Working Paper 21-097

Judging Foreign Startups

Nataliya Langburd Wright
Rembrand Koning
Tarun Khanna

Harvard Business School

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[1]

Nataliya Langburd Wright (Harvard Business School)

Rembrand Koning (Harvard Business School)

Tarun Khanna (Harvard Business School)

January 13, 2023

Research Summary
Can accelerators pick the most promising startup ideas no matter their provenance? Using unique data from a global accelerator where judges are randomly assigned to evaluate startups headquartered across the globe, we show that judges are less likely to recommend startups headquartered outside their home region by 4 percentage points. Back-of-the-envelope calculations suggest this discount leads judges to pass over 1 in 20 promising startups. Despite this systematic discount, we find that—in contrast to many past studies—judges can discern startup quality and are no better at evaluating local firms. These differences emerge because the pool of startups accelerator judges evaluate is both broader and less "local," suggesting that judging ability depends on the composition of the companies they are tasked with evaluating.

Managerial Summary
Accelerators often seek the most promising startup ideas. Yet, they can only do so if their judges can discern the quality of startups, both local and foreign to them, without systematic bias. We used unique data from a global accelerator where judges are assigned to evaluate startups headquartered across the globe and find that, while judges can detect the quality of both local and foreign startups, they discount startups foreign to them, hindering their ability to accept the best startup ideas. As venture capitalists increasingly source startups from accelerators, this foreign discounting can result in investors passing over promising ideas. However, simple measures like reducing the threshold for startups evaluated by foreign judges may help reduce judges' foreign discounting and enable picking the best companies.

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

I. Introduction

Startups, like corporations, are increasingly globalized in terms of their markets, investments, and workforce, partially due to the advent of technology that reduces the cost of

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

expanding internationally (Alcácer et al., 2016; Alvarez-Garrido and Guler, 2018; Brynjolfsson et al., 2019; Ghemawat and Altman, 2019; Kerr, 2016; Lu and Beamish, 2001; Oviatt and McDougall, 2005). As a result, entrepreneurial gatekeepers, particularly accelerators[2] which have diffused around the world (Cohen et al., 2019b), 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 (X1 Group, 2018; Y Combinator, 2020). At the same time, gatekeepers have missed out on promising international startup opportunities. The same Y Combinator rejected Canada-based online apparel company, Stylekick, that ended up reaching 80 countries, being translated into 14 languages, and ultimately acquired by Shopify (Business Insider, 2018; Mitra, 2018).

Can accelerators choose the most promising startups from this increasingly global pool? Indeed, accelerators are now soliciting applications from across the globe (Cohen et al., 2019b). However, these organizations may not be able to discern the quality of the startups that apply (Gans et al., 2008; Kerr et al., 2014b; Luo, 2014). Further, they may be particularly inaccurate in discerning the potential of foreign startups 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, like the gender, race, and expertise biases documented across a range of entrepreneurial and innovation settings (e.g., Hegde and Tumlinson, 2014; Lee and Huang, 2018; Li, 2017; Niessen-Ruenzi and Ruenzi, 2019).

[2] An accelerator is defined as a "fixed-term, cohort-based program for startups, including mentorship and/or educational components, that culminates in a graduation event" (Cohen et al., 2019b).

These concerns are especially acute at the earliest stages of the startup selection process when accelerators make decisions with little more than a quick pitch or text description, often because the number of startups screened makes it too costly to conduct in-depth due diligence on each. In these earliest stages, bias and uninformedness are especially problematic because when judges pass on a startup, they also never get a chance to learn more about the firm and correct any initial mistakes. The sheer number of startups in the earliest selection pool of an accelerator makes "spray and pray" approaches infeasible (Ewens et al., 2018). Offering support to the thousands of startups would be incredibly expensive, leading accelerators to necessarily rely on meaningful filtering and selection (Cohen et al., 2019b).

Thus, understanding whether judges are informed about the quality of local and foreign startups at the earliest screening stage of accelerator decision-making is both necessary if we hope to understand why home bias occurs and how accelerators might address it. Prior research in non-accelerator contexts on foreign discounting[3] 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 (Coval and Moskowitz, 1999; 2001; Disdier and Head, 2008; Sorenson and Stuart, 2001). As mentioned above, home bias by accelerators could result from 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 selecting foreign firms. However, such an approach will backfire if the underlying mechanism is instead rooted in the inability of judges to distinguish foreign winners

[3] We use the terms "foreign discounting," "foreign bias," and "home bias" to refer to the fact that judges are more likely to give lower scores to foreign startups after accounting for startup quality. We do not claim that the presence of bias indicates judges are necessarily xenophobic. As we discussed in our Theoretical Framework section, there are numerous potential mechanisms that can lead to lower evaluations of foreign startups that are unrelated to startup quality and potential.

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. This approach might involve assigning judges to only evaluate startups from their home region. Such judges can more quickly and cheaply determine quality without burdensome due diligence. Redesigning how scores are aggregated into decisions may not make a meaningful difference in this scenario. In short, the underlying mechanisms that lead to foreign discounting in startup screening have strong implications for how accelerators should design the first stage of 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 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. This mechanism ultimately determines how organizations should respond. 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 to causally identify if judges exhibit home bias and pinpoint the mechanisms underlying this effect. In the first round of this competition—where judges evaluate text applications—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 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. This staged judging process, where judges first evaluate a brief pitch or application before deciding which startups to interview and conduct further due diligence on, is widely used at accelerators including Y Combinator and Techstars (Cohen et al., 2019b).

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 contexts 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 that 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, experience more user growth, as well as have higher employee, valuation, and revenue growth, contrary to prior work showing that judges can struggle to pick startup winners from losers (e.g., Scott et al., 2020). Further, when we conduct back-of-the-envelope calculations, we find that judges passed over 148 promising foreign startups, equating to roughly 1 in 20 startups in our sample. This evidence suggests that simple changes to how accelerators aggregate judges' evaluations may mitigate the impact of home bias on outcomes.

These findings, at first glance, are at odds with prior work from other contexts showing that experts cannot detect the quality of early-stage firms (e.g., Scott et al., 2020) and that when investors can detect quality differences, it is because they have a local information advantage (e.g., Coval and Moskowitz, 2001). Recent work suggests investors increasingly use a "spray and pray" approach to learn about startup quality after the making small up-front investments instead of heavily screening which firms to invest in (Ewens et al., 2018). Yet, this work has largely focused on investments and decisions on pre-screened and relatively successful firms. When we restrict our sample to conceptually replicate this prior work, we recover the patterns found in this work. When we only include a more selective range of startups, for example, firms with founders who attended an elite university as in Scott et al. (2020), we find that judges are less capable of evaluating which startups are promising and which are not. Similarly, when we use the application text to restrict our sample to more localized firms as in Coval and Moskowitz (2001), we find that, unlike in our full sample of globally oriented technology startups, judges do possess a local information advantage. These patterns suggest that the quality of a judge depends not only on their innate skills and preferences but also on the composition of the pool of startups they are tasked with evaluating.

Our findings make three primary contributions. First, they show that home bias exists in the accelerator setting. Unlike in trade and investment settings (e.g., Coval and Moskowitz, 2001) where home bias has often been studied, we find that judges in the accelerator setting are generally informed but biased against foreign firms when screening early-stage startup ideas. As our conceptual replication of prior work shows, this result does not reflect innate characteristics of the judges, but rather is a combination of judge behavior and the pool of startups being evaluated. Specifically, the pool of startups in the accelerator setting—versus previously studied settings—tend to follow more globally oriented business models that may be easier to evaluate across countries than firms analyzed in prior research on home bias. This result suggests that future work on evaluation should focus both on *who* evaluates and, equally importantly, *what* ends up being evaluated. Indeed, our findings suggest that the widening of the pool of startups (e.g., in terms of educational backgrounds) that judges consider, along with the increasingly standardized business models that these startups adopt, may well imply that accelerators are "better" at screening startups than are investors and mentors studied in prior research (Howell, 2020; Kerr et al., 2014a; Nanda et al., 2020; Scott et al., 2020). Interestingly, as investors increasingly pursue "spray and pray" approaches to learning about startup quality (Ewens et al., 2018), accelerators can serve as complementary sources of early screening to narrow down the "sprayable" pool of startups.

Second, our results suggest that geographic discounting 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 discount foreign startups, and if most of these gatekeepers still reside in entrepreneurial hubs like in the US, this may potentially result in a gap in startups from non-hub regions. Especially because startups excluded at the first stage undergo no further due diligence, the presence of early bias has the potential to distort the sorts of firms that receive support and succeed. These startups otherwise may not be able to get the same type of support from investors who are increasingly pursuing "spray and pray" models that tradeoff providing support to portfolio startups with investing in more startups to learn about their quality (Ewens et al., 2018). And this discounting does not just impact which startups succeed but also may impact who benefits from their innovations (Koning et al., 2020; 2021). Indeed, if accelerators overlook ideas from these non-hub markets, then there may be too few startups serving customers' needs in these non-hub, often non-western, regions.

Third and finally, we highlight a potential limitation of accelerators when it comes to helping foreign startups gain access to key entrepreneurial ecosystems, driven by selection effects. While various studies focus on the treatment effects of accelerator programs, finding positive performance gains for startups (Cohen et al., 2019a; Fehder and Hochberg, 2014; Gonzalez-Uribe and Leatherbee, 2018; Hallen et al., 2020; Howell, 2017; Yin and Luo, 2018; Yu, 2020), our results suggest that the impact of accelerators may be muted for foreign startups because these organizations discount them. This finding shows the value of evaluating selection processes—in addition to treatment effects—in accelerators to fully understand their role in entrepreneurial growth. That said, our results also suggest that relatively minor tweaks to how an accelerator aggregates decisions can address this home bias.

II. Theoretical Framework

Evaluating Startup Quality

Evaluating early-stage startup quality is especially 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 (Aggarwal et al., 2015; Gompers et al., 2020; Hoenig and Henkel, 2015; Kaplan et al., 2009; Sørensen, 2007). Second, there are few precedents to anchor startup evaluations. Great startup ideas are inherently novel, and only a subset of those 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 (Arrow, 1962; Gans et al., 2008; Luo, 2014). Consistent with these priors, research shows that entrepreneurial judges often lack the ability to evaluate the quality of startups, and instead, experiment with small investments into startups to learn of their value (Ewens et al., 2018; Kerr et al., 2014a; 2014b; Nanda et al., 2020; Scott et al., 2020).

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 (Coval and Moskowitz, 2001; Dahl and Sorenson, 2012; Malloy, 2005). 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 foreign startups. For example, an Israeli judge 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 US 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 ones (Coval and

Moskowitz, 2001; Malloy, 2005), and information frictions are higher for foreign acquirers (Conti et al., 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" (Franke et al., 2006; Huberman, 2001; Lin et al., 2013). In the context of demographics, prior research has found substantial evidence of bias against entrepreneurs from different genders and races (Hegde and Tumlinson, 2014; Lee and Huang, 2018; Niessen-Ruenzi and Ruenzi, 2019). Similarly, in the geographic context, studies in financial and trade markets have detected a home bias for local portfolio stocks or trade partners (Coval and Moskowitz, 1999; 2001; Disdier and Head, 2008).

The literature puts forth at least three reasons why home bias might emerge even if judges are no better at evaluating the quality of local startups. First, judges may cognitively prefer what is more familiar or culturally proximate. For example, a startup from a similar geography as a judge may have a subtle way of framing its pitch that draws on local customs that are especially likely to resonate with the judge (Bell et al., 2012; Chadha et al., 2022; Huberman, 2001). Second, judges may simply be xenophobic against particular nationalities or geographic regions, causing them to give lower scores to startups from foreign places (Arikan and Shenkar, 2013). Inversely, judges may prefer that their own regions benefit from entrepreneurial growth and innovation, leading judges to give higher evaluations to local startups (Bell et al., 2012). No matter which mechanism dominates, in each case, judges give lower scores to foreign startups for reasons unrelated to their ability to detect the startup's quality.

Hypothesis Development

These different mechanisms—evaluation uncertainty, contextual expertise, and bias—generate six scenarios that each call for different strategic responses by accelerators and startups. 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 et al., 2016).

However, research ranging from work on contextual intelligence to the benefits of investing in and running firms in one's home region (Coval and Moskowitz, 2001; Dahl and Sorenson, 2012; Malloy, 2005), 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 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 reflects meaningfully different mechanisms. While cell B suggests that organizations would be better off re-allocating attention away from the selection process altogether, 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 can 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 et al., 2019), as are coding practices (Haefliger et al., 2008). Cell F shows that 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., Boudreau et al., 2016; Li, 2017). 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' quality, we have sufficient information to separate the different mechanisms.

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 US, 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. This round is akin to the earliest screening stages of major accelerators that involve the evaluation of inbound text applications. 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 Y Combinator or Techstars. The startups in the program have collectively raised over $6.2 billion, generated over $3 billion in revenue, and created over 157,000 jobs since the accelerator's inception.

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. Unlike the first round that we analyze, later rounds involve interviews between judges and the startup team, pitches, and further due diligence by judges with expertise in the startup's domain. 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—the earliest screening round—where judges—representing executives (60 percent), investors (13 percent), and other professionals (27 percent)—across these international regions initially screen startups from around the world. Judges are well-seasoned. On average, the judges in our sample had already evaluated 56 startups for the accelerator before the evaluation rounds we analyze. Furthermore, 26 of these 56 startups were foreign to the judge. As such, our estimates do not merely reflect how inexperienced evaluators might decide but also capture how more experienced judges screen startups.

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 20. All applications are in English.[4] While the applications do not specifically inform judges of the startup's location, judges may infer it easily through the description of the startup, founder(s), and market. Through a word search analysis of the application text, we find that the home region of the startup is explicitly mentioned in 42 percent of startups' market, traction, and team text. This percent is likely an underestimate because it does not take into consideration implicit mention to the home region, for example, via mention of the past employer and educational institutions of the team. Appendix 8 (Table A8) shows robustness checks that startups are not strategically disclosing their location based on their quality or location. 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 and applicants move onto the

[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.

next round when at least 50 percent of judges recommend the startup should move on.[5] Judges also provide subscores on a scale of 1-10 on the following criteria: startup team, industry & competitors, and business model & financials. The program does not give judges a quota in terms of the number of startups they can recommend. Further, judges must agree to terms that indicate that they "do not expect anything in return," including "future contact" from the startups they evaluate.

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 (US & Canada), and Israel.

These broad regional categories will lead us to underestimate biases within regions. For example, a UK judge evaluating a Latvian startup would appear as a regional match in our data, though we can imagine that the judge would consider the startup foreign and so potentially discount it. Similarly, measurement error due to some judges being assigned to a home program in which they do not work or reside (e.g., a Chicago-based judge is assigned to the Latin American program) should also bias our estimates towards zero. In Appendix 1, we use additional data on judge locations and Monte Carlo simulations to show that our research design, coupled with the large sample of judge-startup evaluations we observe, allow us to detect foreign bias estimates even in the face of substantial measurement error.

[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. While most startups move on to the next round when 50 percent of judges recommend them, there are a small number of exceptions to this rule.
[6] The accelerator does not collect data on judges' location of residence. It only collects the home accelerator program of each judge.

The startups in this global round are of a similar type as those participating in landmark accelerator programs around the world, such as Y Combinator and Techstars. They are largely technology-driven and growth-oriented. Indeed, 39 percent of them are in high tech, 27 percent in general sectors (e.g., retail, consumer products), 17 percent in healthcare/life sciences, 13 percent in social impact, and 4 percent in energy/clean tech. Roughly a fifth of them mention a hub city—such as Silicon Valley, Boston, or London—as identified by the Startup Genome Project (2021), in their market, traction, and team application text (Table A9c). The same share also mentions an elite university in their team application text. About 12 percent of the startups mention an MBA and 9 percent mention a PhD education in their team application text.

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

[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.

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 et al., 2021; Howell, 2017; Yu, 2020). This is a common measure because these investors' decisions reflect both selection and treatment effects that should result in startups with financing having higher startup performance. On the selection side, early-stage investors conduct rigorous due-diligence on portfolio companies prior to investing that may enable them to understand the quality of ventures (Gompers et al., 2020). On the treatment side, investors provide added value (Bernstein et al., 2016) and a stamp of approval (Lerner et al., 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 et al., 2019). Another increasingly 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 (Cao et al., 2021; Hallen et al., 2020; Koning et al., 2022).

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. Beyond these measures, in Appendix 13, we show that the findings hold when we use additional measures of startup quality including valuation, employee counts, and estimated revenue growth 3-4 years after the accelerator program.

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 percent), 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] Judges separately provide this score on the judging form, so while this measure is correlated with the substantive subscores discussed above, it is not perfectly so. This is the main measure used by the accelerator to determine whether startups move to the next round. However, there are exceptions to this cutoff. In these exceptions, the scores on the numerical dimensions (e.g., customer pain/solution and business model/financials) along with other factors can play a part in the startup's acceptance into the program.

Independent Variables

Foreign Startup – Our key covariate captures whether the judge and startup are from the same region (e.g., both from Europe, the US/Canada, Israel, or Latin America). We construct a binary

---

[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.

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.

Logged Page Visits (Pre) – We include logged website page visits three 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

[10] All logged values are of (1+x) because of the frequency of zeros in our dataset.

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 each evaluation. The raw 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 two exceptions. The first is pre-accelerator user traction: local startups have a higher value on average by 6 percentage points (p=0.000). The second is post-accelerator logged financing: local startups have a higher value on average by 5 percentage points (p=0.002). These exceptions occur because US and Canadian startups, which are more likely to be local to judges since the majority of our data are from US 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]

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

V. Empirical Specification

To assess whether judges systematically give lower or higher scores to foreign startups, we fit the following model (Li, 2017; Malloy, 2005):

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

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 also use several fixed effects to account for differences in startup quality across regions and countries. As with our judge fixed effects, we interact all our fixed effects with the program year to account for the fact that startups can apply in multiple years. In our first specification, μ_{it} in Equation 1 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 each 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 UK-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. β indicates that 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 uninformed or informed. However, 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 also are interested in δ and ϕ. A positive 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 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 based on 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 percent) 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 startup evaluations where the judge and startup do not match geographies versus those that do.

Figure A2 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 US-based. Thus, startups that are foreign are more likely to be those that are non-US based, and non-US based startups may be worse quality on average than US-based firms.

We account for these regional quality differences in our regression models. To begin, Column 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 (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. A potential concern with these estimates is that it could be that only US judges are biased against foreign (i.e., non-US) startups. While judge fixed effects will account for differences in harshness among US and other judges, we also show in Appendix 3 that US, EU, and Israeli judges are all more likely to recommend local over foreign startups. This suggests that our findings are not idiosyncratic to US judges.

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 about one-third as much as having some traction suggests that the foreign bias effect is non-trivial. We confirm that the regression results are not driven by differences in the probability of judges giving incomplete subscores to foreign relative to local startups. Table A5 in Appendix 5 shows that judges are equally as likely to give foreign and local startups incomplete subscores.

