Identifying and Impacting "Gazelles": Evidence from Business Accelerators[1]

February 2019

Juanita González-Uribe **Santiago Reyes**

LSE **IADB**

High-growth young firms disproportionately contribute to growth but identifying "gazelles" remains challenging. Can business accelerators identify and impact high-growth firms? Our context is an accelerator that selects participants based on scores from randomly assigned judges, and provides them with *entrepreneurship schooling*, but *no cash*. We show that judges can identify gazelles, but that correcting for heterogeneity in judges' scoring generosity would have improved selection. Exploiting exogenous differences in judges' scoring generosity, we show that schooling causally impacts growth, particularly for high-growth applicants. The results are most relevant for accelerators that attract ventures with traction and suggest that entrepreneurial capital matters.

JEL Classification: G24, L26, M13
Keywords: Business Accelerators, Entrepreneurial Finance, Young Firms

[1]Corresponding authors: Juanita Gonzalez-Uribe (j.gonzalez-uribe@lse.ac.uk) and Santiago Reyes (sreyes@IADB.ORG). We thank Vicente Cunat, Dirk Jenter, David McKenzie, Daniel Paravisini, and seminar participants at ESSEC, IADB, LSE, Nova, Oxford, and Universidad de los Andes for comments. We also thank Isabela Echeverry and Esteban Piedrahita at Cali Chamber of Commerce for access to the data and helpful comments. Marcela Torres and Lina Zarama provided excellent research assistance.

The high positive skewness in the growth distribution of firms is a well-documented fact in the economics literature (cf. Birch and Medoff 1994, Henrekson and Johansson 2008). High-growth young firms, typically referred to as "gazelles," disproportionately contribute to growth. For example, more than 20% of job creation in a typical year in the U.S. can be attributed to less than 5% of start-ups (Haltiwanger, Jarmin, and Miranda 2016).[2] In spite of their importance, identifying and spurring these transformational entrepreneurs remains a challenge (Hall and Woodward 2010, Fafchamps and Woodruff 2016, Nanda 2016, McKenzie and Sansone 2017). Start-up failure rates are notoriously high, and many examples exist of renowned investors forgoing ex-post high-value investments.[3]

One increasingly popular method to identify and influence high-growth firms is the business accelerator.[4] An accelerator is a fixed-term, cohort-based, "business school for entrepreneurs." While accelerators have been touted by the popular press as critical to the development of start-up ecosystems, some economists, policymakers, and even business people remain skeptical.[5] One of the main reasons for this skepticism is the belief among academics and many investors that high-potential entrepreneurs are born with skills that schools cannot teach. Instead, the narrative in business accelerators argues that many entrepreneurs have no "entrepreneurial capital" to successfully grow positive net present value opportunities, even if injected with cash. Entrepreneurial capital can be defined as the organizational abilities to manage an effective operations scale-up in emerging ventures (cf. González-Uribe and Leatherbee 2018a), which goes beyond managerial capital for established firms (cf. Bruhn, Karlan, and Schoar 2010) and includes socioemotional skills and access to network of contacts, among other factors. Can accelerators identify high-growth firms? Do firms with high-growth potential that are constrained in their access to entrepreneurial capital exist? Can business accelerators really impact such firms?

---

[2] The term "gazelle" was originally coined by David Birch in the early 90s (see Landström 2005). He defined gazelles as companies that grow 20 percent or more annually for four years, at least doubling their revenue in the process, and that began with at least $100,000 in sales. Since then, several other definitions for gazelles have been used in the literature; see Henrekson and Johansson (2008) for a meta-analysis of the literature.

[3] See, for example: https://www.bvp.com/portfolio/anti-portfolio.

[4] The proliferation of business accelerators is well documented (e.g., Fehder and Hochberg 2014). So is the increasing prevalence of public funds for these programs: an estimated 40% of businesses accelerators receive some form of government support (e.g., Bone, Allen, and Haley 2017).

[5] See, for example: https://singularityhub.com/2018/02/05/are-accelerators-the-secret-to-building-truly-great-startup-hubs/#sm.001bzbtyezzaetj10ql2gcki7b36t

In this paper, we investigate these questions through the context of an ecosystem business accelerator in the central west region of Colombia.[6] "ValleE" was launched in 2015, targeted new and young businesses (0-3 years) with high-growth potential, and attracted almost 300 applicants to the first cohort. Only the top 35 applicants were competitively selected based on average scores from randomly assigned, partially overlapping three-judge panels.[7] Judges evaluated online applications that included easy-to-extract "hard information" (e.g., gender of founders) but also harder to codify "soft information" from detailed business plans. The judges were provided with uniform criteria to independently judge applicants, and the selection cut-off was pre-determined due to budget restrictions. The winners were chosen to receive *entrepreneurship schooling* in the form of business training, access to mentors, and visibility, but *no seed capital*. Three annual follow-up surveys, together with administrative data from the business registry, enable the tracking of the trajectory of impacts 12, 24, and 36 months after participation.

We have three main findings. The first finding shows that judges can identify high-growth firms, but that correcting for judge fixed effects in the selection process would have led to higher quality participants. Namely, we find that future revenue is predicted by adjusted scores that correct for judge fixed effects but not by the unadjusted average scores on which the program based its selection of participants. Adjusted scores have remarkable predictive power: applicants ranking among the top quartile of adjusted scores are 20% more likely to become gazelles (even after controlling for treatment). By gazelles, we mean applicants that were among the top 10% according to sales in 2017, which we show exhibit an annual revenue growth of 68% in the three years following their application.[8] Average scores are not good growth predictors because they do not control for heterogeneity in the *scoring generosity* of judges, which we show is substantial: 34% of judges tend to award scores that are one

[6] Ecosystem accelerators are one of the three types of business accelerators. These programs are generally sponsored by governments, universities, or nonprofits and their aim is to stimulate the entrepreneurship ecosystem rather than profit (see Clarysse, Wright, and Van Hove 2015).

[7] There were 50 judges overall.

[8] To implement this classification, we split the sample into two groups according to age at application: (i) more than 1 year relative to incorporation and (ii) less than 1 year since incorporation or not incorporated. Our definition is similar to that in related work (e.g., McKenzie and Sansone 2017). However, there is no general definition of gazelles in the literature. See Section 2.3 for a discussion of our criteria and a comparison to other definitions of the term.

standard deviation above or below the average score of the other judges, whereas in 500 placebo samples we show only 6.8% do so (cf. Fee, Hadlock, and Pierce 2013). As a result, a third of the participants ended up being selected because they were scored by relatively more generous judges rather than because of their inherent better quality. The judge fixed effects likely reflect heterogeneity in the subjective meaning of scores across judges; for example, in a scale from 0 to 1, a mediocre project might score 0.7 for "generous" judge A but only 0.4 for "ungenerous" judge B. This type of heterogeneity is reminiscent of variation in "leniency" tendencies across judges described in other contexts such as in the dismissal rates of U.S. bankruptcy filings (see Chang and Schoar 2013 and Dobbie and Song 2015).

Using the random allocation of applicants across judges with different scoring generosities as an instrumental variable for acceleration, our second finding identifies average causal impacts of the accelerator on the growth of marginal projects—i.e., applicants whose selection decision is altered by the judge assignment and would (not) have been selected but for the strictness (generosity) of the judges. Acceleration increases annual revenues by $66.31M COP (circa $20K USD)—a 157% increase relative to baseline sales.

Finally, we show that the accelerator's impact increases with the quality of applicants, as measured by their estimated ex-ante propensity score for acceleration (a function of adjusted scores and observed application covariates). Treatment effects are highest for firms in the top quartile of acceleration propensity and positive for all applicants except for "bottom quality" ones (i.e., those that have a propensity for acceleration that is below 0.4). These additional results provide novel, compelling evidence of heterogeneous acceleration treatments and cast doubt on critics' contention that accelerators only teach fledging businesses how to attract funds rather than add value to high-potential ventures. To produce this evidence, we exploit further the random allocation of judges across applicants, and use an approach based on propensity matching and local polynomials following Xie, Brand, and Jann (2012).

The main empirical challenge in this paper is the small sample size, which can affect statistical inference. To address this concern, we complement results from our main empirical analysis with results from local randomization methods (Rosenbaum 2002), which conduct exact finite-sample inference,

and which remain valid even when the number of observations is small. We propose an intuitive adaptation of these methods to our setting, which can also be generalized to other contexts.

We conclude our work by discussing the implications for accelerators, including best practices on selection methods. To illustrate the importance of correcting for heterogeneity in the scoring generosity of judges, we use a back-of-the-envelope calculation of our findings. Our calculation shows that the accelerator's annual revenues would have been 31%-40% higher had the accelerator controlled for judge heterogeneity in the selection process. Beyond our scope are issues about optimal program size, whether access to entrepreneurial capital should be subsidized, and distinguishing which component of schooling exerts the largest influence on entrepreneurial capital (e.g., training, mentoring, visibility).

We contribute to a number of streams in the academic literature. First, our work contributes to the literature on constraints to young firm growth. Most existent work focuses on the role of financial constraints (e.g., McKenzie 2017). Our contribution is to show that constraints to entrepreneurial capital also appear to substantially limit growth in young firms. This finding is consistent with prior work on the importance of managerial capital, and the degree to which it is missing in underdeveloped economies (Bruhn, Karlan, and Schoar 2010, Bloom and Van Reenen 2010, Karlan and Valdivia 2011, and Drexler, Fischer, and Schoar 2014). Our focus on high-growth young firms is novel; existing research instead focuses on established and often nonscalable firms such as local microenterprises or subsistence businesses (cf. McKenzie and Woodruff 2014).

We also contribute to the growing literature on business accelerators (Cohen and Hochberg 2014, Fehder and Hochberg 2014, and González-Uribe and Leatherbee 2018a). Our contribution is threefold. First, we show that expert judges can identify high-growth firms in a way that adds to the predictive power of the hard information in the applications. This first result contributes to the debate on whether and when expert opinion has predictive power (Shanteau 1992, Kahneman and Klein 2009, Fafchamps and Woodruff, 2016; McKenzie and Sansone 2017). Second, we show that business accelerators can impact venture growth even if they provide no cash. Instead, prior work could not isolate the effect of schooling from that of seed capital: Gonzáez-Uribe and Leatherbee, (2018a) show that the provision of schooling services *bundled* with cash is more effective than a cash injection on its

own. Finally, we provide the first estimate for the treatment effect *curve* as a function of applicants' growth potential. In contrast, related work exploiting qualifying thresholds to assess impacts faces challenges in extrapolating local average results beyond restricted subsamples (e.g., González-Uribe and Leatherbee, 2018a; González-Uribe, Hu, and Koudjis 2018). ln terms of external validity, we argue that our results are most relevant for other ecosystem business accelerators that, like ValleE, target high-potential entrepreneurs, attract incorporated businesses with already some traction in the market, and have access to high-quality staff, mentors, and judges.

Finally, our work contributes to the literature on empirical methods by proposing a new methodology to estimate causal impacts when selection is based on the random allocation of applicants across partially overlapping judge panels that differ in their scoring generosity. The instrumental variables strategy we use builds upon prior work exploiting the leniency of individual judges to study the impacts of incarceration (Kling 2006, Aizer and Doyle 2013; Di Tella and Schargrodsky 2013; and Dobbie, Goldin, and Yang 2016), foster care (Doyle 2007, 2008), corporate bankruptcy (Chang and Schoar 2008), temporary employment (Autor and Houseman 2010), and disability insurance (French and Song 2011; Maestas, Mullen, and Strand 2013). Two important distinguishing features from our setting are that applicants are evaluated by a panel of judges rather than by one single judge and that evaluation outcomes are scores rather than dichotomous variables. These setting features allow us to characterize judge heterogeneity in more detail, similar to prior work using fixed effects models in panel data to assess the importance of managers in corporations (cf. Bertrand and Schoar 2003; Fee, Hadlock, and Pierce 2013) and general partners in limited partnerships (Ewens and Rhodes-Kropf 2015). It also allows to control for unobservable heterogeneity across applicants in our impact estimations.

The rest of this paper proceeds as follows. In Section 1, we describe the context and data. In Section 2, we investigate the ability of judges to identify high-growth firms. We detail the empirical strategies to identify the average impact and the impact curve as a function of applicants' potential in Section 3 and present the main results. We present the randomization inference results as a robustness check in Section 4. Section 5 estimates the revenue costs to the accelerator from failing to correct for judge heterogeneity and discusses the external validity of findings. Section 6 concludes the paper.

1. Institutional Setting

Our focus is on ValleE, a local ecosystem business accelerator launched during 2015 after an intense local advertising campaign using social media and radio in the city of Cali, the third most important city in Colombia in terms of population.[9] The accelerator is the brainchild of the Regional Network of Entrepreneurship in ValleE del Cauca (a private organization that aims to encourage entrepreneurship in the ValleE del Cauca region), and is operated by the city Chamber of Commerce, a private entity that has been delegated public duties such as the management of the Colombia business registry.[10] As is common among ecosystem accelerators, ValleE's main objective is to encourage local growth through the expansion of existing high-potential businesses and the creation of new transformational ventures (cf. Clarysse, Wright, and Van Hove 2015).[11]

Like other business accelerators worldwide, ValleE is a fixed-term, cohort-based program that selects participants based on the relative quality of applications submitted online, as evaluated by a panel of judges (cf. Leatherbee and González-Uribe 2018b). As explained in more detail in Section 1.2, participants are selected based on *average scores* from partially overlapping 3-judge panels in order to satisfy pre-determined budget and space restrictions, as well as judges' time constraints. Any person proposing the creation of a new business or the scale of an existing young (0-3 years) business located in the region is in principle eligible for the program. However, the program focuses on transformational entrepreneurs, and many applicants are de facto incompatible and thus rejected (as explained in more detail in Section 3).

Also similar to traditional business accelerators, ValleE provides participants with entrepreneurship schooling (which we describe in more detail below). The distinguishing feature in our setting, however, is that the program offers no cash (as is nevertheless common among the subset of

[9] Ecosystem business accelerators are popular in low and lower-middle income countries: 37.9% of the ecosystem accelerators in the Entrepreneurship Database at Emory University are located in Africa (17.9%), Latin America (10.3%), and India (10.3%).

[10] Chambers of Commerce oversee the private sector development policies in their region. They are key connecting actors that execute programs aimed at improving regional competitiveness.

[11] The top two impact objectives among ecosystem accelerators are employment generation (35%) and community development (30%). Source: The Entrepreneurship Database at Emory University.

ecosystem accelerators worldwide, cf. Clarysse, Wright, and Van Hove 2015).[12] The perception is that for many young businesses the foremost constraint to growth is their lack of entrepreneurial capital, which contrasts with the usual academic narrative. The starting point of most entrepreneurial finance models are entrepreneurs with access to and ability to execute positive NPV opportunities that cannot convince investors to give them funds for a myriad of potential reasons (unrelated to the notion of entrepreneurial capital) such as investors' fear of moral hazard or adverse selection. The narrative in ecosystem business accelerators questions this starting point and argues that in fact many entrepreneurs with access to positive NPV opportunities have no actual ability to successfully execute those opportunities. As a consequence, the businesses of these entrepreneurs would not grow even if they received cash injections.

Like traditional business accelerators, the entrepreneurship schooling offered by ValleE to participants has several aspects (with varying degrees of structure), including instruction via bootcamps, the provision of advice and potential contacts via mentoring, and increased visibility.

Bootcamps are highly structured and simultaneously attended by all participants in the offices of the Chamber of Commerce. They consist of roughly 8 weekly hours of organized content (100 hours overall in a space of three months) delivered by local and national experts that are hired by the Chamber of Commerce. Bootcamps combine lecture-based conceptual sessions together with case-based sessions discussing real-life practical examples, and cover the topics of business modelling, early-stage financing, market validation, prototyping, accounting, and pitching. Two types of mentoring are provided. The first type consists of bi-monthly meetings to discuss business strategy with high-level advisors assigned based on industry, which include renowned CEOs in the region as well as managers at the Chamber of Commerce. Assigned advisors may provide introductions to potential clients or industry contacts, which are likely to be high-impact, as the selected CEOs and Chamber of Commerce managers are well connected within the local ecosystem. The second type of mentoring sessions are handled by program coordinators who take a more hands-on approach: sessions are conducted weekly and are of varying duration. Coordinators are junior to advisors and focus on helping entrepreneurs

[12] Circa 55% of the ecosystem accelerators in the Entrepreneurship Database at Emory University provide no seed capital. Source: https://www.galidata.org/accelerators/.

throughout the day-to-day operations rather than designing avenues for growth. Finally, ValleE provides several opportunities to increase visibility: participants are showcased on the Chamber of Commerce website, in their monthly publications, and exhibited at different events. At the end of their term, participating businesses "graduate" through a "demo day" competition (i.e., a formal presentation of the companies to potential investors). Other potential benefits parallel the benefits of traditional business schools, such as self-validation from selection, certification, and peer effects, among others (see González-Uribe and Leatherbee 2018a and Lerner and Malmendier 2013).

