Alexa Haushalter, Mika Limcaoco & Nicolas Nuñez
Computational Social Science

An Investigation on Local Bias in United States Startup Funding

Introduction

Considering the recent "tech boom" that has overtaken the United States, it should come as no surprise that social science research in Startups and Venture Capitalists is quickly gaining momentum. With this in mind, we decided to add to the already growing body of research on this topic by investigating the presence of "local bias" in U.S. startup funding. We define "local bias" as the phenomenon that occurs when investors have a lower standard for investing in local firms rather than non-local firms. In our case, we use local to refer to the surrounding region, which we will define more explicitly when we discuss our dataset. Research on local bias in startup funding is particularly interesting considering the large amounts of literature on local bias amongst traditional (i.e. bank and security) investors. For instance, a paper by French and Poterba (1991) documented that U.S. equity traders tended to invest around 94% of their funds to domestic securities, even though the American equity market comprised of less than 48% of the global equity market. In papers such as those by Coval and Moskowitz (1990 and 2001), researchers have found an even more proximal form of local bias, wherein American mutual fund managers preferred to own stocks of companies located nearby.

Despite the expansive body of research on investor bias, there is limited evidence to suggest that this bias exists amongst VCs and startups, especially considering the unique nature of VC-startup interactions/investments. After all, with the rise of tech-enabled connectivity, shouldn't tech-savvy startups be able to overcome geographic barriers? VC-startup investments also require two-sided matching, whereby the entrepreneurs must select the VCs and the VCs must select the entrepreneurs. VCs also require a significant amount of frequent in-person meetings with startups throughout the entire investment process, unlike in investment banking which often does not have high levels of in-person involvement.

Anecdotal articles by the New York Times have asserted the existence of phenomena such as the "20-minute rule" that "guides fateful decisions in Silicon Valley." Craig Johnson, an established VC director with over 30 years of experience in early-stage financing, advocates the "20-minute rule," and many venture capitalists adhere to it. The rule demands that "if a start-up company seeking venture capital is not within a 20-minute drive of the venture firm's offices, it will not be funded." More quantitative articles on TechCrunch have reported that nearly 50% of investments by VCs have been located within 233 miles of the VC's headquarters. This data comes from tracking investments from 1980 to 2009. In their 2010 paper, Douglas Cumming and Na Dai found that distance does not matter for the eventual performance of VC, although more reputable VCs exhibited less local bias. The VCs that exhibited less local bias tended to already have a stronger IPO track record, and were usually larger and older. The 2009 paper *Buy Local? The Geography of Successful and Unsuccessful Venture Capital Expansion* by Henry Chen, Paul Gompers, Anna Kovner, and Josh Lerner found that firms open new satellite offices based on the success rate of venture capital-backed investments in an area. This suggests that on at least an informal level, startups are aware of local bias, as they target areas with high concentrations of successful VCs.

The review of the literature suggests that the existence of local bias could be attributed to (a) proximity: startups have an easier time pitching to investors that are close by, (b) networking: startups have easier access to VC companies due to a closer network, (c) town pride: investors may want to support their local entrepreneurs, and (d) personal benefit: investors may have a desire for new services and products to develop in their area.

However, it should be noted that the NYT and TechCrunch findings were more anecdotal, and Cummings' and Dai's paper collected its data from VenturExpert which is significantly less comprehensive than Crunchbase. The Crunchbase data we utilize (from 1995-2015) is more expansive than that of VenturExpert. Additionally, the *Buy Local?* Paper suggested that the success of VCs may not solely be attributed to a VC's local bias towards startups; it may be due to the presence of being around other successful VCs. Considering the significant amount of external factors that have been reported to affect VC's investments in startups, we thought it would be appropriate to investigate the phenomena of local bias through a quantitative analysis of a comprehensive dataset from Crunchbase.

Dataset

We obtained our dataset from Crunchbase, an open data platform containing worldwide investment data for startups. The dataset has more than 168,000 data points from 1995 to 2015, each specifying a single investment. For each data point, we have information on the company name, company industry, company location, investor name, investor location, funding round (venture or seed, round A to H), and money raised. Company and investor locations are divided into city, state, country, and most importantly for our analysis, region. The region combines cities into small clusters based on geography. For example, "New York City" combines cities in the greater Manhattan area, and "SF Bay Area" combines cities in Silicon Valley. The grouping into regions benefits our analysis enormously, in that we can consider investments within a region as local. If we considered our analysis on the city level instead, our results may be misleading due to counterintuitive classifications. For example, a Palo Alto investor investing in a Mountain View company would be considered non-local, when in reality we believe most people would agree investments between Palo Alto and Mountain View are "local".