[Insert Table 3]

Table 4 is similar to Table 3, but it 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 negative, but it 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]

Further, this foreign discounting result is robust to alternative measures of foreignness, different sub-sample restrictions, and regional quality controls. We show in Appendix 4 (Tables A4a-b) that the foreign discounting effect holds when we use raw weighted and non-weighted measures of judges' final recommendation score. In Appendix 6 (Table A6b), we also show that that our foreign discounting effect holds when we measure foreignness using (1) geographic distance between the judge's HQ region and the startup's country of operation, (2) whether the region is explicitly mentioned in the startup's application text, and (3) how "regional" a startup appears based on the text in its application. In Appendix 7, we further demonstrate that the foreign discounting effect holds when we exclude investor judges who might prefer local startups because they represent a more promising investment opportunity than more distant firms (Table A7a). We also show in this section that our results hold when we exclude Latin American startups, which suggests that differences in English ability and training do not account for our findings (Table A7b). In Appendix 9, we show that the foreign discounting effect holds when we directly control for measures of a country's startup quality including GDP per capita, patent applications, venture capital availability, and hub status (Table A9a). We also show our results hold when we directly control for founder quality measures, including whether the team has a PhD, MBA, or elite university affiliation (Table A9b). In Appendix 10 (Table 10a), we confirm that the foreign bias result holds when judge-startup industries match or not. We further show that the results hold when we control for whether the startups are headquartered in a hub (Table A11). Finally, while the focus of our paper is on isolating discounting in the first stage of the accelerator evaluation and screening process, we also show that our findings generalize when estimated on a larger sample of accelerator data in which judges are far from randomly assigned. In Appendix 12 (Table A12), we show that our findings hold across all rounds and years of the program and that foreign bias occurs even in the later rounds of the program when judges interview and evaluate the startup team in person.

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 can 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.[12] 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

[12] Startups may participate in any of the four regions of the accelerator (US, Israel, Latin America, or Europe) and may not be necessarily foreign to this location, even if they were foreign to a judge's location in the initial screening.

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 ($foreign_{ij} x performance_i$) (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-accelerator 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. We also confirm that the result holds if we exclude startups that participated in the accelerator all-together as shown in Column 3. We get similar results when using logged financing six months after the program as our measure of startup quality as shown in Columns 4-7. There is no difference in the relationship between startup quality and judge scores by local startup origin, no matter if we control for or exclude startups who participated in the accelerator.

[Insert Table 5]

As with our foreign bias results, our findings here appear quite robust. Our findings hold no matter the measure of startup quality that we use. In Appendix 13, we show that our findings hold when we use pre-accelerator page traction, page visits, and financing as our quality measures (Table A13a). Our findings also hold if we instead use post-accelerator valuation, employee, revenue growth, and a composite index measure of startup success (Table A13b). The findings also are consistent if we split our sample by foreignness: the r-squared statistics are similar for foreign and local startup samples when we regress judges' scores on startup quality and quality on score, as shown in Appendix 14 (Table A14a). In Appendix 17 (Table A17), we show that judges can detect quality of local and foreign startups with similar precision no matter if their region is explicitly stated or not in the application, suggesting that the startup location provides marginal (if any) informational value to judges.

Reconciling Results with Prior Work

These results suggest that judges can detect the quality of all startups with relatively equal precision, though they discount foreign startups, reflecting cell F in Figure 1. Yet, prior work either suggests that judges cannot detect quality of startups at all and instead experiment with small investments, as shown in cells A and B (Ewens et al., 2018; Kerr et al., 2014b; Nanda et al., 2020; Scott et al., 2020) or have a local information advantage as shown in cells C and D (Coval and Moskowitz, 2001; Malloy, 2005). Why do our results contrast with this prior work?

Crucially, our sample differs in two important respects from this past research. First, by focusing on the earliest screening stage of the evaluation process, judges evaluate a much broader range of startups. In contrast to the global and heterogenous sample of startups analyzed by our accelerator's judges, the sample in Scott et al. (2020)'s study are all startups with founders from MIT. Samples evaluated by venture capital research too tend to comprise pre-selected and high-quality Silicon Valley founders. This suggests that the judges in our sample may well be more informed because they are evaluating startups that vary more in their quality than the already pre-selected firms analyzed in prior work. Second, our sample is dominated by globally oriented technology startups. Indeed, every startup in our sample applied to the global round of an online accelerator, suggesting in their choice that they are likely less "localized" than most firms, and especially less localized than the non-traded goods-producing, small, or remote firms analyzed in Coval and Moskowitz (1999; 2001).

To test our explanation for the first difference, that our pool is much more diverse than prior research, we split our sample into startups with founders affiliated with an elite university (based on the application text), whether the startup is financed at the time of application, and whether the startup mentions being part of a hub city in its application text. These splits let us separate startups that have already been screened (founders affiliated with elite schools, already financed, and startups that have decided to work from a hub) to those that have not. For each sample, in Appendix 15 (Table A15), we show regression results similar to Table 5.[13] Figure 4 shows coefficient plots from these regressions, with the estimates reflecting how much of the judge's score is responsive to differences in startup quality. Consistent with our arguments, we find that judges are worse at picking winners from losers among the pre-screened samples. The coefficients in the pre-screened pools are closer to zero, suggesting scores are less reflective of differences in quality. Thus, our results do not contradict prior works, such as Ewens et al. (2018), Kerr et al. (2014b), Nanda et al. (2020), and Scott et al. (2020). Instead, these results show that accelerators may be better and have an easier time screening good from bad startups because they cast wider nets.

[Insert Figure 4]

[13] These regressions do not include an interaction term between foreign startup and the quality term because we are interested here in isolating the ability of judges to detect quality of startups overall (as opposed to their relative ability to detect quality of local versus foreign firms, which we later evaluate in Appendix 18).

Intriguingly, we also find in Appendix Table A15 that our foreign bias estimate might increase when judges evaluate pre-screened startups, with the foreign discounting coefficient being larger for elite university-affiliated, financed, and hub-affiliated startups than those that are not. This suggests that when judges assess startups that have already met a higher quality threshold, they might rely more on the startup's location. Without easily detectable quality differences, judges may default to picking between startups based on their location.

Further, once the accelerator does narrow down to the approximately 120 startups that it accepts into the final program, judges' ability to detect quality may decline making "spray and pray" or other experimentation techniques employed in venture capital (Ewens et al., 2018) valuable to learn about startup quality. Indeed, in Appendix 16 (Table A16), we show that judges' ability to detect quality of startups declines when evaluating companies accepted into the accelerator program, relative to those in the top-of-the-funnel global round.

To test our second discrepancy, why judges lack a local information advantage in our setting, we again split our sample. This time we restrict our sample to startups that are particularly "localized" following Coval and Moskowitz (1999; 2001)'s approach, as it is for these firms that local information advantage is likely to matter. To measure a startup's localness, we use the application text and exploit the fact that some words are often used by startups from particular regions. For example, terms like "Jerusalem" and "IDF" are particularly used by Israeli startups and not startups from other regions. Appendix 6 provides details. Specifically, for every word in our corpus, we calculate the log-odds ratio that is used in one particular region versus any other region. By aggregating these word-level log-odds ratios, we can calculate a standardized score for how "North American," "Israeli," "Latin American," and "European" each startup application is. To get our final sample of "localized" startups, we restrict our sample to firms where (1) the startup's home region score is greater than X standard deviations and (2) the startup's region score is less than X standard deviations from all other non-home regions. We set X to be 0.5, 0.75, and 1 standard deviations, each reflecting an increasingly localized sample of startups. These two restrictions ensure that the startup is both very localized to its own home region, but also does not happen to read like it is from any other region.

In Appendix 18 (Table A18b), we replicate our Table 5, but only including startups that meet these localization cutoffs. The models include our measures for whether a startup is foreign, our proxy for startup quality, and an interaction term between the two. If judges are worse at evaluating foreign startup quality, the coefficient on quality should be positive and the interaction term negative. Indeed, as Table A18b shows, as we restrict the sample to the most localized startups, we see that judges remain able to detect quality differences, but only for local startups.

To shed further light on this pattern, Figure 5 plots the key coefficient, the interaction term between startup quality and whether the startup is foreign, for "localization" cutoffs ranging from 0.5 to 1 standard deviation. If judges are worse at evaluating foreign startups when the sample of firms only includes very localized firms, then the estimates should gradually become more negative. Indeed, the plot shows exactly this, with the interaction term dropping from 0 to a negative estimate at about 0.75 standard deviation. Consistent with the idea that most startups are globally focused in our sample, just under 5 percent of startups in our sample are "local enough" to meet to the 0.75 cutoff.

[Insert Figure 5]

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

Our finding thus far show that judges give lower scores to foreign startups on average. However, it is possible that this discounting has little impact on which startups move on to the next round. For example, perhaps judges discount high-quality foreign startups who, though rated somewhat lower, still end up the selected for the next round no matter the discount. 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.

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 we removed their foreign bias. Unfortunately, each judge does not tell us nor the accelerator how foreign biased they are. Fortunately, the fact that nearly all judges evaluate multiple foreign *and* multiple local startups lets us estimate a "foreign bias" fixed effect for the vast majority of judges in our sample. This judge-level effect is simply the difference in the average rating a judge gives to foreign versus local startups. Moreover, we can estimate this judge-level bias while simultaneously estimating startup fixed effects. This lets us isolate bias net of any average quality differences between startups that are foreign or local to the judge. These estimates of an individual judge's bias allow us to then "debias" each judge's scoring and so test if different startups would have made it to the second round if selection relied on these debiased scores instead of the judge's actual decisions.

Specifically, we first regress each judge's total score on fixed effects of each startup application, a fixed effect that absorbs each judge's evaluation of local startups, and a fixed effect for each judge's evaluation of foreign startups. In comparison to when we include judge fixed effects in Equation 1, we are not merely accounting for each judge's overall "harshness," but instead accounting for each judge's individual harshness towards foreign and local firms. We then use these fixed effects, instead of an estimated "foreign discounting" coefficient, to unpack and address judge bias. Consistent with our primary findings that judges discount startups by 0.06 standard deviations, we find the average judge fixed effects for the foreign startups they evaluate is 0.07 standard deviations lower than for the local startups they evaluate. We also find that some judges appear especially biased, with the 25th percentile judge discounting foreign startups -0.35 standard deviations more than local startups and the 5th percentile -0.8 standard deviations.

We then use these individual fixed effects to "debias" each judge's score. For example, imagine a judge is relatively harsh, giving foreign startups they evaluate scores that are -0.5 lower than the average judge and local startups -0.3 lower, even after accounting for startup fixed effects. In this case, we would estimate that this judge has an individual foreign bias of -0.2 (-0.5 minus -0.3). To "debias" this judge we would add 0.2 to the score for each foreign startup they evaluated. More generally, we repeat this procedure for each judge to account for the distribution of biases in our data. As mentioned above, this offset is net of startup quality since we include fixed effects for each startup when estimating the judge bias fixed effects. To convert these scores into the recommendations the accelerator uses to select startups, we have our "debiased" judges select the same number of recommended startups as we observe in the actual data, but we select those with the highest scores according to our debiased estimates. We use this assumption of the same number of recommendations because judges can recommend as few or as many startups as they would like. There is no numerical score cutoff that leads to a recommendation. Finally, we follow the accelerator's rules and mark a startup as moving onto the next round if 50 percent or more of the judges recommend the startup.

Using our debiased scores, we find that removing home bias would lead to 148 startups moving from non-recommended to recommended and 86 moving in the other direction. Together home bias appears to lead to mistaken decisions—if the goal is to only select the highest quality startups—for 234 startups, just over 6 percent of applicants. Moreover, as we show in Appendix Figure A19, these "missed" startups have promise. The estimated quality of these "missed" startups is similar to the majority of actually selected startups. While the highest-quality startups make it to the next round regardless, home bias appears to cause the accelerator to miss out on a non-trivial number of promising ventures.

VII. Conclusion and Implications

We find that judges can equally discern the quality of local and foreign startups in the earliest stage of the evaluation process. 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 having some venture financing. Back-of-the-envelope estimates suggest that this bias potentially excludes about 1 in 20 promising entrepreneurial ideas. These results reveal that judges are informed about the quality of both local and foreign startups, but they still discount foreign firms.

However, we also find that observed judge behavior depends on the pool of startups that judges are tasked with evaluating. Judges are worse at evaluating quality when the startups have already been screened and when foreign startups have more localized business models. However, as accelerators increasingly consider a wider pool of global ventures, and startups continue to adopt standardized technology-driven business models, our findings suggest accelerators and similar business plan competitions may increasingly play an important, if biased, role in screening early-stage startups (Chatterji et al., 2019; Haefliger et al., 2008; Howell, 2020). Our findings contrast with past work showing that venture capital firms struggle to screen promising ventures from bad ideas (Kerr et al., 2014a; Nanda et al., 2020; Scott et al., 2020). That said, our findings that observed judge effectiveness and bias depend on the pool of startups evaluated reconcile this difference. Venture capitalists likely struggle to predict success because they evaluate a pre-screened pool of startups, screening that is increasingly first done by accelerators that can separate winners from losers at the very earliest stages of the entrepreneurial process.

Our results also highlight how accelerators increasingly complement venture capitalists' "spray and pray" approaches (Ewens et al., 2018). Specifically, our findings reveal that accelerators effectively screen the thousands upon thousands of startup applications they receive. When they can no longer pick winners from losers, they can then refer the startups they accelerated to venture capitalists who can use "spray and pray" approaches to learn about startup quality through sequential experimentation (Ewens et al., 2018; Hallen et al., 2023; Howell, 2020). However, this linkage also suggests that foreign bias by accelerator judges may lead investors to pass over promising foreign startups since those foreign startups never even make it into the accelerator to begin with. Even if investors are not discounting foreign firms (c.f., Lin and Viswanathan, 2016), bias by judges, mentors, and other gatekeepers earlier in the entrepreneurial process may well explain why startups in many parts of the world fail to scale (Wright, 2023).

This logic also suggests that the foreign bias we identify here may impact the direction of innovation. If accelerators pass over startups from remote regions, which are more likely to be foreign to accelerators, they reduce the probability that innovations addressing the needs of those markets will survive and grow. Even if these foreign startups employ globally standardized business models and practices, their innovations and target customers may still disproportionately benefit the home market. This distortion is similar to effects seen in studies of bias in gender and race contexts (e.g., Koning et al., 2020; 2021).

Turning to practice, our results also suggest that accelerators may benefit from opening their initial screening processes to startups more globally, given their ability to discern startup quality at the top-of-the-funnel no matter the startup's location. Accelerators have the potential to identify firms that might not have received any support otherwise. That said, later rounds of evaluation, where there is likely an opportunity to use local references and networks, may still require localized capabilities to best pick which global startups are most promising. Crucially, however, any global approach depends on accelerators revising their processes to reduce the impact of bias—in this case foreign bias—that all too often enters the evaluation of diverse and heterogenous samples (Brooks et al., 2014; Cao et al., 2021).

For entrepreneurs, our findings suggest caution when acting on feedback from accelerators. While past studies show that accelerators, by providing signals on startup's quality to the entrepreneur, are an important source of learning (Cohen et al., 2019a; Howell, 2020; Lyons and Zhang, 2018; Yu, 2020), such signals may lead entrepreneurs astray when they originate from a non-representative sample or from biased actors (Cao et al., 2021). In our case, judges' foreign discounting implies that the signals from accelerators may be distorted for firms from regions under-represented amongst accelerator judges. Given the increasingly recognized importance of entrepreneurial learning in startup performance (Koning et al., 2022), the fact that there is less bias in local signals also provides a novel mechanism to explain why ventures tend to perform better when located in a founder's native region (Dahl and Sorenson, 2012).

Overall, we find that startups face a “liability of foreignness” (Zaheer, 1995). Notably, we do not find that judges face a disadvantage in evaluating foreign startups. Instead, we find that judges can discern the quality of startups across regions in the early screening stage. This may be because technology and business models have standardized into a “playbook” that is comparable across countries, for example, with the proliferation of codified management (Bloom and Reenen, 2007; Chatterji et al., 2019) and technology practices (Haefliger et al., 2008). Further, the existence of such a playbook may reduce the need for private information (Coval and Moskowitz, 2001; Malloy, 2005) or contextual intelligence (Khanna, 2014) to evaluate foreign opportunities. Future work should continue to explore how the changing nature of startups and their strategies impact gatekeepers’ ability to screen promising ventures from bad ideas.

References

Aggarwal, R., Kryscynski, D., & Singh, H. (2015). Evaluating venture technical competence in venture capitalist investment decisions. *Management Science*, *61*(11), 2685-2706.

Alcácer, J., Cantwell, J., & Piscitello, L. (2016). Internationalization in the information age: A new era for places, firms, and international business networks? *Journal of International Business Studies*, 47(5), 499-512.

Alvarez-Garrido, E., & Guler, I. (2018). Status in a strange land? Context-dependent value of status in cross-border venture capital. *Strategic Management Journal*, *39*(7), 1887-1911.

Arikan, I., & Shenkar, O. (2013). National animosity and cross-border alliances. *Academy of Management Journal*, *56*(6), 1516-1544.

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.*

Bell, R. G., Filatotchev, I., & Rasheed, A. A. (2012). The liability of foreignness in capital markets: Sources and remedies. *Journal of International Business Studies*, *43*(2), 107-122.

Bernstein, S., Giroud, X., & Townsend, R. R. (2016). The impact of venture capital monitoring. *The Journal of Finance*, *71*(4), 1591-1622.

Bloom, N., & Van Reenen, J. (2007). Measuring and explaining management practices across firms and countries. *The Quarterly Journal of Economics*, *122*(4), 1351-1408.

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

Brooks, A. W., Huang, L., Kearney, S.W., & Murray, F. E. (2014). Investors prefer entrepreneurial ventures pitched by attractive men. *Proceedings of the National Academy of Sciences*, *111*(12), 4427-4431.

Brynjolfsson, E., Hui, X., & Liu, M. (2019). Does machine translation affect international trade? Evidence from a large digital platform. *Management Science*, *65*(12), 5449-5460.

Business Insider. (2018). These founders got rejected from Y Combinator four times. Here's how they finally got in. https://www.businessinsider.in/These-founders-got-rejected-from-Y-Combinator-four-times-Heres-how-they-finally-got-in-/articleshow/66644585.cms.

Cao, R., Koning, R. M., & Nanda, R. (2021). Biased sampling of early users and the direction of startup innovation. National Bureau of Economic Research Working Paper, No. w28882.

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

Chadha, S., Kleinbaum, A. M., & Wood, A. (2022). Social Networks are Shaped by Culturally Contingent Assessments of Social Competence. Working Paper.

Chatterji, A., Delecourt, S., Hasan, S., & Koning, R. (2019). When does advice impact startup performance? *Strategic Management Journal*, *40*(3), 331-356.

Cohen, S. L., Bingham, C. B., & Hallen, B. L. (2019a). 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. (2019b). The design of startup accelerators. *Research Policy*, 48(7), 1781-1797.

Conti, A., Guzman, J., & Rabi, R. (2020). Herding 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.

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.

Ewens, M., Nanda, R., & Rhodes-Kropf, M. (2018). Cost of experimentation and the evolution of venture capital. *Journal of Financial Economics*, *128*(3), 422-442.

