, *22*, 6 (August 1998), 1019–1061.

7. Baumeister, R.F., and Leary, M.R. The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, *117*, 3 (May 1995), 497–529.

8. Bond, R.M.; Fariss, C.J.; Jones, J.J. Kramer, A.D.I.; Marlow, C.; Settle, J.E.; and Fowler, J.H. A 61-million-person experiment in social influence and political mobilization. *Nature*, *489*, 7415 (September 2012), 295–298.

9. Braithwaite, J. *Crime, Shame and Reintegration*. Cambridge, UK: Cambridge University Press, 1989.

10. Carter, C., and Catlett, J. Assessing credit card applications using machine learning. *IEEE Expert*, *2*, 3 (September 1987), 71–79.

11. Chen, A.; Lu, Y.; Chau, P.Y.K; and Gupta, S. Classifying, measuring, and predicting users' overall active behavior on social networking sites. *Journal of Management Information Systems*, *31*, 3 (Winter 2014), 213–253.

12. Chehrazi, N., and Weber, T.A. Dynamic valuation of delinquent credit-card accounts. *Management Science*, *61*, 12 (December 2015), 3077–3096.

13. Christakis, N.A., and Fowler, J.H. Social contagion theory: Examining dynamic social networks and human behavior. *Statistics in Medicine*, *32*, 4 (February 2013), 556–577.

14. CNNIC (China Internet Network Information Center). *China Statistical Report on Internet Development*. Beijing, China, January 2017.

15. Cohen-Cole, E., and Duygan-Bump, B. Household bankruptcy decision, the role of social stigma vs. information sharing. Working paper, Federal Reserve Bank of Boston, November 2008.

16. Crocker, J.; Major, B.; and Steele, C. Social stigma. In D.T. Gilbert, S.T. Fiske, and G. Lindzey (eds.), *The Handbook of Social Psychology*. Vols. 1 and 2. New York, NY: McGraw-Hill, 1998, pp. 504–553.

17. de Waal, F.B.M. *Good Natured: The Origins of Right and Wrong in Humans and Other Animals*. Cambridge, MA: Harvard University Press, 1997.

18. Dore, R. Goodwill and the spirit of market capitalism. *British Journal of Sociology*, *34*, 4 (December 1983), 459–482.

19. Fawcett, T. An introduction to ROC analysis. *Pattern Recognition Letters*, *27*, 8 (June 2006), 861–874.

20. Freedman, S.M., and Jin, G.Z. Learning by doing with asymmetric information: Evidence from Prosper.com. No. w16855. National Bureau of Economic Research, Cambridge, MA, March 2011.

21. Goh, K.Y.; Heng, C.S.; and Lin, Z.J. Social media brand community and consumer behavior: Quantifying the relative impact of user- and marketer-generated content. *Information Systems Research*, *24*, 1 (March 2013), 88–107.

22. Gross, D.B., and Souleles, N.S. An empirical analysis of personal bankruptcy and delinquency. *Review of Financial Studies*, *15*, 1 (Spring 2002), 319–347.

23. Harder, V.S.; Stuart, E.A.; and Anthony, J.C. Propensity score techniques and the assessment of measured covariate balance to test causal associations in psychological research. *Psychological Methods*, *15*, 3 (September 2010), 234–249.

24. Henley, W.E., and Hand, D.J. A k-nearest-neighbour classifier for assessing consumer credit risk. *Statistician*, *45*, 1 (April 1996), 77–95.

25. Hildebrand, C.; Haubl, G.; Herrmann, A.; and Landwehr, J.R. When social media can be bad for you: Community feedback stifles consumer creativity and reduces satisfaction with self-designed products. *Information Systems Research*, *24*, 1 (March 2013), 14–29.

26. Ho, D.E.; Imai, K.; King, G.; and Stuart, E.A. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis*, *15*, 3 (Summer 2007), 199–236.

27. Huang, J., and Ling, C.X. Using AUC and accuracy in evaluating learning algorithms. *IEEE Transactions on Knowledge and Data Engineering*, *17*, 3 (January 2015), 299–310.