We wanted to focus our project domestically, so we filtered the dataset to analyze only investments between U.S. investors and U.S. companies. Filtering left us with 75,000 data points, approximately halving the size of the original dataset. We were concerned that removing all investments with a foreign investor or a foreign company would bias our dataset, so we ran additional analyses to confirm that removing the foreign investments would not cause significant errors in the distribution. First, we confirmed that the distribution of companies across regions of the U.S. stayed the same whether we included investments from U.S. investors only or investments from both U.S. and foreign investors. Likewise, we verified that the distribution of investors across regions of the U.S. stayed the same whether we included investments going towards U.S. companies only or investments going towards both U.S. and foreign companies. Finally, we analyzed the distribution of investments between U.S. companies and foreign investors across regions of the U.S., as well as the distribution of investments between U.S. investors and foreign companies across regions of the U.S. We confirmed that it matched the distribution of investments between U.S. companies and U.S. investors across regions of the U.S. Since these distributions are equal, we can confirm that removing the foreign investments will not significantly bias our dataset nor our conclusions about local bias.

Our analysis involves two main dimensions: local versus non-local investments and successful versus unsuccessful companies. As discussed above, we classify local investments as those where the investor region and the company region are the same, and non-local investments as those where the investor region and the company region are different. Unfortunately, classifying company success is not as easy as classifying local versus non-local investments. "Success" is an ill-defined term when it comes to a start-up company. It is not hard-coded into the dataset, leaving us to make a decision on what makes a company successful. Based on the data we had available, we decided the most reasonable metric for determining success was funding round. We decided that if a company reached Round D of funding, they were "successful." We recognize this definition has many flaws, especially due to the time-series nature of our data. It could be a company has not reached Round D yet because it is new or progressing slowly, but will reach Round D in the future. It could also be that a company did not reach Round D because the company simply does not need additional funding to move forward. Furthermore, it could be that a company reached Round D but then ultimately went bankrupt. Does this company deserve to be classified as a success? We considered removing investments for companies that started obtaining funding recently to reduce errors with companies simply not having enough time to reach Round D, but we did not want to remove the recent data points since they provide the most up-to-date and accurate information. We also considered defining success based on amount of money raised, but this seemed unfair and inaccurate considering a bad company could raise a lot of money but spend it mindlessly while a good company could triumph with a comparatively small amount of money. Some companies need less money to be successful, for example due to a lower cost of living in that company's region or lower costs associated with that company's industry. Based on external research and a bit of data analysis, we decided to choose Round D as a good balance for classifying companies as successful or not successful.

A local bias exists when investors hold a lower standard for investing in local firms than for investing in non-local firms. The purpose of our investigation is to combine information on the rates of local versus non-local investing with the rates of local and non-local startup success to conclusively determine if certain regions have a local bias in startup funding.

Local Investors versus U.S. Investors

American venture capitalists and startup founders are not equally distributed across the regions of the United States. In fact, as can be observed in Figure 1, the West dominates the startup funding sphere: 48.8% of all American startup companies come from this zone (we use the word 'zone' to distinguish these large areas from the relatively smaller areas, which we denominate 'regions'), and 36.2% of all venture capitalist investors reside in the West. The West is followed by the Northeast zone, with 26.8% of all American startup companies, and 34.4% of all American investors; next is the Midwest, the Southeast and finally the Southwest with less than 15% each in both categories. The purpose of these

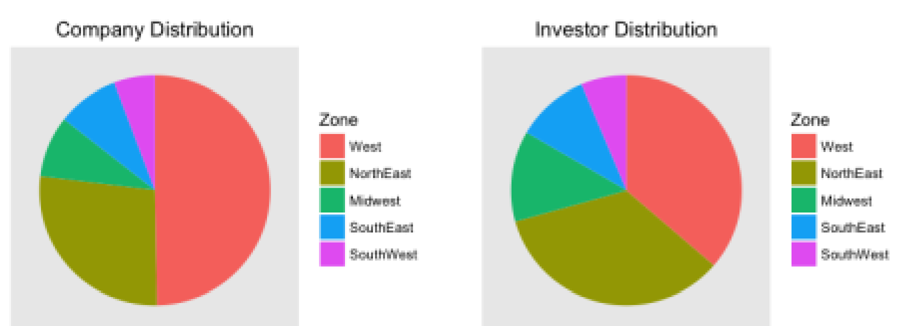

Figure 1. Distribution of Companies and Investors across American Zones

plots is to remind us that startup funding rates vary dramatically across zones in the United States.