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.

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.

Hallen, B. L., Cohen, S. L., & Park, S. H. (2023). Are Seed Accelerators Status Springboards for Startups? Or Sand Traps? *Strategic Management Journal.*

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.

Howell, S. T. (2020). Reducing information frictions in venture capital: The role of new venture competitions. *Journal of Financial Economics*, 136(3), 676-694.

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. (2014a). 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. (2014b). 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. (2022). Experimentation and Start-up Performance: Evidence from A/B Testing. *Management Science*.

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).

Koning, R., Samila, S., & Ferguson, J. P. (2021). Who do we invent for? Patents by women focus more on women's health, but few women get to invent. *Science*, 372(6548), 1345-1348.

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.

Lyons, E., & Zhang, L. (2018). Who does (not) benefit from entrepreneurship programs?. *Strategic Management Journal*, 39(1), 85-112.

Malloy, C. J. (2005). The geography of equity analysis. *The Journal of Finance*, *60*(2), 719-755.

Mitra, S. (2018). Building a Fast Growth, Cutting-Edge Insurance Brokerage: Karn Saroya, CEO of Cover. *One Million by One Million Blog.* https://www.sramanamitra.com/2018/10/22/building-a-fast-growth-cutting-edge-insurance-brokerage-karn-saroya-ceo-of-cover-part-1/.

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.
Startup Genome Project. (2021). Rankings 2021: Top 30 + Runners-up. Retrieved from https://startupgenome.com/article/rankings-2021-top-30-plus-runners-up.
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.
Wright, N. L. (2023). Where Strategy Matters: Evidence from a Global Startup Field Study. *Available at SSRN 4335210.*
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	7232	0.01	1.01	-3.31	2.36	9107	-0.12	1.05	-3.31	2.36	0.13***
Overall Raw Score	7706	2.92	1.16	0.00	5.00	9902	2.75	1.13	0.00	5.00	0.16***
Recommend	7706	0.61	0.49	0.00	1.00	9902	0.56	0.50	0.00	1.00	0.05***
Subscore: Customer Needs and Acquisition	7692	6.25	1.85	1.00	10.00	9833	6.06	1.92	1.00	10.00	0.18***
Subscore: Customer Pain and Solution	7694	6.82	1.84	1.00	10.00	9840	6.63	1.95	1.00	10.00	0.18***
Subscore: Financial Business Model	7675	5.72	1.98	1.00	10.00	9787	5.53	2.07	1.00	10.00	0.19***
Subscore: Industry and Competitor	7690	6.11	1.85	1.00	10.00	9827	5.93	1.94	1.00	10.00	0.17***
Subscore: Overall Impact	7686	6.21	1.93	1.00	10.00	9820	6.03	2.00	1.00	10.00	0.18***
Subscore: Regulation and IP	7261	5.91	2.15	1.00	10.00	9175	5.65	2.25	1.00	10.00	0.27***
Subscore: Team and Advisors Investors	7678	6.51	2.01	1.00	10.00	9805	6.31	2.09	1.00	10.00	0.20***
Startup Quality Measures											
Log Pre-Accelerator Total Page Visits	3917	1.37	2.77	0.00	12.50	5816	1.46	2.88	0.00	12.50	-0.09
Log Pre-Accelerator Financing	7706	0.45	1.41	0.00	6.03	9902	0.41	1.33	0.00	6.03	0.04
Log Post-Accelerator Total Page Visits	7706	2.87	3.52	0.00	12.82	9902	2.93	3.61	0.00	12.82	-0.06
Log Post-Accelerator Financing	7706	0.30	1.10	0.00	5.95	9902	0.25	0.98	0.00	5.92	0.05**
Has User Traction	7706	0.59	0.49	0.00	1.00	9902	0.53	0.50	0.00	1.00	0.06***
Has Financing	7706	0.12	0.33	0.00	1.00	9902	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	US & Canada	Israel	Total
Europe	229	791	206	1,226
	239.3	*783.7*	*203*	
US & 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	US & 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*
US & 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 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 evaluating success for local and foreign startups alike, and (3) still discount foreign startups

	(1)	(2)	(3)	(4)	(5)	(6)
			Judge's Total Score			
Foreign Startup	-0.065**	-0.056*	-0.039	-0.053**	-0.047*	-0.040*
	(0.025)	(0.024)	(0.025)	(0.019)	(0.020)	(0.020)
Log Post-Accelerator Page Visits	0.050***	0.036***	0.043***			
	(0.004)	(0.004)	(0.004)			
Foreign Startup * Log Post-Accelerator Page Visits	0.000	0.003	-0.001			
	(0.005)	(0.005)	(0.005)			
Log Post-Accelerator Financing				0.170***	0.027*	0.178***
				(0.008)	(0.012)	(0.040)
Foreign Startup *Log Post-Accelerator Financing				-0.010	0.009	-0.026
				(0.011)	(0.015)	(0.055)
Accelerator Participation		0.682***			0.701***	
		(0.032)			(0.042)	
Foreign Startup * Accelerator Participation		-0.109**			-0.109*	
		(0.041)			(0.055)	
Observations	16,320	16,320	14,475	16,320	16,320	14,475
Judge x Year	Yes	Yes	Yes	Yes	Yes	Yes
Startup Country x Year	Yes	Yes	Yes	Yes	Yes	Yes
Startup x Year	No	No	No	No	No	No
Accelerator Participation	Yes	Yes	No	Yes	Yes	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

Figures

Figure 1: Predicted relationships between judge scores and startup quality

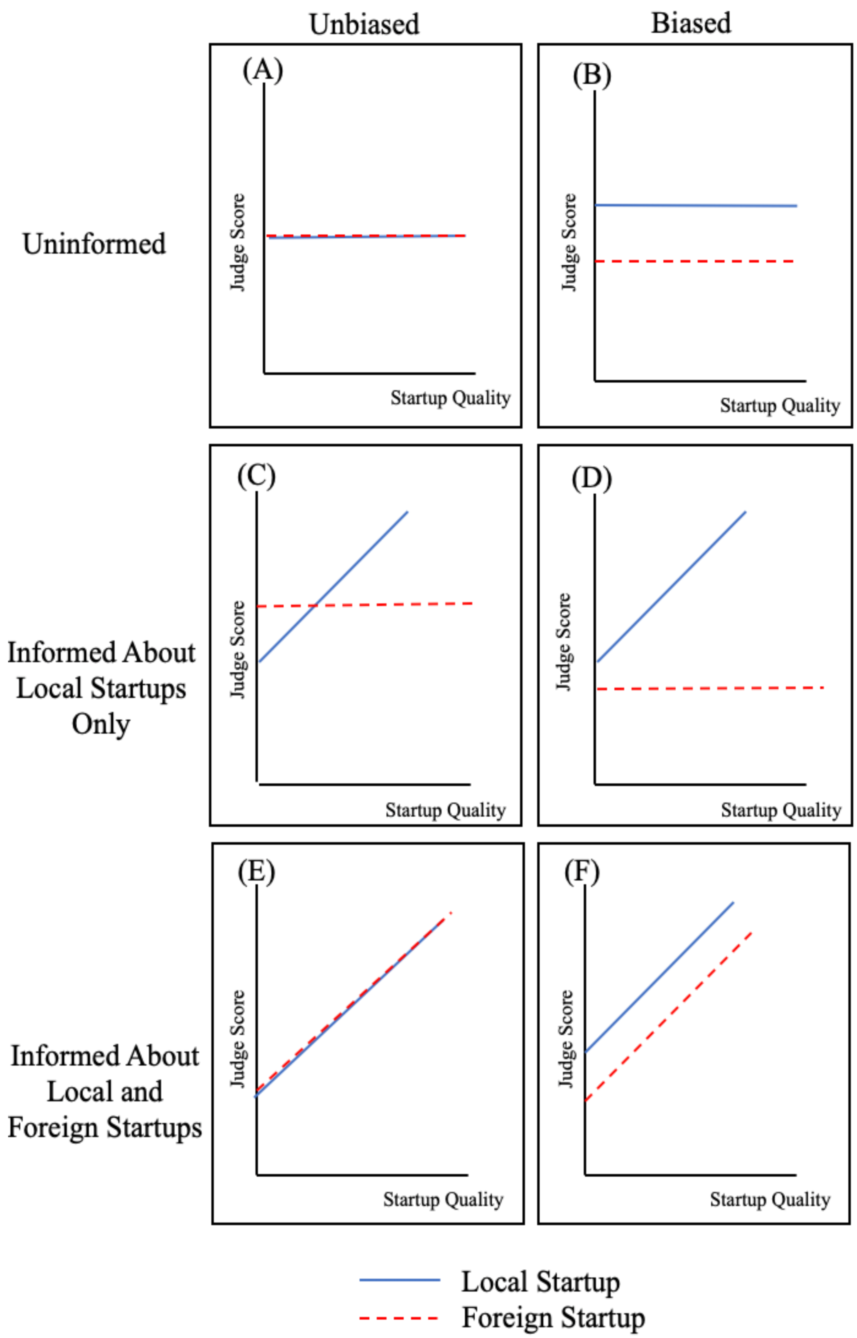
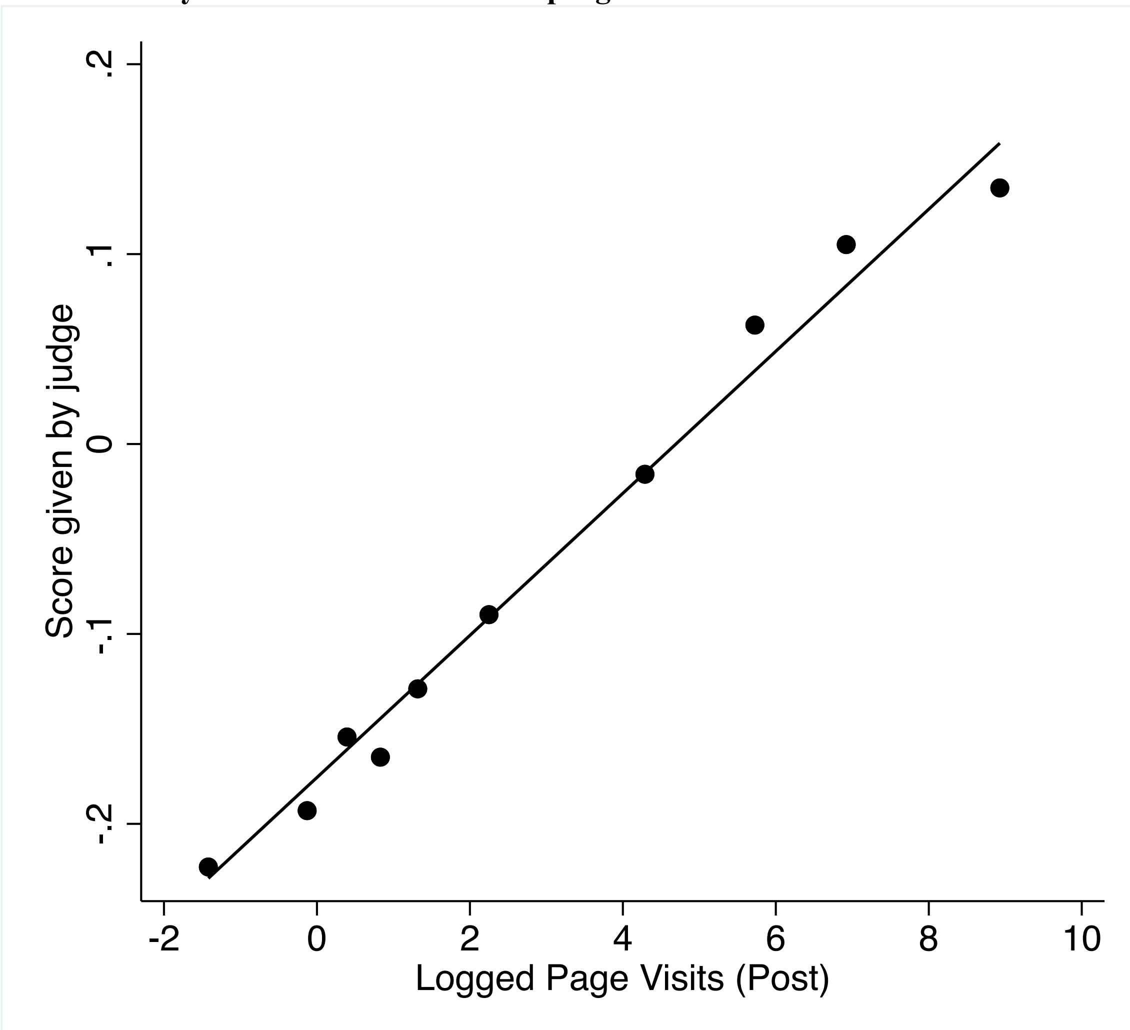
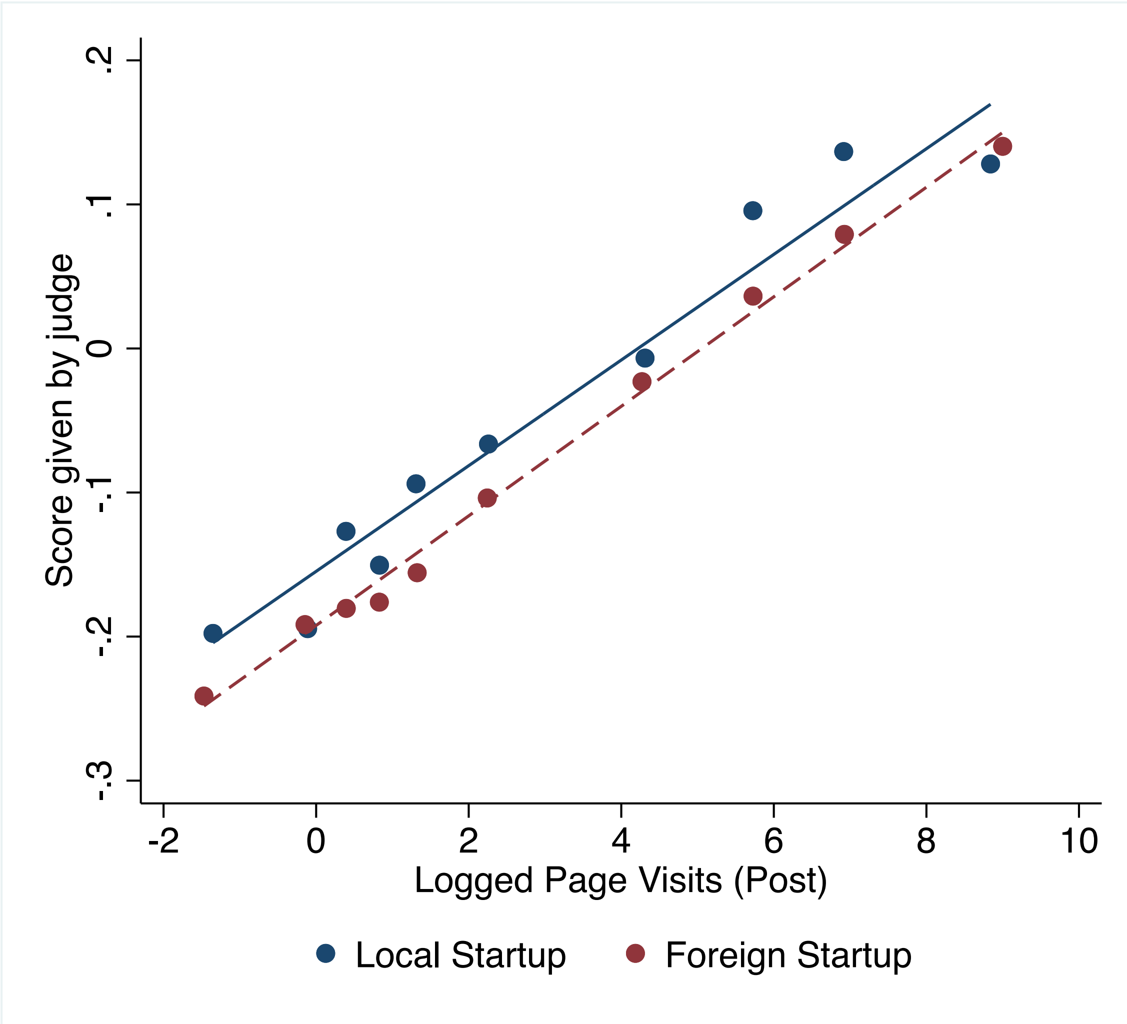


Figure 2: Binscatter showing that judges give higher scores to startups with more growth one- to- two- years after the accelerator program



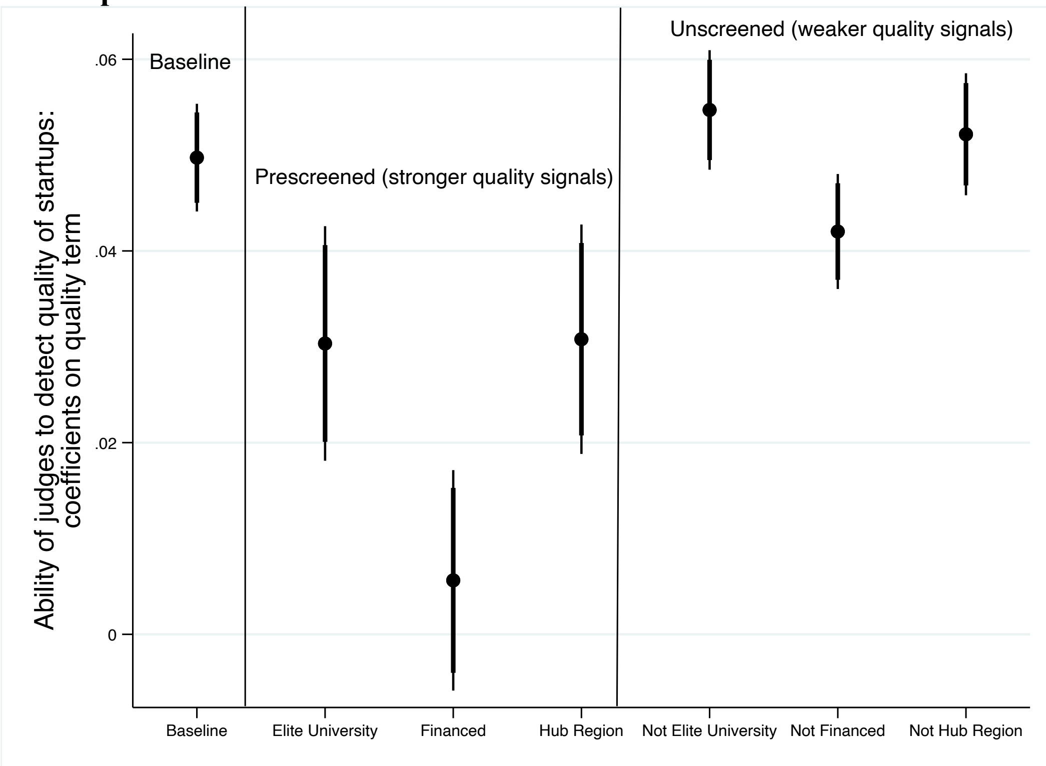
Binscatter after accounting for judge x year and startup country x year fixed effects and accelerator participation. We use 10 bins.

