Association for Information Systems
AIS Electronic Library (AISeL)

CONF-IRM 2018 Proceedings

International Conference on Information Resources Management (CONF-IRM)

5-2018

Financial Capital or Social Capital: Evidence From the Survival Analysis of Online P2P Lending Platforms

Meijing Lan
The University of Nottingham Ningbo, meijing.lan@nottingham.edu.cn

Xiuping Hua
The University of Nottingham Ningbo, xiuping.hua@nottingham.edu.cn

Xiaoqua Liu
The University of Nottingham Ningbo, xiaoquan.liu@nottingham.edu.cn

FINANCIAL CAPITAL OR SOCIAL CAPITAL: EVIDENCE FROM THE SURVIVAL ANALYSIS OF ONLINE P2P LENDING PLATFORMS

Meijing Lan
The University of Nottingham
Ningbo, China
Guangdong University of Finance
meijing.lan@nottingham.edu.cn

Xiuping Hua
The University of Nottingham
Ningbo, China
xiuping.hua@nottingham.edu.cn

Xiaoquan Liu
The University of Nottingham
Ningbo, China
xiaoquan.liu@nottingham.edu.cn

Abstract

In this paper, we draw upon the bank survival literature and that in the information management area in identifying the key factors behind the survival of Chinese online P2P lending platforms. In particular, we are interested in determining whether the traditional financial capital or the social capital, associated with the online nature of these innovative lending platforms, plays a more essential role. We implement a flexible proportional odds model with a baseline spline function to analyze survival patterns and also consider potential fractional polynomial transformation and time-dependent effect of variables. Using a hand-collected dataset of 6190 platforms from June 2007 to June 2017, we provide robust evidence that although financial capital variables play an important role in driving platform survival, they are less significant or become insignificance in the presence of social capital variables. These findings contribute to both the literature and the development of this innovative and fast-growing industry of financial inclusion.

Keywords

Social Network; Online Reviews; Proportional Odds Model; Cubic Polynomials.

1. Introduction

During the *Great Default* between June 2007 to June 2017, 61% (or 3766) of 6190 online peer-to-peer (P2P) lending platforms in China has defaulted. By end of June 2017, half a million investors have suffered from defaulted loans with an estimate loss of 4.39 billion USD.[i] The large number of default has triggered public outcries for more financial capital requirement including increased registered capital, compulsory insurance plan, setting up reserve fund, and so forth, suggested by the main findings in the banking literature, especially the strand on bank survival (Calomiris & Mason, 2003), as the P2P lending platform is considered a form of financial intermediary (Datta & Chatterjee, 2008).

Meanwhile, the online nature of P2P platforms indicates such research crosses over to the literature of e-commerce and information system, which is one of the contributions this paper makes. Electronic financial service agents behave as financial intermediaries by assisting and providing sources of trust in online transactions (Datta & Chatterjee, 2008). An important concept related to default in the e-commerce and information system field is *trustworthiness*. The lack of trust is shown to be the greatest barrier to online transactions (Kim et al., 2004). Companies that successfully build trust are able to better connect with customers and other economic players, and achieve better chance for survival (Hoffman et al., 1999; Karimov et al., 2011). So what is *trust*? How can we evaluate it? According to the social capital theory, trust is interconnected with social capital: social capital is not only a generator of trust (Putnam et al., 1994), but also a motivational result of trust (Adler & Kwon, 2002). An organization with higher level of social capital is able to generate higher level of trust, which in turn accumulate more social capital (Lins et al., 2017). The same theory carries over to the online P2P lending industry.

Given the different theories and focuses on survival in the banking literature versus the e-commerce literature, we raise two research questions: First, for the online P2P lending industry, are financial capital and social capital both significant factors for survival? If they are, is one type of capital more essential than the other? Answers to these questions are relevant to investors, regulators, and platform managers, and substantively advance our understanding of this new form of financial innovation. For P2P investors, a healthy and viable P2P lending industry helps mitigate or avoid financial losses. For regulators, the ability to identify major risk factors allows them to put forth more relevant and effective policy and reduce specific or regional financial risk. For online P2P platform managers, our findings will enable them to learn from the history and improve their chance for survival.

As far as we are aware, our study is the first that examines P2P lending platform default in the wider crowdfunding industry. Hence we fill a gap in the crowdfunding literature that so far focuses mainly on the market economic mechanism or behaviour of market participants (e.g., Mollick, 2014; Zhang & Liu, 2012). Methodologically, we extend the economics and e-commerce literature which predominantly adopts the Cox (1972) model (e.g., Wang et al., 2013; Wheelock & Wilson, 2000) by constructing a flexible proportional odds model with a baseline spline function. The model is able to not only identify influential variables to a platform's survival but also predict with accuracy future survival pattern. Fractional polynomial transformation is also considered to uncover potential nonlinearity in variables.

The rest of the paper is organized as follows. Section 2 outlines our econometric model. Section 3 describes the data. Section 4 presents our empirical results. Section 5 concludes.

