HARVARD | BUSINESS | SCHOOL

Judging Foreign Startups

Nataliya Langburd Wright
Rembrand Koning
Tarun Khanna

Working Paper 21-097

Judging Foreign Startups

Nataliya Langburd Wright
Harvard Business School

Rembrand Koning
Harvard Business School

Tarun Khanna
Harvard Business School

Working Paper 21-097

Funding for this research was provided in part by Harvard Business School.

Judging foreign startups*

April 9, 2021

Nataliya Langburd Wright (Harvard Business School)

Rembrand Koning (Harvard Business School)

Tarun Khanna (Harvard Business School)

ABSTRACT

Can accelerators and investors pick the most promising startup ideas no matter their provenance? We investigate this question using unique data from a global accelerator in which judges across international regions are randomly assigned to evaluate startups headquartered across the globe. Our analysis of this natural experiment shows that judges are less likely to recommend startups headquartered outside their home region by 4 percentage points. Additional information about a startup does not appear to attenuate this bias. Back-of-the-envelope calculations suggest that this discount leads judges to pass over 1 in 10 high-potential startups. Our results reveal that judges are able to discern high from low quality startups regardless of their origin, but they systematically discount foreign ideas relative to local ones.

Keywords: Entrepreneurship and Strategy, Global Strategy, Entrepreneurial Financing, Innovation, International

*nlangburdwright@hbs.edu, rkoning@hbs.edu, tkhanna@hbs.edu. We thank Susan Cohen, Jorge Guzman, Ramana Nanda, and Maria Roche for valuable comments. We also would like to thank Julia Comeau for excellent copy-editing. All errors remain our own.

I. Introduction

Startups, like corporations, are increasingly globalized in terms of their markets, investments, and workforce participation (Kerr, 2016; Ghemawat and Altman, 2019; Lu and Beamish, 2001; Oviatt and McDougall, 2005), partially due to the advent of technology that reduces the cost of expanding internationally (Brynjolfsson, Hui, Liu, 2019). As a result, entrepreneurial gatekeepers, ranging from investors to accelerators, increasingly evaluate a global pool of startups and must choose the most promising to provide support and funding (Balachandran and Hernandez, 2020). For example, Silicon Valley-based Y Combinator funded Ukraine-based Petcube, an interactive pet monitor startup that went on to become a unicorn, valued at over $1 billion (Y Combinator, 2020; X1 Group, 2018). At the same time, gatekeepers have missed out on promising international startup opportunities; for example, Silicon Valley-based Bessemer Venture Partners passed over Australian-based Atlassian, a project management software company, which is now worth over $43 billion (Bessemer Venture Partners, 2020).[1] Furthermore, Skype, a video chat software company co-founded by an Estonian entrepreneur, faced numerous rejections from foreign venture capitalists, including Canadian-based BCE capital, before ultimately becoming a unicorn valued at $8.5 billion when acquired by Microsoft (Haley, 2005; Damouni and Rigby, 2011).

Can accelerators and investors choose the most promising startups from this increasingly global pool? Judges may not be able to discern the quality of startups overall because early stage ideas are highly uncertain (Kerr, Nanda, and Rhodes-Kropf, 2014; Gans, Hsu, and Stern, 2008; Luo, 2014). They may be particularly inaccurate in discerning the potential of foreign startups

[1] The valuation is in terms of market capitalization.

because they lack the contextual expertise and information – ranging from knowledge of institutions to differences in consumer tastes – necessary to sort winners from losers. Moreover, judges may carry a bias for or against foreign startups, similar to the gender, race, and expertise biases documented across a range of entrepreneurial and innovation settings (e.g. Lee and Huang, 2018; Niessen-Ruenzi and Ruenzi, 2019; Hegde and Tumlinson, 2014; Li, 2017). Overall, judges may be either uninformed about the quality of both foreign and local startups, only informed about local startups, or informed about foreign and local startups equally. They also may be biased.

Understanding whether judges are informed about the quality of local and foreign startups is essential to understanding why home bias occurs and how organizations should address it. Prior home bias research shows that trade partners, financial analysts, and investors are more likely to select companies that are nearby, but these studies often conflate crucial differences in the mechanisms underlying the effect (Disdier and Head, 2008; Coval and Moskowitz, 1999; 2001; Sorenson and Stuart, 2001). As mentioned above, home bias by accelerators and startup investors could be the result of a simple preference for home-grown startups irrespective of each startup's potential. Under this mechanism, an accelerator could simply counter its bias by lowering its threshold for accepting or investing in foreign firms. However, such an approach will backfire if the underlying home bias mechanism is instead rooted in the inability of judges to distinguish foreign winners from losers. In this situation, judges pick the most promising local ventures whereas their choices of foreign firms are potentially no better than random draws. No matter the threshold, judges will always end up selecting lower quality foreign ventures than local ones. In this case, remedying the underlying "bias" requires finding judges who can discern winners from losers, perhaps by assigning judges to only evaluate startups from their home region. Redesigning how scores are aggregated into decisions will not matter. In short, the underlying mechanisms that lead to home bias have strong implications for how accelerators and investors should design their selection processes.

However, teasing apart these mechanisms is non-trivial. First, estimating judge home bias effects, in and of itself, is not easy. Estimates that rely on the location of selected startups, as well as the investors and accelerators who select them will nearly always confound supply-side forces (the judge's choice of who to pick) and demand-side ones (the founder's choice of where to apply). Further, even when the distribution of potentially selected startups is fully observed (e.g. in venture competitions), startups may selectively choose whether to enter local or foreign competitions, and judges are often non-randomly assigned which startups to assess. In these cases, estimates are again biased because higher-quality startups might disproportionally select into local competitions, or harsher judges might be assigned to foreign ventures. Finally, even if judges and startups from different countries are randomly assigned to one another, showing that judges discount foreign startups is insufficient to reveal the underlying mechanism, which ultimately determines optimal organizational responses. Specifically, teasing apart whether home bias is rooted in uniform discounting or differences in a judge's ability to evaluate requires not just random assignment of judges but also measures of each startup's quality.

Here we analyze data from an accelerator's global venture competition in 2017 and 2018 that meet these criteria and so allow us not only to causally identify if judges exhibit home bias, but also pinpoint the mechanisms underlying this effect.[2] In the first round of this competition, 1,040 judges from North America (the United States and Canada), Latin America, Europe, and Israel evaluated 3,780 startups from across the globe. Crucially, in this first round, the

[2] Accelerators are defined as "fixed-term, cohort-based program for startups, including mentorship and/or educational components, that culminates in a graduation event" (Cohen, Fehder, Hochberg, and Murray, 2019).

accelerator randomly assigned judges to evaluate startups no matter their origin, and no startups could opt out of being evaluated by judges from particular regions.

We find that judges are less likely to recommend startups from a foreign region by 4 percentage points after accounting for observed and unobserved differences in startup quality with startup-level fixed effects. The magnitude is meaningful. It is roughly a third of the effect of a startup going from having no users to some user traction and a tenth of the size of the effect of having raised venture financing. These magnitudes are consistent with prior work documenting home bias in other settings ranging from financial markets to trade (Coval and Moskowitz, 1999; Disdier and Head, 2008).

Our analysis reveals that this effect is driven by a consistent discounting of foreign startups by local judges and not by differences in the ability of judges to better pick winners from losers amongst local firms relative to foreign firms. Surprisingly, we instead find judges are equally good at evaluating startup quality whether the startup is from their home region or not. In fact, judges give higher scores to local and foreign startups that go on to raise financing and experience more user growth, contrary to prior work showing that judges can struggle to pick startup winners from losers (e.g. Scott, Shu, and Lubynsky, 2020). Further, this foreign discount does not diminish when stronger signals of startup quality (whether startups have raised financing) are present, suggesting that the judge's home bias is unlikely to stem from statistical discrimination. Overall, evidence points to explanations rooted in a taste-based preference for local firms over foreign ones. Indeed, when we conduct back-of-the-envelope calculations, we find that judges passed over 324-512 promising foreign startups, equating to roughly 1 in 10 startups in our sample. Overall, we find strong evidence that judges can evaluate startup quality no matter the firm's provenance, but they uniformly discount foreign firms, suggesting that simple changes to how accelerators and incubators aggregate judges' evaluations may mitigate the impact of home bias on outcomes.

This study makes three primary contributions. First, our findings suggest that entrepreneurial gatekeepers are adept at evaluating startup quality, in contrast to prior work showing that accelerators, investors, and judges struggle to separate the most from the least promising startups (Nanda, Samila, and Sorenson, 2020; Scott, Shu, and Lubynsky, 2020; Kerr, Lerner, and Schoar, 2014). Crucially, we find little evidence that gatekeepers have a local advantage in screening startups. To be clear, local judges might hold other advantages when it comes to investing in or nurturing these startups, but our findings suggest screening is unlikely to explain such differences. Further, the fact that judges were equally good at screening foreign and local startups suggests that the diffusion of standardized business models (Y Combinator, 2019)[3], management practices (Chatterji, Delecourt, Hasan, and Koning, 2019), and technology (Haefliger, Von Krogh, and Spaeth, 2008) might be making it easier to evaluate startups no matter their origin.

Second, our results suggest that geographic bias may distort the composition and direction of entrepreneurship and innovation in ways that research has shown in terms of gender and race (e.g. Lee and Huang, 2018). If gatekeepers are biased against foreign startups, and if the majority of these gatekeepers still reside in entrepreneurial hubs like in the U.S., this may potentially result in a gap in startups from non-hub regions. And this bias does not just impact which startups succeed, but also may impact who benefits from their innovations (Koning, Samila, and Ferguson, 2020). Indeed, if accelerators overlook ideas from these non-hub markets, then there may be too few startups serving the needs of customers in those foreign markets.

---

[3] For example, Silicon Valley-based Y Combinator highlights "nine business models and metrics that investors want" (Y Combinator, 2019).

Third and finally, we highlight a potential limitation of accelerators when it comes to helping foreign startups gain access to key entrepreneurial ecosystems. While various studies find that accelerator programs result in positive performance gains for startups (Cohen, Bingham, and Hallen, 2019; Hallen, Cohen, and Bingham, 2020; Yu, 2020; Howell, 2017; Gonzalez-Uribe and Leatherbee, 2018; Fehder and Hochberg, 2014; Yin and Luo, 2018), our results suggest that the impact of accelerators may be muted for foreign startups because these organizations discount them. That said, our results also suggest that relatively minor tweaks to how a firm aggregates decisions might address this foreign bias.