1.1. Sample

ValleE provided us with all the application data, including application scores by each judge and final selection decisions, for the program's first cohort.[13] All selected applicants in this cohort participated in the accelerator for three months, during May, June, and July of 2015. Our sample consists of 135 projects (35 participants and 100 nonparticipants) that applied to the accelerator in March of 2015 and were deemed to have high-growth potential by the staff as explained in Section 2.

Our sample size exceeds that of similar papers exploring the impact of business training on firms (e.g., e.g., Bloom et al., 2013; Mano et al. 2012), but is small enough for concerns to be raised, for instance regarding our ability to detect the impact of the accelerator if such an effect exists. We return to this issue in Section 4 below, where we conduct exact finite-sample inference using a randomization inference approach.

Based on the program's records, we constructed several hard information variables to use as controls in our empirical strategy that are easy to extract from applications: the age of the firm (*Firm Age*); the founder's sector experience in years (*Experience*); indicator variables for male applicants, serial entrepreneurs, and projects with founding teams (*Male*, *Serial*, *Team*); projects located in Cali (*Cali*); founder's education (*High school*, *Technical degree*, *College*, *Graduate*); founder's motivation

[13] The judges' identities were not provided by ValleE for confidentiality reasons. For the purpose of our investigation, we were provided with anonymized information that includes judge identifiers in order to track different projects evaluated by the same judge.

to start a business (*Stable income, Own boss, Opportunity)*; and industry and location indicators. Baseline information on *Revenues, Profits,* and *Employees* are also included.

Table 1 reports summary statistics for the main variables in the application forms. On average, applicants have 5.6 years in sector experience, are male (79%), educated (67% have a bachelor's or master's degree), are likely to be serial entrepreneurs (61%), and have an entrepreneurial team (88%). The average number of founders is 3, and the average number of employees is 4. The likelihood of positive revenues is 45%, and median (positive) annual sales are $52M COP—aprox. $13,000 USD. Most applicants are in the services industry (56%), have participated in other entrepreneurship contests (59%), and applied with business ideas (53%) rather than already established firms (47%). Applicants classified as having "business ideas" include informal businesses that at the time of application were trading but had not been incorporated in the business registry of Colombia. Accordingly, the average revenues and employment at baseline for these businesses were positive but modest ($4.61M COP [$1,000 USD] and 2.7, respectively; see Table 1).

Our sample is comparable to that in prior work on early stage ventures. Haltiwanger, Jarmin, and Miranda (2013) document that 33% of young firms (less than a year old) in the U.S. have between one and four employees, and Puri and Zarutskie (2012) show that the distribution of VC-backed firms is concentrated in the services industry. Our sample is also comparable to that used in prior work on ecosystem business accelerators: González-Uribe and Leatherbee (2018a) show that applicants to Start-up Chile, a renowned ecosystem accelerator sponsored by the Chilean government, are likely to be male (86%), have between two and three employees, and are predominantly from services industries such as E-commerce (18%). Relative to average applicants to ecosystem accelerators worldwide (based on information from the Entrepreneurship Database (ED) program at Emory University), the average applicant to ValleE is more educated (47% of ED applicants have a bachelor's or master's degree), less likely to be female (29% of ED applicants are female), and has a more mature business (19% of ED applicants report positive revenues prior to application) that is similarly sized (ED applicants have an

average of 3.5 employees, a 43.2% likelihood of positive profits, and median (positive) sales of $12,000 USD).[14]

1.2. Accelerator Selection Process

Selection into ValleE is a four-part process. First, aspiring participants submit an online application that requests information about the entrepreneurs and their detailed business plans. Next, the accelerator filters applicants to exclude projects that are deemed to have no high-growth potential (e.g., taxi drivers, shopkeepers). Filtered applications are then randomly assigned to a panel of three judges. The total number of judges is generally larger than three (e.g., there were 50 judges in the first cohort), and thus panels only partially overlap across applicants. The judges evaluate the applications according to five criteria: (i) clarity of the business model proposal, (ii) innovation, (iii) scalability, (iv) potential profitability, and (v) entrepreneurial team. The main reasons behind using judge panels (rather than individual judges) are to minimize the burden on individual judges (given their time constraints), and mitigate the chance that one judge determines the treatment status of any given project, as this could lead to unwanted biases such as judges favoring projects from their own industries, regions, or communities. Finally, the staff at the accelerator makes the final decision by picking the top 35 ranking applicants based on the average score of the judge panels. The capacity threshold of 35 participants is due to budget and space limits, which were determined prior to the launch of the program.

Several features of the selection process make it unlikely that the judges can precisely manipulate the rankings (for instance, to help an applicant friend qualify). While judges are provided with standard written guidelines for the evaluation (see Appendix 1), they are unaware of the weight of each of the five components in the final score. In addition, the three judges in each evaluation panel independently score projects, and are not aware of the identity of the other judges in the panel; no judge sees all applications. Importantly, applicants do not know who their judges are, nor do they know their position in the rankings; thus it is impossible for applicants to manipulate the ranking process.[15]

[14] See https://www.galidata.org/accelerators/.
[15] Entrepreneurs were never given their ranking or scores in order to avoid any negative psychological effects or create rivalry among participants.

In the first cohort of ValleE (our sample source), there were 255 applicants who submitted a complete application online. Of these, only 135 businesses were deemed to have "high potential for growth" and therefore correspond to our analysis sample.[16] The maximum length allowed for business plans submitted with the applications and read by the judges was 5 pages. The pool of 50 judges included independent business consultants, industry experts, and two staff members. The average number of projects scored by any given judge was 8, and the minimum (maximum) was 5 (14). Following the selection rule of the program, the top 35 ranking applicants based on judges' average scores were selected in the first cohort, and all selected applicants participated (see Figure 1, Panel A). Table 2 shows statistically significant differences at application between accelerated and nonaccelerated applicants: participants have bigger founding teams, which are slightly more educated, have more sectorial experience, and are more likely to be serial entrepreneurs. The economic significance of most of these differences is, however, small, in part due to the filter applied by the program to remove the non-transformational entrepreneurs from the sample.

1.3. Post-Application Performance

In this section, we explain how we collected information on applicant revenues—our main outcome measure. Table 3 presents summary statistics.

We use two complementary strategies to address the numerous challenges faced by researchers when measuring accelerator applicants' performance (cf., Leatherbee and González-Uribe 2018b). First, we partnered with ValleE and collectively designed a performance survey that was sent by the program to all applicants every year for three years after application to the program. The yearly surveys included questions regarding revenue; our main objective was to measure improvements relative to the revenues reported at the time of application. They also included questions on other firm outcomes that we used to run robustness checks (i.e., employment and profits; see Appendix 7).

[16] The characteristics of the final 135 projects differ slightly from the 120 businesses removed by the initial filter, which were more likely to have a female founder, have less educated founders, and refer to nonpecuniary benefits (e.g., being their own boss) as the main motivation behind their business.

Survey response rates were 77%, 67%, and 60%, respectively, in years 2016, 2017, and 2018; with participants having slightly higher rates (82%, 77%, and 65%) than nonparticipants (75%, 64%, and 58%). These annual survey response rates were much higher than that found in prior work on business accelerators (e.g., 10% in González-Uribe and Leatherbee 2018a), and imply similar survey attrition rates to those in related papers exploring the effect of business training in microenterprises and small- and medium-sized firms (e.g., 25% in Karlan and Valdivia 2011; 26% in Calderon, Cunha, and de Giorgi 2013; 28% in Klinger and Schundeln, 2011). They also allow us to measure the trajectories of outcomes over a longer period, which is important because the short- and long-term impacts of many policies can differ substantially (cf. King and Berhman 2009 and de Mel, McKenzie, and Woodruff, 2013). This increased period of measurement improves upon existing work, as the median number of follow-up surveys in papers evaluating the impact of business training interventions is one (McKenzie and Woodruff 2014).

To mitigate standard survey-related concerns (e.g., selective survey attrition), our second strategy collects information from the business registry in Colombia, which is managed by Chambers of Commerce. One unique feature of this business registry is that it records annual renewals of firms' operating licenses, in addition to the standard initial business registrations. The renewal of an operating license for firms in Colombia is mandatory *de jure* and *de facto*, as companies are required to submit this license to validate their operations with their banks and corporate clients, among others. Using these data, we discern which applicant ideas eventually turned into actual businesses (i.e., start-up rates) from those that did not launch. We also distinguish which applicant-established firms continued operating (i.e., survival rates) from those that instead did not renew their operating license (i.e., closure rates). On average across the survey years, most (74%) survey attrition can be explained by real business closures rather than the refusal of ongoing businesses to answer the survey questions. The refusal rate is no different between participants and nonparticipants, which suggests that refusal is not endogenous to participation (e.g., non-respondents are busy firms, rather than mistrustful nonparticipants) and helps

mitigate concerns regarding the impacts of the program on quality of outcome data besides the potential effect of acceleration (cf. McKenzie and Woodruff 2014).[17]

A second unique feature of the Colombian business registry is that firms have to provide revenue information as part of their registration and license renewal processes with the Chambers of Commerce. We use these data in two ways. First, we validate self-reported revenues. We find little discrepancy between self-reported and registry-based revenues with a 95% coincidence rate (and with 60% of the discrepancies being due to mistakes—e.g., missing or extra zeros). Moreover, discrepancies do not vary across participants and nonparticipants, which further mitigates concerns of data quality from the survey variables, and which lends credence to the robustness checks we ran using other survey data that we cannot validate with registry information (see Appendix 7). Second, we adjust the revenue outcome variable by completing missing observations for ongoing businesses that did not answer the survey, and by replacing with zeros nonresponses of closed firms (as measured by license expiration). This constructed revenue metric helps further address the remaining concerns with the survey data (e.g., potential survey and survivorship bias).

2. Identifying Potential Gazelles

In this section, we start by showing that average scores are not correlated with high growth in a way that adds to the predictive power of the hard information in application responses. We then demonstrate that the lack of predictability is associated with the failure of average scores to account for judge heterogeneity in scoring generosity—i.e., the tendency of some judges to systematically award high scores to projects. Finally, we introduce a methodology to correct for judge heterogeneity that dramatically increases the predictive power of judges' scores and can be easily adapted by other programs that select participants based on partially overlapping judge panels.

2.1 Average Scores and Performance

[17] P-values for differences in response rates between treatment and control groups are 0.352, 0.167, and 0.427 for follow-up surveys 1 to 3 years after treatment.

We regress projects' revenues against application scores to analyze the ability of judges to identify high-growth firms. We estimate variants of the following regression using observations during the 2013-2017 period for all projects that applied to the accelerator:

$$(1)\quad k_{it} = \alpha + \beta AverageScore_i + \rho Average\ Score_i \times After_t + \theta Z_{it} + \mu_t + \varepsilon_{it}\ ,$$

where k_{it} is revenue for project *i* at time *t*, $After_t$ is a dummy that equals 1 during 2015-2017, Z_{it} is a vector of controls, and $Average\ Score_i$ is the *normalized* project's average score that varies between 0 and 1, corresponding, respectively, to the "bottom score" project (scored at 1 by all judges) and the "top score" project (scored at 5 by all judges). We include several controls at the founder and project levels (i.e., the hard information described in see Section 1.1) as well as interactions of these characteristics with the variable $After_t$. Finally, we also include time (μ_t) fixed effects. The main coefficient of interest is ρ, which measures the average additional level of annual revenue within three years of potential participation in the accelerator for the top scoring project relative to the bottom scoring project, over and above secular growth in revenue that is projected by the controls. Table 4 summarizes the results.

The results in Panel A of Table 4 show that average scores do not predict entrepreneurs with high growth in a way that adds to the predictive power of the hard information in applications. Column 2 shows that average scores enter negatively (although not significantly) in the growth regression, and Column 4 shows an estimate of ρ that, while positive, is not significant. Further, a comparison between Columns 1 and 2 shows that including average scores in growth regressions does not improve R^2. The same conclusion is visible in Panel B of Table 4, which reports the Shapley-Owen decomposition of the R^2 across different groups of explanatory variables for selected columns of Table 4. Column 2 in the panel shows that average scores explain only 1.2% of the R^2, which is much lower than the 16.9% contribution of entrepreneurs' characteristics.

2.2. Judge Fixed Effects

The apparent lack of predictability of average scores may be an artefact of the small sample size. An alternative explanation is that judges' heterogeneity in scoring generosity distorts the predictive

power of average scores. By scoring generosity we mean the tendency of some judges to systematically award high scores. Heterogeneity in scoring generosity can exist even though the accelerator provides uniform criteria for scoring across judges, because the subjective meaning of scores can differ across individuals.[18] That is, on a scale of 0 to 1, a mediocre project might receive a score of 0.7 from "generous" judge A, but a score 0.4 from "ungenerous" judge B, even though both judges would still agree to rank the project last among a group of better projects. This issue can arise even if judges have the ability to individually correctly order applicants (i.e., assign higher scores for the projects with more potential), because the final ranking does not reflect the individual orderings but is rather a function of the individual scores. Finally, this concern can hold true even though judges are randomly assigned, because judge panels only partially overlap and are very small (only three individuals); hence, for a given project the scores of overly generous judges may not tend to cancel out those of overly strict judges, as the probability that both types of judges will be randomly assigned to the same project does not tend to 1.

To explore whether judge heterogeneity in scoring generosity explains the poor predictive power of average scores, we begin by decomposing individual scores into company and judge fixed effects using the following regression:

$$(2)\ \ Score_{ij} = \alpha_i + \gamma_j + \varphi_{ij}\ ,$$

where α_i are project fixed effects and γ_j are judge fixed effects. Each judge fixed effect is estimated with at least 5 observations, as the minimum number of projects evaluated by any judge is 5 (see Section 1.2). These judge fixed effects are meant to capture heterogeneity across judges in their scoring generosity. By contrast, the project fixed effects can be understood as the quality of the project that all judges agree on. In what follows, we refer to the project fixed effects as adjusted scores. We plot results in Figures 2 and 3, rather than report regression estimates, to ease exposition. Figure 2 plots the adjusted score against the average score. Figure 3 plots the distribution of fixed effects across judges.

[18] Subjective meaning can vary because judges have different frames of reference—i.e., sets of criteria of stated values in relation to which measurements of judgments are made. For example, some judges may have higher points of reference with regards to the quality of projects that are based on their own experience as entrepreneurs, risk preferences, or general outlook on life (e.g., optimistic or negative).

There are three main findings from estimating equation (2). First, judge heterogeneity in scoring generosity is statistically significant: the *F*-test on the joint significance of the judge fixed effects is 5.49 (p value of 0.00). By contrast, if judge heterogeneity was irrelevant (or nonsystematic), then judge fixed effects would not be jointly significant (as judges are randomly assigned by design). One caveat regards the validity of *F*-tests in the presence of high serial correlation (Wooldridge 2002).[19] To address this concern, we scramble the data 500 times, each time randomly assigning judges' scores to different applicants while holding constant the number of projects evaluated by each judge and making sure that each project receives three scores, in the same spirit as in Fee, Hadlock, and Pierce (2013). Then we proceed to estimate the "scrambled" projects' and judges' fixed effects and test the joint significance of the latter in each scrambled sample. The distribution of the scrambled *F*-tests is plotted in Figure 4 (Panel A). Lending credence to the statistically significant judge heterogeneity in our setting, we reject the null of "no joint significance of the judge fixed effects" in only 3.99% of the placebo assignments (the largest estimated placebo *F*-test is 1.84).