2. Econometric framework

2.1. Parametric model construction

Originated from biostatistics, survival analysis is applied widely to analyze firm survival. Following Royston and Parmar (2002), we use a flexible parametric survival model with cubic splines, to address the disadvantages of the popular Cox (1972) model, which is criticized for constant hazard ratio and not being able to predict survival time, hazard rates, or absolute risk (Orth, 2013). Furthermore, we choose the flexible proportional odds (PO) model since it generalizes the baseline survival function to avoid poor fit and over fit, respectively, of standard parametric models and the Cox models (Gelfand et al., 2000). Also, it assumes that hazard ratios converge to one if $t \to \infty$ (Bennett, 1983), which is consistent with our dataset. In our application, the log odds that failure events will occur in the interval $(0,\ t)$ can be written as follows:

$$\ln O(t;\mathbf{x}) = \ln O_0(t) + \boldsymbol{\varphi}'\mathbf{x} = \ln O_0(t) + \boldsymbol{\alpha}'\boldsymbol{Fin} + \boldsymbol{\beta}'\boldsymbol{Soc} + \boldsymbol{\theta}'\boldsymbol{Control}, \tag{1}$$

where $\mathbf{x}$ stands for the covariate matrix and ***Fin***, ***Soc***, and ***Control*** are covariate matrices for financial capital, social capital and control variables respectively. Define the baseline distribution function $\ln O_0(t) = \boldsymbol{s}(\ln t;\gamma) = \boldsymbol{s}$, we have the PO model as follows:

$$\ln O(t;\mathbf{x}) = \boldsymbol{s} + \boldsymbol{\varphi}'\mathbf{x} = \boldsymbol{s} + \boldsymbol{\alpha}'\boldsymbol{Fin} + \boldsymbol{\beta}'\boldsymbol{Soc} + \boldsymbol{\theta}'\boldsymbol{Control}. \tag{2}$$

Our survival curves $S(t;\mathbf{x})$ and $h(t;\mathbf{x})$ are expressed as follows:

$$S(t;\mathbf{x}) = 1 - Pr(T \leq t) = [1 + \exp(\boldsymbol{s} + \boldsymbol{\varphi}'\mathbf{x})]^{-1}, \tag{3}$$

$$h(t;\mathbf{x}) = \frac{ds}{dt}\exp(\boldsymbol{s} + \boldsymbol{\varphi}'\mathbf{x})\,[1 + \exp(\boldsymbol{s} + \boldsymbol{\varphi}'\mathbf{x})]^{-1}. \tag{4}$$

Following Bouvatier and Delatte (2015) and Royston and Parmar (2002), the baseline function is a restricted cubic spline function, which exhibits less restrictions on the functional form and mitigates misspecification problem (see Harrell (2001)). A restricted cubic splines function is a smoothed piecewise cubic polynomials function restricted to be linear beyond the two boundary knots, k_{min} and k_{max}, the minimum and maximum of uncensored survival times of platforms. The restricted cubic spline function $\boldsymbol{s}(\ln t;\gamma)$ with m interior knots,

$k_{min}<k_1<\ldots<k_m<k_{max}$, and two boundary knots, k_{min} and k_{max}, can be written as

$$\boldsymbol{s} = \boldsymbol{s}(\ln t;\gamma) = \gamma_0 + \gamma_1 Basis_0 + \gamma_2 Basis_1 + \cdots + \gamma_{m+1} Basis_m,$$ of which,

$Basis_0 = \ln t$, and $Basis_j = (\ln t - k_j)_+^3 - \lambda_j(\ln t - k_{min})_+^3 - (1-\lambda_j)(\ln t - k_{max})_+^3$ for

for $j=1,\ldots,m$. And the "+" function is defined as

$$(\ln t - k)_+ = \begin{cases} \ln t - k, & \text{if } \ln t > k \\ 0, & \text{otherwise} \end{cases}, \text{ and } \lambda_j = \frac{k_{max}-k_j}{k_{max}-k_{min}}.$$

The baseline spline function parameters are orthogonalized to be uncorrelated with each other, have mean zero and unit standard deviation to improve their numerical stability. For the flexible PO(d) model, d is the degree of freedom (d.f.) corresponding to d-1 interior knots. The number of knots can be determined via the Akaike information criterion (AIC) or the Bayes information criterion (BIC) to avoid underfitting or overfitting and the interior knots are evenly placed at every 100/d percentiles as suggested by Royston and Parmar (2002).

In the defaults strand of the literature, time-dependent effect of covariates are usually considered (Gupta et al., 2017). The time-dependent effect is captured by the interaction between the covariate and the spline variables as follows:

$$\ln O(t;\mathbf{x}) = \boldsymbol{s}(\ln t;\gamma) + \sum_{j=1}^{D} \boldsymbol{s}(\ln t;\delta_j)\mathrm{x}_j + \boldsymbol{\varphi}'\mathbf{x} = \boldsymbol{v}\left(\ln t;\gamma + \sum_{j=1}^{D}\delta_j\mathrm{x}_j\right) + \boldsymbol{\varphi}'\mathbf{x} = \boldsymbol{v}(\ln t;\varpi) + \boldsymbol{\varphi}'\mathbf{x}. \quad (5)$$

where D is the number of time-dependent effect variables, δ_j is the parameter estimates of the spline function of time-dependent effect variable, x_j is the time-dependent effect variable. We constrain knots number and placements to those of the baseline spline function.