II. Theoretical Framework

Evaluating Startup Quality

Predicting startup quality is difficult because of at least three information challenges. First, the success of startup ideas hinges on the interaction of complex factors, including the technology itself, the business model, customer demand, competition, and the founding team (Gompers, Gornall, Kaplan, and Strebulaev, 2020; Sørensen, 2007; Kaplan, Sensoy, and Strömberg, 2009; Aggarwal, Kryscynski, and Singh, 2015; Hoenig and Henkel, 2015). Second, there are few precedents to anchor startup predictions. Great startup ideas are inherently novel, and only a subset of those actually succeed in practice (Hall and Woodward, 2010). Third, entrepreneurs may only provide incomplete information about their ideas, as disclosure can eliminate incentives to "pay" for the now "free" to appropriate idea (Gans, Hsu, and Stern, 2008; Luo, 2014; Arrow, 1962). Consistent with these priors, research shows that investors and mentors often lack the ability to predict the quality of startups (Nanda, Samila, and Sorenson, 2020; Scott, Shu, and Lubynsky, 2020; Kerr, Lerner, and Schoar, 2014).

Contextual Intelligence

Given these challenges in discerning startup quality, when (if at all) can evaluators distinguish winners from losers? Evaluators may be able to do so when they have expertise (Li, 2017) or intuition (Huang and Pearce, 2015) that compensates for the imperfect information they have on any new venture. Indeed, prior research suggests that expertise is a product of the local region where investors and inventors live and work (Dahl and Sorenson, 2012; Malloy, 2005; Coval and Moskowitz, 2001). However, this locally-developed expertise may not be transferable to foreign contexts because of differences in institutions, culture, language, and markets (Khanna, 2014). Evaluators, therefore, may only be able to use this locally derived expertise to better assess the quality of local, but not of foreign startups. For example, an Israeli investor might be able to use her expertise of Israel's military structure to understand the relative quality of founders of an Israeli company with military experience and not a U.S. company with founders who have military experience. Consistent with this view, prior work has shown that financial analysts are worse at picking foreign stock winners, relative to local stock winners (Malloy, 2005; Coval and Moskowitz, 1999), and information frictions are higher for foreign acquirers (Conti, Guzman, and Rabi, 2020).

Bias in Evaluations

However, reliance on local expertise to evaluate startups may also induce biases. Prior work shows that judges prefer what is more "familiar" (Huberman, 2001; Franke, Gruber, Harhoff, and Henkel, 2006; Lin, Prabhala, and Viswanathan, 2013). In the demographic context, prior research has found substantial evidence of bias against entrepreneurs from different genders and races (Lee and Huang, 2018; Niessen-Ruenzi and Ruenzi, 2019; Hegde and Tumlinson, 2014). Similarly, in the geographic context, studies in financial and trade markets have detected a home bias for local portfolio stocks or trade partners (Disdier and Head, 2008; Coval and Moskowitz, 1999; 2001).

Hypothesis Development

These different mechanisms -- evaluation uncertainty, contextual expertise, and bias – generate six scenarios that each call for different strategic responses. Figure 1 sketches how each of these scenarios reveals a different relationship between startup quality (x-axis) and a judge's evaluation score (y-axis) for startups foreign to the judge (dashed line) and local to the judge (solid line).

[Insert Figure 1]

In the first row of Figure 1, we show the pessimistic cases where judges cannot pick winners from losers. No matter whether judges are biased (cell B) – systematically preferring local or foreign startups — or unbiased (cell A), the selected pool of startups consists of a random share of high and low quality firms. In this worst case scenario, organizations should reduce their attention to screening startups and perhaps re-allocate resources to monitoring selected startups in the hopes of improving firms' future performance (Bernstein, Giroud, and Townsend, 2016).

However, research ranging from work on contextual intelligence to the benefits of investing in and running firms in one's home region (Dahl and Sorenson, 2012; Malloy, 2005; Coval and Moskowitz, 2001), suggests that judges can pick winners from losers locally even if they cannot evaluate the quality of foreign startups. The second row of Figure 1 illustrates this scenario. Cell C shows that when judges have a local information advantage, and are not biased against foreign startups, they will give higher quality local startups higher scores. However, they will not necessarily give higher scores to lower quality local startups. In fact, with better local information, it is likely that judges will give low quality local startups low scores while erroneously evaluating low quality foreign startups as better than they actually are. The result is that the lines intersect in cell C. However, if judges are also biased, this shifts the line for local startups upwards as seen in cell D. While judges still give higher scores to better local startups, all local startups will be judged as better than any given foreign firm. The result is that in cell D and cell B, we see consistent foreign discounting, but each reflect meaningfully different mechanisms. While cell B suggests that organizations would be better off re-allocating attention away from the selection process all-together, cells C and D suggest that organizations would be better off assigning judges to evaluate local but not foreign startups.

Lastly and most optimistically, judges might be able to evaluate the quality of both local and foreign firms, as shown in the third row of Figure 1. Startups may follow a similar enough playbook that separating good from bad investments across countries is not significantly harder than within countries. For example, work has shown the benefits of good management appear universal for corporations and startups across the globe (Bloom and Reenen, 2007; Chatterji, Delecourt, Hasan, and Koning, 2019), as are coding practices (Haefliger, Von Krogh, and Spaeth, 2008). As shown in cell F, bias interferes with picking the most promising startups because judges may pass over higher quality foreign startups for lower quality local startups. In this case, organizations can simply revise their processes to reduce bias either in aggregate (e.g. by lowering the threshold for selecting a foreign firm versus a local firm) or at an individual judge level (e.g. by introducing nudges) to counter this discount.

The framework presented in Figure 1 builds on information-bias tradeoffs discussed in other studies of evaluation (e.g. Li, 2017; Boudreau, Guinan, Lakhani, and Riedl, 2016). Our simple two-by-three reveals that knowing whether judges give lower scores to foreign startups –

as is the case in cells B, C, D, and F – is insufficient to understand how an organization might change to address foreign discounting. However, with knowledge of startups' realized success or ex-ante measures that let the analyst measure differences in firm quality, we have sufficient information to separate the different mechanisms that operate in each cell.

III. Context: Global Accelerator Competition

To unbundle these scenarios, we use data from a large global accelerator's new venture competition. The accelerator operates in four regions around the world: the U.S., Europe, Israel, and Latin America. There are four rounds in the accelerator program. In the first round (the global round), startups virtually apply to several of the regional locations of the accelerator program. In the latter rounds, the accelerator assigns startups (based on their preferences and judge scores) to one of its regional locations, and judges generally local to that area evaluate the startups. The pool consists of mostly high tech startups, similar to startups in other top accelerator programs like Silicon Valley-based Y Combinator or Techstars. Roughly a third of startups make it from the initial applicant pool into the second round, a third from the second to the third round, and a quarter from the third to the final round. Startups who make it to the third round (approximately 10 percent of the initial applicant pool) participate in the full in-person accelerator program, including the educational curriculum, mentorship program, and other networking events. The top 10-20 rated startups across the globe at the conclusion of the last round gain both credibility and monetary prizes worth tens-of-thousands of dollars. Across 2013-2019, these four rounds consist of 87,977 startup-judge level observations, including 11,188 unique startups and 3,712 unique judges.

We focus on the global round of the competition where judges – representing executives (60%), investors (13%), and other professionals (27%) – across these international regions initially screen startups from around the world. Judges evaluate an application that includes self-reported information on the company's background and funding, industry & competitors, and business model & financials. We show the full application template in Appendix A7. All applications are in English.[4] While the applications do not specifically inform judges of the startup's location, judges may infer it fairly easily through the description of the startup, founder(s), and market. Judges review these applications online. Each judge evaluates roughly 20 startups, and each startup receives evaluations from 5 judges on average. Judges recommend whether a startup should move to the next round of the competition.[5] Judges also provide subscores on a scale of 1-10 on the following criteria: startup team, industry & competitors, and business model & financials.

To infer judges' location, we use data on the location of the accelerator the judge is affiliated with.[6] As judges need to evaluate startups in person during the later rounds of the competition, they tend to be assigned to a physically proximate accelerator. We therefore categorize judge locations as corresponding to the accelerator's locations: Europe, Latin America, Northern America (U.S. & Canada), and Israel. We manually checked a random subsample of 136 judges from our data to test if the residence regions of the judges match those of their home programs. Of this subsample, 76 percent (103/136) of judges resided in their home program region. Crucially, this measurement error should bias our estimates of foreign bias towards zero. Further, the broad regional categories will lead us to underestimate biases within regions. For example, a U.K. judge evaluating a Latvian startup would appear as a regional

---

[4] While English applications may mask quality of startups whose founders have a different native language with different writing styles, such a language requirement is common for startup accelerator program applications.
[5] Judges provide a 0-5 score on whether they recommend the startup to the next round of the competition; scores above 2 result in startups moving to the next round.
[6] The accelerator does not collect data on judges' location of residence. It only collects the home accelerator program of each judge.

match in our data, though we can imagine that the judge would consider the startup foreign and so potentially discount it.

IV. Data

Our data come from the accelerator's 2017 and 2018 cycles. During these two years, judges were randomly assigned to startups during the initial global round. This random assignment allows us to overcome the possibility that startups self-select into local programs. Such selection would make it impossible to separate judge from startup effects. Our 2017-18 data consist of 20,579 startup-judge level observations, including 4,420 unique startups and 1,043 unique judges. We remove startups whose headquarter regions do not match any of the judges' home programs to exclude the startups that are foreign to all judges in our sample and therefore lack a local judge score as a basis of comparison.[7] We also remove judges who lack a home program that is part of the main accelerator.[8] This brings our final sample to 17,608 startup-judge level observations, including 3,780 unique startups and 1,040 unique judges.

Measuring Startup Quality

Measuring startup quality is not only difficult for judges, but also for researchers. Early stage startups rarely have revenue or profits that are common metrics of company performance. Instead, entrepreneurship studies turn to other intermediate milestones to proxy early stage companies' performance and quality. One common measure is financing from angel investors or venture capitalists (Cao, Koning, and Nanda, 2020; Howell, 2017; Yu, 2020). This is a common measure because of these investors' selection and treatment effects that may result in higher startup performance. On the selection side, early stage investors conduct rigorous due-diligence

[7] Our results are robust to including or excluding startups whose headquarter regions do not match those of any of the judges' home programs.

[8] Our results are robust to including or excluding judges whose home program is not one of the main accelerator programs.

on portfolio companies prior to investing that may enable them to understand the quality of ventures (Gompers, Gornall, Kaplan, and Strebulaev, 2020). On the treatment side, investors provide added value (Bernstein, Giroud, and Townsend, 2016) and a stamp of approval (Lerner, Schoar, Sokolinski, and Wilson, 2018) to startups that enable them to gain subsequent financing and increase their chances of a successful exit, either an acquisition or initial public offering (Catalini, Guzman, and Stern, 2019). Another common indicator is user traction, reflecting how much visibility and use a startup is getting from customers and other gatekeepers. Website page visits are becoming a common indicator for the latter in entrepreneurship studies to proxy startup performance (Koning, Hasan, and Chatterji, 2019; Cao, Koning, and Nanda, 2020; Hallen, Cohen, and Bingham, 2020).

