Can Capital Defy the Law of Gravity? Investor Networks and Startup Investment*

Christian Catalini†

MIT

Xiang Hui‡

MIT

July 5, 2017

Abstract

In early crowdfunding platforms, backers would directly fund projects without relying on traditional experts to select and curate projects for them. This approach becomes problematic when equity is involved, since the degree of asymmetric information and the risk of moral hazard are higher than in reward-based crowdfunding. Platforms have therefore experimented with market design solutions targeted at counterbalancing these risks. We study how online syndication by professional investors changes the allocation of capital on the leading US platform. Using novel data on investments and startup valuations (2013-2016), we find that the introduction of intermediaries increases capital flows to non-hub regions, a result that relies on syndicate leads having pre-existing professional ties in these areas. Moreover, the early-stage deals closed through an intermediary in these new regions tend to be associated with better performance, suggesting that expert networks play a key role in arbitraging investment opportunities and expanding access to capital across US regions.

---

*We thank Erik Brynjolfsson, Avi Goldfarb, Kevin Laws, Meng Liu, Hong Luo, Ramana Nanda, Scott Stern and Jane Wu for helpful discussions and comments. The researchers acknowledge the support of the Junior Faculty Research Assistance Program at the MIT Sloan School of Management, and the MIT Initiative on the Digital Economy (http://ide.mit.edu/).

†Christian Catalini is Assistant Professor of Technological Innovation, Entrepreneurship and Strategic Management at MIT Sloan School of Management: catalini@mit.edu.

‡Xiang Hui is a Post-Doctoral Associate at MIT Sloan School of Management, Initiative on the Digital Economy: xianghui@mit.edu

1 Introduction

Early crowdfunding platforms were based on a premise of complete disintermediation from expert networks: The crowd would directly fund projects based on the information shared online by the entrepreneurs, bypassing in the process traditional gatekeepers. This approach becomes problematic when equity is involved because of asymmetric information between startup founders and investors, and because of the lack of incentives for individual investors to perform due diligence on a startup when investing small amounts of capital.

To avoid the unraveling of the market and to attract high quality deals, platforms have experimented with market design solutions targeted at providing professional investors with incentives to perform due diligence, curation and monitoring of early-stage deals on behalf of the crowd. In exchange for their services and to encourage them to share and syndicate their offline deal flow on the platform, professional investors are rewarded in a similar way to venture capitalists with carried interest (a share of profits).[1] The process of online syndication aligns incentives between the crowd, professional investors and the platform hosting the deals, as everyone only profits if the entrepreneurs that are funded are ultimately successful.[2]

While potentially beneficial for the overall efficiency of equity crowdfunding platforms (because it creates a market for due diligence and monitoring), the switch from direct investment by the crowd to deals syndicated by professional investors risks skewing access to capital in favor of regions with strong entrepreneurial ecosystems. In the syndicated model, professional investors act as intermediaries by pre-screening online deals and "certifying" them through their offline reputation. As a result, if angel investments are predominantly local, then online capital, after the introduction of syndication, will disproportionately flow to regions with a strong, pre-existing presence of professional investors, reinforcing pre-existing agglomeration.

The objective of this paper is to explore the trade-offs generated by the introduction of intermediaries in online markets for early-stage capital. We first develop a simple theoretical model to compare how investors and startups in hub regions versus not are affected by syndication, and then use unique data on capital flows and outcomes from the leading US platform to test our predictions.

Our key finding is that syndication is associated with increased capital flows towards non-

[1]In addition to the carry, venture capital firms also charge management fees on the capital they invest on behalf of limited partners.

[2]In reward-based equity crowdfunding (e.g. Kickstarter, Indiegogo), platforms typically earn a fee on the total amount successfully raised through the website, irrespective of long run outcomes. Such a model is particularly ineffective when equity is involved, as it does not incentivize platforms to surface and match only high quality deals, but encourages them to increase the total volume of transactions, irrespective of quality.

hub regions. We test the robustness of this novel finding using multiple empirical methods: An event-study approach around the introduction of this experimental feature on the platform (which can be considered as a natural experiment on the pre-existing crowd of investors); propensity score matching on investor observables; and a difference-in-differences estimation combined with matching. Whereas all these approaches have their weaknesses, they are all consistent in terms of the direction and magnitude of the effect: after syndication, non-hub regions receive up to 25% more funding, and investment in startups from California decreases by a comparable amount.

Results support the view that investors' and syndicate leads' professional ties to a region, more than geography per se, explain the observed flows. While investments on the platform - consistent with the early-stage finance literature - tend to be disproportionately local, this local bias is drastically diminished when we control for the geographic distribution of an individual's professional network across regions. In this context, the presence of professional ties in an area appears to be a better proxy for an individual's ability to source, evaluate, and secure deals from that area relative to what can be captured using location information alone. Professional investors are highly mobile individuals, and their professional ties likely reflect valuable relationships they may have formed in the past through temporary or extended forms of co-location outside of their current home region.

Under syndication, moreover, the relevant professional network for explaining where capital is allocated is not the network of the investors from the crowd, but the network of the intermediary. This is consistent with syndicate leads using their ties to source and evaluate deals across regions, and with the crowd delegating these tasks to them in exchange for a share of future returns. In the data, syndicate leads with professional networks that span outside of the main entrepreneurial hubs are disproportionately responsible for moving capital towards new regions.

The local premium in online investment is actually stronger when intermediaries select startups relative to when the crowd makes investment decisions on its own. This is consistent with syndicate leads being more likely than the crowd to take advantage of lower due diligence and monitoring costs locally to reduce information asymmetry and the risk of moral hazard with startup founders. As intermediaries, they have an incentive to invest in screening both because of the carried interest they can earn if a startup has an exit, and because of the reputational cost they would face in case of failure.

Consistent with information asymmetry constituting a key obstacle to online investment, when observable startup quality is already very high ex-ante, the local investment premium is almost non-existent. Similarly, syndicate leads with a stronger, observable reputation attract capital from any US region, suggesting that investors can evaluate this set of high profile intermediaries directly online. Syndicate leads with a less established reputation, instead, tend to receive capital mostly from local investors. This supports the idea that for this set of individuals information that travels mostly locally and through professional networks may be valuable in separating high ability individuals from others. Moreover, the effect of syndication on capital flows does not vary with the distance between the syndicate lead and the startup, which is inconsistent with monitoring costs being a key friction in this context.

Last, using unique data on startup valuations, we explore the performance of intermediated versus direct online investments, finding that the investments in new regions enabled by syndication are on average associated with better performance. This is consistent with syndicate leads arbitraging investment opportunities across regions by connecting online investors with high quality startups outside of traditional hubs which may have had a more difficult time raising funding offline. Top performing intermediaries are not necessarily the ones with the most established professional network – a proxy for their access to deals and investment ability – but tend to have ties to regions where capital is more scarce.

Our paper contributes to three related streams of research. First, the economic literature on local bias, which spans from entrepreneurial finance, angel investment and venture capital (Lerner [1995], Gompers [1995], Sorenson and Stuart [2001], Cumming and Dai [2010], Chen et al. [2010]), to more traditional stock and equity markets (Coval and Moskowitz [2001], Huberman [2001], Seasholes and Zhu [2010], Van Nieuwerburgh and Veldkamp [2009], French and Poterba [1991], Cooper and Kaplanis [1994], Coval and Moskowitz [1999], Graham et al. [2009]), and, more recently, to the conditions under which online markets can help overcome offline frictions (Blum and Goldfarb [2006], Hortaçsu et al. [2009], Lin and Viswanathan [2015], Forman et al. [2009], Brynjolfsson et al. [2009], Overby and Forman [2014]). This stream is closely related to work showing the presence of both rational and irrational herding on online platforms (Zhang and Liu [2012], Burtch et al. [2013], Agrawal et al. [2015], Freedman and Jin [2011], Senney [2016]).

Second, we build on research that studies the role of professional ties, such as "family and friends" (Parker [2009], Agrawal et al. [2015], Hampton and Wellman [2003]), professional networks (Hochberg et al. [2007], Hsu [2007], Cumming and Johan [2013]), online communities (Mollick [2014]) and diaspora networks (Nanda and Khanna [2010]) in early-stage entrepreneurial activity and related outcomes. Of particular importance for high growth startups are the offline networks that venture capitalists form with each other in order to diversify risk and syndicate deals across geographies (Sorenson and Stuart [2008], Lerner [1994]).

Third, we build on the nascent literature on the economics of equity crowdfunding (Belleflamme et al. [2014], Agrawal et al. [2014, 2016a], Vulkan et al. [2016]), and on the ability of the crowd to expand access to capital to new segments of entrepreneurs and projects (Kim and Hann [2014], Agrawal et al. [2016b], Sorenson et al. [2016]), and to compete with experts in the selection of high impact projects (Mollick and Nanda [2015], Mollick [2013]).

The paper proceeds as follows: In the next section we introduce a simple theoretical framework to guide our empirical predictions. Section 3 describes the data and empirical strategy. Section 4 discusses our key results and Section 5 concludes.

2 Theoretical Framework

The objective of this section is to provide a simple framework for describing how the introduction of intermediaries on equity crowdfunding platforms may influence the allocation of capital across regions. We start by characterizing direct online investments by the crowd, and by comparing the optimal investment decision of different individuals based on their ability to access investment opportunities and perform due diligence on startups located in hub versus non-hub regions. We then allow for intermediaries in the form of syndicate leads scouting and curating deals on behalf of the crowd in exchange for a share of profits. The theoretical framework guides our empirical predictions and helps us identify some of the key mechanisms that may drive changes in the investment behavior of the crowd when intermediaries are introduced.

2.1 Direct Investment by the Crowd

We model individual investment decisions using a static, single-agent optimization framework. Investors indexed by i are either based in a top entrepreneurial ecosystem (hub, or $L_i = H$) or in a peripheral region (non-hub, or $L_i = NH$). As a result of agglomeration, hub regions are assumed to have, on average, higher quality startups: i.e. if we were to randomly draw a startup from a hub versus a non-hub region, the quality of the first is likely to be higher because of Marshallian agglomeration economies (economies of scale, labor market pooling, knowledge spillovers).

Investors are profit maximizers, and their returns depend on three factors: their access to deals, ability to perform due diligence, and monitoring costs. Every period, investors observe their investment parameters and decide to either invest in a startup or not to invest. Conditional on investment, their return from investing directly (i.e. without an intermediary) is given by:

$$\Pi_i^D = max\{\gamma(n_i^H), \gamma(n_i^{NH}), \rho\} - \kappa_1 d_{id} \tag{1}$$

The first term captures investors' returns due to their ability to invest in high quality startups. Higher returns could stem from investors having access to higher quality deals through their professional network, from their ability to screen startups and perform due diligence more effectively, or from their ability to better monitor startups ex-post. Investors select between the following options: 1) they can leverage their connections within a hub region to select a startup in a hub and obtain $\gamma(n_i^H)$; 2) they can leverage their connections outside of a hub region and obtain $\gamma(n_i^{NH})$; or 3) they can randomly select a startup from a hub region and get ρ. ρ captures the difference between the average return from startups located in hubs versus non-hubs. $\gamma(n_i^H)$ and $\gamma(n_i^{NH})$ are respectively the highest investment return that investors can make given their degree of connectedness in hub versus non-hub regions. The professional network strength measures $-n_i^H$ and $n_i^{NH}-$ are empirically represented by the number of individuals who are reciprocally connected to the investor on the platform. We assume that $\gamma(n)$ increase in n, i.e. investors with more connections in the focal region have access to better investment opportunities in the same area (e.g. because of lower search costs). For tractability, we assume that $n_i^H \sim U[0, 1]$ and $n_i^{NH} \sim U[0, 1]$. We allow for an arbitrary correlation between investors' locations and their ties in different regions: $n_i^H|(L_i = H) \sim U[\Delta_1, 1]$ and $n_i^{NH}|(L_i = NH) \sim U[\Delta_2, 1]$, where δ_1 and δ_2 are both between 0 and 1. The *max* operator captures the idea that investors choose the option that yields the highest profit.

The second term $\kappa_1 d_{id}$ captures the cost of monitoring a startup after an investment has been made, conditional on investing. d_{id} is the geographic distance between investors and startups. We assume that monitoring costs increase in d_{id} governed by a marginal effect of $\kappa_1 > 0$. For example, monitoring a distant startup could incur additional travel or information costs.

It is important to note that the size of an investor's pre-existing professional network is likely to not only be positively correlated with their access to deals, but also with their ability to conduct due diligence, to sign up entrepreneurs, and to monitor and mentor startups (i.e. n_i is not orthogonal to ability). Investors of higher ability will attract more inbound requests for investments from local and distant startups both because of their broader professional network, and because of their talent. In the paper, we do not directly separate investors' ability from the reach of their professional network and related access to deal flow, although in some of the regressions we use observable measures of investor quality to at least partially distinguish between these mechanisms.

Proposition 1 *(Geography of Direct Investments) Direct online investments are subject to a local premium (LP^1) and a hub premium (HP) if $\rho > \gamma\left(\frac{\Delta_2+2\eta(\kappa_1)(1-\Delta_2)}{1+\Delta_2}\right)$. The probability of investing outside of a hub weakly increases with the share of an investor's professional network outside of hubs.*

Proof. The local premium (LP^1) results from two sources: a disproportionate share of an individual's professional network being local, and the monitoring cost term $\kappa_1 d_{id}$. HP comes from the fact that investors with low n_i prefer a randomly drawn startup from a hub region – provided the quality differential between startups in hub regions and in non-hub regions is high enough – over investing in a startup they can source and evaluate through their professional network.[3] $\gamma'(\cdot) > 0$ implies that the probability of investing in non-hubs strictly increases with an investor's ties to non-hubs if $n_i^{NH} \geq \gamma^{-1}(\rho)$; this probability is constant with respect to n_i^{NH} if $n_i^{NH} < \gamma^{-1}(\rho)$. ■

2.2 Introduction of Intermediaries on the Platform

We extend the model by allowing for intermediaries on the platform in the form of syndicate leads. Syndicate leads source deals and share them with the online crowd in exchange for a share of future returns (the 'carry'). They are indexed by s, and differ on the same dimensions as investors in terms of their location and the size of their professional network. Since the aim of our paper is to study the changes in capital flows induced by the introduction of syndication, we define the crowd's return function under syndication as:

$$\Pi_i^S = (1-\tau)\left[max\{\widetilde{\gamma}(n_i^H), \widetilde{\gamma}(n_i^{NH}), \rho\} - \kappa_2 d_{sd}\right] \tag{2}$$

Syndicate leads charge carry τ for their services, and investors from the crowd get $(1-\tau)$ of the leads' investment return. Like other investors, syndicate leads take the profit-maximizing option among the three presented before: investing in hub regions through their network, investing in non-hubs through their network, or taking a random draw from a hub. We assume the syndicated investment return $\widetilde{\gamma}(\cdot)$ – which increases in a syndicate lead's network $\widetilde{n_s}$ – also increases with the network of the investor from the crowd n_i. This could be driven by better-connected investors from

[3] Formal proofs are in Appendix Section 6.1.

the crowd having access to better syndicates (e.g. if a deal is oversubscribed, only the best investors will be allocated a share in it), or by better-connected investors being able to screen syndicate leads more effectively in the first place (e.g. because they not only rely on the leads' public reputation, but also on information coming through their professional network). $\kappa_2 d_{sd}$ captures the syndicate leads' cost of monitoring startups.

