Fostering Entrepreneurship through Crowdfunding: What Drives Local Biases?*

Jian Ni†
Virginia Tech

Yi Xin‡
California Institute of Technology

Fangzhu Yang§
Johns Hopkins University

September 8, 2023

Abstract

We present empirical evidence consistent with strong local biases among funders in online crowdfunding markets. What drives these biases? We find that in addition to funders' local preference, information frictions play a more important role in driving the local biases, preventing high-quality entrepreneurial projects from being identified and funded by investors outside their regions. Exploring the role of platform-design features, we find that, with the presence of strong local biases, not revealing locational information of projects could be welfare improving. Business strategies could be tailored for different project categories to help mitigate non-local funders' informational disadvantages and foster local entrepreneurship.

Keywords: Entrepreneurship, crowdfunding, local bias, information frictions, local preference

*The usual disclaimer applies.

†Pamplin College of Business, Virginia Tech, Data and Decision Sciences (D&DS) Building, VA 24060. Email: jiann@vt.edu.

‡Division of the Humanities and Social Sciences, California Institute of Technology, 1200 E California Blvd, MC 228-77, Pasadena, CA 91125. Email: yixin@caltech.edu.

§Department of Economics, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218. Email: fangzhu@jhu.edu.

1 Introduction

Micro-entrepreneurs have been increasingly utilizing crowdfunding platforms to raise funds for their projects (Belleflamme et al., 2014; Bruton et al., 2015).[1] This increase is particularly salient in markets where access to credit is difficult or traditional banking infrastructure is weak (Narasimhan et al., 2015; Bao et al., 2018; Sharma et al., 2019). Although online crowdfunding platforms enable entrepreneurs and funders to interact conveniently with fewer geographic constraints, information asymmetry and preference toward local entrepreneurial projects may continue to play an important role in these markets.[2] For example, Baltimore's local nonprofit organization–Community Wealth Builders–connects small, mostly minority-owned businesses with crowdfunding platforms such as Honeycomb credit. Since the COVID pandemic, more than 70 small businesses in the Baltimore area have successfully raised funds through this channel and the funders are mostly local people, including many existing customers and investors who "love Baltimore, shop local, and wear purple on Fridays".[3]

While receiving financial support from the local community benefits the micro-entrepreneurs, information frictions and biases towards local projects can prevent high-quality entrepreneurial projects from being identified and funded by a greater population of investors outside their regions. The goal of our paper is to investigate whether local biases are indeed pervasive in online crowdfunding markets, and if so, to what extent they are driven by local preferences and information frictions. Furthermore, we want to better understand what managerial strategies can be performed to reduce local biases and help expand credit access for micro-entrepreneurs. The answer to these questions would be informative in designing policies to stimulate entrepreneurial activities, which are an important driver of job creation, innovation, and long-term economic growth.

Exploiting a unique dataset on individuals' funding behavior from a large crowdfunding plat-

[1] A recent survey by Jayachandran (2020) shows that expanding access to microcredit (through cash transfers, grants, micro-loans, etc.) increases profits and fosters micro-entrepreneurship.

[2] Agrawal et al. (2014) discuss the prevalence of information asymmetry in crowdfunding markets in a theoretical framework.

[3] https://www.baltimoresun.com/business/bs-bz-crowdfunding-aids-baltimore-businesses-20220914-2ans3psp3jdprnxyouv6qcy4ce-story.html

form in China, we find evidence of strong local bias even in the online scenario: funders are significantly more likely to invest in projects that are from the same region, all else equal.[4] We then focus on two channels that could be driving this bias. First, funders may have inherent preferences for projects from their hometown (i.e., preference channel), which may originate from non-pecuniary factors such as social capital, cultural identity, and enthusiasm in supporting local community. Second, funders might exhibit informational advantages when evaluating projects from their region, such as being more familiar with the local environment and specialty products and receiving private information about the quality of the projects (i.e., information channel).[5] Our paper provides quantitative estimates of the impact of local preference and information frictions on inducing local biases. We find the latter accounts for two-thirds of the total effect in the focal setting.

The crowdfunding platform we study operates similarly to Kickstarter, where individuals promote various projects (related to agricultural products, charity purposes, arts, entertainment, IT, etc.) on the website to raise funds. Once the project is funded, the fundraisers are responsible for completing the projects and distributing their promised rewards. The dataset contains detailed information on the characteristics of each project and the historical investment behavior of the funders, including past projects they funded and the amount they contributed to them. We also observe geographical information for both the projects and funders at the provincial level.

To disentangle the information and the preference channels, we use text-mining techniques to classify projects into two groups: location-related (LR; e.g., selling agricultural products) and non-location-related projects (NLR; e.g., developing mobile applications). Intuitively, agents' funding behavior for LR projects is affected by both preference and information channels, whereas funding for NLR projects, which do not rely on local expertise, is mainly driven by the preference channel. For both types of projects, our regression analysis shows funders are more likely to invest in those

[4] Similar patterns have been documented in other online marketplaces, such as e-commerce markets (Hortacsu et al., 2009), peer-to-peer lending websites (Lin and Viswanathan, 2016), and online fund-raising platforms for musicians (Agrawal et al., 2015), etc.

[5] For example, Ivkovic and Weisbenner (2005) find that individual households are able to process and exploit locally nonpublic information to achieve superior stock returns and the effects are even stronger in scenarios where information asymmetries are likely to be more prevalent (e.g., less well-known stocks).

from their own provinces. In addition, we also find local biases for LR projects are more salient, indicating the importance of the information channel.

Motivated by the reduced-form evidence, we estimate a structural model of agents' funding behavior on the crowdfunding platform. In this model, a potential funder attaches a premium to projects from her own region but also faces uncertainty about their quality. We assume that for LR projects, the agent receives a more precise signal about the quality of the project if the locations of the project and the investor match, whereas for NLR projects, the agent receives the same signal regardless of whether the locations match. Under this model, agents are more likely to invest in same-location projects because of the local-preference premium. Risk-averse agents derive additional benefits from investing in LR projects in the same province, due to informational advantages, which further exacerbates local bias.[6] Our estimation results suggest agents tend to be less interested in investing in LR projects within online crowdfunding markets. However, we find a local-preference premium that is about five times larger than the disutility from investing in LR projects. If a project is LR and from the same province as the investor, the variance of the quality signals is 88.25% of the variance in other cases, which leads to a higher investment probability, due to informational advantages.

With these structural estimates, we conduct two sets of counterfactual experiments. First, we isolate the effects of each channel by shutting them down sequentially. By comparing the market outcomes when only the information channel is shut down and when both are shut down, with the original market outcomes, we quantify the effects of the preference and the information channel separately on inducing local biases. We find that for LR projects, shutting down the two channels reduces the number of funders per project by 1.16 (19.8%), and 64.32% of the reduction is attributed to the information channel. In addition, we find that when the information channel is shut down, the amount contributed by same-province investors decreases by 50.80%, which is almost twice as large as the impact of the preference channel, resulting in more than a 264,000 CNY

[6] The individuals participating in the crowdfunding markets are often risk averse, have lower net worth and income (SEC Investor Bulletin, 2017), and are more prevalent in the emerging markets. These characteristics make them differ from angel investors, who are often high-net-worth individuals in developed markets (Derdenger and Srinivasan, 2019).

reduction in the platform's revenue.

The second set of counterfactual experiments focus on evaluating the impact of platform-design policies that aim to reduce local biases and improve the informativeness of the project descriptions. We consider hypothetical scenarios where the location information of the project, the number of comments received by the project, and the option to use video presentations for advertisement are removed from the website. Interestingly, we find potential welfare improvement in the form of additional contributions when we remove the projects' location information, which can help attract more investors from different regions that otherwise may not have contributed. We also find that removing comments from the website disincentivizes funders from making contributions, because the comments are potentially an informative signal about the quality of the projects from a social-learning perspective. Lastly, we find the impact of removing the video from project descriptions is heteogeneous across project types. For example, whereas video removal increases the number of investors for projects related to agriculture, an opposite effect occurs for design-related projects. This finding highlights the importance of designing more flexible and targeted marketing strategies for different types of projects in online crowdfunding markets.

Related Literature Our paper is closely related to the literature that studies the geographic patterns of trading and investment in online marketplaces (Hortacsu et al., 2009; Agrawal et al., 2015; Lin and Viswanathan, 2016). These papers, rooted in the home-bias literature,[7] offer various explanations underlying the local-bias phenomenon in online settings, including behavioral biases, social capital, cultural factors, and the possibility of direct contract enforcement. Our paper follows a rationality-based framework and focuses on disentangling the effects of local preference versus information frictions on funders' local-bias behavior. We depart from other work that estimates regressions of home-bias measures on proxies of information frictions[8] by developing and estimating a structural model to quantify the extent to which informational advantages held by local investors drive their local bias.

[7] For a more detailed review of the equity home bias literature, see the survey by Cooper et al. (2013).

[8] Examples include proportion of public US listing (Ahearne et al., 2004), capital market openness (Bekaert and Wang, 2009), familiarity (e.g., using the same language, see Chan et al. (2005)).

Our work also builds on the literature on designing microfinance and crowdfunding platforms to stimulate investment and foster micro-entrepreneurship. For example, Strausz (2017) uses a theoretical framework to analyze the design of deferred payments in controlling entrepreneurial moral hazard. Xin (2020) investigates the benefits of reputation and feedback systems in facilitating transactions in peer-to-peer lending markets. Deb, Oery, and Williams (2021) analyze the impact of disclosing a crowdfunding campaign's progress on its final success, and they also find allowing seed money at the start might lead to worse outcomes. Xu and Ni (2022) investigate whether and how crowdfunding outcomes could impact entrepreneurs' product-launch decisions. Our paper adds to this literature by exploring the role of platform design in mitigating investors' biases toward local projects and reducing information frictions between funders and entrepreneurs. Our finding suggests that when strong local biases exist, omitting locational information of projects could lead to more efficient market outcomes.

The rest of the paper is organized as follows. We first summarize the institutional background and data patterns in Section 2. Empirical evidence on local biases is provided in Section 3. We present the structural model in Section 4. The identification and estimation results are shown in Section 5. In Section 6, we conduct counterfactual experiments, and Section 7 concludes.

2 Institutional Background and Data

The dataset we use is from one of the largest online crowdfunding marketplaces in China.[9] The platform was founded in 2013 and is considered one of the most impactful crowdfunding platforms in China. Its operation is similar to that of Kickstarter: individuals propose projects on the platform to raise funds, and funding is on an all-or-nothing basis. During the sample period when the data were collected, the projects on this platform were sorted based on their posting dates. The website provided a filter based on project category, but no algorithm accounted for the potential in-

[9] Funders on this platform might also invest in projects on other competing crowdfunding platforms. Lacking of data from other platforms, we are unfortunately unable to model funders' investment decisions across different websites.

vestors' preferences, location, or past funding behavior to recommend a project.[10] Investors decide whether to contribute after reviewing the details of a particular project. Once the project is successfully funded, the website charges a 3% commission fee, and the entrepreneurs are responsible for completing the projects and distributing the rewards they initially promised (e.g., agricultural products, music albums, or services) to the funders.

We collect project-level characteristics and information of funders associated with each project for all listings proposed between March 2013 and February 2017. In particular, the dataset includes the amount of money requested by the entrepreneurs, whether the entrepreneurs' identities have been verified, categories of the projects, the presence of a video advertisement, and locational information of each project. We also observe proxies of the popularity of each project, such as the number of supporters, comments, and people who expressed interest in the project. We observe a unique identifier for each of the funders, which allows us to build a panel dataset of agents' funding behavior. Each funder may be linked to multiple projects, and we observe the amount of contribution money and the time that contribution was made. Like other crowdfunding platforms, the entrepreneurs' family and friends likely also invest in the project. However, because the platform does not collect or reveal whether family or friends endorsed the project, we cannot identify the relationship between the fundraisers and the funders.

Overall, our dataset contains 4,876 proposed listings in the sample period, with a majority (97.77%) of them listed between 2014 and 2016. Table 1 provides summary statistics for the project-level characteristics. The table shows a large proportion of projects were funded (around 95.22%), and the average percentage of requested funds across projects exceeds 100%. We find that, on average, each listing has 87 funders, which is consistent with the crowdfunding feature of the market.[11] The average time period for a project to be active online is around 31 days.

In the data, province-level location information is available for most projects (97.58%). En-

[10] A recent paper by Fu et al. (2021) finds that using machine learning algorithms to select projects in crowd lending markets leads to a higher rate of return for investors and provides more funding opportunities.