We measure both pre-accelerator and post-accelerator measures of financing and website page visits in our analysis. Pre-accelerator measures allow us to assess whether judges can evaluate the quality of startups at the time of evaluation. Post-accelerator measures allow us to evaluate whether judges can evaluate the future potential of startups. Judges seek to evaluate both types of quality. In both cases, we test if judges give higher ratings to higher performing local startups but not to higher performing foreign ones.

Dependent Variables

Score **–** Our first dependent variable is a composite z-score created from the z-scored subscores judges give to startups. These underlying subscores include: customer pain and solution, customer needs and acquisition, financial/business model, industry competition, overall impact, regulations and intellectual property, team (including advisors and investors), and the overall recommendation. These subscores correspond to the sections in the applications startups initially complete. All but the last range from a scale of 1-10. The latter is on a scale of 0-5. While not all judges complete every subscore evaluation, the vast majority do. Of the 17,608 recommendation evaluations in our data, for 16,339 (93%) we have complete subscore information.

Recommend **–** Our second dependent variable is a binary variable indicating whether a judge recommended the startup to advance to the next round of the competition.[9]

Independent Variables

Foreign Startup - Our key covariate captures whether the judge and startup are from the same region (e.g. both from Europe, the U.S./Canada, Israel, or Latin America). We construct a binary variable indicating whether a judge is evaluating a foreign startup ("1" indicates a foreign startup, "0" indicates a local startup).

Logged Financing Value (Post) - We use logged financing value six months after the program.[10] This variable indicates the logged amount of USD startups received from investors six months after the program.

Logged Page Visits (Post) - We also use logged monthly page visits after the accelerator program in 2019 (the latest data we have available).

Financing (Pre) - We use logged financing value (in USD) that startups received from investors before the program.

Whether Has Financing **–** We include a binary variable indicating whether a startup received financing before the program to indicate financing traction.

[9] We constructed this as equal to 1 if the judge's score was over 2 (on a scale of 0-5) and 0 otherwise, as the accelerator uses this cutoff to determine whether a startup makes it to the next round of the competition.
[10] All logged values are of (1+x) because of frequency of zeros in our dataset.

Logged Page Visits (Pre) - We include logged website page visits 3 months before the initial application review period of the accelerator.

Whether Has User Traction - We use a binary variable on whether a startup reached at least 100 website page visitors on average per month over the last three months before the program to indicate user traction.

In our context, when startups lack page visit or financing data, they generally have so few visits or little financing that corresponding databases like SimilarWeb (that collects companies' page visits) and Crunchbase (that collects startups' funding rounds) do not track them. We therefore set missing page visit or financing values to zero. In robustness checks, we confirm that whether a startup has financing and page view data are positively correlated with their evaluations, suggesting that the missing values are the result of startup shutdown or slow maturity.[11]

Accelerator Participation - We also account for whether a startup participated in the accelerator interacted with whether a startup is local or foreign to the judge. This variable allows us to control for the potential treatment effects of the accelerator that may confound our ability to assess whether judges are able to detect the post-accelerator performance quality of startups. We include it in specifications involving post-accelerator financing and page visit variables.

Descriptive Statistics for Evaluations - Table 1 shows summary statistics for our main sample from the global round of the competition, including 17,608 startup-judge level observations, 3,780 unique startups, and 1,040 unique judges. These summary statistics break up our main dependent variables (judge score measures) and independent variables (startup quality measures) by whether a startup is local or foreign to the judge in a given evaluation. The raw data

---

[11] Our results are robust to imputation or lack of imputation of zeros in the page visits data. We do not have sufficient sample size to evaluate results without imputation of zeros for the financing data.

comparing means of scores given to foreign and local startups show that for the most part there is no difference in the quality measures between local and foreign firms with the exception of pre-accelerator traction, where local startups have a higher value on average by 6 percentage points (p=0.000), as well as post-accelerator logged financing, where local startups have a higher value on average by 5 percentage points (p=0.002). This occurs because U.S. and Canadian startups, which are more likely to be local to judges since the majority of our data are from U.S. startups and judges, have higher user traction and financing. This difference in traction suggests that controlling for differences in startup quality will be crucial. Table 1 also reveals that judges are less likely to recommend foreign startups and rate them as lower quality.

[Insert Table 1]

V. Empirical Specification

To assess whether judges systematically give lower or higher scores to foreign startups, we fit the following model:

$$(1)\ score_{ijt} = \ \propto + \beta foreign_{ij} + \ judge_{jt} + \mu_{it} + \ \epsilon_{ijt}$$

Where $score_{ijt}$ is either a z-scored average or a binary variable on whether judge j recommends startup i to the next round in year t. $foreign_{ij}$ is our binary variable indicating whether the region of startup i is different from that of judge j. Our main coefficient of interest is β, indicating whether judges discount startups from outside their home region.

We include a battery of fixed effects to identify judge effects from differences in startup quality. We account for judge harshness and judges participating across multiple years of the program through judge-year fixed effects ($judge_{jt}$), so that our analysis focuses on judge evaluations of startups within the same year.

We phase in several fixed effects to account for startup-level differences across years (since startups can apply in multiple years), as shown by μ_{it}. In our first specification, μ_{it} is equal to startup region-year fixed effects. These fixed effects measure startup evaluations within a particular region (e.g. Europe, Latin America, Israel, and Northern America) in a given year to account for differences in quality across regions.

We then tighten our specification, with μ_{it} equal to startup country-year fixed effects. These fixed effects focus our analysis on startup evaluations within a particular country in a year to account for differences in quality across countries (within regions). These fixed effects allow us to account for quality differences between, for example, a U.K.-based startup and a Latvia-based startup within Europe.

In our most stringent specification, we focus on evaluations at the startup level in a given year (across multiple judge evaluators), so that μ_{it} is equal to individual startup-year fixed effects. These fixed effects enable us to account for differences in individual startup quality within countries. We cluster robust standard errors at the judge and startup levels. If β is statistically significant, it means the judges discount or boost foreign startups relative to local ones. Returning to the two-by-three in Figure 1, this rules out cells A and E where judges are unbiased and either uniformed or informed. However, a significant β can be consistent with the remaining cells.

To assess whether foreign discounting is driven by judges being better at evaluating local startups or because of bias, we estimate a model similar in spirit to Li (2017) that measures the sensitivity of judges' scores to local vs. foreign startups' performance measures. This model allows us to discern the remaining scenarios in Figure 1, including whether judges are informed and biased (cell F), informed only about local startups and biased (cell D), informed only about local startups and unbiased (cell C), or uninformed about all startups and biased (cell B).

$$(2)\ score_{ijt} = \propto + \beta foreign_{ij} + \delta performance_i + \phi foreign_{ij} x performance_i + judge_{jt} + startupcountry_{it} + \epsilon_{ijt}$$

Where $performance_i$ indicates logged page visits for the startup one-year (for the 2018 cycle) or two-years (for the 2017 cycle) after the program. In addition to β, we are also interested in δ and ϕ. A positive and significant δ indicates that judges are able to discern winners from losers among startups overall. If δ is positive, then future performance correlates with judge scores. A negative and significant ϕ indicates that judges are less sensitive to the quality of foreign versus local startups. A concern with our approach is that the accelerator itself impacts the post-accelerator performance of startups, which confounds the judges' selection of startups with the treatment effect of the accelerator. Further, this treatment effect might differ for startups from different regions. To account for these possible treatment effects, we control for startups' participation in the accelerator program, and this participation interacted with whether the startup is foreign or local to the judge.

VI. Results

Are foreign judges actually randomly assigned?

Our ability to measure the presence and impact of foreign discounting hinges on the assumptions that startups and judges are randomly assigned. To check random assignment, we use chi-squared tests shown in Tables 2a-b. These chi-squared tests allow us to measure whether there is a difference between a predicted distribution of startup-judge regions under random assignment versus the actual distribution of pairs observed in the data. In 2017, there is no difference (p=0.809) between the predicted distribution of startup-judge region assignments under random allocation and the observed distribution. Thus, we cannot reject the null hypothesis that startup-judge assignments on the basis of geography are random. In 2018, we see that we can reject this null hypothesis because of the perhaps non-random assignment of Israeli judges to European startups (p=0.006), a fairly small share (0.26%) of our sample, representing 25 judge-startup pairings out of 9,733 total in 2018. However, when we take out Israeli judges, we see a similar situation as in 2017 (p=0.256). The distribution is again consistent with random assignment. Our results hold if we include or exclude these Israeli judges from our data. These patterns suggest that the natural experiment that is at the heart of our story is in fact randomized.

[Insert Tables 2a-b]

Is there foreign discounting of startups?

We now turn to whether judges discount foreign startups. In Table 1, summary statistics of scores for startups that match the geography of the judge show that, on average, the main composite score, recommend, and subscores are lower for startups that do not match geographies versus those that do.

Figure A1 also reveals that the distribution of scores from judge evaluations of foreign startups are lower on average than those of local startups. We confirm in a two-sample Kolmogorov-Smirnov test that the two distributions are different from one another (p=0.000). However, this graph may reflect the fact that most judges in our sample are U.S.-based. Thus, startups that are foreign are more likely to be those that are non-U.S. based, and non-U.S. based startups may be worse quality on average than U.S.-based firms.

We account for these regional quality differences in our regression models. To begin, model 1 in Table 3 shows that when we only control for judge-year fixed effects, judges give 0.2 standard deviation lower scores to foreign vs. local startups (p=0.000). Column 2 adds in startup region-year fixed effects to account for regional variations among startups. Our estimate shrinks to -0.06 standard deviation (p=0.002). Columns 3-4 add more restrictive startup country-year and startup-year fixed effects, respectively. Our results are virtually identical. These results show that there is little in the way of systematic differences between startups within regions. Overall, Table 3 shows that regional differences in startup quality account for about two-thirds of the foreign discounting effect, and judges account for one-third.

Column 5 includes measures for whether a startup has user traction and financing at the time of the application. Controlling for these pre-accelerator quality measures allows us to benchmark the judge bias effect against the effect of key startup milestones. The home bias effect (-0.06, p=0.001) is about 30 percent of the size of a startup having user traction and about 8 percent of the size of the effect of a startup having raised a round of financing at the time of the application. The fact that the whim of a judge matters above one-third as much as having some traction suggests that the foreign bias effect is non-trivial.

[Insert Table 3]

Table 4 is similar to Table 3, but uses our binary measure of whether a judge recommended a startup to the next round of the competition as the dependent variable. Judges are less likely to recommend foreign vs. local startups to the next round by 9 percentage points (p=0.000) before accounting for startup quality differences. This coefficient remains significant and negative, but falls to 4 percentage points (p=0.000) when accounting for startup region-year fixed effects (column 2), startup country-year fixed effects (column 3), and startup-year fixed effects (column 4), indicating that judge preferences account for about 40 percent of the foreign bias effect.

[Insert Table 4]

Together, these results reveal that judges consistently give lower evaluation scores to foreign versus local startups.

Is foreign discounting the result of judges being better evaluators of local startups?