Second, judge heterogeneity in scoring generosity is also *economically* significant. Figure 3 shows that 34% of judges tend to systematically award individual scores that are one standard deviation above or below the average score of the other judges. In contrast, only 6.8% of judges systematically award overly high or low scores in the 500 placebo assignments (Panel B in Figure 4; Fee, Hadlock, and Pierce 2013). The most generous (strict) judge adds (subtracts) an average of 0.26 (0.28) to any given project she scores (relative to a mean normalized score of 0.7). Put differently, the most generous (strict) judge favors (penalizes) applications at 39% (42%) of the median adjusted score (0.66). This economic significance is also evident in Figure 2, which shows that a projection of average score on adjusted scores does not result in a 45-degree line, as would be expected if judge fixed effects were irrelevant. While the correlation between these two scores is high at 0.83, the figure shows that some projects were "lucky" in that they randomly drew a generous judge panel that assigned an average score

[19] In the parallel literature, when seeking to identify the "style" of managers using an endogenous assignment of (movers) managers to multiple companies (e.g., Bertrand and Schoar 2003), concerns have been raised regarding the validity of *F*-tests in the latter settings on the grounds of (a) the particularly acute endogeneity in samples of job movers and (b) the high level of serial correlation in most of the variables of interest (see Fee, Hadlock, and Pierce 2013). The first reason for concern is not at play in our setting, as judges are randomly assigned by design, but the second concern may still apply.

substantially above the firm fixed effect. Some projects instead appear "unlucky" in that they randomly drew a strict judge panel that assigned an average score substantially below the firm fixed effect. Relying on a panel of judges rather than on individual judges helps mitigate the effect of judge heterogeneity but does not fully correct it, because judge panels are small (see Appendix 2).

The third result is that judge fixed effects do not appear to be driven by outliers or capturing noise. We see very small differences across the different "leave one out" estimates of judge fixed effects; as Figure 5 shows, the average standard deviation per judge is 0.003, and the maximum is 0.006. Figure 5 plots the distribution of the standard deviation of the estimated leave one out judge fixed effects by judge (with each point representing one judge). For a given project *i*, the leave one out approach uses all observations except project *i* to estimate the judge fixed effects. A maximum (minimum) of 14 (5) different leave one out judge fixed effects are estimated for the judge that evaluated the most (least) applicants. We also construct leave one out adjusted scores by subtracting the leave one out judge fixed effects from the average score, which we use in Section 2 (as is standard in the parallel literature exploiting heterogeneity in judge leniency). The correlation between the adjusted score constructed in this way and the project fixed effects is very high at 0.98.

2.3 Adjusted Scores and Future Performance

Having shown compelling evidence of significant judge heterogeneity, we now turn to exploring whether this heterogeneity can help explain the low predictive power of average scores. To that end, we estimate versions of equation (1) where we replace the average score with the adjusted score.

Column 5 of Panel A in Table 4 shows that adjusted scores, in contrast to average scores, are highly positively correlated with future performance; the top adjusted score project has an additional 86 million COP ($20K USD) in annual revenue relative to the bottom adjusted score project and has over and above secular growth in revenue projected by the controls (compared to a baseline level of revenues in 2014 of 25 million COP; Table 3). The correlation is strongest for business ideas (see Appendix 3), and is not explained by a potential treatment effect, but rather reflects the predictive power of judges: Column 6 (Panel A, Table 4) shows that the correlation continues to hold if we add as a

control an indicator variable for participants.[20] Moreover, the Shapley-Owen decomposition (Column 3, Panel B, Table 4) shows that adjusted scores contribute substantially to the model—four times more than average scores (1.2%, Column 2, Panel B, Table 4)—and represent almost a third of the contribution of the entire set of entrepreneurs' characteristics (15.4%, Column 3, Panel B, Table 4).

We explore further the predictive power of judges to try and get at the question we ask in the title of this paper. Table 5 shows evidence that adjusted scores can predict high growth; applicants among the top quartile of adjusted scores are 20% (Column 1, Panel A) more likely to become "gazelles" by the end of the sample period. The result holds when we control for several covariates (Columns 2 and 3, Panel A) and for acceleration (Column 4, Panel A).

For this additional exercise, we classify applicants as "gazelles" if they are among the top 10% of largest firms by 2017. There is no general definition of gazelles; Henrekson and Johansson (2008) show a large variation in definitions in their meta-analysis of the empirical literature. Our classification follows other papers in the gazelles literature that define gazelles using performance thresholds (Kirchhoff 1994, Picot and Dupuy 1998, Schreyer 2000, Fritsch and Weyh 2006, McKenzie and Sanson 2017). Our threshold is based on size (revenues in 2017) to avoid the difficulties of measuring growth rates in our sample: a large fraction of our applicants have zero sales at baseline (57%; see Table 1). To implement this classification, we split the sample into two groups according to age at application: (i) more than 1 year relative to incorporation (77 applicants) and (ii) less than 1 year since incorporation or not incorporated (58 applicants). We then define as gazelles the top 10% of firms in each group according to sales in 2017 (9 and 7 applicants in each group, respectively; see Appendix 4). To produce these results, we estimate a probit model of an indicator variable for gazelles against an indicator of quality as measured by firms in the top quartile of adjusted scores.

Panel B in Table 5 shows that the applicants we identify as gazelles exhibit remarkable average growth rates. These gazelles grow their revenue by 376% relative to baseline sales, which contrasts with the negative average contraction rate of 19% for nongazelles over the sample period. The implied

[20] This is as expected: if the correlation was fully explained by treatment, then average scores should be more predictive of future performance than adjusted scores—after all, participation in the accelerator is defined by the average, rather than, the adjusted score.

annual growth rates in the three years following their application for these high-growth firms is 68.24% on average, which exceeds the growth rate requirements for gazelles in most other definitions used in the literature (e.g., 20% for Birch and Medoff 1994 and Birch et al. 1995; 50% for Ahmad and Petersen 2007; Deschryvere 2008, and Autio et al. 2000).

Overall, these results show that adjusted scores predict performance in a way that adds to the predictive power of the hard information in applications responses. Judges have the ability to extract soft information from the business plans they evaluated, which is useful in predicting venture growth. Results in these tables also show that judges' scores, however, are not exhaustive: the hard information from the applications has important predictive power, as shown in Panel B of Table 4 (a point that we come back to in Section 5).

The main implication for the accelerator is that the program could have picked better participants had it controlled for judge heterogeneity in scoring generosity. Panel B in Figure 1 illustrates this point by showing that, with respect to adjusted scores, the program selected applicants substantially below the top 35 target. Table 6 provides supportive evidence of potential selection improvements by showing that "marginal participants" had lower revenues in 2013, founders with less entrepreneurial experience, and smaller founding teams than "nonmarginal participants." Marginal participants correspond to the 12 participants that rank among the top 35 projects according to average scores but not according to adjusted scores. Nonmarginal participants are the 23 participants that rank in the top 35 according to both average and adjusted scores.

It is possible that the accelerator could have increased the *impact* of its program as well had it avoided these mistakes in selection. This would hold true if rejected applicants with high-growth potential benefited more from the accelerator services than the marginal projects accepted by mistake. We return to this question in Sections 3.5 and 5, where we explore heterogeneity of impact across applicants with different growth potential, and quantify the revenue costs of selection mistakes, respectively.

3. Estimating the Impact of the Business Accelerator on Venture Growth

In this section, we exploit the random allocation of projects across judges with different scoring generosities to show compelling evidence of causal and heterogeneous impacts of the accelerator on project growth. We begin by showing that scoring generosity strongly predicts selection into the accelerator. We then explain how we exploit this exogenous predictor of acceleration using an instrumental variables (IV) approach. Finally, we explore potential heterogeneous treatment effects by adapting the methodology of Xie, Brand, and Jann (2012) to our setting.

3.1 Scoring Generosity and Selection Probability

To show that scoring generosity strongly predicts selection into the accelerator, we begin by classifying projects into quartiles of scoring generosity based on the sum of the fixed effects of the three judges that evaluated each project (see Appendix 2). Relative to a mean average score of 0.7, the breakpoints for the scoring generosity quartiles are -0.03, 0.001, and 0.05, and the max (min) scoring generosity is 0.21(-0.13).[21] After classifying applicants in this manner, we then show a positive relation between acceleration probabilities and scoring generosity by estimating acceleration rates across quartiles. Results are reported in Table 7.

Two main facts about the scoring generosity and acceleration are visible in the table. First, scoring generosity significantly correlates with acceleration probability; on average, moving from the bottom to the top quartile of scoring generosity roughly doubles participation rates (17.64% relative to 36.36% relative; see column 1 in Table 7). Because judges were randomly assigned, these differences in acceleration are unlikely to be explained by project heterogeneity across quartiles of scoring generosity. We provide supportive evidence in Figure 6 where we plot acceleration probabilities against adjusted scores across the quartiles of scoring generosity. The table shows substantial gaps between the probability curves across the different quartiles, with the probability curve of the top (bottom) quartile shifted farthest to the left (right) in the figure. The plot demonstrates that, holding constant the quality of projects (as measured by adjusted score), the probability of acceleration is always highest (lowest) for projects assigned to the top (bottom) quartile of scoring generosity. To illustrate this, if we focus on

[21] These numbers imply that projects classified in the top (bottom) quartile of judge generosity received between 0.05 and 0.21 (0.13 and 0.003) additional (fewer) points than their project fixed effects.

projects of median quality (adjusted score equal to 0.66), the figure shows that the probability of selection increases from 0.01% to 78.96% when we move from projects in the bottom to the top quartile of scoring generosity (see column 3 in Table 7).

The second fact is that the apparent impact of scoring generosity on acceleration is not symmetric for high and low project qualities, as measured by the adjusted score. Column 2 in Table 7 shows that the probability of participation for projects in the bottom quality quartile (i.e., adjusted score below 0.59) is very low, even for projects that were assigned to the top quartile of scoring generosity (5.31%). Instead, Column 5 in the table shows that projects at the 90th percentile of quality are only 26.89% likely to be selected if they were allocated to the bottom quartile of scoring generosity, relative to 99.99% for those allocated to top quartile. The impact is also nonlinear: For projects in the 75th percentile of quality, moving from the bottom to the second lowest quartile of scoring generosity increases the participation probability by 27.50%, whereas moving from the top to the second highest quartile of scoring generosity decreases the participation rate by 13.87%.

3.2 Exploiting Scoring Generosity as an Exogenous Predictor of Acceleration

We estimate the casual impact of acceleration through a two-stage least squares (2SLS) regression using quartiles of scoring generosity as an instrumental variable for acceleration. The second stage estimating equation is:

$$(3)\ k_{it} = \alpha + \vartheta_i + \rho Acceleration_i \times After_t + \theta X_{it} \times After_t + \mu_t + \varepsilon_{it},$$

where ϑ_i are project fixed effects, μ_t are time fixed effects, and X_{it} includes several controls (the hard information from applications; see Section 1.1), which are interacted with $After_t$, as the main effect of the time-invariant controls is absorbed by the ϑ_i. We also include the interaction between the adjusted score and $After_t$ to control for differential trends across different quality projects. The first-stage estimating equation associated with equation (3) is:

$$(4)\quad Acceleration_i \times After_t = \alpha + \vartheta_i + \beta \sum_{i=2,3,4} Quartile_i \times After_t + \theta X_{it} \times After_t + \mu_t + \epsilon_{it},$$

where $Quartile_i$ is a dummy indicating the ith quartile of scoring generosity (the left out category is the bottom quartile). We present results using bootstrap standard errors clustered at the applicant level to account for any serial correlation across applicants, and for the fact that the adjusted score is a generated regressor (Wooldridge, 2002; Young, 2018).

Using the quartiles of scoring generosity (interacted with $After_t$) to instrument for acceleration yields a consistent two-stage least squares estimate of ρ as the number of applicants grows to infinity, but is potentially biased in finite samples. This bias is the result of the mechanical correlations between an applicant's own outcomes and the estimation of that applicant's judge fixed effects. Following the parallel literature exploiting judge leniency (Kling 2006, and related papers thereafter), we address the own observation problem by using the leave one out measures of judge scoring generosity introduced in Section 2 to build our instruments for acceleration. In unreported results, we verify that results are similar using the raw measure.[22]

The ρ estimate measures the local average treatment effect of the accelerator for applicants whose participation is altered by judges' scoring generosity. Three conditions must hold to interpret these estimates as the average (local) causal impact of acceleration: (1) scoring generosity is associated with participation in the accelerator, (2) scoring generosity only impacts venture outcomes through the probability of participating in the accelerator (i.e., the "exclusion restriction"), and (3) the impact of scoring generosity on the probability of acceleration is monotonic across applicants.

We present further evidence in support of the first assumption in Table 8 by regressing the probability of acceleration against quartiles of scoring generosity and the adjusted score and including application controls. The results show that for a given adjusted score applicants in the top quartile of scoring generosity are 49% more likely to be accelerated than applicants in the bottom quartile of scoring generosity (Column 1). The results are similar across applicant business ideas and applicant established firms (Columns 2 and 3, respectively). Table 8 also formally tests the relevance of the instrument and reports the F-test of joint significance of the quartiles of scoring generosity in the first stage (i.e., equation 4), showing that the instruments are not weak (Stock and Yogo 2005).

[22] Results are available upon request; they are not reported to conserve space.

Regarding the exclusion restriction, we argue that it is likely to hold for a number of reasons. The most natural concern of favoritism (that firms with higher growth potential were assigned the most generous teams of judges) can be ruled out by design, as judges were randomly allocated. Any remaining concerns regarding the unintentional assignment of generous judges to high quality firms are not consistent with the patterns shown in Figure 2—i.e. projects with high adjusted scores do not systematically have higher average scores than expected. These concerns are also not consistent with the fact that observable characteristics are similar across applicants assigned to judge panels with low and high scoring generosity (see Appendix 5). Differences in the interaction between applicants and judges across applicants in different quartiles of scoring generosity are unlikely because only two of the 50 judges are ValleE staff members, the rest of the judges do not interact with participants as part of the program, and the judges' identities are not revealed to applicants throughout the process. Finally, because applicants are not made aware of their scores, nor how generous their panel, psychological reactions are also unlikely (e.g., feelings of grandeur or depression). Ultimately, however, the assumption that scoring generosity only systematically affects applicants' performance through acceleration is fundamentally untestable, and our estimates should be interpreted with this identification assumption in mind.

Finally, we present evidence in support of the monotonicity assumption in Appendix 6. The monotonicity assumption implies that being assigned to a more (less) generous panel of judges does not decrease (increase) the likelihood of selection into the accelerator depending on the projects' characteristics. The monotonicity assumption would be violated if judges differ in the types of applications they grade more generously. For example, the monotonicity assumption could be invalidated if some judges score business ideas more generously than established firms. We provide supportive evidence of the monotonicity assumption in Appendix 6, where we plot scoring generosity measures that are calculated separately for four restricted subsamples: i) using only business ideas, ii) using only established firms, iii) excluding the bottom quartile projects according to adjusted score, and iv) excluding the top quartile according to the same metric. Consistent with the monotonicity assumption, we find that judges exhibit remarkably similar scoring generosity tendencies across

observably different applicants. The plots show a strong correlation between the actual fixed effects and the fixed effects from the restricted samples.

3.3. Local Average Impact Results

Table 9 shows compelling evidence of causal impacts of acceleration. Within three years of applying, there are annual increases in revenue of 66M COP ($20K USD) relative to a 35M COP ($9K USD) baseline, as shown in Column 3. In robustness checks, we show that the results are similar if we use alternative survey-based metrics (see Appendix 7).

The IV estimates the local average treatment effect (cf., Imbens and Angrist, 1994) for the projects at the margin of selection—i.e., applicants whose selection decision was altered by the judge assignment (and would [not] have been selected but for the strictness [generosity] of the judges). Complementary analysis suggests that the local average effect is not driven by a few outliers: Figure 7 shows a shift in the sales distribution 3 years after application to the accelerator (2017 versus 2014), for projects at the top-quartile of scoring generosity that is not evident for projects in the bottom-quartile of scoring generosity.

We contrast the IV estimate with the naïve OLS estimate of equation (4) that compares average performance across participants and nonparticipants. A comparison between columns 2 and 3 in Table 9 reveals that a positive difference exists between the IV and the OLS estimates (66.31 versus 42.91). One potential explanation for this positive difference is that the projects at the margin of acceptance are most sensitive to the accelerator (cf. Card, 2001). We explore this potential explanation further in the next section, by adapting the methodology of Xie, Brand, and Jann (2012) to assess the evidence of heterogeneous treatments in our setting.

A comparison between columns 6 and 9 in Table 9 reveals that the increase in revenue for marginal applicants is driven by established firms (column 9) and is not significant for business ideas (column 6). Marginal established firms had additional annual revenue of 116M COP ($29K USD) during 2015–2017 (or 2.4 times their initial revenue), whereas the estimate for marginal business ideas is negative, albeit not statistically significant. Results in Table 10 provide a possible explanation for this difference in estimated average impact. The table shows that accelerated entrepreneurs that applied

with ideas (and not established firms) were less likely to start a firm during the first year after the program, though they often closed that gap during the second year. The first year after treatment, 39% of the rejected applicants created a firm, while only 9% of participants did. By 2016, those numbers increased to 56% for the controls and 50% for the participants. In the third and last year, 66% and 58% of the firms created were established firms, respectively. The reason for this delay could be explained by the bootcamps, which included discussions on the value of delaying firm creation until product markets are identified. To produce these reduced form results, we regress an indicator variable of firm registration at the Chamber of Commerce against the interaction between the indicator variable for acceleration and the different year fixed effects.