These startup funding asymmetries across regions of the United States are important when discussing local bias because local startup funding in the San Francisco Bay Area is very different from local startup funding in Cedar Rapids, Iowa. Figure 2 not only exposes the heterogeneity of the localities in the United States, but also suggests that there are significant differences when analyzing investors from a specific region versus analyzing American investors as a whole. Specifically, the plot presents evidence that the rate at which investors from a specific region invest in companies from their same region is higher than the rate at which American investors invest in this specific region. We emphasize that the plot is about investment rates in different regions, which cannot be conclusively linked to a local bias without an underlying analysis of additional factors, particularly startup success rates in these regions.

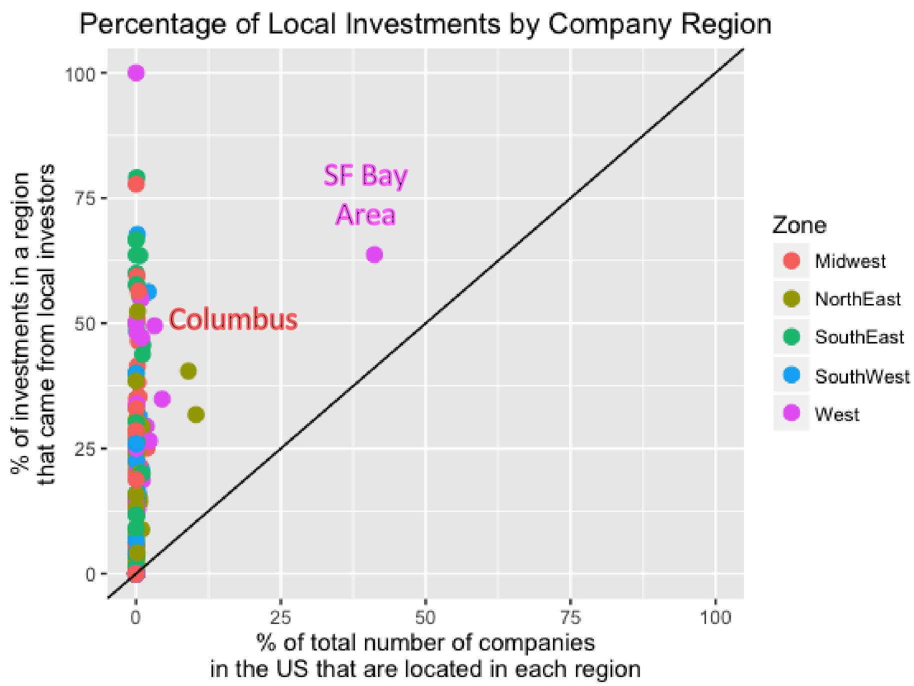

Figure 2. Percentage of Local Investments by Company Region

The plot's x-axis contains the percentage of total investments that go to each region, and the y-axis shows the percentage of investments from each region that are given to local companies. The pink point close to the center of the graph represents the San Francisco Bay Area. 40.88% of total investments between 1995 and 2015 were given to companies from the SF Bay Area, while 63.44% of investments from SF Bay Area investors go to companies in that region. If the distribution of preferences for all of the American investors were equivalent to the distribution of preferences for SF Bay Area investors, then we would expect the SF Bay Area investors to invest in the Bay Area at the same rate as the population of American investors, that is, at a rate of 40.88%. Since SF Bay Area investors actually invest at a rate that is 1.55 times larger than the rate of the population of American investors, there is a potential for local bias in the SF Bay Area. If we observe a smaller region, this effect is amplified: consider the city of Columbus, OH. While only 0.213% of total investments from American investors went to companies in Columbus, 50.79% of investments from investors from Columbus, OH were delivered to companies from this same city. The rate at which investors from Columbus finance companies from Columbus is 238 times larger than the rate at which American investors finance companies from Columbus. These analyses represent investments in terms of the number of times an investor invested in a company. In Figure 3, we present a similar analysis, except we represent investments in terms of the amount of money contributed instead of in absolute terms.