We now turn to testing if this foreign bias is the result of differences in judges' expertise or is rooted in a preference for local vs. foreign firms. To begin, we assess whether judges are able to select winners from losers amongst all startups no matter their origins. Figure 2 shows a binscatter graph depicting the relationship between startups' website page visits 1-2 years after the program (x-axis) and the scores given by judges (y-axis), after netting out judge-year and startup country-year fixed effects, as well as startups' participation in the accelerator. The graph shows that better performing startups are given higher scores. Judges can pick winners from losers in the full sample.

[Insert Figure 2]

To what extent is this ability to detect the quality of startups driven by evaluations of local startups? To answer this question, in Figure 3, we split the evaluations into startups that are foreign to the judge (dotted line) and startups that are local to the judge (solid line). We see that both lines have a positive slope, suggesting that judges can separate high potential startups from those destined to fail. The fact that the solid line depicting local startup evaluations is above the dashed line across the quality spectrum suggests that judges give an across-the-board penalty to foreign startups no matter their quality. Further, the solid and dashed lines are similarly sloped. It does not appear that judges are better able to pick winners from losers among local versus among foreign startups. Figure 3 matches cell F in Figure 1 and so suggests that judges are informed about local and foreign startups, but are simply biased against foreign firms.

[Insert Figure 3]

We next turn to regressions to further confirm that judges are not any better at evaluating local startups. Column 1 in Table 5 reveals that there is no difference in the relationship between startup quality and judge scores by local startup origin, as seen in the coefficient on the interaction term between foreign startups and logged post-page visits ($foreign_{ij}xperformance_i$) (p=0.921). Consistent with Figure 3, we do indeed find that judge scores correlate with startup quality, shown by the positive coefficient on the main effect for logged post-page visits. In column 2, we control for accelerator participation and the possibility that accelerator participation matters more for foreign firms. While accelerator participation has a positive effect on post-accelerator startup page visits, and while this effect is slightly greater for local startups, it does not meaningfully account for the foreign discounting effect nor a judge's ability to evaluate startup potential.

These results suggest that judges are able to detect the quality of all startups with relatively equal precision. This is in in contrast to work that shows investors have a local information advantage (Malloy, 2005; Coval and Moskowitz, 2001).

[Insert Table 5]

Why then do judges consistently discount foreign versus local startups? On the one hand, while informed about quality on average, judges might be more uncertain in their evaluations of foreign firms and so more likely to discount them. On the other hand, judges might simply prefer local startups. To address this question, we separate our sample of startups into firms that have stronger versus weaker quality signals – in particular whether they have raised financing and/or have user traction. If the first channel accounts for foreign discounting, then the presence of strong quality signals (e.g. financing) should mitigate the foreign discounting effect.

Surprisingly, in Table 6, we find that judges' bias against foreign startups is stronger among startups with financing (-0.2, p=0.000) than among those without financing (-0.04, p=0.028). Using our full sample, in column 3, we confirm that the magnitude of the foreign bias is larger by including an interaction term between whether a startup is foreign and whether it has raised financing (p=0.000).[12] We find similar results for user traction in Table A1.

[Insert Table 6]

These results point to foreign discounting being rooted in differences in judges' preferences rather than differences in uncertainty. The fact that the bias is larger for startups that raised funding is puzzling. One possible explanation is that the financing information reduces noise that is inherent in judges' original evaluations. While this noise may be positive or negative, it might push estimates toward zero even when there is a large effect, similar to classical measurement error in regressions. With the provision of quality information – financing in this case – there is less idiosyncratic noise around judge evaluations perhaps making it easier to recover the estimated discounting effect.

Overall, our results are consistent with a model in which judges are informed about the quality of startups and are biased against foreign firms.

Robustness

We run a number of robustness checks. We first show that our foreign bias result holds in the broader sample of data from the accelerator, which suggests our results generalize beyond the

---

[12] We also check whether the adjusted R-squared differs across the financed and non-financed samples. If judges rely less on common geography when other quality signals are available, then the adjusted R-squared should be lower for the financed than for the non-financed sample. We find that the adjusted R-squared is similar for financed (0.59) and for non-financed (0.63) samples.

virtual event we analyze in the 2017-2018 global rounds. Specifically, we use data from 2013-2019 across all four rounds of the accelerator's competition. We find that judges give a 0.06 standard deviation (p=0.000) discount to foreign startups and are 4 percentage points (p=0.000) less likely to recommend foreign startups to the next round of the competition in this larger sample (Table A2). While these rounds do not exhibit random assignment, the fact that this estimate from the full data on 75,340 evaluations[13], is consistent with earlier results on foreign bias lends further credence to the idea that judges discount foreign startups.

We also confirm that our estimate is robust to other measures of startup quality. Figures A2-A4 show that the ability of judges to pick winners and losers among foreign versus local startups holds with startup pre-accelerator financing and page visit measures, as well as a post-accelerator financing measure. We also confirm the results hold in regression models.[14]

Lastly, we confirm that the foreign discounting effect is not the result of judges from a particular region being especially harsh to foreign firms. In Figure A5, we show the average local and foreign recommendations for North American, European, Latin American, and Israeli judges. In Panel A, we see that U.S. and European judges are less likely to recommend foreign startups relative to local ones, though this does not hold for Latin American and Israeli judges. This likely is because of the lower quality of Latin American startups[15] as well as the relatively small share of both Israeli and Latin American firms in our sample.[16] When we control for a minimum threshold of quality by limiting our sample to financed startups in Panel B, we find

[13] These evaluations only include startups who come from the same regions as judges' home regions to ensure that startups are both local and foreign to judges in our sample.
[14] The authors are happy to show these regression results upon request. They are omitted from this manuscript because of space limitations.
[15] Latin American startups have the lowest probability of being recommended to the next round relative to startups from other regions. These startups are over 20 percent less likely to be recommended to the next round compared to other startups in our main sample.
[16] Latin American startups comprise less than 15 percent of all startups in our main sample, and Israeli startups comprise 9 percent of all startups in our main sample.

that judges from all regions but Latin America strongly discount foreign startups. The effect is not the result of a single country being particularly harsh towards foreign firms.

Does foreign discounting cause judges to pass on promising foreign startups?

Our results above show that across the quality distribution, judges give lower scores to foreign startups. However, it is possible that this discounting has little impact on which startups judges select for the next round. For example, perhaps judges discount high quality foreign startups who, though rated lower, are still selected for the next round. Conversely, judges may discount low quality foreign startups who would not make it to the next round regardless. In these extreme cases, foreign discounting would not impact the marginal decision. However, for startups in the middle of the quality distribution, this foreign discounting may have a substantial impact.

To estimate the number of "missed foreign startups," for whom foreign discounting does make a marginal difference, we estimate what judge decisions would be if they only relied on quality dependent measures and not on the startup's foreign status. To isolate the quality dependent portion of the judges' scores, we regressed judge decisions on our startup quality measures. While crude, this model allows us to recover the judges' weights on different measures of startup quality – both ex-ante and ex-post – and so construct counterfactual rankings as if judges are unbiased but still selected the same number of startups.[17] We then compare this ranking to two alternatives. The first is the *actual* recommendation of the judge. The second is "biased" counterfactual rankings that use the quality measures and whether the startup is foreign to generate deliberately foreign-biased recommendations. The first alternative tells us how much relying only on quality measures would increase the number of foreign startups. The second

[17] If foreign startups are lower quality, then judge could still discount them. However, our argument is that judges have a direct bias against foreign startups that is not mediated by quality.

reveals how many foreign startups are missed when we introduce foreign bias on top of “unbiased" quality-based evaluations.

In these back-of-the-envelope counterfactuals we find that foreign bias impacts the number of foreign startups that are recommended to the next round of the competition. We find that moving to evaluations only based on quality leads to 512 more foreign startups being recommended, accounting for 14 percent of the startups in our sample. When we introduce foreign bias onto the quality-based recommendations, 324 fewer foreign startups are recommended. These differences suggest that foreign bias leads judges to overlook 9-to-14 percent of startups that at least based on our quality measures should have been recommended to the next round.

VII. Conclusion and Implications

We find that judges can discern the quality of local and foreign startups with similar ability. However, they discount foreign startups no matter their potential. Judges are less likely to recommend foreign startups by 4 percentage points, equivalent to roughly one third of the effect of having some user traction, or a tenth of the effect of going from no financing to some venture capital or angel financing. Stronger quality signals (e.g. financing) do not appear to attenuate this bias. Back-of-the-envelope estimates suggest that this bias results in the potential exclusion of over 1 in 10 highly promising entrepreneurial ideas. These results reveal that judges are informed about quality of both local and foreign startups, but they are biased against foreign firms. Therefore, while they have an ability to select the most promising companies, they systematically discount foreign startups relative to local ones.

Our results suggest that startups from remote locations may fail when they try to move or scale to “hubs” – despite lower costs to such movements because of the advent of IT and cloud technology. Indeed, recent work reveals such home bias in online investment platforms (Lin and Viswanathan, 2016). The bias may contribute to explanations to why ventures tend to perform better when located in the native regions of founders (Dahl and Sorenson, 2012) and why investment from “high-status” venture capital firms is most beneficial when the home country of the venture capital firm is more connected to that of the venture (Alvarez-Garrido and Guler, 2018).

Foreign bias also may impact the direction of innovation. If accelerators select out startups from remote regions, which are more likely to be foreign to accelerators or investors, they reduce the probability that innovations addressing the needs of those markets will survive and grow. This distortion is similar to effects seen in studies of bias in gender and race contexts (e.g. Koning, Samila, and Ferguson, 2020).

While we find that startups face a “liability of foreignness” (Zaheer, 1995), with the across-the-board discount given by foreign judges, we notably do not find that judges face a disadvantage in evaluating foreign startups. Instead, we find that judges can discern quality of startups across regions. This may be because startup practices have standardized into a “playbook” that is comparable across countries, for example, with the proliferation of codified management (Bloom and Reenen, 2007; Chatterji, Delecourt, Hasan, and Koning, 2019) and technology practices (Haefliger, Von Krogh, and Spaeth, 2008). The existence of such a playbook may reduce the need for private information (Malloy, 2005) or contextual intelligence (Khanna, 2014) to evaluate foreign opportunities.

Together, our results show that accelerators and investors may benefit from looking more globally for startups, given their ability to discern startup quality. This global search need not be done in localized programs, as is the current standard in accelerator and venture capital models with local subsidiaries around the world. For example, California-based venture capital firm Sequoia has offices in eight other cities around the world. Rather, the global search may occur directly out of the headquarters location. This global approach, however, can only yield a high quality pool of startups if organizations revise their processes to reduce bias against foreign startups either in aggregate, lowering the threshold for selecting a foreign firm versus a local firm, or at an individual judge level. Further research should measure the effects of such interventions.

References

Aggarwal, Rohit, David Kryscynski, and Harpreet Singh. (2015). Evaluating venture technical competence in venture capitalist investment decisions. *Management Science*, *61*(11), 2685-2706.