Proposition 2 *(Geography of Syndicated Investments) When investors from the crowd select syndicate leads they will exhibit a local premium (LP^2). Similarly, if syndicate leads perform due diligence on the startups they invest in, and this is less costly when co-located, then they will exhibit a local premium in their selection of startups (LP^3). Under syndication, the overall hub premium (HP) decreases if $\widetilde{\gamma}'(\cdot) \geq \gamma'(\cdot)$, and a syndicate lead's share of professional ties to non-hubs is not smaller than that of investors from the crowd.*

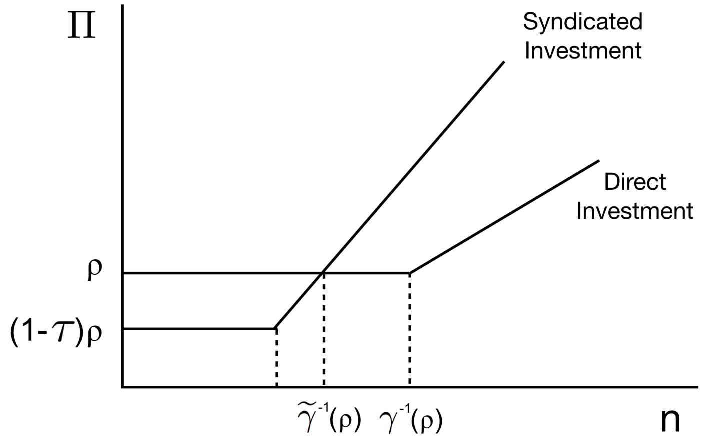
Proof. *(Informal proof)* We have LP^2 and LP^3 because professional ties are disproportionately local, and monitoring costs increase with distance (both for selecting syndicate leads and for selecting startups). If $\widetilde{\gamma}'(\cdot) \geq \gamma'(\cdot)$, under syndication, investors derive more value from their professional network from non-hub regions. HP decreases for two reasons: first, some investors who used to take a random draw from a hub region now switch to syndicated investments (some of these funds may end up in non-hub regions depending on the geography of the syndicate lead's network). Second, if monitoring syndicates is less costly than monitoring startups, then with intermediaries, non-hub investors may be inclined to delegate to a local, non-hub syndicate lead relative to investing in a hub startup directly. On the other hand, if $\widetilde{\gamma}'(\cdot) < \gamma'(\cdot)$ and $\kappa_2 < \tilde{\gamma}(\gamma^{-1}(\rho + \kappa_1)) - \rho$, investors with strong professional networks will still prefer direct investment to using a syndicate.[4] ■

2.3 Extension: Heterogeneous Quality

In Appendix Section 6.1.3, we extend the model by introducing heterogeneous startup and syndicate lead quality into our framework. In particular, we discuss both quality that is observable on the platform, as well as unobservable without direct interaction (e.g. face-to-face due diligence). We also explicitly allow investors to obtain an estimate of the true, unobservable quality of a startup through face-to-face due diligence, and explore how the geography of capital flows changes with the strength of the information available on the platform.

[4] Formal proofs are in Appendix Section 6.1.

Figure 1: Introduction of Syndication



Notes: This figure plots investment returns for syndicated and direct investments. For simplicity we abstract away here from hubs versus non-hubs. The x-axis represents an investor's number of professional ties. τ is the average carry charged by a syndicate lead. γ is the investment return that would be obtained by relying only on the professional ties of the investor from the crowd. $\widetilde{\gamma}$ is the investment return through syndication. ρ is the investment return from a random draw from a hub region.

A key prediction of the extension is that if syndicate leads are actually better at face-to-face due diligence, and are able to obtain a less noisy signal of true quality than the crowd, then the difference in local premium between syndicate leads and investors from the crowd should be larger for startups that have ex-ante a low, observable signal of quality. Because of a comparative advantage in offline screening, intermediaries are more likely to surface startups that would otherwise look less promising based on observables.

3 Data and Empirical Strategy

We use online investment reservation data from AngelList from August 2013 to October 2016. AngelList is the leading equity crowdfunding platform operating under Title II of the US JOBS Act in terms of the number of accredited investors, startups and deals that are performed online. According to the platform, the majority of the deals are seed stage investments (56%), although follow on investments at the Series A (17.7% of the activity) and Series B (15.5%) level are increasing. More than 300 syndicate leads are on the platform, spanning multiple geographic locations and industry sectors: from IT, software and e-commerce to health care, fintech, hardware, logistics and analytics. As of 2016, over $440M have been invested online through the platform in over 1,000 startups. These online, early-stage investments are often followed by larger, traditional venture capital rounds lead by some of the top US VC firms.[5] Venture capital firms are increasingly co-investing with the crowd to later have the option to lead larger, follow-on rounds. Although online investment by accredited investors was only allowed since 2013, by 2016 the platform already had two 'unicorn' exits from its first investments: Dollar Shave Club, acquired by Unilever for $1B, and Cruise Automation, acquired by General Motors for $1B.[6]

The data we use in the paper consists of detailed information on investors, syndicate leads, startups, and capital flows. One key feature of our dataset is that it allows us to observe investors' professional networks, locations, investment amounts, and whether they use an intermediary or decide to invest directly. We also have access to 2017 valuation data for almost all of the online deals, and use this information to compare the unrealized returns of different types of deals on the platform.

Table 1 presents descriptive statistics for our sample. In Panel A, we see that during our observation period (2013–2016) there are 112,529 investment reservations on the platform. Most of these come from investors located in California (65%) or New York (16%). The share of syndicated investments has been steadily increasing on the platform since the introduction of the feature, and by the end of our sample the vast majority of online deals are syndicated. On average, 43% of reservations come from investors located in the same US state as the startup. This percentage is substantially higher within California (61%). As of 2017, 56% of syndicated investments have above the median performance (*'Share High Performance'*), compared to 48% for direct investments.

The geographic distribution of startups (Panel B) and syndicate leads (Panel D) is similar to the one presented in Panel A for reservations. Investors from the crowd (Panel C) instead, are more geographically dispersed: 52% of investors are from California, 15% from New York, 4% from Massachusetts and the remaining 30% is from other regions. The total amount of capital reserved is approximately $870M, with a per startup average of $730K. Observable startup quality, which ranges from 0 to 10, is a platform measure of profile integrity, completeness and overall quality.

Comparing Panels C and D, we see that the average number of reciprocal, professional ties is 179 for investors, and 451 for syndicate leads. The median observable syndicate lead quality – which similar to startup quality, is a platform proxy for the overall quality of a profile – is 7. On average, syndicate leads charge 13% carry on a per-deal basis on the deals they curate.

[5] E.g., Accel Partners, Andreessen Horowitz, Bessmer Venture Partners, First Round Capital, Greylock Partners, Khosla Ventures, Sequoia Capital, Union Square Ventures.

[6] The internal, unrealized rate of return for all 2013 investments conducted on the platform has been higher than comparable metrics for top quartile venture capital funds of the same vintage, which is also consistent with these investments representing earlier stage, riskier deals relative to VC rounds.

Table 1: Summary Statistics

	All	CA	NY	MA	Other States
Panel A. Investment Reservations Characteristics					
	All	CA	NY	MA	Other States
Number of Investment Reservations	112,529	72,849	18,204	3,393	17,818
Share of Investment Reservations	100%	65%	16%	3%	16%
Mean Amount per Reservation	7,757	8,050	8,305	10,286	5,519
Share of Syndicated Reservations	95%	94%	95%	99%	97%
Share of Same State Reservations	43%	61%	14%	9%	6%
Share High Performance (Direct Investment)	48%	46%	76%	N/A	25%
Share High Performance (Syndicated)	56%	52%	65%	63%	61%
Panel B. Startup Characteristics					
	All	CA	NY	MA	Other States
Number of Startups (>=1 Reservation) in	1,188	757	188	69	174
Share of Startups (>=1 Reservation) in	100%	64%	16%	6%	15%
Share of Syndicated Startups	79%	76%	94%	81%	82%
Total Dollars Received	870,878,876	586,457,027	151,191,112	34,899,734	98,331,003
Mean Dollars Received	733,063	774,712	804,208	505,793	565,121
Median Observable Startup Quality	5	5	5	4	5
Panel C. Investor Characteristics					
	All	CA	NY	MA	Other States
Number of Investors (>=1 Reservation) in	9,060	4,726	1,325	318	2,691
Share of Investors (>=1 Reservation) in	100%	52%	15%	4%	30%
Mean Number of Reciprocal Followers	179	117	24	6	33
Panel D. Syndicate Lead Characteristics					
	All	CA	NY	MA	Other States
Number of Syndicate Leads (>=1 Reservation) in	325	222	42	30	31
Share of Syndicate Leads (>=1 Reservation) in	100%	68%	13%	9%	10%
Median Observable Syndicate Lead Quality	7	7	7	8	7
Syndicate Carry	13	14	13	12	13
Minimum Backer Amount	6,709	6,128	5,825	9,650	5,231
Number of Annual Deals	6	6	7	3	6

3.1 Empirical Strategy

We adopt several econometric approaches to estimate the effect of syndication on the geography of online capital flows. We first perform a simple, cross-sectional comparison of the likelihood that an investment reservation is targeted at a startup in a hub versus a non-hub region. We estimate variations of:

$$Y_r = \alpha + \beta Synd_r + \epsilon_r,$$

where Y_r is a dummy equal to one if investment reservation r is in a startup located in state Y; $Synd_r$ captures whether the investment is syndicated versus not; and ϵ_r is an idiosyncratic error term. The estimated $\widehat{\beta}$ captures the correlation between the use of syndication and the likelihood that the investment will end up in the focal region. In the regressions, we also try to separate information that travels disproportionately locally from information that may also travel through professional networks, and control for the investor being co-located in the same state as the startup, as well as the share of an investor's pre-existing professional ties[7] that is clustered in the startup's region. We build the same measures for investors from the crowd and for syndicate leads and estimate:

$$Y_r = \alpha + \beta Synd_r + \gamma_1 SameState\_i + \gamma_2 ShareTies\_i + \gamma_3 SameState\_s + \gamma_4 ShareTies\_s + \epsilon_r,$$

where suffix $\_i$ indicates an investor from the crowd and $\_s$ indicates a syndicate lead.

Of course, the correlations from the cross-sectional analysis capture both the 'treatment effect' of syndication on capital flows as well as investor unobserved heterogeneity, which likely affects their selection into syndicated investments in the first place. Hence, we also rely on an event-study approach around the introduction of syndication, combined with investor fixed effects, to test the robustness of our result to controlling for investor unobservables. The regression model for the event-study is:

$$Y_{rit} = \beta Synd_{rit} + \mu_i + \psi_t + \epsilon_{rit},$$

where Y_{rit} is a dummy equal to one if investment reservation r is in state Y, and μ_i and ψ_t are respectively investor and month fixed effects. These regressions mirror the previous set and include controls for investors' and syndicate leads' locations and professional ties. Key advantages of this approach over the cross-sectional one are that the introduction of syndication is an exogeneous event for the pre-existing online crowd of investors (it was not pre-announced), and that this specification allows us to control for time-invariant, investor-specific unobservables through investor fixed effects.

A weakness of the event-study approach is that it lacks a control group, since the policy affects all investors on the platform. Thus, we construct a comparison set using two different matching estimators: propensity score matching (PSM), and a matching, difference-in-differences estimator (MDID). The intuition behind both of these methods is to compare geographic outcomes for syndicated investments versus direct investments for investors with similar observable characteristics. We match investors using multiple variables: whether they are based in California, New York, or in a non-hub region, whether they have more than 500 connections on LinkedIn, have investment experience, had an investment that resulted in an IPO or acquisition, and whether they graduated from a Top 25 MBA program according to US News.

Ideally, we would perform a direct match on all of our observables, however such a procedure

[7] Captured one or six months before an online reservation is made depending on the specification.

rapidly runs into the 'curse of dimensionality' problem. Hence, we perform a propensity score match (PSM)[8] using two steps: 1) the parametric estimation of the propensity of using syndication through a probit model; and 2) the estimation of the effect of syndication with a cross-sectional comparison between matched syndicated and direct investors with similar estimated propensity scores. To further test the robustness of the findings, we use multiple matching methods: nearest neighbour matching, radius matching, kernel matching, stratification matching, and inverse probability weights. Estimates are consistent across all these approaches to matching.

Last, in addition to the PSM estimation, we also use a matching, difference-in-differences (MDID) estimator in the form of:

$$\widehat{\beta}^{MDID} = \frac{1}{N} \sum_{i \in \{I^1 \cap S^*\}} \left\{ \Delta Y_{it} - \sum_{j \in \{I^0 \cap S^*\}} W_{ij} \Delta Y_{jt} \right\}$$

where I^1 is the set of treated investors; I^0 is the set of control investors; S^* is the region of common support revealed by the propensity score; W is the weight placed when comparing control unit j with treatment unit i (which depends on the matching method).

The idea behind the MDID estimation is to compare temporal changes in the treated units against those of the matched, non-treated units. The key advantage of MDID over the PSM estimation is that MDID also allows for time-invariant, unobservable characteristics of the investors to affect selection into treatment. The identification assumption of the MDID approach is that there are no time-varying, unobservable effects that are correlated both with selection into syndication and the decision of where to invest.

4 Results

We first present descriptive results on the geography of capital flows under direct investment. In the absence of intermediaries, non-hub regions underperform entrepreneurial hubs in terms of the capital they attract relative to their share of startup activity (Section 4.1). We then explore how the introduction of syndication influences these flows, and find that it expands access to capital outside of traditional hubs (Section 4.2). To understand the mechanism behind this effect, we perform a number of additional analyses and introduce proxies for ex-ante startup and syndicate lead quality

[8] Rosenbaum and Rubin [1983] prove that when matching based on all variables Z is valid (i.e. outcome is independent of program participation conditional on Z), then matching based on the propensity score $Pr(D = 1|Z)$ is also valid.

into our regressions to study heterogeneous effects (Section 4.3). Last, we use startup valuation data to assess how the investments enabled by syndication perform relative to direct investment deals (Section 4.4).

4.1 Geography under Direct Investment

In this section, we explore capital flows under the direct investment model, and use investor reservation data for all deals that do not involve a syndicate lead intermediating between the crowd and the startup. AngelList started as a website for listing angel investors and only later added startup and founder profiles. Direct investment was launched on the platform in 2013 after AngelList received a 'no-action letter'[9] from the SEC, which allowed it to accept online investments from accredited investors.[10]

As can be seen in Panel A of Table 2, under direct investment, California receives a larger share of investments (74%) relative to its share of startups (65%), whereas capital inflows to New York (17% of the total) are roughly proportional to the region's share of startups (16%). In terms of investment reservations received relative to its share of startups, the third top region, Massachusetts, is similar to non-hub regions. Therefore for the rest of the paper we treat California and New York as hubs, and all other states as non-hubs.