[11] The average number of funders per listing in Table 1 is computed based on the total number of investors for each project reported by the website. For analysis of local bias behavior, we focus on funders with location information available. The summary statistics for those investors are provided in Table 2.

trepreneurs are required to provide detailed personal information (e.g., address, phone number, etc.) to the platform, which will be verified upon initiation of the fund-raising process. Fundraisers also tend to provide detailed descriptions and tags for their projects (which the platform highly encourages) to attract potential investors. Among the different project categories, 24% were proposed for charity purposes, and about 18% were related to agricultural products. The rest of the projects were related to arts, entertainment, publishing, technology, and other categories.

Table 1: Summary Statistics: Project Level

Variable	Mean	Std. Dev.	Min	Max	Obs
	Funding information				
Funded	0.9522	0.2133	0	1	4,876
Amount requested (CNY)	34,487.4561	249,057.9736	41	1.00E+07	4,876
Amount raised (CNY)	29,409.3394	107,871.3067	1	3763600	4,876
Percent funded (%)	145.4524	117.3713	5.00E-05	1106.2000	4,876
Number of investors	87.2785	241.8580	0	6,319	4,876
Project duration (days)	31.2754	19.7093	1	141	4,876
	Project details				
Number of certificates	0.3726	0.5250	0	3	4,876
Number of projects by the fundraiser	10.2607	38.4082	1	197	4,876
Number of comments	24.1731	58.7942	0	1,211	4,876
Provide province info	0.9758	0.1537	0	1	4,876
Soft information					
– Provide video	0.2734	0.4457	0	1	4,876
– Provide tags	0.9262	0.2615	0	1	4,876
– Provide description	0.9979	0.0452	0	1	4,876
– Length of title	58.0381	19.7762	6	120	4,876
– Length of tags	22.5619	17.9584	0	112	4,876
– Length of description	157.5295	60.0323	0	367	4,876
	Project categories				
Charity	0.2404	0.4273	0	1	4,876
Agriculture	0.1778	0.3824	0	1	4,876
Entertainment	0.1432	0.3503	0	1	4,876
Arts	0.1224	0.3278	0	1	4,876
Publishing	0.1091	0.3118	0	1	4,876
Technology	0.0519	0.2218	0	1	4,876
Design	0.0498	0.2176	0	1	4,876
Others	0.1054	0.3071	0	1	4,876

In Table C.1 in Appendix C, we report the number of projects across each province and calculate the Herfindahl-Hirschman Index (HHI) based on the market share of projects within each province to measure the degree of market concentration. We estimate that the HHI is 10.08%, suggesting a relatively low degree of market concentration geographically. Although we generally find projects on this platform are distributed across the country, a large number of projects are still coming from more economically developed regions, such as Beijing, Guangdong, and Shanghai.

In the dataset, we observe 20,068 funders with location information available,[12] which we focus on for our empirical analysis of local-bias behavior. Summary statistics of individual-level variables for these funders are provided in the top panel of Table 2. On average, funders with location information contributed to two projects on the website during the sample period, with an average contribution amount equal to 638 CNY. We define *active days* as the number of calendar days between an individual's first and last contribution on the platform. Consistent with consumers' behavior in many other online marketplaces, funders were active for a relatively short period (roughly 40 days). We find that, conditional on a funder investing in a project, about 27% of their contributions went to projects located in the same regions they were from. Given that projects on this website come from provinces across the country (Table C.1 of Appendix C), this finding indicates funders are potentially more likely to fund projects from their own regions.

We also summarize the number of projects invested in, the average amount invested, and the duration on the platform for investors who did not disclose province information (see the bottom panel of Table 2). Comparing this group of investors with those who revealed their location, we find the latter group is, on average, more active on this crowdfunding platform (i.e., invest in more projects and stay longer). However, the medians of the number of projects funded, the amount invested, and the number of active days are similar between the two groups of users. We also compare the percentage of investments made into different project categories by investors with and without province information in Figure 1. The first category includes projects proposed for charitable purposes or for producing agricultural products, and the second category includes projects related to design, arts, entertainment, and IT. Intuitively, unlike art/IT projects, assessing agricultural/charity-related projects requires a more thorough understanding of the local environment, making them more LR. We can see from Figure 1 that the percentages of investments made

[12] Investors are not required to provide detailed personal information to the platform.

into LR or NLR projects by investors with and without province information do not differ qualitatively. This finding mitigates the concern about systematic differences in local-bias behavior between the two groups of investors. Section 6 provides additional comments on the biases that may exist in our empirical findings when using only a subsample of investors who voluntarily reveal their location information.

Table 2: Summary Statistics: Individual Level

Variable	Mean	Median	Std. Dev.	Min	Max	Obs
	Investors with province information					
Number of projects invested	2.0470	1	5.6827	1	351	20,068
Avg. amount invested (CNY)	637.8114	60	7032.6570	0.01	450000	20,068
Active days	39.7717	1	108.1120	1	1098	20,068
Number of same-prov investments	0.4139	0	1.0163	0	54	18,884
Percentage of same-prov	0.2713	0	0.4268	0	1	18,884
	Investors without province information					
Number of projects invested	1.158	1	1.1503	1	297	345,029
Avg. amount invested (CNY)	284.7137	50	2358.567	0.01	455020	345,029
Active days	10.7025	1	53.3772	1	1548	345,029

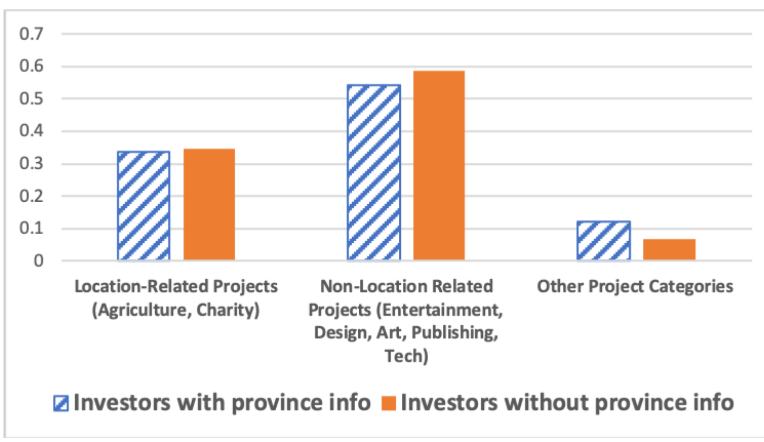

Figure 1: Percentage of investments made into different project categories by investors with or without province information.

3 Empirical Evidence on Local Biases

In this section, we provide empirical evidence on the existence of local biases (Section 3.1) and two channels that drive them (Section 3.2). We discuss robustness checks of our empirical results in Section 3.3.

3.1 Existence of Local Biases

Funders in our sample were active on the platform for about 40 days on average. These investors viewed projects proposed by the entrepreneurs and made funding decisions accordingly. We find that funders tend to focus on recently posted projects on this website and most contributions were made soon after the projects first became available online.[13] This is likely due to the fact that projects were listed in reverse chronological order and most projects were closed within 1-2 months. Given these empirical features of our setting, we assume that investors considered all projects that are available between "the month before their first investment" and "the month of the last investment".[14]

For each pair of a project and a potential investor, we construct an indicator $Invest_{ij}$, which equals 1 if project j was funded by investor i, and 0 otherwise. To demonstrate the existence of local biases, we estimate a logit regression model:

$$Invest_{ij} = \mathbf{1}\{Z_j\beta + b_1 SameProv_{ij} + \varepsilon_{ij} \geq 0\}, \tag{3.1}$$

where Z_j is a vector of control variables, including the amount requested, the number of projects proposed by the fundraiser, the number of comments posted, and project-category dummies. We are interested in estimating b_1, which is the coefficient on $SameProv_{ij}$—an indicator that equals 1 if project j and individual i are from the same province, and 0 otherwise.

The estimation results for Equation (3.1) are reported in the first column of Table 3. We find a significantly positive coefficient on *SameProv*, implying that funders are more likely to contribute to projects in the region where they are from. This finding provides direct evidence of funders' local-bias behavior.

[13] 50% of the contributions were made within 10 days since the projects were first posted online, and over 90% of the contributions were made within 40 days.

[14] For robustness, we also explore alternative approaches to constructing investors' consideration sets and the probability that each project was considered in our empirical analysis. The details are provided in Sections 3.3 and 4.

Table 3: Logit Regression: Baseline

VARIABLES	(1) Invest	(2) SVC Invest	(3) kNN Invest	(4) NB Invest
log(Amount requested)	0.147***	0.146***	0.147***	0.146***
	(0.00393)	(0.00392)	(0.00392)	(0.00392)
# of projects by the fundraiser	-0.00255***	-0.00255***	-0.00262***	-0.00254***
	(0.000330)	(0.000334)	(0.000335)	(0.000335)
# of comments	0.00284***	0.00285***	0.00285***	0.00285***
	(2.37e-05)	(2.36e-05)	(2.36e-05)	(2.36e-05)
# of certificates	0.0486***	0.0515***	0.0513***	0.0535***
	(0.0119)	(0.0119)	(0.0119)	(0.0119)
Provide video	-0.0932***	-0.0906***	-0.0909***	-0.0907***
	(0.0121)	(0.0121)	(0.0121)	(0.0121)
SameProv	1.055***	0.877***	0.880***	0.877***
	(0.0124)	(0.0146)	(0.0147)	(0.0147)
LocRelated		-0.159***	-0.133***	-0.158***
		(0.0142)	(0.0142)	(0.0142)
SameProv $\times$ LocRelated		0.690***	0.660***	0.658***
		(0.0263)	(0.0261)	(0.0260)
Constant	-7.078***	-7.030***	-7.039***	-7.026***
	(0.0388)	(0.0391)	(0.0392)	(0.0393)
Control for category	Y	Y	Y	Y
Observations	12,848,906	12,848,906	12,848,906	12,848,906
AIC	514082.8	513433.1	513478.2	513475.5
BIC	514283.9	513663	513708.1	513705.4

3.2 Preference vs. Information

We focus on two channels that could be inducing the local-bias phenomenon. First, funders may have some inherent “taste” for projects from their hometown (i.e., preference channel), which may originate from non-pecuniary factors such as cultural identity (e.g., attending the same school) or enthusiasm in supporting local community. Second, funders may have informational advantages if the projects are from their own region (i.e., information channel). These funders are potentially more familiar with the local environment and specialty products, which helps them better assess the quality of the proposed project.[15]

[15] For example, if an entrepreneur starts a campaign to raise funds for her bakery shop, an investor from the same area would know better about the location of this business (e.g., whether parking is easy, are competing shops located nearby, etc.).

To disentangle the information and the preference channels empirically, we consider two types of projects: location-related (LR; e.g., selling agricultural products) and non-location-related (NLR; e.g., developing mobile applications) projects. Intuitively, agents' funding behavior for LR projects is affected by both preference and information channels, whereas funding for NLR projects, which do not rely on local expertise, is mainly driven by the preference channel. The crucial step for identification is to accurately classify projects into LR and NLR projects. We adopt text-mining techniques based on the text description of projects for the classification task.

We collect keywords from descriptions and tags of each listing and adopted natural language processing techniques (Gentzkow et al., 2019; Netzer et al., 2019) and create a dictionary of 32 keywords that bear the most essential relevance to the LR/NLR classifications adopting domain experts' knowledge. By mapping the keywords from descriptions and tags of each project to those keywords defining LR/NLR projects, we then proceed to train classifiers to determine whether each project is LR or not. Details on data processing and implementation of the machine learning methods are provided in Appendix A.

In Figure 2, we present the proportion of LR projects across different categories originally specified by the crowdfunding platform. We can see from this figure that the classification results using support vector classifier (SVC), k-nearest neighbors (kNN), and naive Bayes (NB) do not differ meaningfully. A higher proportion of LR projects are in agriculture, charity, and entertainment among all eight categories. This observation is consistent with the idea that evaluating the merits of an agricultural project requires more information about the local environment and weather conditions than art/tech projects. Funders may have more robust knowledge about the quality of an agricultural project if they are from the same area where it is produced.