3.4. Impacting high-potential firms

The results so far show compelling evidence of average treatment effects among the marginal companies whose acceleration is altered by judge assignment. One question that remains regards the interaction between growth potential and treatment: critics often contend that accelerators can only teach fledging businesses how to "pitch" but cannot provide value-added to high-potential ventures. In this section, we take further advantage of the random allocation of applicants across judges with different scoring generosities to explore the evidence on this contention. Our main objective is to estimate a treatment *curve* that describes how acceleration affects venture performance as a function of applicants' growth potential.

One unique feature of our setting allows this additional exercise: the marginal projects are not clustered around the top-35 eligibility threshold—e.g., Panel B of Figure 1 shows how some marginal participants rank as low as the 80th company. In contrast, in parallel papers exploiting eligibility thresholds using regression discontinuity approaches to assess acceleration impacts, the marginal projects are by construction positioned close to eligibility cut-offs (see González-Uribe and Leatherbee 2018a; González-Uribe, Hu, and Koudijs 2018). As a consequence, extrapolating results far from the threshold in these papers remains a challenge.

We exploit this feature of our setting using an adaptation of the methodology of Xie, Brand, and Jann (2012) to estimate a treatment curve. The estimated curve is plotted in Figure 8, and shows

novel evidence of heterogeneous acceleration effects. Under the hypothesis that acceleration affects all applicants equally (i.e., independent of their growth potential), the treatment curve would be a constant function of an applicant's growth potential. Instead, Figure 8 shows an increasing relation between the impact of acceleration and the applicant's growth potential. Strikingly, the figure shows that the type of heterogeneity runs opposite to common claims. There is no evidence of treatment effects for bottom quality applicants (i.e., those applicants with a propensity for acceleration below 0.4), whereas positive effects are visible for projects with higher growth potential. For projects with a propensity for acceleration above 0.4, the relation between treatment and growth potential is increasing and approximately linear; annual revenues are estimated to increase by 183M COP ($21K USD) for every 10 percentage point increase in propensity for acceleration.

To estimate the curve in Figure 8, we begin by running matching regressions, where we match accelerated participants and rejected applicants based on their adjusted score and observed covariates at application. By construction, accelerated applicants and rejected applicants (so-matched) differ in observables on judge scoring generosity only. The matching algorithm we use is kernel matching (with a radius of 0.05), a nonparametric matching estimator that uses weighted averages of all individuals in the control group to construct the matched outcome (cf. Heckman, Ichimura, and Todd 1997). One advantage of this algorithm (over others based on one-to-one to matching such as nearest neighbor matching) is the lower variance which is achieved because more information is used. A drawback is that observations can be used that are bad matches, which we mitigate by restricting applicants to those in the common support. Results are presented in Appendix 8. The majority of participants fall within the common support, and the average absolute difference in propensity of acceleration is 0.014.

Next, we estimate individual treatment effects by each level of propensity score for acceleration (within the common support) as the difference in post-application annual revenues between participants and matched applicants, where the kernel weights are used to weight the outcomes of the matched applicants (cf. Smith and Todd 2005).[23] Appendix 8 shows that the average of the individual treatment

[23] We use a symmetric, nonnegative, unimodal kernel; hence, higher weight is placed on applicants who are close in terms of propensity score of a participant and lower weight is placed on more distant observations.

effects is 59.10, which is (by design) very close to the local average treatment effect that we estimated in Section 3.3 using the IV approach.

Finally, we transform our data so that the individual treatment effects constitute the observed data subject to further modelling. We then apply local polynomial regressions to the transformed data to connect the estimated differences per propensity score of acceleration within the common support, where the confidence intervals ensue from the asymptotic normality of the local polynomial estimators (cf. Fan and Gijbels 1996). The identification assumption behind this curve is that for any given propensity for acceleration, accelerated participants and matched applicants only differ in their "judge assignment luck", and thus that absent differences in the scoring generosity of judges both types of companies would have had the same treatment status (i.e., both accepted or both rejected).

The results in this section constitute the first rigorous evidence on heterogeneous impacts from acceleration that we are aware of. Taken at face value, Figure 8 shows that, contrary to common claims, the value-added from acceleration is highest for high-potential applicants, which suggests that accelerators have the ability to impact applicants with "gazelle potential". The pattern in Figure 8 also helps explain the positive difference between the IV and the OLS estimates reported in Table 9. Taken together, the results in Table 9 and Figure 8 imply that the average treatment effect we estimate for the marginal applicants using the IV approach is driven by the high-potential applicants that are left out of the program because they were assigned to relatively stricter judges.

The main caveat from the results in this section is the small sample size, and our estimates should be interpreted with this limitation in mind. For example, small sample issues may help explain why we find no evidence of impact for bottom quality applicants (i.e., the subsample used to estimate this part of the curve is very small; see Panel B in Figure 1). Taken at face value, the main implication from the results in this section is that failing to correct for judge heterogeneity in scoring generosity curtails these programs' potential for impact, which can lead to sizable revenue losses. We return to this point in Section 5.1 below.

4. Robustness Checks

The main concern with our impact results is the potential small sample bias from the small cross section; that is, the possibility that our effects pick up the 5% chance we would see an effect when no such effect exists. This concern is minimized by the fact that our results are robust to using different outcomes variables (see Appendix 7), and to different measures of scoring generosity and different sets of controls.

Nevertheless, to further address this concern we use a randomization inference (RI) approach that conducts exact finite-sample inference, and which remains valid even when the number of observations is small (cf. Rosenbaum 2002). This approach is somewhat similar to bootstrap, but different in spirit. In particular, when estimating bootstrapped p-values the econometrician is looking to address her uncertainty over the specific sample of the population she drew, while randomization inference helps the econometrician address instead her uncertainty over which units within her sample are assigned to the treatment.

There are two steps to our RI approach. In the first step we identify a subsample of applicants where we argue treatment can be assumed to be "as good as randomly assigned." We select this subsample by taking (i) all accelerated applicants with lower adjusted scores than the highest adjusted score of a nonaccelerated applicant and (ii) all nonaccelerated applicants with adjusted scores higher than the lowest adjusted score of the accelerated projects. Overall, we find 62 projects that match our definition, 28 being ideas and 34 established businesses. For this subsample, we estimate the treatment effect as the relative increase in revenue for participants versus nonparticipants, and then test the sharp null hypothesis of no treatment effect by applying standard exact randomization inference tools (see among others: Rosenbaum 2002, 2010; Imbens and Rosenbaum 2005). In particular, we scramble the data 5,000 times, each time randomly assigning different companies to be placebo participants. For each permutation we estimate a placebo effect equal to the average conditional difference between placebo participants and placebo rejected applicants, as estimated using equation (3). We then compare the placebo effect with our estimated treatment effect and keep track of the number of times that our estimate is bigger (in absolute value) than the placebo difference. We then say that we reject the null of no treatment effect if in more than 95% of the permutations our treatment estimate (absolute value) is smaller than the placebo effects. Results are summarized in Table 11 and illustrated in Figure 9.

The results in Table 11 suggest that our main results (Table 9) are unlikely to be driven by small sample bias: our placebo effects for established firms are larger than our real estimates in only 3.5% of the permutations (Column 2, Panel B). The identification assumptions behind this randomization inference test are that the distribution of the score is the same for all observations in the subsample and that initial outcome variables are statistically similar between groups. We present supportive evidence in Panel A of Table 11, where we show that adjusted scores and revenue among treated and nontreated entrepreneurs are similar in the subsample. The differences in sectorial experience and in initial number of employees (for established firms) are controlled for in the regressions presented in Panel B.

5. Discussion

Our findings show compelling evidence that entrepreneurship schooling in accelerators can have significant impacts on venture performance, especially for companies with high-growth potential. The implied magnitudes of our average findings for revenues are large relative to similar work on business training interventions for traditional microenterprises. For example, Calderon, Cunha, and De Giorgi (2013) and de Mel, McKenzie, and Woodruff (2014) find a 20% and a 41% increase in revenues (within 12 and 18 months), which contrasts with our estimated 157% revenue increase relative to baseline. The difference in magnitudes is consistent with ValleE's objective: to target high-potential entrepreneurs that are looking to scale their projects beyond a traditional microenterprise. These differences can also reflect additional added-value sources of schooling in business accelerators beyond business instruction, including advice from mentors and staff, access to a network of like-minded individuals, and certification and self-validation effects from selection, among others. Distinguishing between these different sources of value-added is one of the main topics for future work in the literature.

In the rest of this section we discuss two final points: revenue costs from failing to correct for heterogeneity in scoring generosity and the external validity of the results.

5.1 Revenue Costs

We use a back-of-the-envelope calculation of our findings to estimate ValleE's costs from failing to correct for heterogeneity in judges' scoring generosity. The results are presented in Table 12.

We find that program revenues could have increased by 31%–40% had the accelerator used different selection mechanisms, where program revenues correspond to the sum of the revenues from all participants.

To produce these estimates, we start by calculating counterfactual revenues to the accelerator under the assumption that the program selected applicants based on adjusted scores rather than average scores. To make this calculation, we split the sample of participants into two groups: marginal (12) and nonmarginal, as explained in Section 2.4. We then estimate counterfactual revenues in 2015–2017 for the marginal projects using the implied revenues from Figure 8 (according to the heterogeneous treatment effects by propensity of acceleration). The counterfactual revenues of nonmarginal projects are the same as observed revenues. Finally, we compare observed and counterfactual revenues. The results are presented in Table 12. They show that counterfactual revenues are 31% higher than actual revenues.

We then consider a second alternative scenario, where the program selected participants based both on adjusted scores and covariates at application. We consider this second scenario because the results in Table 4 (Panel B) show that entrepreneurs' characteristics (even after controlling for adjusted scores) are highly correlated with firm growth. Therefore, an algorithm that takes into account both, the hard (observables) and soft (adjusted scores) information, rather than just adjusted scores, is likely to generate further selection improvements. To calculate counterfactual revenues in this scenario, we use the propensity score calculated in Section 3.4 to create a ranking of the best 35 entrepreneurs according to the ex-ante observables and the adjusted scores. According to this new ranking, 26 out of the 35 accelerated entrepreneurs are not marginal in that they would also had been selected by this method, whereas 9 are classified as marginal. Performing the same exercise as above (i.e., replacing the revenues of marginal projects for those of the rejected applicants plus the implied treatment effect), we find potential revenue increases of 40%, as shown in Panel B of Table 12.[24]

[24] In unreported analysis, we show that the additional revenue improvements in the second counterfactual stem from keeping four applicants that rank below the top 35 adjusted scores but rank high in propensity scores, and from eliminating one company that ranks low in propensity score but is among the top 35 adjusted scores.

Our results have implications for business accelerators, as well as other interventions that select participants based partly on scores from partially overlapping judge panels. They imply that sizable revenue improvements can exist if they control for heterogeneity across judges in scoring generosities. Further potential improvements can also be had if selection processes rely on hard information too, rather than only on judges' evaluations. It is possible that the costs from failing to correct for judge heterogeneity in scoring generosity extend beyond the accelerator and imply more general welfare inefficiencies. This will happen if a number of conditions are satisfied, including if (i) the schooling opportunities forgone by the accelerator are not arbitraged away by other accelerators (or other economic agents) and (ii) the provision of schooling through the program is more efficient (including potential spillover effects) than provision through alternative mechanisms (i.e., mentoring by angel investors). Given the high-quality resources at ValleE relative to other support organizations available in the region, efficiency costs are likely, but given data limitations, a judicious welfare quantification is beyond the scope here.

5.2 External Validity

The main potential issue that can affect the external validity of results is that ValleE is a unique program, and thus results from this setting may not be generalizable to other accelerators. However, ValleE is actually very similar to the average ecosystem accelerator on many dimensions. For example, the location of the program (not the capital city) is a common trait among ecosystem accelerators. Roughly 38% of these programs are located in underdeveloped regions. 40% are in the United States, but outside Silicon Valley, Massachusetts, New York, or Washington D.C.; the rest are in Europe but typically not in the capital cities. In terms of services, those offered at ValleE are similar to the offerings of these programs worldwide (cf. Clarysse, Wright, and Van Hove 2015). This is not to say that some differences between ValleE and other ecosystem accelerators do not exist. Perhaps the most distinguishing features of the program include its access to highly qualified staff, mentors, and judges. We are also careful to emphasize the differences between average applicants to ValleE and other ecosystem accelerators, as mentioned in Section 2. We argue that the external validity of our findings

is likely confined to other ecosystem accelerators that attract young businesses with traction and have access to high-quality resources, including staff, mentors, and judges.

6 Conclusion

We investigate whether business accelerators can identify and impact high-growth firms in the context of an ecosystem accelerator in Colombia that provided participants with entrepreneurship schooling (e.g., business training, access to mentors, visibility), but no cash. We find compelling evidence that these programs can both identify and impact firms with high growth potential, using a novel methodology that exploits the random allocation of applicants across judges with different scoring generosities. Our findings provide first-time evidence of heterogeneity in accelerator impacts: entrepreneurship schooling has sizable effects on the performance of applicants with high-potential and no apparent effects on low-quality projects. This novel evidence casts doubt on critics' contention that accelerators can only teach low-quality fledging businesses how to attract funds, rather than add value to high-potential ventures. We quantify revenue improvements of 31%–40% from changing selection processes to account for judge heterogeneity in scoring generosity. Our results have direct implications for the design of selection processes in business accelerators and other programs based on the evaluation of partially overlapping judge panels. The findings are most relevant for accelerators that attract young businesses with traction, and provide access to high-quality mentors, staff, and evaluators. They suggest that entrepreneurial capital matters for young businesses, and that business accelerators can mitigate constraints on this type of capital in high-potential firms.

References

Aizer, Anna, and Joseph Doyle, Jr. 2013. "Juvenile Incarceration, Human Capital, and Future Crime: Evidence from Randomly-Assigned Judges." NBER Working Paper No. 19102.

Ahmad, N. and Petersen, R., 2007, High-Growth Enterprises and Gazelles – Preliminary and Summary Sensitivity Analysis, OECD-FORA, Paris

Autio, E., Arenius, P., & Wallenius, H., 2000, Economic impact of gazelle firms in Finland. Working Papers Series 2000:3, Helsinki University of Technology, Institute of Strategy and International Business, Helsinki.

Autor, David, and Susan Houseman. 2010. "Do Temporary-Help Jobs Improve Labor Market Outcomes for Low-Skilled Workers? Evidence from "Work First." American Economic Journal: Applied Economics, 2(3): 96-128.

Bertrand, M. and Schoar, A. 2003. Managing with style: The effect of managers on firm policies. Quarterly Journal of Economics 118: 1169–1208.

Birch, D. L., Haggerty, A., & Parsons, W. (1995). *Who's creating jobs?*. Boston: Cognetics Inc.

Birch, D.L. and Medoff, J. 1994. Gazelles. In L.C. Solmon & A.R. Levenson, (Eds), Labor markets, *employment policy and job creation*, pp. 159-167. Boulder, CO and London: Westview Press.

Bloom, N., Eifert, B., Mahajan, A., McKenzie, D., and Roberts, J. 2013. Does management matter? evidence from India. *Quarterly Journal of Economics* 128: 1–51.

Bloom, N. and van Reenen, J. 2010. Why do management practices differ across firms and countries? *Journal of Economic Perspectives* 24: 203–24.

Bone, J., Allen, O., and Haley, C. 2017. *Business incubator and accelerators: The national picture.* Department for Business, Energy and Industrial Strategy, Nesta, BEIS Research Paper Number 7.

Bruhn, M., Karlan, D., and Schoar, A. 2010. What capital is missing in developing countries? *The American Economic Review* 100: 629–33.

Calderon, G., Cunha, J.M., and De Giorgi, G. 2013. *Business literacy and development: Evidence from a randomized controlled trial in rural Mexico.* 2013. NBER Working Paper No. 19740.

Card, D. 2001. Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica* 69: 1127–60.

Clarysse, B., Wright, M., and Van Hove, J. 2015. *A look inside accelerators: Building businesses.* Research Paper, Nesta, London, UK.

Chang, T., and Schoar, A. 2013. *Judge specific differences in chapter 11 and firm outcomes*. MIT Working Paper.