2.2. Nonlinear transformation of variables

There is widespread interest in considering potential nonlinearity transformation of variables in the literature (Nikolaeva et al., 2015), since nonlinearity can more precisely model the variables and better discover the phenomena/theory (Nikolaeva et al., 2015). But rather than use logarithmic or quadratic functions with predefined shapes, we use fractional polynomial (FP) function with flexible shapes following Royston and Sauerbrei (2008) and Sauerbrei and Royston (1999). In specific, for a given variable x, the nonlinear relationship can be modeled by a fractional polynomial function with m degree and powers p_m as follows:

$$\mathrm{FP}_m(\mathrm{x}) = \beta_1 \mathrm{x}^{p_1} + \cdots + \beta_m \mathrm{x}^{p_m} = \mathrm{x}^*. \quad (6)$$

The set of powers p is $\{-2,-1,-0.5,0,0.5,1,2,3\}$ and $p=0$ is the natural logarithm. For a 2 degree FP where $p_1=p_2$, $\mathrm{FP}_2(\mathrm{x})=\beta_1\mathrm{x}^{p_1}+\beta_2\mathrm{x}^{p_1}\ln\mathrm{x}$. In our PO model we use the FP with one power term, FP(1). The decision of FP transformation depends on whether it decreases the deviance (minus two maximized log likelihood), and the degree of power term is determined by the best-fitting FP_m that generates the minimum deviance.

2.3. Model estimation

The PO(*d*) model is estimated via the full maximum likelihood. The log-likelihood function for platform *i* is denoted as L_i, and the likelihood for the whole sample is $\prod_{i=1}^{n} L_i$. Let $\mathrm{o}_i=\boldsymbol{v}(\ln t;\varpi)+\boldsymbol{\varphi}'\mathbf{x}$, then its first derivative is $\mathrm{o}_i{}'=d\mathrm{o}_i/dt_i=d\boldsymbol{v}(\ln t_i;\varpi)/dt_i$. The likelihood function for a PO(*d*) is as follow:

$$L_i=\begin{cases}\mathrm{o}_i{}'\exp(\mathrm{o}_i)(1+\exp\mathrm{o}_i)^{-2}, & \text{for an uncensored observation}\\ (1+\exp\mathrm{o}_i)^{-1}, & \text{for a censored observation.}\end{cases} \tag{7}$$

And our econometric analysis is conducted via the following steps: First, choose the d.f. for the PO model via the AIC and/or BIC values to identify the most appropriate proportionality parameterization. Second, perform a backward elimination process and remove the least significant variables. Third, apply a backward selection process to identify nonlinear variables and choose the best fit FP transformation using closed-test algorithm. Fourth, extend the model by including possible time-dependent effect. A forward selection process is implemented using the 1% significance level. Fifth, plot the smoothed martingale residuals to assess the goodness of fit. Finally, investigate covariate effects graphically.

3. Data

3.1. Data source

The social capital data are hand collected from the two most vibrant third-party professional sites, WDZJ (www.wdzj.com) and P2Peye (www.p2peye.com), which serve investors and platforms by providing free service such as investor community discussion forum, platform data disclosure, platform directory, and industry statistics. Based on Alexa website traffic ranking in China, these two websites are the top two most visited directory websites in online P2P lending industry. The financial capital data come from three sources. WDZJ and P2Peye disclose risk transfer and mitigation data based on information extracted from platforms' official websites. Company registry information are collected from the government official database, National Enterprise Credit Information Publicity System (www.gsxt.gov.cn).

Prefecture-level economic capital data are from the database of National Bureau of Statistics of China (http://data.stats.gov.cn/english/), and from city statistics bureau respectively. The final data source is the official websites for platforms for supplement or for cross check.

3.2. Variables

3.2.1. Platform survival

Our study examines P2P platforms founded between June 2007, when the first P2P online lending platform was established, and June 2017. Survival is defined if a P2P lending platform's website is still accessible, its contact numbers can be reached and its business is normally and independently operating. Our definition for survival is consistent with the literature on firm survival (Lyles et al., 2004); and the event indicator is 0 for these survived platforms, and 1 for failed platforms. For platforms that are ongoing by the end of our sample period, its lifespan is denoted as the difference between end of June 2017 and its entry date, but the event indicator will be 0, so right censoring will occur in this case, but PO model can deal with the issue. Besides, on condition that a platform was merged or acquired by another company, depending on whether the new entity is still running in online P2P lending industry, we categorize it differently: event indicator is 1 if the new entity is still running business in online P2P lending industry, and 0 if the platform changes to another industry.