We analyze the effects of the preference and the information channels on funders' local-bias behavior by estimating the following logit regression model:

$$Invest_{ij} = \mathbf{1}\{Z_j\beta + b_1 SameProv_{ij} + b_2 LR_j + b_3 SameProv_{ij} \times LR_j + \varepsilon_{ij} \geq 0\}, \tag{3.2}$$

where LR_j is an indicator for whether project j is classified as LR using the machine learning meth-

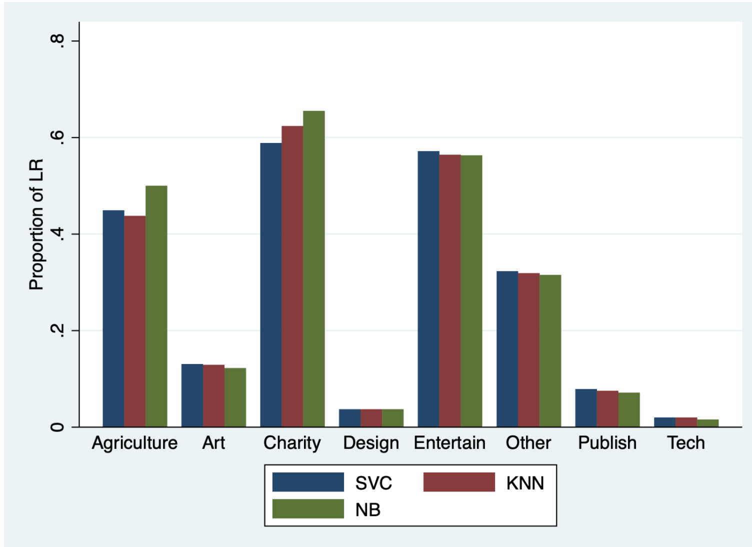

Figure 2: Proportion of projects classified as LR by project category

ods. The coefficient of interest in this regression is the interaction term (*SameProv* $\times$ *LR*), which describes the additional effect of *SameProv* on investment behavior conditional on the project being LR. We expect this coefficient to be positive, because if the information channel indeed plays a role, the local bias behavior would be even more severe for LR projects.

The last three columns in Table 3 report the estimates of Equation (3.2). The results are consistent with our conjecture: the coefficient on *SameProv* $\times$ *LR* is significantly positive. This finding indicates that local bias is more pervasive among LR projects, likely driven by the information channel. We also observe that, for NLR projects, which are less related to the local environment, the coefficient on *SameProv* is significantly positive. This result suggests that funders may have inherent preference for projects from their home region, highlighting the importance of the preference channel.

3.3 Robustness Checks

We provide robustness checks for the regression results in the previous section. We first estimate the logit model in Equation (3.2) when different consideration sets are constructed for the investors.

In columns (1)–(3) of Table 4, we present the results where we assume investors only consider the available projects between the months of their first and last investment. This alternative method builds up a smaller consideration set than the baseline case. Next, in columns (4)–(6), we assume investors consider all projects available between "the month before the first investment" and "the month after the last investment." This set is the largest among all the consideration sets. Comparing the estimates in Table 4 with those in the last three columns of Table 3, we can see the coefficients on *SameProv*, *LR*, and the interaction term are similar across all specifications. This result suggests our main finding on the local biases and the importance of the two channels is robust to different investment options the investors consider.

Table 4: Logit Regression: Different Consideration Sets

	Small consideration set			Large consideration set		
	(1)	(2)	(3)	(4)	(5)	(6)
	SVC	kNN	NB	SVC	kNN	NB
VARIABLES	Invest	Invest	Invest	Invest	Invest	Invest
log(Amount requested)	0.139***	0.140***	0.139***	0.147***	0.147***	0.147***
	(0.00394)	(0.00393)	(0.00394)	(0.00392)	(0.00391)	(0.00392)
# of comments	0.00290***	0.00289***	0.00289***	0.00286***	0.00286***	0.00286***
	(2.42e-05)	(2.43e-05)	(2.42e-05)	(2.32e-05)	(2.32e-05)	(2.32e-05)
# of certificates	-0.0104	-0.0104	-0.00824	-0.0235**	-0.0239**	-0.0215*
	(0.0119)	(0.0119)	(0.0119)	(0.0118)	(0.0118)	(0.0118)
Provide video	-0.101***	-0.101***	-0.101***	-0.0869***	-0.0869***	-0.0869***
	(0.0122)	(0.0122)	(0.0122)	(0.0121)	(0.0121)	(0.0121)
# of projects by the fundraiser	-0.00226***	-0.00233***	-0.00225***	-0.00241***	-0.00248***	-0.00240***
	(0.000329)	(0.000330)	(0.000330)	(0.000335)	(0.000336)	(0.000336)
SameProv	0.865***	0.868***	0.865***	0.885***	0.889***	0.885***
	(0.0146)	(0.0147)	(0.0148)	(0.0146)	(0.0146)	(0.0147)
LocRelated	-0.153***	-0.128***	-0.155***	-0.161***	-0.133***	-0.159***
	(0.0142)	(0.0142)	(0.0142)	(0.0142)	(0.0142)	(0.0142)
SameProv × LocRelated	0.689***	0.659***	0.656***	0.688***	0.657***	0.657***
	(0.0264)	(0.0262)	(0.0261)	(0.0263)	(0.0261)	(0.0260)
Constant	-6.686***	-6.694***	-6.681***	-7.223***	-7.232***	-7.220***
	(0.0393)	(0.0393)	(0.0395)	(0.0390)	(0.0391)	(0.0392)
Control for project category	Y	Y	Y	Y	Y	Y
Observations	10,024,914	10,024,914	10,024,914	15,931,496	15,931,496	15,931,496
AIC	494117.6	494161.6	494161.9	530028.5	530075	530067.9
BIC	494343.6	494387.5	494387.9	530261.9	530308.3	530301.2

Table 5 reports the estimation results for the logit model in Equation (3.2) across project categories separately. We account for potential competition from other projects and seasonal fluctuations by including control variables, such as the number of same-category projects available on that day, and day-of-the-week and month dummies. Comparing the estimated coefficients on *SameProv* across project categories, we find a consistent pattern whereby potential investors prefer projects from their own province. However, the estimated effect varies across categories. For example, the estimated local-preference premium for agriculture and charity projects is larger than for other categories. In addition, a statistically significantly positive coefficient exists on the interaction term $SameProv \times LR$ for five out of eight categories (for the other three categories, the coefficients are insignificant), which highlights the importance of the information channel.

We also notice that the impact of providing a more advanced type of project description (e.g., video) varies across project categories. For example, whereas providing video for the design and entertainment projects has a significantly positive effect on the funding probability, the impact for agricultural and charity categories is zero to negative. A possible reason driving this result is that the quality of the video can rely heavily on the project's category. For example, entrepreneurs proposing a design/art project should presumably be better at showcasing their products using advanced media tools than entrepreneurs in a field such as agriculture. This finding suggests different managerial recommendations may be needed for different categories, and we explore this idea in the counterfactual analysis provided in Section 6.2.

Given the panel structure of the data, we consider other robustness checks, such as adding dummies of the project province and individual fixed effects to the logit regression, and the results (shown in Tables C.2–C.3 in Appendix C) do not change materially.

4 Model

In this section, we develop a structural model of agents' investment behavior in online crowdfunding markets. The model, motivated by the empirical evidence in Section 3, seeks to capture funders' local-bias behavior driven by both the preference and information channels.

Let $i = 1, 2, \cdots N$ be the index of each funder in the data, and j be the index of each project. Consider a project j that was proposed by an entrepreneur on day s_j. Let Z_j denote the observed characteristics of project j, which include the amount requested, the number of projects proposed

Table 5: Logit Regression: Different Categories

VARIABLES	Invest	Invest	Invest	Invest
Project category	**Charity**	**Agri.**	**Enter.**	**Publish**
log(Amount requested)	0.266***	0.0857***	0.112***	0.174***
	(0.0103)	(0.00896)	(0.0137)	(0.0103)
# of projects by the fundraiser	-0.0463***	0.0273***	-0.00341***	0.0203***
	(0.00325)	(0.00541)	(0.000399)	(0.00585)
# of comments	0.00577***	0.00910***	0.00286***	0.00364***
	(0.000173)	(0.000218)	(7.51e-05)	(6.17e-05)
# of certificates	0.0278	0.226***	0.699***	-0.0296
	(0.0288)	(0.0366)	(0.0598)	(0.0273)
Provide video	-0.0297	0.00911	0.124***	0.0543*
	(0.0276)	(0.0311)	(0.0385)	(0.0313)
SameProv	1.457***	1.150***	0.828***	0.429***
	(0.0451)	(0.0429)	(0.0530)	(0.0302)
LocRelated	0.179***	0.0313	-0.540***	-0.249***
	(0.0298)	(0.0280)	(0.0533)	(0.0613)
SameProv × LocRelated	0.0391	0.278***	0.874***	0.393***
	(0.0602)	(0.0661)	(0.0789)	(0.105)
# of same-category projects	-0.0178***	-0.0242***	-0.00221	-0.0624***
	(0.000825)	(0.000969)	(0.00219)	(0.00208)
Control for week & month dummies	Y	Y	Y	Y
Observations	843,375	686,319	306,481	513,488
AIC	73459.24	71782.65	32060.70	71222.07
Project category	**Art**	**Design**	**Tech**	**Other**
log(Amount requested)	-0.201***	0.0338**	0.0689***	0.193***
	(0.0164)	(0.0161)	(0.0111)	(0.00883)
# of projects by the fundraiser	-0.0275***	-0.0453***	-0.00973***	0.0479***
	(0.00946)	(0.00449)	(0.00349)	(0.00570)
# of comments	0.00779***	0.00450***	0.00271***	0.00274***
	(0.000164)	(0.000186)	(6.82e-05)	(7.70e-05)
# of certificates	-0.0286	0.365***	0.485***	0.195***
	(0.0892)	(0.0720)	(0.0440)	(0.0490)
Provide video	0.0285	0.140***	-0.0376	-0.443***
	(0.0560)	(0.0440)	(0.0366)	(0.0370)
SameProv	0.918***	0.621***	0.383***	0.976***
	(0.0475)	(0.0506)	(0.0405)	(0.0440)
LocRelated	-0.0128	-0.709***	-0.0761	-0.267***
	(0.0675)	(0.211)	(0.161)	(0.0364)
SameProv × LocRelated	0.163	0.444	0.840***	0.605***
	(0.135)	(0.398)	(0.277)	(0.0697)
# of same-category projects	-0.0203***	-0.0378***	-0.0234***	-0.0848***
	(0.00383)	(0.00521)	(0.00548)	(0.00432)
Control for week & month dummies	Y	Y	Y	Y
Observations	290,547	130,033	167,564	289,289
AIC	32883.77	26229.41	40289.69	44705.05

Note: For the estimation results in the table, we use the SVC classification method to define location-relatedness for projects in each category. For each category, we also controlled for the number of same-category projects on that day to account for competition from other projects. In addition, we controlled for day-of-the-week and month dummies to capture fluctuations over time.

by the fundraiser, the number of comments and certificates, and whether a video is provided for the project. Based on the machine-learning classification algorithm described previously, we also observe whether evaluating its quality requires location-specific information for each project. We define

$$H_j = \begin{cases} 1 & \text{if project } j \text{ is LR} \\ 0 & \text{if project } j \text{ is NLR} \end{cases} .$$

In addition, we observe the location of all projects and funders. Let $L(i)$ and $L(j)$ represent the location of funder i and project j, respectively. If the funder and project are from the same region, $L(i) = L(j)$.

Following our discussions in Section 3, we assume investors actively consider whether to make contributions to available projects on the platform between "the month before her first investment" and "the month of her last investment." Formally, let $\mathscr{T}_i$ represent the set of all periods when i is an active investor, and let $\underline{t}_i$ and $\bar{t}_i$ be the first and last date i invested in the sample, respectively. Note $\{\underline{t}_i, \underline{t}_i + 1, \cdots, \bar{t}_i\} \subseteq \mathscr{T}_i$.

We account for the fact that investors might pay more attention to recently posted projects, by allowing the probability that an investor considers a project heterogeneous. Let $C_{ij} = 1$ represent the case where i considers whether to invest in project j, and $C_{ij} = 0$ otherwise. If a project is proposed between $\underline{t}_i$ and $\bar{t}_i$, we assume the probability that agent i considers this project is equal to 1. For projects proposed before $\underline{t}_i$, the probability of consideration depends on the time gap between the starting date of project j and $\underline{t}_i$. Moreover, we assume investors do not consider any projects after the date of their last investment observed in the data. Specifically,

$$Pr(C_{ij} = 1 | s_j, \underline{t}_i, \bar{t}_i) = \begin{cases} p(\underline{t}_i - s_j) & \text{if } s_j < \underline{t}_i \text{ and } s_j \in \mathscr{T}_i \\ 1 & \text{if } \underline{t}_i \leq s_j \leq \bar{t}_i \\ 0 & \text{otherwise} \end{cases} . \tag{4.1}$$

In Equation (4.1), the function $p(n)$ is empirically estimated by the ratio between the number of investments made on the nth date after the project was proposed and the number of investments made on the the project launch date. The implicit assumption we impose is that the probability of considering a project equals 1 when it was first proposed. The longer the project remains on the platform, the more the probability that the investors consider it decreases. We estimate $p(\cdot)$ to be a decreasing function, which implies investors pay less attention to projects that have been active for a longer period.