Alvarez-Garrido, Elisa, and Isin Guler. (2018). Status in a strange land? Context-dependent value Of status in cross-border venture capital. *Strategic Management Journal*, *39*(7), 1887-1911.

Arrow KJ. (1962). Economic welfare and the allocation of resources for invention. Nelson R, ed. *The Rate and Direction of Inventive Activity: Economic and Social Factors* (pp. 609-626). Princeton, NJ: Princeton University Press.

Balachandran, S., & Hernandez, E. (2020). Mi Casa Es Tu Casa: immigrant entrepreneurs as pathways to foreign venture capital investments. Available at SSRN 3331264.

Bernstein, Shai, Xavier Giroud, and Richard R. Townsend. (2016). The impact of venture capital monitoring. *The Journal of Finance*, *71*(4), 1591-1622.

Bessemer Venture Partners. (2020). Anti-Portfolio. Retrieved at: https://www.bvp.com/anti-portfolio/.

Bloom, Nicholas, and John Van Reenen. (2007). Measuring and Explaining Management Practices across Firms and Countries. *Quarterly Journal of Economics, 122*(4), 1341-1408.

Boudreau, Kevin J., Eva C. Guinan, Karim R. Lakhani, and Christoph Riedl. (2016). Looking across and looking beyond the knowledge frontier: Intellectual distance, novelty, and resource allocation in science. *Management Science, 62*(10), 2765-2783.

Brynjolfsson, Erik, Xiang Hui, and Meng Liu. (2019). Does machine translation affect international trade? Evidence from a large digital platform. *Management Science, 65*(12), 5449-5460.

Cao, Ruiqing, Rembrand Koning, and Ramana Nanda. (2020). Biased Sampling of Early Users and the Direction of Startup Innovation. Harvard Business School working paper 21-059, Harvard Business School, Boston, MA. Available at: https://www.hbs.edu/ris/Publication%20Files/21-059_0ec587e4-4b74-4f21-9f8c-3f8920649ab9.pdf.

Catalini, C., Guzman, J., & Stern, S. (2019). *Hidden in plain sight: venture growth with or without venture capital* (No. w26521). National Bureau of Economic Research.

Chatterji, Aaron, Solène M. Delecourt, Sharique Hasan, and Rembrand Koning. (2019). When does advice impact new venture performance?. *Strategic Management Journal, 40*(3), 331-356.

Cohen, S. L., Bingham, C. B., & Hallen, B. L. (2019). The role of accelerator designs in mitigating bounded rationality in new ventures. *Administrative Science Quarterly*, *64*(4), 810-854.

Cohen, S., Fehder, D. C., Hochberg, Y. V., & Murray, F. (2019). The design of startup accelerators. Research Policy, 48(7), 1781-1797.

Conti, Annamaria, Jorge Guzman, and Ron Rabi. (2020). Information Frictions in the Market for Startup Acquisitions. *Available at SSRN 3678676*.

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

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

Damouni, Nadia and Bill Rigby. (2011). Microsoft to buy Skype for pricey $8.5 billion." Retrieved from:

https://www.reuters.com/article/us-skype-microsoft/microsoft-to-buy-skype-for-pricey-8-5-billion-idUSTRE7490F020110510.
Dahl, M. S., & Sorenson, O. (2012). Home sweet home: Entrepreneurs' location choices and the performance of their ventures. *Management science*, *58*(6), 1059-1071.
Disdier, A. C., & Head, K. (2008). The puzzling persistence of the distance effect on bilateral trade. *The Review of Economics and statistics*, *90*(1), 37-48.
Fehder, D. C., & Hochberg, Y. V. (2014). Accelerators and the regional supply of venture capital investment. *Available at SSRN 2518668*.
Franke, N., Gruber, M., Harhoff, D., & Henkel, J. (2006). What you are is what you like—similarity biases in venture capitalists' evaluations of start-up teams. *Journal of Business Venturing*, *21*(6), 802-826.
Gans, J. S., Hsu, D. H., & Stern, S. (2008). The impact of uncertain intellectual property rights on the market for ideas: Evidence from patent grant delays. *Management science*, *54*(5), 982-997.
Ghemawat, P., & Altman, S. A. (2019). The state of globalization in 2019, and what it means for strategists. *Harvard Business Review*, 2-8.
Gompers, P. A., Gornall, W., Kaplan, S. N., & Strebulaev, I. A. (2020). How do venture capitalists make decisions?. *Journal of Financial Economics*, *135*(1), 169-190.
Gonzalez-Uribe, J., & Leatherbee, M. (2018). The effects of business accelerators on venture performance: Evidence from start-up chile. *The Review of Financial Studies*, *31*(4), 1566-1603.
Haefliger, S., Von Krogh, G., & Spaeth, S. (2008). Code reuse in open source software. *Management science*, *54*(1), 180-193.
Haley, Colin C. (2005). The ones that got away. Retrieved from: https://www.enterprisenetworkingplanet.com/unified_communications/Skype-and-Vonage-The-Ones-That-Got-Away-3550606.htm.
Hall, R. E., & Woodward, S. E. (2010). The burden of the nondiversifiable risk of entrepreneurship. *American Economic Review*, *100*(3), 1163-94.
Hallen, B. L., Cohen, S. L., & Bingham, C. B. (2020). Do accelerators work? If so, how?. *Organization Science*, *31*(2), 378-414.
Hegde, D., & Tumlinson, J. (2014). Does social proximity enhance business partnerships? Theory and evidence from ethnicity's role in US venture capital. *Management Science*, *60*(9), 2355-2380.
Hoenig, D., & Henkel, J. (2015). Quality signals? The role of patents, alliances, and team experience in venture capital financing. *Research Policy*, *44*(5), 1049-1064.
Howell, S. T. (2017). Financing innovation: Evidence from R&D grants. *American Economic Review*, *107*(4), 1136-64.
Huang, L., & Pearce, J. L. (2015). Managing the unknowable: The effectiveness of early-stage investor gut feel in entrepreneurial investment decisions. *Administrative Science Quarterly*, *60*(4), 634-670.
Huberman, G. (2001). Familiarity breeds investment. *The Review of Financial Studies*, *14*(3), 659-680.
Kaplan, S. N., Sensoy, B. A., & Strömberg, P. (2009). Should investors bet on the jockey or the horse? Evidence from the evolution of firms from early business plans to public companies. *The Journal of Finance*, *64*(1), 75-115.
Kerr, W. R. (2016). Harnessing the best of globalization. *MIT Sloan Management Review*, *58*(1), 59.
Kerr, W. R., Lerner, J., & Schoar, A. (2014). The consequences of entrepreneurial finance: Evidence from angel financings. *The Review of Financial Studies*, *27*(1), 20-55.
Kerr, W. R., Nanda, R., & Rhodes-Kropf, M. (2014). Entrepreneurship as experimentation. *Journal of Economic Perspectives*, *28*(3), 25-48.
Khanna, T. (2014). Contextual intelligence. *Harvard business review*, *92*(9), 58-68.
Koning, R., Hasan, S., & Chatterji, A. (2019). Experimentation and startup performance: Evidence from A/B testing. NBER working paper 26278, National Bureau of Economic Research, Cambridge, MA. Available at: https://www.nber.org/papers/w26278.
Koning, R., Samila, S., & Ferguson, J. P. (2020, May). Inventor Gender and the Direction of Invention. In *AEA Papers and Proceedings* (Vol. 110, pp. 250-54).
Lee, M., & Huang, L. (2018). Gender bias, social impact framing, and evaluation of entrepreneurial ventures. *Organization Science*, *29*(1), 1-16.
Lerner, J., Schoar, A., Sokolinski, S., & Wilson, K. (2018). The globalization of angel investments: Evidence across countries. *Journal of Financial Economics*, *127*(1), 1-20.
Li, D. (2017). Expertise versus Bias in Evaluation: Evidence from the NIH. *American Economic Journal: Applied Economics*, *9*(2), 60-92.

Lin, M., Prabhala, N. R., & Viswanathan, S. (2013). Judging borrowers by the company they keep: Friendship networks and information asymmetry in online peer-to-peer lending. *Management Science*, *59*(1), 17-35.
Lin, M., & Viswanathan, S. (2016). Home bias in online investments: An empirical study of an online crowdfunding market. *Management Science*, *62*(5), 1393-1414.
Lu, J. W., & Beamish, P. W. (2001). The internationalization and performance of SMEs. *Strategic management journal*, *22*(6-7), 565-586.
Luo, H. (2014). When to sell your idea: Theory and evidence from the movie industry. *Management Science*, *60*(12), 3067-3086.
Malloy, C. J. (2005). The geography of equity analysis. *The Journal of Finance*, *60*(2), 719-755.
Nanda, R., Samila, S., & Sorenson, O. (2020). The persistent effect of initial success: Evidence from venture capital. *Journal of Financial Economics*, *137*(1), 231-248.
Niessen-Ruenzi, A., & Ruenzi, S. (2019). Sex matters: Gender bias in the mutual fund industry. *Management Science*, *65*(7), 3001-3025.
Oviatt, B. M., & McDougall, P. P. (2005). Toward a theory of international new ventures. *Journal of international business studies*, *36*(1), 29-41.
Scott, E. L., Shu, P., & Lubynsky, R. M. (2020). Entrepreneurial uncertainty and expert evaluation: An empirical analysis. *Management Science*, *66*(3), 1278-1299.
Sørensen, M. (2007). How smart is smart money? A two-sided matching model of venture capital. *The Journal of Finance*, *62*(6), 2725-2762.
Sorenson, O., & Stuart, T. E. (2001). Syndication networks and the spatial distribution of venture capital investments. *American journal of sociology*, *106*(6), 1546-1588.
X1 Group. (2018). 16 Unicorn Startups with R&D Offices in Ukraine. *Medium*. Retrieved from: https://medium.com/x1group/16-unicorn-startups-with-r-d-offices-in-ukraine-6243cd0ddd6
Y Combinator. (2019). Nine Business Models and the Metrics Investors Want (SUS 2019). Retrieved from https://www.ycombinator.com/library/8E-nine-business-models-and-the-metrics-investors-want-sus-2019.
Y Combinator. (2020). Top Companies. Retrieved from https://www.ycombinator.com/topcompanies/
Yin, B., & Luo, J. (2018). How do accelerators select startups? Shifting decision criteria across stages. *IEEE Transactions on Engineering Management*, *65*(4), 574-589.
Yu, S. (2020). How do accelerators impact the performance of high-technology ventures?. *Management Science*, *66*(2), 530-552.
Zaheer, S. (1995). Overcoming the liability of foreignness. *Academy of Management journal*, *38*(2), 341 363.