In Figure 3, the x-axis defines the percentage of total startup investment money that is invested in each region. The y-axis defines the percentage of total investment money from each region that is sent to local companies. As above, the pink point presented at the center of the graph corresponds to the SF Bay Area. According to CrunchBase data, 43.83% of the total amount of money invested in startups between 1995 and 2015 went to companies in the SF Bay Area, while 60.62% total investment money from SF Bay Area investors was invested in SF Bay Area companies. We see that the same effect applies whether we quantify investments in terms of U.S. dollars or in terms of the number of investments. When we analyze Columbus, OH in this context, we see that 0.065% of total money invested by U.S. investors goes toward Columbus companies, while 19.34% of total money invested by Columbus investors is devoted to Columbus companies. In summary, these plots show that investors invest in their own region at higher rates than U.S. investors invest in that specific region, on average.

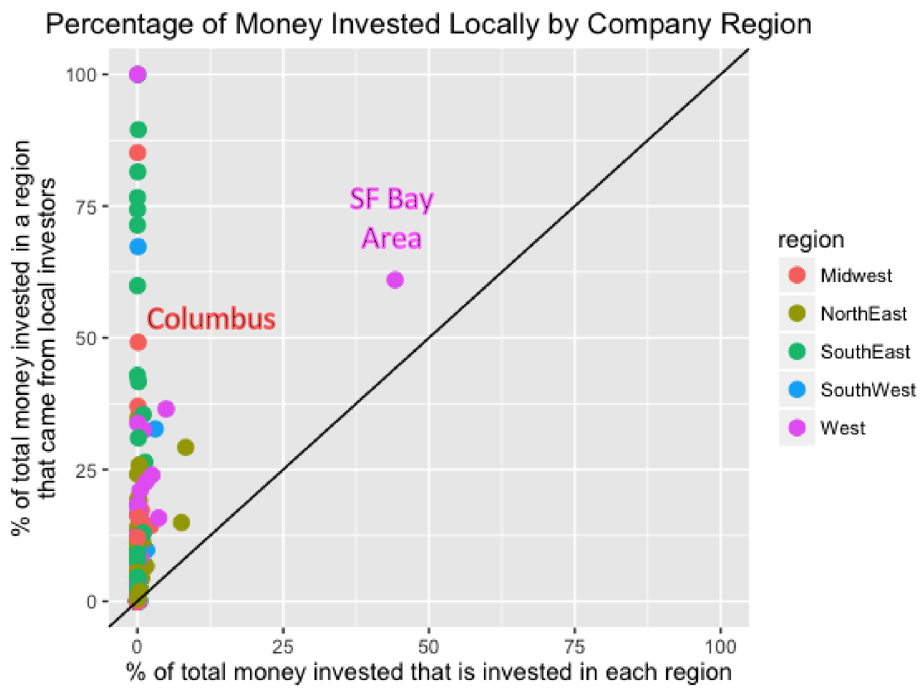

Figure 3. Percentage of Money Invested Locally by Company Region

There are several explanations for these differences in investment rates. The first is that investors in different regions tend to have different distributions for the regions of companies that pitch to them. In fact, it is likely that the rate at which companies from Columbus, Ohio pitch to investors from Columbus, Ohio is much higher than the rate at which companies outside of Columbus, Ohio pitch to Columbus, Ohio investors. Since these investors are more easily targeted by local companies, it is more likely for them to invest locally, which is one explanation for the trend observed in the above plots. The notion of omitted variable bias provides another explanation for the trend. In particular, we have omitted startup success rate for each region up to this point. In the next section, we analyze how success rates across different regions can determine if the differences in investment rates in local companies observed earlier is evidence for local bias.