Conditional on considering a project, we now turn to model funders' decisions on whether to contribute. In this paper, we assume the agent has a CARA utility function, namely,

$$u(x;\gamma_i) = -\exp(-\gamma_i x),$$

where γ_i is the risk-aversion parameter, which we allow to be time-invariant and individually heterogeneous. Assume $\gamma_i \sim N(\mu_\gamma, \sigma_\gamma^2)$. The payoff agent i receives for project j is specified in the following equation:

$$x_{ij} = Z_j\beta + \phi H_j + \xi_j + \alpha \mathbf{1}\{L(i) = L(j)\} + \varepsilon_{ij}, \tag{4.2}$$

where ξ_j represents the unobserved quality of project j, and α represents the preference premium received by the funders if the projects are from their own region. ε_{ij} denotes the idiosyncratic shocks funder i receives for project j. We assume ε_{ij}'s are i.i.d. across funders and projects and follow a Type I extreme value distribution.

Given the payoff specification in Equation (4.2), our model contains local biases (i.e., funders are more likely to fund projects from their regions) for two reasons. First, funders may have an inherent preference for projects that come from their home province, which is captured by the parameter α. Second, funders may not observe the quality of the projects directly and instead receive a noisy signal about its quality. The signal can be more informative if the projects are LR and the funders are from the same region. The second channel is reflected in the distribution of ξ_j, which we assume follow Equation (4.3):

$$\xi_j \sim \begin{cases} N(0, \kappa\sigma^2) & \text{if } H_j = 1 \text{ and } L(i) = L(j) \\ N(0, \sigma^2) & \text{otherwise} \end{cases}. \tag{4.3}$$

$\kappa \in (0, 1]$ is a parameter that captures the extent to which the quality of the signals can be improved if the project is LR and the funders and projects are from the same region, that is, $H_j = 1$ and $L(i) = L(j)$.

When making the investment decisions, the funders compare their expected returns from a project with their outside options. Let ν_{ij} denote the outside option funder i draws for project j and assume they are i.i.d. across funders and projects and follow an extreme value distribution. In our model, agent i invests in project j if $E[u(x_{ij})] > u(\nu_{ij})$, where

$$E[u(x_{ij})] = \int -\exp(-\gamma_i x_{ij}) dF_\xi(\xi_j).$$

We derive the choice probabilities by first considering the case where projects are LR and are from the same province as the investor, i.e., $H_j = 1$ and $L(i) = L(j)$:

$$\begin{aligned} E[u(x_{ij})|Z_j, H_j = 1, L(i) = L(j)] &= -\exp(-\gamma_i(Z_j\beta + \phi + \alpha + \varepsilon_{ij})) \int \exp(-\gamma_i \xi_j) dF_\xi(\xi_j) \\ &= -\exp(-\gamma_i(Z_j\beta + \phi + \alpha + \varepsilon_{ij}) + \frac{\kappa\sigma^2\gamma_i^2}{2}). \end{aligned} \tag{4.4}$$

In Equation (4.4), the second equality holds by integrating out the unobserved quality ξ_j, which follows a normal distribution with mean 0 and variance equal to $\kappa\sigma^2$. The probability that agent i invests in project j conditional on the project being LR and from the same region is:

$$\begin{aligned} & Pr(invest_{ij} = 1|Z_j, H_j = 1, L(i) = L(j)) \\ & = Pr\left(\exp(-\gamma_i(Z_j\beta + \phi + \alpha + \varepsilon_{ij}) + \frac{\kappa\sigma^2\gamma^2}{2}) < \exp(-\gamma_i\nu_{ij})\right), \\ & = \frac{\exp(Z_j\beta + \phi + \alpha - \frac{\kappa\sigma^2\gamma_i}{2})}{1 + \exp(Z_j\beta + \phi + \alpha - \frac{\kappa\sigma^2\gamma_i}{2})}. \end{aligned} \tag{4.5}$$

We provide the derivations of the last equality of Equation (4.5) in Appendix B. The expected utility an agent receives from a project has a familiar mean-variance structure. Intuitively, when the variance of the project's unobserved quality is higher, agents are less likely to invest. Given the same level of uncertainty, if the agent is more risk-averse, she is also less likely to invest. Following a similar approach, we can also derive the funders' choice probabilities in other cases, such as when the variance of the unobserved quality is σ^2 instead of $\kappa\sigma^2$.

We close the model section with a few remarks. In this paper, we assume funders evaluate potential investment opportunities separately and the reasons are as follows. First, the projects proposed on the crowdfunding marketplaces are generally in an early development stage and are therefore less likely to generate products the consumers consider close substitutes. Funders on these platforms tend to be more interested in receiving heterogeneous and innovative goods/services promised by the fundraisers instead of monetary payoffs from their investment (as is the case for peer-to-peer lending markets). Second, we find the amounts the funders contribute are often small (the 50th and 90th percentiles of the distribution are 45 CNY and 500 CNY, respectively), lowering the likelihood of the funders' choices being constrained by their budget. In addition, the vast number of available projects on the website (there are on average more than 100 projects available each day) make comparing all opportunities before contributing mentally taxing and time consuming for investors.[16]

Our model abstracts away from strategic interactions among potential funders. Unlike many peer-to-peer lending markets in which the listing is immediately closed if the amount the borrower requests is reached, the amount funded on this platform has no upper limit. As shown in Table 1, on average, the percent of requested funds for each project exceeds 100%, implying investors do not need to compete for a potentially good investment opportunity. However, herding may occur on the online crowdfunding platform, where funders are more likely to contribute to projects that are popular among other investors (see Zhang and Liu (2012)). We account for this potential behavior in our empirical analysis by including control variables, such as the number of comments for the project as a proxy for its general popularity among investors in our sample.

[16]To further examine whether funders trade off between projects, we study investors' behavior change in response to the number of projects available on the platform. Specifically, we regress (1) the amount contributed to each project by each funder and (2) the logarithm of the daily number of investments made by the funder on the logarithm of the number of projects available on the platform when the investments were made. The results are provided in Table C.4 in Appendix C. We find that as the number of choices increases, the amount contributed by the investors to each project is not significantly reduced; the funders' number of investments increases as more projects become available on the platform. These results suggest projects on this platform are not competing against each other and that funders' budget constraints are likely not binding.

5 Identification and Estimation

5.1 Identification

We now discuss the identification of the structural parameters of the model described in Section 4. The main goal of this paper is to separately identify the effect of the information and preference channels in agents' decision-making process, making the key parameters of interest α and κ in agents' utility specification. α measures the inherent preference for projects from the same province, and κ represents the degree to which uncertainty is reduced for agents investing in LR and same-province projects.

The intuition for identifying the preference channel is as follows. The information channel does not play a significant role for NLR projects, such as producing art products or developing software packages. Due to the nature of these projects, the funders have difficulty gleaning additional information about its quality irrespective of whether the funders are from the same or different regions than the project. When the same-province funders do not have informational advantages, the difference in funding probabilities conditional on $L(i) = L(j)$ and $L(i) \neq L(j)$ is induced by their inherent preferences, therefore helping us identify α.[17]

The identification of the information channel follows a difference-in-differences analog. Suppose our control group includes all NLR projects, and the treatment group consists of all LR projects. We quantify the effects of the information channel using the following equation:

$$\begin{aligned}\Delta = &\Big(Pr(Invest_{ij} = 1|H_j = 1, L(i) = L(j)) - Pr(Invest_{ij} = 1|H_j = 1, L(i) \neq L(j))\Big) \\ &- \Big(Pr(Invest_{ij} = 1|H_j = 0, L(i) = L(j)) - Pr(Invest_{ij} = 1|H_j = 0, L(i) \neq L(j))\Big).\end{aligned} \tag{5.1}$$

The first term in Equation (5.1) measures the total degree of local biases for LR projects, which can be decomposed into two parts: the preference and the information channel. The second term in Equation (5.1) captures the effect of the preference channel for NLR projects. If the prefer-

[17] Agents may also derive signaling value of the NLR projects through location (e.g., movie production in Hollywood). However, if the signaling value is shared among potential funders irrespective of their location, the difference in investment probabilities between same-province and different-province funders will still identify the preference channel.

ence premium is the same for the control and treatment groups, the difference of the differences in Equation (5.1) quantifies the remaining variations induced solely by the information frictions, which helps us identify the value of κ.

Following our derivations of the choice probabilities in Section 4, we have the following set of objects identified from individuals' investment behavior:

$$
\begin{aligned}
&(1)\text{ when } H_j = 1 \text{ and } L(i) = L(j), \text{ we identify:} \quad \Delta_1 = Z_j\beta + \phi + \alpha - \frac{\kappa\sigma^2\gamma_i}{2};\\
&(2)\text{ when } H_j = 1 \text{ and } L(i) \neq L(j), \text{ we identify:} \quad \Delta_2 = Z_j\beta + \phi - \frac{\sigma^2\gamma_i}{2};\\
&(3)\text{ when } H_j = 0 \text{ and } L(i) = L(j), \text{ we identify:} \quad \Delta_3 = Z_j\beta + \alpha - \frac{\sigma^2\gamma_i}{2};\\
&(4)\text{ when } H_j = 0 \text{ and } L(i) \neq L(j), \text{ we identify:} \quad \Delta_4 = Z_j\beta - \frac{\sigma^2\gamma_i}{2}.
\end{aligned}
$$

Variations in the project-level characteristics Z_j identify β, $\Delta_2 - \Delta_4$ identifies agents' utility parameter for LR projects, and $\Delta_3 - \Delta_4$ quantifies the preference for projects from the same region. Once β, ϕ, and α are recovered, it is straightforward to identify the disutility generated from uncertainty about the project quality, i.e., $\frac{\kappa\sigma^2\gamma_i}{2}$ and $\frac{\sigma^2\gamma_i}{2}$, from Δ_1 and Δ_2, respectively. Although using these two terms does not allow us to separately identify the variance of the unobserved quality distribution from the risk-aversion parameter γ_i, we can still pin down the value of κ. In other words, we can only identify the relative change in the variance of the unobserved quality when the LR projects are from the same region as the funders. In the estimation, we normalize the value of σ^2 to 1 so that the mean and variance of the risk-aversion parameter for investors are identified.

5.2 Estimation Results

Following our identification strategies discussed in Section 5.1, we estimate the model primitives using maximum likelihood. In the data, we observe $(Z_j, H_j, L(j))$ for each project j, where Z_j represents a vector of project characteristics (including project category dummies), H_j is a dummy variable that equals 1 if the project is LR, and $L(j)$ denotes the location (at the provincial level) of project j.[18] For each investor i, we also observe the geographic-location information, which is

[18] In the estimation, we use the SVC classification method for LR projects.

denoted by $L(i)$. For each project-investor pair, we observe data on whether a transaction occurs. If an agent i invests in project j, $Invest_{ij} = 1$, and 0 otherwise. In summary, the data we have contain $(Z_j, H_j, L(j))$ for all j, $L(i)$ for all i, and $Invest_{ij}$ for all i and j. The vector of parameters to be estimated is denoted by $\theta = (\beta, \phi, \alpha, \mu_\gamma, \sigma_\gamma^2, \kappa)$.

To construct the likelihood function, we first derive the probability of observing an investor i with risk-aversion parameter γ_i who contributes to a project j by the following equation:

$$
\begin{aligned}
q(Z_j, s_j, \underline{t}_i, \bar{t}_i, H_j, L(i), L(j); \gamma_i) =& Pr(C_{ij} = 1 | s_j, \underline{t}_i, \bar{t}_i) \\
& \times Pr(invest_{ij} = 1 | Z_j, H_j, L(i), L(j); \gamma_i).
\end{aligned}
$$

The log-likelihood function can, therefore, be written as follows:

$$
\begin{aligned}
LL(\theta) = \sum_{i=1}^{N} \log \Bigg(\int \prod_j \Big[& q(Z_j, s_j, \underline{t}_i, \bar{t}_i, H_j, L(i), L(j); \gamma_i)^{\mathbf{1}\{invest_{ij}=1\}} \times \\
& (1 - q(Z_j, s_j, \underline{t}_i, \bar{t}_i, H_j, L(i), L(j); \gamma_i))^{\mathbf{1}\{invest_{ij}=0\}} \Big] dF(\gamma_i; \mu_\gamma, \sigma_\gamma) \Bigg).
\end{aligned} \tag{5.2}
$$

In this paper, we provide two sets of estimation results. We first estimate the model assuming the risk-aversion parameters are the same across all individual investors. The estimated structural parameters and their standard errors are shown in the third column of Table 6. From the estimation results, we can see agents prefer to invest in projects with more supporter comments and higher amounts requested by the fundraiser. However, funders are less interested in projects whose fundraisers have more listings proposed in the past or those associated with videos. We also find agents are less willing to invest in LR projects, which is consistent with the findings from the reduced-form analysis. This finding can also be attributed to funders in online crowdfunding markets often being younger than the general investor population, and potentially being more interested in IT and art projects.