3.2.2. Social capital variables

To operationalize the concept of social capital, we propose two measures based on the seminar work of Adler and Kwon (2002): online social network and online social reviews. The online social network is essential for a platform to establish and generate social relations, whereas online social reviews help strengthen the social relations (Gunilla & Mariam, 2004; Wang et al., 2013). Online social network is measured as friends number (Wang et al., 2013; Yang & Li, 2016) and online information sharing activity (Gunilla & Mariam, 2004). We use subscriber number (*Subscribers*) as a proxy for friends number. Posts number (*Posts*) is used to measure online information sharing activity. Both variables are scaled by platform age considered that older platforms tend to have more subscribers other things being equal.

Online social reviews can be measured as detailed service quality ratings and online reviews (Wang et al., 2013). Investors are able to share their investment experience by leaving detailed service quality ratings (*DSQR*), which are categorized into four aspects of platform service: cash withdrawal time, capital utility rate, customer service, and user experience. Each category is graded from 1 to 5 with the highest score being the best. We use the average of the four ratings (*mDSQR*) to capture this variable following Wang et al. (2013). The variance of DSQR (*vDSQR*) is used for robustness checks. Investors are also able to put down reviews for a particular platform and also attribute the reviews to one of three categories: "Poor; do not recommend", "Neutral", and "Excellent; recommend". Thus we derive the ratio of negative/neutral/positive ratings by taking the percentage of negative/neutral/positive ratings over the total number of ratings. We only include negative ratings (*Neg_rating*) for two reasons. First, the literature has shown that good news is expected and neutral news is not influential (Fiske, 1980). The other reason is to avoid potential multicollinearity. When using percentage negative ratings, a zero percent can indicate either the platform is very good and do not receive negative ratings, or no investors has rated the platform. To differentiate between these two scenarios, we add a dummy variable (*No_rating*).

3.2.3. Financial capital variables

Financial capital is the financial resources available to a company. In the online P2P lending platform context, we identify financial capital as the capital providing safety net, risk transfer, risk mitigation and the prefecture-level economic support to a platform following the bank failure literature (Betz et al., 2014; Calomiris & Mason, 2003). The registered capital (*Capital*) is expected to serve as safety net and cover various costs since platforms usually have heavy operational costs (Cubillas et al., 2017). For risk transfer, platforms usually seek external guarantee such as insurance companies for sustainability improvement. We use a dummy variable (*Out_guaran*) to indicate if a platform has outside guarantees and a variable (*Out_guaran_amt*) to show log external guarantee amount. Another dummy (*Loan_nego*) for risk transfer is to see whether a secondary market allowing investors to negotiate loans is built. Risk mitigation measures can help relieve a platform's financial distress. Two common measures platforms take are to require borrower collateral against their borrowing (*Brrw_cllatrl*) or to set up risk fund reserve (*Riskfund*). Platforms also try to mitigate investor risk exposure by committing to pay back investors principal (*Prin_guaran*) or advancing capital to investors if a loan turns bad (*Advance*). The GDP per capita (*GDP_capita*) is used as a proxy for prefecture-level economic capital because people live in more well-off cities are more likely to invest more and more tolerant to losses. Although P2P lending platforms operate online, they are governed and supported by local regulatory authority.

3.2.4. Control variables

A number of control variables are included to avoid potential variable omission problem, including dummies to indicate whether platforms provide hotline (*Hotline*), interactive system (*App*) or transaction system (*Autobid*), and variables to illustrate self-disclosure level (*News*) and firm size (*Employees*).

3.3. Descriptive analysis

Figure 1 shows platforms lifespan distribution in our sample. It is apparent that the industry is dominated by young companies. Only around 2% of platforms last more than five years, and only 0.2% of platforms survive more than eight years. Descriptive and correlation statistics are available upon request. Most social capital and financial capital variables are right-skewed, indicating that many platforms are underperformers in the industry, with some very strong performers on the other end of the distribution. Correlation in most cases is between -0.50 and 0.50 and not cause for collinearity concern.