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



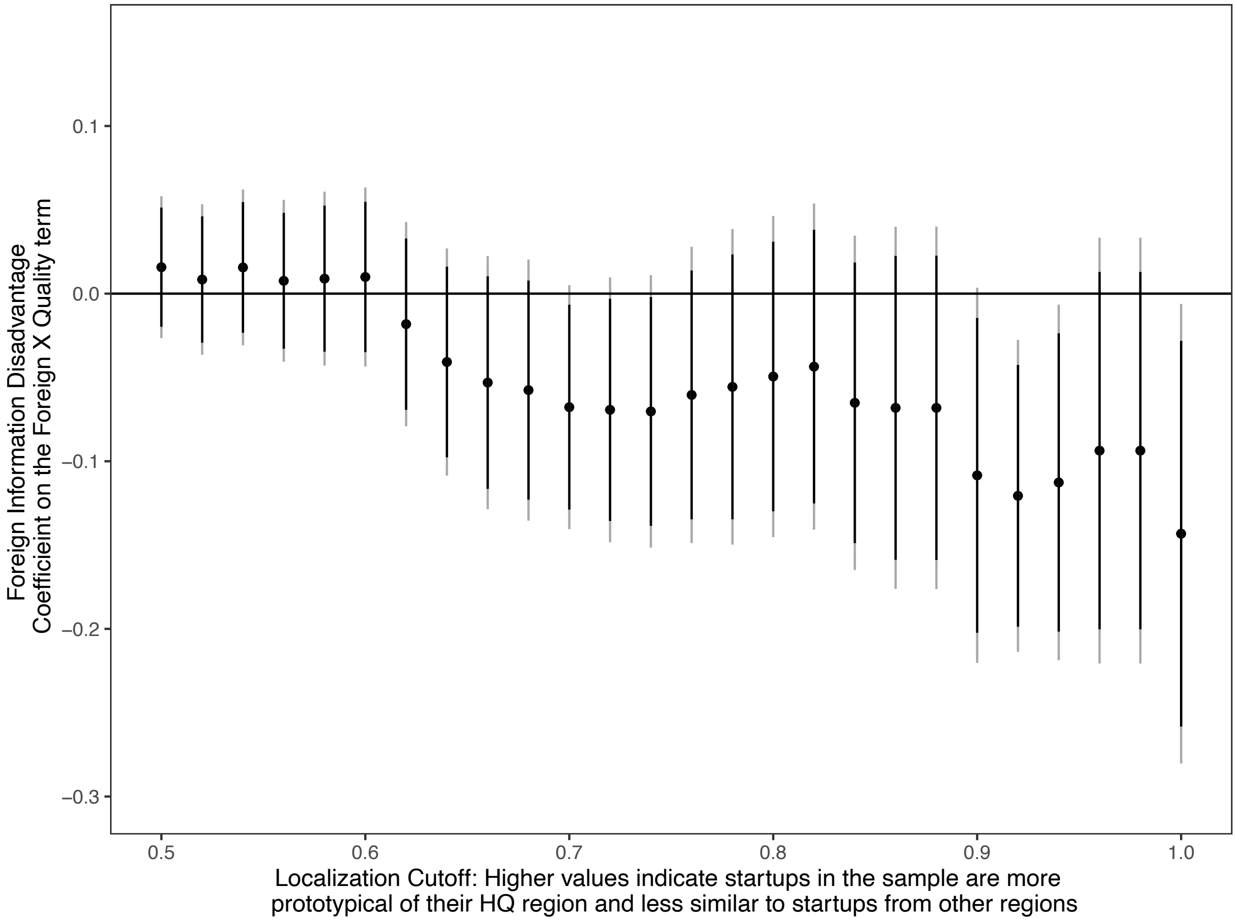
Binscatter after accounting for judge x year and startup country x year fixed effects and accelerator participation. We use 10 bins.

Figure 4. Coefficient plot showing that judges are less sensitive to the quality of startups with elite university affiliation, financing, and in a hub region. The bars show 90 percent and 95 percent confidence intervals.



The plot shows coefficients from regressing judges' score on post-accelerator log page visits, controlling for participation in the program across different sub-samples of startups. It shows 90 and 95 percent confidence intervals.

Figure 5. Coefficient plot showing that judges have a local information advantage only for localized startups. The bars show 90 percent and 95 percent confidence intervals.



APPENDIX

Judging Foreign Startups

Contents:

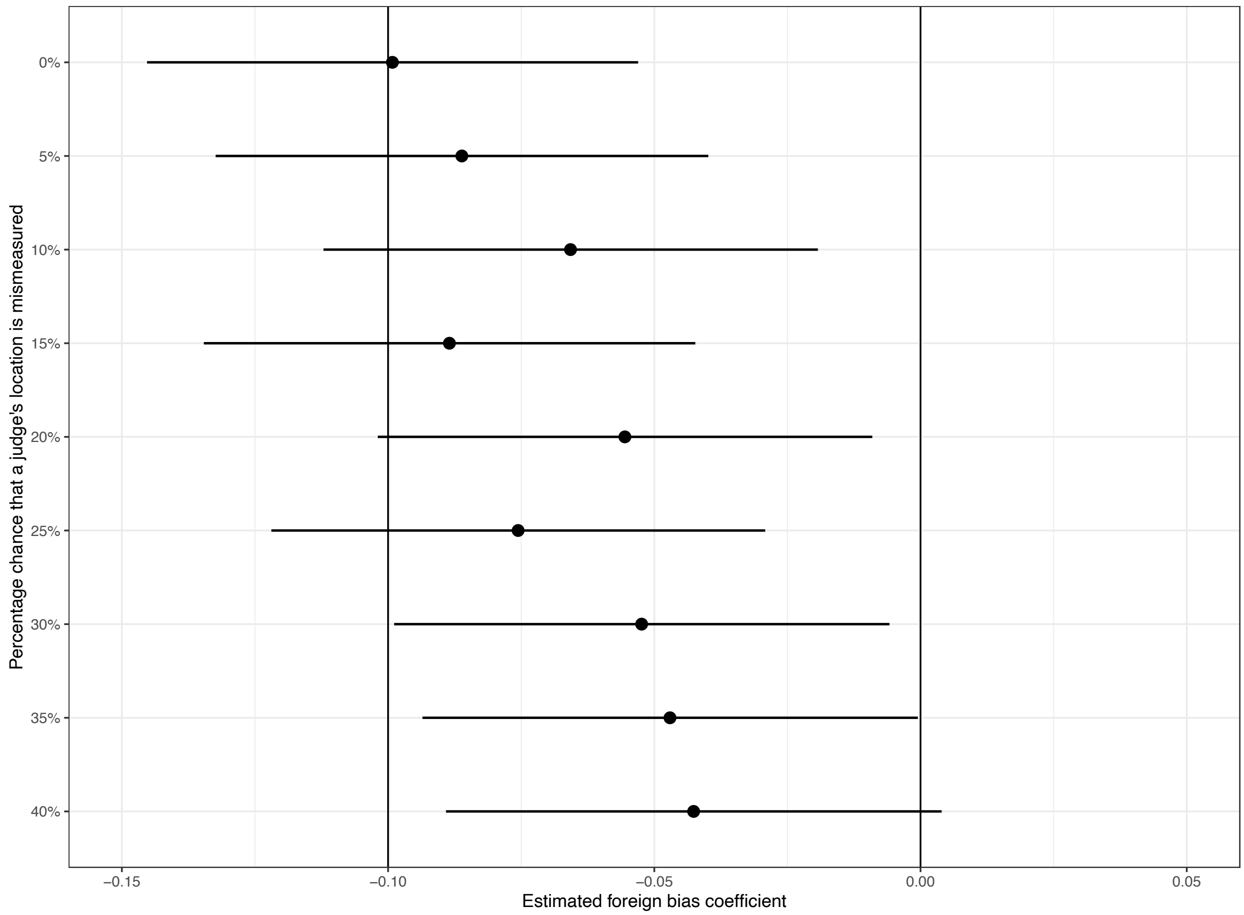
A1. Monte Carlo simulations showing that 0-40 percent judge location measurement error still allows us to detect foreign discounting and local information advantage effects

Our data includes information on judges' home regions in the program. There may be concern that some judges may not actually be from the home regions, which they are associated with in the accelerator, resulting in measurement error that confounds our results. To address this concern, we conduct Monte Carlo simulations to assess how measurement error in judge location impacts our coefficient estimates of both the amount of foreign discounting and local information advantage. In our simulations, we assume that there are 1,000 judges (roughly what is present in our actual analysis). Each judge evaluates 15 startups (again similar to our actual sample). This gives us 15,000 judge-startup pairs, comparable to the 16,320 observations we analyze in Table 3. For simplicity, we assume there are two types of judges: US judges and EU judges.

We then evaluate the impact of measurement error on two different models of how judges evaluate startups. In the first model, we assume that startups vary in their quality, judges vary in their harshness, and judges simply discount startups from outside their own region by -0.1 standard deviations, no matter the startup's quality. This effect is slightly larger than our estimate of -0.06, but if there is measurement error in our data, then this figure—as the simulations below show—is likely an underestimate of the true effect. Given this model, we then randomly shuffle judge's locations to vary the mismeasurement rate from 0 percent (all judge locations are perfectly measured) to 40 percent (2 in 5 judges are recorded as being from the wrong region). We then run regressions exactly like Equation 1 in the paper where we regress the score on whether the judge is measured to be foreign or not while including judge and region fixed effects. With this simulation effort, our goal is to test if our coefficient estimates match the true foreign discount rate of -0.1 standard deviations.

Figure 1.1, displayed below, shows the estimated foreign bias coefficient estimates given data with different measurement error rates. As the error rate goes up, our estimate is increasingly biased towards zero, consistent with intuition and arguments concerning classical measurement error. With an error rate of 30 percent, the estimated effect size is -0.05; at 40 percent we find a positive but noisy effect.

Figure A1a. Classical measurement error: Foreign bias coefficients



Beyond the impact on the main foreign bias coefficient, we also evaluate how measurement error impacts a model where judges are assumed not just to be biased, but also more informed about the quality of local as against foreign startups. In this model, which conceptually mirrors Equation 2 in the paper, we assume that all judges, on average, rate a startup one standard deviation higher when the startup's quality goes up by one standard deviation. That said, we also assume judges from a different region are worse at evaluating the quality of these foreign startups (e.g., US judges are worse at evaluating the quality of EU startups and vice versa). Again, following equation 2, we assume that the interaction between "foreign" and "quality" has a value of -0.1, indicating that foreign judges in our simulated data are a tenth of a standard deviation less responsive to differences in quality for foreign as against local firms. We refer to this interaction term as the "ignorance" coefficient. This value reflects how much less sensitive judges are to foreign startup quality relative to local startup quality. Figures A1b-c below show the estimated "bias" and "ignorance" coefficients as we increase the amount of measurement error. Again, we see estimates on both coefficients shrink towards zero.

Figure A1b. Classical measurement error incorporating information asymmetries: Foreign bias coefficients

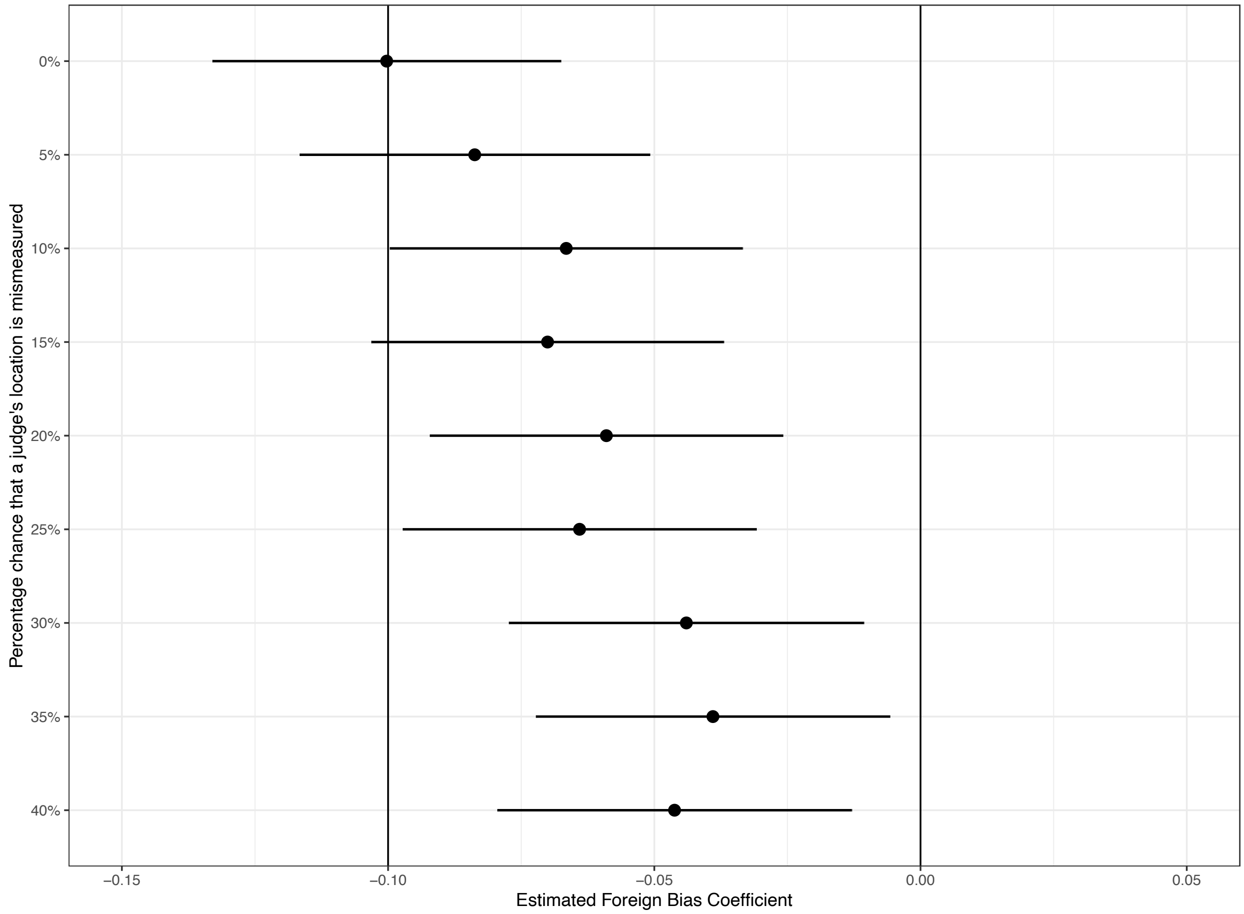
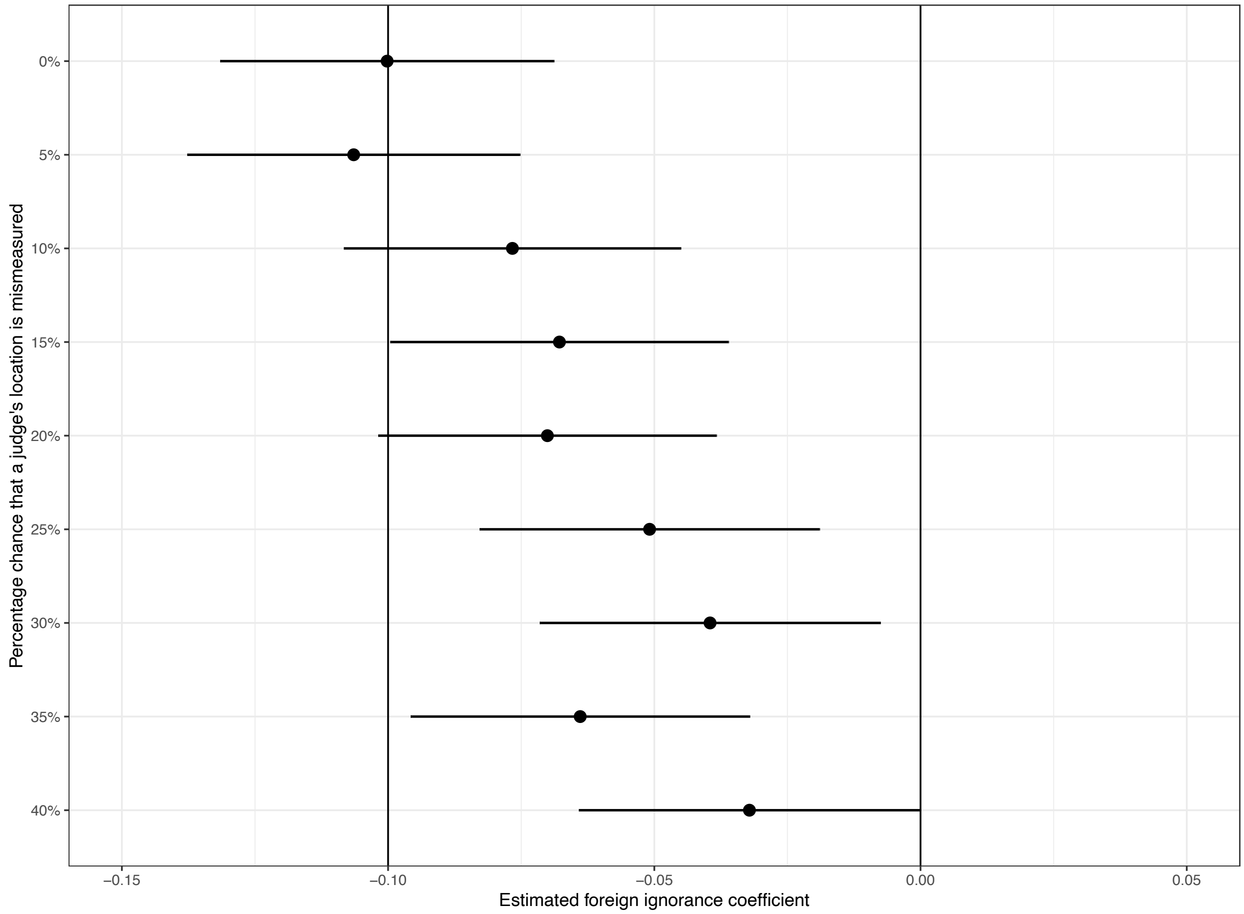
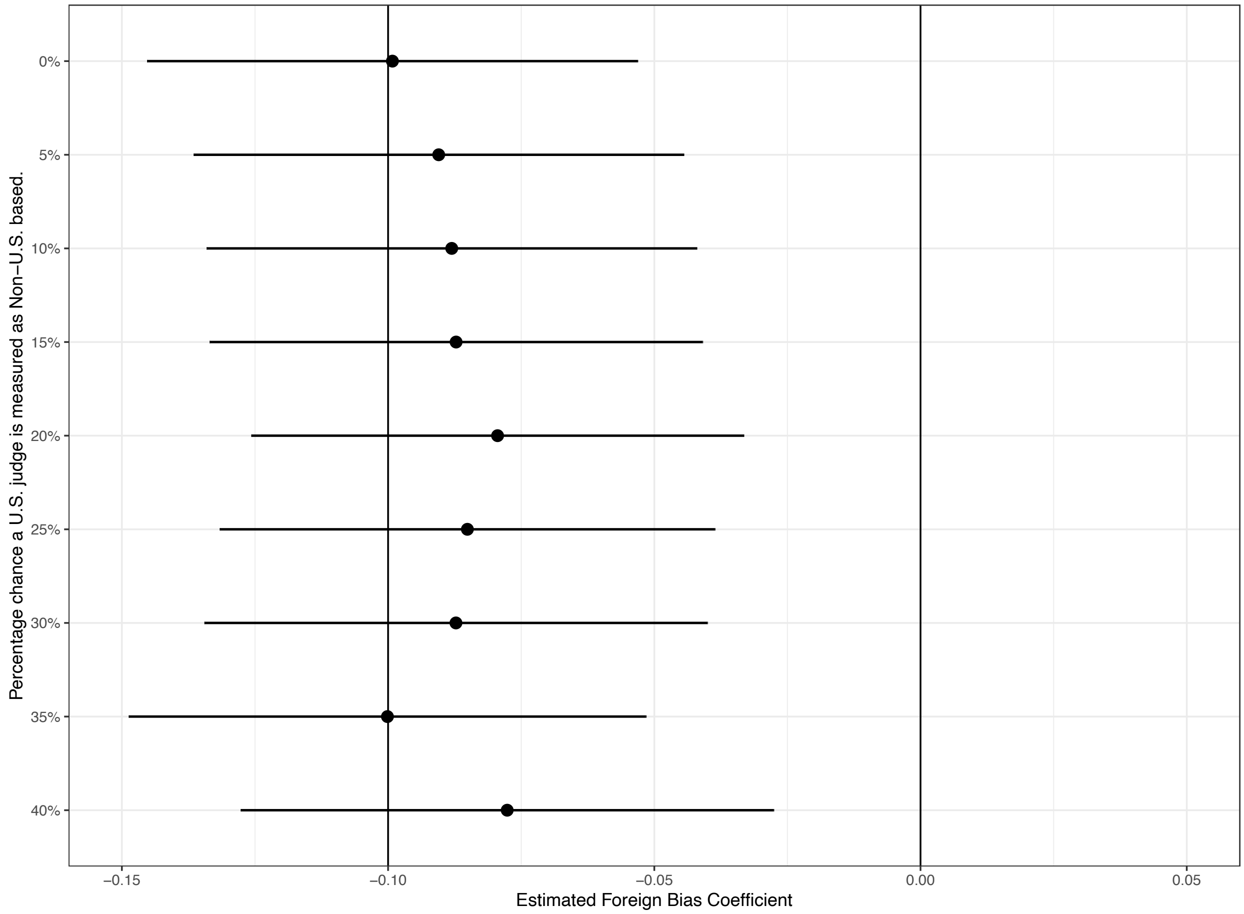