Cohen, S.G. and Hochberg, Y.V. 2014. *Accelerating startups: The seed accelerator phenomenon.* Working Paper.

Deschryvere, M., 2008, High-growth firms and job creation in Finland. Discussion Paper No. 1144, Research Institute of the Finnish Economy (ETLA), Helsinki.

Di Tella, R., and Schargrodsky, E. Criminal recidivism after prison and electronic monitoring. *Journal of Political Economy* 121 (2013): 28-73

Dobbie, W., Goldin, J., and Yang, C. 2018. "The Effects of Pre-Trial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges." American Economic Review, 108, 2, Pp. 201-240.

Doyle, Joseph. 2007. "Child Protection and Child Outcomes: Measuring the Effects of Foster Care." American Economic Review, 97(5): 1583-1610.

Doyle, Joseph. 2008. "Child Protection and Adult Crime: Using Investigator Assignment to Estimate Causal Effects of Foster Care." Journal of Political Economy, 116(4): 746-770.

Mel, S. de, D. McKenzie, and Woodruff, C. 2008. Returns to capital in microenterprises: Evidence from a field experiment. *The Quarterly Journal of Economics* 123: 1329–72.

———. 2014. Business training and female enterprise start-up, growth, and dynamics: Experimental evidence from Sri Lanka. *Journal of Development Economics* 106: 199–210.

Dobbie, Will, and Jae Song. 2015. "Debt Relief and Debtor Outcomes: Measuring the Effects of Consumer Bankruptcy Protection." American Economic Review, 105 (3): 1272-1311.

Drexler, A., G. Fischer, and Schoar, A. 2014. Keeping it simple: Financial literacy and rules of thumb. *American Economic Journal: Applied Economics* 6: 1–31.

Ewens, M. and Rhodes-Kropf , M. 2015. Is a VC Partnership Greater than the Sum of its Partners? The Journal of Finance, 2015, Vol. 70-3, 1081-1113

Fafchamps, M. and Woodruff, C. 2016. Identifying gazelles: expert panels vs. surveys as a means to identify firms with rapid growth potential. Policy Research Working Paper, World Bank, WPS 7647.

Fan, J. and Gijbels, I. (1996) Local Polynomial Modelling and Its Applications. Chapman and Hall, London.

Fee, C., Hadlock, C. and Pierce, J. 2013, Managers with and without Style: Evidence Using Exogenous Variation, The Review of Financial Studies, Volume 26, Issue 3, 1, Pages 567–601

Fehder, D.C. and Hochberg, Y.V. 2014. *Accelerators and the regional supply of venture capital investment.* Working Paper.

French, Eric, and Jae Song. 2011. "The Effect of Disability Insurance Receipt on Labor Supply." Federal Reserve Bank of Chicago Working Paper WP-2009-05.

Goldfarb, B., Kirsch, D., and Miller, D.A. 2007. Was there too little entry during the Dot Com Era? *Journal of Financial Economics* 86: 100–144.

González-Uribe, J. and Leatherbee, M. 2018a. The effects of business accelerators on venture performance: Evidence from Start-Up Chile. *Review of Financial Studies* 31: 1566-1603.

González-Uribe, J. and Leatherbee, M. 2018b. Selection issues in 'Accelerators' published in Editors: Mike Wright, Imperial College Business School and Israel Drori, VU, Amsterdam, 2018.

González-Uribe, J, Hu, Z and Koudjis, P. 2018. Corporate Accelerators, Working paper.

Hall R. and Woodward, S., 2010. The burden of the nondiversifiable risk of entrepreneurship. *American Economic Review* 100: 1163-94

Haltiwanger, J.C., Jarmin, R.S. and Miranda, J. 2013. Who creates jobs? Small versus large versus young. *The Review of Economics and Statistics* XCV: 347–61.

Heckman, James, Ichimura, Hidehiko and Todd, Petra E., 1997, Matching As An Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme, Review of Economic Studies, 64, issue 4, p. 605-654

Henrekson, M. and Johansson, D. 2008. *Gazelles as job creators—A survey and interpretation of the evidence.* IFN Working Paper No. 733

Imbens, G.W. and Angrist, J.D. 1994. Identification and estimation of local average treatment effects. Econometrica 62(2): 467–75.

Imbens G, Rosenbaum P. 2005. Randomization Inference with an Instrumental Variable. Journal of the Royal Statistical Society, Series A 168 (1) :109-126.

Kahneman, D., and G. Klein. 2009. "Conditions for Intuitive Expertise: A Failure to Disagree." American Psychologist 64 (6): 515–26.

Karlan, D and Valdivia, M. Teaching Entrepreneurship: Impact of Business Training on Microfinance Clients and Institutions. The Review of Economics and Statistics. Vol. 93, No. 2 (May 2011), pp. 510-527

King, E. Behrman, J. Timing and Duration of Exposure in Evaluations of Social Programs, *The World Bank Research Observer*, Volume 24, Issue 1, 1 February 2009, Pages 55–82, https://doi.org/10.1093/wbro/lkn009

Kirchhoff, B.A., 1994, Entrepreneurship and Dynamic Capitalism. Westport, Conn.: Praeger.

Kling, J.R. 2006.. Incarceration length, employment, and earnings. American Economic Review 96: 863-876.

Klinger, B. and Schundeln, M. 2011. Can entrepreneurial activity be taught? Quasi-experimental evidence from central America. World Development 39: 1592–1610.

Landström, H., 2005, *Pioneers in entrepreneurship and small business research*. Berlin: Springer.

Lerner, J. and Malmendier, U. 2013. With a Little Help from My (Random) Friends: Success and Failure in Post-Business School Entrepreneurship, Review of Financial Studies, Society for Financial Studies, vol. 26(10), pages 2411-2452.

Mano, Y, Akoten, J, Yoshino, Y, and Sonobe, T. 2011. Teaching KAIZEN to small business owners: an experiment in a metalworking cluster in Nairobi, Working paper

Maestas, Nicole, Kathleen Mullen, and Alexander Strand. 2013. “Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt.” American Economic Review, 103(5): 1797-1829.

McKenzie, David. 2017. "Identifying and Spurring High-Growth Entrepreneurship: Experimental Evidence from a Business Plan Competition." American Economic Review, 107 (8): 2278-2307.

McKenzie, D. J. & Sansone, D. 2017. *Man vs. machine in predicting successful entrepreneurs: evidence from a business plan competition in Nigeria*. Policy Research Working Paper Series 8271, The World Bank.

McKenzie, D. and Woodruff, C. 2008. Experimental evidence on returns to capital and access to finance in Mexico. *The World Bank Economic Review* 22: 457–82.

——— 2014. What are we learning from business training and entrepreneurship evaluations around the developing world? *World Bank Research Observer* 29: 48–82.

Nanda, R. 2006. Financing high-potential entrepreneurship. *IZA World of Labor 2016*: 252.

Puri, M. and Zarutskie, R. 2012. On the life cycle dynamics of venture capital and non venture-capital financed firms. *The Journal of Finance* 67: 2247–93.

Rosenbaum, P.R. 2002 Observational studies. 2nd Edition, Springer, New York. doi:10.1007/978-1-4757-3692-2

Rosenbaum, P. R. 2010. Design of observational studies. New York, NY: Springer-Verlag

Shanteau, J. 1992. “Competence in Experts: The Role of Task Characteristics.” Organizational Behavior and Human Decision Processes 53: 252–62.

Smith, J. and Todd, P., 2005. Does matching overcome LaLonde's critique of nonexperimental estimators?, Journal of Econometrics, Elsevier, vol. 125(1-2), pages 305-353.

Stock J, Yogo M. 2005. Testing for Weak Instruments in Linear IV Regression. In: Andrews DWK Identification and Inference for Econometric Models. New York: Cambridge University Press. pp. 80-108.

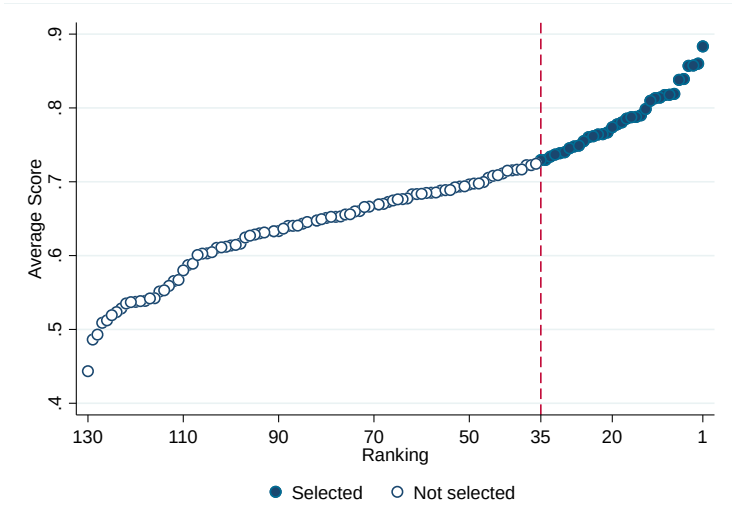
Wooldridge, J. 2002, Econometric Analysis of Cross Section and Panel Data, MIT Press

Young, A. 2018. Consistency without Inference: Instrumental Variables in Practical Application, Working paper, LSE.

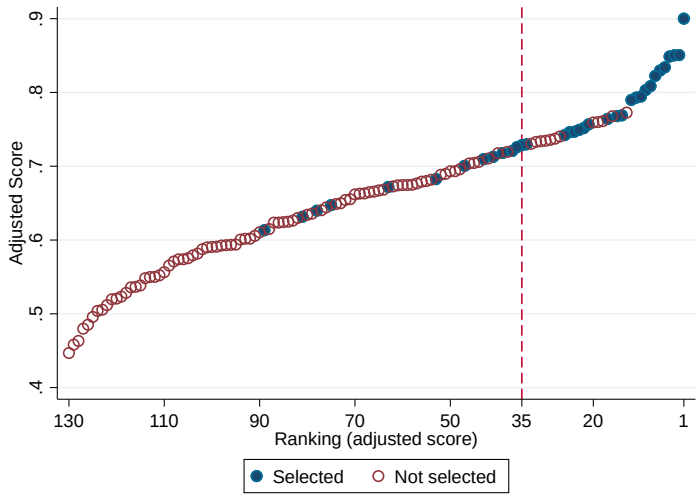
Xie, Yu, Jennie Brand, and Ben Jann. 2012. "Estimating Heterogeneous Treatment Effects with Observational Data." Sociological Methodology, 42(1): 314-347.

Figure 1. Distribution of applicant scores and selection into the accelerator

Panel A—Average scores

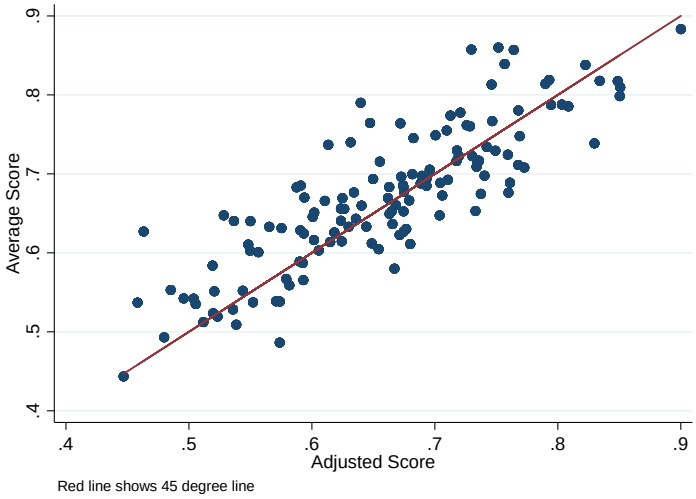


Panel B—Adjusted Scores



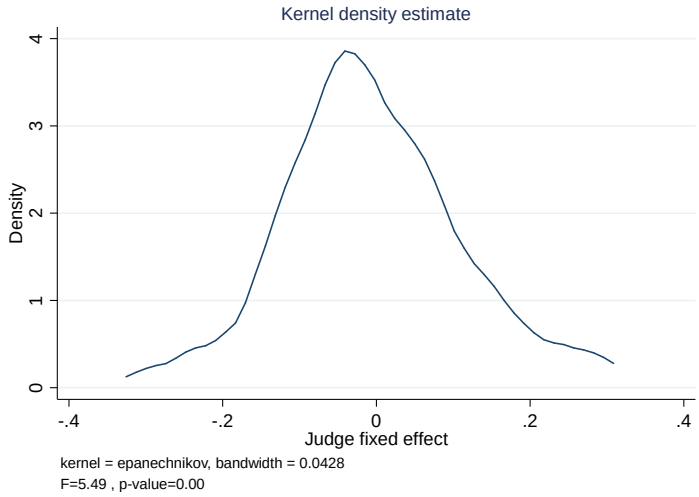
Panel A plots average scores against rankings based on the average score. Panel B plots adjusted scores, estimated as projects' fixed effects from equation (2), against rankings based on the adjusted score. In each panel, each dot represents an applicant; the solid (open) dots indicate the applicants that were (were not) selected into the accelerator.

Figure 2. Average scores and adjusted scores



This figure plots average scores against adjusted scores. Each dot represents an applicant. The red line shows the 45-degree line. Applicants with adjusted scores above the 45-degree line were "lucky" in that they drew a generous judge panel, while applicants with average scores below the 45-degree line were "unlucky" and drew a strict judge panel. The correlation between average scores and adjusted scores is 0.825.

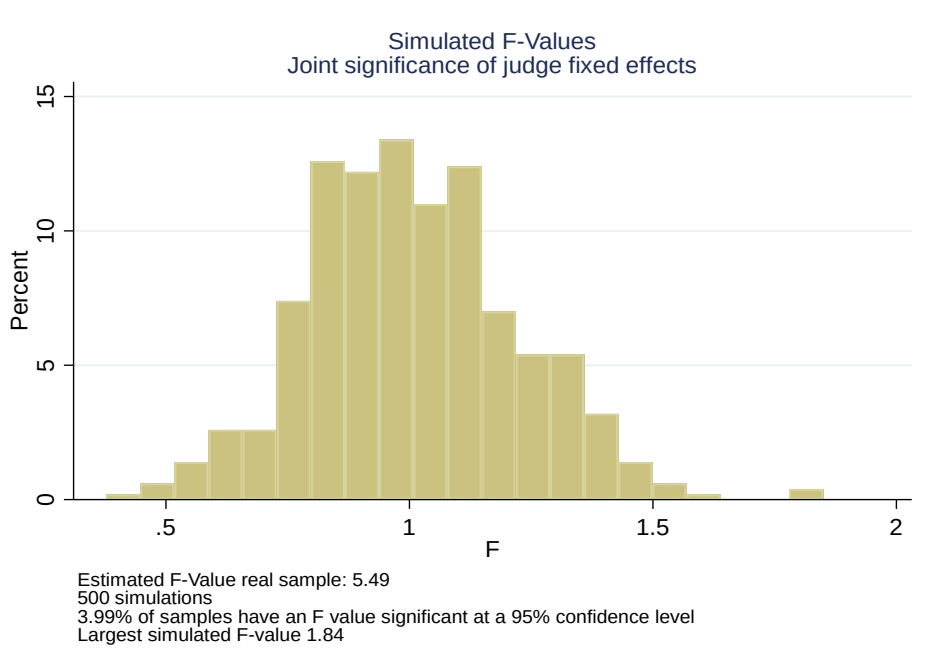
Figure 3. Distribution of judges' fixed effects



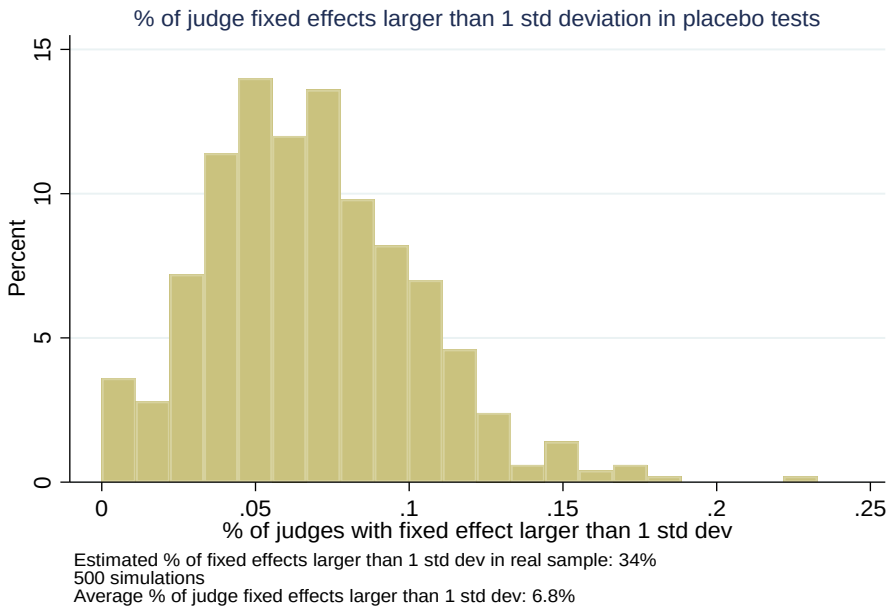
This figure plots the distribution of the estimated judge fixed effects from equation (2), which regresses project scores (by individual judges) against applicant fixed effects and judge fixed effects. Each project was evaluated by 3 randomly selected judges. Judges evaluated an average of 8 projects. The minimum (maximum) number of projects evaluated by a judge was 5 (14). The table reports the statistics of an *F*-test showing that the judge fixed effects are jointly significant (*p*-value of 0.00).