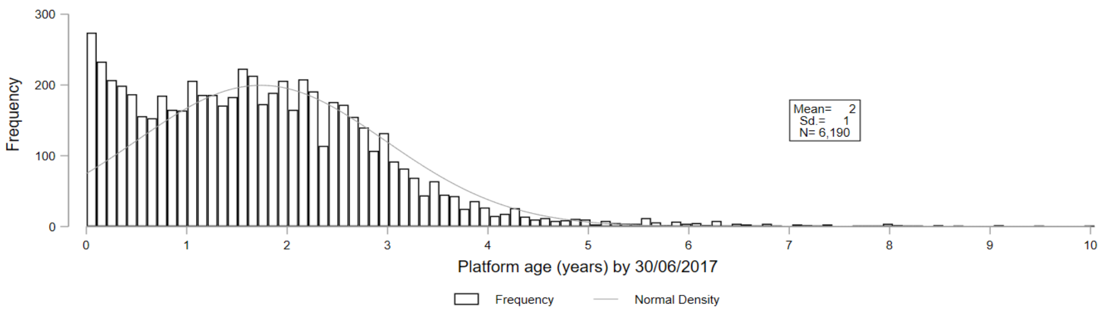

Figure 1: Histogram of platform distribution

4. Empirical analysis

4.1. Estimation results

Our baseline results are reported in Table 1 using the PO(3) model without considering nonlinear transformation of variables or time-dependent effect. We prefer the PO(3) model which combines low AIC/BIC value with model parsimony, and the problem of underfitting or overfitting is avoided. Three main results are apparent. First, positive social relations between platforms and their customers decreases the default probability and increases the lifespan. Second, Model (2) shows that platforms benefit generally from their financial safety net, risk transfer and mitigation measures, and prefecture-level economic development, but the money back policy (*Prin_guaran*) puts platforms at a higher risk of failure. Third, when both social and financial capital variables are included in Model (3), the significance of all but two coefficients for financial capital variables are reduced and some of them become insignificant; while the sign and significance of social capital variables remain.

Interestingly, the number of posts is positively related to a platform's failure, which seems counter-intuitive since more posts is usually interpreted as more social visibility and relations (Wang et al., 2013). We run a text description analysis and extract the 30 most frequently used words in online posts for platforms less than 2 years old, the sample mean lifespan, for failed and surviving platforms separately. Results show that for the failed platforms the most often used words are negative including fraud, rights problem and so forth whereas the corresponding words for surviving platforms are more positive such as return, experience, and security. Furthermore, more posts are published for failed platforms than surviving ones. We believe that in this context the saying *no news is good news* is evidenced

	(1)		(2)		(3)	
	Coefficient	z-statistics	Coefficient	z-statistics	Coefficient	z-statistics
Subscribers	-1.568***	(-25.50)			-1.207***	(-17.60)
Posts	1.731***	(26.80)			1.813***	(26.25)
mDSQR	-0.014	(-0.59)			-0.057**	(-2.27)
Neg_rating	1.064***	(10.26)			0.884***	(8.08)
No_rating	1.121***	(14.12)			0.917***	(10.78)
Capital			-0.086***	(-5.87)	0.0142	(0.82)
Loan_nego			-1.111***	(-18.23)	-0.357***	(-4.91)
Out_guaran			-0.503***	(-9.28)	-0.061	(-0.98)
Out_guaran_amt			-0.037***	(-4.94)	-0.016**	(-2.15)
Brrw_cllatrl			-0.935***	(-6.14)	0.295*	(1.74)
Riskfund			-0.526***	(-8.87)	-0.041	(-0.63)
Prin_guaran			0.450***	(7.30)	0.811***	(10.83)
Advance			-0.022	(-0.36)	0.097	(1.46)
GDP_capita			-0.508***	(-8.87)	-0.238***	(-4.04)
Basis0	1.748***	(72.08)	1.355***	(71.67)	1.812***	(70.09)
Basis1	-0.286***	(-18.55)	-0.209***	(-15.66)	-0.303***	(-18.66)
Basis2	-0.067***	(-4.70)	-0.086***	(-7.40)	-0.063***	(-4.15)
Controls	NO		NO		YES	
Constant	-1.278***	(-17.31)	7.115***	(10.52)	2.356***	(3.30)
N	6190		5859		5842	
LogL	-6658		-7551		-6029	
R_D^2	0.518		0.147		0.559	
C	0.782		0.653		0.796	
D	0.564		0.307		0.591	

Note: *, **, and *** denote significance at the 10%, 5% and 1% level, respectively.

Table 1. Parameter estimates for baseline PO(3) model

.

The long list of significant coefficients in the baseline results makes it difficult for investors, policymakers and platform managers to clearly see the message. Hence we follow the model estimation procedure outlined in Section 2.3 and undertake a prognostic test. Results are summarized in Table 2. In Table 3, we include all significant variables from the prognostic test. The results are consistent with our baseline models, further substantiating the stability and robustness of our model. Model (5) is our ultimate model for P2P lending platform survival. In addition, the smoothed martingale residuals of coefficient estimation see no systematic departure from zero, indicating a good fit for Model (5).

Variables	Deviance difference, closed-test procedure		FP transformation	Time-dependent effect	
	FP(1) versus null	FP(1) versus linear	Powers or exclusion	Include or not	d.f.
Subscribers	548.723***	0	1	no	—
Posts	1502.350***	272.745***	0.5	no	—
mDSQR	7.579*	0.478	1	no	—
Neg_rating	57.347***	5.446**	2	no	—
No_rating	131.183***	0	1	include	2
Capital	5.364	—	exclude	no	—
Loan_nego	20.651***	0	1	no	—
Out_guaran	1.531	—	exclude	no	—
Out_guaran_amt	4.682	—	exclude	no	—
Brrw_cllatrl	4.846	—	exclude	no	—
Riskfund	0.042	—	exclude	no	—
Prin_guaran	99.616***	0	1	include	3
Advance	0.948	—	exclude	no	—
GDP_capita	12.378***	0	1	no	—
Hotline	32.197***	0	1	no	—
App	36.945***	0	1	no	—
Autobid	0.142	—	exclude	no	—
News	81.899***	28.321***	-2	no	—
Employees	58.695***	7.787***	-1	include	1

Note: *, **, and *** denote significance at the 10%, 5% and 1% level, respectively.