Although non-hub regions account for 19% of startups, they only attract 9% of the capital. In addition, regardless of where investors are located, size-adjusted capital flows to California and New York are consistently larger than those to non-hub regions. This is consistent with the presence of the hub premium discussed in the theoretical framework (HP): in the absence of intermediaries, online investors are more likely to invest in startups from regions known for their entrepreneurial ecosystems, possibly also because these startups have more recognizable signals of quality (e.g. affiliation with a top accelerator program or notable investor, which facilitate investment over distance). These signals of quality should facilitate investment irrespective of geographic distance. The table also supports the presence of a local premium (LP^1), which is not unusual in early-stage deals: the share of investments targeted towards non-hub regions is higher if the investors are from the same region.

Last, consistent with Proposition 1 in the theoretical framework, in Panels B1 and B2 we see that the share of non-hub investments is slightly increasing with an investor's share of professional

[9]https://www.sec.gov/divisions/marketreg/mr-noaction/2013/angellist-15a1.pdf

[10]For an investor to be considered accredited by the SEC, the individual must have net worth of $1M (not including the primary residence), or income of $200K in the last two years with the expectation of similar income going forward.

ties within non-hub regions, independent of where the investor is located. Whereas investors in the top quartile of the distribution target 11% of their investments towards non-hub regions, the share is only 8% for everyone else. In Panel B2, the dependent variable is equal to 1 if the investor ever invests in a non-hub, and the right hand side variables capture her share of professional ties within non-hub regions. We use the raw share in Column 1, and bins defined by quartiles, deciles and percentiles of the share distribution across investors in Columns 2 to 4 to create a scale-free estimate. Consistent with the theory, all estimates are positive and statistically significant.

Table 2: Direct Investment

Panel A. Geography of Direct Investments				
	CA	NY	MA	Other
Share of Startups in	65%	16%	3%	16%
Share of Direct Investment to	74%	17%	0%	9%
Share of Direct Investment from CA to	76%	17%	0%	8%
Share of Direct Investment from NY to	77%	13%	1%	9%
Share of Direct Investment from MA to	69%	20%	0%	10%
Share of Direct Investment from Other to	72%	16%	0%	11%
Panel B1. Share of Non-Hub Investments by Quartiles of the Share of Non-Hub Professional Ties				
	Q1	Q2	Q3	Q4
Share of Non-Hub Investments	8%	7%	8%	11%
Panel B2. Regression. Outcome Variable: Invest in a Non-Hub (1/0)				
	(1)	(2)	(3)	(4)
	Share	By Quartiles	By Deciles	By Percentiles
Ties in Non-Hubs	0.083***	0.009**	0.004**	3.3E-4***
	(0.035)	(0.004)	(0.002)	(1.5E-4)
Observations	3,978	3,978	3,978	3,978

4.2 Geography under Syndication

Table 3 introduces syndication into our analysis of geography and capital flows. The unit of analysis for the OLS regressions is an investor-startup pair, and the dependent variable is equal to 1 if the startup is located in California (Panel A), New York (Panel B), or any other region (Panel C). Standard errors are clustered at the deal level. Adding months fixed effects leads to qualitatively similar results (see Table A-1).

Column (1) in Panels A and C shows that the use of syndication is correlated with a 15%

reduction in investment in startups in California, and a 14% increase in non-hub regions. This supports the idea that syndication is associated with a decrease in the concentration of capital flows. Even in a context where deals are conducted online, positive coefficients on *'Investor in Same State'* are consistent with the presence of a local premium (LP^1). The local premium is more pronounced in non-hub regions, which is a reflection of non-hub investors also investing in hub deals, but hub investors not diversifying outside of their home region.[11]

In Column (2), we introduce in the regressions the investors' share of professional ties in the startup's state. This share is taken one month before the investment takes place, and in Appendix Table A-2, we show robustness to extending this window further to six months. Adding this proxy for an investor's professional network halves local premiums in California and New York, and decreases the coefficient on *'Investor in Same State'* in the other regions by approximately 80%. This suggests that more than half of the investors' local premium comes from the disproportionately local nature of their professional ties. We suspect that these ties are highly correlated with an investor's ability to access deals in the focal regions, although they are clearly also positively associated with the investor's underlying ability (i.e. better investors will have larger and broader networks, and will also be better at screening and monitoring deals). Co-location, whether temporary or not, is likely to be a pre-requisite for the formation of professional ties in a region in the first place.

Moving from Column (2) to Column (3), we add a control for the location of the syndicate lead. The weights on the investors' local premium and on the association between the share of their network in the startup's state and investment are further reduced. Furthermore, consistent with syndicate leads also exhibiting a local premium (LP^3), there is a strong, positive association between the location of the syndicate lead and the location of the focal startup (*'Syndicate Lead in Same State'*).

Although syndicate leads are mostly concentrated in hub regions, their professional networks span outside of California and New York. In Column (4), we introduce the share of a syndicate lead's ties in the startup's state as a proxy for their connections within the relevant ecosystem. Interestingly, the syndicate lead's local premium drops substantially when their professional network is taken into account, suggesting that here, geography mostly proxies for offline relations between

[11] The coefficients for *'Investor in Same State'* are not directly comparable between the three panels. In all regressions, the comparison group includes investment reservations where the investor and the startup are in different states, as well as cases where the investor and startup are in the same state, but this is not the focal state of the panel. Since Panel C aggregates multiple states, this last group is likely to be smaller. Our results are robust to running splits of the sample state by state. As can be seen in Figure A-1, many of the states are associated with a higher local premium relative to hub regions.

the syndicate leads and entrepreneurs in these regions. The role of the investor network is also reduced, which is consistent with the idea that investor ties may help select syndicate leads, but conditional on using an intermediary, the relevant network for sourcing startups is the syndicate lead's network. We will explore this further in Table 6.

The share of ties in the same state (both for the investor and the syndicate lead) seem to play a more important role outside of hub regions, which is consistent with hubs attracting capital over distance more easily (e.g. because startups in these regions have stronger, observable quality signals), with non-hub startups possibly having a harder time attracting capital from distant investors, or a combination of both.

Figure A-1 plots the coefficients shown in Table 3 for multiple states. In particular, Figure A-1a corresponds to the changes in coefficients on *'Investor in Same State'* from Column (1) to (2); Figure A-1b to those for *'Syndicate Lead in Same State'* from Column (3) to (4); and Figure A-1c for *'Investor Share of Ties in Same State'* from Column (2) to (4). The graphs show that the results from Table 3 on the correlation between capital flows and investors' and syndicate leads' locations and professional ties are robust across states.

Overall, in hub regions, once we control for the location of syndicate leads and their network in the focal region, the crowd does not exhibit any local premium, which is consistent with intermediaries taking over the selection process on behalf of the crowd. In non-hub regions instead, possibly because of greater information asymmetry, investors' professional ties in the focal region still matter for investment decisions.

4.2.1 The Introduction of Syndication as a Natural Experiment

Syndicates were launched as an experiment on the platform to see if offering a carry to angel investors in exchange for sharing their deal flow with the online crowd would incentivize them to move some of their deals online. Their success progressively turned them into the most used investment mode on the platform.

In Table 4, we exploit the period immediately before and after the introduction of syndication as a natural experiment, and use investor fixed effects to focus on within-investor variation and keep the investor pool constant in the analysis. The fixed effects control for unobservable investor characteristics that may drive their use of syndication versus direct investment.[12] Whereas syndi-

[12] We also performed the analysis removing investor fixed effects to capture the change in overall capital flows (a composition of the behavior of pre-existing and new investors), finding qualitatively similar results.

Table 3: Introduction of Syndication

Panel A. Dependent Variable: Startup from California (1/0)

	(1)	(2)	(3)	(4)
Syndicated	-0.150***	-0.145***	-0.406***	-0.152***
	(0.049)	(0.050)	(0.060)	(0.056)
Investor in Same State	0.063***	0.030***	0.012*	0.002
	(0.008)	(0.007)	(0.006)	(0.004)
Investor Share of Ties in Same State		0.123***	0.048***	0.015*
		(0.016)	(0.011)	(0.008)
Syndicate Lead in Same State			0.328***	0.129***
			(0.051)	(0.048)
Syndicate Lead Share of Ties in Same State				1.264***
				(0.115)
Adj R-squared	0.007	0.009	0.082	0.469
N	93087	83938	83533	69631

Panel B. Dependent Variable: Startup from New York (1/0)

	(1)	(2)	(3)	(4)
Syndicated	-0.000	-0.007	-0.017	-0.048
	(0.045)	(0.046)	(0.047)	(0.036)
Investor in Same State	0.025***	0.013***	0.010**	0.003
	(0.005)	(0.005)	(0.004)	(0.004)
Investor Share of Ties in Same State		0.049***	0.039***	0.015
		(0.014)	(0.013)	(0.010)
Syndicate Lead in Same State			0.112*	-0.004
			(0.064)	(0.042)
Syndicate Lead Share of Ties in Same State				1.661***
				(0.136)
Adj R-squared	0.000	0.001	0.007	0.451
N	93087	83938	83533	69631

Panel C. Dependent Variable: Startup from Other Region (1/0)

	(1)	(2)	(3)	(4)
Syndicated	0.144***	0.146***	0.116***	0.032*
	(0.024)	(0.023)	(0.023)	(0.017)
Investor in Same State	0.816***	0.146***	-0.064**	-0.035
	(0.020)	(0.033)	(0.027)	(0.023)
Investor Share of Ties in Same State		1.923***	1.094***	0.499***
		(0.106)	(0.152)	(0.083)
Syndicate Lead in Same State			0.747***	0.227**
			(0.029)	(0.103)
Syndicate Lead Share of Ties in Same State				1.570***
				(0.138)
Adj R-squared	0.054	0.086	0.216	0.622
N	93087	83938	83533	69749

Notes: Standard errors clusterd at the deal level. *** indicates significance at p = 0.01; ** indicates p = 0.05 ; * indicates p = 0.1.

cation has the potential to expand the set of investors who feel comfortable investing online, Table 4 examines how it changes pre-existing investors' propensity to allocate capital to hub versus non-hub regions. All regressions include month fixed effects to control non-parametrically for the changing propensity to invest across regions, and standard errors are clustered at the deal level. Relative to the results presented in Table 3, the event-study approach used in Table 4 takes advantage of the fact that the introduction of syndication is exogenous to the pre-existing online crowd of investors, as there was no pre-announcement on the platform.

Results, which include only the 10 months[13] around the launch of syndication, are consistent with our previous findings on the democratization of capital flows: the use of syndication decreases investment in startups in California by 33%, and increases investment in startups in non-hub regions by 25% (the effect for New York is positive, but non-significant). The local premium of syndicate leads is more than halved once we control for their share of professional ties in a region, similar to what we saw in Table 3. Across all three panels, there is a large, positive association between the location of the syndicate lead's professional network and the location of the startup, which is consistent with the lead's ties being the relevant ones for sourcing and securing deals under the intermediated model.

We should note that compared to Table 3, here we cannot control for whether the investors are in the same state of the startup because the dummy is absorbed in the investor fixed effects. We can still control for investors' pre-existing share of professional ties in the same state because they vary over time. The coefficient for these variables is negative in California and positive in other regions. This is consistent with our theoretical prediction on the effect of intermediaries: the reduction in the hub premium drives some hub investors to diversify outside of their home regions, and some non-hub investors to invest in their own area instead of a hub.

4.2.2 Propensity Score and Matching with Difference-in-Differences Estimation

To provide further robustness to our main finding that syndication expands access to capital to non-hub regions, in this section, instead of relying on investor fixed effects as in Section 4.2.1, we use two matching estimators to build a control group: propensity score matching (PSM), and a matching, difference-in-difference estimator (MDID).

We match investors using multiple variables: whether they are based in California, New York or in a non-hub region, whether they have more than 500 connections on LinkedIn, have invested previously, had an investment that resulted in an IPO or acquisition, and whether they graduated

[13]There are only 10 months of data between the introduction of syndication and the date when the SEC allowed platforms like AngelList to enable online investments.

Table 4: Natural Experiment: Introduction of Syndication, Controlling for Investor Fixed Effects

Panel A. Dependent Variable: Startup from California (1/0)

	(1)	(2)	(3)	(4)
Post	-0.330**	-0.325**	-0.402***	-0.115
	(0.131)	(0.131)	(0.132)	(0.145)
Investor Share of Ties in Same State		-0.174	-0.238**	-0.166**
		(0.114)	(0.108)	(0.083)
Syndicate Lead in Same State			0.232***	0.102*
			(0.066)	(0.055)
Syndicate Lead Share of Ties in Same State				1.343***
				(0.093)
Investor Fixed Effects	✓	✓	✓	✓
Month Fixed Effects	✓	✓	✓	✓
Adj R-squared	0.067	0.069	0.111	0.488
N	3959	3735	3603	3472

Panel B. Dependent Variable: Startup from New York (1/0)

	(1)	(2)	(3)	(4)
Post	0.074	0.071	0.056	0.055
	(0.059)	(0.060)	(0.062)	(0.052)
Investor Share of Ties in Same State		0.021	-0.052	0.085
		(0.129)	(0.145)	(0.126)
Syndicate Lead in Same State			0.467**	0.182**
			(0.197)	(0.080)
Syndicate Lead Share of Ties in Same State				1.322***
				(0.172)
Investor Fixed Effects	✓	✓	✓	✓
Month Fixed Effects	✓	✓	✓	✓
Adj R-squared	0.082	0.086	0.112	0.362
N	3959	3735	3603	3472

Panel C. Dependent Variable: Startup from Other Region (1/0)

	(1)	(2)	(3)	(4)
Post	0.256**	0.225**	0.157*	0.091
	(0.116)	(0.104)	(0.091)	(0.070)
Investor Share of Ties in Same State		3.504***	2.243***	0.872***
		(0.413)	(0.575)	(0.286)
Syndicate Lead in Same State			0.741***	0.054
			(0.063)	(0.272)
Syndicate Lead Share of Ties in Same State				1.885***
				(0.312)
Investor Fixed Effects	✓	✓	✓	✓
Month Fixed Effects	✓	✓	✓	✓
Adj R-squared	0.061	0.146	0.265	0.616
N	3959	3735	2274	2174

Notes: We control for month fixed effects and investor fixed effects in all regressions. Standard errors clustered at the deal level. *** indicates significance at p = 0.01; ** indicates p = 0.05 ; * indicates p = 0.1.

from a Top 25 MBA program according to US News. Ideally, we would perform a direct match on all of our observables, however such a procedure rapidly runs into the 'curse of dimensionality problem'. Hence, we perform a propensity score match (PSM). A proper identification of a treatment effect using PSM hinges on the balance of the relevant observables between treated and control groups, hence we build three matched stratas to maximize balance across our variables. Of course, investors may still be unbalanced on dimensions that are unobservable to us, but the hope is that given the wide set of dimensions we have access to, we are able to capture a sizable part of their differences in investment ability, experience etc.

In Panel A of Table 5, we regress each matching variable on strata fixed effects and their interaction with a dummy indicating whether an investor used syndication. If the matching is valid, we should see non-significant coefficients for all these interactions. Our balance between the treatment (syndication) and control group (direct investment) is valid at the 1% level, and valid at the 10% level for the vast majority of the variables. We repeat the matching using five estimators: nearest neighbor matching, radius matching, kernel matching, stratification matching, and inverse probability weights. PSM estimation results are presented in columns (2)–(7) in Panels B–D. In Column (1) we report the OLS estimates as a benchmark. Column (7) further controls for the investor being co-located in the same state as the startup, as well as month fixed effects.