Tables

Table 1: Summary Statistics at the Evaluation Level

	Local Startup					Foreign Startup					
	Judge-Startup from the Same Region					Judge-Startup from Different Region					Local-Foreign
	No. Obs.	Mean	SD	Min	Max	No. Obs.	Mean	SD	Min	Max	Diff. in Means
Judge Score Measures											
Composite Score	7,232	0.01	1.01	-3.31	2.36	9,107	-0.12	1.05	-3.31	2.36	0.13***
Overall Raw Score	7,706	2.92	1.16	0.00	5.00	9,902	2.75	1.13	0.00	5.00	0.16***
Recommend	7,706	0.61	0.49	0.00	1.00	9,902	0.56	0.50	0.00	1.00	0.05***
Subscore: Customer Needs and Acquisition	7,692	6.25	1.85	1.00	10.00	9,833	6.06	1.92	1.00	10.00	0.18***
Subscore: Customer Pain and Solution	7,694	6.82	1.84	1.00	10.00	9,840	6.63	1.95	1.00	10.00	0.18***
Subscore: Financial Business Model	7,675	5.72	1.98	1.00	10.00	9,787	5.53	2.07	1.00	10.00	0.19***
Subscore: Industry and Competitor	7,690	6.11	1.85	1.00	10.00	9,827	5.93	1.94	1.00	10.00	0.17***
Subscore: Overall Impact	7,686	6.21	1.93	1.00	10.00	9,820	6.03	2.00	1.00	10.00	0.18***
Subscore: Regulation and IP	7,261	5.91	2.15	1.00	10.00	9,175	5.65	2.25	1.00	10.00	0.27***
Subscore: Team and Advisors Investors	7,678	6.51	2.01	1.00	10.00	9,805	6.31	2.09	1.00	10.00	0.20***
Startup Quality Measures											
Log Pre-Accelerator Total Page Visits	3,917	1.37	2.77	0.00	12.50	5,816	1.46	2.88	0.00	12.50	-0.09
Log Pre-Accelerator Financing	7,706	0.45	1.41	0.00	6.03	9,902	0.41	1.33	0.00	6.03	0.04
Log Post-Accelerator Total Page Visits	7,706	2.87	3.52	0.00	12.82	9,902	2.93	3.61	0.00	12.82	-0.06
Log Post-Accelerator Financing	7,706	0.30	1.10	0.00	5.95	9,902	0.25	0.98	0.00	5.92	0.05**
Has User Traction	7,706	0.59	0.49	0.00	1.00	9,902	0.53	0.50	0.00	1.00	0.06***
Has Financing	7,706	0.12	0.33	0.00	1.00	9,902	0.11	0.32	0.00	1.00	0.01

Notes: The table reports descriptive statistics for the sample of 17,608 startup-judge pairings from the 2017 and 2018 global rounds.
* $p<0.05$ ** $p<0.01$ *** $p<0.001$

Table 2a: Chi-squared Table for the 2017 global round showing distribution of judges to startups is no different than what we would expect from random chance

Pearson chi2(4) = 1.5988 Pr = 0.809

	Judge Subregion			
Startup Subregion	Europe	U.S. & Canada	Israel	Total
Europe	229	791	206	1,226
	239.3	*783.7*	*203*	
U.S. & Canada	1,008	3,322	860	5,190
	1,013.00	*3,317.60*	*859.4*	
Israel	300	921	238	1,459
	284.8	*932.6*	*241.6*	
Total	1,537	5,034	1,304	7,875

Non-italicized numbers indicate observed frequency. Italicized numbers indicate the expected frequency of the cell counts if they were randomly assigned based on the marginal distributions.

Table 2b: Chi-squared table for the 2018 global round showing distribution of judges to startups is no different than what we would expect from random chance when excluding outliers

With Israeli Judges: Pearson chi2(9) = 22.9832 Pr = 0.006
Without Israeli Judges = Pearson chi2(6) = 7.7603 Pr = 0.256

	Judge Subregion			
Startup Subregion	Europe	Latin America	U.S. & Canada	Israel
Europe	568	153	1,389	25
	595.8	*177.7*	*1,348.20*	*13.4*
Latin America	705	213	1,539	11
	688.7	*205.4*	*1,558.40*	*15.5*
U.S. & Canada	1,406	432	3,134	23
	1,393.90	*415.7*	*3,154.10*	*31.3*
Israel	37	12	84	2
	37.7	*11.2*	*85.2*	*0.8*
Total	2,716	810	6,146	61

Non-italicized numbers indicate observed frequency. Italicized numbers indicate the expected frequency of the cell counts if they were randomly assigned based on the marginal distributions.

Table 3: Regressions showing that judges give lower scores to startups from outside their home region even when we control for judge and startup fixed effects

	(1)	(2)	(3)	(4)	(5)
	Judge's Total Score				
Foreign Startup	-0.204***	-0.061**	-0.061**	-0.061***	-0.058**
	(0.021)	(0.020)	(0.020)	(0.016)	(0.018)
Has User Traction					0.201***
					(0.029)
Has Financing					0.712***
					(0.023)
Observations	16,320	16,320	16,320	16,264	16,320
Judge x Year	Yes	Yes	Yes	Yes	Yes
Startup Region x Year	No	Yes	No	No	No
Startup Country x Year	No	No	Yes	No	Yes
Startup x Year	No	No	No	Yes	No

Of the 17,608 recommendation evaluations in our data, for 16,339 (93%) we have complete subscore information. Standard errors (shown in parentheses) are clustered at the judge and startup levels. Fixed effects shown below observations.
* p<0.05 ** p<0.01 *** p<0.001

Table 4: Regressions showing that judges are less likely to recommend startups from outside their home region even when we control for judge and startup fixed effects

	(1)	(2)	(3)	(4)	(5)
	Judge Recommends Startup?				
Foreign Startup	-0.091***	-0.036***	-0.038***	-0.039***	-0.037***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Has User Traction					0.088***
					(0.015)
Has Financing					0.345***
					(0.010)
Observations	17,593	17,593	17,593	17,590	17,593
Judge x Year	Yes	Yes	Yes	Yes	Yes
Startup Region x Year	No	Yes	No	No	No
Startup Country x Year	No	No	Yes	No	Yes
Startup x Year	No	No	No	Yes	No

Standard errors (in parentheses) are clustered at the judge and startup levels. Fixed effects shown below observations.
* p<0.05 ** p<0.01 *** p<0.001

Table 5: Regressions showing judges (1) give higher scores to more successful startups, (2) are equally good at predicting success for local and foreign startups alike, and (3) still discount foreign startups

	(1)	(2)
	Judge's Total Score	
Foreign Startup	-0.065**	-0.056*
	(0.025)	(0.024)
Log Post-Page Visits	0.050***	0.036***
	(0.004)	(0.004)
Foreign Startup * Log Post Accelerator Page Visits	0.000	0.003
	(0.005)	(0.005)
Foreign Startup * Accelerator Participation		-0.109**
		(0.041)
Accelerator Participation		0.682***
		(0.032)
Observations	16,320	16,320
Judge xYear	Yes	Yes
Startup Country xYear	Yes	Yes
Startup xYear	No	No

Standard errors (in parentheses) are clustered at the judge and startup level. Fixed effects shown below observations.
* p<0.05 ** p<0.01 *** p<0.001

Table 6: Regressions showing the foreign startup discount is larger for financed vs. non-financed startups

	(1)	(2)	(3)
		Judge's Total Score	
Foreign Startup	-0.038*	-0.191***	-0.042*
	(0.017)	(0.045)	(0.017)
Foreign Startup * Has Financing			-0.148***
			(0.040)
Observations	14,269	1,560	16,264
Financed Startups	No	Yes	Yes
Non-Financed Startups	Yes	No	Yes
Judge x Year	Yes	Yes	Yes
Startup x Year	Yes	Yes	Yes

Standard errors (shown in parentheses) are clustered at the judge and startup levels. Fixed effects shown below observations.
* p<0.05 ** p<0.01 *** p<0.001