Figure 4. Placebo assignment of judges' scores

Panel A—Distribution of *F*-values

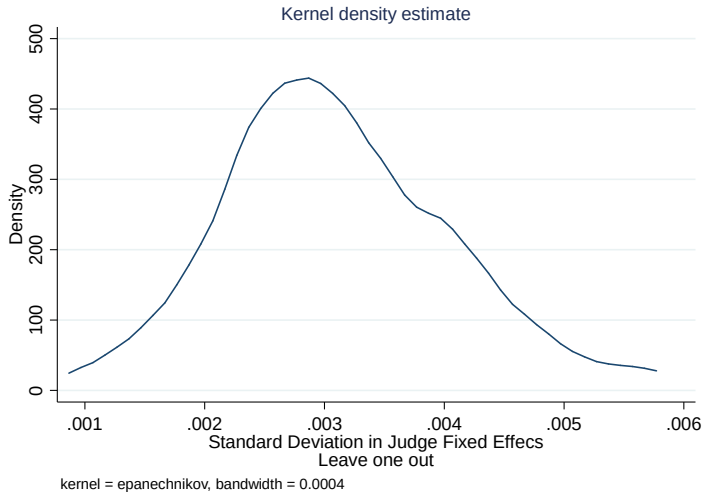


Panel B—Fixed effects one standard deviation above/below project effect



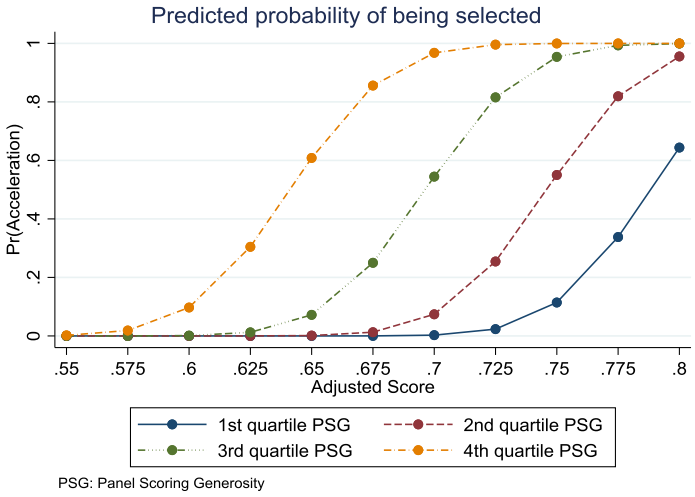
This figure plots the distribution of *F*-tests on the joint significance of the judge fixed effects in 500 placebo assignments.

Figure 5. Standard deviation of judge fixed effects (per judge)



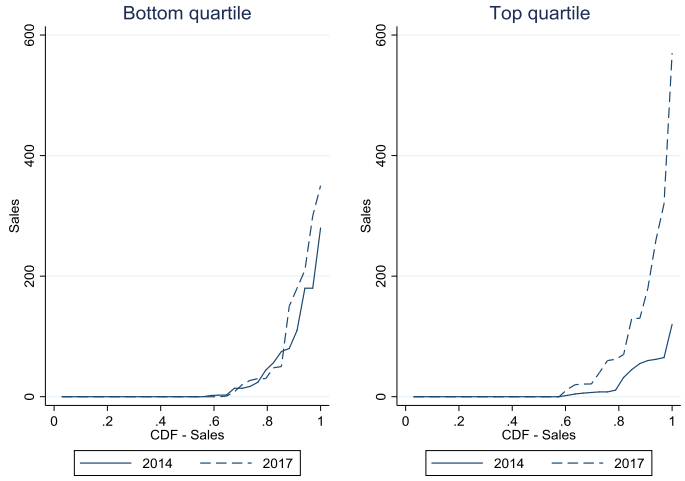
This figure plots the standard deviation of all the "leave one out" estimates of the judge fixed effects per judge. For each judge, we estimate as many fixed effects as the number of applicants the judge evaluated. We produce each estimate by sequentially leaving out of the sample each of the projects that the judge evaluated. The average standard deviation per judge of the leave one out fixed effects is 0.003 and the maximum is 0.006.

Figure 6. Probability of acceleration and scoring generosity



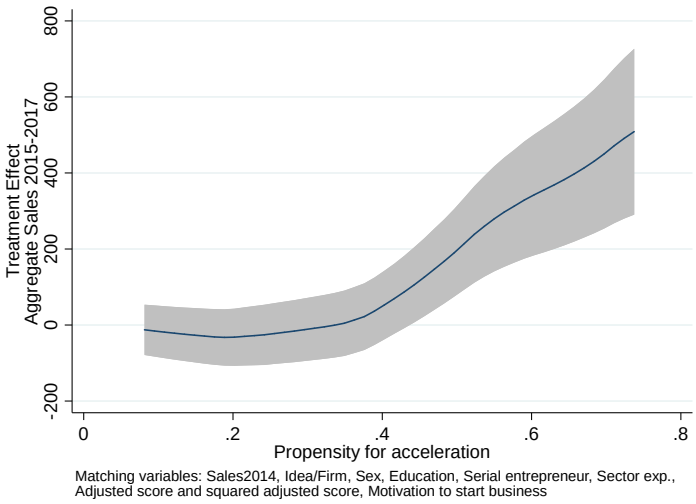
This figure plots the probability of acceleration against adjusted score by each quartile of scoring generosity. The top (bottom) quartile of scoring generosity corresponds to the most (least) generous judge panels.

Figure 7. Cumulative distribution of sales



This figure plots the cumulative distribution of revenues for 2014 and 2017 for projects in the bottom quartile of scoring generosity (left panel) and in the top quartile of scoring generosity (right panel).

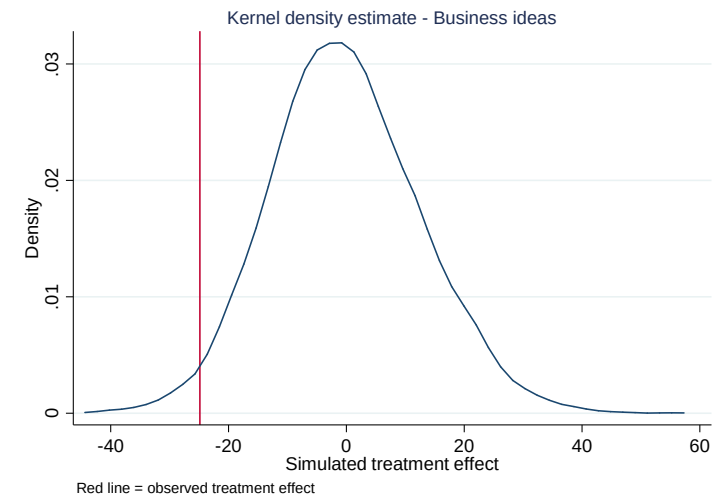
Figure 8. Heterogeneous acceleration effects



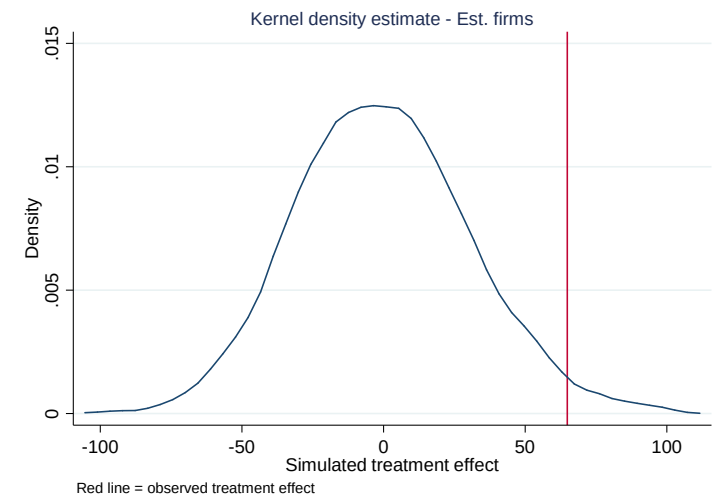
This figure plots the estimated treatment as a function of applicants' propensity for acceleration. The figure plots the estimated differences between accelerated participants and matched applicants with the same propensity score for acceleration (based on their adjusted score and observed covariates at application) but with differing treatments because they were randomly assigned to judges with different scoring generosities. We used a kernel matching algorithm and local polynomial regressions, based on the methodology of Xie, Brand and Jann (2012).

Figure 9. Randomization inference

Panel A. Business ideas



Panel B. Established firms



This figure plots results from the randomization inference exercise. Panel A (B) plots the distribution of the estimated acceleration effects from the 5,000 placebo assignments for the applicants that applied as business ideas (established firms).

Table 1. Sample composition

Variable	All Sample			Business Ideas	Established Firms
	Mean	Min	Max	Mean	Mean
Gender: Male	79%	0	1	75%	84%
Education: High school	12%	0	1	17%	6%
Education: Technical degree	21%	0	1	22%	21%
Education: College	52%	0	1	39%	67%
Education: Masters or PhD	15%	0	1	22%	6%
Location: Cali	85%	0	1	88%	83%
Motivation: Have stable income	12%	0	1	13%	11%
Motivation: Own boss	1%	0	1	0%	2%
Motivation: Business opportunity	87%	0	1	88%	87%
Dedication: Sporadic	6%	0	1	10%	2%
Dedication: Half-time	21%	0	1	25%	17%
Dedication: Full-time	73%	0	1	65%	81%
Sector experience (years)	5.6	0	30	4.7	6.6
Serial entrepreneur	61%	0	1	61%	62%
Has entrepreneurial team	88%	0	1	85%	92%
# people on team	3.0	1	10	2.8	3.3
Sector: Agriculture	16%	0	1	13%	19%
Sector: Manufacturing	21%	0	1	24%	17%
Sector: Water and Electricity	3%	0	1	4%	2%
Sector: Construction	3%	0	1	3%	3%
Sector: Commerce	2%	0	1	1%	3%
Sector: Services	56%	0	1	56%	56%
Participated in other contests	59%	0	1	56%	63%
% Established Firms	47%	0	1	0%	100%
Year founded (established firms)	2013	2010	2015	.	2013
Sales 2013 (million pesos)	10.62	0	290	1.27	21.48
Sales 2014 (million pesos)	25.80	0	300	4.61	50.01
Total employees 2014	4.0	0	25	2.7	5.6
Observations		135		72	63

The table presents the composition of the sample and selected summary statistics of the variables in the application forms. The sample includes all 135 applicants that were evaluated by judge panels. The subsample of established firms (business ideas) corresponds to applicants that at the time of the application had (had not) registered as a business with the Chamber of Commerce.

Table 2. Differences between accelerated and non-accelerated applicants

Variable	Business Ideas			Established Firms		
	Rejected	Accelerated	*P*-value Diff in means	Rejected	Accelerated	*P*-value Diff in means
Gender: Male	72%	87%	0.25	81%	90%	0.39
Education: High school	19%	7%	0.25	7%	5%	0.77
Education: Technical degree	21%	27%	0.65	28%	5%	0.04**
Education: College	39%	40%	0.92	58%	85%	0.04**
Education: Masters or PhD	21%	27%	0.65	7%	5%	0.77
Location: Cali	84%	100%	0.10	84%	80%	0.72
Motivation: To have stable income	11%	20%	0.33	12%	10%	0.85
Motivation: Own boss	0%	0%	.	2%	0%	0.50
Motivation: Opportunity	89%	80%	0.33	86%	90%	0.67
Dedication: Sporadic	11%	7%	0.66	2%	0%	0.50
Dedication: Half-time	25%	27%	0.87	16%	20%	0.72
Dedication: Full-time	65%	67%	0.90	81%	80%	0.90
Sector experience (years)	5.2	2.9	0.14	5.2	9.7	0.00***
Serial entrepreneur	53%	93%	0.00***	51%	85%	0.01***
Has entrepreneurial team	81%	100%	0.07*	91%	95%	0.56
# people on team	2.7	3.1	0.39	3.3	3.2	0.78
Sector: Agriculture	11%	20%	0.33	16%	25%	0.42
Sector: Manufacturing	28%	7%	0.08*	19%	15%	0.73
Sector: Water and Electricity	2%	13%	0.05**	0%	5%	0.14
Sector: Construction	2%	7%	0.31	0%	10%	0.04**
Sector: Commerce	2%	0%	0.61	5%	0%	0.33
Sector: Services	56%	53%	0.85	60%	45%	0.26
Participated in other contests	51%	73%	0.12	58%	75%	0.20
% Established Firms	0%	0%	.	100%	100%	.
Year founded (est. firms)	.	.	.	2013	2013	0.16
Sales 2013 (million pesos)	1.25	1.34	0.96	13.37	39.84	0.07
Sales 2014 (million pesos)	3.84	7.53	0.32	47.54	55.34	0.70
Total employees 2014	2.6	2.9	0.66	5.9	4.9	0.42

The table reports differences between accelerated and non-accelerated applicants, separately for established firms and business ideas. .*, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 3—Summary statistics

Variable	All Sample		
	N	Mean	SD
Sex: Male	135	79.3%	0.407
Location: Cali	135	85.2%	0.357
Sectoral experience (years)	135	5.59	5.643
Serial entrepreneur	135	61%	0.488
Has entrepreneurial team	135	88%	0.324
Motivation: To have stable income	135	12%	0.324
Motivation: Own boss	135	1%	0.086
Motivation: Business opportunity	135	87%	0.333
Education: High school	135	12%	0.324
Education: Technical degree	135	21%	0.412
Education: College	135	52%	0.502
Education: Masters or PhD	135	15%	0.357
Average score	135	0.67	0.090
Adjusted score	135	0.66	0.094
Sales 2014	135	25.80	56.14
Profits 2014	135	8.20	15.90
Total employees 2014	135	4.03	3.93
Sales 2015	135	51.66	124.25
Profits 2015	104	14.04	44.42
Total employees 2015	104	5.35	6.90
Sales 2016	135	58.32	128.77
Profits 2016	92	16.65	39.86
Total employees 2016	92	4.68	4.65
Sales 2017	135	50.64	118.61
Profits 2017	86	8.73	21.62
Total employees 2017	86	4.09	4.88

The table presents summary statistics of the main variables used in the analysis. The upper panel includes variables from the applications. The lower panel includes performance variables constructed using the application response (data before 2015), survey responses (employees, revenues, and profits 2015–2017) and the Colombian business registry (revenues 2015–2017).