Table 2. Selection of influential variables, nonlinear transformation and time-dependent effect

	FP Transformation	(4)		(5)	
		Coefficient	z-statistics	Coefficient	z-statistics
Subscribers	—	-1.246***	(-19.99)	-1.282***	(-19.29)
*Posts**	$(Posts)^{0.5}$	2.925***	(32.12)	2.852***	(31.71)
mDSQR	—	-0.069***	(-2.74)	-0.082***	(-3.07)
*Neg_rating**	$(Neg\_rating)^{2}$	0.875***	(7.87)	1.032***	(8.65)
No_rating	—	0.958***	(11.52)	1.029***	(11.52)
Loan_nego	—	-0.330***	(-4.57)	-0.324***	(-4.22)
Prin_guaran	—	0.748***	(10.93)	0.614***	(7.59)
GDP_capita	—	-0.225***	(-3.88)	-0.218***	(-3.76)
Hotline	—	-0.368***	(-4.9)	-0.323***	(-4.55)
App	—	-1.342***	(-5.08)	-1.317***	(-4.99)
*News**	$(News)^{-2}$	5.56e-09***	(9.22)	5.52e-09***	(8.95)
*Employees**	$(Employees)^{-1}$	0.055***	(8.02)	0.043***	(6.32)
Basis0	—	1.856***	(72.12)	2.124***	(26.12)
Basis1	—	-0.327***	(-20.79)	-0.124*	(-1.86)
Basis2	—	-0.051***	(-3.35)	-0.088***	(-5.23)
Basis_No_rating0	—			-0.337***	(-3.91)
Basis_No_rating1	—			-0.144**	(-2.01)
Basis_Prin_guaran0	—			0.618***	(6.19)
Basis_Prin_guaran1	—			-0.113	(-1.46)
Basis_Prin_guaran2	—			0.240***	(4.99)
Basis_Employees0	—			-0.014***	(-2.79)
Constant	—	0.170	(0.25)	0.003	(0.00)
N		5859		5859	
LogL		-5897		-5813	
R_D^2		0.586		0.636	
C		0.801		0.806	
D		0.601		0.613	

Note: *, **, and *** denote significance at the 10%, 5% and 1% level, respectively.

Table 3. Prognostic model with nonlinearity transformation and time-dependent effect

To summarize, the consistency in coefficient estimates across Model (1) to Model (5) reinforces our key message on P2P lending platform survival: these online platforms still rely on financial resources for their survival in a way similar to more traditional financial service entities. However, operational entirely online has added an additional dimension to their core characteristics thus making social capital a more essential driver for survival. In the online context, when social capital is absent, financial capital offers valuable indication of platform operation. However, when social capital is available, it provides more dynamic and detailed information on customer population, ratings and feedback, hence proves more influential to the decision-making of customers, leading to its more significant impact on platform survival. These results enrich and advance our understanding of the determinants for the survival of Chinese online P2P lending platforms.

4.2. Graphical analysis

Figure 2 illustrates the time-varying effect of a variable exerts on the hazard rate. The hazard ratio is between the hazard rate of a variable plus one standard deviation and the hazard rate of the variable itself. For dummy variables, the ratio is between the hazard rate of 1 and that of zero. For social capital variables, subscriber number has a long run effect, whereas others have comparatively short-term effect. For example, the effect of posts number is quite large at t=0 but drops swiftly and massively within two years. For financial capital variables, prefecture-level economic capital exhibits a long-term effect whereas the other two exhibit shorter-term effect. Money back policy (*Prin_guaran*) improves survivability at a very early stage to around four months as it mitigates the platform's liability of newness (Freeman et al., 1983). However, the benefit gradually diminishes for the first few years of operation, but if the platform can survive long enough, this negative effect subsequently fades away.