Investment Success Rates

Based on the higher-than-average levels of investment in an investor's own region we noted in an earlier plot, there is an apparent local bias. However, we need to control for company success rates in a region to truly understand whether a local bias exists. For example, it could be possible that SF Bay Area investors invest more often in SF Bay Area companies than U.S. investors do because SF Bay Area companies are more successful. To begin our analysis controlling for company success rates, we first compare the success rates for the ten largest regions in the U.S. Here, we define size by the number of companies in our dataset coming from a specific region. Figure 4 shows the percent of companies within each region that reached the fourth round of funding, Round D. As discussed earlier, we are using reaching Round D as a measure of a company's success. Interestingly, we see in the plot that out of the ten largest cities, the SF Bay Area has the highest success rate. We note that out of 210 regions in the U.S., the SF Bay Area has the 19th highest success rate. If we only consider regions with at least 60 companies, the SF Bay Area ranks second. These findings suggest that the SF Bay Area investors may simply be investing more money and more frequently locally because the SF Bay Area companies outperform companies from other regions on average.

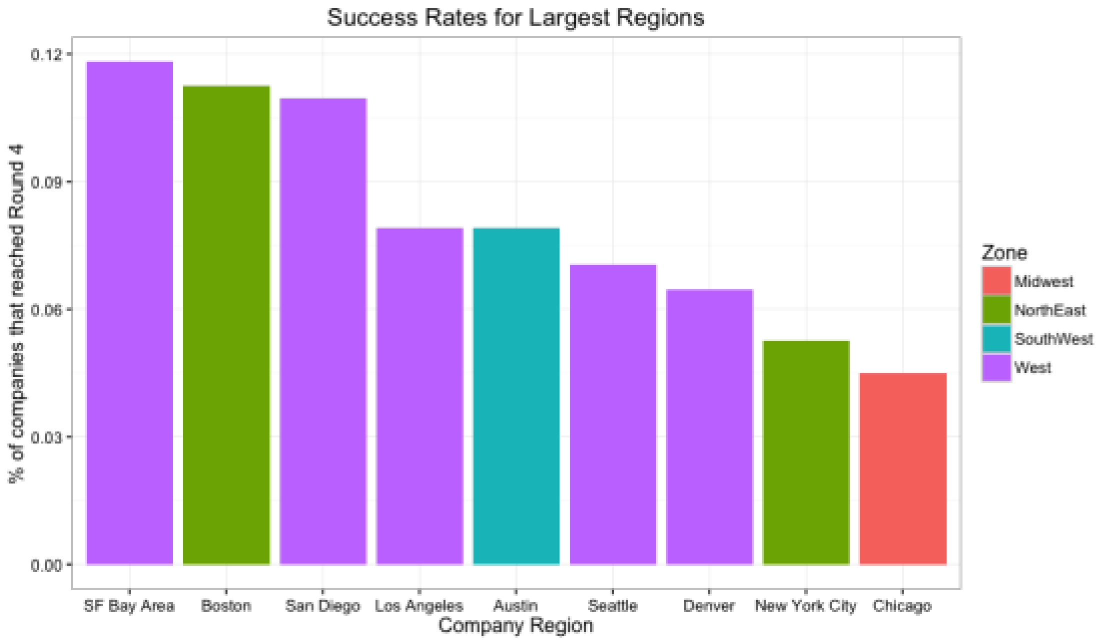

Figure 4. Success Rates for Largest Regions

To control for company success rates in every region of the U.S., we plotted the percent of local investments that reached Round D against the percent of non-local investments that reached Round D for investments made by investors in each region in the U.S. Clearly, this plot looks very different from the previous plots that showed an apparent local bias, highlighting the importance of adjusting for company success rates in each region. In this plot, points on the black 45 degree line indicate consistency: local success rates and non-local success rates are nearly equal for the average investor in these regions. Points above the 45 degree line indicate regions that may be exhibiting a local bias. For the average investor in regions above the 45 degree line, local success rates are lower than non-local success rates. This means the average quality of a local investment in this region is lower than the average quality of a non-local investment in this region. In other words, the caliber required for a company to obtain an investment from an investor in these regions is lower for local companies compared to non-local companies, on average. Likewise, points below the 45 degree line indicate regions that may be exhibiting a non-local bias. For the average investor in regions below the 45 degree line, local success rates are higher than non-local success rates. This suggests the caliber required for a company to obtain an investment from an investor in these regions is actually higher for local companies compared to non-local companies, on average. If an investor wants to maximize profits, they need to balance their investments to achieve equal success rates for local and non-local investments (move towards the 45 degree line) and improve investment strategies to achieve higher success rates overall (move along the 45 degree line to the top right).