Figure A1c. Classical measurement error incorporating information asymmetries: Foreign information coefficients



Overall, these findings suggest that uniform measurement error should not cause us to find a result when there isn't one. However, we might worry that the rate of measurement error varies across geographies. Perhaps many of the judges from non-US home programs are actually from the US, but judges from the US home program are nearly always from the US. To evaluate how differential rates of measurement error impact our results, we also present results showing what happens if we increase measurement error in only one region. Specifically, we take US judges and assign a given percentage of them to be labeled as "EU" judges. For EU judges, we introduce no measurement error. Figure A1d shows the estimated coefficients for the "foreign bias only" model as we increase the mislabeling rate for US-based judges from 0 percent to 40 percent. While the coefficients shrink, the decline is less pronounced than before because only half of our sample is mismeasured. The figure suggests our estimates are likely a lower bound for the foreign bias effect.

Figure A1d. Classical measurement error: Foreign bias coefficients with asymmetrical error



Turning to the impact on a model with both foreign bias and foreign ignorance we find similar patterns. Figures A1e-f again show that as the error rate increases, our estimates tend towards zero. It is not the case that when mismeasurement only impacts one region that we overstate the size of the ignorance or bias effects. Instead, effect sizes tend towards zero as is the case with symmetric measurement error. We do see in these graphs that the bias in the ignorance coefficient towards zero grows faster than the bias in the foreign bias coefficient. However, neither coefficient ends up overestimated.

Figure A1e. Classical measurement error incorporating information asymmetries: Foreign bias coefficients with asymmetrical error

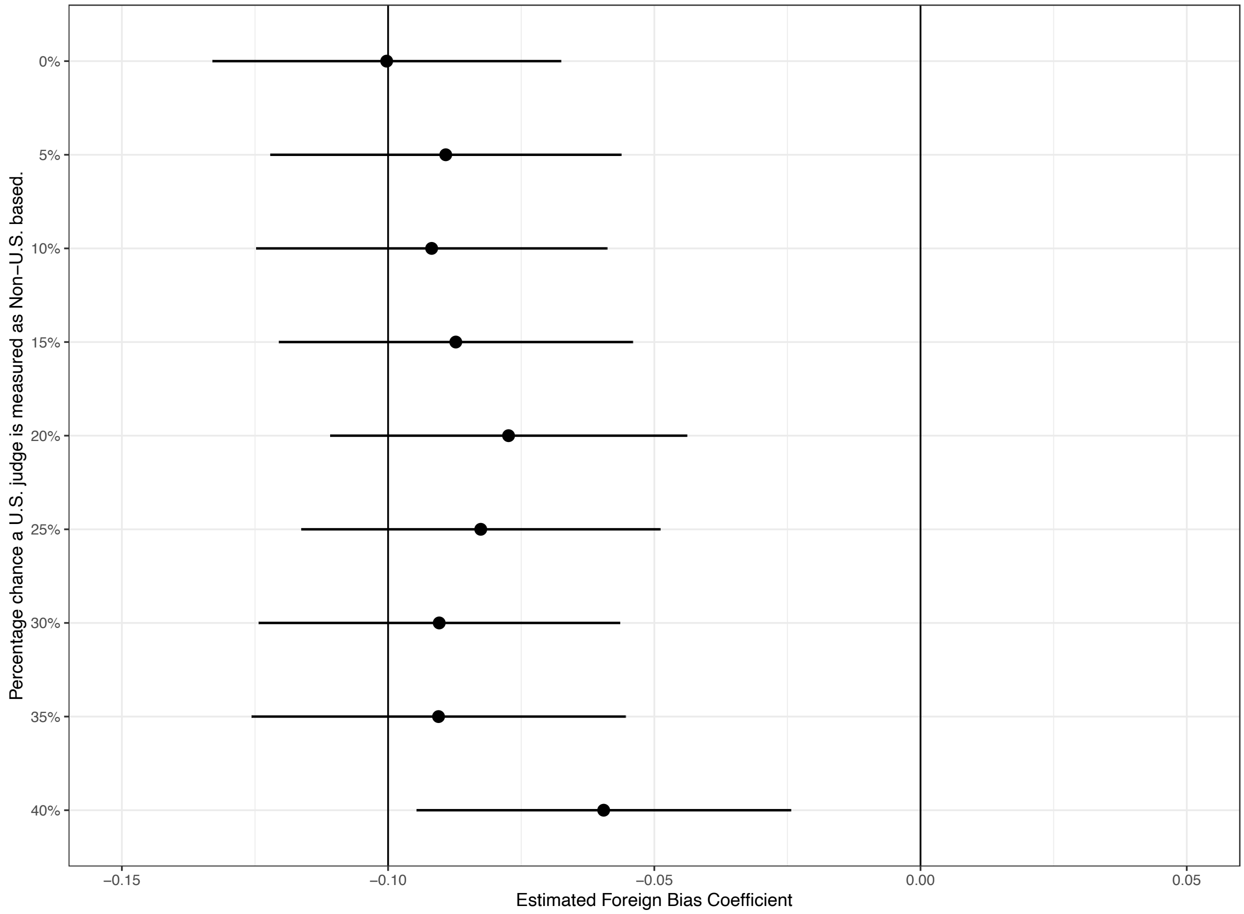
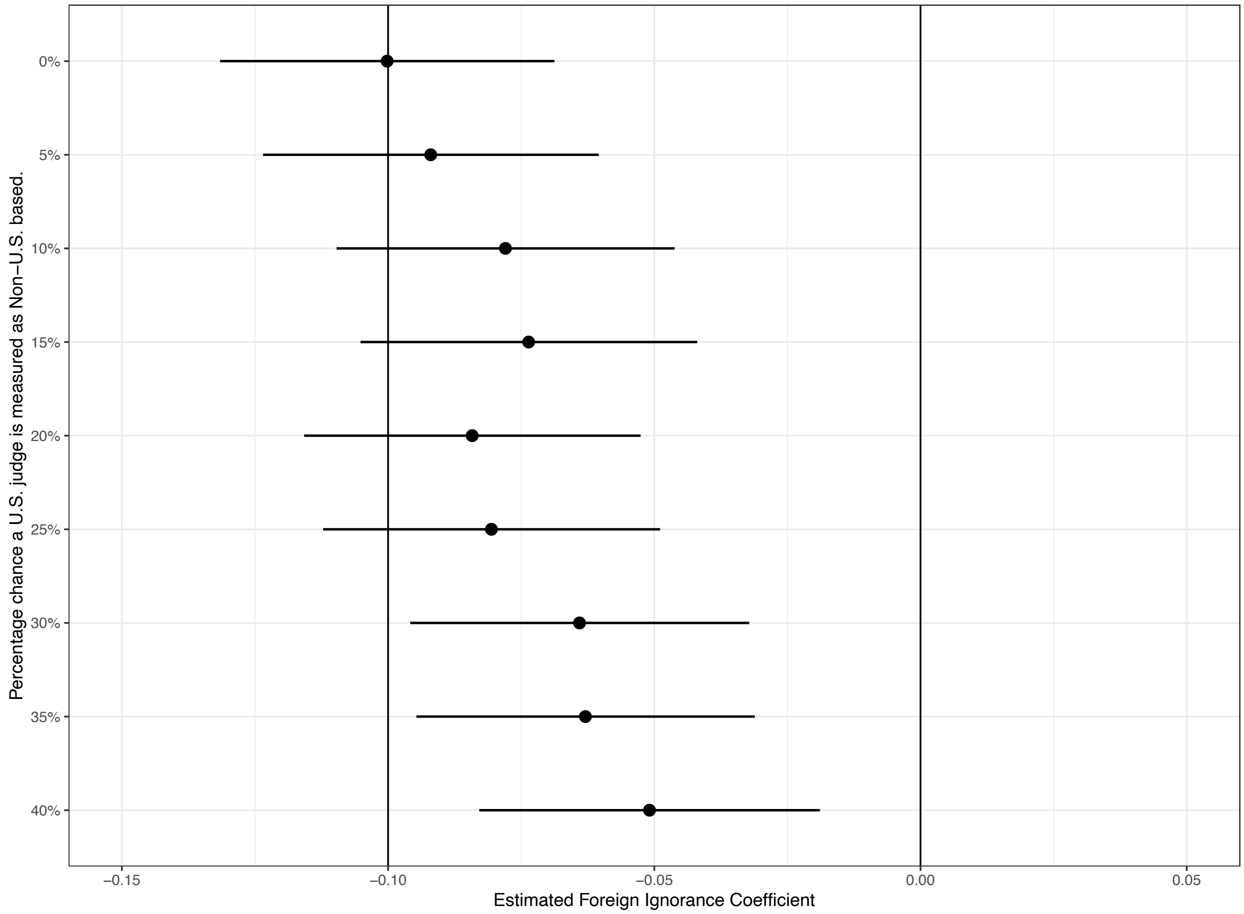


Figure A1f. Classical measurement error incorporating information asymmetries: Information coefficients with asymmetrical error

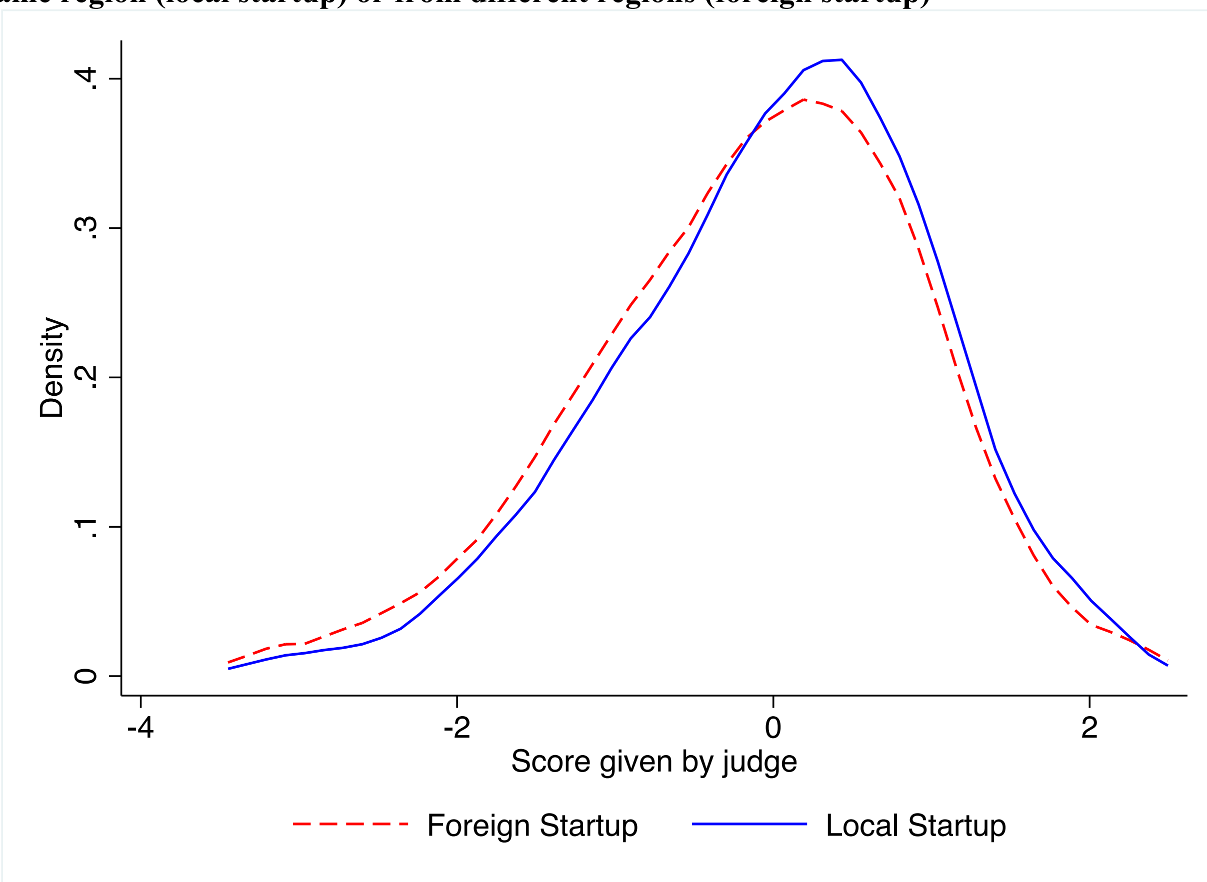


Overall, while we wish we had richer data on judge location, our simulation results show that even if there is relatively severe measurement error in judge location, we should still be able to detect effects, though the estimates are likely to be lower bounds on the true effect. This robustness is largely due to our large sample size, which lets us recover effects even in the face of relatively large measurement error rates. Again, the simulations suggest that, if anything, our findings are underestimates of the true foreign discounting effects.

A2. Kernel density plots showing that judges give lower scores to foreign startups

Tables 3-4 and Figure 3 show that judges discount foreign startups on average. However, this effect might only apply at the left tail of the scoring distribution, which would suggest that the bias only matters for startups that would never be selected anyway, limiting the economic significance of the foreign bias effect. To assess whether this is the case, we plot the distribution of scores by whether the startup is local or foreign to the judge. Figure A2 presents a kernel density plot of these distributions. The scores for local startups first order stochastically dominate (i.e., are always to the right) of the scores given to foreign startups. Foreign discounting matters for startups with both bad and good scores.

Figure A2. Kernel density plot of scores by whether the judge and startup are from the same region (local startup) or from different regions (foreign startup)



A3. Judges across home regions are biased against foreign startups.

Is our foreign bias effect simply the result of judges from a particular geography being biased? For example, perhaps only US judges dislike startups from other countries. Or is the foreign bias broadly based? To adjudicate between these alternatives, in Figure A3, we show the average local and foreign recommendations for North American, European, Latin American, and Israeli judges. In Panel A, we see that US 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[14] as well as the relatively small share of both Israeli and Latin American firms in our sample.[15] When we control for a minimum threshold of quality by limiting our sample to financed startups, shown in Panel B, we find that judges from all regions, but Latin America, discount foreign startups. The effect is not merely the result of a single country being particularly harsh towards foreign firms.

To further test whether non-US judges differ from US judges, in Table A3 we run regressions similar to our primary Table 3 in the paper, but we interact our "foreign startup" dummy with whether the judge is from Latin America, Europe, or Israel. Even though the judge's home region is fixed, since the "foreign startup" dummy varies within a judge, these interaction terms are still identified when we include judge fixed effects. Since we have fixed effects for judges, the main "judge region" dummies that one would normally include drop in the regression. In this regression, the coefficient on "foreign startup" corresponds to the level of bias exhibited by US judges. Each interaction term then reflects if the bias is different for judges from Latin America, Europe, or Israel.

Column 1 in Table A3 shows the results of regressing the judge's score on whether the startup is foreign along with our judge region interaction terms. This model includes region fixed effects. Consistent with our results in Table 3, we find a negative effect on the foreign startup dummy variable. The interaction terms are noisy, and the coefficients for Latin American and Israeli judges are close to zero. While the coefficient for European judges is positive and large, the confidence intervals overlap with zero. Further, in models 2 and 3 when we include country and startup fixed effects, respectively, we see the size of the coefficients shrink towards zero. Overall, these results show that our findings are not driven by judges from one particular region.

[14] 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.
[15] 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.

Figure A3. 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.