The estimate for agents' preference premium on projects from their own region is significantly positive, demonstrating that all else equal, agents are more likely to invest in projects from their province. This preference premium is approximately five times as large as the disutility from investing in LR projects. Our estimation results also suggest that if a project is LR and from the same province as the funders, the variance of the unobserved project quality equals 88.25% of the variance in all other cases. This finding highlights the important role the information channel plays in explaining local biases. When the project is related to the local environment, funders from the same province receive less noisy signals about its true quality, thus making them more willing to contribute.

In the last column of Table 6, we present the estimation results when the individual's risk-aversion parameter is heterogeneous and assumed to be drawn from a normal distribution $N(\mu_\gamma, \sigma_\gamma^2)$. This model specification takes better advantage of the data's panel structure. We find that allowing for heterogeneous risk preferences across potential investors does not significantly change the estimates of parameters in agents' payoff function.[19] Allowing the risk-aversion parameter to be heterogeneous results in the disutility of investing in location-related projects being slightly larger and the preference premium on same-province projects being smaller. The estimated variance of the distribution for γ is quite small relative to its mean (which is close to the risk-aversion parameter in the first specification), implying the risk attitudes for investors in our dataset are not very dispersed.

We also estimate coefficients separately for different product categories[20] and provide the results for agricultural, charity, technology, and design projects in Table 7. Estimation results for all other categories are provided in Table C.5 in the appendix. As reported by the estimates in Table 7, we find consistent preference premiums on projects from the investors' own province, although its magnitude varies across different project categories. For example, the preference premium for agricultural and charity projects is much larger than that for tech and design projects. The extent to which the variance of the unobserved product quality can be reduced when a funder evaluates an LR and same-province project also varies across categories. Surprisingly, we find the variance of the unobserved project quality can be reduced by 33.23% for LR tech projects if the investors

[19]When allowing the risk-aversion parameter to be heterogeneous, we find a positive effect of having more certificates and providing video on increasing agents' utility. The signs and magnitudes of other estimated coefficients are similar across the two specifications.

[20]To reduce computational burden, when estimating parameters for different categories separately, we assume the risk-aversion parameter is homogeneous across agents.

Table 6: Estimation Results

Variables	Notations	Const. RA	Heter. RA
# of certificates	β_1	-0.1279^{***}	0.1367^{***}
		(0.0121)	(0.0001)
# of projects by the fundraiser	β_2	-0.0060^{***}	-0.0104^{***}
		(0.0004)	(0.0002)
# of comments	β_3	0.0044^{***}	0.0050^{***}
		(0.0000)	(0.0000)
log(Amount requested)	β_4	0.1329^{***}	0.0866^{***}
		(0.0036)	(0.0001)
Provide video	β_5	-0.0365^{***}	0.0976^{***}
		(0.0121)	(0.0002)
Category Charity	β_6	-0.4924^{***}	-0.3949^{***}
		(0.0194)	(0.0002)
Category Agriculture	β_7	-0.3372^{***}	-0.0512^{***}
		(0.0198)	(0.0001)
Category Publishing	β_8	-0.1739^{***}	-0.0361^{***}
		(0.0198)	(0.0003)
Category Entertainment	β_9	-0.8304^{***}	-0.9840^{***}
		(0.0242)	(0.0001)
Category Technology	β_{10}	0.0663^{***}	0.1508^{***}
		(0.0224)	(0.0000)
Category Art	β_{11}	-0.6536^{***}	-0.8345^{***}
		(0.0238)	(0.0001)
Category Design	β_{12}	-0.0911^{***}	-0.0663^{***}
		(0.0245)	(0.0001)
LR project	ϕ	-0.1608^{***}	-0.3970^{***}
		(0.0149)	(0.0002)
Preference premium on projects from same prov	α	0.8353^{***}	0.6594^{***}
		(0.0146)	(0.0003)
Ratio between variances	κ	0.8852^{***}	0.8454^{***}
		(0.0044)	(0.0002)
Risk aversion	γ	12.4235^{***}	N/A
		(0.0732)	N/A
Risk-aversion mean	μ_γ	N/A	12.0515^{***}
		N/A	(0.0002)
Risk-aversion std	σ_γ	N/A	0.7512^{***}
		N/A	(0.0003)
Log-Likelihood		-265857	-263030

and projects are from the same area. For LR charity projects, however, the variance remains nearly unchanged despite the investors and projects being from the same region. We also find interesting heterogeneous effects of providing soft information on attracting investors. For example, providing video content helps attract more investors in the tech and design categories but has the opposite effect for agricultural and charity projects.

Table 7: Estimation Results for Agricultural, Charity, Tech, and Design Categories

Variables	Notations	Agriculture	Charity	Tech	Design
# of certificates	β_1	-0.3741^{***}	-0.1815^{***}	0.1268^{***}	0.2513^{***}
		(0.0253)	(0.0265)	(0.0338)	(0.0567)
# of projects by the fundraiser	β_2	0.0232^{***}	-0.0646^{***}	-0.0303^{***}	-0.0548^{***}
		(0.0057)	(0.0058)	(0.0040)	(0.0050)
# of comments	β_3	0.0093^{***}	0.0062^{***}	0.0039^{***}	0.0059^{***}
		(0.0003)	(0.0002)	(0.0001)	(0.0002)
log(Amount requested)	β_4	0.0755^{***}	0.2999^{***}	0.0529^{***}	0.0261
		(0.0089)	(0.0096)	(0.0102)	(0.0161)
Provide video	β_5	-0.0137	-0.0709^{***}	0.0286	0.1599^{***}
		(0.0303)	(0.0275)	(0.0351)	(0.0428)
Location-related project	ϕ	0.0299	0.1404^{***}	-0.0967	-0.5501^{**}
		(0.0279)	(0.0294)	(0.1617)	(0.2148)
Preference premium on projects form same prov	α	1.0945^{***}	1.4921^{***}	0.3692^{***}	0.5792^{***}
		(0.0425)	(0.0437)	(0.0413)	(0.0507)
Ratio between variances	κ	0.9342^{***}	0.9986^{***}	0.6677^{***}	0.7601^{***}
		(0.0130)	(0.0082)	(0.0680)	(0.0949)
Risk aversion	γ	10.2354^{***}	14.5080^{***}	7.8804^{***}	7.6200^{***}
		(0.1697)	(0.1868)	(0.1897)	(0.2880)
Log-likelihood		-38947.91	-40637.73	-23403.10	-14892.28

We then simulate funders' decisions using the structural estimates and compare the simulated behavior with the actual investment behavior observed in our dataset. By comparing the average number of contributions per project and per funder using actual and simulated data in panels (A) and (B) in Table 8, we find our estimates match the data patterns reasonably well.

6 Counterfactual Analysis

Given our structural estimates, we conduct counterfactual experiments to (1) quantify the effects of the information and the preference channels on agents' local-bias behavior separately and (2) evaluate different policies that may help facilitate transactions on the platform and provide managerial recommendations for crowdfunding marketplaces.

6.1 Decomposition: Information versus Preference

Two counterfactual scenarios are considered to disentangle the effects of the information and the preference channels on local biases. First, we shut down the information channel by setting $\kappa = 1$,

Table 8: Model Fit

(A) Average investors per project

	Actual Data	Simulated Data
All proj	8.3474	8.4802
All proj: same prov investors	1.9409	1.9536
All proj: diff prov investors	6.4065	6.5267
Location-related proj	6.2130	5.8542
Location-related proj: same prov investors	1.6148	1.5465
Location-related proj: diff prov investors	4.5982	4.3077
Non-location-related proj	9.5802	9.9970
Non-location-related proj: same prov investors	2.1293	2.1887
Non-location-related proj: diff prov investors	7.4509	7.8084

(B) Average investments per investor

	Actual Data	Simulated Data
All projects	1.9791	2.0106
Location-related projects	0.5393	0.5082
Non-location-related projects	1.4398	1.5024
Same province projects	0.4602	0.4632
Different province projects	1.5189	1.5474

resulting in agents losing their informational advantages for LR projects from their own regions. The decrease in the number of contributions compared with the levels observed in the data helps quantify the effect of the information channel on rationalizing local bias. Second, we shut down both the information and preference channel by setting $\kappa = 1$ and $\alpha = 0$. Comparing the market outcome when the two channels are shut down with the one observed in the data, we quantify the total effects of the information and the preference channels on producing local bias.

In Table 9, we summarize the average number of funders per project under three scenarios: (1) actual data, (2) when the information channel is shut down, and (3) when both channels are shut down. Overall, when shutting down either channel, we find the number of contributions per project drops, reflecting that same-province funders are less willing to invest in the project. For location-related projects, we find that when the information channel is shut down, the number of same-province funders decreases by 0.78. Simultaneously shutting down the preference channel leads to an additional reduction (0.44) in the number of same-province funders per project. In summary, for location-related projects, the information channel accounts for 63.93% of the decrease in the number of same-province funders, and the preference channel accounts for 36.07%.[21]

From the bottom panel of Table 9, we can see shutting down the information channel has no significant impact on the number of funders for NLR projects. This result occurs because funders cannot learn additional information about its quality despite coming from the same area. For NLR projects, when the preference channel is shut down, the number of same-province funders decreases from 2.19 to 1.04, suggesting the preference channel plays an important role in motivating agents to invest in projects from their own regions.

Table 9: Decompose Two Channels: Average Number of Funders per Project

	Data	Shut down info	Shut down both
	All projects		
Number of funders	8.48	8.17	7.31
Number of funders from same prov	1.95	1.67	0.78
Number of funders from diff prov	6.53	6.50	6.53
	LR projects		
Number of funders	5.85	5.12	4.69
Number of funders from same prov	1.55	0.77	0.33
Number of funders from diff prov	4.31	4.36	4.36
	NLR projects		
Number of funders	10.00	9.93	8.83
Number of funders from same prov	2.19	2.19	1.04
Number of funders from diff prov	7.81	7.74	7.79

We also compute the amount of money that would have been raised for each project if, hypothetically, the information channel were shut down or both channels were shut down. We first estimate a regression of the amount contributed by funders on each project (in logarithm) on a series of projects' characteristics and a dummy variable describing whether the project and individual come from the same province.[22] The results of this regression are provided in Table C.6 in Appendix C. These regression coefficients help us obtain the predicted values for the amount

[21]The counterfactual results shown in Table 9 are based on a group of investors who voluntarily provided their location information. One caveat is that these investors are possibly more attached to their hometown, which is associated with them having a higher local-preference premium than the general investor population. Our analysis using this group of investors can overestimate the importance of the preference channel as a driver of local-bias behavior on the crowdfunding platform.

[22]We also include dummy variables for each investor's province to control for heterogeneity across provinces. In addition, day-of-the-week and month dummies for the investment date are included to control for seasonal fluctuations of investment behavior.

contributed for each counterfactual investment record of each individual. Aggregating across all funders, we obtain the amount of money raised for each project under the counterfactual scenarios and compare it with the amount requested by the entrepreneurs. If the aggregate amount raised exceeds the amount requested, the project is considered to be successfully funded.

One data limitation that we encountered when simulating the total amount of money raised for each project was that only a subset of funders had available province information. The group of funders we use for estimation is a subsample of all funders, making data on investment behavior for each agent in the estimation sample insufficient to compute the total investment amount unless "extrapolation" is implemented. We address this issue by computing the ratio between the total amount raised for each project using the full sample and the sample of funders with province information and then taking an average across all projects. In the following counterfactual experiments, we divide the requested amount for each project by this average ratio and obtain the *scaled* requested amount for the set of investors with province information. Comparing the predicted amount of money contributed by the agents who provided province information and the scaled requested amount helps us ascertain whether a project is successfully funded in counterfactual scenarios.

Table 10 presents the average amount raised from investors with province information and the project funding rate under different counterfactual scenarios. For each type of project, we also separately report the average amount raised from same-province and different-province funders. In the first column of Table 10, we report the market outcome using the simulated amount invested by agents with location information when both the information and the preference channels are present. Columns (2)-(3) report the market outcome when the information channel is shut down and when both channels are shut down, respectively.

From Table 10, we can see shutting down the information channel reduces the project funding rate from 87.57% to 85.85% for LR projects. A shutdown of the preference channel leads to an additional decrease in the funding probability (although the magnitude is small). Another observation that can be gleaned from Table 10 is that the information channel has a significant effect on the amount of money raised for projects from agents from the same province. When the information channel is shut down, the amount contributed by same-province funders is decreased by 15.24%. For LR projects, this effect is even more pronounced as one would expect – the amount contributed by same-province investors is decreased by 50.80% after the information channel is shut down.