Results show that there is between a 5.8% and 17.8% reduction in the share of investment to California startups, no significant change in investment to New York startups, and a 5.9% to 16.6% increase in share of investment targeted at other regions.

In addition to the propensity score analysis, we also perform a matching, difference-in-differences (MDID) estimation. As can be seen in Columns (8) and (9) in Panels B–D of Table 5, there is a 7.5% to 14.2% decrease in the share of investments to startups in California after the introduction of syndication, and a 5.6% to 9.4% increase in the share of investment to non-hub regions (depending on whether we control for the co-location between the investor and the startup, and month fixed effects).

Taken together, results from Tables 4 and 5 strongly suggest that there is a causal component to how intermediation through syndicate leads changes the geography of capital flows by moving funding away from California and towards new regions.

Table 5: Propensity Score and Matching Difference-in-Difference Estimations

Panel A. Matching Quality									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	CA	NY	Other	Follower	Invested	Entrep.	Success	Exp.	MBA25
Block 1 ×	0.064	-0.033	-0.031	-0.038	0.039	0.071*	0.006	0.054	0.054*
Syndicated	(0.044)	(0.028)	(0.041)	(0.036)	(0.031)	(0.041)	(0.026)	(0.050)	(0.028)
Block 2 ×	-0.010	-0.019	0.029	-0.004	-0.026*	-0.011	-0.001	-0.021	-0.020
Syndicated	(0.022)	(0.014)	(0.021)	(0.018)	(0.016)	(0.021)	(0.013)	(0.025)	(0.013)
Block 3 ×	-0.029	0.210**	-0.181	0.031	-0.014	0.013	0.061	-0.069	-0.172
Syndicated	(0.154)	(0.097)	(0.143)	(0.125)	(0.107)	(0.142)	(0.091)	(0.175)	(0.129)
Panel B. Dependent Variable: Startup from California (1/0)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	OLS	NN	Radius	Kernel	Strata	IPW	IPW	MDID	MDID
ATT	-0.149***	-0.143**	-0.174***	-0.178***	-0.110***	-0.118**	-0.058*	-0.142**	-0.075*
	(0.049)	(0.056)	(0.045)	(0.041)	(0.020)	(0.049)	(0.035)	(0.070)	(0.045)
Same State							✓		✓
Month FEs							✓		✓
Panel C. Dependent Variable: Startup from New York (1/0)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	OLS	NN	Radius	Kernel	Strata	IPW	IPW	MDID	MDID
Syndicated	-0.001	0.042	0.011	0.030	-0.008	0.019	-0.010	0.051	0.019
	(0.045)	(0.049)	(0.040)	(0.028)	(0.015)	(0.043)	(0.027)	(0.072)	(0.045)
Same State							✓		✓
Month FEs							✓		✓
Panel D. Dependent Variable: Startup from Other Region (1/0)									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	OLS	NN	Radius	Kernel	Strata	IPW	IPW	MDID	MDID
Syndicated	0.152***	0.102***	0.166***	0.150***	0.120***	0.100***	0.059**	0.094**	0.056**
	(0.025)	(0.031)	(0.025)	(0.023)	(0.014)	(0.028)	(0.027)	(0.038)	(0.027)
Same State							✓		✓
Month FEs							✓		✓

Notes: We also control for monthly fixed effects in these regressions. Standard errors clustered at the deal level. *** indicates significance at p = 0.01; ** indicates p = 0.05; * indicates p = 0.1.

4.3 Understanding the Mechanism

Having provided robust evidence that the introduction of syndication is associated with an increase in the amount of capital allocated outside of hub regions, in this section we try to isolate the underlying mechanism, and explore the conditions under which the crowd is moving funds differently because of intermediaries.

We first study if the way the online crowd selects startups differs from the way it selects syndicate leads (Table 6). We then compare how syndicate leads and the crowd select startups to fund (Table 7). The tests performed in this set of tables are complementary, as in syndicated investment there are two steps involved: first, the crowd selects syndicates, and then the syndicates select the startups to invest in on their behalf. The observed effects for intermediated deals are therefore likely to be a composition of these two sets of choices.

Last, building on the insights from our extension of the theoretical framework – which incorporates the role offline due diligence on startups and syndicate leads can play in this context (see Section 6.1.3) – we study heterogeneous effects across the observable quality distribution of syndicate leads, startups, and investors.

4.3.1 The Crowd's Problem: How the Crowd Selects Startups versus How it Selects Intermediaries

The geography of capital flows in the presence of intermediation is a combination of the geography of investors from the crowd choosing syndicate leads, and the geography of syndicate leads choosing deals. We study the former first and ask the following questions: does the crowd exhibit a local premium when selecting syndicate leads (LP^2)? If so, how does it compare to the local premium we observe when the crowd selects startups (LP^1)?

In Table 6, the dependent variable captures the location of the startup if the deal is a direct investment (Panel A), and the location of the syndicate lead if the deal is syndicated (Panel B). In other words, the dependent variable always reflects who the crowd is directly selecting. The key explanatory variable is whether the investor from the crowd is in the same state as the entity being selected (startup or syndicate lead depending on the investment mode).

If we compare columns (1), (3) and (5) between Panels A and B in Table 6, we see that the investor's local premium (*'Investor Same State'*) is stronger when the crowd selects syndicate leads relative to when it selects startups ($LP^2 > LP^3$). In Panel C, we report the t-statistics for the differences between the two local premiums, which is statistically different from zero in all columns with the exception of New York. This means that proximity is more relevant for the process of selecting syndicates than for the process of selecting startups. Information about the quality of syndicate leads could disproportionately travel through local professional networks, or alternatively some startups, specially from hub regions, may have access to quality signals (e.g. affiliation with a top accelerator or notable investor) that allow them to raise capital more easily over distance. In the second case, switching from direct investment to intermediated investment reduces the hub premium.

In Columns (2), (4) and (6) we further control for the investor's share of ties in the focal region. Interestingly, the larger local premium observed in Panel B is largely explained by this new term, suggesting that the selection of a syndicate lead is more strongly correlated with the crowd having professional ties to the region of the lead than the selection of startups. Through syndication, investors from the crowd may be able to leverage information – captured by the network measures – that is not available to them for founders. Whereas investors are unlikely to be connected to every startup founder they would like to invest in, they are more likely to be able to identify, through their professional network and offline interactions, a few lead investors they trust to make investment decisions on their behalf.

The introduction of intermediaries may therefore shift the type of information the crowd uses from public information available on the platform to information they can retrieve through their professional ties and interactions. Our intuition is that this switch has the potential to expand the type of deals the crowd is willing to experiment with, as syndicate leads may have a more visible and easily interpretable reputation (e.g. through their past investment track-record) than first-time entrepreneurs. By choosing syndicates instead of startups, the crows is able to leverage their network and expertise to possibly make better online investment decisions.

Table 6: How the Crowd Selects Startups versus How It Selects Syndicate Leads

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Investors Selecting Startups						
	Startup in California		Startup in New York		Startup in Other	
Investor Same State	0.036	0.025	0.040**	0.016	0.025**	-0.051**
	(0.024)	(0.021)	(0.018)	(0.020)	(0.013)	(0.024)
Investor Share Ties		0.053		0.081		0.204***
Same State		(0.052)		(0.060)		(0.059)
Adj R-squared	0.001	0.002	0.001	0.001	0.001	0.010
N	4848	3849	4848	3849	4848	3849
Panel B. Investors Selecting Syndicate Leads						
	Startup in California		Startup in New York		Startup in Other	
Syndicate Lead Same State	0.335***	0.133***	0.116*	-0.003	0.418***	-0.142**
	(0.051)	(0.049)	(0.064)	(0.043)	(0.058)	(0.066)
Syndicate Lead Share Ties		1.254***		1.649***		1.618***
Same State		(0.116)		(0.134)		(0.127)
Adj R-squared	0.081	0.470	0.007	0.454	0.128	0.622
N	92776	76493	92776	76493	92776	76493
	5.34	1.15	12.31	3.49	4.15	-3.68
Panel C. t-statistics of the Differences Between Panel A and Panel B						
Same State	5.34	2.03	1.15	-0.38	6.6	-1.3
Share Ties Same State		9.48		10.67		10.12

Notes: Standard errors are clustered at the deal level. *** indicates significance at p = 0.01; ** indicates p = 0.05 ; * indicates p = 0.1.

4.3.2 The Startup Selection Problem: How Syndicate Leads versus the Crowd Select Startups

The previous section showed that the crowd behaves differently when it selects syndicate leads compared to when it selects startups. But are syndicate leads behaving any different than investors from the crowd when screening deals on their behalf? A key assumption of the intermediated model is that the ability to earn a carry, combined with the reputational cost of being associated with low quality deals create enough of an incentive for syndicate leads to select high quality startups, perform due diligence, and monitor them.

To explore this, in Table 7 we compare the local premium we observe when the crowd selects startups directly relative to the one we observe when syndicate leads perform selection on its behalf. The independent variables capture the relationship between the startup and the agent making the final capital allocation decision, i.e. an investor in direct investment (Panel A) versus a syndicate lead if the deal is intermediated (Panel B).

When comparing Columns (1), (3), and (5) between Panels A and B, we find a larger local premium when syndicate leads select startups (LP^3) relative to when investors select startups directly (LP^1). This suggests that proximity to the startup plays a more important role in the screening process done by syndicate leads than in the selection – possibly performed entirely online – done by the crowd. Whereas the crowd only provides capital, syndicate leads also face a reputational cost if a specific deal turns out to be a failure (since they are the ones vetting and endorsing startups and founders). Moreover, in exchange for the carry, leads are expected to perform some combination of due diligence, monitoring, and mentorship. A higher local premium for syndicated deals is consistent with the leads investing in at least some of these activities and having face-to-face interactions with the founders both before and after an investment is made. Proximity between the syndicate leads and the startups is likely to lower search and monitoring costs, reduce asymmetric information, and lower the risk of moral hazard. Furthermore, it increases the chances that high quality entrepreneurs will accept having their company syndicated by a given lead.

As before, when we introduce the share of a syndicate lead's ties in the focal region, the local premium drops substantially (Columns 2, 4 and 6 in Panel B), and turns negative in non-hub regions. Relative to the crowd, when selecting startups, syndicate leads seem to rely more on not only co-location, but also on their professional network (compare Panels A and B). This supports the idea that the syndicate leads' local premium observed before is mostly driven by the geography of their professional network. Professional ties in a region may provide leads with exposure to investment opportunities (lower search costs), better information on founders (lower asymmetric information), and even lower risk of moral hazard (as the entrepreneurs may face a greater reputational cost within their network if the lead is well connected within it). At the same time, the fact that *'Syndicate Lead Same State'* is still positive and significant, suggests that the process of screening and monitoring startups may still benefit from co-location. We will explore this further in Table 10.[14]

Taken together, Table 6 and Table 7 suggest that a possible explanation for why syndication expands access to capital outside of the hubs might be because it allows for a reduction of asymmetric information and better search through professional networks. Prior to syndication, most investors are unlikely to be directly connected to the startups they are considering investing in, or may not feel comfortable enough to invest in the ones that they have ties to. As a result, they are more likely to invest in a startup from a hub region, where the average quality is higher and possibly more observable through external quality signals on the platform.

Furthermore, in the absence of syndication, investors with access to high quality, offline deals do not have any incentive to share them on the platform. Syndication allows them to appropriate the returns from their offline search, due diligence, monitoring and mentoring activities. Meanwhile, in order to take advantage of the syndicates' expertise, the crowd still needs to be able to evaluate them, and investors' professional ties may proxy for its overall ability to effectively do so. As a result, many investors in non-hub regions who used to invest in hub startups switch to using a syndicate lead – who is more likely to be co-located with the startup to take advantage of lower screening and monitoring costs – as their agent. This decreases the concentration of capital going towards hub investments because syndicate leads can now use their reputation to signal the ultimate quality of the startups they will select, irrespective of where the startups are located. If high quality startups are more geographically dispersed than early-stage capital, syndicate leads leverage their professional network to arbitrage opportunities across regions.

[14]Note that the presence of more ties in a specific region is likely to be positively correlated with better access to deals in that area, but also higher ability to select startups in the same region (e.g. because of domain expertise in the sectors the region specializes in). This implies that once we control for professional ties, the remaining local premium is a lower bound on the syndicate leads' comparative advantage in the selection and monitoring of local startups.

Table 7: How Syndicate Leads Select Startups Compared to the Crowd

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Investors Selecting Startups						
	Startup in CA		Startup in NY		Startup in Other	
Investor Same State	0.036 (0.024)	0.025 (0.021)	0.040** (0.018)	0.016 (0.020)	0.025** (0.013)	-0.051** (0.024)
Investor Share Ties Same State		0.053 (0.052)		0.081 (0.060)		0.204*** (0.059)
Adj R-squared	0.001	0.002	0.001	0.001	0.001	0.010
N	4848	3849	4848	3849	4848	3849
Panel B. Syndicate Leads Selecting Startups						
	Startup in CA		Startup in NY		Startup in Other	
Syndicate Lead Same State	0.120*** (0.009)	0.056*** (0.006)	0.035*** (0.005)	0.015*** (0.004)	0.116*** (0.011)	-0.021*** (0.005)
Syndicate Lead Share Ties Same State		0.234*** (0.018)		0.089*** (0.015)		0.354*** (0.030)
Adj R-squared	0.021	0.033	0.002	0.003	0.024	0.043
N	89578	81280	89578	81280	89578	81280
Panel C. t-statistics of the Differences Between Panel A and Panel B						
Variable 1	3.26	-9.52	-0.25	-2.69	5.47	1.21
Variable 2		17.35		32.7		2.27

Notes: Standard errors clustered at the deal level. *** indicates significance at p = 0.01; ** indicates p = 0.05 ; * indicates p = 0.1.

4.3.3 The Role of Monitoring: Heterogeneous Effects by Geographic Distance

The democratization result we have established could come from a reduction in asymmetric information, a reduction in moral hazard or a combination of both. The former is informational and is about the ability to separate, ex-ante, high quality startups from low quality ones. For example, a syndicate lead might be able to assess more accurately the true quality of a startup relative to an investor from the crowd. The latter refers to ex-post costs investors incur to monitor the founders they have invested in. Syndicate leads may have more experience, or may simply experience economies of scale in monitoring startups relative to the crowd.

While we do not observe individuals' motivations and costs, we might be able to find indirect evidence in support of the relative importance of these mechanisms by exploring how our results change with distance. Of course, distance will likely affect both the extent of asymmetric information – since part of the information about a deal may disproportionately flow locally – as well as monitoring costs. Our test should be therefore considered as an imperfect way to separate these two channels. Our assumption is that monitoring requires regular and possibly more frequent interactions than initial due diligence, hence distance related costs should affect monitoring more than asymmetric information.

In Table 8, we explore how the association between syndication and the changes in capital flows we observe varies by distance between investors and startups. If monitoring costs are the key factor driving capital allocation decisions, we would expect the coefficient on *'Syndicated'* to increase as investors are further away from a startup. We adopt a difference-in-differences approach: in even columns, we interact the syndicated dummy with the log distance between states. The interaction term captures whether the democratizing effect of syndication increases with distance (i.e. if the syndication 'technology' is more useful when monitoring costs would otherwise be prohibitively high).