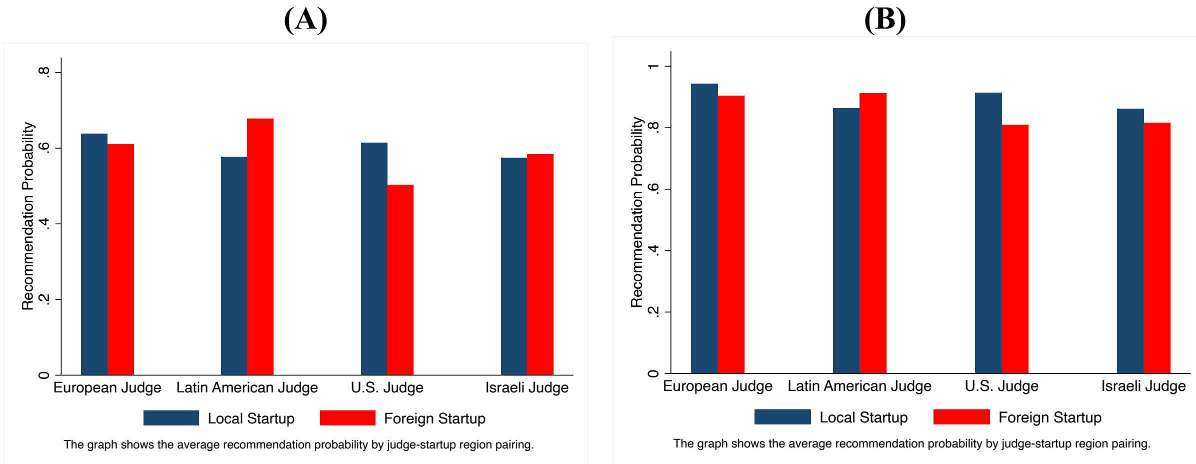


Table A3. Israeli, European, and Latin American judges are similarly discounting foreign startups as are other judges.

	(1)	(2)	(3)
	Judge's Total Score		
Foreign Startup	-0.109**	-0.108**	-0.087**
	(0.038)	(0.039)	(0.033)
Foreign Startup x Latin American Judge	0.099	0.101	0.052
	(0.106)	(0.105)	(0.084)
Foreign Startup x European Judge	0.132	0.130	0.085
	(0.076)	(0.075)	(0.063)
Foreign Startup x Israeli Judge	0.017	0.013	-0.029
	(0.099)	(0.099)	(0.089)
Observations	16,320	16,320	16,264
Judge x Year	Yes	Yes	Yes
Startup Region x Year	Yes	No	No
Startup Country x Year	No	Yes	No
Startup x Year	No	No	Yes

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

A4. Foreign bias results are robust to raw weighted and non-weighted measures of judges' final recommendation score.

Tables 3-4 show that judges discount foreign startups when using a composite measure of judge scores (across all sub-categories discussed in Section III) and a binary recommend measure. A concern with these tables is that they show noisy measures of judges' evaluation of startups because they are not using the ultimate *continuous* measure of final recommendation that judges give to startups. To address this concern, we show in Tables A4a-b the same regressions as in Table 4, but now using two continuous versions of the judges' final recommendation score as the dependent variables. The first is a weighted continuous measure of judges' final recommendation score to startups (on a 0, 2, 4, 12, 16, and 20 scale) with scores 12 and above indicating a positive recommendation (Table A4a). The second is a non-weighted version (on a scale of 0, 1, 2, 3, 4, 5) with scores 3 and above indicating a positive recommendation (Table A4b). The results are consistent with those in Table 4. Judges discount foreign startups even when we control for judge and startup fixed effects.

Table A4a. Regressions showing that judges score lower startups from outside their home region even when we control for judge and startup fixed effects: **Weighted raw recommendation score**

	(1)	(2)	(3)	(4)
		Judge's Total Score		
Foreign Startup	-1.347***	-0.636***	-0.631***	-0.622***
	(0.120)	(0.114)	(0.102)	(0.107)
Has Traction				1.223***
				(0.181)
Has Financing				4.936***
				(0.138)
Observations	17,593	17,593	17,590	17,593
Judge x Year	Yes	Yes	Yes	Yes
Startup Region x Year	No	Yes	No	No
Startup Country x Year	No	No	Yes	No
Startup x Year	No	No	No	Yes

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 A4b. Regressions showing that judges score lower startups from outside their home region even when we control for judge and startup fixed effects: Non-weighted raw recommendation score

	(1)	(2)	(3)	(4)
	Judge's Total Score			
Foreign Startup	-0.277***	-0.130***	-0.131***	-0.127***
	(0.025)	(0.023)	(0.021)	(0.022)
Has Traction				0.258***
				(0.037)
Has Financing				0.969***
				(0.030)
Observations	17,595	17,595	17,592	17,595
Judge x Year	Yes	Yes	Yes	Yes
Startup Region x Year	No	Yes	No	No
Startup Country x Year	No	No	Yes	No
Startup x Year	No	No	No	Yes

Standard errors (in parentheses) are clustered at the judge and startup levels. Fixed effects shown below observations. The table includes two additional data points (0.01% of the total sample) due to labeling changes.
* p<0.05 ** p<0.01 *** p<0.001

A5. Judges are equally likely to give incomplete subscores to foreign and local startups.

Table 3 shows that judges score foreign startups lower than local ones. This table excludes evaluations which received missing subscores from judges. We might be concerned that judges are more likely to give missing subscores to foreign startups. Perhaps this may be because they have a harder time understanding them and therefore "punt" the decision by leaving the score blank. If this were the case, then there would be missing foreign startups in the analysis that may confound our main result. Since foreign startups are given lower scores on average than are local startups, these missing foreign values would likely bias our foreign discounting result upward. In Table A5, we regress whether a judge gives an incomplete score on whether the startup is foreign to judge. We find that judges are equally likely to give incomplete scores to local and foreign startups. This suggests that incomplete subscores are not driven by whether startups are foreign to judges. Therefore, incomplete subscores are unlikely to confound our results.

Table A5. **Probability that judges give incomplete subscores**

	(1)
	Incomplete Subscore
Foreign Startup	0.002
	(0.005)
Observations	17,590
Judge x Year	Yes
Startup x Year	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

A6. Alternative foreignness measures produce similar results.

Our primary measure of whether a startup is foreign is based on the headquarters region of the startup. However, given that the startup application does not list the HQ location of the startup, it remains possible that the discounting could be related to factors unrelated to the startup's location. Further, even if we find evidence that judges are biased against foreign locations, we know that "foreignness" to the judge is likely non-binary but continuous. There likely is a spectrum of how foreign a startup is to a judge. Finally, judges might be picking up on the fact that ideas and founder experiences from different countries might be different. For example, an Israeli judge is likely to understand the pros and cons of backing a team that was part of the IDF's Unit 8200, whereas a Latin American judge might not.

We address each of these three concerns here. First, we used a simple dictionary matching procedure to check if the application text explicitly mentioned the country name when describing the market the startup was targeting, the description of the team, and the startup's coming traction. Admittedly, this approach will miss indirect location indicators, for example, if the application mentions the team is based in the city of Ghent indicating operations in Belgium. However, it will allow us to identify startups that are obviously and explicitly local or foreign to the judge. In Column 1 of Table A6b, we show that using this much more stringent measure still results in judges discounting foreign startups. In Table A6c Column 1, we further show that the foreign bias effect that we pick up with the headquarters of a startup is stronger when the home region is explicitly mentioned in the text according to this measure. In fact, when we limit the sample to startups which do not reveal their home region in their application text, we see that the foreign discounting effect weakens substantially (Column 2).

To address the second concern that distance is not binary but continuous, we generate a simple measure using the geographic distance between the startup's HQ location and the center of the judge's home region, using a measure developed by Berry et al. (2010). This measure serves two functions. First, for a US judge, it would classify startups from Germany as more distant than a startup from London. Second, given the European office of the accelerator is based in Switzerland, for a European judge, startups from Germany would be classified as less foreign than from London. To account for the skew in distances, we log geographic distance as is common in studies of international trade and strategy. Table A6b Column 2 shows the results using our logged distance measure. Consistent with our results in Table 3, we find that judges give lower scores to startups more distant in geographic space even when using a more continuous measure.

To address the third and final concern—that startups from different regions are different in their ideas and approaches—we use natural language processing (NLP) tools to classify the text of startups from different regions as more or less "of that" region. Specifically, we take the application text for each startup and convert it into a "bag of words." We follow standard practice and first remove all punctuation and capitalization before converting the text to a bag of words. We also remove standard stop words. To remove idiosyncratic words, we only retain words that appear in at least 10 startup descriptions and are used at least 20 times. For example, this approach avoids us picking up on startup names that might be unique to a firm. Using these words, we then estimate for each word in our corpus of startup text whether the word is more or less likely to occur in a given region or not. We do so using a weighted log-odds ratio procedure as described in Monroe et al. (2008). The end result is that for each word in our corpus, we know how much more likely (and unlikely) it is to appear in North American, Latin American, European, and Israeli startup applications. Table A6a shows N example words that are most and least likely to appear from each region along with the estimated log-odds. These log-odds give us a quantitative estimate of how regional or localized the startup's application is. To generate measures at the startup level, we simply sum these scores across the application text and divide by the total number of words to account for differences in text length. We then standardize each of these variables to get measures for how North American, Latin American, European, and Israeli each startup is.

Table A6a. Example "localized" words by startup home region

North American words	Raw log-odds ratio	Israeli words	Raw log-odds ratio
opioid	3.61	jerusalem	6.77
sbir	3.48	technion	4.98
nih	3.46	idf	4.37
northeast	3.17	wix	2.79
dartmouth	3.08	jewish	2.91

European words	Raw log-odds ratio	Latin American words	Raw log-odds ratio
epfl	5.59	pesos	6.73
gmbh	3.57	mercado	5.58
organisation	3.3	cdmx	5.6
hec	3.76	colombia	3.7
chf	5.58	argentina	3.16

To construct a single foreignness measure, we use the score for the startup that matches the location of the judge. That is, for an Israeli judge, we use how "non-Israeli" a startup looks as our measure of foreignness for all startups irrespective of location. We do similar procedures for judges from the three other regions. The resulting unidimensional score tells us how foreign a startup's application is relative to startup's home region. In Column 3 of Table A6b, we again find that judges discount startups that appear more foreign.

Overall, the results in Table A6b show that judges discount foreign startups no matter how we measure foreignness.

A6b. Foreign discounting using other measures of foreign startup

	(1)	(2)	(3)
	Judge's Total Score	Judge's Total Score	Judge's Total Score
Foreign Startup (Application Word Search)	-0.039**		
	(0.013)		
Log Geographic Distance		-0.008***	
		(0.002)	
Foreign Startup (Application NLP)			-0.0265
			(0.0074)
Observations	17,590	16,257	17,590
Judge x Year	Yes	Yes	Yes
Startup x Year	Yes	Yes	Yes

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

A6c. Foreign discounting weakens when the region is not explicit in the application text.

	(1)	(2)	(3)
	Judge's Total Score		
Foreign Startup	-0.067**	-0.039	-0.039*
	(0.026)	(0.021)	(0.020)
Foreign Startup * Home Region Explicit in Application			-0.054*
			(0.027)
Observations	6,764	9,356	16,264
Judge x Year	Yes	Yes	Yes
Startup x Year	No	No	No
Region Explicit in App	Yes	No	Yes
Region Not Explicit in App	No	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

A7. Results hold when excluding investor judges and Latin American startups.

Our data includes startups and judges from heterogenous backgrounds. On the one hand, this improves the generalizability of our analysis. On the other, it opens up the possibility that particularly "weird" subgroups drive our findings. Specifically, while most judges are not investors, some are. For these investor judges, we might worry that they give lower scores to foreign startups because they would rather select a local startup that will be easier for them to invest in. On the startup side, we might worry that judges are simply responding to differences in writing quality. Since applications are in English, startups who are from international contexts where English training and education are weaker may be at a disadvantage. This is likely the case for startups from Latin America as English is both used and taught much more often in North America, Europe, and even Israel.

In Table A7a, we show that the foreign bias effect holds when we exclude judges who are investors from our sample. This is not all together surprising since the accelerator guidelines explicitly tell the judges that they should not expect to personally gain by participating.

In Table A7b, we exclude Latin American startups from our sample and again find that the foreign bias effect holds. This largely rules out the idea that poor writing coming from a single region is responsible for our results.

Table A7a. Foreign discounting effect excluding investor judges

	(1)	(2)
	Judge's Total Score	Judge Recommends Startup?
Foreign Startup	-0.063***	-0.044***
	(0.018)	(0.010)
Observations	13,930	15,013
Judge x Year	Yes	Yes
Startup x Year	Yes	Yes

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

Table A7b. Foreign discounting effect excluding Latin American startups

	(1)	(2)
	Judge's Total Score	Judge Recommends Startup?
Foreign Startup	-0.058**	-0.035***
	(0.018)	(0.010)
Observations	13,418	14,505
Judge x Year	Yes	Yes
Startup x Year	Yes	Yes

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

A8. Robustness to check that startups are not strategically disclosing their region in their applications

One concern about the foreign bias results in Tables 3-4 is that startups are strategically mentioning their region in their applications. This would mean that the foreign bias results are partly driven by startups' decisions to disclose their locations, confounding our ability to evaluate judge decision-making. Thankfully, the design of the program mitigates this problem since startups do not know the location of the judges who will be judging them in the rounds we analyze. Thus, a US startup has no control over whether an EU based judge will or will not judge them. This makes it challenging to strategically game disclosure.

That said, you might be concerned that lower-quality startups are less aware of potential foreign bias and so are more likely to disclose their region explicitly. This would suggest that quality measures may bias the region identifiers. We check this in Table A8 where we regress whether a startup mentions their region explicitly in their application on startup quality measures. Column 1 shows that different metrics of quality—whether startups have user traction, financing, or are located in a hub—do not predict whether startups explicitly mention their region in their application.

We also might be concerned that non-US startups think there is a US premium (given that the accelerator is US-headquartered) and thus are more likely to NOT disclose their locations. Column 2 shows that non-US startups are actually *less* likely to disclose their region by about 4 percentage points.

Together, these results suggest that neither quality nor being US-based predict whether startups explicitly mention their regions in their applications, reducing concerns about startups strategically disclosing their location.

Table A8. Startup quality does not predict whether region is explicitly mentioned in applications

	(1)	(2)
	Whether Region Explicit in App	
Log Pre-Accelerator Page Visits	0.009	
	(0.012)	
Log Pre-Accelerator Financing	0.010	
	(0.015)	
Has Traction	-0.025	
	(0.066)	
Has Financing	-0.073	
	(0.088)	
Whether Hub Region	0.047	
	(0.027)	
Whether US HQ		-0.037*
		(0.016)
Observations	9,724	17,594
Judge x Year	Yes	Yes
Startup Region x Year	No	No
Startup Country x Year	Yes	No
Startup x Year	No	No

Standard errors (shown in parentheses) are clustered at the judge and startup levels. Fixed effects shown below observations. The table includes one additional data point (0.01% of the total sample) due to labeling changes.

* p<0.05 ** p<0.01 *** p<0.001

A9. Robustness to country and founder quality measures

To rule out quality differences across countries and startups in Table 3, we include region, country, and startup fixed effects. While these fixed effects account for any time-invariant quality differences between startups or countries, they provide little insight into what aspects of a country or a startup judges rate higher. Are startups from wealthier countries rated better? From more innovative countries? From places with more VC funding? In hubs like Silicon Valley?

To address these questions, in Table A9a, we directly control for these differences. We include GDP Per Capita (World Bank)[16], patent applications to the USPTO (OECD)[17], venture capital availability (the World Economic Forum's Global Competitiveness Index)[18], and whether the startup affiliates with a hub in its application text (using the Startup Genome (2021) classification shown in Table A9c where we also show the distribution of startups in our sample). As expected, judges are more likely to rate startups from wealthier, move innovative, VC-rich, and hub regions more highly. That said, even when we control for these variables directly, we still find a foreign discounting effect.

Similarly, while startup fixed effects account for all time-invariant differences in founder quality, they provide little insight into what aspects of a founder are valued by judges. To directly test how differences in founder background impact judging, we generate measures for whether a founder has a PhD, MBA, or an affiliation with an elite university. We use the text describing the founding team to generate these measures. Thus, if a founder has a PhD but does not mention it in the text, we would mark her as not having a PhD. To generate our measure of whether the founder affiliates with an elite university, we check if the text contains the name of one of the top 10 elite universities in each of the regions in our sample (Europe, North America, and Latin America), according to QS World University Rankings (2021). Table A9b shows models including these controls. Unsurprisingly, founders with MBAs, PhDs, or elite university affiliations are given higher scores. As above, we still find a foreign discounting effect.

[16] Data may be accessed here: https://data.worldbank.org/indicator/NY.GDP.PCAP.CD.
[17] Data may be accessed here: https://stats.oecd.org/Index.aspx?DataSetCode=PATS_COOP.
[18] Data may be accessed here:
https://tcdata360.worldbank.org/indicators/h8a7ea3d1?indicator=529&viz=line_chart&years=2007,2017.

Table A9a. Foreign discounting effect controlling for country HQ quality

	(1)	(2)	(3)	(4)
	Judge's Total Score			
Foreign Startup	-0.133***	-0.067***	-0.104***	-0.195***
	(0.021)	(0.020)	(0.020)	(0.021)
Log Startup HQ GDP Per Capita	0.180***			
	(0.020)			
Log Startup HQ Patent Apps		0.053***		
		(0.005)		
Startup HQ VC Availability			0.166***	
			(0.017)	
Startup Hub				0.114***
				(0.025)
Observations	16,304	16,308	16,306	16,320
Judge x Year	Yes	Yes	Yes	Yes
Startup x Year	No	No	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 A9b. Foreign discounting effect controlling for founder quality

	(1)	(2)	(3)
		Judge's Total Score	
Foreign Startup	-0.059**	-0.061**	-0.062**
	(0.020)	(0.020)	(0.019)
Founder(s) have PhD	0.359***		
	(0.032)		
Founder(s) have MBA		0.193***	
		(0.028)	
Founder(s) Affiliated with Elite University			0.301***
			(0.024)
Observations	16,320	16,320	16,320
Judge x Year	Yes	Yes	Yes
Startup Country x Year	Yes	Yes	Yes

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 A9c. Distribution of startups in our sample by hub

Hub Name	Number of Startups	Share of Startups	Hub Rank in Startup Genome Project (2021)
Amsterdam-Delta	4	0.11	13
Atlanta	4	0.11	26
Austin	8	0.21	20
Bangalore	1	0.03	23
Berlin	15	0.4	22
Bern-Geneva	52	1.38	36
Boston	339	8.97	5
Chicago	32	0.85	14
Dallas	1	0.03	31
Denver-Boulder	6	0.16	27
Dublin	3	0.08	36
Hong Kong	5	0.13	31
London	50	1.32	3
Los Angeles	5	0.13	6
Melbourne	3	0.08	36
Montreal	3	0.08	31
Munich	13	0.34	31
New York City	38	1.01	2
Paris	30	0.79	12
Philadelphia	6	0.16	28
San Diego	2	0.05	21
Seattle	5	0.13	10
Seoul	4	0.11	16
Shanghai	7	0.19	8
Shenzhen	1	0.03	19
Silicon Valley	16	0.42	1
Singapore	9	0.24	17
Stockholm	3	0.08	17
Sydney	1	0.03	24
Tel Aviv	12	0.32	7
Tokyo	6	0.16	9
Toronto-Waterloo	5	0.13	14
Washington, D.C.	5	0.13	11
Other	3,086	81.64	

Startups are identified into hubs based on whether they explicitly mention the hub in the market, team, or traction application fields.
Source: Startup Genome Project (2021)

A10. Foreign bias results hold whether the judge-startup industry matches or not.