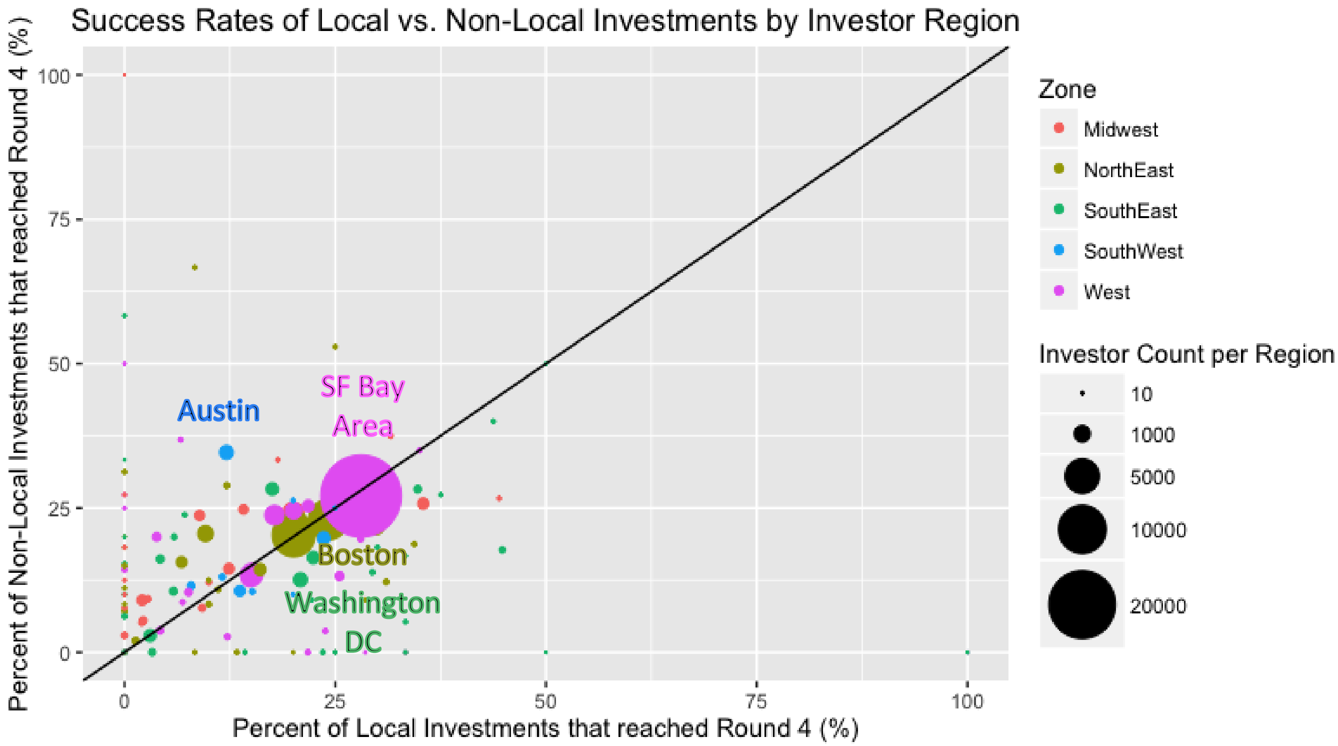

Figure 5. Success rates of Local vs. Non-Local Investments by Investor Region

Examining the plot further, we note that the largest circles, representing regions in the U.S. with the largest number of investors, fall on the 45 degree line. Furthermore, the largest regions also lie further to the right on the 45 degree line than the vast majority of other regions in the plot. Combined, these two facts suggest that the average investor in these regions is superior to the average investor in other regions; the average investor in these regions does not exhibit a local bias and has relatively high success rates overall. The largest pink dot corresponds to the SF Bay Area. Based on its location on this plot (comparatively high on the 45 degree line), the SF Bay Area seems to be the most successful investment region in the U.S. This is not surprising, given the fertility of entrepreneurship in Silicon Valley and the experience of investors in the region. For these same reasons, it is not surprising that regions with more investors seem to be the more successful, based on this plot.