Figures

Figure 1: Predicted relationships between judge scores and startup quality

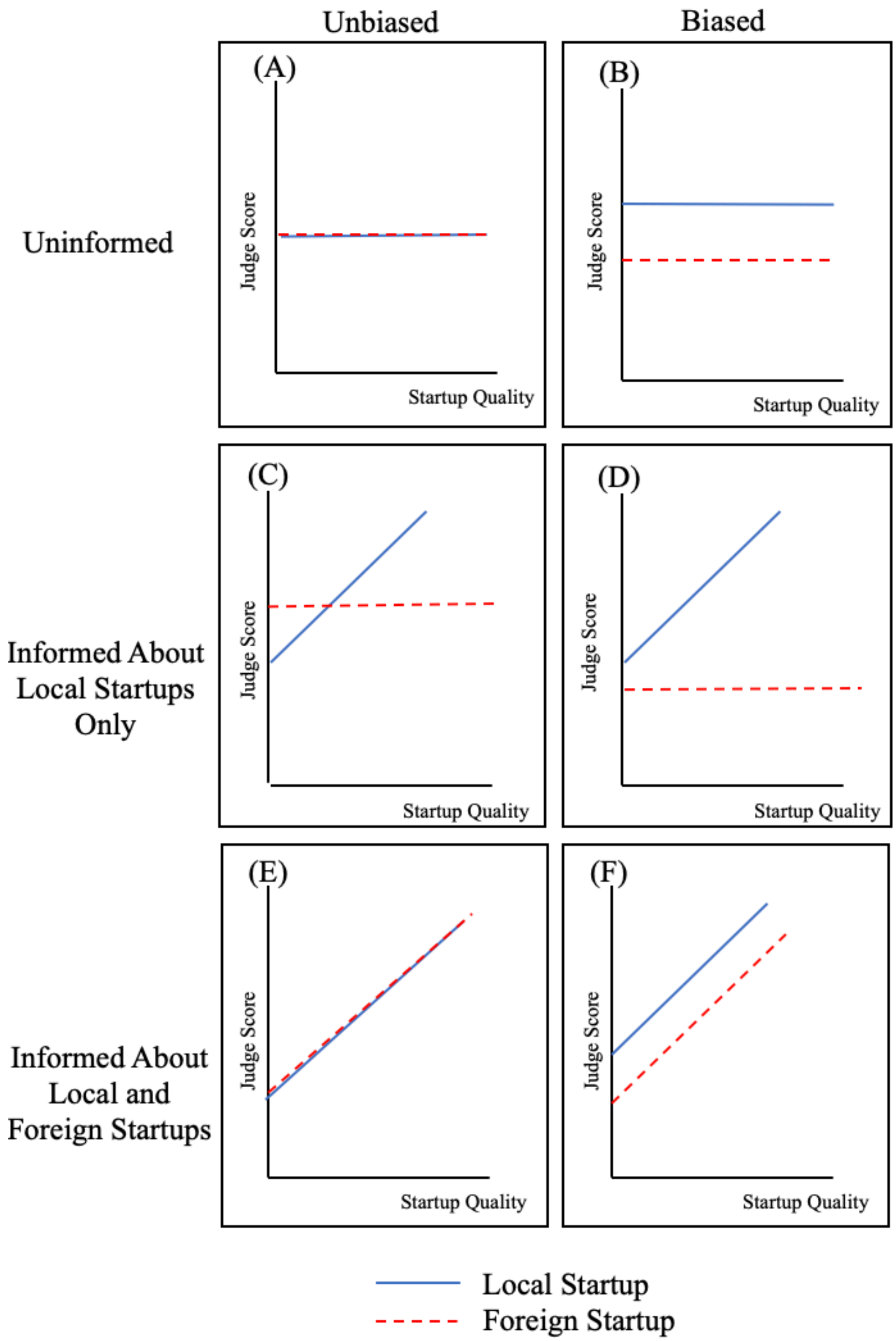
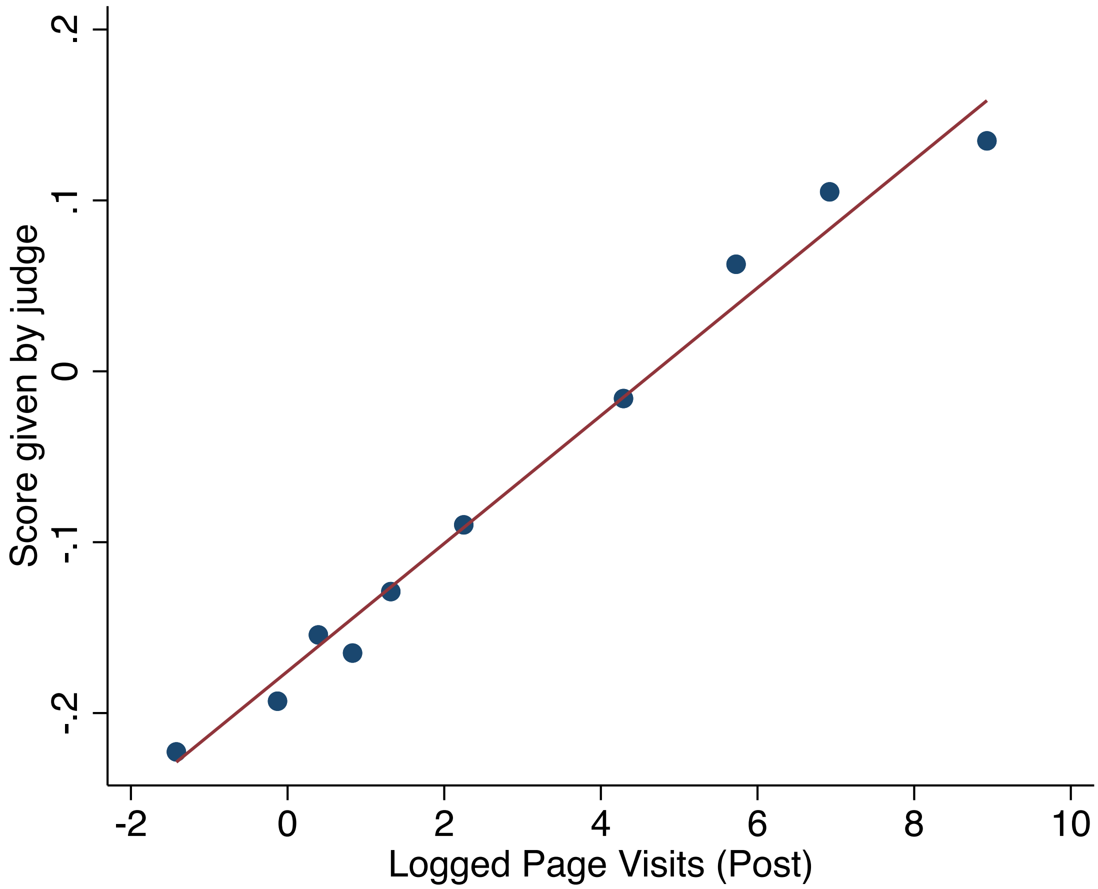
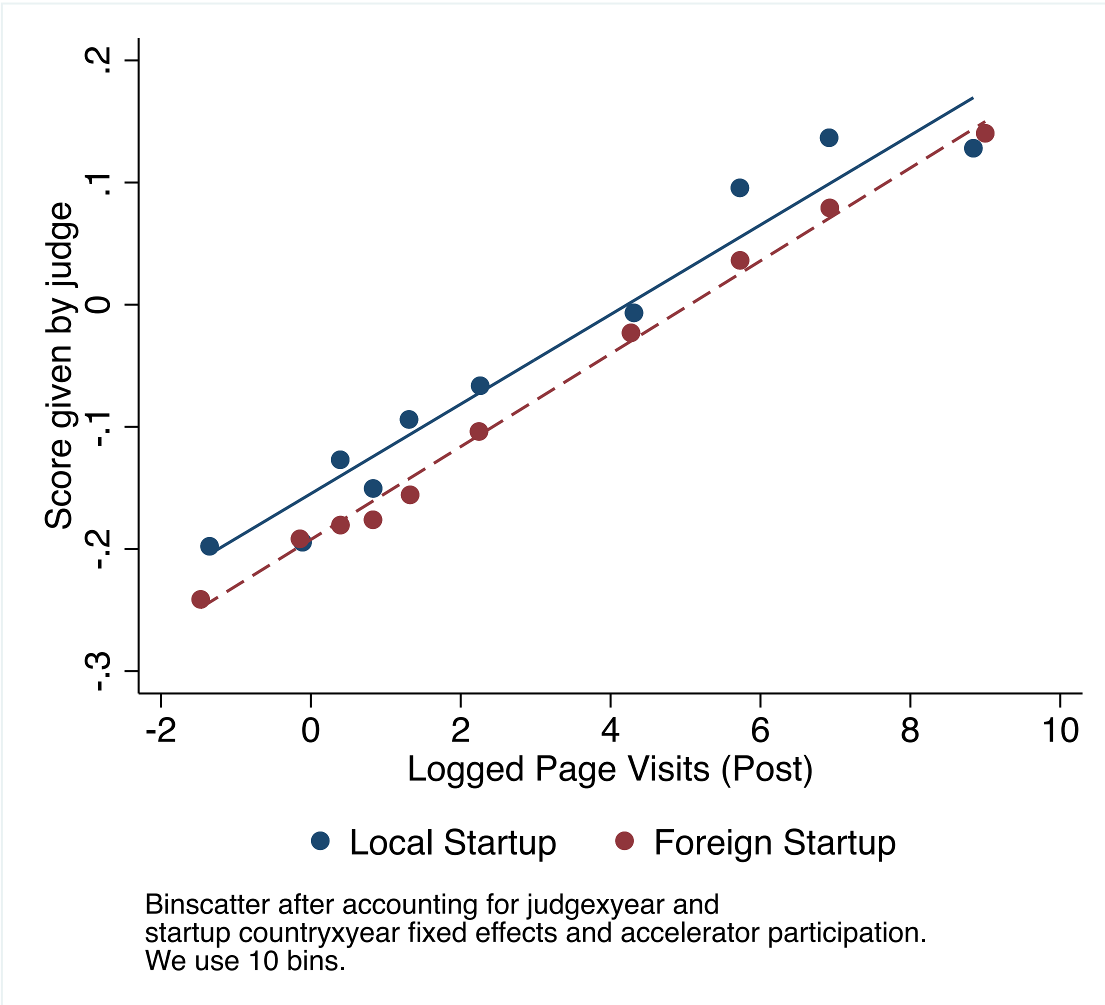


Figure 2: Binscatter showing that judges give higher scores to startups with more growth one- to- two- years after the accelerator program



Binscatter after accounting for judgexyear and
startup countryxyear fixed effects and accelerator participation.
We use 10 bins.

Figure 3: Binscatter showing that judges give higher scores to startups with more growth one- to- two- years after the program, but that they consistently discount foreign startups no matter their eventual success



Appendix

Tables

Table A1: Regressions showing that judges discount foreign vs. local startups with user traction, more so than foreign vs. local startups without user traction

	(1)	(2)	(3)
	Judge's Total Score		
Foreign Startup	-0.047*	-0.075**	-0.049*
	(0.024)	(0.024)	(0.023)
Foreign Startup * Has User Traction			-0.021
			(0.032)
Observations	7,201	8,889	16,264
Startups w/ Page Traction	No	Yes	Yes
Startups w/ No Page Traction	Yes	No	Yes
Judge x Year	Yes	Yes	Yes
Startup x Year	Yes	Yes	Yes

Standard errors (shown in parentheses) are clustered at the judge and startup levels. Fixed effects shown below observations.
* p<0.05 ** p<0.01 *** p<0.001

Table A2: Regressions showing that judges are less likely to recommend startups from outside their home region even when we control for judge and startup fixed effects in the full sample

	(1)	(2)
	Judge's Total Score	Judge Recommends Startup?
Foreign Startup	-0.064***	-0.040***
	(0.012)	(0.007)
Observations	69,639	75,153
Judge x Year	Yes	Yes
Startup x Year	Yes	Yes
Program x Year	Yes	Yes

The table shows evaluations across rounds in 2013-2019. Standard errors (shown in parentheses) are clustered at the judge and startup levels. Fixed effects shown below observations.
* p<0.05 ** p<0.01 *** p<0.001

Figures

Figure A1: Kernel density plot of scores by whether the judge and startup are from the same region (local startup) or from different regions (foreign startup)