One concern with our foreign bias results in Tables 3-4 is that only less or more informed judges are driving foreign bias results. One way we can measure the informedness of judges relative to one another is by whether their industries match with those of the startups. If this were the case, then we might expect that when judges match industries with startups, they would be less biased against foreign startups because they would have other ways to discern quality of the startups through their industry expertise. They also would be better able to detect the quality of startups overall.

As shown in Table A10a, when we adapt the specification from Tables 3-4, judges are no less likely to be biased against foreign startups when their industries match, as shown by the interaction term between whether a startup is foreign and whether judge-startup industries match in Column 3.

Further, in Table A10b, when we adapt the specification from Table 5, we find that judges are no better at detecting quality of startups when their industries match (Column 2), versus when they do not match (Column 3). Of course, these results might be partly capturing selection on industry matches (unlike geographic matches). The accelerator partly allocates judges to startups on the basis of industry matches. Further, industry categories are broadly construed. For example, both consumer application and cybersecurity technology startups are classified as “high technology” (the largest industry category comprising nearly 40 percent of judges and startups). This means that even within an industry category, there is variation in expertise areas.

There may also be concern that the foreign bias measure is actually reflecting industry bias. This might be the case because industries are often concentrated in certain geographies. To test for this confounding effect, we assess whether judges are biased against startups from different industries. To do so, we apply a similar specification as in equation 1, but replace whether a startup is foreign to the judge, to whether the startup is from a different region as the judge. In Table A10c, we show that judges do not discount startups from a different industry. If anything, the coefficient is actually positive (though noisy). This result may reflect that startups are following standardized enough business models that judges from across industries can understand them.

Together, these results suggest that more informed judges as proxied by industry are not driving our results.

Table A10a. Judges are similarly biased against foreign startups no matter whether their industry matches with that of the startups.

	(1)	(2)	(3)
	Judge's Total Score		
Foreign Startup	-0.018	-0.070**	-0.037
	(0.034)	(0.023)	(0.022)
Foreign Startup * Different Industry			-0.039
			(0.027)
Observations	5,179	8,847	16,264
Judge x Year	Yes	Yes	Yes
Startup x Year	Yes	Yes	Yes
Same Industry	Yes	No	Yes
Different Industry	No	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 A10b. Judges can tell the quality of foreign vs. local startups with similar precision no matter whether their industry matches with that of the startups.

	(1)	(2)	(3)
	Judge's Total Score		
	Baseline	Same Industry	Different Industry
Foreign Startup	-0.065**	-0.068	-0.057
	(0.025)	(0.039)	(0.032)
Log Post-Accelerator Page Visits	0.050***	0.051***	0.048***
	(0.004)	(0.005)	(0.005)
Foreign Startup * Log Post-Accelerator Page Visits	0.000	0.005	0.000
	(0.005)	(0.007)	(0.006)
Observations	16,320	6,471	9,625
Judge x Year	Yes	Yes	Yes
Startup Country x Year	Yes	Yes	Yes
Startup x Year	No	No	No
Same Industry	Yes	Yes	No
Different Industry	Yes	No	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 A10c. Judges do not discount startups from outside of their industry.

	(1)	(2)
	Judge's Total Score	Recommend
Different Industry	0.022	0.010
	(0.015)	(0.009)
Observations	16,264	17,591
Judge x Year	Yes	Yes
Startup x Year	No	No

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

A11. Hub location does not explain the foreign discounting effect.

Figure A3 shows that judges discount foreign startups when located in hubs like the US, Europe, and Israel, but not necessarily elsewhere like in Latin America. Indeed, Latin American startups differ from US, European, and Israeli startups because, even if they are located in a major city, that city will be a less developed tech hub than Silicon Valley, London, or Tel Aviv is. This raises the concern that if our effects are driven by Latin American startups then our findings might reflect differences in how judges rate startups from hubs vs. non-hubs rather than foreign discounting. To test this possibility, in Appendix 7, we show that our foreign discounting effect remains even when we exclude Latin American startups from our sample.

To address the broader concern that we are picking up discounting of non-hub startups, we first classify every startup in our sample as operating in a hub or not. As described in Appendix 9, we used data from PitchBook and the Startup Genome Project (2021) to map the city a startup is in to whether it is classified as a startup hub. In Table A11 (Column 2), we show that controlling for whether a startup is in a hub or not does not alter our foreign discounting effect.

Even if hub location does not explain the foreign discounting effect, perhaps it is the case that foreign discounting occurs for startups only inside or outside of hubs. To understand whether hub location has such a moderating effect, in Table A11 (Column 3), we add an interaction term between whether a startup is in a hub and whether it is foreign. This interaction term in the third row turns out to not be meaningful, suggesting that hub location does not moderate foreign discounting.

Together, these analyses suggest that hub location is neither explaining the foreign discounting effect, nor moderating it.

Table A11: Whether a startup is in a hub does not remove or moderate the foreign discounting effect.

	(1)	(2)	(3)
	Judge's Total Score		
Foreign Startup	-0.060***	-0.060**	-0.051*
	(0.016)	(0.020)	(0.021)
Hub		0.066**	0.089**
		(0.025)	(0.031)
Foreign x Hub			-0.047
			(0.039)
Observations	16,264	16,321	16,321
Judge x Year	Yes	Yes	Yes
Startup Region x Year	Yes	No	No
Startup Country x Year	Yes	Yes	Yes
Startup x Year	Yes	No	No

Standard errors (shown in parentheses) are clustered at the judge and startup levels. Fixed effects shown below observations. The table includes one additional data point (0.01% of the total sample) due to labeling changes.
* p<0.05 ** p<0.01 *** p<0.001

A12. The findings generalize to all years and rounds of the accelerator program.

Our main results show that judges discount foreign startups when provided only textual information about the startups during the first round of judging in 2017 and 2018. The benefits of focusing on this restricted sample are meaningful. Judges were randomly assigned to startups on the basis of home region allowing us to estimate a foreign discount effect without worry that the estimate is merely the result of selection bias. Further, since judges only evaluated the text application, we know the full set of information that the judges based their decision on.

However, there is no such thing as a free lunch. The focus on this restricted sample may raise concern that our results might be particular to this first round of judging in 2017 and 2018. Fortunately, we have access to a broader sample of judge-startup evaluations spanning all years of the accelerator (2013-2019) and from all four rounds of the program. These latter rounds include in-person interaction between judges and startups. While judges were far from randomly assigned in this sample, the larger dataset allows us to check if our results generalize beyond our unique dataset.

Column 1 in Table A12 shows that judges still discount foreign startups in this broader sample that includes all rounds from 2013-2019. Furthermore, in Column 2, we show that the foreign discounting effect holds even if we only focus on the later rounds of the competition. The effect remains negative. While beyond the scope of our study, this suggests that interviewing or interacting with founders—as is common in the VC due diligence process—will not eliminate foreign bias.

Table A12. Foreign discounting effect across all years and rounds of the accelerator program

	(1)	(2)
	Judge's Total Score	
	All rounds (1-4)	Excluding first round
Foreign Startup	-0.064***	-0.118*
	(0.012)	(0.053)
Observations	69,639	24,883
Judge x Year	Yes	Yes
Startup Country x Year	Yes	Yes
Program x Year	Yes	Yes

The table shows evaluations 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

A13. Robustness using alternative measures of startup quality

An important concern with the results in Table 5, which show that judges are no more informed about local as against foreign startups, is that the findings might be specific to the particular quality proxies that we use. If our proxy for startup quality is too noisy, the lack of a difference in how informed the judges are might simply be the result of our noisy measure, not the fact that the judges are actually equally informed.

To address this concern, here we replicate our result in Table 5 using a variety of different quality proxies. Further, we use quality measures from both before and after the program. This lets us rule out the possibility that the post-program measures are somehow influenced by the judges who evaluated the startup and thus are biased. Our measures of quality from before the program include a binary measure of whether a startup has financing at the time of the application, a binary measure indicating whether a startup has user traction (as measured with having at least 100 page visits on average per month over the last three months before the program), logged financing value prior to the program, and logged page visits prior to the program.

However, our true focus is not necessarily observable quality at the time of application and is instead whether the startup will be a success, a much harder quantity to measure before the program. Fortunately, once we control for whether the startup is admitted into the program, it seems reasonable to assume judge scores have essentially no influence on future startup performance. Thus, we can use the realized outcomes for the startup in our sample to measure their quality. Specifically, using data from Pitchbook, we collect measures as of 3-4 years after the program on the logged number of employees, revenue growth, and logged valuation. We also construct a quality index by taking the three variables previously mentioned, as well our two post-accelerator quality measures from the main paper (logged page visits and funding after the program), and then standard normalize each variable. We then add these variables together and again normalize to construct our final index.

Table A13a replicates Table 5 but uses our four measures of pre-program startup quality. In all cases, we find that judges give higher scores to higher quality startups, this estimate is the same for local and foreign startups, and judges discount foreign startups. Figures A12a and A12b present graphical evidence for our two continuous measures of startup quality. The binscatters show that judges are biased against foreign startups as the red dotted line is above the blue solid line across the quality distribution of startups. They are also able to the detect the quality of local and foreign startups with equal precision as shown by the similarly positively sloped red dotted and blue solid lines.

Table A13b replicates A13a but uses our four post-program realized success measures of startup quality. We find similar patterns as with our pre-program measures. Column 4, which uses our index measure, shows that a one standard deviation in quality leads to a 0.1 standard deviation increase in the score a judge will give to a startup. This suggests that while judges are informed, they are far from oracles.

Table A13a. The ability of judges to evaluate startup quality: Pre-accelerator quality measures

	(1)	(2)	(3)	(4)
		Score		
Foreign Startup	-0.045*	-0.038	-0.053**	-0.039
	(0.019)	(0.027)	(0.020)	(0.027)
Has Financing	0.773***			
	(0.029)			
Foreign Startup * Has Financing	-0.097**			
	(0.038)			
Has User Page Traction		0.264***		
		(0.036)		
Foreign Startup * Has User Traction		-0.042		
		(0.036)		
Log Pre-Accelerator Financing			0.154***	
			(0.007)	
Foreign Startup *Log Pre-Accelerator Financing			-0.012	
			(0.009)	
Log Pre-Accelerator Page Visits				0.036***
				(0.006)
Foreign Startup * Log Pre-Accelerator Page Visits				-0.006
				(0.007)
Observations	16,320	16,320	16,320	9,063
Judge x Year	Yes	Yes	Yes	Yes
Startup Country x Year	Yes	Yes	Yes	Yes

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 A13b. The ability of judges to evaluate startup quality: Post-accelerator quality measures

	(1)	(2)	(3)	(4)
		Score		
Foreign Startup	-0.052*	-0.048*	-0.047*	-0.044*
	(0.020)	(0.020)	(0.020)	(0.020)
Log Post-Accelerator Employees	0.115***			
	(0.014)			
Foreign Startup * Log Post-Accelerator Employees	0.013			
	(0.017)			
Post-Accelerator Revenue Growth		0.027		
		(0.043)		
Foreign Startup * Post-Accelerator Revenue Growth		0.031		
		(0.054)		
Log Post-Accelerator Valuation			0.123***	
			(0.030)	
Foreign Startup * Log Post-Accelerator Valuation			-0.003	
			(0.038)	
Post-Accelerator Performance Index				0.101***
				(0.016)
Foreign Startup * Post-Accelerator Performance Index				0.027
				(0.020)
Accelerator Participation	0.694***	0.768***	0.743***	0.625***
	(0.032)	(0.030)	(0.031)	(0.037)
Foreign Startup * Accelerator Participation	-0.091*	-0.096*	-0.096*	-0.117*
	(0.040)	(0.039)	(0.040)	(0.048)
Observations	16,320	16,320	16,320	16,320
Judge x Year	Yes	Yes	Yes	Yes
Startup Country x Year	Yes	Yes	Yes	Yes

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

Figure A13a. 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

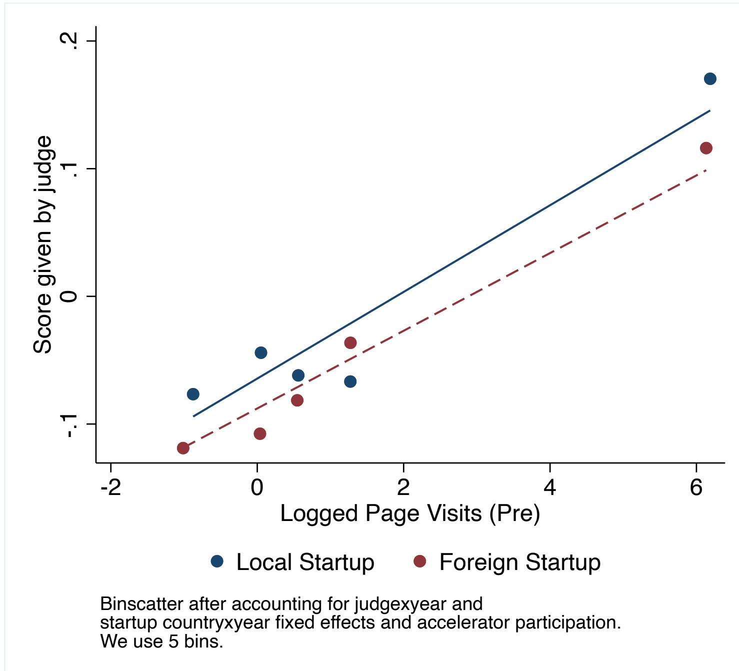
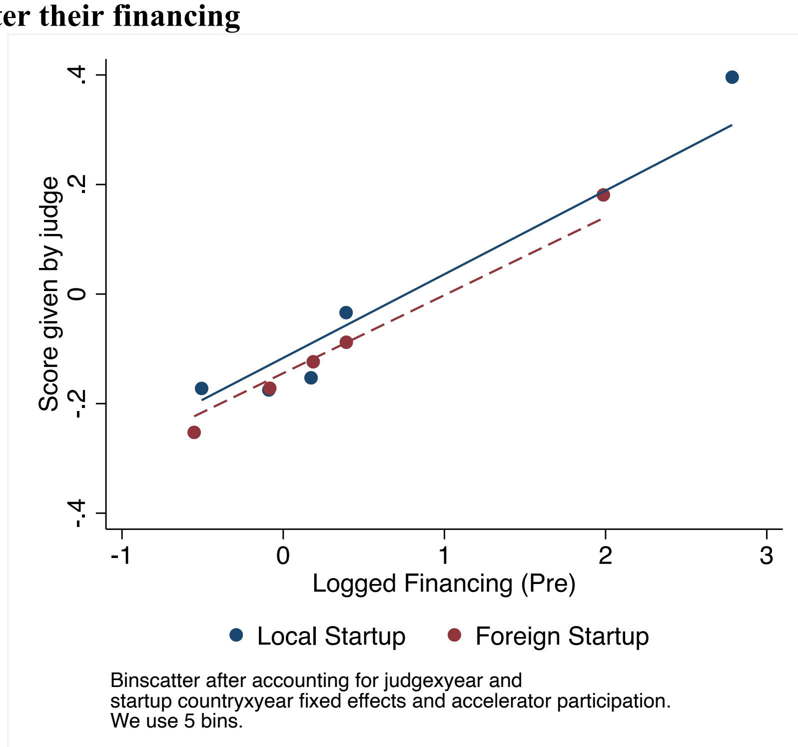


Figure A13b. Binscatter showing that judges can give higher scores to startups with more financing before entering the accelerator program, but consistently discount foreign startups no matter their financing



A14. Judges are equally responsive to quality when we split our sample by foreignness.

Table 5 shows that judges are able to tell winners from losers with equal precision among local and foreign startups. One concern with this estimate is that the extent to which the judge's score accounts for startup quality might vastly differ among foreign vs. local startups. Conversely, our measures of startup quality may account for the judge's score to different extents among local and foreign startups. For example, a European judge may be able to account for much more of a European startup's quality than for a US startup's quality because of their contextual knowledge of the European market. To test this, we compare the R-squared statistics of regressions assessing the relationship between judge scores and startup quality for local versus foreign startups. Specifically, we show the R-squared statistics for models regressing the judge's score on quality (Table A14a) and quality on score (Table A14b) are similar when we split the sample by foreignness.

In Table A14a, Models 1-2 include no additional fixed effects and show judges give higher quality scores to both local and foreign startups. Moreover, the R-Squared values are nearly identical. In Models 3-4, we include judges-year fixed effects. Again, the within-R2s are nearly identical. Models 5-6 phase in country-year fixed effects. We find similar patterns.

Table A14b mirrors Table A14a but swaps the dependent and independent variables. Since judges are randomly assigned to evaluate startups, there is no need to include judge fixed effects. Further, in this model, judge fixed effects account for outcome differences in the startups the judge evaluates and not differences in judge harshness. That said, the results below are essentially unchanged when including judge fixed effects. Again, we find that judges can crudely predict which startups will succeed and which will not.

At least in our full sample, it does not appear that judges are any better at evaluating local startups as against foreign ones. Overall, we find R-squared estimates that, depending on the model, range from 3-to-6 percent.

Table A14a. The R-squared values for the extent to which judges are informed are similar when we split our sample into startups foreign to the judge and startups local to the judge.

	(1)	(2)	(3)	(4)	(5)	(6)
			Judge's Total Score			
Log Post-Accelerator Page Visits	0.367***	0.4104***	0.3862***	0.4082***	0.3876***	0.3775***
	(0.0299)	(0.0287)	(0.0267)	(0.0251)	(0.0268)	(0.0255)
Observations	7,232	9,107	7,232	9,107	7,232	9,107
Sample	Local	Foreign	Local	Foreign	Local	Foreign
Judge x Year	No	No	Yes	Yes	Yes	Yes
Startup Country x Year	No	No	No	No	Yes	Yes
R-Squared	0.03279	0.03757	0.44205	0.47552	0.44633	0.49148
Within R-Squared			0.05086	0.05701	0.05087	0.04856

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 A14b. Regressions showing that judges' scores predict future startup performance.

	(1)	(2)	(3)	(4)
	Log Post-Accelerator Page Visits			
Judge's Total Score	0.0893***	0.0916***	0.0922***	0.082***
	(0.0068)	(0.0065)	(0.0066)	(0.0065)
Observations	7,232	9,107	7,232	9,107
Sample	Local	Foreign	Local	Foreign
Startup Country x Year	No	No	Yes	Yes
R-Squared	0.03279	0.03757	0.05712	0.07963
Within R-Squared			0.03516	0.03016

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

A15. Judges are worse at detecting quality of startups that have raised financing, are from hubs, or have founders affiliated with elite universities.

Table 5 shows that judges can detect startup quality. This finding contradicts other studies that show that judges often cannot detect quality of startups, especially startups in consumer and enterprise high-technology sectors (Scott et al., 2020). However, in this past work judges often evaluate startups form a highly pre-selected pool. For example, startups in Scott et al. (2020)'s study are all from an elite university (MIT) and so reflect a much higher quality pool than is present in our study, which includes both MIT founders and those with less prestigious educational credentials. Perhaps differences in the variability of the pool of startups being evaluated explain when judges can separate the best startups from the worst.