28. Imbens, G.W., and Wooldridge, J.M. Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, *47*, 1 (March 2009), 5–86.

29. Iyer, R.; Khwaja, A.I.; Luttmer, E.F.; and Shue, K. Screening peers softly: Inferring the quality of small borrowers. *Management Science*, *62*, 6 (June 2015), 1554–1577.

30. Knoke, D. Organizational networks and corporate social capital. In R.T.A. Leenders and S.M. Gabbay (eds.), *Corporate Social Capital and Liability*. Boston, MA: Springer, 1999, pp. 17–42.

31. Leana, C.R., and Van Buren, H.J. Organizational social capital and employment practices. *Academy of Management Review*, *24*, 3 (July 1999), 538–555.

32. Lin, M.; Prabhala, N.R.; and Viswanathan, S. Judging borrowers by the company they keep: Friendship networks and information asymmetry in online peer-to-peer lending. *Management Science*, *59*, 1 (January 2013), 17–35.

33. Luo, X.M., and Zhang, J. How do consumer buzz and traffic in social media marketing predict the value of the firm? *Journal of Management Information Systems*, *30*, 2 (Fall 2013), 213–238.

34. Manove, M.; Padilla, A.J.; and Pagano, M. Collateral versus project screening: A model of lazy banks. *RAND Journal of Economics*, *32*, 4 (Winter 2001), 726–744.

35. Matook, S.; Cummings, J.; and Bala, H. Are you feeling lonely? The impact of relationship characteristics and online social network features on loneliness. *Journal of Management Information Systems*, *31*, 4 (Spring 2015), 278–310.

36. Pope, D.G., and Sydnor, J.R. What's in a picture? Evidence of discrimination from Prosper.com. *Journal of Human Resources*, *46*, 1 (January 2011), 53–92.

37. Putnam, R.D. Bowling alone: America's declining social capital. *Journal of Democracy*, *6*, 1 (January 1995), 65–78.

38. Qiu, L.; Rui, H.; and Whinston, A.B. Effects of social networks on prediction markets: Examination in a controlled experiment. *Journal of Management Information Systems*, *30*, 4 (Spring 2014), 235–268.

39. Qiu, L.; Rui, H.; and Whinston, A.B. The impact of social network structures on prediction market accuracy in the presence of insider information. *Journal of Management Information Systems*, *31*, 1 (Summer 2014), 145–172.

40. Ravina, E. Love and loans: The effect of beauty and personal characteristics in credit markets. SSRN 1101647, February 2013.

41. Rosenbaum, P.R., and Rubin, D.B. The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*, 1 (April 1983), 41–55.

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

43. Rubin, D.B. The design versus the analysis of observational studies for causal effects: Parallels with the design of randomized trials. *Statistics in Medicine*, *26*, 1 (January 2007), 20–36.

44. Stiglitz, J.E., and Weiss, A. Credit rationing in markets with imperfect information. *American Economic Review*, *71*, 3 (June 1981), 393–410.

45. Tang, Q.; Gu, B.; and Whinston, A.B. Content contribution for revenue sharing and reputation in social media: A dynamic structural model. *Journal of Management Information Systems*, *29*, 2 (Fall 2012), 41–76.

46. Theseira, W. Competition to default: Racial discrimination in the market for online peer-to-peer lending. Ph.D. dissertation, University of Pennsylvania, Philadelphia, April 2009.

47. *The Economist*. The age of the appacus: In China, fintech shows the way. February 25, 2017, 55.

48. Thorne, D., and Anderson, L. Managing the stigma of personal bankruptcy. *Sociological Focus*, *39*, 2 (May 2006), 77–97.

49. West, D. Neural network credit scoring models. *Computers and Operations Research*, *27*, 11 (October 2000), 1131–1152.

50. Wiginton, J.C. A note on the comparison of logit and discriminant models of consumer credit behavior. *Journal of Financial and Quantitative Analysis*, *15*, 3 (September 1980), 757–770.

51. Xie, K., and Lee, Y.J. Social media and brand purchase: Quantifying the effects of exposures to earned and owned social media activities in a two-stage decision making model. *Journal of Management Information Systems*, *32*, 2 (Fall 2015), 204–238.