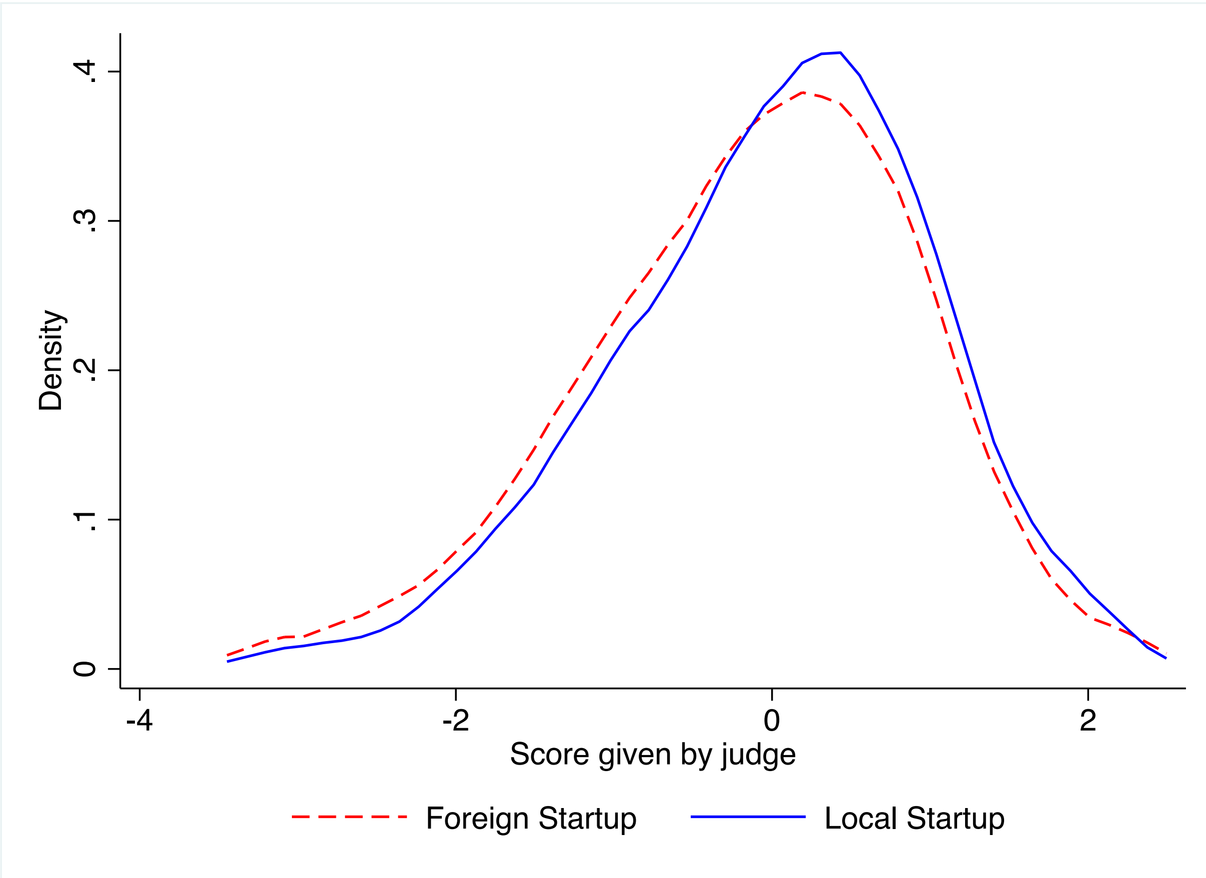


Figure A2: Binscatter showing that judges can give higher scores to startups with more growth before entering the accelerator program, but consistently discount foreign startups no matter their growth

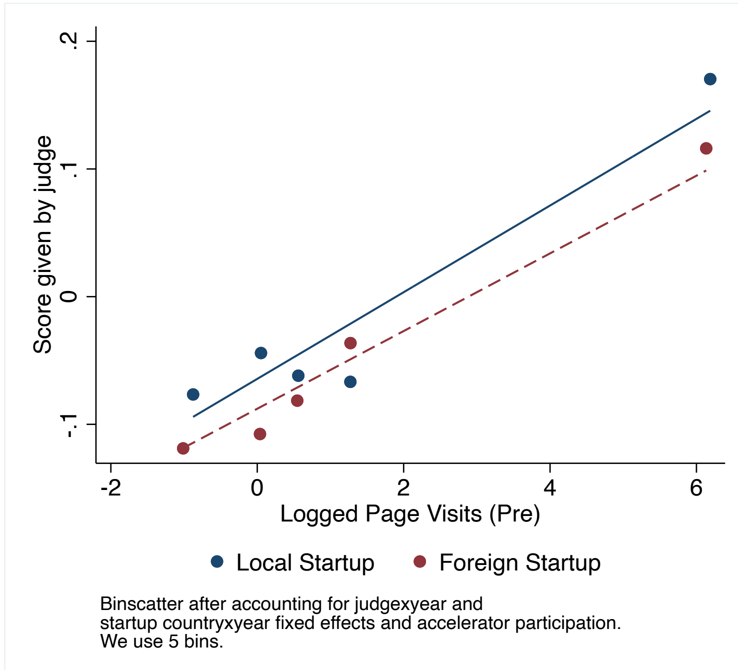


Figure A3: Binscatter showing that judges can give higher scores to startups with more financing before entering the accelerator program, but consistently discount foreign startups no matter their financing

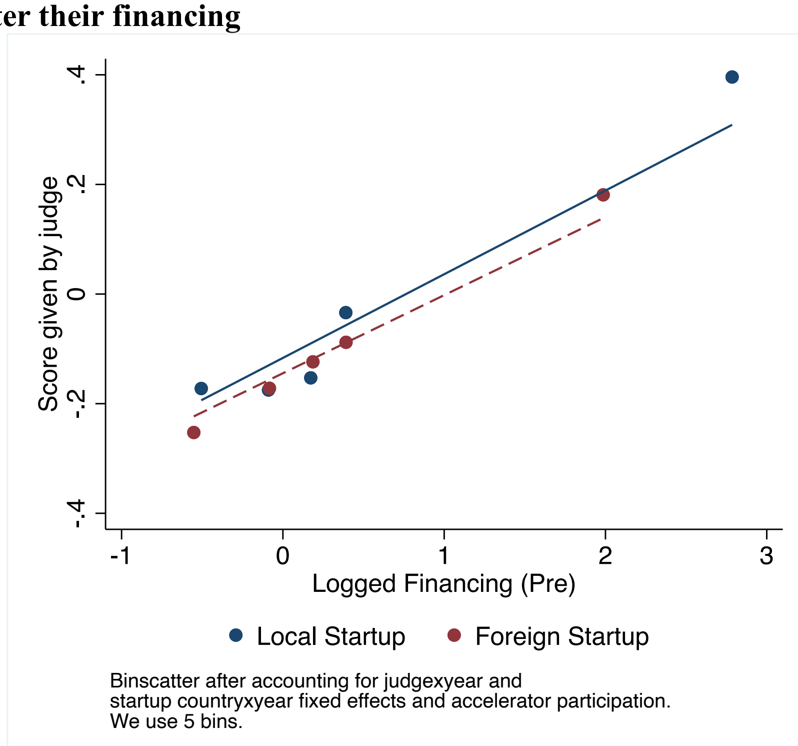


Figure A4: Binscatter showing that judges can give higher scores to startups with more financing after entering the accelerator program, but consistently discount foreign startups no matter their eventual financing

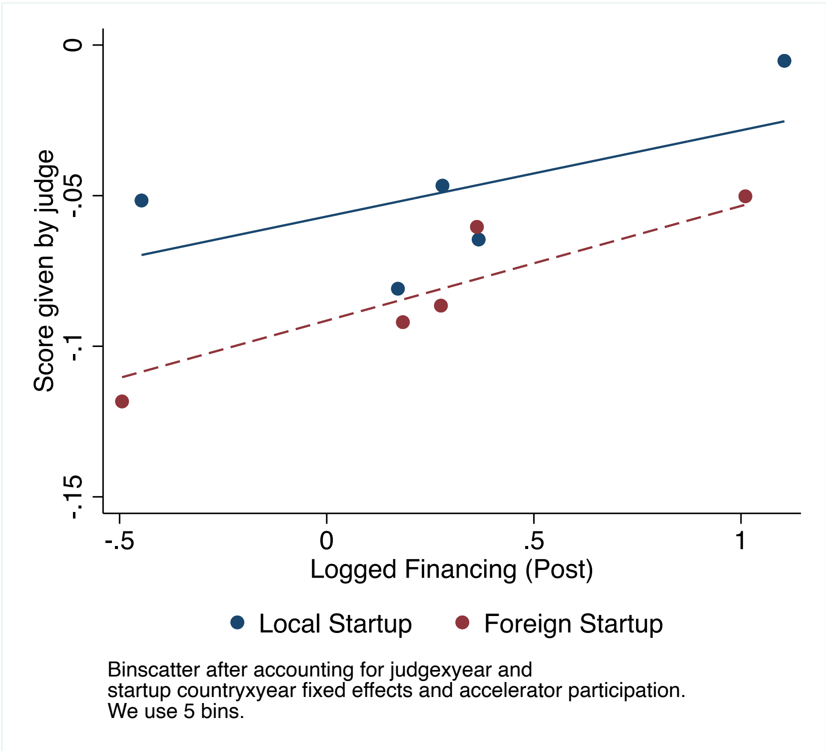
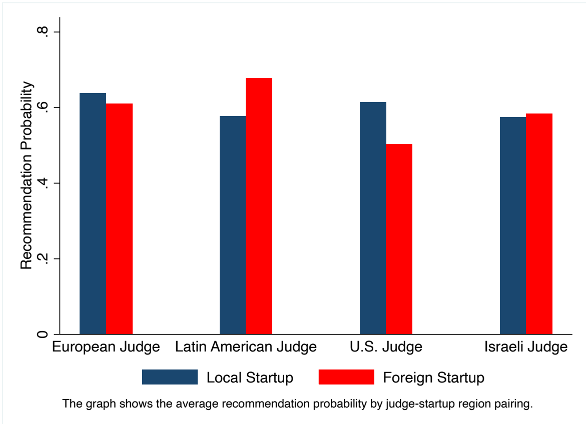
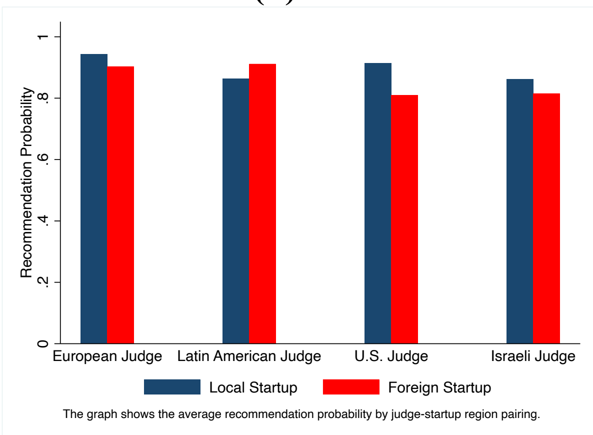


Figure A5: Bar graph showing that US, EU, and Israeli judges are more likely to recommend local over foreign startups to the next round of the competition. Panel A is for all startups and Panel B is only for startups that have raised financing.

(A)



(B)



A7: Application Questions

Company Background

- Full time employees – the number of full time employees currently in your company.
- Part time employees – the number of part time employees currently in your company.
- Interns/volunteers – the total number of interns or volunteers in your company.

Customer Pain and Solution

- Problem – please describe what problem (customer pain point) you are trying to solve.
- Solution – what is your solution?

Overall Impact

- Define the 1 year and 5 year impact that you home to accomplish – use whatever metrics are most appropriate for you (e.g. revenue, profit, jobs, societal benefits).

Customer Needs and Acquisition

- How would you define your potential market and what is the addressable market size?
- What traction have you made to date with market validation?
- Marketing – what will be your messaging to users/customers and how do you plan to spread it?
- Sales and distribution – how will you reach your customers?

Industry and Competitors

- Which organizations compete with your value offering now, and who might do so in the future?
- Which organizations complement your offering in the market?
- What are the primary advantages relative to existing or potential competitors?

Business Model/Financials

- What are the key drivers of business economics (price points, margins, etc.)?

Regulation and IP

- What intellectual property or regulatory requirements exist for your business or in your industry?

Founding Team and Advisors/Investors

- Please share some background information on your team members.
- Please tell us about current or anticipated advisors and investors.