In the table, we do not see statistically significant coefficients on the interaction terms, suggesting that lower monitoring costs through the intermediary may not be the primary mechanism through which syndication is adding value to the marketplace. To account for a possible California–New York "corridor effect" which would work against finding an effect on distance, we repeat the estimation excluding investors in these two states: results are qualitatively unchanged. As an additional robustness, we separate startups based on their observable, ex-ante quality: If monitoring is more valuable for startups with lower initial quality, then we would expect to see a stronger effect of syndication within this subgroup. However, in Table A-3 we do not observe such a pattern. Overall, monitoring costs do not seem to constitute the key friction addressed by intermediaries in this context. In the next section, we therefore turn to observable measures of startup and syndicate lead quality to separate cases where asymmetric information is more versus less likely to be a concern.

4.3.4 The Role of Asymmetric Information: Heterogeneous Effects by Observable Quality of Syndicate Leads and Startups

If the local premium the crowd exhibits under both investment modes is due to information that is more easily available locally and through professional networks (but is otherwise not available on the platform), then as the observable quality of startups and syndicates on AngelList increases, investment decisions by the crowd should be less constrained by geography and pre-existing professional networks. In the extreme case where observable quality is extremely high, information asymmetry should be less of a concern, and investors from all regions should be able to identify a startup or a syndicate lead as a top performer.

We explore this hypothesis through two separate tables: Table 9 focuses on understanding how

Table 8: Heterogeneity by Geographic Distance

	(1)	(2)	(3)	(4)	(5)	(6)
	Startup in CA		Startup in NY		Startup in MA	
Syndicated	-0.107**	-0.113**	-0.001	-0.005	0.024***	0.058***
	(0.048)	(0.047)	(0.042)	(0.043)	(0.007)	(0.011)
Log(Distance)		-0.002		-0.002**		-0.000
		(0.001)		(0.001)		(0.000)
Syndicated ×		-0.002		0.001		-0.005***
Log(Distance)		(0.002)		(0.001)		(0.001)
Adj R-squared	0.002	0.007	-0.000	0.000	0.001	0.007
N	98403	93995	98403	93995	98403	93995

	(1)	(2)	(3)	(4)	(5)	(6)
	Startup in CO		Startup in TX		Startup in IL	
Syndicated	0.042***	0.087***	-0.000	0.001	0.005	0.023
	(0.012)	(0.031)	(0.016)	(0.024)	(0.008)	(0.015)
Log(Distance)		-0.003		-0.001		-0.000
		(0.002)		(0.001)		(0.000)
Syndicated ×		-0.007**		-0.000		-0.002*
Log(Distance)		(0.003)		(0.001)		(0.001)
Adj R-squared	0.002	0.014	-0.000	0.002	0.000	0.003
N	98403	93995	98403	93995	98403	93995

Notes: Standard errors clusterd at the deal level. *** indicates significance at p = 0.01; ** indicates p = 0.05 ; * indicates p = 0.1.

the local premium of the crowd varies with the observable reputation of syndicate leads, and Table 10 uses the same approach but splits the sample by the observable quality of the startups involved. These measures, which are based on profile integrity, completeness, and quality on the platform, should be considered as rough proxies of the true, unobservable quality of syndicate leads and startups.

Consistent with investors from the crowd delegating to syndicate leads based on public information on the platform (e.g. track-record, expertise, etc.) and also on additional information they may be able to retrieve through their networks or direct interactions, the local premium is higher when the observable quality of the syndicate lead is lower. In Table 9, the coefficient for *'Investor in Same State'* is substantially larger in the bottom panel of the table, which only includes syndicate leads in the bottom three quartiles of the observable quality distribution. Syndicate leads with a strong, observable signal of quality attract capital from any US region, while their lesser known counterparts mostly receive funds from local investors who may be able to evaluate them more accurately through complementary offline information.

Table 10 repeats the exercise for startups. The idea is to compare the localization of capital flows in cases where asymmetric information between the founders and the crowd is likely to be higher versus lower. In the table, *'Same State'* refers to the co-location between investors and startups in direct investments, and between syndicate leads and startups in syndicated investments, as these are the relationships that should matter for the screening of startups across the two investment modes. As before, we control for the investors' and syndicate leads' share of ties in the focal states as a proxy for their professional connections to the region. Local premiums are stronger for startups with lower observable signals of quality. Moreover, while this is true for both investment modes, it is actually more pronounced under syndication (compare *'Same State'* and *'Syndicated $\times$ Same State'*). This supports the idea that syndicate leads may be able to extract more information about the startups than online investors, possibly through offline due diligence and face-to-face interactions. As we have seen in the previous tables, this increases the size of the market, allowing capital to be deployed outside of the usual regions.

For startups with high ex-ante signals of quality, the local premium is instead small and insignificant, which is consistent with the idea that when alternative proxies for quality are available on the platform and asymmetric information is lower, online capital flows are less constrained by geography and pre-existing professional networks.

Table 9: Geography of Selecting Syndicate Leads by Syndicate Lead's Observable Quality

Top Quartile of Observable Syndicate Lead Quality			
	Startup in CA	Startup in NY	Startup in Other
Investor in Same State	0.008***	0.017***	0.010***
	(0.002)	(0.005)	(0.003)
Adj R-squared	0.000	0.001	0.001
N	46356	46356	46356
Bottom Three Quartiles of Observable Syndicate Lead Quality			
	Startup in CA	Startup in NY	Startup in Other
Investor in Same State	0.194***	0.054***	0.201***
	(0.008)	(0.007)	(0.013)
Adj R-squared	0.039	0.003	0.044
N	43222	43222	43222

Notes: Standard errors clustered at the deal level. *** indicates significance at p = 0.01.

Table 10: Heterogeneity by Observable Startup Quality

Top Quartile of Observable Startup Quality

	Startup in California	Startup in New York	Startup in Other
Syndicated	-1.026***	-0.270***	0.016***
	(0.072)	(0.053)	(0.033)
Same State	0.011	0.015	0.074
	(0.027)	(0.021)	(0.020)
Syndicated ×	0.002	-0.061	0.236
Same State	(0.052)	(0.047)	(0.068)
Adj R-squared	0.534	0.460	0.581
N	50576	50576	50576

Bottom Three Quartiles of Observable Startup Quality

	Startup in California	Startup in New York	Startup in Other
Syndicated	-1.001***	-0.145***	-0.008
	(0.122)	(0.052)	(0.031)
Same State	0.062**	0.017	0.111***
	(0.029)	(0.040)	(0.043)
Syndicated ×	0.262***	0.139	0.401***
Same State	(0.088)	(0.123)	(0.103)
Adj R-squared	0.094	0.030	0.607
N	39656	39656	39656

Notes: We control for investors' share of ties and syndicate leads' share of ties in the focal states. Standard errors clusterd at the deal level. *** indicates significance at p = 0.01.

4.4 Investment Performance

In this last section, we study how investment performance differs between syndicated and direct investments, as well as how it varies with the observable quality of the startups, syndicate leads and investors involved. For each deal, we calculate if an investment has above versus below the median (unrealized) returns within its cohort using 2017 startup valuation data. This helps us account for the fact that older deals had more time to mature and experience both positive and negative updates in valuation.

In Table 11, we start by analyzing differences in performance across regions. In Column 1 of Panel A, we see that syndicated deals outperform direct investment within non-hub regions, where they are 36.9% more likely to have above the median returns. Column 2 limits the sample to the 10 months before and after the introduction of syndication, and adds month fixed effects. Results are qualitatively unchanged.

In Panel B of Table 11, we further decompose the result by the location of investors and startups. For simplicity, we only split the pairs by hub (California and New York) versus non-hub regions.

According to the table, higher performing deals are disproportionally represented by syndicated deals in non-hub startups, irrespective of the location of the investors from the crowd. This finding is consistent with past literature that has shown that the welfare gains from online marketplaces are often linked to arbitrage across regions and markets (Brynjolfsson and Smith [2000], Forman et al. [2009], Ghose and Yao [2011]).

When we explore heterogeneity by syndicate lead in Table 12, only syndicates of above median observable quality (Column 1) are associated with higher unrealized returns relative to direct investments without an intermediary. The result is again driven by syndicated investments in startups from non-hub regions. Interestingly, this effect is stronger for startups that are – on observables – of below median quality (Column 4). This again supports the idea that syndicate leads do indeed extract more information from founders, possibly through face-to-face interactions and due diligence, relative to what is available to everyone else on the platform. As reflected by the crowd's willingness to pay a share of future returns (the carry) in exchange of the lead services, this information is valuable for the functioning of the online market.

Last, when comparing Columns 5 and 6, we do not find any differences in performance based on how experienced investors from the crowd are,[15] which is consistent with investors having similar ability to search and screen syndicated deals on the platform. Results are robust to the inclusion of month fixed effects (shown in Table A-4).

[15]We define investors as experienced if: 1) they have invested before; or 2) have entrepreneurial experience; or 3) have graduated from a top MBA program according to US News. Different combinations of these variables lead to qualitatively similar results.

Table 11: Investment Outcome by Startup and User Location

Panel A. Outcome by Startup State	(1) All	(2) 20M Window	*Panel B. Outcome by Investor-Startup Region*	(3) All	(4) 20M Window
Startup in California × Syndicated	0.065	0.119	Hub Inv. × Hub Startup × Syndicated	0.037	0.041
	(0.082)	(0.140)		(0.077)	(0.133)
Startup in New York × Syndicated	-0.109	-0.301	Hub Inv. × Non-Hub Startup × Syndicated	0.421***	0.543**
	(0.119)	(0.215)		(0.136)	(0.242)
Startup in Other Region ×Syndicated	0.369***	0.547**	Non-Hub Inv. × Hub Startup × Syndicated	0.043	0.029
	(0.138)	(0.236)		(0.077)	(0.133)
Startup in California	0.21	0.252	Non-Hub Inv. × Non-Hub Startup × Syndicated	0.295*	0.542**
	(0.146)	(0.174)		(0.154)	(0.234)
Startup in New York	0.512***	0.607***	Hub Inv. × Hub Startup	0.219	0.335**
	(0.162)	(0.215)		(0.159)	(0.161)
			Hub Inv. × Non-Hub Startup	-0.074	0.038
				(0.078)	(0.034)
			Non-Hub Inv. × Hub Startup	0.206	0.342**
				(0.159)	(0.160)
Month Fixed Effects		✓	Month Fixed Effects		✓
Adj R(squared	0.015	0.081	Adj R(squared	0.006	0.071
N	85966	13263	N	86231	13263

Notes: Standard errors are clustered at the deal level. *** indicates significance at p = 0.01; ** indicates p = 0.05 ; * indicates p = 0.1.

Table 12: Outcome by Syndicate Lead Quality, Startup Quality, and Investor Experience

	Syndicate Lead Quality		Startup Quality		Investor Experience	
	Above Median	Below Median	Above Median	Below Median	Low	High
	(1)	(2)	(3)	(4)	(5)	(6))
Startup in California × Syndicated	0.04	0.1	-0.328**	0.168*	0.075	0.075
	(0.092)	(0.083)	(0.137)	(0.091)	(0.082)	(0.085)
Startup in New York × Syndicated	-0.004	-0.221*	-0.239	-0.061	-0.112	-0.116
	(0.129)	(0.129)	(0.177)	(0.144)	(0.121)	(0.125)
Startup in Other Region ×Syndicated	0.617***	0.227	0.242	0.445***	0.444***	0.469***
	(0.141)	(0.141)	(0.222)	(0.166)	(0.124)	(0.115)
Startup in California	0.21	0.21	0.227	0.243	0.269**	0.291**
	(0.146)	(0.146)	(0.233)	(0.173)	(0.132)	(0.126)
Startup in New York	0.512***	0.512***	0.453*	0.556***	0.580***	0.609***
	(0.163)	(0.163)	(0.233)	(0.198)	(0.151)	(0.147)
Adj R-squared	0.088	0.011	0.112	0.011	0.016	0.018
N	46,670	42,583	17,169	68,797	67,747	41,333

Notes: Standard errors are clustered at the deal level. *** indicates significance at p = 0.01; ** indicates p = 0.05 ; * indicates p = 0.1.

5 Conclusions

We document the changes in capital flows that follow the introduction of intermediaries on the leading US equity crowdfunding platform. Using novel data on investment reservations, individuals and startups, we show that intermediaries allow online capital to be allocated outside of traditional startup hubs, expanding the set of founders that can rely on this new source of early-stage capital to fund their startup. After the change in market design, we also find that the deals from non-hub regions that are surfaced by intermediaries are associated with higher performance relative to other deals on the platform. This is consistent with syndicate leads taking advantage of their expertise and professional network to arbitrage investment opportunities across US regions, ultimately connecting less competitive markets for early-stage capital with an active online crowd of potential investors.

Our paper has a number of limitations. In particular, whereas the introduction of syndication on the platform can be considered as a natural experiment for the pre-existing crowd of investors, we have to rely on investor observables to build a credible control group for the individuals that use syndication. Although the evidence we provide is strongly consistent with a causal interpretation, unobserved heterogeneity may still affect our estimates. Moreover, our ability to identify the exact mechanisms at work is also limited, as we do not observe true investor, startup and syndicate lead quality, but have to rely on imperfect proxies for them. Similarly, we cannot measure offline due diligence and monitoring activities directly, but have to rely on geographic distance to partially separate the role they play in the patterns we observe.

Nevertheless, given the increasing importance of this new source of early-stage capital for high growth startups, we believe our findings are useful for understanding the matching process taking place on these platforms, and for identifying how professional investors can contribute through their access to deal flow and expertise to the functioning and scaling of these online markets beyond hub regions. Our geography results also have the potential to inform policy, as they highlight the conditions under which online capital is more versus likely to flow to regions that do not have a strong angel and venture capital presence. On the one hand, we find that the platform is able to expand access to capital to high quality entrepreneurs from non-hub regions that would possibly not have been funded otherwise. On the other hand, we also show that this result rests on the availability of intermediaries familiar with the target region to begin with: In the absence of co-located syndicate leads, or at least leads that have pre-established, professional ties to a region, capital is unlikely to flow outside of traditional entrepreneurial hubs, reinforcing agglomeration.

References

Ajay Agrawal, Christian Catalini, and Avi Goldfarb. Some simple economics of crowdfunding. *Innovation Policy and the Economy*, 14(1):63–97, 2014.

Ajay Agrawal, Christian Catalini, and Avi Goldfarb. Crowdfunding: Geography, social networks, and the timing of investment decisions. *Journal of Economics & Management Strategy*, 24(2): 253–274, 2015.

Ajay Agrawal, Christian Catalini, and Avi Goldfarb. Are syndicates the killer app of equity crowdfunding? *California management review*, 58(2):111–124, 2016a.

Ajay Agrawal, Christian Catalini, Avi Goldfarb, and Hong Luo. Slack time and innovation. *SSRN Working Paper No. 2599004*, 2016b.

Paul Belleflamme, Thomas Lambert, and Armin Schwienbacher. Crowdfunding: Tapping the right crowd. *Journal of business venturing*, 29(5):585–609, 2014.

Bernardo S Blum and Avi Goldfarb. Does the internet defy the law of gravity? *Journal of international economics*, 70(2):384–405, 2006.