Table 10: Decompose Two Channels: Amount Raised and Project Successful Rate

	Simulated Data	Shut down info	Shut down both
	All projects		
Amt Raised	4629.05	4321.49	3472.69
Amt Raised from Same Prov	1973.72	1672.83	819.09
Amt Raised from Diff Prov	2655.33	2648.66	2653.60
Proj Successful Rate (%)	92.54	92.13	90.76
	LR projects		
Amt Raised	3110.47	2342.80	1847.31
Amt Raised from Same Prov	1605.96	790.08	341.16
Amt Raised from Diff Prov	1504.51	1552.72	1506.14
Proj Successful Rate (%)	87.57	85.85	85.12
	NLR projects		
Amt Raised	5495.08	5449.91	4399.63
Amt Raised from Same Prov	2183.45	2176.25	1091.64
Amt Raised from Diff Prov	3311.63	3273.67	3307.99
Proj Successful Rate (%)	95.38	95.72	93.97

Note: Average amount raised from investors with province information is reported in this table, and the unit of amount is CNY. The extrapolation ratio is computed based on the total amount raised for each project using the full sample and the sample of investors with province information.

Next, we evaluate the impact of the information and the preference channels on the website's revenue. Based on the summary statistics reported in Table 1, the average amount raised for a project is 29,409 CNY. During the sample period, 4876 projects are available online, of which 40% are LR. The platform charges a 3% commission fee of the raised amount for successfully funded projects. Because the amount raised and the funding rate for projects drops upon shutting down the two channels, the amount of revenues collected by the website also decreases. In Table 11, we separately estimate the changes in revenue when the preference and the information channels are shut down for LR and NLR projects. We estimate that shutting down the information channel for LR projects leads to a more than 264,000 reduction in the platform's revenue, which is twice as large as the effects of when the preference channel is shut down. For NLR projects, shutting down the preference channel has a significant negative impact on the revenue collected by the platform, because the amount raised drops precipitously. Overall, we estimate that shutting down the preference channel decreases the revenue of the platform by around 603,000 CNY.

Table 11: Decompose Two Channels: Changes in Revenue (CNY)

	Shut Down Info	Shut Down Preference
LR Projects	-264,746.87	-163,060.25
NLR Projects	-13,684.25	-603,206.82

Note: We approximate the revenue collected by the platform by the number of projects funded $\times$ amount raised per project $\times$ 3%. Based on our estimates in Table 10, for LR projects, the revenue collected by the platform from investors with province information when both channels are allowed equals $4876 \times 40\% \times 87.57\% \times 3110.47(CNY) \times 3\% = 159,377.24(CNY)$, where, 4876 is total number of projects, of which 40% are LR. The project funding rate is 87.57%, and the average amount raised per project is 3110.47. To approximate the total amount of revenue received by the platform, we multiply the revenue from investors with location information by the extrapolation ratio $6.35(\approx 29409/4629.05)$. The platform's revenue when the information channel is shut down, and when both channels are shut down can be computed using similar methods for both LR and NLR projects.

6.2 Evaluating Platform-Design Policies

Motivated by the empirical evidence on the existence of strong local biases among funders and the information channel playing an important role in driving this bias, we now focus on evaluating the impact of platform-design policies that can be used to reduce local biases and improve the informativeness of the project descriptions. Specifically, we consider three counterfactual scenarios: (1) the location information of the project is removed from the website, (2) the number of comments the project receives is hidden, and (3) the platform does not allow the fundraiser to post videos. These counterfactual experiments are conducted for different project categories separately to determine whether the managerial implications of these interventions could vary across project types. Table 12 considers three counterfactual scenarios for agricultural and design-related projects and summarizes the average investors per project, amount raised, and the project success rate. Results for other project categories are provided in Tables C.7–C.9 of Appendix C.

Removing location information of the project has different effects on investors depending on whether they are from the same province. From Table 12 (see column (2) for Agriculture and

Table 12: Evaluating Platform-Design Policies: Agriculture and Design

	Average investors per project							
	Agriculture				**Design**			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Data	Remove location	Remove comments	Remove video	Data	Remove location	Remove comments	Remove video
All proj	8.19	9.69	5.93	8.21	12.04	14.87	8.74	11.19
same prov investors	1.50	0.61	1.05	1.43	2.30	1.58	1.60	2.12
diff prov investors	6.69	9.08	4.88	6.79	9.74	13.28	7.13	9.07
LR proj	8.12	9.37	5.84	8.09	4.00	4.56	3.44	4.44
same prov investors	1.34	0.52	1.02	1.34	2.11	0.33	0.89	1.33
diff prov investors	6.78	8.84	4.82	6.76	1.89	4.22	2.56	3.11
NLR proj	8.24	9.95	6.01	8.31	12.35	15.27	8.94	11.45
same prov investors	1.63	0.68	1.08	1.50	2.30	1.63	1.63	2.15
diff prov investors	6.62	9.27	4.93	6.81	10.05	13.63	7.31	9.30

	Amount raised and project successful rate							
	Agriculture				**Design**			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Data	Remove location	Remove comments	Remove video	Data	Remove location	Remove comments	Remove video
All proj	3172.18	3178.56	1993.16	3147.35	4657.80	5196.76	3081.59	4326.33
same prov investors	1110.96	481.82	657.20	1060.58	1491.56	1069.10	1032.13	1447.21
diff prov investors	2061.22	2696.74	1335.97	2086.77	3166.24	4127.66	2049.46	2879.13
Proj successful rate (%)	97.17	97.03	96.09	97.03	99.56	99.12	97.80	99.12
LR proj	3499.06	3364.44	2092.89	3440.15	1984.32	1206.34	1347.06	1621.79
same prov investors	1132.23	456.57	726.07	1121.53	1515.04	293.21	777.58	934.31
diff prov investors	2366.83	2907.87	1366.82	2318.62	469.28	913.13	569.48	687.48
Proj successful rate (%)	98.76	98.14	97.83	97.52	100.00	87.50	87.50	100.00
NLR proj	2920.97	3035.71	1916.52	2922.33	4755.46	5342.53	3144.95	4425.13
same prov investors	1094.62	501.22	604.27	1013.73	1490.70	1097.44	1041.43	1465.94
diff prov investors	1826.36	2534.49	1312.26	1908.6	3264.76	4245.09	2103.52	2959.19
Proj successful rate (%)	95.94	96.18	94.75	96.66	99.54	99.54	98.17	99.09

column (6) for Design), we see that whereas removing location information disincentivizes same-province investors to contribute, the number of different-province investors per project and the average amount these investors contribute increase. When the location information is removed from the website, the same-province investors lose their informational advantages, and their preference premium towards local projects also disappears. On the other hand, this intervention results in uncertainty about where the projects are from for different-province investors. These investors now form expectations on whether the project comes from the same area as them, using the empirical distribution of projects from each province. This behavior can lead to a higher chance of these investors believing the projects are from their own regions, incentivizing them to contribute more to the projects. This platform-design policy has practical managerial implications for the crowdfunding marketplaces. In light of our finding that investors on this platform exhibit strong local biases toward the project from their own region, removing the location information of the projects could be welfare improving. This helps to mobilize a larger group of investors who are not from the same provinces as the project to contribute, which can eventually increase the amount raised for the project and its success rate.

In the second experiment, we compare the existing market outcome with a market outcome in which the comments projects receive are hidden from investors. The results shown in Table 12 (see column (3) for Agriculture and column (7) for Design) suggest fewer investors exist per project, less money is raised, and the project success rate is lower when the number of comments is removed. Intuitively, investors could use the number of comments received by the project as a proxy for the project's popularity among other investors and form their beliefs about its quality on this basis. Our finding that removing comments from the website disincentivizes funders to make contributions highlights the importance of social-learning behavior in online crowdfunding markets.

Our last counterfactual experiment studies the impact of removing the video from project descriptions across different types of projects. Intuitively, providing more detailed and accurate information about the project through video presentations should help investors better evaluate the quality of the project and help accelerate its fundraising. However, the quality of the video produced can vary widely. A poorly created video could have the opposite effect by creating bad impressions for the investors and discouraging them from contributing to the project. This possibility is consistent with columns (4) and (8) in Table 12, where we find that for agriculture-related projects, preventing fundraisers from posting a video for their projects increases the number of investors per project, whereas the opposite is true for design-related projects. This finding highlights the importance of designing flexible and targeted marketing strategies for different types of projects in online crowdfunding markets. For example, the widespread encouragement of entrepreneurs to showcase their products via video presentations may not be cost-effective, as was shown in the case for agricultural products. Instead, platforms or regulators could consider conducting interviews, preparing onsite visits, or providing more tools to facilitate communications between funders and entrepreneurs (e.g., organizing online Q&A sessions, developing instant-messaging tools) to help entrepreneurs better advertise their projects.[23]

7 Conclusions

This paper investigates the existence of local biases and the channels that drive them in online crowdfunding marketplaces. We present empirical evidence consistent with strong local biases among funders by exploiting a unique dataset from a large crowdfunding platform in China. We then develop a structural model of agents' investment behavior and quantify the importance of information asymmetry and preference toward local projects on inducing these local biases. We find the former accounts for two-thirds of the total effect.

Our counterfactual experiments have strong policy implications for the development of micro-entrepreneurship in markets where credit access is difficult or banking infrastructure is weak. With strong local bias, we find that not revealing location information of the projects could potentially be welfare improving. Providing more detailed and accurate information about the local projects through various marketing strategies may mitigate the informational disadvantages faced by non-local funders and stimulate investment. However, flexible and category-specific marketing strategies still need to be considered.

Our study quantifies the economic impact of market frictions in crowdfunding marketplaces. In our model, we focus on the funders' participating decisions while abstracting away from the strategic interactions among fundraisers and the specific amounts the funders might decide to invest. Though our empirical results provided the basis of this modeling choice and because of the

[23] Bao and Ni (2020) show such communication, if not carefully designed, in the form of cheap talk, does not always improve the crowdfunding outcome, though it can help increase the funding prospects. They propose a pricing mechanism to integrate costly signaling with cheap talk to improve the market outcome.

institutional setting of small contributions made by most funders, a richer model that accounts for substitution across different projects might produce additional empirical insights. Although we find on average that funders only invest in two projects within the four years of our data sample, potential investment dynamics might exacerbate the market-friction issues. A dynamic model in which funders could infer project quality through the participation of local versus non-local investors would be worth considering. Our analysis focuses on funders' investment decisions on a particular crowdfunding platform. With potential data access to multiple competing platforms, studying funders' cross-platform investment strategies would be possible. We leave these different considerations for future research.

A Text Mining from Project Descriptions and Tags

In this section, we provide details on employing text mining techniques ((Gentzkow et al., 2019)) to classify projects into groups (LR vs. NLR) based on their contents. We begin by collecting keywords from descriptions and tags of each project. This results in a vector of 1000 keywords that bear relevance to the LR and NLR classifications. In order to reduce the dimensionality of this extensive keyword space, we then interview domain experts and develop a refined set of 32 keywords that capture the essential elements pertaining to the location-relatedness of projects. By encoding each project with the keyword dictionary, we proceed to train classifiers to determine whether a project is LR or not with the aid of pre-labeled projects by human coders. We choose the classification algorithm that achieves the highest prediction accuracy during the training-validation process. We begin by outlining the steps for data processing and follow by describing the implementation of different methods.

A.1 Data Processing

First, we collect all keywords from descriptions and tags of each project in the original data, splitting the keywords by spaces, removing duplicated words, punctuation marks, and numerical numbers. We transform the unique words in the texts into a vector and calculate the term frequency-inverse document frequency (tf-idf) for each word. We then build a dictionary containing 32 keywords that are most essential to the LR and NLR classification. [24] By mapping the keywords we obtained from descriptions and tags of each project to those keywords defining LR/NLR projects, we can determine whether each project is LR or not. We shuffle and split the data into two parts, one with 3,077 projects and the other with 3,018 projects. The first part is used for building the model, so each project in this group is classified into LR or NLR projects through human labeling. We use the model to classify projects in the second group.[25]

[24] The keywords include name of the locations (if available), nature, food, technology, animal/plants, sports, and so on. The full list is available from the authors upon request.

[25] We drop projects if they have no keywords assigned into the pre-specified dictionary. We end up with 5488 projects (2,777 for training and 2,711 for testing).

A.2 Model Building and Parameter Selection

Following data preparation, we build the model using three methods: support vector classifier (SVC), k-nearest neighbors (kNN), and naive Bayes (NB). The details for model implementation and parameter selection are described as follows.