To test this hypothesis, we restrict our sample to startups that look more like the firms analyzed in Scott et al. (2020). We do so in three ways. First, we measure affiliation with an elite university by whether the application text team portion mentions the name of one of the top 10 elite universities in each of the regions in our sample (Europe, North America, and Latin America), according to QS World University Rankings (2021). Second, we restrict our sample to only startups that had raised funding at the time of their application. Third and finally, we measure if a startup is connected to a hub region like Silicon Valley by whether the application mentions the name of a hub region as defined by the Startup Genome Project (2021).

Table A15 presents regression results with samples split by each of these three measures. We include our primary measure of startup quality, logged post-accelerator page visits, along with our foreign dummy variable. We do not include the interaction term for two reasons. Following Scott et al. (2020), we do not include the interaction as our focus here is on whether judges can detect quality in these subsamples, not on whether they are more or less informed about local startups.

We find that judges' scores are less responsive to quality differences between startups in the "high-quality" sub-samples as evidenced by the lower magnitude of the coefficient on "log post-accelerator page visits." The coefficient plot in Figure 4 reveals that these estimates are meaningfully different from our baseline estimate. Unlike the pronounced drop we see in the ability of judges to detect quality differences, we find that the foreign discounting effect remains relatively stable across the different subsamples. While the effect becomes noisy for startups with founders from elite universities, it is actually larger for both financed and startups connected to hub regions. Though beyond the scope of our paper, this difference could be because judges turn to geography to discern one startup from another when it is otherwise hard to choose among a sample of relatively high-quality startups.

Table A15. Foreign discounting effect across different quality sub-samples of startups

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Judge's Total Score						
	Baseline	Founders from Elite Universities	Founders Not from Elite Universities	Financed Startups	Non-Financed Startups	Startups Targeting Hubs	Startups Not Targeting Hubs
Foreign Startup	-0.064**	-0.054	-0.069**	-0.112*	-0.040*	-0.107*	-0.054*
	(0.020)	(0.050)	(0.021)	(0.046)	(0.020)	(0.049)	(0.022)
Log Post-Accelerator Page Visits	0.050***	0.031***	0.055***	0.006	0.042***	0.031***	0.052***
	(0.003)	(0.006)	(0.003)	(0.006)	(0.003)	(0.006)	(0.003)
Observations	16,320	2,899	13,124	1,596	14,323	2,734	13,265
Judge x Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Startup Country x Year	Yes	Yes	Yes	Yes	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

A16. Judges are worse at selecting winners from losers among startups in later stages of the competition.

Table 5 shows that judges can pick winners from losers among the startups. Can they still do so when they get to later rounds of the competition with a higher quality pool of startups? In Table A16, we apply the same specification as for Table 5, but for the latter two rounds of the accelerator. Column 2 shows that judges' ability to detect startup quality declines relative to the baseline results in Table 5 for post-accelerator page visits by 0.03 standard deviation (Columns 1-2) and by 0.135 standard deviation for post-accelerator financing (Columns 3-4). These results suggest that judges might be able to screen in early rounds, but as the sample of startups becomes higher quality, their ability declines, consistent with results from Table A15. As a result, they may need to turn to alternative experimentation (e.g., "spray-and-pray") measures—as found in the venture capital context (Ewens et al., 2018)—to detect startup quality.

Table A16. Judges' ability to detect startup quality among startups accepted into the accelerator program declines.

	(1)	(2)	(3)	(4)
	Judge's Total Score			
Log Post-Accelerator Page Visits	0.050***	0.020**		
	(0.003)	(0.006)		
Log Post-Accelerator Financing			0.165***	0.030***
			(0.007)	(0.007)
Observations	16,320	3,898	16,320	9,753
Judge x Year	Yes	Yes	Yes	Yes
Startup Country x Year	Yes	Yes	Yes	Yes
Startup x Year	No	No	No	No
Initial Round	Yes	No	Yes	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

A17. Judges can equally detect quality of local vs. foreign startups no matter whether startups' regions are mentioned in their applications.

While Table 5 shows that judges can detect startup quality of local and foreign startups with similar precision, it may be the case that foreign discounting is absorbing a local information advantage. Thus, judges might be able to detect the quality of both local and foreign startups only if their region is made explicit in their applications, suggesting that the geography of startups is already "priced" into the judgement. In Table A17, we show that judges can detect quality of startups with even better precision when the region is not explicitly stated in the application of startups. Indeed, the coefficient on logged post-accelerator page visits increases by 0.01 standard deviation when the region is not made explicit in the application (Column 3) relative to when it is explicit (Column 2). Meanwhile, there remains no local information advantage when the region is not made explicit. This suggests that location does not provide an informational value that is already incorporated into the judging. If anything, it seems to worsen judges' sensitivity to the quality of startups, perhaps because it leads them to rely on their location preferences.

Table A17. Judges can similarly detect the quality of foreign vs. local startups no matter whether startups explicitly mention their region in their applications.

	(1)	(2)	(3)
	Judge's Total Score		
	Baseline	Region Explicit in Startup App	Region Not Explicit in Startup App
Foreign Startup	-0.065**	-0.082*	-0.054
	(0.025)	(0.036)	(0.034)
Log Post-Accelerator Page Visits	0.050***	0.042***	0.053***
	(0.004)	(0.005)	(0.005)
Foreign Startup * Log Post-Accelerator Page Visits	0.000	0.006	-0.003
	(0.005)	(0.007)	(0.006)
Observations	16,320	6,786	9,395
Judge x Year	Yes	Yes	Yes
Startup Country x Year	Yes	Yes	Yes
Startup x Year	No	No	No
Region Explicit in App	Yes	Yes	No
Region Not Explicit in App	Yes	No	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

A18. Judges have a local information advantage when assessing very localized startups.

In Table 5, we show that judges detect the quality of foreign and local startups with equal precision. This finding contrasts from past studies (e.g., Coval and Moskowitz, 1999; 2001; Malloy, 2005) that show that judges have a local information advantage. One possible reason for this difference is that the industries and business models (mainly high technology) that represent the majority of our startups are geographically agnostic. They follow standardized models, for example, originating from Silicon Valley, such as software-as-a-service. Indeed, Table A18a shows that the top words appearing in applications of startups across the four regions in our sample are similar. The presence of these common words—such as “market,” “sales,” and “platform”—suggests that startups may be pursuing increasingly standardized approaches across geographies. In contrast, the companies in older studies finding a local information advantage tend to be more localized, in terms of producing non-traded goods, being smaller size, or being located in remote regions (Coval and Moskowitz, 1999; 2001).

To confirm if the difference in sample composition of localized companies may account for the difference in results in our study relative to past studies on local information advantage, we split our sample into startups that include terms in their application text that are more specific to their home region versus not (using the NLP approach described in Appendix 6) and the rest. In Table A18b, we show that judges have more of a local information advantage when evaluating startups in such geographically sensitive sectors relative to other startups. However, this localized group of startups is an extremely small share of the sample (less than 5 percent). As a result, our overall result shows the lack of a local information advantage among startups.

Table A18a. Top words in application by startup home region

	Europe			Israel		
	Word	**Count**	**Log Odds Ratio**	**Word**	**Count**	**Log Odds Ratio**
1	**market**	2166	0.151	**market**	1008	0.179
2	**business**	1414	0.193	**product**	571	0.21
3	**companies**	1084	0.0338	**companies**	567	0.206
4	**sales**	1027	0.0683	**business**	538	-0.00536
5	**product**	1026	-0.0374	**experience**	515	0.312
6	**data**	1025	0.0986	**users**	478	0.169
7	**marketing**	984	0.0559	**people**	473	0.172
8	**platform**	923	-0.0653	**marketing**	456	0.0915
9	revenue	919	-0.0607	**platform**	452	0.044
10	**people**	916	0.0122	based	384	0.337
11	**customers**	888	0.0953	**time**	384	0.054
12	**experience**	831	-0.0578	**customers**	381	0.0419
13	**time**	818	-0.0027	**data**	373	-0.134
14	food	806	0.776	**social**	373	0.000594
15	development	805	0.199	**company**	354	0.114
16	team	783	0.103	**online**	351	0.151
17	**users**	782	-0.195	solution	350	0.509
18	**technology**	769	0.0173	ip	343	0.203
19	**online**	725	0.0679	**technology**	343	0.016
20	**social**	715	-0.19	israel	331	4.64

	Latin America			North America		
	Word	Count	Log Odds Ratio	Word	Count	Log Odds Ratio
1	de	1329	3.93	**market**	5777	-0.14
2	**market**	1227	-0.044	**business**	3495	-0.236
3	mexico	1119	4.43	revenue	3455	0.317
4	**companies**	1009	0.43	**product**	3439	0.0485
5	**business**	979	0.227	**data**	3235	0.182
6	**people**	977	0.577	**sales**	3060	-0.0326
7	**platform**	772	0.205	**platform**	3015	-0.0729
8	**sales**	768	0.201	**users**	2953	0.018
9	**experience**	737	0.282	**companies**	2932	-0.315
10	**social**	719	0.301	**marketing**	2857	-0.119
11	**marketing**	690	0.115	**social**	2627	-0.0383
12	**users**	660	0.0901	**time**	2604	-0.011
13	**company**	621	0.312	**customers**	2554	-0.0664
14	**product**	572	-0.231	**experience**	2536	-0.216
15	**online**	569	0.258	**technology**	2491	0.0749
16	products	549	0.377	**people**	2443	-0.38
17	**time**	541	-0.00681	team	2274	-0.0273
18	**customers**	538	-0.0168	ip	2178	0.171
19	usd	530	2.32	**company**	2145	-0.151
20	services	523	0.481	cost	2133	0.0826

Table A18b. As we restrict the sample to more localized startups, we see that judges become better at evaluating startup quality and worse at evaluating foreign startup quality.

	(1)	(2)	(3)
	Judge's Total Score		
Foreign Startup	-0.2835*	-0.2151	-0.3726
	(0.1400)	(0.1998)	(0.2328)
Log Post-Accelerator Page Visits	0.0401*	0.0608*	0.2001**
	(0.0180)	(0.0304)	(0.0576)
Foreign Startup *Log Post-Accelerator Page Visits	0.0158	-0.0645	-0.1432*
	(0.0216)	(0.0455)	(0.0699)
Observations	1,225	485	157
Localness cutoff	0.5 S.D.	0.75 S.D.	1 S.D.
Regional Score Controls	Yes	Yes	Yes
Judge x Year	Yes	Yes	Yes
Startup Country x Year	Yes	Yes	Yes

The first column includes all startups that have a score of 0.5 standard deviations for their home region and are below 0.5 standard deviations for all other regions. The second column raises this cutoff to 0.75. The third column to 1. Standard errors (in parentheses) are clustered at the judge and startup level. Fixed effects shown below observations.
* $p<0.05$ ** $p<0.01$ *** $p<0.001$

A19. Additional analyses estimating "missed startups."

In the primary manuscript, we discuss how we use rich fixed effects models to estimate that judges reject about 5 percent of promising startups due to foreign bias that, in the absence of such bias, would likely have been selected. These fixed effect models include both startup fixed effects and "foreign discounting" fixed effects for each judge.

Here, we use the startup fixed effects to show that the "missed" startups are not especially low quality, but come from the core of the accepted startup quality distribution. To do so, in Figure A19, we plot the kernel density of estimated startup quality, as measured using the startup fixed effects, in blue. The x-axis shows the "quality" in terms of the judge's standardized z-score and so is easily interpretable in terms of standard deviations. The dotted red line shows the estimated quality for the startups missed due to foreign bias. These are the startups we think should have been selected for the next round if the accelerator when judges' scores are "debiased." The figure reveals that missed startups are not just the "bad" startups that would have been rejected regardless. Instead, as shown by the larger overlap between the two distributions, these missed startups are of similar quality to those that make it to the next round. While the very best startups make it to the next round no matter, this plot points to the idea that judges appear to be discounting promising companies.

To triangulate our back-of-the-envelope results, we conduct two additional approaches to the calculation estimating missed out startups. These calculations compare the startups judges would have selected if they only relied on quality-dependent measures and not on the startup's foreign status to the startups they selected when considering 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 pre-accelerator and post-accelerator—and so construct counterfactual rankings as if judges are unbiased but still selected the same number of startups.[19] We then compare this ranking to two alternatives. The first is the *actual* recommendation of the judge. The second is the "biased" counterfactual ranking that uses 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 reveals how many foreign startups are missed when we introduce foreign bias on top of "unbiased" quality-based evaluations.

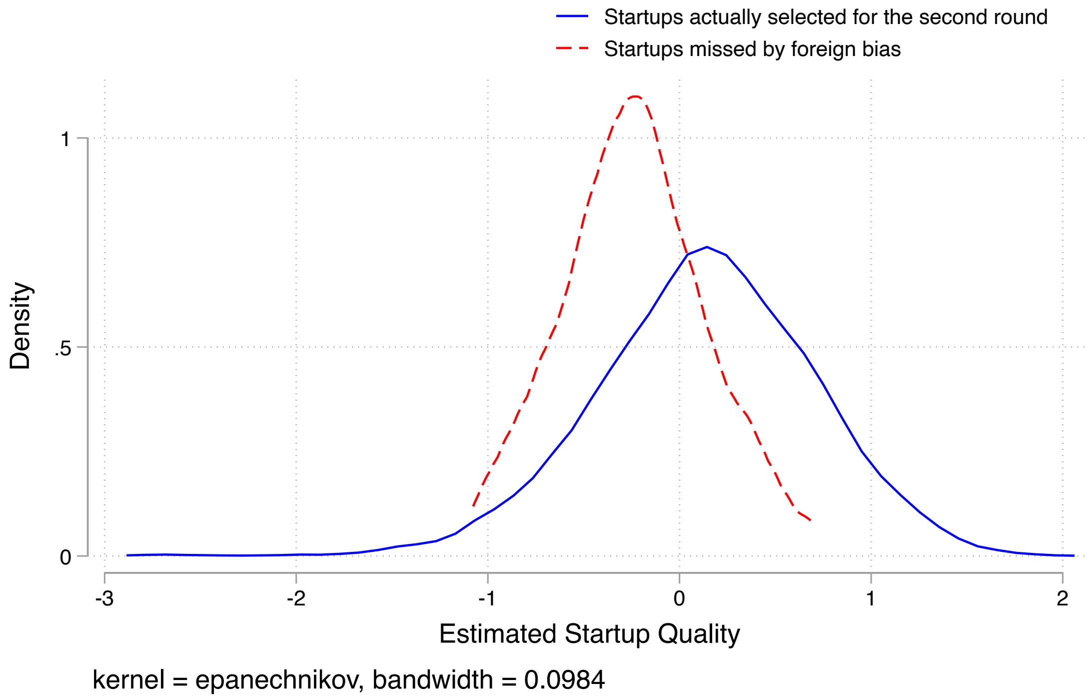
In these additional back-of-the-envelope counterfactuals, we find that foreign bias impacts the number of foreign startups that are recommended to the next round of the competition. We find that moving to evaluations only based on quality leads to 512 more foreign startups being recommended, accounting 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. When we use only the criteria-based score, 312

[19] 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.

fewer 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.

These calculations show that a higher share of startups—9-to-14 percent—are passed over, suggesting our main estimate of 5 percent is conservative.

Figure A19. Distribution of startups actually selected versus estimated to be missed with foreign discounting



A20. Application questions that startups answer and judges evaluate in the accelerator program

We show the application questions that startups fill out and judges evaluate in the venture competition below.

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.

A21. Robustness when including elite university and hub controls

Table 5 shows that judges can equally detect startups that are foreign and local to them. While the regressions control for page visits, financing, and headquarters country as proxies of startup quality, we might be concerned that other measures of startup quality like whether executives affiliate with an elite university or whether the startup is headquartered in a hub might confound this result. To address this concern, we show in Table A21 the same results as Table 5, but with controls for whether a startup executive is affiliated with an elite university and whether the startup is headquartered in a hub. The results are similar. Judges are no better at detecting of local startups as they are of foreign ones.

Table A21: Regressions showing judges (1) give higher scores to more successful startups, (2) are equally good at evaluating success for local and foreign startups alike, and (3) still discount foreign startups when controlling for elite university affiliation and hub location

	(1)	(2)	(3)	(4)	(5)	(6)
			Judge's Total Score			
Foreign Startup	-0.068**	-0.058*	-0.044	-0.052**	-0.046*	-0.040*
	(0.024)	(0.024)	(0.025)	(0.019)	(0.019)	(0.020)
Log Post Accelerator Page Visits	0.049***	0.036***	0.042***			
	(0.004)	(0.004)	(0.004)			
Foreign Startup * Log Post Accelerator Page Visits	0.002	0.004	0.000			
	(0.005)	(0.005)	(0.005)			
Log Post Accelerator Financing				0.163***	0.024*	0.158***
				(0.008)	(0.012)	(0.038)
Foreign Startup *Log Post Accelerator Financing				-0.015	0.008	-0.022
				(0.011)	(0.015)	(0.060)
Accelerator Participation		0.654***			0.681***	
		(0.032)			(0.041)	
Foreign Startup * Accelerator Participation		-0.124**			-0.119*	
		(0.041)			(0.054)	
Elite University	0.297***	0.252***	0.266***	0.260***	0.242***	0.257***
	(0.023)	(0.022)	(0.025)	(0.023)	(0.022)	(0.025)
Hub	0.043	0.036	0.056*	0.046	0.038	0.058*
	(0.024)	(0.023)	(0.026)	(0.024)	(0.024)	(0.026)
Observations	16,321	16,321	14,476	16,321	16,321	14,476
Judge x Year	Yes	Yes	Yes	Yes	Yes	Yes
Startup Country x Year	Yes	Yes	Yes	Yes	Yes	Yes
Startup x Year	No	No	No	No	No	No
Accelerator Participation	Yes	Yes	No	Yes	Yes	No

Standard errors (in parentheses) are clustered at the judge and startup level. Fixed effects shown below observations. The table includes one additional data point (0.01% of the total sample) due to labeling changes. * p<0.05 ** p<0.01 *** p<0.001

References

Berry, H., Guillén, M. F., & Zhou, N. (2010). An institutional approach to cross-national distance. *Journal of International Business Studies*, *41*(9), 1460-1480.

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.

Ewens, M., Nanda, R., & Rhodes-Kropf, M. (2018). Cost of experimentation and the evolution of venture capital. *Journal of Financial Economics*, *128*(3), 422-442.

Malloy, C. J. (2005). The geography of equity analysis. *The Journal of Finance*, *60*(2), 719-755.

Monroe, B. L., Colaresi, M. P., & Quinn, K. M. (2008). Fightin'words: Lexical feature selection and evaluation for identifying the content of political conflict. *Political Analysis*, *16*(4), 372-403.

QS World University Rankings. (2021). QS World University Rankings 2022. Retrieved from https://www.topuniversities.com/university-rankings/world-university-rankings/2022.

Scott, E. L., Shu, P., & Lubynsky, R. M. (2020). Entrepreneurial uncertainty and expert evaluation: An empirical analysis. *Management Science*, *66*(3), 1278-1299.

Startup Genome Project. (2021). Rankings 2021: Top 30 + Runners-up. Retrieved from https://startupgenome.com/article/rankings-2021-top-30-plus-runners-up.