Erik Brynjolfsson and Michael D Smith. Frictionless commerce? a comparison of internet and conventional retailers. *Management science*, 46(4):563–585, 2000.

Erik Brynjolfsson, Yu Hu, and Mohammad S Rahman. Battle of the retail channels: How product selection and geography drive cross-channel competition. *Management Science*, 55(11):1755–1765, 2009.

Gordon Burtch, Anindya Ghose, and Sunil Wattal. An empirical examination of the antecedents and consequences of contribution patterns in crowd-funded markets. *Information Systems Research*, 24(3):499–519, 2013.

Henry Chen, Paul Gompers, Anna Kovner, and Josh Lerner. Buy local? the geography of venture capital. *Journal of Urban Economics*, 67(1):90–102, 2010.

Ian Cooper and Evi Kaplanis. Home bias in equity portfolios, inflation hedging, and international capital market equilibrium. *Review of Financial Studies*, 7(1):45–60, 1994.

Joshua D Coval and Tobias J Moskowitz. Home bias at home: Local equity preference in domestic portfolios. *The Journal of Finance*, 54(6):2045–2073, 1999.

Joshua D Coval and Tobias J Moskowitz. The geography of investment: Informed trading and asset prices. *Journal of political Economy*, 109(4):811–841, 2001.

Douglas Cumming and Na Dai. Local bias in venture capital investments. *Journal of Empirical Finance*, 17(3):362–380, 2010.

Douglas J Cumming and Sofia A Johan. *Venture capital and private equity contracting: An international perspective*. Academic Press, 2013.

Chris Forman, Anindya Ghose, and Avi Goldfarb. Competition between local and electronic markets: How the benefit of buying online depends on where you live. *Management science*, 55(1): 47–57, 2009.

Seth M Freedman and Ginger Zhe Jin. Learning by doing with asymmetric information: evidence from prosper. com. Technical report, National Bureau of Economic Research, 2011.

Kenneth R French and James M Poterba. Investor diversification and international equity markets. Technical report, National Bureau of Economic Research, 1991.

Anindya Ghose and Yuliang Yao. Using transaction prices to re-examine price dispersion in electronic markets. *Information Systems Research*, 22(2):269–288, 2011.

Paul A Gompers. Optimal investment, monitoring, and the staging of venture capital. *The journal of finance*, 50(5):1461–1489, 1995.

John R Graham, Campbell R Harvey, and Hai Huang. Investor competence, trading frequency, and home bias. *Management Science*, 55(7):1094–1106, 2009.

Keith Hampton and Barry Wellman. Neighboring in netville: How the internet supports community and social capital in a wired suburb. *City & Community*, 2(4):277–311, 2003.

Yael V Hochberg, Alexander Ljungqvist, and Yang Lu. Whom you know matters: Venture capital networks and investment performance. *The Journal of Finance*, 62(1):251–301, 2007.

Ali Hortaçsu, F Martínez-Jerez, and Jason Douglas. The geography of trade in online transactions: Evidence from ebay and mercadolibre. *American Economic Journal: Microeconomics*, 1(1):53–74, 2009.

David H Hsu. Experienced entrepreneurial founders, organizational capital, and venture capital funding. *Research Policy*, 36(5):722–741, 2007.

Gur Huberman. Familiarity breeds investment. *Review of financial Studies*, 14(3):659–680, 2001.

Keongtae Kim and Il-Horn Hann. Crowdfunding and the democratization of access to capital: a geographical analysis. *Robert H. Smith School Research Paper. Available at SSRN: http://ssrn.com/abstract*, 2334590, 2014.

Josh Lerner. Venture capitalists and the oversight of private firms. *The Journal of Finance*, 50(1): 301–318, 1995.

Joshua Lerner. The syndication of venture capital investments. *Financial management*, pages 16–27, 1994.

Mingfeng Lin and Siva Viswanathan. Home bias in online investments: An empirical study of an online crowdfunding market. *Management Science*, 62(5):1393–1414, 2015.

Ethan Mollick. The dynamics of crowdfunding: An exploratory study. *Journal of business venturing*, 29(1):1–16, 2014.

Ethan Mollick and Ramana Nanda. Wisdom or madness? comparing crowds with expert evaluation in funding the arts. *Management Science*, 62(6):1533–1553, 2015.

Ethan R Mollick. Swept away by the crowd? crowdfunding, venture capital, and the selection of entrepreneurs. 2013.

Ramana Nanda and Tarun Khanna. Diasporas and domestic entrepreneurs: Evidence from the indian software industry. *Journal of Economics & Management Strategy*, 19(4):991–1012, 2010.

Eric Overby and Chris Forman. The effect of electronic commerce on geographic purchasing patterns and price dispersion. *Management Science*, 61(2):431–453, 2014.

Simon C Parker. *The economics of entrepreneurship*. Cambridge University Press, 2009.

Paul R Rosenbaum and Donald B Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, pages 41–55, 1983.

Mark S Seasholes and Ning Zhu. Individual investors and local bias. *The Journal of Finance*, 65 (5):1987–2010, 2010.

Garrett T Senney. The geography of bidder behavior in peer-to-peer lending markets. Technical report, mimeo, 2016.

Olav Sorenson and Toby E Stuart. Syndication networks and the spatial distribution of venture capital investments 1. *American journal of sociology*, 106(6):1546–1588, 2001.

Olav Sorenson and Toby E Stuart. Bringing the context back in: Settings and the search for syndicate partners in venture capital investment networks. *Administrative Science Quarterly*, 53 (2):266–294, 2008.

Olav Sorenson, Valentina Assenova, Guan-Cheng Li, Jason Boada, and Lee Fleming. Expand innovation finance via crowdfunding. *Science*, 354(6319):1526–1528, 2016.

Stijn Van Nieuwerburgh and Laura Veldkamp. Information immobility and the home bias puzzle. *The Journal of Finance*, 64(3):1187–1215, 2009.

Nir Vulkan, Thomas Åstebro, and Manuel Fernandez Sierra. Equity crowdfunding: A new phenomena. *Journal of Business Venturing Insights*, 5:37–49, 2016.

Juanjuan Zhang and Peng Liu. Rational herding in microloan markets. *Management science*, 58 (5):892–912, 2012.

6 Appendix

Table A-1: Robustness: Introduction of Syndication, Controlling for Month Fixed Effects

	(1)	(2)	(3)	(4)
Panel A. Dependent Variable: Startup from California (1/0)				
Syndicated	-0.121**	-0.115**	-0.365***	-0.161***
	(0.055)	(0.055)	(0.064)	(0.054)
Investor in Same State	0.068***	0.036***	0.020***	0.003
	(0.007)	(0.005)	(0.005)	(0.003)
Investor Share of Ties in Same State		0.119***	0.038***	0.013*
		(0.015)	(0.009)	(0.007)
Syndicate Lead in Same State			0.372***	0.132***
			(0.049)	(0.040)
Syndicate Lead Share of Ties in Same State				1.268***
				(0.081)
Month Fixed Effects	✓	✓	✓	✓
Adj R-squared	0.067	0.070	0.156	0.533
N	93087	83938	83533	69631
Panel B. Dependent Variable: Startup from New York (1/0)				
Syndicated	-0.055	-0.058	-0.072	-0.058
	(0.043)	(0.045)	(0.044)	(0.038)
Investor in Same State	0.022***	0.016***	0.013***	0.003
	(0.004)	(0.004)	(0.004)	(0.003)
Investor Share of Ties in Same State		0.033***	0.025**	0.008
		(0.012)	(0.010)	(0.008)
Syndicate Lead in Same State			0.109*	-0.003
			(0.065)	(0.041)
Syndicate Lead Share of Ties in Same State				1.573***
				(0.125)
Month Fixed Effects	✓	✓	✓	✓
Adj R-squared	0.110	0.115	0.122	0.525
N	93087	83938	83533	69631
Panel C. Dependent Variable: Startup from Other Region (1/0)				
Syndicated	0.170***	0.166***	0.127***	0.042*
	(0.039)	(0.037)	(0.038)	(0.023)
Investor in Same State	0.799***	0.148***	-0.059**	-0.029
	(0.024)	(0.031)	(0.026)	(0.021)
Investor Share of Ties in Same State		1.858***	1.045***	0.484***
		(0.112)	(0.145)	(0.081)
Syndicate Lead in Same State			0.767***	0.221**
			(0.030)	(0.095)
Syndicate Lead Share of Ties in Same State				1.564***
				(0.126)
Month Fixed Effects	✓	✓	✓	✓
Adj R-squared	0.099	0.130	0.260	0.644
N	93087	83938	83533	69749

Notes: Standard errors clusterd at the deal level. *** indicates significance at p = 0.01; ** indicates p = 0.05 ; * indicates p = 0.1.

Table A-2: Syndicated Investment: Using Deep Lag of Professional Ties

	(1)	(2)	(3)	(4)
Panel A. Dependent Variable: Startup in California				
Syndicated	-0.150***	-0.146***	-0.412***	-1.000***
	(0.049)	(0.052)	(0.062)	(0.084)
Investor in Same State	0.063***	0.035***	0.014**	0.001
	(0.008)	(0.007)	(0.006)	(0.005)
Investor Share of Ties in Same State		0.100***	0.038***	0.017**
		(0.014)	(0.011)	(0.008)
Syndicate Lead in Same State			0.332***	0.157***
			(0.052)	(0.054)
Syndicate Lead Share of Ties in Same State				1.161***
				(0.127)
Adj R-squared	0.007	0.008	0.081	0.415
N	93087	77284	76916	58629
Panel B. Startup in New York				
Syndicated	-0.000	-0.003	-0.012	-0.234***
	(0.045)	(0.048)	(0.049)	(0.047)
Investor in Same State	0.025***	0.014***	0.011**	-0.000
	(0.005)	(0.005)	(0.005)	(0.004)
Investor Share of Ties in Same State		0.038**	0.029**	0.017
		(0.015)	(0.014)	(0.011)
Syndicate Lead in Same State			0.109*	0.000
			(0.065)	(0.047)
Syndicate Lead Share of Ties in Same State				1.601***
				(0.132)
Adj R-squared	0.000	0.000	0.006	0.423
N	93087	77284	76916	58629
Panel C. Startup in Other Region				
Syndicated	0.144***	0.146***	0.116***	0.028
	(0.024)	(0.023)	(0.024)	(0.018)
Investor in Same State	0.816***	0.156***	-0.084***	-0.032
	(0.020)	(0.033)	(0.027)	(0.030)
Investor Share of Ties in Same State		1.882***	1.124***	0.468***
		(0.103)	(0.149)	(0.094)
Syndicate Lead in Same State			0.753***	0.332***
			(0.029)	(0.103)
Syndicate Lead Share of Ties in Same State				1.579***
				(0.141)
Adj R-squared	0.054	0.082	0.213	0.634
N	93087	77284	76916	58629

Notes: We use reciprocal follows in the six month before reservation date. Standard errors clusterd at the deal level.

*** indicates significance at p = 0.01; ** indicates p = 0.05 ; * indicates p = 0.1.

Table A-3: Heterogeneity by Geographic Distance

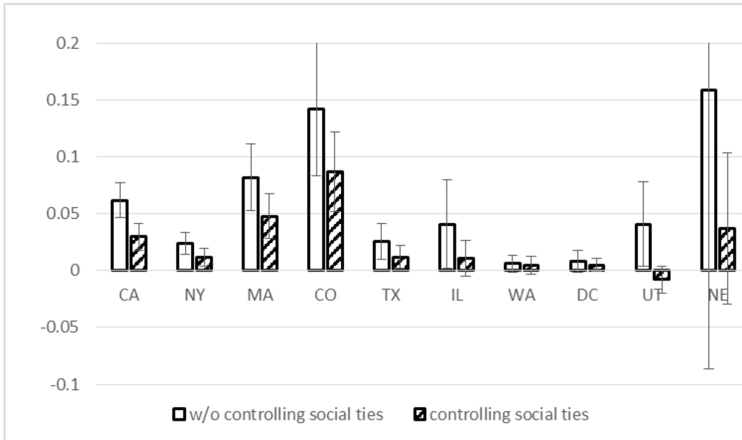
Panel A. Above Median Startup Quality

	Startup in CA		Startup in NY		Startup in MA	
	(1)	(2)	(3)	(4)	(5)	(6)
Syndicated	-0.126*	-0.127*	-0.020	-0.028	0.030**	0.055***
	(0.065)	(0.065)	(0.056)	(0.058)	(0.013)	(0.017)
Log(Distance)		-0.004***		-0.003**		0.000
		(0.001)		(0.001)		(0.000)
Syndicated ×		-0.000		0.002		-0.003***
Log(Distance)		(0.002)		(0.001)		(0.001)
Adj R-squared	0.004	0.009	0.000	0.000	0.002	0.004
N	37056	35361	37056	35361	37056	35361

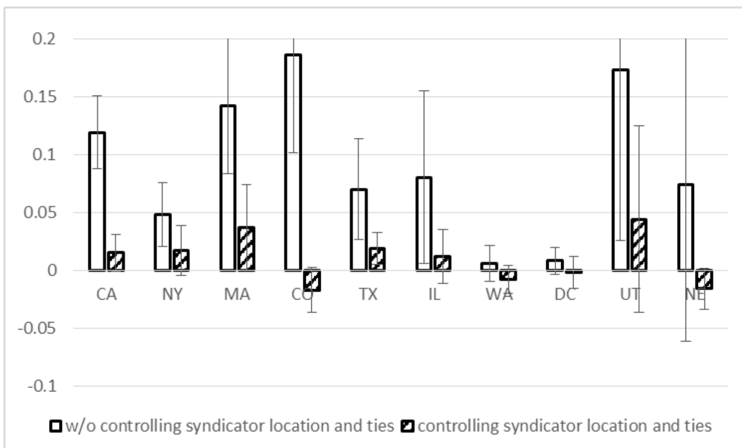
Panel B Below Median Startup Quality

	Startup in CA		Startup in NY		Startup in MA	
	(1)	(2)	(3)	(4)	(5)	(6)
Syndicated	-0.097	-0.107	0.014	0.013	0.020***	0.059***
	(0.070)	(0.068)	(0.062)	(0.063)	(0.007)	(0.015)
Log(Distance)		-0.000		-0.002		-0.001
		(0.002)		(0.002)		(0.001)
Syndicated ×		-0.003		0.001		-0.005***
Log(Distance)		(0.002)		(0.002)		(0.001)
	0.002	0.005				
Adj R-squared	0.002	0.005	0.000	0.001	0.001	0.011
N	61347	58634	61347	58634	61347	58634

(a) Investors' Local Premium With vs. Without Network Controls



(b) Syndicate Leads' Local Premium With vs. Without Network Controls



(c) Investor Share of Ties in Same State With vs. Without Network Controls

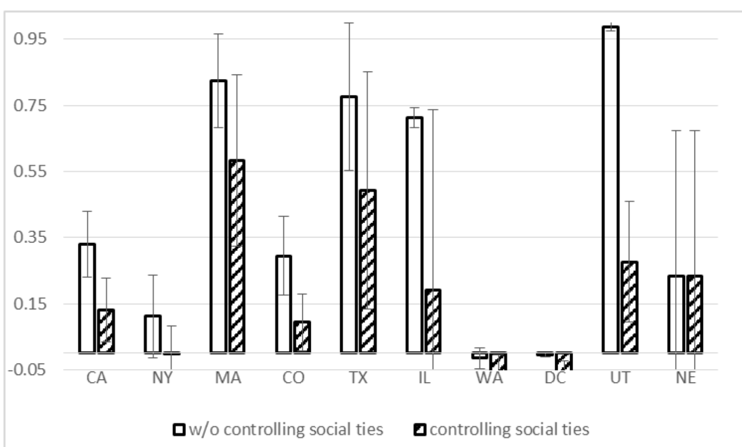


Figure A-1: Local Premiums and Professional Ties

Notes: Figure 2(a) corresponds to the coefficients on *'Investor in Same State'* from Column (1) to (2) in Table 3. Figure 2(b) corresponds to the coefficients on *'Syndicate Lead in Same State'* from Columns (3) to (4). Figure 2(c) corresponds to the coefficients on *'Investor Share of Ties in Same State'* from Columns (2) to (4).