Table 4. Applicant scores and project growth

Panel A-Regression results

	(1)	(2)	(3)	(4)	(5)	(6)
Average score		-6.861		-28.44		
		(22.42)		(28.12)		
Adjusted score			45.54**		-5.991	-52.37
			(17.58)		(19.12)	(64.64)
Average score ×After				36.18		
				(45.11)		
Adjusted score × After					86.16**	87.33**
					(37.46)	(35.42)
Location: Cali	22.59***	22.76*	20.85*	22.81*	20.98*	20.87*
	(7.498)	(12.02)	(12.09)	(11.66)	(12.18)	(12.10)
Gender (1=Male)	15.34**	15.50	15.28	15.44	15.18	13.48
	(6.883)	(10.50)	(10.43)	(10.58)	(10.62)	(8.744)
Has entrepreneurial team	-0.229	-0.271	-4.485	-0.275	-4.479	-2.905
	(8.387)	(7.008)	(7.534)	(7.243)	(8.595)	(8.274)
Serial entrepreneur	9.480	9.613	5.698	9.586	5.631	4.213
	(7.307)	(11.51)	(11.39)	(11.26)	(11.32)	(11.83)
Sector experience	-0.255	-0.266	-0.301	-0.263	-0.293	-0.309
	(0.763)	(0.809)	(0.888)	(0.891)	(0.970)	(0.981)
Motivation: Own boss	-30.63	-31.65	-29.13	-31.62	-28.96	-23.49
	(22.33)	(22.68)	(20.79)	(21.82)	(20.36)	(23.35)
Motivation: Business opportunity	19.95***	20.00	20.69	19.94	20.56	22.38
	(6.365)	(14.67)	(14.07)	(14.32)	(13.70)	(14.63)
Education: Technical degree	-17.65**	-17.43	-17.25	-17.49	-17.37	-19.37
	(8.251)	(11.14)	(10.90)	(12.15)	(12.79)	(12.95)
Education: College	9.047	9.532	7.822	9.490	7.778	5.143
	(8.177)	(9.881)	(8.335)	(11.11)	(10.66)	(8.757)
Education: Masters or PhD	21.27*	21.48	19.93	21.49	19.95	20.20
	(11.75)	(20.68)	(20.80)	(22.01)	(20.90)	(22.22)
Constant	-49.64**	-45.60**	-71.08***	-30.56	-37.35	-9.736
	(22.65)	(17.76)	(23.79)	(18.30)	(22.06)	(32.73)
Observations	675	675	675	675	675	675
R-squared	0.209	0.209	0.210	0.210	0.212	0.215
Control for Acceleration						Yes

Panel B—Shapley Owen participation

	(1) Observables	(2) Observables + *Average score*	(3) Observables + *Adjusted score*
Firm's age	61.5%	61.0%	59.7%
Entrepreneur characteristics	17.4%	16.9%	15.4%
Firm's characteristics	11.5%	11.3%	11.0%
Context (time fixed effects)	9.7%	9.7%	9.6%
Score		1.2%	4.4%
R-squared	0.209	0.209	0.210

The table presents results from estimating equation (1). The outcome variable is revenue. The variables average score and adjusted score correspond to the average score from the panel judges and the adjusted score that removes the judge fixed effects. All columns include time fixed effects and industry fixed effects. Standard errors are clustered at the applicant level and bootstrapped for all columns including the adjusted score as a covariate (columns 3, 5 and 6). *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 5. Predicting gazelles

Panel A—Probit models

	(1)	(2)	(3)	(4)
Top quartile (by adjusted score)	0.204**	0.189**	0.220**	0.200**
	(0.0818)	(0.0814)	(0.0974)	(0.0838)
Controls for top quartile of average score		Yes	Yes	Yes
Controls for covariates at application			Yes	Yes
Control for treatment				Yes
Pseudo R^2	0.0873	0.0881	0.184	
Observations	135	135	135	135

Panel B—Revenue growth rates

	Initial sales (2014)	Final sales (2017)	Growth from baseline	Implied annual growth
Gazelles	63.5	302.3	376%	68.24%
Non-gazelles	20.7	16.8	-19%	-6.70%

Panel A in the table presents results from probit regressions; reported coefficients correspond to marginal effects. The main explanatory variable is an indicator for applicants in the top quartile of adjusted scores. Regression controls vary as specified in each column. The covariates at application include indicator variables for established firms, gender, serial entrepreneurs, and founding team. They also include fixed effects for sectorial experience and entrepreneurs' education. The dependent variable is an indicator for gazelles: the top 10% applicants according to 2017 revenue (and splitting sample into business ideas and established firms). Panel B summarizes revenue growth rates across gazelles and non-gazelles.

Table 6. Characteristics of accelerated applicants

	Marginal participants	Nonmarginal participants	*P*-value Difference in means
Gender: Male	83%	87%	0.73
Education: High school	4%	4%	0.98
Education: Technical degree	21%	9%	0.25
Education: College	54%	70%	0.29
Education: Masters or PhD	21%	17%	0.77
Location: Cali	88%	91%	0.68
Motivation: To have stable income	17%	9%	0.42
Motivation: Own boss	0%	0%	.
Motivation: Business opportunity	83%	91%	0.42
Dedication: Sporadic	4%	4%	0.98
Dedication: Half-time	17%	22%	0.67
Dedication: Full-time	79%	74%	0.68
Sector experience (years)	7.4	6.4	0.61
Serial entrepreneur	79%	96%	0.09*
Has entrepreneurial team	88%	100%	0.08*
# people on team	2.5	3.2	0.06*
Sector: Agriculture	21%	22%	0.94
Sector: Manufacturing	13%	13%	0.96
Sector: Water and electricity	4%	9%	0.54
Sector: Construction	4%	9%	0.54
Sector: Commerce	0%	0%	.
Sector: Services	58%	48%	0.48
Participated in other contests	63%	83%	0.13
% Established firms	54%	57%	0.87
Sales 2013 (million pesos)	5.42	33.37	0.09*
Sales 2014 (million pesos)	27.60	43.48	0.43
Total employees 2014	4.1	4.3	0.83
Number of projects	23	12	

This table presents average differences between accelerated applicants that were (were not) affected by the scoring generosity of judges. Accelerated applicants that were affected by the scoring generosity (i.e., marginal applicants) are those businesses with average scores above the participation threshold, but with adjusted scores below the threshold. Accelerated applicants that were not affected by the scoring generosity are those businesses with both average and adjusted scores above the participation threshold.

Table 7. Unconditional probability of acceleration and scoring generosity

Quartile of panel judge generosity	Overall (No controls)	Project in 25th percentile (adjusted score= 0.59)	Median project (adjusted score= 0.66)	Project in 75th percentile (adjusted score= 0.73)	Project in 90th percentile (adjusted score= 0.77)
1 (Unlucky)	17.64%	0.00%	0.01%	3.32%	26.89%
2	23.53%	0.00%	0.54%	30.82%	76.38%
3	26.74%	0.03%	16.58%	85.91%	98.91%
4 (Lucky)	36.36%	5.31%	78.96%	99.78%	99.99%

This table shows the probability of acceleration across a double sort of applicants by adjusted score (columns) and quartile of scoring generosity (rows). Column 1 reports results from a probit regression of accelerated, a dummy that indicates applicants that participated in the accelerator, against dummy variables indicating the quartiles of scoring generosity. Columns 2 to 5 report results from the same probit but control for adjusted score.

Table 8. Probability of acceleration and scoring generosity

	(1)	(2)	(3)
	All	Business Ideas	Established Firms
2nd Quartile × After	0.203***	0.277***	0.101
	(0.0385)	(0.0472)	(0.0613)
3rd Quartile × After	0.272***	0.162***	0.352***
	(0.0389)	(0.0512)	(0.0580)
4th Quartile × After	0.490***	0.478***	0.482***
	(0.0406)	(0.0519)	(0.0626)
Adjusted Score * After	3.726***	3.161***	3.683***
	(0.201)	(0.259)	(0.318)
Constant	0.000	0.000	0.000
	(0.0144)	(0.0172)	(0.0206)
Observations	675	360	315
R-squared	0.652	0.678	0.739
Number of ids	135	72	63
F	57.58	35.80	39.09
Prob. > F	0.000	0.000	0.000

The table presents results from estimating equation (4). The outcome variable is accelerated × after. The variables accelerated and after correspond to dummy variables indicating accelerated applicants , and years after application to the accelerator, respectively. The bottom quartile of judge scoring generosity is omitted from the regression. All columns include applicant fixed effects and several controls, including adjusted score, firm's age, entrepreneur's age, entrepreneur's education, location, and sectorial and entrepreneurial experience. All columns include time fixed effects, and interactions between the controls and the variable after. Standard errors are bootstrapped and clustered at the applicant level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

.

Table 9. Acceleration and project growth

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		All			Business ideas			Established firms	
	Fixed effects	fixed effects	IV	Fixed effects	Fixed effects	IV	Fixed effects	Fixed effects	IV
Accelerated × After	40.94*	42.91**	66.31**	8.962	-3.715	-40.53	62.42**	99.80**	116.8**
	(24.23)	(20.80)	(32.31)	(16.92)	(23.08)	(38.09)	(30.32)	(40.47)	(56.96)
Constant	10.54***	10.54***	10.54***	1.266*	1.266**	1.266*	21.14***	21.14***	21.14***
	(3.166)	(3.512)	(3.284)	(0.660)	(0.637)	(0.661)	(6.519)	(7.716)	(6.399)
Observations	675	675	675	360	360	360	315	315	315
R-squared	0.074	0.095		0.087	0.100		0.092	0.156	
Number of ids	135	135	135	72	72	72	63	63	63
Controls × After	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes

The table presents results from estimating equation (3). The outcome variable is revenue. The variables accelerated and after correspond to dummy variables indicating applicants that were accelerated and years after application to the accelerator, respectively. All columns include applicant fixed effects and several controls, including adjusted score, firm's age, entrepreneur's age entrepreneur's education, location, and sectorial and entrepreneurial experience. All columns include time effects. Some specifications also include interactions between the controls and the variable after, as specified in each column under the row "Controls × After". Standard errors are bootstrapped and clustered at the applicant level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 10. Delay in firm creation

	(1)
Accelerated × 2015	-0.259*
	(0.133)
Accelerated × 2016	0.297*
	(0.160)
Accelerated × 2017	0.069
	(0.211)
Constant	0.773
	(0.450)
Observations	107
R-squared	0.140

The table presents results from regressing a dummy indicating firm registration at the Chamber of Commerce against interactions of the Accelerated indicator variable and year fixed effects. The estimation includes several controls, including adjusted score, firm's age, entrepreneur's age, entrepreneur's education, location, and sectorial and entrepreneurial experience. Standard errors are bootstrapped and clustered at the applicant level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 11. Randomization inference

Panel A--Mean differences pre-acceleration in subsample

	Business Ideas			Established Firms		
	Not accelerated	Accelerated	*P*-value Diff in means	Not accelerated	Accelerated	*P*-value Diff in means
Observations	19	9		20	14	
Adjusted score	0.69	0.71	0.25	0.70	0.72	0.35
Gender: Male	68%	89%	0.26	85%	93%	0.50
Location: Cali	95%	100%	0.50	90%	79%	0.37
Has entrepreneurial team	84%	100%	0.22	100%	93%	0.24
Sectorial experience (years)	6.7	2.9	0.05**	5.1	10.9	0.00***
Serial entrepreneur	74%	89%	0.38	70%	79%	0.59
Advanced education (college or grad)	68%	67%	0.93	75%	93%	0.19
Firm's age				2.8	3.1	0.42
Total employees 2014	2.4	2.0	0.64	7.6	4.5	0.09*
Sales 2014 (million pesos)	3.77	6.00	0.68	64.79	64.26	0.99

Panel B—Acceleration and revenue

	(1) Business ideas	(2) Established firms
Treatment * After	-24.86*	64.84**
p = c/n	0.053	0.035
SE (p)	0.003	0.0026
Controls and time fixed effects	Yes	Yes
Number of permutations	5000	5000
Note: c = #{|T| >= |T(obs)|}		

The table reports results for the randomization inference exercise. The sample is restricted to 62 applicants that belong to one of two subsamples: (i) applicants that were not accelerated but whose adjusted score is higher than the lowest adjusted score of the accelerated projects and (ii) accelerated applicants whose average score is lower than the highest average score among nonaccelerated applicants. Panel A summarizes differences between accelerated and nonaccelerated businesses in the restricted sample, for business ideas and established firms separately. Panel B summarizes the randomization inference results of regressing post-application revenue against a dummy indicating whether the applicant was accelerated, time fixed effects, and controls.

Table 12. Correcting for heterogeneity in scoring generosity and program revenues

Panel A—Adjusted scores

Real status	Counterfactual status	Number	Real revenue (2015–2017)	Sum estimated individual impact	Counterfactual revenue
Accelerated	Accelerated	23	$7,773	$10,380	$7,773
Accelerated	Rejected	12	$2,121	$924	
Rejected	Accelerated	12	$2,510	$2,651	$5,161=$2,510+$2,610
Revenue accelerator			$9,894		$12,934
Revenue improvement controlling for judge fixed effects					$3,040=$12,934-$9,894
% Revenue improvement					31%=$3,040/$9,894

Panel B—Propensity for acceleration

Real status	Counterfactual status	Number	Real revenue (2015–2017)	Sum estimated individual impact	Counterfactual revenue
Accelerated	Accelerated	26	$9,169	$11,419	$9,169
Accelerated	Rejected	9	$724	-$115	
Rejected	Accelerated	9	$1,363	$3,274	$4,637=$1,363+$3,274
Revenue accelerator			$9,894		$13,806
Revenue improvement controlling for judge fixed effects					$3,912=$13,806-$9,894
% Revenue improvement					40%=$3,912/$9,894

The table report our estimates of the revenue costs to the accelerator for not correcting for judge heterogeneity in scoring generosity. Panel A (B) includes the estimates for the counterfactual scenario where the accelerator selects participants based on the adjusted score (propensity score for acceleration, which is a function of the adjusted score and covariates at application).

. ONLINE APPENDIX

Appendix 1—Evaluation guidelines for ValleE judges

Por favor complete el siguiente formato de evaluación para cada uno de los proyectos a evaluar. Por favor use el ID que le fue asignado al igual que el ID de la empresa que va a evaluar y que encontrará en la ficha técnica de cada proyecto.

1) Por favor ingrese su ID de evaluador*

2) Por favor ingrese el ID del proyecto que está evaluando*

3) Nombre del proyecto que está evaluando*

4) Por favor califique de 1 a 5, qué tan bien definidos están los componentes del modelo de negocio. Por favor para esta pregunta califique ÚNICAMENTE la claridad en los aspectos y no el valor o potencial del modelo de negocio

5) Por favor califique de 1 a 5 el nivel de innovación del modelo de negocio, siendo 1 un modelo de negocio con muy bajo nivel de innovación (varios productos similares) y 5 un modelo de negocio con muy alto nivel de innovación (producto, servicio y/o mercado nuevo)*

6) Por favor califique de 1 a 5 el nivel de escalabilidad del proyecto, siendo 1 un modelo de negocio con muy poco escalable y 5 un modelo de negocio con alta probabilidad de escalabilidad.

Un proyecto es escalable cuando puede acceder a un siguiente nivel en términos de: volumen de ventas, cobertura geográfica, cobertura de otro nicho de clientes, crecimiento sostenible de empleados, entre otros

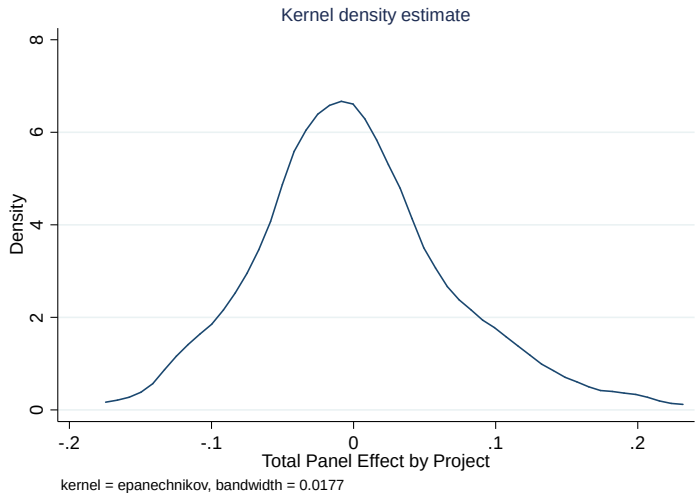
7) Por favor califique de 1 a 5 el potencial de rentabilidad del modelo de negocio, siendo 1 un modelo de negocio con muy baja probabilidad de generar rentabilidades mayores al promedio de su sector y 5 un modelo de negocio con altas probabilidades de generar rentabilidades mayores al promedio de su sector

8) Por favor califique de 1 a 5 las capacidades del equipo emprendedor para desarrollar el proyecto, siendo 1 un equipo con bajas capacidades y 5 un equipo experimentado y con las capacidades para desarrollarlo

Appendix 2—Judge Panels and scoring generosity

Relying on a panel of judges rather than on individual judges to rank projects helps mitigate the effect of judge heterogeneity but does not fully correct it, because judge panels are small (only three judges). We provide supporting evidence by comparing the distribution of *individual* scoring generosity (Figure 3) with the distribution of *overall* scoring generosity (Figure A2, below), defined as the sum of the corresponding judge fixed effects for each project. A comparison between the figures reveals that the distribution of the overall scoring generosity is more concentrated around 0 than the distribution of individual judge fixed effects, yet economically significant heterogeneity across judge panels remain, as 28% of judge panels tend to award individual scores that are one standard deviation above or below the average score of the other judge panels. The most generous (strict) judge panel adds (subtracts) an average of 0.12 (0.21) to any given project the group scores (relative to a mean normalized score of 0.66).