Support Vector Classifier An SVC finds hyperplanes in a basis expansion of numerical array (representing raw text) that partitions the observations into sets with an equal response. In a two-category case, the SVC model represents the training documents as points in a space, such that the documents in separate categories can be divided by a clear gap. A new document from outside the training set gets mapped into the same space and classified into a specific category based on which side of the gap it falls on. After initial exploration, we settle with the linear kernel for the support vector machine classifier. Our choice of the hyper-parameters - loss function, regularization methods, and the value of regularization parameters - are as follows. We use an exhaustive grid search to find an optimized parameter solution on the training data set. After trials on the test set, we choose to use the hinge loss function with L2 regularization and set the regularization constant to 0.0001. For the hyperparameter C and γ in the model, an exhaustive grid search based on n-fold ($n = 3$) cross-validation is conducted to find the parameters giving the highest accuracy. After fitting the training data (70% of data part 1), the best parameters are $C = 100.0$, and $\gamma = 10^{-6}$ with a score of 0.703. Using the model with the optimal parameters, we make predictions on the testing data (30% of data part 1) and compare the results with the true values.

k-Nearest Neighbors The kNN approach classifies test objects based on their closest k training samples in the feature space. Each new document in the text data is represented as a row vector. The distances between this vector and all vectors in the training data are calculated first. The new document is assigned into the category determined by the majority of the labels of its k-closest neighbors from the training set. The measure of closeness/similarity between each pair of vectors can be calculated by the standard cosine similarity. The parameter we tune is the number of nearest neighbors we consider. Similar to the previous method, we conduct a grid search based on n-fold ($n = 3$) cross-validation. The range for the number of neighbors is in linear space from 10 to 100, with 30 data points. After fitting the training data (70% of data part 1), the optimal choice for the number of neighbors equals 28 with a score of 0.703. Using the model with the optimal parameters, we make predictions on the testing data (30% of data part 1) and compare the results with the true values.

Naive Bayes The NB method uses word-based probabilities and assumes the appearance of each feature (word) in a document is conditionally independent. In this paper, we use Bernoulli naive Bayes. For the hyperparameter α in the model, an exhaustive grid search based on n-fold ($n = 3$) cross-validation is conducted to find the parameter that produces the highest accuracy. After fitting the training data (70% of data part 1), the optimal choice for the α is 1 with a score of 0.65. Again, we make predictions on the testing data (30% of data part 1) and compare the results with the true values.

B Derivation of Choice Probabilities

In this section, we provide details on the derivation of agents' choice probabilities shown in Equation (4.5). Recall that

$$\begin{aligned}
&Pr(invest_{ij} = 1 | Z_j, H_j = 1, L(i) = L(j)) \\
&= Pr\left(\exp(-\gamma_i(Z_j\beta + \phi + \alpha + \varepsilon_{ij}) + \frac{\kappa\sigma^2\gamma^2}{2}) < \exp(-\gamma_i \nu_{ij})\right),
\end{aligned} \tag{B.1}$$

where both ε_{ij} and ν_{ij} follow an extreme value distribution. Notice that if a random variable V follows an extreme value distribution, and $Y = \exp(a - bV)$, then $Y \sim Weibull(1/b, \exp(a))$. Let

$$Y_1 = \exp(-\gamma_i(Z_j\beta + \phi + \alpha + \varepsilon_{ij}) + \frac{\kappa\sigma^2\gamma_i^2}{2}) \sim Weibull(1/\gamma_i, \exp(-\gamma_i(Z_j\beta + \phi + \alpha) + \frac{\kappa\sigma^2\gamma_i^2}{2})),$$

$$Y_2 = \exp(-\gamma_i \nu_{ij}) \sim Weibull(1/\gamma_i, 1).$$

The probability that funder i chooses to invest in project j therefore equals:

$$\begin{aligned}
&Pr(invest_{ij} = 1 | H_j = 1, L(i) = L(j)) = Pr(Y_1 < Y_2) = \int_0^\infty F_{Y_1}(y_2) dF_{Y_2}(y_2) \\
=&\int_0^\infty \left(1 - \exp(-(y_2 \exp(-c))^{1/\gamma_i})\right) f_{Y_2}(y_2) dy_2 \\
=&1 - \int_0^\infty \frac{1}{\gamma_i} y_2^{1/\gamma_i - 1} \exp(-(y_2^{1/\gamma_i}(\exp(-c/\gamma_i) + 1))) dy_2 \\
=&1 - \int_0^\infty \exp(-(y_2^{1/\gamma_i}(\exp(-c/\gamma_i) + 1))) dy_2^{1/\gamma_i} \\
=&1 - (\exp(-c/\gamma_i) + 1)^{-1} \int_0^\infty \exp(-(y_2^{1/\gamma_i}(\exp(-c/\gamma_i) + 1))) dy_2^{1/\gamma_i}(\exp(-c/\gamma_i) + 1) \\
=&1 - (\exp(-c/\gamma_i) + 1)^{-1}(-\exp(-y)|_0^\infty) = \frac{\exp(-c/\gamma_i)}{1 + \exp(-c/\gamma_i)} \\
=&\frac{\exp(Z_j\beta + \phi + \alpha - \frac{\kappa\sigma^2\gamma_i}{2})}{1 + \exp(Z_j\beta + \phi + \alpha - \frac{\kappa\sigma^2\gamma_i}{2})},
\end{aligned} \tag{B.2}$$

where

$$\begin{aligned}
&F_{Y_1}(y) = 1 - \exp(-(y\exp(-c))^{\frac{1}{\gamma_i}}), \quad \text{with } c = -\gamma_i(Z_j\beta + \phi + \alpha) + \frac{\kappa\sigma^2\gamma_i^2}{2}; \\
&F_{Y_2}(y) = 1 - \exp(-y^{\frac{1}{\gamma_i}})
\end{aligned}$$

are the cumulative density functions for Y_1 and Y_2, respectively.

C Additional Tables

Table C.1: Number of Projects from Each Province

Province	# of projects	Province	# of projects
Anhui	56	Liaoning	44
Beijing	1,291	Neimenggu	31
Chongqing	97	Ningxia	21
Fujian	142	Others	5
Guangdong	527	Qinghai	26
Gansu	161	Sichuan	208
Guangxi	78	Shandong	199
Guizhou	73	Shanghai	294
HeilongJiang	39	Shaanxi	110
Hainan	18	Shanxi	42
Hebei	86	Tianjin	62
Henan	144	Xianggang	14
Hubei	156	Xinjiang	35
Hunan	110	Xizang	24
Jilin	21	Yunan	147
Jiangsu	252	Zhejiang	196
Jiangxi	49		
Total	4,758	HHI	10.08%

Note: The dataset contains 4,876 listings in total, among which 4,758 projects have province-level location information available.

Table C.2: Logit Regression: Add Project Province Fixed Effects

VARIABLES	(1) Invest SVC	(2) Invest kNN	(3) Invest NB
log(Amount requested)	0.153***	0.153***	0.153***
	(0.00401)	(0.00400)	(0.00401)
# of projects by the fundraiser	-0.00239***	-0.002348**	-0.00240***
	(0.000330)	(0.000331)	(0.000330)
# of comments	0.00282***	0.00282***	0.00282***
	(2.37e-05)	(2.38e-05)	(2.37e-05)
# of certificates	0.0302**	0.0295**	0.0320***
	(0.0121)	(0.0122)	(0.0121)
Provide video	-0.0960***	-0.0960***	-0.0964***
	(0.0122)	(0.0122)	(0.0122)
SameProv	0.920***	0.923***	0.921***
	(0.0156)	(0.0157)	(0.0157)
LocRelated	-0.191***	-0.156***	-0.182***
	(0.0147)	(0.0146)	(0.0146)
SameProv $\times$ LocRelated	0.714***	0.680***	0.678***
	(0.0266)	(0.0265)	(0.0263)
Constant	-6.859***	-6.867***	-6.851***
	(0.0579)	(0.0580)	(0.0580)
Control for Proj Prov Dummies	Y	Y	Y
Observations	12,848,906	12,848,906	12,848,906
AIC	512582.4	512641.4	512634.5
BIC	513272.1	513331.1	513324.2

Table C.3: Logit Regression: Add Individual Fixed Effects

VARIABLES	(1) Invest	(2) Invest	(3) Invest
	SVC	kNN	NB
log(Amount requested)	0.145***	0.145***	0.144***
	(0.00361)	(0.00361)	(0.00361)
# of projects by the fundraiser	-0.00431***	-0.00435***	-0.00429***
	(0.000319)	(0.000319)	(0.000319)
# of comments	0.00293***	0.00293***	0.00293***
	(2.63e-05)	(2.63e-05)	(2.62e-05)
# of certificates	0.0663***	0.0658***	0.0679***
	(0.0173)	(0.0173)	(0.0172)
Provide video	-0.0713***	-0.0718***	-0.0719***
	(0.0121)	(0.0121)	(0.0121)
SameProv	1.089***	1.094***	1.090***
	(0.0164)	(0.0165)	(0.0166)
LocRelated	-0.160***	-0.135***	-0.160***
	(0.0145)	(0.0146)	(0.0146)
SameProv $\times$ LocRelated	0.609***	0.574***	0.572***
	(0.0268)	(0.0266)	(0.0265)
Control for individual dummies	Y	Y	Y
Observations	12,661,705	12,661,705	12,661,705
AIC	434080.8	434132.7	434125.5
BIC	434296.1	434348	434340.8

Note: 375 groups (187,201 obs) dropped because of all positive or all negative outcomes.

Table C.4: Regression of Amount Invested on Number of Available Projects

VARIABLES	(1) Amt contributed	(2) Amt contributed	(3) log(Daily # of investments)
log(Amount requested)	359.6***	167.0***	
	(41.11)	(24.07)	
log(# of available projects)	-174.1	64.44	0.0239***
	(232.5)	(49.33)	(0.00857)
Constant	-1,943*	-1,147***	-0.0257
	(1,070)	(360.2)	(0.0427)
Funder fixed effects	N	Y	Y
Observations	39,717	39,717	33,798
R-squared	0.094	0.005	0.001

Note: The regressions in columns (1)–(2) use all funder-project pairs; column (3) uses all daily-level observations for each funder. Other project-level characteristics, such as the number of projects proposed by the same fundraisers, the number of comments, and project-category dummies, are also controlled in these regressions.

Table C.5: Estimation Results for Entertainment, Art, Publishing, and Other Categories

Variables	Notations	Entertain	Art	Publishing	Others
# of certificates	β_1	0.6197^{***}	-0.1478^{*}	-0.2043^{***}	-0.2551^{***}
		(0.0564)	(0.0858)	(0.0255)	(0.0464)
# of projects by the fundraiser	β_2	-0.0045^{***}	-0.0224^{**}	0.0328^{***}	0.0544^{***}
		(0.0005)	(0.0097)	(0.0057)	(0.0058)
# of comments	β_3	0.0048^{***}	0.0081^{***}	0.0040^{***}	0.0052^{***}
		(0.0001)	(0.0002)	(0.0001)	(0.0001)
log(Amount requested)	β_4	0.1174^{***}	-0.2147^{***}	0.1750^{***}	0.2037^{***}
		(0.0124)	(0.0165)	(0.0105)	(0.0085)
Provide video	β_5	0.0299	0.0303	0.0030	-0.5259^{***}
		(0.0397)	(0.0562)	(0.0313)	(0.0374)
Location-related project	ϕ	-0.5919^{***}	-0.0474	-0.2112^{***}	-0.1505^{***}
		(0.0565)	(0.0665)	(0.0613)	(0.0376)
Preference premium on projects form same prov	α	0.6852^{***}	0.9079^{***}	0.3330^{***}	0.9198^{***}
		(0.0537)	(0.0473)	(0.0297)	(0.0438)
Ratio between variances	κ	0.8107^{***}	0.9663^{***}	0.9346^{***}	0.8989^{***}
		(0.0159)	(0.0503)	(0.0183)	(0.0123)
Risk aversion	γ	10.8769^{***}	5.3974^{***}	11.5159^{***}	11.5949^{***}
		(0.2496)	(0.2766)	(0.2161)	(0.1662)
Log-likelihood		-17699.56	-18028.09	-40738.02	-26109.59

Table C.6: Predicting Amount Invested by Agents

VARIABLES	(1) log(Amt Invested)
log(Amount requested)	0.3424***
	(7.715e-03)
# of projects by the fundraiser	0.01049***
	(6.689e-04)
# of comments	0.0005141***
	(7.460e-05)
# of certificates	-0.2696***
	(0.02817)
Provide video	-0.1025***
	(0.02664)
SameProv	0.8305***
	(0.03417)
Location-Related	0.1572***
	(0.0333)
Same Prov $\times$ Location-Related	0.08628
	(0.06033)
Category Charity	-0.4668***
	(0.04531)
Category Agriculture	-0.1642***
	(0.046)
Category Publishing	0.2687***
	(0.04519)
Category Entertainment	0.35***
	(0.05451)
Category Technology	-0.03629***
	(0.04998)
Category Art	0.7632***
	(0.05381)
Category Design	0.204***
	(0.05407)
Constant	0.9798***
	(0.1547)
Control for Individal Prov Dummies	Yes
Control for Project Dummies	Yes
Control for Month & Week Day Dummies	Yes
Observations	39717
R-squared	0.139