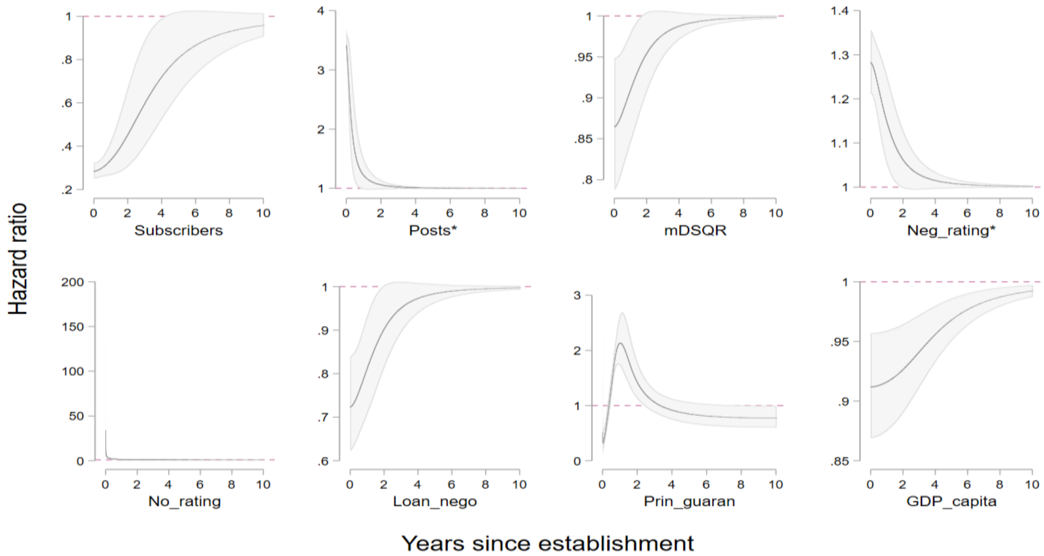

Figure 2. Effect of financial and social capital varaibles on hazard ratio, with 95% CI band

The economic intuition of social and financial capital variables indicates that in order to enhance platform survival probability more efficiently, a platform can either increase its social capital by obtaining more subscribers, having less negative posts, and soliciting better investor rating, or increase its financial capacity by opening a secondary market, growing independent from money back policy, and locating in economically better off cities.

4.3. Prediction analysis

The prognostic index $(\widehat{\boldsymbol{\varphi}'\mathbf{x}})$ estimated via our model can illustrate the level of a platform's default risk and help predict its survival. Figure 3 plots platform survival time against the prognostic index, with the 10^{th} and the 90^{th} percentile representing low and high default risk, respectively. For firms with medium-level risk, the survival time for bottom and top 10^{th} suggests that 80% platforms with medium risk operate between 9 to 67 months before they default. Information of this nature is very useful, especially to investors for making informed investment decision, and to regulators for introducing relevant and directed policy.

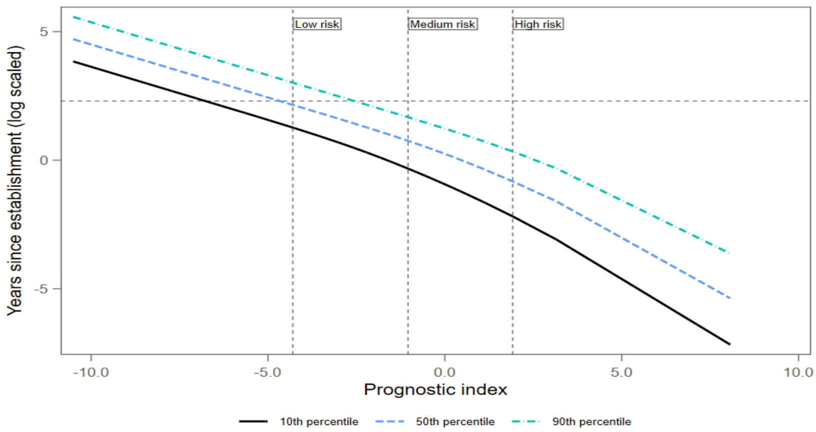

Figure 3. Survival time in 10th, 50th, and 90th percentiles

4.4. Robustness check

To show that our results are not due to specific choice for variables or estimation method, we conduct four robustness tests, including dividing the outsider guarantee variable (*Out-guaran*) into four specific common types, using the variance of DSQR rather than the mean, applying an extended Cox model (Cox, 1972), and using variables without FP transformation; we find qualitatively similar results. Results are available upon request.

5. Conclusion

We assess the role of social capital variables and financial capital variables in determining survival patterns of online P2P lending platforms via implementing a flexible proportional odds model with a baseline spline function as well as considering potential nonlinearity and time-dependent effect of variables. Our prognostic model captures 12 influential variables out of 19 available variables, and it shows nice fit and robust prediction power. The model is simple but powerful in providing reference to investors and policy makers in monitoring platforms. It is flexible enough to accommodate new variable and data at higher frequency.

We also uncover a host of important findings. First, social capital is a key driver to platform survival and is more significant than financial capital in economic terms. For investors and policy makers, social capital is helpful for distinguishing high and low risk lending platforms. Financial capital is important to platforms but registered capital can be misleading to investors as it is not a significant indicator for platform survival. For platform managers, not all social capital deserves their effort in the long run. We find that increasing subscriber number and achieving better customer ratings are the key to survive longer. In addition, opportunities of information exchange between platform and customers and between customers themselves enhance survival probability. We also argue that gathering a large number of posts at an early stage is not a good sign for survival as investors are more likely to punish those platforms by posting negative comments and sharing poor experience.

---

[i] The information is disclosed by the WDZJ database, which is introduced in Section 3.1.