Figure A2. Overall scoring generosity



This figure plots the distribution of the estimated overall scoring generosity by project, which corresponds to the sum of the judge fixed effects from the estimates of equation (2).

Appendix 3—Sorting ability of judges

We characterize in more detail the evidence on the sorting ability of judges by cutting the data and regressing residual sales (post-application: 2015–2017) against adjusted scores. Residual sales correspond to the residuals from regressions of post-application revenue (i.e., 2015–2017 period) against all the growth determinants from applications (i.e., all variables included in column 1 of Panel A in Table 4). Table A.3 summarizes the results. There are three findings regarding the apparent sorting ability of judges.

The first finding shows that judges seem better at evaluating the prospects of business ideas rather than the prospects of already established firms. A comparison between Columns 2 and 3 in Table A.3 shows that the adjusted scores predict the revenue of projects that apply as business ideas, whereas the estimated coefficient for established firms is not statistically significant. An alternative interpretation is that we do not have sufficient statistical power to reject the no-prediction null. However, the difference in observations between the two sample cuts is small (216 against 189), whereas the increase in the variance of the estimator is substantial.

The second finding is that judges' sorting ability appears higher at the top, rather than at the bottom, of the distribution of projects. Column 4 shows that when we leave out the bottom quartile of the distribution by adjusted scores, going from the lowest ranked to the top ranked project increases future annual sales by 197 million COP. Instead, Column 5 shows no predictive power for the adjusted score once we drop the top quartile of participants (as indicated by the adjusted scores). This asymmetry in the predictive power of judges is consistent with prior work evaluating judges' predictions in business plan competitions (e.g., Fafchamps and Woodruf 2016). However, Fafchamps and Woodruf, (2016) show that in their sample the asymmetry runs in the opposite direction: judges in business plan competitions are better at cleaving off the bottom of the distribution, but not as effective at distinguishing within the top. One potential explanation behind the difference in results, is the preliminary filter in ValleE: recall that 120 (out of 255 total applications) were not assigned to judges for evaluation as their potential for growth was deemed to be too low. Had judges also evaluated the filtered projects, it is possible that adjusted scores would have had a high correlation with future performance for the bottom of the distribution (which presumably would include the projects that were filtered out).

Table A.3 Applicant scores and project growth: sample cuts

	Residual sales 2015 to 2017					
	All	Business ideas	Established firms	Excl. bottom Q.	Excl. top Q.	Excl. treated and bottom Q.
	(1)	(2)	(3)	(4)	(5)	(6)
Adjusted score	59.30*	57.74*	78.43	197.4***	-65.93	50.78**
	(35.38)	(33.37)	(59.24)	(55.95)	(78.85)	(23.63)
Observations	405	216	189	303	306	198
R-squared	0.003	0.008	0.002	0.013	0.002	0.030

Panel A reports the Shapley-Owen participation of each set of explanatory variables in equation (1) on the *R*-squared across the different regression models in Table 3. Note that some variables that are not significant in the estimation may still have a high Shapley-Owen participation value because of high correlation among explanatory variables in the model. Panel B reports the Shapley-Owen participation of each set of explanatory variables in equation (1) on the *R*-squared of the model in column (5) of Table 3 across different subsamples as detailed on the title of each column.

Appendix 4—Classification of gazelles

Type of applicant	Age at application	Number	Initial sales (2014)	Final sales (2017)	Growth from baseline	Implied annual growth
Gazelles	Business idea, <1 year incorporation	9	14.0	205.0	1364%	144.6%
Nongazelles	Business idea, <1 year incorporation	68	8.2	3.0	-63%	-28.5%
Gazelles	>1 year incorporation	7	127.0	427.0	236%	49.8%
Nongazelles	>1 year incorporation	51	28.2	44.3	57%	16.2%

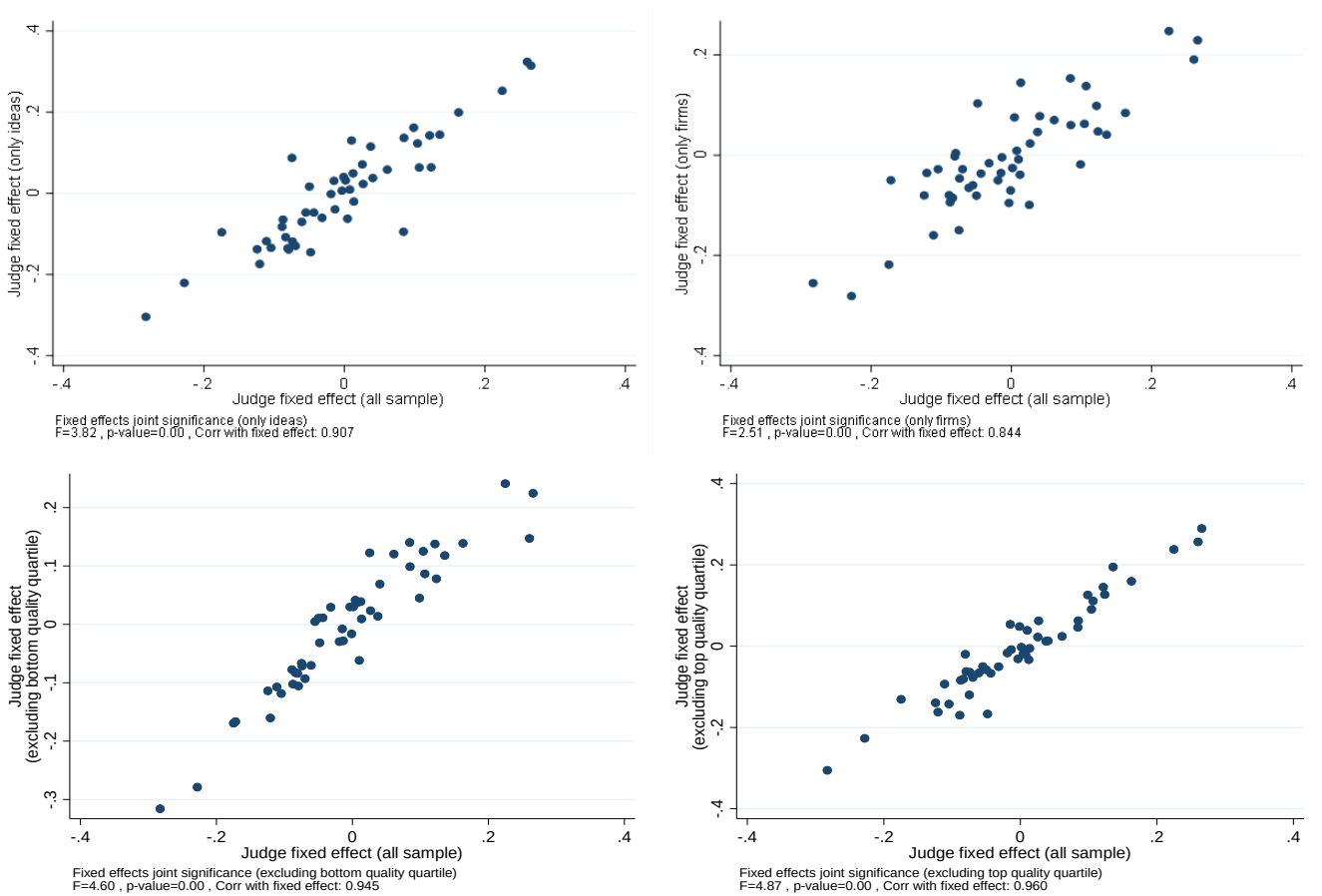
The table shows the revenue growth rates of gazelles and nongazelles across two types of applicants: (i) established businesses that have been registered with the Chamber of Commerce for more than one year at application and (ii) business ideas that were not incorporated at application and established businesses that had been registered with the Chamber of Commerce for less than one year at application.

Appendix 5—Balanced sample across scoring generosity quartiles

Variable	Q 1	Other Q	*p*-value diff. in means	Q 2	Other Q	*p*-value diff. in means	Q 3	Other Q	*p*-value diff. in means	Q 4	Other Q	*p*-value diff. in means
Sex: Male	76%	80%	0.65	76%	80%	0.65	76%	80%	0.65	88%	76%	0.16
Education: High school	3%	15%	0.06	24%	8%	0.01	15%	11%	0.56	6%	14%	0.24
Education: Technical degree	24%	21%	0.74	18%	23%	0.53	18%	23%	0.53	27%	20%	0.36
Education: College	53%	51%	0.88	44%	54%	0.30	56%	50%	0.59	55%	51%	0.72
Education: Masters or PhD	21%	13%	0.28	15%	15%	0.98	12%	16%	0.57	12%	16%	0.62
Location: Cali	91%	83%	0.26	85%	85%	0.98	85%	85%	0.98	79%	87%	0.24
Motivation: To have stable income	21%	9%	0.07	12%	12%	0.99	3%	15%	0.06	12%	12%	0.96
Motivation: Own boss	3%	0%	0.08	0%	1%	0.56	0%	1%	0.56	0%	1%	0.57
Motivation: Business opportunity	76%	91%	0.03	88%	87%	0.87	97%	84%	0.05	88%	87%	0.93
Dedication: Sporadic	12%	4%	0.10	9%	5%	0.41	3%	7%	0.40	0%	8%	0.10
Dedication: Half-time	24%	21%	0.74	21%	22%	0.88	21%	22%	0.88	21%	22%	0.97
Dedication: Full-time	65%	75%	0.24	71%	73%	0.76	76%	71%	0.56	79%	71%	0.36
Sectoral experience (years)	6.1	5.4	0.51	4.6	5.9	0.23	5.3	5.7	0.72	6.4	5.3	0.37
Serial entrepreneur	62%	61%	0.97	53%	64%	0.24	62%	61%	0.97	70%	59%	0.27
Has entrepreneurial team	91%	87%	0.53	85%	89%	0.56	91%	87%	0.53	85%	89%	0.50
# people on team	3.2	2.9	0.43	2.9	3.0	0.73	2.9	3.0	0.73	3.0	3.0	0.91
Sector: Agriculture	18%	15%	0.70	9%	18%	0.21	15%	16%	0.88	21%	14%	0.31
Sector: Manufacturing	18%	22%	0.61	21%	21%	0.98	26%	19%	0.34	18%	22%	0.68
Sector: Water and electricity	0%	4%	0.24	9%	1%	0.02	0%	4%	0.24	3%	3%	0.98
Sector: Construction	3%	3%	0.99	0%	4%	0.24	3%	3%	0.99	6%	2%	0.23
Sector: Commerce	0%	3%	0.31	3%	2%	0.74	3%	2%	0.74	3%	2%	0.72
Sector: Services	62%	53%	0.40	59%	54%	0.66	53%	56%	0.73	48%	58%	0.35
Participated in other contests	62%	58%	0.73	62%	58%	0.73	59%	59%	0.95	55%	61%	0.53
% Established firms	47%	47%	0.96	44%	48%	0.73	53%	45%	0.40	42%	48%	0.58
Year founded (established firms)	2013	2013	0.58	2013	2013	0.79	2013	2013	0.38	2014	2013	0.07
Sales 2013 (million pesos)	17.68	8.22	0.20	7.68	11.62	0.60	14.48	9.36	0.49	2.52	13.27	0.15
Sales 2014 (million pesos)	31.82	23.77	0.47	21.40	27.28	0.60	34.93	22.72	0.27	14.72	29.38	0.19
Total employees 2014	5.0	3.7	0.10	3.7	4.1	0.55	4.2	4.0	0.80	3.2	4.3	0.19

The table compares applicants' characteristics (at application) across the different quartiles of panel judge scoring generosity.

Appendix 6—Robustness checks monotonicity



The figure plots scoring generosity measures that are calculated separately for different restricted samples (as specified in the y-axis of each subplot) against the corresponding judge fixed effects estimated for the full sample.

Appendix 7—Employment and profits

Table A7 summarizes the impact of the accelerator using profits and employment as outcome variables. We construct these additional performance measures using survey responses. The main advantage of this exercise is to provide a robustness check for our findings on acceleration impacts using alternative metrics. The main drawback is that the Colombian Registry has no information on these variables, and thus results based on these performance metrics can be subject to survey reporting bias (which is why we do not use these variables in our main set of results). With this caveat in mind, Table A7 shows evidence of large acceleration effects for employment. Acceleration increases number of workers by 4.4, relative to a baseline average of 4.0 employees per firm (Column 4). In terms of profits, the evidence is more nuanced; the point estimate is positive, but not statistically significant. When we restrict the sample to established firms, however, the evidence on profits is compelling: profits increase by 26M COP ($6.5K USD), 1.4 times relative to baseline (Column 3). Our results on employment are similar to those in the literature; Glaub et al. (2012) estimate that treated firms have roughly twice as many workers as control firms after five to seven months of a three-day training intervention.

Table A7. Acceleration, profitability, and employment

	(1)	(2)	(3)	(4)	(5)	(6)
	Profits			Total Employment		
Sample	All	Business ideas	Established firms	All	Business ideas	Established firms
Accelerated × After	18.50	-2.261	26.05**	4.416**	0.138	6.901**
	(11.99)	(11.55)	(10.59)	(2.014)	(2.821)	(3.315)
Year 2015	43.11**	-8.556	111.7*	10.96**	4.404	16.87*
	(20.66)	(16.96)	(63.66)	(4.860)	(4.854)	(10.15)
Year 2016	43.60**	-5.025	108.7*	10.20**	4.516	15.12
	(20.35)	(16.95)	(62.63)	(4.772)	(4.944)	(10.33)
Year 2017	40.24**	-4.955	101.5	9.528**	3.891	14.39
	(20.37)	(15.96)	(62.64)	(4.714)	(4.744)	(10.41)
Constant	6.116***	0	13.11***	4.030***	2.667***	5.587***
	(1.361)	(0)	(2.587)	(0.348)	(0.361)	(0.601)
Observations	540	288	252	540	288	252
Number of applicants	135	72	63	135	72	63

The table presents results from estimating equation (3), instrumenting Accelerated × After with quartiles of panel judge leniency. The outcome variable is indicated on top of the columns. The variables *Accelerated* and *After* correspond to dummy variables indicating applicants that were accelerated and years after application to the accelerator, respectively. All columns include applicant fixed effects and several controls, including adjusted score, firm's age, entrepreneur's age entrepreneur's education, location, and sectorial and entrepreneurial experience. The specifications also include interactions between the controls and the variable *After*. Standard errors are clustered at the applicant level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Appendix 8—Propensity score matching

Panel A—Probit

	(1) Accelerated
Established firm	0.00681
	(0.0875)
Male	0.153**
	(0.0711)
Sector experience	0.00500
	(0.00651)
Serial entrepreneur	0.109
	(0.0808)
Adjusted score	-9.030
	(11.07)
Squared adjusted score	8.950
	(8.555)
Education: Technical degree	0.0836
	(0.177)
Education: College	0.128
	(0.125)
Education: Masters or PhD	-0.0498
	(0.132)
Has team	0.127
	(0.0839)
Motivation: Business opportunity	0.240
	(0.176)
Sales 2014	-0.000358
	(0.000607)
Pseudo R^2	0.4709
Observations	135

Panel B—Common support and differences in propensity scores

	Off support	On support	Total
Untreated	1	99	100
Treated	16	19	35
Total	17	118	135
Average absolute difference in propensity scores		0.014	

Panel C—Kernel matching estimator

	All	Ideas	Firms
Average (Weighted) difference in revenue	59.10***	4.48	122.01***
	(25.79)	(33.51)	(50.97)
	354		

The matching procedure relies on a kernel matching of propensity scores. The matching begins with a probit regression at the applicant level of a binary variable indicating acceleration against different controls, including adjusted scores. Panel A presents the coefficient estimates and the adjusted R^2. They reveal that the regression captures a significant amount in selection, as indicated by the R^2 of 0.47. We then use the predicted probabilities from this estimation, the propensity scores for acceleration, to perform a kernel match with a radius of 0.05 that forces the matches to be in the common support. This requirement results in 16 accelerated companies for which we am unable to find a corresponding match (i.e., they have propensity scores that are too high relative to the rejected applicants). We report the number of applicants in the common support and the number of matched participants in Panel B. Panel B shows that the majority of differences in the estimated propensity scores for acceleration between the accelerated companies and their matches are inconsequential. The average absolute difference between the matched propensity scores is 1.4%. Panel C reports average annual differences in revenues post-application (2015–2017) between accelerated participants and their matches. Bootstrap standard errors are reported in parentheses. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.