Table C.7: Evaluating Platform Design Policies: Charity and Entertainment

	Average investors per project							
	Charity				**Entertainment**			
	(1) Data	(2) Remove location	(3) Remove comments	(4) Remove video	(5) Data	(6) Remove location	(7) Remove comments	(8) Remove video
All proj	5.96	5.75	5.08	5.99	4.66	4.31	3.58	4.38
same prov investors	1.47	0.38	1.26	1.39	1.77	0.70	1.52	1.61
diff prov investors	4.48	5.37	3.82	4.59	2.88	3.62	2.06	2.76
Location-related proj	5.94	5.88	5.13	5.90	3.01	2.03	2.67	2.65
same prov investors	1.31	0.34	1.09	1.24	1.68	0.36	1.45	1.48
diff prov investors	4.63	5.55	4.03	4.67	1.33	1.67	1.22	1.17
Non-location-related proj	5.98	5.55	5.02	6.11	6.91	7.44	4.83	6.74
same prov investors	1.70	0.44	1.51	1.62	1.9	1.16	1.62	1.79
diff prov investors	4.27	5.11	3.51	4.49	5.01	6.28	3.21	4.95

	Amount raised and project successful rate							
	Charity				**Entertainment**			
	(1) Data	(2) Remove location	(3) Remove comments	(4) Remove video	(5) Data	(6) Remove location	(7) Remove comments	(8) Remove video
All proj	1886.42	1520.17	1520.76	1850.81	4329.48	3113.17	3262.83	4089.31
same prov investors	775.00	210.84	634.69	727.68	2583.09	943.52	2202.21	2393.00
diff prov investors	1111.42	1309.34	886.08	1123.13	1746.39	2169.65	1060.62	1696.31
Proj successful rate (%)	94.69	94.59	94.29	93.88	81.80	77.89	79.08	78.23
Location-related proj	1872.63	1555.64	1556.66	1813.76	4035.57	2056.76	3456.51	3610.20
same prov investors	711.37	187.04	589.29	672.17	3027.99	746.29	2591.42	2734.71
diff prov investors	1161.27	1368.60	967.37	1141.60	1007.58	1310.47	865.09	875.48
Proj successful rate (%)	95.24	95.75	94.73	94.90	71.72	64.72	68.51	66.76
Non-location-related proj	1907.11	1466.98	1466.92	1906.38	4740.96	4592.14	2991.67	4760.08
same prov investors	870.46	246.54	702.79	810.94	1960.23	1219.64	1657.32	1914.61
diff prov investors	1036.66	1220.44	764.14	1095.44	2780.74	3372.50	1334.35	2845.47
Proj successful rate (%)	93.88	92.86	93.62	92.35	95.92	96.33	93.88	94.29

Table C.8: Evaluating Platform Design Policies: Publish and Technology

	Average investors per project							
	Publish				**Tech**			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Data	Remove location	Remove comments	Remove video	Data	Remove location	Remove comments	Remove video
All proj	13.86	15.83	10.50	13.84	18.92	22.70	12.61	18.55
same prov investors	3.37	2.60	2.60	3.26	3.25	2.79	1.98	3.20
diff prov investors	10.49	13.23	7.91	10.58	15.67	19.92	10.63	15.46
Location-related proj	11.12	10.05	9.21	11.02	10.40	11.80	12.60	13.20
same prov investors	3.83	1.67	3.21	3.69	4.40	1.40	4.60	5.40
diff prov investors	7.29	8.38	6.00	7.33	6.00	10.40	8.00	7.80
Non-location-related proj	14.10	16.33	10.61	14.08	19.09	22.92	12.61	18.65
same prov investors	3.33	2.68	2.54	3.22	3.23	2.81	1.93	3.16
diff prov investors	10.77	13.65	8.07	10.77	15.86	20.11	10.68	15.50

	Amount raised and project successful rate							
	Publish				**Tech**			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Data	Remove location	Remove comments	Remove video	Data	Remove location	Remove comments	Remove video
All proj	9560.91	9898.77	6201.45	9422.97	7119.16	7784.47	3060.65	6923.29
same prov investors	4283.32	3290.27	2896.95	4046.55	2378.33	2074.07	935.37	2368.30
diff prov investors	5277.59	6608.49	3304.50	5376.42	4740.84	5710.4	2125.28	4554.99
Proj successful rate (%)	96.55	95.94	95.74	94.52	97.50	97.92	96.25	98.33
Location-related proj	9638.34	6087.95	7511.62	8407.83	4595.21	3768.13	4780.50	5428.15
same prov investors	6559.62	2372.49	4926.45	5222.63	2885.61	856.27	2659.62	3176.88
diff prov investors	3078.72	3715.46	2585.16	3185.2	1709.61	2911.86	2120.88	2251.27
Proj successful rate (%)	94.74	97.37	94.74	94.74	100.00	75.00	75.00	75.00
Non-location-related proj	9554.44	10217.03	6092.03	9507.75	7161.94	7852.55	3031.50	6948.63
same prov investors	4093.21	3366.92	2727.45	3948.33	2369.73	2094.71	906.14	2354.60
diff prov investors	5461.23	6850.11	3364.57	5559.42	4792.21	5757.83	2125.36	4594.03
Proj successful rate (%)	96.70	95.82	95.82	94.51	97.46	98.31	96.61	98.73

Table C.9: Evaluating Platform Design Policies: Art and Other

	Average investors per project							
	Art				**Other**			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Data	Remove location	Remove comments	Remove video	Data	Remove location	Remove comments	Remove video
All proj	5.61	7.65	4.13	5.29	9.99	10.29	7.71	11.14
same prov investors	1.27	0.62	0.95	1.14	2.71	1.17	2.26	2.91
diff prov investors	4.34	7.03	3.18	4.15	7.28	9.12	5.45	8.23
Location-related proj	4.84	6.52	3.23	4.84	10.11	9.96	8.90	11.74
same prov investors	1.09	0.51	0.87	1.21	3.07	0.88	2.56	3.38
diff prov investors	3.75	6.01	2.36	3.63	7.04	9.08	6.33	8.36
Non-location-related proj	5.73	7.82	4.26	5.36	9.93	10.45	7.11	10.83
same prov investors	1.30	0.64	0.96	1.13	2.52	1.31	2.10	2.67
diff prov investors	4.43	7.18	3.31	4.23	7.41	9.14	5.01	8.16

	Amount raised and project successful rate							
	Art				**Other**			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Data	Remove location	Remove comments	Remove video	Data	Remove location	Remove comments	Remove video
All proj	3579.48	4250.27	2463.60	3367.91	5153.24	4486.64	3999.10	5835.29
same prov investors	1313.09	679.09	957.56	1235.88	2272.92	955.07	1925.14	2545.44
diff prov investors	2266.39	3571.17	1506.03	2132.04	2880.32	3531.57	2073.96	3289.86
Proj successful rate (%)	94.11	96.21	93.05	94.74	93.75	93.25	93.75	96.50
Location-related proj	3747.42	4145.51	2339.15	3863.79	6383.26	5044.01	5383.55	7468.04
same prov investors	1556.20	878.09	1131.56	1868.53	3080.10	883.45	2532.72	3431.44
diff prov investors	2191.22	3267.41	1207.59	1995.26	3303.16	4160.57	2850.83	4036.59
Proj successful rate (%)	90.16	95.08	88.52	90.16	96.45	93.62	95.04	97.16
Non-location-related proj	3554.74	4265.7	2481.93	3294.85	4483.61	4183.20	3245.40	4946.43
same prov investors	1277.27	649.77	931.93	1142.66	1833.48	994.06	1594.37	2063.09
diff prov investors	2277.47	3615.93	1550	2152.19	2650.13	3189.14	1651.03	2883.34
Proj successful rate (%)	94.69	96.38	93.72	95.41	92.28	93.05	93.05	96.14

References

Agrawal, Ajay, Christian Catalini, and Avi Goldfarb. 2014. "Some Simple Economics of Crowdfunding." *Innovation Policy and the Economy,* 14(1): 63–97.

Agrawal, Ajay, Christian Catalini, and Avi Goldfarb. 2015. "Crowdfunding: Geography, Social networks, and the Timing of Investment Decisions." *Journal of Economics & Management Strategy,* 24(2): 253–274.

Ahearne, Alan G., William L. Griever, and Francis E. Warnock. 2004. "Information Costs and Home Bias: An Analysis of US Holdings of Foreign Equities." *Journal of International Economics,* 62(2): 313–336.

Bao, Weining, and Jian Ni. 2020. "Paying to Brag? A Structural Model of Cheap Talk in Online Crowdlending Market." *Working Paper,* Johns Hopkins University.

Bao, Weining, Jian Ni, and Shubhranshu Singh. 2018. "Informal Lending in Emerging Markets." *Marketing Science,* 37(1): 123–137.

Bekaert, Geert, and Xiaozheng Sandra Wang. 2009. "Home Bias Revisited." Available at SSRN 1344880 (2009).

Belleflamme, Paul, Thomas Lambert, and Armin Schwienbacher. 2014. "Crowdfunding: Tapping the Right Crowd." *Journal of Business Venturing,* 29(5): 585–609.

Bruton, Garry, Susanna Khavul, Donald Siegel, and Mike Wright. 2015. "New Financial Alternatives in Seeding Entrepreneurship: Microfinance, Crowdfunding, and Peer-to-peer Innovations." *Entrepreneurship Theory and Practice,* 39(1): 9–26.

Chan, K., V. Covrig, and L. Ng. 2005. "What Determines the Domestic Bias and Foreign Bias? Evidence from Mutual Fund Equity Allocations Worldwide." *The Journal of Finance,* 60(3): 1495—34.

Cooper, Ian, Piet Sercu, and Rosanne Vanpee. 2013. "The Equity Home Bias Puzzle: A Survey." *Foundations and Trends in Finance,* 7(4): 289–416.

Deb, Joyee, Aniko Oery, and Kevin R. Williams. 2021. "Aiming for the Goal: Contribution Dynamics of Crowdfunding." *NBER Working paper*, No. w25881.

Derdenger, Timothy P., and Kannan Srinivasan. 2019. "Funding of Innovative Products: The Impact of Non-Pecuniary Income on Choice." *Working Paper,* Carnegie Mellon University.

Fu, Runshan, Huang, Yan and Singh, Param Vir. 2021. "Crowds, Lending, Machine, and Bias." *Information Systems Research,* 32(1): 72–92.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy. 2019. "Text as Data." *Journal of Economic Literature,* 57(3): 535–574.

Hortacsu, Ali, F. Martínez-Jerez, and Jason Douglas. 2009. “The Geography of Trade in Online Transactions: Evidence from eBay and Mercadolibre.” *American Economic Journal: Microeconomics* 1(1): 53–74.

Ivkovic, Z., Weisbenner, S., 2005. “Local Does as Local Is: Information Content of the Geography of Individual Investors’ Common Stock Investments.” *Journal of Finance* 60(1): 267–306.

Jayachandran, Seema. 2020. “Microentrepreneurship in Developing Countries.” *Working Paper*, Northwestern University.

Lin, Mingfeng, and Siva Viswanathan. 2016. “Home bias in Online Investments: An Empirical Study of an Online Crowdfunding Market.” *Management Science,* 62(5): 1393–1414.

Narasimhan, Laxman, Kannan Srinivasan, and K. Sudhir. 2015. “Editorial – Marketing Science in Emerging Markets.” *Marketing Science,* 34(4): 473–479.

Netzer, Oded, Alain Lemaire, and Michal Herzenstein. 2019. “When words sweat: Identifying signals for loan default in the text of loan applications.” *Journal of Marketing Research,* 56(6): 960-980.

Sharma, Amalesh, V. Kumar, and Koray Cosguner. 2019. “Modeling Emerging-Market Firms’ Competitive Retail Distribution Strategies” *Journal of Marketing Research,* 56(3): 439–458.

Strausz, Roland. 2017. “A Theory of Crowdfunding: A Mechanism Design Approach with Demand Uncertainty and Moral Hazard.” *American Economic Review,* 107(6), 1430–76.

Xin, Yi. 2020. “Asymmetric Information, Reputation, and Welfare in Online Credit Markets.” *Working Paper,* California Institute of Technology.

Xu, Yan, and Jian Ni. 2022. “Entrepreneurial Learning and Disincentives in Crowdfunding Markets.” *Management Science, Articles in Advance, Available at https://doi.org/10.1287/mnsc.2021.4248*.

Zhang, Juanjuan, and Peng Liu. 2012. “Rational Herding in Microloan Markets.” *Management Science,* 58(5): 892–912.