Table A-4: Robustness: Outcome by Syndicate Lead Quality, Startup Quality, and Investor Experience

	(1)	(2)	(3)	(4)	(5)	(6)
	Syndicate Lead Quality		Startup Quality		Investor Experience	
	> Median	< Median	> Median	< Median	> Median	< Median
Startup in CA ×	-0.005	0.029	-0.388***	0.103	0.031	0.037
Syndicated	-0.096	-0.098	-0.131	-0.105	-0.092	-0.095
Startup in NY ×	-0.149	-0.376***	-0.202	-0.2	-0.248**	-0.274**
Syndicated	-0.131	-0.131	-0.168	-0.153	-0.121	-0.128
Startup in Other	0.530***	0.122	0.238	0.362*	0.346**	0.414**
×Syndicated	-0.18	-0.178	-0.25	-0.199	-0.162	-0.163
Startup in CA	0.122	0.165	0.276	0.209	0.197	0.255*
	-0.16	-0.164	-0.237	-0.189	-0.148	-0.15
Startup in NY	0.475***	0.560***	0.409	0.576***	0.572***	0.662***
	-0.174	-0.178	-0.261	-0.203	-0.167	-0.166
Month Fixed Effects	✓	✓	✓	✓	✓	✓
Adj R-squared	0.174	0.077	0.249	0.061	0.065	0.071
N	46651	42564	17150	68797	67728	41327

Notes: Standard errors are clustered at the deal level. *** indicates significance at p = 0.01; ** indicates p = 0.05 ; * indicates p = 0.1.

6.1 Formal Proofs and Model Extension

6.1.1 Formal Proofs of Hub Premium and Local Premium in Proposition 1

We first show the existence of a hub premium (HP) for investors both in hub and non-hub regions under some parameter restrictions. For investors in the hub region, the probability of investing in the hub region is

$$
\begin{aligned}
Pr(H|L_i = H) &= Pr\left(\gamma(n^H) > \gamma(n^{NH}) - \kappa_1 | L_i = H\right) \\
&\quad + Pr\left(\gamma(n^{NH}) - \kappa_1 < \rho \bigcap \gamma(n^H) < \gamma(n^{NH}) - \kappa_1 | L_i = H\right) \\
&> Pr\left(\gamma(n^H) > \gamma(n^{NH}) | L_i = H\right) \\
&\quad + Pr\left(\gamma(n^{NH}) < \rho \bigcap \gamma(n^H) < \gamma(n^{NH}) | L_i = H\right) \\
&= Pr(n^H > n^{NH} | L_i = H) \\
&\quad + Pr(\gamma(n^{NH}) < \rho | L_i = H, n^H < n^{NH}) Pr(n^H < n^{NH} | L_i = H) \\
&= \frac{1}{2}(1 + \Delta_1) + \gamma^{-1}(\rho) \left(1 - \frac{1}{2}(1 + \Delta_1)\right) \\
&= \frac{1}{2} + \frac{1}{2}\Delta_1 + \frac{1}{2}\gamma^{-1}(\rho) - \frac{1}{2}\Delta_1 \gamma^{-1}(\rho)
\end{aligned}
$$

Note that the second last relationship comes from the following two derivations:

$$
\begin{aligned}
Pr(n^H > n^{NH} | L_i = H) &= Pr(n^H > n^{NH} | L_i = H, n^{NH} > \Delta_1) Pr(n^{NH} > \Delta_1 | L_i = H) \\
&\quad + Pr(n^H > n^{NH} | L_i = H, n^{NH} \leq \Delta_1) Pr(n^{NH} \leq \Delta_1 | L_i = H) \\
&= \frac{1}{2}(1 - \Delta_1) + \Delta_1 \\
&= \frac{1}{2}(1 + \Delta_1).
\end{aligned}
$$

and

$$
\begin{aligned}
Pr(\gamma(n^{NH}) < \rho | L_i = H, n^H < n^{NH}) &= Pr(n^{NH} < \gamma^{-1}(\rho) | L_i = H, n^H < n^{NH}) \\
&\quad * Pr(n^H < n^{NH} | L_i = H) \\
&= \gamma^{-1}(\rho) \left(1 - \frac{1}{2}(1 + \Delta_1)\right).
\end{aligned}
$$

On the other hand, the probability of investing in a non-hub region for an investor in a hub region is given as:

$$\begin{aligned} Pr(NH|L_i = H) &= 1 - Pr(H|L_i = H) \\ &= \frac{1}{2} - \frac{1}{2}\Delta_1 - \frac{1}{2}\gamma^{-1}(\rho) + \frac{1}{2}\Delta_1\gamma^{-1}(\rho) \end{aligned}$$

To show the existence of HP for investors in hub regions, we need to show that $Pr(H|L_i = H) - Pr(NH|L_i = H) > 0$. It is enough to show that $Pr(H|L_i = H) > \frac{1}{2}$:

$$\begin{aligned} Pr(H|L_i = H) &> \frac{1}{2} + \frac{1}{2}\Delta_1 + \frac{1}{2}\gamma^{-1}(\rho) - \frac{1}{2}\Delta_1\gamma^{-1}(\rho) \\ &= \frac{1}{2} + \frac{1}{2}\Delta_1(1 - \gamma^{-1}(\rho)) + \frac{1}{2}\gamma^{-1}(\rho) \\ &> 0 \end{aligned}$$

Having established the existence of a hub premium for investors in hub regions, we show the existence of a hub premium for investors in non-hub regions. The probability of non-hub investors investing in a hub region is:

$$\begin{aligned} Pr(H|L_i = NH) &= Pr\left(\gamma(n^H) - \kappa_1 > \gamma(n^{NH})|L_i = NH\right) \\ &\quad + Pr\left(\gamma(n^{NH}) < \rho \bigcap \gamma(n^H) - \kappa_1 < \gamma(n^{NH})|L_i = NH\right) \\ &= Pr(n^H > n^{NH} + \eta(\kappa_1)|L_i = NH) \\ &\quad + Pr(\gamma(n^{NH}) < \rho|L_i = NH, n^H < n^{NH})Pr(n^H < n^{NH}|L_i = NH) \\ &= 1 - \left[(\frac{1}{2} + \eta(\kappa_1))(1 - \Delta_2) + \Delta_2\right] + \gamma^{-1}(\rho)\frac{1}{2}(1 + \Delta_2) \\ &= \frac{1}{2} - \frac{1}{2}\Delta_2 - \eta(\kappa_1) + \Delta_2\eta(\kappa_1) + \frac{1}{2}\gamma^{-1}(\rho) + \frac{1}{2}\gamma^{-1}(\rho)\Delta_2 \end{aligned}$$

Note that the last relationship comes from the following two derivations:

$$\begin{aligned}
Pr(n^{NH}+\eta(\kappa_1)>n^H|L_i=NH) &= Pr(n^{NH}+\eta(\kappa_1)>n^H|L_i=NH,n^H>\Delta_2) \\
&\quad *Pr(n^H>\Delta_2|L_i=NH) \\
&\quad +Pr(n^{NH}+\eta(\kappa_1)|>n^H|L_i=NH,n^H\leq\Delta_2) \\
&\quad *Pr(n^H\leq\Delta_2|L_i=NH) \\
&= (\frac{1}{2}+\eta(\kappa_1))(1-\Delta_2)+\Delta_2
\end{aligned}$$

and

$$\begin{aligned}
Pr(\gamma(n^{NH})<\rho|L_i=NH,n^H<n^{NH}) &= Pr(n^{NH}<\gamma^{-1}(\rho)|L_i=NH,n^H<n^{NH}) \\
&\quad *Pr(n^H<n^{NH}|L_i=NH) \\
&= \gamma^{-1}(\rho)\frac{1}{2}(1+\Delta_2).
\end{aligned}$$

On the other hand, the probability of investing in a non-hub region for an investor in a non-hub region is given as:

$$\begin{aligned}
Pr(NH|L_i=NH) &= 1-Pr(H|L_i=NH) \\
&= \frac{1}{2}+\frac{1}{2}\Delta_2+\eta(\kappa_1)-\Delta_2\eta(\kappa_1)-\frac{1}{2}\gamma^{-1}(\rho)-\frac{1}{2}\gamma^{-1}(\rho)\Delta_2
\end{aligned}$$

To show the existence of HP for investors in non-hub regions, we need to show that $Pr(H|L_i=NH)-Pr(NH|L_i=NH)>0$. It is enough to show that $Pr(H|L_i=NH)>\frac{1}{2}$ if the average quality difference between hub and non-hub deals is not too small:

$$\begin{aligned}
Pr(H|L_i=NH) &= \frac{1}{2}-\frac{1}{2}\Delta_2-\eta(\kappa_1)+\Delta_2\eta(\kappa_1)+\frac{1}{2}\gamma^{-1}(\rho)+\frac{1}{2}\gamma^{-1}(\rho)\Delta_2 \\
&> \frac{1}{2} \\
&if \quad \rho>\gamma\left(\frac{\Delta_2+2\eta(\kappa_1)(1-\Delta_2)}{1+\Delta_2}\right)
\end{aligned}$$

Having established a hub premium for investors both in hub and non-hub regions, we now show the existence of a local premium (LP^1) for deals in hub and non-hub regions under some parameter restrictions. For deals in a hub region, the difference between the probability of them attracting local versus distant investors is:

$$\begin{aligned}
Pr(H|L_i = H) - Pr(H|L_i = NH) &= \frac{1}{2} + \frac{1}{2}\Delta_1 + \frac{1}{2}\gamma^{-1}(\rho) - \frac{1}{2}\Delta_1\gamma^{-1}(\rho) \\
&\quad - \left[\frac{1}{2} - \frac{1}{2}\Delta_2 - \eta(\kappa_1) + \Delta_2\eta(\kappa_1) + \frac{1}{2}\gamma^{-1}(\rho) + \frac{1}{2}\gamma^{-1}(\rho)\Delta_2\right] \\
&= \frac{1}{2}\Delta_1(1 - \gamma^{-1}(\rho)) + \frac{1}{2}\Delta_2(1 - \gamma^{-1}(\rho)) + (1 - \Delta_2)\eta(\kappa_1) \\
&> 0
\end{aligned}$$

For deals in non-hub region, the difference between the probability of attracting local versus distant investors is:

$$\begin{aligned}
Pr(NH|L_i = NH) - Pr(NH|L_i = H) &= \frac{1}{2} + \frac{1}{2}\Delta_2 + \eta(\kappa_1) - \Delta_2\eta(\kappa_1) - \frac{1}{2}\gamma^{-1}(\rho) - \frac{1}{2}\gamma^{-1}(\rho)\Delta_2 \\
&\quad - \left[\frac{1}{2} - \frac{1}{2}\Delta_1 - \frac{1}{2}\gamma^{-1}(\rho) + \frac{1}{2}\Delta_1\gamma^{-1}(\rho)\right] \\
&= \frac{1}{2}\Delta_2(1 - \gamma^{-1}(\rho)) + (1 - \Delta_2)\eta(\kappa_1) + \frac{1}{2}\Delta_1(1 + \gamma^{-1}(\rho)) \\
&> 0
\end{aligned}$$

6.1.2 Formal Proofs of Hub Premium and Local Premium in Proposition 2

First, we show the existence of a local premium (LP^2) when investors choose syndicate leads in both hub and non-hub regions under some parameter restrictions. For syndicate leads in a hub region, the difference between the probability of them being chosen by local versus distant investors is:

$$\begin{aligned}
Pr(H|L_i = H) - Pr(H|L_i = NH) &= \frac{1}{2} + \frac{1}{2}\Delta_1 + \frac{1}{2}\tilde{\gamma}^{-1}(\rho) - \frac{1}{2}\Delta_1\tilde{\gamma}^{-1}(\rho) \\
&\quad - \left[\frac{1}{2} - \frac{1}{2}\Delta_2 - \eta(\kappa_2) + \Delta_2\eta(\kappa_2) + \frac{1}{2}\tilde{\gamma}^{-1}(\rho) + \frac{1}{2}\tilde{\gamma}^{-1}(\rho)\Delta_2\right] \\
&= \frac{1}{2}\Delta_1(1 - \tilde{\gamma}^{-1}(\rho)) + \frac{1}{2}\Delta_2(1 - \tilde{\gamma}^{-1}(\rho)) + (1 - \Delta_2)\eta(\kappa_2) \\
&> 0
\end{aligned}$$

For syndicate leads in non-hub region, the difference between the probability of them being selected by local versus distant investors is:

$$\begin{aligned}
Pr(NH|L_i = NH) - Pr(NH|L_i = H) &= \frac{1}{2} + \frac{1}{2}\Delta_2 + (\kappa_2) - \Delta_2\eta(\kappa_2) - \frac{1}{2}\tilde{\gamma}^{-1}(\rho) - \frac{1}{2}\tilde{\gamma}^{-1}(\rho)\Delta_2 \\
&\quad - \left[\frac{1}{2} - \frac{1}{2}\Delta_1 - \frac{1}{2}\tilde{\gamma}^{-1}(\rho) + \frac{1}{2}\Delta_1\tilde{\gamma}^{-1}(\rho)\right] \\
&= \frac{1}{2}\Delta_2(1 - \tilde{\gamma}^{-1}(\rho)) + (1 - \Delta_2)\eta(\kappa_2) + \frac{1}{2}\Delta_1(1 + \tilde{\gamma}^{-1}(\rho)) \\
&> 0
\end{aligned}$$

Having showed the existence of a local premium (LP^2) when investors choose syndicate leads in both hub and non-hub regions, we now show that syndicate leads exhibit a local premium (LP^3) in their selection of startups. The proof follows the same structure used before in deriving LP^1.