References

Adler, P. S., & Kwon, S. W. (2002). Social capital: rospects for a new concept. *The Academy of Management Review, 27*(1), 17-40.

Bennett, S. (1983). Analysis of survival data by the proportional odds model. *Statistics in Medicine, 2*(2), 273-277.

Betz, F., Oprică, S., Peltonen, T. A., et al. (2014). Predicting distress in European banks. *Journal of Banking & Finance, 45*(Supplement C), 225-241.

Bouvatier, V., & Delatte, A.-L. (2015). Waves of international banking integration: a tale of regional differences. *European Economic Review, 80*, 354-373.

Calomiris, C. W., & Mason, J. R. (2003). Fundamentals, panics, and bank distress during the depression. *The American Economic Review, 93*(5), 1615-1647.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological), 34*(2), 187-220.

Cubillas, E., Fernández, A. I., & González, F. (2017). How credible is a too-big-to-fail policy? International evidence from market discipline. *Journal of Financial Intermediation, 29*, 46-67.

Datta, P., & Chatterjee, S. (2008). The economics and psychology of consumer trust in intermediaries in electronic markets: the EM-Trust Framework. *European Journal of Information Systems, 17*(1), 12-28.

Fiske, S. T. (1980). Attention and weight in person perception: the impact of negative and extreme behavior. *Journal of Personality and Social Psychology, 38*(6), 889-906.

Freeman, J., Carroll, G. R., & Hannan, M. T. (1983). The liability of newness: age dependence in organizational death rates. *American Sociological Review*, 692-710.

Gelfand, A. E., Ghosh, S. K., Christiansen, C., et al. (2000). Proportional hazards models: a latent competing risk approach. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 49*(3), 385-397.

Gunilla, W.-W., & Mariam, G. (2004). Explaining knowledge sharing in organizations through the dimensions of social capital. *Journal of Information Science, 30*(5), 448-458.

Gupta, J., Gregoriou, A., & Ebrahimi, T. (2017). Empirical comparison of hazard models in predicting SMEs failure. *Quantitative Finance,* doi:10.1080/14697688.2017.1307514, 1-30.

Harrell, F. E. (2001). *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*: Springer, New York, USA.

Hoffman, D. L., Novak, T. P., & Peralta, M. (1999). Building consumer trust online. *Communications of the ACM, 42*(4), 80-85.

Karimov, F. P., Brengman, M., & Van Hove, L. (2011). The effect of website design dimensions on initial trust: a synthesis of the empirical literature. *Journal of Electronic Commerce Research, 12*(4), 272-301.

Kim, H.-W., Xu, Y., & Koh, J. (2004). A comparison of online trust building factors between potential customers and repeat customers. *Journal of the Association for Information Systems, 5*(10), 13-61.

Lins, K. V., Servaes, H., & Tamayo, A. N. E. (2017). Social capital, trust, and firm performance: the value of corporate social responsibility during the financial crisis. *The Journal of Finance, 72*(4), 1785-1824.

Lyles, M. A., Saxton, T., & Watson, K. (2004). Venture survival in a transitional economy. *Journal of Management, 30*(3), 351-375.

Mollick, E. (2014). The dynamics of crowdfunding: an exploratory study. *Journal of Business Venturing, 29*(1), 1-16.

Nikolaeva, R., Bhatnagar, A., & Ghose, S. (2015). Exploring curvilinearity through fractional polynomials in management research. *Organizational Research Methods, 18*(4), 738-760.

Orth, W. (2013). Multi-period credit default prediction with time-varying covariates. *Journal of Empirical Finance, 21*, 214-222.

Putnam, R. D., Leonardi, R., & Nanetti, R. Y. (1994). *Making democracy work: civic traditions in modern Italy*: Princeton university press.

Royston, P., & Parmar, M. K. (2002). Flexible parametric proportional‐hazards and proportional‐odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in Medicine, 21*(15), 2175-2197.

Royston, P., & Sauerbrei, W. (2008). *Multivariable model-building: a pragmatic approach to regression anaylsis based on fractional polynomials for modelling continuous variables* (Vol. 777): John Wiley & Sons.

Sauerbrei, W., & Royston, P. (1999). Building multivariable prognostic and diagnostic models: transformation of the predictors by using fractional polynomials. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 162*(1), 71-94.

Wang, Y., Wang, S., Fang, Y., et al. (2013). Store survival in online marketplace: an empirical investigation. *Decision Support Systems, 56*, 482-493.

Wheelock, D. C., & Wilson, P. W. (2000). Why do banks disappear? The determinants of US bank failures and acquisitions. *Review of Economics and Statistics, 82*(1), 127-138.

Yang, X., & Li, G. (2016). Factors influencing the popularity of customer-generated content in a company-hosted online co-creation community: a social capital perspective. *Computers in Human Behavior, 64*(Supplement C), 760-768.

Zhang, J., & Liu, P. (2012). Rational herding in microloan markets. *Management Science, 58*(5), 892-912.