We also note that many points fall far from the 45 degree line. For example, the blue dot above the 45 degree line is Austin, Texas. Its location on the plot indicates that the average investor in this region may be exhibiting a local bias. This is intriguing, considering Texas is typically known for having strong state pride; it likely emphasizes supporting local entrepreneurship, consciously or subconsciously. On the other hand, Washington, D.C. falls below the 45 degree line, indicating that it may be exhibiting a non-local bias. This likely relates to Washington, D.C. national and global focus, rather than local focus, considering it is where the national government is located. We note that Washington, D.C. is closer to the line than Austin, indicating the bias is stronger for Austin.

For most zones, the regions included in those zones fall on both sides of the 45 degree line at approximately equal rates, similar to the distribution for all regions as shown in the plot above. However, the story is different for the Midwest. In Figure 6, we see that the vast majority of regions in the Midwest fall above the 45 degree line, indicating a possible local bias. Similar to Texas, many states in the Midwest have strong local and state pride. Additionally, the Midwest is known for friendly nature. Thus, it is interesting, though not surprising, that most Midwest regions fall above the 45 degree line; investors in the Midwest want to support their friends and neighbors, and invest in local companies.

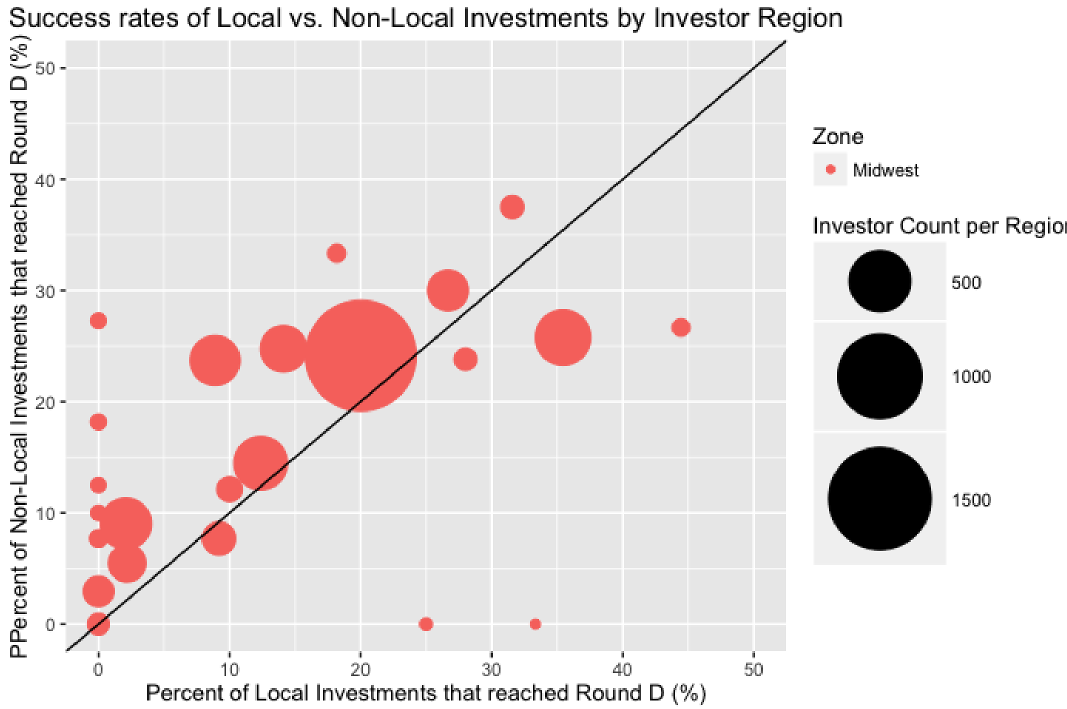

Figure 6. Success rates of Local vs. Non-Local Investments by Investor Region

Investment Threshold

A useful modeling strategy to conceptualize bias involves the adoption of an action threshold. The model assumes that investors decide whether to invest on a company or not by distilling all the information that is available to them to a single quantity *P*, which represents the ability of the company to generate value for the investor in the future. Once the investor has aggregated all the relevant company information and contextual information to one quantity, they compare this quantity to their investment threshold *T,* which determines whether or not this company will receive funding from this investor. In this model's context, there is a bias when local companies and non-local companies have different thresholds. Namely, if the threshold for local companies is lower (respectively, higher) than that of non-local companies, then there is a local bias (respectively, non-local bias).