Last, we show that the hub premium is smaller for syndicated investments relative to direct investments. For investors in a hub region, the probability of investing in a hub region under direct investment is:

$$\begin{aligned}
Pr^{Dir}(H|L_i = H) &= Pr\left(\gamma(n^H) > \gamma(n^{NH}) - \kappa_1 | L_i = H\right) \\
&\quad + Pr\left((n^{NH}) - \kappa_1 < \rho \bigcap \gamma(n^H) < \gamma(n^{NH}) - \kappa_1 | L_i = H\right) \\
&= (\frac{1}{2} + \eta(\kappa_1))(1 - \Delta_1) + \Delta_1 + \gamma^{-1}(\rho + \kappa_1)\frac{1}{2}(1 + \Delta_1)
\end{aligned}$$

Note that the last relationship comes from the following two derivations:

$$\begin{aligned}
Pr^{Dir}\left(\gamma(n^H) > \gamma(n^{NH}) - \kappa_1 | L_i = H\right) &= Pr(n^H > n^{NH} - \eta(\kappa_1) | L_i = H) \\
&= Pr(n^H > n^{NH} - \eta(\kappa_1) | L_i = H, n^{NH} > \Delta_1) \\
&\quad * Pr(n^{NH} > \Delta_1 | L_i = H) \\
&\quad + Pr(n^H > n^{NH} - \eta(\kappa_1) | L_i = H, n^{NH} \leq \Delta_1) \\
&\quad * Pr(n^{NH} \leq \Delta_1 | L_i = H) \\
&= (\frac{1}{2} + \eta(\kappa_1))(1 - \Delta_1) + \Delta_1
\end{aligned}$$

and

$$\begin{aligned} Pr^{Dir}\left((n^{NH})-\kappa_1<\rho\bigcap\gamma(n^H)<\gamma(n^{NH})-\kappa_1|L_i=H\right) &= Pr(n^{NH}<\gamma^{-1}(\rho+\kappa_1)|L_i=H,n^H<n^{NH}) \\ &\quad *Pr(n^H<n^{NH}|L_i=H) \\ &= \gamma^{-1}(\rho+\kappa_1)\frac{1}{2}(1+\Delta_1). \end{aligned}$$

Now, for investors in a hub region, the probability of investing in a hub region under syndicated investment is:

$$\begin{aligned} Pr^{Synd}(H|L_i=H) &= Pr\left(\tilde{\gamma}(n^H)>\tilde{\gamma}(n^{NH})-\kappa_2|L_i=H\right) \\ &\quad +Pr\left(\tilde{\gamma}(n^{NH})-\kappa_2<\rho\bigcap\tilde{\gamma}(n^H)<\tilde{\gamma}(n^{NH})-\kappa_2|L_i=H\right) \\ &= (\frac{1}{2}+\eta(\kappa_2))(1-\Delta_1)+\Delta_1+\tilde{\gamma}^{-1}(\rho+\kappa_2)\frac{1}{2}(1+\Delta_1) \end{aligned}$$

Note that the last relationship comes from:

$$\begin{aligned} Pr^{Synd}\left(\tilde{\gamma}(n^H)>\tilde{\gamma}(n^{NH})-\kappa_2|L_i=H\right) &= Pr(n^H>n^{NH}-\eta(\kappa_2)|L_i=H) \\ &= Pr(n^H>n^{NH}-\eta(\kappa_2)|L_i=H,n^{NH}>\Delta_1) \\ &\quad *Pr(n^{NH}>\Delta_1|L_i=H) \\ &\quad +Pr(n^H>n^{NH}-\eta(\kappa_2)|L_i=H,n^{NH}\leq\Delta_1) \\ &\quad *Pr(n^{NH}\leq\Delta_1|L_i=H) \\ &= (\frac{1}{2}+\eta(\kappa_1))(1-\Delta_1)+\Delta_1 \end{aligned}$$

and

$$\begin{aligned} Pr^{Synd}\left((n^{NH})-\kappa_2<\rho\bigcap\tilde{\gamma}(n^H)<\tilde{\gamma}(n^{NH})-\kappa_2|L_i=H\right) &= Pr(n^{NH}<\tilde{\gamma}^{-1}(\rho+\kappa_2)|L_i=H,n^H<n^{NH}) \\ &\quad *Pr(n^H<n^{NH}|L_i=H) \\ &= \tilde{\gamma}^{-1}(\rho+\kappa_2)\frac{1}{2}(1+\Delta_1). \end{aligned}$$

To see that the hub premium is smaller for syndicated investments relative to direct investment for investors in a hub region if monitoring costs for syndicators are not too large, note that:

$$
\begin{aligned}
Pr^{Synd}(H|L_i = H) - Pr^{Dir}(H|L_i = H) &= (\frac{1}{2} + \eta(\kappa_2))(1-\Delta_1) + \Delta_1 + \tilde{\gamma}^{-1}(\rho+\kappa_2)\frac{1}{2}(1+\Delta_1) \\
&\quad - \left[(\frac{1}{2} + \eta(\kappa_1))(1-\Delta_1) + \Delta_1 + \gamma^{-1}(\rho+\kappa_1)\frac{1}{2}(1+\Delta_1)\right] \\
&> 0 \\
if &\quad \kappa_2 < \tilde{\gamma}(\gamma^{-1}(\rho+\kappa_1)) - \rho
\end{aligned}
$$

For investors in a non-hub region, we have established that the probability of investing in a hub region under direct investment is:

$$
\begin{aligned}
Pr^{Dir}(H|L_i = NH) &= Pr\left(\gamma(n^H) - \kappa_1 > \gamma(n^{NH})|L_i = NH\right) \\
&\quad + Pr\left(\gamma(n^{NH}) < \rho \bigcap \gamma(n^H) - \kappa_1 < \gamma(n^{NH})|L_i = NH\right) \\
&= \frac{1}{2} - \frac{1}{2}\Delta_2 - \eta(\kappa_1) + \Delta_2\eta(\kappa_1) + \frac{1}{2}\gamma^{-1}(\rho) + \frac{1}{2}\gamma^{-1}(\rho)\Delta_2)
\end{aligned}
$$

For investors in a non-hub region, the probability of investing in a hub region under syndicated investment is similarly given as:

$$
\begin{aligned}
Pr^{Synd}(H|L_i = NH) &= Pr\left(\tilde{\gamma}(n^H) - \kappa_2 > \tilde{\gamma}(n^{NH})|L_i = NH\right) \\
&\quad + Pr\left(\tilde{\gamma}(n^{NH}) < \rho \bigcap \tilde{\gamma}(n^H) - \kappa_2 < \tilde{\gamma}(n^{NH})|L_i = NH\right) \\
&= \frac{1}{2} - \frac{1}{2}\Delta_2 - \eta(\kappa_2) + \Delta_2\eta(\kappa_2) + \frac{1}{2}\tilde{\gamma}^{-1}((\rho) + \frac{1}{2}\tilde{\gamma}^{-1}(\rho)\Delta_2)
\end{aligned}
$$

To see that the amount of hub premium is smaller for syndicated investments relative to direct investment for investors in a non-hub region, note that:

$$
\begin{aligned}
Pr^{Synd}(H|L_i = NH) - Pr^{Dir}(H|L_i = NH) &= \frac{1}{2} - \frac{1}{2}\Delta_2 - \eta(\kappa_2) + \Delta_2\eta(\kappa_2) + \frac{1}{2}\tilde{\gamma}^{-1}((\rho) + \frac{1}{2}\tilde{\gamma}^{-1}(\rho)\Delta_2) \\
&\quad - \left[\frac{1}{2} - \frac{1}{2}\Delta_2 - \eta(\kappa_1) + \Delta_2\eta(\kappa_1) + \frac{1}{2}\gamma^{-1}(\rho) + \frac{1}{2}\gamma^{-1}(\rho)\Delta_2)\right] \\
&> 0 \\
if &\quad \kappa_2 < \tilde{\gamma}(\gamma^{-1}(\rho+\kappa_1)) - \rho
\end{aligned}
$$

6.1.3 An Extension of the Theoretical Framework

The objective of this section is to provide an extension of our theoretical framework (Section 2) by explicitly incorporating the process of face-to-face due diligence, and observable startup and syndicate lead quality into the model.

We start by writing out return function for investors from the crowd in direct investment:

$$\Pi_i^D = max\{f_1(n_i^{Hub}), f_2(n_i^{NHub}), \phi^H + \rho\} - \kappa_1 d_{id}.$$

The return function of investing directly using their professional network is

$$\begin{aligned} f_1(n_i^{Hub}) &= \gamma_1(n_i^{Hub}) max\{\phi^H + \rho, (\phi^L + \rho + \epsilon_d)\mathbb{1}_{d_{id}=0}\} \\ f_2(n_i^{NHub}) &= \gamma_2(n_i^{NHub}) max\{\phi^H, (\phi^L + \epsilon_d)\mathbb{1}_{d_{id}=0}\}, \end{aligned}$$

Compared to the return function from Section 2, we introduce a few changes. First, the return functions $f_1(\cdot)$ and $f_2(\cdot)$ are now the product of two components, the old $\gamma_1(\cdot)$ and $\gamma_2(\cdot)$, and a new component which captures investors' choices for startups with a high quality signal (ϕ^H) or with a low quality signal (ϕ^L). Note that the network terms enters multiplicatively as a scaling factor – an "iceberg" search friction – and the investment return is scaled down by this factor. Startup quality is the sum of its observable signal, plus a term that captures its unobservable quality. Without loss of generality, we define ϵ_d to be the unobservable quality term of low-signal local startups relative to that of high-signal local startups, and $\epsilon_d \in \{-\alpha, 0, \alpha\}$. This term explicitly captures the possible existence of local information on the quality of a startup. Local investors observe a noisy signal of a startup's unobservable quality $\widehat{\epsilon_i}$ with precision p.

$$\widehat{\epsilon_i} = \begin{cases} \epsilon_d & \text{with probability p,} \qquad \epsilon_d \in \{-\alpha, 0, \alpha\} \\ 0 & \text{with probability 1-p} \end{cases}$$

This leads to a new proposition:

Proposition 3 *(Geography of Direct Investments by Startup Quality) In direct investments, the local premium (LP^1) is higher for startups with a low quality signal.*

Proof. *(Informal proof)* A high quality signal is observed and desirable to both local and non-local investors. Local investors, however, have a chance to learn the unobservable quality of local startups from local incidental information. This increases the value of investing in low-signal startups for local investors relative to distant investors. Therefore, the local premium is higher for startups with a low quality signal. ■

We then turn to syndicated investments. These investments involve two steps, namely investors choosing syndicate leads and then syndicate leads choosing startups. We start by explicitly modelling the second part:

$$\mathcal{P}_s = max\{g_1(n_s^{Hub}), g_2(n_s^{NHub}), \phi^H + \rho\} - \kappa_2 d_{sd}.$$

The return function for a syndicate lead s, $\mathcal{P}_s$, is similar to that for investors in that the syndicate lead can choose to invest using her professional network, or through a random draw from startups with a high signal in a hub region. A key difference is that syndicate leads have potentially different return functions from their network, g_1 and g_2, which are given as

$$\begin{aligned} g_1(n_i^{Hub}) &= \delta_1(n_i^{Hub}) max\{\phi^H + \rho, (\phi^L + \rho + \epsilon_d)\mathbb{1}_{d_{id}=0}\} \\ g_2(n_i^{NHub}) &= \delta_2(n_i^{NHub}) max\{\phi^H, (\phi^L + \epsilon_d)\mathbb{1}_{d_{id}=0}\}, \end{aligned}$$

where the search friction depends on their network, $\delta_1(\cdot)$ and $\delta_2(\cdot)$. Syndicate leads differ on three dimensions: their location, their search cost based on an unobservable number of relevant network ties, and their unobservable ability to perform due diligence. Therefore, for a given location, there are four types of syndicate leads based on combinations of their observable signals and unobservable characteristics.

For syndicate leads with high ability for conducting due diligence, the signal for unobservable startup quality $\widetilde{\epsilon_s}$ is more precise. The signals for the two types are given as:

$$\widetilde{\epsilon_s^L} = \begin{cases} \epsilon_d & \text{with probability p,} \qquad \epsilon_d \in \{-\alpha, 0, \alpha\} \\ 0 & \text{with probability 1-p} \end{cases}$$

$$\widetilde{\epsilon_s^H} = \begin{cases} \epsilon_d & \text{with probability p+}\Delta\text{p,} \qquad \epsilon_d \in \{-\alpha, 0, \alpha\} \\ 0 & \text{with probability 1-p-}\Delta\text{p} \end{cases}$$

This leads to a new proposition linking the geography of syndicated investments to the strength of startup signals:

Proposition 4 *(Geography of Syndicate Leads' Choice of Startups by Startup Quality Signal)*

When syndicate leads choose startups, the local premium (LP^3) is higher for startups with a low quality signal.

Proof. *(Informal proof)* High quality signals are observed and desirable to both local and non-local syndicate leads. However, local syndicate leads have a chance to learn the unobservable quality of local startups from local incidental information. This increases the value of choosing low-signal startups for local syndicate leads, relative to non-local syndicate leads. Therefore, the local premium (LP^3) is higher for startups with low signals. ■

Having discussed how syndicate leads choose startups, we now turn to the first step: decisions made by investors from the crowd selecting syndicate leads. Conditional on using syndication, investors can choose syndicate leads from hubs or non-hub regions, and with high or low signals. The investor's return functions from syndication Π_i^S is given as:

$$\Pi_i^S = (1-\tau) max\{\widetilde{f}_1(n_i^{Hub}), \widetilde{f}_2(n_i^{NHub})\},$$

where

$$\begin{aligned}
\widetilde{f}_1(n_i^{Hub}) &= \widetilde{\gamma}_1(n_i^{Hub}) max\{n_s^H, (n_s^L + \widetilde{\epsilon}_i)\mathbb{1}_{d_{is}=0}\} \\
\widetilde{f}_2(n_i^{NHub}) &= \widetilde{\gamma}_2(n_i^{NHub}) max\{n_s^H, (n_s^L + \widetilde{\epsilon}_i)\mathbb{1}_{d_{is}=0}\},
\end{aligned}$$

Syndicate leads have high or low observable signals based on their number of professional ties, n_s^H and n_s^L. They also differ in unobservable quality $\widetilde{\epsilon_s^H}$ and $\widetilde{\epsilon_s^L}$. This generates a few new propositions related to geography, startup signals, and syndicate lead signals.

Proposition 5 *(Geography of Investors' Choice of Syndicate Leads) The amount of local premium observed when investors select syndicate leads (HP^2) is higher for syndicate leads with low signals.*

Proof. *(Informal proof)* Syndicate leads with a stronger professional network have stronger signal and are appealing to both local and non-local investors. However, local investors may also be able to get a signal of syndicate leads' unobservable quality (i.e. their ability to perform due diligence and source high quality deals). This increases the value and therefore the likelihood they will select local syndicate leads with low signals in cases where these leads have high unobservable quality, but are only detected as high types by local investors. ■

Proposition 6 *(Geography of Syndicate Leads' Choice of Startups and Investors' Choice of Star-*

tups). The local premium is stronger when syndicate leads allocate capital to startups (LP^3) relative to when the online crowd does so directly (LP^1).

Proof. *(Informal proof)* Syndicate leads with high unobservable quality are better at due diligence, i.e. they have higher precision in their ability to observe the true quality of startups relative to investors from the crowd. When local startups have higher quality (observable signal plus unobservable quality), then syndicate leads are more likely to invest in them. Otherwise, syndicate leads are less likely to invest locally. ■

Proposition 7 *(Difference between Syndicate Leads' and Investors' Local Premiums) The difference in local premiums when syndicate leads allocate funds and when investors allocate funds ($LP^3 - LP^1$) is larger for startups with low signals.*

Proof. *(Informal proof)* Syndicate leads with high unobservable quality have higher precision in detecting high quality local statups relative to investors. Therefore syndicate leads are able to arbitrage high-quality local startups with low signal when these are indeed high quality deals. ■