We can interpret the information from Figure 5 in terms of investment thresholds by assuming that our definition of success is an accurate proxy for a company's ability to generate value for an investor. Consider a region on the 45 degree line such as the SF Bay Area. We see that both local and non-local investments from Bay Area investors tend to succeed 25% of the time, which suggests that Bay Area investors have a consistent threshold for both local and non-local companies. When the success rates of local and non-local investments differ, it is possible that the threshold for funding the less successful investments is lower, which offers an explanation as to why the results were worse.

It is essential to point out that even if the success rates are different, this does not necessarily imply that local bias exists. This is because the valuation P of a company in a specific investor's perspective is a function of the information that the investor possesses. We imagine that this signal is noisy, and therefore the valuation P is a random variable with some associated variation. Hence, some apparently locally biased regions might have held the local and non-local companies to the same threshold, but random chance affected one of the groups more than another, leading to worse success rates. Another reason local bias might not necessarily exist is that investors from a certain region are not taking into consideration barriers for their local companies to succeed, such as reduced demand caused by smaller populations in non-metropolitan areas, or under-resourced infrastructure in smaller regions, such that their estimate for P is inaccurate. We expect an inaccurate estimate of P for less-experienced investors and from smaller regions, since these errors in estimation tend to wash out when aggregating across hundreds of investors. Hence, we conclude that while interpreting success rates as a proxy for investment thresholds can reveal insights on the existence of local bias, we cannot be certain that investors from regions above the 45 degree line in Figure 5 expect less from their local than their non-local investments.

The investment threshold model allows us to connect the investor's acumen with their likelihood of success by assuming that experienced investors will have better estimates for a company's ability to generate value. This model also allows us to draw a distinction between discrepancies in success rates and the presence of a bias. This overview of an investment threshold model can be extended by specifying distributions for the signal P_{ij}, which represents the ability of company j to generate value for investor i.

Conclusion

Overall, we do not find evidence for a systematic local bias across the United States. Figure 5 controls for company success rates within each region, allowing us to draw sensible conclusions. In general, the largest regions fall almost exactly on the 45 degree line in this plot, indicating that regions with more entrepreneurship exhibit little to no bias. However, many regions also lie off the 45 degree line and appear to be exhibiting bias, whether it be local or non-local. Finally, we note that the majority of regions in the Midwest fall above the 45 degree line (Figure 6), indicating that on average, Midwest investors tend to exhibit more local bias than investors in other zones.

Due to the large differences in across regions of the U.S., we encourage investors to ask themselves whether or not they are exhibiting bias, and whether this is what they desire. If the success rates of local companies differ significantly from the success rates of non-local companies for a specific investor's investments, that investor is likely exhibiting bias. If the investor's goal is to maximize profit, they should attempt to reduce their bias and increase overall success rates. Specifically, if the investor is exhibiting local bias and wants to maximize profits, they should increase investments in high-quality non-local companies and decrease investments in low-quality local companies. Likewise, if the investor is exhibiting non-local bias and wants to maximize profits, they should decrease investments in low-quality non-local companies and increase investments in high-quality local companies. We note, however, that exhibiting no bias does not match the goals of all investors. For example, some investors may want to exhibit local bias if supporting local entrepreneurship is more important to them than maximizing profits. Thus, we suggest that investors simply analyze whether or not they are exhibiting bias, in order to then assess if they are exhibiting the level of bias they desire or if they should change their investment strategies.

One of the largest limitations of our dataset is that it only contains information on startup funding rounds that were successful; that is, on those where the investor chooses to fund the startup. Analyzing the rate of unsuccessful pitches from both local and non-local startups would be an interesting extension to this investigation. This is because knowing the rate of rejection of funding for local versus non local startups would allow us to quantify the fraction of total pitches that are local versus non-local for each region, thereby informing us on regional variations on the exposure of investors to local companies. Additionally, it would be interesting to try to find better ways to define success, possibly as a combination of metrics. Likewise, it might be worthwhile to consider new ways to define local, possibly based on number of miles between the investor and company. We consider successful versus unsuccessful and local versus non-local as binary variables, but it would be interesting to consider an analysis of these as continuous variables instead. Overall, we did not find evidence for a systematic local bias in U.S. startup funding, but additional research could be valuable to shed new light on the topic.